Department of Economics and Finance
Degree Program in Economics and Finance major in Finance

ANALYZING THE VARIANCE RISK PREMIUM: EMPIRICAL EVIDENCE FROM FINANCIAL MARKETS

Course of Econometric Theory

Supervisor:
Prof. Paolo Santucci de Magistris

Co - Supervisor:
Prof. Emilio Barone

Candidate:
Lorenzo Lanfrancotti
ID: 767631

Academic Year 2023/2024

Contents

Introduction

Volatility influences all aspects of financial markets. The concept of "Risk" has been associated with the concept of Volatility as it quantifies the potential magnitude of price movements within a given period. Larger price swings, whether upward or downward, indicate higher risk, as the predictability of returns decreases. From an investor's perspective, this uncertainty impacts portfolio performance and investment decisions, making volatility a key driver of risk premiums. Consequently, investors require compensation, in the form of higher expected returns, to face the uncertainty associated with high volatility. When investing in financial markets, investors deal with at least two types of uncertainty, the volatility of returns and fluctuations in the volatility of returns themselves. The compensation for this second form of risk is captured by the variance risk premium (VRP). It is the difference between the market's expectation of future variance, as implied by option prices, and the actual realized variance observed in financial markets. This premium exists because investors want extra compensation for the uncertainty surrounding future volatility, beyond the usual risks tied to price fluctuations. The behavior of the VRP is linked to broader economic and financial conditions. During times of financial turmoil, such as the Great Recession of 2007-2009 and the COVID-19 pandemic, a sharp increase in market uncertainty leads to significant fluctuations in the VRP as investors rush to hedge against extreme market movements. In contrast, during stable market conditions, the VRP typically declines, reflecting a reduced need for volatility protection. These dynamics make VRP a critical variable in risk management, derivative pricing and portfolio optimization.

The concept of VRP was formally introduced by Carr and Wu (2009), who defined it as the difference between a risk neutral measure of variance, denoted by the $\mathbb{Q}$ measure, and a realized measure of variance, denoted by the $\mathbb{P}$ measure. Using $E_t[\cdot]$ to denote the conditional expectation with respect to the information available at time t , the VRP is expressed as:

$$VRP_t = E_t^{\mathbb{Q}}[\sigma^2_{r,t+1}] - E_t^{\mathbb{P}}[\sigma^2_{r,t}].$$

In practice, the Implied Variance (IV) serves as a proxy for the variance under the risk-neutral measure $\mathbb{Q}$, while the Realized Variance (RV) approximates the variance under the physical measure $\mathbb{P}$. Empirical studies indicates that VRP present sharp temporal variations and is typically negative. This shows that investors are willing to accept lower expected returns or even incur costs to hedge against upward movements in volatility, which tend to be linked to market stress and declining asset prices.

This thesis is part of an increasing body of research focused on the role of the VRP in predicting short-term fluctuations in financial returns. As highlighted in Zhou (2018), the VRP exhibits significant predictive power for equity, bond, currency and credit spread returns, with its predictive ability peaking over a few months before gradually weakening at longer horizons. Its predictability for short horizons is meant to complement traditional established financial indicators such as the P/E ratio, the term spread, the interest rate differential and the leverage ratio. In addition, VRP can also serve as a proxy for macroeconomic uncertainty or volatility, which varies independently of consumption growth, the target of the long-run risk model proposed by Bansal and Yaron (2004). The empirical evidence suggests that incorporating a time-varying component of economic uncertainty into a general equilibrium framework, especially for the case of recursive preferences, can improve understanding of how uncertainty risk is priced in financial markets. The development of a dynamic model for VRP is central, given it showcases significant time-series dependencies and develops continually in response to new information. Explicitly modeling VRP helps capture its persistence as well as its short and long term fluctuations. This, in turn, gives us a clearer understanding of how it influences market uncertainty across different time horizons. This particular approach provides insights into how variance risk affects asset returns and general financial states. Furthermore, an econometric specification of VRP increases its usefulness in forecasting applications, rendering it a useful instrument for risk management, asset pricing and macroeconomic analysis.

Starting from and building on these empirical insights, this thesis aims to contribute to the ongoing discussion by analyzing the evolution of the VRP and evaluating the effectiveness of econometric models in capturing its behavior.

The first chapter will explain the derivation of the Variance Risk Premium, the formulation of its components and the data and methodological tools used in the analysis.

In the second chapter, the models applied for the analysis are described in more detail. Both simple Autoregressive models AR(p) and Heterogeneous Autoregressive models (HAR), introduced by Corsi (2009), are employed to study the dynamics of the VRP. For the Autoregressive models, Maximum Likelihood Estimation (MLE), which in this context is equivalent to Ordinary Least Squares (OLS), was used for parameter estimation. For the different HAR-type models, only OLS regression was applied.

In the third chapter, the empirical analysis begins with a detailed examination of the time series of implied variance (IV), realized variance (RV) and variance risk premium (VRP) for the S&P 500, focusing on their persistence and memory properties. This initial analysis provides valuable insights into the temporal patterns of these variables, helping to guide the selection of appropriate econometric models. Then the analysis proceeds with both in-sample and out-of-sample estimation, followed by a forecast evaluation to compare models performance. The estimated models are AR(1), AR(2) and HAR models applied directly to the VRP time series, as well as an alternative approach that decomposes VRP in its components. In this last model, HAR models are estimated separately for RV and IV and their forecasts are combined to construct an indirect VRP forecast. This decomposition is done to capture the different statistical properties of realized and implied variance and to assess whether their joint modeling improves forecasting accuracy. To compare forecasting performance, standard error metrics Root Mean Squared Error ($RMSE$), Mean Absolute Error (MAE) and R^2 are computed. In addition, the Diebold-Mariano (DM) test is used to formally assess whether differences in predictive accuracy between models are statistically significant.

The main result of this study is that the AR(2) model consistently outperforms all other models in forecasting VRP, as indicated by its lowest $RMSE$ value and highest R^2 in out-of-sample tests. This finding suggests that, despite the ability of HAR models in capturing long-run effects, a simpler autoregressive model with only two lags is sufficient to describe the short-term dynamics of VRP. The better performance of the AR(2) specification highlights the fact that VRP fluctuations are dominated to a large extent by short-term dependencies rather than persistent long-term patterns. This implies that the variance risk premium reacts quickly to market conditions and that efforts to incorporate long-memory structures do not necessarily lead to better forecasts. Instead, models that focus on capturing short-term dependencies tend to provide the best predictions, highlighting the high-frequency nature of variance risk premium adjustments in financial markets.

Literature Review

The variance risk premium has received considerable attention in the financial literature as a key indicator of market sentiment, risk aversion and macroeconomic uncertainty. Despite defined properly by Carr and Wu (2009), the first studies on this topic focused on it components. Whaley (2000) highlights the fact that VIX is widely recognized by investors as an indicator of market sentiment, reflecting investors' concerns about future uncertainty. A major focus of financial research has been the gap between implied and realized volatility, which is commonly used as a proxy for VRP. This difference has been interpreted as an indicator of risk aversion, capturing how much investors demand to compensate for exposure to volatility risk, as in Rosenberg and Engle (2002), Bakshi and Madan (2006), Bollerslev et al. (2011) and Bekaert and Hoerova (2016). Another perspective links the VRP to economic uncertainty, suggesting that it reflects market reactions to fluctuations in broader macroeconomic conditions, as in Bollerslev et al. (2009), Drechsler and Yaron (2011) and Drechsler (2013). Empirical evidence underscores the predictive power of the VRP for risk premia in various asset classes, including equities, bonds, currencies and credit markets. In particular, its predictive ability peaks at short horizons, such as a few months, and declines over longer horizons. This short-term predictability complements established predictions such as P/E ratios for equities, forward rates for bonds and leverage ratios for credit markets. Moreover, the VRP improves the explanatory power of these traditional measures, suggesting that it captures a common component of risk premia across financial markets. For equities, time-varying economic uncertainty, modeled by recursive preferences, explains the role of the VRP in predicting returns Bollerslev et al. (2009). While this study does not primarily address other asset classes, it is worth highlighting some significant related findings. Grishchenko et al. (2022) emphasize that in bond markets, the interest rate variance risk premium is predominantly driven by short-term risk, complementing traditional forward-rate factors tied to both short and long-term economic dynamics. In currency markets, Londono and Zhou (2017) reveal that variations in the currency variance risk premium are systematically linked to global inflation uncertainty, as explained by consumption-based asset pricing models. Similarly, Zhang et al. (2009) argue that in credit markets, incorporating stochastic asset volatility into standard structural models is crucial for explaining observed credit spreads and their predictability.

Chapter 1

Variance Risk Premium Construction

1.1 Derivation of VRP

The Variance Risk Premium (VRP) measures the difference between the risk-neutral expectation for future variance and the realized variance over the same period. As already said, this premium arises because investors demand compensation for bearing uncertainty about volatility fluctuations. A standard approach to deriving the VRP follows the framework introduced by Carr and Wu (2009). This method uses the concept of variance swap, an OTC financial derivative contract that enables market participants to speculate on or hedge against volatility risks associated with an underlying asset, such as an exchange rate, interest rate, or stock index. Variance swaps have a zero net market value at entry time. Upon maturity, the payoff is calculated as the difference between the realized variance over the contract's duration and a constant known as the variance swap rate.

$$[RV_{t,T} - SW_{t,T}]L, \tag{1.1}$$

where $RV_{t,T}$ is the realized variance between time t and T, $SW_{t,T}$ is the fixed variance swap rate determined at t and paid at T, and L denotes the notional amount, converting variance payoff into monetary payouts. By no-arbitrage conditions, the variance swap rate must be equal to the risk neutral expected value of realized variance:

$$SW_{t,T} = E_t^Q[RV_{t,T}], \tag{1.2}$$

where $E_t^{\mathbb{Q}}[\cdot]$ denotes the expectations under the risk neutral measure $\mathbb{Q}$. This follows from standard arbitrage-free pricing arguments, as the swap must be fairly priced to prevent arbitrage opportunities in financial markets. To compute variance swap rates, Carr and Wu assume that the future price follows a stochastic differential equation with both continuous and jump components. Under this assumption, the quadratic variation of log returns over a given period $[t, T]$ can be replicated by a portfolio of European Options with a continuum of strike prices. The key results is that the risk-neutral expected value of realized variance can be approximated using out-of-the-money European options across all strikes $K > 0$ at the same maturity T:

$$E_t^{\mathbb{Q}}[RV_{t,T}] = \frac{2}{T-t} \int_0^{\infty} \frac{\Theta_t(K,T)}{B_t(T)K^2} dK + \epsilon, \tag{1.3}$$

where $B_t(T)$ denotes the time t price of a bond paying one dollar at T, $\Theta_t(K,T)$ is the time t value of an out-of-the-money option with strike $K > 0$ and maturity $T \geq t$ and ϵ denotes the approximation error, which is zero when future prices process is continuous. While instead the future prices can jump, the approximation error ϵ is of order $O((\frac{dF_t}{F_{t-}})^3)$ and is determined as:

$$\epsilon = \frac{-2}{T-t} E_t^{\mathbb{Q}} \int_t^T \int_{\mathbb{R}_{\not{0}}} [e^x - 1 - x - \frac{x^2}{2}] \upsilon_s(x) dx ds \tag{1.4}$$

Defining the physical probability measure with $\mathbb{P}$, Carr and Wu relate the variance swap to realized variance under physical measure through the following equation:

$$SW_{t,T} = E_t^{\mathbb{P}}[M_{t,T} RV_{t,T}]/E_t^{\mathbb{P}}[M_{t,T}] = E_t^{\mathbb{P}}[m_{t,T} RV_{t,T}], \tag{1.5}$$

where $E_t^{\mathbb{P}}[\cdot]$ is the expectation under real-world measure $\mathbb{P}$, $M_{t,T}$ denotes the stochastic discount factor and $m_{t,T} = M_{t,T}/E_t^{\mathbb{P}}[M_{t,T}]$. Using no arbitrage argument, they decompose the variance swap rate in two components:

$$SW_{t,T} = E_t^{\mathbb{P}}[m_{t,T} RV_{t,T}] = E_t^{\mathbb{P}}[RV_{t,T}] + \mathrm{Cov}_t^{\mathbb{P}}(m_{t,T} RV_{t,T}), \tag{1.6}$$

The first term $E_t^{\mathbb{P}}[RV_{t,T}]$ represents the conditional mean of the realized variance time series. The second term captures the covariance between the stochastic discount factor and the realized variance. The Variance Risk Premium is then defined as the negative of this covariance term. Rearranging the formula is possible to find the empirical computation of the VRP:

$$VRP_t = RV_{t,T} - SW_{t,T}, \tag{1.7}$$

this difference represents the ex-post profit and loss form holding a long variance swap position. The VRP is typically negative, implying that investors demand compensation to hedge against volatility.

1.2 Implied Variance

Having derived the Variance Risk Premium (VRP) as the difference between realized and risk-neutral expected variance, we now turn to the empirical estimation of its components. The first step involves obtaining a measure of implied variance, which serves as a proxy for the market's expectation of future variance under the risk-neutral measure $\mathbb{Q}$. This expectation is embedded in option prices and can be extracted using model-free approaches, as exemplified by the VIX index. There will now presented three of the most common methods use to obtain an estimation of the implied variance.

A simple yet widely used approach involves deriving the implied volatility by inverting the standard Black-Scholes formula (developed by Black and Scholes (1973)) for the price of a European call option and solving it for σ. The implied variance is then obtained by squaring the implied volatility:

$$C(S,t) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2), \tag{1.8}$$

where:

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \quad d_2 = d_1 - \sigma\sqrt{T-t}.$$

Here: $C(S,t)$ is the Price of the call option at time t, S is the current stock price, K is the Strike price, T is Time to maturity, r is the Risk-free interest rate, σ is the Volatility of the stock, and $\Phi(x)$ is the Cumulative distribution function of the standard normal distribution. One of the assumption of the Black-Scholes model is that the stock prices follow a geometric Brownian motion, which implies constant volatility. This assumption may not hold in practice, as financial markets often exhibit time varying and stochastic volatility.

A second method, as demonstrated by Carr and Madan (1998), Demeterfi et al. (1999), Britten-Jones and Neuberger (2000), consists in obtaining a risk-neutral expectations of future implied variance, and it is particularly suited for capturing the expected total return variation between t and $t+1$ conditional on time t. The model is expressed as:

$$IV_t \equiv 2\int_0^{\infty} \frac{C_t(t+1, \frac{K}{B(t,t+1)}) - C_t(t,K)}{K^2}\, dK \;=\; E_t^Q[\sigma^2_{t,t+1}], \tag{1.9}$$

where: $C_t(T,K)$ is the price of a European Call option with maturity T and strike price K, and $B(t,T)$ is the price of a time t zero-coupon bond with maturity T. The computation relies on an ever-increasing number of calls with strikes that goes from zero to infinity. Obviously the concrete computation has to be done with a non-infinite number of strikes, but even with a few different option prices it is possible to obtain a good approximation.

The third method refers to the previous derivation of the Variance Risk Premium based on the work of Carr and Wu (2009). In their study, they use the variance swap as a proxy for the volatility measure under the risk-neutral measure $\mathbb{Q}$. In empirical research, it is common to use the implied variance with a 30-day horizon, such as the VIX index, as an approximation of this measure. However, variance swaps with an exact 30-day maturity are not always available in the market. To address this, Carr and Wu propose constructing a synthetic variance swap rate by applying linear interpolation between the two nearest available variance swap maturities They compute the variance swap rate at fixed 30-day horizon as:

$$SW_{t,T} = \frac{1}{T-t}\left[\frac{SW_{t,T1}(T1-t)(T2-T) + SW_{t,T2}(T2-t)(T1-T)}{T2-T1}\right], \tag{1.10}$$

where T_1 and T_2 denote the two maturity dates, and T denotes the interpolated maturity date such that $T-t$ is 30 days.

1.3 Realized Variance

While the foundational model for Realized Variance originates from the mathematical formulation of asset price dynamics,this study follows the advancements made by Barndorff-Nielsen and Shephard (2002), who provided significant refinements to its practical implementation. Specifically, the model is grounded in the stochastic behavior of asset prices, which can be described as a Itô semi-martingale. Let p_t denote the log-price of an asset at time t, which evolves as a stochastic process. A commonly used model is the stochastic differential equation:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \tag{1.11}$$

where: μ_t represents the deterministic drift term, σ_t denotes the instantaneous volatility, W_t is a standard Brownian motion. In this framework, the integrated volatility (σ_t^{2*}) over a time interval $[0, T]$ is defined as:

$$\sigma_t^{2*} = \int_0^T \sigma_t^2 \, dt, \tag{1.12}$$

Integrated volatility captures the total variation of the stochastic component of p_t, but it is not directly observable due to the continuous nature of the integral. The realized variance offers a practical, discrete approximation of the integrated volatility by summing squared returns over equally spaced intervals within $[0, T]$. Given high-frequency price observations at n time points $[t_o, t_1, t_2, ..., t_n]$, the log-returns are defined as:

$$r_i = p(t_i) - p(t_{i-1}), \ \ with \, i = 1, ..., n.$$

The realized variance is calculated as:

$$RV_t = \sum_{i=1}^{n} r_i^2, \tag{1.13}$$

As the sampling frequency increases ($n \to \infty$ and $\Delta t_i \to 0$), the Realized Variance converges in probability to the integrated volatility:

$$RV_t \xrightarrow{p} \sigma_t^{2*} = \int_0^T \sigma^2(s) ds, \tag{1.14}$$

This consinstency property, rigorously established by Barndorff-Nielsen and Shephard (2002), ensures that RV_t is a reliable estimator of Variance.
As established in the literature (e.g. Andersen et al. (2001a), Andersen et al. (2001b), Barndorff-Nielsen and Shephard (2002), Meddahi (2002)) the RV constructed using high-frequency intraday data provides a more accurate ex-post measure of the true and nonobservable return variation compared to traditional sample variances based on daily or lower-frequency data. Using intraday data allows for more precise volatility estimation because it enhances accuracy by capturing intra-day price movements, reduces estimation bias from overnight gaps in asset prices and allows for better modeling of volatility clustering and eventual jumps. However, in practical applications, various market microstructure limit the maximum sampling frequency that can be used to estimate RV, such as the bid-ask bounce effects and latency effects and price discreteness, which dis-

tort variance estimates. To obtain a balance between the estimation error minimization and all the noise of high frequencies prices quotation, many studies, based on the results obtained by Andersen et al. (2000), suggest that a five minute sampling frequency is a good choice.

1.4 Dataset

In this thesis, we focus on variance rather than standard deviation. This has been done to ensure consistency in terms of variance, thus ensuring a correct interpretation of the results and facilitating econometric modeling.

In analyzing the Implicit Variance we rely on the VIX index as a proxy of our IV_t. Its computation is done using option prices on the S&P500 index, based on a variance swap model using a wide range of out-of-the-money call and put options. It is not based on a specific pricing model, as the Black-Scholes, but instead is a model-free approach that measures the 30-day expected volatility of the S&P500 index. A basic interpretation of VIX's values allows one to understand the uncertainty and investors confidence in the short horizon[1]. For the construction of our model-free Realized Variance RV_t, we used intraday prices with 5 minute intervals. The dataset used consists of the daily values of the VIX and the realized variance for the period 2000-2024. For the period from 2000 to 2020, the open source data provided by Candila (2024) was used. For the subsequent period from 2021 to 2024, the VIX data were obtained from Investing.com, while the five-minute historical quotes of the S&P500 were obtained from Bloomberg.

[1] A detailed description of the VIX computation is available in the White Paper provided in the CBOE website

Chapter 2

Econometric Models

In this chapter, the econometric models applied to time series analysis will be presented in detail. First, the Autoregressive (AR) model will be introduced, along with its Maximum Likelihood Estimation (MLE) procedure for parameter estimation. It is important to note that MLE estimation coincides with the Ordinary Least Squares (OLS) estimation under standard assumptions. Following this, the Heterogeneous Autoregressive (HAR) model, proposed by Corsi (2009), will be discussed. Specifically, the classical HAR-RV model, which focuses on volatility as the main variable, but instead here will be presented using variance.

2.1 Autoregressive model of order p - AR(p)

The first approach employed to model the VRP is the Autoregressive model (AR). This model assumes that the current value of the variables depends on its past values plus a random schock, capturing the temporal dependence commonly observed in a lot of financial time series. The AR model is defined as follows:

$$X_t = \delta + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \qquad with \quad \epsilon_t \sim WN(0, \sigma^2), \tag{2.1}$$

Using the Lag operator L, defined as $L^k x_t = x_{t-k}$, it is possible to rewrite the $AR(p)$ process as:

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)x_t = \delta + \epsilon_t,$$

where the polynomial in the lag operator is: $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$. This implies that for the process to be stationary requires that the roots of the characteristic equation associated with $\phi(L)$ lie outside the unit circle in the complex plane, or, equivalently, the modulus of each root must satisfy ($|z| > 1$). The characteristic equation is obtained solving: $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0$, and here z represents the roots of the polynomial. To granting stationarity then all roots of $\phi(z) = 0$ must lie outside the unit circle ($|z| > 1$). Intuitively, this implies that the influence of past values on the current value diminishes over time, ensuring that the time series does not exhibit trends or an explosive behavior.

For a generic Autoregressive Process of order p, the Autocorrelation function (ACF) typically decays gradually, with an exponentially or sinuisoidal mode depending on whether the roots of the characteristic polynomial are real or complex. The Partial Autocorrelation function (PACF) is a key diagnostic tool for identifying the correct order p for a model, typically cutting off sharply after p lags.

2.1.1 AR model estimation

The parameters that are included in an AR model are typically estimated using methods such as the Ordinary Least Squares (OLS), Maximum Likelihood Estimation (MLE) and the Yule-Walker Equations. In this work, the focus will be on the MLE, assuming Gaussian white noise, i.e. $\epsilon \sim N(0, \sigma^2)$ (which, under this assumption, is equal to OLS estimation). Given this, the conditional probability density function of X_t, given its past values, is:

$$f(X_t|X_{t-1}, X_{t-2}, ..., X_{t-p}, \phi, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\epsilon_t^2}{2\sigma^2}), \tag{2.2}$$

Since the observations are conditionally independent, the total likelihood of the data (for t=p+1,...T), starting from (2.2) is:

$$L(X, \phi, \sigma^2) = \prod_{t=p+1}^{T} f(X_t|X_{t-1}, X_{t-2}, ...X_{t-p}, \phi, \sigma^2),$$

$$L(X, \phi, \sigma^2) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\epsilon_t^2}{2\sigma^2}), \tag{2.3}$$

Taking the logarithm of (2.3), we simplify the optimization process, converting the product in a summation and then expanding each term:

$$\mathcal{L}(X,\phi,\sigma^2) = \log(\prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp{(-\frac{\epsilon_t^2}{2\sigma^2})}),$$

$$\mathcal{L}(X,\phi,\sigma^2) = \sum_{t=p+1}^{T} \log(\frac{1}{\sqrt{2\pi\sigma^2}}) + \sum_{t=p+1}^{T} \log(\exp(-\frac{\epsilon_t^2}{2\sigma^2})),$$

$$\mathcal{L}(X,\phi,\sigma^2) = \sum_{t=p+1}^{T} [-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{\epsilon_t^2}{2\sigma^2}], \tag{2.4}$$

and breaking the summation in (2.4) into individual terms:

$$\mathcal{L}(X,\phi,\sigma^2) = -\frac{T-p}{2}\log(2\pi) - \frac{T-p}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+1}^{T} \epsilon_t^2, \tag{2.5}$$

Let's introduce the parameter vector θ, defined as $\theta = (\phi, \sigma^2)$. Let's call Θ its parameter space, defined as $\Theta = \{(\phi, \sigma^2) \in \mathbb{R}^{\mathbb{P}} \times \mathbb{R}^+\}$. Having now the Log-Likelihood function, we can optimize the (2.5) to get the parameter estimation:

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \log \mathcal{L}(X,\phi,\sigma^2),$$

2.2 HAR-RV

The Heterogeneous Autoregressive Realized Volatility (HAR-RV) model, as pioneered by Corsi (2009), signifies a substantial progression in the field of financial market volatility modeling. It is founded on the so called "Heterogeneous Market Hypothesis" introduced by Müller et al. (1993), which posits that financial markets comprise agents with divergent time horizons and trading behaviours. These differences give rise to distinct variance components that are shaped by interactions among heterogeneous market participants. This multi-scale approach adopted by the HAR-RV model enables it to effectively capture the complexities of volatility dynamics in a parsimonious yet effective manner.

The partial variance $\tilde{\sigma}_t^{2(.)}$ represents the variance attributed to a specific market component. The proposed model is characterized as an additive cascade of partial variances, where each component exhibits a structure similar to an AR(1) process. The model then assumes a hierarchical process where the future partial variances at any given level is influenced by two factors: the historical variance observed at the same scale (the AR(1) component) and the partial variance from the next higher level of the hierarchy, corresponding to a longer time horizon. To facilitate the analysis, the model focuses on three distinct components of variance, each associated with different time horizons: one day ($1d$), one week ($1w$), and one month($1m$). These components are represented as $\tilde{\sigma}_t^{2(d)}$,$\tilde{\sigma}_t^{2(w)}$ and $\tilde{\sigma}_t^{2(m)}$. The hierarchical structure of the model facilitate the capture the interdependence between variances observed over different time horizons, thereby providing a detailed approach to understanding market dynamics.

The model for the unobserved partial variance process, denoted by $\tilde{\sigma}_t^{2(.)}$ at each time period, is considered as a function of past realized variance experienced at same time scale. This is done whilst taking into account the asymmetric propagation of variances and the expectation of the next period values of the longer-term partial variances. The model is defined:

$$\tilde{\sigma}_{t+1m}^{2(m)} = c^{(m)} + \phi^{(m)} RV_t^{(m)} + \tilde{\omega}_{t+1m}^{(m)}, \tag{2.6}$$

$$\tilde{\sigma}_{t+1w}^{2(w)} = c^{(w)} + \phi^{(w)} RV_t^{(w)} + \gamma^{(w)} E_t[\tilde{\sigma}_{t+1m}^{2(m)}] + \tilde{\omega}_{t+1w}^{(w)}, \tag{2.7}$$

$$\tilde{\sigma}_{t+1d}^{2(d)} = c^{(d)} + \phi^{(d)} RV_t^{(d)} + \gamma^{(d)} E_t[\tilde{\sigma}_{t+1m}^{2(w)}] + \tilde{\omega}_{t+1w}^{(d)}, \tag{2.8}$$

Where $RV_t^{(d)}$, $RV_t^{(w)}$ and $RV_t^{(m)}$ are the daily, weekly and monthly Realized Variances previously defined, while the variance innovations $\tilde{\omega}_{t+1d}^{(d)}$, $\tilde{\omega}_{t+1w}^{(w)}$ and $\tilde{\omega}_{t+1m}^{(m)}$ are serially independent zero mean nuisance variates with truncated left tail to guarantee the positivity of partial variances. Through straightforward recursive substitutions of the partial variances, this cascade model can be expressed as:

$$\sigma_{t+1d}^{2(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \tilde{\omega}_{t+1d}^{(d)}, \tag{2.9}$$

The equation (2.9) can be interpreted as a three factor stochastic variance model, where the factors are directly linked to past realized variances measured at different frequencies. Building up this process for latent variance, it is straightforward to derive the structure of a time series model based on realized variance. Specifically, the ex-post latent daily variance can be expressed as:

$$\tilde{\sigma}_{t+1d}^{2(d)} = RV_{t+1d}^{d} + \omega_{t+1d}^{(d)}, \tag{2.10}$$

Equation (2.10) establishes a connection between the ex-post variance estimate RV^d_{t+1d} and the corresponding contemporaneous measure of latent variance σ^d_{t+1d}. By substituting (2.10) in (2.9), it is possible to obtain a simplified time series representation of the cascade model:

$$RV^d_{t+1d} = c + \beta^{(d)} RV^{(d)}_t + \beta^{(w)} RV^{(w)}_t + \beta^{(m)} RV^{(m)}_t + \omega_{t+1d}, \tag{2.11}$$

where $\omega_{t+1d} = \tilde{\omega}^{(d)}_{t+1d} - \omega^{(d)}_{t+1d}$

In order to estimate the HAR model correctly, it is necessary to clarify certain points. In the empirical analysis, the focus will be on Realized Variance as opposed to Realized Volatility. This methodological decision does not affect the construction of the model, and thus the previously outlined approaches may be applied. The following rules will be applied to compute the monthly and weekly realized variance:

$$RV^{(w)}_t = \frac{1}{5} \sum_{i=1}^{5} RV^{(d)}_{t-i},$$

$$RV^{(m)}_t = \frac{1}{22} \sum_{i=1}^{22} RV^{(d)}_{t-i},$$

One of the advantages of the HAR model is that its parameters can be estimated using Ordinary Least Squares (OLS). OLS regression method offers several benefits, such as computation efficiency, simple implementation and interpretable results.

2.3 Tests

To ensure the reliability of the analysis, the stationarity of the time series is tested using the Augmented Dickey-Fuller (ADF) test. The test determines whether there is a unit root in the data, as this indicates that the time series is not stationary. Testing for stationarity is important because many econometric models require this condition in order to avoid bias arising from time trends or persistent data shocks. The ADF test was applied to all the time series, to confirm whether these series exhibit mean-reverting behavior or require transformations such as first differencing. Model performance metrics were used to compare the forecasting ability of econometric models in capturing VRP dynamics. The predictive models were evaluated by Root Mean Squared Error ($RMSE$) and Mean Absolute Error (MAE) and R^2 calculation for each model. These measures allow a quantitative comparison of how well each model fits the data and predicts future values. A Diebold-Mariano (DM) test was used to determine which of the tested models provided superior predictive accuracy. The DM test compares the forecast errors of two competing models and determines whether the difference in their predictive performance is statistically significant. In time series applications, the DM test is a valuable approach because it performs the model comparison in an efficient manner that compensates for autocorrelation and heteroskedasticity in the forecast error patterns. In the Appendix section, the document provides a detailed explanation with mathematical representations for ADF, $RMSE$, MAE, R^2 and the DM test, together with their basic assumptions and their technical implementation.

Chapter 3

Empirical Analysis

3.1 Time series analysis

Beginning our empirical analysis, we will now examine the time series components of the variance risk premium. The S&P 500 data have been obtained as explained in the previous chapter, and the time series are now discussed. First, the descriptive statistics are presented:

Table 3.1: Descriptive statistics for the period 2000-2024

	Mean	Median	Std. Deviation	Skewness	Kurtosis
VRP	9.0005x10^{-5}	7.3168x10^{-5}	1.4589x10^{-4}	-12.1545	448.1220
RV	9.4793x10^{-5}	4.0603x10^{-5}	2.3160x10^{-4}	12.1777	260.5009
IV	1.8483x10^{-4}	1.2601x10^{-4}	2.0630x10^{-4}	4.9966	39.8206

Table 3.1 provides a better understanding of the distribution of the series. For both Implied Variance (IV) and Realized Variance (RV), the mean is higher than the median, indicating a positively skewed distribution. This is supported by the values of the skewness, especially for RV. In contrast, the Variance Risk Premium (VRP) exhibits negative skewness, as indicated by its mean, median, and most notably, its skewness value. The extreme negative skewness of VRP (-12.1545) aligns with market participants' tendency to hedge against volatility risk, leading to sharp declines in VRP during periods of financial stress. The kurtosis values indicate the presence of heavy tails, confirming that all series exhibit leptokurtic behavior, which is a typical characteristic of financial markets

where volatility tends to cluster and undergo sudden, high-impact fluctuations. The extremely high kurtosis of VRP (448.1220) supports the idea that variance risk premiums are particularly sensitive to abrupt movements. Apart from extreme events, volatility clustering shows that periods of high volatility tend to be followed by sustained high volatility, whereas periods of low volatility exhibit prolonged stability. This persistence in the volatility process allows for the use of Heterogeneous Autoregressive (HAR) models, which can capture the influence of past variance at different time horizons. The fact that RV and IV remain low for extended periods but occasionally spike dramatically suggests that market variance is shaped by both short-term shocks and long-term trends. The strong negative skewness and high kurtosis of VRP also indicate that it tends to decline sharply in response to uncertainty, reflecting the corresponding shifts in investors' risk perception and hedging behavior.

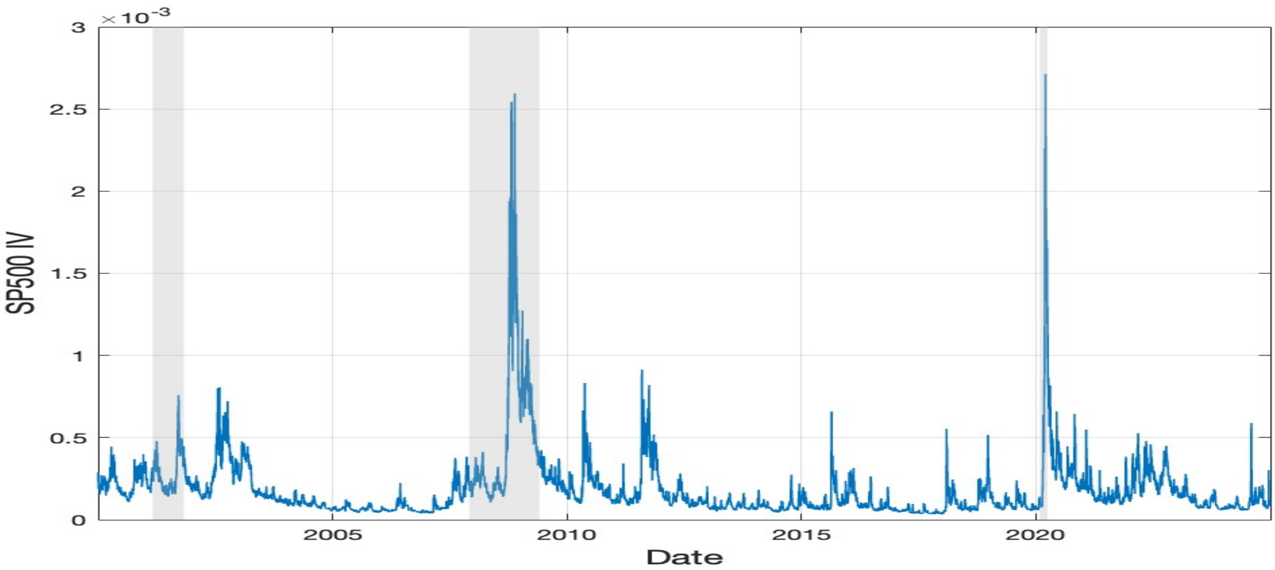
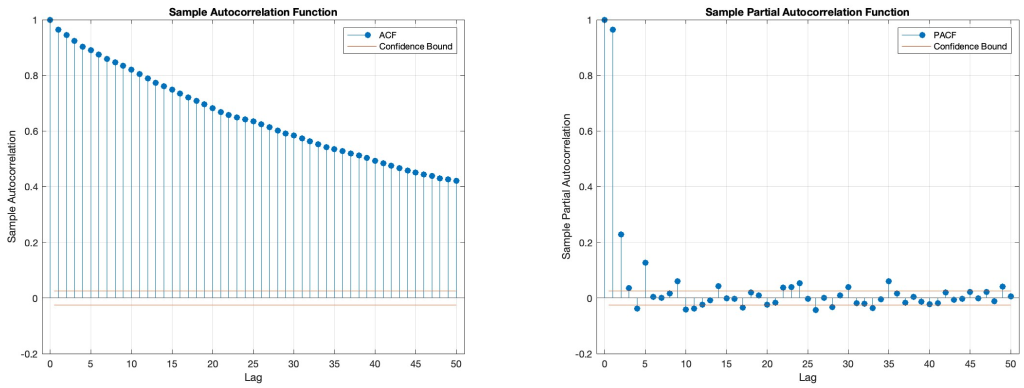


Figure 3.1: S&P500 Implied Variance

In Figure 3.1, the evolution of the S&P500 Implied Variance over time can be observed. Notable spikes occur during periods of increased uncertainty, such as the 2007–2009 Great Recession or the COVID-19 pandemic crisis. These peaks indicate moments when the market's perception of risk increased, leading to higher option premiums. From the analysis of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), shown in Figure 3.2, it is clear that IV exhibits a long-memory process, with persistent correlations between implied variance and its past values. The PACF shows a sharp decay after lag 2, and, together with the behavior of the ACF, supports the selection of a lower-order autoregressive model as a baseline for capturing time dependencies.



(a) Implied Variance ACF

(b) Implied Variance PACF

Figure 3.2: Implied Variance ACF and PACF

Continuing the analysis, Figure 3.3 illustrates the evolution of the S&P500 Realized Variance over time. Similarly to IV, it exhibits sharp peaks in periods of financial turmoil. However, RV displays more extreme fluctuations compared to IV. This difference arises because RV directly measures variance based on actual market movements, whereas IV reflects investors' forward-looking expectations of variance, indicated in option prices. As a result, RV reacts more quickly to market shocks, showing sharp and sudden spikes in response to significant financial events. In contrast, IV adjusts more gradually, serving as a smoothed measure of perceived risk that incorporates market sentiment and risk premia over time.

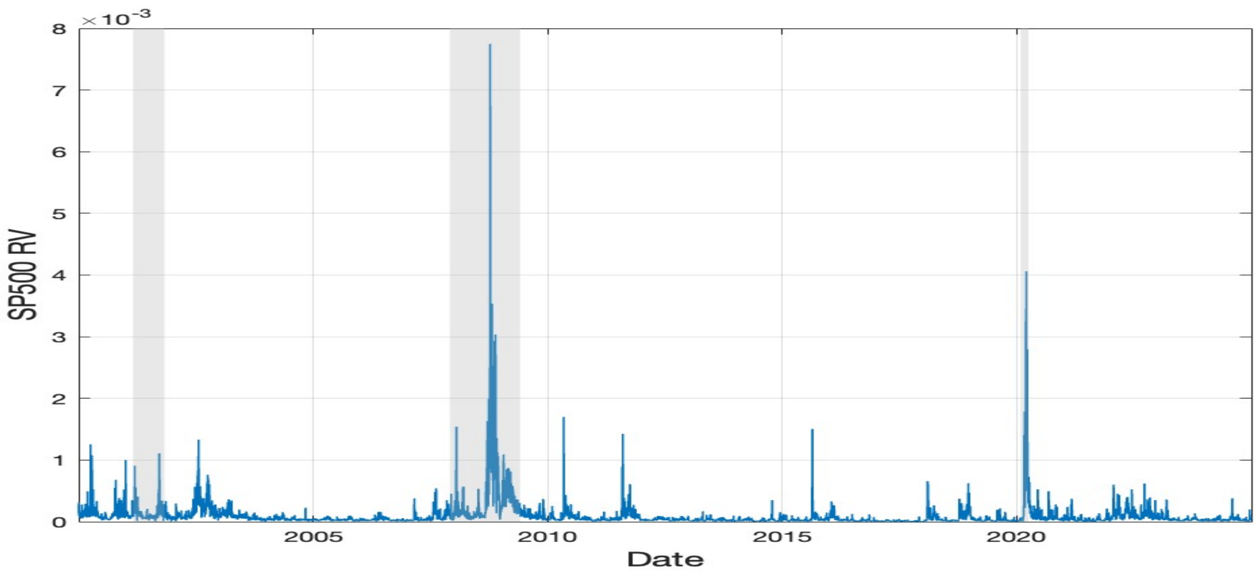
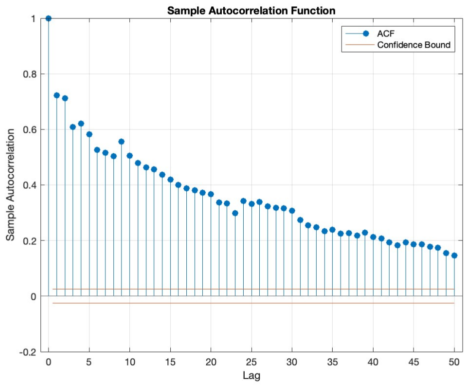
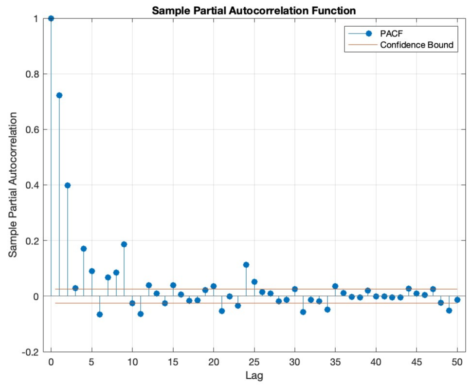


Figure 3.3: S&P500 Realized Variance

The divergence between RV and IV shows the distinct nature of variance dynamics. Realized Variance (RV) is inherently reactive, responding immediately to the events of the markets, while Implied Variance (IV), being expectation based, adjusts at a slower rate. This highlights the unique role these measures play in financial markets, showing how each serves a different purpose: RV measures the direct impact of price fluctuations, while IV reflects the broader assessment of future risk by the market, including both immediate uncertainty and investor sentiment. Similar results have been found for the Realized Variance in the analysis of the ACF and PACF functions. Looking at Figure 3.4, RV exhibits strong memory, though with lower persistence compared to IV. The autocorrelation function (ACF) decays slowly, while the partial autocorrelation function (PACF) decays sharply after lag 2. This combined behavior suggests that an autoregressive model would be a good starting point for capturing the time dependencies in Realized Variance.



(a) Realized Variance ACF



(b) Realized Variance PACF

Figure 3.4: Realized Variance ACF and PACF

Now, we proceed with the analysis of the S&P500 Variance Risk Premium (VRP). Figure 3.5 shows significant volatility peaks and clustering over time. For a more complete overview, some key financial events are marked in the graph. The dot-com bubble of 2000–2002, which also includes the events of September 11, 2001, marks a period of persistent volatility clustering at the beginning of the 21st century. The second cluster corresponds to the 2007–2009 Great Recession, a period characterized by extreme intraday market fluctuations. In particular, the 10% variation in the S&P500 on October 10, 2008, is clearly visible, showing the extreme volatility of that period. The rest of the spikes in volatility align with other shocks in the market, including the 2010 and 2015 flash crashes and the 2011 European sovereign debt crisis. More recently, a third major period of increased volatility occurred during the global coronavirus pandemic (2019–2023), along with the early phase of the Russian invasion of Ukraine (2022). This latter period introduced considerable uncertainty into financial markets due to rising energy prices and inflation, amplifying market volatility even more.

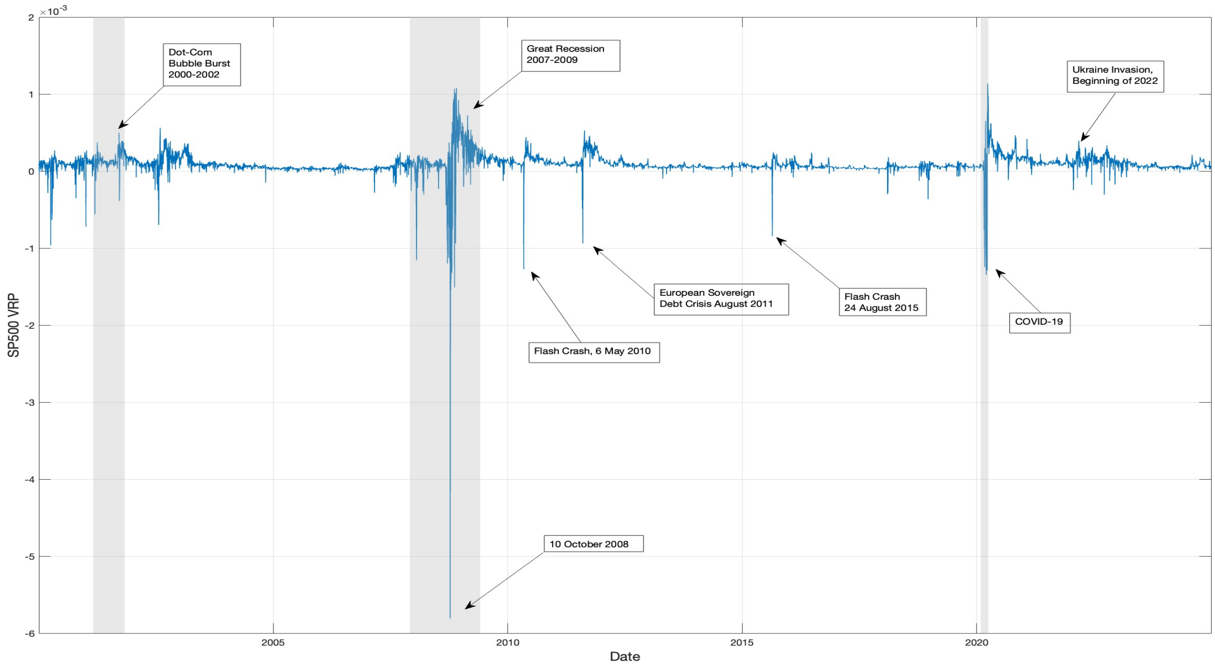
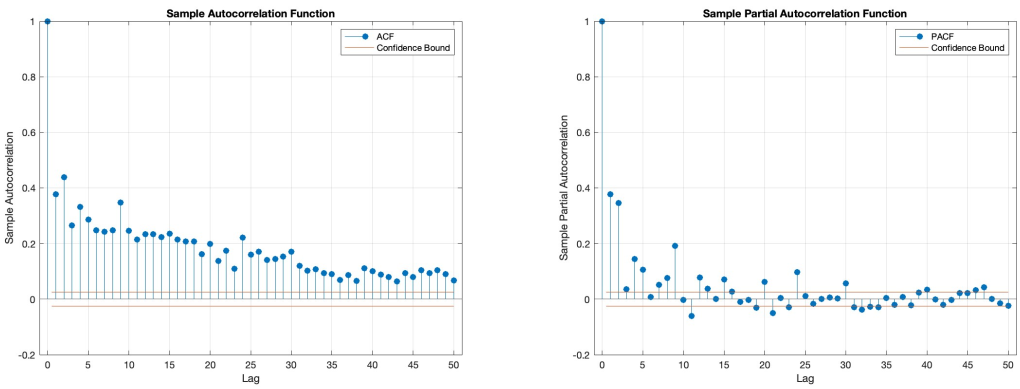


Figure 3.5: Evolution of SP500 VRP

VRP exhibits a moderate decay in is autocorrelaton function. While there is noticeable persistence, it is weaker compared to implied variance and realized variance, which aligns with he nature of VRP as the difference between the two. As expected, the PACF behave exactly as the others, decaying after lag 2.



(a) Variance Risk Premium ACF

(b) Variance Risk Premium PACF

Figure 3.6: Variance Risk Premium ACF and PACF

The Augmented Dickey-Fuller (ADF) test was conducted to asses the stationarity of all three time series. The test was run under the null hypothesis of non-stationarity of the series, implying the presence of a unit root, against the alternative of stationarity with a constant and, where necessary, a linear time trend. The results confirmed that each series is stationary at the 5% significance level, which ensures stability for econometric modeling without requiring further transformations. This guarantees the reliability of subsequent analyses and model estimations based on these data.

3.2 In Sample Analysis

Now, the results of the model estimations will be presented. At first, an $AR(1)$ and $AR(2)$ model were estimated, where the order p was determined based on the ACF and PACF of the VRP time series. As discussed earlier, VRP exhibits long memory, but its partial autocorrelation function (PACF) decays sharply after lag 2, suggesting that an autoregressive process is an appropriate model for capturing the data dynamics. The parameters of these models were estimated using Maximum Likelihood Estimation (MLE). Subsequently, three different HAR models were estimated. The first one is the $HAR-VRP$ model, defined as follows:

$$VRP_{t+1d}^{d} = c + \beta^{(d)}VRP_t^{(d)} + \beta^{(w)}VRP_t^{(w)} + \beta^{(m)}VRP_t^{(m)} + \omega_{t+1d} \tag{3.1}$$

The $HAR-VRP$ model provides a simple framework to capture multi-scale persistence of the VRP. It incorporates daily, weekly and monthly components, and therefore accounts for the heterogeneous behaviour of market participants with different investment horizons. The second and third model, respectively an $HAR-RV$ and an $HAR-IV$ model, will be constructed to get forecast of future RV and IV, and then this two value will be used to get a VRP forecast. These models are defined as follows:

$$RV_{t+1d}^{d} = c + \beta^{(d)}RV_t^{(d)} + \beta^{(w)}RV_t^{(w)} + \beta^{(m)}RV_t^{(m)} + \omega_{t+1d} \tag{3.2}$$

$$IV_{t+1d}^{d} = c + \beta^{(d)}IV_t^{(d)} + \beta^{(w)}IV_t^{(w)} + \beta^{(m)}IV_t^{(m)} + \omega_{t+1d} \tag{3.3}$$

Equation 3.2 corresponds to the model originally introduced by Corsi (2009). The decision to estimate RV and IV separately, rather than modeling VRP as a single dependent variable, allows to satisfy the need to mitigate potential biases arising from their distinct time series properties. By treating them individually, it becomes possible to accurately decompose their variance persistence across different time horizons, providing a clearer understanding of their respective contributions to VRP dynamics. Although Implied Variance (IV) represents a market-based expectation of future variance, it is not directly known at time t+1 and remains subject to dynamics that can be modeled and forecasted. The primary objective of applying a HAR model to IV is to provide an estimate of tomorrow's implied variance given today's available information. Specifically, while IV encapsulates market expectations, these expectations themselves evolve over time, responding to new information, changes in risk perception, and structural shifts in market conditions. The HAR model, which captures variance dynamics over multiple horizons, provides a structured framework to model the persistence and responsiveness of implied variance across different time scales. Thus, the decision to apply an HAR model to IV is motivated by the need to generate a data-driven estimate of future implied variance, rather than assuming that IV follows a purely exogenous process. Once the forecasts $\hat{RV}_t$ and $\hat{IV}_t$ have been obtained, an estimate of the variance risk premium, $V\hat{R}P_t$, can be computed in this way:

$$V\hat{R}P_t^* = \hat{IV}_t - \hat{RV}_t$$

where $\hat{IV}_t$ and $\hat{RV}_t$ are the predicted values from the $HAR-IV$ and $HAR-RV$ models, respectively.

	$AR(1)$	$AR(2)$
ω	0.001	0.001
	(0.000)	(0.000)
ϕ_1	0.378	0.247
	(0.002)	(0.002)
ϕ_2		0.3451
		(0.003)

	$HAR-VRP$	$HAR-RV$	$HAR-IV$
c	1.705×10^{-5}	9.226×10^{-6}	4.659×10^{-6}
	(2.368×10^{-6})	(2.127×10^{-6})	(9.539×10^{-7})
$\beta^{(d)}$	0.094	0.313	0.757
	(0.015)	(0.015)	(0.014)
$\beta^{(w)}$	0.401	0.484	0.221
	(0.029)	(0.023)	(0.018)
$\beta^{(m)}$	0.3157	0.105	-0.002
	(0.03)	(0.023)	(0.01)

Table 3.2: In Sample model estimations results for SP&500 VRP

However, it is important to note that the estimated models are not all based on the same time series, which raises concerns about the suitability of using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for model comparison. While it is technically possible to compute these metrics, they mainly assess the goodness of fit of individual models rather than the accuracy of the final VRP forecasts. For this reason, model comparison is conducted using forecasting evaluation metrics such as Root Mean Squared Error ($RMSE$), Mean Absolute Error (MAE), and the R^2 coefficient, which provide a more appropriate measure of predictive accuracy. These metrics allow for a more reliable assessment of the models' ability to forecast VRP, ensuring that the results reflect out-of-sample performance rather than in-sample optimization.

Table 3.2 shows the results for the in-sample analysis. The parameters of the models exhibit statistical significance, given that all estimated coefficients have p-values well below the significance level $\alpha = 5\%$, providing strong evidence against the null hypothesis. By examining the estimation results, notable patterns emerge. The estimation outcomes for the $AR(1)$ and $AR(2)$ models provide valuable insights into the persistence of the time series. Constructing a model with just a single lag offers a clear perspective on the extent to which VRP depends on its past values, highlighting the presence of temporal dependence in its evolution. In the case of the HAR models, the results from the $HAR - IV$ model indicate that short-term implied variance has a significantly stronger influence on future predictions compared to long-term implied variance. This suggests that market expectations of volatility are highly responsive to recent fluctuations rather than being shaped by more persistent, long-term trends. Such behavior aligns with the idea that implied variance primarily reflects investor sentiment and forward-looking risk perception, incorporating new information almost instantaneously. This finding is in contrast with the behavior of realized variance, which exhibits a more gradual adjustment to market dynamics and is less sensitive to fluctuations occurring within specific temporal intervals. Unlike implied variance, realized variance captures the accumulation of past variance, making it a more stable measure that evolves in response to actual market movements rather than anticipatory sentiment.

	$AR(1)$	$AR(2)$	$HAR-VRP$	$HAR-RV$	$HAR-IV$	VRP^*
$RMSE$	1.351×10^{-4}	1.268×10^{-4}	1.28×10^{-4}	1.568×10^{-4}	6.715×10^{-5}	1.339×10^{-4}
MAE	5.198x10^{-5}	4.552x10^{-5}	4.245x10^{-5}	4.899x10^{-5}	2.910x10^{-5}	4.354x10^{-5}
R^2	0.143	0.245	0.233	0.543	0.894	0.1599

Table 3.3: Comparative Measures for In-Sample model Estimation

Based on the results in Table 3.3, it is clear that the $AR(2)$ model performs best according to $RMSE$ and R^2, whereas MAE identifies $HAR-VRP$ as the most accurate model. While $RMSE$ prioritizes overall accuracy by penalizing larger errors more heavily, MAE provides a more balanced measure by treating all deviations equally, making it less sensitive to extreme values. This distinction may help explain the divergent rankings between the $AR(2)$ and $HAR-VRP$ models, and can be observed in Figure 3.7. Similarly, the R^2 statistic, which measures the proportion of variance explained by the model, also identifies $AR(2)$ as the best-performing model, stressing on its strong ability to capture the underlying data patterns.

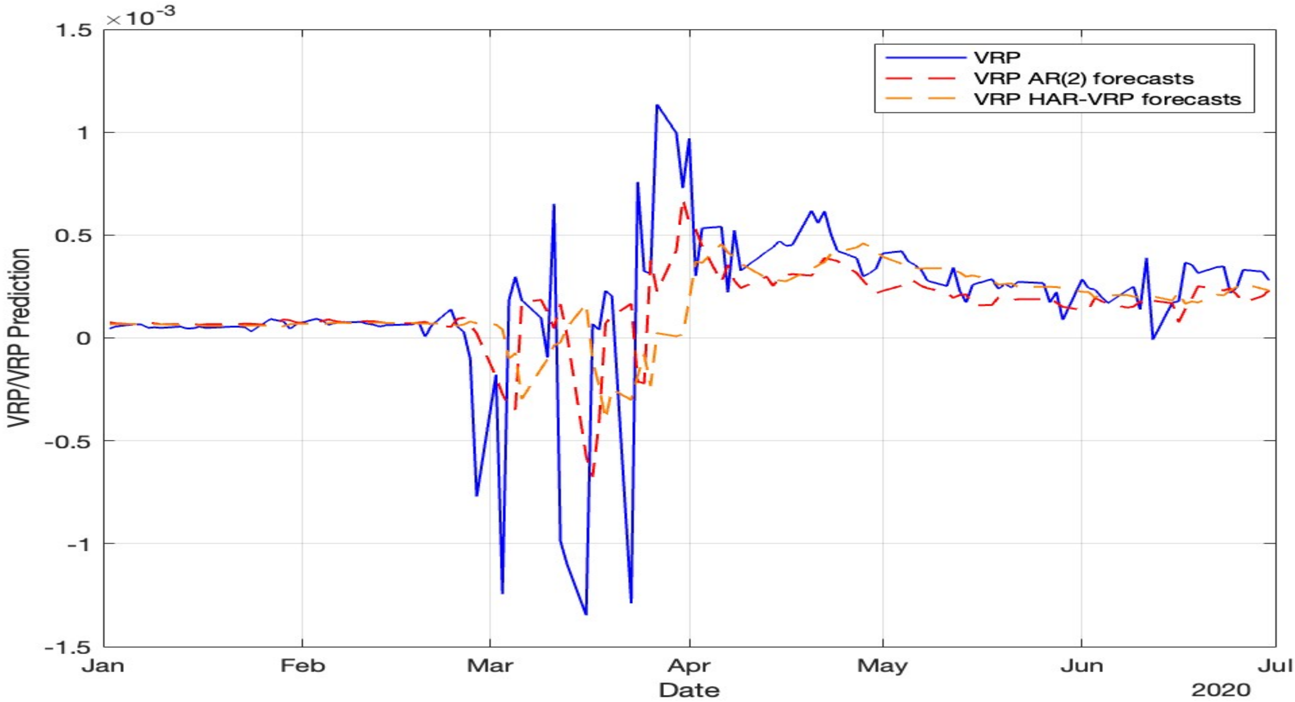


Figure 3.7: Model forecats performance during COVID-19 crisis

3.3 Out of sample Analysis

Following the evaluation of in-sample model performance, the next step involves conducting an out-of-sample analysis to asses the predictive accuracy of the models on unseen data. To achieve this, the dataset was split into two parts: the training set, which accounted for 50% of the observations, was used to estimate the model parameters, while the test set, consisting of the remaining 50%, was reserved for forecast evaluation. The VRP, RV, and IV time series consist of a total of 6,323 observations, covering the 2000–2024 period. This means that the training and test sets have respectively 3,162 and 3,161 observations. The estimation results for the same five models previously presented are illustrated in Table 3.4, followed by the comparative performance measures contained in Table 3.5.

	$AR(1)$	$AR(2)$
ω	0.001	0.001
	(0.000)	(0.000)
ϕ_1	0.336	0.219
	(0.003)	(0.004)
ϕ_2		0.346
		(0.004)

	$HAR-VRP$	$HAR-RV$	$HAR-IV$
c	1.953×10^{-5}	1.098×10^{-5}	4.205×10^{-6}
	4.07×10^{-6}	3.869×10^{-6}	1.486×10^{-6}
$\beta^{(d)}$	0.061	0.281	0.764
	(0.021)	(0.021)	(0.019)
$\beta^{(w)}$	0.392	0.433	0.191
	(0.042)	(0.035)	(0.025)
$\beta^{(m)}$	0.349	0.201	0.027
	(0.045)	(0.03)	(0.015)

Table 3.4: Out of Sample model estimations results for SP&500 VRP

Before proceeding with the analysis of the data, a preliminary note is to be provided. The results of the out-of-sample analysis for the comparative statistics yield lower values than their corresponding in-sample counterparts. This outcome is highly counterintuitive and does not align with the established theoretical expectations in forecast analysis, where out-of-sample errors are generally expected to be larger due to increased model uncertainty when applied to new data. However, a vital factor to consider is the substantial difference in variance between the training and test datasets. In particular, the variance of the training set is 3.3408×10^{-8}, while the variance of the test set is significantly lower at 9.0168×10^{-9}. This difference suggests that the model was estimated over a period in which the VRP time series exhibited much higher volatility. In particular, the training set includes periods of extreme market turbulence, such as the 2008 Global Financial Crisis and the Eurozone Sovereign Debt Crisis of 2011. This difference in variance implies that the model was trained under conditions of strong financial instability, making the in-sample estimation particularly sensitive to extreme fluctuations. As per Figure 3.5, the training set is influenced by substantial and more frequent volatility spikes, while the test set covers a period of relatively lower market uncertainty. Consequently, the out-of-sample performance benefits from the model being applied to a more stable dataset, leading to lower forecast errors despite theoretical expectations suggesting otherwise.

	$AR(1)$	$AR(2)$	$HAR-VRP$	$HAR-RV$	$HAR-IV$	VRP^*
$RMSE$	8.365×10^{-5}	7.826×10^{-5}	8.067×10^{-5}	1.098×10^{-4}	5.823×10^{-5}	8.57×10^{-5}
MAE	4.166×10^{-5}	3.425×10^{-5}	3.023×10^{-5}	3.705×10^{-5}	2.487×10^{-5}	3.081×10^{-5}
R^2	0.224	0.321	0.283	0.543	0.859	0.192

Table 3.5: Comparative Measures for Out-of-Sample model Estimation

The out-of-sample forecast evaluation provides important insights into the predictive performance of the models. The results for the $AR(1)$ and $AR(2)$ models highlight the fact that the second outperforms the first one, as evidenced by a lower $RMSE$ and a higher R^2. This result suggests that the Variance Risk Premium exhibits dependence beyond just a single lag, meaning that incorporating an additional past observation improves predictive accuracy. Despite this, the overall forecasting ability of the AR models remains limited, as indicated by the relatively low R^2 value, suggesting that the VRP dynamics are not fully captured by simple autoregressive models.

The $HAR-VRP$ model, estimated directly on the VRP dataset, performs better than the $AR(1)$ model but does not outperform $AR(2)$. The slightly higher R^2 compared to the $AR(2)$ model suggests that incorporating multiple time horizons provides some additional predictive power over simple autoregressive models. However, the improvement is not substantial enough to establish it as a clearly better approach, as it may overestimate the distinct dynamics of realized and implied variance, potentially introducing complexity without significantly improving predictive accuracy. A particular aspect is that the MAE metric, also in this case, identifies the $HAR-VRP$ model as the best performer among those that directly forecast VRP. This finding underscores the robustness of MAE as an error measure compared to MSE and $RMSE$. While MSE and $RMSE$ penalize larger errors more heavily due to squaring the residuals, MAE considers all deviations equally, making it less sensitive to extreme values and outliers. The fact that $HAR-VRP$ achieves the lowest MAE suggests that it produces more stable and reliable forecasts on average, even if it does not minimize large deviations as effectively as other models.

The VRP^* model, constructed using the forecasts from the $HAR-RV$ and $HAR-IV$ models, does not demonstrate a significant improvement in predictive accuracy compared to direct estimation using the $HAR-VRP$ model or a classic autoregressive model. This suggests that reconstructing VRP from separate variance forecasts does not necessarily yield better out-of-sample predictions. Although not directly comparable to the other models, the $HAR-RV$ and $HAR-IV$ models yield interesting insights. The $HAR-RV$ model performs considerably worse in an out-of-sample analysis. The much higher $RMSE$ and a relatively lower R^2 compared to the $HAR-IV$ model indicate that realized variance alone does not provide useful insights for predicting future values. This result is not surprising, as realized variance tends to be affected by sudden jumps and high persistence, making it difficult to capture its long-term behavior through standard autoregressive models. A particularly interesting result is that the $HAR-IV$ model emerges as a strong predictor in an out-of-sample context, reinforcing the idea that market expectations about variance contain more relevant information than past realized variance when forecasting future variance dynamics. This suggests that market expectations embedded in implied variance play a dominant role in influencing the future trajectory of VRP, making it the most informative predictor among the two components. Now the forecasts plots for each of the proposed models will be presented:

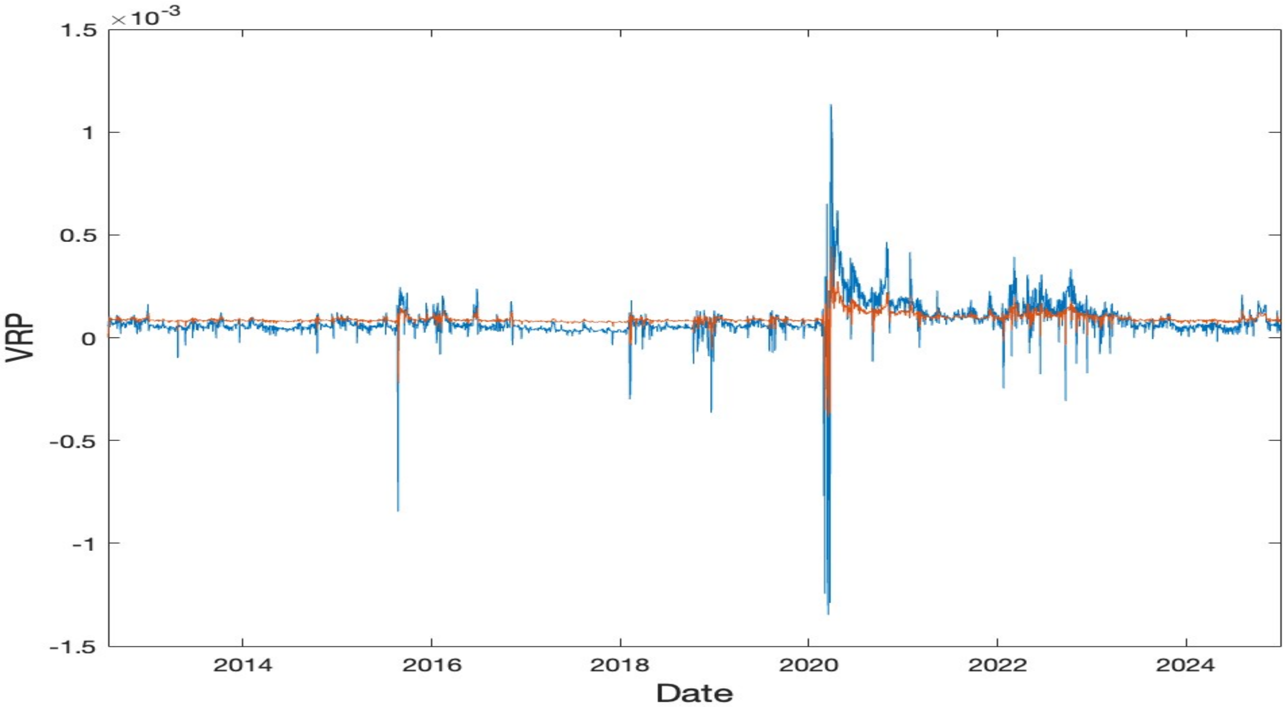


Figure 3.8: Out of sample forecast with AR(1) model

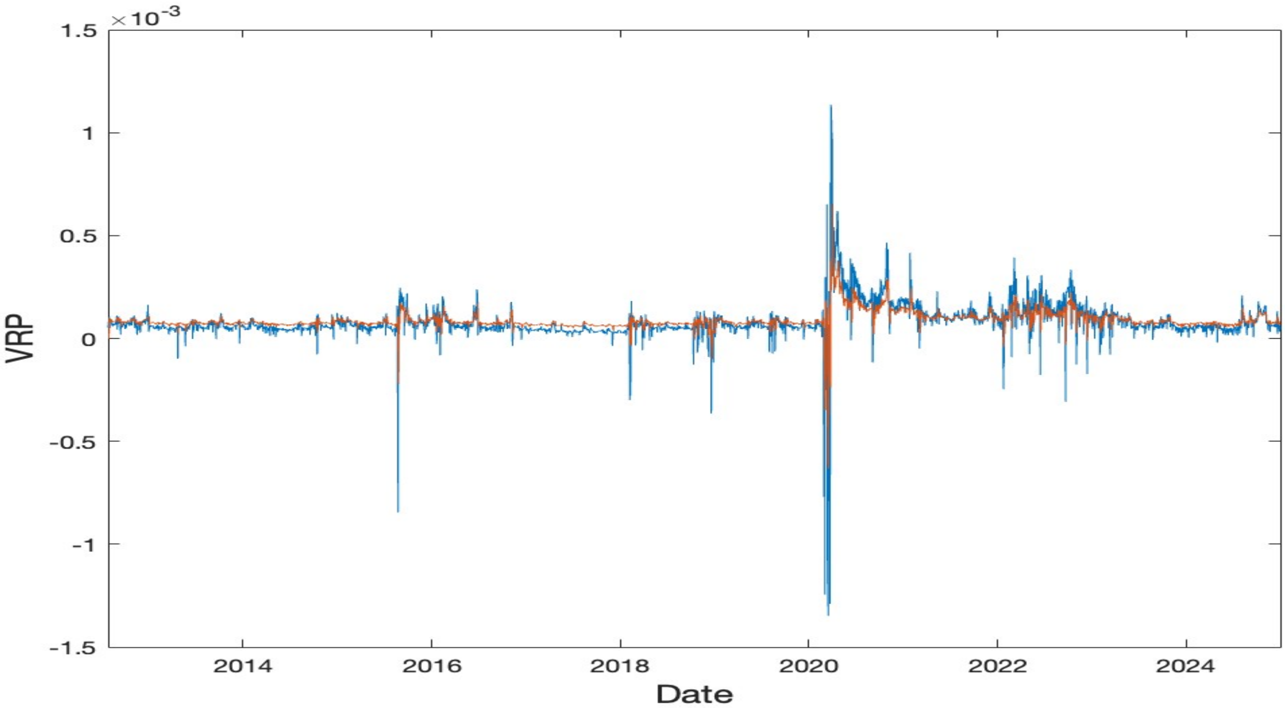


Figure 3.9: Out of sample forecast with AR(2) model

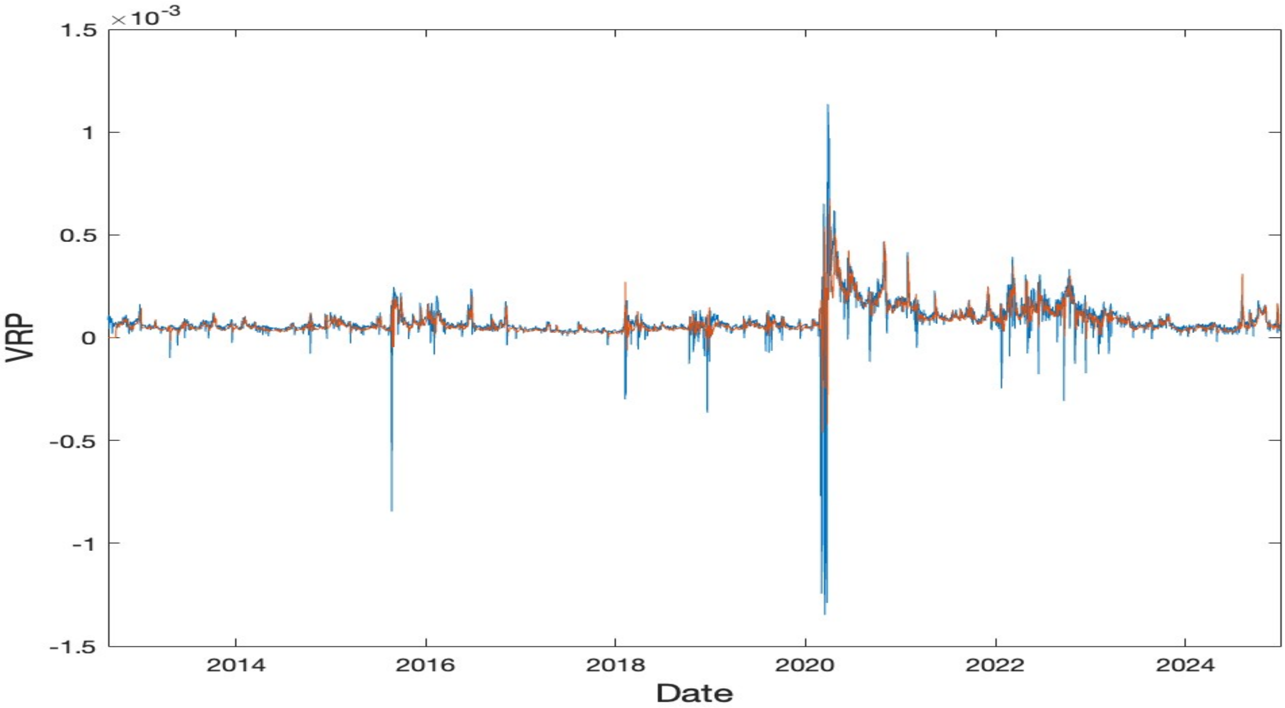


Figure 3.10: Out of sample forecast with HAR-VRP model

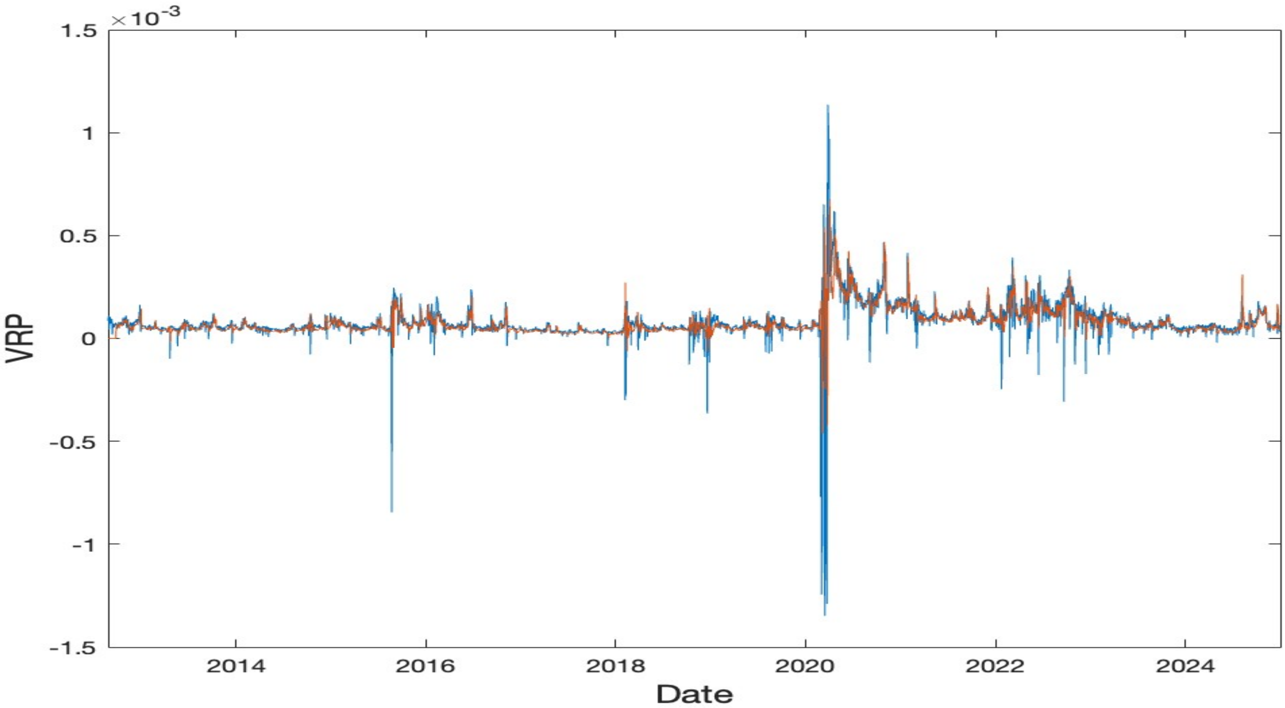


Figure 3.11: Out of sample forecast with VRP* model

3.3.1 Diebold-Mariano test

The results[1] presented in Table 3.6 do non yield statistical differences in predictive accuracy, with the only exception of AR(1)-AR(2) comparison, where the first model seems to perform better.

	DM test	p-value
$AR(1)/AR(2)$	2.0758	0.038
$AR(1)/HAR-VRP$	0.9658	0.3342
$AR(1)/VRP*$	-0.2588	0.7958
$AR(2)/HAR-VRP$	-0.5305	0.5958
$AR(2)/VRP*$	-1.2306	0.2186
$HAR-VRP/VRP*$	-1.1440	0.2527

Table 3.6: Results of Diebold-Mariano test for "out-of-sample" residuals

This is in contrast with the results given by comparative metrics such as $RMSE$, where AR(2) is indicated as the best one regarding predictive accuracy. This contradiction can be due to the different methodology used for these evaluation metrics. $RMSE$ and MAE measure the overall magnitude of forecast errors, with lower values indicating that, on average, a model produces more accurate predictions by minimizing the difference between actual and predicted values. However, the DM test evaluates whether the forecasts errors of two models are statistically different from each other, rather than just comparing their magnitude. One possible explanation for this difference in results is that AR(2) may produce more stable forecasts with lower error variance, while AR(1) errors might exhibit a different distribution that leads to relative better performance when compared directly to AR(2) model under DM test. The results of the other DM test, as already said, do not present a statistically significant difference in the models' forecast accuracy.

[1]Values for DM test have been obtained using the Matlab function called *dmtest_modified* made by Semin Ibisevic (2011) and Jaime Trujillo (2016)

Conclusion

This thesis was developed around a fundamental research question: Given that the Variance Risk Premium (VRP) is widely used as a predictive signal for asset returns, can econometric models be effectively constructed and evaluated to systematically capture its behavior? The ability to model and forecast VRP is particularly important in financial markets, as it represents the compensation investors require for bearing uncertainty about future volatility fluctuations. By identifying appropriate econometric methodologies, this study aims to assess whether VRP can be formally modeled and whether its predictive power can be leveraged to improve return forecasting and financial decision-making.

Empirical research, including the work of Zhou (2018), suggests that VRP exhibits strong predictive power for asset returns, peaking over short to medium term horizons. Specifically, its ability to forecast returns is most pronounced over a few months before gradually weakening. This characteristic implies that VRP may serve as a valuable risk indicator in short-term financial modeling, particularly in applications related to volatility trading, portfolio optimization, and risk management. In this study, we focus specifically on the equity market, using S&P500 data as a case study, given its importance as a benchmark for global financial markets and the availability of robust measures of implied and realized variance.

To systematically analyze the behavior of VRP, we implemented two widely used classes of econometric models: Autoregressive (AR(p)) models and Heterogeneous Autoregressive (HAR) models. Within these frameworks, we estimated four specific model specifications: $AR(1)$, $AR(2)$, $HAR - VRP$, and an alternative specification VRP^*, designed to capture different aspects of VRP dynamics, including its persistence and multi-scale structure. The empirical findings reveal that among the models tested, the AR(2) model achieved the lowest $RMSE$ and the highest R^2 in the out-of-sample evaluation, indicating that it provides the most accurate forecasts of VRP relative to the alternative specifications. This result suggests that VRP exhibits a notable degree of short-term predictability, reinforcing its role as an important financial signal for return forecasting. These results contribute to the broader literature on variance risk premiums, confirming that VRP could be used effectively as a meaningful and efficient predictor of asset returns, particularly over short time horizons. Given its impact in particular on derivative pricing and volatility risk assessment, future research could explore extending VRP based forecasting models into asset allocation strategies and examining its interaction with broader macroeconomic variables. Additionally, exploring alternative modeling techniques that was not used in this analysis, such as nonlinear models, regime-switching approaches, or machine learning-based methods, could further enhance our ability to capture VRP dynamics and improve its practical applications in financial markets.

Appendix

A.1 Augmented Dickey-Fuller Test

The Augmented Dickey-Fuller test is a widely used statistical procedure to determine whether a given time series is stationary, particularly in the presence of a unit root, which is an indicator of non-stationarity. A unit root implies that the shocks to the time series have permanent effects, making the time series non-stationary and unsuitable for many economic analyses that rely on stationarity assumptions. the ADF test extend the classical Dickey-Fuller (DF) test incorporating additional lagged differences to address serial correlation in the residuals of the model. Dickey and Fuller (1979), Dickey and Fuller (1981). The ADF test is based on estimating an augmented autoregressive equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \epsilon_i,$$

where γy_{t-1} tests for the presence of a unit root. The Null Hypothesis (H_0) of the ADF test states that the time series has a unit root ($\gamma = 0$), while the Alternative Hypothesis (H_1) indicates stationarity ($\gamma < 0$). The test statistic for the ADF test is calculated as:

$$ADF\ Statistic = \frac{\hat{\gamma}}{SE(\hat{\gamma})},$$

where $\hat{\gamma}$ is the estimated coefficient of y_{t-1} and $SE(\hat{\gamma})$ is its standard error.

A.2 Model Performance Metrics

In this section will be illustrated the metrics used in the assessment of model estimation and forecasting analysis.

The Root Mean Squared Error (RMSE) measures the mean of the squared errors, which represents the average squared difference between the predicted values and the actual values.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)},$$

where y_i and $\hat{y}_i$ are the actual and predicted values, respectively. N is the number of observations.
The $RMSE$ represents the standard deviation of the residuals (prediction errors). A lower $RMSE$ indicates that the model's predictions are closer to the actual values. Because $RMSE$ squares the errors before averaging, it penalizes larger errors more than smaller ones. This makes $RMSE$ useful when large deviations are particularly undesirable (e.g., in risk forecasting or financial modeling). However, $RMSE$ is sensitive to outliers, as the squared term amplifies the effect of large residuals. If a dataset contains extreme values, $RMSE$ may not always be the best metric for evaluating model performance.

The Mean Absolute Error (MAE) measures the average magnitude of the absolute differences between observed and predicted values. Its formula is:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \,,$$

where y_i and $\hat{y}_i$ are the actual and predicted values, respectively. N is the number of observations.
MAE provides a direct and intuitive measure of prediction errors in the same units as the dependent variable. Unlike $RMSE$, it does not square the residuals, meaning all errors contribute equally to the final metric. A lower MAE indicates that the model has, on average, smaller absolute errors. MAE is less sensitive to outliers than $RMSE$ since it treats all errors equally, regardless of magnitude. This makes it particularly useful in cases where all prediction errors should be weighted equally, rather than emphasizing large deviations.

The R^2 statistics, also known as coefficient of determination, is a statistical measure that quantifies how well a regression model explains the variability of the dependent variable. It indicates the proportion of the total variance in the observed data that is accounted for by the independent variables in the model.
Mathematically, the R^2 is defined as:

$$R^2 = 1 - \frac{RSS}{TSS},$$

where: RSS is the Residual Sum of Squares, defined as $RSS = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, and it represents the portion of the variance that is not explained by the model, capturing the differrence between actual and predicted values. TSS is the Total Sum of Squares, defined as $\sum_{i=1}^{N}(y_i - \bar{y})^2$, and it represents the total variability in the dependent variable, measuring how much the observed values y_i deviate from their mean $\bar{y}$. The values that R^2 can assume goes from 0 to 1, where higher values indicates better fit for the model.

A.3 Diebold-Mariano Test

In order to compare the predictive ability of the different models used in the empirical analysis, the Diebold-Mariano (DM) test is presented here. Originally introduced in Diebold and Mariano (1995), the test enables the assessment of whether one model significantly outperforms another with respect to forecast accuracy.

The DM test is constructed based on the forecast errors produced by the two models under comparison. For each time period, the forecast error is calculated as the difference between the observed value and the predicted value from the model. Denoting the forecast errors as $e_t^{(1)}$ and $e_t^{(2)}$ for models 1 and 2 respectively, the test evaluates the difference in forecast accuracy based on a specific loss function. The test statistic is derived from the sequence of loss differentials $d_t = L(e_t^{(1)}) - L(e_t^{(2)})$, where L represents the chosen loss function. Under the null hypothesis H_0, the two models are assumed to have equal predictive accuracy, which implies that the expected value of d_t is zero. The test then evaluates whether the sample mean of d_t deviates significantly from zero. The test hypotheses are:

$$H_0 : E[d_t] = 0, \quad \forall t$$
$$H_1 : E[d_t] \neq 0,$$

The standard DM test statistic is given by:

$$DM = \frac{\overline{d}}{\sqrt{\hat{\sigma}_d^2 / T}},$$

where $\overline{d}$ is the mean loss differential and $\hat{\sigma}_d^2$ is an estimate of the variance of the loss differential. If the test statistic exceeds the critical value (in absolute terms), the null hypothesis of equal predictive accuracy is rejected, indicating that the first model significantly outperforms the other. While the DM test is asymptotically normal, in small samples it has been shown to exhibit size distortions, leading to an increased likelihood of incorrectly rejecting the null hypothesis. To address this issue, Harvey et al. (1997) propose a small sample correction that adjusts the test statistic to improve its reliability in finite samples. The corrected test statistic is given by:

$$DM_H = \left(\frac{T + 1 - 2h + h/T}{T} \right)^{1/2} DM,$$

where T is the sample size and h is the forecast horizon. Unlike the standard DM test statistic, which follows an asymptotic normal distribution, the corrected statistic follows a t-distribution with $T - 1$ degrees of freedom. This adjustment accounts for the bias in the variance estimator and provides a more accurate assessment of forecast accuracy differences, especially when the sample size is limited. This correction ensures that inference drawn from the DM test remains valid and reduces the risk of over-rejecting the null hypothesis in finite samples.

A.4 Residual Analysis

In this section, the results of the residuals analysis for the in-sample model estimation will be presented, for all the model based on the VRP. For the $VRP*$ model, the residuals have been computed as the difference between the actual VRP and the estimation obtained using the two HAR model based on RV and IV. Analyzing the residual plots, it is clear that, in all the cases, there are some visible trends and clusters. This means that the models are struggling to capture the complete dynamics of the data, in particular during period of high variance. The ACF and PACF of the residuals show that the models are not able to fully capture the real dynamics of the data. In theory, if a model fit well the data, the residuals should no exhibit temporal correlations, but it can be seen in Figure 5-8, there are not case in which this happens. AR(2) can be individuated as the one in which the residuals exhibit smaller correlation. The Ljung-Box test was performed on the residuals of all four models, and every time it was possible to reject the null hypothesis of uncorrelated residuals.

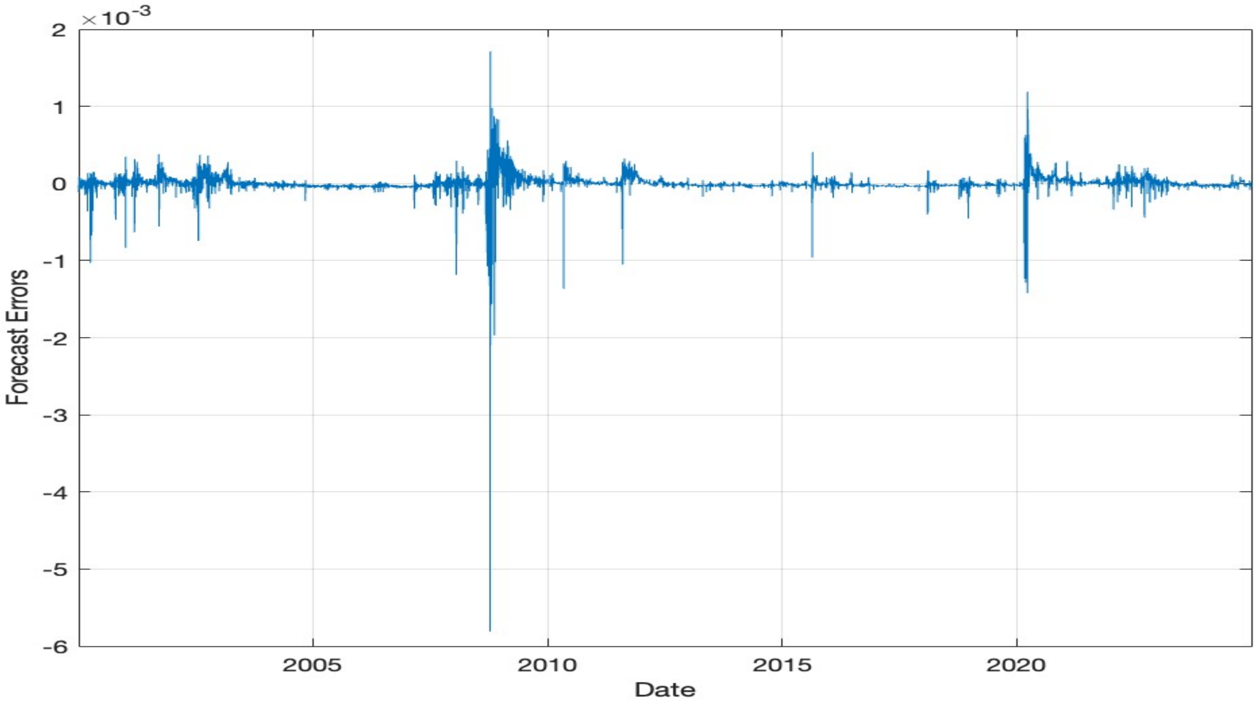


Figure 12: Forecasts Errors for AR(1) model

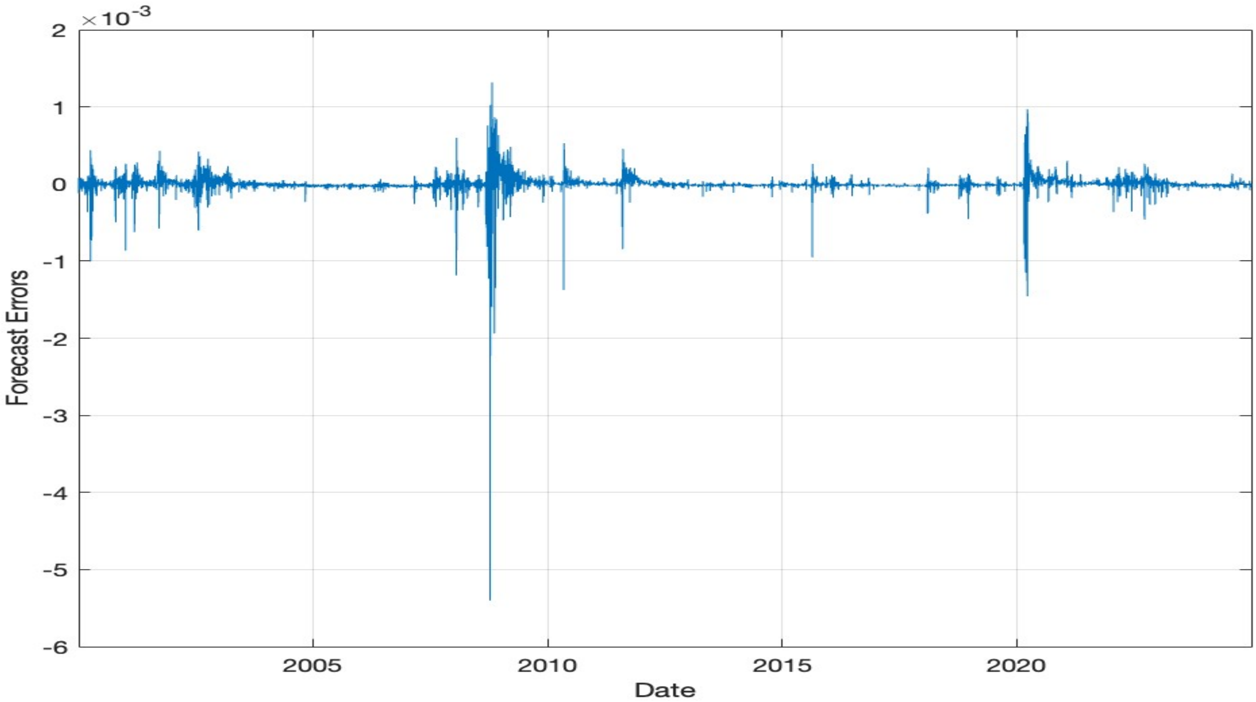


Figure 13: Forecasts Errors for AR(2) model

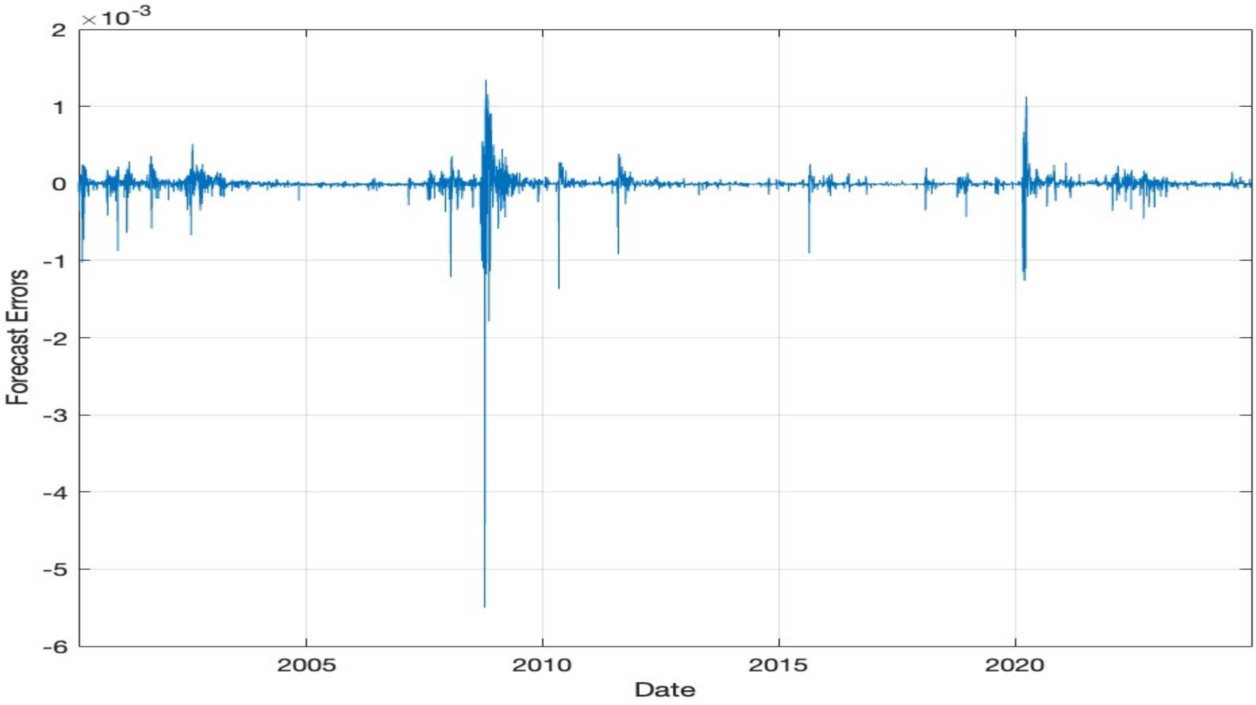


Figure 14: Forecasts Errors for HAR-VRP model

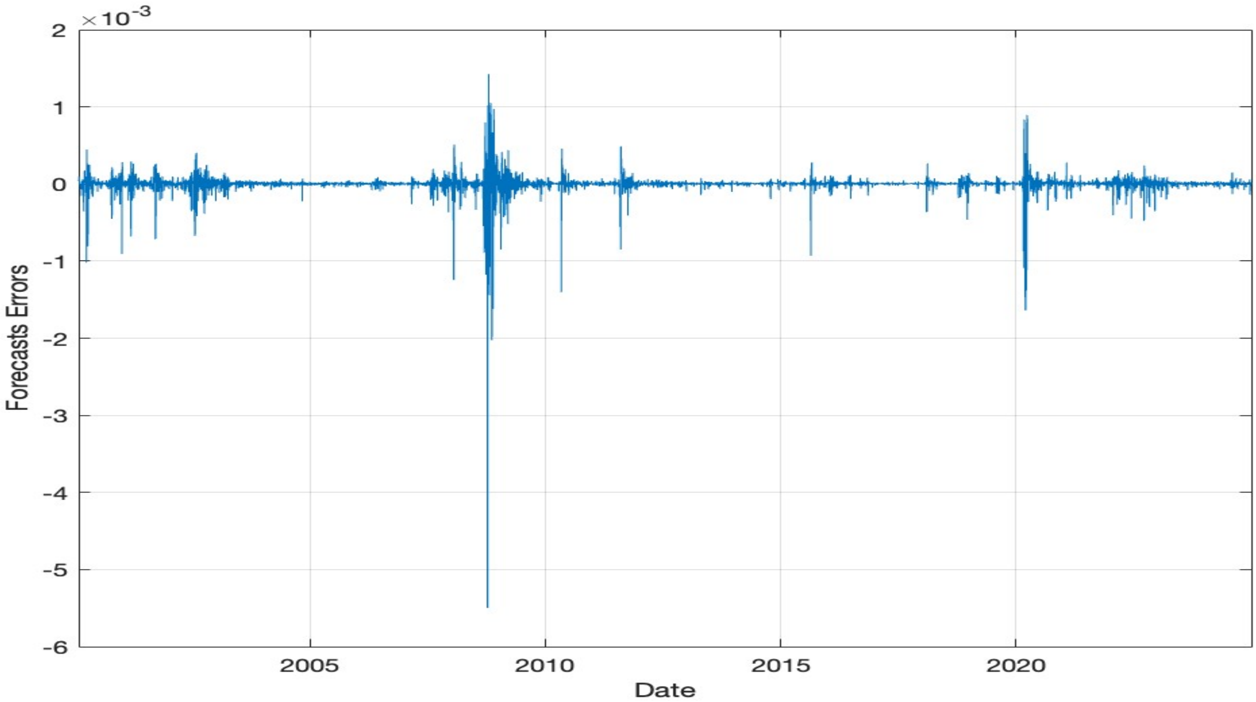
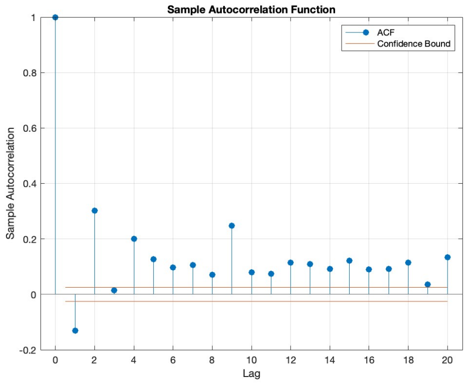
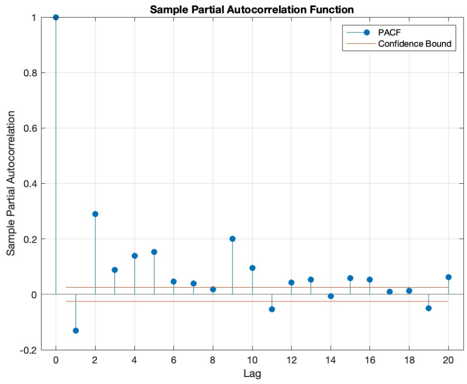


Figure 15: Forecasts Errors for VRP* model

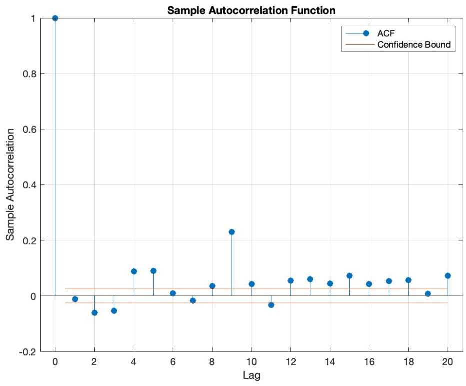


(a) AR(1) forecast errors ACF

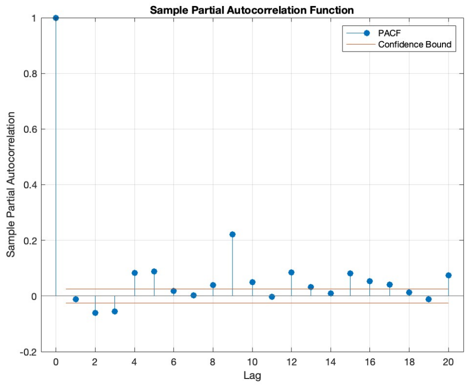


(b) AR(1) forecast errors PACF

Figure 16: AR(1) forecast errors ACF and PACF

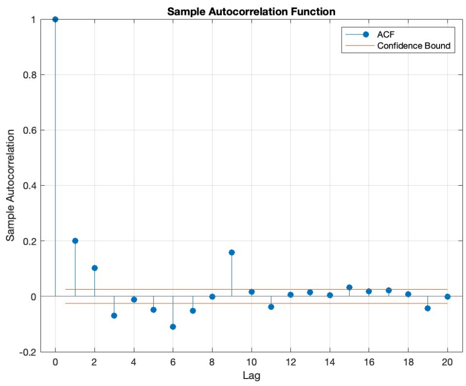


(a) AR(2) forecast errors ACF

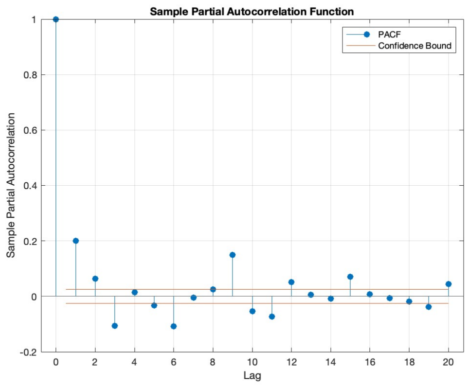


(b) AR(2) forecast errors PACF

Figure 17: AR(2) forecast errors ACF and PACF

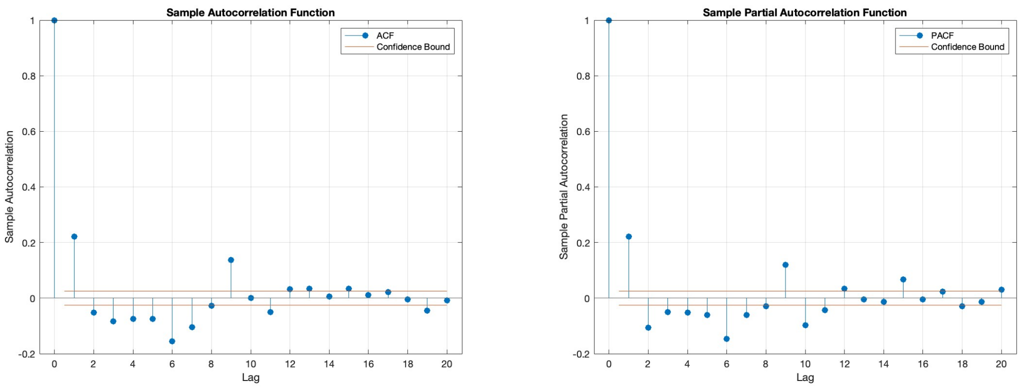


(a) HAR-VRP forecast errors ACF



(b) HAR-VRP forecast errors PACF

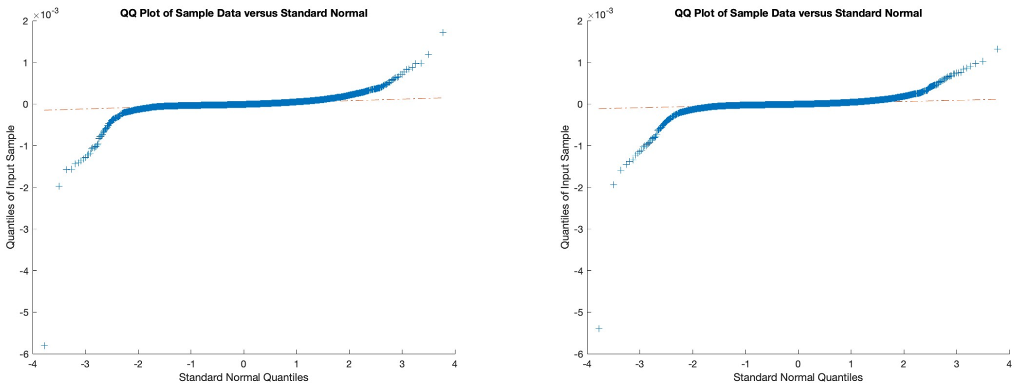
Figure 18: HAR-VRP forecast errors ACF and PACF



(a) VRP* forecast errors ACF

(b) VRP* forecast errors PACF

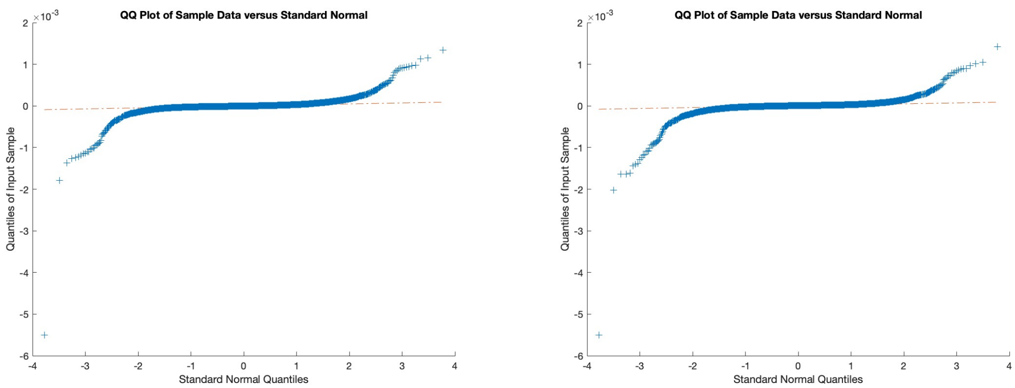
Figure 19: VRP* forecast errors ACF and PACF



(a) AR(1) QQ-Plot

(b) AR(2) QQ-Plot

Figure 20: Forecast Errors QQ-Plot for AR(1) and AR(2) models



(a) HAR-VRP QQ-Plot

(b) VRP* QQ-Plot

Figure 21: Forecast Errors QQ-Plot for HAR-VRP and VRP* models

Looking at the QQ-Plots, in all the four case there are strong deviations in the tails, suggesting that the models are not able to capturing extreme variations.

These plots suggest that none of the estimated models could fully understand and capture the behavior of the S&P500 VRP, meaning that further studies can improve our research. In general, the results indicate that the models could be improved considering alternative error distributions, or using more advanced modeling techniques that better capture the characteristics of the VRP.

Bibliography

Andersen, T., T.Bollerslev, F.X.Diebold, and P.Labys (2000). Great realizations. *Risk 13:105-8.*

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock returns. *Journal of Financial Economics*, 61(1):43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Bakshi, G. and Madan, D. B. (2006). A theory of volatility spreads. *Management Science*, 52:1945–1956.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59:1481–1509.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280.

Bekaert, G. and Hoerova, M. (2016). What do asset prices have to say about risk appetite and uncertainty? *Journal of Banking Finance.*

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bollerslev, T., Gibson, M., and Zhou, H. (2011). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, 160:102–118.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22(11).

Britten-Jones, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55:839–866.

Candila, V. (2024). *rumidas: Univariate GARCH-MIDAS, Double-Asymmetric GARCH-MIDAS arnd MEM-MIDAS models.* R package version 0.1.2.

Carr, P. and Madan, D. (1998). *Towards a theory of volatility trading*, chapter 29, pages 417–427.

Carr, P. and Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, 22(3).

Corsi, F. (2009). A simple approximate long memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Demeterfi, K., Derman, E., Kamal, M., and Zou, J. (1999). A guide to volatility and variance swaps. *The Journal of Derivatives*, 6:9–32.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, pages 1057–1072.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3).

Drechsler, I. (2013). Uncertainty, time-varying fear, and asset prices. *Journal of Finance*, 68:1843–1889.

Drechsler, I. and Yaron, A. (2011). What's vol got to do with it? *Review of Financial Studies*, 24:1–45.

Grishchenko, O. V., Song, D., and Zhou, H. (2022). Term structure of interest rates with short-run and long-run risks. *The Journal of Finance and Data Science*, 8:255–295.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13:281–291.

Londono, J. M. and Zhou, H. (2017). Variance risk premiums and the forward premium puzzle. *Journal of Financial Economics*, 124:415–440.

Markets, C. G. (2022). The global volatility index - vix : Methodology.

Meddahi, N. (2002). Theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17(5):479–508.

Müller, U., Dacorogna, M., Dav, R., Pictet, O., Olsen, R., and Ward, J. (1993). Fractals and intrinsic time - a challenge to econometricians. In *XXXIXth International AEA Conference on Real Time Econometrics*, Luxembourg. 14–15 Oct 1993.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels. *Journal of Financial Economics*, 64:341–372.

Whaley, R. E. (2000). The investor fear gauge. *Journal of Portfolio Management*, 26(3):12–17.

Zhang, B. Y., Zhou, H., and Zhu, H. (2009). Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *The Review of Financial Studies*, 22:5099–5131.

Zhou, H. (2018). Variance risk premia, asset predictability puzzles, and macroeconomic uncertainty. *Annual Review of Financial Economics*, 10(1):481–497.