



Economics and Business

Variance Risk Premium as a Tail Risk Indicator

Mathematical Finance

Supervisor:

Prof. Sara Biagini

Candidate:

Giorgio Rossilli, Matricola 303841

Academic Year 2025–2026

Acknowledgments

I would like to express my deepest gratitude to my thesis supervisor, Prof. Sara Biagini, for her guidance and my family for their support. In particular, I would like to thank my father Dario Rossilli, my grandfather Paolo Santamaria Maurizio, and my dear friend Gaia Esposito for giving me the strength and courage to pursue my goals.

Abstract

This paper is based on the Heston, S. L. (2018) [1] model for Recovering Variance Risk Premium, and it aims to prove how the model-free Implied Variance derived from Option prices and more specifically the Variance Risk Premium can be better indicators for Tail Risk and Tail Events compared to historical variance alone. Starting from the Heston, S. L. (1993) [2] model for stochastic variance we can derive VRP at the instantaneous level to then use a shortcut to risk-neutral expectations calculated through the Fair Variance Swap Rate according to the Demeterfi, K. et al. (1999) [3] Goldman Sachs notes on variance swaps.

To prove the underlying assumption of the paper empirically we will perform tests across 4 different industries: Tech, Pharmaceutical, Defense, and Energy. We will use ETFs and their corresponding Option contracts to represent each different industry as it can be difficult to find indices that represent them accurately. We will identify crash events and substantial shocks like the rise of new technologies, wars, pandemics or energy crisis and prove that a time varying Variance Risk Premium is a more robust indicator compared to historical variance that can sometimes hide information.

Contents

Abstract	iii
I Theoretical Model for Variance Risk Premium	1
1 Introduction	2
1.1 Motivation	2
1.2 Structure of the Thesis	2
2 Stochastic Volatility and probability measure	5
2.1 Heston (1993) Model	5
2.2 Change of Measure and Girsanov	7
3 Asset Pricing Kernel	9
3.1 Introduction to the Stochastic Discount Factor	9
3.2 Deriving Market Prices of Risk	16
4 Deriving the Variance Risk Premium	19
4.1 Proportionality to the square root of variance	19
4.2 Solving the Variance Risk Premium	21
4.3 From Instantaneous to over an Horizon VRP	22
5 Joint Restriction Under Path Independence	23
5.1 Equity Risk Premium and Variance Risk Premium	23
6 Fair Variance Swap Rate	26

6.1	Deriving the Fair Variance Swap Rate from Options	26
6.2	Redefining VRP	30
II	Empirical Analysis	31
7	Empirical Calibration	32
7.1	Introduction to Process	32
7.2	Data Gathering	33
7.3	HAR and Random Forest Modeling	36
7.4	Splitting the Realized Variance Forecast	38
7.5	Logistic Regressions	40
8	Empirical Results	41
8.1	Introduction to Empirical Results	41
8.2	General Results	42
9	Conclusion	48
	Bibliography	50

Part I

Theoretical Model for Variance Risk Premium

1 Introduction

1.1 Motivation

In financial markets variance has always been used as an indicator for tail events. Nowadays the VIX index is known as the "fear gauge" as it represents future expectations of investors. In this paper thanks to the Heston, S. L. (1993) [2] stochastic volatility model and his work on variance premia we will first derive and define the variance risk premia starting from Heston's equations and then we will use the Fair Variance Swap Rate to calculate the risk- neutral expectations of variance. The variance risk premium has often been defined as the difference between implied variance and realized variance under the same time horizon. Heston has instead described it as the difference between risk-neutral expectations and physical expectations on variance along with proving how it intertwined with the concept of Equity Risk Premium. The main goal of this paper is to investigate whether the Variance Risk Premium is a better indicator then historical Variance for tail events.

1.2 Structure of the Thesis

At its core Variance Risk Premium (VRP) is a measure of the extra amount investors pay for Option contracts due to their fear of market crashes. This process generally pushes Implied Variance above expected Realized Variance. In addition, due to the forward-looking nature of implied variance to have an efficient measure of VRP one can only estimate expected future Realized Variance as it is unobservable. The Heston, S. L. (1993) [2] stochastic volatility model inferred that the variance of an

underlying asset is not constant, rather it follows a random mean-reversion process. Heston elaborated how S_t (asset price) remains non-stationary whereas variance is, due to its mean reversion properties, a stationary variable. This model presented itself as a more realistic closed-form way to price Options compared to the Black-Scholes (1973) model that was limited at that time because it assumed constant variance. In fact, if one were to construct a volatility surface from the Black-Scholes it would result as completely flat across all strike prices. Therefore, I decided to implement the model in Heston, S. L. (2018) *Recovering the Variance Premium* [1] to construct the VRP formulation in this paper. The VRP encompasses both expectations and realization of variance as its mathematical representation at a fundamental level is the difference between implied and Realized Variance, that is interpreted as how much investors are paying, or better overpaying, for protection. This definition is quite trivial as we could also infer that the VRP is instead how much investors are willing to pay to make a speculative bet on the market to make profit, if and only if, the right-tail skew is unusually high, meaning investor pay up for OTM Calls. In the latter case VRP is actually representing leveraged bullishness by investors which are trying to hit the “lottery ticket”.

Although, for simplicity the tail events at which we will be looking are limited exclusively to market crashes, which we shall define as unusually low cumulative returns. As previously inferred, a simple representation of VRP (Variance Risk Premium) is the difference between Implied Variance derived from Option contracts of an underlying asset, and the Realized Variance of the returns of such asset. Although, the Heston representation of the same problem is a Variance Premium described as the difference between physical and risk-neutral expectations on variance. Moreover, Heston represents the instantaneous VRP as follows:

$$E^P[dv] - E^Q[dv] = -\lambda v_t dt \quad (1.1)$$

Where $E[.]$ is the conditional expectation with respect to all available information up until time t , sometimes denoted by $\mathcal{F}_t$, and λ , a scalar calibrated to the Option markets, is essentially a multiple that represents the VRP per unit of variance at the

instantaneous level. In addition, $E[dv]$ represents the physical expectation, whereas $E^Q[dv]$ represents the expectation under the risk-neutral probability measure.

The model in Heston, S. L. (2018), *Recovering the Variance Premium* [1], makes use of a stochastic discount factor, or pricing kernel, that depends on the price of the underlying asset and relies on the assumption of path independence. In this paper we review this model and implement a theoretical framework to empirically test it in a discrete setting. Furthermore, we address the change of measure from physical to risk-neutral expectations, as well as how to derive the latter from the fair variance swap rate following Demeterfi, K. et al. (1999) [3] Goldman Sachs Notes on Variance Swaps. To gain insight on the model we also discuss the link between the equity premium and the variance premium in order to fully understand how the VRP behaves according to Heston, S. L. (2018) [1].

After carefully defining the theoretical backbone of this paper, we move on to the empirical testing. we will perform our tests on four different sectors: defence, technology, energy, and pharmaceuticals. These sectors have all experienced high volatility due to wars, the advent of new technologies, energy crises, and pandemics. Thus, the empirical analysis is conducted using ETFs that are representative of the aforementioned industries and their corresponding Option chains, from which we derive market-implied variance. To achieve this, we employ a discretized version of the models presented in this paper, since the financial markets we analyse provide discrete data for the relevant variables.

To do so, we first move from an instantaneous formulation of the VRP to a discrete formulation over a 21 trading-days annualized time horizon. Then we discretize the formulation of the fair variance swap rate so that it can be computed through strips of out-of-the-money (OTM) Options with the same maturity across different strikes. Moreover, we use an HAR model to forecast Realized Variance over the same time horizon. In conclusion, we will statistically test the predictive value of VRP compared to historical volatility and other indicators through logistic regressions and their corresponding statistical measures, such as the Brier score, pseudo- R^2 , and AUC.

2 Stochastic Volatility and probability measure

2.1 Heston (1993) Model

According to the Heston, S. L. (1993) [2] stochastic volatility model, we define the following differential equation to describe asset prices and variance under the physical measure:

$$dS_t = (r + \mu v_t) S_t dt + \sqrt{v_t} S_t dW_t^{1,P}, \quad (2.1)$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_t^{2,P}, \quad (2.2)$$

$$dW_t^{1,P} dW_t^{2,P} = \rho dt \quad (2.3)$$

This formulation is particularly useful due to the characteristic that the two wiener processes are correlated at the instantaneous level, thus connecting the two otherwise independent sources of randomness in a single model. In this model variance is no longer constant but it is a continuous random mean reversion process meaning variance reverts to its long-term average. An even more important observation is that we for the mean reversion process we must place a condition known as the Feller condition, according to the CIR model, to keep the variance strictly positive.

The Feller condition is as follows:

$$2\kappa\theta > \sigma^2 \quad (2.4)$$

From the asset price differential equation under physical expectations (2.1), we can already shed some light on how the equity premium is represented. If we divide both sides of Equation (2.1) by S_t and then take expectations, we are left with only the first term on the left-hand side of the equation, $(r + \mu v_t) dt$, since the expectation of the Wiener process increment is zero by the principles of stochastic calculus.

This process yields the Equity Premium which by definition is as follows:

$$E \left[\frac{dS_t}{S_t} \right] - r dt = \mu v_t dt. \quad (2.5)$$

Intuitively at time t we can infer that the instantaneous equity risk premium is proportional to instantaneous variance with slope μ . In fact, according to Heston, S. L. (2018) *Recovering the Variance Premium* [1], path independence doesn't identify equity risk premium alone rather it jointly restricts it to the variance premium. To better understand this model, we are now going to move to the risk-neutral measure from the physical measure through Girsanov's Theorem. Under the risk-neutral measure, we have a drift $r - q$ with q being the continuous dividend yield of the asset, inference which we make simply through the no-arbitrage intuition that the asset must drift at $r - q$. This is because the risk-neutral measure respects perfectly time value of money principles. Now with Girsanov we must find a market price of risk 2-dimensional vector to move from the physical measure to the risk-neutral measure.

2.2 Change of Measure and Girsanov

Under the risk-neutral probability measure Q Heston defines the following SDE's for the asset price and variance, which are consistent with the no-arbitrage principle, as follows:

$$dS_t = (r - q)S_t dt + \sqrt{v_t} S_t dW_t^{1,Q}, \quad (2.6)$$

$$dv_t = \kappa^Q(\theta^Q - v_t) dt + \sigma\sqrt{v_t} dW_t^{2,Q}, \quad (2.7)$$

$$dW_t^{1,Q} dW_t^{2,Q} = \rho dt \quad (2.8)$$

From here, we can define a 2-dimensional brownian vector as follows:

$$dW_t := \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix} \quad (2.9)$$

with once again the correlation between the two brownian motions being ρ :

$$dW_t^1 dW_t^2 = \rho dt \quad (2.10)$$

Now by applying Girsanov's theorem, we can express the change of measure from the physical measure P to the risk-neutral measure Q through a market price of risk vector Λ_t . Thus, we define the market price of risk vector as follows:

$$\Lambda_t := \begin{pmatrix} \lambda_{1,t} \\ \lambda_{2,t} \end{pmatrix} \quad (2.11)$$

In general, the market price of risk is the extra premium investors demand per unit of risk. In corporate finance it is often associated with the Sharpe ratio. Although, in this scenario it is simply an excellent tool to derive the VRP.

From here, we can define a Radon-Nykodym density process where L_t is a density explaining us how much to “tilt” the probability of each outcome when moving from physical probability measure to risk-neutral probability measure:

$$L_t = \exp \left(- \int_0^t \Lambda_s^\top dW_s - \frac{1}{2} \int_0^t \|\Lambda_s\|^2 ds \right) \quad (2.12)$$

Assumption 2.1. *From the above Radon-Nykodym density process we can infer that the risk-neutral probability measure can be defined as:*

$$\frac{dQ}{dP} = L_t \quad (2.13)$$

From the Assumption (2.1) coupled with the market price of risk vector Λ_t we can define the brownian motion under Q as follows:

$$W_t^Q := W_t^P + \int_0^t \Lambda_s ds = \begin{pmatrix} W_t^{(1),P} + \int_0^t \lambda_{1,s} ds \\ W_t^{(2),P} + \int_0^t \lambda_{2,s} ds \end{pmatrix} \quad (2.14)$$

or equivalently:

$$dW_t^{1,P} := dW_t^{1,Q} - \lambda_{1,t} \quad (2.15)$$

$$dW_t^{2,P} := dW_t^{2,Q} - \lambda_{2,t} \quad (2.16)$$

Now by substituting in the asset price SDE (2.1) and mean-reverting variance SDE (2.2) we can change from physical to risk-neutral probability measure through the above findings. Although, onwards we will see how this is not our true objective but rather a tool to ultimately define the VRP as previously mentioned.

3 Asset Pricing Kernel

3.1 Introduction to the Stochastic Discount Factor

We will see how Girsanov plays an important role for the variance process rather than the stock process. Through Girsanov we will see how the risk-neutral variables relate to physical variables, process which will be key to derive the VRP.

Assumption 3.1. *One of the key assumptions of Heston, S. L. (2018) Recovering the Variance Premium [1] is the concept of path independence that is the idea that the pricing kernel depends only on the current state and not on the path taken to get there.*

From here, before we apply Girsanov we must look at the non-negative pricing kernel consistent with both physical and risk-neutral probability measures, that we know to exist for the Fundamental Theorem of Asset Pricing. From Heston, S. L. (2018) *Recovering the Variance Premium* [1] we know that there is a unique pricing kernel that takes the form of:

$$M_t = M_0 S_t \exp \left(\delta_t + \eta \int_0^t v_s ds + \xi v_t \right) \quad (3.1)$$

- M_0 is a positive constant
- δ_t is a deterministic function of time
- η , γ and ξ are constants

The kernel converts any asset's payoff at t into its current price by weighing future states by their marginal utility, meaning riskier states receive higher weights making payoffs worth less today at t . The primary insight is that the kernel M_t when multiplied by any assets price S_t must be a martingale.

Heston derived each constant under the no-arbitrage principle, such that the constants and the deterministic function of time are jointly restricted as follows:

$$\xi = \frac{\rho\sigma\mu - \kappa^Q + \kappa}{(1 - \rho^2)\sigma^2} \quad (3.2)$$

$$\gamma = -\mu - \rho\sigma\xi \quad (3.3)$$

$$\eta = -\frac{1}{2}(\gamma^2 + 2\rho\sigma\gamma\xi + \sigma^2\xi^2) - \left(\mu - \frac{1}{2}\right)\gamma + \kappa\xi \quad (3.4)$$

$$\delta = -(1 + \gamma)r - \xi\kappa^Q\theta^Q \quad (3.5)$$

In Equation (3.1) the parameters capture different dimensions of investor risk preferences such as γ which is the market price of equity risk per unit of the square root of instant variance, η which is the weight for accumulated variance, ξ which is the weight of instantaneous variance, and δ which is the constant time discount rate component of the equation. In fact, we will derive the variance premium through Girsanov and the pricing kernel, proving that the market price of risk vectors are multiples of the square root of variance. In addition, we will see how we can define VRP through a constant which is a normalized scalar for the market price of variance risk.

Remark 3.1. The pricing kernel in Equation (3.1) does not follow the Assumption of path independence (3.1) thus we must convert it into a path independent kernel before applying girsanov. To do so we must eliminate the accumulated variance in the pricing kernel such that it relies only on state variable S_t and v_t .

To obtain the adjusted SDF what we can do is set the equation of the pricing kernel as follows:

$$M_t = M_0 S_t \exp(\delta_t) h(v_t) \quad (3.6)$$

where $h(v_t)$ is an eigenfunction that must perform two distinct jobs: compensate for having dropped the accumulated variance term $e^{\eta \int_0^t v_s ds}$ and also replace the state dependent component $e^{\xi v_t}$ with the variance process dv_t . Thus we must find a function $h(v_t)$ such that $e^{\delta_t} h(v_t)$ behaves, in terms of its martingale properties, identically to $e^{\delta_t + \eta \int_0^t v_s ds + \xi v_t}$. In fact, $h(v_t)$ must satisfy the following ODE:

$$\mathcal{L}h(v_t) = c(v_t)h(v_t) \quad (3.7)$$

Where $\mathcal{L}$ is a second order differential operator and $c(v_t)$ is a function of v_t . In this scenario $h(v_t)$ is a function which the operator $\mathcal{L}$ must map back to a scalar of itself. This is because v_t is a diffusion process and when you apply Ito's Lemma to any smooth function $h(v_t)$ you will get that the drift component of the differential equation $dh(v_t)$ is unavoidably a second order differential operator acting on $h(v_t)$ evaluated at exactly v_t . In addition, the martingale condition forces $\mathcal{L}h(v_t)$ to be proportional to $h(v_t)$ itself. From here, since we know according to Ito's Lemma that for a stochastic process to be a martingale, its drift must be equal to zero, we must first find the drift of $\frac{dM_t}{M_t}$ from Equation (3.6) and then set it equal to zero. Therefore, we have for $M_t = M_0 \cdot A_t \cdot C_t$ that $\frac{dM_t}{M_t}$ according to Ito's Lemma is:

$$\frac{dM_t}{M_t} = \frac{dA_t}{A_t} + \frac{dC_t}{C_t} + \frac{dA_t}{A_t} \cdot \frac{dC_t}{C_t} \quad (3.8)$$

where we define A_t and C_t as follows:

$$A_t = S_t^\gamma \quad (3.9)$$

$$C_t = \exp(\delta_t) h(v_t) \quad (3.10)$$

First of all, we must first find $\frac{dA_t}{A_t}$. Therefore, we need to find dA through the following Ito's Lemma solution:

$$dA = \frac{\partial A}{\partial S} dS + \frac{1}{2} \frac{\partial^2 A}{\partial S^2} (dS)^2 \quad (3.11)$$

From Equation (3.11) we can find the first and second derivative of A with respect to S as follows:

$$\frac{\partial A}{\partial S} = \gamma S^{\gamma-1} \quad (3.12)$$

$$\frac{\partial^2 A}{\partial S^2} = \gamma(\gamma-1) S^{\gamma-2} \quad (3.13)$$

Now by substituting the above derivatives and the asset price SDE (2.1) under physical probability measure along with its corresponding quadratic variation into the Ito's Lemma solution (3.11) we can find dA . The first step of this substitution process is to first find the quadratic variation of the asset price SDE which under Ito's Lemma properties is as follows:

$$(dS)^2 = \left((r + \mu v_t) S_t dt + \sqrt{v_t} S_t dW_t^{(1),P} \right)^2 = v_t S^2 dt$$

Putting everything together we can find dA as follows:

$$\begin{aligned} dA &= \gamma S^{\gamma-1} \left[(r + \mu v_t) S_t dt + \sqrt{v_t} S_t dW_t^{(1),P} \right] + \frac{1}{2} \gamma (\gamma-1) S^\gamma v_t dt \\ &= \gamma S^\gamma (r + \mu v_t) dt + \gamma S^\gamma \sqrt{v_t} W_t^{(1),P} + \frac{1}{2} \gamma (\gamma-1) S^\gamma v_t dt \end{aligned} \quad (3.14)$$

Such that we can find $\frac{dA}{A}$ as follows:

$$\frac{dA}{A} = \gamma (r + \mu v_t) dt + \gamma \sqrt{v_t} W_t^{(1),P} + \frac{1}{2} \gamma (\gamma-1) v_t dt \quad (3.15)$$

Now we can move on to finding $\frac{dC_t}{C_t}$. Therefore, we need to apply the product rule to get dC_t as follows:

$$dC_t = \delta_t dt + e^{\delta_t} h(v_t) dt \quad (3.16)$$

Then dividing by C_t we get:

$$\frac{dC_t}{C_t} = \delta dt + \frac{dh(v_t)}{h(v_t)} \quad (3.17)$$

From here, we must find $dh(v_t)$ which is a function of v_t . once again applying the product rule we find that $dh(v_t)$ is as follows:

$$dh(v_t) = h'(v_t) dv_t + \frac{1}{2} h''(v_t) (dv_t)^2 \quad (3.18)$$

From the mean-reverting differential equation for variance under the physical measure (2.2) we find that:

$$(dv_t)^2 = \sigma^2 v_t dt \quad (3.19)$$

Such that:

$$\begin{aligned} dh(v_t) &= h'(v_t) \left[\kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_t^{(2),P} \right] + \frac{1}{2} h''(v_t) \sigma^2 v_t dt \\ &= \left[\kappa(\theta - v_t) h'(v_t) + \frac{1}{2} \sigma^2 v_t h''(v_t) \right] dt + \sigma \sqrt{v_t} h'(v_t) dW_t^{(2),P} \end{aligned} \quad (3.20)$$

Therefore, substituting back in Equation (3.17) we have:

$$\frac{dC_t}{C_t} = \left[\delta + \frac{\kappa(\theta - v_t) h'(v_t) + \frac{1}{2} \sigma^2 v_t h''(v_t)}{h(v_t)} \right] dt + \frac{\sigma \sqrt{v_t} h'(v_t)}{h(v_t)} dW_t^{(2),P} \quad (3.21)$$

To complete the identity of Equation (3.8) we must find the cross term $\frac{dA}{A} \cdot \frac{dC}{C}$. The cross term according to the rules of stochastic calculus picks up only the stochastic components as the others become zero during the multiplication. Thus we have that the cross term yields the following:

$$\begin{aligned}
 \frac{dA_t}{A_t} \cdot \frac{dC_t}{C_t} &= \gamma \sqrt{v_t} dW_t^{(1),P} \cdot \frac{\sigma \sqrt{v_t} h'(v_t)}{h(v_t)} dW_t^{(2),P} \\
 &= \gamma \sigma v_t \frac{h'(v_t)}{h(v_t)} \cdot dW_t^{(1),P} dW_t^{(2),P} \\
 &= \gamma \sigma v_t \frac{h'(v_t)}{h(v_t)} \cdot \rho dt
 \end{aligned} \tag{3.22}$$

From here, by putting everything together we find that the drift of $\frac{dM_t}{M_t}$ is as follows:

$$\left[\gamma(r + \mu v_t) + \frac{1}{2} \gamma(\gamma - 1) v_t + \delta_t + \frac{\kappa(\theta - v_t) h'(v_t) + \frac{1}{2} \sigma^2 v_t h''(v_t)}{h(v_t)} + \rho \sigma \gamma v_t \frac{h'(v_t)}{h(v_t)} \right] dt$$

Now we set the drift equal to zero and we multiply by the function $h(v_t)$ to get rid of the denominators, thus yielding:

$$\frac{1}{2} \sigma^2 v h''(v_t) + [\kappa(\theta - v) + \rho \sigma \gamma v] h'(v_t) = -[\gamma(r + \mu v_t) + \delta_t + \frac{1}{2} \gamma(\gamma - 1) v_t] h(v_t) \tag{3.23}$$

From which we get that the unique solution for $h(v_t)$ is:

$$h(v_t) = e^{\xi v_t} \tag{3.24}$$

This yields a Stochastic Discount Factor that ensures there exists not only at least one Equivalent Martingale Measure (EMM), but most importantly a unique EMM. The latter is because Heston constructed the SDF as unique from the start since he knows what the dynamics are under both P and Q for both the asset price and the variance process, thus knowing starting and ending points for the change of probability measure. This means that there is a unique mapping between P and Q probability measures through the following adjusted SDF under path independence:

$$M_t = M_0 S_t^\gamma \exp(\delta_t + \xi v_t) \quad (3.25)$$

Through the Fundamental Theorem of Asset Pricing, FTAP, we know that under the risk-neutral probability Q , our Equivalent Martingale Measure, all discounted traded assets are martingales. In addition, since we know that the pricing kernel is valid for both P and Q dynamics, we will have that $S_t M_t$ is a martingale under P and therefore $E_t^P[S_t M_t | \mathcal{F}_t] = S_t M_t$.

Corollary 3.1. *Given the pricing kernel under the assumption of path independence, for unique values of δ and ξ also the pricing kernel becomes unique such that we have one and only one Equivalent Martingale Measure Q making the market complete under FTAP.*

Thus, $dS_t M_t$ has no drift under P and therefore $S_t M_t$ is a martingale under P and the pricing kernel in Equation (3.25) is uniquely determined by the equity parameter γ and the risk-neutral parameters. In fact, once we set δ , ξ and the equity parameter γ we have, as previously stated, a unique mapping between the two probability measure P and Q .

3.2 Deriving Market Prices of Risk

The market price of variance risk and the market price of equity risk are derived through the pricing kernel under path independence, more specifically by solving for the pricing kernel as the product of M_0 and two other functions of which one is exponential. Thus, we can re-write the pricing kernel as follows:

$$M_t = M_0 \cdot A_t \cdot B_t \quad (3.26)$$

where A_t and B_t are defined as follows:

$$A_t = S_t^\gamma \quad (3.27)$$

$$B_t = \exp(\delta_t + \xi v_t) \quad (3.28)$$

Assumption 3.2. *We know that by definition the market price of risk vector is identified from the diffusion part of $-\frac{dM}{M}$.*

Thus, now that we broke the equation for the pricing kernel into multiple components, we can find the market price of risk vector by finding $-\frac{dM}{M}$ by finding $\frac{dA}{A}$ and $\frac{dB}{B}$.

We have already previously derived $\frac{dA}{A}$ when we derived the pricing kernel free of accumulated variance as shown in Equation (3.15). Thus we know that $\frac{dA}{A}$ is as follows:

$$\frac{dA}{A} = \gamma(r + \mu v_t) dt + \gamma \sqrt{v_t} W_t^{(1),P} + \frac{1}{2} \gamma(\gamma - 1) v_t dt$$

We can now move on to finding $\frac{dB}{B}$ through a similar process. First of all we conveniently re-write B_t as $B_t = e^{X_t}$ where X_t is defined as follows:

$$X_t = \delta_t + \xi v_t \quad (3.29)$$

Therefore, we must first find dX differentiating term by term in the following way:

$$dX = d\delta_t + \xi dv_t \quad (3.30)$$

Starting from Equation (3.30) we can now substitute the mean-reverting variance SDE Equation (2.2) under physical probability measure to find dX as follows:

$$\begin{aligned} dX &= \delta dt + \xi \left(\kappa (\theta - v_t) dt + \sigma \sqrt{v_t} dW_t^{(2),P} \right) \\ &= [\delta + \xi \kappa (\theta - v_t)] dt + \xi \sigma \sqrt{v_t} dW_t^{(2),P} \end{aligned} \quad (3.31)$$

Going back to Equation (3.28) we can now find dB with the Ito's Lemma solution as follows:

$$dB = e^{X(t)} dX + \frac{1}{2} e^{X(t)} (dX)^2 \quad (3.32)$$

Fortunately, since this is an exponential function the only parts that require substitution are dX , which we already derived in Equation (3.31), and its square. Considering that under the properties of stochastic calculus the square of the time increment is equal to zero and the square of the brownian motion differential becomes dt the quadratic variation of dX is as follows:

$$(dX)^2 = (\xi \sigma \sqrt{v_t})^2 dt = \xi^2 \sigma^2 v_t dt$$

Therefore, putting everything together which is substituting dX and its quadratic variation we have that:

$$dB = e^{X(t)} \left\{ [\delta + \xi \kappa (\theta - v_t)] dt + \xi \sigma \sqrt{v_t} dW_t^{(2),P} \right\} + \frac{1}{2} e^{X(t)} [\xi^2 \sigma^2 v_t dt] \quad (3.33)$$

Now we can find $\frac{dB}{B}$ where $B_t = e^{X_t}$ cancels out as follows:

$$\frac{dB}{B} = \left[\delta + \xi \kappa (\theta - v_t) + \frac{1}{2} \xi^2 \sigma^2 v_t \right] dt + \xi \sigma \sqrt{v_t} dW_t^{(2),P} \quad (3.34)$$

From here, we can find through Ito's Lemma the solution for $\frac{dM}{M}$ as follows:

$$\frac{dM}{M} = \frac{dA}{A} + \frac{dB}{B} + \frac{dA}{A} \cdot \frac{dB}{B} \quad (3.35)$$

The only missing part to the above solution is the product of $\frac{dA}{A}$ found in Equation (3.15) and $\frac{dB}{B}$ found in Equation (3.34) in which the stochastic components follow the correlation properties in Equation (2.3) thus giving the following solution:

$$\gamma\sqrt{v_t}dW_t^{(1),P} \cdot \xi\sigma\sqrt{v_t}dW_t^{(2),P} = \gamma\xi\sigma v_t\rho dt \quad (3.36)$$

Therefore, putting everything together we can find $\frac{dM}{M}$ which will contain two stochastic components and a drift component as a result of all the derivations so far. Thus we have the following solution:

$$\begin{aligned} \frac{dM}{M} = & \left[\gamma(r + \mu v_t) + \frac{1}{2}\gamma(\gamma - 1) + \delta + \xi\kappa(\theta - v_t) + \frac{1}{2}\xi^2\sigma^2 v_t + \gamma\xi\sigma v_t\rho \right] dt \\ & + \gamma\sqrt{v_t}dW_t^{(1),P} + \xi\sigma\sqrt{v_t}dW_t^{(2),P} \end{aligned} \quad (3.37)$$

From here, we can find the two market prices of risk by evaluating at the diffusion part of Equation (3.37) according to Assumption (3.2) as follows:

$$-\frac{dM}{M} \Big|_{diffusion} = \lambda_{1,t}dW_t^{(1),P} + \lambda_{2,t}dW_t^{(2),P} \quad (3.38)$$

where $\lambda_{1,t}$, market price of equity risk, and $\lambda_{2,t}$, market price of variance risk, are defined as follows:

$$\lambda_{1,t} = -\gamma\sqrt{v_t} \quad (3.39)$$

$$\lambda_{2,t} = -\xi\sigma\sqrt{v_t} \quad (3.40)$$

4 Deriving the Variance Risk Premium

4.1 Proportionality to the square root of variance

From Equations (3.39) and (3.40) we can infer that it is the square root diffusion of the variance process that forces $\sqrt{v_t}$ into both market prices of risk. In fact, given any pricing kernel that is exponential-affine in v_t and power function S_t , this is completely unavoidable. The result $\lambda_{1,t} = -\gamma\sqrt{v_t}$ and $\lambda_{2,t} = -\xi\sigma\sqrt{v_t}$ is one of the most important structural consequences of the Heston model. Both market prices of risk are proportional to the square root of instantaneous variance $\sqrt{v_t}$, which means that in periods of high volatility (large v_t) investors demand a proportionally larger premium for bearing both equity risk and variance risk. This is consistent with the well-documented empirical regularity that risk premia are counter-cyclical: they are highest during market stress precisely when volatility is elevated.

The fact that $\lambda_{2,t}$ scales with $\sigma\sqrt{v_t}$ rather than with $\sqrt{v_t}$ alone reflects the structure of the variance process: the diffusion of v_t is $\sigma\sqrt{v_t}dW_t^{(2),P}$, so the market price of volatility risk must compensate per unit of that diffusion. Rompolis, L. S., and Travalis, E. (2010) [4] provide empirical corroboration for this structure, showing that risk-neutral moments extracted from Option prices are consistent with a pricing kernel whose sensitivity to variance shocks scales with $\sqrt{v_t}$, lending support to the Heston, S. L. (1993) [2] model specification used here.

Corollary 4.1. *From these derivations we have proven how the market price of risk and more specifically the market price of variance risk is proportional to $\sqrt{v_t}$ such that:*

$$\lambda_{2,t} \propto \sqrt{v_t} \quad (4.1)$$

Thus we can set $\lambda_{2,t}$ as follows:

$$\lambda_{2,t} = \frac{\lambda \sqrt{v_t}}{\sigma} \quad (4.2)$$

Where λ is a scalar constant governing its magnitude, and more broadly a free parameter calibrated to Options market.

Now we can substitute $\lambda_{2,t}$ into the variance SDE (2.2) according to the change of measure Equation (2.16) to find the constraints that will allow us to derive the fundamental Heston solution to instantaneous VRP. Therefore, by substitution we have that:

$$\begin{aligned} dv_t &= \kappa (\theta - v_t) dt + \sigma \sqrt{v_t} \left(dW_t^{(2),Q} - \frac{\lambda \sqrt{v_t}}{\sigma} \right) \\ &= [\kappa \theta - (\kappa + \lambda) v_t] dt + \sigma \sqrt{v_t} dW_t^{(2),Q} \end{aligned} \quad (4.3)$$

From here, by equating the above equation with our target volatility SDE under risk-neutral measure (2.7) we get:

$$[\kappa \theta - (\kappa + \lambda) v_t] dt + \sigma \sqrt{v_t} dW_t^{(2),Q} = [\kappa^Q \theta^Q - \kappa^Q v_t] dt + \sigma \sqrt{v_t} dW_t^{(2),Q} \quad (4.4)$$

Since the two stochastic components in Equation (4.4) cancel each other out, we must simply equate the constants and coefficients of v_t .

By applying the above logic we find the following two constraints under path independence:

$$\kappa\theta = \kappa^Q\theta^Q \quad (4.5)$$

$$\kappa^Q = \kappa + \lambda \quad (4.6)$$

These two equations are crucial to derive the instantaneous VRP that is the difference between the expectations of the SDE for the volatility mean reversion process under physical (2.2) and risk-neutral expectations (2.7). Finally, we have the necessary elements to construct our instantaneous VRP under the Assumption of path independence (3.1).

4.2 Solving the Variance Risk Premium

The instantaneous VRP is given by the difference between the risk-neutral expectations of variance and the physical expectation of variance respectively. Thus, since the expected value of the brownian motion is zero, by taking the expectations of the two variance SDEs (2.7) and (2.2), we get:

$$E^P[dv_t] = \kappa(\theta - v_t)dt \quad (4.7)$$

$$E^Q[dv_t] = \kappa^Q(\theta^Q - v_t)dt \quad (4.8)$$

Now with the two above expectations and the constraints (4.5) and (4.6), we can proceed to solve for the instantaneous VRP as follows:

$$E^P[dv_t] - E^Q[dv_t] = -\lambda v_t dt \quad (4.9)$$

or equivalently by changing signs:

$$E^Q[dv_t] - E^P[dv_t] = \lambda v_t dt \quad (4.10)$$

4.3 From Instantaneous to over an Horizon VRP

The problem with the definition of VRP according to Equation (4.10) is that the VRP is at the instantaneous level at this point and not over a specific time horizon. Thus, if we want a finite time horizon formula for the VRP based on Equation (4.10) that we just derived, we can define it as follows:

$$VRP_{t,T} = \mathbb{E}^{\mathbb{Q}}[RV_{t,T}] - \mathbb{E}^{\mathbb{P}}[RV_{t,T}] \quad (4.11)$$

Equation (4.11) represents a clean and simple definition of VRP which is consistent with the mathematical model of Heston, S. L. (2018) *Recovering the Variance Premium* [1], while giving the possibility to empirically calculate in a much simpler way the VRP and perform statistical testing on its relevance as an indicator for tail events. Moreover, to calculate expected variance under risk-neutral expectations we can make use of convention provided by Demeterfi, K. et al. (1999) [3] Goldman Sachs notes on variance swaps.

Definition 4.2. The fair variance swap rate is the rate that makes the net present value of all the respective variance swaps future cash flows equal zero at inception.

This definition is in turn useful because it is consistent with the time value of money properties of the risk-neutral measure.

5 Joint Restriction Under Path Independence

5.1 Equity Risk Premium and Variance Risk Premium

The central theoretical result of Heston, S. L. (2018) *Recovering the Variance Premium* [1] is that, under the path independence Assumption (3.1), the equity risk premium (ERP) and VRP are not independent of one another, rather they are jointly restricted by the parameters of Heston, S. L. (1993) [2] stochastic volatility model for Option pricing. We already saw as the coefficient for ERP scales linearly with instantaneous variance in Equation (2.5).

Let's start by defining ERP as the extra return investors demand for holding risk compared to being compensated at the risk-free rate; this phenomenon is driven by the drift μ . On the other hand, as we previously mentioned, VRP is the compensation investors demand (or better the premium they pay) for the fear that volatility will spike. Empirically we know that $\rho < 0$ for stock returns and volatility meaning that a crash in prices usually happens at the same time volatility spikes; this is known as the leverage effect which is helpful to explain how VRP leaks into ERP since when an investor holds a stock he indulges in both price risk and volatility risk. In fact, if we view VRP as the cost of hedging volatility, we can infer that if it becomes too expensive to hedge volatility using Options investors will also demand a higher equity premium just to hold the underlying asset. Thus, VRP is not only a tail event

indicator in the short term, as we will prove in the empirical part of the present paper, but it could also be a bullish signal in the medium term since we know that volatility, according to the Heston, S. L. (1993) [2] stochastic volatility model, eventually mean-reverts to its long term average resulting in investors being compensated with ERP. In fact, the leverage effect is a coincident relationship. To better understand this concept, we can use a spring as an analogy where volatility is the tension in the spring, price is the height of the spring, and the risk premium is the potential energy. Therefore, let's assume that at $t=0$ bad news comes such that VRP spikes and the stock price crashes as if the spring is being compressed; now at $t=1$ the cost of being in the market went up as VRP is high and price is low, creating a new equilibrium such that the spring is fully compressed. Empirically, at $t=2$ for investors to stay ERP must rise due to the scary environment and high VRP, and this is when the spring releases its potential energy.

In fact, through the use of the pricing kernel Equation (3.25) from Heston, S. L. (2018) *Recovering the Variance Premium* [1] restricts the two under the Assumption of path independence (3.1) with:

$$-\lambda = -\kappa^Q - \rho\sigma \left(\mu + \frac{1}{2} \right) - \sigma \sqrt{\left(\frac{\kappa^Q}{\sigma} + \rho \left(\mu + \frac{1}{2} \right) \right)^2 - \mu \left(\frac{2\kappa^Q \rho}{\sigma} + \mu - 1 \right)} \quad (5.1)$$

$$\mu = \frac{\rho(\lambda - \kappa^Q)}{\sigma} + \frac{1}{2} - \frac{\sqrt{\rho(\lambda - \kappa^Q) + \frac{\sigma}{2}}^2 - \lambda^2 + (\rho\sigma + 2\kappa^Q)\lambda}{\sigma} \quad (5.2)$$

Where in Equation (5.1) and in Equation (5.2) the risk-neutral long-run variance will be greater than the physical long-run variance, $\theta^Q > \theta$, and risk-neutral variance will be more persistent than physical variance, such that $\kappa^Q < \kappa$. Thus, here Heston devises that the two parameters are not free but once one is determined the other is jointly constricted to a value as shown by the above equations.

Since we already know what the starting point (SDEs under P) and the end point (SDEs under Q) Heston provides a unique SDF that effectively bridges the two probability measures under the Assumption of path independence (3.1) thus eliminating the accumulated variance term as seen in Equation (3.1). This effectively makes ERP as described by Equation (2.5) and VRP as described by Equation (4.10) functions of the same underlying parameters. More specifically Heston solves for the Pricing kernel getting:

$$\mu = -(\gamma + \rho\sigma\xi) \quad (5.3)$$

$$\lambda = \kappa - \kappa^Q = -\rho\sigma\mu + (1 - \rho^2)\sigma^2\xi \quad (5.4)$$

Such that the two slopes of the two risk premia are linear functions of the parameters ξ and γ and therefore jointly restricted as shown by Equations (5.3) and (5.4).

Assumption 5.1. *Given the risk-neutral parameters, once we fix the path independent kernel, thus choosing γ and ξ , we no longer have the freedom to choose VRP independently as its slope is pinned down as a specific function of μ .*

In fact, with out path independence not only we would not have market completeness described by this model, and more specifically a unique pricing kernel, but we would also be able to specify arbitrary discount factors thus making λ and μ anything we desire. This is perhaps one of the most important aspects of the path independence Assumption (3.1) as it ties Options and equity returns by one unique pricing kernel that depends solely on the parameters γ and ξ .

The economic interpretation of such behaviour is that if investors marginal utility is represented by a path independent kernel, then high ERP implies a specific strictly negative VRP in the immediate term, yet a positive relationship in the medium term as explained by the spring analogy above. This is because when VRP widens it typically predict higher subsequent returns on equity since staying in the market has become more risky, thus giving it predictive value.

6 Fair Variance Swap Rate

6.1 Deriving the Fair Variance Swap Rate from Options

According to Demeterfi, K. et al. (1999) [3] Goldman Sachs notes on variance swaps, a continuous-time variance swap on an asset (such as an ETF) pays at T (maturity):

$$Payoff = RV_{t,T} - K_{var}(t, T) \quad (6.1)$$

Where $RV_{t,T}$ is the realized integrated variance over $[t, T]$.

Therefore, we know that in a no-arbitrage pricing scenario, under the risk-neutral measure Q , the fair variance swap rate K_{var} is approximately equal to the risk-neutral expectations of variance $E^Q[RV_{t,T}]$ over the time horizon $(T - t)$ such that:

$$E^Q[RV_{t,T}] \approx K_{var}(t, T) \quad (6.2)$$

This holds regardless of the underlying model, as long as the variance swap is correctly replicated by Options. Thus, to actually empirically calculate the fair variance swap rate we need to construct strips of OTM Options that will then feed into the following time-continuous equation for the K_{var} :

$$K_{var}(t, T) \approx \frac{2e^{r(T-t)}}{T-t} \left(\int_0^F \frac{P(K)}{K^2} dK + \int_F^\infty \frac{C(K)}{K^2} dK \right) \quad (6.3)$$

Such that we would be using OTM puts for $K < F$ and OTM calls for $K > F$ to compute K_{var} . where F is the forward price of the underlying computed as:

$$F = S_t e^{(r-q)(T-t)} \quad (6.4)$$

Now only one problem remains, which is that since we will be dealing with discrete variables a continuous-time process is not a viable solution. Thus, we shall be discretizing in the following way through the use of strike steps ΔK_i for Call and Put Options:

$$\int_0^F \frac{P(K)}{K^2} dK \approx \sum_{K_i < F} \frac{P(K_i)}{K_i^2} \Delta K_i \quad (6.5)$$

$$\int_F^\infty \frac{C(K)}{K^2} dK \approx \sum_{K_i > F} \frac{C(K_i)}{K_i^2} \Delta K_i \quad (6.6)$$

Transforming our process to compute K_{var} in its discretized form as follows:

$$K_{var}(t, T) \approx \frac{2e^{r(T-t)}}{T-t} \left(\sum_{K_i < F} \frac{P(K_i)}{K_i^2} \Delta K_i + \sum_{K_i > F} \frac{C(K_i)}{K_i^2} \Delta K_i \right). \quad (6.7)$$

Naturally for this process we must sort strikes as $K_1 < K_2 < \dots < K_N$ and define ΔK_i as the spacing between adjacent strikes. This way $K_{var}(t, T = t + 21)$ will be our annualized expected variance (under the risk-neutral measure) over a 21 trading-days horizon.

Remark 6.1. There is an additional boundary term in the formula according to Demeterfi, K. et al. (1999) [3], which is:

$$\frac{2}{T-t} \left[r_t(T-t) + \left(\frac{S_t}{F} e^{r(T-t)} - 1 \right) - \ln \frac{F}{S_t} \right]$$

that when $q_t \neq 0$ becomes:

$$\frac{2}{T-t} \left[q_t(T-t) + (1 - e^{q_t(T-t)}) \right]$$

This term is negligible for the purposes of our calculations in the empirical part.

In practice, we use a discrete approximation to the model free replication integral (following VIX logic) which follows Equation (6.7). An important clarification is that in the empirical part of this paper we will be looking at annualized variables on a 21 trading-days window for both the physical and risk-neutral leg, thus calculations will be adjusted accordingly. In addition, it is important to note that fair variance swap rate, compared to IV derived from Black-Scholes, is better suited to calculate risk-neutral expectations of Realized Variance, as it uses all OTM Options across strikes capturing the skew/smile.

Thus, to build our 21 trading-days K_{var} we need to choose the Option maturity T whose time to expiry is closest to 21 trading-days. In addition, for that maturity we need a full strip of plain-vanilla Options including all available OTM puts with strikes $K < F_t$ and all available OTM calls with strikes $K > F_t$ along with current price, risk free rate and dividend yield to compute the Forward price F_t , and lastly we must set $T - t = \frac{21}{252}$.

Now that we have defined the expected variance under risk-neutral expectations we can move on to defining expected variance under the physical measure. As previously stated, under the physical measure variance follows a random mean reversion process explained by Equation (2.2). Naturally, K_{var} , which is consistent with our definition of expected variance under risk-neutral measure, is forward looking. Such that if we were to construct the VRP today we would need to forecast future Realized Variance over the same horizon as K_{var} . Hence, we must take expectations on Realized Variance under the physical measure through some forecasting model.

In the empirical part of this paper, these expectations will translate into forecasted Realized Variance through an auto regressive HAR model using daily, weekly, monthly and quarterly lags along with other significant regressors. This AR model is standard practice for forecasting volatility and variance and when back testing the statistical relevance of each indicator (including VRP), to then compare the statistical measures of each indicator, we will need to forecast variance every time since otherwise the test would not hold statistical significance as we do not know at time t the future 21 trading-days Realized Variance.

For now, we will focus on the following way of describing the physical expectations of Realized Variance:

$$E^P[v_{t+\Delta} | \mathcal{F}_t] = e^{-\kappa\Delta}v_t + (1 - e^{-\kappa\Delta})\theta \quad (6.8)$$

Equation (6.8) is the expectation under P for the instantaneous variance with all the information up to now. More broadly, we define time-continuous Realized Variance as the time integrated variance over $[t, T]$ interval (21 days), which is:

$$RV_{t,T} = \frac{252}{21} \int_t^T v_u du \quad (6.9)$$

From here, one can compute the the time-continuous $E^P[RV_{t,T}]$ using the following affine structure:

$$E_t^P[RV_{t,T}] = \int_t^T E_t^P[v_u] du = \int_0^{T-t} (e^{-\kappa s}v_t + (1 - e^{-\kappa s})\theta) ds \quad (6.10)$$

Once again, it is important to note that in the empirical part of this paper we will be working with discrete variables and not continuous variables such that we will have a discretized version of the above formulas to calculate Realized Variance over the 21-day horizon. Thus, we can better define $RV_{t,T}$ with the following formula:

$$RV_{i,t \rightarrow t+21} = \frac{252}{21} \sum_{u=t+1}^{t+21} r_{i,u}^2 \quad (6.11)$$

Such that for each sector i we define a 21 trading-days forward annualized Realized Variance, without demeaning as an approximation because it would not have a relevant impact on the results. Naturally, this process can only be done ex-post since we do not know what future Realized Variance will be at time t . Thus, we will use the above formula to calculate historical Realized Variance and forecast forward Realized Variance, as previously stated, with an HAR model in the empirical part of this paper.

6.2 Redefining VRP

Now after our last iterations we can finally redefine VRP as the difference between the annualized fair variance swap rate over a 21 trading-days horizon and the annualized Realized Variance over the same 21 trading-days horizon, that we will forecast with our AR model.

Thus, we redefine VRP as:

$$\text{VRP}_{t,T} = K_{\text{var}}(t, T) - E_t^P[\text{RV}_{t,T}] \quad (6.12)$$

Where the left term represents the model-free Implied Variance extracted from the relevant Option strips and the right term is the expected forecasted Realized Variance over the same period.

Part II

Empirical Analysis

7 Empirical Calibration

7.1 Introduction to Process

To test the validity of VRP as a tail event indicator we must first define what a tail event is and collect the relevant data such as cumulative returns, historical variance and Option strips of the relevant assets we are interested in. We will be looking at ETFs that are representative of the 4 industries we have taken into our interest: Tech, Pharmaceutical, Defense, and Energy. We shall use the following ETFs to represent each industry: ITA (for Defense), XLV (for Pharma/Healthcare), XLE (for Energy), and XLK (for Tech). From these ETFs we will extract the relevant Option contracts to derive our variance, under risk-neutral expectations and their prices to calculate their historical Realized Variance and then forecast it for the VRP construction with an HAR model. We will apply a 21 trading-days horizon and the 3m Fred Treasury as the discount rate. Subsequently, we will test the validity of the VRP as a tail event indicator using univariate and multivariate logistic regressions and checking their R^2 , Brier score and AUC along with their coefficients. Empirically the variance in general has an upside and a downside, which in the risk-neutral measure come from the two sides of the Option markets. Therefore we will be looking at variance asymmetrically in the empirical part of this paper. In fact, we will see how also the upside of VRP can give important information for crashes since the both left wings of the Implied Variance rise before crashes. In addition, as previously stated, we will only be closely looking for crashes and not rallies.

7.2 Data Gathering

In this first part of our empirical analysis we will focus on gathering the relevant historical data such as cumulative returns to identify our tail events. To be consistent with the rest of our model we will use log-returns for each sector i . The log-returns will be calculated with a general formula which is:

$$r_i = \ln \left(\frac{P_t}{P_{t-1}} \right), \quad \text{for each sector } i.$$

Subsequently for each sector i we define a 21 trading-days forward cumulative log-return like the following:

$$R_{i,t \rightarrow t+21} = \frac{252}{21} \sum_{u=t+1}^{t+21} r_u$$

Remark 7.1. The log-returns (from 2010 to 2025 calculated with ETF daily closes) will also be used to calculate Realized Variance to train our forecasting model (more specifically an HAR model) that will be used to calculate the physical leg of the VRP. The forecasts might not be very accurate since volatility and variance empirically do not progress linearly. Thus we will do our best to account for jumps and regime changes.

The cumulative log-returns will be used to calculate tail events, which will be defined as follows:

$$\text{Tail}_{i,t} = \mathbf{1} \left\{ R_{i,t \rightarrow t+21} < q_{p,t} \right\} \quad (7.1)$$

where a tail event exist when the above equation is equal to 1 and does not exist when equal to zero. In particular, the above Equation (7.1) checks the 21 trading-days cumulative returns against their 5th percent quantile, thus labeling tail events as horizons which have negative or poor returns with respect to their history.

From this labeling process we can then test the conditional probability of a tail event given that we know variance, or VRP. This can be done through the analysis of the Brier score and AUC of Univariate and Multivariate logistic regressions.

Before moving on to this process we must gather the corresponding strikes and premiums of OTM Options for each sector to construct our fair variance swap rate, K_{var} . I have gathered this data through ivolatility.com (a provider of Option chains data for ETFs). In addition we gather the 3M FRED interest rate for our discount factor component and the trailing 12 month cash dividend yield q for each sector/ETF from Yahoo fianance. From here, we must calculate the fair variance swap rate, or more specifically model-free Implied Variance, and use it as a proxy for the risk-neutral leg of the VRP. For the sake of simplicity we will use the following cutoffs for the strike prices to get the deepest OTM Options:

- OTM calls: $F_t < K \leq 1.4F_t$
- OTM puts: $0.6F_t \leq K < F_t$

To calculate each K_{var} observation we must pick Options with same expiry closest to $t+H$ with $H = 21$ trading-days. Then we gather S_t from closing price for each ETF, r_t and q_t from FRED and Yahoo Finance and calculate F_t with the following equation:

$$F = S_t e^{(r-q)(T-t)} \quad (7.2)$$

From here, we apply the discretized Demeterfi Equation (6.7), that we derived from the conitnuous equation in Demeterfi, K. et al. (1999) [3], to calculate each K_{var} observation. In this K_{var} equation it important to note that the Option mids $P(K)$ and $C(K)$ we got from the Option strips information are all present-value market prices, therefore Demterfi, K. et al. (1999) [3] compounds them with $e^{r(T-t)}$ to cancel initial discounting since they follow the form:

$$P(K) = e^{-r(T-t)} E[\max(K - S_t, 0)] \quad \text{and} \quad C(K) = e^{-r(T-t)} E[\max(S_t - K, 0)] \quad (7.3)$$

From here, to find K_{var} we must find the discretized trapezoid integrals of the Option chains across all strikes following the aforementioned cutoff filters.

In order to find the discretized Trapezoid integrals of the Option chains with spot Option mid at strike K_i we use the following equations:

$$\mathcal{I}_t^{put} = \sum_i \frac{K_{i+1} - K_i}{K_i^2} P_i \quad (7.4)$$

$$\mathcal{I}_t^{call} = \sum_i \frac{K_{i+1} - K_i}{K_i^2} C_i \quad (7.5)$$

From here, we define the Put/Call wings of OTM Options for the K_{var} as follows:

$$K_{var,t}^{put} = \frac{2}{T-t} e^{r(T-t)} \mathcal{I}_t^{put} \quad (7.6)$$

$$K_{var,t}^{call} = \frac{2}{T-t} e^{r(T-t)} \mathcal{I}_t^{call} \quad (7.7)$$

Such that $K_{var,t}$ when $q_t = 0$ is defined as:

$$K_{var,t} = K_{var,t}^{put} + K_{var,t}^{call} \quad (7.8)$$

where as $K_{var,t}$ when $q_t \neq 0$ includes the boundary term described in Remark (6.1), such that it is defined as:

$$K_{var,t} = \frac{2}{T-t} [q_t(T-t) + (1 - e^{q_t(T-t)}) + e^{r(T-t)}(\mathcal{I}_t^{put} + \mathcal{I}_t^{call})] \quad (7.9)$$

We quote about 300 observations of K_{var} per ETF therefore we interpolate values between quote/observation days in the time-series analysis.

Remark 7.2. the put-wing sum weighs strikes near F_t such that we have in some periods many near-ATM OTM puts with wide bid-ask spreads and inflated mids that therefore can spike K_{var} . We will see how this effect is particularly visible in for the ETF for the Pharmaceutical/Healthcare sector XLV. Although, the K_{var} model free method still remains efficient for our empirical studies in this paper.

7.3 HAR and Random Forest Modeling

To compute VRP once we have K_{var} we must find Realized Variance over the same horizon. I decided to implement a Heterogeneous Autoregressive model that follows the rationale of the model proposed by Corsi, F. (2004) [5]. The HAR model we implement, differently from Corsi (2004), F. [5], takes into account continuous and jump components across different time frames: daily, weekly (5 trading days), monthly (22 trading days), and quarterly (66 trading days). The reason this model looks across different time frames is because we assume investors have different investing time horizons and therefore we will have a Continuous and a Jump regressor for each time frame previously mentioned. In addition, we will have a VIX and an SPY Realized Variance (of the last 22 trading days) regressors to improve the prediction accuracy of the model. An important note is the fact that the entire model is in log space therefore after we obtained the log forward Realized Variance prediction we must re-convert it to the same level units of K_{var} in order to find the effective Physical leg and have coherent results when we compute the VRP. The model's coefficients are trained on the 10 years between the start of 2010 and the start of 2020 whereas intuitively each regressor only uses up to 66 trading days of data.

Lets start by analyzing how the continuous and jump components work inside the model. For our model we first must calculate daily annualized Realized Variance to then compute the Bipower Variation (BPV) that accounts for the continuous components and then from that derive the jump component (J) through the following formulas:

$$RV_{t,d} = 252 r_t^2. \quad (7.10)$$

$$BPV_t = \frac{\pi}{2} 252 |r_t| |r_{t-1}|. \quad (7.11)$$

$$J_t = \max(RV_{t,d} - BPV_t, 0). \quad (7.12)$$

Log continuous and jump components are calculated at daily scales and at weekly, monthly and quarterly scales as rolling means of the daily results for the past 5, 22, and 66 days respectively.

In particular, we will use as regressors the natural logarithms of the smoothed BPV and J components, where smoothed indicates the rolling-mean characteristic of the weekly, monthly, and quarterly scales. Thus we defined these log smoothed components as follows:

$$\ell C_{\Omega} = \ln(\text{smoothed BPV}), \ell J_{\Omega} = \ln(\text{smoothed jump}), \Omega \in \{d, w, m, q\} \quad (7.13)$$

Therefore yielding 8 regressors: $\ell C_d, \ell C_w, \ell C_m, \ell C_q, \ell J_d, \ell J_w, \ell J_m, \ell J_q$.

In addition, as previously mentioned, we have other two regressors for the VIX and the SPY Realized Variance that are defined as follows:

$$\ell \text{VIX}_t = \ln((\text{VIX}_t/100)^2) \quad (7.14)$$

$$\ell \text{SPY_rv}_t = \ln(\text{annualized 22-day RV of SPY returns}) \quad (7.15)$$

From here, putting everything together we get the following HAR-CJ-X regression that targets the log forward Realized Variance over the next $H = 21$ trading-days:

$$\begin{aligned} y_t = & \beta_0 + \beta_{C_d} \ell C_d + \beta_{J_d} \ell J_d + \beta_{C_w} \ell C_w + \beta_{J_w} \ell J_w + \beta_{C_m} \ell C_m + \beta_{J_m} \ell J_m \\ & + \beta_{C_q} \ell C_q + \beta_{J_q} \ell J_q + \beta_{\text{VIX}} \ell \text{VIX}_t + \beta_{\text{SPY_rv}} \ell \text{SPY_rv}_t \end{aligned} \quad (7.16)$$

The objective of the final regression model will then output the log forward Realized Variance over horizon $H = 21$ trading-days as previously mentioned such that:

$$y_t \equiv \ln \text{RV}_{t \rightarrow t+H}^{\text{fwd}} \quad (7.17)$$

where,

$$\text{RV}_{t \rightarrow t+H}^{\text{fwd}} = \frac{252}{H} \sum_{k=1}^H r_{t+k}^2 \quad (7.18)$$

In addition to our HAR modeling, to achieve a better forecast, we also constructed a Random Forest model with 400 trees and maximum depth of 6 (for less overfitting to noise) taking an average of the 400 tree predictions to blend in with our HAR model. Finally, we applied a Jensen correction of residual variance to account for Jensens inequality under the uncertainty of our model.

Therefore, we get our final prediction as follows accounting also for the fact that we must compare it with K_{var} which is not under log form:

$$\widehat{RV}_t^{\text{fwd}} = \exp\left(\frac{1}{2} \ln \widehat{RV}_t^{\text{OLS}} + \frac{1}{2} \ln \widehat{RV}_t^{\text{RF}} + \frac{\hat{\sigma}^2}{2}\right) \quad (7.19)$$

7.4 Splitting the Realized Variance Forecast

To perform efficient tests on the VRP we must split it into Upside and Downside VRP to get clearer picture of how VRP can act as a tail event indicator. Therefore, we are not going to run separate HAR and random forest models, but we are going to split forward Realized Variance, RV_t^{fwd} , and define Upside (UVRP) and Downside (DVRP) VRP with semivariances and K_{var}^{put} and K_{var}^{call} . We start by defining backward $H = 21$ trading-days semivariances up to t as follows:

$$\text{RSV}_t^- = \frac{252}{H} \sum_{k=0}^{H-1} r_{t-k}^2 \mathbf{1}_{\{r_{t-k} < 0\}} \quad (\text{negative returns only}) \quad (7.20)$$

$$\text{RSV}_t^+ = \frac{252}{H} \sum_{k=0}^{H-1} r_{t-k}^2 \mathbf{1}_{\{r_{t-k} > 0\}} \quad (\text{positive returns only}) \quad (7.21)$$

we also define $H = 21$ trading-days backward historical Realized Variance to later perform tests on its validity as a tail event indicator as follows:

$$\text{RV}_t^{\text{hist}} = \frac{252}{H} \sum_{k=0}^{H-1} r_{t-k}^2. \quad (7.22)$$

From here, we can define our forward looking semivariances to construct our UVRP and DVRP as follows:

$$\widehat{\text{RSV}}_t^{-,\text{fwd}} = \widehat{\text{RV}}_t^{\text{fwd}} \cdot \frac{\text{RSV}_t^-}{\text{RSV}_t^- + \text{RSV}_t^+} \quad (7.23)$$

$$\widehat{\text{RSV}}_t^{+,\text{fwd}} = \widehat{\text{RV}}_t^{\text{fwd}} \cdot \frac{\text{RSV}_t^+}{\text{RSV}_t^- + \text{RSV}_t^+} \quad (7.24)$$

In addition to the UVRP and DVRP, we will also construct a log-ratio VRP and model-free Implied Volatility derived from K_{var} to check if they provide any additional value according to our statistical model. Thus, let's define our different VRP measures and model-free Implied Volatility as follows:

$$\text{VRP}_t = K_{\text{var},t} - \widehat{\text{RV}}_t^{\text{fwd}} \quad (7.25)$$

$$\text{VRP}_t^{\text{ratio}} = \ln \left(\frac{K_{\text{var},t}}{\widehat{\text{RV}}_t^{\text{fwd}}} \right) \quad (7.26)$$

$$\text{DVRP}_t = K_{\text{var},t}^{\text{put}} - \widehat{\text{RSV}}_t^{-,\text{fwd}} \quad (7.27)$$

$$\text{UVRP}_t = K_{\text{var},t}^{\text{call}} - \widehat{\text{RSV}}_t^{+,\text{fwd}} \quad (7.28)$$

$$\text{IV}_t = \sqrt{K_{\text{var},t}} \quad (7.29)$$

Remark 7.3. In our definition of tail events we are looking at crashes, although, UVRP is still useful and adds important information. This is because in stressed markets, often both wings become expensive due to fear and lottery demand. Additionally, symmetric VRP tends to collapse level, skew, and wings into one number, which might not be efficient.

Now that we have constructed our symmetric VRP and our two put and call wing wedges, UVRP and DVRP, we can now test their validity as tail event indicators.

7.5 Logistic Regressions

To verify which is the best tail event indicator we will perform univariate and multivariate logistic regressions and verify their R^2 , Brier score, and AUC. We will start by defining the logistic regression that we will perform as follows:

$$\mathbb{P}(Tail_{i,t} = 1) = \Lambda(\alpha + \gamma_1 x_{1,t} \dots), \quad \Lambda(u) = \frac{1}{1 + e^{-u}} \quad (7.30)$$

We will perform univariate logistic regressions for the following indicators:

$$\{VRP_t, \widehat{RV}_t^{\text{fwd}}, RV_t^{\text{hist}}, IV_t, VRP_t^{\text{ratio}}, DVRP_t, UVRP_t\}$$

In addition, we will perform also the following multivariate logistic regressions:

$$\mathbb{P}(Y_t = 1) = \Lambda(\beta_0 + \beta_1 VRP_t^{\text{ratio}} + \beta_2 DVRP_t + \beta_3 UVRP_t),$$

$$\mathbb{P}(Y_t = 1) = \Lambda(\delta_0 + \delta_1 IV_t + \delta_2 \widehat{RV}_t^{\text{fwd}}),$$

$$\mathbb{P}(Y_t = 1) = \Lambda(\eta_0 + \eta_1 IV_t + \eta_2 \widehat{RV}_t^{\text{fwd}} + \eta_3 RV_t^{\text{hist}}).$$

Running this horse race of logistic regressions and comparing their statistical measures we will effectively find the overall best tail event indicator or at least valuable information regarding the predictive power of each indicator.

8 Empirical Results

8.1 Introduction to Empirical Results

Applying the empirical model discussed previously we obtained results on the statistical relevance of each indicator singularly and also of the indicators together. The purpose of the multivariate regressions was to apply a holistic approach to better understand the nuances that regard tail events such as skew, level, wings, and the fact that Options richness is asymmetrical so much that even for crashes Upside VRP contains valuable predictive information. We have also described the intuition behind why VRP can result more useful compared to Implied Variance (or Implied Volatility) and historical Realized Variance alone, as it encompasses the very premium investors are willing to pay for greed or fear. In this paper we mainly addressed how the variance risk premium can work as fear gauge possibly better than volatility indexes like the VIX.

Remark 8.1. It is important to note that the results we have obtained may be biased due to inaccurate prediction of the Physical Leg of VRP since both the HAR model and the random forest may be biased for a variety of reasons, like the mere fact that variance doesn't evolve linearly as described by regression models. Nonetheless, we applied these models to forecast forward Realized Variance as a mere proxy to actual Realized Variance in the same time frame as the risk-neutral leg which is unobservable at time t . In fact, one could also look at the valuable information ex post VRP may contain by actual Realized Variance and model-free Implied Variance over the same period.

8.2 General Results

The HAR-CJ-X and Random Forest model have decent R^2 values for all indicators, specifically: 0.5439 for ITA, 0.4241 for XLV, 0.6162 for XLE, and 0.5801 for XLK. The Random Forest has generally provided better R^2 values across all ETFs although the HAR-CJ-X model also yielded good results, in particular the HAR component of the Realized Variance prediction model added important general market information from the VIX and the SPY Realized Variance; in fact, the VIX consistently showed the highest coefficient in the regression across all ETFs. On the other hand, univariate and multivariate logistic regressions yielded interesting results, in particular, symmetric VRP alone yielded similar results to IV, RV^{hist} and forecasted RV, varying from one ETF to another as we can see from Table (8.1), Table (8.2), Table (8.3), and Table (8.4).

Although, on the multivariate regressions there was a significant difference: across all 4 ETFs the multivariate regression of VRP ratio (log), DVRP and UVRP had significantly higher AUC and lower brier scores indicating the best possible combination to predict tail events. It is crucial that we take a holistic approach when trying to predict tail events as encompassing all the information available into one single number/indicator would introduce bias. Additionally, the multivariate regression with the VRP ratio (log), UVRP and DVRP yielded consistently low p-values except in ITA which was empirically an outlier in this study. ITA still presented the best AUC and brier scores in the multivariate regression VRP ratio (log), UVRP and DVRP but the p-value for that regression was significantly higher compared to the other ETFs which showed for the same regression a p-value near 0.001.

XLE and XLK remain the ETFs which majorly support the thesis of VRP as an effective Tail Event indicator, although, also ITA and XLV provide important evidence on how taking a holistic approach and splitting VRP into two asymmetrical legs can still provide useful information on the market. In addition, it is important to note, as explained in Remark (7.2), that the measure of VRP may be not exact both for a modeling error of K_{var} in cases where there is an abundance of near-ATM OTM puts and for the fact that we interpolate between K_{var} quote days.

What is most clear by this study is that VRP had the most action during the first two years of the studies almost across all ETFs as we can see from Figure (8.1), Figure (8.2), Figure (8.3) and Figure (8.4). It is clear that the market was reacting during this period due to COVID, though it is interesting to note that it was the action of the VRP that showed a stressed market and not necessarily the value of VRP itself as often forecasted RV would follow the K_{var} measure thus compressing VRP in most moments. This is because K_{var} jumps first and then RV and its forecast quickly catchup therefore the wedge can be compressed as both RV (forecasted and actual) and K_{var} move in the same direction.

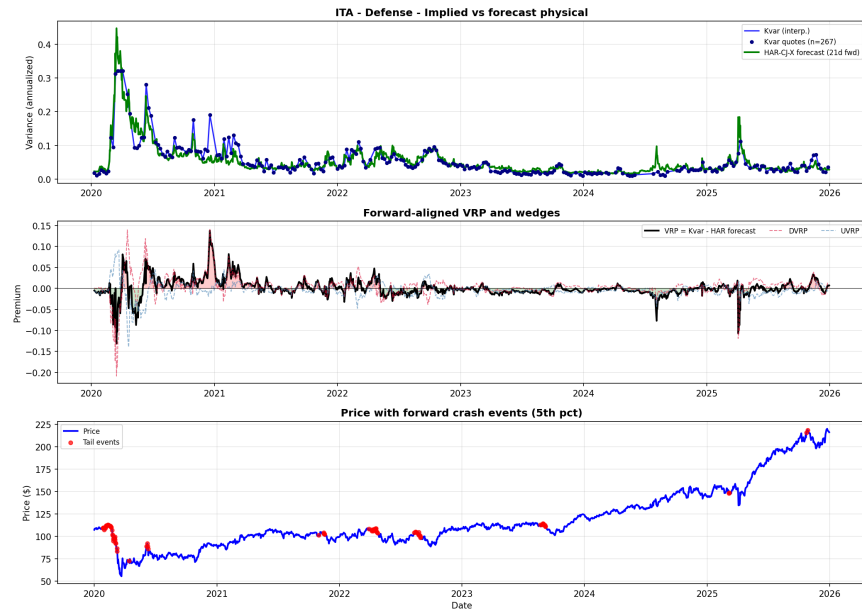


Figure 8.1: ITA Diagnostics

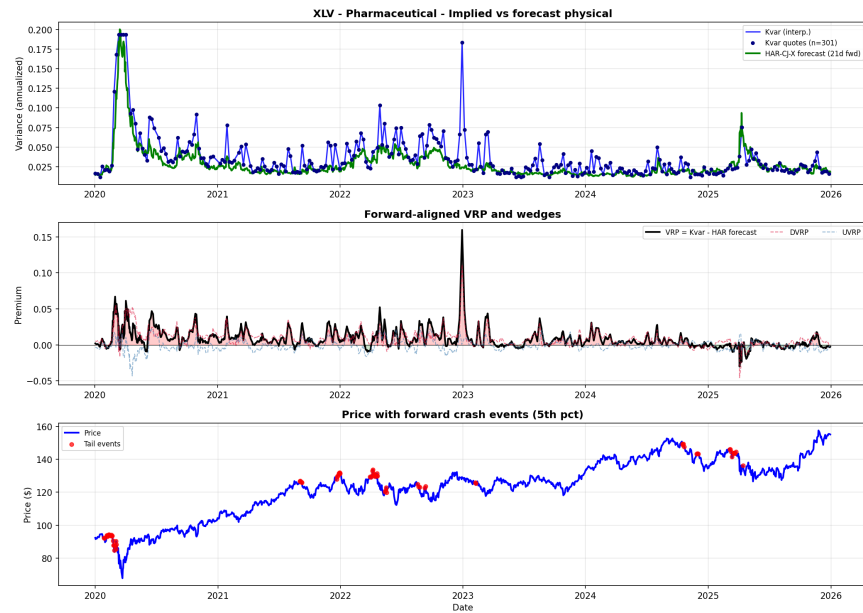


Figure 8.2: XLV Diagnostics

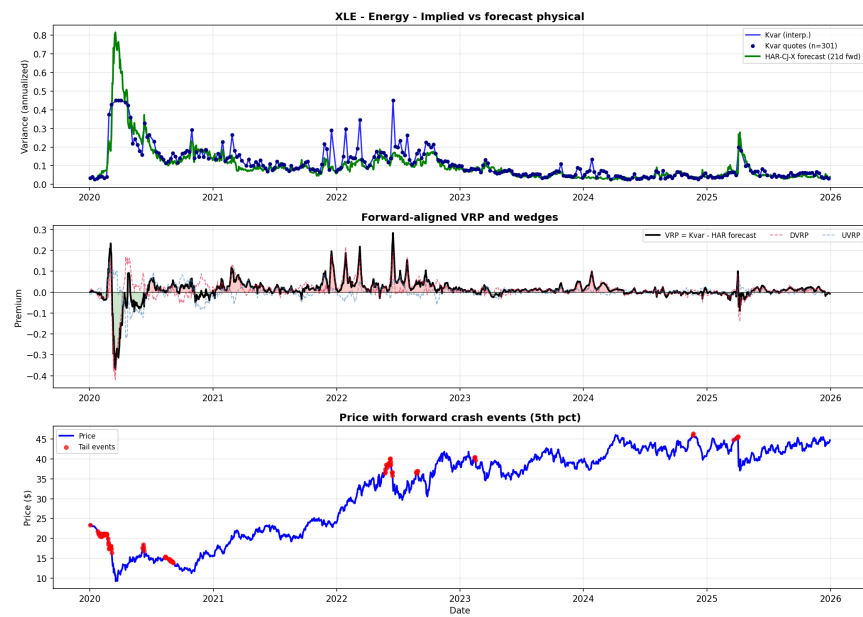


Figure 8.3: XLE Diagnostics

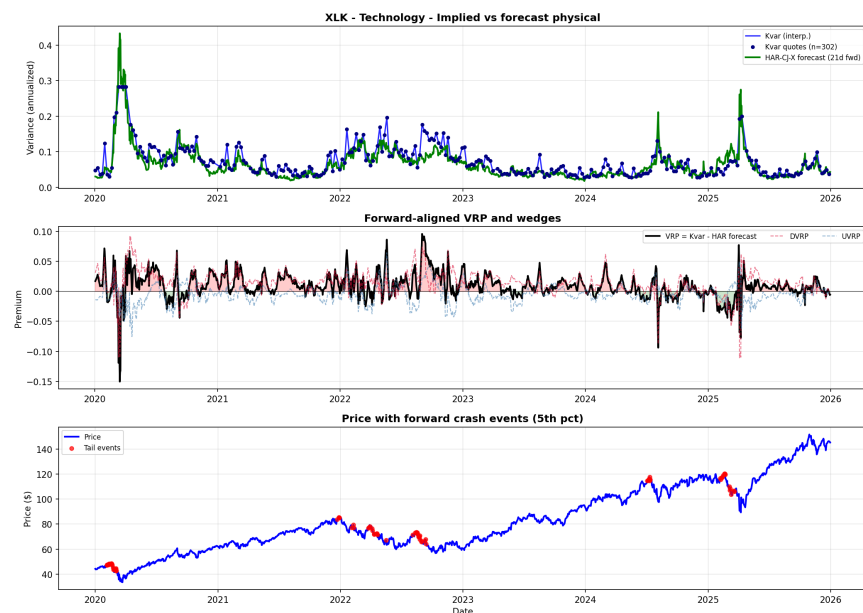


Figure 8.4: XLK Diagnostics

Table 8.1: Logit diagnostics for ITA (Defense): R^2 , Brier score, and AUC by model.

Model	R^2	Brier	AUC
VRP	0.0010	0.04744	0.5716
Forecasted_RV	0.0048	0.04741	0.5519
Historical_RV	0.0001	0.04747	0.5475
IV	0.0035	0.04740	0.5343
VRP_Ratio	0.0012	0.04745	0.5578
DVRP	0.0065	0.04723	0.5717
UVRP	0.0072	0.04724	0.5282
Ratio_DVRP_UVRP	0.0105	0.04714	0.5756
IV_Forecasted_RV	0.0048	0.04741	0.5542
IV_HistRV	0.0087	0.04727	0.5587

Notes. R^2 is McFadden pseudo- R^2 . Bold Brier = lowest; bold AUC = highest.

Table 8.2: Logit diagnostics for XLV (Pharmaceutical/Health Care): R^2 , Brier score, and AUC by model.

Model	R^2	Brier	AUC
VRP	0.0007	0.04731	0.4691
Forecasted_RV	0.0033	0.04730	0.6206
Historical_RV	0.0007	0.04732	0.4161
IV	0.0036	0.04725	0.5416
VRP_Ratio	0.0045	0.04722	0.5535
DVRP	0.0003	0.04731	0.4891
UVRP	0.0007	0.04731	0.4923
Ratio_DVRP_UVRP	0.0293	0.04657	0.6223
IV_Forecasted_RV	0.0038	0.04726	0.5631
IV_HistRV	0.0116	0.04708	0.5607

Notes. R^2 is McFadden pseudo- R^2 . Bold Brier = lowest; bold AUC = highest.

Table 8.3: Logit diagnostics for XLE (Energy): R^2 , Brier score, and AUC by model.

Model	R^2	Brier	AUC
VRP	0.0234	0.04572	0.5197
Forecasted_RV	0.0036	0.04672	0.6632
Historical_RV	0.0009	0.04672	0.4028
IV	0.0205	0.04625	0.5903
VRP_Ratio	0.0069	0.04644	0.5430
DVRP	0.0129	0.04642	0.5589
UVRP	0.0032	0.04659	0.4910
Ratio_DVRP_UVRP	0.1562	0.04028	0.7476
IV_Forecasted_RV	0.0296	0.04557	0.5893
IV_HistRV	0.0522	0.04463	0.6121

Notes. R^2 is McFadden pseudo- R^2 . Bold Brier = lowest; bold AUC = highest.

Table 8.4: Logit diagnostics for XLK (Technology): R^2 , Brier score, and AUC by model.

Model	R^2	Brier	AUC
VRP	0.0048	0.04719	0.6105
Forecasted_RV	0.0088	0.04732	0.6881
Historical_RV	0.0002	0.04732	0.3436
IV	0.0061	0.04722	0.5723
VRP_Ratio	0.0529	0.04550	0.6600
DVRP	0.0003	0.04731	0.5083
UVRP	0.0062	0.04719	0.5941
Ratio_DVRP_UVRP	0.1357	0.04394	0.7733
IV_Forecasted_RV	0.0088	0.04732	0.6941
IV_HistRV	0.0145	0.04695	0.5676

Notes. R^2 is McFadden pseudo- R^2 . Bold Brier = lowest; bold AUC = highest.

9 Conclusion

We have mathematically deconstructed VRP starting from the Heston, S. L. (1993) [2] stochastic volatility model, not to reach one final VRP formula to statistically test it against other indicators like RV^{hist} and K_{var} , but to gain deeper understanding of what it represents with regards to tail event prediction, which requires a holistic approach that cannot be encompassed in one single measure. Across ETFs results have varied significantly with some increasing the strength of our initial assumption that VRP contains more information compared to historical Realized Variance and model-free Implied Variance alone, and others proving that deconstructing VRP in VRP ratio, UVRP and DVRP is a more efficient way to predict tail events. Symmetric VRP, which is the difference between K_{var} and the forecasted Realized Variance, essentially compares Options richness against physical variance, thus yielding the premium investors pay for either fear or greed. The holistic part of our approach comes from the understanding that when a crash is coming both wings become richer or at least the put wing for protection, whereas when a rally is coming, the call wing in particular becomes richer because investors have bullish expectations. In fact, using VRP ratio, UVRP, DVRP instead of symmetric VRP is the best way to try and predict market direction; together these three indicators give a clearer picture, compared to symmetric VRP alone, as they provide both information on put wing fear and call wing richness vs their respective forecast of Realized Variance. In the Pharmaceutical and Defense sector the wedge data was definitely noisier, as shown by Figure (8.1) and Figure (8.2). In fact, the joint view with VRP ratio and the two asymmetric wedges (UVRP and DVRP) was certainly useful to get a clearer picture for the aforementioned sectors; in contrast in the energy and tech sectors we

have seen clearer and more significant proof of the superiority of the joint model of VRP ratio with UVRP and DVRP compared to the joint models of IV and historical or forecasted Realized Variance. An important consideration is the fact that for the Defense and Pharmaceutical sectors the forecasting model for Realized Variance had a weaker R^2 , therefore introducing bias when performing the logistic regression test. Therefore, with better forecasting techniques, probably the joint regression with VRP ratio, UVRP and DVRP would have produced more statistically significant results in proving the higher efficiency and accuracy in predicting tail events through an asymmetric VRP analysis. In conclusion, this paper supports the assumption that VRP and more specifically the asymmetric analysis of VRP with its two level wedges on put and call wings can provide more information to predict tail events compared to standard historical Realized Variance and model-free Implied Variance (or model-free implied volatility). In addition, the correct alignment between the two horizons of K_{var} and $\widehat{RV}^{fwd}$ is essential for a correct estimate of VRP along with the accuracy of the estimation of $\widehat{RV}^{fwd}$ itself.

Bibliography

- [1] Heston, S. L. (2018). Recovering the variance premium. University of Maryland (working paper covering the sample period 2007–2018).
- [2] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2), 327–343.
- [3] Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999). More than you ever wanted to know about volatility swaps. *Goldman Sachs Quantitative Strategies Research Notes*, March 1999.
- [4] Rompolis, L. S., & Travalis, E. (2010). Retrieving risk neutral moments and expected quadratic variation from option prices. Athens University of Economics and Business.
- [5] Corsi, F. (2004). A simple long memory model of realized volatility. Institute of Finance, University of Lugano (working paper, this version 18 August 2004; first version November 2002).