What does the volatility risk premium say about liquidity provision and demand for hedging tail risk?∗ †

Jianqing Fan
jqfan@princeton.edu

Michael B. Imerman‡
mbi212@lehigh.edu

Wei Dai
weidai@princeton.edu

October 28, 2015

JEL Classification: C01, C58, G12

Keywords: volatility risk premium, integrated volatility, ultra-high-frequency data, microstructure noise, Fourier transform, tail risk, Big Data risk analytics

∗Jianqing Fan is the Frederick L. Moore Professor of Finance, Bendheim Center for Finance, Princeton University, Princeton, NJ 08544; Michael B. Imerman is an Assistant Professor, Perella Department of Finance, Lehigh University, Bethlehem, PA 18015; and Wei Dai is a Ph.D. Student, Department of Operations Research & Financial Engineering, Princeton University, Princeton, NJ 08544.

†We would like to thank participants at the 2015 China International Conference in Finance (CICF), the 2015 FMA meetings in Orlando, the 20th Annual Conference on Pacific Basin Finance, Economics, Accounting, and Management at Rutgers University and the Second Measuring Risk Conference at Princeton University as well as the Finance Department Seminar Series at Rutgers Business School. We are especially grateful for comments from Yacine Aït-Sahalia, Peter Carr, Hui Chen, Hao Zhou, and Hong Zhang (CICF discussant).

‡Contact Author Email: mbi212@lehigh.edu; Phone: (610) 758-6380.
What does the volatility risk premium say about liquidity provision and demand for hedging tail risk?

ABSTRACT

This paper provides a data-driven analysis of the volatility risk premium, using tools from high-frequency finance and Big Data analytics. We argue that the volatility risk premium, loosely defined as the difference between realized and implied volatility, can best be understood when viewed as a systematically priced bias. We first use ultra-high-frequency transaction data on SPDRs and a novel approach for estimating integrated volatility on the frequency domain to compute realized volatility. From that we subtract the daily VIX, our measure of implied volatility, to construct a time series of the volatility risk premium. To identify the factors behind the volatility risk premium as a priced bias we decompose it into magnitude and direction. We find compelling evidence that the magnitude of the deviation of the realized volatility from implied volatility represents supply and demand imbalances in the market for hedging tail risk. It is difficult to conclusively accept the hypothesis that the direction or sign of the volatility risk premium reflects expectations about future levels of volatility. However, evidence supports the hypothesis that the sign of the volatility risk premium is indicative of gains or losses on a delta-hedged portfolio.
1 Introduction

Whether realized volatility is greater than or less than implied volatility is an empirical question, and one that has been studied over time (see Mixon, 2009). However, the literature contains conflicting evidence about implied volatility as an unbiased estimator of future realized volatility (e.g. Canina and Figlewski, 1993; Christensen and Prabhala, 1998). Theory suggests that implied volatility should be a biased estimate of future realized volatility since implied volatility includes the market price of risk; that is, implied volatility is the expected “actual” (or statistical) volatility plus a risk premium. In mathematical finance this is formalized in terms of a change of measure. The volatility risk premium is defined as the difference between the expected future volatility under the physical measure (ex-ante forecast of realized volatility) and the expected future volatility under the risk-neutral measure (implied volatility from option prices). Therefore, the existence of a non-zero volatility risk premium indicates that not only is implied volatility a biased estimator of future realized volatility but, furthermore, that the bias is systematically priced.

The volatility risk premium has been an active area of research in financial economics for some time now. Whereas existing studies typically begin with an asset pricing model or some framework of stochastic or time-varying volatility, we take a step back from the theoretical foundation of volatility in financial markets and perform a purely data-driven analysis of the volatility risk premium, leveraging the insights of Big Data analytics. That is, we start with a massive data-set of transaction level prices: our sample includes over half a billion trades in SPDRs, the ETF that tracks the S&P 500 index, from 2006 to 2011. We use this data to estimate the realized volatility of the market using a robust methodology with minimal assumptions. Then, we compare this realized volatility to a measure of model-free implied volatility, daily over the same five year period. Our proxy for the model-free implied volatility is the VIX volatility index, which is commonly used in other studies of the volatility risk premium. Since our objective is to quantify the volatility risk premium in a model-free, nonparametric manner, we compare the computed ex-post realized volatility to the contemporaneous level of VIX. As a result, we essentially return to
the fundamental idea of the volatility risk premium as a bias – not a random bias, but a systematically priced bias. Through our data-driven analysis we seek to better understand the economic determinants of this bias.

Our study of the volatility risk premium contributes to a recent trend in financial research and risk analysis. In any quantitative field, there are two approaches to conducting research: model-based and data-driven. This dichotomy is perhaps more pronounced now than ever in the quantitative fields of financial economics – i.e. asset pricing, derivatives, and risk management. Traditionally, the researcher would construct a model based on theory and then use data to empirically verify or validate the model and justify the economic intuition it conveys. However, with the abundance of financial data being generated every day and the increasing popularity of “Big Data” along with data mining and machine learning techniques making their way into the financial engineer’s toolbox, a new data-driven approach to research in these areas is gaining popularity. We view our study as an extension of this research philosophy to better understand how the market prices volatility.

In particular, our data-driven analysis highlights the role of intermediaries that provide liquidity to investors who seek to hedge their downside tail risk – that is, a supply and demand framework – for pricing volatility in the market. We begin with a statistical analysis of the data, using large-scale data collected from several different sources. This introduces the additional challenge of having to collate data from multiple platforms, in different formats, often on differing time scales. Such challenges are commonplace in “Big Data” research and give rise to problems such as spurious correlation, time asynchronicity, and noise accumulation (see Fan, 2013; Fan et al., 2014). However, using tools from high-frequency finance and big data analytics, we are able to obtain a clean and more precise estimate of the true integrated volatility without any reliance on a model or parametric assumptions. Our consistently estimated realized volatility is then compared to a model-free measure of implied volatility, proxied by the VIX index, to get our time-series of the volatility risk.

For a discussion about this dichotomy as it arose between statistics and mathematics, please see Shepp (2007).
premium. We make the distinction between this ex-post formulation of the volatility risk premium with other studies and note that our resulting time series is the daily realization of the volatility risk premium. We still interpret this as the market price of volatility risk, but consistent with our data-driven philosophy, frame our measure of the volatility risk premium as a *systematically priced bias*.

This bias should be negative on average since the realized volatility is expected to be less than the option implied volatility, the latter containing the risk premium. The literature suggests that since traditional risk factors are unable to explain the negative volatility risk premium, there is some independent risk factor that is driving its dynamics (Carr and Wu, 2009). Despite the volatility risk premium being negative in expectation, or on average, we find that there is a large amount of variation in the volatility risk premium over time including several instances with extreme positive spikes. Following our framework that the volatility risk premium is a systematically priced bias, our empirical design approaches the problem from the perspective of a *deviation analysis*, decomposing it into magnitude and direction. Although this may be basic relative to modern tools in statistical learning theory, such an analysis has never been performed to understand how the financial markets price risk. This allows us to use straightforward regression techniques to gain valuable insights into the magnitude (absolute value) and direction (sign) of the systematically priced bias associated with the volatility risk premium, and to our knowledge we are the first to do so.

We control for micro-level noise in our specific methodology for computing realized volatility at high frequency. Then examining the absolute deviation of the realized volatility from the implied volatility mitigates the effects of the extreme swings in the volatility risk premium (what might be termed as “outliers” in traditional finance research) producing more robust results. Additionally, the empirical specification using absolute deviation, or the magnitude of the volatility risk premium, highlights

---

2We were able to confirm the inability of traditional risk factors to explain the volatility risk premium in results that are unreported in this version of the paper.
the potential “missing risk factor” by accounting for supply and demand imbalances in the market for hedging tail risk.

We find that the magnitude of the volatility risk premium, that is the absolute value of the bias between the realized and implied volatility, reflects these supply and demand imbalances in the index option market where investors buy protection on downside tail risk. The demand effects are captured by a statistically significant relationship between open interest for put options on the S&P 500 and the magnitude of the volatility risk premium. The supply effects, are captured by the TED spread and the credit spread. The former often appears as a proxy for liquidity (or illiquidity) in financial markets and is viewed a general measure of financial instability. The latter, which is most significant during the Financial Crisis in our results, interestingly has the interpretation of dealers deleveraging and shrinking their balance sheets by selling off risky positions. This is consistent with the evidence provided by Adrian and Shin (2010).

As for the direction of the volatility risk premium, practitioners believe that the volatility risk premium’s sign is indicative about future levels of realized volatility. When the volatility risk premium is negative, implied volatility is higher than realized volatility and market participants believe that volatility is likely to increase in the future. On the contrary, when the volatility risk premium is positive, implied volatility is less than realized volatility and market participants believe that volatility is likely to decrease in the future. This is related to the contentious idea of implied volatility’s ability to forecast future realized volatility in the literature and, more recently, the “Expectation Hypothesis” discussed in Aït-Sahalia et al. (2012). In fact, Aït-Sahalia et al. (2012) provides a way to test this hypothesis about the direction of the volatility risk premium.

An alternative explanation in the finance literature says that the sign of the volatility risk premium represents the gains or losses on market makers’ delta-hedged positions. This was first proposed by Bakshi and Kapadia (2003) within the context of
stochastic volatility and jump-diffusion models. Using our model-free, data-driven analysis, we are able to find much stronger evidence in favor of this explanation.

The remainder of this paper is structured as follows. The next section, Section 2, presents the methodology for computing realized volatility using ultra-high-frequency data. Section 3 discusses the economic basis for the volatility risk premium and why implied volatility can, and will, deviate from realized volatility. We also propose our testable hypotheses based on several stylized facts about the volatility risk premium and the market for hedging downside tail risk from the financial economics literature. Section 4 details the data collection and methodology for constructing the volatility risk premium. Our empirical analysis and results are presented in Section 5. Section 6 concludes. We have two Appendices: Appendix A reviews different statistical methodologies for computing realized volatility with emphasis on the estimation of integrated volatility with high-frequency data; Appendix B presents simulations that demonstrate the benefit of using ultra-high-frequency data in estimating integrated volatility and the extent to which our method performs better than alternatives.

2 Realized Volatility and Frequency Domain Estimation of Integrated Volatility

In this section we discuss the frequency domain estimation of realized volatility from Olhede et al. (2009). While this does not represent a methodological innovation of our work, we argue that it is a powerful methodology for computing realized volatility under fairly general conditions with minimal parametric assumptions and is therefore able to give cleaner estimates using noisy ultra-high-frequency data and yield new insights into how volatility is priced in the market. More specifically, the procedure computes a consistent and unbiased estimator of integrated volatility at ultra-high-frequencies under very general specifications of the microstructure noise process. The

---

3In a Black-Scholes-Merton world (Black and Scholes, 1973; Merton, 1973) gains or losses on the delta-hedge perfectly offset the loss or gain on the short option position. However, that is not the case with jumps and, in fact, there is evidence that downward jumps explain much of the pricing of volatility risk.
frequency domain methodology, and its statistical properties, is compared to some alternative methodologies for estimating volatility with high-frequency data in Appendix A.

2.1 The methodology

Without loss of generality but for the sake of formalization, we can start with a latent true (log)price $X_t$ that follows an Ito process

$$dX_t = \mu_t dt + \sigma_t dW_t$$ \hfill (1)

where $W_t$ is a standard Brownian Motion and $\mu_t$ and $\sigma_t$ are time-varying drift and volatility, respectively, that may or may not follow stochastic processes themselves.\footnote{Some methodologies can also be applied to jump-diffusion processes rather than just pure diffusion processes.} However, what we observe is the transaction price, or its logarithm, $Y_t$ at times $\{t_i\} \in [0, T]$, which are related to $X_t$ according to

$$Y_{ti} = X_{ti} + \epsilon_{ti},$$ \hfill (2)

where $\epsilon_{ti}$ represents microstructure noise.

The frequency domain methodology uses a discrete Fourier transform to go from the time domain, on which financial data is typically portrayed, to the frequency domain. For a generic time series $U_{t_j}, j = 1, \cdots, N$, the discrete Fourier transformation of the increment process $\Delta U_{t_j} = U_{t_{j+1}} - U_{t_j}$ is

$$J_k^{(U)} = \sqrt{\frac{1}{N}} \sum_{j=1}^{N} \Delta U_{t_j} e^{-2\pi i j f_k}, f_k = \frac{k}{T}. \hfill (3)$$
Using this notation, the integrated volatility can be written in terms of the variance of $J^{(X)}_k$: 

$$
\int_0^T \mathbb{E}\{\sigma_s^2\}ds = \sum_{k=0}^{N-1} \mathbb{E}|J^{(X)}_k|^2 + O(\Delta t). \tag{4}
$$

On the other hand, applying the transformation on the observable transaction price $Y_t$ at times $\{t_i\} \in [0, T]$ leads to a noise contribution at each frequency,

$$
\sum_{k=0}^{N-1} \mathbb{E}|J^{(Y)}_k|^2 = \sum_{k=0}^{N-1} \left( \mathbb{E}|J^{(X)}_k|^2 + a^2|2\sin(\pi f_k \Delta t)|^2 \right), \tag{5}
$$

where for now we assume $\epsilon_t$ is a white noise process with variance $a^2$. It can be seen that the high-frequency coefficients are more heavily contaminated by the noise. Ideally we would like to shrink $\mathbb{E}|J^{(Y)}_k|^2$ by

$$
L_k = \frac{\mathbb{E}|J^{(X)}_k|^2}{\mathbb{E}|J^{(X)}_k|^2 + a^2|2\sin(\pi f_k \Delta t)|^2} \tag{6}
$$

and therefore remove the noise locally at each frequency. In practice we still need to estimate the multiscale ratio $L_k$. One way is to use the Whittle log-likelihood to estimate the unknown quantities in Equation (6),

$$
l \left( \sigma_X^2, a^2 \right) = -\sum_{k=1}^{N/2-1} \log \left( \sigma_X^2 + a^2|2\sin(\pi f_k \Delta t)|^2 \right) - \sum_{k=1}^{N/2-1} \frac{|J^{(Y)}_k|^2}{\sigma_X^2 + a^2|2\sin(\pi f_k \Delta t)|^2}, \tag{7}
$$

and it follows that

$$
\hat{L}_k = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_X^2 + \hat{a}^2|2\sin(\pi f_k \Delta t)|^2}. \tag{8}
$$

It can be shown that the final de-biased estimator

$$
\langle X, X \rangle_T^{(\hat{L}_k)} = \sum_{k=0}^{N-1} \hat{L}_k|J^{(Y)}_k|^2 \tag{9}
$$

\footnote{To simplify the notation, we assume the drift term is zero as it does not affect the asymptotic behaviour (for a more complete version, see Olhede et al. (2009)).}
is a consistent estimator of the integrated volatility:

$$\langle X, X \rangle^{(L_k)}_T = \int_0^T \sigma_t^2 \, dt + O_p(\Delta t^{1/4}).$$  \hfill (10)

### 2.2 Autocorrelated Noise

This frequency domain methodology also allows us to easily model autocorrelated noise and disentangle the noise effect at each frequency in the same way. If we assume that $\epsilon_{t_j}$ is an autocorrelated stationary time series, it is convenient to model it as a moving average process of order $q$,

$$\epsilon_{t_j} = \eta_{t_j} + \sum_{k=1}^q \theta_k \eta_{t_{j-k}},$$  \hfill (11)

where $\{\eta_{t_j}\}$ is a white noise process with variance $\sigma^2_{\eta}$. This MA($q$) specification leads to a new likelihood function

$$l\left(\sigma^2_X, \sigma^2_{\eta}, \{\theta_k\}_{k=1}^q\right)$$

$$= - \sum_{k=1}^{N/2-1} \log \left(\sigma^2_X + \sigma^2_{\eta} |1 + \sum_{k=1}^q \theta_k e^{2i\pi fk}|^2\right) - \sum_{k=1}^{N/2-1} \frac{|J^Y_k|^2}{\sigma^2_X + \sigma^2_{\eta} |1 + \sum_{k=1}^q \theta_k e^{2i\pi fk}|^2}.$$  \hfill (12)

Therefore the multiscale ratio is given by

$$\hat{L}_k = \frac{\hat{\sigma}^2_X}{\hat{\sigma}^2_X + \hat{\sigma}^2_{\eta} |1 + \sum_{k=1}^q \hat{\theta}_k e^{2i\pi fk}|^2}.$$  \hfill (13)

where the order of the moving average process – i.e. the appropriate number of lags in the autocorrelated noise – is determined through model selection using the corrected Akaike information criteria (AICC),

$$AICC(q) = -2l\left(\hat{\sigma}^2_X, \hat{\sigma}^2_{\eta}, \{\hat{\theta}_k\}_{k=1}^q\right) + 2\frac{(q + 2)N}{N - q - 3}.$$  \hfill (14)
2.3 Advantages

The frequency domain method for estimating integrated volatility has several desirable features, both in terms of the statistical properties and the practicality in applying to real financial data. From the financial modeling and data analysis point-of-view, working in the frequency domain provides an elegant way to address more general specifications of the microstructure noise process. It bears repeating that maintaining minimal parametric assumptions and high degree of flexibility are some of the features that make this methodology attractive for our purposes, considering the data-driven philosophy of our approach.

Many methods for estimating integrated volatility assume that the noise process, $\epsilon_t$, is i.i.d. or uncorrelated. However, in practice this is an unreasonable assumption.\footnote{While most previous approaches assume i.i.d. microstructure noise, recent work by Aït-Sahalia et al. (2011) addresses the complicated issue of estimating volatility from ultra-high-frequency data with dependent microstructure noise. Our methodology is similar in that it also permits estimation of integrated volatility for more general classes of microstructure noise, but with less parametric restrictions.} Autocorrelated microstructure noise may be a more reasonable assumption, since large disturbances this second may be highly correlated with large disturbances last second, especially if there is a lot of noise in the market. This may give the impression that the market is more turbulent or volatile, when in fact the persistent volatility in our observed time series, $Y$, is coming from the microstructure noise. Thus, we need a clean way to strip away the true volatility of the price process in the presence of microstructure noise at ultra-high-frequencies. This is why we use the frequency domain estimation method in computing integrated volatility. In fact, we find that the data indicates the latent noise process has on average lag-1 autocorrelation, and the time varying order of autocorrelation ranges from 0 to 5.

Once we get our de-biased, clean measure of realized volatility in the market, we can compare it to a measure of implied volatility to get a time series of the volatility risk premium on which we perform our deviation analysis. The next section discusses these concepts – implied volatility, the volatility risk premium – and constructs our testable hypotheses.
3 Implied Volatility and the Volatility Risk Premium

The notion of implied volatility is well understood and widely used by options traders and financial engineers. In this section we briefly review the concept of implied volatility and discuss in greater detail the volatility risk premium which is imbedded in implied volatilities but not included in realized volatilities. Therefore, if we wish to look at the difference between realized volatility and implied volatility, it gives us a measure of the volatility risk premium. We further view this deviation between realized and implied volatilities in an ex-post sense as a priced bias in the options markets.

While it is well known that the Black-Scholes-Merton option pricing model relies on unrealistic assumptions and therefore cannot reasonably price options, the model is still widely used by traders to infer the level of volatility associated with a particular option on a given asset (e.g. stocks, indexes, or currencies). The idea is that since volatilities are unobservable but option prices are observed and convey traders expectations about the future riskiness of the underlying asset over the life of the option, the financial engineer is faced with a mathematical inverse problem. Given the output of the model – option price – solve for the value of the volatility parameter that sets the model value equal to the market price of the option. This is the *implied volatility*. Now, of course, any option pricing model can be used; it is just that the Black-Scholes-Merton model is the most basic and convenient. Regardless, implied volatilities computed in this fashion are by definition model-dependent and constrained by parametric restrictions.

Researchers including Rubinstein (1994), Dupire (1994), and Derman and Kani (1994) have, to much success, extended this idea of implied volatility to extract market information across entire classes of options on a given asset (i.e. different strikes and/or maturities; often referred to as the “volatility smile” or “volatility surface”) and fit a deterministic function of asset price, strike price, and time to expiry.
More recently, the work of Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) broke away from the reliance on models and derived and implemented a model-free implied volatility using only current option prices. It is along these lines that we would like to use implied volatility in our data-driven analysis.

The implied volatility is forward-looking and represents the market’s expectation of volatility over the life of the option. Mathematically, it can be thought of as an expectation under the risk-neutral or pricing measure. The volatility as computed from the underlying asset price movement can be thought of as being generated under the physical or statistical measure. The difference between the two represents the market price of volatility risk, or what is referred to as the “volatility risk premium”. In some instances it is easier to work with the squared volatility, which leads to the variance risk premium defined as follows.

$$VRP_t = \mathbb{E}^P_t \left[ \int_t^T \sigma^2_u du \right] - \mathbb{E}^Q_t \left[ \int_t^T \sigma^2_u du \right].$$

(15)

Taking the square root of each of the expectations in Equation (15) gives the volatility risk premium which we will denote as a lowercase $\text{vrp}_t$. While we present all of our results in terms of the volatility risk premium, everything is robust with respect to the variance risk premium. The reason we chose to use the former is that the results are easier and more natural to interpret in terms of vol units.

In many studies the first term in Equation (15) is computed as an ex-ante conditional expectation of the future realized volatility or variance given the current value through an econometric model. We intentionally choose not to do this, but rather use an ex-post measure of the realized volatility. The reason is computing the ex-ante conditional expected realized volatility introduces statistical problems such as overfitting, model error, and possible misspecification bias. Instead we compare the ex-post realized volatility (averaged over a one month period) to the model-free im-

---

7While the majority of volatility risk premium studies use the ex-ante conditional expectation, we are not alone in using the ex-post realized volatility. In a recent publication, Buraschi et al. (2014) define the volatility risk premium in the same way as us just with the reverse sign (see p. 124). We follow Carr and Wu (2009) who take the perspective that the negative sign reflects investors’ willing to pay to hedge their volatility risk.
plied volatility (covering the same horizon) to get a realization of the volatility risk premium on each trading day over the sample period. In symbols, this is

$$vrp_t = \sqrt{\int_t^T \sigma_u^2 du} - \sqrt{E_t^Q \left[ \int_t^T \sigma_u^2 du \right]}.$$  \hspace{0.5cm} (16)

Therefore another distinction from the uppercase $VRP_t$ in Equation (15) and the lowercase $vrp_t$ in Equation (16) is that the latter is our bias representation of the volatility risk premium. Furthermore, once we substitute the level of the VIX volatility index in for the risk-neutral expectation, it would be the realization of the volatility risk premium which is similar to the realized P&L on a volatility swap (see Demeterfi et al., 1999).

Recent studies have been able to establish some interesting empirical properties of the volatility/variance risk premium. Bollerslev et al. (2009) and Bollerslev et al. (2011) examine the predictability of the variance risk premium on stock market returns noting that the predictive power is better than other financial and macroeconomic factors that are typically used in stock market return forecasting. Zhou (2011) studies the predictability of the variance risk premium across financial markets through equity returns, bond returns, and credit spreads with the $VRP$ predictability maximized typically in the one to four month horizon.

Using high-frequency index futures data, Wu (2011) considers the role that jumps play finding that both the jump arrival rate and the absolute value of the negative variance risk premium are proportional to the variance level. The interpretation, which has implications for our hypotheses below, is that when volatility is high the volatility risk premium is either very positive or very negative; that is the bias between realized and implied volatility increases when the level of uncertainty is heightened.

Thus, the volatility risk premium appears to be a priced risk factor in the capital markets and investors are willing to pay a premium to hedge their downside risk, especially when uncertainty is high. However, our knowledge is still very limited about the determinants of the volatility risk premium and we do not have sound
empirical evidence documenting what exactly the volatility risk premium says about the mechanics of the market for pricing and hedging risk.

We construct our testable hypotheses based on several stylized facts from the recent literature in financial economics. First, the volatility risk premium represents option market makers’ willingness to absorb inventories and provide liquidity (Gárleanu et al. (2009), Nagel (2012)). Also, investors are net buyers of index options (Gárleanu et al. (2009)). To the extent that investors use index put options to hedge their downside tail risk, then we should be able to use option market data to draw inferences about investors’ demand for hedging downside tail risk and intermediaries’ willingness to meet this demand (i.e. provide liquidity). The volatility risk premium can, therefore, naturally be interpreted as the compensation that option market makers receive for this intermediation and liquidity provision (supply) to meet hedging demand. Adrian and Shin (2010) find evidence of this interpretation in the expansion and contraction of financial intermediaries’ balance sheets.

We begin our analysis by postulating that the direction (sign) contains different information than the magnitude (absolute value) of the \( \nu_{rp} \), further justifying our empirical strategy of a deviation analysis. By some accounts, traders view the direction of the volatility risk premium as indicative of nothing more than the market’s expectation of future levels of volatility. Therefore, pulling out the sign leaves us with the absolute deviation of the realized volatility from implied volatility. Based on the aforementioned stylized facts about the volatility risk premium, we expect that the magnitude of the volatility risk premium reflects the supply and demand imbalances in the market for hedging downside tail risk. When demand for hedging downside tail risk increases, market makers will take the short side (sell volatility) but must be compensated appropriately. The price of volatility increases and implied volatility rises relative to realized levels. When demand for hedging downside tail risk decreases, there will be a selloff of volatility and market makers will take the other side, but only at a substantial discount. Implied volatility falls relative to realized levels. Therefore, the magnitude captures the extent to which market makers must be compensated to provide liquidity to the options markets, either as a premium or
discount if intermediaries are selling volatility to meet hedging demand or buying it back in response to a reduction in hedging demand. Our hypotheses about the absolute deviation or the magnitude of the volatility risk premium follow from this logic.

**Magnitude Hypotheses:**

**H1:** The magnitude of the volatility risk premium reflects investors’ demand for hedging tail risk.

**H2:** The magnitude of the volatility risk premium reflects the willingness of option market makers to absorb inventory and provide liquidity.

We should note that the hypotheses need not be mutually exclusive. Hypothesis H1 can be thought of as a demand-side effect and Hypothesis H2 can be thought of as a supply-side effect. We may, therefore, find that supply and demand forces work with or against each other to determine the magnitude of the \( \text{vrp} \) at a given time.

The direction, on the other hand, may or may not be independent of the magnitude component; that is, it may reflect entirely different information about the market or it could be telling us something different about the role of intermediaries that provide liquidity to options traders. Therefore, we have another set of hypotheses about the sign of the volatility risk premium. Some traders view the volatility risk premium being negative or positive as nothing more than a reflection of expectations about future levels of volatility. However, there is an alternative hypothesis that the suggests the sign of the volatility risk premium contains information about the performance of market makers’ hedged positions as a function of different market conditions. Both of these perspectives have been proposed in the literature (see Aït-Sahalia et al. (2012), Bakshi and Kapadia (2003), respectively).

**Direction Hypotheses:**

**H3:** The sign of the volatility risk premium contains information about future levels of realized volatility relative to implied volatility.

**H4:** The sign of the volatility risk premium reflects the delta-hedged gains or losses for option market makers.
We perform our deviation analysis on the systematic bias between implied and realized volatility by decomposing the \( \text{vrp} \) in Equation (16) into magnitude and sign as follows

\[
\text{vrp}_t = \sqrt{\int_t^T \sigma^2_u du} - \sqrt{\mathbb{E}_t^{\mathcal{G}} \left[ \int_t^T \sigma^2_u du \right]} \times \text{sgn}(\text{vrp}_t).
\]  

(17)

To see how the market prices this systematic bias we then conduct several statistical and econometric tests on each of the components of \( \text{vrp} \) in Equation (17) within the context of the hypotheses H1-H4.

In the next section we discuss our data in detail: the collection, cleaning, and processing; as well as the construction of our volatility risk premium time series and the other financial market and economic variables that are used in our econometric analysis of the volatility risk premium.

4 Data and Methodology

4.1 Data Collection

The data used for our empirical analysis came from several different sources, on multiple platforms, and were analyzed using a variety of softwares. This is a common feature of Big Data analytics and requires careful processing and collating to ensure that the data are in consistent formats, with large-scale computations often being done in parallel (see Fan et al., 2014). First, we began by cleaning and processing the ultra-high-frequency transaction data for the SPDR ETF. Then we used the cleaned price data to estimate the integrated volatility on the frequency domain via the Fast Fourier Transform (FFT) algorithm. The computed integrated volatility was then merged with a daily time series of the VIX index, and the difference between the two time series gave us the volatility risk premium. Finally, we had to collect, clean, and merge with the economic, financial market, and risk factor variables from
their respective databases. This data was used in our econometric analyses of the determinants and drivers of the volatility risk premium.

Transaction price data for the SPDR ETF (ticker SPY) was obtained from the TAQ database within WRDS. The sample period we studied goes from July 2006 to June 2011. Over these five years there were a total of 523,814,632 trades. For our integrated volatility estimation method to work best, we need as many observations as possible. Trade volume decreases considerably as we go further back in time which is why we stop at 2006. The first year of data (2006-2007) has approximately one-quarter the number of trades as the final year of data (2010-2011). Additionally, this sample period contains about the same number of observations Pre-Crisis, Crisis, and Post-Crisis for better comparison across subperiods.

For data cleaning and processing purposes, we filtered the data based on the “Correction Indicator” (CORR) and “Sale Condition”. We kept only transactions where CORR=00; these represent regular trades that were not cancelled or corrected. This resulted in only 0.003% of the data being removed from the sample, leaving us with 523,796,850 trades remaining. We also eliminated any “Special Condition Trades” which introduced suspicious and irregular patterns in the transaction price sequences (i.e. large jumps that were immediately reversed). This resulted in 1.8% of the data being removed from the sample leaving us with 514,270,624 trades remaining.

Since multiple trades can occur in any given second, we next introduced an aggregation step in the data processing. This would allow us to have a second-by-second time series of SPY prices. We tried two methods for aggregation: median and size-weighted average price and did not find significant aberrations. Finally, we had to include an expansion step to account for seconds where no trades were executed. To address these instances we used piecewise constant interpolation; i.e. if there was no trade at second $t$ then we filled it with the last executed price $t-1$ (“last tick”).

---

8We illustrate this principle with simulations in Appendix B. The simulations show that, under our method, sampling at higher frequencies allows for the most precise estimation of integrated volatility. Our method performs better than naive subsampling rules that are typically used in high-frequency studies, and as noted earlier, has the added benefit that the microstructure noise can be autocorrelated and so we need not restrict ourselves to the case where microstructure noise is independent over time.
This resulted in 29,461,859 second-by-second data points covering 1,259 trading days. This was the data that was used to compute our daily time series of monthly realized volatility (on a rolling 21 trading day basis) via the frequency domain estimation methodology.

The daily opening level of the VIX volatility index was obtained from the CBOE database. The VIX is a model-free implied volatility extracted from near-term out-of-the-money put and call options on the S&P 500 index.

The explanatory variables in our regressions also come from multiple sources. First, we have the traditional risk factors from the Fama-French Three Factor Model (Fama and French, 1993); the data for the Fama-French factors are available from Kenneth French’s website. We also include the credit spread, also known as the default risk premium, which is the difference in yield on Baa-rated and Aaa-rated corporate debt. The yields on corporate debt, by Moody’s rating, are available from the FRED database maintained by the Federal Reserve Bank of St. Louis. Use of the credit spread as a risk factor in asset pricing studies goes back to Chen et al. (1986) and has the interpretation as a measure of investor risk aversion. It has subsequently been used in volatility risk premium studies such as Zhou (2011). We will see that there is also a supply-side interpretation of the highly significant effects that the credit spread has on the volatility risk premium (and its magnitude), especially during the Crisis subperiod. The TED spread is included to capture liquidity effects in the financial markets and as a measure of distress in the financial system. The TED spread is the difference between 3-month Eurodollar rates and 3-month Treasury rates, both of which are also available through the FRED database. The interpretation of the TED spread follows from the logic that as uncertainty in the financial system heightens, financial institutions charge more to each other for short-term borrowing—this is reflected in Eurodollar rates; at the same time they require

---

9For details on the methodology used in constructing the VIX volatility index please see CBOE (2009). A similar methodology is employed in Jiang and Tian (2005) where the information content of model-free implied volatility is studied. The CBOE volatility index is studied in Carr and Wu (2006) and Jiang and Tian (2007).

10http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
better collateral, which drives up demand for Treasury Bills and pushes down their rates (Brunnermeier, 2009).

As a final explanatory variable we include the daily open interest for put options on SPY, which is obtained from the OptionMetrics database. We use this as a proxy for investors’ demand for hedging tail risk, following the combined logic of several recent options studies. Cao and Han (2013) use open interest as a proxy for option demand pressure. Additionally, Gârleanu et al. (2009) find that investors are net buyers of index options. Since S&P 500 index put options give investors a way to hedge against market-wide crashes, the open interest provides a natural proxy for investors’ demand to purchase protection and hedge downside tail risk. While options on the SPDR ETF (SPY) are different from index options on the S&P 500 (SPX), they essentially provide the same protection for investors and have some features that may make them more attractive (see Kelly et al., 2012). In fact, our choice to use SPDR options might be even more consistent with our desired proxy as demand for hedging downside tail risk, since the former are American-style options with physical settlement (SPX index options are European-style options with cash settlement only) and therefore give the investor more flexibility and robustness in protecting themselves against market crashes. Furthermore, when we discuss market makers’ delta-hedging, SPDRs would be a more effective hedge on SPY put options than on SPX index options.

The explanatory variables are summarized in Table 1. Descriptive statistics for all variables are given in Table 2, which will be referred to throughout the discussion.

4.2 Construction of the Volatility Risk Premium

In this section we discuss our construction of the volatility risk premium from the market data. Consistent with our representation of the vrp as a “bias”, we calculate it as the deviation of the realized volatility from the expected volatility implied by option prices. Therefore we first compute the realized volatility as the estimated integrated volatility using the Fourier method, described in Section 2, on the ultra-
high-frequency transaction data for the SPDR S&P 500 ETF (ticker \textit{SPY}). This gives us a daily time series of the realized volatility; however, there is a lot of variation in the day-to-day realization of this quantity which will contribute to additional statistical noise in our attempt to quantify the systematic bias that is the \textit{v}rp. To better represent this systematic bias we smooth the time-series of realized volatility by taking a rolling average of the next 21 trading days so as to cover the same month as the contemporaneous \textit{VIX} index, which is our measure of model-free implied volatility. Define this rolling average as \( \{RV_t\}_{t=2006.07}^{2011.06} \), where \( RV_t = \frac{1}{21} \sum_{i=1}^{21} RV_{t+i-1} \) and \( RV_i \) represents the realized volatility for day \( i \) computed using the frequency domain methodology.

The time series of the \textit{VIX} Open value, \( \{VIX_t\}_{t=2006.07}^{2011.06} \) is then subtracted from the average realized volatility to measure the extent that the implied volatility represents a biased expectation of the future realization:

\[
vrp_t = RV_t - VIX_t.
\]

We use the Open value (rather than the Close) of \textit{VIX} so as to be consistent with our realized volatility estimate in terms of the 21-trading-day period for which we are looking at on any given day.\[11\]

A time series of our computed \textit{v}rp over the sample period is plotted in Figure 1. Looking at the figure, two observations stand out immediately: first, the risk premium is negative throughout most of the sample period (we confirmed, statistically, that the \textit{v}rp is not a mean zero process by conducting a series of one-sample t-tests for the whole sample period as well as in each subperiod); second, there are a few instances where the \textit{v}rp goes positive– most notably in the third quarter of 2008. That large positive spike which extended for a period of more than two months seemed to be anomalous to what most of the literature says, so how are we to explain this anomaly in the data? We wanted to rule out the possibility that it is a fictitious by-product of the frequency domain estimation methodology. Fortunately, we were able to confirm

\[11\]For robustness, we ran the regressions using the \textit{VIX} Close value and found that the results (sign, significance) are consistent for the variables of interest in our study.
positive vrp including the third quarter of 2008 using other methods for computing realized volatility (i.e. the TSRV estimator of Zhang et al. (2005) and the pre-averaging estimator of Jacod et al. (2009)).

One possible explanation for this large positive spike is that the option markets underpriced the actual volatility level during that period of time. Unexpected shocks such as Lehman Brothers’ failure and subsequent government interventions kept the markets on edge, and it was impossible to know the magnitude of such a market tsunami and its impacts on realized volatilities, a priori. The idea that the government would provide a backstop against any large financial catastrophe, known as the “Fed put”, was arguably priced into the market keeping implied volatilities low relative to realized volatilities. Therefore, there was a strong bias in one direction with implied volatility underestimating the realized volatility. This eased a bit after Lehman did fail, but when officials were quick to step in thereafter, it remained to keep implied volatilities lower than perhaps they should have been given the circumstances. After the government programs such as TARP and QE were in place and it became clear there would be no “quick fix”, implied volatilities rose relative to realized levels (even though both were rising steadily during this entire period because of the high degree of overall uncertainty) thus reversing the bias.

There is evidence that during this time options dealers went from suppliers of liquidity in the market for hedging tail risk to demanders of liquidity (see Chen et al., 2015). This might help explain the joint effect of sign and magnitude of the volatility risk premium, which we briefly discuss at the end of the paper. First, however, we seek to identify and understand the effect that supply and demand imbalances in this market have on the magnitude of the bias, or the absolute deviation of the realized volatility from the implied volatility. We will see that when the demand for hedging tail risk increases, the deviation between realized and implied volatility widens as well. Similarly, when there is a supply shock (an extreme case would be when dealers switch from suppliers of liquidity to demanders of liquidity) we see that this deviation widens. We find that our proxies for supply and demand side variables are significant, both statistically and economically; furthermore, both the partial effects (changing
supply or demand holding all else constant) and marginal effects (letting the other factors vary as well) produce consistent results. Based on this supply and demand framework, we interpret the magnitude of the volatility risk premium as the price for hedging downside tail risk. The direction of the volatility risk premium, or the sign of the bias, we find reflects different information which may also be related to the role of intermediaries (i.e. options dealers) in the market for hedging downside tail risk.

5 Empirical Analysis

Now we empirically examine the properties of the volatility risk premium as a priced bias between realized and expected volatility. Following our deviation analysis approach, we isolate the magnitude of the bias (absolute value) from the direction (sign) and are able to identify how they reflect different economic forces. First, we analyze the magnitude of the volatility risk premium – the absolute deviation – in relationship to traditional risk factors as well as additional factors to capture supply and demand imbalances in the market for hedging downside tail risk. As is common practice in empirical finance studies, we first control for the known market risk factors, and then include proxies for investor demand for hedging tail risk and financial intermediaries’ willingness to take on risk, provide liquidity, and meet that demand (i.e. the supply side). Investors who wish to hedge their downside tail risk or the risk of a market-wide crash, can buy put options on the S&P 500. As discussed in the previous section, they can do this using either index put options (SPX) or put options on the SPDR ETF which tracks the S&P 500 (SPY put options). We use open interest on the latter as our proxy for investors’ demand for hedging tail risk. The supply side effects are more subtle. On the one hand, there is the well known TED spread which is often a proxy for market liquidity and a measure of distress in the financial sector (see Brunnermeier, 2009). Within the framework of our analysis it could even be interpreted as a net cost of funding for financial institutions who borrow overnight and pledge T-bills as collateral. We do find that the TED spread has moderate explanatory power at varying points in time over the sample period. The explanatory variable that has
more significance and is linked to supply side forces is the credit spread. We find that the credit spread is highly significant and explains the variation in the magnitude of the volatility risk premium quite well. While the credit spread has traditionally been used in empirical finance studies as a proxy for risk aversion (which is not unrelated to the volatility risk premium), we use an interpretation that has appeared in the recent financial economics literature that relates the widening or tightening of credit spreads with “de-leveraging” or increased risk-taking, respectively, of large dealer banks. This is consistent with the hypothesis that the magnitude of the deviation between the realized and expected volatility reflects supply-side imbalances. The evidence in favor of this hypothesis is compelling, especially during the Crisis period.

After a detailed examination of the factors that are priced into the magnitude of the bias that is the volatility risk premium, we turn our attention to the direction or sign. The visual evidence in Figure 1 clearly suggests that the $\text{vrp}$ is usually negative, and statistical tests reject the null hypothesis that it is a mean zero process.\footnote{The exact results are not reported to save space, but they are available from the authors upon request.} However, it is not clear why, if the volatility risk premium is usually negative, it turns positive most notably in the midst of the Financial Crisis (e.g. December 2007 through January 2008, and August 2008 through October 2008). One hypothesis is that the sign of the $\text{vrp}$ reflects the market’s expectation about the future trajectory of realized volatility; i.e. whether it is expected to increase or decrease. While there is some evidence to support this hypothesis, the presence of proven statistical issues calls the results into question. However, we do find evidence in favor of the other hypothesis which is inspired by finance literature on derivatives pricing and volatility which suggests that the sign of the volatility risk premium reflects gains or losses on delta hedges.

5.1 Magnitude Regressions

Our econometric specification is to regress the $\text{magnitude}$ component of the volatility risk premium on the variables to see the extent to which the absolute value of deviation...
between realized and implied volatility is explained by supply and demand imbalances in the market for hedging tail risk. The general regression specification is

\[ |\text{vrp}_t| = \beta_0 + \beta_1 \text{Mkt}_t + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \beta_4 \text{CS}_t + \beta_5 \text{TED}_t + \beta_6 \text{POI}_t + \epsilon_t. \]  \quad (19)

The results for the whole sample period are shown in Table 3. First, we see that the adjusted R-squared is 58% which indicates that, over our entire sample period, the factors are able to explain more than half of the variation in the magnitude of the volatility risk premium. We control for traditional financial market factors with the Fama-French three factors; the only significant factor is the market risk premium (significant at the < 1% level with a p-value of \(7.56 \times 10^{-6}\)). The most significant explanatory variables in the regression are the credit spread and the put option open interest, suggesting that supply and demand imbalances, as per hypotheses H1 and H2, are the driving forces behind the absolute deviation between realized and implied volatility. Both are significant at the < 1% level with p-values of \(3.24 \times 10^{-15}\) and \(2.86 \times 10^{-15}\), respectively.

The economic significance lends convincing support for our hypotheses. The coefficient estimate for \(\text{CS}\) is 5.17; so when the credit spread widens by 100bps, the deviation between the realized and the implied volatility increases by 517bps. The traditional risk aversion story supports this; when market-wide risk aversion is heightened, the BBB yields will rise further above AAA yields, it is natural to expect that the options market will also price this in and increase the premium for hedging downside risk which will result in implied volatilities rising relative to realized volatility. If the \(\text{vrp}\) is negative (i.e. realized volatility is less than implied volatility) this will result in an increase in the absolute deviation between the two. As we will explain below, there is a more compelling explanation for the highly significant – both statistical and economic – effect of credit spreads on the magnitude of the deviation between realized and implied volatility that is robust to whether the \(\text{vrp}\) is positive or negative and it reflects precisely to what we have been referring as supply-side effects. We return to this shortly in our discussion of the “de-leveraging effect” (see Section 5.1.1).
It is important to mention here that when credit spreads widen, other variables can change too. Therefore, it is also useful to perform marginal regressions on the individual factors. The logic behind marginal regressions is as follows. Suppose \( Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon \); when \( X_1 \) increases from \( x_1 \) to \( x_1 + \Delta_1 \), a more plausible scenario is for \( X_2 \) to change from \( \text{E}[X_2|X_1 = x_1] \) to \( \text{E}[X_2|X_1 = x_1 + \Delta_1] \) rather than stay the same. If we treat the conditional expectation \( \text{E}[X_2|X_1] \) as linear in \( X_1 \) (i.e., \( \text{E}[X_2|X_1] = a + bX_1 \)), then the change in \( y \) is \( \Delta_y = (\beta_1 + \beta_2 b)\Delta_1 \) rather than just \( \beta_1 \Delta_1 \), which is equivalent to the marginal regression of \( Y \) on \( X_1 \). That is, if we regress \( Y \) on \( X_1 \) only, then the coefficient before \( X_1 \) is exactly \( \beta_1 + \beta_2 b \). We ran marginal regressions of this type on all of the factors for the entire sample period and each of the subperiods for robustness. We find that the results are consistent with those reported for the multivariate regressions. Some marginal regression results are reported in the text of the paper for robustness, but in the interest of space none have been reported in tables.

The other highly significant variable in the magnitude regression is the put option open interest, which is our measure of demand for hedging downside tail risk. For the entire sample period, the estimated coefficient of put option open interest is \( 6.71 \times 10^{-7} \) and is statistically significant at the < 1% level (p-value of \( 2.86 \times 10^{-15} \)). The interpretation and economic significance of this result is that, over the entire sample period, a 1 million unit increase in hedging demand results in an increase in the magnitude of the volatility risk premium of approximately 67bps. This is what can be considered a demand-side effect, holding supply constant. If we were to allow the supply and other factors to change, then the marginal regressions tell a similar story: a 1 million unit increase in hedging demand results in an increase in the magnitude of the volatility risk premium of approximately 57bps (the coefficient on the \( POI \) marginal regression is \( 5.70 \times 10^{-7} \) and is significant at the < 1% level). For the entire sample period, the data supports H1, both statistically and economically.

Next, we performed the magnitude regression in Equation (19) for three subperiods – “Pre-Crisis”, “Crisis”, and “Post-Crisis” – to see if we can identify any patterns that might coincide with the dramatic changes in financial markets as a result of the
Financial Crisis of 2007-2009. It is difficult to assign start and end points to financial crises, since they are not as clearly defined as business cycles. To best address this, we use the official NBER recession dates which puts our “Crisis” period from December 2007 to June 2009. The results for the magnitude regressions for the three subperiods are reported in Tables 4, 5, and 6 for “Pre-Crisis”, “Crisis”, and “Post-Crisis”, respectively.

An initial comparison of the results in Tables 4, 5, and 6 with those in Table 3 reveals two preliminary observations. First, there is some variation in the explanatory power of the factors over the three subperiods. The R-squared in the “Pre-Crisis” subperiod is at the same level as the R-squared for the whole sample period (around 58%); the R-squared for the “Crisis” subperiod is higher than the whole sample period (64% versus 58%); and the R-squared for the “Post-Crisis” subperiod is lower than the whole sample period (28% versus 58%). It is rather interesting that the best fit is achieved during the Financial Crisis. Over the entire sample period, the most significant explanatory variables are credit spread and put option open interest, thus supporting hypotheses H1 and H2; and there seems to be a trend that these two factors become increasingly more significant over time. In the “Pre-Crisis” subperiod, neither the credit spread nor put option open interest are significant, but the TED spread is the sole significant explanatory variable (at the < 1% level). During the “Crisis” subperiod, the credit spread is highly significant (at the < 1% level) and open interest is significant at the 10% level. In the “Post-Crisis” subperiod, both credit spread and put option open interest are highly significant (at the < 1% level), and the TED spread reappears as minimally significant (< 10% level). Thus, the data suggests that supply and demand factors play an increasing role in explaining the bias between realized volatility and implied volatility. Additionally, since the quality and quantity of data increased over this period, it might also suggest the additional economic insight that can be obtained from in-depth analysis of large scale financial data.

When examining the results from the “Post-Crisis” subperiod one notices that the R-squared (28%) drops off compared to either of the other subperiods or the entire
sample period. However, despite the lower R-squared, it is important to note that both the credit spread and the put option open interest are highly significant. Thus, the economic interpretation of the coefficients is consistent with the findings from the whole sample period and supports our hypotheses. The R-squared is interesting since it indicates perhaps a structural change after the financial crisis. Since the R-squared measures the proportion of the variation in the y-variable (here the magnitude of the vr$p$) that is explained by the x-variables (the risk factors and economic variables) it makes sense to simply look at the descriptive statistics to see if any patterns emerge to which we might attribute a structural change. In the “Post-Crisis” subperiod, both the vr$p$ and its magnitude have less variation (standard deviations of 4.19 and 3.52, respectively) than the entire sample period (standard deviations of 6.91 and 5.63, respectively), so a possible explanation is that some factor(s) lost relatively more of their variability and became less correlated with the vr$p$. We know that early on in the sample period, the TED spread appeared to have good explanatory power for the magnitude component, but that the significance seems to disappear over time. We also see that the mean and standard deviation of the TED spread become very small in the “Post-Crisis” subperiod (mean of 38bps and standard deviation of 17bps). The TED spread was our proxy for liquidity risk and financial market stability, which has a secondary or tertiary effect when analyzing the regression results. However, it is possible that after the Crisis, the low TED spread, with its minimal variation, ceased to be a good proxy for liquidity risk and financial stability. We also saw that over time the put option open interest, our proxy of demand for hedging tail risk, seemed to become increasingly more important. While the mean put option open interest increased over time (4.13mm, 6.59mm, and 10.51mm, for the “Pre-Crisis”, “Crisis”, and “Post-Crisis” subperiods, respectively), which could just be indicative of the growing market for index options and options on ETF’s, its standard deviation falls during the “Crisis” subperiod (from 1.57 to 0.92) and then more than doubles in the “Post-Crisis” period (2.05). It is possible that because of the government’s implicit and explicit backstop – the so-called “Fed put” – demand played less of a role during the Crisis, but as we emerged, new demand for hedging tail risk came to be a key driver in explaining the market price of volatility risk. Therefore, our structural
change could be the increased role of put option open interest and investors’ demand for hedging tail risk rather than liquidity and overall financial stability as being a key factor.

The highest overall explanatory power of the credit spread during the “Crisis” subperiod suggests that supply-side effects drove results during the Financial Crisis. This can, in part, be explained by less variation in the put option open interest (see summary stats and also figure). This may be because of the aforementioned “Fed put” and implicit (or explicit) government bailout, or it could be that dealers went from being liquidity providers to liquidity suppliers themselves. This could help explain the change in the sign of vrп. But there is still a compelling explanation in terms of the magnitude component where the data tells an interesting story that we refer to as the de-leveraging effect.

5.1.1 The “De-leveraging Effect”

During the “Crisis” subperiod, the R-squared of the magnitude regression indicates that the factors are able to explain 64% of the variation in the magnitude component of the volatility risk premium. The most significant explanatory variable is the credit spread, which not only provides validation for the risk aversion interpretation of the volatility risk premium, but also supports our hypothesis that the magnitude of the bias between realized and implied volatility reflects the willingness of market makers to absorb inventory and take risk onto their balance sheet. This is what we refer to as the supply-side effect. When there are supply-side shocks, market makers would be less willing to take on risk and provide liquidity and therefore the wedge between realized and implied volatility grows. Either the dealers would charge a larger premium for providing the liquidity (thus driving the vrп more negative) or become demanders of liquidity (driving the vrп largely positive).

So what is it, exactly, that the credit spread variable is picking up? For many years empirical finance studies have used credit spread as a proxy for risk aversion. We propose an alternative, but certainly not contradictory, interpretation that credit spreads, or the difference between yields on speculative and investment grade debt,
captures a related supply-side effect that is present for the entire sample period and seems to dominate during the “Crisis” subperiod. When a financial institution is concerned with the risk on its balance sheet, one way to reduce the overall risk exposure is to de-leverage. De-leveraging can be achieved through the right-hand-side of the balance sheet by altering the capital structure: buying back debt, issuing equity, or both. However, there is evidence that large financial institutions have a preference to de-leverage through the left-hand-side of the balance sheet: i.e. reducing its holding of risky assets (Adrian and Shin, 2010). Bai and Collin-Dufresne (2011) show that the credit spread explains this effect quite well, particularly during the 2007-2009 Financial Crisis. They present evidence that the large financial institutions classified as primary dealers in the credit markets sold off their holdings of risky corporate debt which would have exerted downward pressure on speculative grade bond prices and increases in yields relative to investment grade debt.

This de-leveraging also justifies the dominant impact of credit spreads in explaining the volatility risk premium during the “Crisis” subperiod in our results, since there is a high degree of overlap in the set of institutions that serve as primary dealers in the credit markets and those that are market makers in index options. During the “Crisis” subperiod, recall that the most statistically significant explanatory variable was the credit spread (see Table 5). The economic significance of the coefficient estimate is that for a 100bps widening of the credit spread we would expect a 624bps increase in the price of volatility risk. As can be seen from Figure 2, credit spreads increased dramatically at the end of 2008. While this can certainly be viewed as an increase in risk aversion, it is also reflective of the massive de-leveraging that occurred after the failure of Lehman Brothers. As large financial institutions reduced their holdings of speculative grade debt they were also reluctant to take on additional risk in other markets. It is reasonable to conclude that during this time dealers in index options increased the price at which they were willing to make a market for hedging downside tail risk, thus increasing the magnitude of the volatility risk premium during this

13Compare the Federal Reserve Bank of New York’s list of primary dealers at http://www.newyorkfed.org/markets/pridealers_current.html with the members of the Options Clearing Corporation that are dealers of index options at http://www.optionsclearing.com/membership/member-information/
time. Because of the inherent leverage in option positions, it is not surprising to see a multiplier effect to the order of 5 to 7 times that of what occurs in the credit markets. This supply-side effect was clearly the driving force behind the increase in the magnitude of the volatility risk premium, not only because it is the only statistically significant explanatory variable during the “Crisis” subperiod, but also because there were no sharp increases in demand for put options on SPRDs. This is illustrated in Figure 3 where open interest fluctuated between 5,000,000 and 10,000,000 when the range for the entire sample period is 2,000,000 to nearly 16,000,000; this can also be seen in Table 2 by comparing the standard deviation for the put open interest variable across subperiods. While it may seem curious that during the most uncertain point in the financial crisis investors were not aggressively trying to hedge their downside tail risk, but there is an intuitive explanation for this: the so-called “Fed put”. After the failure of Lehman Brothers and the subsequent bailout of the financial industry, it became clear that the federal government would provide a backstop either implicitly or explicitly. Therefore, there was little need for investors to pay the high price to hedge their downside tail risk. After the financial crisis, however, demand for hedging downside tail risk returned (see Figure 3) and, indeed, put option open interest was highly significant in the “Post-Crisis” subperiod (see Table 6). This is in addition to the credit spread which still represents the supply-side effect.

Of course, we can also examine what impact this supply-side effect would have on the magnitude of the volatility risk premium if we allow the demand-side effect and other factors to change with it. The marginal regressions indicate that a 100bps increase in the CS results in a 696bps increase in the absolute deviation between realized and implied volatility during the “Crisis” subperiod (the coefficient on the CS marginal regression is 6.96 and is significant at the < 1% level). Thus, when we allow the other factors to change, as they likely did considering the dynamic market events during this time, the de-leveraging effect is even more pronounced.

Wu (2011) shows that the absolute value of the variance risk premium is proportional to the level of volatility in the market. Therefore, the variance risk premium (and consequently the volatility risk premium) is either very negative or very posi-
tive when volatility levels are high. This provides added justification for examining the absolute value of the volatility risk premium to make inferences about the price of volatility risk. More specifically, it suggests that uncertainty in the market increases the bias between realized and implied volatility and the price that must be paid to hedge this risk. In this section we were able to provide empirical evidence explicitly linking this to demand for hedging tail risk and liquidity provision to the volatility market (hypotheses H1 and H2) including results that reflect the massive de-leveraging going on in the markets during the Financial Crisis. Next we examine the direction, or sign, of the volatility risk premium.

5.2 Sign Tests

While a negative volatility risk premium is justified theoretically and for the most part supported by the data, the large positive spike in $v_{rp}$ in the middle of the Financial Crisis as well as several other positive spikes throughout the entire sample period provides a paradox. To help reconcile this paradox within our priced bias interpretation, we now examine the information in the direction, or sign, of the bias as proposed by Hypotheses H3 and H4. Recall, H3 reflects the view of some derivative traders that the direction of the volatility risk premium reflects the market’s expectation of future changes in volatility. When the $v_{rp}$ is negative, then the realized volatility in the equity market is less than the implied volatility extracted from option prices. Volatility is priced higher in the forward-looking options market indicating that market participants expect realized volatility to increase in the future. When the $v_{rp}$ is positive, then the realized volatility in the equity market is greater than the implied volatility extracted from option prices. Volatility is priced lower in the forward-looking options market indicating that market participants expect realized volatility to decrease in the future.

We seek to test Hypothesis H3 using a modified version of the regression proposed in Aït-Sahalia et al. (2012) to test the Expectation Hypothesis. First, to establish a
baseline we run the following specification of the Expectation Hypothesis:

\[
\overline{RV}_t = \alpha + \beta_1 VIX_{t-21} + \epsilon_t. \tag{20}
\]

Here, the y-variable is the average ex-post realized volatility using the frequency domain estimation methodology for the current 21 trading-day period; the x-variable is the 21 trading-day lagged VIX. The idea is to see whether implied volatility does convey information about the market’s expectations about the future realized volatility levels. If implied volatilities are unbiased and efficient estimates of future realized volatility, then we could use the VIX index to predict what the future level of realized volatility will be one month in the future. This is the essence of the Expectation Hypothesis.

The results for this regression can be found in Table 7. The coefficient \( \beta_1 \) is significant at the < 1% level and the R-squared indicates that the lagged VIX is able to explain 33% of the future realized volatility. This suggests that implied volatility does have some predictive power for realized volatility, or in other words, it represents to some extent the market’s expectation about future realized volatility. We note that the t-statistic and p-value associated with \( \beta_1 \) just tells us that the coefficient is statistically different from zero. It is easy to show that \( \beta_1 \) is also statistically less than 1\(^{14}\). The implication is that given the current level of VIX, to predict the future realized volatility you would first discount the level of VIX (by approximately 0.45) and then add a constant (\( \alpha = 5.28\% \)). Furthermore, the results in Table 7 indicate that for VIX > 9.60, implied volatility tends to overestimate future realized volatility; this says that, except when VIX is very low, there should be a negative volatility risk premium. However, we know from examining our time series, that the vrp was positive when VIX was approaching historical highs. Therefore, perhaps a simple linear model such as in Equation (20) does not tell the entire story about the Expectations Hypothesis when it comes to volatilities.

\(^{14}\)Construct the new test statistic \( \hat{t} = \frac{\hat{\beta}_1 - 1}{\text{se}_{\hat{\beta}_1}} = \frac{0.45 - 1}{0.07} = -7.86. \)
To test hypothesis H3, we introduce a binary variable, sgn(vrp) which equals -1 if vrp < 0 and +1 if vrp > 0 and run the modified regression specified as:

$$RV_t = \alpha + \beta_1 VIX_{t-21} + \beta_2 sgn(vrp_{t-21}) + \epsilon_t.$$  

(21)

This attempts to identify whether or not the direction, or the sign, of the volatility risk premium provides additional information about future levels of realized volatility. The results in Table 8 show that including the sign of the vrp improves the explanatory power as the adjusted R-squared is 55% with everything - $\alpha$, $\beta_1$, $\beta_2$ - significant at the < 1% level. Note that $\beta_1$ is still statistically less than 1, but the relationship between the lagged level of VIX and the future level of realized volatility is no longer linear.

The results can be used to predict next month’s realized volatility in terms of the sign of the current volatility risk premium and conditional on the current level on VIX. Suppose the VIX is currently 21 (roughly the median value for our entire sample period). Then our prediction for next month’s realized volatility level depends on whether the vrp is currently positive or negative. If the vrp is negative, then the results of the regression specified by Equation (21) predicts next month’s realized volatility to be approximately 13.43%; and, thus, when the vrp is negative, implied volatility overestimates expected future realized volatility. However, if the vrp is positive, the forecast changes to 29.97% and now implied volatility underestimates expected future realized volatility. This appears to confirm hypothesis H3; similar analysis of higher and lower VIX values (e.g., using 30 and 13, roughly the median value for the “Pre-Crisis” and “Crisis” subperiods, respectively) leads to the same conclusion.

These results should be met with some degree of skepticism. Aït-Sahalia et al. (2012) show that as the forecasting horizon increases, the parameter estimates of the Expectation Hypothesis regression become biased and inefficient. So, while our results do perhaps provide some evidence in favor of hypothesis H3, the statistical inferences may not be sound. Furthermore, it does not say much about the economic meaning of a positive or negative volatility risk premium.
Alternatively, Bakshi and Kapadia (2003) provide an explanation that is more consistent with our motivating theme about market-making and intermediation in the options market. They show, both theoretically and with empirical evidence on index options, that a negative volatility risk premium is representative of the underperformance of a delta-neutral portfolio, where the trader sells calls and purchases $\Delta$ units of the underlying as a hedge (or sells puts and short sells $\Delta$ units of the underlying as a hedge). Since we are examining the vrp in terms of market makers who provide liquidity to investors that wish to hedge downside tail risk with put options on the market (the S&P 500 index or SPDR ETF), the market maker must short sell the underlying (e.g. SPDRs) in order to maintain delta-neutrality. Consequently, the market maker will have a gain on the delta-hedge when the S&P 500 is down and a loss on the delta-hedge when the S&P 500 is up. Table 9 shows the annualized returns on the S&P 500 index when the volatility risk premium is positive (Panel A) and negative (Panel B). We can see that for every period that the vrp is positive, the S&P 500 has negative returns. This is consistent that, in the less frequent instances when the vrp is positive, traders making a market in SPY put options have a profit on their delta-neutral hedge. More often than not, when the vrp is negative, traders making a market in SPY put options are losing money on their delta-neutral hedge as evidenced by the majority of periods that show positive returns on the S&P500, as well as the overall average annualized return of 10.99% when the vrp is negative. This suggests strong evidence in favor of H4.

In order to give more econometric rigor and statistical significance of this relationship posited by hypothesis H4, we ran the following regression:

$$S&P_{return_t} = \beta_0 + \beta_1 \text{sgn}(\text{vrp}_t) + \epsilon_t$$

(22)

where the dependent variable, $S&P_{return_t}$, is the annualized daily return on the S&P500 index and on day $t$ and the independent variable, $\text{sgn}(\text{vrp}_t)$, is the sign of the contemporaneous volatility risk premium; The results for the regression specified by Equation 22 can be found in Table 10. The coefficient estimate for $\beta_1$ is negative and significant at the $< 5\%$ level. We can therefore conclude from this regression.
that returns on the S&P 500 index statistically depend on whether the volatility risk premium is positive or negative. Furthermore, the negative coefficient estimate supports the delta-hedged gains argument of Bakshi and Kapadia (2003), but within the context of put options on the market: when the \( v_{rp} \) is positive, returns on the S&P 500 (or SPDRs) can be expected to be negative (and the delta hedge of being short the underlying will make money), whereas when the \( v_{rp} \) is negative, returns on the S&P 500 (or SPDRs) can be expected to be positive (and the delta hedge of being short the underlying will lose money).

It turns out that this interpretation of the sign of the bias between realized and implied volatility is not unrelated to the theme of intermediation in the market for hedging downside tail risk. There is a possible way to unify both the magnitude and direction of the priced bias interpretations of the volatility risk premium, based on encouraging new evidence in a paper by Chen et al. (2015). Using detailed data on trading in deep out-of-the-money S&P 500 put options Chen et al. (2015) are able to track variation in market-makers’ willingness and ability to provide liquidity to investors seeking insure against market crashes (what we refer to as “hedging downside tail risk”). They find that during the Financial Crisis, market makers went from being liquidity providers to liquidity demanders. To the extent that this period coincides with the period where the \( v_{rp} \) goes sharply positive in Figure 1, it is possible that this flip in sign along with the spike in magnitude represents the restricted supply of liquidity to the market for hedging tail risks in conjunction with dealers’ desire to switch to the other side of the market and becoming liquidity demanders. Since our proxy for demand is open interest, which was rather constant during this period, our analysis might not be picking up that effect. This would just further solidify the intermediary’s role in driving the deviation between realized and implied volatility. However, it would require considerable additional examination to properly weave these two aspects together and, thus, we leave this for future research. Our deviation analysis of the magnitude and direction of the volatility risk premium separately as two components of a systematically-priced bias provides the first steps needed to better understand how the market prices volatility.
6 Conclusion

“Big Data” has the potential to transform research in many areas, including financial economics and risk analysis. We use a massive data set, collected from numerous sources, to perform a unique study of how the market prices volatility. Our research questions equate the volatility risk premium to a systematically priced bias between ex-post realized volatility and ex-ante expected volatility implied by options. Unlike most other studies of the volatility risk premium, rather than start with a theoretical model of volatility, we begin with intensive data-driven methods leveraging the insights of Big Data analytics.

First, we collected price and volume data on every transaction in SPDRs, the ETF that tracks the S&P 500 index, over a five year period from 2006 through 2011 yielding over half a billion observations. We then use a novel technique to estimate the integrated volatility using the processed ultra-high-frequency data in the frequency domain. This methodology allows us to distinguish the true volatility of the price process and the microstructure noise, even when the noise is correlated over time. The result is a consistent, de-biased estimate of the integrated volatility as our measure of realized volatility. In constructing the time series of the volatility risk premium, we smooth the daily realized volatility by taking a 21-trading-day rolling average and subtract from it the daily value of \( \text{VIX} \) volatility index for the same month.

Insofar as the option implied volatility represents the market’s expectation of future volatility, this formulation of the volatility risk premium is very much like a statistical bias. We decompose this bias into magnitude (absolute value) and direction (sign) and analyze them separately. Based on stylized facts about the volatility risk premium, we construct four testable hypotheses about its economic meaning and determinants. The general theme is that the volatility risk premium cannot be explained by traditional risk factors, but rather are related to supply and demand forces in option markets and the role of market makers in providing liquidity to investors who seek to hedge their downside tail risk. This is all viewed within the lens of the volatility risk premium being systematically priced bias.
The results indicate that the size of this bias represents the price that market makers require to meet the demand of investors who wish to hedge their downside tail risk and compensates for supply and demand imbalances in this market. In fact, we find compelling evidence that during the Financial Crisis, supply-side forces dominated as financial intermediaries shed risky positions and were reluctant to take more risk onto their balance sheets. Demand-side forces dried up as the implicit guarantees and “Fed put” made hedging tail risk less attractive for investors. This is substantiated by the highly significant credit spread, which reflects market makers de-leveraging during the Crisis, and reduced significant in put option open interest, which is our proxy for investors’ demand for hedging tail risk.

Practitioners view the sign of the volatility risk premium, on the other hand, as the market’s expectation about future levels of volatility. This is similar to the Expectation Hypothesis discussed in Aït-Sahalia et al. (2012) and, while we are able to find some evidence in favor of this hypothesis in the data, statistical issues raise doubt on the validity of the inferences and there is no clear economic interpretation within the conceptual framework we established. An alternative hypothesis links the sign of the volatility risk premium to the gains and losses on traders’ delta-hedged positions when making a market for index options. Bakshi and Kapadia (2003) were the first to propose this interpretation, and we are able to find fairly conclusive evidence in favor of it in the market for S&P 500 put options. That is, market makers provide liquidity to investors seeking to hedge their downside tail risk – via put options on SPDRs – will delta-hedge these positions by short-selling shares of the underlying ETF. We find that returns on the S&P 500 are negative over all consecutive trading days where the volatility risk premium is positive, indicating a delta-hedged gain for the market maker. We find that the returns on the S&P 500 are positive over all but two series of consecutive trading days where the volatility risk premium is negative, indicating a delta-hedged loss for the market maker. Regression results corroborate this finding.

Overall, the ability of our data-driven analysis to identify economic insights into how the market prices volatility is very encouraging for researchers interested in us-
ing similar approaches for other quantitative studies in financial economics and risk
analysis. While we do utilize the results from the existing literature along with eco-

nomic intuition to highlight some stylized facts about the volatility risk premium and
come up with testable hypotheses, our analysis does not rely on any specific theo-

retical model and has minimal parametric assumptions. The trend over our sample
period seems to indicate that the increase in high-frequency data will allow for more
precise and accurate measurement of the volatility risk premium as a systematically
priced bias. This will then allow for even better identification of the determinants of
the volatility risk premium and highlight the role that intermediation in the market
for volatility and hedging downside risk as well as the role that supply and demand
imbalance play in driving the deviation between realized and implied volatilities.
Appendix

A Methods for Estimating Integrated Volatility with High Frequency Data

In this Appendix we follow the formulation and notation in Section 2.1 and compare some alternative methodologies for estimating integrated volatility with high frequency data.

Before delving any deeper, let us first look at a simple case where \( \sigma_t \) is constant, and the noise is i.i.d. Gaussian with mean 0 and variance \( a^2 \). Under these simplified assumptions, the log-likelihood function of \( \delta Y_i = Y_{t_i} - Y_{t_{i-1}} \) is

\[
    l(\sigma^2, a^2) = -\frac{1}{2} \log \det(\Omega) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \delta Y' \Omega^{-1} \delta Y, \tag{23}
\]

where \( Y_{t_i}'s \) are observed at regularly spaced times with interval \( \Delta = t_i - t_{i-1} \), \( n = \frac{T}{\Delta} \) is the number of sampled data points, and \( \Omega \) is the covariance matrix of \( \delta Y \). Choosing \( \sigma \) and \( a \) to maximize Equation (23) gives the Maximum Likelihood Estimator, or MLE, of volatility. It can be shown that the MLE is consistent for both the volatility component \( \sigma^2 \) and the noise component \( a^2 \) at rates \( O_p(n^{-1/4}) \) and \( O_p(n^{-1/2}) \), respectively. Moreover, misspecification of the marginal distribution of \( \epsilon \) does not have adverse consequences (A"it-Sahalia et al. (2005), Xiu (2010)).

The assumption that volatility is constant is probably not very reasonable – there is considerable evidence of time-varying volatility. As a result, we are often interested in estimating the integrated volatility over a period of time.\(^{15}\) This is done by making use of the quadratic variation, \( \langle X, X \rangle_T \), of the stochastic process described by Equation (1):

\[
    \langle X, X \rangle_T = \int_0^T \sigma_t^2 \, dt. \tag{24}
\]

\(^{15}\)There is also research on estimating the instantaneous volatility process \( \sigma_t \) either parametrically or nonparametrically. See, e.g., Andersen et al. (2004).
A naïve estimator would be the Realized Volatility (RV) estimator

\[ [Y, Y]_T = \sum_{i=1}^{n} (Y_{t_{i+1}} - Y_{t_i})^2, \quad (25) \]

which is a consistent estimator in a noise-free model. However, since the observable price process given by Equation (2) is contaminated by the microstructure noise this RV estimator is both biased and inconsistent.

Statistical theory indicates that we should be able to improve the accuracy and precision of our estimate by increasing the rate at which we sample the data; hence the value of ultra-high-frequency data.\(^\text{16}\) This would be the case if we could observe \(X_t\) directly; but microstructure noise introduces an added dimension of complexity to the problem. In fact, assuming iid noise, the bias of the RV estimator is \(2nE[\epsilon^2]\). This tells us that as we increase the frequency of the price data, the effect from noise becomes more overwhelming.

One way to address the problem of noise when sampling at too high of a frequency is to sample sparsely and use the corresponding RV estimator. This practice, known as the subsampling approach, was first introduced by Zhou (1996). However, even when sampling sparsely at the optimally-determined frequency, the fact that large portions of data are discarded violates basic statistical principles. Furthermore, Zhang et al. (2005) argue that sampling over longer horizons merely reduces the impact of microstructure, rather than quantifying and correcting its effect for volatility estimation.

One of the earliest solutions to incorporate the full data sample is Two Scales Realized Volatility (TSRV) as proposed by Zhang et al. (2005). The TSRV estimator is based on subsampling, averaging, and bias-correction. They sample sparsely over subgrids of \(n\) observations to get \(K\) subsamples on a slower time scale. For each such sample the RV estimator is \([Y, Y]_{T}^{(\text{sparse}, k)}, k = 1, \cdots, K\), and averaging them yields

\(^{16}\)See Appendix B where we use simulations to illustrate the benefit of using ultra-high-frequency data.
the estimator \([Y, Y]_T^{(\text{avg})}\). The final de-biased estimator is:

\[
\hat{\langle X, X \rangle}_T^{(\text{tsrv})} = [Y, Y]_T^{(\text{avg})} - \frac{n}{\bar{n}}[Y, Y]_T,
\]

(26)
after accounting for the bias, where \(\bar{n} = \frac{n}{K}\). Choosing the optimal sampling step \(K = cn^{2/3}\) yields the convergence rate \(n^{-1/6}\). The TSRV estimator is shown to outperform the standard RV estimator empirically in the study by Aït-Sahalia and Mancini (2008).

A closely related estimator is Multiple Scale Realized Volatility (MSRV), which is proposed and derived in Zhang (2006). As a generalization of TSRV, the MSRV estimator combines \(M\) different time scales with weights, when chosen optimally, can achieve the optimal convergence rate \(n^{-1/4}\). Based on a different smoothing idea, Fan and Wang (2007) introduces a different estimator achieves the same rate, but allows jumps in the price processes.

Realized kernels, which are based on linear combination of autocovariances, represent another popular class of estimators. Barndorff-Nielsen et al. (2008) designed several realized kernels which are robust to endogenous sampling and noise. The realized kernel estimators can achieve convergence rates up to that of MSRV. Further work in this direction includes Barndorff-Nielsen et al. (2009) and Barndorff-Nielsen et al. (2011).

The pre-averaging approach of Jacod et al. (2009) uses all or most of the data, and removes noise by local average smoothing, i.e., averaging over a moving window. The result is a rate optimal (with convergence rate \(n^{-1/4}\)) consistent estimator of integrated volatility in the presence of microstructure noise.

Returning to the parametric approach, if we do not assume that volatility is constant and noise normally distributed with variance \(\sigma^2\), but nevertheless use the log-likelihood function in Equation (23), the resulting estimator is the quasi maximum likelihood estimator (QMLE). Interestingly, it is still a consistent estimator at the most efficient rate \(n^{-1/4}\). Statistical properties of the QMLE are derived in Xiu (2010).
B Simulations

We compare the performance of Fourier method and naive subsampling at different sampling frequency on simulated data using a Heston (1993) model:

\[
\begin{align*}
\frac{dX_t}{t} &= (\mu - \nu_t/2)dt + \sigma_t dB_t, \\
\frac{d\nu_t}{t} &= \kappa(\alpha - \nu_t)dt + \gamma \nu_t^{1/2} dW_t,
\end{align*}
\]

where \( \nu_t = \sigma_t^2 \). The parameters are set as follows: \( \mu = 0.05, \kappa = 5, \alpha = 0.04, \gamma = 0.5 \), and the correlation between the two Brownian motions \( B_t \) and \( W_t \) is \( \rho = -0.5 \). The initial values are \( X_0 = 0 \) and \( \nu_0 = 0.04 \). We take \( T \) as one day, and simulate data with \( \Delta_t = 0.1s \), which yields a sample path of length \( N = 234,000 \) in one trading day. We first calculate the underlying true integrated volatility by a Riemann sum approximation of the integral, i.e.: \( T \sum_{i=1}^{N} \sigma_i^2 = \int_0^T \sigma_t^2 dt \). Then we add AR(2) noise \( \epsilon_i = 0.6\epsilon_{i-1} - 0.4\epsilon_{i-2} + \eta_i \), to get the observed data \( Y_i = X_i + \epsilon_i \), where \( \eta_i \)'s are i.i.d. \( \mathcal{N}(0, \sigma^2_\eta) \) and we set \( \sigma_\eta = 5 \times 10^{-4} \).

We estimate the integrated volatility using two methods, the Fourier method and the naive subsampling, which yields \( \langle X, X \rangle_T^{\text{Fourier}} \) and \( \langle X, X \rangle_T^{\text{subsampling}} \). We calculate the RMSE (root-mean-square error) of the estimates to the truth over 200 simulated sample paths. To further illustrate the effect of high frequency data, we evaluate two methods from \( \Delta_t = 1s \) up to \( \Delta_t = 150s \). Figure 4 shows the RMSE of the Fourier method and the naive subsampling against decreasing sampling frequencies.

The takeaway of this figure is two folds. First, the Fourier method can effectively filter the correlated microstructure noise, and works better than naive sampling method. We did not implement other more sophisticated methods for comparison, as the simulation is not to illustrate the superiority of the Fourier method, but rather to justify the use of high frequency data. Second, if we can filter the microstructure noise, higher frequency gives us a better estimate as we are able to utilize more data hence more information.

\footnote{These are the same as those used in the Olhede et al. (2009) simulations.}
References

Adrian, T. and Shin, H. S. (2010), “Liquidity and leverage,” *Journal of Financial Intermediation*, 19, 418 – 437.

Aït-Sahalia, Y., Karaman, M., and Mancini, L. (2012), “The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis,” Available at SSRN: http://ssrn.com/abstract=2136820.

Aït-Sahalia, Y. and Mancini, L. (2008), “Out of sample forecasts of quadratic variation,” *Journal of Econometrics*, 147, 17 – 33.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005), “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise,” *Review of Financial Studies*, 18, 351–416.

— (2011), “Ultra high frequency volatility estimation with dependent microstructure noise,” *Journal of Econometrics*, 160, 160 – 175.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2004), “Parametric and Non-parametric Volatility Measurement,” in L.P. Hansen and Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics*, North-Holland.

Bai, J. and Collin-Dufresne, P. (2011), “The Determinants of the CDS-Bond Basis During the Financial Crisis of 2007-2009,” Available at SSRN: http://ssrn.com/abstract=2024531.

Bakshi, G. and Kapadia, N. (2003), “Delta-Hedged Gains and the Negative Market Volatility Risk Premium,” *Review of Financial Studies*, 16, 527–566.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), “Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76, 1481–1536.

— (2009), “Realized kernels in practice: trades and quotes,” *Econometrics Journal*, 12, C1–C32.
Black, F. and Scholes, M. (1973), “The pricing of options and corporate liabilities,” 
Journal of Political Economy, 81, 637–654.

Bollerslev, T., Gibson, M., and Zhou, H. (2011), “Dynamic estimation of volatility 
risk premia and investor risk aversion from option-implied and realized volatilities,” 
Journal of Econometrics, 160, 235 – 245.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009), “Expected Stock Returns and 
Variance Risk Premia,” Review of Financial Studies, 22, 4463–4492.

Britten-Jones, M. and Neuberger, A. (2000), “Option Prices, Implied Price Processes, 
and Stochastic Volatility,” The Journal of Finance, 55, 839–866.

Brunnermeier, M. K. (2009), “Deciphering the Liquidity and Credit Crunch 2007-2008,” Journal of Economic Perspectives, 23, pp. 77–100.

Buraschi, A., Trojani, F., and Vedolin, A. (2014), “When Uncertainty Blows in the 
Orchard: Comovement and Equilibrium Volatility Risk Premia,” The Journal of 
Finance, 69, 101–137.

Canina, L. and Figlewski, S. (1993), “The informational content of implied volatility,” 
Review of Financial studies, 6, 659–681.

Cao, J. and Han, B. (2013), “Cross section of option returns and idiosyncratic stock 
volatility,” Journal of Financial Economics, 108, 231–249.

Carr, P. and Wu, L. (2006), “A Tale of Two Indices,” The Journal of Derivatives, 13, 
13–29.

— (2009), “Variance Risk Premiums,” Review of Financial Studies, 22, 1311–1341.

CBOE (2009), “The CBOE Volatility Index - VIX,” White paper, Chicago Board 
Options Exchange, Chicago, IL.
Chen, H., Joslin, S., and Ni, S. X. (2015), “Demand for Crash Insurance, Intermediate Constraints, and Risk Premia in Financial Markets,” Available at SSRN: http://ssrn.com/abstract=2024416.

Chen, N.-F., Roll, R., and Ross, S. A. (1986), “Economic Forces and the Stock Market,” The Journal of Business, 59, pp. 383–403.

Christensen, B. J. and Prabhala, N. R. (1998), “The relation between implied and realized volatility,” Journal of Financial Economics, 50, 125–150.

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

Derman, E. and Kani, I. (1994), “Riding on a smile,” Risk, 7, 32–39.

Dupire, B. (1994), “Pricing with a smile,” Risk, 7, 18–20.

Fama, E. F. and French, K. R. (1993), “Common risk factors in the returns on stocks and bonds,” Journal of Financial Economics, 33, 3 – 56.

Fan, J. (2013), “Features of big data and sparsest solution in high confidence set,” Past, Present, and Future of Statistical Science, 507–523.

Fan, J., Han, F., and Liu, H. (2014), “Challenges of Big Data analysis,” National Science Review, 1, 293–314.

Fan, J. and Wang, Y. (2007), “Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data,” Journal of the American Statistical Association, 102, 1349–1362.

Gărleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009), “Demand-Based Option Pricing,” Review of Financial Studies, 22, 4259–4299.

Heston, S. L. (1993), “A closed-form solution for options with stochastic volatility with applications to bond and currency options,” Review of Financial Studies, 6, 327–343.
Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009), “Microstructure noise in the continuous case: The pre-averaging approach,” *Stochastic Processes and their Applications*, 119, 2249 – 2276.

Jiang, G. J. and Tian, Y. S. (2005), “The Model-Free Implied Volatility and Its Information Content,” *Review of Financial Studies*, 18, 1305–1342.

— (2007), “Extracting Model-Free Volatility from Option Prices,” *The Journal of Derivatives*, 14, 35–60.

Kelly, B. T., Lustig, H., and Van Nieuwerburgh, S. (2012), “Too-systemic-to-fail: What option markets imply about sector-wide government guarantees,” Fama-Miller Working Paper; Chicago Booth Research Paper No. 11-12. Available at SSRN: http://ssrn.com/abstract=1762312.

Merton, R. C. (1973), “Theory of rational option pricing,” *The Bell Journal of Economics and Management Science*, 4, 141–183.

Mixon, S. (2009), “Option markets and implied volatility: Past versus present,” *Journal of Financial Economics*, 94, 171–191.

Nagel, S. (2012), “Evaporating Liquidity,” *Review of Financial Studies*, 25, 2005–2039.

Olhede, S., Sykulski, A., and Pavliotis, G. (2009), “Frequency Domain Estimation of Integrated Volatility for Ito Processes in the Presence of Market-Microstructure Noise,” *SIAM - Multiscale Modeling and Simulation*, 8, 393–427.

Rubinstein, M. (1994), “Implied binomial trees,” *The Journal of Finance*, 49, 771–818.

Shepp, L. (2007), “Statistical thinking: From Tukey to Vardi and beyond,” in *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, Institute of Mathematical Statistics, pp. 268–273.

Wu, L. (2011), “Variance dynamics: Joint evidence from options and high-frequency returns,” *Journal of Econometrics*, 160, 280 – 287.
Xiu, D. (2010), “Quasi-maximum likelihood estimation of volatility with high frequency data,” *Journal of Econometrics*, 159, 235 – 250.

Zhang, L. (2006), “Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach,” *Bernoulli*, 12, pp. 1019–1043.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data,” *Journal of the American Statistical Association*, 100, pp. 1394–1411.

Zhou, B. (1996), “High-Frequency Data and Volatility in Foreign-Exchange Rates,” *Journal of Business & Economic Statistics*, 14, pp. 45–52.

Zhou, H. (2011), “Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty,” Available at SSRN: http://ssrn.com/abstract=1783624.
Figure 1: Time Series of the Volatility Risk Premium
Figure 2:
Top: Time series of the absolute value of the volatility risk premium or the *magnitude component* of the vrp
Bottom: Time series of the credit spread, defined as the yield on Baa-rated and Aaa-rated corporate debt.
Note: The vertical lines separate the three subperiods.
Figure 3: Open interest for put options on the S&P500 index over the sample period. Put open interest is our proxy for investor demand for hedging downside tail risk.
Figure 4: RMSE (shown in logarithmic scale) of two methods
| Variable Name                      | Description                                                                 |
|-----------------------------------|-----------------------------------------------------------------------------|
| \(r_m - r_f\) (Mkt)              | Fama-French market risk factor; market return minus risk-free rate           |
| SMB                               | Fama-French size factor                                                    |
| HML                               | Fama-French value factor                                                   |
| Credit Spread (CS)                | Difference in yield on Baa-rated and Aaa-rated corporate debt               |
| TED Spread (TED)                  | Difference between 3-month Eurodollar rate and 3-month Treasury rate        |
| Put Open Interest (POI)           | Daily open interest for put options on SPY                                  |

Table 1: Explanatory variables

|                  | Whole-Sample (N = 1238) | Pre-Crisis (N = 338) | Crisis (N = 375) | Post-Crisis (N = 525) |
|------------------|-------------------------|----------------------|------------------|-----------------------|
|                  | Mean | Med | Std  | Mean | Med | Std  | Mean | Med | Std  | Mean | Med | Std  |
| vrp              | -8.10 | -8.32 | 6.91 | -4.53 | -3.86 | 3.82 | -9.17 | -9.05 | 10.20 | -9.63 | -10.00 | 4.19 |
| abs(vrp)         | 9.03 | 8.68 | 5.63 | 4.86 | 3.88 | 3.39 | 11.58 | 10.58 | 7.34 | 9.90 | 10.00 | 3.52 |
| VIX              | 24.20 | 21.88 | 11.70 | 15.56 | 13.54 | 5.16 | 34.58 | 28.19 | 14.68 | 22.34 | 21.84 | 4.86 |
| Mkt              | 0.02 | 0.10 | 1.60 | 0.04 | 0.11 | 0.86 | -0.09 | -0.01 | 2.45 | 0.09 | 0.14 | 1.10 |
| SMB              | 0.02 | 0.02 | 0.64 | -0.02 | -0.04 | 0.42 | -0.02 | -0.03 | 0.88 | 0.03 | 0.05 | 0.54 |
| HML              | -0.01 | -0.01 | 0.73 | -0.03 | -0.03 | 0.26 | -0.00 | 0.01 | 1.13 | 0.01 | -0.01 | 0.55 |
| CS               | 1.35 | 1.09 | 0.68 | 0.90 | 0.90 | 0.05 | 2.07 | 1.56 | 0.80 | 1.12 | 1.07 | 0.25 |
| TED              | 0.87 | 0.44 | 0.88 | 0.61 | 0.40 | 0.51 | 1.78 | 1.38 | 1.03 | 0.38 | 0.33 | 0.17 |
| POI (×10^6)      | 7.58 | 7.04 | 3.15 | 4.13 | 3.86 | 1.57 | 6.59 | 6.50 | 0.92 | 10.51 | 10.65 | 2.05 |

Table 2: Descriptive statistics for all variables across all periods.
| Variable | Estimate | Std Error | t-value | Pr(>|t|) | Significance |
|----------|----------|-----------|---------|----------|--------------|
| (Intercept) | -3.64 | 1.01 | -3.62 | 0.30e-03 | *** |
| Mkt | 0.42 | 0.09 | 4.50 | 7.56e-06 | *** |
| SMB | -0.11 | 0.15 | -0.72 | 0.47 | |
| HML | -0.13 | 0.14 | -0.94 | 0.35 | |
| CS | 5.17 | 0.65 | 7.98 | 3.24e-15 | *** |
| TED | 0.69 | 0.53 | 1.30 | 0.19 | |
| POI | 6.17e-07 | 8.39e-08 | 8.00 | 2.86e-15 | *** |

**Table 3:** Regression of $|vrp|$ on explanatory variables: Whole Sample Period. Standard errors and t-statistics are computed using the Newey-West correction. Significance levels are < 1%, < 5%, and < 10% for ***, **, and *, respectively.

| Variable | Estimate | Std Error | t-value | Pr(>|t|) | Significance |
|----------|----------|-----------|---------|----------|--------------|
| (Intercept) | -4.74 | 7.42 | -0.64 | 0.52 | |
| Mkt | 0.16 | 0.15 | 1.01 | 0.31 | |
| SMB | 0.17 | 0.33 | 0.52 | 0.61 | |
| HML | 0.42 | 0.62 | 0.67 | 0.50 | |
| CS | 7.55 | 8.10 | 0.93 | 0.35 | |
| TED | 5.63 | 0.88 | 6.41 | 4.93e-10 | *** |
| POI | -1.63e-07 | 2.95e-07 | -0.55 | 0.58 | |

**Table 4:** Regression of $|vrp|$ on explanatory variables: Pre-Crisis Subperiod. Standard errors and t-statistics are computed using the Newey-West correction. Significance levels are < 1%, < 5%, and < 10% for ***, **, and *, respectively.
| Variable | Estimate | Std Error | t-value | Pr(>|t|) | Significance |
|----------|----------|-----------|---------|----------|--------------|
| (Intercept) | -10.04 | 4.03 | -2.49 | 0.01 | ** |
| Mkt | 0.48 | 0.11 | 4.25 | 2.74e-05 | *** |
| SMB | 0.14 | 0.18 | 0.79 | 0.43 | |
| HML | -0.16 | 0.19 | -0.85 | 0.39 | |
| CS | 6.24 | 0.74 | 8.44 | 7.35e-16 | *** |
| TED | 0.90 | 0.67 | 1.34 | 0.18 | |
| POI | 1.08e-06 | 5.81e-07 | 1.87 | 0.06 | * |
| Adjusted R² | | | | 0.64 | |

*Table 5:* Regression of $|vrp|$ on explanatory variables: Crisis Subperiod. Standard errors and t-statistics are computed using the Newey-West correction. Significance levels are < 1%, < 5%, and < 10% for ***, **, and *, respectively.

| Variable | Estimate | Std Error | t-value | Pr(>|t|) | Significance |
|----------|----------|-----------|---------|----------|--------------|
| (Intercept) | -6.40 | 3.14 | -2.04 | 0.04 | ** |
| Mkt | 0.05 | 0.17 | 0.32 | 0.75 | |
| SMB | -0.25 | 0.27 | -0.96 | 0.34 | |
| HML | 0.29 | 0.34 | 0.86 | 0.39 | |
| CS | 11.67 | 2.27 | 5.13 | 4.04e-07 | *** |
| TED | -5.22 | 3.01 | -1.73 | 0.08 | * |
| POI | 4.95e-07 | 1.69e-07 | 2.92 | 3.67e-03 | *** |
| Adjusted R² | | | | 0.28 | |

*Table 6:* Regression of $|vrp|$ on explanatory variables: Post-Crisis Subperiod. Standard errors and t-statistics are computed using the Newey-West correction. Significance levels are < 1%, < 5%, and < 10% for ***, **, and *, respectively.
Table 7: Regression to test the Expectation Hypothesis. Standard errors and t-statistics are computed using the Newey-West correction. Significance levels are < 1%, < 5%, and < 10% for ***, **, and *, respectively.

| Variable       | Estimate | Std Error | t-value | Pr(>|t|)  | Significance |
|----------------|----------|-----------|---------|-----------|--------------|
| (Intercept)    | 5.28     | 1.71      | 3.098   | 2.00e-03  | ***          |
| VIX_{t-21}     | 0.45     | 0.07      | 6.80    | 1.60e-11  | ***          |

Adjusted $R^2$: 0.33

Table 8: Regression to test the Modified Expectation Hypothesis (H3). Standard errors and t-statistics are computed using the Newey-West correction. Significance levels are < 1%, < 5%, and < 10% for ***, **, and *, respectively.

| Variable       | Estimate | Std Error | t-value | Pr(>|t|)  | Significance |
|----------------|----------|-----------|---------|-----------|--------------|
| (Intercept)    | 11.62    | 3.28      | 3.54    | 4.10e-04  | ***          |
| VIX_{t-21}     | 0.48     | 0.05      | 8.76    | < 2.2e-16 | ***          |
| sgn(vrp_{t-21})| 8.27     | 3.13      | 2.64    | 8.39e-03  | ***          |

Adjusted $R^2$: 0.55
### Positive vrp Dates

| Dates           | Trading Days | Avg(vrp) | S&P500 Return |
|-----------------|--------------|----------|---------------|
| 2007/02/02 - 2007/02/27 | 17           | 1.49     | -47.81%       |
| 2007/07/12 - 2007/07/26 | 11           | 2.76     | -53.34%       |
| 2007/12/21 - 2008/01/02 | 7            | 1.17     | -30.65%       |
| 2008/08/15 - 2008/10/08 | 38           | 11.69    | -169.62%      |
| 2010/04/08 - 2010/05/04 | 19           | 3.63     | -8.37%        |

Average S&P500 return when vrp is positive: **-88.13%**

(a) Panel A: S&P500 returns (annualized) when vrp is positive

### Negative vrp Dates

| Dates           | Trading Days | Avg(vrp) | S&P500 Return |
|-----------------|--------------|----------|---------------|
| 2006/07/31 - 2007/02/01 | 128          | -3.88    | 24.50%        |
| 2007/02/28 - 2007/07/11 | 93           | -3.72    | 22.89%        |
| 2007/07/27 - 2007/12/20 | 103          | -8.18    | -1.71%        |
| 2008/01/03 - 2008/08/14 | 156          | -6.52    | -15.89%       |
| 2008/10/09 - 2010/04/07 | 375          | -13.40   | 18.56%        |
| 2010/05/05 - 2011/06/29 | 291          | -9.88    | 10.42%        |

Average S&P500 return when vrp is negative: **10.99%**

(b) Panel B: S&P500 returns (annualized) when vrp is negative

**Table 9:** Test of Delta-Hedged Gain/Loss Hypothesis
| Variable     | Estimate | Std Error | t-value | Pr(>|t|) | Significance |
|--------------|----------|-----------|---------|---------|--------------|
| (Intercept)  | -0.39    | 0.23      | -1.66   | 0.10    | *            |
| sgn(vrp)     | -0.50    | 0.22      | -2.21   | 0.03    | **           |

Adjusted $R^2$ 3.47e-03

**Table 10:** Regression of S&P 500 return on the direction of vrp. Standard errors and t-statistics are computed using the Newey-West correction. Significance levels are < 1%, < 5%, and < 10% for ***, **, and *, respectively.