INSTITUTO
UNIVERSITÁRIO
DE LISBOA

UNIVERSIDADE
DE LISBOA

The Variance Risk Premium

Mafalda Amaro Caneira

Master in Financial Mathematics

Supervisors:
PhD, João Pedro Vidal Nunes, Full Professor, Iscte - IUL
PhD, João Pedro Bento Ruas, Assistant professor, Iscte - IUL

October, 2024

Department of Finance

Department of Mathematics

The Variance Risk Premium

Mafalda Amaro Caneira

Master in Financial Mathematics

Supervisors:
PhD, João Pedro Vidal Nunes, Full Professor, Iscte - IUL
PhD, João Pedro Bento Ruas, Assistant professor, Iscte - IUL

October, 2024

Resumo

Esta tese investiga o comportamento empírico do prémio de risco de variância (VRP) no índice S&P 500. O PRV é definido como a diferença entre a variância realizada esperada sob a medida física e a variância neutra ao risco derivada dos preços das opções. Tradicionalmente, assume-se que a VRP é negativa, o que indica que os investidores estão dispostos a pagar um prémio para se protegerem contra a incerteza futura da volatilidade dos activos. No entanto, estudos recentes têm sugerido desvios a esta premissa teórica, particularmente durante períodos de maior incerteza no mercado. Este estudo visa explorar a validade deste prémio de risco de variância negativa em diferentes condições de mercado, utilizando dados de variância realizados de alta frequência e aplicando modelos ARFIMA para estimar a dependência de longo prazo nas séries de variância. Os resultados fornecem novas perspectivas sobre a persistência da variância realizada e as suas implicações para o prémio de risco da variância em diferentes condições de mercado.

Palavras-chave: Prémio de Risco de Variância, Variância Realizada, Variância Neutra ao Risco, Memória Longa, ARFIMA, Mercados Financeiros, Cobertura de Volatilidade.

Códigos de Classificação JEL: C53, G17.

Abstract

This thesis investigates the empirical behavior of the variance risk premium (VRP) in the S&P 500 index. The VRP is defined as the difference between the expected realized variance under the physical measure and the risk-neutral variance derived from option prices. Traditionally, it is assumed that the VRP is negative, which indicates that investors are willing to pay a premium to hedge against future uncertainty in asset volatility. However, recent studies have suggested deviations from this theoretical premise, particularly during periods of heightened market uncertainty. This research aims to explore the validity of this negative variance risk premium across different market conditions, using high-frequency realized variance data and applying ARFIMA models to estimate long-range dependence in the variance series. The findings provide new insights into the persistence of realized variance and its implications for the variance risk premium in varying market conditions.

Keywords: Variance Risk Premium, Realized Variance, Risk-Neutral Variance, Long Memory, ARFIMA, Financial Markets, Volatility Hedging.

JEL Classification Codes: C53, G17.

Contents

1 Introduction

In financial theory, the concept of the variance risk premium (VRP) has garnered significant attention in financial economics. It plays a crucial role in understanding how investors perceive and manage risk.

The variance risk premium, defined as $E_t[RV_{t,T}] - E_t^*[RV_{t,T}]$, represents the difference between the expected realized variance (RV) under the physical measure and the risk-neutral variance. The prevailing theoretical premise posits that this premium should consistently be negative on average i.e. that implied variance is higher than realized variance. This expectation stems from the notion that investors are willing to pay a premium to hedge against the uncertainty of future asset volatility. Volatility, seen as an adverse risk, is something investors generally seek to avoid, even if it incurs a cost. Thus, the negative sign of the variance risk premium reflects the cost investors are prepared to bear to mitigate volatility risk.

Despite this widely accepted theoretical framework, recent empirical studies have raised questions about the constancy and universality of this negative variance risk premium. Cheng (2019) presents an argument that the negative VIX premium, a close relative of the variance risk premium, tends to "fall or stay flat when risk rises" before rebounding. This observation suggests a deviation from the expected behavior during periods of heightened uncertainty or risk. Similarly, Bekaert and Hoerova (2014) document phenomena indicating that the variance risk premium may not always increase in response to rising risk, challenging the conventional understanding.

These empirical challenges to the theoretical premise are not merely academic curiosities. They have profound implications for financial theory and practical risk management. If the variance risk premium is not consistently negative or does not behave predictably in response to increased risk, the strategies that investors and risk managers rely on, may need to be re-evaluated. Traditional models of asset pricing and risk management, which assume a consistently negative variance risk premium, might not adequately capture the complexities of real-world financial markets.

The primary objective of this thesis is to investigate the empirical validity of the premise that the variance risk premium is always negative. Specifically, this research aims to explore whether this premise holds true across different market conditions and historical periods. Furthermore, it seeks to identify and understand the conditions under which the variance risk premium might deviate from its expected negative value.

To address these research questions, the thesis will do a comprehensive empirical analysis using historical market data. This analysis will involve calculating the variance risk premium for 1-minute time periods and different market conditions to observe its behavior. The results of this investigation will then be compared to theoretical predictions to identify any discrepancies.

By providing a detailed examination, this research aims to contribute to a better understanding of risk and volatility in financial markets. The new findings could offer valuable insights for investors and risk managers, helping them to develop more effective strategies for hedging against volatility. Moreover, the thesis seeks to enrich financial theory by challenging and refining existing models of asset pricing and risk management in light of empirical evidence.

In conclusion, this thesis endeavors to bridge the gap between theoretical expectations and empirical realities concerning the variance risk premium. Through rigorous analysis and critical examination, it aims to enhance our understanding of how investors perceive and manage volatility risk, ultimately contributing to more robust financial models and practices.

2 Literature Review

The variance risk premium (VRP) is a key concept in understanding market dynamics and investor behavior. Carr & Wu (2009) provide a robust framework for understanding and quantifying the VRP using options data. They define the VRP as the difference between realized variance and risk-neutral variance, which is often extracted from market prices of options through variance rates. The risk-neutral variance represents the expected variance under the market's risk-neutral measure, which can be synthesized from a portfolio of European options with varying strike prices and maturities. Carr & Wu's approach builds on the concept of a variance swap, where the fixed swap rate equals the risk-neutral expected value of realized variance. By comparing the realized variance ex-post to the swap rate, the variance risk premium can be quantified.

Carr and Wu's empirical investigation highlights that the VRP is significantly negative for major indices like the S&P 500, demonstrating the market's high demand for volatility protection, particularly during periods of heightened uncertainty.

Tetlock (2023) expandes on this work and introduces an alternative perspective that emphasizes the role of investor sentiment and behavioral biases in shaping the VRP. Tetlock argues that the VRP reflects not only compensation for expected volatility but also investor overreactions to market signals and economic news. This behavioral aspect adds an additional layer of complexity to the traditional understanding of the VRP, suggesting that market participants may be willing to pay a premium not solely based on objective risk assessments but also due to psychological factors driving their risk aversion.

Together, these perspectives highlight the evolution of VRP research from purely statistical and market-based analyses to an inclusion of behavioral insights, enriching our understanding of how variance risk is priced in financial markets.

2.1 Evolution of VRP Research

When considering the methodologies for the VRP calculation, I can categorize the authors into two different schools. The traditional understanding uses straightforward statistical and econometric comparisons. The models used involve calculating implied variances from options markets and comparing with realized variances using historical data.

Several studies have quantified the Variance Risk Premium across several asset classes and market data and conditions. Some authors like Carr & Wu (2009), who focus on comparing risk-neutral variance derived from options with realized variance derived from high-frequency data and Bakshi and Kapadia (2003) investigated whether the volatility risk premium is negative by examining the statistical properties of delta-hedged option portfolios.

Different authors don't only calculate the different variances but they add economic drivers into their work. Their analysis incorporates economic uncertainty components like risk aversion and behavioral expectations.

Bollerslev, Tauchen, and Zhou (2009) combine VRP with, most notably, the P/E ratio. This new method results in greater return predictability. This suggests that volatility and consumption risk have a significant importance in determining returns. In more recent works, Bekaert and Hoerova (2014) introduce a macroeconomic interpretation by linking the Variance Risk Premium to broader economic uncertainty and risk aversion, extending the analysis beyond financial markets.

These two different approaches emphasize the evolution of the VRP research, from purely empirical measurement to an economically informed one.

2.2 Different methodologies in VRP measurement

Focusing on the investigations where the VRP measurement involves the comparison between realized variance and risk-neutral variance from option prices. Bollerslev, Tauchen, and Zhou (2009), Cheng (2019), Carr & Wu (2009) and Andersen et al. (2001) employ the use for intraday (high-frequency) data. Its use provides a much more accurate ex post observations on the actual return variations than the traditional samples based on daily frequency.

However, Bekaert and Hoerova (2014) and Carr and Madan (2002) are among those who assess volatility models using daily data. They state that this approach helps understand long-term trends in volatility and risk premium.

Other studies, such as those using ARFIMA models (Crato & Ray, 1996), underlie the importance of accounting for long-memory effects in volatility processes, as these can significantly affect the persistence of variance and consequently the correct measurement of the VRP. The use of ARFIMA models enables a better forecasting of variance by capturing both short-term and long-term dependencies in the data, which is crucial when calculating risk premiums in financial markets.

2.3 Implications

For traders dealing with volatility products, such as options or VIX futures, the VRP is a critical metric. A high VRP implies that options may be overpriced due to heightened demand for volatility hedging, creating opportunities for traders to engage in volatility arbitrage by selling overvalued options or variance swaps.

The consensus in the literature is that the VRP tends to be negative on average, meaning that investors are willing to pay a premium to hedge against future increases in volatility. However, during periods of market stress, this premium can widen, reflecting heightened risk aversion and uncertainty regarding future volatility.

3 Realized variance

3.1 Data

To proceed to the calculation of the realized variance, I use the intraday SPX prices with 1-minute interval, between 2 January 2018 and 31 May 2022 from LSEG Data & Analytics. These results are the 1-minute interval values of the S&P 500 index.

In its entirety, the data has 429,777 quotes. Each day is composed of 390 price points from the trading hours from 9:31 to 16:00.

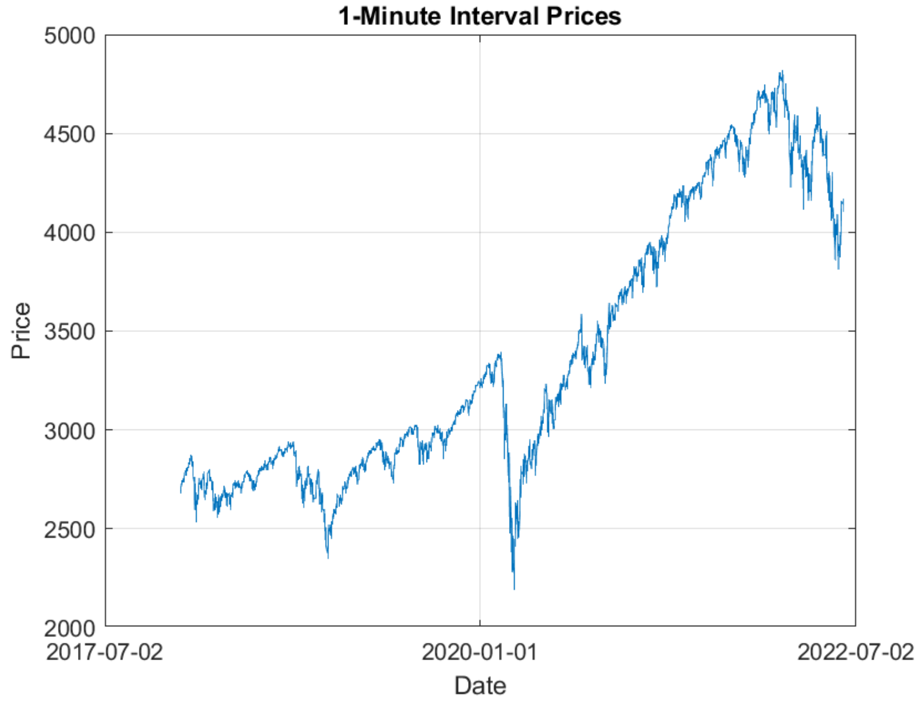

Figure 1: Display of the values of intraday data

3.2 Methodology

Following the methodology used by Tetlock (2023), the daily realized variance was calculated by partitioning each day's 1-minute price data into five distinct groups. By using this cyclical method, I avoid bias from specific times of the day that might have unique volatility characteristics. Sampling every fifth price also reduces the impact of short-term serial correlation, breaking the sequence of consecutive prices and stabilizing variance because intraday prices can exhibit heteroscedasticity. Each group was constructed by sampling every fifth price point within the trading day. Specifically:

- The first group included prices at times 1, 6, 11, 16, ..., 386.
- The second group included prices at times 2, 7, 12, 17, ..., 387.
- The subsequent groups were formed similarly, offset by one minute each.

3.3 Computation

To compute the realized variance for each trading day I calculated, for each group, the sum of squared log-returns. The returns were calculated as the natural log of the ratio between consecutive prices within each group. I then summed the variances of the five groups and divided this sum by five, yielding the daily average realized variance. This cyclical sampling method allows for a robust estimation of daily variance, accounting for the variability and potential auto correlation in high-frequency financial data.

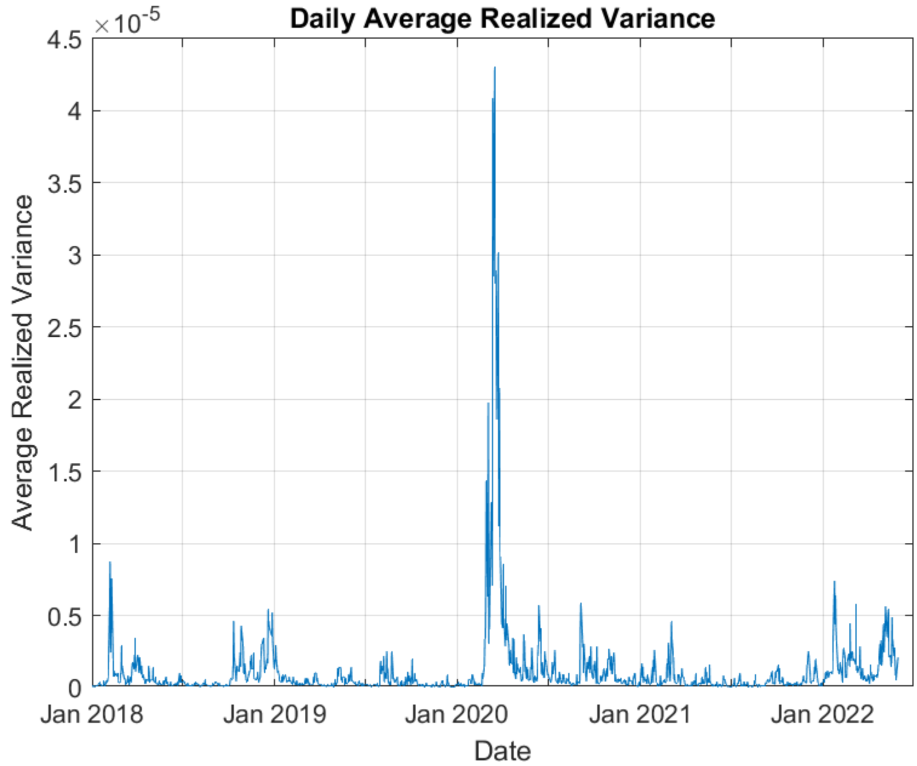

Figure 2: Display of the Daily Realized Variance for 1-minute return

3.4 Demeaned Average Realized Variance

After obtaining the average realized variance for each day, I proceeded to normalize these values. The mean of all daily average realized variances over the analysis period was calculated. Subsequently, I demeaned the average realized variance by subtracting this mean from each daily value. This step standardizes the variances, facilitating the comparison of daily risk relative to the overall average volatility of the S&P 500 index during the specified period.

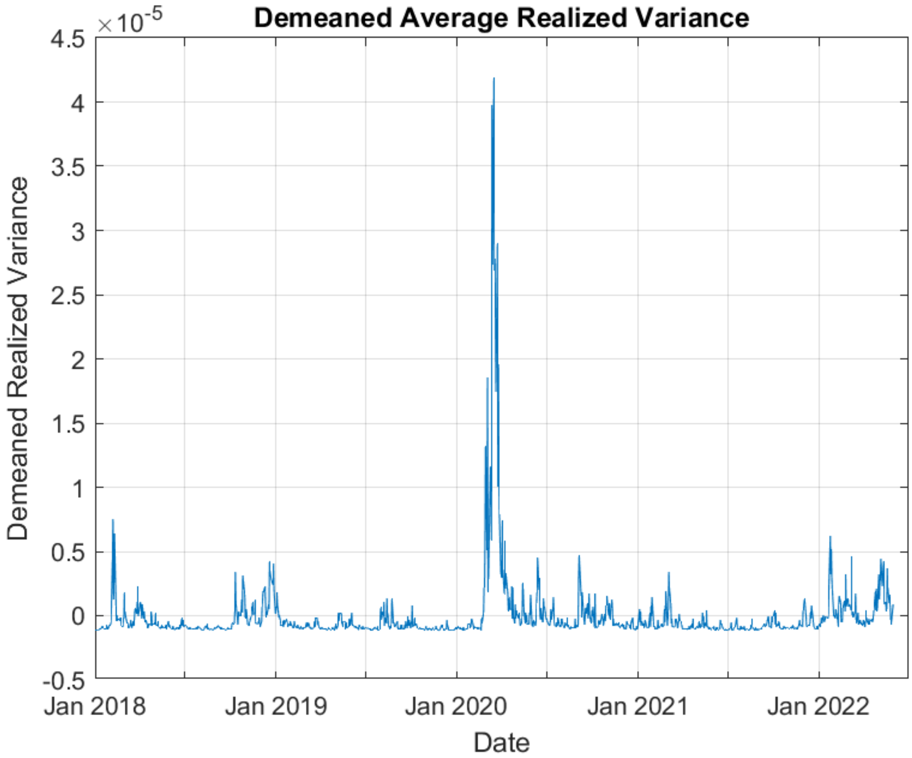

Figure 3: Displays the Daily Demeaned Realized Variance

4 Risk-neutral expected variance

The risk-neutral expected variance represents the expected value of the variance of an asset's return under the risk-neutral measure.

The risk-neutral measure is denoted as $\mathbb{Q}$ and it is a probability measure to price derivatives. The expected returns under this measure are equal to the risk-free rate. It is known that in order to be profitable for the investors, the returns needs to compensate the risk.

$\mathbb{Q}$ assumes that any financial asset generates a rate of return equal to the risk-free interest rate. This probability measure would be used by a risk neutral investor and not a risk averse one. Knowing that for $T \geq t$, X is a financial asset, $\mathbb{P}$ is a probability measure we can define $\mathbb{E}_P(X_T|F_t) := \mathbb{E}_t^*(X_T)$ as the expected value of the unknown future value that X will assume at time T, computed under the probability measure $\mathbb{P}$ and on the information at time t (σ-algebra F_t).

Another definition that we need is regarding money-market accounts. We define r as a risk-free rate with continuous compounding, B_0 as the initial value of the account and constant interest rates the value the t-time value is $B_t = B_0 e^{rt}$.

Applying these definitions we can define risk-neutral measure $\mathbb{Q}$:

$$\mathbb{E}_t^*(X_T) = X_t e^{r(T-t)},$$

Risk neutral expected variance is approximated by the value of a specific portfolio of option named the variance swap rate, it derives from option prices and it is a proxy for the risk-neutral expected variance of return volatility. Analyzing option data and synthesizing variance swap rates can be used to estimate the historical behavior of the variance risk premia. This last mentioned variance quantifies the compensation investors require for taking certain risks about future volatility.

4.1 Variance Swap

A Variance Swap is a financial derivative that allows one party to trade future realized variance against current implied variance and that is often expressed by the VIX index. There are two components of Variance Swaps

- Fixed Leg - Implied Variance

It is set at the outset of the contract and it is based on the implied variance. The VIX is the marktet's 30-day forward-looking volatility expectation and it is reckoned through S&P 500 option prices.

- Floating Leg - Realized Variance

It is determined at the maturity of the contract and it is based on the realized variance of the S&P 500 over the life of the swap. It is typically calculated using high-frequency data like 1-minute returns from the S&P 500 index.

Following the approach by Carr & Wu (2009), because it has a price of zero at inception, the Swap Rate can be represented as the expected realized Variance under the risk-neutral measure $\mathbb{Q}$

$$SW_{t,T} = \mathbb{E}_t^{\mathbb{Q}}[RV_{t,T}] \tag{1}$$

4.2 Payoff of a Variance Swap

The payoff of a variance swap contract starting at time t and expiring at time $T > t$ is $P_{t,t+\tau} = VIX^2_{t,t+\tau} - RV_{t,t+\tau}$ where $VIX^2_{t,t+\tau}$ is the VIX at the beginning of the contract and $RV_{t,t+\tau}$is the realized variance from t to t+τ.

4.3 The Carr and Madan (2001, Appendix 1) spanning formula

I estimate the risk-neutral moments of excess market returns $\tilde{R}_{m,T}$ using the prices of call and put options on the market. This estimation follows the general framework established by Bakshi and Madan (2000) and further specialized by Carr and Madan (2001). According to this formula, the risk-neutral expected excess market returns, raised to the j-th power, can be expressed as:

$$R_{f,t}^{-1}\mathbb{E}_t^*\tilde{R}_{m,T}^j = \frac{j!}{S_t^j}[\int_{F_{t,T}}^{\infty}(K-F_{t,T})^{j-2}C(K)dK + \int_0^{F_{t,T}}(K-F_{t,T})^{j-2}P(K)dK] \tag{2}$$

In this equation, $R_{f,t}$ represents the risk-free return over the time-interval [t, T], such as the return on government bonds, and $\mathbb{E}_t^*\tilde{R}_{m,T}^j$ denotes the jth moment of excess market returns under the risk-neutral probability measure. In finance, moments are statistical measures used to characterize the distribution of returns. The risk-neutral moment of order j for maturity T at time t, $\mathbb{E}_t^*\tilde{R}_{m,T}^j$ has a substantial significance analyzing market risk. For instance, the second moment $\mathbb{E}_t^*\tilde{R}_{m,T}^2$ represents the variance of market returns, providing a measure of market risk, while higher moments, such as the third $\mathbb{E}_t^*\tilde{R}_{m,T}^3$ represents the skewness, which describes the asymmetry and tail risks of the return distribution, showing whether the distribution of returns is biased toward gains or losses.

In this framework, S_t refers to the market SPX index value, K is the strike price for call and put options, priced as $C(K)$ for calls and $P(K)$ for puts. The forward price for the market with maturity T is given by $F_{t,T}$.

The right-hand side of the equation (2) can be approximated by summing an infinite number of infinitely small areas across a range. Because I am using a finite number of call options, with strike prices above the forward price $F_{t,T}$, the function $C(K)$ is only known at certain strike prices K_i. The same will happen for the put options, only its strike prices will need to be lower than the forward price $F_{t,T}$. The integral will be approximated by summing the discrete values.

When one has a finite set of data points the continuous integral over the range $[F_{t,t},\infty[$ is approximated by the discrete sum

$$\int_{F_{t,T}}^{\infty} f(K)\,dK \approx \sum_{i=1}^{n_c} f(K_i)\Delta K_i \tag{3}$$

where ΔK_i denotes the spacing among successive strike price K_i. n_c and n_p represent the number of call and put strike prices, respectively.

Reflecting equation (3) over the right-hand side of equation (2) I obtain

$$\int_{F_{t,T}}^{\infty} (K - F_{t,T})^{j-2} C(K)\, dK \approx \sum_{i=1}^{n_c} (K_i - F_{t,T})^{j-2} C_i \Delta K_i \tag{4}$$

for the call options with $K_i > F_{t,T}$ and similarly

$$\int_{0}^{F_{t,T}} (K - F_{t,T})^{j-2} P(K)\, dK \approx \sum_{i=1}^{n_p} (K_i - F_{t,T})^{j-2} P_i \Delta K_i \tag{5}$$

for the put options with $K_i < F_{t,T}$.

The final result is

$$\mathbb{E}_t \tilde{R}_{m,T}^j = R_{j,t} \frac{j!}{S_t^j} \left[\sum_{i=1}^{n_c} (K_i - F_{t,T})^{j-2} C_i \Delta K_i + \sum_{j=1}^{n_p} (K_j - F_{t,T})^{j-2} P_j \Delta K_i \right] \tag{6}$$

I refer to the right-hand side of equation (6) as a discounted risk-neutral moment, denoted by $\tilde{M} j_{t,T}$.

4.4 Data

In reference to the data used for the risk-neutral calculation I retrieved daily SPX options between the dates of 2 of January 20218 up to 31 of May 2022 from LSEG Data & Analytics. I have, for each day, different strike values for call options and an equivalent number for puts.

4.5 Methodology

I adopt many filtering methods to eliminate securities with low liquidity and unreliable prices. Firstly I use monthly out-of-the-money options, meaning call options with strike price greater than the forward price and put options with a strike price smaller than the forward price. In addition, I select only the daily lowest time to maturity, assuming it exceeds seven days. The options data is filtered to focus on contracts that are near to expiration (but not less than seven days to maturity), allowing for more accurate estimates of short-term market expectations.

The discounted risk-neutral moment that I derived is the $\tilde{M}2_{t,T}$. The following Figure 4 displays the the calculated moments from (6) with the filtered data.

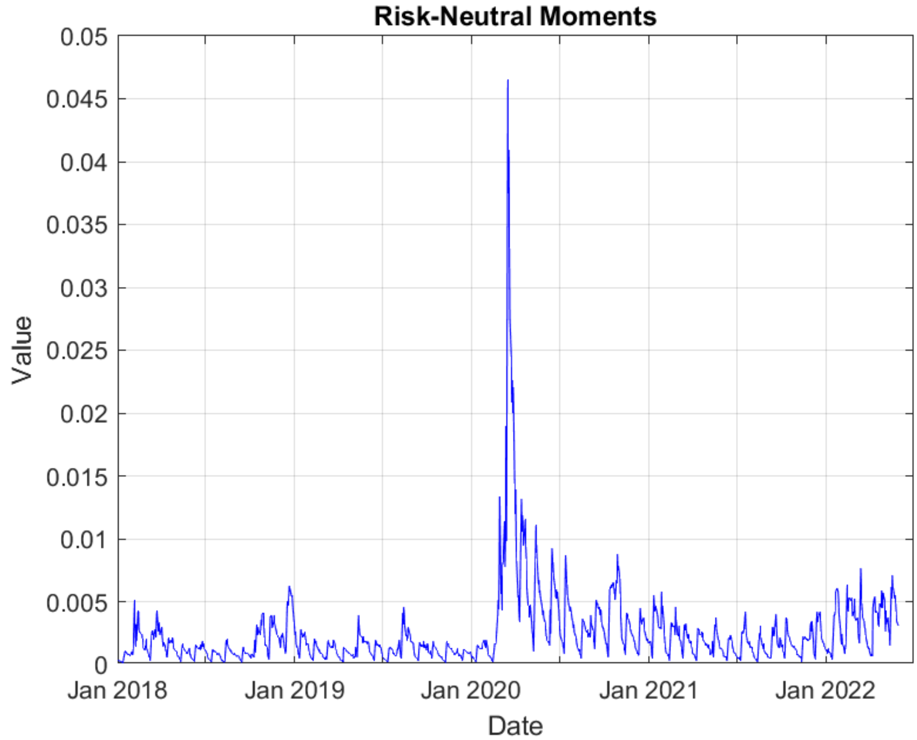

Figure 4: Risk-Neutral τ-day Variance

5 Forecast of Realized Variance

5.1 ARFIMA Model

The fractionally integrated model exhibits a established long memory property of return variance. This long-memory process is stationary whose auto-correlation functions decay at a low pace. We can describe it as a time series model that generalize ARIMA(p, q) (autoregressive integrated moving average) allowing non-integer values of the differencing parameter denoted by d. It can also be defined by ARFIMA (F stands for Fractionally), allowing the model to capture long-range dependencies in time series data. This is particularly useful for modeling data with long memory characteristics, where shocks have a persistent effect that decays over time at a hyperbolic rate. Crato & Ray (1996) results suggest that while ARFIMA models can improve forecasts for long-memory processes, ARMA models often perform similarly in practice unless the time series is strongly persistent, requiring a large sample size for accurate estimation.

The ARFIMA(p,d,q) is defined as

$$(1-L)^d \phi(L)(y_t - \mu) = \theta(L)\varepsilon_t \tag{7}$$

where L is the lag operator, its key parameter is the order of fractional differencing, $0 \leq d \leq 0.5$, $\phi(L)$ is the auto-regressive polynomial of order p and models the influence of past daily realized variance values on the current variance. $\theta(L)$ is the moving average polynomial of order q and captures the influence of past shocks on current variance. ε_t is white noise and μ is the mean of the process.

The fractional differencing operator, $(1-L)^d$, accounts for long memory in the variance, meaning that the influence of past variance values decays at a slow, hyperbolic rate. It is defined as:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k \tag{8}$$

in which $\binom{d}{k}$ is the binomial coefficient. In order to estimate d I set p and q to 0. The equation (7) will now look, simplified, like

$$(1-L)^d (y_t - \mu) = \varepsilon_t \tag{9}$$

Here, $\phi(L) = 1$ and $\theta(L) = 1$ imply that there are no auto-regressive or moving average components. If we denote y_t as the daily realized variance r_{vt}, and assume the mean is constant or negligible, we can simplify further into:

$$(1-L)^d rv_t = \varepsilon_t \tag{10}$$

5.2 ARFIMA Estimator

I use ARFIMA packages from CRAN, designed to fit and analyze ARFIMA models. I use it first for estimating the fractional differencing parameter d in time series data. The R fracdiff() function, which I employ for this estimation, takes as input a numeric vector representing the time series data and a range of potential values to be estimated. It returns an object that includes the estimated d, the variance of the residuals, fitted model values, and the log-likelihood of the fitted model, which indicates how well the chosen model explains the observed data.[1] It is relevant to highlight that the fracdiff() function does not inherently transform the time series to stationary; instead, it estimates the parameter d for the existing time series. A fractional differencing parameter d close to 0 indicates short memory, while values closer to 0.5 (or more) suggest long memory. Users can examine this output for detailed information about the model fit, including confidence intervals for the estimated parameters.

First, I define a historical window length of 252 days, which represents approximately one calendar year of trading. I start the forecast of the RV on the day 253. I apply the rolling window approach ensuring that, as new data points are available, they are added into the historical dataset used for the modeling. This dataset expands and translates into a more accurate forecast due to more reference points being available.

For each date after the initial 252 days, the historical series composed by values of the average daily realized variance is demeaned by subtracting the mean of the series from each data point. This step removes long-term trend and standardizes the variances, facilitating the comparison of daily risk relative to the overall average volatility of the S&P 500 index during the specified period.

I proceed with fitting the fractional differencing model ARFIMA set with p and q to 0 and I estimate the fractional differencing parameter d for each date, that captures the degree of long memory in the time series within the [0, 0.5] range.

For each date that corresponds to a discounted risk-neutral moment, it also has a respective maturity date τ. The ARFIMA model is then used to forecast variance for a given time horizon τ. For each date, the model generates a forecast for τ days ahead and calculates the sum of those values.

The values for the differencing parameter d range between [0.4671379; 0.4982318].

[1]The core idea behind Maximum likelihood estimation (MLE) is to find the parameter that measures how well the chosen model explains the observed data.

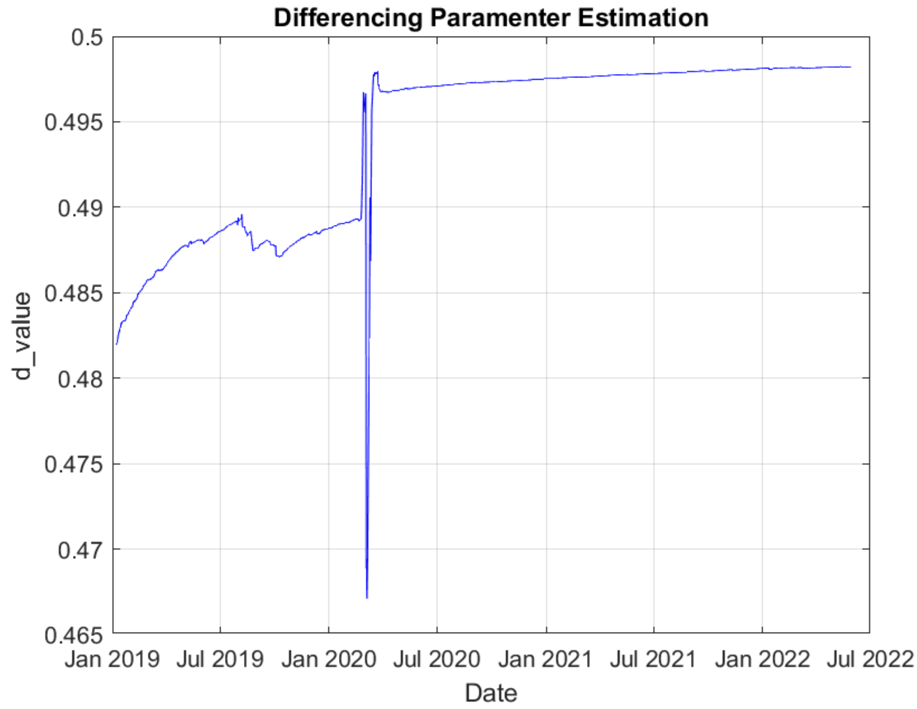

Figure 5: Different values of d throughout the forecasted period

These parameter values translate into a substantial long-range dependence in the data used. The time series exhibits a long memory which is interpreted as the current values of the time series are being influenced by the prior values far in the past, creating a slowly decaying auto-correlation structure. This persistence suggests that volatility, or variance, will not revert to the mean quickly, but will continue to be affected by past shocks over extended periods. Given that $0 \leq d \leq 0.5$, the series is said to exhibit long memory, however it is still stationary, hence the shocks to the system will have unremitting reactions yet fade over time. In my case, d indicates correlations between observations decay gradually, although still bounded. The series is weakly stationary, meaning that the variance, the mean and the covariance are consistent over time.

Concluding, the values of d imply that past values still greatly influence the present ones. To predict future values for the series requires to account for the long history of the prior ones.

5.3 Expected variance over larger horizons

When forecasting variances over larger horizons, the long memory properties imply that the variance forecast will remain elevated for a longer time compared to a series with short memory. Particularly, even distant future variances will keep reflecting the impact of prior volatility events. As the fractional differencing parameter approaches 0.5, this slow decay means that the variance forecasts will exhibit persistence and may not decrease significantly over time.

The series will be more predictable in terms of sustained levels of volatility. However, it also implies that sudden changes in volatility will take time to dissipate, which is crucial when assessing risk or making long-term predictions.

To compute the forecast I used the R forecast() function. The main inputs are h and $level$. The parameter h indicates the number of steps ahead for which forecast are to be calculated and *level* is a vector that identifies the confidence levels for the prediction intervals.

Figure 6 provides a clearer way to assess how future data points are expected to behave based on the ARFIMA model.

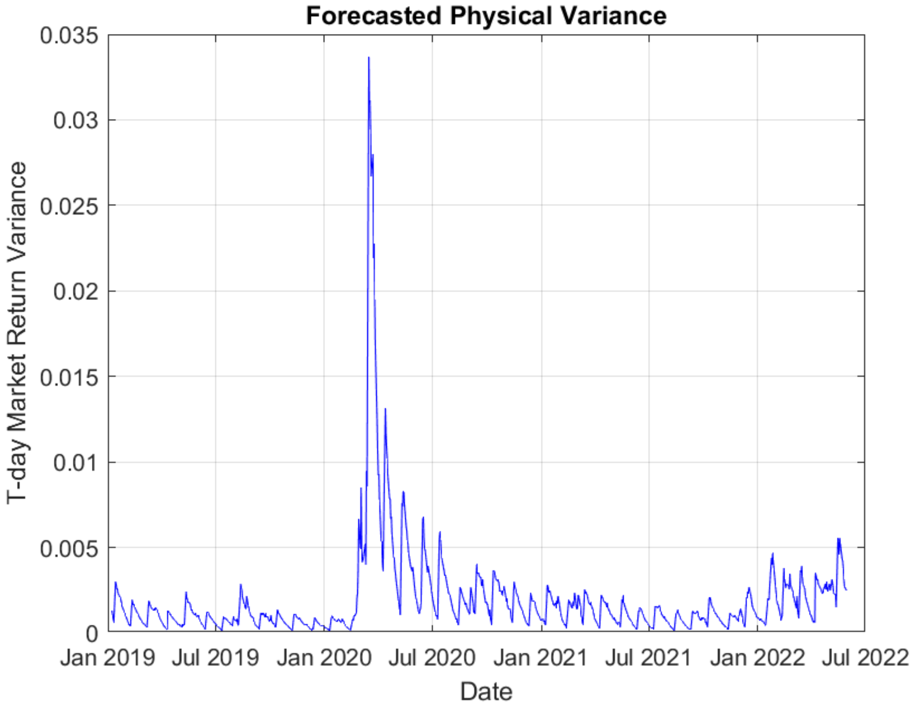

Figure 6: Expected τ-day Variance

6 Variance Risk Premium

There is a fundamental connection between the risk premium in the stock market and the variance and higher-order risk premiums in the option market. Tetlock (2023) interprets the growth-optimal portfolio and the implied risk premium as arising from a heterogeneous agent model in resemblance to the classic behavioral models of Shiller (1984) and Campbell and Kyle (1993).

The model describes two types of investors: growth-optimal investors (who behave like rational log utility maximizers) and behavioral investors (who deviate from this optimal behavior due to various psychological factors or constraints).

The high average equity premium originates from the risk aversion of behavioral investors. They are averse not only to stock market risk but also to variance risk. As a result, they are unwilling to hold a proportionate share of these risks in their portfolios. Consequently, the growth-optimal investors must bear a disproportionate share of market and variance risks. To compensate for this additional risk-bearing, the market equilibrium demands a higher equity premium.

Kansal and Singh (2018) explain that people overestimate their knowledge and abilities, so this overconfidence can result in over-trading, believing that they possess insights and skills. This can lead into extra cost and bad portfolio performance. The investor sentiments also adds complexity to the field of finance. Therefore, during periods of economic and market uncertainty, the behavioral investors might become pessimistic and consequently reduce their demand for stocks. If these investor do not buy or sell the stocks, the price of stocks falls. This decline increases the expected return on stocks, meaning, the equity premium.

With this higher equity premium, in order to get the equilibrium back, it needs to lead the growth-optimal investors into buying more stock, the risk that the behavioral investors eschew from.

This framework provides a bridge between behavioral finance and asset pricing theory. It shows how psychological factors and constraints affecting a subset of investors can influence market-wide prices and expected returns, even in the presence of rational, unconstrained investors.

I use Tetlock (2023) equations for my variance risk premium calculations, assuming that in a frictionless, complete market, all securities are priced by a pricing kernel, M_t, that reflects the market's risk preferences. This kernel is related to the growth-optimal (GO) portfolio, which maximizes long-run expected wealth:

$$M_t = [R_{f,t} + \sum_{k=1}^{\infty} w_{k,t}(\tilde{R}_{m,T}^k - c_k)]^{-1}, \tag{11}$$

where $w_{k,t}$ represents the time-varying weights in the GO portfolio applied to market returns, $R_{m,T}$ and c_k are related to the risk-neutral expectations. M_t reflects the relationship between physical expectations (realized outcomes) and risk-neutral expectations (derived from option prices).

The risk-neutral valuation of any security with return of R_T, its given by

$$\mathbb{E}_t^* R_T = R_{f,t}\mathbb{E}_t[M_T R_T] \tag{12}$$

This price was obtained by discounting under the risk-neutral measure and by using the pricing kernel. This relationship relates observed market prices to actual economic expectations.

Alternatively,

$$\mathbb{E}_t R_T = R_{f,t}^{-1}\mathbb{E}_t^*[M_T^{-1} R_T] \tag{13}$$

Taking $R_T = R_{m,T}^2$, I get the variance risk premium, as the subtraction between physical and risk-neutral variances

$$\mathbb{E}_t \tilde{R}_{m,T}^2 - \mathbb{E}_t^* \tilde{R}_{m,T}^2 = R_{f,t}^{-1}\mathbb{E}_t^*[(M_t^{-1} - R_{f,t})R_{m,T}^2] \tag{14}$$

The pricing kernel is expanded as a polynomial of market returns and includes high-order moments like variance, skewness and kurtosis

$$M_T^{-1} \approx R_{f,t} + \sum_{k=1}^{K} w_{k,t}(R_{m,T}^k - \mathbb{E}_t^* R_{m,T}^k) \tag{15}$$

Focusing only in the second moment (variance) and combining (14 and (15) we get variance risk premium as:

$$\mathbb{E}_t \tilde{R}_{m,T}^2 - \mathbb{E}_t^* \tilde{R}_{m,T}^2 = -R_{f,t}^{-1} \sum_{k=1}^{K} w_{k,t}(\mathbb{E}_t^* \tilde{R}_{m,T}^{k+2} - \mathbb{E}_t^* \tilde{R}_{m,T}^k \mathbb{E}_t^* \tilde{R}_{m,T}^2) \tag{16}$$

7 Empirical results

The estimation of the expected variance premium requires estimates of the risk-neutral and physical variance of market excess returns and its value its given by equation (16). The Variance Risk Premium (VRP) represents the difference between the realized variance (observed in actual returns) and the implied variance (anticipated by options markets) over a given period.

Figure 7 summarizes the estimates of risk-neutral ($\tilde{M}2_{t,T}$) and the predicted market variance.

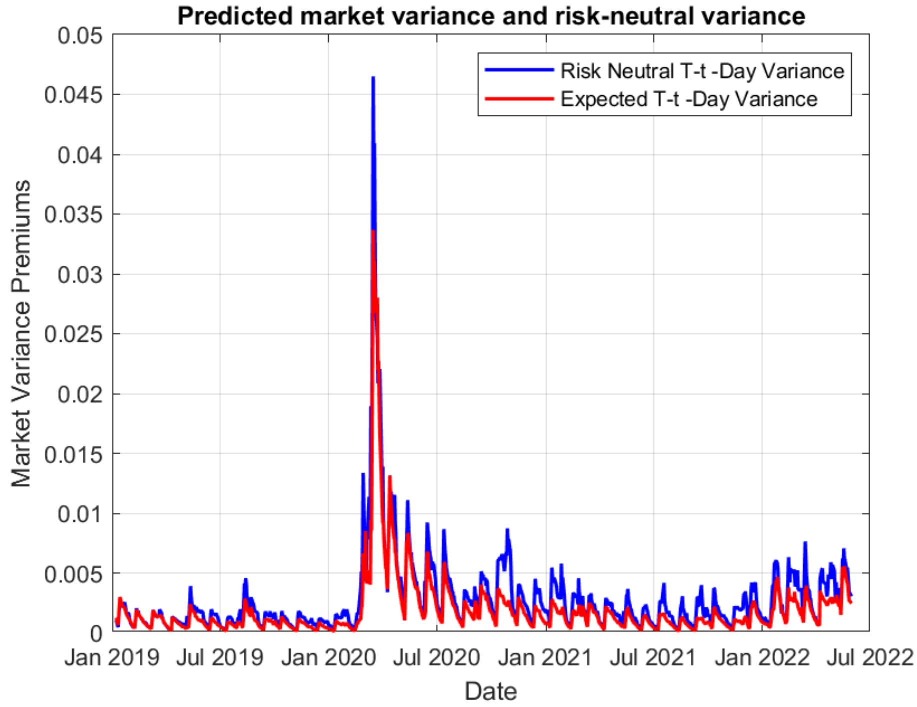

Figure 7: Predicted variance from the fractionally integrated model and risk-neutral variance from option prices at the τ-day horizon. The data are from January 2019 to May 2022.

The fractionally integrated model of expected market variance provides a natural explanation to the puzzle posed by Bollerslev et al. (2009), Bekaert and Hoerova (2014), and Cheng (2019), who find that the variance premium seems to be positive in periods of sudden market turmoil, like September 2008. This behavior is inconsistent with investor aversion to variance risk.

Bollerslev and Todorov (2011) focus on the importance of tail events and the variance risk premium, highlighting how rare but severe market movements affect the pricing of risk in financial markets. They are often referred to as jumps. The market turmoil in 2020 provides a real example of such a jump, where their framework can be used to understand investor behavior, risk perceptions, and pricing during times of extreme volatility.

Investors demanded higher compensation for holding risky assets due to elevated uncertainty and fear surrounding the economic impact of COVID-19. This increased demand for compensation, or risk premium, is exactly what Bollerslev and Todorov (2011) describe. The uncertainty during this period in time resulted in a significant increase in the implied volatility index, VIX, which is indicative of the cost that investors were willing to pay to hedge against further volatility.

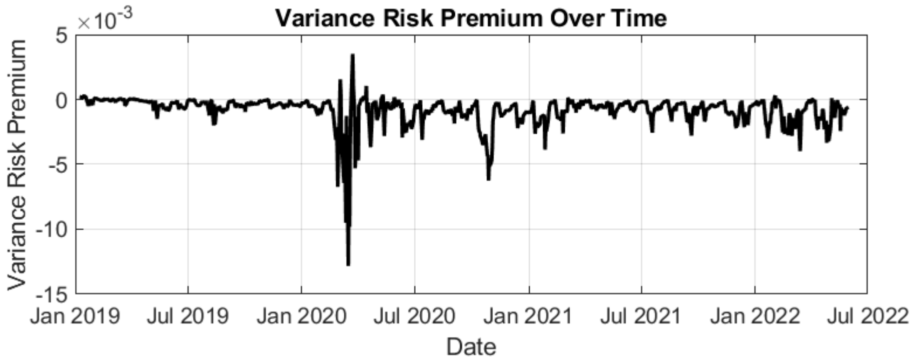

Figure 8: Difference between predicted variance and risk-neutral variance

Figure 8 reveals an extraordinary event. The smallest observed VRP is -0.0128, while the largest VRP is 0.0035. On average, the VRP is -0.0008, indicating a predominantly negative premium, which aligns with the notion that investors generally pay to hedge against future volatility. Specifically, 94.35% of the VRP values are negative, with only 5.65% being positive. This suggests a significant aversion to variance risk among market participants, consistent with the behavior of risk-averse investors seeking protection against potential market downturns.

In detail, Figure 8 displays that the smallest value -0.0128 happened on 16 of March 2020. On this day CNN reported that the S&P 500 plummet 12%, and it was also reported by CNBC as the worst day since the 1987 crash. The extreme negative value on this day reflects the panic and risk aversion of the investors. The risk-neutral variance rose sharply on this day, as shown in Figure 7, compared with the realized variance. The rush to hedge led to a rise in risk-neutral variance and, therefore, an extreme negative value of the VRP.

While a positive VRP is less common, on 23 of March 2020, it reached 0.0035. On this day the realized variance exceeded the implied variance. This sharp change in VRP, in a short period of time, suggests that investor's perception of risk shifted. According to CNBC the Federal Reserve intervened on this day announcing new programs to help with the market functioning. This announcement helped reduce perceived risk, and thus, the risk-neutral variance (which had surged) started to normalize while realized variance remained high, producing a temporary positive premium.

These findings support the work of Bollerslev, Tauchen, and Zhou (2009). They highlighted that the stock market predictability is more pronounced over long horizons. The fact that most of VRP results are negative strengthens the argument that market participants need a consistent compensation for holding assets with uncertain future volatility, especially if it is for long period of times.

My results are also consistent with Carr and Wu (2009). They found that the VRP in generally negative for major indices like S&P 500 indicating that market participants are willing to pay a premium to hedge against market volatility.

Bekaert and Hoerova (2014) provide more support emphasizing the bond between the VRP and economic uncertainty. They argue that the premium is driven by aversion to risk and also a increased perceived likelihood of adverse economic events.

Bakshi and Kapadia's (2003) research regarding the negative returns on delta-hedged option portfolios found that the underperformance is amplified during period of higher volatility. I can interpret in this context that the negative VRP is interpreted as the market's recognition of the higher costs associated with hedging against volatility.

Cheng (2019) also adds an important perspective regarding the dynamics of the VRP. In his work, he discovers that the VIX premium does not always respond predictably during market turmoil and higher risk. Analysing March 2020 through Cheng lenses, it suggests that shifts in hedging demand or investor sentiment may contribute to these "jumps". The VRP underscores the investor behavior under stress.

From a different perspective, when Banks and financial institutions sell put options, they are taking on the risk of increased volatility, however they sell them to generate income and other motivations like in Kelly and Jiang (2014). This is one of the strategies to collect premiums from investors who are trying to protect themselves against downside risk. When the volatility in the market is high, the implied variance, reflected by the options prices, tends to be elevated; this translates that the put options premiums are higher and, therefore, attractive for banks to sell.

Variance Swaps are another financial instrument that allows investors to trade future realized variance against current implied variance and the banks use the swaps to hedge their exposure and pay implied volatility, as in Boyle and Boyle (2001). The dependence on implied volatility means that variance swaps are hardest to price during periods when they are most needed. typically during times of market stress as Martin (2011) analyses. This combination of selling puts and engage in variance swaps can contribute to a negative VRP. They receive compensation (through put premiums) for absorbing this risk and then use variance swaps to potentially gain from the difference between implied and realized volatility, profiting from the investors often overstated fear of future volatility.

Overall, the negative values of VRP are aligned with the existing literature. Volatility is considered and priced as a risk and requires compensation, whether hedging options or adjustments in expected asset returns. This premise is a reflection of the aversion to variance risk, especially during uncertainty time as described by Feunou et al. (2018) and also Drechsler and Yaron (2011).

8 Conclusions

This thesis sheds new light on the properties and economic implications of the variance risk premium (VRP) in financial markets. This work interprets the tendency of the variance risk premium (VRP) to be negative as either a reflection of a premium or, more likely, an indication that there is consistent market demand for insurance across all levels of future expected volatility. This is consistent with the literature as shown in Carr & Wu (2009), Bekaert and Hoerova(2014) or Bakshi Kapadia (2003) who investigate bivariate relationship between implied realized variance, concerning investor risk aversion but also regarding economic uncertainty.

Additionally, the examination of extreme market disturbance such as March 2020 highlights how much VRP acts as an investor sentiment and a view on the general state of the same markets. The large negative VRP in periods of crisis suggests an increase in the demand for out-of-the-money options consistently with evidence from Cheng (2019) and Bollerslev et al. (2009). The shifts between negative and positive VRP, for example, caused by interventions such as the one of Federal Reserve in March 2020, demonstrate how policy actions affect investor risk perception which then causes changes in market volatility.

This thesis also explores how banks and other financial institutions are selling put options and variance swaps. These strategies illustrate a very calculated approach toward the generation of income through capturing the positive VRP with appropriate risk management. The interrelation among these strategies illustrates complex dynamics related to the way financial markets and intermediaries, providing liquidity and absorbing market volatility manage risk.

In other words, the variance risk premium is important to investors' behavior, market expectations, and the cost of volatility risk. The negative nature of VRP was dominating in most cases, especially during normal market conditions. The originality of the present thesis consists, but is not limited to, the empirical analysis of the behavior of the VRP during market stress. In addition, future research may want to go into further detail about the impact of macroeconomic policies on the VRP and explore broader implications for asset pricing and risk management in increasingly volatile markets.

9 References

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens, 2001, The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43-76.

Bakshi, Gurdip, and Nikunj Kapadia, 2003, Delta-hedged gains and the negative market volatility risk premium, *Review of Financial Studies* 16, 527-566.

Bekaert, G. and M. Hoerova, 2014, The VIX, the Variance Premium and Stock Market Volatility, *Journal of Econometrics* 183, 181—192.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance. Risk Premia." *Review of Financial Studies* 22, 4463-4492.

Bollerslev, Tim, and Viktor Todorov, 2011, Tails, fears, and risk premia, *Journal of Finance* 66, 2165-2211.

Boyle, Phelim P., and Feidhlim Boyle. 2001. "Derivative Pricing in Discrete Time." *Journal of Economic Dynamics and Control* 25 (7-8): 1067-1095.

Carr, P. and D. Madan, 2002, Towards a Theory of Volatility Trading (Cambridge University Press).

Carr, P., & Wu, L., (2009), Variance risk premiums. *Review of Financial Studies*, 22 , 1311–1341.

Cheng, I.-H., 2019, The VIX Premium, *Review of Financial Studies* 32, 180—227.

CNBC. "Dow Drops Nearly 3,000 Points, as Coronavirus Collapse Continues; Worst Day Since '87." *CNBC*, March 16, 2020.

CNBC. "The Federal Reserve Just Pledged Asset Purchases with No Limit to Support Markets." *CNBC*, March 23, 2020.

CNN Business. "Dow Plunges 3,000 Points—Its Worst Point Drop in History—as Trump's Remarks Fail to Reassure Investors." *CNN*, March 16, 2020.

Crato, Nuno, and Bonnie K. Ray, 1996, "Model Selection and Forecasting for Long-Range Dependent Processes." *Journal of Forecasting* 15: 107–125.

Drechsler, Itamar, and Amir Yaron. 2011. "What's Vol Got to Do With It: Variance Risk Premiums and Asset Prices." *Journal of Finance* 66 (1): 67-99.

Feunou, Bruno, Mohammad R. Jahan-Parvar, Chayawat Ornthanalai, and Cedric Okou. 2018. "Variance Risk Premiums and the Predictive Power of Financial Uncertainty." *Journal of Financial and Quantitative Analysis* 53 (6): 2471-2512.

Kansal, P. and Singh, S. (2018), "Determinants of overconfidence bias in Indian stock market", *Qualitative Research in Financial Markets*, Vol. 10 No. 4, pp. 381-394.

Kelly, Bryan, and Hao Jiang. 2014. "Tail Risk and Asset Prices." *Review of Financial Studies* 27 (10): 2841-2871.

Martin, Ian. 2011. "Simple Variance Swaps." *NBER Working Paper No.* 16884. National Bureau of Economic Research.

R Project. "The Comprehensive R Archive Network (CRAN)." Accessed September, 2024.

Shiller, R.J., (1984), Stock Prices and Social Dynamics, *Brookings Papers on Economic Activity*, 1984, 457-498.

Tetlock, Paul C., 2023, "The Implied Equity Premium." Columbia Business School.