Financial Analysts Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ufaj20

Option Pricing via Breakeven Volatility

Blair Hull, Anlong Li & Xiao Qiao

To cite this article: Blair Hull, Anlong Li & Xiao Qiao (2023) Option Pricing via Breakeven Volatility, Financial Analysts Journal, 79:1, 99-119, DOI: 10.1080/0015198X.2022.2100234

To link to this article: https://doi.org/10.1080/0015198X.2022.2100234

Published online: 23 Aug 2022.

Submit your article to this journal

Article views: 2037

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ufaj20

Financial Analysts Journal | A Publication of CFA Institute
https://doi.org/10.1080/0015198X.2022.2100234

Research

Option Pricing via Breakeven Volatility

Blair Hull, Anlong Li, and Xiao Qiao

Blair Hull is a founder and chairman of Hull Tactical Asset Allocation, LLC, Chicago, IL. Anlong Li is a senior financial engineer at Hull Tactical Asset Allocation, LLC, Chicago, IL. Xiao Qiao is an assistant professor at the City University of Hong Kong and Hong Kong Institute for Data Science, Hong Kong, China. Send correspondence to Xiao Qiao at xiaoqiao@cityu.edu.hk.

The fair value of an option is given by breakeven volatility, the value of implied volatility that sets the profit and loss of a delta-hedged option to zero. We calculate breakeven volatility for 400,000 options on the S&P 500 and build a predictive model for these volatilities. A two-stage regression approach captures the majority of the observed variation. By providing a link between option characteristics and breakeven volatility, we establish a non-parametric approach to pricing options without the need to specify the underlying price process. We illustrate the economic value of our approach with a simulated trading strategy based on breakeven volatility predictions.

Keywords: options; prediction; trading strategy; volatility

Disclosure: The authors report no funding sources. The authors alone are responsible for the content and writing of the paper. Hull Tactical Asset Allocation, LLC is a registered investment adviser that deploys strategies, such as the one described in this paper in investment products including separately managed accounts and an exchange-traded fund. All errors are our own.

PL Credits: 2.0

Introduction

The advent of option pricing models (Black and Scholes 1973; Cox, Ross, and Rubinstein 1979) launched a revolution in modern finance that saw explosive growth in options markets and massive interest in option pricing in the subsequent decades. Although the option pricing literature contains a multitude of models, the approaches often share a common idea: The value of an option is equal to that of a replication portfolio with the same eventual payoffs (Merton 1973). The most natural implementation of this simple and elegant idea is to build a distribution of outcomes from a large set of replication portfolios using historical data, which can then be used to price additional options. However, the implementation of this approach turns out to be fraught with issues. Zou and Derman (1999) point out that such replication can be time-consuming, and hedging errors due to inaccurate volatility forecasts or infrequent hedging make the exercise difficult. As a result, most researchers opt for a more structural approach: Specify a return process for the underlying asset, derive the pricing relationships, then calibrate the model to data.

We set out to build a non-parametric option pricing model based on Merton's (1973) original insight. In the past 20 years, computing power has grown rapidly and option databases have become more comprehensive. These developments enable us to overcome the limitations cited in Zou and Derman (1999). The first step towards an option pricing model is to build a measure for the appropriate value of options. Using historical prices, we calculate the fair value of an option, the value of implied volatility (IV) that sets the profit and loss of a delta-hedged option position to zero. This value is called breakeven volatility (BEV).

The profit and loss from a dynamically hedged option position recover the difference between implied volatility and realized volatility, which indicates a

The authors thank William Goetzmann (Executive Editor), Luis Garcia-Feijoo (Managing Editor), Daniel Giamouridis (Co-Editor), two anonymous referees, Petra Bakosova, Yeguang Chi, Jake DeSantis, Ruslan Goyenko, Steve Heston, Ray Iwanowski, Craig Iseli, Kris Jacobs, Jason Wei, Lei Wu, Qi Wu, and Zonghao Yang for helpful comments. We are especially grateful to Ben Albert for extensive discussions and data processing.

© 2022 CFA Institute. All rights reserved.

positive or negative risk premium (Bakshi, Cao, and Chen 1997; Christoffersen et al. 2018). In computing breakeven volatility, we set the profit and loss to zero, removing risk premia embedded in market prices. To the extent market implied volatility of an option differs from its breakeven volatility, the market price contains either a risk premium or a discount. If the IV were higher than the BEV, a delta-hedged short position in this option would result in a positive profit. If the IV were lower than the BEV, a delta-hedged long position would show a positive profit. In this sense, breakeven volatility provides the fair value of an option to both sides of the contract (Dupire 2006).

We build a large set of breakeven volatility values for options on the S&P 500 Index (SPX), and we construct a predictive model that connects option characteristics to BEV. By providing a link between moneyness, time to expiration, and other observable characteristics and the fair value of the option, the predictive relation can be interpreted as a non-parametric option pricing model.[1] Given a new option with its own set of characteristics, we can quickly determine its fair value.

Parametric approaches are favored over non-parametric ones when the underlying asset's return dynamics are known, but this is rarely the case in practice. The non-parametric approach to option pricing can be data-intensive, but it offers a promising alternative to standard parametric pricing models when parametric restrictions are violated. Since non-parametric models do not rely on restrictive assumptions, such as log-normality or sample-path continuity, they are robust to the specification errors that plague parametric models (Ait-Sahalia and Lo 1998).

We construct delta-hedged positions for nearly 400,000 S&P 500 options with a variety of strike prices and time to expiration ranging from 5 to 74 trading days. The empirical distribution of breakeven volatility is similar to, but distinct from, that of implied volatility. We construct volatility smirks for both measures of volatility, and we uncover the following pattern: Breakeven and implied volatility are almost identical in the at-the-money (ATM) region, but as we move further away from the ATM strike prices, the two measures diverge. Implied volatility is higher than breakeven volatility for strike prices lower than the underlying price, as well as for strikes exceeding the underlying. The shapes of the volatility smirks indicate that delta-hedged short positions tend to earn positive profits for out-of-the-money (OTM) puts, an empirical regularity confirmed in the literature (Coval and Shumway 2001; Bakshi and Kapadia 2003).

We build a predictive model of breakeven volatility using 12 variables and their transformations. In addition to typical parameters used in option pricing, such as moneyness and time to expiration, we also include variables that have the potential to bring added predictive power, such as the CBOE VIX Volatility Index (VVIX), the difference between the delta of the option and the ATM delta, and the sensitivity of the option price with respect to volatility. We use a two-stage regression approach to build a statistical model that connects the predictor variables and breakeven volatility. The first stage produces forecasts of logged breakeven volatility, and the second stage makes an adjustment such that the predicted breakeven volatility values are unbiased.

Our statistical model provides a reliable fit to breakeven volatility. The two-stage model captures 90% of the variation in breakeven volatility. Model diagnostics reveal that the model performs best between volatility values of 5 and 50%, but has difficulties if breakeven volatility is below 5%. Our model predictions exhibit convexity in volatility space, and option prices are consistent with an underlying return distribution that exhibits negative skew and excess kurtosis. While the above features can be carefully engineered into a parametric model, they arise naturally in our approach as we learn from data.[2]

We test the economic value of our statistical model through a simulated trading strategy that exploits the difference between implied volatility and our forecast of breakeven volatility. For a particular option, if its implied volatility were lower than the predicted BEV, the market price of the option is too low, and we take a delta-hedged long position. If the implied volatility were higher than the predicted BEV, the market price is too high, so we take a delta-hedged short position. We trade options whose breakeven volatility values differ more than $1 compared to the mid-point price.

In our testing period from January 2015 to November 2020, our statistical model achieves an impressive out-of-sample R-squared of 0.71. The BEV-based strategy identifies 6,783 trading opportunities, earning 6.8% per year with an annual volatility of 4.5%. With a Sharpe ratio of 1.50, this strategy outperforms three alternative options strategies in terms of expected returns per unit of risk. The BEV strategy also provides a more attractive risk-return tradeoff compared to a buy-and-hold strategy in the S&P 500, which has a Sharpe ratio of 0.79 in our sample period.

We are not the first to use historical data to compute terminal distributions and option prices. Stutzer (1996) and Zou and Derman (1999) also use historical data to compute option prices. Rather than looking at the replicated value of an option, these papers compute a risk-neutral distribution from the empirical distribution of the underlying asset. Stutzer (1996) notes the shortcoming of not incorporating market information related to option prices into his approach, stating "there is need for a more detailed comparison of … non-parametric model values to actual transaction prices." We answer Stutzer's call to action by using actual transaction prices to inform us about the fair values of options. Furthermore, Zou and Derman's (1999) goal is relative pricing: Take exchange-traded option prices as given, build a pricing model, then calculate the prices of less liquid, over-the-counter options. In contrast, our goal is absolute pricing—to determine the fair value of individual options based on their observable characteristics.

Some papers have used breakeven volatility as a diagnostic tool for market prices. Dupire (2006) builds breakeven volatility surfaces and explains that BEV is a fair volatility for both sides of an option contract. He highlights using BEV surfaces as a tool to understand market volatility surfaces. Mitoulis (2019) carries out breakeven volatility calculations using simulations, comparing delta-hedging results using Black and Scholes (1973) and Heston (1993) models. Neither paper attempts to forecast BEV values or explore trading implications of the differences between implied and breakeven volatilities. As far as we know, we are the first to build a non-parametric option pricing model via a predictive model of breakeven volatility.

A crucial distinction between our paper and the existing literature is the calculation of breakeven volatility. Zou and Derman (1999), Dupire (2006), and Mitoulis (2019) all construct a constant BEV throughout the life of an option of a given strike price and time to maturity. This calculation requires perfect foresight of the path of the option and is not feasible in real-time. Additionally, the resulting BEV surfaces often admit arbitrage. Our approach relaxes the constant breakeven volatility assumption and allows for a term structure of BEV, which is then solved by backward recursion. In doing so, we use the implied volatility path as well as the stock price path for the calculation. Breakeven volatility values calculated from our approach do not allow for arbitrage.

Our paper is also related to the literature on non-parametric approaches to option pricing. A strand of literature first developed in the 1990s. Hutchinson, Lo, and Poggio (1994) use neural networks to approximate the Black-Scholes formula. Ait-Sahalia and Lo (1998) propose a kernel density estimator of the risk-neutral distribution, effectively providing a link between option characteristics and the second derivative of the volatility smirk. These pioneering papers led to more recent literature on using machine learning methods to approximate the relationship between model inputs and outputs. Liu, Oosterlee, and Bohte (2019) propose a neural network model to approximate the Black-Scholes and Heston (1993) models. Manzo and Qiao (2021) use neural networks to approximate nine different credit risk models. Compared to these papers, we do not try to approximate existing option pricing models—we let the data dictate how option characteristics and prices are related. Furthermore, our target option price is based on a fair valuation calculation rather than the market price.

Breakeven Volatility

Data. We obtain option prices, implied volatility, and deltas from SpiderRock Platform Services LLC. SpiderRock is an options trading platform provider and data vendor based in Chicago that serves trading desks at large banks, hedge funds, and proprietary trading firms. SpiderRock's historical data and option analytics include implied volatility, option Greeks, risk metrics, and volatility surfaces derived from the live data from SpiderRock's trading systems.

Since OptionMetrics is commonly used in academic studies, we would like to point out two differences between SpiderRock and OptionMetrics data (Please refer to Appendix A for more details). First, daily options and underlying prices are recorded differently across the two databases. OptionMetrics uses option prices at 3:59 p.m. Eastern Time and stock settlement prices at market close at 4 p.m., whereas SpiderRock uses the same option price at 3:59 p.m.

but also uses stock prices at 3:59 p.m. Second, SpiderRock data imposes put-call parity whereas OptionMetrics maintain separate volatility surfaces for puts and calls. Even if OptionMetrics were to synchronize the option and stock quotes, one may still prefer using forward prices derived from put-call parity in implied volatility calculation.[3] For these reasons, we choose to use SpiderRock data. These data are used for the computation and prediction of breakeven volatility values in an attempt to use real-time market data that traders rely on to create trading signals.

Variables used to predict breakeven volatility are obtained from SpiderRock and Bloomberg. In particular, the strike price, days until expiration, implied volatility, and other information related to the underlying are from SpiderRock. The CBOE Volatility Index (VIX), the CBOE VIX Volatility Index, and variables used to construct a realized volatility forecast of the S&P 500 are from Bloomberg. Our sample is from January 2013 to November 2020.

Calculation of Breakeven Volatility. The breakeven volatility is the value that sets the profit and loss from a delta-hedged option position to be zero. Suppose we want to calculate the breakeven volatility for a call option. The delta-hedged profit and loss are given as follows:[4]

$$PnL_t(\sigma) = c_T - c_t - \sum_{\tau=1}^{T} \Delta_{\tau-1}(S_\tau - S_{\tau-1}) \qquad (1)$$

where c_T is the payoff of the call option at maturity, c_t is the call price at the time of initiating the position, Δ_τ is the option delta at time τ, and S_τ is the price of the underlying asset at time τ. The profit and loss of the delta-hedged option depend on the volatility of the underlying. By changing the volatility, we can trace out $PnL_t(\sigma)$ as a function of volatility. The breakeven volatility is the value that sets $PnL_t(\sigma)$ to zero. We can also compute breakeven volatility in a similar manner using put options.

To operationalize the above expression, we compute the breakeven volatility as follows:

$$PnL_{T-1}(\sigma_{T-1}) = c_T - c_{T-1}(\sigma_{T-1}) - \Delta_{T-1}(S_T - S_{T-1})$$

...

$$PnL_t(\sigma_t) = c_{t+1} - c_t(\sigma_t) - \Delta_t(S_{t+1} - S_t) \qquad (2)$$

For day τ, we solve for the breakeven volatility σ_τ such that PnL_τ is set equal to zero. For a call option, we start our computation when the call has five days until expiration, which is taken to be T in the above expressions. The call value at time T is taken to be the market value. For the previous day, at time $T - 1$, we compute the fair value of the call option that sets the profit and loss of the delta-hedged position on that day to be zero. We then solve for the fair value of the call option until we reach a time to expiration of 74 trading days, so we can build a term structure of breakeven volatility for a variety of time to expiration ranging from 5 to 74 days.[5] We compute the breakeven volatility for call and put options available from SpiderRock that have 5–74 days until expiration, including every option with a non-zero bid price.[6]

We transform the fair values into volatility space using the Black and Scholes (1973) model. A transformation of prices into volatilities aids a comparison between breakeven volatility and implied volatility. Volatilities are also more well-behaved compared to dollar amounts, which facilitates building a predictive model. Because the Black-Scholes equation is only used to transform option prices into a more convenient space, much like an affine transformation or taking logarithms, this transformation does not require that the Black and Scholes (1973) model is the correct pricing model (Shimko 1993).

To delta-hedge the call options, we take the deltas from SpiderRock, calculated using the Cox-Ross-Rubinstein (CRR, Cox, Ross, and Rubinstein 1979) binomial tree model.[7] We compute the fair value of the option until five trading days before maturity. We do not hold the option until expiration due to two reasons. First, due to the high gamma exposure in the few days immediately before expiration, volatility calculations can be quite noisy, and the associated volatility curves are not smooth functions of strike prices.[8] Second, option traders often do not roll over their positions on the expiration, but rather a several days before expiration. Our range of 5–74 trading days until maturity covers the typical time to expiration options traders engage in.[9]

Figure 1 provides a time series plot of the number of options we use to calculate breakeven volatility. On average, we have 278 strikes each day. This value is somewhat biased upwards, by the extreme days with more than 750 strikes; the median number of strikes we use is 244. Although the equity index options market is already quite liquid at the beginning of our sample in 2013, the number of strikes steadily

Figure 1. Number of Options Used to Calculate Breakeven Volatility

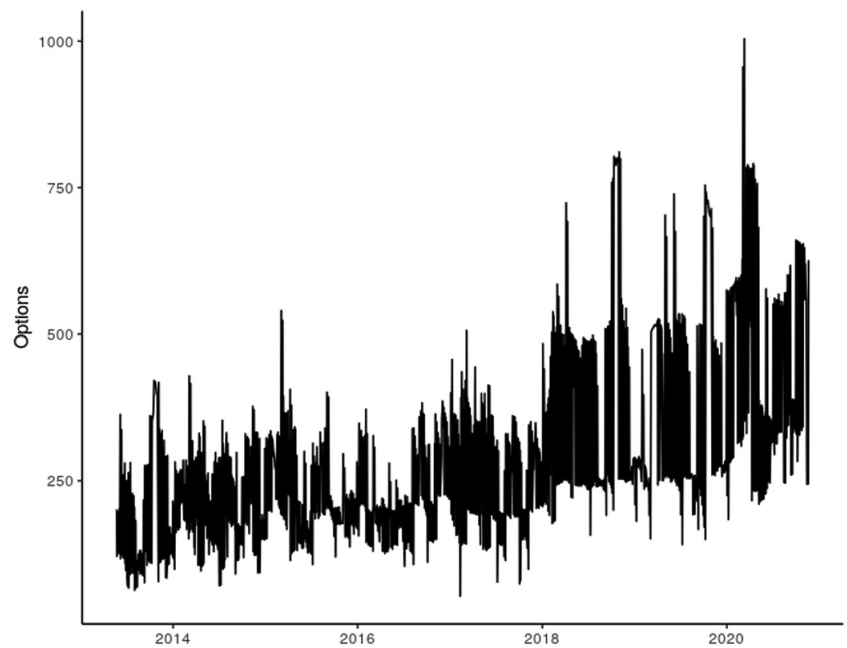


Table 1. Summary Statistics

S&P 500

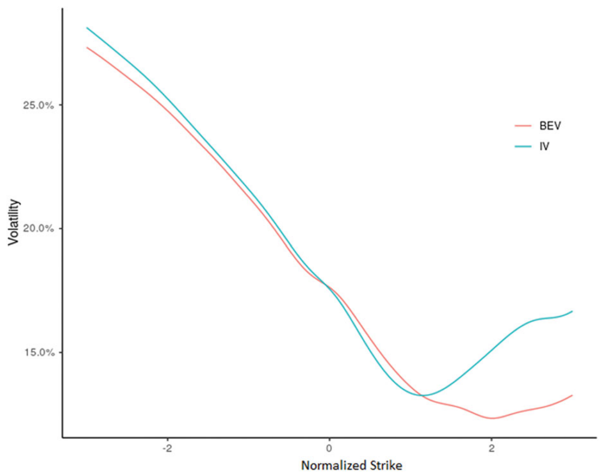
Variable	Mean	*SD*	10th	25th	Median	75th	90th
BEV	25.4%	16.2%	9.5%	14.1%	22.1%	32.2%	45.0%
IV	27.2%	15.6%	10.9%	15.4%	24.1%	34.8%	46.5%
BEV-IV	−1.8%	9.8%	−7.2%	−4.2%	−2.3%	−0.3%	3.0%

This table presents summary statistics for breakeven volatility and implied volatility of the S&P 500 Index. 10th, 25th, 75th, and 90th represent the quantile values. Our sample is from January 2013 to November 2020.

increased towards the later part of our sample. In total, we compute breakeven volatility for 396,899 options.

Panel A of Table 1 presents the summary statistics for breakeven volatility. From January 2013 to November 2020, the average value of all the computed breakeven volatilities is 25.4%, and the median is 22.1%. We observe a broad range of values—the 10th percentile of the empirical distribution is just 9.5%, whereas the 90th percentile is 45.0%. The standard deviation of breakeven volatilities is 16.2%. The distributional statistics for breakeven volatility are similar, but not identical to, those for implied volatility. In our sample period, the implied volatility of options on the S&P 500 index average 27.2%, with a median value of 24.1%. The standard deviation of implied volatility, 15.6%, is somewhat lower compared to breakeven volatility. The percentiles of implied volatility are also similar compared to those of breakeven volatility. We also present the difference between BEV and IV to understand their joint distribution. On average, BEV is lower than IV by −1.8%. Their difference has a standard deviation of 9.8%. BEVs are generally lower than IVs, as indicated by the negative percentile values up to the 75th, which reflects that in our sample, many options appear to

Figure 2. Volatility Smirks of BEV and IV



This figure shows the volatility smirks constructed from breakeven volatility and implied volatility for the S&P 500 Index, averaged across the full sample period and smoothed. The sample period is from January 2013 to November 2020.

be trading at a premium relative to their fair values.

Existing papers focusing on isolating a variance risk premium (VRP) often do so in the form of excess returns of delta-hedged positions (Bakshi, Cao, and Chen 1997; Goyenko and Zhang 2020; Bali et al. 2021). A delta-hedged short position that earns an average positive return indicates implied volatility is on average higher than subsequent realized volatility. By setting the delta-hedged profit and loss to zero, the breakeven volatility is a measure that sets the VRP to zero *for each individual option*. In this way, BEV calculations are more flexible than a single estimate of realized volatility, since we retain the flexibility of having different volatility values for options with the same maturity but different strike prices. Therefore, breakeven volatilities provide richer information set on the realized distribution.

Volatility Smirks. The idea of constructing replicating portfolios using historical data has been discussed in the literature, but implementation has been limited. Zou and Derman (1999) note that theoretically, the appropriate implied volatility for a given option is determined by the cost of replicating that option throughout its lifetime. However, the authors do not pursue this idea because "such replication can be time-consuming, and the hedging errors due to inaccurate volatility forecasting and infrequency of hedging make the exercise difficult." Greatly improved computing power in the past 20 years, as well as more comprehensive option datasets, overcome much of Zou and Derman's (1999) concerns. Mitoulis (2019) carries out breakeven volatility calculations using simulations but stops short of a more complete empirical analysis. Having constructed BEV values for a large set of S&P 500 options, we compare the behavior of breakeven volatility and implied volatility.

Figure 2 compares volatility smirks using breakeven and implied volatilities. We average across all dates and maturities to trace out the unconditional distributions for these volatility measures. We plot the volatility curves as a function of the normalized strike, the ratio of the log moneyness of the option divided by its scaled volatility.[10] Defined this way, the normalized strike provides a better measure to compare options that have very different volatility levels or time to expiration. There is a limited number of listed call options with high strike prices, which causes the volatility calculations to be truncated on the right side, whereas the left side extends much further. This asymmetry around the number of strikes around the at-the-money value is commonly found in the literature (e.g., Figlewski and Malik 2014). Since the 1987 stock market crash, equity index options have shown a persistent skew to OTM puts. The most common explanation for this pattern is that market participants want to protect against the possibility of a large market crash, but they do not necessarily purchase deep OTM calls in anticipation of a large upward move. This sort of market participant preference can explain why we observe many more strike values below the current price of the underlying.

In the vicinity of the at-the-money region, breakeven volatility and implied volatility are largely similar. When we get farther away from ATM strike prices, the two volatility measures diverge. In the regions where the strike price exceeds the underlying price—the normalized strike is negative—implied volatility is higher than breakeven volatility, indicating that delta-hedged short positions in puts or calls with these strikes tend to have positive profits on average. Similarly, in the regions where the strike price is lower than the underlying price, the breakeven volatility is also lower than the implied volatility. The implications for the profitability of delta-hedged positions from Figure 2 are consistent with documented empirical facts in the literature. Coval and Shumway (2001) and Bakshi and Kapadia (2003) find that short delta-hedged positions are profitable for out-of-the-money puts.

A Data-Driven Valuation Process

In the previous section, we used historical data to build a database of nearly 400,000 breakeven volatilities for options of different moneyness and time to expiration. While this exercise presents us with the fair values of options, the breakeven volatility calculation uses future information, so it cannot be computed in real-time. Suppose an investor wants to know what an option with a time to expiration of 70 days should be worth. She cannot calculate the breakeven volatility for this option without knowing its future price path. To compute the fair value of an option in real-time, we require a model that establishes a link between the current observable characteristics of an option to its breakeven volatility. To this end, we build a predictive model for breakeven volatility using a variety of input variables. The mapping from input variables to the breakeven volatility learned by our model constitutes an option pricing model.

Volatility Predictors.

We consider the following set of predictors for breakeven volatility. Whereas some predictors are motivated by financial theory, others are included for their likely potential to contain strong predictive power. Since our primary goal is to make a robust prediction for breakeven volatility, we are not restricted to only using variables that have a strong economic motivation. All variables are calculated at the daily frequency.

1. **VIX** captures the market's expectation for the implied volatility of the S&P 500 Index over the next 30 days. It is commonly used by investors to gauge market sentiment. We use the log of the CBOE Volatility Index at the market close.
2. **VVIX** is a volatility of volatility measure that captures the expected volatility of the 30-day forward price of the VIX. We use the closing value of the CBOE VVIX Index.
3. **rv** is a GARCH forecast of the volatility of the S&P 500 returns over the same horizon as the time to expiration of the option, including an adjustment for the overnight component of volatility.
4. **rt** provides a measure for the time to expiration. It is calculated as the square root of the number of trading days to expiration divided by 252.
5. **lsk** is a measure of the moneyness of the option calculated as the natural log of the price of the underlying over the strike price. This variable is motivated by the empirical finding that volatility is not constant across moneyness (Zou and Derman 1999), and it allows the breakeven volatility to vary as a function of moneyness.
6. **ATM.IV** is the implied volatility of the option closest to at-the-money, calculated using the Black and Scholes (1973) model.
7. **ImpliedVol** is the market implied volatility of the option calculated using the Black and Scholes (1973) model.
8. **RR** is the difference between the delta of the option and the ATM delta. This variable provides another way to capture higher volatilities in the wings compared to ATM options.
9. **vega** is the sensitivity of the option price with respect to volatility. We use the Black-Scholes vega, equal to $S_t N'(d_1)\sqrt{T-t}$, where S_t is the price of the underlying, $N'(\cdot)$ is the probability density function of standard normal distribution, d_1 comes from the Black-Scholes formula,[11] and $T-t$ is the time to expiration.
10. **gamma** is the rate of change in the delta with respect to changes in the underlying price, or the second derivative of the option price with respect to the price of the underlying. We calculate the Black-Scholes gamma, $\frac{N'(d_1)}{S_t\sigma\sqrt{T-t}}$, where σ is the volatility of the underlying.
11. **vrt** is defined as the VIX multiplied by the square root of time to expiration. This variable is a construction of $\sigma\sqrt{T-t}$ from the Black-Scholes model. Breaking up the Black-Scholes model into its constituent parts allows for additional flexibility in predicting breakeven volatility.
12. **vvt2** is defined as the VIX squared multiplied by the time to expiration. This variable is another piece in the Black-Scholes model, $\frac{\sigma^2}{2}(T-t)$.

We also include some transformations of the above variables. lsk2 is the squared value of lsk, motivated by the empirical observation that volatility measures are a convex function of moneyness. ImpliedVol2 is equal to ImpliedVol squared. RR2 and RR3 are the squared and cubed values of RR, motivated by interpolation methods used in the literature to build volatility curves that try to capture the curvature in the volatility surface through a cubic

Table 2. Summary Statistics of Predictors

Variable	Mean	SD	10th	25th	Median	75th	90th
VVIX	97.67	18.31	79.53	85.83	93.97	106.01	118.94
VIX	18.08	8.93	11.53	12.85	15.00	20.10	27.96
rt	0.39	0.10	0.24	0.32	0.40	0.47	0.52
lsk	0.15	0.21	−0.07	0.00	0.12	0.26	0.40
lsk2 (×10)	0.68	1.73	0.00	0.03	0.17	0.70	1.63
ATM.IV	16%	8%	10%	12%	14%	18%	25%
ImpliedVol	27%	16%	11%	15%	24%	35%	46%
ImpliedVol2	10%	14%	1%	2%	6%	12%	22%
RR	0.21	0.35	−0.45	−0.01	0.40	0.47	0.49
RR2	0.17	0.08	0.02	0.11	0.20	0.23	0.24
RR3	0.04	0.07	−0.09	0.00	0.07	0.11	0.11
vega	1.50	1.55	0.09	0.25	0.88	2.40	3.88
vega2	4.63	7.87	0.01	0.06	0.78	5.77	15.08
gamma (x100)	1.14	1.72	0.10	0.17	0.39	1.28	3.43
vrt (x10)	0.63	0.34	0.29	0.41	0.57	0.75	1.01
vvt2 (x100)	0.25	0.36	0.04	0.08	0.16	0.28	0.51
rv	0.24	0.18	0.14	0.16	0.19	0.27	0.39

This table presents summary statistics for the different variables used to predict breakeven volatility. The sample period is from January 2013 to November 2020.

spline (Malz 2014; Neuberger 2012). vega2 is vega squared.

Table 2 shows the summary statistics for the predictor variables, and Table 3 provides the pairwise correlations. This table highlights the diversity of our predictors; most variables are weakly correlated with others. For example, the realized volatility forecast rv is basically uncorrelated with the time to expiration rt or the moneyness measures lsk and lsk2, but it is positively correlated to volatility variables, such as the at-the-money implied volatility, VIX, VVIX, as well as vrt and vvt2. Variable transformations are often highly correlated with the original variables: lsk and lsk2 have a correlation of 0.77, ImpliedVol and ImpliedVol2 have a correlation of 0.93, and vega and vega2 have a correlation of 0.94. Some variable pairs show large negative correlations. Gamma has a −0.79 correlation with RR, whereas RR2 has correlations of −0.90 and −0.79 with vega and vega2.

We include several volatility measures as predictors, including VIX, ATM.IV, and rv. As one would expect, these variables are highly correlated with one another. To limit multicollinearity, we reduce the correlations among these predictors through orthogonalization, via a Gram-Schmidt process (Cheney and Kincaid 2009). We run an ordinary least squares regression of VIX on ATM.IV, then we take the residual of the regression, VIX.res, as the new predictor in place of VIX. We repeat the same process for rv to obtain rv.res. By construction, VIX.res and rv.res are uncorrelated with ATM.IV. Additionally, they are also only 0.33 correlated with each other. We do not orthogonalize ImpliedVol or ImpliedVol2, since these have greater variation depending on the moneyness of the option and tend to bring significantly different information compared to the previous three volatility variables.

ATM.IV, VIX.res, and rv.res are all included as predictors in our model. Presumably, the idiosyncratic component of each volatility measure contributes to predicting breakeven volatility, rather than just the common component. If only the common component of the volatility measures mattered for prediction, it would not be necessary to retain all three volatility measures—some form of average would suffice. For example, we could construct a derived predictor as the largest principal component (Hull and Qiao 2017). Because the three volatility measures capture different information related to breakeven volatility, we choose to include all three in the prediction model.

The first row of Table 3 contains correlations among the predictor variables and breakeven volatility, our prediction target. Correlation is one measure of how

Table 3. Correlation Matrix of Predictors

	VVIX	VIX	rt	lsk	lsk2	ATM.IV	ImpVol	ImpVol2	RR	RR2	RR3	vega	vega2	gamma	vrt	vvt2	rv
BEV	0.35	0.39	-0.06	0.66	0.51	0.38	0.81	0.74	0.47	0.30	0.49	-0.29	-0.19	-0.46	0.28	0.26	0.34
VVIX	1.00	0.84	-0.01	-0.04	0.09	0.79	0.36	0.34	-0.14	-0.11	-0.17	0.18	0.16	-0.12	0.66	0.65	0.75
VIX	0.84	1.00	-0.03	-0.06	0.09	0.98	0.41	0.38	-0.17	-0.12	-0.20	0.16	0.15	-0.11	0.81	0.81	0.92
rt	-0.01	-0.03	1.00	0.09	0.11	0.05	-0.12	-0.12	-0.03	-0.19	-0.06	0.39	0.36	-0.21	0.53	0.33	0.08
lsk	-0.04	-0.06	0.09	1.00	0.77	-0.06	0.81	0.74	0.72	0.42	0.76	-0.46	-0.34	-0.52	-0.01	-0.05	-0.02
lsk2	0.09	0.09	0.11	0.77	1.00	0.10	0.67	0.77	0.26	0.29	0.30	-0.30	-0.21	-0.21	0.14	0.11	0.16
ATM.IV	0.79	0.98	0.05	-0.06	0.10	1.00	0.40	0.37	-0.17	-0.14	-0.21	0.19	0.18	-0.13	0.84	0.82	0.91
ImpVol	0.36	0.41	-0.12	0.81	0.67	0.40	1.00	0.93	0.55	0.40	0.59	-0.41	-0.29	-0.51	0.27	0.27	0.39
ImpVol2	0.34	0.38	-0.12	0.74	0.77	0.37	0.93	1.00	0.37	0.30	0.41	-0.32	-0.22	-0.33	0.24	0.24	0.38
RR	-0.14	-0.17	-0.03	0.72	0.26	-0.17	0.55	0.37	1.00	0.24	0.96	-0.31	-0.26	-0.79	-0.16	-0.15	-0.16
RR2	-0.11	-0.12	-0.19	0.42	0.29	-0.14	0.40	0.30	0.24	1.00	0.29	-0.90	-0.79	0.00	-0.21	-0.16	-0.08
RR3	-0.17	-0.20	-0.06	0.76	0.30	-0.21	0.59	0.41	0.96	0.29	1.00	-0.36	-0.28	-0.78	-0.20	-0.19	-0.18
vega	0.18	0.16	0.39	-0.46	-0.30	0.19	-0.41	-0.32	-0.31	-0.90	-0.36	1.00	0.94	-0.02	0.36	0.27	0.15
vega2	0.16	0.15	0.36	-0.34	-0.21	0.18	-0.29	-0.22	-0.26	-0.79	-0.28	0.94	1.00	-0.05	0.35	0.27	0.14
gamma	-0.12	-0.11	-0.21	-0.52	-0.21	-0.13	-0.51	-0.33	-0.79	0.00	-0.78	-0.02	-0.05	1.00	-0.20	-0.14	-0.12
vrt	0.66	0.81	0.53	-0.01	0.14	0.84	0.27	0.24	-0.16	-0.21	-0.20	0.36	0.35	-0.20	1.00	0.92	0.82
vvt2	0.65	0.81	0.33	-0.05	0.11	0.82	0.27	0.24	-0.15	-0.16	-0.19	0.27	0.27	-0.14	0.92	1.00	0.85
rv	0.74	0.91	0.09	-0.08	0.08	0.90	0.34	0.31	-0.17	-0.13	-0.20	0.19	0.18	-0.12	0.82	0.84	1.00

This table presents the pairwise correlations for the predictor variables and breakeven volatility. Positive correlations are shown in shades of red, whereas negative correlations are shown in shades of blue. The shade of the cell indicates the strength of the correlation. The sample period is from January 2013 to November 2020.

well each variable would do in univariate predictive regressions. Almost all the predictors exhibit considerable correlations with our prediction target, although the relationship can be positive or negative. VVIX, lsk, RR, and vrt all have a positive relationship with breakeven volatility; vega, vega2, and gamma have a negative relationship with BEV. Suggestive economic interpretations of a positive or negative correlation with BEV can be found in our description of the predictor variables.

Statistical Model. Our goal is to make good predictions of breakeven volatility. To the extent linear models can offer good predictions for the target variable, they are preferred over more complex models given their simplicity and interpretability. Our prediction target is $BEV_{i,\ K,\ t,\ T}$ for option i with strike price K and time to expiration $T-t$. Because the BEV calculation requires knowledge of the path of the option and the underlying, it is only known on the expiration date T. We use predictor variables at time t, $X_{i,m,t}$, to form a conditional expectation of the actual BEV value at time T. To simplify notation, we will denote the target and predictor variables as BEV_i and $X_{i,m}$ with the understanding that they are known at different times.

The raw BEV values are truncated at zero and show significant right skew. In the first stage, we transform the dependent variable, BEV, into its logarithms. By working in the log space of BEV, we obtain a distribution well-approximated by a normal distribution, and the predicted breakeven volatility value will be guaranteed to be positive. Our first-stage regression models logged BEV as a linear function of the predictor variables:

$$\log(BEV_i) = \beta_0 + \sum_{m=1}^{M} X_{i,m}\beta_m + \epsilon_i \tag{3}$$

where BEV_i is the breakeven volatility linked to a particular set of parameters including time to expiration and moneyness, $X_{i,m}$ are the M predictors associated with BEV_i, and ϵ_i is the prediction error.

For a given set of predictor variables, Equation (3) provides unbiased estimates for the expectation of logged BEV, $E[\log(BEV)]$. However, our goal is to find the best prediction for the expectation of BEV, $E[BEV]$. The approach to retransform a predicted quantity back to its original scale is known as a "smearing adjustment" in statistics (Duan 1983; Taylor 1986). We cannot simply exponentiate our prediction from Equation (3), because $e^{E[log(BEV)]} \leq E[BEV]$ by Jensen's inequality.[12] This point has been emphasized by Duan (1983), who asserts that unbiased and consistent predictions on a transformed scale (using a monotonic function) do not transform into unbiased or consistent quantities on the untransformed scale.

If the residuals from Equation (3) were normally distributed, there is a closed-form transformation of the prediction of Equation (3), $E[\log(BEV)]$, into $E[BEV]$:

$$E[BEV] = e^{E[\log(BEV)]} e^{\frac{1}{2}\eta^2} \tag{4}$$

where η is the standard deviation of the residuals ϵ_i of Equation (3). For residuals that are approximately normally distributed, the above expression provides a fairly precise adjustment from the logged values to the original values (Duan 1983). However, the adjustment does not work well if the residuals were not normally distributed. In the case of non-normal residuals, Duan (1983) proposes the following:

$$E[BEV] = e^{E[\log(BEV)]} \frac{1}{N} \sum_{j=1}^{N} e^{\epsilon_j} \tag{5}$$

where N is the number of data points used to estimate Equation (3), and ϵ_js are the residuals. Duan (1983) posits that like Equation (4), our desired quantity $E[BEV]$ and the exponentiated prediction from Equation (3) $e^{E[log(BEV)]}$ differ by a constant multiple. Whereas normal theory implies a multiplicative factor of $e^{\frac{1}{2}\eta^2}$, Duan's is $\frac{1}{N}\sum_{j=1}^{N} e^{\epsilon_j}$.

Empirically, the residuals from the first-stage regression are not normally distributed. If we were to apply Equation (4), the resultant breakeven volatility values differ significantly from the actual values. Equation (5) provides a better adjustment factor, but the resulting BEV values still make poor predictions for the actual values. To overcome this retransformation issue, we use a more general, supervised approach to estimate the adjustment factor in a second-stage regression:[13]

$$BEV_i = \gamma e^{E[\widehat{\log(BEV_i)}]} + e_i \tag{6}$$

where $\widehat{log(BEV_i)}$ are the predicted values from the first-stage regression in Equation (3). Equation (6) is estimated without an intercept, adhering to the constant multiple relationships derived in Duan (1983), and γ is an estimate of the multiplicative factor that wedges between $E[BEV]$ and $e^{E[log(BEV)]}$. The final prediction of the breakeven volatility comes from the predicted component of Equation (6), $\hat{\gamma}e^{E[\widehat{log(BEV_i)}]}$. Combining the two stages, the following expression links the predictor variables with the final prediction of BEV:

$$E[BEV_i] = \hat{\gamma} e^{\widehat{\beta_0} + \sum_{m=1}^{M} X_{i,m}\widehat{\beta_m}} \tag{7}$$

The in-sample results of the first-stage regression are shown in Table 4. The predictors are all statistically significant at the 1% level, indicating strong predictive power of these variables for logged values of breakeven volatility. Taken together, the predictor variables can explain 53% of the variation in BEV. The adjusted R-squared is nearly identical to the unadjusted R-squared because we have many observations compared to the number of regressors.[14]

In-sample R-squared can overstate the predictive power of our model. To more accurately assess how well our model can predict breakeven volatility, we use a 10-fold cross-validation approach to calculate the coefficient of determination. We partition our data into 10 subsamples of roughly equal size. For each subsample, we fit our statistical model on the remaining 90% of data and use the fitted model to make predictions on the subsample. We repeat this procedure to obtain 10 validation R-squareds and take their average to be the cross-validated R-squared. The cross-validated R-squared is very close to the full-sample R-squared, indicating that our prediction model performs with a high degree of stability on this dataset. Indeed, the 10 validation R-squareds

Table 4. Predicting Logged Breakeven Volatility

	Estimate	SE	*t*-Stat
(Intercept)	1.85	0.02	116.7
VVIX	0.187	0.009	20.6
VIX.res	0.002	0.001	2.6
rt	−0.59	0.03	−19.7
lsk	0.69	0.02	30.0
lsk2	−0.11	0.01	−7.6
ATM.IV	1.71	0.06	28.1
ImpliedVol (coef × 10)	2.08	0.04	52.7
ImpliedVol2	−0.01	0.00	−44.8
RR	0.15	0.01	15.5
RR2	0.90	0.03	26.8
RR3	1.29	0.06	22.3
vega	2.05	0.37	5.5
vega2	0.54	0.04	12.3
gamma	−67.53	1.44	−46.8
vrt	4.33	0.20	21.7
vvt2	−33.74	0.85	−39.5
rv.res	0.43	0.01	34.1
R-squared		0.53	
Adjusted R-squared		0.53	
10-Fold CV R-squared		0.53	
Number of observations		396,899	

This table shows the first stage of the predictive model for breakeven volatility. We use linear regression to predict the logged values of breakeven volatility.

$$\log(BEV_i) = \beta_0 + \sum_{m=1}^{M} X_{i,m}\beta_m + \epsilon_i$$

where BEV_i is the breakeven volatility linked to a particular set of parameters including time to expiration and moneyness, $X_{i,m}$ are the M predictors associated with BEV_i, β_m are regression coefficients. The coefficients associated with VVIX, ATM.IV, ImpliedVol, ImpliedVol2, vega, vega2, and rv.res are multiplied by 100 for ease of exposition. Standard errors and t-statistics adjusted for heteroskedasticity are shown.

Table 5. Smearing Adjustment of Breakeven Volatility Prediction

	Estimate	SE	*t*-Stat
FirstStagePrediction	1.03	0.001	1310
R squared		0.90	
Adj. R squared		0.90	
10-Fold CV R-squared		0.66	
Number of observations		396,899	

This table shows the second stage of the predictive model for breakeven volatility. We regress breakeven volatility values on the predictions from the first stage.

$$BEV_i = \gamma e^{E[\widehat{\log(BEV_i)}]} + e_i$$

where BEV_i is the breakeven volatility and $\widehat{log(BEV_i)}$ is the predicted value from the first-stage regression. γ is an estimate of the smearing adjustment. Standard errors and *t*-statistics are adjusted for heteroskedasticity.

show a narrow range between 0.51 to 0.56, with a standard deviation of just 0.012. The regression coefficients show a high level of stability across validation folds.

If high correlations among the predictors lead to multicollinearity issues, the statistical integrity of the model may be compromised. While ordinary least squares estimates remain unbiased, their variances can be large. Ridge regression helps to resolve multicollinearity by allowing for better convergence of the variance-covariance matrix and introducing a small degree of bias to produce more reliable estimates. We check the robustness of our prediction using ridge regression, whose penalization parameter we select using 10-fold cross-validation. We find that the optimal penalization parameter is zero, indicating that the optimally-selected ridge model is equivalent to ordinary least squares. We further explore a range of possible values for the penalization parameter, and we confirm that the model with the lowest mean squared error is the one with the penalization parameter set to zero. These results suggest that multicollinearity does not pose a major issue for our model.[15]

Table 5 presents the smearing adjustment according to Equation (6). As implied by the equation, this second-stage regression does not include an intercept term. The smearing adjustment indicates $E[BEV] = 1.03\ e^{E[log(BEV)]}$, consistent with the implications of Jensen's Inequality $e^{E[log(BEV)]} \leq E[BEV]$. This adjustment procedure shows a close relationship between the predicted log breakeven volatility values and actual BEV—the full-sample R-squared is 0.90, and so is the adjusted R-squared. The 10-fold cross-validated R-squared is somewhat lower but still suggests that the majority of the variation in breakeven volatility is captured by our model.

Our approach attempts to estimate breakeven volatilities as a fixed function of certain characteristics of options, such as the current level of the underlying and moneyness. If our model captures the data well, its functional form should be relatively stable over time, although the actual breakeven volatility predictions can change over time or across strike prices if the characteristics on which the prediction is based on change.

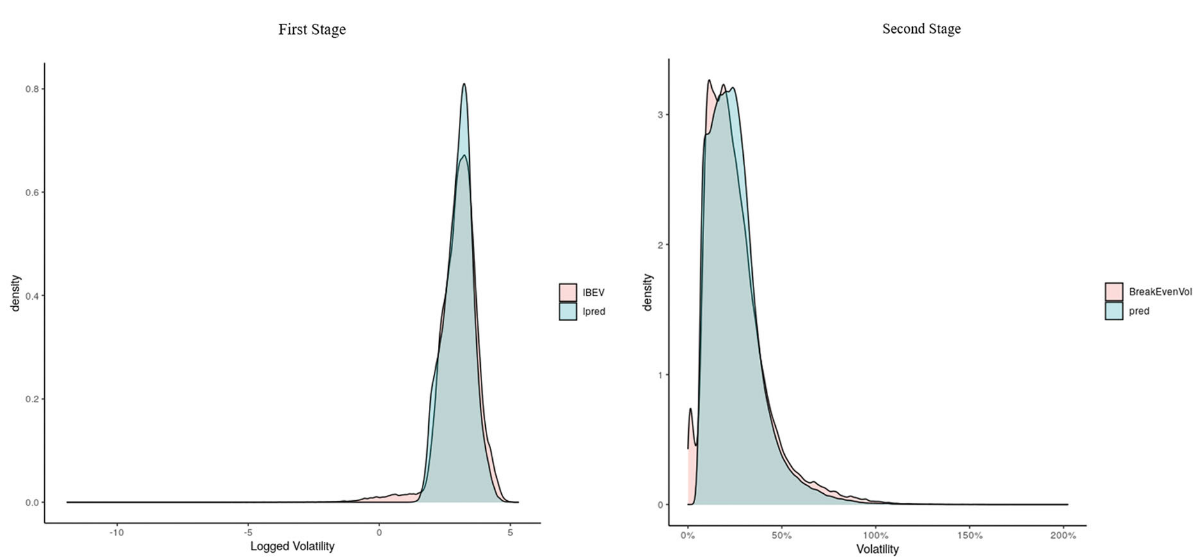
Our predictive model takes a data-driven approach to quantifying the value of an option, which stands in contrast to commonly used structural approaches to option pricing. Existing approaches typically specify key parameters necessary for calculating options prices—volatility, time to expiration, volatility of volatility, etc., then use these parameters to derive the value of an option. In comparison, a data-driven approach learns the relationship between the variables that are important for option valuation, rather than specifying the model parameters in advance.

Any pricing model must provide a link between a set of input variables and output variables, typically taken to be the option prices or volatilities. For example, the Black-Scholes model provides a mapping between the model parameters (the price of the underlying, volatility of the underlying, time to expiration, strike price, and interest rate) and the dollar value of an option. The Heston (1993) model stipulates a different mapping for a set of model parameters and the price of an option. Our approach also establishes a link between the predictor variables and the breakeven volatility, a measure of the fair value of an option. However, the relationship between the predictor variables and BEV is learned from a large database of historical BEV values and predictors, which allows for more flexibility compared to the examples above.

Researchers have explored non-parametric option pricing models in the past. Hutchinson, Lo, and Poggio (1994) use neural networks to approximate the Black-Scholes formula, but the paper does not consider a non-parametric estimation of the relationship between option characteristics and option prices. Ait-Sahalia and Lo (1998) propose a kernel regression estimator of the risk-neutral distribution conditional on option characteristics, such as the price of the underlying, strike price, and time to expiration. In the same spirit, we also try to capture the relationship between predictor variables and option values in a data-driven manner. Our approach is different in that our target is breakeven volatility, which serves as an estimate for the fair value of an option.

Model Diagnostics. Although economic intuition may guide us in thinking about the univariate relationships among predictors and breakeven volatility, it is much more difficult to determine

Figure 3. Probability Densities of Actual and Predicted Values



The figures below show density plots of the first and second-stage regressions. In each plot, kernel densities of the prediction target and their predicted values are shown. In the first-stage regression, the prediction target is logged breakeven volatilities. In the second stage, the target is the level of breakeven volatility.

the signs of these relationships in a multivariate setting. Rather than focus on the contribution of individual predictors, we conduct model diagnostics to identify the strengths and weaknesses of our predictive model. We investigate the statistical validity of the model, as well as the economic implications of the volatility predictions. To start, we inspect the differences between the prediction target and the predicted values. In Figure 3, we plot the kernel densities of the prediction target and the predicted values in the same figure. A comparison of the two densities can reveal where our model is doing well and where it is failing. In the first-stage regression, the prediction target is the logarithm of breakeven volatilities. The predictions match the target values reasonably well in the middle part of the distribution but deviate from the target in the tails. In both the left and right tails, there are too few predicted values compared to the target distribution, and there are too many predicted values hovering in the center of the distribution. The model's predictive power appears to be the strongest when the breakeven volatility is between 5 and 50%.

In the second-stage regression, the left-hand variable is breakeven volatility and the right-hand variable is a transformation of logged BEV values. Like the first-stage regression results, the centers of the two distributions mostly overlap, although it is clear that the predicted values are more concentrated in the middle of its distribution than the tails. The model predictions have the greatest difficulty with the left tail; there are too few predicted values compared to the actual breakeven volatility values, especially if BEV is <5%. The predicted distribution and the actual distribution of breakeven volatilities are more closely matched in their right tails, albeit we still observe too few predictions compared to the target density.

We verify that the predicted volatilities do not violate arbitrage conditions. Orosi (2015) derives static no-arbitrage conditions for call options: Prices must be a decreasing and convex function of the strike price. We check the parallel conditions for puts: Prices are an increasing and convex function of the strike price. These conditions can be verified by checking whether put prices violate arbitrage conditions implied by putting spreads and butterfly spreads. A put spread combined buying and writing put options with the same expiration but different strike prices. Suppose we have two put options p_1 and p_2, with the same expiration date but different strike prices, K_1 and K_2. To prevent risk-free profits, we need the following condition to hold:

$$p_2 \geq p_1 \text{ for } K_2 > K_1 \tag{8}$$

If the above condition were violated, risk-free profit would be available. If $p_2 \leq p_1$, a portfolio that combines a long position in p_2 and short position in p_1 is costless to initiate, and will offer a non-negative payoff in all possible states of the world. The no-arbitrage condition for put spreads compels put prices to be an increasing function of the strike price, an intuitive condition that keeps the volatility curve well-behaved.

For put prices to be a convex function of the strike price, the no-arbitrage condition for butterfly spreads must be satisfied. A butterfly spread uses three contracts: Buy a put at the lowest strike, buy a put at the highest strike, and write two puts at an intermediate strike. These positions form a portfolio that earns the maximum payoff if the underlying does not move—a short volatility position. The no-arbitrage condition for butterfly spreads is the following:

$$p_2 \leq p_1 \frac{K_3 - K_2}{K_3 - K_1} + p_3 \frac{K_2 - K_1}{K_3 - K_1} \tag{9}$$

for

$$K_3 > K_2 > K_1$$

In the above expressions, K_1, K_2, and K_3 are strike prices of increasing value. p_1, p_2, and p_3 are the put prices associated with the respective strikes. This condition states that the put price associated with the intermediate strike, p_2, must be less than a weighted average of the put prices associated with the highest and lowest strikes.[16] The absence of arbitrage from butterfly spreads implies that put prices must be a convex function of the strike price, such that the shape of the volatility curve matches its well-known empirical profile (Bollen and Whaley 2004).

For each unique combination of date and time to expiration, options with different strike prices trace out a volatility curve. We examine the predicted BEV values using the above conditions, and we find that none of the volatility curves admit arbitrage.

Volatility Arbitrage

In the previous section, we built a predictive model for breakeven volatility, the fair value of an option that does not admit any risk premia. Implied volatility from market prices includes the risk attitudes of market participants and therefore embeds risk premia. To the extent our predictive model can make accurate forecasts of breakeven volatility, we can potentially convert the difference between implied and breakeven volatility into trading profits. In this way, the economic value of our model can be understood through a simulated trading strategy. The purpose of this section is to evaluate whether the difference between implied volatility and predicted breakeven volatility can generate profitable trading opportunities.

Simulated option trading strategies have often been used in the literature to determine the difference between implied and realized volatility. Coval and Shumway (2001) and Bakshi and Kapadia (2003) find large returns to delta-hedged options and interpret this result as a volatility risk premium. Bollen and Whaley (2004) attempt to match the profits of delta-hedged positions with net buying pressure in a demand-based asset pricing framework. Zou and Derman (1999) ask, "How is an investor to know which strike and expiration provide the best value? What metric can option investors use to gauge their estimated excess return?" Existing approaches to option strategies do not offer answers to these questions, as they are not able to determine which options offer relatively more attractive investment opportunities or what expected returns we should expect from delta-hedged positions.

Our non-parametric option pricing model based on the prediction of breakeven volatility readily answers the above questions. The investor can make breakeven volatility predictions for several options and compare her predictions against implied volatilities. If the predicted BEV and IV for an option are different, the investor may initiate a position with positive expected returns. If implied volatility were lower than the predicted breakeven volatility, the market price of the option is too low, and the investor can "buy cheap" by taking a long delta-hedged position. Conversely, if implied volatility were higher than the predicted breakeven volatility, the market price of the option is too high, and the investor can "sell dear" by taking a short delta-hedged position.

Simulated Trading Strategy. Trading the difference between a volatility prediction and implied volatility is commonly called "volatility arbitrage" (Ammann and Herriger 2002). It is not arbitrage in the strict sense, but rather refers to whether the investor can make a sufficiently accurate forecast for future realized volatility to lock in a trading profit. To illustrate the economic value of our predictive model, we construct a trading strategy that exploits the difference between our predicted breakeven volatility values and implied volatility from market prices.

We keep the trading strategy simple because we want to test the accuracy of our statistical predictions. If we were to proceed with a more complex trading strategy—for example, adding in a layer of portfolio optimization—it becomes more difficult to evaluate the accuracy of our breakeven volatility forecasts. In a strategy that combines BEV forecasts with sophisticated portfolio optimization, we would be testing the transformation of our trading signals into portfolio positions as well as the accuracy of the BEV forecasts, resulting in a "joint hypothesis problem." By keeping the simulated trading strategy simple, we can test the accuracy of the BEV forecasts with limited confounding factors. Therefore, our empirical choices in the simulated trading strategy closely mirror those we made in the construction of our predictive model.

We use an initial training period from January 2013 to December 2014 to fit our model. We fix the model parameters and make predictions for breakeven volatilities throughout 2015, and we refit the model at the end of 2015 using all the data from 2013 through 2015. We then make new predictions using the updated model parameters. Each subsequent year, the model is refit using an expanding window. These predictions only use the information available at the time of the forecast, so they can be made in real-time. The final set of predictions goes from January 2015 to November 2020.

In the modeling section, we showed that our statistical model achieves high in-sample R-squared. However, good in-sample performance does not always imply generalizable results. This point is emphasized in return predictability studies where the goal is to produce the best predictions of future market returns (e.g., Welch and Goyal 2008; Campbell and Thompson 2008). Our expanding window predictions constitute out-of-sample (OOS) forecasts of breakeven volatility, and we evaluate the out-of-sample predictive performance of our model compared to the historical mean. We assess the volatility forecasts using the mean squared forecast error (MSFE), calculated for our prediction and the historical mean:

$$\hat{e}_{0,i} = BEV_i - \overline{BEV}_i \tag{10}$$

$$\hat{e}_{1,i} = BEV_i - \widehat{BEV}_i \tag{11}$$

where BEV_i is the actual breakeven volatility, $\widehat{BEV}_i$ is the predicted value of breakeven volatility, and $\overline{BEV}_i$ is the historical mean estimated using the same information set as $\widehat{BEV}_i$. The mean squared forecast error is calculated as the sum of squared forecast errors across the full set of N forecasts:

$$\widehat{MSFE}_k = \frac{1}{N}\sum_{i=1}^{N}\hat{e}_{k,i}^2,\ \ k = 0,\ 1 \tag{12}$$

Based on the Clark and West (2007) procedure, Dong et al. (2022) develop a test for a difference in the population of the MSFEs:

$$H_0 : MSFE_0 \leq MSFE_1,\ \ H_1 : MSFE_0 > MSFE_1 \tag{13}$$

Campbell and Thompson (2008) propose an alternative method to evaluate competing forecasts via an out-of-sample R-squared statistic:

$$R^2_{OS} = 1 - \frac{\widehat{MSFE}_1}{\widehat{MSFE}_0} \tag{14}$$

Dong et al. (2022) show that the Clark and West (2007) statistics to test Equation (13) is equivalent to testing whether the out-of-sample R-squared of Campbell and Thompson (2008) is positive or negative. We follow Dong et al. (2022) to evaluate the predictive power of our model. If $\widehat{BEV}_i$ captures more variation of actual breakeven volatility than the historical mean, the sum of squared errors in the numerator would be smaller than the denominator, indicating our predictive model forecasts BEV better than its historical average. In this case, the out-of-sample R-squared R^2_{OS} would be positive, corresponding to the alternative hypothesis in Equation (13). If $\widehat{BEV}_i$ captures less variation of BEV_i than $\overline{BEV}_i$, the out-of-sample R-squared would be negative, corresponding to the null hypothesis in Equation (13).

Every OOS prediction of breakeven volatility allows for a comparison with implied volatility. Starting in January 2015, each day, we scan through all available OTM options with time to expiration between five and 74 days, and we identify those with a predicted profit remaining >\$1.[17] Profit remaining is calculated as the difference between the breakeven dollar value and the midpoint between the bid and ask prices. For a particular option, if the breakeven price were lower than the mid-price, we sell one contract. If the breakeven price were higher than the mid-price, we purchase one contract. All trades are delta-hedged and held until five days to expiration, at which time the positions are closed out. Portfolio returns are calculated by summing up the daily profit and loss and dividing by the aggregate notional value of the contracts.[18]

Results

In the simulated trading strategy, the predictive model makes out-of-sample forecasts of breakeven volatility. We first assess these OOS forecasts before evaluating the trading strategy. The OOS forecasts provide a valuable perspective for understanding the validity of our predictive model. The out-of-sample R-squared is 0.71, indicating that our predictive model for breakeven volatility significantly outperforms the historical average estimate—the mean squared forecast error is 71% smaller for our forecasting model compared to the historical mean.

Table 6 presents a confusion matrix of the implied trading strategy from the breakeven volatility forecasts. The confusion matrix compares the number of long and short investment decisions implied by the model predictions with the decisions that an investor would make if she could perfectly observe the actual BEV values. For example, suppose an option with 30 days to expiration has an IV of 25% and BEV of 28%. With 30 days to go, we would not observe the true BEV value. Suppose our predicted BEV value is 24%. In this case, we would take a short position in the option, whereas if we could observe the true BEV value, we would have taken a long position. Since the options literature has documented a robust

Table 6. Confusion Matrix of BEV Predictions

	Predicted	
	Short	Long
Actual		
Short	3002	1467
Long	0	2314

This table shows the out-of-sample BEV predictions for those observations that show favorable trades. Each year, we retrain our forecasting model for breakeven volatility using as much historical data as possible, and we use this model to make predictions throughout the year. If the breakeven price were lower than the mid-price, we sell one contract. If the breakeven price were higher than the mid-price, we purchase one contract.

Table 7. Strategy Performance

A. Raw values

	BEV	Short Options	Short Puts	Short Calls	Buy-and-Hold S&P
Average returns	6.8%	4.2%	3.1%	1.1%	10.4%
Volatility	4.5%	4.3%	3.0%	1.6%	13.0%
Sharpe ratio	1.50	0.99	1.04	0.68	0.79
Max drawdown	−3.2%	−5.1%	−4.7%	−2.1%	−19.8%

B. Option strategies scaled to the same volatility as the S&P 500

	BEV	Short Options	Short Puts	Short Calls	Buy-and-Hold S&P
Average returns	19.6%	12.9%	13.5%	8.9%	10.4%
Volatility	13.0%	13.0%	13.0%	13.0%	13.0%
Sharpe ratio	1.50	0.99	1.04	0.68	0.79
Max drawdown	−9.3%	−15.7%	−20.3%	−17.7%	−19.8%

C. Including transaction costs

	One Tick	Two Ticks	Three Ticks
Average returns	4.4%	3.2%	2.1%
Volatility	4.5%	4.5%	4.5%
Sharpe ratio	0.98	0.72	0.46
Max drawdown	−4.9%	−7.2%	−11.5%

This table presents the performance statistics of our simulated trading strategy and several other investment strategies. "BEV" is a strategy that uses breakeven volatility forecasts to select and invest in options. "Short Options" is a strategy that takes short positions in all available options. "Short Puts" and "Short Calls" only invest in puts or calls separately. Panel A shows the statistics for the raw returns of the strategies. Panel B scales the options strategies to have the same volatility as that of the S&P 500. Panel C includes transaction costs for the BEV strategy (each tick is five cents, see text for details).

variance risk premium (Carr and Wu 2009; Bollerslev, Tauchen, and Zhou 2009), our trading strategy will have a default bias towards short positions. In this sense, the short positions are the "positive" class in the confusion matrix, and the long positions are the "negative" class.

Our out-of-sample predictions cover 6,783 observations that display an investing edge. Of these, 3,002 are true positives—the model short positions match the actual BEV implied short positions, and 2,314 are true negatives. Therefore, the accuracy of the model, measured as the total correct predictions divided by the total number of observations, is 78%. Precision measures the ratio of correctly predicted positive observations to the total predicted positive observations. This number comes out to be 100%—in this particular sample, every predicted short position with an expected profit remaining of more than $1 turned out to be correct. Recall, that the ratio of correctly predicted positive observations to all the actual positive observations, is 67%, indicating that the number of correctly predicted short positions is 67% of the total actual number of short positions an investor would take if she could perfectly observe breakeven volatilities. F1 score, the harmonic mean of precision and recall, provides a holistic evaluation of the predictive power of the model that balances false positives and false negatives. The F1 score is 0.80 in our case. By looking at the predictions from our regression model through the lens of classification, we gain an additional perspective on its performance.

We provide a summary of the performance of our BEV-based trading strategy in Table 7. We include three additional options strategies for comparison: A portfolio of delta-hedged short positions in all available options, a portfolio of delta-hedged short positions in all puts, and a portfolio of delta-hedged short positions in all calls. Finally, we include a buy-and-hold S&P 500 strategy. The BEV-based strategy earns an annual return of 6.8% with a volatility of 4.5%, achieving the highest Sharpe ratio and the highest ratio of expected to maximum drawdown among all the options strategies. A buy-and-hold strategy in the S&P 500 Index performed well in our sample period, earning 10.4% per year with 13% volatility. Its Sharpe ratio of 0.79 shows an attractive risk-return tradeoff between 2015 and 2020, higher than its historic longer-term value.

Due to the differences in volatility, it is difficult to compare average returns and maximum drawdowns across strategies. In Panel B of Table 7, we scale all the option strategies such that their annual volatilities are all equal to that of the S&P 500. Putting all the strategies on the same scale makes clear the distinct investment opportunities available to investors. At annual volatility of 13%, the BEV strategy earns 19.6% per year with a maximum drawdown of −9.3%, providing the highest average return and the lowest tail risk. In comparison, the other options strategies earn between 8.9% and 13.5%, with maximum drawdowns between −15.7 and −20.3%. The buy-and-hold S&P strategy has a maximum drawdown of −19.8%.

Margins and transaction costs are major factors that can impact the return of option strategies (Zhan et al. 2022). While our return calculation makes a conservative assumption about margin requirement—fully collateralized positions—we did not include transaction costs thus far. We explore the impact of transaction costs for our BEV strategy in Panel C of Table 7. The tick size for SPX options is five cents. For far out-of-the-money options, the bid-ask spread is often five cents. For at-the-money options, the typical bid-ask spread ranges from 20 to 30 cents. To incorporate costs, we assume we can trade the E-mini futures contract (which we use to hedge) one tick worse than the midpoint price, and we assume three values for trading SPX options: One tick, two ticks, and three ticks.[19] According to Muravyev and Pearson (2020), the average quoted bid-ask spread is 8.1 cents for stock index options, and the average effective spread is 6.2 cents. For sophisticated trades that take into account the expected future price movement of the options, the average effective spread is just 1.3 cents. The authors point out that proprietary traders or institutional investors who have algorithms or use brokerage firm execution algorithms pay closer to 1.3 cents than 6.2 cents. In light of the findings of Muravyev and Pearson (2020), our transaction cost assumptions are on the conservative side.

Panel C shows that if we trade one tick worse than the midpoint price for SPX options, the average return reduces to 4.4%, the Sharpe ratio reduces to 0.98, and the maximum drawdown rises to −4.9%. Incurring larger transaction costs further reduces the average return and Sharpe ratio, and increases the maximum drawdown. Unless the investor can limit her transaction costs, a strategy based on breakeven volatility will not offer attractive risk-adjusted returns.

Conclusion

In this paper, we build a non-parametric option pricing model whose output is the fair value of an option as measured by breakeven volatility, the value of implied volatility that sets the profit and loss of a delta-hedged option to zero. We compute breakeven volatilities for a large set of S&P 500 index options, and we use these historical values to construct a predictive model. A two-stage regression approach captures the majority of the variation in breakeven volatility, and the resulting predictions satisfy no-arbitrage conditions. The predictions from our model can be used to formulate a volatility arbitrage strategy that exploits the difference between implied volatility and the predicted breakeven volatility.

There are several interesting future directions to explore. A natural and immediate extension would be a careful comparison of the different ways of delta hedging. We used deltas associated with individual options, but alternative choices may include using the at-the-money delta or a delta calculated from a realized volatility forecast. A deeper analysis of the relationship between breakeven volatility, realized volatility, and realized skewness would contribute to the existing literature on modeling and forecasting higher moments of asset distributions. Garleanu, Pedersen, and Poteshman (2009) show that option prices are associated with demand pressure from option end-users, which may serve as an additional predictor variable in an option pricing model. An investigation of the connection between breakeven volatility and the Recovery Theorem of Ross (2015) could shed light on the viability of Ross's assumptions.

For a fixed time to expiration, the second derivative of option prices with respect to strike prices gives the risk-neutral distribution (Breeden and Litzenberger 1978). To the extent breakeven volatilities are different for options with different strike prices, the second derivative of the prices associated with BEVs provides another distribution. Because prices associated with breakeven volatilities do not admit any risk premia, this distribution does not embed the risk attitudes of market participants. A breakeven volatility surface that changes as a function of moneyness and time to expiration may be of interest for additional investigation.

Another interesting research direction is to expand beyond equity indices to individual equity options. It would be worth exploring whether one can use the same predictive model for BEVs in different sectors, and what sort of adjustments may be necessary. Researchers have documented distinct patterns for variance risk premium and option skewness for equity index options compared to individual equity options. Presumably, these discrepancies would translate to differences between breakeven volatility and implied volatility. One may also consider expanding the analysis to options traded on other asset classes, such as bond indices or commodity futures. Consistent empirical findings across asset classes can prevent overfitting and promote a common explanation for the behavior of option prices.

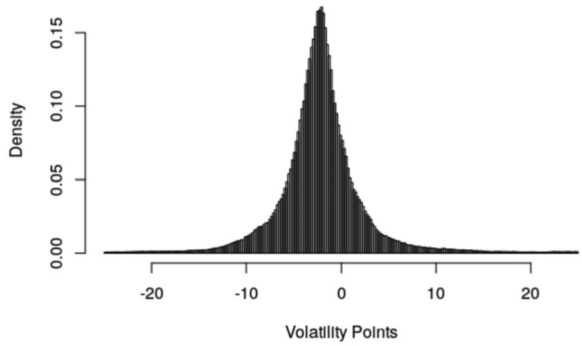
Appendix A

The table below outlines some differences between the closing price methodology for the SpiderRock platform and OptionMetrics. We refer to the IvyDB Europe Reference Manual v2.4 for OptionMetrics, and we refer to the SpiderRock technical notes for calculating volatility surfaces and forward rates. While both firms capture market data and use volatility surface calculations to report these values, they have different approaches. OptionMetrics has long been the standard for academic and institutional research, whereas SpiderRock is widely used by practitioners to support valuation in live trading environments and recently started to provide full datasets going back at 10 years.

Data Item	OptionMetrics Approach	SpiderRock Approach
Marking closing prices and characteristics	Stock and option close: settlement/last/bid/ask mid waterfall Closing prices use the last trade or closing prices for the close. Implied volatility calculations use underlying prices that are time synchronized with options bid, ask, or last quotes	Stock close: bid/ask mid logic Option: bid/ask/surface SpiderRock closing prices include the surface price are taken 1 min before the option market close. Implied volatility is synchronized with the underlying market
Closing theoretical surface price for options	Option prices used to calculate the implied volatility are selected as the first available price in the following: (1) settlement, (2) last trade, (3) bid-ask average, (4) bid, (5) ask	Call and put bid and ask markets are fitted directly as a component of the volatility surface shape fitting process. The instantaneous surface price is determined by aligning the surface shape with the underlying mid-price and ATM volatility in real-time.
Surface-fitting approach to determine the surface price	Vega-weighted kernel fit to options implied volatility space. Standardized options are interpolated from all options (all terms and strikes) for that day.	Price fits have an adaptive number of spline points. The result is a multi-point spline that describes a single volatility curve for both calls and puts for each expiry. Curves for normalized maturities are also generated as part of this process.
Alignment of call and put surfaces	Surface fits for calls and puts are fit independently. Calls and puts are interpolated with a weighted scheme that penalizes when option types are different.	Call and put prices are fit independently, and the price data are transformed into a single volatility curve for both calls and puts.
Forward calculation and dividend rate calibration	Forward fit to normalized maturities using a forward rate curve.	Forward prices are calculated using a continuous forward dividend rate curve with an adjustment based on aligning call and put volatility as part of the curve fitting process.
Calculation of time to expiration	Continuous calendar time	A hybrid time convention that takes both trading day and continuous calendar time into account

Appendix B

The figure shows a histogram of the difference between breakeven volatility and implied volatility for options on the S&P 500 Index. The sample is from January 2013 to November 2020.



Appendix C

This table presents a predictive model for the difference between breakeven volatility and implied volatility:

$$BEV_i - IV_i = \beta_0 + \sum_{k=m}^{M} X_{i,m}\beta_m + \epsilon_i$$

In this specification, rather than first predicting log(BEV) followed by a transformation, we directly try to predict whether an option is overvalued or undervalued relative to its breakeven volatility. We include the same predictor variables used in the first stage regression as shown in Table 4.

Left-Hand Variable: BEV-IV

	Estimate	SE	*t*-Stat
(Intercept)	−7.46	0.16	−47.0
VVIX	0.043	0.001	48.1
VIX.res	−0.188	0.006	−30.3
rt	−11.22	0.30	−37.3
lsk	24.99	0.24	106.0
lsk2	−8.82	0.14	−61.4
ATM.IV	0.53	0.01	88.4
ImpliedVol (coef × 10)	−6.03	0.04	−156.8
ImpliedVol2 (coef × 100)	1.22	0.03	45.0
RR	−0.75	0.10	−7.6
RR2	36.38	0.34	105.9
RR3	16.68	0.59	28.1
Vega	2.48	0.04	64.8
vega2 (coef × 10)	−1.03	0.05	−22.4
gamma	−137.00	14.23	−9.6
vrt	15.81	2.01	7.9
vvt2	−52.55	0.88	−60.0
rv.res (coef × 10)	2.11	0.01	146.0
R-squared		0.13	
Adjusted R-squared		0.13	

Editor's Note

Submitted 2 February 2022

Accepted 7 July 2022 by William N. Goetzmann

Notes

1. Traditional option pricing models typically try to match the market price of options, whereas our pricing model provides the fair value. In this sense, traditional models are positive models that capture how the world is, whereas our model is normative.

2. Parametric models, such as Black-Scholes can be constrained in matching certain empirical patterns. Stochastic volatility models may have difficulty eliminating pricing errors in short-term option prices because the distribution of the underlying does not have enough kurtosis. Jump models may have issues with longer-term options because they revert too quickly to the Black-Scholes prices as the time to expiration increases. Nonparametric models can be more flexible in capturing the features of the data, which must be pre-specified and built into parametric models.

3. For stock indices, deviation from put-call parity is typically small. For individual stocks that are hard-to-borrow, deviation from put-call parity could be much larger.

4. This expression is approximate. To be precise, the summation is missing a borrowing or lending term equal to $(\Delta_{t-1}S_{t-1} - c_{t-1})$. At the daily frequency, this term is very close to zero, especially in a low-interest environment that our data span.

5. The iterative approach we take was used for a different purpose by Dumas et al. (1998), who try to determine whether the delta-hedging component matched the changes in option prices. They show that the Black-Scholes model provides the best approximation for option price changes.

6. Our calculation of breakeven volatility requires knowing the future path of the option, so it cannot be implemented in real time. To compute BEV in a real-time implementation, we would require a predictive model that uses currently available information to forecast the eventual BEV values.

7. There are a number of alternatives ways of hedging. One could choose to hedge at market open rather than at market close, use delta calculated from the at-the-money volatility or forecasted volatility, hedge only when the net delta exceeds a certain threshold, or adjust the hedge ratio depending on whether the order flow is believed to be informed. Different hedging methods correspond to different objectives, and there is no single best method. On a practical note, the BEV calculation should match the specific trading desks' preferred hedging policy.

8. More frequent delta hedging (i.e., intraday) can alleviate the issue of large changes in gamma in the volatility calculations near maturity, especially for ATM options.

9. Extending our calculation to the expiration date will not significantly change our results. We do find that numerical computation of breakeven volatilities in the last several days just before expiration to be more difficult, as more volatility calculations do not converge in the days immediately before expiration. More frequent delta hedging can improve this convergence issue.

10. $Normalized\ Strike = \frac{\log\left(\frac{Se^{r(T-t)}}{K}\right)}{ATM.IV\sqrt{T-t}}$ for underlying price S, strike price K, the at-the-money implied volatility $ATM.IV$, and time to expiration $T-t$. The annualized interest rate r is set to 0.5%.

11. d_1 in the Black-Scholes model is given by the following expression:

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right]$$

where S_t is the price of the underlying, K is the strike price, r is the interest rate, σ is the volatility of the underlying, and $T-t$ is the time to expiration.

12. The exponential function, $e^{(\cdot)}$, is convex. Jensen's inequality states that for a convex function f, $f(E[X]) \leq E[f(X)]$.

13. We compare our supervised approach with the parametric methods of Duan (1983) and we find that our approach achieves the lowest mean squared error.

14. With 396,899 observations and 17 predictors, the adjusted R-squared is 1.00004 times larger compared to the unadjusted R-squared.

15. We thank an anonymous referee for this suggestion.

16. If Equation (9) were violated, a costless arbitrage profit may be earned by writing a put with strike price K_2, buying $\frac{K_3-K_2}{K_3-K_1}$ units of the put with strike price K_1, and buying $\frac{K_2-K_1}{K_3-K_1}$ units of the put with strike price K_3.

17. We also explore thresholds of \$0.50, \$2, and \$3, and we find our results are qualitatively unchanged. A higher trading threshold has two offsetting effects: We have more conviction on each trade, but we may miss some profitable opportunities that do not exceed the threshold. In our sample, these two effects apparently have similar magnitudes such that changing the threshold has only a marginal effect on the strategy performance.

18. Our return calculation is on the conservative side. By taking the notional value as the denominator, we have made the implicit assumption that the option positions are fully collateralized with 100% margin requirement. In practice, margin requirements typically range between 5 and 30% for puts and calls on the S&P 500 Index.

19. Private correspondence with option traders indicates that unless one wants immediate execution, transacting one tick from the midpoint price is achievable with very high probability.

References

Ait-Sahalia, Y., and A. W. Lo. 1998. "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices." *The Journal of Finance* 53 (2): 499–547. doi:10.1111/0022-1082.215228.

Ammann, M., and S. Herriger. 2002. "Relative Implied-Volatility Arbitrage with Index Options." *Financial Analysts Journal* 58 (6): 42–55. doi:10.2469/faj.v58.n6.2485.

Bakshi, G., C. Cao, and Z. Chen. 1997. "Empirical Performance of Alternative Option Pricing Models." *The Journal of Finance* 52 (5): 2003–2049. doi:10.1111/j.1540-6261.1997.tb02749.x.

Bakshi, G., and N. Kapadia. 2003. "Delta-Hedged Gains and the Negative Market Volatility Risk Premium." *Review of Financial Studies* 16 (2): 527–566. doi:10.1093/rfs/hhg002.

Bali, T. G., H. Beckmeyer, M. Moerke, and F. Weigert. 2021. "Option Return Predictability with Machine Learning and Big Data." Available at SSRN 3895984.

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (3): 637–654. doi:10.1086/260062.

Bollen, N. P., and R. E. Whaley. 2004. "Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?" *The Journal of Finance* 59 (2): 711–753. doi:10.1111/j.1540-6261.2004.00647.x.

Bollerslev, T., G. Tauchen, and H. Zhou. 2009. "Expected Stock Returns and Variance Risk Premia." *Review of Financial Studies* 22 (11): 4463–4492. doi:10.1093/rfs/hhp008.

Breeden, D. T., and R. H. Litzenberger. 1978. "Prices of State-Contingent Claims Implicit in Option Prices." *The Journal of Business* 51 (4): 621–651. doi:10.1086/296025.

Campbell, J. Y., and S. B. Thompson. 2008. "Predicting Excess Stock Returns out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies* 21 (4): 1509–1531. doi:10.1093/rfs/hhm055.

Carr, P., and L. Wu. 2009. "Variance Risk Premiums." *Review of Financial Studies* 22 (3): 1311–1341. doi:10.1093/rfs/hhn038.

Cheney, E. W., and D. R. Kincaid. 2009. *Linear Algebra: Theory and Applications*. Jones & Bartlett Learning.

Christoffersen, P., R. Goyenko, K. Jacobs, and M. Karoui. 2018. "Illiquidity Premia in the Equity Options Market." *The Review of Financial Studies* 31 (3): 811–851. doi:10.1093/rfs/hhx113.

Clark, T. E., and K. D. West. 2007. "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics* 138 (1): 291–311. doi:10.1016/j.jeconom.2006.05.023.

Coval, J. D., and T. Shumway. 2001. "Expected Option Returns." *The Journal of Finance* 56 (3): 983–1009. doi:10.1111/0022-1082.00352.

Cox, J. C., R. A. Ross, and M. Rubinstein. 1979. "Option Pricing: A Simplified Approach." *Journal of Financial Economics* 7 (3): 229–263. doi:10.1016/0304-405X(79)90015-1.

Dong, X., Y. Li, D. E. Rapach, and G. Zhou. 2022. "Anomalies and the Expected Market Returns." *The Journal of Finance* 77 (1): 639–681. doi:10.1111/jofi.13099.

Duan, N. 1983. "Smearing Estimate: A Nonparametric Retransformation Method." *Journal of the American Statistical Association* 78 (383): 605–610. doi:10.1080/01621459.1983.10478017.

Dumas, B., J. Fleming, and R. E. Whaley. 1998. "Implied Volatility Functions: Empirical Tests." *Journal of Finance* 53 (6): 2059–2106. doi:10.1111/0022-1082.00083.

Dupire, B. 2006. "Fair Skew: Break-Even Volatility Surface." *Technical Report*, Bloomberg L.P.

Figlewski, S., and F. Malik. 2014. "Options on Leveraged ETFs: A Window on Investor Heterogeneity." Available at SSRN 2477004.

Garleanu, N., L. H. Pedersen, and A. M. Poteshman. 2009. "Pricing Options in an Extended Black-Scholes Economy with Illiquidity: Theory and Empirical Evidence." *Review of Financial Studies* 22 (10): 4259–4299.

Goyenko, R., and C. Zhang. 2020. "The Joint Cross Section of Option and Stock Returns Predictability with Big Data and Machine Learning." Available at SSRN.

Heston, S. L. 1993. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Current Options." *Review of Financial Studies* 6 (2): 327–343. doi:10.1093/rfs/6.2.327.

Hull, B., and X. Qiao. 2017. "A Practitioner's Defense of Return Predictability." *The Journal of Portfolio Management* 43 (3): 60–76. doi:10.3905/jpm.2017.43.3.060.

Hutchinson, J. M., A. W. Lo, and T. Poggio. 1994. "A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks." *The Journal of Finance* 49 (3): 851–889. doi:10.1111/j.1540-6261.1994.tb00081.x.

Liu, S., C. W. Oosterlee, and S. M. Bohte. 2019. "Pricing Options and Computing Implied Volatilities Using Neural Networks." *Risks* 7 (1): 16. doi:10.3390/risks7010016.

Malz, A. M. 2014. "A Simple and Reliable Way to Compute Option-Based Risk-Neutral Distributions." *FRB of New York Staff Report 677*.

Manzo, G., and X. Qiao. 2021. "Deep Learning Credit Risk Modeling." *The Journal of Fixed Income* 31 (2): 101–127. doi:10.3905/jfi.2021.1.121.

Merton, R. C. 1973. "Theory of Rational Option Pricing." *The Bell Journal of Economics and Management Science* 4 (1): 141–183. doi:10.2307/3003143.

Mitoulis, N. 2019. "Breakeven Volatility." *Technical Report*, University of Cape Town.

Muravyev, D., and N. D. Pearson. 2020. "Option Trading Costs Are Lower than You Think." *The Review of Financial Studies* 33 (11): 4973–5014. doi:10.1093/rfs/hhaa010.

Neuberger, A. 2012. "Realized Skewness." *Review of Financial Studies* 25 (11): 3423–3455. doi:10.1093/rfs/hhs101.

Orosi, G. 2015. "Estimating Option-Implied Risk-Neutral Densities: A Novel Parametric Approach." *The Journal of Derivatives* 23 (1): 41–61. doi:10.3905/jod.2015.23.1.041.

Ross, S. 2015. "The Recovery Theorem." *The Journal of Finance* 70 (2): 615–648. doi:10.1111/jofi.12092.

Shimko, D. 1993. "The Bounds of Probability." *Risk* 6 (4): 33–37.

Stutzer, M. 1996. "A Simple Nonparametric Approach to Derivative Security Valuation." *The Journal of Finance* 51 (5): 1633–1652. doi:10.1111/j.1540-6261.1996.tb05220.x.

Taylor, J. M. 1986. "The Retransformed Mean after a Fitted Power Transformation." *Journal of the American Statistical Association* 81 (393): 114–118. doi:10.1080/01621459.1986.10478246.

Welch, I., and A. Goyal. 2008. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies* 21 (4): 1455–1508. doi:10.1093/rfs/hhm014.

Zhan, X., B. Han, J. Cao, and Q. Tong. 2022. "Option Return Predictability." *The Review of Financial Studies* 35 (3): 1394–1442. doi:10.1093/rfs/hhab067.

Zou, J., and E. Derman. 1999. *Strike-Adjusted Spread: A New Metric for Estimating the Value of Equity Options*. SSRN. https://ssrn.com/abstract=170629