The Risk Premium That Never Was: A Fair Value Explanation of the Volatility Spread

Richard J. McGee*

Centre for Computational Finance and Business Analytics, University of Southampton, UK

Frank McGroarty

Centre for Computational Finance and Business Analytics, University of Southampton, UK

Abstract

We present a new framework to investigate the profitability of trading the volatility spread, the upward bias on implied volatility as an estimator of future realized volatility. The scheme incorporates the first four option-implied moments in a growth-optimal payoff that is statically replicated using a portfolio of options. Removing the upward bias on implied volatility worsens the likelihood score of risk neutral densities obtained from S&P500 index options when they are used as forecasts of the underlying index return distribution. It also results in negative expected capital growth when they are used in a volatility arbitrage scheme. Our empirical finding is that the upward bias on implied volatility does not represent a long term return premium, rather it is required to mitigate the large losses associated with tail events when trading volatility in options markets.

Keywords: Finance, Volatility Spread, Variance Premium, Tail Risk, Growth Optimal Portfolios

1. Introduction

An upward bias in option-implied volatility (IV) as an estimate of future realized volatility (RV) has been consistently documented by researchers in a range of options markets (see e.g. Day and Lewis (1992), Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Jorion (1995), Bates (1996), Christensen and Prabhala (1998) and Jiang and Tian (2005)). Dating back to Lamoureux and Lastrapes (1993), the upward bias on IV has traditionally been measured using a regression-based approach. The regression-based test for bias is usually of the form:

$$RV(t,\tau) = a + \lambda IV(t,\tau) + \epsilon(t,\tau), \tag{I}$$

where, RV is the realized volatility of the underlying over the lifetime of the option from time t to time τ and IV is the implied volatility for the option at time t. A test for a=0, λ=1 is used to test for a bias on IV as an estimate of RV. In this article we use the second moment of the risk neutral density implied by a full strip of options as our measure of IV and

*Corresponding author
Email address: R.J.McGee@soton.ac.uk (Richard J. McGee)

define the volatility spread as the difference between this value and the volatility realized over the life of the option from time t to expiry, τ.

The upward bias on implied volatility is often described as a premium (see e.g. Carr and Wu (2009)), and has been show to have an associated return premium in equity markets (see Bali and Hovakimian (2009)). The spread has also been linked with uncertainty (Drechsler (2013)) and has been widely used as an economic uncertainty measure (see e.g. Bloom (2009), Bekaert et al. (2013) and Longstaff et al. (2007)).

Estimating the model in Equation I through ordinary least squares regression assumes that the cost of the volatility forecast errors, $\epsilon(t,\tau)$, is symmetrical and ignores the impact of the higher moments of the return distribution, such as skewness and kurtosis. If these assumptions were valid then a bias on implied volatility would represent a statistical arbitrage opportunity for an options trader. We test for a return premium associated with the volatility spread in the S&P 500 options markets, over a period of 17.5 years and a data set of over 2,150 option sets.

We use a new framework for investigating the economic importance of the volatility spread. The framework allows the spread to be examined in terms of its contribution to a full forecast density, including the first four moments of the risk neutral distribution. This is a key aspect when investigating the spread because real-world volatility arbitrage schemes have exposure to the higher moments of the underlying return distribution, and this will be reflected in the price of volatility in options markets. Delta hedging schemes, for example, have an exposure to non-zero excess skewness and kurtosis. So much so that the disappearance of the market for single name variance swaps in 2008 is attributed to the difficulty in accurately delta hedging the associated options (see Carr and Lee (2009)).

The proposed trading scheme offers a number of benefits over alternative volatility arbitrage schemes such as delta hedging or butterfly spreads:

- The scheme uses a static replication portfolio to trade volatility. Unlike delta hedging there are no continuous hedging trades and the bid/ask spread is only paid on one transaction as options are held to expiry.

- The resulting option portfolio is the growth optimal portfolio[1] given a forecast distribution that is closer to the true underlying distribution than the market risk neutral distribution is. Investing in this portfolio maximizes expected growth for a given informational advantage. Other static portfolios, such as a butterfly spread, are not growth optimal. If we view a butterfly spread as a portfolio of options, the portfolio weights on a butterfly spread are arbitrarily selected and are not optimal.

- The scheme constructs risk neutral densities and optimal portfolios using options across a range of strikes. This is consistent with the mostly widely observed measure of implied volatility in the marketplace, the VIX. It is shown in Jiang and Tian (2005) that there is more information in an IV measure constructed over the full range of strikes than in the IV of an at-the-money option. The implied volatility that we test is a market representative measure of implied volatility. This avoids the need to control for moneyness that is required when testing the volatility spread by hedging single options as is done in Bakshi and Kapadia (2003).

[1] it can also be made optimal for a range of power law utility risk aversion levels.

A key aspect of our analysis is that we trade the options that are themselves used to estimate option implied volatility. Increased option prices lead to increased implied volatility. Conversely, if there is selling pressure on options to avail of the spread trade, this would reduce the option prices through increased supply and the implied volatility should decrease and reduce the spread. The evidence from options markets is that the spread persists and this mechanism is not occurring. In this article we analyze the spread trade to explain this persistence and to determine if it is explained by a risk premium, whereby an excess return is received as compensation for the non-diversifiable risk of trading the spread.

2. Materials and Methods

2.1. Option Data

Option data is sourced from the Option Metrics Ivy database. The period covered is from January 1996 to May 2013. Up until November 2005 only monthly option expiries were available for the S&P500. After this date bi-weekly option expiries were available, up until June 2012, when weekly expiries then became available. The main spikes in volume can be seen at weekly, monthly and quarterly time intervals, with monthly expiries dominating the trading volume.

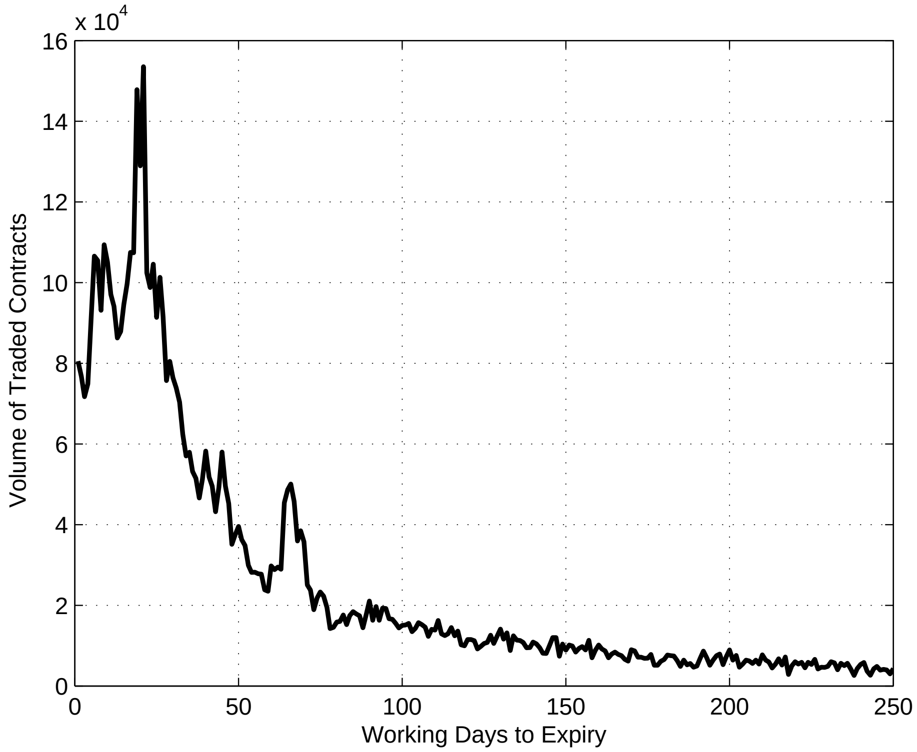


Figure 1: Average volume of S&P 500 options contracts traded for each number of working days to expiry, over the period January 1996 to May 2013. The time to expiry is calculated in working days from each trade date, t_i, to each expiry date, τ_i.

Figure 1 shows the average volume traded against the number of work days between the trade date and the expiry date of the option sets. This is calculated over all option sets in the database. Volume spikes around option sets with between 19 and 21 working days to expiry. We select our data set to include the most liquid option strips; all options within a range of between 15 and 25 working days to expiry from January 1996 to May 2013 are included.

A number of filters are applied to the data. Options with zero traded volume on a given trade day are omitted if the Open Interest for the corresponding option is also zero. Options for which no implied volatility was calculable in option metrics are removed [2]. After filtering there are 2,450 option strips remaining in the sample.

CBOE S&P500 option settlement prices are calculated using opening prices on the previous trading day to the expiry date. The time to expiry is taken as the time from closing at 3:30PM on the trade date to 8:30AM on the settlement date (the Friday before the expiry date). Dividend and interest rates used are obtained from the Ivy DB, details of the calculation of these are given in the Ivy DB reference manual. Historical S&P500 index opening and closing prices are obtained from Bloomberg.

2.2. Risk Neutral Density (RND) Estimation

We require a risk neutral density for each strip of options in our data set, where a strip of options consists of a set of closing put and call option prices with varying strikes and the same expiry date in the range 15 to 25 business days from the date of the closing price quotes. There is a large literature that describes many different techniques to extract the RND from option prices (see e.g. Aït-Sahalia and Lo (1998), Jackwerth (1999), Chernov and Ghysels (2000) and Bondarenko (2003)). We employ a three step method to fit the RND:

1. As a first step we apply the approach described in Figlewski (2010) to obtain a non-parametric RND. This method incorporates prior work by Shimko (1993) and the seminal result in Breeden and Litzenberger (1978).
2. We then curve fit a Normal Inverse Gaussian (NIG) density to the non-parametric density. Due to the flexibility of the NIG we find that the two densities are virtually identical across our data set[3]. In Ghysels and Wang (2014) and in Eriksson et al. (2009), the distribution was also shown to work extremely well in modeling RNDs and in pricing options.
3. We use the parameters of the curve fit NIG density as a starting point in an optimization to select the NIG inputs that return prices closest to the mid point of the bid/ask for the set of options used to calculate the RND.[4]

In step 3, for each option strip we obtain a set of option prices under our candidate NIG

[2] See the Option Metrics Reference Guide for details on how they calculate implied volatility.

[3] We also tested the Pearson system of distributions but found that these could not fit the non-parametric option implied RNDs as well as the NIG.

[4] We also tried the popular method of Bakshi et al. (2003) to obtain m, v, s and k analytically. The average option pricing error using an NIG based on these parameters was much larger than that obtained using our 3 stage method.

distribution, Q:

$$C_{RNi}(t,\tau) = e^{-r(\tau-t)} \int_{-\infty}^{\infty} q(x,m,v,s,k)max(0,x-k_i)dx,$$
$$P_{RNi}(t,\tau) = e^{-r(\tau-t)} \int_{-\infty}^{\infty} q(x,m,v,s,k)max(0,k_i-x)dx,$$

where, t is the date of the closing option price quotes, τ is the expiry time of the option set, k_i is the strike price of option i and x is the underlying index level. These are standard option pricing equations that calculate the option price as the expected future payoff of the option ($max(0,x-k_i)$ for a call and $max(0,k_i-x)$ for a put) under the risk neutral density, $q(x,m,v,s,k)$, discounted to time t at a rate equal to the risk-free rate less the dividend rate of the index ($r = rf - d$) [5].

The usual inputs to the NIG consist of a location parameter, μ; a tail heaviness parameter, α; an asymmetry parameter, β; and a scale parameter, δ. For the purposes of our analysis, and to isolate the effect of the second moment, it is preferable to deal with the distributional moments directly - mean, variance, skewness and kurtosis: m, v, s and k. To do this we include a mapping from distributional moments to the NIG inputs in the likelihood function (see Appendix). This allows us to deal directly with those moments in a new likelihood function $q(x,m,\sigma,s,k)$.

The optimizer varies the m,v, s and k inputs to return a set of option prices as close to the market values as possible, where closeness is defined as a moneyness-weighted pricing error, D:

$$D = \frac{\sum_i^{NC} \frac{|C_{RNi}-C_{Mi}|m_i}{C_{mi}}}{NC} + \frac{\sum_i^{NP} \frac{|P_{RNi}-P_{Mi}|m_i}{P_{mi}}}{NP},$$

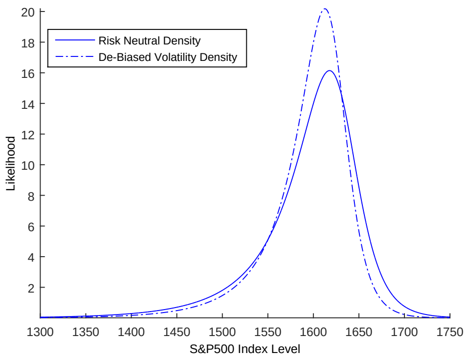
where: $m_i = \frac{k_i}{S_t}$ is the moneyness of the option struck at k_i, $C_{mi}(P_{mi})$ is the midpoint of the Bid and the Ask price for the Call (Put) struck at k_i, and NC (NP) is the number of out-of-the-money calls (puts) in the option strip. The metric, D, gives more weight to pricing accuracy for the more liquid options with a moneyness close to one as these are the options with the highest average trading volume. We filter out sets of options for which the average moneyness-weighted distance from the Bid/Ask midpoint across the option strip, D, is greater than half the Bid/Ask spread. From an initial set of 2,450 option sets this leaves 2,154 for our analysis. An example RND obtained using the technique is illustrated in Figure 2a.

2.3. Growth Optimal Options Portfolios

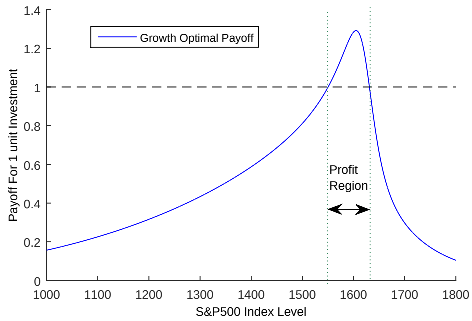
In this section we consider how to invest in options in an optimal manner when the risk neutral density implied by option prices, q as obtained in Section 2.2, differs from the objective distribution of the underlying index returns, p. The risk neutral density includes both the market's objective forecast of the distribution of the underlying and market preferences that may favor payoffs in certain market states over payoffs in other states.

The difference in information implied by the difference between the forecast, p, and the market implied distribution, q, implies a capital growth rate for an investor trading against the market when their forecast distribution is closer to the true distribution of the underlying returns. This growth rate was determined in Barron and Cover (1988), where

[5] the annualized dividend rate for the S&P 500 is obtained from Option Metrics database.



(a) Risk Neutral and De-Biased Volatility Densities.



(b) Growth Optimal Payoff Function

Figure 2: An example trade using closing S&P 500 option prices on 30 April 2013 for options with expiry on 31 May 2013. The risk neutral density is illustrated in (a) along with a density constructed using a de-biased version of the risk neutral volatility in addition to the original risk neutral mean, skewness and kurtosis. The growth optimal payoff function in (b) is the ratio of the two densities. The function is broadly similar in shape to a butterfly spread, with a central peak sloping away on either side. The payoff of the function on the settlement of the options is 1.004 representing a return of 0.4% on an investment of 1 unit. This is just one of over 2,150 option sets that we evaluate in our empirical study of the volatility spread.

it is also shown in a footnote that the growth optimal payoff is given by investing in the ratio of the two distributions: $\frac{p}{q}$. This ratio is framed as an optimal derivative in Edelman (2000). In this article we construct the payoff $\frac{p}{q}$ using static replication with a portfolio of options taken from the same set of options used to estimate the risk neutral density.

Specifically, we consider the case where the risk neutral density implied by a set of option prices has an upwardly biased volatility estimate. In this paper we create a forecast density by adjusting the second moment of the risk neutral density to remove an upward bias. We vary the λ values in Equation I to remove the upward bias and reconstruct a density with the other three of the first four moments matching the risk neutral density versions. The result is a distribution function with the same mean, skewness and kurtosis as the risk neutral density but with a de-biased second moment. We correct the risk neutral density for the volatility bias and test the unbiased volatility estimate in the presence of the other moments of the distribution. The impact of the volatility spread can then be tested in the presence of the risk neutral skewness and kurtosis as they are included by construction.

2.3.1. A Probabilistic Scoring Rule: Quantifying Capital Growth Opportunities

We now have two pricing densities for each option set, q_i and p_i, the difference between the two being an adjustment to remove the volatility spread. To investigate the impact of the premium we would now like to compare the two densities on expiry of the 2,154 option sets in our sample. We adopt the following probabilistic scoring rule:

$$W = \prod_i \frac{p_i(x(\tau_i))}{q_i(x(\tau_i))}, \tag{II}$$

where $q_i(x(\tau_i))$ is the likelihood score of the risk neutral density evaluated at the index level at expiry time, τ_i, of the option set i, and $p_i(x(\tau_i))$ is the likelihood score of the de-biased volatility version of the risk neutral density. W, therefore, corresponds to a compounded version of the ideal payoff from Barron and Cover (1988), where a value on the expected capital growth rate, G, achievable with this ratio was given as:

$$G \leq D(p(x)||q(x)),$$

where the Kullback-Leibler distance, D, is defined as:

$$D(p(x)||q(x)) = \int_{r=-\infty}^{r=\infty} p(x) \cdot \log\left(\frac{p(x)}{q(x)}\right) dx.$$

D is the expected log of the likelihood ratio of the true distribution to the market implied distribution, where the expectation is taken under the true distribution. It corresponds to the maximum capital growth rate achievable by trading the true distribution against the market implied distribution.

In this article we are interested in the financial value of the information that option-implied volatility has an upward bias. We test whether this information can be exploited for economic gain. W in Equation II, corresponds to the final compounded wealth after 2,154 trades when an investor starting with a unit investment invests in the growth-optimal payoff to exploit their information that the implied volatility has an upward bias.

Maximizing the rule in Equation II is equivalent to selecting the NIG density parameters through Maximum Likelihood Estimation (MLE), as the divisor is set by the market prices and is a constant from the perspective of a trader. The fact that MLE is consistent with maximizing log utility is a known result and is discussed in Johnstone (2011).

A similar rule is commonly applied in econometrics in likelihood ratio tests; a technique used to select between nested models. We can consider a hypothesis test where the volatility estimate is unadjusted versus where it has two components: the original market parameter σ_{iv} and a second correction parameter that allows a scaling of the market-implied volatility:

$$H_0 : v = \sigma_{iv}$$
$$H_1 : v_c = \sigma_{iv} - (1 - \lambda)\sigma_{iv}$$

The models are nested and we can apply a likelihood ratio test to determine whether the Null hypothesis of the adjusted model being superior in performance can be rejected. The market model is a restricted version ($\lambda = 1$) of our adjusted model and we can apply Wilk's theorem, see Wilks (1938). This gives the distribution of the log of the likelihood ratio as being χ^2 with degree of freedom equal to the number of restrictions in the model (one). This allows the calculation of p-values for the out-performance of our modified IV models against the market RND (or vice versa) in addition to the calculation of an implied capital growth rate from Equation II.

2.3.2. Incorporating Risk Aversion

Optimization based on the scoring rule outlined in the previous section is consistent with maximizing log utility. In this section we extend the rule to incorporate a range of Constant Relative Risk Aversion (CRRA) levels in a power law utility function:

$$U(w) = \frac{1}{1-\gamma} w^{1-\gamma} \tag{III}$$

$U(w)$ is the utility an agent gets from having the wealth level, w. The level of risk aversion, γ, determines an agent's risk vs reward preferences. When $\gamma = 1$ the agent utility becomes a log function and the agent is only concerned with maximizing their expected capital growth. An approach to introduce risk aversion to our scheme is given in Guasoni and Robertson (2012), we can replace the original forecast, p, with a new forecast, p_{RA}, given by:

$$p_\gamma(x) = \frac{p(x)^\gamma q(x)^{1-\gamma}}{\sum p(x)^\gamma q(x)^{1-\gamma}}, \tag{IV}$$

where $p_{RA}(x)$ is a new risk-adjusted density forecast that the agent trades against the market implied distribution, $q(x)$, incorporating their risk preferences as captured by the constant of relative risk aversion,, γ. The resulting payoff $\frac{p_{RA}(x)}{q(x)}$ is the optimal payoff for an investor with $CRRA = \gamma$. In order to test our findings across risk aversion levels we modify our forecast densities, as per Equation IV, for each of five levels of risk aversion: $\gamma \in \{1, 2, 3, 4, 5\}$. This results in a set of returns corresponding to the optimal investment in the volatility spread trade for investors with each level of risk aversion. In each case we analyze the effect on each investors utility of removing the upward bias on implied volatility by varying the volatility scaling factor, λ. This results in a payoff function that is specific to the level of volatility scaling and investors' relative risk aversion:

$$\Psi_\gamma(\lambda) = \frac{\frac{p(\lambda)^\gamma q^{1-\gamma}}{\sum p(\lambda)^\gamma q^{1-\gamma}}}{q} \tag{V}$$

The payoff function $\Psi_1(0.8)$ is illustrated in Figure 2b for an example option set. This payoff is constructed as per Equation V, the payoff has a cost of \$1. On expiry of the option set the trader receives the value of the payoff evaluated at the level of the S&P 500 index on the settlement date of the option set. A volatility arbitrageur gains if the index does not move far from the index level at which the trade is entered.

2.4. Static Replication

The optimal payoff of the ratio of our de-biased RND to the original RND is not a traded security. In order to invest in it we statically replicate the payoff using a portfolio of the original options used to extract the RND. Our choice of the NIG distribution has a nice property in this regard as the ratio of two NIG functions is twice-continuously differentiable. It is shown in Bakshi and Madan (2000) that any twice-continuously differentiable function can be spanned by a set of vanilla put and call options; the underlying and a zero coupon bond. We find that we can replicate the payoffs very closely using the underlying options alone. An example of an RND, the de-biased RND, the optimal payoff and the statically replicated version of the payoff are given in Figures 2a, 2b and 4. The corresponding portfolio constituents are given in Table 1. The statically replicated version of the payoff can be seen to very closely match the target payoff.

3. Empirical Results

3.1. Unbiased Volatility Estimates

While we are primarily interested in testing the removal of the bias in Equation I, as it is most commonly used measure in the literature, we also consider a number of competing volatility forecasts to substitute for IV and to test in a density forecast against the market RND. In each case we replace the implied volatility, $\sigma_{rn}(t, \tau)$, with candidates for the

Table 1: A portfolio of call and put options that statically replicates the optimal volatility arbitrage payoff, as given in Figure 4. For ease of presentation, the risk neutral cost of the target payoff is scaled up from \$1 to \$1,000. The net cost of the replication portfolio is \$1,013.2. The theoretical price of \$1,000 comes from a risk neutral distribution constructed around the midpoint of the bidask spread, the real portfolio is net long options and incurs some transaction costs due to the spread.

	Puts			Calls		
Strike	Weight	Bid Price	Ask Price	Weight	Bid Price	Ask Price
1380	-	-	-	1.111	11.30	11.70
1390	-	-	-	0.309	13.60	14.00
1400	0.199	0.650	0.800	0.007	16.40	16.90
1410	-0.287	0.700	0.900	-	-	-
1420	0.014	20.00	20.30	-	-	-
1425	-	-	-	0.934	24.00	24.60
1430	-	-	-	0.583	29.00	29.60
1440	-	-	-	0.142	38.10	39.10
1450	-	-	-	1.799	41.70	42.70
1460	-	-	-	0.168	49.40	50.90
1470	-	-	-	0.053	58.10	59.50
1475	-	-	-	0	86.60	88.00
1480	-	-	-	0.023	193.60	194.80
1490	-0.841	0.950	1.15	1.353	183.70	185.00
1500	-	-	-	8.841	105.80	107.30
1510	-	-	-	-0.033	52.50	53.60
1520	-	-	-	-0.089	44.40	45.30
1525	-	-	-	-0.057	36.80	37.60
1530	-	-	-	-0.033	33.20	33.80
1540	-	-	-	-0.366	29.70	30.30
1550	-0.001	1.10	1.30	-0.002	23.20	23.80
1560	-0.204	1.20	1.40	-8.705	17.50	17.90
1570	-0.056	1.25	1.50	-6.886	12.50	13.00
1575	0.447	1.50	1.70	-	-	-
1580	0.005	1.75	1.95	-8.064	8.50	8.90
1590	-0.048	2.00	2.25	-3.976	6.90	7.30
1600	0.094	2.35	2.60	0.090	5.40	5.90
1610	-0.188	2.50	2.80	0.594	3.30	3.70
1620	-0.400	2.75	3.00	-	-	-
1625	-6.81	3.20	3.50	5.195	2.00	2.30
1630	-0.547	3.80	4.10	3.860	1.20	1.45
1640	0.140	4.50	4.90	1.405	0.75	0.95
1650	0.100	5.40	5.80	0.239	0.60	0.80
1660	0.267	5.90	6.30	-	-	-
1670	0.184	6.40	6.90	-	-	-
1675	0.657	7.70	8.20	-	-	-
1680	0.504	9.30	9.80	1.924	0.50	0.70

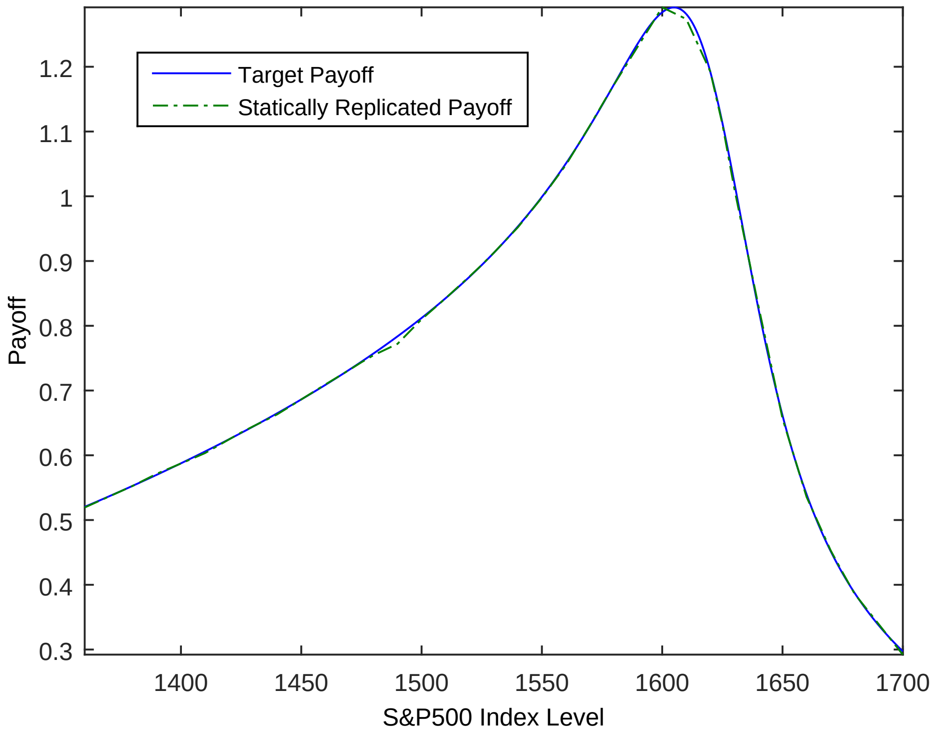


Figure 3: The result of static replication of the growth optimal payoff with a portfolio of options. The portfolio is obtained using an optimizer that attempts to replicate the payoff as cheaply as possible trading the full range of put and call options used to construct the original risk neutral density. The cost of the portfolio is 1.013, reflecting the impact of the real bid ask spread on the theoretical price of 1.0 obtained under the risk neutral density based on the mid-point of the option quotes. [6] The portfolio corresponding to the replicated payoff is given in Table 1.

expected future volatility under the physical measure over the lifetime of the option set, $\sigma_{FC}(t, \tau)$.

The IV can be represented as the expected volatility under the physical measure plus a volatility spread:

$$\sigma_{iv} = E^P[\sigma(t, \tau)] + VS \tag{VI}$$

This can be rearranged to represent the expected volatility under the physical measure as the IV corrected for the volatility spread:

$$E^P[\sigma(t, \tau)] = \sigma_{iv} - VS \tag{VII}$$

We have framed the analysis in these terms as we can then attribute an expected capital growth rate directly to the volatility spread. It is the implied capital growth rate when replacing σ_{iv} with $E^P[\sigma(t, \tau)]$ in a density that is traded against the original RND.

Table 2: Table indicating the average capital growth rate achieved per density forecast, by modifying the Risk Neutral Density (RND) over 2,154 forecasts. The volatility, skewness and kurtosis of the resulting log returns are also reported. The first 6 schemes replace the Implied Volatility (IV) with alternative volatility forecasts. $\sigma_{0.96IV}$ is the optimal reduced IV forecast obtained by scaling IV by a factor of 0.96; σ_{G11} is the GARCH(1,1) forecast obtained by fitting a GARCH(1,1) model to the previous 1,000 daily returns and forecasting ahead T days, where T is the time to option expiry. σ_T is historical volatility over the previous T days. σ_{22}; σ_{65} and σ_{252} are historical volatility estimates over the indicated number of working days. The Zero Skew scheme replaces the option implied skew with zero, while holding all other moments at the RND levels. The Zero Excess Kurtosis scheme replaces the option-implied kurtosis with 3.

Scheme	$\mu(\%)$	$\sigma(\%)$	Skewness	Kurtosis
$\sigma_{0.96IV}$	0.212	5.933	-8.367	173.629
σ_{G11}	-3.240	33.630	-4.550	50.632
σ_T	-13.005	68.136	-4.794	50.133
σ_{22}	-12.653	66.701	-6.245	93.166
σ_{65}	-10.565	79.063	-20.396	681.077
σ_{252}	-13.171	84.967	-18.891	600.411
Zero Skew	-12.238	39.503	0.119	2.697
Zero Kurtosis	0.235	28.249	-6.534	81.121

3.1.1. Historical Volatility

In this section we test the historical volatility estimated over some commonly used time periods: 22 trading days (monthly); 65 trading days (quarterly); 252 trading days (yearly) and over $(\tau - t)$ trading days, where $(\tau - t)$ is the time to expiry of the option set. We use the Quadratic Variation to measure realized daily variance over the previous T_h days and scale to match the time to expiry of the option set $(\tau - t)$:

$$\sigma_h(t, \tau) = \sqrt{\sum_{i=1}^{i=T_h} r_i^2 . \frac{(\tau - t)}{T_h}}$$

where:

$r_i$is the daily log return of the index.

The results can be seen in Table 2. The best performing historical volatility measure is quarterly (σ_{65}), but even this measure loses 10.6% on average per option set. The 252 day historical estimate performs very badly, losing 13.2% on average per option set. These results show that simple historical volatility estimates cannot compete with option-implied volatility when used in density forecasts.

3.1.2. GARCH(1,1)

In this section a GARCH(1,1) forecast is used in place of the option-implied volatility. For each option set, the GARCH model is fit using the previous 1,000 daily returns prior to the trade date and the resulting model is used to forecast the following T days, where T is the time to expiry of the option set. The average cost of trading GARCH(1,1) against the IV is 3.24% per forecast across the sample (see Table 2 and Figure 4).

3.1.3. Reduced Implied Volatility

In this section we adopt a simple model for the implied volatility as an upwardly-biased version of expected volatility under the physical measure. In order to obtain a forecast of future realized volatility over T, the time to expiry of an option set, we scale the implied volatility by a shrinkage factor, λ.

$$\sigma_r = \lambda \sigma_{iv} + \epsilon_t \tag{VIII}$$

An ex-post linear regression, as per Equation VIII, over the whole sample results in a λ value of 0.8. This value has the best score in the least squares sense as a point estimate of volatility, but when considered in a complete density as a likelihood score it significantly under-performs the original density using the unadjusted market-implied volatility.

A likelihood ratio test, as described in Section 2.3.1, of the two densities over the 2,154 option expiries, selects the market-implied density as the superior model with a p value of 1e-14. The likelihood score for the RND including the upward bias on implied volatility beats an RND with the bias removed with strong statistical significance.

We test the performance of a range of volatility forecasts $\sigma_{FC} = \lambda \sigma_{iv}$ with λ values in the range [0.8, 1.2]. The results are given in Table 3. The optimal value over the whole sample is $\lambda = 0.96$. This results in an average expected growth rate of 21 bps per option set. We select the reduced IV ($\sigma_{FC} = 0.96\sigma_{RN}$) as the best performing volatility forecast measure for our remaining analysis.

3.1.4. Skewness and Kurtosis

In this section we use our framework to perform an analysis on the contribution of the option-implied skewness and kurtosis to the RND. The zero skew scheme holds the mean, variance and kurtosis at their original risk neutral values and sets skewness to zero. The high cost of ignoring skew is evidenced with an average return of -12.2% per forecast. The zero excess kurtosis scheme holds the mean, variance and skewness at their original risk neutral values and sets excess kurtosis to zero (kurtosis to 3). This results in a small positive return of 0.23% per forecast. The results of the two schemes can be seen in Table 2 and in Figure 4. The results demonstrate the relative importance of skewness in financial markets.

3.2. Return Analysis

In this section we use the reduced IV from Section 3.1.3 as our best de-biased volatility estimate.

3.2.1. Controlling for Risk Aversion

We estimate the returns to trading a reduced IV scheme across a range of volatility scaling parameters, λ, and for a range of levels of risk aversion $\gamma \in \{1, 2, 3, 4, 5\}$, where risk aversion is incorporated as described in Section 2.3.2.

The results are given in Table 3. The Certainty Equivalent, CEQ, for each λ and γ is calculated in a standard fashion as the w value in Equation III that results in the same utility, $U(w)$, as the average utility over the 2,154 returns for each scheme. The best certainty equivalent is obtained with the same λ value for each level of risk aversion, $\lambda = 0.96$. The maximum expected capital growth for the scheme is obtained with this setting and $\gamma = 1$. This is the growth-optimal risk aversion level corresponding to log utility. Even in the growth-optimal case the average capital growth rate per forecast over the sample is just 21 bps, an amount unlikely to overcome realistic transaction costs.

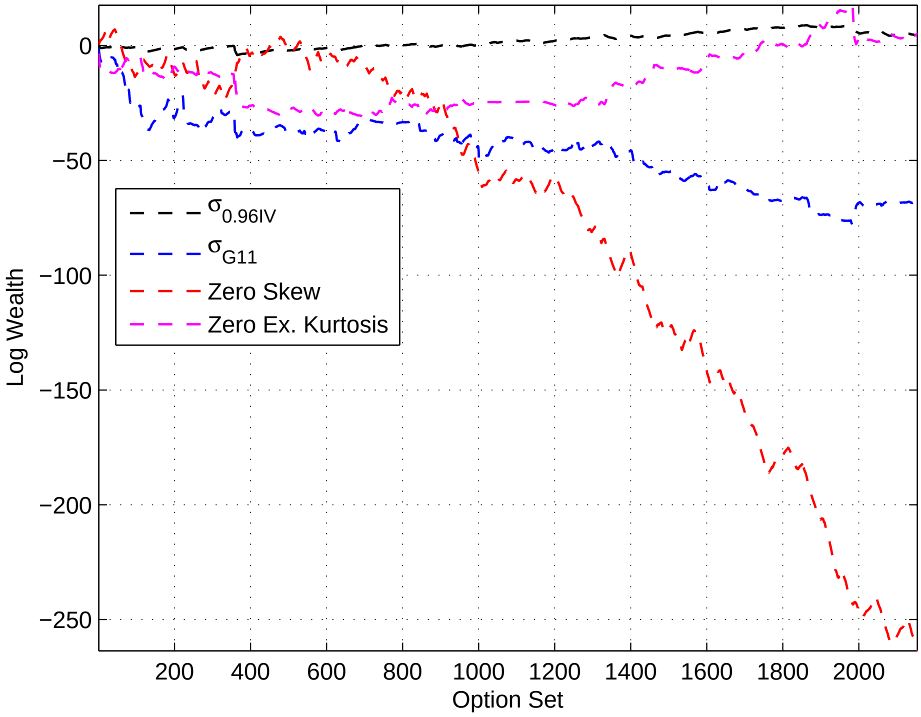


Figure 4: The resulting log of wealth over 2,154 option sets where a modified RND is traded against the original RND. The $\sigma_{0.96IV}$ scheme adjusts the RND, replacing the implied volatility σ_{iv} with a scaled value of $0.96\sigma_{iv}$. The σ_{G11} scheme replaces the IV with the output forecast of a GARCH(1,1) model that is parametrized using the previous 1,000 daily returns of the index and used to forecast T days ahead, where T is the number of days to expiry of the option set. The Zero Skew scheme replaces the option-implied skewness value with 0. The Zero Excess Kurtosis scheme replaces the option-implied kurtosis with a value of 3.

The optimal scaling factor, $\lambda = 0.96$, is consistent across all levels of risk aversion tested. In our implementation risk aversion is added by moving the resulting forecast density closer to the market forecast density in the transformation given in Equation IV. In the presence of this transformation the optimal parameter choice is the log optimal one. This empirical finding supports the use of maximum likelihood in financial modeling in spite of its equivalence to maximizing log utility, as the optimal parameter for other risk aversion levels match the maximum likelihood parameter. It suggests that a two stage approach of first optimizing parameters through maximizing likelihood followed by incorporating risk aversion in a second investment stage is optimal for investors with power law utility functions.

3.2.2. Asymmetrical Losses

In this section we examine whether the returns to a volatility arbitrage scheme are symmetrical in the volatility spread, $VS = IV - RV$. Are the losses when realized volatility, RV, spikes higher than implied volatility, IV, offset by the gains when IV overestimates RV by the same amount?

Table 3: Table showing the mean return (μ) and Certainty Equivalent (CEQ) for an investor with capital of 100 when trading a scaled factor of implied volatility in a density against the market RND, in a sample of 2,154 option sets from January 1996 to May 2013. The optimal scaling factor, λ, on the option-implied volatility is a factor of 0.96. The table shows that the result is consistent across risk aversion levels.

λ	$\gamma = 1$		$\gamma = 2$		$\gamma = 3$		$\gamma = 4$		$\gamma = 5$	
	μ(%)	CEQ	μ(%)	CEQ	μ(%)	CEQ	μ(%)	CEQ	μ(%)	CEQ
0.80	-3.70	96.37	-0.60	55.68	-0.09	31.50	0.06	31.61	0.11	35.28
0.82	-2.66	97.38	-0.34	81.44	0.02	61.26	0.11	56.50	0.14	57.26
0.84	-1.80	98.21	-0.13	92.92	0.10	85.36	0.15	81.64	0.16	80.64
0.86	-1.12	98.89	0.02	97.20	0.15	95.33	0.17	94.35	0.16	94.00
0.88	-0.60	99.41	0.13	98.87	0.18	98.47	0.17	98.33	0.16	98.32
0.90	-0.21	99.79	0.19	99.59	0.19	99.51	0.17	99.50	0.15	99.52
0.92	0.04	100.04	0.21	99.92	0.18	99.89	0.15	99.88	0.13	99.89
0.94	0.18	100.18	0.20	100.07	0.15	100.03	0.12	100.01	0.10	100.01
0.95	0.21	100.21	0.18	100.10	0.13	100.06	0.11	100.04	0.09	100.03
0.96*	0.21	100.21*	0.16	100.11*	0.11	100.07*	0.09	100.05*	0.07	100.04*
0.97	0.19	100.19	0.13	100.10	0.09	100.07	0.07	100.05	0.06	100.04
0.98	0.15	100.15	0.09	100.08	0.06	100.05	0.05	100.04	0.04	100.03
0.99	0.09	100.09	0.05	100.04	0.03	100.03	0.02	100.02	0.02	100.02
1.00	0	100.00	0	100.00	0	100.00	0	100.00	0	100.00
1.01	-0.10	99.90	-0.05	99.95	-0.034	99.96	-0.03	99.97	-0.02	99.98
1.02	-0.23	99.77	-0.11	99.88	-0.070	99.92	-0.05	99.94	-0.04	99.95
1.03	-0.37	99.63	-0.17	99.81	-0.11	99.88	-0.08	99.91	-0.06	99.93
1.04	-0.52	99.48	-0.23	99.74	-0.15	99.83	-0.11	99.87	-0.08	99.90
1.05	-0.70	99.31	-0.30	99.66	-0.19	99.77	-0.14	99.83	-0.11	99.87
1.06	-0.89	99.12	-0.37	99.57	-0.23	99.72	-0.17	99.79	-0.13	99.83
1.07	-1.09	98.92	-0.45	99.48	-0.27	99.66	-0.20	99.75	-0.15	99.80
1.08	-1.30	98.71	-0.52	99.38	-0.32	99.60	-0.23	99.70	-0.18	99.76
1.09	-1.53	98.48	-0.60	99.28	-0.36	99.53	-0.26	99.65	-0.20	99.72
1.10	-1.77	98.24	-0.68	99.17	-0.41	99.46	-0.29	99.60	-0.22	99.69
1.11	-2.03	97.99	-0.77	99.07	-0.46	99.40	-0.32	99.55	-0.25	99.65
1.12	-2.29	97.74	-0.85	98.95	-0.51	99.33	-0.36	99.50	-0.27	99.61

Table 4: Demonstration of the asymmetry in correlation between the returns from trading the volatility spread and the volatility spread itself, VS, across a range of scaling factors on the Implied Volatility, λ. The ρ^+ and ρ^- values are defined as: $\rho^+ = corr(r_t|_{VS>0}, VS|_{VS>0})$, $\rho^- = corr(r_t|_{VS<0}, VS|_{VS<0})$. The P Value of the difference is calculated using the using the Fisher transform (see Fisher (1915)). The value of $\lambda = 0.81$ corresponds to the least squares estimate, the value of $\lambda = 0.96$ corresponds to the maximum likelihood estimate.

λ	ρ^+	ρ^-	$\lvert\rho^+ - \rho^-\rvert$	P Value
0.810	-0.017	0.294	0.311^{***}	(< 0.000)
0.850	-0.015	0.324	0.339^{***}	(< 0.000)
0.900	-0.004	0.401	0.405^{***}	(< 0.000)
0.960	-0.005	0.369	0.374^{***}	(< 0.000)

To test this we calculate the correlation of scheme returns to the volatility spread when the spread is positive:

$$\rho^{+} = corr(r_t|_{VS>0}, VS|_{VS>0}) \tag{IX}$$

and when the volatility spread is negative:

$$\rho^{-} = corr(r_t|_{VS<0}, VS|_{VS<0}) \tag{X}$$

Results are given in Table 4. We test across a range of λ values and across a range of risk aversion levels, γ. In all cases there is a large difference in correlations when the spread is negative versus when it is positive. The uPside of the volatility arbitrage is small relative to the large losses that can occur when the spread is negative. For each ρ^{+}/ρ^{-} pair in Table 4 we use the Fisher transform (see Fisher (1915)) to obtain confidence intervals on the correlation estimates and to obtain a p value on their difference. In all cases p values of zero are obtained indicating statistical significance in the difference between the two values.

The source of the asymmetry is illustrated in Figures 5 & 6. Figure 5 shows that the returns to the statistical arbitrage scheme are not linear in volatility spread. When the market index does not move a reduced volatility forecast beats the market implied forecast and vice versa, however the shape of the likelihood ratio plane is curved. The statistical performance (and the implied economic performance) are not linear in the spread between the implied volatility and the forecast volatility. Figure 6 illustrates the area in the realized return space that the index needs to hit for the statistical arbitrage scheme to be profitable. It is a narrow flat region with quite low returns and the steepness of the drop off in return increases with decreasing λ (labeled as Implied-Volatility Scaling Factor in the plot).

3.2.3. Variation in Return Dependence on the Volatility Spread and Jump Risk

In this section we test the dependence of the returns of the trading scheme on the spread itself and on jump risk fears as proxied by the risk neutral skewness and kurtosis (this is a commonly used substitution, see e.g. Bakshi and Kapadia (2003)). We run the following regression, to test the change in the exposures as captured by the β values with variation of the volatility scaling parameter λ:

$$r_t = \alpha + \beta_1(IV - RV) + \beta_2 SKEW_{RND} + \beta_3 KUR_{RND} + \epsilon_t \tag{XI}$$

The results of the regression are given in Table 5, and presented graphically in Figure 7. The exposure of the trade returns to the volatility spread are strongly significant and positive across all values of λ. This is as expected as the scheme is constructed to trade the spread.

The variation in exposure to the higher moments of the RND is also clear in the results. The exposure to the risk neutral kurtosis can be seen to vary in λ, decreasing as λ approaches 1.0, but is only statistically significant for small λ values close to the least squares λ estimate of 0.8. The exposure to the risk neutral skewness can also be seen to vary in λ, and is statistically significant for λ values below 0.92.

Overall, decreasing λ from 1.0 increases the exposure to the volatility spread (IV-RV), the option-implied skewness and the option-implied kurtosis. Alternatively, adding an upward bias to IV decreases the exposure to each compared to the exposure of a trading scheme based on the unbiased least square estimate of implied volatility (0.8*IV). These results confirm empirically the theoretical results in Bakshi and Madan (2006), where the variance premium is associated with jump fears proxied by the higher moments of the RND.

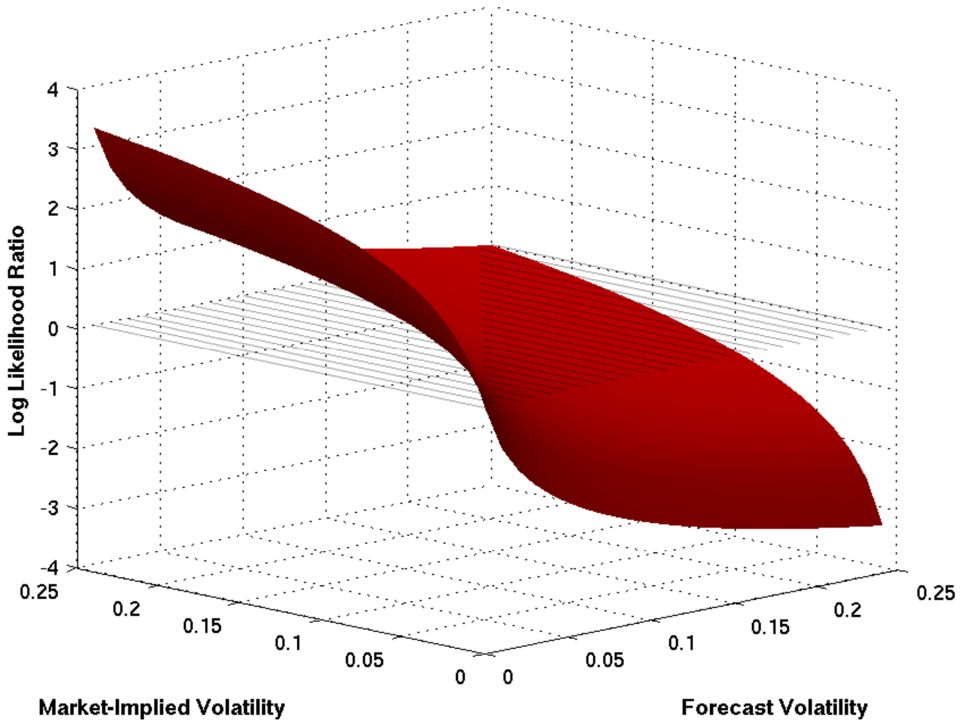


Figure 5: Plot of the Log Likelihood Ratio consisting of the Normal Inverse Gaussian (NIG) likelihood for the forecast volatility over the NIG likelihood for the implied volatility. An index return of zero is used to evaluate the ratio over different implied and forecast volatility pairs. This is the ideal return for our volatility arbitrage scheme. When the forecast is lower than implied the ratio is positive as the likelihood of a zero return is higher under a lower volatility forecast, and vice versa. The ratio is not linear in volatility as can be seen by the the steepness of the changes in the log ratio at the extremities of the plot. Mean, skewness and kurtosis values for both the forecast and the market implied likelihoods are set to the same median values, those implied for options in the data set with 30 days to expiry.

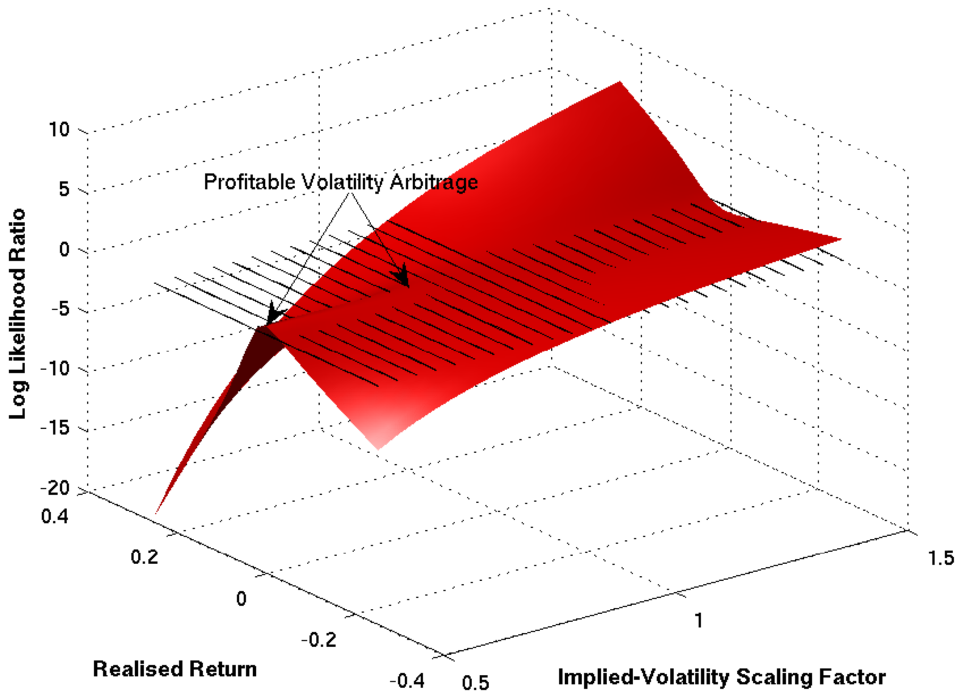


Figure 6: Plot of the Log Likelihood Ratio consisting of the Normal Inverse Gaussian (NIG) likelihood for the forecast volatility over the NIG likelihood for the implied volatility, where the forecast volatility is a scaled version of the market-implied volatility. Market-implied volatility is held constant at the median value for options with 30 days to expiry and the index return is allowed to vary in the range [-40% +40%]. The profitable region for a forecast based on reduced implied volatility is indicated in the graphic. It can be seen that, as the scaling factor is reducing toward 0.5, the slope of the drop off of the log likelihood ratio outside of the profitable regions increases. This captures the increasing costs of being wrong when taking a more extreme view on the volatility relative to the market implied volatility.

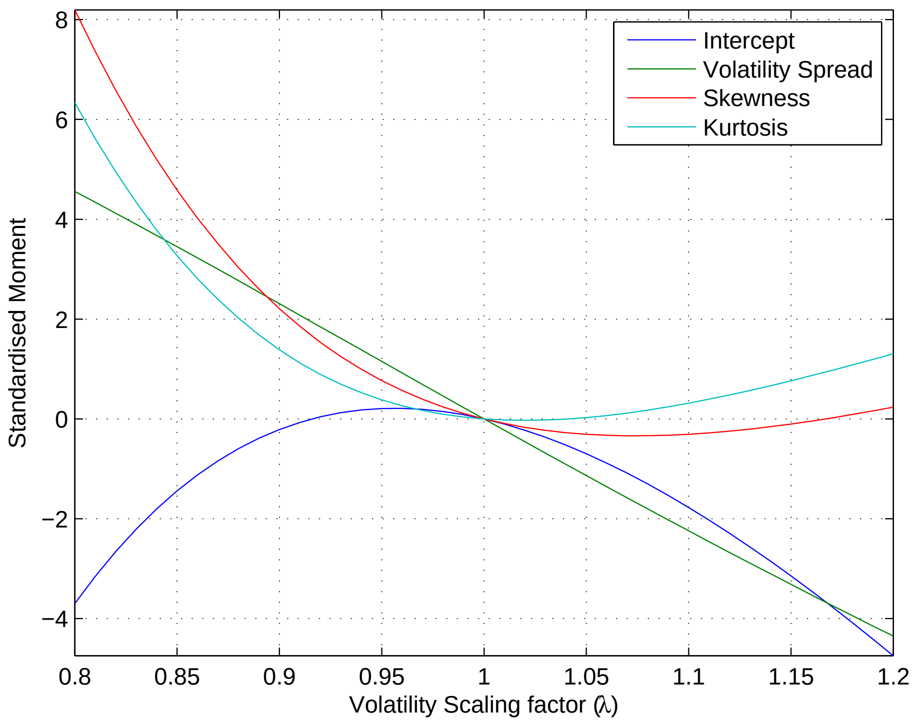


Figure 7: Beta coefficients in a linear regression of the scheme returns:
$r_t = \alpha + \beta_1(IV - RV) + \beta_2 SKEW_{RND} + \beta_3 KUR_{RND} + \epsilon_t$
As the scaling factor, λ, on implied volatility approaches one the exposure to option-implied skewness can be seen to reduce. The upward bias in option implied volatility is reducing the exposure to losses when the Risk Neutral Density includes jump fears as surrogated by the option-implied skewness and kurtosis.

Table 5: Beta coefficients in a linear regression of the scheme returns:
$r_t = \alpha + \beta_1(IV - RV) + \beta_2 SKEW_{RND} + \beta_3 KUR_{RND} + \epsilon_t$
As the scaling factor, λ, on implied volatility approaches one the exposure to option-implied skewness can be seen to reduce. The upward bias in option implied volatility is reducing the exposure to losses when the Risk Neutral Density includes jump fears as surrogated by the option-implied skewness and kurtosis.

λ	α	$IV - RV$	$SKEW_{RND}$	KUR_{RND}
0.800	-3.698^{***}	4.557^{***}	8.192^{**}	6.335^{*}
	(0.000)	(0.000)	(0.015)	(0.059)
0.820	-2.656^{***}	4.121^{***}	6.583^{**}	4.949^{*}
	(0.001)	(0.000)	(0.021)	(0.082)
0.840	-1.803^{***}	3.677^{***}	5.200^{**}	3.781
	(0.010)	(0.000)	(0.029)	(0.112)
0.860	-1.121^{*}	3.226^{***}	4.022^{**}	2.810
	(0.052)	(0.000)	(0.041)	(0.153)
0.880	-0.596	2.769^{***}	3.032^{*}	2.018
	(0.202)	(0.000)	(0.057)	(0.205)
0.900	-0.213	2.309^{***}	2.210^{*}	1.385
	(0.563)	(0.000)	(0.079)	(0.270)
0.920	0.042	1.847^{***}	1.537	0.894
	(0.881)	(0.000)	(0.107)	(0.348)
0.940	0.180	1.384^{***}	0.995	0.526
	(0.366)	(0.000)	(0.143)	(0.439)
0.960	0.212^{*}	0.921^{***}	0.569	0.264
	(0.093)	(0.000)	(0.188)	(0.540)
0.980	0.149^{**}	0.460^{***}	0.242	0.094
	(0.013)	(0.000)	(0.241)	(0.650)
1.020	-0.227^{***}	-0.457^{***}	-0.169	-0.029
	(0.000)	(0.000)	(0.373)	(0.880)
1.040	-0.524^{***}	-0.910^{***}	-0.275	-0.003
	(0.000)	(0.000)	(0.450)	(0.994)
1.060	-0.885^{***}	-1.358^{***}	-0.328	0.068
	(0.000)	(0.000)	(0.532)	(0.897)
1.080	-1.304^{***}	-1.803^{***}	-0.337	0.177
	(0.000)	(0.000)	(0.619)	(0.794)
1.100	-1.774^{***}	-2.242^{***}	-0.307	0.316
	(0.000)	(0.000)	(0.707)	(0.699)
1.140	-2.851^{***}	-3.103^{***}	-0.156	0.666
	(0.000)	(0.000)	(0.885)	(0.534)

4. Controlled Experiment

In this section we use the the two factor stochastic volatility and jump model of Bates (see Bates (1991)) in a controlled experiment to replicate the approach in the empirical section of the paper in a defined model framework. The model includes a jump-diffusive data generating process for the underlying returns with a clearly specified pricing kernel and associated prices for diffusive market risk, diffusive stochastic-volatility risk, and asset-value jump risk. The model is given by:

$$
\begin{aligned}
&dF/F = (\mu - \lambda_t \bar{k} + c_{v1} V_{1t} + c_{v2} V_{2t})dt + \sqrt{V_{1t}} dZ_1 + \sqrt{V_{2t}} dZ_2 + k dq, \\
&dV_{it} = (\alpha_i - \beta_i V_{it})\, dt + \sigma_{vi} \sqrt{V_{it}} dZ_{vi}, i = 1, 2, \\
&Cov(dZ_i, dZ_{vi}) = \rho_i dt, i = 1, 2, \\
&Cov(dZ_1, dZ_2) = Cov(dZ_{v1}, dZ_{v2}) = 0,
\end{aligned}
$$

where, Z_i and Z_{vi}, i=1, 2, are Wiener processes with the correlation structure specified above, $\lambda_t = \lambda_0 + \lambda_1 V_{1t} + \lambda_2 V_{2t}$ is the instantaneous conditional jump frequency, k is the random percentage jump conditional on a jump occurring, with time-invariant lognormal distribution $ln(1+k) \sim N[ln(1+k) - \frac{1}{2}\delta^2, \delta^2]$ and q is a Poisson counter with instantaneous intensity $\lambda_t : Prob(dq = 1) = \lambda_t dt$.

Under the model skewness can be generated through jumps and also through the correlation of volatility and returns. In Bates (1991), Bates parameterizes the model to S&P 500 option prices over the period 1988 -2003. The resulting parameter values give rise to an interpretation of the two volatility factors V_{1t} and V_{2t}:

- V_{1t} is a 'volatility-and-skewness' factor that heavily affects implicit skewness and leptokurtosis at all maturities through two channels: its almost total determination of the instantaneous risk-neutral jump frequency λ_t , and an assessed strong 'volatility feedback' channel that predicts a strong tendency for jump and non-jump risk to rise whenever the market falls.
- V_{2t} by contrast primarily affects implicit variances, with relatively little impact upon higher cumulants.

Under the model the volatility risk premia are defined as:

$$\phi_{vi} \equiv -Cov(dV_{it}, dJ_w/J_w)/dt = \xi_i V_{it} \equiv (\beta_i^* - \beta_i) V_{it} \qquad \text{(XII)}$$

where J_w is the marginal utility of nominal wealth for the representative investor, and the ξ_i's are free 'risk premium' parameters estimated by the divergence between the β_i^* parameters inferred from option prices and the β_i parameters from time series analysis.

We analyze the payoff $\Psi_\gamma(\lambda)$ under the Bates model to provide more clarity on our empirical results under a defined model framework. We perform the volatility scaling for different λ values by varying the β_i parameters in the model. This links the analysis directly to an adjustment of the variance premia under the definition of variance risk premia in the model (as per Equation XII).

4.1. Experimental Framework

We parameterize the Bates 2 factor model to the risk neutral density that resulted in one of the worst payoffs for a volatility arbitrageur in our empirical analysis. The RND corresponds to an option set trading on the 12th of September 2008 with expiry on October 18th 2008. The period covers the fall out around the Lehman bankruptcy and the return on the S&P500 index over that timeframe was a fall of over 28%, placing the return very much in the left tail of the risk neutral distribution.

As a first step we parameterize the Bates model to the actual RND obtained in Section 2.2. A Matlab toolbox was generously provided (see Fusari (2015)) to generate arbitrage-free option prices corresponding to a set of parameters for the Bates model. A genetic algorithm optimization program is used to select the 15 input parameters for the Bates model to fit the risk neutral density implied by the corresponding set of option prices to the target density [7]. The RND for each candidate set of option prices is obtained in the same way as described in Section 2.2.

The starting point for the optimization is the set of constrained parameters (SVJDC) in Table 2 of Bates (1991). We show that, by varying the volatility risk premia in Equation XII alone, it is possible to get the reduced volatility version of the RND and a corresponding payoff, $\Psi_\gamma(\lambda)$, both closely matching their equivalents, obtained in the empirical section. The NIG parameters and the matching Bates parameters are reported in Table 6.

We apply the same procedure to obtain two alternative forecast densities with reduced variance premia. The starting point for these optimizations is the 15 parameters selected by the GA for the unmodified RND. Two new optimizations are run but this time only the two β_i parameters are optimized and the remaining parameters are held at the original RND values. The first optimization varies the β_i to implement the least squares estimate of the scaling factor, $\lambda = 0.8$. The second uses the optimal value used in the empirical section, $\lambda = 0.96$.

The two payoffs, $\Psi_1(0.8)$ and $\Psi_1(0.96)$, are then evaluated at expiry of the option set at the realized index return level of -28%. The density functions are illustrated in Figure 8 and the corresponding parameters are listed in Table 6. The losses incurred can be seen to be significantly increased by completely removing the variance premium ($\Psi_1(0.8) = 0.36$ Vs. $\Psi_1(0.96) = 0.64$).

4.2. Discussion of results

Our results are consistent with a volatility spread being added to density forecasts to mitigate the cost of tail events. A possible explanation for this is the scenario whereby a trader is uncertain of the true parameters of their density forecasting model (such as the Bates model above). There is a large array of candidate option pricing models (see e.g. Fabozzi et al. (2016), Fusai et al. (2016), Bao et al. (2012), Date and Islyaev (2015), Kou (2002)) that a trader could use and they each require calibration. A trader is faced with Knightian uncertainty (see Knight (1921)) as to whether they have A. selected the correct model or B. correctly calibrated that model. A fuzzy logic approach to pricing options under uncertainty is used in Yoshida (2003).

A definition of uncertainty commonly used in the economic literature (see e.g. Jurado et al. (2015)) is of the form:

$$U_t^y(h) = \sqrt{E\left[(y_{t+h} - E\left[y_{t+h}\right])^2 | I_t\right]} \tag{XIII}$$

The uncertainty about a h-step ahead forecast of a variable or parameter, y, is given by the expected standard deviation of the forecast error of that variable, where the expectation is calculated using information available at the forecast time, I_t.

[7] The fitness function used was $-\sum |\frac{p_{bates}(x)}{p_{NIG}(x)} - 1|$

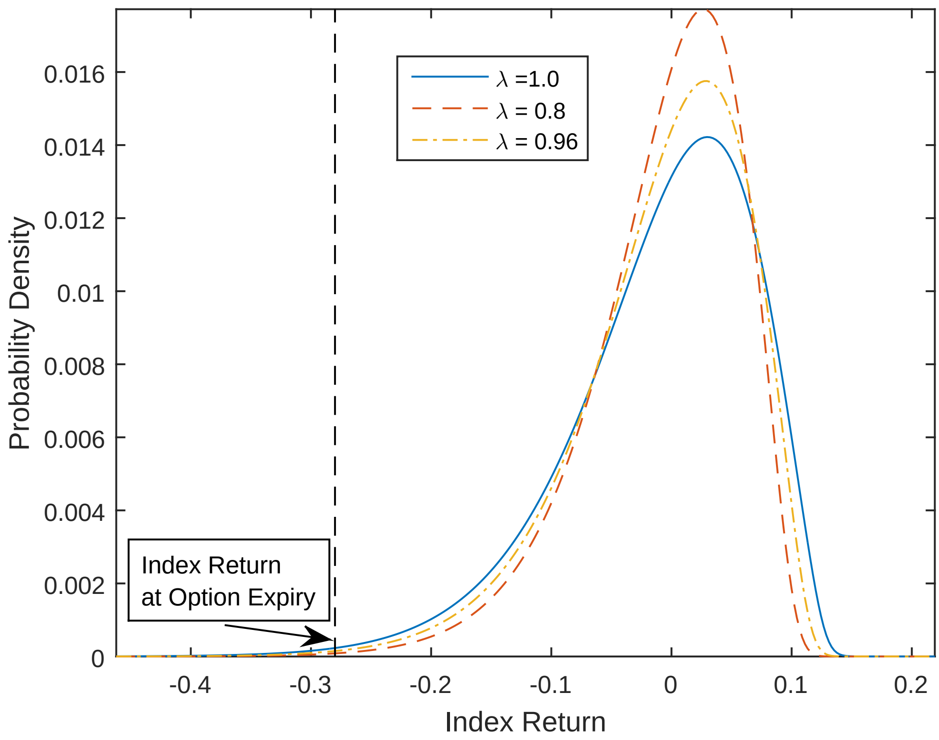


Figure 8: This figure illustrates how the $\Psi_\gamma(\lambda)$ analysis in the empirical section can be replicated under the Bates 2-factor stochastic volatility and jump model through varying the β_i parameters associated with variance premia in the model.

Table 6: Parameters for the Bates 2 factor model resulting from a genetic algorithm curve-fitting the RND obtained from a set of arbitrage-free prices generated under the model to the RND for a set of options traded on the 12th of September 2008 with expiry on the 18th of October 2008. The process was repeated for three densities, the original RND and two reduced variance versions ($\lambda = 0.8, 0.96$). The variance was reduced by adjusting the β_i parameters of the Bates model, these are associated with risk premia in the model. The resulting payoffs evaluated at expiry ($\Psi_1(1.0), \Psi_1(0.8), \Psi_1(0.96)$) closely match the payoffs obtained using the NIG in the empirical analysis and highlight that adding an upward bias to volatility significantly reduces the cost of a tail event (from a loss of 64% of capital to a loss of 36% of capital).

Parameter	Volatility Scaling Factor		
	$\lambda = 1.0$	$\lambda = 0.8$	$\lambda = 0.96$
Bates 2-Factor Stochastic Vol. Params			
α_1	0.003	-	-
β_1^*	0.068	5.43	2.154
σ_{v1}	0.594	-	-
ρ_1	-0.968	-	-
α_2	0.229	-	-
β_2^*	3.887	17.333	9.908
σ_{v2}	0.542	-	-
ρ_2	-0.99	-	-
Bates 2-Factor Jump Params			
λ_0	0.000	-	-
λ_1	2.741	-	-
λ_2	0.000	-	-
k^*	-0.069	-	-
δ	0.075	-	-
NIG Params			
mean	-0.002	-	-
volatility	0.079	0.063	0.076
skewness	-0.974	-	
kurtosis	4.8	-	
Volatility Arbitrage Payoff			
$\Psi_1(\lambda)$	1.0	0.36	0.64

Our results are consistent with traders being uncertain of their estimated parameters, in particular those pertaining to the tail of the forecast distribution. This is Knightian uncertainty – uncertainty about the forecasting model itself as opposed to the expected level of variation in the forecast error of the variable output by the model.

As tail events are rare, we might expect the estimation of the related tail parameters to be subject to a higher degree of uncertainty, $U_t^y(h)$. Trading the volatility spread leaves an agent exposed to errors in the tails of their forecast distribution. We have shown empirically that removing the upward bias on the second moment of the risk neutral density increases exposure to tail events and results in larger capital losses. Using the Bates model we show that our volatility scaling approach can be interpreted as a trader adjusting their variance premium through the model's β_i parameters. Our association of the spread with the cost of uncertain tail events is consistent with findings in Gruber et al. (2015), where the variance premia at short horizons of one month are almost completely explained by a time-varying premium for pure jump variance risk.

5. Conclusions

We present a new framework to trade the volatility spread resulting from the upward bias on option implied volatility as an estimator of future realized volatility. The scheme is constructed to be optimal for investors with power law utility and can be tailored to individual risk aversion levels. A portfolio of options is used to statically replicate a growth optimal payoff that trades a distribution with an unbiased second moment against the market-implied distribution.

The framework allows us to examine the implications of removing the volatility spread in a trading scheme that incorporates the impact of the higher moments of both the underlying return distribution and of the market risk neutral distribution. We find that removing the upward bias on forecast volatility is a negative expected growth strategy when trading against the market implied risk neutral density. The losses due to low probability tail events are more costly when an unbiased volatility forecast density is traded against the market implied density. The cost of these events over the data sample is not recouped by gains through exploiting the upward bias on implied volatility over the same period.

Our results are consistent with the volatility spread being added to mitigate the cost of unpredictable tail events. The process of volatility arbitrage moves probability density from the tails of a forecast density into the centre and this increases losses when tail events occur. We find that, despite its upward bias, the market pricing of implied volatility is efficient to the extent that trading the upward bias does not generate a long term return premium over the period of our study.

Appendix A. The Normal Inverse Gaussian Density

$$
\begin{aligned}
f(x, \alpha, \beta, \mu, \delta) &= \frac{(\frac{\delta * \alpha}{\pi}) \exp \delta \sqrt{\alpha^2 - \beta^2}}{\sqrt{\delta^2 + \bar{x}^2} K_1(1, (\alpha \sqrt{\delta^2 + \bar{x}^2})) \exp(\beta * (x - \mu))}, \\
&\text{where:} \\
K_1(\nu, z) &= \text{a modified bessel function of the second kind}, \\
\mu &= \text{a location parameter}, \\
\beta &= \text{an asymmetry parameter}, \\
\delta &= \text{a scale parameter}, \\
\alpha &= \text{a tail heaviness parameter}.
\end{aligned}
$$

A distribution's mean (m), variance (σ), skew (s) and kurtosis (k) values map to the NIG α, β, μ and δ parameters as follows:

$$
\begin{aligned}
\alpha &= \sqrt{\frac{3k - 4s^2 - 9}{\sigma^2 (k - 5/3s^2 - 3)^2}}, \\
\beta &= \frac{s}{\sigma (k - \frac{5}{3}s^2 - 3)}, \\
\mu &= m - 3s \frac{\sigma}{3k - 4s^2 - 9}, \\
\delta &= 3^{\frac{3}{2}} \frac{\sqrt{\sigma^2 (k - 5/3s^2 - 3)}}{3k - 4s^2 - 9}.
\end{aligned}
$$

Acknowledgements

The authors would like to thank the conference participants at the 2015 Infiniti meeting and the 2015 Forecasting Financial Markets meeting for their comments. This publication uses data provided with the financial support of Science Foundation Ireland under grant number 08/SRC/FM1389.

References

Yacine Aït-Sahalia and Andrew W Lo. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, 53(2):499–547, 1998.

Gurdip Bakshi and Nikunj Kapadia. Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies*, 16(2):527–566, 2003.

Gurdip Bakshi and Dilip Madan. Spanning and derivative-security valuation. *Journal of Financial Economics*, 55(2):205–238, 2000.

Gurdip Bakshi and Dilip Madan. A theory of volatility spreads. *Management Science*, 52 (12):1945–1956, 2006.

Gurdip Bakshi, Nikunj Kapadia, and Dilip Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, 16 (1):101–143, 2003.

Turan G Bali and Armen Hovakimian. Volatility spreads and expected stock returns. *Management Science*, 55(11):1797–1812, 2009.

Qunfang Bao, Shenghong Li, and Donggeng Gong. Pricing vxx option with default risk and positive volatility skew. *European Journal of Operational Research*, 223(1):246–255, 2012.

Andrew R Barron and Thomas M. Cover. A bound on the financial value of information. *Information Theory, IEEE Transactions on*, 34(5):1097–1100, 1988.

D. Bates. The crash of '87: Was it expected? the evidence from options markets. *Journal of Finance*, (46):1009–1044, 1991.

D. S. Bates. *Handbook of Statistics*, chapter Testing Options Pricing Models, pages 567–611. Elsevier, 1996.

Geert Bekaert, Marie Hoerova, and Marco Lo Duca. Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788, 2013.

Nicholas Bloom. The impact of uncertainty shocks. *econometrica*, 77(3):623–685, 2009.

Oleg Bondarenko. Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, 116(1):85–112, 2003.

Douglas T Breeden and Robert H Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, pages 621–651, 1978.

Linda Canina and Stephen Figlewski. The informational content of implied volatility. *Review of Financial studies*, 6(3):659–681, 1993.

Peter Carr and Roger Lee. Volatility derivatives. *Annual Review of Financial Economics*, 1(1):319–339, 2009.

Peter Carr and Liuren Wu. Variance risk premiums. *Review of Financial Studies*, 22(3): 1311–1341, 2009.

Mikhail Chernov and Eric Ghysels. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56(3):407–458, 2000.

Bent J Christensen and Nagpurnanand R Prabhala. The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125–150, 1998.

Paresh Date and Suren Islyaev. A fast calibrating volatility model for option pricing. *European Journal of Operational Research*, 243(2):599–606, 2015.

Theodore E Day and Craig M Lewis. Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52(1):267–287, 1992.

Itamar Drechsler. Uncertainty, time-varying fear, and asset prices. *The Journal of Finance*, 68(5):1843–1889, 2013.

David Edelman. On the financial value of information. *Annals of Operations Research*, 100 (1-4):123–132, 2000.

Anders Eriksson, Eric Ghysels, and Fangfang Wang. The normal inverse gaussian distribution and the pricing of derivatives. *The Journal of Derivatives*, 16(3):23–37, 2009.

Frank J Fabozzi, Tommaso Paletta, Silvia Stanescu, and Radu Tunaru. An improved method for pricing and hedging long dated american options. *European Journal of Operational Research*, 254(2):656–666, 2016.

S. Figlewski. *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, chapter Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio. Oxford University Press, 2010.

Ronald A Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, pages 507–521, 1915.

Gianluca Fusai, Guido Germano, and Daniele Marazzina. Spitzer identity, wiener-hopf factorization and pricing of discretely monitored exotic options. *European Journal of Operational Research*, 251(1):124–134, 2016.

Nicola Fusari. Matlab option pricing toolbox. 2015.

Eric Ghysels and Fangfang Wang. Moment-implied densities: Properties and applications. *Journal of Business and Economic Statistics*, 32(1):88–111, 2014.

Peter H Gruber, Claudio Tebaldi, and Fabio Trojani. The price of the smile and variance risk premia. *Swiss Finance Institute Research Paper*, (15-36), 2015.

Paolo Guasoni and Scott Robertson. Portfolios and risk premia for the long run. *Annals of Applied Probability*, 22(1):239–284, 2012.

Jens Carsten Jackwerth. Option implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives*, 7(2):66–82, 1999.

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

DJ Johnstone. Economic interpretation of probabilities estimated by maximum likelihood or score. *Management Science*, 57(2):308–314, 2011.

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

Kyle Jurado, Sydney C Ludvigson, and Serena Ng. Measuring uncertainty. *The American Economic Review*, 105(3):1177–1216, 2015.

Frank H Knight. Risk, uncertainty and profit. *New York: Hart, Schaffner and Marx*, 1921.

Steven G Kou. A jump-diffusion model for option pricing. *Management science*, 48(8): 1086–1101, 2002.

Christopher G Lamoureux and William D Lastrapes. Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *Review of Financial Studies*, 6(2):293–326, 1993.

Francis A Longstaff, Jun Pan, Lasse H Pedersen, and Kenneth J Singleton. How sovereign is sovereign credit risk? Technical report, National Bureau of Economic Research, 2007.

David Shimko. Bounds of probability. *Risk*, 6(4):33–37, 1993.

Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

Yuji Yoshida. The valuation of european options in uncertain environment. *European Journal of Operational Research*, 145(1):221–229, 2003.