Copenhagen Business School
September 11, 2021

Variance risk premia estimation and stock return predictability

submitted by

Louis Galowich (134136)

Lukas Gejgus (111038)

supervised by

Johan Stax Jakobsen, PhD

Master's Thesis, MSc in Applied Economics and Finance

Pages: 94 (120)

Characters: 139,613 (165,451)

Abstract

In this paper we explore how to derive the variance risk premium which is a premium paid by investors to shield against large price swings. Formally, it is defined as the difference between the expectation of the risk-neutral and physical return variation and represents the payoff of a variance swap. We rely on the VIX volatility index and high frequency realized variance as proxies for risk-neutral and physical variance, respectively. An empirical challenge in the computation of the variance risk premium is modelling a forecast of the actual (or physical) expected volatility accounting for the distribution properties and stylized facts of the realized measure. To this end, we employ forecasting models based on a heterogenous autoregressive (HAR) framework to generate the one-month ahead estimate of the realized variance. Lastly, we show that the variance risk premium can be a predictor for stock returns at short horizons (i.e. less than one year). Although significant, the proportion of the variance in future returns explained by the variance risk premium is rather small in a univariate regression analysis. The degree of predictability is more impressive within a multivariate regression including other commonly employed predictor variables.

Keywords: Variance Risk Premium, VIX, Realized Variance, Heterogenous Autoregressive Models, Stock Return Predictability

Contents

List of Figures

1. Introduction

1.1 Overview

Hardly any other field of research in finance has received as much attention as financial markets return volatility. Research on volatility has had substantial implications for security valuation, risk management and investment. Volatility is fundamental to option pricing models, notably as an input variable in the famous Black-Scholes option pricing model. Compared to historical volatility, volatility implied by option prices is often believed to contain superior information about future volatility (Jiang and Tian, 2005). Afterall, given the forward-looking nature of options, its implied volatility should provide the market's expectation of future fluctuations of the underlying asset. Several early studies (e.g. Canina & Figlewski, 1993) conclude that option implied volatility is an important, although biased, element of the true expected future volatility. However, these studies often focused on a narrow range of at-the-money options and exclusively relied on Black-Scholes implied volatility. Implied volatility backed out from the standard Black-Scholes model is flawed as it imposes the counterfactual assumption that returns on the underlying asset have constant implied volatility. Efforts have been devoted to modeling time-varying volatility into option pricing models, but it has proved difficult to create an empirically sound procedure. Advances in this field were made by Britten-Jones and Neuberger (2000) who outlined a model-free volatility procedure obtained from the cross-section of actively traded options. Jiang and Tian (2005) revisit the forecasting abilities of implied volatility within a model-free specification and find that it is a more efficient estimate of future realized volatility than the Black-Scholes implied volatility. They further note that model-free implied volatility is on average larger than the realized volatility; an observation that constitutes the premise of this paper. The information content in model-free volatility has become a prominent research area and has had wide implications. The influential VIX volatility index from the

Chicago Options Exchange (CBOE) is based on a model-free volatility configuration. Carr and Wu (2006) show that the squared VIX is analogue to a variance swap rate which equals the risk-neutral expected return variance. Given that options allow investors to lock in expectations about future price developments, the VIX theoretically contains information about the actual expected volatility and a premium accounting for the variance from the expected volatility. By extension, it would be possible to extract any variance risk premium by subtracting the expectation of the actual (or physical) variance from the VIX. The variance risk premium can be viewed as the premium investors are willing to pay to shield against future variance.

1.2 Research Question

An empirical challenge in the computation of the variance risk premium is modelling a forecast of the actual expected volatility. The desirable theoretical properties of high frequency realized variance has resulted in it becoming a suitable proxy for actual contemporaneous volatility. While the computation of the variance risk premium is of interest on its own due to varying approaches in deriving both the risk-neutral and actual future variance, Bekaert and Hoerova (2014), among others, argue that the variance risk premium is a measure of aggregate risk aversion that could be a predictor for future stock return.

The combination of multiple financial and economic fields of research provides both a challenging and interesting framework. Throughout this paper we review the latest developments in the research on variance risk premia and aim to provide novel insights on contested assumptions. More specifically, we aim to answer the following research questions:

How can the variance risk premium of the S&P 500 be empirically estimated?

Is the variance risk premium a predictor of future stock returns?

To this end, we contrast the performance of different realized variance forecasting models against a simple non-estimation technique and test the stock return predictability in the presence of other predictor variables.

1.3 Limitations

There are a number of initial limitations we need to impose to set the framework of this paper.

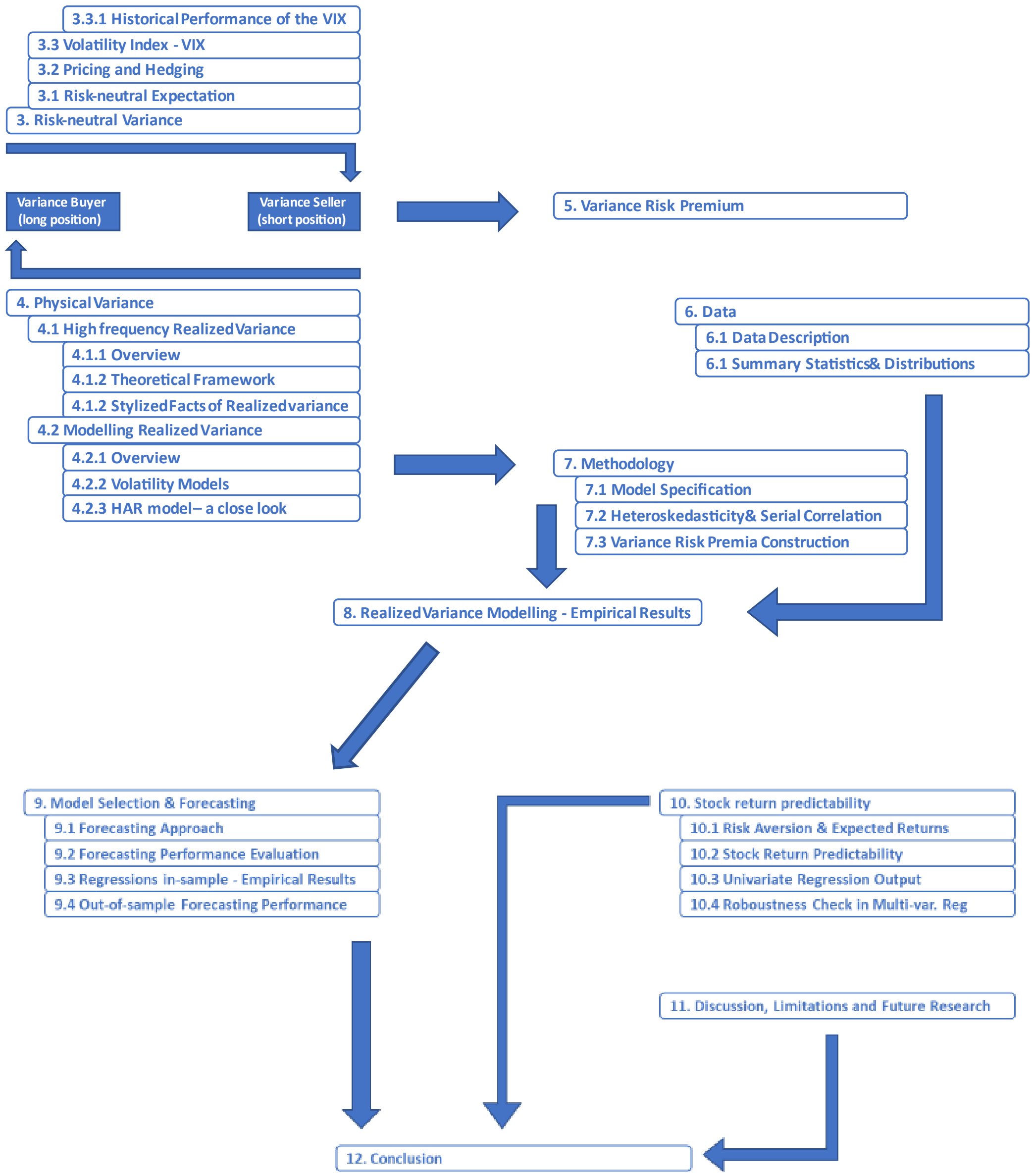
In our analysis section we are going to exclusively focus on the U.S stock market. More specifically, we analyze the S&P 500 which is often perceived as a proxy for the overall U.S equity market. We have decided to restrict our analysis on the U.S market due to the abundance of available data allowing a straightforward replication of our analysis. Furthermore, the majority of research on variance risk premia has been performed with U.S data providing a comparable benchmark with our results.

Another limitation is imposed in regard to the selection of the realized variance forecasting techniques. The research on forecasting models is too vast to provide a comprehensive comparison on all available techniques. Instead, we will compare the most relevant techniques and provide a qualitive argumentation in favor of the selected models. The selected forecasting models are then benchmarked in a quantitative performance analysis.

Inevitably, this paper will build on a lot of theory and thus all relevant concepts that are material to this paper will be explained in detail. However, it may at times be assumed that the reader is familiar with some common concepts which will not be derived or explained in full. However, at all times are relevant academic sources included should additional material be sought by the reader.

This paper is structured as follows. In section (2) we show the theoretical overlaps between variance swaps and the variance risk premium. Section (3) and (4) introduce the two key components of the variance risk premium namely, the risk-neutral and the physical variance, respectively. Section (5) ties the theory sections together and builds the path for the empirical analysis. After presenting the data in section (6), we dive into the methodology (7) for the forecasting models and ultimately the variance risk premium calculation. In section (8) we report the regression outputs of the models estimated using full samples. Section (9) presents the procedures for out-of-sample forecasting and model selection. In section (10) the capacity of the derived variance risk premia to predict stock returns is tested. The limitations of our study are discussed in section (11) and we conclude this paper in the section (12).

Figure 1: Thesis outline and flow



Notes: The figure outlines the thesis section flow and how the section are interconnected

2. Variance Swaps

2.1 Definition

A variance swap is a derivative contract that is traded over the counter, meaning the terms of the contract are privately established between two or more counterparties outside of a public exchange. It pays the difference between the annualized variance of the underlying security and the annualized variance strike which is agreed upon at the inception of the trade (De Weert, 2008, p. 139). This class of contracts first appeared in the mid-1990s as a way to facilitate direct variance trading (Bossu, 2014). Prior to that, options were primarily used as vehicles to execute bets on volatility. However, option vega[1] and gamma[2] exposures are concentrated around the strike price and tend to diminish as the price of the underlying moves away from it. The resulting local characteristic of option volatility exposure may lead to cases when a trader is right on volatility but still losses money on the strategy (De Weert, 2008). Variance swaps were introduced to tackle this issue.

The variance swap payoff (VSP) for the buyer is defined as:

$$VSP = (\sigma_R^2 - \sigma_K^2)N \tag{1}$$

where σ_R is the annualized realized volatility for the duration of the contract, σ_R^2 is the annualized realized variance for this duration, σ_K^2 is the annualized variance strike and N is the variance notional. Each contract must precisely describe the method for calculating the realized volatility. Generally, it is specified as the square root of the sum of the squared log-returns multiplied by an annualization factor. In other words, annualized realized volatility is a population standard deviation under a zero-mean assumption[3]

[1] Vega of an option measures sensitivity of the option´s premium to the changes of implied volatility.
[2] Gamma captures the sensitivity of the option´s delta to changes in stock price of the underlying.
[3] Mean is not subtracted from the observed returns which is equivalent to assuming the known mean return is 0. This is a fairly common practice when structuring volatility derivatives.

(Derman & Miller, 2016). Mathematically, σ_R for N daily log-returns on the underlying S between time $t = 0$ and $t = T$ is calculated as (Bossu, 2014):

$$\sigma_R = \sqrt{\frac{252}{N}\sum_{i=0}^{N-1}\left(\ln\frac{S_{i+1}}{S_i}\right)^2} \tag{2}$$

Since traders are more comfortable thinking in terms of volatility rather than variance, the notional is commonly first stated as the vega amount with the variance notional derived from it. The variance notional, also referred to as variance units, is specified through the vega notional as:

$$N = \frac{Vega\ notional}{2\sigma_K} \tag{3}$$

It represents a dollar amount that converts the variance difference into a dollar payoff. The vega notional is a term commonly used in volatility swap trading. It is equal to the payoff of a volatility swap if the difference between the realized volatility and the strike is one volatility point.

To avoid the need for upfront cash exchange, a regular variance swap has a net market value of zero at the inception of the trade. Thereby, when the contract is constructed the variance strike, also known as the variance swap rate, which ensures the expected payoff of zero is chosen (De Weert, 2008). According to no arbitrage reasoning, the variance swap rate must be equal to risk neutral expected value[1] of the realized variance (Carr & Wu, 2009, p.5).

$$\sigma_K^2 = E_t^Q\left[\sigma_{R,t,T}^2\right] \tag{4}$$

The theory on variance swaps marks the outset for this paper and will provide the guiding theme throughout the theory section. More specifically, in the theory section we explain and derive the risk-neutral and physical variance, the two core components of equation (1). We start by examining risk-neutral expectations and explaining its economic intuition, followed by the theoretical formulation of the risk-neutral variance. We conclude the first section by introducing a widely applied proxy for risk neutral variance namely, the VIX. The section on physical variance is kicked off by examining the concept of high frequency realized variance and related stylized facts. Next, we explore realized variance estimation models and build the path for the analysis section.

3. Risk-neutral Variance

We continue our theory section by introducing the risk-neutral variance and outlining its theoretical formulation. Ultimately, this section is concluded by discussing a widely applied proxy for risk-neutral variance namely, the VIX volatility index. Again, the risk-neutral variance is one of two core components of equation (1).

3.1 Risk-neutral Expectation

An investor is risk-neutral when she is indifferent between a certain payment and a risky investment with an equal expected payoff. Such an investor only considers a price of the asset to be fair if the expected discounted price of the asset at some future date equals the current price. A price process satisfying this condition is called a martingale (Hilpisch, 2015, p. 54).

Formally, martingales are processes "whose future variations have no specific direction based on the process history." (Knopf & Teall, 2015, p.128). A filtration F adapted stochastic[4] process is a martingale under probability measure Q if it fulfills:

$$For\ all\ t, s \geq 0, t + s \leq T: E_t^Q[S_{t+s}] = S_t \tag{5}$$

A probability measure is a function that contains information regarding the probability of the observable event's occurrence and satisfies the measure properties below.

$$1. \forall E \in \mathrm{F} : P(E) \geq 0 \tag{6}$$

$$2. P\left(U_{i=1}^{I} E_i\right) = \sum_{i=1}^{I} P(E_i)\ for\ disjoint\ sets\ E_1, E_2, \dots E_I \in \mathrm{F} \tag{7}$$

[4] Stochastic process is a family of random variables indexed by some mathematical set. It can alternatively be thought of as a probability distribution over a space of paths.

$$3. P(\Omega) = 1 \quad (8)$$

where E denotes a set in the algebra F which is a family of sets in the state space Ω representing all possible states ω.

It is important to note that the risk-neutral probability is conceptually very different to the real world physical probability. For the purposes of risk-neutral pricing, probabilities of various state paths are implied from the prices of traded securities, payoffs of which depend on these paths. As the market actors are on aggregate risk-averse, they tend to overweight probabilities of bad states when pricing securities (Hatfield, 2009). This phenomenon of distinct bad state probability assignment between risk-neutral and physical worlds is of high relevance to the design of this study as will be outlined in the following sections.

3.2 Pricing and Hedging

Variance swaps can be replicated and therefore also priced with a static portfolio of put and call options with equal maturity which are subsequently delta-hedged. Appendix 1 provides a derivation of the fair value of the realized annualized variance by utilizing the relationship between variance swaps and a log contract. A log contract is an exotic derivative first introduced by Neuberger (1994). It pays the natural logarithm of the ratio between the underlying's terminal price and its forward price (Bossu , 2014). Neuberger (1994, p.78) claims that; "A log contract can be delta-hedged without making any forecast of volatility, and the hedged position is a pure and simple volatility play."

In short, it can be shown that the realized variance can be replicated by continuously maintaining a position of $\frac{2}{S_t}$ in the underlying[5] in addition to holding a certain quantity of log contracts (Bossu, 2014 p.63). A log contract is not a traded instrument however, its payoff can be constructed by shorting a

[5] Where S_t is the spot price of the underlying at time t.

forward contract on S struck at the forward price F, being long all long puts with $K < F$ and holding all long calls struck at $K > F$ both in quantities $\frac{dK}{K^2}$.

The fair value of the annualized variance is then:

$$\sigma_K^2 = \frac{2}{T} E\left(\ln\frac{S_T}{F}\right) = \frac{2e^{rT}}{T}\left[\int_0^F \frac{1}{K^2} p(K)dK + \int_F^\infty \frac{1}{K^2} c(K)dK\right] \quad (9)$$

where r is the interest rate corresponding to maturity T, $p(K)$ is the price of the put struck at K and $c(K)$ is the price of the call struck at K.

As in the real world, only a finite number of strikes are available, thus the following proxy formula is needed:

$$\sigma_K^2 \approx \frac{2e^{rT}}{T}\left[\sum_{i=1}^{n} \frac{p(K_i)}{K_i^2}\Delta K_i + \sum_{i=n+1}^{n+m} \frac{c(K_i)}{K_i^2}\Delta K_i\right] \quad (10)$$

Multiple studies (e.g. Carr & Wu, 2009; Driessen et al, 2009) have pointed to significantly negative average payoffs on synthetized swaps that have the realized variance of U.S. stock market indices as the underlying. In other words, data implies there is a premium demanded from a party seeking to take the long position in the variance swap. The long position is held by the buyer of the contract who receives the payment when the annualized realized variance over the period ends up being higher than the variance strike. Following sections describe what factors are thought to cause the presence of this premium according to the existing literature.

3.3 Volatility Index - VIX

The VIX is a non-tradable volatility index established by the Chicago Board Options Exchange (CBOE). It is computed from the price cross-section of S&P 500 index option contracts and measures the 30-day (i.e.

30 calendar days and 22 trading days) ahead expected volatility of the S&P 500 (Whaley, 2009). Given that S&P 500 options are among the most actively traded derivatives and the S&P 500 itself is viewed as a proxy of the United States (US) stock market, the VIX is perceived as a sound measure of the US market's expectation of future volatility (Sarwar, 2012). Edwards and Preston (2017, p.2) note that the VIX can be thought of as "a crowd-sourced estimate for the degree to which the market is uncertain about the future". Hence, the VIX is often labelled the investor fear gauge since a high VIX level alludes to high uncertainty in the U.S stock market. The fear gauge is based on a "model-free" calculation since it is constructed from a multitude of actively traded out-of-the-money S&P 500 put and call options across different strike prices and is thus not extracted from a model, unlike the Black-Scholes implied volatility (BSIV) (Whaley, 2009). The VIX has emerged as a substitute for the renowned BSIV since the model's assumptions of constant volatility and lognormal return distribution are not supported by actual market data (Figlewski, 2016). Empirical studies (e.g. Carr and Wu, 2009; Duan and Yeh, 2010) suggest that the VIX approximates the 30-day ahead variance swap rate of the S&P 500 which equates to the risk-neutral expected value of the return variance of the S&P 500, as already addressed in previous sections. For that reason, the VIX has been widely used in academic research as a proxy for the risk-neutral expected market variance, most notably in applications related to stock return predictability and market volatility forecasting (among others, Bollerslev et al., 2009; Bekaert and Herova, 2014; Giot, 2005).

The general formula for the VIX as provided by CBOE (2019):

$$\sigma^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2 \tag{11}$$

where, F is the unique strike price for which the absolute difference between the put and call prices are minimized and K_0 represents the strike price nearest to F, with F being strictly larger than K_0. K_i and ΔK_i

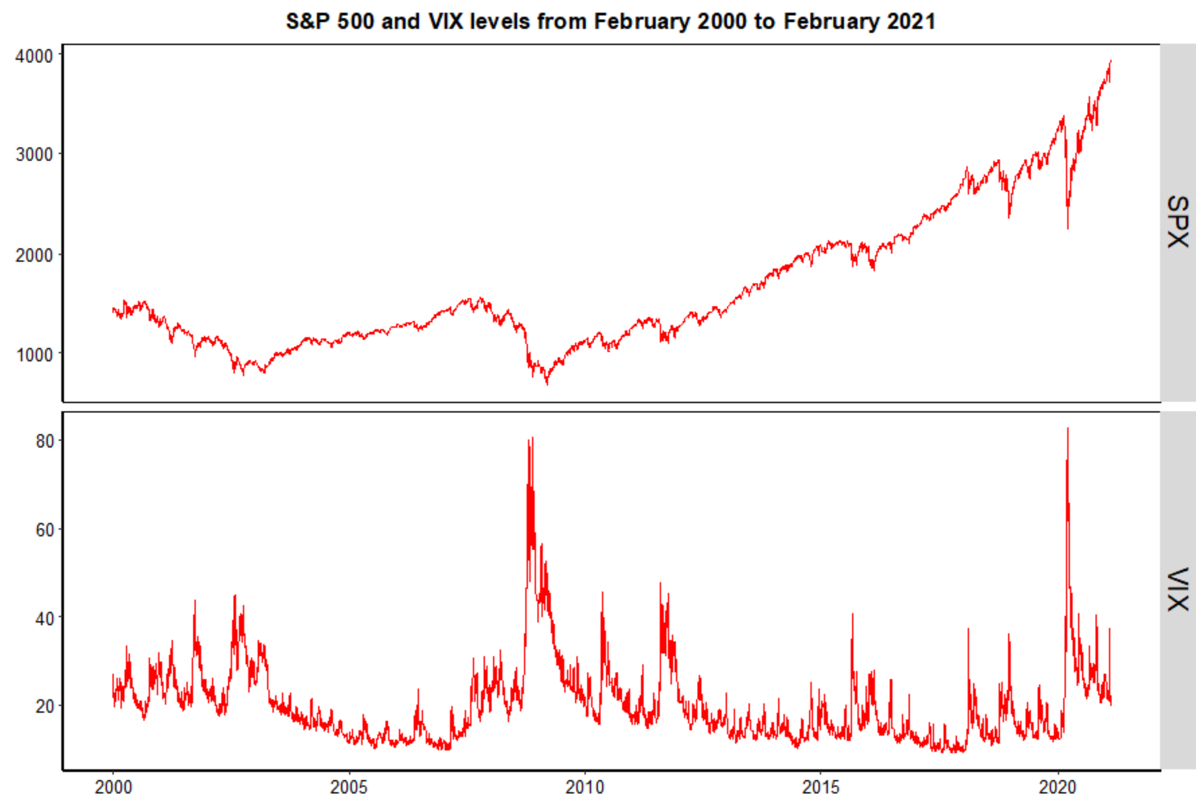
represent the strike price of the ith out-of-the-money contract and the average interval between two strike prices, respectively. R entails the risk-free interest rate and is generally the t-bond yield closest to the expiration date of the contracts. $Q(K_i)$ provides the average bid-ask spread for every option and T are the minutes to expiration.

Near- (strictly more than 23 days) and next-term (strictly less than 37 days) call and put options across a wide range of strike prices serve as components for the formula above. Once a week, contracts are discarded and new ones with different maturities for both near- and next-term options are used. The near- and next-term contracts are then weighted to create the VIX. The number of options and strike prices considered to calculate the VIX change dynamically based on the liquidity of option contracts at various strike prices (CBOE, 2019). Specifically, options with zero bid prices and the two contracts preceding a no bid contract are excluded. These readjustments allow to extract the timeliest market information and reflect the market's expected volatility on a real-time basis. Taking the option price cross-section across a wide range of strike prices allows to extract the risk-neutral density which reflects investors beliefs about all the volatility properties of the underlying asset including potential risk premium (Figlewski, 2016). Naturally, the actual future volatility may not exactly correspond to the forward-looking VIX, however it is precisely the discrepancy between the risk-neutral expectations and the physical expectations that has attracted a wide range of research and is also of interest in the present paper.

3.3.1 Historical Performance of the VIX

It can be observed that the S&P 500 and the VIX movements are noticeably negatively correlated. The graph below shows that the relationship between S&P 500 and contemporaneous VIX levels (Figure 2). It can be observed the VIX and S&P 500 move in opposite directions. Level changes in the VIX are the result of greater relative changes in the option prices for puts than for calls or vice versa. That is, negative (positive) shocks turn investors bearish (bullish) and hence put option prices soar (fall) relative to call option prices, ultimately leading to an increase (decrease) in the VIX (Low, 2004). Thus, the index presents an informative metric to assess the relationship of the markets risk perception and the market conditions (Low, 2004).

Figure 2: Negative correlation between S&P 500 and VIX



Notes: The figure plots daily levels of Standard and Poor´s 500 index (top) and the VIX volatility index (bottom) between the 3rd of January 2000 and 12th of February 2021.

That expected volatility and price move together is hardly news, in his seminal paper, Sharpe (1964) predicted that as expected volatility increases (decreases), all else equal, investors require a greater (lower) rate of return and thus stock prices fall (rise). Correspondingly, the VIX shows extreme spikes during high market turmoil accompanied with considerable changes of the S&P 500 levels, for example during the financial crisis and the beginning of the global Covid-19 pandemic. It is, however, noteworthy that the relation between the S&P 500 and the VIX is not one-to-one, as might be reasoned by Sharpe (1964), but rather displays an asymmetric behavior. Figure 2 reveals that a drop in the S&P 500 causes a larger change, in absolute terms, in the VIX than a rise of the S&P 500 of equal magnitude would. Again, the asymmetric behavior between market volatility and stock prices has sparked an area of research on its own, dating back to, for example, Schwert (1990). In relation to the VIX, the most plausible explanation for this observation relates to the use of options as insurance tools. Over recent years, the S&P 500 option market has become dominated by portfolio hedgers who buy put options to shield their portfolios from downside risk during market turmoil (Whaley, 2009). The demand for portfolio insurance (i.e. puts) strongly influences the VIX since it serves as an input variable for its calculation. Hence, VIX levels reflect the market downside risk fears rather than investor excitement (Whaley, 2009). In other words, during bear markets investors are concerned about the future and prefer to insure their portfolios leading to spikes in the VIX, while they have a lower incentive to hedge during bull markets resulting in much more subtle VIX movements. By extension, since the VIX is derived from readily available options market information it provides an overall market sentiment indicator (Fassas & Papadamou, 2018).

Despite its popularity, the accuracy of the VIX as a risk-neutral measure has been contested, particularly regarding the construction of the VIX and the influence of market structure noise as well as sudden extreme jumps (Jiang & Tian 2005; Andersen et al., 2015; Carr and Wu, 2006). In theory, the fair value of the return variance considers continuous strikes prices as illustrated by equation (9) however, the VIX necessarily relies on discrete strike prices which may results in discretization errors. Furthermore, equation (9) considers option strike prices ranging from F to infinity, this cannot be translated into the VIX measure especially since deep out-of-the-money options often have no bids. While some have argued that these errors can introduce substantial bias others have advocated that the VIX remains a good approximation of the risk-neutral expectation of total return variation (e.g. Bollerselv et al., 2009).

Nonetheless, the computation of the VIX underlies a strong theoretical foundation with no easily accessible alternative. In the present paper, the VIX will serve as a measure for the risk-neutral expected market volatility to ensure the comparability of findings with other papers.

4. Physical Variance

The second component of equation (1) that we examine in greater detail is the physical variance. The section on physical variance is kicked off by examining the concept of high frequency realized variance and its related stylized facts. Next, we explore realized variance estimation models and build the path for the analysis section.

4.1 High frequency Realized Variance

4.1.1 Overview

Broadly speaking, volatility measures are practical tools to assess the market's uncertainty surrounding a security's price (Drechsler and Yaron, 2011). Return volatility measures come in many different shapes and forms. The most common measure of return variance is based on the squared daily returns:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(r_i)^2 \tag{12}$$

However, early papers (e.g. Jorion 1995) have noted that the squared daily returns performed poorly in out-of-sample forecasting. Initially, these deficiencies were believed to be inherent to the forecasting model employed (e.g. GARCH). Andersen and Bollerslev (1998) refute these accusations, arguing that the problem rather relates to the squared inter-day returns being a relatively noisy proxy of the conditional return variance. Advances in data collection and its increased accessibility have fueled the use of high-frequency data in financial research, resolving some of the main constraints of low frequency data. Hansen and Lunde (2011) argue that the added benefits of high frequency data are directly related to finer sampling leading to increased precision in the dynamic properties of the variable which, ultimately allows for better forecasting capabilities. They further underline that high frequency data has sparked the development of new volatility models that show potential for superior forecasting performance over other established forecasting models; a discussion that will be further addressed in section (4.2). The most common substitute for inter-day return volatility is realized volatility, generally computed as the sum of intra-day squared log returns.

4.1.2 Theoretical Framework

To define the realized variance and understand its properties it is useful to review its computation. Consider a time interval $[a, b]$ which includes a time index t_i such that $t_i \in [a, b]$. The time interval is divided into a total of τ equidistant partitions. In other words, the interval between each t_i is of equal length for all $i = 1,2, \dots \tau$. It follows that the sampling frequency is given by $m = \frac{a-b}{\tau-1}$. For each t_i, one price observation is made and denoted $[p_{ti}]_{i=1}^{\tau}$. The logarithmic returns are then expressed as $r_{ti} = \log(p_{ti}) - \log(p_{ti-1})$. The realized variance over the interval $[a, b]$ can then be formulated as,

$$RV_{[a,b]} = \sum_{i=1}^{\tau} r_{ti}^2 \qquad (13)$$

The theoretical underpinnings of realized variance are of interest since as the interval of the sampled returns becomes infinitesimal small, $\tau \to \infty$, the sum converges in probability towards the quadratic variation (QV) of the process (Degiannakis, 2015). In turn, for all stochastic volatility models the QV provides the integrated volatility (IV) (Hansen and Lunde, 2011). The integrated volatility (IV) is a measure of the conditional variance, σ_t^2. Both the IV and the population volatility are not directly observable and need to be obtained by an accurate observable proxy. Again, under the assumption made above, the realized variance is, at least in theory, a proxy for the conditional variance and therefore often used in research as an input for variance forecasting (Hansen and Lunde. 2011). Inevitably, the sampling frequency cannot converge towards infinity since as the sampling frequency is further compressed substantial biases are introduced leading to deviations from the latent efficient price (Hansen and Lunde.

2011). The noises are often linked to bid-ask bounce and incorrect recordings among others. As such, the realized variance is, at best, only a good approximation of the true conditional variance but tends to outperform in precision the coarser inter-day squared returns (Bollerselv et al., 2009, Andersen et al., 2001). In practice, 5-minute intervals have emerged as the gold standard in econometric applications as it provides a trade-off between accuracy and microstructure market noises (McAleer and Medeiros, 2008). Liu et al. (2015) empirically test different time intervals and conclude that 5-minute intervals capture much of the favorable benefits of high-frequency data while smaller sampling intervals provide only marginal improvements. As a result, the present paper defines realized variance as the sum of 5-minute squared logarithmic returns which will be retrieved from the Realized Library provided by the Oxford-Man Institute of Quantitative Finance.

4.1.3 Stylized Facts of Realized Variance

Multiple studies (Bollerslev, et al., 2018; Andersen et al., 2001) have found that realized variance distributions are generally extremely right-skewed and leptokurtic. In other words, when the realized variance is considered in its level form, the third and fourth moments of the probability distributions are significantly larger than normal. However, as pointed out by Andersen et al. (2001), the logarithmic distributions of realized variances approximately resemble a normal distribution.

Financial time series generally showcase highly persistent volatility dynamics meaning that the impact of volatility shocks in the series volatility autocorrelation function dissipates over a very long period of time (Andersen & Bollerslev, 2018). Numerous studies focusing on the high-frequency stock return and currency exchange data (e.g. Andersen & Bollerslev, 1997; Andersen, Bollerslev, Diebold & Labys, 2001; Andersen et al., 2001) have proposed that the long-range dependence is best described by slow hyperbolic rate of decay. In their recent paper, Bollerslev et al. (2018) analyze the realized volatility characteristics for 58 assets across 4 major asset classes and conclude that the general dynamic patterns and decay rates in autocorrelations for daily realized volatilities averaged across the different asset classes are very similar and all display long-range dependence. A multitude of approaches were introduced to achieve the goal of approximating the long-memory process. These span from the models utilizing fractional integration exemplified by FI-GARCH (Baillie, Bollerslev and Mikkelsen, 1996) or ARFRIMA to models using the superposition of short-memory frequencies like Heterogeneous Autoregressive model (HAR) (Bauwens, Hafner, Laurant, 2012). The models most pivotal for our study are introduced and described in the section (4.2).

Leverage effect refers to one of the most prominent regularities in the financial time series. It has been observed that negative stock return innovations increase the stock return volatility more than the positive ones (Smith, 2015). The inverse relationship between the stock return and (future) volatility has been so enduring across multiple studies that it is now considered to be one of the stylized facts of stock return data (Zumbach, 2013). The phenomenon was first described by Black (1976) who theorized that this dynamic can be explained by the "direct causation" effect of financial leverage, hence the name leverage effect. A stock price decrease generally implies a drop in company´s equity value leading to increase in the leverage as the level of debt in the capital structure remains fixed in the short term. The heightened financial leverage in turn leads to higher equity return volatility. Black (1976) also provided an alternative explanation by introducing the "reverse causation" effect which describes a causal relationship from volatility changes to stock returns (Hasanhodzic & Lo, 2011). Increase in the expected future volatility must lead to the drop in the stock price in order to increase the stock´s expected return and thereby induce investors to continue holding it. Following this line of thought led to the development of an extensive separate branch of literature focusing on time-varying risk premia.

Black's leverage hypothesis has been empirically tested numerous times over the years, with multiple authors (Christie, 1982; Ericsson et al. 2016) finding evidence in its favor while many others (French, Schwert, and Stambaugh 1987; Gallant et al., 1992; Duffee, 1995) have provided evidence against it. Hasanhodzic and Lo (2011) argue that leverage effect cannot be explained by financial leverage alone as they observed volatility asymmetry in stock returns when studying the sample containing only all equity financed firms. The validity of Hasanhodzic and Lo's (2011) conclusions may be questioned on the merit that the authors did not take into account the presence of operating leverage[6]. More recently, Smith (2015) used a methodology revolving around a EGARCH model to study the sources of the leverage effect. His results were consistent with the predictions of Black's leverage effect hypothesis (Smith, 2015).

Due to the open-endedness of the debate regarding the leverage hypothesis validity, financial literature has suggested alternative causes of leverage effects. Avramov et al. (2006) link the asymmetry in stock returns and volatility relationship to information asymmetry between informed and uniformed traders. Authors argue that liquidity driven (herding) trades executed by uninformed traders following price declines drive the volatility up whereas the contrarian trades made by the informed market actors reduce the volatility following stock price increases (Avramov, 2006). Dennis et al. (2006) distinguish between systematic and idiosyncratic volatility which allows them to differentiate between the impact of market-wide factors and firm-level factors explaining the asymmetric volatility phenomenon. The authors identify a broader set of market-level factors that cause the leverage effect besides the firm's financial leverage.

[6] Operating leverage is defined as fixed costs that remain constant over a short term (Sanusi, 2017).

4.2 Modelling Realized Variance

4.2.1 Overview

Volatility modeling has received a lot of attention in financial research which has led to the emergence of numerous models of which the most relevant ones will be explained and analyzed in this section. A volatility model aims to generate a forecast of future volatility. In section (5) we will explain in detail the reasons we require volatility forecast for the computation of the variance risk premium. Prior to that, we will discuss different volatility models and focus in particular on direct forecasting models with the realized variance as a dependent variable. Importantly, the stylized facts described in the previous section play a crucial role in selecting a fitting volatility model.

4.2.2 Volatility Models

The appropriate model selection needs to be assessed in light of the inherent properties of the data dynamics. As discussed in section (4.1.3) the challenge of modelling realized variance depends on accurately capturing its stylized facts.

Arguably one of the most influential volatility models for conditional volatility are the autoregressive conditionally heteroskedastic class models first introduced by Engle (1982). The two most popular models are the ARCH and GARCH models. As opposed to the traditional ARCH model, GARCH includes, in addition to lagged past squared innovations, ε_{t-i}, also lagged conditional variance terms, h_{t-i},:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_j h_{t-i} \tag{14}$$

The rationale to expand the ARCH model was to reduce the rapid decay of its autocorrelation function to improve the empirical fit to financial variance time series since volatility is generally highly persistent across time (Andersen et al. 2009). While similar properties can be achieved through a high-lag ARCH model, the more parsimonious GARCH model is often preferred in application. Although the autocorrelation function of GARCH models decrease at a slower rate compared to their ARCH counterparts, they continue to decrease exponentially with the lag length, even with a high β value. Engle and Bollerslev (1986) observed that if $\alpha + \beta = 1$, the GARCH contains a unit root which implies that past shocks influence the conditional variance indefinitely. This model is labeled Integrated GARCH or IGARCH for short. The indefinite persistence of the IGARCH is however not empirically sound since volatility processes are generally mean reverting. The contrasting difference in variance persistence between the GARCH and IGARCH model has led to the emergence of a more flexible alternative namely, the Fractionally Integrated GARCH (FIGARCH). Baille et al. (1996) introduced the FIGARCH model which displays hyperbolic decay of its autocorrelation function and is thus compatible with the time dependence observed in financial market volatility.

Empirical evidence suggests that forecasting models based on high frequency measures outperform standard GARCH models which are generally based on daily returns to model the latent volatility (e.g. Çelik et al., 2014). In other words, models that directly model realized variance tend to showcase superior performance (Andersen et al., 2003). We therefore explore other venues to directly model realized variance in order to take full advantage of the information content in high frequency realized variance data.

A model that allows to directly model realized variance is the ARFIMA model which has been widely applied to capture the long-range dependencies of realized volatility (Degiannakis & Floros, 2013). A stationary ARFIMA process of order (p, d, q), with p and q being non-negative integers and $d \in [0,0.5]$, may be expressed as:

$$\alpha(L)(1-L)^d(x_t - \mu) = \beta(L)\varepsilon_t \tag{15}$$

where μ represents the mean of x_t[7], ε_t is independently identically distributed white noise, L are backshift operators such that $\alpha(L)$ and $\beta(L)$ are defined as $1 - \sum_{j=1}^{p} \alpha_j L^j$ and $1 + \sum_{j=1}^{q} \beta_j L^j$, respectively. From (15) it is apparent that the long-term dependence is introduced via the fractional difference filter $(1-L)^d$. The exponent d, generally defined as the long memory parameter, dictates the dependence (i.e. short or long). As such, if $d = 0$ or $d = 1$ the ARFIMA model reduces to a ARMA and ARIMA model, respectively. Long persistence results from a fractional parameter (e.g. $d = 0.4$) applied to the left side of equation (15). (Brockwell & Davis, 2016)

Hosking (1981) shows that under the condition $d \in [0,0.5)$, the autocorrelation of the ARFIMA model is approximately k^{2d-1} as k tends to infinity. The function implies that the autocorrelation process decays at a hyperbolic rate with the length of the lags (i.e. k) which is in line with the stylized fact of realized variance data. Empirical studies show that ARFIMA models tend to outperform standard GARCH models (Franke et al., 2019).

A comparatively recent model is the Mixed Data Sampling (MIDAS) model which has been developed primarily to predict volatility (Andreau et al., 2011). In the context of realized variance (RV) the MIDAS model may be defined as:

$$RV_{t+h,t} = \mu + \phi \sum_{k=0}^{k^{max}} \omega(k,\theta) X_{t-k} + \varepsilon_t \tag{16}$$

Unlike the ARFIMA model, MIDAS regressions do not necessarily rely on autoregressive components, such that the lagged independent variable, X_{t-k}, on the right does not necessarily represent lagged value of the regressand. For example, the independent variables included in MIDAS should be appropriate

[7] x_t can be the realized variance or its logarithmic transformation.

predicators of quadratic variation such as daily squared returns, realized power or daily range. The weight functions are selected based on a parameterized function such as a linear step function. (Ghysels & Valkanov, 2012)

Important characteristics of MIDAS are that 1) the model provides the flexibility to analyze the performance of the individual regressors as they can be directly compared against each other to determine if one regressor dominates another 2) and the inclusion of data sampled at varying frequencies, all within a unified framework (Ghysels & Valkanov, 2012). The MIDAS model draws parallels with heterogenous autoregressive (HAR) regressions, introduced by Corsi (2004). As such, Forsberg and Ghysels (2007) build a MIDAS regression that considers discrete steps of partial sum high frequency variables which have been inspired by Corsi's HAR model (Foroni & Marcellino, 2013). They conclude that MIDAS and HAR show very similar performance for realized variance modelling.

The HAR model is a parsimonious and simple to estimate model that is analogue to long memory models. The HAR-RV model developed by Corsi (2004) adds distinct AR models into a single model to capture the long-range dependence of realized variance. The computational details are outlined in the following section. A key benefit of the HAR model is that it underlies a strong economic intuition. To understand this intuition, it is worthwhile to outline its theoretical underpinnings rooted in the heterogeneous market hypothesis.

Müller et al. (1997) utilized a fractal approach to derive the theory of heterogeneous markets. The concept of fractals, first introduced by Mandelbrot (1983), revolves around analyzing objects on different scales or degrees of resolution and then comparing the results (Müller et al, 1997). One can think of this process as using yardsticks of different sizes to measure the length of a coastline. The length of the seashore is then a function of the yardstick size. Applying this reasoning to time series, different time yardsticks represented by various data sampling frequencies are used to analyze the object.

The rationale to consider distinct sampling frequencies is based on the premise that distinct agents in heterogeneous markets have different time horizons and trading frequencies. Müller et al. (1997) argue that the different transaction frequencies reflect different reactions to the same market news. In his view, "The market is heterogeneous, with a fractal structure of the participants' time horizons as it consists of short-term, medium-term and long-term components." (Müller, 1997, p. 12). This has consequences for the structure of market volatility memory as it is composed of multiple exponential declines with different time constants.

Contrary to homogeneous markets where an increased number of market actors leads to more rapid convergence of prices to real market values on which all rational investors agree, an increased number of participants in heterogeneous market results in distinct market actors settling for different prices in different situations driving up volatility. The authors motivate this claim by pointing to empirically observed positive correlation between volatility and market presence. In their study, the lowest volatility as well as market presence was observed during the weekends while both variables were the highest during early European afternoons which coincide with North American mornings.

4.2.3 Heterogenous Autoregressive Model – A close look at the HAR model

As outlined above, market participants create different volatility movements such that they can be distinctively split to create an 'additive cascade of partial volatilities' (Corsi 2004, p.9). Corsi's HAR model considers three horizons, daily (d), weekly (w) and monthly (m) which corresponds to 1, 5 and 22 working days, respectively. Formally, the models for each of the three different time horizons are:

$$\tilde{\sigma}_{t+1m}^{m} = c^{m} + \alpha^{m} RV_{t}^{m} + \tilde{\varepsilon}_{t+1m}^{m} \quad (17)$$

$$\tilde{\sigma}^{w}_{t+1w} = c^{w} + \alpha^{w} RV^{w}_{t} + \gamma^{w} E_{t}[\tilde{\sigma}^{m}_{t+1m}] + \tilde{\varepsilon}^{w}_{t+1w} \tag{18}$$

$$\tilde{\sigma}^{d}_{t+1d} = c^{d} + \alpha^{d} RV^{d}_{t} + \gamma^{d} E_{t}[\tilde{\sigma}^{w}_{t+1w}] + \tilde{\varepsilon}^{d}_{t+1d} \tag{19}$$

The daily and weekly volatilities are a function of their past realized variance and the expected value of the longer horizon partial volatility. The monthly partial variance translates to a simple first order autoregressive model. It is apparent that the three equations can be reduced to a single model through substitution of the larger time horizons into the daily variance.

$$\sigma^{d}_{t+1d} = c + \beta^{d} RV^{d}_{t} + \beta^{w} RV^{w}_{t} + \beta^{m} RV^{m}_{t} + \tilde{\varepsilon}^{d}_{t+1d} \tag{20}$$

The daily latent volatility needs to be related to the forecast of the realized volatility.

$$\sigma^{d}_{t+1d} = RV^{d}_{t+1} + \varepsilon^{d}_{t+1d} \tag{21}$$

As addressed in section (4.1.2), the realized variance is only an approximate measure of the latent volatility hence it is necessary to account for potential errors which are included in the error term, ε. Substituting equation (21) into (20) it can be rewritten as

$$RV^{d}_{t+1} = c + \beta^{d} RV^{d}_{t} + \beta^{w} RV^{w}_{t} + \beta^{m} RV^{m}_{t} + \varepsilon_{t+1d} \tag{22}$$

where, $\varepsilon_{t+1d} = \tilde{\varepsilon}^{d}_{t+1d} - \varepsilon^{d}_{t+1d}$

The realized variances for each partial volatility model are obtained as follows,

$$RV^{h}_{t} = \frac{1}{h} \sum_{i=i}^{h} RV_{t+i} \tag{23}$$

with $h = 1, 5, 22$ representing the daily, weekly and monthly time scale, respectively. For each time horizon the realized variance is then the average of the realized variance series.

The HAR model as outlined above is at its core an autoregressive (AR) model with 22 lags onto which restrictions are imposed. Audrino and Knaus (2016) show that the HAR model can be reformulated as

$$RV_{t+1}^{d} = a + \sum_{i=1}^{22} \phi_i^{HAR} RV_{t-i+1}^{d} + \varepsilon_{t+1} \tag{24}$$

with restrictions,

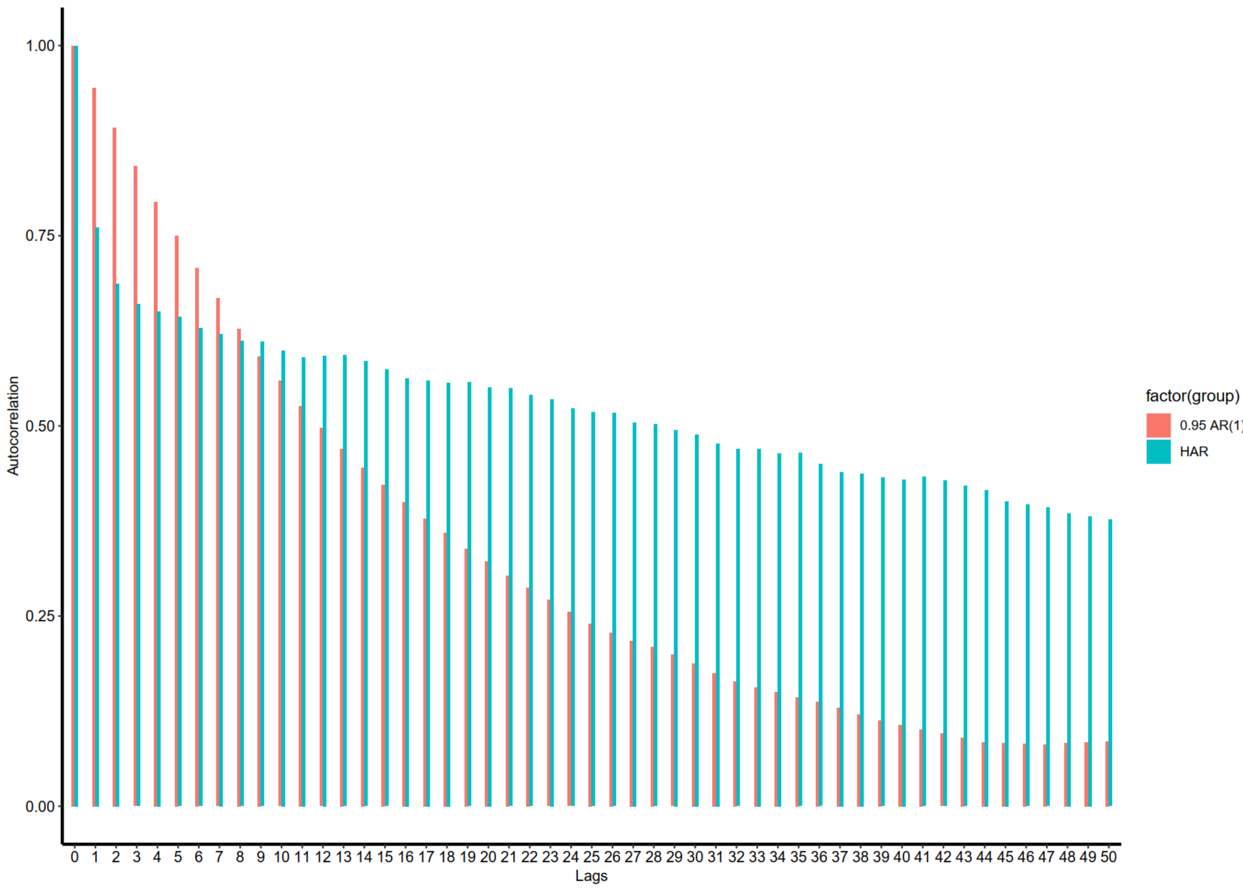
$$\phi_i^{HAR} = \begin{cases} b^d + \frac{1}{5}b^w + \frac{1}{22}b^m \; for \; i = 1 \\ \frac{1}{5}b^w + \frac{1}{22}b^m \; for \; i = 2, \dots, 5 \\ \quad \frac{1}{22}b^m \;\; for \; i = 6, \dots, 22 \end{cases} \tag{25}$$

Effectively, the HAR fits a partial autocorrelation function into a linear combination of realized variances that can be estimated through a simple ordinary least squares (OLS) regression (Bauwens 2012).

A key benefit of the cascade composition of the HAR model is that, unlike a GARCH (1,1) model, it distributes the weight over distinct observations that contain varying information about future volatility developments. For example, for a large unexpected spike in volatility, a GARCH (1,1) model will assign a lot of weight on the last observation which may corrupt the next forecast while the HAR model spreads the weight across distinct horizons (Reschenhofer et al., 2020).

The HAR model follows a similar rationale as outlined by LeBaron (2001) who showed that a linear combination of three autoregressive (AR) functions appears to model the long-term persistence observed with financial volatility, while a single AR function can match these properties. To underline the persistence observed in the HAR model, we simulate an AR and HAR function and contrast their autocorrelation function. Figure (3) illustrates the striking difference in autocorrelation decay between the two models. As discussed, the simple AR(1) model decays exponentially even with a high coefficient (i.e. 0.95) while the HAR[8] model effectively models persistence in the data.

Figure 3: Modelling autocorrelation decay – AR(1) vs. HAR



Notes: The figure plots the autocorrelation functions of a simulated autoregressive process of order 1 (red) and a Heterogenous Autoregressive process (blue) with coefficients 0.35, 0.3, and 0.3 for the daily, weekly, and monthly observations, respectively. The processes are simulated based on random normally distributed observations.

[8] For illustrative purposes we calibrate the coefficients in the HAR model to 0.35, 0.3 and 0.3 for the daily, weekly and monthly coefficients, respectively.

5. Variance Risk Premium

Now that all the relevant concepts have been introduced, we can frame the bigger picture and explain how these concepts are interrelated before moving on to the empirical analysis.

High variance represents a price risk to an investor's holdings and is often accompanied by market crashes, which is generally unfavorable to investors. In turn, it would be reasonable to expect that investors are willing to pay a price (or premium) to shield their positions against large price swings. The premium may be thought of as volatility insurance where premiums are high when there is a lot of volatility and low (or zero) when the market is steady. In other words, the premium investors pay should be time-varying and depends on expected market volatility. Hence, during turbulent times, investors going long on a variance swap will bear a premium to protect themselves against large future variance.

To analyze this hypothesis the VIX and realized variance we discussed in the previous sections will play a central role. The difference between the expectation of the risk-neutral and physical return variation provides the variance risk premium and represents the payoff of a variance swap (equation 1) (Anderson et al., 2001):

$$VRP_t = E_t^{\mathbb{Q}}(Var_{t,t+1}) - E_t^{\mathbb{P}}(Var_{t,t+1}) \tag{26}$$

These two measures are not observable and need to be replaced by proxies. As already outlined in previous sections the VIX and realized variance are appropriate proxies for the risk-neutral and physical return variance, respectively. Formally, the variance risk premium can then be defined as:

$$VRP_t = {VIX_t}^2 - E_t(RV_{t+1}) \tag{27}$$

where VIX is observable at time t and $E_t(RV_{t+1})$ is the estimate of the one-step ahead realized variance. The corresponding estimation model for the realized variance forecast will be introduced in section (7.1).

The value of the variance risk premium (i.e. the difference of equation (27)) is generally positive since the VIX includes the insurance characteristics of options which market participants seek during turbulent times. Note that some academic articles write the variance risk premium in the inverse order and thus report a negative variance risk premium. Given that the variance risk premium is mostly negative, we report it as in equation (27) to facilitate readability.

Other papers (e.g. Bollerselv et al., 2009) analyzing the variance risk premium have proceeded with the assumption that the realized variance follows a martingale diffusion process such that the current realized variance subsumes all information about its future realization, simplifying to $E(RV_{t+1}) = RV_t$:

$$VRP_t = {VIX_t}^2 - RV_t \tag{28}$$

Clearly, the benefit of this approach is that all variables are observable at time t such that no estimation model for the realized variance is required. Bekaert and Hoerova (2014) oppose the latter approach, pointing out that the data does not support such assumptions and could bias the calculations of the variance risk premium. For the purpose of comparability between different findings and to test the relative performance, the present paper will extract the variance risk premium based on both approaches. As addressed in section (4.2.3) we rely on the HAR models and variations thereof to model the future realized variance. The exact configurations of the models are described in the Methodology section.

6. Data

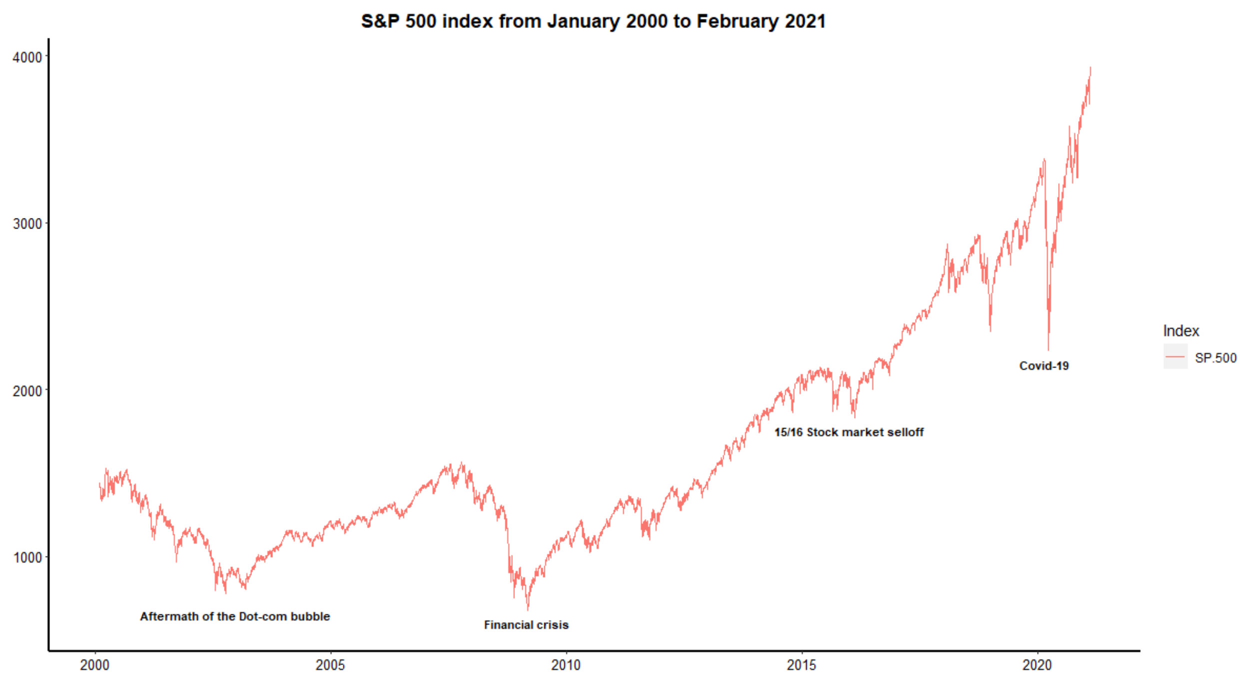
6.1 Data Description

In this section the variables comprising our dataset are described and preliminarily analyzed. Our core dataset contains daily closing spot prices of the S&P 500 index which we use to calculate daily percentage and log-returns. In addition, the data also includes daily values of the S&P 500 realized variance computed based on observations conducted with 5-minute frequency as well as historical adjusted daily closing levels of the CBOE Volatility index VIX. The stock index and realized variance data were sourced from the Realized Library of the Oxford-Man Institute of Quantitative Finance while the CBOE Volatility index (VIX) figures were retrieved from the Yahoo Finance website.

VIX levels are reported in annualized volatility terms. As we are interested in modelling monthly realized variance and variance risk premia, the VIX variable must be adjusted. Following Dreschler & Yaron (2009) we square the VIX figures to transform them into the variance space and then divide the product by 12 to obtain a monthly quantity. Realized variance data is on the other hand reported as the daily sum of the squared 5-minute returns. In order to make this variable comparable to our constructed monthly squared VIX quantity, we multiply the observations by 10^4 to convert them into squared percentages (Dreschler & Yaron, 2009). The monthly realized variance for a particular date is then obtained by summing the last 22 of these squared percentages. Mathematically it is expressed as,

$$RV_t^m = \sum_{i=1}^{22} RV_{t-22+i}^d \tag{29}$$

Figure 4:S&P 500 index development

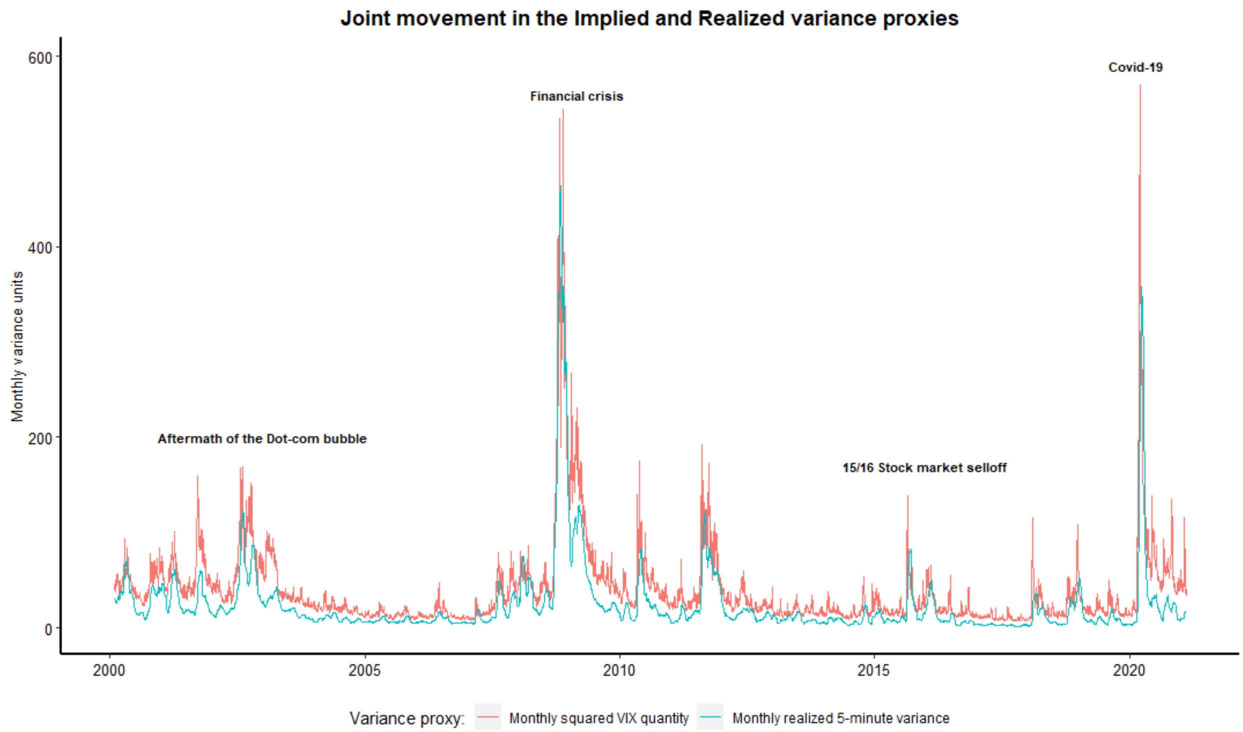


Notes: The figure plots daily levels of Standard and Poor´s 500 cash index between the 3rd of January 2000 and 12th of February 2021. Periods of major index declines are highlighted with annotations underneath the line.

Overall, the dataset covers the period from the 3rd of January 2000 until the 12th of February 2021 which translates to 5,293 daily observations. Neither VIX nor realized variance data are reported during the weekends or public holidays. Moreover, the realized variance summation procedure of equation (29) leads to exclusion of the first 21 observations which leaves us with a sample size of 5,272 observations for each observed variable.

As can be observed in the Figure 5 the sample period for the most part consists of expansions with steadily rising or recovering stock market. However, multiple recessions and depressions are also observed during the studied timespan. These include for example the aftermath of the Dot-com bubble in early 2000s, the financial crisis of 2008 and the recent 2020 stock market crash triggered by the Covid-19 pandemic.

Figure 5: Historical movement of implied and realized variance proxies



Notes: Simultaneous movement in the two variance proxies with major increases during the annotated recessions. Monthly squared VIX quantity is derived by squaring the annualized VIX index and then dividing the product by 12. Monthly realized 5-minute variance is obtained by summing 22 consecutive daily realized variances based on 5-minute observations which have been multiplied by 10^4.

As already outlined in the section (3.3.1), variance and returns tend to move in opposite directions. Figure 5 corroborates that this is the case for our sample as well, with realized variance hovering at low levels during the bull-markets and soaring in the recessions. Realized variance peaked at unprecedented levels during the 2008 financial crisis which caused the liquidity in volatility derivatives to dry up (Bekaert & Hoerova, 2013). Expectedly, the monthly squared VIX index, an indicator which is considered to be the risk-neutral expectation of the realized variance moves in tandem with the monthly 5-minute realized variance as evidenced by correlation of 0.88 between the two variables[9].

[9] Strength of the relationship between the two proxies is also graphically displayed by Figure A.1 in Appendix 2.

6.2 Summary Statistics and Distributional Properties

Table 1 provides the basic descriptive statistics for the studied variables. The summary provides information about the first four moments of the variables probability distributions (i.e. mean, standard deviation, skewness and excess kurtosis). In addition, number of observations, value of the middle observation, average distance between each data value and a mean as well as minimum and maximum are also reported. Since we are interested in modelling monthly realized variance, most emphasis is put on the description of variance variables at the monthly scale, however, we summarize the daily variance and return variables for the sake of completeness.

Table 1: Summary statistics

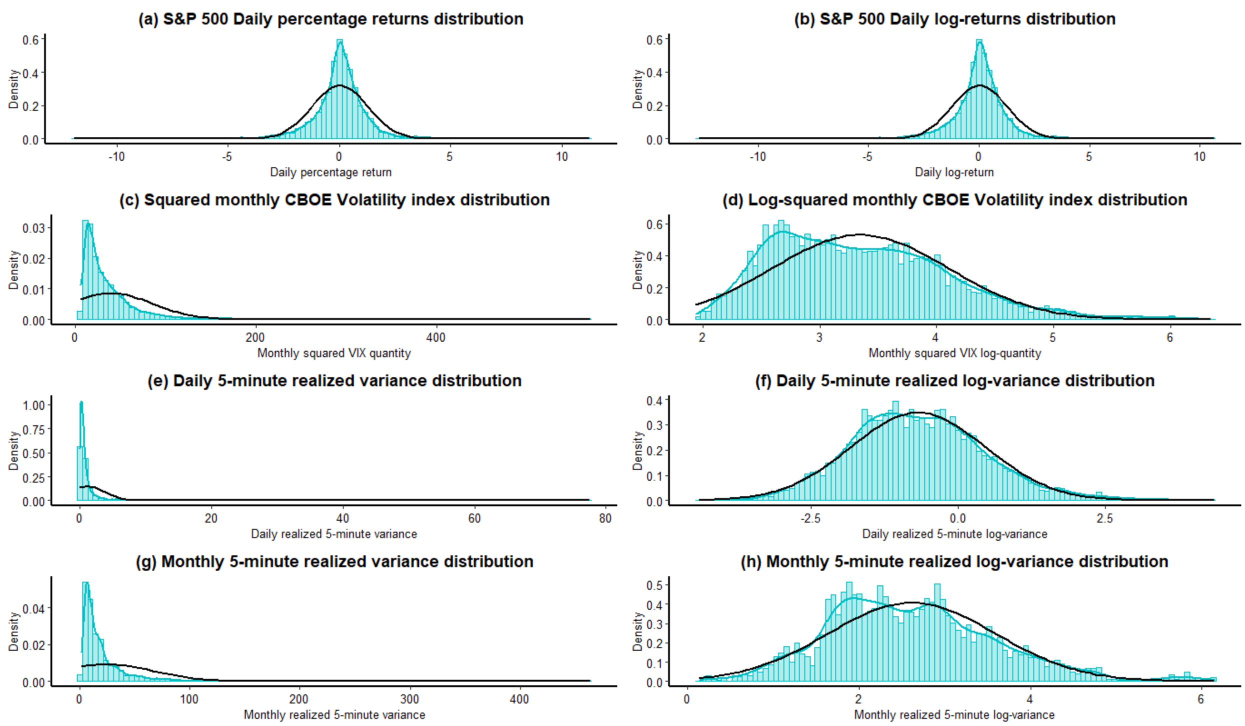
	Returns		Variance					
	Daily % return	Daily log-return	VIX^2 monthly	RV 5 daily	RV 5 monthly	ln(VIX^2 monthly)	ln(RV 5 daily)	ln(RV 5 monthly)
N	5,272	5,272	5,272	5,272	5,272	5,272	5,272	5,272
Mean	0.027	0.019	39.841	1.111	24.468	3.347	-0.678	2.616
Standard deviation	1.244	1.246	46.591	2.649	42.963	0.754	1.149	0.980
Median	0.062	0.062	25.975	0.474	12.612	3.257	-0.747	2.535
Mean absolute deviation	0.773	0.771	18.394	0.451	10.289	0.807	1.118	0.958
Minimum	-11.9	-12.67	6.962	0.012	1.141	1.940	-4.406	0.132
Maximum	11.229	10.642	569.803	77.477	464.186	6.345	4.350	6.140
Skewness	-0.153	-0.387	4.743	10.773	5.83	0.703	0.383	0.557
Kurtosis	10.508	10.705	32.253	193.503	41.99	0.372	0.395	0.490

Notes: Summary statistics of return and variance variables for the sample ranging from 3rd of January 2000 to 12th of February 2021.

Figure 6 (a) displays the probability distribution of daily S&P 500 percentage returns which has a mean of approximately 0.03 and standard deviation of 1.24. The exhibited distribution is left skewed as indicated by its negative third moment. The distribution also has a sharper peak and heavier tails relative to the normal distribution. Information regarding the tail extremity of a probability distribution function can be conveyed through its fourth moment known as the kurtosis. The kurtosis statistic reported in Table 1 is an excess kurtosis of the distribution which is defined as its fourth moment minus the value of 3. Excess kurtosis of 10.51 implies a leptokurtic distribution with relatively fatter tails thus confirming the eye test and conforming to stylized facts of financial time series (Zumbach, 2013). Distributional properties of the

S&P 500 log-returns whose probability distribution is shown in Figure 6 (b) are overall very similar to its perecentage counterpart with even more negative skewness and more extreme excess kurtosis.

Figure 6: Distributional properties of return and variance variables



Notes: Histograms, probability density functions and gaussian curves for the comparison in black to outline the distributional properties of daily returns as well as daily and monthly variance proxies in both percentage/level and log. Level variance charts 6(c), (e) and (g) showcase extremely right-skewed distributions while corresponding log charts 6(d), (f), (h) plot distributions that are much better approximated by the Gaussian curve.

Figure 6 (c) plots the probability distribution of squared monthly CBOE Volatility index (VIX). The histogram is right skewed with most of the density concentrated to the left of the mean and with significantly drawn-out right tail. In terms of the third moment, the distribution has a positive skewness of 4.74. As for the tail extremity, considerable mass of the observations laying more than three standard deviations to the left of the mean, results in an excess kurtosis of 32.25. Thereby the curve represents a leptokurtic distribution with fatter tails relative to the normal distribution. The variable constructed as natural logarithm of squared monthly CBOE Volatility index has a bimodal probability distribution displayed in Figure 6 (d) with a mean of 3.35 and standard deviation of 0.75. The distribution's third and fourth moments are closer to normality compared to the level variable with skewness and excess kurtosis much closer to 0.

The probability density function of the daily realized variance is illustrated graphically in Figure 6 (e). The distribution has the expected value of 1.11 and standard deviation of 2.65. A skewness of 10.77 implies a significant right skew and prolonged right tail. Major outliers observed during the above-mentioned recessions contribute to gargantuan excess kurtosis of 193.5. The observation for the sample maximum of 77.48 from 10th of October 2008 provides an illustrative example of the extremity of the outliers. This value is located more than 29 standard deviations away from the sample average meaning its occurrence would be exceedingly improbable and close to impossible had the probability distribution for the variable been normal. The logarithmic distribution of the daily realized variance which is almost Gaussian can be found in Figure 6 (f). The last summarized variables are monthly realized variance and its natural logarithm represented by Figures 6 (g) and 6 (h), respectively. Relative to its daily counterpart, the probability distribution of monthly realized variance is less skewed and less tail-heavy. The distribution for the logarithm is again multimodal but the third and fourth moments are nearly Gaussian.

7. Methodology

7.1 Model Specification

Following the methodology of Bekaert and Hoerova (2014) we consider monthly realized variance of the S&P 500 returns as the one-step ahead estimate of the realized variance that is used to derive the variance risk premium (VRP). The basic model HAR-RV is specified as:

$$RV_{t+22}^{m} = c + \beta^{d} RV_{t}^{d} + \beta^{w} RV_{t}^{w} + \beta^{m} RV_{t}^{m} + \varepsilon_{t+22d} \tag{30}$$

The regressand RV_{t+22}^{m} is defined as the sum of daily realized variances over the 22 trading days. Since the dependent variable RV_{t+22}^{m} is in fact just a future value of RV_{t}^{m} we lose the last 22 rows of the data-frame to the estimation procedure. This leaves us with 5,250 observations that we use throughout the analysis. Mathematically;

$$RV_{t+22}^{m} = \sum_{i=1}^{22} RV_{t+22-i+1}^{d} \tag{31}$$

We allow for the intercept c in the functional form. As has already been explained in detail in previous sections, the first three independent variables represent the squares of "the past realized volatilities viewed at different frequencies" (Corsi, 2004, p.10). We decided to pattern our model upon the tradition observed in the literature and chose the usual daily (h=1), weekly (h=5) and monthly (h=22) time scales for the first three volatility components in the regression. Considering that our model forecasts monthly variance, we express all realized variance variables in monthly units as follows:

$$RV_{t}^{(h)} = \frac{22}{h} \sum_{j=1}^{h} RV_{t-j+1} \tag{32}$$

To illustrate, the weekly realized variance expressed in monthly units at time t is constructed by summing daily realized variances observed from time $t-4$ until (including) time t and then multiplying the sum by $\frac{22}{5}$. Correspondingly, the daily realized variance expressed in monthly units is obtained by substituting 1 for h in equation (32). As for the monthly quantities, no further transformation in addition to initial variance summation procedure covered in section (6.1) is needed.

Corsi et al. (2012) extend the heterogeneous framework to include leverage effects as well. The authors were inspired by multiple studies (Corsi et al., 2005; Scharth & Medeiros, 2009; Fernandes et. al., 2009) which argued that the impact of negative price innovations on future volatility levels exhibits long-range dependence resembling the persistence in volatility itself. The inclusion of lagged negative returns with varying frequencies was thereby logically proposed as a way of modelling this dynamic feature. We mimic this procedure and supplement our base HAR-RV model with three additional factors representing the impact of daily (h=1), weekly (h=5) and monthly (h=22) negative returns on the level of the dependent variable estimate.

Our heterogeneous autoregressive model augmented with leverage effects components is labelled L-HAR-RV and is specified as follows:

$$RV_{t+22}^{m} = c + \beta^{d} RV_{t}^{d} + \beta^{w} RV_{t}^{w} + \beta^{m} RV_{t}^{m} + \delta^{d} r_{t}^{d-} + \delta^{w} r_{t}^{w-} + \delta^{m} r_{t}^{m-} + \varepsilon_{t+22d} \qquad (33)$$

The variables for modelling of leverage effects at different frequencies are constructed as:

$$r_{t}^{h-} = \min\left(r_{t}^{h}, 0\right) \qquad (34)$$

where we start with the construction of the vector containing daily leverage effects and then apply the same methodology for conversion of daily and weekly leverage effects to monthly units. Mathematically;

$$r_t^{(h)} = \frac{22}{h}\sum_{j=1}^{h} r_{t-j+1} \tag{35}$$

The last of our models estimated in levels is labelled VIX-L-HAR-RV. As the name suggests the model is an extension of the L-HAR-RV with the addition of squared monthly VIX quantity as an extra independent variable. The motivation to include monthly quantity of CBOE´s volatility index stems from VIX's definition of market's expectation of future volatility. Some authors have argued that implied volatility contains information about the future realized volatility levels (Christensen and Prabhala, 1998). In addition, the VIX itself has been employed as a predictor of realized volatility in multiple studies (Bekaert and Hoerova, 2014; Busch, Christensen and Nielsen, 2011). Nevertheless, due to the presumed existence of a variance risk premium it is generally acknowledged that the VIX is not an unbiased predictor of realized volatility.

$$RV_{t+22}^m = c + \alpha VIX_t^2 + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + \delta^d r_t^{d-} + \delta^w r_t^{w-} + \delta^m r_t^{m-} + \varepsilon_{t+22d} \tag{36}$$

Recalling Figures 6 (d), (f) and (h) which showed that log variances in our sample have distributions that are much better approximated by the Gaussian curve, we also estimate all three models in Log-Log form. We again follow the example of Bekaert and Hoerova (2014) and take the natural logarithms of every variable in the regressions outside of the ones representing leverage effects which are already measured in logarithm. Finally, the models log HAR-RV, log L-HAR-RV and log VIX-L-HAR-RV are specified as:

$$\ln(RV_{t+22}^m) = c + \beta^d \ln(RV_t^d) + \beta^w \ln(RV_t^w) + \beta^m (RV_t^m) + \varepsilon_{t+22d} \tag{37}$$

$$\ln(RV_{t+22}^m) = c + \beta^d \ln(RV_t^d) + \beta^w \ln(RV_t^w) + \beta^m \ln(RV_t^m) + \delta^d r_t^{d-} + \delta^w r_t^{w-} + \delta^m r_t^{m-} + \varepsilon_{t+22d} \tag{38}$$

$$\begin{aligned}\ln(RV_{t+22}^m) = c + \alpha \ln(VIX_t^2) + \beta^d \ln(RV_t^d) + \beta^w \ln(RV_t^w) \\ + \beta^m \ln(RV_t^m) + \delta^d r_t^{d-} + \delta^w r_t^{w-} + \delta^m r_t^{m-} + \varepsilon_{t+22d}\end{aligned} \tag{39}$$

For the models to be comparable, the estimates must be converted to the same scale. However, Jensen´s inequality dictates that:

$$E[f(RV_{t+22}^m)] \neq f(E[RV_{t+22}^m]) \quad (40)$$

unless the function f is linear. As the function applied in our application is concave, we obtain a strict inequality $E[f(RV_{t+22}^m)] < f(E[RV_{t+22}^m])$ barring degenerate cases. This forces us to take into consideration the distributional aspects of $\ln(RV_{t+22}^m)$ in order to get an unbiased estimate of RV_{t+22}^m when converting logarithmic estimates to level. Some authors (Bekaert and Hoerova, 2014) have tackled this challenge by assuming that the predicted variable is log-normally distributed and applying the corollary (Poulsen, 2010):

$$If\ \ln(RV_{t+22}^m) \sim N\left(\mu', \sigma'^2\right)\ then\ E(RV_{t+22}^m) = e^{\mu' + \frac{\sigma'^2}{2}} \quad (41)$$

Others (Jakobsen, 2018) solved the logarithm to level transformation problem through a simulation procedure where the empirical distribution of estimated residuals obtained by re-sampling is employed.

Although it is fair to question the appropriateness of the log-normality assumption, due to the computational constraints, we have decided to lean on log-normality assumption and transform the estimates of monthly logarithmic variance to levels as follows:

$$\mathrm{E}\,[RV_{t+22}^m] = e^{(\mathrm{E}\,[rv_{t+22}^m] + \frac{1}{2}var(rv_{t+22}^m)} \quad (42)$$

where,

$$rv_{t+22}^m = \ln(\widehat{RV_{t+22}^m}) \quad (43)$$

The logarithmic model is used to calculate the conditional expectation of monthly logarithmic variance estimates rv_{t+22}^m and the variance term is computed as the sample variance of rv_{t+22}^m.

7.2 Addressing Heteroscedasticity and Serial Correlation of Residuals

While the OLS estimator is still consistent and unbiased in the presence of the heteroskedasticity, the existence of non-constant variance of the error term brings serious problems for the standard inference procedures and efficiency of the OLS (Greene, 2014). To investigate whether the error terms in our models are heteroscedastic we conduct a Breush-Pagan test. Breusch and Pagan (1979) formulated a Lagrange multiplier (LM) test of the null hypothesis of homoscedasticity where the hypotheses can be formally defined as follows:

$$H_0 : Var(u_i|X) = \sigma^2 \quad for\ all\ i \tag{44}$$

$$H_1 : Var(u_i|X) = {\sigma_i}^2 \tag{45}$$

We perform the studentized version of the Breusch-Pagan test recommended by Koenker (1981) that addresses the original test's sensitivity to the normality assumption by incorporating a more robust estimator of the error variance. The residuals extracted from each of the six estimated models are squared and then regressed on the matrices of the explanatory variables from the corresponding models. The ratio of the explained sum of squares and total sum of squares of this auxiliary regression is then used to calculate the LM statistic which in case of normality and under the null hypothesis of homoskedasticity follows a limiting chi-squared distribution with the degrees of freedom dictated by the number of independent variables in the particular model. There has been evidence that in the absence of normality this modified statistic provides a more powerful test relative to the base version (Greene, 2014).

$$LM_{BP} = nR_h^2 \tag{46}$$

Where n is the number of observations and R_h^2 is the coefficient of determination for the auxiliary regression. The test results reported in the Table 2 provide strong evidence against the null of homoscedasticity in the error term.

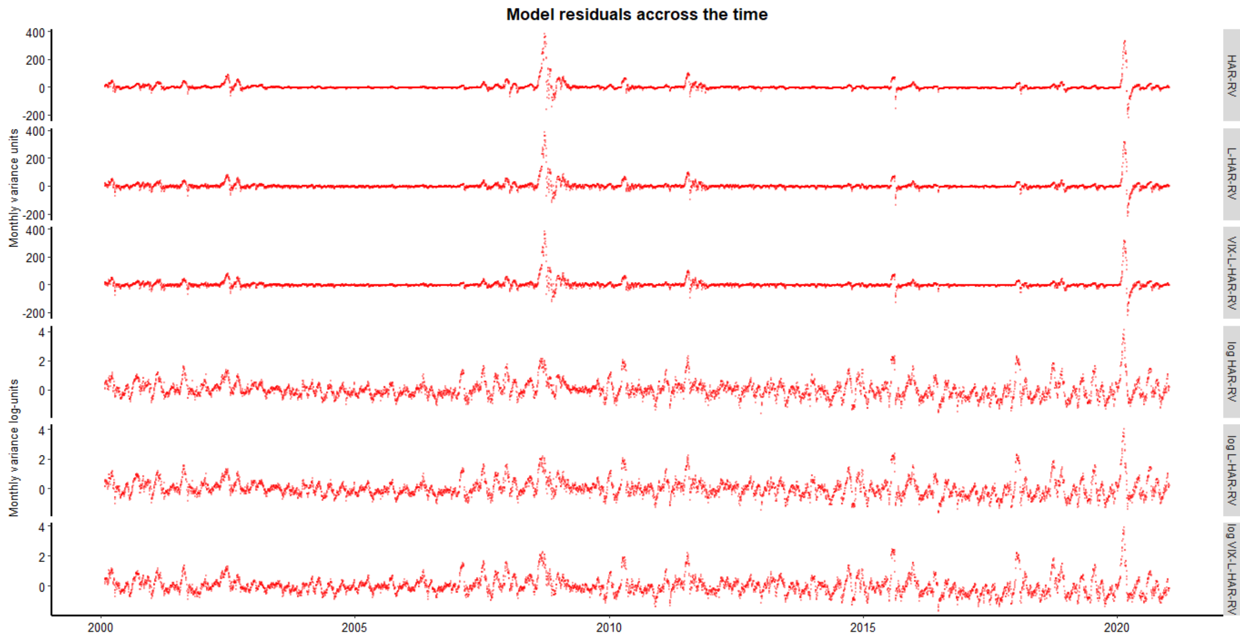
Table 2: Heteroskedasticity and serial correlation testing

	Breusch-Pagan test			Breusch-Godfrey test		
	BP statistic	df	p-value	BG statistic	df	p-value
HAR-RV	457.030	3	< 2.2e-16	5,098.400	1	< 2.2e-16
L-HAR-RV	511.520	6	<2.2e-16	5,088.400	1	< 2.2e-16
VIX-L-HAR-RV	550.110	7	< 2.2e-16	5,097.400	1	< 2.2e-16
log HAR-RV	42.751	3	2.78E-09	5,018.600	1	< 2.2e-16
log L-HAR-RV	63.308	6	9.55E-12	5,033.500	1	< 2.2e-16
log VIX-L-HAR-RV	76.936	7	5.80E-14	5,070.300	1	< 2.2e-16

Notes: Provided are the statistics of tests for heteroskedasticity and serial correlation in the model residuals. Order of 1 was chosen for the Breusch-Godfrey test. The results above are for the models estimated in full sample (3rd of January 2000 until 12th of February 2021).

Figure 7 plots the errors for each of the estimated models and displays a blatant clustering of the data points which indicates the presence of autocorrelation. Thereby, in addition to testing for general heteroscedasticity in the error term we also test for the serial correlation of the residuals by performing a Breusch-Godfrey test. Similar to the heteroscedasticity test described above, the procedure is again a LM test.

Figure 7: Residuals of models estimated using full sample



Notes: Plotted are residuals of each of the models estimated in the full sample (3rd of January 2000 until 12th of February 2021). Clustering which is especially noticeable in the residuals of models estimated in level form is an indication of serial correlation being present.

The model residuals are regressed on the original independent variables of a corresponding model and the lagged residuals of order ranging from 1 up to p. The LM statistic is then computed by multiplying the number of observations n and the coefficient of determination R_a^2 from this auxiliary regression.

$$LM_{BG} = nR_a^2 \tag{47}$$

The test statistic is asymptotically chi-squared distributed with p degrees of freedom (Greene, 2014). The results of the Breusch-Godfrey tests against the first-order autocorrelation of the residuals are reported in Table 2. The strong evidence against the null of no autocorrelation in the error term at lag one is not surprising considering our use of overlapping data in the regressions.

In the presence of heteroskedasticity and serial correlation the covariance matrix of the error term Ω can be represented as:

$$\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \dots & \sigma_{1T} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \dots & \sigma_{2T} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} & \dots & \sigma_{3T} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \dots & \sigma_{4T} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1T} & \sigma_{2T} & \sigma_{3T} & \sigma_{4T} & \dots & \sigma_T^2 \end{bmatrix} \tag{48}$$

which implies that the conditional variance of the least squares estimator is no longer the usual expression for the homoscedastic case (Greene, 2014):

$$Var[b|X] = \sigma^2 \left(X'X\right)^{-1} \tag{49}$$

but rather the general expression:

$$Var[b|X] = \left(X'X\right)^{-1} X'(\Omega) X \left(X'X\right)^{-1} \tag{50}$$

Since the central term Ω cannot be simplified to $\sigma^2 I$, where X is a design matrix of predictor values and I is the identity matrix. As the structure of Ω is not directly observed we need to estimate it. In this paper, we employ autocorrelation consistent covariance estimator introduced by Newey and West (1987) which has over the years become very popular in econometric research. The authors utilize the estimated residuals $\hat{u}_t$ and apply a weighting scheme with the assigned weighting terms diminishing in value as one moves away from the diagonal. This treatment ensures that the resulting matrix is positive semi-definite:

$$\widehat{\Omega}_{NW} = \begin{bmatrix} \hat{u}_1^2 & w_1\hat{u}_1\hat{u}_2 & w_2\hat{u}_1\hat{u}_3 & 0 & \dots & 0 \\ w_1\hat{u}_1\hat{u}_2 & \hat{u}_2^2 & w_1\hat{u}_2\hat{u}_3 & w_2\hat{u}_2\hat{u}_4 & \ddots & \vdots \\ w_2\hat{u}_1\hat{u}_3 & w_1\hat{u}_2\hat{u}_3 & \hat{u}_3^2 & w_1\hat{u}_3\hat{u}_4 & \ddots & 0 \\ 0 & w_2\hat{u}_2\hat{u}_4 & w_1\hat{u}_3\hat{u}_4 & \hat{u}_4^2 & \ddots & w_2\hat{u}_{T-2}\hat{u}_T \\ \vdots & \ddots & \ddots & \ddots & \ddots & w_1\hat{u}_{T-1}\hat{u}_T \\ 0 & \dots & 0 & \dots & \dots & \hat{u}_T^2 \end{bmatrix} \tag{51}$$

The weight of covariances located on a particular off-diagonal is computed as:

$$w_l = 1 - \frac{l}{(L+1)} \tag{52}$$

where L is the so-called truncation parameter or a bandwidth which is usually defined as a function of the sample size. Unfortunately, there is no universal rule specifying the formula for L. For example, Greene (2014) recommends the rule of thumb $L \approx T^{\frac{1}{4}}$ while others (i.e. Stock & Watson, 2016) advise $L = 0.75T^{\frac{1}{3}}$. Finally, we calculate the variance of our estimated parameter $\hat{\beta}$:

$$Var_{NW}(\hat{\beta}) = (X'X)^{-1}X'(\widehat{\Omega}_{NW})X(X'X)^{-1} \tag{53}$$

Recent literature stresses that incorporating overlapping data when forecasting in long-horizons relative to the length of samples provides only a minor benefit (Boudoukh et al., 2019). According to the authors, the commonly used methods of correcting for serial correlation including Newey-West estimator provide fictious comfort to the practitioners and that standard error estimates obtained by these methods are in such cases unreliable. To account for the bias and noise caused by the features of our sample and method, we impose higher than usual "rule of thumb" number of Newey-West lags when constructing the heteroskedasticity and autocorrelation consistent standard errors for our models.

7.3 Variance Risk Premia Construction

To construct the variance risk premia the one month ahead realized variance estimates of different models are subtracted from the squared monthly quantity of VIX. Formally:

$$VRP_t = {VIX_t}^2 - \widehat{RV_{t+22}^{m}} \tag{54}$$

This exercise is analogue to determining the premium demanded from a party taking the long position in the variance swap. One can recall equation (1) where the variance swap payoff for the party buying the variance is the difference between the actual realized variance and a variance strike multiplied by variance notional. In our application ${VIX_t}^2$ acts as the variance strike and the model fitted values $\widehat{RV_{t+22}^m}$ approximate the realized variance. We express the variance risk premia in monthly variance units instead of monetary terms and thereby no variance notional is required. As these payoffs tend to be generally negative, we expect a positive variance risk premia (Carr & Wu, 2009).

8. Realized Variance Modelling - Empirical Results

We report the regression output for the level and logarithmic models in Tables 3 and 4, respectively. All of the models were estimated over the full sample of 5250 daily observations. The standard errors in brackets underneath the coefficients are Newey-West estimator with 44 lags. The applied heteroskedasticity and autocorrelation consistent standard errors are thereby considerably larger than the usual OLS standard errors, but they allow us to perform statistical inference. For comparison, the formula recommended by Greene (2014) yields 8.5 lags and the rule of thumb proposed by Stock & Watson (2016) results in 13 Newey-West lags. We consider this rather large derogation from the usual rules of thumb as appropriate reflecting on the remarks of Müller (2014) who states that for strongly autocorrelated time series the Newey-West estimators with the usual bandwidths are severely biased downwards and this bias can only be tackled by choosing a very large number of lags which itself leads to issues caused by high sampling variability of the estimator. A rather extensive branch of the literature has been developed attempting to address the poor small sample performance of these procedures by taking the sampling variability of the long run variance estimators into account. While we do not pursue this avenue further and rely only on somehow arbitrary increase in the size of the bandwidth[10], the interested reader is directed to works of Sun et al.,(2008), Müller (2014) and Kiefer and Vogelsang (2005) for more information on the subject.

[10] As it was performed by Bekaert & Hoerova (2014).

Table 3: Full sample regressions in level form

Regressions in Levels

	Dependent variable:		
		RV_{t+22}^{m}	
	HAR-RV	L-HAR-RV	VIX-L-HAR-RV
	(1)	(2)	(3)
VIX_t^2			0.031
			(0.176)
RV_t^d	0.154^{***}	0.089^{***}	0.086^{***}
	(0.035)	(0.024)	(0.025)
RV_t^w	0.296^{***}	0.229^{***}	0.226^{**}
	(0.093)	(0.087)	(0.095)
RV_t^m	0.185^{*}	0.209^{*}	0.194
	(0.105)	(0.122)	(0.161)
r_t^{d-}		-0.191^{***}	-0.184^{**}
		(0.049)	(0.083)
r_t^{w-}		-0.737^{**}	-0.713^{*}
		(0.293)	(0.384)
r_t^{m-}		0.177	0.205
		(0.524)	(0.533)
Constant	8.905^{***}	4.965^{*}	4.781^{*}
	(2.140)	(2.992)	(2.761)
Observations	5,250	5,250	5,250
R^2	0.470	0.493	0.493
Adjusted R^2	0.469	0.492	0.492
Residual Std. Error	31.360 (df = 5246)	30.671 (df = 5243)	30.670 (df = 5242)
F Statistic	$1{,}548.916^{***}$ (df = 3; 5246)	849.882^{***} (df = 6; 5243)	728.676^{***} (df = 7; 5242)
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: Sample period spans from the 3rd of January 2000 to 12th of February 2021. All regressions are based on daily observations leading to considerable overlapping. Therefore, the standard errors in brackets are constructed using 44 Newey-West lags. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme above.

The original HAR-RV model of Corsi (2004), whose coefficients are reported in column one of Table 3 explained approximately 47% of all variation in the dependent variable. Three out of four estimated coefficients are statistically significant at the 1% level with only the coefficient for the monthly realized variance not meeting the 1% threshold but still being significant at the 90% confidence level. It is notable that each of the coefficients was estimated with the expected positive sign meaning that the increase in the past volatility observed over a particular horizon drives up the estimated value. The explanatory variable representing the weekly realized variance displays the largest effect among the volatility components appearing in the regression. This observation holds true across all estimated models, level and logarithmic alike, with the exception of log L-HAR-RV where the monthly component had the dominant impact on the estimates of the one-month ahead realized variance.

Including the leverage effect variables increases the adjusted R^2 by an additional 2.2%. The coefficients calculated for the daily and weekly leverage effects are both significant at 1% and 5%, respectively, while the monthly component of the leverage effects is not statistically significant. Paralleling the past realized volatility components, we again obtain the expected sign for the significant coefficients of leverage effect variables. As discussed in section (7.1) leverage effect variables are constructed by summing the vector of negative daily logarithmic returns. The negative coefficient therefore multiplies a variable that is by design non-positive. This captures the inverse relationship between the stock returns and variance covered in section (4.1.3) whereby the decreases in the stock price correspond to increases in the modelled realized variance.

The addition of the squared monthly VIX quantity as a regressor in the level form does not have any noteworthy impact on model's R^2 but marginally affects the magnitudes and level of significance of other coefficients. Most notably the coefficient for monthly horizon of realized variance is no longer statistically significant. As for the added variable itself, its coefficient is not statistically significant either.

Table 4: Full sample regressions in logarithmic form

Regressions in Log

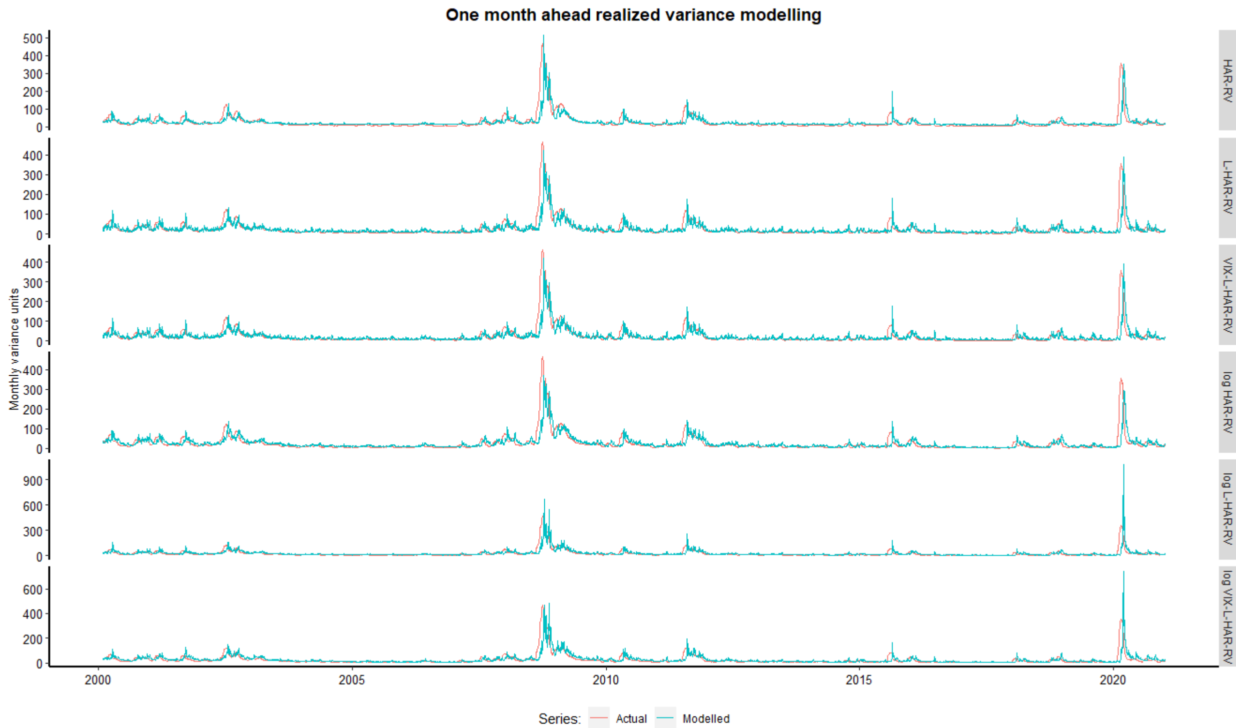
	Dependent variable:		
		$\ln(RV_{t+22}^{m})$	
	log HAR-RV	log L-HAR-RV	log VIX-L-HAR-RV
	(1)	(2)	(3)
$\ln(VIX_t^2)$			0.544^{***}
			(0.094)
$\ln(RV_t^d)$	0.202^{***}	0.144^{***}	0.090^{***}
	(0.018)	(0.017)	(0.016)
$\ln(RV_t^w)$	0.289^{***}	0.246^{***}	0.175^{***}
	(0.046)	(0.045)	(0.045)
$\ln(RV_t^m)$	0.279^{***}	0.356^{***}	0.129
	(0.061)	(0.072)	(0.087)
r_t^{d-}		-0.003^{***}	-0.001^{***}
		(0.0003)	(0.0004)
r_t^{w-}		-0.010^{***}	-0.006^{**}
		(0.002)	(0.003)
r_t^{m-}		0.009	0.008
		(0.008)	(0.008)
Constant	0.666^{***}	0.676^{***}	-0.196
	(0.101)	(0.124)	(0.174)
Observations	5,250	5,250	5,250
R^2	0.625	0.633	0.655
Adjusted R^2	0.624	0.632	0.654
Residual Std. Error	0.601 (df = 5246)	0.595 (df = 5243)	0.577 (df = 5242)
F Statistic	$2,908.263^{***}$ (df = 3; 5246)	$1,505.901^{***}$ (df = 6; 5243)	$1,419.684^{***}$ (df = 7; 5242)
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: Sample period spans from the 3rd of January 2000 to 12th of February 2021. All regressions are based on daily observations leading to considerable overlapping. Therefore, the standard errors in brackets are constructed using 44 Newey-West lags. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme above.

Estimating Corsi’s (2004) basic model in the logarithmic form leads to major improvements in the model’s fit. This is reflected by the adjusted R^2 of the model increasing to 62%. In addition, each of the estimated coefficients is significant at the 1% level. As already mentioned, the weekly variance component has the largest coefficient however, it is only marginally bigger than the monthly component. Incorporating the leverage effects further improves the model’s capability to explain variations in the modelled variable. Paralleling the log L-HAR-RV’s to its level form counterpart, only the effects of daily and weekly leverage effects are statistically significant with weekly coefficient dominating in terms of magnitude.

In contrast to the level regressions where the inclusion of squared monthly VIX variable failed to make a significant difference in model’s explanatory power, in the logarithmic regressions the same regressor helps to boost the adjusted R^2 by an extra 2.1% relative to its nested model.

Figure 8: One month ahead estimates of monthly RV – full sample estimation



Notes: Fitted values of regressions estimated using full sample (from the 3rd of January 2020 to 12th of February 2021) in turquoise are compared to the actual level of the 22 days ahead realized variance in red. The plot is divided into 6 facets corresponding to each of the estimated models. Series for logarithmic models (log HAR-RV, log L-HAR-RV and log VIX-L-HAR-RV) are constructed by transforming the fitted values logarithmic estimates to level following the equation (42).

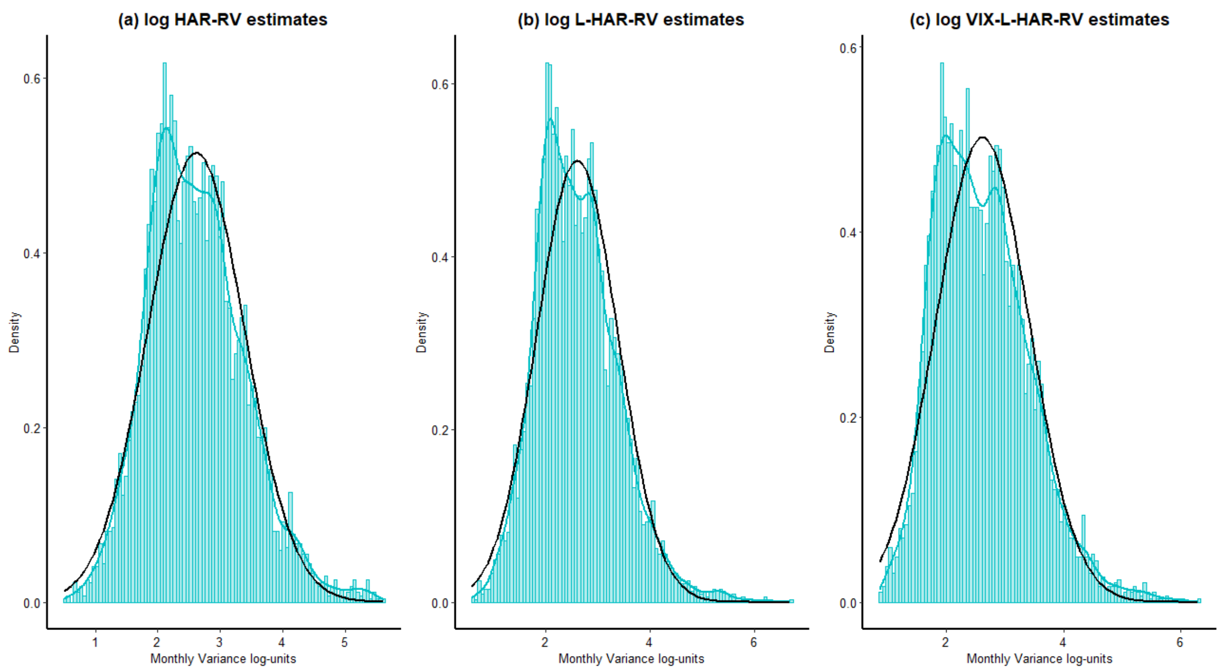
The fitted values[11] for each of the estimated models are compared to the actual one month-ahead levels of the realized variance in Figure 8. Overall, it appears that our models capture the long-range dependence and other stylized facts of the modelled variable reasonably well. However, it is notable that both log L-HAR-RV and log VIX-L-HAR-RV tend to massively overpredict the level of the realized variance during the crisis periods. This might, at first glance, seem surprising considering the same is not observed for the

[11] For logarithmic estimates their transformations to the level.

estimates of the less complex logarithmic model, log HAR-RV, and that the best fit was achieved by these two most comprehensive log-models.

Plot A.2 in Appendix 3 as well as the relatively high R^2's reported for the two models corroborate that the root of this issue is not in the estimation itself but rather in the conversion of $\ln(\widehat{RV_{t+22}^m})$ to level form. In section (7.1), we assumed that the modelled variable is lognormally distributed when applying equation (42) to transform the log estimates to level. As is revealed by Figure 9, the probability density functions for the estimates generated by the two models in question deviate from normality considerably more than the ones for the fitted values of the basic logarithmic model. While the three estimated logarithmic variables have effectively identical first two moments, the distributions of log L-HAR-RV and log VIX-L-HAR-RV estimates are significantly more skewed to the right and more leptokurtic. We are especially concerned about the mass of modelled values located in the right tail. To illustrate, the maximum value estimated by the log HAR-RV model is located 3.85 standard deviations away from the mean compared to 5.23 standard deviations for the maximum estimate of the log L-HAR-RV model. The right tails being more drawn out in combination with a larger share of the two distributions being located in these tails cause major issues during the logarithm to level transformation.

Figure 9: Distributions of logarithmic estimates



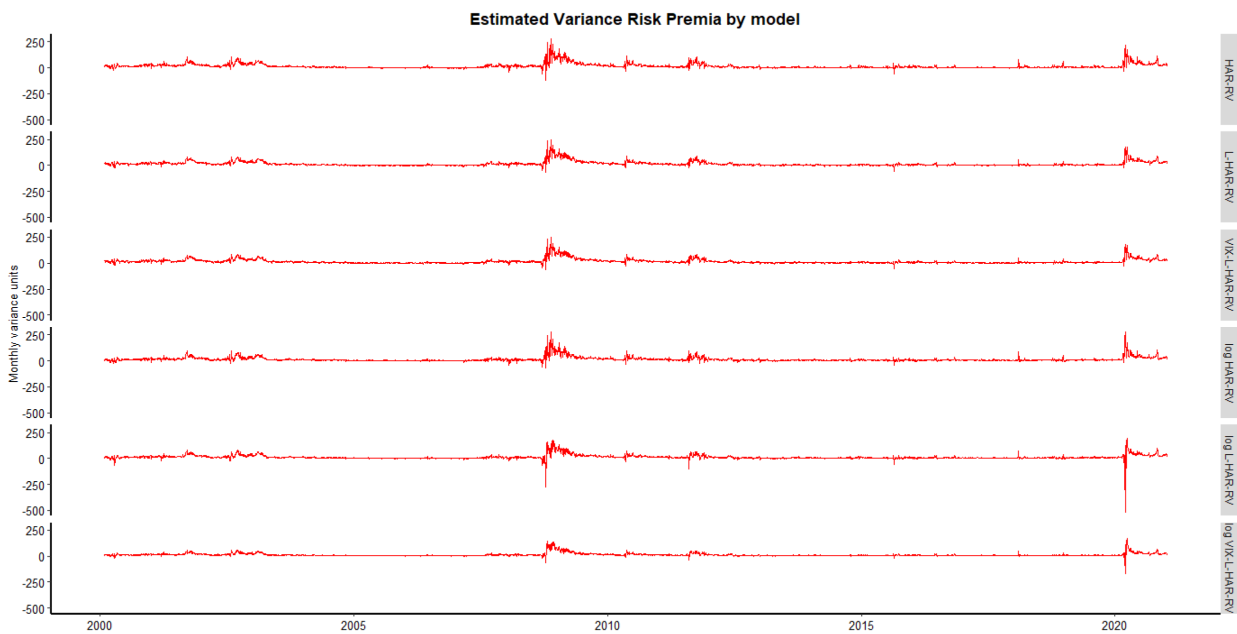
Notes: Histograms & probability density plots for the logarithmic estimates. The fitted values are estimated utilizing full sample (from the 3rd of January 2020 to 12th of February 2021). Black curve represents the gaussian probability density function with the same first two moments as the distributions of the modelled values.

Figure 10 presents the variance risk premia constructed as per equation (54) corresponding to each of the estimated realized variance models. In line with our theoretical expectations, the variance risk premia increase during times of crises and heightened economic uncertainty. On the other hand, we observe a reversion of the variance risk premia to lower levels during recovering, bull and stable markets. In addition, the expected value of the variance risk premium is positive across all models[12]. This is analogous to a negative expected payoff for a buyer of the underlying in a variance swap which is again aligned with the theory of variance risk premia. However, the constructed variance risk premium does sometimes turn negative, a problem that is not isolated to our study specifically but one which can be found in multiple

[12] Full set of descriptive statistics can be found in Table A.1 of Appendix 4.

similar papers (Bollerslev et al., 2009; Bekaert and Hoerova, 2014). These states appear during the absolute extreme peaks of realized variance and as argued by Bekaert and Hoerova (2014) they do not have a sound economic interpretation. The authors conjecture that negative variance risk premia are caused by the inherent shortcomings of the chosen model. It appears that the realized variance has different dynamics of mean reversion[13] in the extreme periods relative to the rest of the economic states. Linear models thereby cannot satisfactorily capture the development in the conditional variance.

Figure 10: VRPs constructed using full sample estimation



Notes: Daily VRP series constructed as the difference between the monthly squared VIX quantity and fitted values of level regressions or transformed fitted values of logarithmic regressions estimated using full sample (from the 3rd of January 2000 to 12th of February 2021) according to the equation (54). The plot is divided into 6 facets corresponding to each of the estimated models.

[13] Bekaert and Hoerova (2014) theorize that during the major crises selected realized variance movements should be able to revert to the mean faster resulting in lower effects on conditional variance.

It should also be noted that while the negative values of the variance risk premium during the extreme periods can be observed for each of our models, the overprediction of the realized variance levels by log L-HAR-RV and log VIX-L-HAR-RV models stemming from the logarithm to level transformation issue, outlined above, further exacerbates this problem. Following the Bekaert and Hoerova's (2014) line of thought, non-linear model type would likely be better suited to handle the behavior of realized variance in the crisis period. Such models are however outside of the scope of this paper.

The estimation of the models over the full sample has provided us with valuable insights regarding the dynamics of the process and ability of different models to capture it. However, it is still not clear which of the models should be preferred by the practitioners if the goal is to project future realized variance and eventually future stock returns as the above analysis is backward looking. The following section introduces the procedure designed to emulate the forecasting efforts of market participants which then allows us to analyze the quasi out of sample performance of different models in attempt to select the recommended one.

9. Model Selection and Forecasting

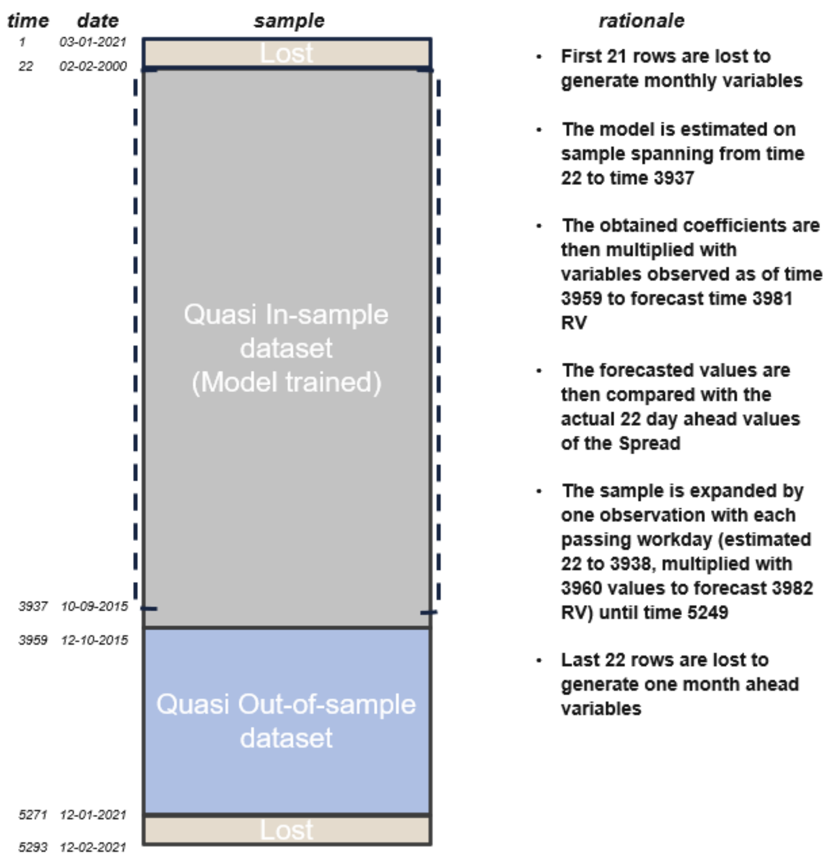
9.1 Forecasting Approach

To identify the best realized monthly variance estimation model, the out-of-sample forecasting performances of the six models specified in the previous sections are compared. The models are trained on the first 75% of the sample corresponding to the period between the 3rd of January 2000 and 09th of October 2015 and the remaining 25% of the dataset representing the period from 12th of October 2015 until 12th of February 2021 is used as a reference for the forecast performance evaluation.

Contrary to Bekaert and Hoerova (2014) who apply the same fixed coefficients estimated based on the first 75% of the data across the whole out-of-sample section of the dataset we continuously expand the sample size available to the out-of-sample model to imitate the process of market players performing the one month ahead forecasts. To this end, we run 1335 regressions for each of the six models and extract the coefficients which are then used for the construction of forecasted values. Our model estimates time $t+22$ realized variance at time t which would cause hindsight bias in the out-of-sample forecast as the value of the regressand is not observable at time t. To eliminate the hindsight bias, coefficients from time $t-22$ are applied to variables observed at time t to forecast the realized variance at time $t+22$[14]. The sample size for the calculation of sample variance of rv_{t+22}^{m} is also recursively extended for each of the "newer" regressions. Figure 11 provides an illustration of this procedure.

[14] This is the reason for estimating 1335 regressions despite only 1313 values are to be forecasted. The coefficients from the last 22 regressions are deleted as we do not observe the one month ahead realized variances at those points in time and therefore the forecast cannot be compared to the actual value.

Figure 11: Forecasting procedure



Notes: The plot above graphically visualizes and explains the forecasting procedure. The figure should serve for illustrative purposes only as the ratios of block sizes representing particular dataset sections do not correspond to the ratios of their sample sizes.

In addition to evaluating the forecasts separately, we also consider their combination. The seminal paper by Bates and Granger (1969) has sparked great interest in the potential benefits of combining projections from alternative sources. Numerous studies have successfully explored combination forecasts (e.g. Clemen, 1998; Stock & Watson, 2004; Marcellino, 2004). The common argument in favor of forecast combinations is the possible existence of diversification gains from the "robustification" of forecasts against misspecification biases and effects of structural breaks (Timmermann, 2006). On the other hand, there is also an extensive body of literature arguing against the method. Some authors (e.g. Diebold, 1989) argue that when all the underlying information sets are known, pooling of these sets is, from an econometric perspective, more desirable than pooling of the forecasts. Other papers (Yang, 2004; Kang, 1986) highlight issues of contamination and instability of combination weights caused by the estimation errors and non-stationarities in the underlying data generating process.

In this paper we use a simple equal weighting combination scheme which has historically proven to be difficult to beat and thereby has been viewed as a natural benchmark (Claeskens, 2016). Interested reader can find a large body of research focusing on the estimation of optimal weights in forecast combination and the topic of the forecast combination puzzle (Graefe et al., 2014; Smith & Wallis, 2009). This theory is however outside of the scope of this paper. The following section introduces the procedure used to select the best performing models which are ultimately included in our combined forecast.

9.2 Forecasting Performance Evaluation

The Model Confidence Set (MCS) of Hansen et. al (2011) is implemented to statistically compare the forecasting performance of the estimated models and martingale[15] model proposed by Bollerslev et.al (2009). The MCS refers to a sequential testing procedure which allows one to identify the group of Superior Set Models (SSM) $M^*$for which the null hypothesis of equal predictive ability (EPA), computed using a chosen loss function, is not rejected. The MCS begins with the set of competing objects, in our case models, $M_0 = \{1, \ldots\ldots, m\}$ and an arbitrary criterion, usually a form of a loss function, that is used for

[15] As mentioned in section (5), Bollerslev et al.'s (2009) martingale model expects one step ahead realized variance to be equal to its value at lag one. Formally for one month ahead case: $E_t(RV_{t+22}^m) = RV_t^m$

the empirical evaluation of these objects. The procedure then considers the loss differentials between the models i and j $d_{i,j,t}$ defined as:

$$d_{ij,t} = L_{i,t} - L_{j,t}, \qquad i,j = 1, \dots, m, \quad t = 1, \dots, T \tag{55}$$

where T is the total number of observations out-of-sample and loss function for model i is calculated as:

$$L_{i,t} = L\left(RV_t^m, \widehat{RV_{i,t}^m}\right) \tag{56}$$

With $\widehat{RV_{i,t}^m}$ being a model i forecast of the actual value RV_t^m observed out-of-sample at time t. As has been stated above, the method allows for flexibility when choosing the utilized loss function which enables the user to compare models based on various aspects. For the purposes of our application, we focus on the punctual out-of-sample forecasting of realized variance[16] and thereby choose the Mean Squared Error (MSE) as an evaluation criterion in the MCS due to its symmetry and robustness (Jakobsen, 2018). Formally:

$$L_{i,t} = \left(RV_{i,t}^m - \widehat{RV_{i,t}^m}\right)^2 \tag{57}$$

According to Patton (2011) a loss function is robust when "the ranking of any two volatility forecasts, by expected loss is the same whether the ranking is done using the true conditional variance or some conditionally unbiased proxy" (Patton, 2011, p. 248). The author then proceeds to evaluate the collection of the loss functions commonly seen in the literature[17] and concludes that only MSE and Quasi-Likelihood (QLIKE) loss functions belong to the class of robust loss functions suitable for the comparison of volatility forecasts.

[16] As discussed at length in section High-frequency realized variance. Theoretical framework the realized variance is a proxy for inherently unobservable integrated volatility.
[17] Full list of the loss functions evaluated by Patton (2011): MSE, QLIKE, MSE-LOG, MSE-SD, MSE-prop, MAE, MAE-LOG, MAE-SD, MAE-prop.

The procedure then continues with the application of the equivalence test δ_M testing the null hypothesis of EPA at a chosen significance level. The MCS assumes that $\mu_{ij} = E(d_{ij,t})$ is finite and time independent. The null can be formulated as:

$$H_{0,M}: \mu_{ij} = 0 \qquad for\ all\ i,j \in M \tag{58}$$

where M is a subset of M_0 and the alternative hypothesis is stated as:

$$H_{A,M}: \mu_{ij} \neq 0 \qquad for\ some\ i,j \in M \tag{59}$$

Special t statistics must be computed for the hypotheses to be tested. To construct the t statistics one has to calculate the relative sample loss statistics $\bar{d}_{ij} = \frac{\sum_{t=1}^{T} d_{ij,t}}{n}$ and $\bar{d}_i = \frac{\sum_{j \in M} \bar{d}_{ij}}{m}$, which represent the relative sample losses between ith and jth model and the sample loss of the ith model relative to the average of all models in M, respectively. We construct the t statistics:

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{Var(\bar{d}_{ij})}}} \qquad for\ i,j \in M \tag{60}$$

and

$$t_i = \frac{\bar{d}_i}{\sqrt{\widehat{Var(\bar{d}_i)}}} \qquad for\ i \in M \tag{61}$$

where $\widehat{Var(\bar{d}_{ij})}$ is the estimated variance of $\bar{d}_{ij}$ and $\widehat{Var(\bar{d}_i)}$ is the estimated variance of $\bar{d}_i$ obtained by bootstrapping[18]. These two statistics are directly linked to the null hypotheses $H_{ij}: \mu_{ij} = 0$ and $H_i: \mu_i = 0$ which[19] are both effectively equivalent to $H_{0,M}$ since the "main" null hypothesis of EPA, $H_{0,M}$,

[18] Detailed characteristics of the bootstrap procedure will be outlined in section (9.4).
[19] Where μ_i is defined as $\mu_i = E(\bar{d}_i)$

encompasses $\{H_{ij}\, for\ all\ i,j \in M\}$ while also being correspondent to $\{H_i\, for\ all\ i \in M\}$. As the equivalence extends to $\{|\mu_{ij}| < 0\, for\ all\ i,j \in M\}$ and $\{\mu_{ij} < 0\, for\ all\ i \in M\}$ authors map the hypothesis into the test statistics which are finally used to test the $H_{0,M}$:

$$T_{R,M} = max\, |t_{ij}|_{i,j\in M} \tag{62}$$

and

$$T_{max,M} = max\, t_{i_{i\in M}} \tag{63}$$

Hansen et al. (2011) point out that the asymptotic distributions of the test statistics are nonstandard due to their dependency on nuisance parameters and propose a bootstrap method that can handle the nuisance parameter problem.

Lastly, the elimination rules e_M ensuring the exclusion of inferior models in case of rejection of the equivalence test are introduced to the procedure. For the first test statistic, $T_{R,M}$, the elimination rule is:

$$e_{R,M} = \arg max_{i\in M}\, sup_{j\in M} t_{ij} \tag{64}$$

where the test statistic for the eliminated model is a supremum of the set ensuring that $t_{e_{R,M}j} = T_{R,M}$. As for the test statistic $T_{max,M}$ the elimination rule equates to:

$$e_{max,M} = \arg max_{i\in M}\, t_i \tag{65}$$

As the rejection of null hypothesis recognizes the hypothesis $\mu_j = 0$ as being false for $j = e_{max,M}$ which leads to the procedure removing the model with the largest contribution to the test statistic.

To summarize, δ_M is applied to the initial set of models, in case of its rejection, e_M is utilized to eliminate the model with the poor sample performance from the set. This is repeated until δ_M is accepted at a chosen confidence level and the group of SSM is defined.

In addition to performing the MCS we also consider multiple standard measures of forecasting accuracy. These include Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage error (MPE), Mean Absolute Percentage Error (MAPE) as well as Mean Squared Error (MSE) which has been used as the evaluation criterion in the MCS itself.

We now transition from the theoretical part of this section dedicated to realized variance forecasting and model selection to the empirical segment. The following section summarizes the models trained on the in-sample portion of the dataset. Afterwards, out-of-sample projections for each of the models are presented and evaluated based on their forecast accuracy.

9.3 Regressions in-sample – Empirical Results

Table 5 shows the in-sample regression coefficients for the three level models including the Newey-West standard errors in parentheses. The coefficients of the standard HAR-RV model have the expected sign and are all significant at least at the 10% significance level across all three level models. The significance of the independent variables across the three horizons displays the long-range dependence of realized variance. In the first model, the weekly realized variance variable has the largest coefficient such that, holding all other variables constant, a one unit change in RV_t^w leads to a 0.306 unit change in the dependent variable. Similarly, Bekaert and Hoerova (2014) observe that the weekly realized variance carries the largest coefficient, although their comparable (levels) model also includes a squared VIX term, unlike our first model. Model 1 displays solid explanatory abilities, explaining 56.5% of the variations in the dependent variable. Expanding the first model by adding leverage effects (i.e. model 2), as is the case for the L-HAR-RV model, marginally improves the adjusted R^2. Significant leverage effects are observed for daily and weekly horizon, while the monthly leverage effect is non-significant. Given that the leverage effect variables are expressed as natural logarithmic returns, the coefficient interpretation is different from the realized variance coefficients. Specifically, all else equal, a 1% increase in one of the leverage effect variables will reduce the dependent variable by $\frac{\beta^{(.)}}{100}$ units, where $\beta^{(.)}$ represents the respective coefficients at the different time horizons. Compared to model 2, the model fit does not improve by including the squared VIX term (model VIX-L-HAR-RV). The coefficient for the squared VIX variable does not have the expected sign, although given the large standard error there is not sufficient statistical evidence to infer that the coefficient is different from zero. The non-significance of some coefficients is largely driven by the corrected standard errors accounting for serial autocorrelation. The regression output of the in-sample logarithmic models is outlined in Table 6. These models are the logarithmic transformations of the models in the paragraph above. It appears that the logarithmic models display a superior fit than their level counterparts since the adjusted R^2 is relatively larger for all log models. Interestingly, out of the three realized variance variables, the monthly realized variance has the largest influence on the dependent variable in the first two models, while for the level models the weekly realized variance dominated. Log model 2 (i.e. log L-HAR-RV) is fairly similar to the levels model although the coefficients showcase greater statistical significance. Log model 3 (i.e. log VIX-L-HAR-RV) includes a significant coefficient for the squared VIX term and now has the expected sign, unlike the corresponding levels model.

Table 5: In sample regressions in level form

Regressions in Levels in sample

	Dependent variable:		
		RV_{t+22}^{m}	
	HAR-RV	L-HAR-RV	VIX-L-HAR-RV
	(1)	(2)	(3)
VIX_t^2			-0.032
			(0.202)
RV_t^d	0.117***	0.067***	0.071***
	(0.022)	(0.017)	(0.026)
RV_t^w	0.306***	0.250***	0.250***
	(0.110)	(0.092)	(0.092)
RV_t^m	0.296***	0.237*	0.255**
	(0.092)	(0.123)	(0.124)
r_t^{d-}		-0.179***	-0.185**
		(0.051)	(0.084)
r_t^{w-}		-0.601**	-0.625
		(0.262)	(0.387)
r_t^{m-}		-0.387	-0.430
		(0.419)	(0.597)
Constant	7.426***	1.073	1.200
	(1.540)	(3.524)	(3.069)
Observations	3,937	3,937	3,937
R^2	0.565	0.589	0.589
Adjusted R^2	0.565	0.588	0.588
Residual Std. Error	28.545 (df = 3933)	27.765 (df = 3930)	27.764 (df = 3929)
F Statistic	1,704.471*** (df = 3; 3933)	938.704*** (df = 6; 3930)	804.803*** (df = 7; 3929)
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: Sample period spans from the 3rd of January 2000 to 09th of October 2015. All regressions are based on daily observations leading to considerable overlapping. Therefore, the standard errors in brackets are constructed using 44 Newey-West lags. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme above.

Table 6: In sample regressions in logarithmic form

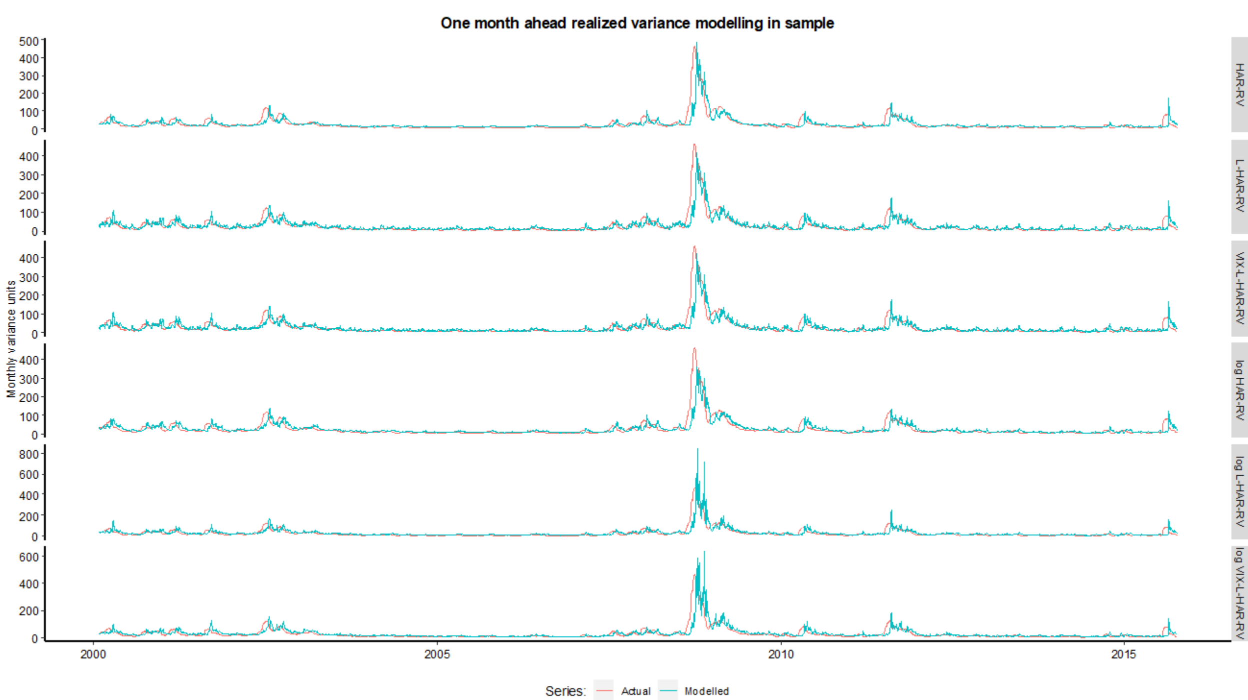
Regressions in Log in sample

	Dependent variable:		
		$\ln(RV_{t+22}^{m})$	
	log HAR-RV	log L-HAR-RV	log VIX-L-HAR-RV
	(1)	(2)	(3)
$\ln(VIX_t^2)$			0.560^{***}
			(0.108)
$\ln(RV_t^d)$	0.181^{***}	0.117^{***}	0.064^{***}
	(0.018)	(0.016)	(0.018)
$\ln(RV_t^w)$	0.268^{***}	0.207^{***}	0.133^{***}
	(0.050)	(0.050)	(0.051)
$\ln(RV_t^m)$	0.341^{***}	0.363^{***}	0.072
	(0.053)	(0.066)	(0.083)
r_t^{d-}		-0.003^{***}	-0.001^{***}
		(0.0004)	(0.0004)
r_t^{w-}		-0.011^{***}	-0.006^{**}
		(0.003)	(0.003)
r_t^{m-}		-0.001	-0.006
		(0.007)	(0.007)
Constant	0.635^{***}	0.762^{***}	0.004
	(0.104)	(0.128)	(0.193)
Observations	3,937	3,937	3,937
R^2	0.646	0.658	0.682
Adjusted R^2	0.645	0.658	0.681
Residual Std. Error	0.535 (df = 3933)	0.525 (df = 3930)	0.507 (df = 3929)
F Statistic	$2,389.947^{***}$ (df = 3; 3933)	$1,261.677^{***}$ (df = 6; 3930)	$1,201.002^{***}$ (df = 7; 3929)
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: Sample period spans from the 3rd of January 2020 to 9th of October 2015. All regressions are based on daily observations leading to considerable overlapping. Therefore, the standard errors in brackets are constructed using 44 Newey-West lags. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme above.

Our in-sample models achieved notably higher levels of R^2 compared to the models estimated using the full sample. The increase was most prominent for the models estimated in levels. This is likely caused by the omission of two highly turbulent periods which now appear in the out-of-sample portion of the dataset, namely a portion of 2015/2016 stock market sell-off and the recent market crash triggered by Covid-19 pandemic. As has been covered in the previous sections, periods of severe crises when realized variance peaks, pose major challenges for our linear models. This is no different for the models estimated in-sample and thereby we still observe overprediction of realized variance levels during the extreme periods across all models. This issue is again further amplified for transformed level values corresponding to log L-HAR-RV and log VIX-L-HAR-RV models due to the breach of lognormality assumption described in the previous sections. The fitted values are compared to actual one month ahead monthly realized variance in Figure 12 and the corresponding chart for the log form can be found in the Figure A.3 of Appendix 5.

Figure 12: One month ahead monthly RV estimates – in sample estimation



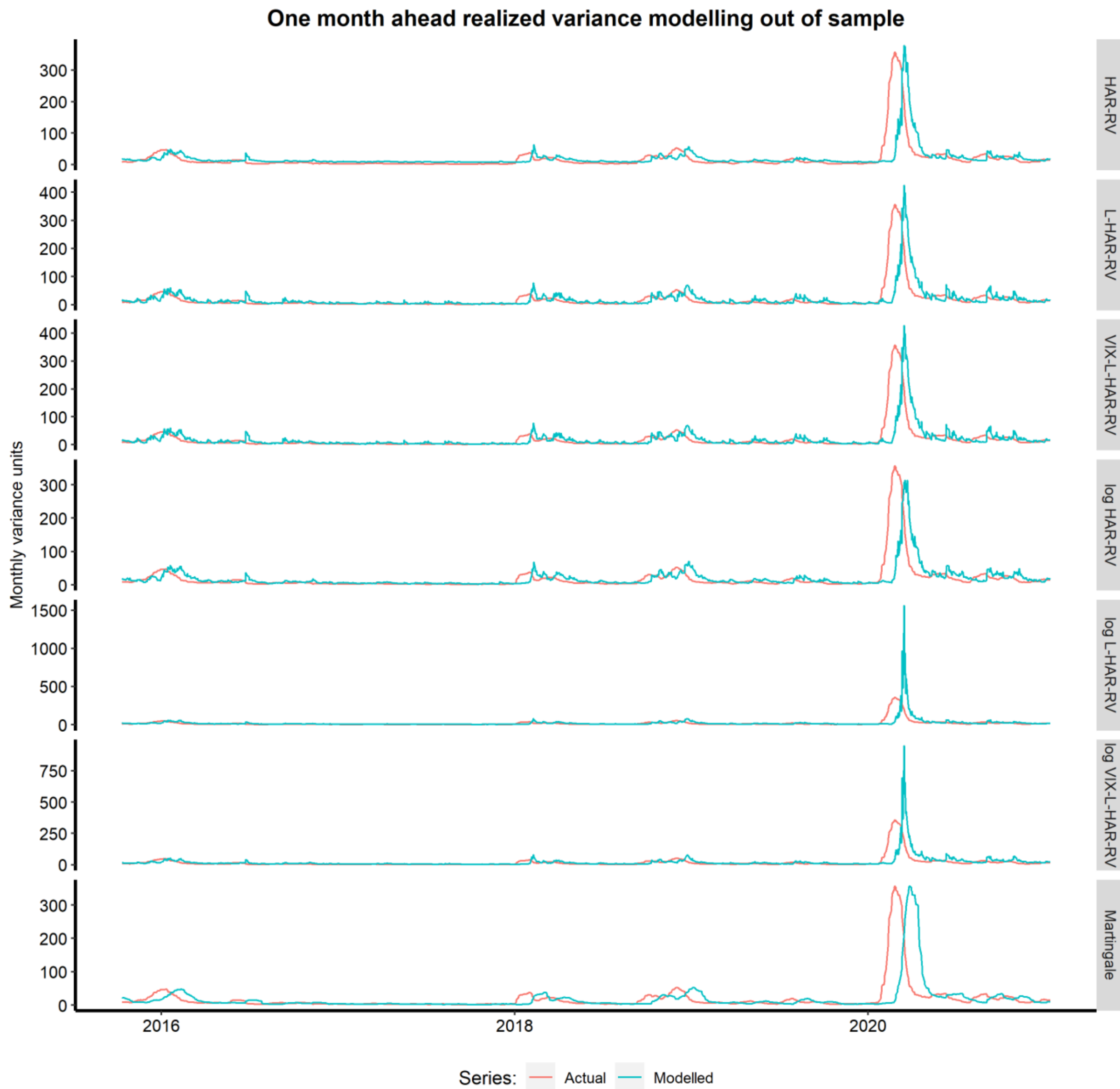
Notes: Fitted values of regressions estimated using 75% of the sample (from the 3rd of January 2020 to 9th of October 2015) in turquoise are compared to the actual level of the 22 days ahead realized variance in red. The plot is divided into 6 facets corresponding to each of the estimated models. Series for logarithmic models (log HAR-RV, log L-HAR-RV and log VIX-L-HAR-RV) are constructed by transforming the fitted logarithmic estimates to level following the equation (42).

The in-sample models summarized above are different to the starting point for our out-of-sample projections. As already mentioned, to avoid the look ahead bias the first out of sample forecast is obtained by multiplying the variables observed on 12th of October 2015 by the coefficients of the models estimated in sample ranging from the initial period to 10th of September 2015 to forecast 11th of November 2015 realized variance. The coefficients from these base models are then recursively updated with each passing day to incorporate new information; as described in section (9.1). It follows that the same coefficients as we have reported for the in-sample models are used to forecast the realized variance as of 11th of December 2015 (22nd out of sample forecast). We report the individual projections as well as their accuracy in the following section. In the following section we select the best performing models and thus conclude section (9). These are subsequently used to construct variance risk premia included in the stock return predictability regressions which are the main focus of the last part of our paper.

9.4 Out-of-sample Forecasting Performance

Figure 13 visualizes the comparison of the out-of-sample forecasts yielded by the considered models and the actual 22-day ahead monthly realized variance. When it comes to the general fit of the forecasts, one can note a substantial lag in the response of the models to developments in the realized variance. This is by virtue of the linear models we have employed to capture the long-range dependence in realized variance. In other words, large positive coefficients on weekly and monthly variance components cause the slow reaction to extreme spikes and by extension slow mean reversion of fitted values when the extreme values drop again.

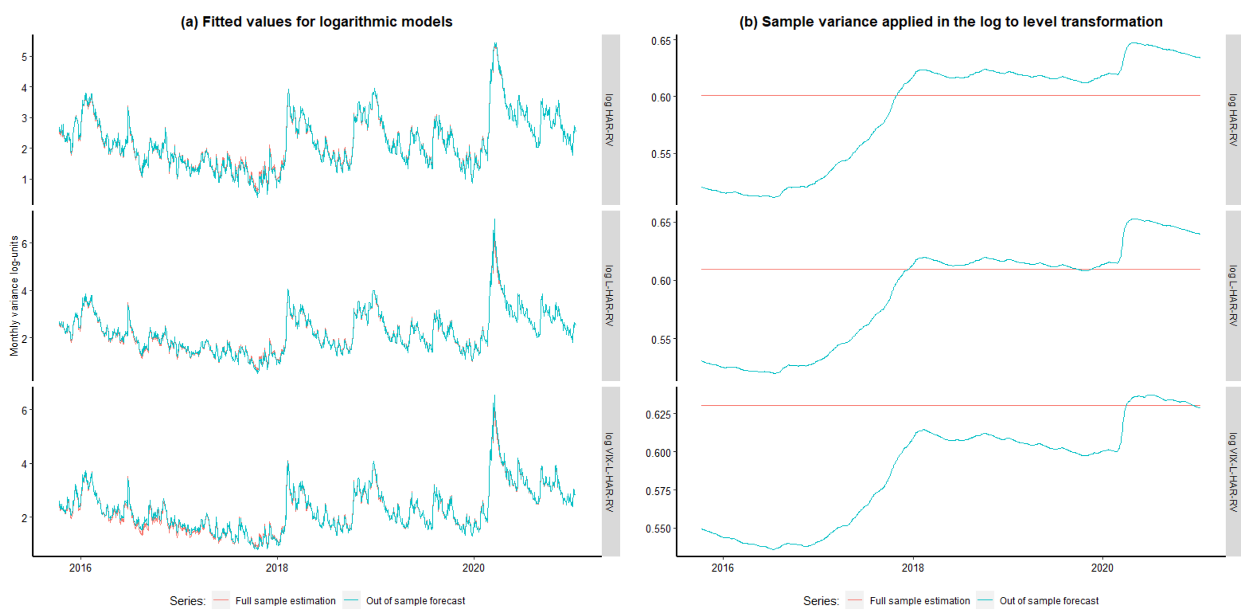
Figure 13: Out of sample forecasts of monthly realized variance



Notes: Forecasted level of RV in turquoise is compared to actual 22 day ahead RV in red. Forecasts are constructed by applying the RV observed at time t to coefficients of regressions using the sample spanning from the initial period to t-22 to avoid the look-ahead bias. Forecasts for logarithmic models are converted to level as per equation (42). In contrast to full sample and in sample estimation, the applied sample variance if the estimate is non constant.

As has been the case throughout all our modelling attempts, we again observe substantial overprediction during the peak periods for nearly all models. Paralleling our efforts of estimating the monthly realized variance using full sample, Covid-19 has proven to be a remarkably challenging period for our logarithmic models particularly those including leverage effects. In fact, the extent to which log L-HAR-RV and log VIX-L-HAR-RV overestimate the modelled values in levels during the extreme periods is materially amplified in the out-of-sample application. Figures 14(a) and 14(b) plot the two variables appearing in the transformation equation (35), namely the conditional expectation of monthly logarithmic variance rv_{t+22}^{m} displayed in panel (a) and its sample variance shown in panel (b). This representation allows us to contrast the two modelling approaches outlined in the paper.

Figure 14: Log to level transformation components



Notes: 14(a) compares out of sample forecasts of rv_{t+22}^{m} in turquoise to fitted values of logarithmic models utilizing full sample in red. 14(b) plots the sample variances of logarithmic estimates generated by the out of sample regressions (in turquoise) and the full sample estimation (in red).

The logarithmic estimates of the monthly realized variance generated by our full sample models are generally relatively close to the out-of-sample forecasts which are a product of recursive regressions. Nevertheless, some of the exceptions when the estimates of rv_{t+22}^{m} for particular models differ substantially across the applications appear during the Covid-19 peak when the two considered models produce forecasts with maxima that are 0.35 and 0.26 higher than the logarithmic estimates yielded by estimation of corresponding models utilizing full sample. All of these values are far in the right tail of the distribution, breaching the lognormality assumption of the corollary (41). Another notable cause of the augmented overprediction during the Covid-19 extreme period by the forecasting procedure for log L-HAR-RV and log VIX-L-HAR-RV, is the higher sample variance of rv_{t+22}^{m} applied in the transformation. Panel (b) in Figure 14 showcases that the sample variance of rv_{t+22}^{m} for out-of-sample estimates overtakes the constant applied when transforming the full sample estimates to level at the time coinciding with the downturn caused by Covid-19 pandemic. Since the logarithmic estimates are converted to level with an exponential function, the transformation is extremely sensitive to further expansion of the distribution's right tails. We demonstrate this effect graphically in Figure A.4 of Appendix 6.

Table 7: Model Confidence Set procedure results

	T_{max}			T_R			
	Rank by T_{max}	t_i	p-value	Rank by T_R	$t_{i,j}$	p-value	MSE
HAR-RV loss	1	-1.168	1.000	3	0.564	0.961	1503.087
L-HAR-RV loss	2	-1.159	1.000	1	-0.173	1.000	1469.808
VIX-L-HAR-RV loss	3	-1.157	1.000	6	1.306	0.535	1471.826
log HAR-RV loss	4	-1.100	1.000	2	0.173	1.000	1479.241
log L-HAR-RV loss	7	0.995	0.575	4	1.036	0.826	4036.577
log VIX-L-HAR-RV loss	5	0.804	0.681	5	1.065	0.826	2341.869
Martingale loss	6	0.819	0.678	7	1.319	0.502	2517.839

Notes: MCS procedure was conducted at 20% significance level choosing Mean Squared Error (MSE) as the evaluation criterion. Table 7 reports the model´s rank, t statistics and corresponding p-values for both $T_{R,M}$ and $T_{max,M}$ test statistics. The evaluated forecasts were for period spanning from the 12th of October 2015 to 12th of February 2021. Bootstrapping was done using block length equal to the maximum number of significant parameters in the AR (p) process on all d_{ij} (default parameter in Catania & Bernardi, 2017) and 5,000 replications.

Table 7 presents the results of the MCS procedure. A significance level of 20% was chosen in our application meaning that our group of SSM includes the best model with 80% confidence. The null hypothesis of EPA $H_{0,M}$ was tested utilizing the test statistic $T_{max,M}$. Block bootstrap method with block length p dictated by the maximum number of significant parameters in the AR(p) process on all d_{ij} and 5,000 replications were used to derive the asymptotic distributions of test statistics and variance terms of the relative loss statistics. MCS failed to reject any of the models from the group of SSM. This is likely linked to the limited information content of the data. The nature of the procedure is such that the models remain in the set until they are proven to be significantly inferior to the surviving ones. Thereby as cautioned by the authors, while the MCS contains only the best models asymptotically, it might still contain several poor models in finite samples. Acknowledging this feature of the procedure, we will not select the best models solely based on the results of MCS but we will also take into account other factors such as economic interpretation and other forecast accuracy measures.

It is important to add that Hansen et al. (2011) recommend estimating the competing models using rolling rather than expanding window to avoid cases when the main assumptions of MCS are not satisfied[20]. Inspired by the approach of Amendola et al. (2020) we address this by reporting the results of MCS procedure considering models estimated with non-expanding window scheme in Appendix 7 to showcase that our empirical results are robust to the choice of the estimation window.

[20] Hansen et al. (2011) alternatively recommend to only estimate the coefficients once and keep them as fixed.

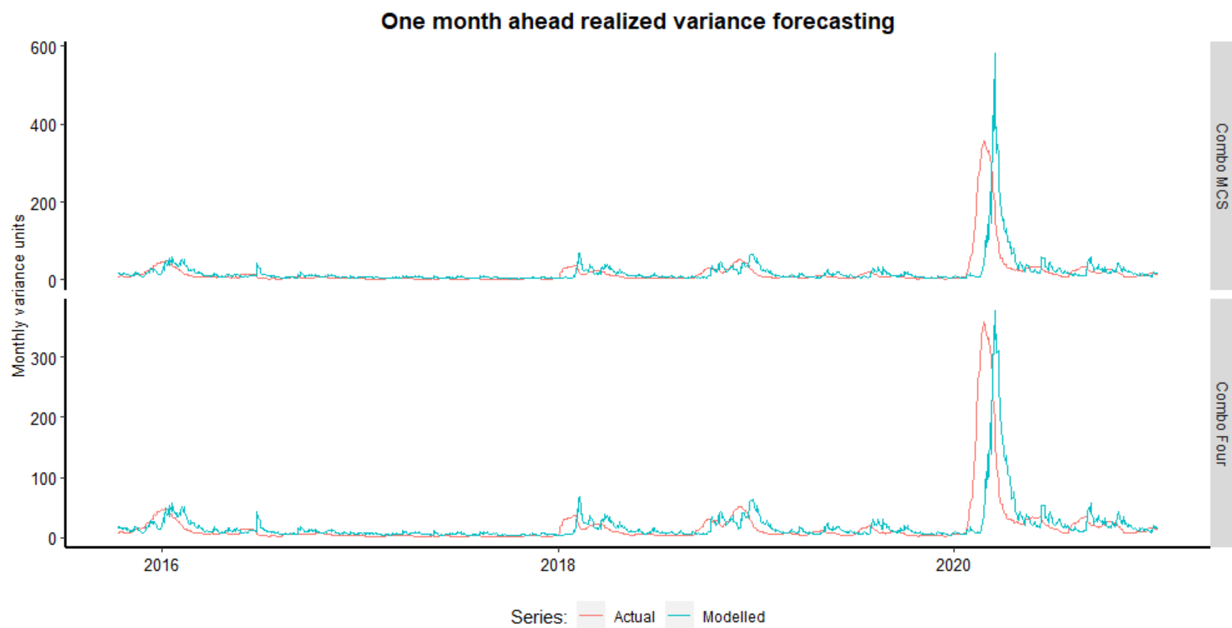
Table 8: Out-of-sample forecasting performance statistics

		ME	RMSE	MAE	MPE	MAPE	MSE
Estimated models	HAR-RV	-1.681	38.770	14.445	-115.834	132.809	1,503.087
	L-HAR-RV	-1.040	38.338	13.886	-70.592	93.018	1,469.808
	VIX-L.HAR-RV	-1.202	38.364	13.940	-71.710	93.798	1,471.826
	log HAR-RV	-1.374	38.461	14.087	-78.807	97.642	1,479.241
	log L-HAR-RV	-4.722	63.534	16.567	-83.793	101.630	4,036.577
	log VIX-L-HAR-RV	-4.438	48.393	15.871	-92.627	107.883	2,341.869
Non-estimated & combined models	Martingale	-0.049	50.178	16.318	-40.045	77.974	2,517.839
	Combo MCS	-1.870	40.499	14.345	-82.074	100.807	1,640.165
	Combo Four	-1.324	38.296	13.953	-84.236	102.851	1,466.595

Notes: The listed accuracy measures are for forecasts of the RV in the out of sample period spanning from the 12th of October 2015 to 12th of February 2021. Functional form of the estimated models is found in section (7.1) while the Martingale specification is defined in section (5). Combined forecasts are constructed using a simple equal weighting scheme.

Table 8 reports the common forecast accuracy measures for all estimated models, two of their linear combinations and a martingale model of Bollerslev et al. (2009). L-HAR-RV achieves the best performance among the estimated models across all statistics edging VIX-L-HAR-RV and log HAR-RV which performed only marginally worse. A slightly more pronounced drop-off in the forecasting accuracy is observed for Corsi's basic model HAR-RV with ME, MPE and MAPE increasing by 62%, 64% and 43% relative to the best performing estimated model. The two logarithmic models containing leverage effects, log L-HAR-RV and log VIX-L-HAR-RV delivered by far the worst results reaching the highest absolute values for all reported measures outside of MPE and MAPE. Most notably, the MSE for the two models increased in comparison to L-HAR-RV by 179% and 59%, respectively. The poor statistics can to a large extent be explained by the extensive overprediction of realized variance levels during the Covid-19 period discussed at length in the last paragraph. As the overprediction intensifies the problem of negative variance risk premium which as argued, does not have any reasonable economic explanation we decided to construct another forecast combination by equally weighting the set of forecasts excluding the ones generated by log L-HAR-RV and log VIX-L-HAR-RV and compare it to the linear combination of all models not rejected by the MCS procedure. This new combination of forecasts from the “restricted” set of models is referred to as Combo Four and the original combined forecast averaging all estimated models is termed Combo MCS.

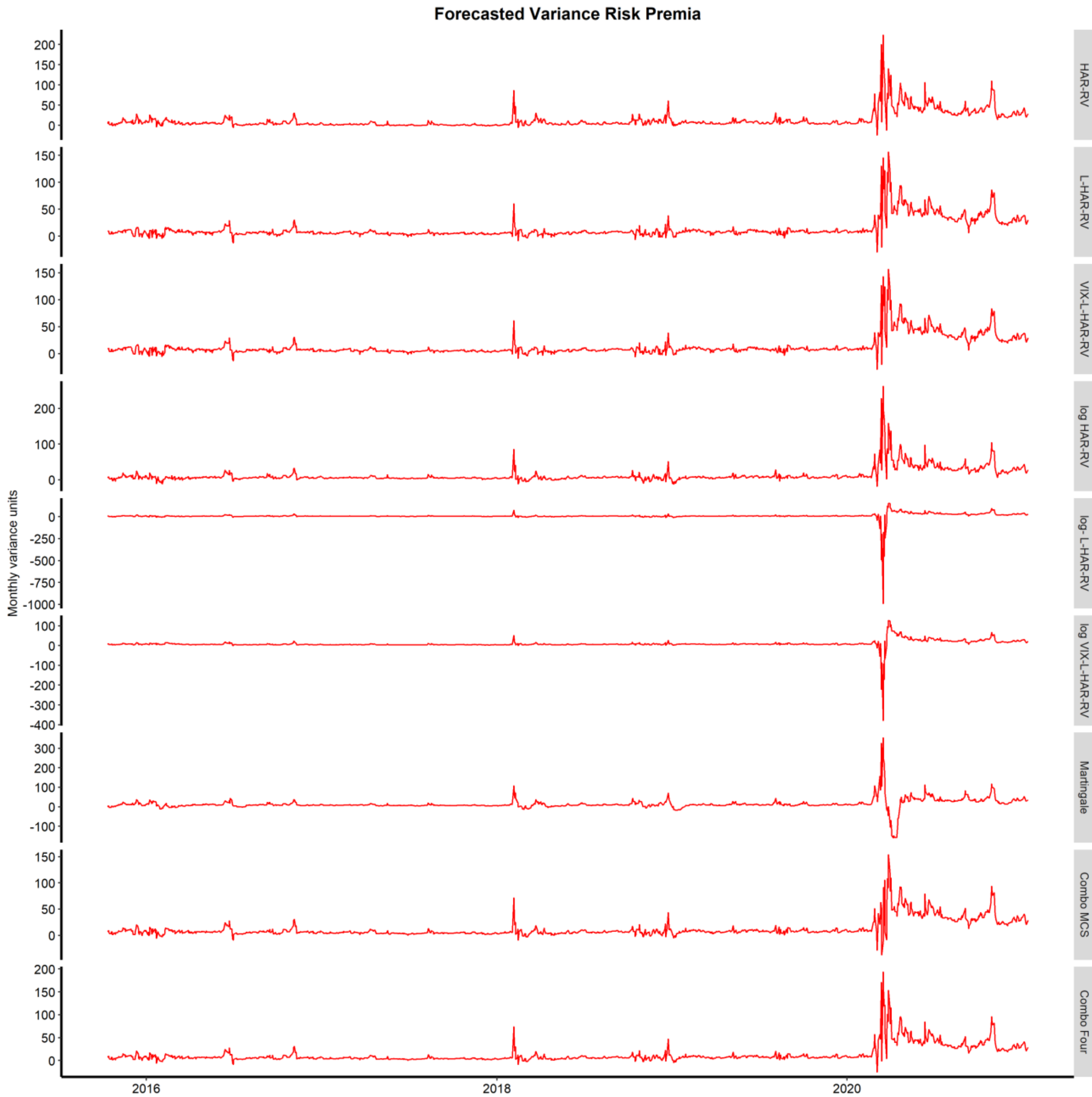
Figure 15: Forecast Combinations



Notes: Combined forecasts of the RV (turquoise) in the out of sample period spanning from the 12th of October 2015 to 12th of February 2021 compared to the actual RV levels (red). Combo MCS vector is generated by summing the products of a constant $\frac{1}{6}$ and the output of each of the estimated models. Combo Four construction is an analogue of the above utilizing only the 4 estimated models (HAR-RV, L-HAR-RV, VIX-L-HAR-RV and log HAR-RV).

Figure 15 visualizes the forecasts produced by the two combination schemes. As expected, the Combo Four alleviates the overprediction during the Covid-19 period which should be reflected in the forecast accuracy measures. Indeed, the Combo Four beats its Combo MCS counterpart in four out of six considered statistics and above all achieves the minimum MSE across all models and combinations. The martingale model, added only for benchmarking purposes generates the lowest ME, MPE and MAPE. Nevertheless, it also produces the second worst MSE and major negative variance risk premium as showcased in Figure 16.

Figure 16: Projected VRPs out of sample



Notes: Daily VRP series for the out of sample period (12th of October 2015 to 12th of February 2021) constructed as the difference between the monthly squared VIX quantity and the forecasted level of RV. The first 6 facets show the VRPs corresponding to estimated models, the 7 th facet visualizes VRPs generated utilizing the martingale model and the last two facets plot VRPs derived using linear combinations of estimated RV models.

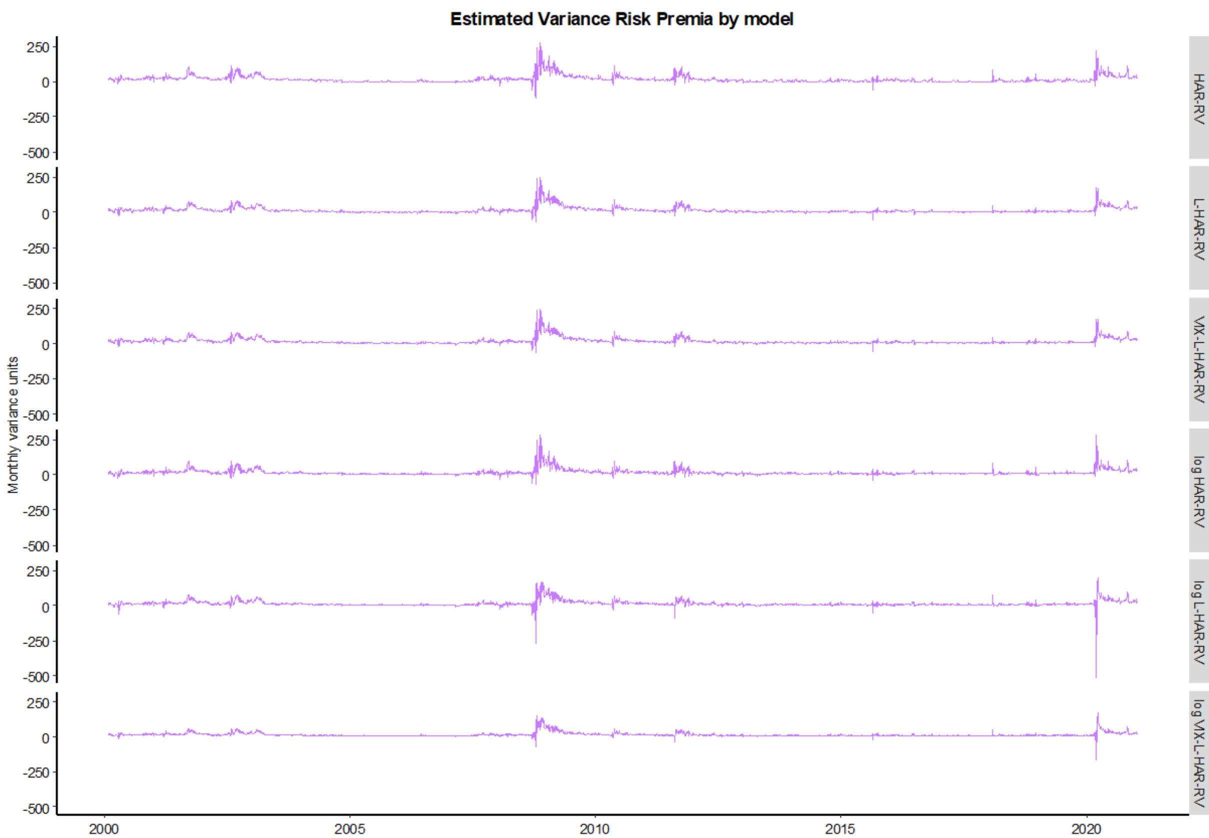
10. Stock Return Predictability

In the previous sections we have outlined different models to obtain the variance risk premium and analyzed their relative forecasting performance. This section builds on the theory and findings of the previous sections and investigates whether the variance risk premium is a predicator for future stock returns.

10.1 Risk Aversion and Expected Returns

Market participants are exposed to at least two types of risks namely, return variance and the variance of return variance. The uncertainty about future variance exposes investors to variance risk. To recap, in section (5), we discussed that a premium on variance risk can be theoretically derived from a variance swap whereby the premium is analogue to an insurance investor are willing to pay to shield against future volatility. Bakshi and Madan (2006) argue that rational risk-adverse market investors dislike extreme loss states and prefer to purchase protection to cover their exposure. They further note that the preference to shield against losses loads probability mass to the risk-neutral tail probabilities resulting in a spread between the risk-neutral and physical probabilities ultimately giving rise to the variance risk premium. As we have empirically demonstrated, the variance risk premium can be obtained from the VIX index by subtracting the forecasted realized variance which cleans the objective variation from the risk-neural variation. Bekaert et al. (2014) argue that the risk-neural variation incorporates probabilities that adjust for the pricing of risk. The variance risk premium is thus often referred to as a measure of aggregate market risk aversion and/or economic uncertainty (Bollerslev et al., 2014, Bekaert et al., 2013). In the same vein, Bollerslev et al. (2011) theoretically demonstrate that within an intertemporal asset pricing model a representative agent's constant relative risk aversion is proportional to the variance risk premium. Overall, relating the variance risk premium to risk aversion has found strong support. Our results appear to corroborate these claims. It is apparent from Figure 17 that the variance risk premium spikes during times of crisis and high uncertainty, most notably during the 2009 financial crisis aftermath and the beginning of the Covid-19 pandemic.

Figure 17: VRPs constructed utilizing RV estimated over full sample



Notes: Variance risk premium one-month ahead estimations based on different realized variance forecasting models. Estimations are based on the full sample.

A high variance risk premium thus implies a high degree of aggregate risk aversion in the market. Intuitively, it can be argued that during times of high-risk aversion (i.e. high variance risk premium), market participants reduce their investments and shift their holdings towards less risky assets. In turn, as risk aversion increases in the economy, investors require higher expected returns (Bollerslev et al., 2009).

These insights have led to a large body of research examining the link between the variance risk premium and stock return predictability (e.g. Bekaert and Hoerova, 2014; Bollerslev et al., 2009). The variance risk premium may be a priced risk factor that could have predictive powers for stock returns over relatively short horizons. However, the degree of predictability varies greatly between studies. We aim to put our variance risk premium models to the test and determine whether they succeed to predict stock returns for different (short) horizons.

10.2 Stock Return Predictability

To determine if the variance risk premium is a predictor for future stock returns, we run simple scaled monthly regressions following the procedure of Bollerslev et al. (2009):

$$\frac{1}{h}\sum_{j=1}^{h} r_{t+j}^{ex} = \alpha_h + \beta_h \psi_t + \varepsilon_{t+h} \tag{66}$$

The regressor ψ_t represents the variance risk premium estimated based on our different models in section (9). Specifically, we run univariate regressions for the best performing estimation models as well as the combined forecast model and the martingale model. The dependent variable r_{t+j}^{exc} denotes the excess return expressed in percentages, which is the difference between the logarithmic monthly returns of the S&P 500 and the three-month treasury bills. The treasury bill percentage rates r_t^{3m} are retrieved from the Federal Reserve Economic Data (FRED) database of the St. Louis Federal Reserve Bank. The excess return is given by:

$$r_{t+j}^{exc} = \left(\ln\left(\frac{S_T}{S_t}\right) - \frac{\ln\left(1 + \frac{r_t^{3m}}{100}\right)}{12} \right) * 100 \tag{67}$$

where $\ln\left(\frac{S_T}{S_t}\right)$ is the one-month logarithmic return of the S&P 500 and the second term, $\frac{\ln\left(1+\frac{r_t^{3m}}{100}\right)}{12}$, provides the continuously compounded three-month treasury rate which is divided by 12 to match the monthly horizon of the returns. As equation (66) indicates, the monthly excess returns are summed and averaged by a factor h to get the h-month excess returns. To align with other papers, we rely on end-of month observations resulting in 239 observations over the full sample. This approach is not without controversy since the overlapping return data may create autocorrelation issues that could bias results (Britten-Jones & Neuberger, 2011). To mitigate this issue, we employ Newey West standard errors with increasing lags relative to the forecasting horizon, such that lag $L = \max\{3, 2 * h\}$[21].

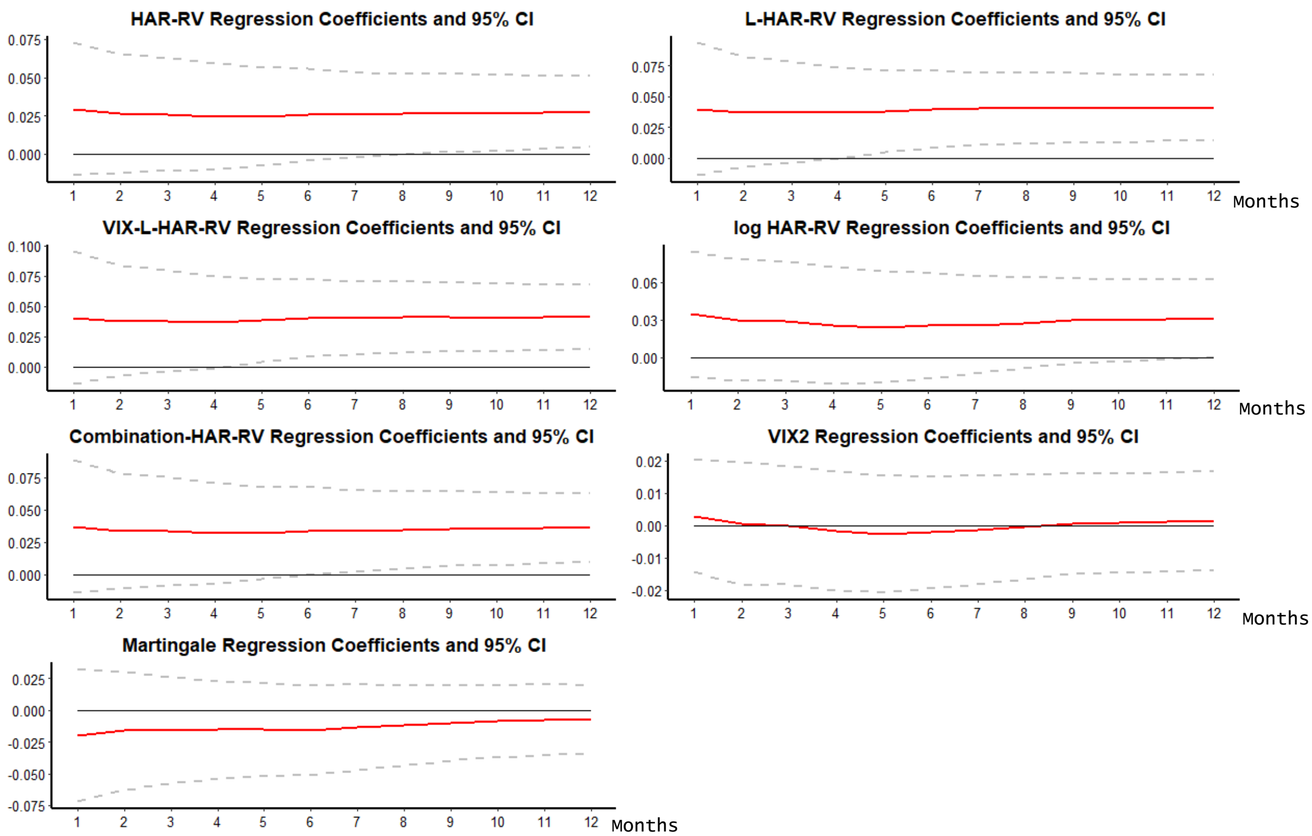
As mentioned, previous research has found that the variance risk premium has the highest predictive power over short monthly horizons as opposed to multi-year horizons. Therefore, we consider the time horizons ranging from one month to twelve months in the forecasting regressions. The quest for short horizon return predictability variables is economically relevant since trading strategies can be implemented more frequently ultimately generating high annual returns (Timmermann, 2008). We also test whether the variance risk premium can predict stock returns of the S&P 500 better than the squared VIX. Banerjee et al. (2007) and Giot (2005) find that squared VIX levels have predictive power over relatively short horizons especially during times of market turbulence. To this end, we construct a univariate regression substituting ψ_t in equation (66) for VIX-squared and contrast the R-square to the variance risk premium regression output.

[21] Similar to Bekaert and Hoerova, 2014.

10.3 Univariate Regression Output

The univariate regression coefficients at the monthly horizons are depicted in Figure (18). The dashed lines represent the 95% confidence bands based on Newey-West standard errors. If both dashed lines are above (or below) the solid horizontal line the coefficients are significant at the 95% confidence level. Thus, the univariate regression outputs based on the log HAR-RV model, VIX^2 and Martingale are never significant at the 5% significance level for all horizons. The variance risk premium estimated from the L-HAR-RV model and VIX-L-HAR-RV showcase significant coefficients from the fourth month onwards. The coefficient of the variance risk premium based on the baseline HAR-RV model turns significant at the 9-month horizon and remains significant for the following three months. The coefficients of the combination model are significant as of the sixth month. Overall, similar to Bekaert and Hoerova (2014), we observe that our estimation models do not yield significant coefficients over short horizons (i.e. between 1 and 3 months) in a univariate regression forecasting exercise.The variance risk premiums based on the HAR specification all have the expected positive coefficient sign for all horizons. This implies that an increase in variance risk premium leads to an increase in the future excess return of the S&P 500, as the theory suggests. The VIX^2 coefficients are roughly zero for all time horizons. Surprisingly, the Martingale model carries a negative sign for all maturities unlike a priori expectations. This observation contrasts results of other studies that find that simple non-estimation models generally perform comparable to more complex variance risk premium estimation models.

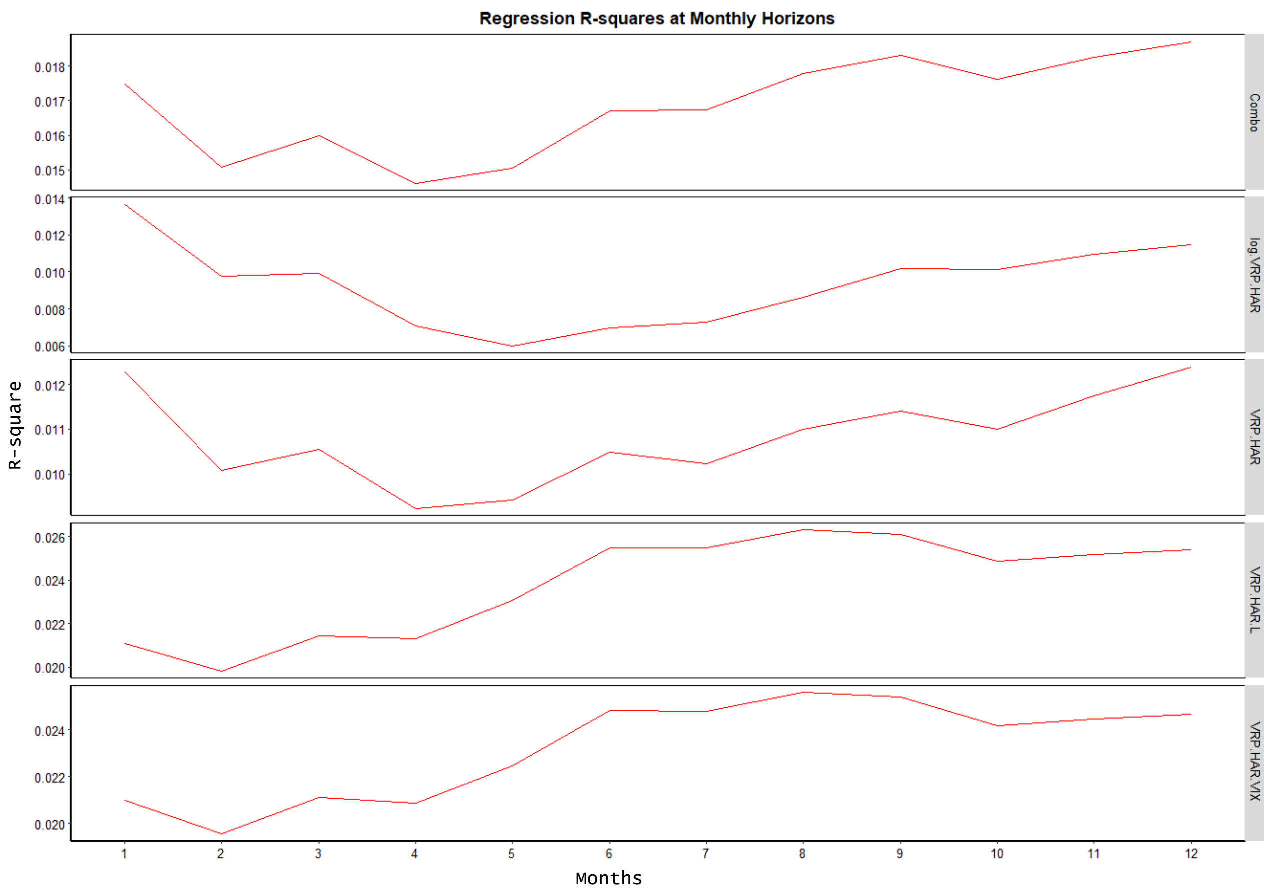
Figure 18: Regression Coefficients and confidence interval bands



Notes: The red line in the graphs display the univariate regression coefficients for each of the 12 months considered in the stock return predictability analysis. The grey dotted lines represent the confidence bands at the 5% significance level.

Figure (19) summarizes the R^2 for the individual regressions ranging from one to twelve months. Figure (19) only includes models that show significant coefficients during at least one of the considered time horizons. The best result is achieved with the HAR extension including leverage effects (i.e. L-HAR-RV) which steadily increases from the second month onwards and peaks at the 8 months horizon with an adjusted R^2 of approximately 2.6% (and an unadjusted R^2 of 3%). The model including the VIX component (i.e. VIX-L.HAR-RV) performs practically identical. The basic HAR-RV model shows comparatively poor predictability results with adjusted R^2's never exceeding 1.2% and is thus outperformed by the extended HAR models for all time horizons. The logarithmically transformed HAR-RV does not display improved results compared to its counterpart level model. The combination model averages the results of the four models and positions between the best performing models and worst performing models. The VIX^2 levels and the Martingale model do not have predictive power for any of the 12 monthly horizons.

Figure 19: Regression R-squares



Notes: The red line graphs the R-square for each of the regressions at every time horizon in the stock return predictability analysis. Only models that have significant coefficients for at least one of the horizons under consideration are displayed.

Arguably, even for the best performing model, the explanatory power of the variance risk premium is relatively small. Bekaert and Hoerova (2014) reported R^2's as high as 4,5% at the 12-month horizon for a model similar to our L-HAR-RV model. More general, comparable empirical studies have documented that the variance risk premium can predict between 2% and 7% of U.S. excess return, generating the highest predictability in the first 12 months and declining thereafter (e.g. Bekaert and Hoerova, 2014; Bollerslev et al., 2009). We find that our best performing models show the greatest degree of predictability at the 8-month horizon (Table 9)[22].

Table 9: Univariate stock return predictability regression output for 8-month horizon

Stock Return Regressions 8-month Horizon

	Dependent variable:						
	r^{ex}_{t+8}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HAR-RV	0.026^{*}						
	(0.014)						
L-HAR-RV		0.041^{***}					
		(0.015)					
VIX-L-HAR-RV			0.041^{***}				
			(0.015)				
log HAR-RV				0.027^{*}			
				(0.016)			
Four Combination					0.035^{**}		
					(0.015)		
VIX^2						-0.001	
						(0.006)	
Martingale							-0.012
							(0.014)
Constant	-0.112	-0.306	-0.308	-0.090	-0.219	0.306	0.466
	(0.355)	(0.359)	(0.361)	(0.359)	(0.361)	(0.385)	(0.362)
Observations	239	239	239	239	239	239	239
R^2	0.015	0.030	0.030	0.013	0.022	0.00003	0.003
Adjusted R^2	0.011	0.026	0.026	0.009	0.018	-0.004	-0.001
Residual Std. Error (df = 237)	4.449	4.415	4.416	4.455	4.434	4.483	4.477
F Statistic (df = 1; 237)	3.646^{*}	7.434^{***}	7.253^{***}	3.073^{*}	5.309^{**}	0.007	0.692
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$						

Notes: All univariate regressions are based on monthly observations. The standard errors in brackets are constructed using Newey-West lags which increase with the forecasting horizon. Significance of coefficients is indicated by the star symbol *,**,***

[22] The univariate regression output for the monthly, quarterly, semi-annually, and annually regressions can be found in Appendix 8.

Although some studies have reported remarkable predictability, as such, Bollerslev et al. (2011) find that the variance risk premium predicts close to 16% in the excess return variation at the quarterly horizon. The varying degree of predictability is likely driven by differences in sample characteristics and discrepancies in regression construction. As such, results between studies are not fully comparable since sample horizons greatly differ with some studies using data ranging back to the early 90s. Furthermore, there is no consensus on the excess return data used in the regression analysis. Bollerslev et al. (2011) used non-overlapping quarterly returns leaving them with only 54 observations which may explain their high predictability. We use overlapping data, as the majority of variance risk premium studies do, giving us 4.5 times more observations than Bollerslev et al. (2011).

One may argue that our predictability exercise suffers from look-ahead bias. That is, since we use the full sample variance risk premium estimations, we inevitably rely on forward looking data. The coarse monthly excess return data on which we regress the variance risk premium estimations does not allow for a fully-fledged out-of-sample analysis since we would be left with a small number of out-of-sample observations that would not allow for a meaningful inference. However, for our intents and purposes in-sample predictability is not necessarily misplaced and does not discredit our findings. Afterall, we are interested in the return predictability of the variance risk premium that exists on average in our sample. Rapach and Wohar (2006) argue that for such inferences in-sample return predictability is conceptually better suited than out-of-sample procedures.

10.4 Robustness Check in Multivariate Regressions

Our univariate regression output shows that the variance risk premium has statistical significance for return predictability, but the single variable models explain only a small part of the variation in excess returns. Ang and Bekaert (2006) have argued that, in general, univariate regressions tend to be misspecified, leaving the majority of the components in return predictability unexplained. The goal is to determine whether the predictive power of the standalone variance risk premium could be the results of omitted economic variables or if it can complement other predictor variables. We consider three economic predictor variables as popularized by Fama and French (1989) in a combined multivariate analysis. The variables under consideration are, first, the dividend yield (DY) defined as the logarithmic difference between dividends and prices of the S&P 500, where dividends are summed over the past 12 months on a moving sum basis. Second, the default yield spread (CS) derived from the difference between Moody's BAA and AAA rated long-term bond yields. Lastly, the term spread (TS) extracted from the difference between ten-year and three months T-bill yields. The first variable is obtained from Robert J. Shiller's database while default and term spreads are retrieved from the Federal Reserve Economic Data (FRED) database of the St. Louis Federal Reserve Bank. All variables are monthly observations and span over the same time horizon than the S&P 500 stock returns considered in the previous subsection. We only consider our best performing variance risk premia model from the univariate analysis (i.e. L-HAR-RV) in the multivariate exercise.

Table 10: Summary statistics for other predictor variables

Summary Statistics	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
CS	239	1.055	0.428	0.550	0.825	1.180	3.380
TS	239	1.681	1.177	-0.770	0.705	2.625	3.790
DY	239	-0.794	0.085	-0.995	-0.836	-0.764	-0.468

Notes: Summary statistics of the three additional predictor variables namely, the default spread, term spread, and dividend yield denoted CS, TS and DY, respectively

Empirical studies have shown that these economic variables can predict returns (e.g. Schwert, 1990; Chen, 1991; Boons, 2016). Fama (1990) argues that the dividend yield, default spread and term spread can capture the variation in expected returns based on business conditions. Boons (2016) shows that those variables can predict macroeconomic activity and are priced risk factor. Thus, similar to the variance risk premium, these economic variables are related to wider macroeconomic changes.

Table 11: Multivariate regression output including common predictor variables

Multivariate Regressions

	Dependent variable:				
	r^{ex}_{t+1}	r^{ex}_{t+3}	r^{ex}_{t+6}	r^{ex}_{t+8}	r^{ex}_{t+12}
	(1)	(2)	(3)	(4)	(5)
L-HAR-RV	0.106^{**}	0.101^{***}	0.106^{***}	0.107^{***}	0.107^{***}
	(0.043)	(0.038)	(0.039)	(0.039)	(0.040)
log(DY)	17.512^{***}	18.214^{***}	19.556^{***}	20.268^{***}	21.016^{***}
	(4.640)	(4.490)	(4.828)	(4.758)	(4.680)
DS	-5.892^{***}	-5.797^{***}	-6.058^{***}	-6.148^{***}	-6.204^{***}
	(2.077)	(2.191)	(2.329)	(2.336)	(2.384)
TS	-0.269	-0.302	-0.359	-0.376	-0.373
	(0.252)	(0.246)	(0.251)	(0.249)	(0.251)
Constant	19.337^{***}	19.908^{***}	21.272^{***}	21.949^{***}	22.609^{***}
	(4.674)	(4.687)	(5.169)	(5.149)	(5.157)
Observations	239	239	239	239	239
R^2	0.110	0.121	0.140	0.144	0.142
Adjusted R^2	0.095	0.106	0.125	0.129	0.127
Residual Std. Error (df = 234)	4.563	4.263	4.152	4.175	4.254
F Statistic (df = 4; 234)	7.253^{***}	8.083^{***}	9.525^{***}	9.827^{***}	9.692^{***}
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: The table reports the output for the multivariate regressions at different time horizons which include standard predictor variables. All multivariate regressions are based on monthly observations. The standard errors in brackets are constructed using Newey-West lag which increase with the forecasting horizon. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme above.

Table 11 shows that when controlling for other common state variables, the variance risk premium estimate based on the L-HAR-RV model remains highly significant and thus is not crowded out. This implied that the variance risk premium incorporates a priced risk factor that is not captured by traditional economic predictors. Other predictor variables could have been included but it would be practically impossible to consider all of them. We therefore limited the selection to three intensively studied predictors. Overall, there appears to be significant evidence that variance risk premium is a predictor for stock returns although the magnitude of the predictability is rather small.

11. Discussion, Limitations and Future Research

At this point it is worthwhile to again reflect on the two complementing research questions we outlined in section (1). We also devote this section to address the limitations of our findings and discuss the challenges we encountered throughout this paper to provide a reference point for scholars who wish to further explore the realm of variance risk premia research.

The first question which we answered is how the variance risk premium can be empirically estimated. To this end, we proposed different models based on an additive cascade partial volatility technique, as popularized by Corsi (2004), whereby we relied on the information content comprised in realized variance at different horizons (i.e. daily, weekly and monthly). We extended the base HAR-RV model, to account for leverage effects which is a well-documented stylized fact of financial variance. We further composed a model including a VIX term which appears to be correlated with movements in realized variance.

We noted that the realized variance times series in its level form displays fat tails deviating substantially from a normal distribution. Logarithmic transformations of the realized variance were much closer to being normally distributed. While the distributional properties of the logarithmic realized variance models appeared promising at first, we encountered a methodological challenge in transforming the logarithmic model results into level form. We ultimately decided to convert the log results under the convenient assumption of lognormality which is in contradiction to the right-skewed distribution of the estimates in our models causing biased converted estimates. As addressed in section (7.1)., alternative procedures exist but are much more burdensome in application.

Although the models in this paper have proven to explain large and statistically significant variations in future realized variance it has become apparent, we could have selected superior models given the dynamic properties of realized variance. As discussed, especially in times of extreme market turbulences, linear models have a difficult time to model the explosive dynamics of realized variance. The first limitation of our findings is thus that we possibly could have found more effective models if we had allowed for non-persistent jump components in our volatility models to deal with extreme spikes and quick mean reversion during crisis periods as observed during the financial crisis aftermath in 2009 and the Covid-19 selloff in 2020. Alternatively, one could use non-linear models which would likely perform better in times of crisis.

An intriguing approach to smooth out any biases in our estimation is to combine the estimation results. Under the premise that combining forecast estimates would result in a diversification effect we equally weighted our four best performing models. In the performance analysis, the diversification effects are evident, with the combination estimates often showing the best performance statistics. However, by relying on equal weights we likely did not fully leverage the potential benefits that can result from combining forecasts. A large literature exists that provides alternative weighting procedures which we did not consider but provide an interesting venue for future research.

The second research question which we answered in the present paper relates to the ability of variance risk premium to predict future stock returns. We did find statistically significant evidence at short (i.e. less than one year) horizons, although rather modest in magnitude. It appears that the variance risk premium complements commonly employed predictor variables. There are however a few important implications and limitations that are worthwhile to address.

In our predictability exercise we relied on monthly overlapping S&P 500 return data. This procedure is commonly observed in predictability regressions. Overlapping data allows to maintain a large data set even for sparse monthly observations. However, overlapping observations are not without controversy and are believed to bias the confidence intervals and create inflated R^2. We proceeded to mitigate the effects by employing heteroskedasticity and autocorrelation consistent Newey-West standard errors. Although, it is difficult to prove that we effectively crowded out the bias. Alternatively, one can use non-overlapping data for the predictive regressions, but this implies that the data set will be comparatively smaller. Given that we committed to monthly return data this was not a viable option for our purposes as it would have reduced the number of observations substantially which would have been insufficient to conduct any meaningful statistical inference. In hindsight, we could have used a narrower time frame (e.g. weekly) for our return data observations in the predictive regressions but this would have required major adjustment in the realized variance modelling section which was initially designed for monthly forecasts.

Lastly, in our stock return predictability section we relied on ex post information to determine whether the variance risk premium carries any predictability for future stock returns at short horizons. As we have already stressed, this exercise was primarily to determine whether a relationship exists on aggregate. Naturally, an investor can not benefit from hindsight in real time trading strategies. It is particularly for this reason that we refrain from suggesting any variance risk premium trading strategies based on ex post information. An interesting venture for future research would be to analyze real time trading strategies modelled exclusively based on ex ante information.

12. Conclusion

Throughout this paper we aimed to answer two research questions. First, we explored how to derive the variance risk premium. We introduced the underlying theoretical background on variance swaps and explained the relating concepts on risk-neutral and physical variance. To bridge the gap from theory to application we analyzed the respective variance proxies namely, the CBOE volatility index VIX and high frequency realized variance.

As the variance risk premium can be obtained from the difference of the squared VIX and the expectation of the one-month ahead realized variance, the empirical challenge resided in constructing an appropriate forecasting model. To this end, we relied on 20 years of high frequency realized variance data which includes numerous volatile periods such as the Dotcom bubble, the recent financial crisis and the Covid-19 pandemic. Importantly, realized variance data incorporates stylized facts such as volatility clustering and leverage effects, that need to be considered in selecting a fitting forecasting model. We ultimately decided to base our models on a heterogenous autoregressive framework, originally put forward by Corsi (2004). The HAR model accounts for distinct volatility movements created by heterogenous market participants and thus underlies a sound economic interpretation. Due to the distributional properties of the realized variance, we deemed it necessary to specify all models in their level and logarithmic form. For benchmark reasons we also considered a non-estimation model as well as a combination forecast to take advantage of possible 'diversification' benefits.

The logarithmic models showed superior model fit compared to their otherwise identical level models, with the log VIX-L-HAR-RV reaching close to 66% model fit in the full sample. However, as addressed in the Limitations section the output transformation from log to level posed important complications, deteriorating the model superiority due to large overpredictions over the actual realizations. Ultimately, we conclude the best out-of-sample forecasting performance was achieved by the L-HAR-RV model.

Generally, over the full sample the modelled variance risk premium behaved as the theory would suggest. We observed low or no variance risk premia in steady periods and high premium during turbulent periods most notably during the 2008 financial crisis and beginning of the 2020 covid crisis. However, all models create negative variance risk premia that do not have any rational economic interpretation. As already addressed in the Limitations, these occurrences are partly due to the linear models employed.

We also examined the abilities of the variance risk premium to predict future stock returns. We found support that the variance risk premium is a predictor for stock returns over short horizons (i.e. less than one year). Although significant, of the proportion of the variance in future returns explained by the variance risk premium is rather small. As such, the variance risk premium obtained from our best model, L-HAR-RV, generated the highest model fit at the 8-month prediction horizon with an R^2 of 3% in our univariate regression analysis. The weak performance is likely a result of omitted variables that influence future stock returns. In fact, in a multivariate regression analysis including other common predictor variables the results are much more impressive explaining up to 14.5% of the variation of the 8 month ahead stock returns.

13. References

Amendola, A., Braione, M., Candila, V., & Storti, G. (2020). A Model Confidence Set approach to the combination of multivariate volatility forecasts. *International Journal of Forecasting*, *36(3)*, 873–891. https://doi.org/10.1016/j.ijforecast.2019.10.001

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, 885-905.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, *89*(4), 701-720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica, 71(2)*, 579-625

Andersen, T. G., Fusari, N., & Todorov, V. (2015). The risk premia embedded in index options. *Journal of Financial Economics*, *117*(3), 558-584.

Andersen, T., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance, 4(2)*, 115–158. https://doi.org/10.1016/S0927-5398(97)00004-2

Andersen, T., & Bollerslev, T. (2018). *Volatility*. Edward Elgar Publishing Limited. https://doi.org/10.4337/9781788110624

Andersen, T., Bollerslev, T., Diebold, F., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics, 61(1)*, 43–76. https://doi.org/10.1016/S0304-405X(01)00055-1

Andersen, T., Bollerslev, T., Diebold, F., & Labys, P. (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association, 96(453)*, 42–55. https://doi.org/10.1198/016214501750332965

Andersen, T., Davis, R., Kreiss, J., & Mikosch, T. (2009). *Handbook of Financial Time Series* (1. Aufl.). Springer-Verlag.

Andreou, E., Kourtellos, A., & Ghysels, E. (2011). Forecasting with Mixed-Frequency Data. In *The Oxford Handbook of Economic Forecasting* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0009

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there?. *The Review of Financial Studies, 20(3)*, 651-707.

Audrino, F., & Knaus, S. D. (2016). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, *35*(8-10), 1485-1521.

Avramov, D., Chordia, T., & Goyal, A. (2006). The Impact of Trades on Daily Volatility. *The Review of Financial Studies, 19*(4), 1241–1277. https://doi.org/10.1093/rfs/hhj027

Baillie, R., Bollerslev, T., & Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 74*(1), 3–30. https://doi.org/10.1016/S0304-4076(95)01749-6

Bakshi, G., & Madan, D. (2006). A theory of volatility spreads. *Management Science*, *52*(12), 1945-1956.

Banerjee, P. S., Doran, J. S., & Peterson, D. R. (2007). Implied volatility and future portfolio returns. *Journal of Banking & Finance, 31*(10), 3183-3199.

Bates, J., & Granger, C. (1969). The Combination of Forecasts. *OR, 20*(4), 451-468. https://doi:10.2307/3008764

Bauwens, L. (2012). HAR Modeling for Realized Volatility Forecasting. In *Wiley Handbooks in Financial Engineering and Econometrics* (pp. 363-382). Hoboken, NJ, USA: John Wiley & Sons.

Bauwens, L., Hafner, C., & Laurent, S. (2012). *Handbook of volatility models and their applications*. John Wiley & Sons, Inc.

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of econometrics*, *183*(2), 181-192.

Bekaert, G., Hoerova, M., & Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, *60*(7), 771-788.

Black, F. (1976). "*Studies of Stock Price Volatility Changes"*, Proceedings of the Business and Economics Section of the American Statistical Association , 177–181.

Bollerslev, T., Gibson, M., & Zhou, H. (2011). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of econometrics, 160*(1), 235-245.

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. (2018). Risk Everywhere: Modeling and Managing Volatility. *The Review of Financial Studies, 31*(7), 2729–2773. https://doi.org/10.1093/rfs/hhy041

Bollerslev, T., Marrone, J., Xu, L., & Zhou, H. (2014). Stock return predictability and variance risk premia: statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, *49*(3), 633-661.

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, *22*(11), 4463-4492.

Boons, M. (2016). State variables, macroeconomic activity, and the cross section of individual stocks. *Journal of Financial Economics*, *119*(3), 489-511.

Bossu, S. (2014). *Advanced Equity Derivatives: Volatility and Correlation*. (1st ed., Wiley Finance Ser).

Boudoukh, J., Israel, R., & Richardson, M. (2019). Long-Horizon Predictability: A Cautionary Tale. *Financial Analysts Journal, 75*(1), 17–30. https://doi.org/10.1080/0015198X.2018.1547056

Breusch, T.S., & Pagan, A.R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica 47*, 1287–1294.

Britten-Jones, M., & Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The journal of Finance*, *55*(2), 839-866.

Britten-Jones, M., Neuberger, A., & Nolte, I. (2011). Improved inference in regression with overlapping observations. *Journal of Business Finance & Accounting*, *38*(5-6), 657-683.

Brockwell, P., & Davis, R. (2016). Time Series Models for Financial Data. In *Introduction to Time Series and Forecasting* (pp. 195–226). Springer International Publishing. https://doi.org/10.1007/978-3-319-29854-2_7

Busch, T., Christensen, B., & Nielsen, M. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics, 160*(1), 48–57. https://doi.org/10.1016/j.jeconom.2010.03.014

Canina, L., & Figlewski, S. (1993). The informational content of implied volatility. *The Review of Financial Studies, 6*(3), 659-681.

Carr, P., & Wu, L. (2006). A tale of two indices. *The Journal of Derivatives, 13*(3), 13-29.

Carr, P., & Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, *22*(3), 1311-1341.

Catania, L., & Bernardi, M. (2017). MCS: Model Confidence Set Procedure. *R package version 5.2.2*. https://CRAN.R-project.org/package=MCS

Cboe (2019). *Cboe VIX: White Paper Cboe Volatility Index.* Cboe. https://cdn.cboe.com/resources/vix/vixwhite.pdf

Çelik, S., & Ergin, H. (2014). Volatility forecasting using high frequency data: Evidence from stock markets. *Economic modelling*, *36*, 176-190.

Chen, N. F. (1991). Financial investment opportunities and the macroeconomy. *The Journal of Finance*, *46*(2), 529-554.

Christensen, B., & Prabhala, N. (1998). The relation between implied and realized volatility. *Journal of Financial Economics, 50*(2), 125–150. https://doi.org/10.1016/S0304-405X(98)00034-8

Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics, 10*(4), 407–432. https://doi.org/10.1016/0304-405X(82)90018-6

Claeskens, G., Magnus, J., Vasnev, A., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting, 32*(3), 754–762. https://doi.org/10.1016/j.ijforecast.2015.12.005

Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Corsi, F. (2004). A Simple Long Memory Model of Realized Volatility. Available at SSRN: https://ssrn.com/abstract=626064 or http://dx.doi.org/10.2139/ssrn.626064

Corsi, F., Audrino, F., & Renó, R. (2012). HAR modeling for realized volatility forecasting. In: *Handbook of Volatility Models and Their Applications*. (pp. 363-382). New Jersey, USA: John Wiley & Sons, Inc. ISBN 9780470872512

Corsi, F., Barone-Adesi, G., & Audrino, F. (2005). *Measuring and Modelling Realized Volatility: from Tick-by-tick to Long Memory*. University of Lugano.

De Weert, F. (2008). *Exotic Options Trading*. (1st ed., The Wiley Finance Ser).

Degiannakis, S. (2015). *Modelling and Forecasting High Frequency Financial Data.* Palgrave Macmillan UK. https://doi.org/10.1057/9781137396495

Degiannakis, S., & Floros, C. (2013). Modeling CAC40 volatility using ultra-high frequency data. Research in *International Business and Finance, 28*, 68-81.

Dennis, P., Mayhew, S., & Stivers, C. (2006). Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon. *Journal of Financial and Quantitative Analysis*

Derman, E., & Miller, M. B. (2016). *The volatility smile: an introduction for students and practitioners*. Wiley.

Diebold, F. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting, 5*(4), 589–592. https://doi.org/10.1016/0169-2070(89)90014-9

Dodge, Y. (2008). *The Concise Encyclopedia of Statistics.* Springer New York. https://doi.org/10.1007/978-0-387-32833-1

Drechsler, I., & Yaron, A. (2011). What's vol got to do with it. *The Review of Financial Studies*, *24*(1), 1-45.

Driessen, J., Maenhout, P., & Vilkov, G. (2009). The Price of Correlation Risk: Evidence from Equity Options. *Journal of Finance, 64*(3), 1377-1406.

Duan, J. C., & Yeh, C. Y. (2010). Jump and volatility risk premiums implied by VIX. *Journal of Economic Dynamics and Control*, *34*(11), 2232-2244.

Duffee, G. R. (1995*)*. Stock returns and volatility a firm-level analysis. *Journal of Financial Economics, 37*, 399–420.10.1016/0304-405X(94)00801-7

Edwards, T. & Preston, H. (2017). *A Practitioner's Guide to Reading VIX*. S&P Dow Jones Indices. https://www.spglobal.com/spdji/en/education-a-practitioners-guide-to-reading-vix.pdf

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 987-1007.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews, 5*(1), 1-50.

Ericsson, J., Huang, X., & Mazzotta, S. (2016). Leverage and asymmetric volatility: The firm-level evidence. *Journal of Empirical Finance, 38*, 1–21. https://doi.org/10.1016/j.jempfin.2016.02.008

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *The journal of finance*, *45*(4), 1089-1108.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of financial economics, 25*(1), 23-49.

Fassas, A. P., & Papadamou, S. (2018). Variance risk premium and equity returns. Research in *International Business and Finance, 46*, 462-470.

Fernandes, M., Medeiros, M., & Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance, 40*, 1–10. https://doi.org/10.1016/j.jbankfin.2013.11.004 (Was still a working paper when Corsi and Reno got inspired, 2009)

Figlewski, S. (2016). What goes into risk-neutral volatility? Empirical estimates of risk and subjective risk preferences. *The Journal of Portfolio Management*, *43*(1), 29-42.

Foroni, C., & Marcellino, M. G. (2013). *A survey of econometric methods for mixed-frequency data*. *Available at SSRN 2268912*.

Forsberg, L., & Ghysels, E. (2007). Why do absolute returns predict volatility so well?. *Journal of Financial Econometrics, 5*(1), 31-67.

Franke, J., Härdle, W., & Hafner, C. (2019). Long Memory Time Series. In *Statistics of Financial Markets* (pp. 321–342). Springer International Publishing. https://doi.org/10.1007/978-3-030-13751-9_14

Gallant, A.R., & Rossi, P.E., & Tauchen, G. (1992). Stock Prices and Volume. *The Review of Financial Studies, 5*(2), 199–242. https://doi.org/10.1093/rfs/5.2.199

Ghysels, E., & Valkanov, R. (2012). Forecasting volatility with MIDAS. *Handbook of volatility models and their applications*, 383-401.

Giot, P. (2005). Relationships between implied volatility indexes and stock index returns. *The Journal of Portfolio Management*, *31*(3), 92-100.

Graefe, A., Armstrong, J., Jones, R., & Cuzán, A. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting, 30*(1), 43–54. https://doi.org/10.1016/j.ijforecast.2013.02.005

Greene, W. (2014). *Econometric Analysis : Global Edition.* (7. ed.). Pearson Education UK.

Hansen, P. R., & Lunde, A. (2011). Forecasting volatility using high frequency data. *The Oxford Handbook of Economic Forecasting (Vol. 1)*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0020

Hansen, P., Lunde, A., & Nason, J. (2011). The Model Confidence Set. *Econometrica, 79*(2), 453–497. https://doi.org/10.3982/ECTA5771

Hasanhodzic, J., & Lo, A. W. (2011). Black's leverage effect is not due to leverage (Working paper). Cambridge, MA: Boston University and Massachusetts Institute of Technology. Retrieved from SSRN 1762363

Hatfield, G. (2009). A Note Regarding "Risk Neutral" and "Real World" Scenarios—Dispelling a Common Misperception. *Product Matters!* February 2009 volume, Schaumburg: Society of Actuaries.

Hilpisch, Y. (2015). *Derivatives analytics with Python: Data analysis, models, simulation, calibration and hedging* (Wiley finance series). Hoboken, New Jersey: John Wiley & Sons.

Hosking, J. (1981). Fractional Differencing. *Biometrika, 68*(1), 165-176. doi:10.2307/2335817 https://radhakrishna.typepad.com/rks_musings/2014/09/whats-wrong-with-vix.html

Jakobsen, J. S. (2018). *Modeling Financial Market Volatility: A Component Model Perspective*. Institut for Økonomi, Aarhus Universitet. ECON PhD Dissertations Nr. 2018-1

Jiang, G. J., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, *18*(4), 1305-1342.

Jorion, P. (1995). Predicting volatility in the foreign exchange market. *The Journal of Finance*, *50*(2), 507-528.

Kang, H. (1986). Unstable Weights in the Combination of Forecasts. *Management Science, 32*(6), 683–695. https://doi.org/10.1287/mnsc.32.6.683

Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory, 21*(6), 1130–1164. https://doi.org/10.1017/S0266466605050565

Kirchgässner, G., Wolters, J., & Hassler, U. (2012). *Introduction to modern time series analysis*. Springer Science & Business Media.

Knopf, P. M., & Teall, J. L. (2015). *Risk neutral pricing and financial mathematics: A primer*. Amsterdam: Academic Press.

Koenker, R. (1981), A Note on Studentizing a Test for Heteroscedasticity. *Journal of Econometrics 17*, 107–112.

LeBaron, B. (2001). Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance*, *1*(6), 621–631. https://doi.org/10.1088/1469-7688/1/6/304

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*(1), 293-311

Low, C. (2004). The fear and exuberance from implied volatility of S&P 100 index options. *The Journal of Business, 77*(3), 527-546

Mandelbrot B. B., (1983), *The Fractal Geometry of Nature*, W.H.Freeman and Company, New York

Marcellino, M. (2004). Forecast Pooling for European Macroeconomic Variables. *Oxford Bulletin of Economics and Statistics, 66*(1), 91–112. https://doi.org/10.1111/j.1468-0084.2004.00071.x

McAleer, M., & Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, *27*(1-3), 10-45.

Miller, M. (2013). Time Series Models. In *Mathematics and Statistics for Financial Risk Management* (pp. 215-236). Hoboken, NJ, USA: John Wiley & Sons.

Mills, T. C. (2015). *Time Series Econometrics: A Concise Introduction*. Springer.

Müller, U. & Dacorogna, M. & Dav, D. & Pictet, O. & Olsen, R.. (1997). *Fractals And Intrinsic Time - A* Challenge To Econometricians. On https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.197.2969&rep=rep1&type=pdf accessed on 31-07-2021

Müller, U. K. (2014). HAC Corrections for Strongly Autocorrelated Time Series. *Journal of Business & Economic Statistics, 32*(3), 311–322. https://doi.org/10.1080/07350015.2014.931238

Neuberger, A. (1994) The Log Contract. *Journal of Portfolio Management, 20*, 74-80. http://dx.doi.org/10.3905/jpm.1994.409478

Newey, W., & West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica, 55*(3), 703–708. https://doi.org/10.2307/1913610

Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics, 160*(1), 246–256. https://doi.org/10.1016/j.jeconom.2010.03.034

Poulsen, R. (2010). Expectations, Non-Linear Functions, and Log-Normal Random Variables. Accessed OnLognormals.pdf (ku.dk) on 01-06-2021

Rapach, D. E., & Wohar, M. E. (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance, 13*(2), 231-247.

Reschenhofer, E., Mangat, M. K., & Stark, T. (2020). Volatility forecasts, proxies and loss functions. *Journal of Empirical Finance, 59*, 133–153. https://doi.org/10.1016/j.jempfin.2020.09.006

Sanusi, M. S. (2017). Investigating the sources of Black's leverage effect in oil and gas stocks. *Cogent Economics & Finance, 5*(1), 1–.https://doi.org/10.1080/23322039.2017.1318812

Sarwar, G. (2012). Is VIX an investor fear gauge in BRIC equity markets?. *Journal of Multinational Financial Management*, *22*(3), 55-65.

Scharth, M., & Medeiros, M. (2009). Asymmetric effects and long memory in the volatility of Dow Jones stocks. *International Journal of Forecasting, 25*(2), 304–327. https://doi.org/10.1016/j.ijforecast.2009.01.008

Schwert, G. W. (1990). Stock returns and real activity: A century of evidence. *The Journal of Finance, 45*(4), 1237-1257.

Schwert, G. W. (1990). Stock volatility and the crash of'87. *The review of financial studies*, *3*(1), 77-102.

Schwert, G. W. (1990). Stock volatility and the crash of'87. *The review of financial studies*, *3*(1), 77-102.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425-442.

Smith, G. (2015). New Evidence on Sources of Leverage Effects in Individual Stocks. *The Financial Review (Buffalo, N.Y.), 50*(3), 331–340. https://doi.org/10.1111/fire.12069

Smith, J., & Wallis, K. (2009). A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics, 71*(3), 331–355. https://doi.org/10.1111/j.1468-0084.2008.00541.x

Stambaugh, R., Schwert, G., & French, K. (1987). Expected stock returns and volatility. *Journal of Financial Economics, 19*(1), 3–29. https://doi.org/10.1016/0304-405X(87)90026-2

Stock, J., & Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting, 23*(6), 405–430. https://doi.org/10.1002/for.928

Stock, J., & Watson, M. (2016). *Introduction to econometrics* (Updated third edition, Global edition.). Pearson.

Sun, Y., Phillips, P. C. B., & Jin, S. (2008). Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing. *Econometrica, 76*(1), 175–194. https://doi.org/10.1111/j.0012-9682.2008.00822.x

Timmermann, A. (2006). Chapter 4 Forecast Combinations. In *Handbook of Economic Forecasting* (Vol. 1, pp. 135–196). Elsevier B.V. https://doi.org/10.1016/S1574-0706(05)01004-9

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting, 24*(1), 1-18.

Whaley, R. E. (2009). Understanding the VIX. *The Journal of Portfolio Management*, *35*(3), 98-105.

Yang, Y. (2004). Combining forecasting procedures: some theoretical results. *Econometric Theory, 20*(1), 176–222. https://doi.org/10.1017/S0266466604201086

Zumbach, G. (2013). *Discrete Time Series, Processes, and Applications in Finance*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31742-2

14. Appendix

Appendix 1: Variance swap replication

Derivation by Bossu and Carr:

We have a log-contract whose payoff is -$\ln S_T \,/F$where F is the forward price. By applying Ito-Doeblin theorem to $\ln S_T$ we obtain.

$$\ln \frac{S_T}{S_0} = \int_0^T \frac{1}{S_t} dS_t - \frac{1}{2}\int_0^T \frac{1}{S_t^2}(dS_t)^2 = \int_0^T \frac{1}{S_t} dS_t - \frac{1}{2}\int_0^T \sigma_t^2 dt$$

σ_t is instant volatility of the underlying asset which is possibly stochastic. It can be shown that realized variance can be replicated as:

$$\int_0^T \sigma_t^2 dt = 2\int_0^T \frac{1}{S_t} dS_t - 2\ln \frac{S_T}{S_0}$$

When assumed that risk-neutral dynamics of the underlying price are of form:

$$dS_t = \ v_t S_t d_t + \ \sigma_t S_t dW_t$$

where v_t is the risk neutral drift, we obtain:

$$E\left(\int_0^T \frac{1}{S_t} dS_t\right) = E\left(\int_0^T v_t dt\right) = \ln \frac{F}{S_0}$$

Therefore, the fair value of variance is equal to the fair value of two log contracts:

$$E\left(\int_0^T \sigma_t^2 dt\right) = E\left(-2\ln \frac{S_T}{F}\right)$$

Log-contracts are not traded. However, any European payoff can be decomposed as a portfolio of calls and puts struck along a continuum of strike prices.

$$-\ln\frac{S_T}{F} = 1 - \frac{S_T}{F} + \int_0^F \frac{1}{K^2}\max(0, K - S_T)\,dK - \int_F^\infty \frac{1}{K^2}\max(0, S_T - K)\,dK$$

The fair value of annualized variance is then

$$\sigma_K^2 = \frac{2}{T}E\left(-\ln\frac{S_T}{F}\right) = \frac{2e^{rT}}{T}\left[\int_0^F \frac{1}{K^2}p(K)dK + \int_F^\infty \frac{1}{K^2}c(K)dK\right]$$

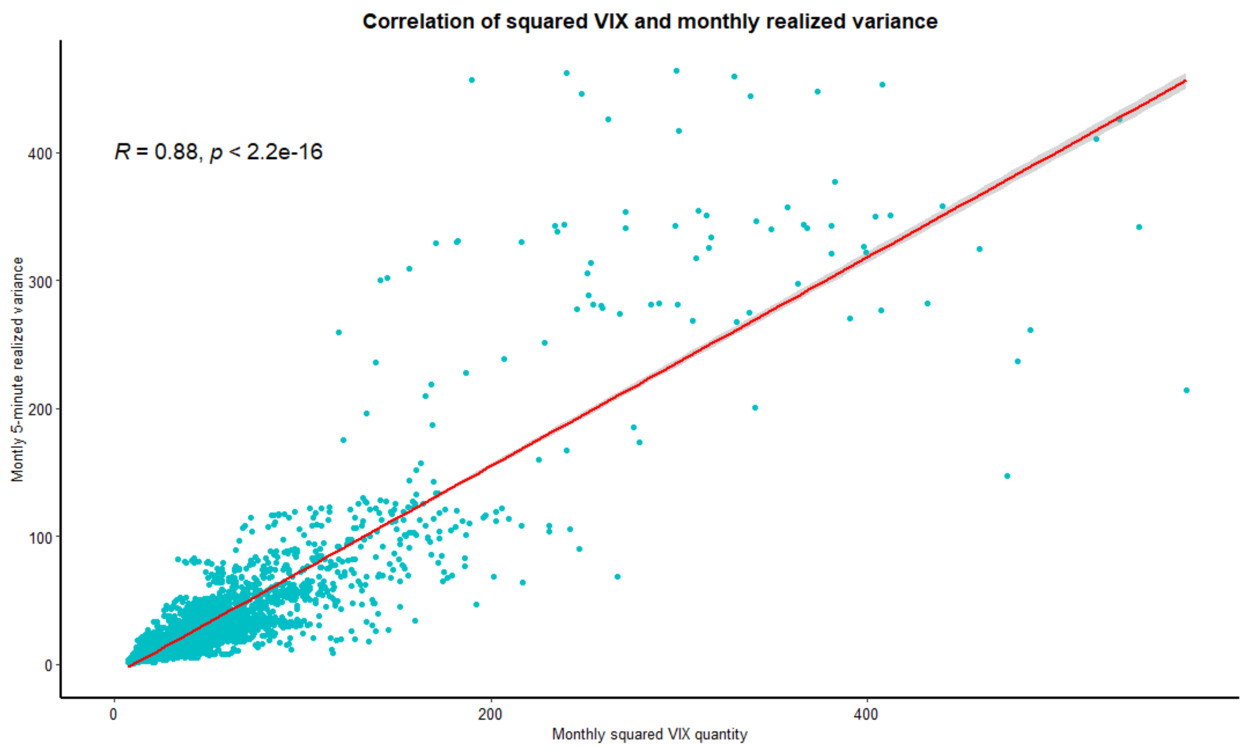
Where r is the interest rate corresponding to maturity T, $p(K)$ is the price of the put struck at K and $c(K)$ is the price of the call struck at K.

In the real world, only a finite number of strikes are traded which is accounted for in the proxy formula:

$$\sigma_K^2 \approx \frac{2e^{rT}}{T}\left[\sum_{i=1}^{n}\frac{p(K_i)}{K_i^2}\Delta K_i + \sum_{i=n+1}^{n+m}\frac{c(K_i)}{K_i^2}\Delta K_i\right]$$

Appendix 2: Correlation of the two monthly variance proxies

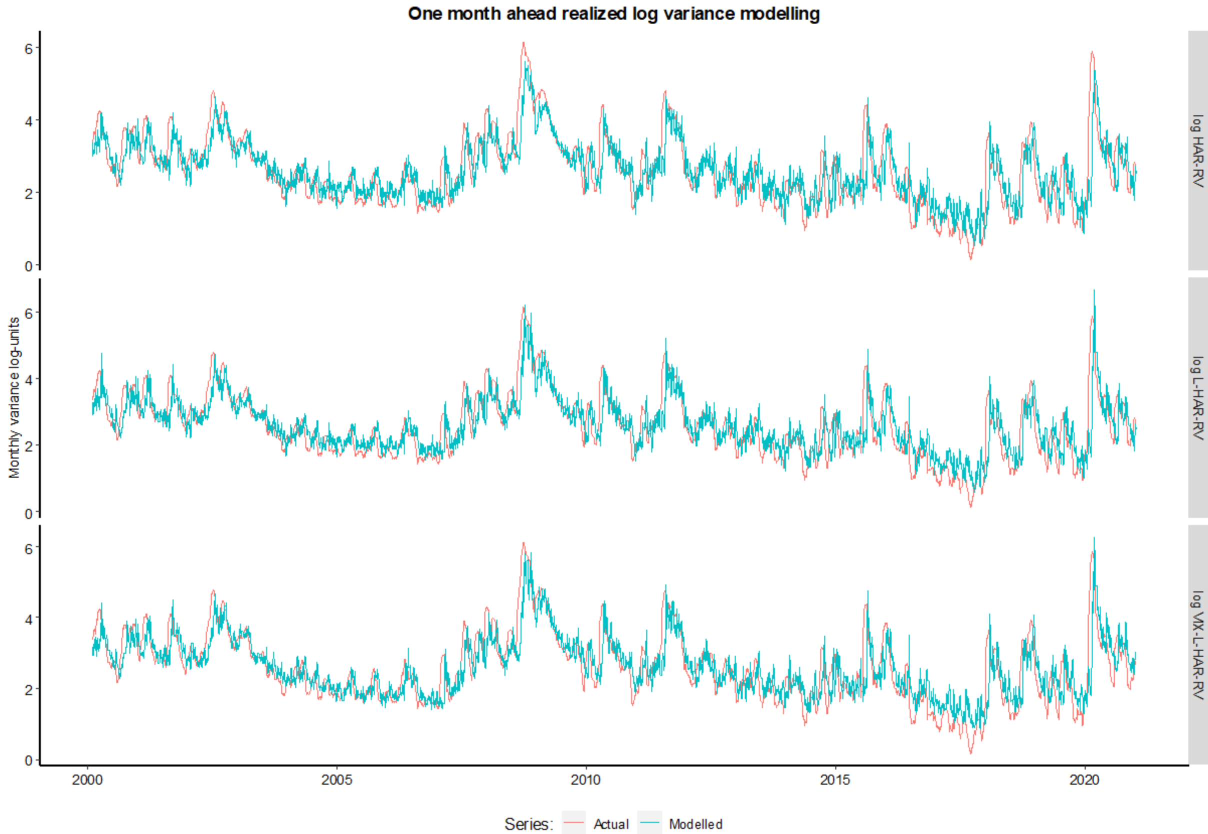
Figure A. 1: Correlation of VIX and Monthly realized variance



Notes: Plot graphically showcases the correlation between the monthly risk neutral (VIX) and physical (Monthly 5-minute RV) variance proxies. Full sample dataset (3rd January 2000 to 12th February) was used to construct the chart.

Appendix 3: Realized log variance modelling full sample

Figure A. 2: One month ahead logarithmic realized variance - full sample estimates



Notes: Fitted values of logarithmic regressions estimated using full sample (from the 3rd of January 2020 to 12th of February 2021) in turquoise are compared to the actual logarithmic value of the 22 days ahead realized variance in red. The plot is divided into 3 facets corresponding to each of the estimated models.

Appendix 4: Variance risk premia summary statistics

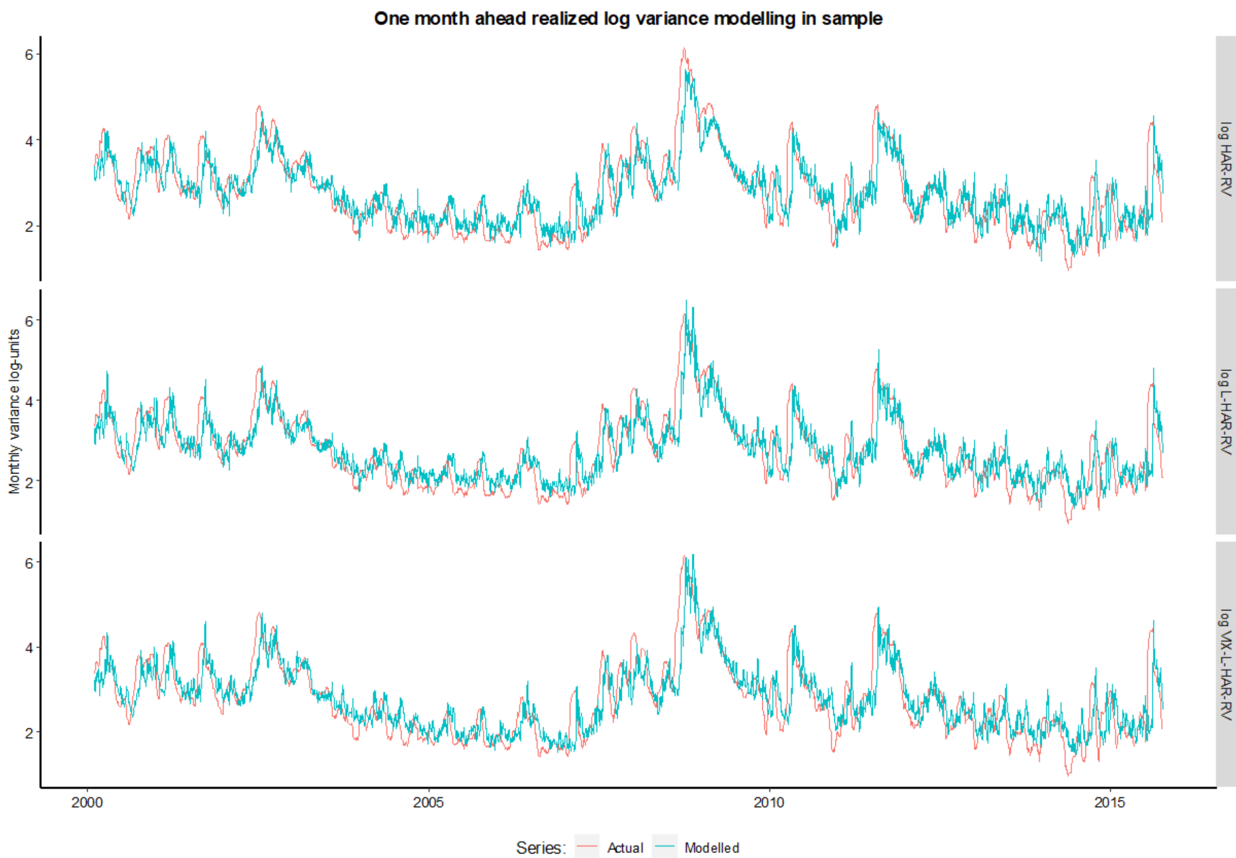
Table A. 1: Variance risk premia summary statistics - full sample estimation

	Level models			Log models		
	VRP HAR-RV	VRP L-HAR-RV	VRP VIX-L-HAR-RV	VRP log HAR-RV	VRP log L-HAR-RV	VRP VIX-L-HAR-RV
n	5,250	5,250	5,250	5,250	5,250	5,250
mean	15.346	15.346	15.346	13.752	12.804	12.466
sd	23.491	21.914	21.596	20.923	21.087	15.015
median	8.281	8.323	8.434	7.748	7.688	8.105
min	-119.723	-65.809	-61.926	-65.732	-515.904	-170.595
max	278.981	250.051	246.265	281.646	193.957	168.983
skew	3.325	3.421	3.443	4.547	-1.88	3.332
kurtosis	19.086	18.736	18.862	33.939	97.934	26.453

Notes: Summary statistics of variance premia estimated by 6 studied models. Full sample (3rd of January 2000 to 12th of February 2021) was used.

Appendix 5: Realized log variance modelling in sample

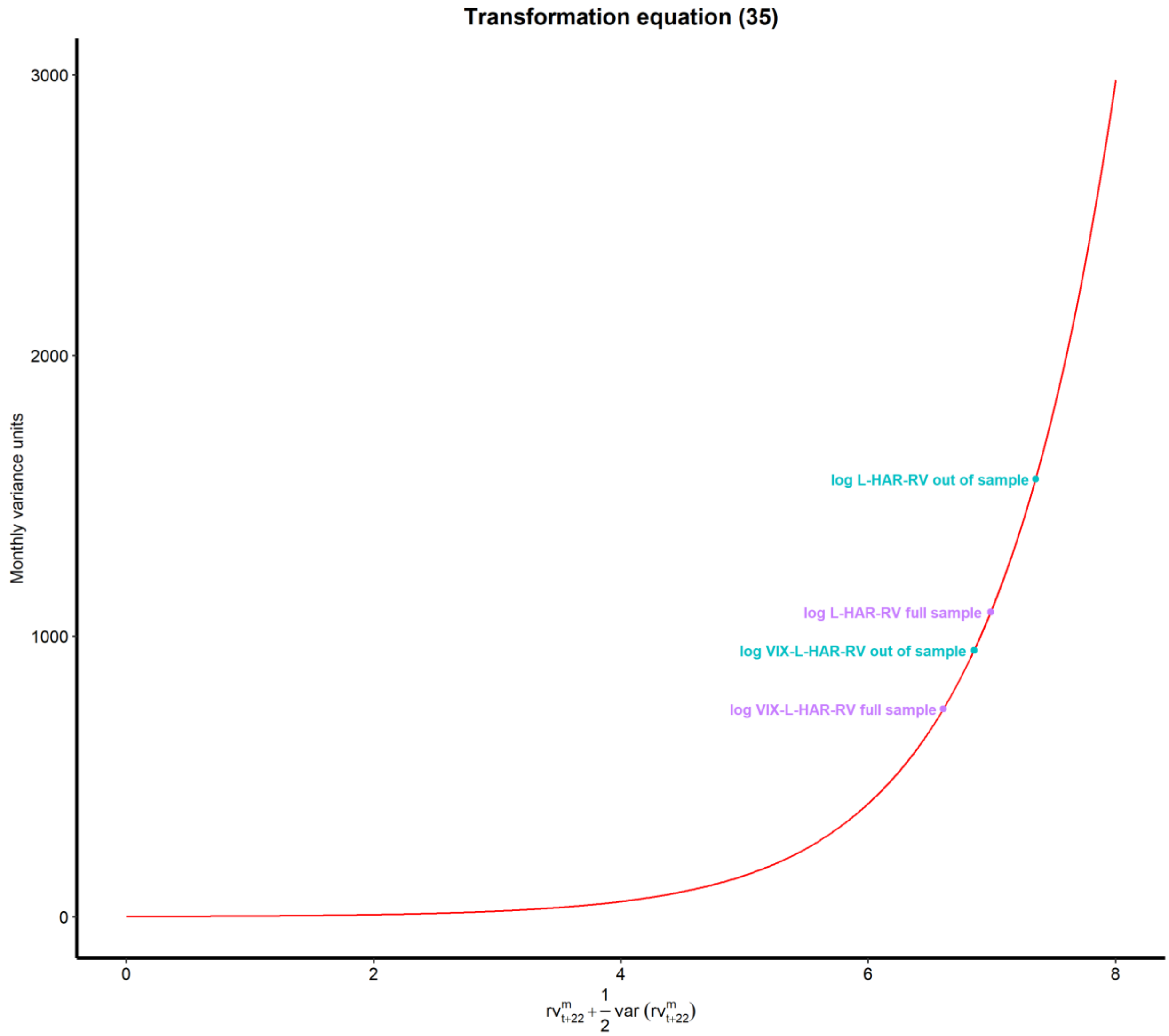
Figure A. 3: One month ahead logarithmic realized variance - in sample estimates



Notes: Fitted values of logarithmic regressions estimated using 75% of the sample (from the 3rd of January 2020 to 9th of October 2015) in turquoise are compared to the logarithm of actual value of the 22 days ahead realized variance in red. The plot is divided into 3 facets corresponding to each of the estimated models.

Appendix 6: Sensitivity of log to level transformation equation

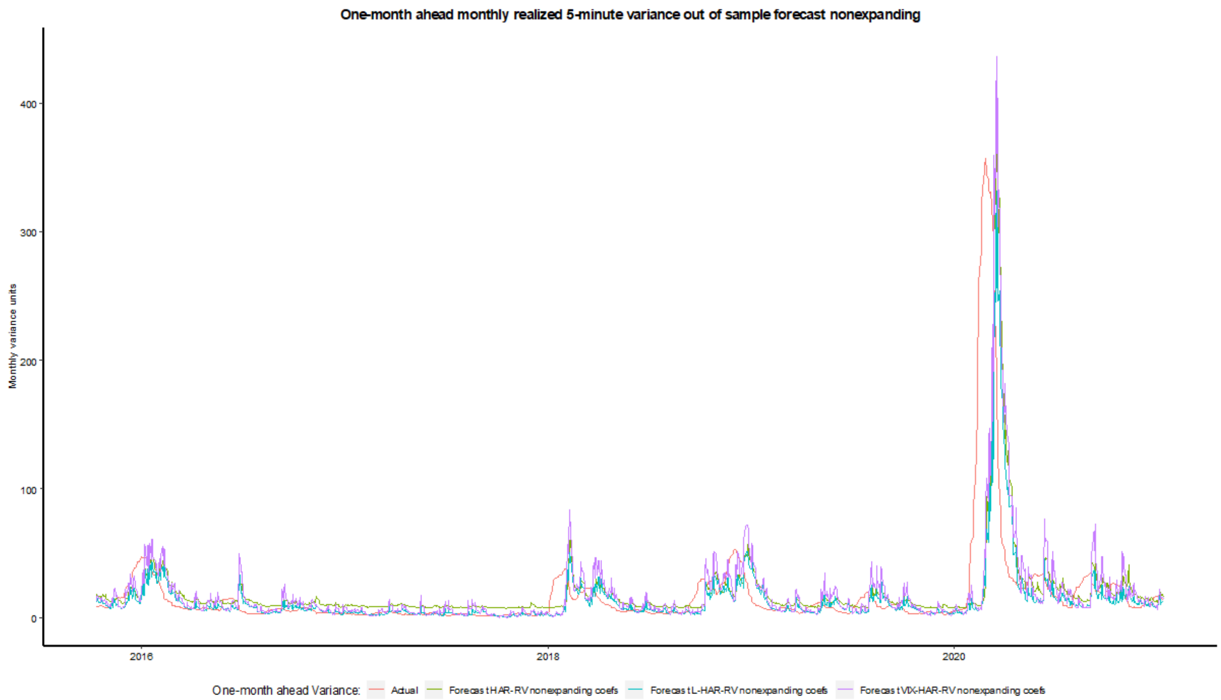
Figure A. 4: Logarithm to level transformation equation sensitivity



Notes: The plot compares the transformed level maxima of models log L-HAR-RV and log VIX-L-HAR-RV between the two estimation approaches (Full sample estimation and Out of sample forecasting procedure introduced in section (9.1)).

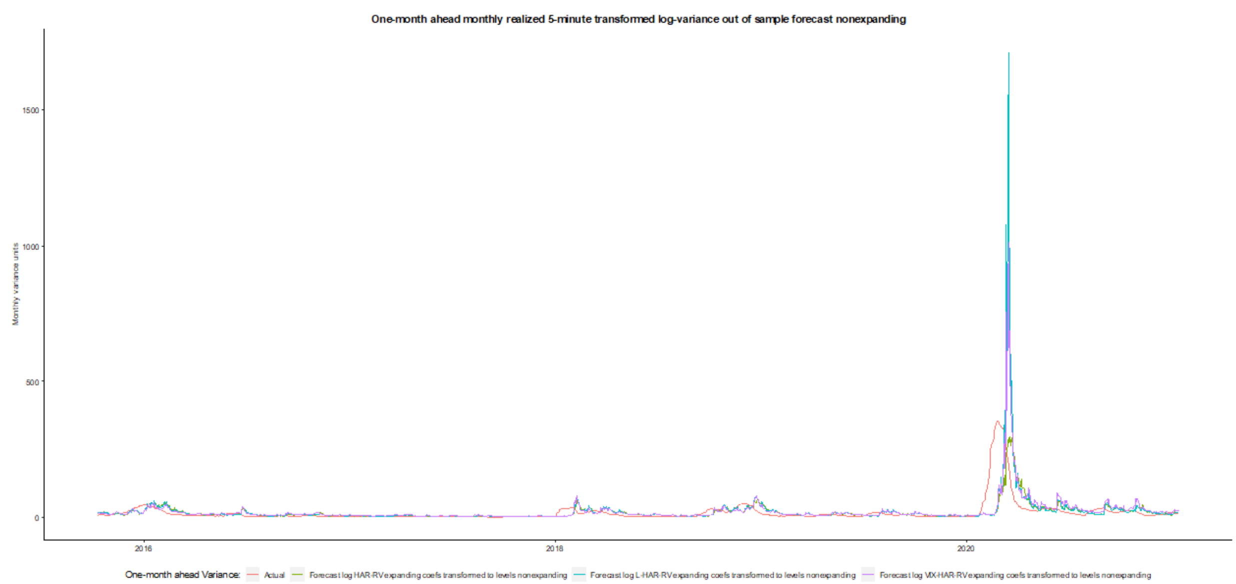
Appendix 7: Non-expanding rolling window forecasting

Figure A. 5: Non-expanding window out of sample forecasts of monthly realized variance of level models



Notes: Forecasted level of realized variance for all level models based on a non-expanding window, contrasted with the actual one-month ahead realized.

Figure A. 6: Non-expanding window out of sample forecasts of monthly realized variance of log models



Notes: Forecasted level of realized variance for all log models based on a non-expanding window, contrasted with the actual one-month ahead realized.

Table A. 2: Model Confidence Set procedure results for non-expanding window procedure

	T_{max}			T_R			
	Rank by T_{max}	t_i	p-value	Rank by T_R	$t_{i,j}$	p-value	MSE
HAR-RV loss	2	-1.06339	1	3	0.989138	0.776333	1509.407
L-HAR-RV loss	1	-1.08353	1	1	-0.74198	1	1376.849
VIX-L-HAR-RV loss	3	-1.03763	1	2	0.741976	1	1488.657
log HAR-RV loss	4	-1.02573	1	6	1.169094	0.46	1452.302
log L-HAR-RV loss	5	1.05066	0.497667	4	1.059451	0.703333	4586.474
log VIX-L-HAR-RV lo:	6	1.083354	0.300667	5	1.084201	0.579667	2488.438

Notes: MCS procedure was conducted at 20% significance level choosing Mean Squared Error (MSE) as the evaluation criterion. The table reports the model´s rank, t statistics and corresponding p-values for both $T_{R,M}$ and $T_{max,M}$ test statistics. The evaluated forecasts were for period spanning from the 12th of October 2015 to 12th of February 2021 on a non-expanding window basis. Bootstrapping was done using block length equal to the maximum number of significant parameters in the AR (p) process on all d_{ij} (default parameter in Catania &

Appendix 8: Stock return predictability regression output

Table A. 3: Univariate stock return predictability regression output for 1-month predictability horizon

Stock Return Regressions 1-month Horizon

	Dependent variable:						
	r^{ex}_{t+1}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HAR-RV	0.029						
	(0.022)						
L-HAR-RV		0.040					
		(0.027)					
VIX-L-HAR-RV			0.040				
			(0.027)				
log HAR-RV				0.035			
				(0.026)			
Four Combination					0.037		
					(0.026)		
VIX^2						0.003	
						(0.009)	
Martingale							-0.020
							(0.026)
Constant	-0.152	-0.284	-0.293	-0.183	-0.244	0.176	0.590
	(0.387)	(0.456)	(0.461)	(0.388)	(0.428)	(0.339)	(0.402)
Observations	239	239	239	239	239	239	239
R^2	0.016	0.025	0.025	0.018	0.022	0.001	0.007
Adjusted R^2	0.012	0.021	0.021	0.014	0.017	-0.003	0.003
Residual Std. Error (df = 237)	4.767	4.746	4.746	4.764	4.755	4.805	4.791
F Statistic (df = 1; 237)	3.962^{**}	6.122^{**}	6.104^{**}	4.299^{**}	5.233^{**}	0.184	1.634
Note:	*p<0.1; **p<0.05; ***p<0.01						

Notes: Regression output for estimation and non-estimation models. Monthly variance risk premium regressed on the one-month ahead S&P 500 stock returns. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme above

Table A. 4: Univariate stock return predictability regression output for 3-month predictability horizon

Stock Return Regressions 3-month Horizon

	Dependent variable:						
	r^{ex}_{t+3}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HAR-RV	0.026						
	(0.019)						
L-HAR-RV		0.038^{*}					
		(0.021)					
VIX-L-HAR-RV			0.038^{*}				
			(0.021)				
log HAR-RV				0.029			
				(0.024)			
Four Combination					0.034		
					(0.025)		
VIX^2						0.0001	
						(0.009)	
Martingale							-0.016
							(0.022)
Constant	-0.111	-0.262	-0.267	-0.113	-0.203	0.281	0.518
	(0.333)	(0.392)	(0.394)	(0.329)	(0.404)	(0.327)	(0.334)
Observations	239	239	239	239	239	239	239
R^2	0.015	0.026	0.025	0.014	0.020	0.00000	0.005
Adjusted R^2	0.011	0.021	0.021	0.010	0.016	-0.004	0.001
Residual Std. Error (df = 237)	4.485	4.461	4.461	4.487	4.473	4.519	4.508
F Statistic (df = 1; 237)	3.535^{*}	6.212^{**}	6.131^{**}	3.381^{*}	4.864^{**}	0.0001	1.141
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$						

Notes: Regression output for estimation and non-estimation models. Monthly variance risk premium regressed on the three-month ahead S&P 500 stock returns. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme

Table A. 5: Univariate stock return predictability regression output for 6-month predictability horizon

Stock Return Regressions 6-month Horizon

	Dependent variable:						
	r^{ex}_{t+6}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HAR-RV	0.026*						
	(0.015)						
L-HAR-RV		0.040**					
		(0.016)					
VIX-L-HAR			0.040**				
			(0.016)				
log HAR-RV				0.025			
				(0.021)			
Four Combination					0.034**		
					(0.017)		
VIX^2						-0.002	
						(0.009)	
Martingale							-0.016
							(0.018)
Constant	-0.106	-0.297	-0.300	-0.066	-0.206	0.366	0.519**
	(0.343)	(0.413)	(0.412)	(0.321)	(0.368)	(0.319)	(0.263)
Observations	239	239	239	239	239	239	239
R^2	0.015	0.030	0.029	0.011	0.021	0.0005	0.005
Adjusted R^2	0.010	0.025	0.025	0.007	0.017	-0.004	0.001
Residual Std. Error (df = 237)	4.416	4.383	4.384	4.424	4.402	4.448	4.437
F Statistic (df = 1; 237)	3.520*	7.222***	7.057***	2.667	5.047**	0.114	1.215
Note:	*p<0.1; **p<0.05; ***p<0.01						

Notes: Regression output for estimation and non-estimation models. Monthly variance risk premium regressed on the six-month ahead S&P 500 stock returns. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme above

Table A. 6: Univariate stock return predictability regression output for 12-month predictability horizon

Stock Return Regressions 12-month Horizon

	Dependent variable:						
	r^{ex}_{t+12}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HAR-RV	0.028^{**}						
	(0.012)						
L-HAR-RV		0.041^{***}					
		(0.013)					
VIX-L-HAR			0.041^{***}				
			(0.014)				
log HAR-RV				0.031^{*}			
				(0.016)			
Four Combination					0.036^{***}		
					(0.014)		
VIX^2						0.001	
						(0.008)	
Martingale							-0.007
							(0.014)
Constant	-0.131	-0.301	-0.303	-0.131	-0.232	0.240	0.401
	(0.388)	(0.444)	(0.443)	(0.349)	(0.405)	(0.333)	(0.298)
Observations	239	239	239	239	239	239	239
R^2	0.017	0.029	0.029	0.016	0.023	0.0002	0.001
Adjusted R^2	0.012	0.025	0.025	0.011	0.019	-0.004	-0.003
Residual Std. Error (df = 237)	4.525	4.496	4.497	4.528	4.511	4.563	4.561
F Statistic (df = 1; 237)	3.988^{**}	7.199^{***}	7.024^{***}	3.764^{*}	5.537^{**}	0.041	0.244
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$						

Notes: Regression output for estimation and non-estimation models. Monthly variance risk premium regressed on the twelve-month ahead S&P 500 stock returns. Significance of coefficients is indicated by the star symbol *,**,*** according to the scheme