CENTRE FOR RESPONSIBLE BANKING & FINANCE
CRBF

University of St Andrews
School of Management

WORKING PAPERS IN RESPONSIBLE BANKING & FINANCE

Cross-Section of Option Returns and the Volatility Risk Premium

By *Simon Fritzsch, Felix Irresberger, Gregor Weiß*

Abstract: This paper presents a robust new finding that delta-hedged equity option returns include a volatility risk premium. To separate volatility risk premia from confounding effects, we estimate conditional quantile curves of implied volatilities using machine learning. We find that a zero-cost trading strategy that is long (short) in the portfolio with low (high) implied volatility conditional on the options' moneyness and realized volatility produces an economically and statistically significant average monthly return. Using conditional quantile curves not only helps in distinguishing volatility risk premia from other effects, most notably realized volatility, it also leads to returns that are higher than those reported in previous work on similar volatility strategies.

WP Nº 21-012

3rd Quarter 2021

University of St Andrews
Scotland's first university

600 YEARS
1413 – 2013

Cross-Section of Option Returns And The Volatility Risk Premium *

Simon Fritzsch[†]
Universität Leipzig

Felix Irresberger[‡]
Durham University

Gregor Weiß[§]
Universität Leipzig

June 10, 2021

ABSTRACT

This paper presents a robust new finding that delta-hedged equity option returns include a volatility risk premium. To separate volatility risk premia from confounding effects, we estimate conditional quantile curves of implied volatilities using machine learning. We find that a zero-cost trading strategy that is long (short) in the portfolio with low (high) implied volatility conditional on the options' moneyness and realized volatility produces an economically and statistically significant average monthly return. Using conditional quantile curves not only helps in distinguishing volatility risk premia from other effects, most notably realized volatility, it also leads to returns that are higher than those reported in previous work on similar volatility strategies.

Keywords: Option returns, Implied volatility, Machine learning, Realized volatility, Volatility Risk Premium, Volatility Mispricing.
JEL Classification Numbers: G11,G13,C14,C58,C45.

*Simon Fritzsch gratefully acknowledges financial support through a PhD scholarship by LBBW Asset Management, Stuttgart.

[†]Universität Leipzig, Wirtschaftswissenschaftliche Fakultät, e-mail: *fritzsch@wifa.uni-leipzig.de.*

[‡]Durham University Business School, e-mail: *felix.irresberger@durham.ac.uk.*

[§]Corresponding author: Universität Leipzig, Wirtschaftswissenschaftliche Fakultät, e-mail: *gregor.weiss@uni-leipzig.de.*

1 Introduction

Volatility is the single most important characteristic of a stock driving the prices of corresponding option contracts. Returns on stock options should thus carry a risk premium for changes in volatility. Likewise, any misestimation of an underlying stock's volatility and its dynamics should lead to a mispricing of options which traders can exploit. Yet, despite its ubiquity in option pricing models, the role volatility (mispricing) and volatility risk play for the cross-section of option returns remains unclear: while some studies have found no evidence for the existence of a volatility risk premium (see, e.g., Carr and Wu, 2009; Driessen et al., 2009), others have shown that volatility mispricing as well as (realized and idiosyncratic) volatility significantly affect the cross-section of option returns (see, e.g., Bollerslev et al., 2009; Goyal and Saretto, 2009; Cao and Han, 2013; Cao et al., 2019; Hu and Jacobs, 2020).

In this paper, we empirically test for the existence of a volatility risk premium (VRP) in the cross-section of option returns. We start our analysis by first documenting potential biases in analyses of the volatility-return relation that arise when relying on standard methods in asset pricing: portfolio sorts and cross-sectional regressions. As an alternative, we propose to use nonparametric methods from the field of machine learning to estimate conditional quantile curves of implied stock option volatilities. We condition on a number of characteristics that would otherwise cloud the effect of implied volatility on option returns. Most importantly, we control for the stock's realized volatility and option moneyness. Doing so helps us to carve out the volatility risk premium in the cross-section of option returns.

Using the cross-section of option returns for US equities between January 1996 and June 2019, we find that call and put option portfolio returns exhibit a strong relation with the volatility risk premium. We sort options on their implied volatility *conditional* on their realized volatility. This yields portfolios with increasing deviations between realized and implied volatilities with average levels of realized volatility remaining constant. We use this to proxy for the volatility risk premium. A strategy that is long (short) in high (low) deviations between realized and implied volatilities yields returns that are both economically and statistically significant. This result holds for call and put delta-hedged and raw option returns for both at the money (ATM) options and options of arbitrary moneyness. For example, average monthly delta-hedged returns of 1-month ATM options are 2.0 % for call and 1.7 % for put contracts with (monthly) Sharpe ratios of 0.842 and 0.796, respectively.

By sorting options on their implied volatility conditional on realized volatility *and* option moneyness, we can easily extend our trading strategy to options of arbitrary moneyness while eliminating potential biases arising from systematic differences in realized volatility or option moneyness (and thus option liquidity). Again, this yields delta-hedged and raw option returns that are highly economically and statistically significant. For example, average monthly delta-hedged returns from a long-short trading strategy based on 1-month options of arbitrary moneyness are 2.4 % for call and 2.5 % for put contracts with (monthly) Sharpe ratios of 0.816 and 0.844, respectively. Our results are robust to controlling for further moments of the underlyings' return distribution, alternative estimators of conditional quantiles, moderate transaction costs (delta-hedged returns) and the inclusion of options on dividend-paying stocks.

Why do we find such strong evidence for the existence of a VRP when results of previous empirical studies have been ambiguous at best? One possible answer lies in our proposed use of nonparametric methods to form factor-mimicking portfolios based on characteristics conditional on a set of control variables. The standard technique in empirical finance for this purpose has been the sorting of portfolios on certain characteristics of assets. It is frequently used to test the assumption of pricing models that expected asset returns are a monotonic function in one or more idiosyncratic characteristics. This common practice of forming uni- or multivariate fractiles dates back to seminal papers on the cross-section of equity returns which, among others, include the works of Basu (1977); Banz (1981); De Bondt and Thaler (1985); Jegadeesh (1990); Fama and French (1992), and Jegadeesh and Titman (1993). Since then, portfolio sorting has been a methodological mainstay in empirical asset pricing, because it does not require the assumption of a linear relation between expected returns and characteristics, and because differences in the returns on the top and bottom fractile portfolios are easily interpreted as the profits from an implementable trading strategy. As intuitive as it may be, however, portfolio sorting does not come without shortcomings. While univariate portfolio sorts on one characteristic do not allow the economist to control for other asset characteristics, multivariate (conditional) sorts quickly become unfeasible for more than two characteristic-based factors due to the curse of dimensionality.

Our approach, in contrast, uses nonparametric methods from the field of machine learning to estimate the conditional quantile curves of implied volatilities while at the same time controlling for several characteristics. Thus it possesses several appealing features that should make it favorable to standard portfolio sorts and non-/semiparametric regression methods alike. First, using machine learning algorithms to estimate quantile curves allows for the data-efficient nonparametric modeling of the multivariate distribution of asset returns and characteristics. In contrast to standard (conditional) portfolio sorts, our method should alleviate at least in part the concern of "empty portfolios" which eventually arises when sorting on too many characteristics (see, e.g., Goyal, 2012).

Second, as a remedy to the "empty portfolio" problem, many researchers have additionally performed multivariate regressions to test whether a certain characteristic is priced. While these cross-sectional regressions allow the inclusion of a large number of covariates, they also suffer from two drawbacks that make them less appealing in our setting. Standard as well as semiparametric regression methods (see, e.g., Connor and Linton, 2007; Connor et al., 2012; Cattaneo et al., 2020) assume additive separability between the explanatory variables in asset pricing models (in addition to a linear relation between characteristics and returns). Employing such standard models leads to a severe bias in the measurement of the VRP due to the nonlinear nature of the relation between implied and realized volatility. Moreover, results from cross-sectional regressions only yield information on long-short strategies that involve trading in *all securities* with potentially highly varying portfolio weights.

Our proposed nonparametric approach circumvents both these problems: we make no assumption on the functional form of the relation between implied and realized volatility, and our approach yields a trading strategy that can easily be implemented.

Our paper is related to an increasing number of empirical studies on the relation between option returns and characteristics of the underlying stocks.[1] Coval and Shumway (2001) were among the first to look at the cross-section of expected option returns. Studying index options, they find that systematic stochastic volatility is priced in option returns. In a related study, Driessen et al. (2009) show that correlation risk is priced in both index and individual options but find no evidence for the existence of a VRP. Conversely, using model-free estimates of implied volatilities, Bollerslev et al. (2009) show in their study that stock returns include a VRP. Finally, Huang et al. (2019) study the pricing of volatility of volatility risk in index options. All of the studies, however do not test for the existence of VRP in expected option returns and usually concentrate on index options rather than individual options.

More recent studies have concentrated on the effects of volatility, volatility risk, and volatility mispricing on the cross-section of expected option returns. In one of the first studies in this field of research, Goyal and Saretto (2009) show that large differences between realized and implied volatilities for at-the-money options are associated with economically and statistically significant monthly returns. Their use of linear differences to proxy for potential volatility mispricing, however, ultimately leads to a portfolio strategy that also (at least partially) invests in realized volatility. Our approach to measure the volatility risk premium by the use of conditional quantile curves builds on their study. After controlling for realized volatility and allowing for arbitrary moneyness, our results confirm the initial findings by Goyal and Saretto (2009) and make an even stronger case for the existence of a VRP. This is important and reassuring at the same time, as more recent studies have shown

[1] Recent studies on the pricing of stock and option characteristics in the cross-section of expected option returns include, but are not limited to, the studies by Baele et al. (2018); Cao et al. (2019); Andreou and Ghysels (2020); Eisdorfer et al. (2020); Cao et al. (2021).

that idiosyncratic (see Cao and Han, 2013) and realized volatility (see Hu and Jacobs, 2020) by themselves drive expected option returns, questioning previous findings on the effect of volatility mispricing on option returns.

Methodologically, our paper is related to a small but growing number of papers that aim to improve standard methods in empirical asset pricing. For example, Patton and Timmermann (2010) were among the first to point out the shortcomings of portfolio sorts and standard tests of monotonicity in asset pricing. Similarly, Connor and Linton (2007), Connor et al. (2012), and Cattaneo et al. (2020) propose semi- and nonparametric models as alternatives for portfolio sorts and cross-sectional regressions. In contrast to our study, however, their models usually concentrate on nonlinear relations between returns and characteristics, and not between covariates. Moreover, none of these studies look at the cross-section of option returns. We complement this field of research by proposing the use of conditional quantile curves as an alternative to traditional portfolio sorts and applying it for the first time to expected option returns. Finally, our paper is also related to a growing number of studies that propose the use of machine learning algorithms in empirical asset pricing. For example, Moritz and Zimmermann (2016) use tree-based conditional portfolio sorts and model-averaging to identify the most relevant factors of the famous "factor zoo" (cf. Cochrane, 2011) that drive stock returns, while Gu et al. (2020) employ trees and neural networks to forecast returns. Our work complements these studies by proposing the use of data-efficient machine learning algorithms to form conditional portfolio sorts in high dimensions.

The rest of the paper is organized as follows. The next section 2 discusses the measurement of the volatility risk premiums in option returns as well as our methodology. Section 3 presents our empirical study. We discuss robustness checks in Section 4. Section 5 concludes.

2 Capturing the volatility risk premium

2.1 Volatility risk premium and volatility mispricing

We start our analysis by revisiting common definitions for the volatility risk premia of individual stocks from the related literature. For example, Cao and Han (2013) define the volatility risk premium of stock i in time period t as

$$VRP_{i,t} = RV_{i,t} - IV_{i,t} \tag{1}$$

where $RV_{i,t}$ is realized return volatility and $IV_{i,t}$ is the implied volatility of the stock extracted from corresponding options (see also Jiang and Tian, 2005; Bollerslev et al., 2009; Carr and Wu, 2009; Driessen et al., 2009). Similarly, Goyal and Saretto (2009) consider the log difference

$$VMP_{i,t} = \log RV_{i,t} - \log IV_{i,t} \tag{2}$$

between realized (historical or current) and implied volatility and interpret large values of $VMP_{i,t}$ as indicative of volatility mispricing.

From the definitions in Equations (1) and (2) it becomes clear that tests of the existence of a VRP in option returns critically depend on whether one is able to control for the level of realized volatility.[2] To this end, previous studies have traditionally relied on (conditional) portfolio sorts and cross-sectional regressions. However, controlling for additional asset characteristics (such as realized volatility) via conditional portfolio sorts quickly becomes infeasible due to the curse of dimensionality (Stone, 1980).[3] Beyond that, portfolio sorts exhibit further shortcomings, some of which we want to illustrate in the following Sec-

[2] The findings of Hu and Jacobs (2020) show that option raw returns are significantly affected by realized volatility.

[3] For example, while sorting on one characteristic and controlling for another via a double-sort might still be practicable, controlling for two or more covariates via triple-, quadruple-sorts etc. is almost always not possible.

tion 2.2. As a remedy, we advocate for replacing conditional portfolio sorts with conditional quantiles, of which we discuss the details in Section 2.3.

2.2 Replacing portfolio sorts by conditional quantiles

To illustrate the potential weaknesses of conditional portfolio sorts, we simulate 200 observations (x_i, y_i) according to $Y = \frac{1}{2} \cdot X + \frac{1}{10} \cdot X \cdot \epsilon$, where ϵ denotes the standard normal distribution and X follows a $Beta(5,5)$ distribution with X and ϵ being statistically independent from each other. That is, we assume a linear relation between X and Y with heteroskedastic errors.

Figure 1 illustrates quintile portfolios from a conditional double-sort of the data (first on x, than on y).

– Insert Figure 1 about here. –

Therefore, we first sort the observations into five portfolios according to their x-values. To obtain portfolios of equal size, this is done by assigning the observations with x-values below the (unconditional) empirical 20 % quantile of all x_i's $i = 1, \ldots, 200$ into portfolio 1, observations between the 20 % and 40 % quantile into portfolio 2, etc. In each of these five portfolios, we then sort on y according to the (unconditional) empirical 20 %, 40 %, etc. quantile of all y-values *within* portfolio 1. This is subsequently repeated for portfolios 2 to 5. Based on this double-sort, we next form long and short portfolios of observations with low and high y-values, respectively, while at the same time controlling for x. Thus, in each of the five portfolios we obtain by sorting on x, we choose the observations corresponding to the y-values in the lower quintile for the long and in the upper quintile for the short portfolio (see the shaded areas in Figure 1).

Figure 2 illustrates that by combining the 20 % quantiles in each of the five portfolios we obtain a step function (black dashed line) which can be understood as an approximation to the true conditional 20 % quantile curve of Y *given* X (black solid line). The same is true for the conditional 80 % quantile curves (red lines). It is obvious from Figure 2 that portfolio sorts provide a quite coarse approximation to the true conditional quantile curves. As a consequence, the long and short portfolios derived from the double-sort exhibit systematic differences in x, see Figure 3. The black dashes on the x-axis correspond to observations in the long portfolio while the red dashes correspond to observations in the short portfolio. If controlling for x had been successful, the black and red dashes would be mixed randomly along the x-axis. However, there are various clusters of red and black dashes with a clear trend of observations in the long portfolio tending to lower and observations in the short portfolio tending to higher x-values. That is, although long and short portfolios were build on a double-sort to control for x, the results of a long (short) strategy that is low (high) in y might still be biased by systematic differences in x.

– Insert Figures 2 and 3 about here. –

To mitigate this bias, we could of course simply increase the number of portfolios and sort the x- and y-values into, say, 10 portfolios each yielding 100 double-sorted portfolios. However, we would still try to approximate the true conditional quantile curves (which are linear in our example) with piecewise constant step functions, which is far from optimal. In addition, if we wanted to control for an additional covariate z, we would be required to do a triple-sort leading to 1000 portfolios. However, this is infeasible based on only 200 observations as this would lead to a large number of empty portfolios. Consequently, we would be forced to reduce the number of portfolios to, say, 5 for each covariate again. Nevertheless, even if we accepted this coarse approximation, another difficulty is posed by the fact that no fragments of observations can enter into the quintile portfolios. By sorting on x and z we get 25 portfolios containing $200/25 = 8$ values each. When finally sorting on y, we would end up with 1.4 observations in the quintile portfolios which is again not feasible. As a consequence, if we decided to consider only one (out of eight) observations in each of the 25 portfolios, we would build long and short portfolios based on 12.5 % of all observations. If we opted for two observations, the quintile portfolios would be formed based on 25 % of observations. That is, conditional portfolio sorts are inflexible with regard to the percentage of observations that enter into the extreme portfolios, especially when the total number of observations is low.

Based on the true conditional quantile functions (as illustrated in Figure 2) we are able to form long (short) portfolios with low (high) y-values while at the same time *perfectly* controlling for x. Additionally, we would be very flexible with regard to the percentage of observations entering into the long and short portfolio. Of course, in practice, we do not know the true conditional quantile curves. Instead we have to estimate them based on the data at hand. However, conditional portfolio sorts do not use the data efficiently when deriving conditional quantile curves. This is highlighted in Figure 2 illustrating that the true quantile curves are approximated with piecewise constant step functions requiring a large number of portfolios to achieve an acceptable fit. As a consequence, one can usually control for only one or at most two characteristics.

As an alternative to portfolio sorts, there exist various methods specifically designed for the estimation of conditional quantiles providing a more data-efficient way of estimation without imposing restrictions on the functional form of the quantile functions. In addition to the advantages mentioned above, this allows for including more control variables.[4] As the objective of this paper is to advocate for replacing characteristic-sorted portfolios with portfolios sorted based on conditional quantile estimates and not to promote a specific estimator, we only give a short introduction into the topic in the next Section.

[4]Of course, due to the curse of dimensionality the number of variables one can efficiently control for with our nonparametric approach will still be limited to a low single figure in most applications. However, our approach can be seen as a way of extending the applicability of portfolio sorts to control for more covariates. For example in Section 4.1, conditionally sorting on options' implied volatility as well as realized volatility, moneyness, and skewness of the underlyings' return distribution would require the usage of quadruple sorts which is infeasible for the major part of our option sample in the empirical study. However, we can easily consider these control variables within our proposed approach.

2.3 Estimation of conditional quantiles

It is standard practice to (linearly) approximate the conditional *mean* function $x \mapsto E(Y|X = x)$. This is done by minimizing the squared errors. By instead minimizing the absolute errors one can derive the conditional *median* function, i.e., the conditional 50 % quantile function. This result generalizes naturally to conditional quantiles at arbitrary confidence levels. The conditional α quantile function is obtained by minimizing the so-called check-error of the residuals, where the check-function ρ_α is given as

$$\rho_\alpha(z) := \begin{cases} -(1-\alpha)z & \text{for } z \leq 0, \\ \alpha z & \text{for } z > 0. \end{cases} \tag{3}$$

That is, depending on the confidence level α, the residuals (deviations of observations from the estimated conditional quantiles) enter into the error term that has to be minimized asymmetrically with weights $1 - \alpha$ and α, respectively. Building on this result Koenker and Bassett (1978) introduce linear conditional quantile estimators. Since the appearance of this seminal paper various other estimators have been proposed. In particular, nonparametric estimators appear promising as they do not require to make any assumptions about the functional form of the quantile curves, see, e.g., Kraus and Czado (2017) and the references therein.[5]

The problem of estimating conditional quantiles nonparametrically has been addressed with different techniques. For example Kraus and Czado (2017) propose an estimator based on likelihood optimal D-vine copulas, in the following referred to as the *copula estimator*. The estimator models multivariate dependencies based on so-called pair-copula constructions, see the original paper for more details.

Starting with the k-nearest neighbor (kNN) estimator due to Bhattacharya and Gan-

[5]Results from our empirical study highlight the necessity to account in particular for non-linear dependencies between implied and realized volatility as well as heteroskedasticity. Note that while in general we speak of conditional quantiles, in the case of conditioning on only one variable, we get conditional quantile *curves*. Therefore, we will use these terms interchangeably.

gopadhyay (1990) conditional quantile estimation has also been addressed within the realm of (unsupervised) machine learning. More recently, Charlier et al. (2015b) introduced an estimator that is derived from the concept of optimal quantization. Loosely speaking, this algorithm efficiently uses the data at hand by identifying clusters of observations of covariates via unsupervised machine learning and deriving empirical quantiles of the response variable within the clusters, see Charlier et al. (2015b,a) for details. In the following we will refer to this estimator as the *quantization estimator*.

Building on the quantization estimator we derive a new estimator by employing a machine learning technique called *leveraging*. Leveraging is an ensemble technique very similar to boosting[6] which according to Meir and Rätsch (2003) "combine[s] simple 'rules' to form an ensemble such that the performance of the single ensemble member is improved". For this purpose we define our *leveraging estimator* in an iterative manner such that in each iteration step we give more weight to those observations for which the latest conditional quantile estimates produce a higher estimation error and less weight to those observations associated with a lower estimation error. Errors are calculated based on the check-function, see Equation (3). More details on the construction of the estimator along with an extensive simulation study will be made available in a separate paper.

We use the leveraging estimator for our main analysis but also include results for the quantization and copula estimator in the robustness checks in Section 4.3.[7] Results from the leveraging estimator are based on our own implementation while for the quantization and the copula estimator we rely on the QuantifQuantile and vinereg (with nonparametric pair copulas) R-package, respectively.

[6]As the concepts of boosting and leveraging are very similar both terms are often used interchangeably in the literature. However, we follow Duffy and Helmbold (2002) and restrict usage of the term boosting to algorithms proved to fulfill a so-called *Probably Approximately Correct (PAC)* learning - property and use the term leveraging for all other related ensemble learning techniques.

[7]Although there is a variety of conditional quantile estimators, the number of methods that can account for two or more covariates is substantially lower. In addition, Figure 6 indicates that conditional quantile curves in the option sample of our empirical study are best tackled by nonparametric methods. We also considered the kernel estimator due to Li and Racine (2008) as well as the kNN estimator but excluded them from the empirical study because of unsatisfactory results in pretests. All computations were performed on the Big-Data-Cluster Galaxy provided by the University Computing Center at Leipzig University.

3 Empirical study

In Section 2.2 we discussed some of the advantages of conditional quantile based portfolio sorts over simple (conditional) portfolio sorts. In this empirical study we build on this approach to study the VRP in the cross-section of options while controlling for the level of realized volatility. We compare our results to those obtained by sorting on the log-difference of RV and IV. In the construction of our data sample as well as in the corresponding trading strategies we closely follow Goyal and Saretto (2009).

3.1 Sample construction

The sample period is from January 1996 to June 2019. Data on US equities (including prices, closing bid and ask quotes, and returns) are retrieved from the Center for Research in Security Prices (CRSP). Option data are obtained from the OptionMetrics IvyDB US database. The data include information on the entire US equity option market (American options) covering in particular closing bid and ask quotes along with option implied volatilities (IV) and greeks (delta, gamma, vega)[8].

For our main empirical analysis we focus on the cross-section of equity options that are at the money (ATM) and one month away from expiration since they are the most frequently traded ones (cf. Goyal and Saretto, 2009). In further analyses we also include in the money and out of the money options. Every month, we form portfolios based on information from the first trading day after monthly option expiration[9].

To minimize the impact of recording errors we apply several standard filters to the data. Following Goyal and Saretto (2009) we exclude all observations with an ask price lower than the bid price, a bid price equal to zero, or a bid-ask spread below the minimum tick size[10].

[8]Implied volatility estimates as well as option greeks are derived from a binomial tree model based on Cox et al. (1979). For further details we refer to the OptionMetrics IvyDB US reference manual.

[9]The expiration day for standard exchange-traded options is the third Friday of the expiration month or the following Saturday.

[10]Before 2007, the minimum tick size is equal to \$0.05 (\$0.10) for options trading below (above) \$3. On January 26, 2007, the SEC introduced the industry wide Penny Pilot Program reducing the minimum

We further remove prices that violate arbitrage bounds. Following Hu and Jacobs (2020) we exclude all call options where the ask price exceeds the price of the underlying (S) or where the ask price is below $S - K$ with K denoting the exercise price of the option. Additionally, we exclude all put options with a bid price above the exercise price or a bid price below $K - S$. Furthermore, to avoid errors due to stock splits and re-capitalizations, we remove all options for which the adjustment factor for the exercise price does not coincide with the adjustment factor for the share price. In order to eliminate options with no liquidity, we exclude options with zero open interest (cf. Driessen et al., 2009). All equity options in our sample are American. We therefore follow Hu and Jacobs (2020) and remove all options with an ex-dividend date during the remaining life of the option contract to reduce the impact of early exercise.[11] Finally, following Cont and da Fonseca (2002) we exclude all options with moneyness values (defined as the ratio K/S) outside of the interval $[0.5, 1.5]$ to limit numerical uncertainty in computing implied volatilities.

This constitutes our option sample for arbitrary moneyness consisting of 2,280,558 calls and 1,758,895 puts on 9,069 and 8,802 different stocks, respectively over 282 timepoints between January 1996 and June 2019. The number of option contracts varies substantially over time. For example, the number of calls ranges between 1,206 (May 1996) and 16,054 (December 2017) with the number of contracts increasing over time.

In our baseline analysis we focus on ATM options. Therefore, for every month and each underlying we select the call and put contracts that are closest to ATM but according to Goyal and Saretto (2009) only consider options with moneyness values in the interval $[0.975, 1.025]$. This constitutes our ATM option sample consisting of in total 267,147 calls and 244,892 puts. There is also substantial variation in the number of option contracts over

tick size for certain equities to \$0.01 (\$ 0.05). This program, today know as Penny Interval Program, has subsequently been extended to cover more equities. For simplicity, we therefore consider a minimum tick size of \$0.05 (\$0.10) before January 26, 2007, and \$0.01 (\$0.05) for all options below (above) \$3 after January 26, 2007, respectively.

[11] We acknowledge that this controls for early exercise of calls while American puts might still exhibit a premium (Goyal and Saretto, 2009; Barraclough and Whaley, 2012). However, there are several studies arguing that the empirical implications of adjustments for early exercise are small, see, e.g., Boyer and Vorkink (2014).

time. For example, the number of calls in the ATM sample varies between 171 (June 1996) and 1683 (January 2018).

We complement our data sample with stock related characteristics. Following Goyal and Saretto (2009), for each month and each stock we calculate the realized volatility (RV) as the standard deviation of the realized daily stock returns over the preceding 12 months.[12] Additionally, we include the third (skewness) and forth (kurtosis) moment of the underlyings' return distribution (over the most recent 12 months).

3.2 Summary statistics

We provide summary statistics for implied (IV) and realized volatility (RV) of ATM calls and puts and as well as calls and puts of arbitrary moneyness in Table I. The volatilities are annualized. We also include summary statistics for option greeks (delta, gamma, vega) as well as skewness and kurtosis of the underlyings' return distribution. The means are obtained by first taking time-series averages of IV and RV for each stock and then computing the cross-sectional averages of these average volatilities. For the other statistics (median, minimum, maximum, standard deviation, skewness, and kurtosis) we proceed analogously so that the provided statistics can be interpreted as summary statistics of an average stock.

– Insert Table I about here. –

For ATM calls and puts, IV and RV are on average very close to each other. For calls the average IV is 48.6 % compared to an average RV of 49.7 % while for puts the average IV is 50.3 % and the average RV is 50.0 %. The distribution of IV is, however, more volatile than the distribution of RV. Additionally, IV is on average more positively skewed and more leptokurtic than RV. The other variables are except for options' delta (0.536 for calls and

[12]Volatility is highly mean-reverting. Therefore, large deviations between current realized volatility (e.g., calculated over the current month) and the long-term average (calculated over a 12 month period) are unlikely to persist. Therefore, we consider the 12 month RV to be a realistic estimate of volatility over the remaining life of the respective options, see also Goyal and Saretto (2009) and the discussion therein. Apart from this, building on the RV over the preceding 12 months allows us to compare the results from our study to those obtained by Goyal and Saretto (2009).

-0.465 for puts) on average very similar to each other. For example, average values for call options are 0.212 (gamma), 3.476 (vega), 0.303 (skewness), and 10.669 (kurtosis).

The differences between ATM options and options of arbitrary moneyness are quite substantial, especially for IV. First of all, the average level of IV is much higher (59.9 % vs. 48.6 % for calls and 59.7 % vs. 50.0 % for puts). In addition, IV is on average more volatile, more positively skewed, and more leptokurtic than in the case of ATM options. This is due to the fact that implied volatilities are in general not constant for different moneynesses (for a given underlying at a given date) but rather exhibit the pattern of a volatility skew, smile, or smirk, see, e.g., Toft and Prucyk (1997). This variation of IV across moneyness highlights the necessity to control for moneyness when building portfolios for options of arbitrary moneyness. On average, the calls and puts in our sample of arbitrary moneyness also exhibit a slightly higher RV then ATM calls and puts. This is explained by the fact that at a given date on average about half of the underlyings do not enter into the ATM sample because there is no corresponding option contract with moneyness in the interval [0.975, 1.025]. These are stocks with a higher RV. Due to the inclusion of out of the money as well as in the money options into the sample of arbitrary moneyness, there are also substantial differences in option greeks. For example, call options of arbitrary moneyness have an average delta of 0.585 (0.536 ATM), gamma of 0.139 (0.212 ATM), and vega of 2.301 (3.476 ATM).

3.3 Portfolio formation

Our portfolio formation is closely related to Goyal and Saretto (2009) who show that ATM delta-hedged call returns and straddle returns increase as a function of the volatility risk premium (VRP), the log-difference of RV and IV (see Hu and Jacobs, 2020).[13] Goyal and

[13] Goyal and Saretto (2009) argue that volatility is highly mean-reverting and therefore large deviations of the current volatility from the long-term average are unlikely to persist. As IV incorporates expectations on future volatility this implies that large deviations between RV (as long-term average) and IV (as forecast on future volatility) are likely to reduce in magnitude. The authors conclude that options with an IV much lower than the corresponding RV are cheap while options with a much higher IV than RV are expensive. This raises the question how large deviations between RV and IV should be quantified. Note that the authors do not take a clear stand on the question if their observed returns are abnormal or arise as compensation

Saretto (2009) determine large deviations between RV and IV based on their log-difference. The underlying assumption is that by applying the transformation $t : \mathbb{R}^2 \to \mathbb{R} : (RV, IV) \mapsto \log\left(\frac{RV}{IV}\right)$ deviations between RV and IV of all (ATM) options at a particular date can properly be compared to each other. The trading strategy is then simply derived by investing long (short) in the options within the lowest (highest) decile of the log-differences, i.e., low and large deviations are identified based on a one-dimensional portfolio sort. This strategy is subsequently shown to earn a statistically and economically significant average monthly return.

Identifying options with large deviations between RV and IV is equivalent to determining options where IV is particularly high or low *given* a particular level of RV. The log-transformation can then be seen as an intriguingly simple approach that essentially translates the two-dimensional problem of identifying value pairs (RV, IV) of options where IV is abnormally high or low *given* their RV to a one-dimensional problem that can be tackled by simple portfolio sorts. However, Goyal and Saretto (2009) report that RV increases when proceeding from decile one to decile ten, see also Table II. As a consequence, the proposed strategy is not only long (short) in large (low) deviations between RV and IV but also long (short) in high (low) realized volatility.

This is due to the implicit assumption of a linear relationship between RV and IV. As an illustrative example we visualize the value pairs (RV, IV) of all ATM call options in January 2010 along with the conditional quantile curves implied by the log-difference of RV and IV in Figure 4.[14] The red dashes (mostly on the left) and black dashes (mostly on the right) on the x-axis illustrate that there are systematic differences in RV between the long and the short portfolio.

---

for some aggregate risk. On the one hand the authors argue that high deviations between RV and IV are indicative of option mispricing. On the other hand they highlight that deviations between RV and IV will be more (less) pronounced for equities with higher (lower) volatility of volatility.

[14]The long portfolio consists of those options fulfilling the inequality $\log RV_i - \log IV_i \geq q_{90\%}$ where $q_{90\%}$ is defined as the *unconditional* empirical 90 % quantile of the log-differences. Analogously, options in the short portfolio fulfill $\log RV_i - \log IV_i \leq q_{10\%}$. This is equivalent to requiring $IV_i \leq RV_i \cdot e^{-q_{90\%}}$ in the long portfolio and $IV_i \geq RV_i \cdot e^{-q_{10\%}}$ in the short portfolio, i.e., the conditional quantile curves are implicitly assumed to be linear functions in RV.

– Insert Figure 4 about here. –

As a further illustration we provide the conditional quantile curves implied by measuring the VRP as the simple difference $RV - IV$ according to Cao and Han (2013). The main difference is that large deviations between RV and IV are no longer determined based on the relative deviation $\frac{RV}{IV}$ in case of the log-differences, but rather by their absolute deviation $RV - IV$. Consequently, the derived quantile curves are linear and parallel to each other. Again, this causes systematic differences in RV between the long and the short portfolio, see Figure 5 for an illustration.

– Insert Figure 5 about here. –

These differences in realized volatility are problematic against the background of recent literature. For example, Cao and Han (2013) show that delta-hedged equity option returns decrease when the underlying stock's idiosyncratic volatility increases. Furthermore, Hu and Jacobs (2020) find that (total) realized volatility drives raw option returns. In the light of these findings it is unclear to which extent the positive returns of the long-short strategy (high minus low VRP) are attributable to differences in the VRP or differences in the average level of RV. This complicates an interpretation of the obtained returns.

This is an ideal field of application of conditional quantile curves allowing us to derive long (short) portfolios with high (low) differences between RV and IV while controlling for RV. This is done by forming long (short) portfolios consisting of options with low (high) IV *conditional* on their underlying's RV.[15] Furthermore, our nonparametric procedure enables us to model the relation between RV and IV without making any assumptions about the functional form, see Figure 6 where we provide the conditional quantile curves of IV given RV for ATM call options in January 2010.[16] The Figure illustrates that the conditional quantile

[15]As long and short portfolios are formed on a fixed date based on the cross-section of ATM options with maturity of one month, we automatically control for moneyness, maturity, and the risk-free interest rate, too. In further analyses on our option sample of arbitrary moneyness we additionally control for option moneyness.

[16]Our approach can easily be extended to control for more covariates. In Section 3.4 we also report results

curves are in fact non-linear. Furthermore, especially for option with high RV, the gradients of the conditional quantile curves are lower than the ones implied by the log-difference.

– Insert Figure 6 about here. –

Analogously to the trading strategy by Goyal and Saretto (2009) our short portfolio is constituted of options with high IVs relative to the RVs of their underlyings. However, the criterion of choosing options in the lowest decile of log RV - log IV is replaced by selecting options with IV above the *conditional* 90 % quantile *given* the RV of the corresponding underlying. More exactly, for option i with realized volatility RV_i and implied volatility IV_i being in the short portfolio, we require $F(IV_i|RV_i) \stackrel{!}{\geq} 90$ % where $F(\cdot|RV_i)$ denotes the conditional cumulative distribution function of IV given a specific level of realized volatility (RV_i). More intuitively, our short portfolio is constituted of the options where the value-pair (RV, IV) is above the 90 % conditional quantile curve in Figure 6. Analogously, the long portfolio is constituted of the options below the 10 % conditional quantile curve.

Figure 6 illustrates that there is no systematic difference in the average RV between the short and long portfolio as visualized by the red and black dashes on the x-axis. In comparison with Figure 4 it can further be seen that there are noteworthy disagreements in which options enter into the short and long portfolios, especially for options with a large RV.

While these figures reflect the relation between RV and IV on a particular date only, Table II presents evidence that by employing conditional quantile curves systematic differences in RV between the decile portfolios can be reduced significantly. The table compares average values of IV and RV in the 10 decile portfolios obtained by sorting ATM options on the log-difference between RV and IV to our approach of sorting options according to their IV conditional on their RV.[17] The comparison is done for calls and puts separately. We further

for our option sample of arbitrary moneyness. To avoid confounding effects due to high or low moneyness ("volatility skew") we additionally condition on option moneyness. In Section 4.1 we provide results when conditioning on further moments of the underlyings' return distribution.

[17]Decile portfolios based on our conditional quantile approach are obtained by including options with

include option greeks (delta, gamma, vega) as well as further moments of the underlyings' return distribution (skewness, kurtosis) in the table.[18]

– Insert Table II about here. –

Results for call and put options are very similar. For brevity, we therefore focus on the analysis of call results. Like Goyal and Saretto (2009) we find that when sorting on log RV - log IV, IV decreases when proceeding from decile 1 to decile 10 by 2.0 percentage points while RV increases by 4.4 percentage points. Thus, the differences in RV between decile 1 and 10 are more than double as high than those in IV. Furthermore, while the decile portfolios are (almost) monotonic in RV, there is no clear pattern for IV. It is therefore unclear to which extent a corresponding long-short strategy (decile 10 - decile 1) is driven by systematic differences in RV rather than differences in the VRP.

In our approach, we control for RV when forming the decile portfolios by sorting option contracts on their IV *conditional* on their RV. This leads to average differences in IV of nearly 6 percentage points that are by construction ensured to decrease monotonically from decile 1 to decile 10. More importantly, differences in RV between decile 1 and 10 are very small (about 0.4 percentage points). Consequently, the influence of different levels of RV on the returns from a long-short strategy is significantly reduced.[19] For both strategies there is not much variation in option greeks (delta, gamma, vega) across portfolios. However, for both strategies the returns of the underlyings in portfolio 10 are more positively skewed and more leptokurtic than those in portfolio 1 with the differences being more pronounced in the portfolios formed on the log-differences. For example, the average skewness in portfolio 10 is

$F(IV|RV) \geq 90$ % in the first portfolio, options with $80\ \% \leq F(IV|RV) < 90$ % in the second portfolio, etc. The portfolios are equal-weighted. Note, that we sort on the conditional IV in descending order to obtain decile portfolios where the differences between RV and IV are increasing.

[18] Means are first calculated (equally weighted) for each month and each portfolio and are then averaged over time.

[19] For puts, the average difference in RV between decile 1 and 10 is a little bit higher (1.8 percentage points). This might be due to the smaller size of our put option sample. However, this difference is relatively small compared to the average difference in IV (7.6 percentage points). When turning to the option sample of arbitrary moneyness (not covered in Table II), the average differences in RV between decile 1 and 10 are 1.1 percentage points for calls and 0.1 percentage points for puts.

0.218 compared to 0.171 in portfolio 1 for the approach based on conditional quantile curves (0.242 vs. 0.160 for portfolios 10 and 1 formed on the log-differences).[20] To account for the fact that results from our long-short strategy might be biased by systematic differences in skewness and kurtosis we control for these characteristics in a robustness check in Section 4.1.

3.4 Trading strategy

At the money option contracts

We start with a trading strategy based on ATM options only so that we are able to compare the results of a trading strategy based on conditional quantile curves to those obtained by sorting according to the criterion by Goyal and Saretto (2009). First of all, in each of the 10 decile portfolios we calculate monthly returns from a raw option strategy.[21] While the portfolios themselves are formed every month on the first trading day (typically a Monday) after the option expiration, we follow Goyal and Saretto (2009) and start trading the day after (typically a Tuesday) to mitigate microstructure biases. We use the mid-point of bid and ask quotes to proxy for the market price of the option at the beginning of the monthly trade (cf., e.g., Coval and Shumway, 2001; Driessen et al., 2009; Goyal and Saretto, 2009; Cao and Han, 2013; Hu and Jacobs, 2020). We hold all options until expiration.[22] For an option expiring in the money, the return is given by the terminal payoff divided by the price of the option contract minus 1. For an option that expires out of the money we set the return to -100 %.

In addition to raw option returns we also calculate returns from a delta-hedged strategy

[20] Eisdorfer et al. (2020) provide evidence that the nominal stock price level matters for option returns. However, for both approaches (log-differences, conditional quantiles) we do not find much variation in the average log stock price across the decile portfolios which is why we do not include the log stock price in further analyses.

[21] Goyal and Saretto (2009) do not implement a raw option strategy but rather provide results for a delta-hedged option strategy as well as for a strategy based on straddles (built of pairs of calls and puts with the same exercise price).

[22] The issue of early exercise is discussed in Section 3.1.

to reduce the directional exposure to the underlying stocks.[23] Returns for these trading strategies are calculated for calls and puts separately and based on equal-weighted portfolios.

Results for both approaches (log-differences and conditional quantiles) are reported in Table III (delta-hedged returns) and Table IV (raw option returns). The tables provide summary statistics on the monthly returns in each of the decile portfolios (long positions) as well as for a long-short strategy (high minus low VRP). The tables illustrate that returns in the portfolios sorted on the VRP increase (almost) monotonically for both puts and calls in the delta-hedged and in the raw option strategy, respectively. This is true for portfolios formed based on conditional quantile curves and portfolios formed according to the log-difference of RV and IV.

– Insert Table III about here. –

A long-short delta-hedged strategy based on conditional quantile curves yields average monthly returns of 2.0 % for calls and 1.7 % for puts with a monthly Sharpe ratio of 0.842 (2.917 annualized) and 0.796 (2.757 annualized), respectively. In comparison, the delta-hedged strategy based on the log-difference of RV and IV earns monthly Sharpe ratios of 0.786 (calls) and 0.612 (puts). That is, we confirm the existence of a volatility risk premium in the cross-section of option returns and find an even higher effect for delta-hedged returns when controlling for the level of realized volatility.

– Insert Table IV about here. –

Higher absolute returns can be earned with raw option strategies. A long-short strategy based on conditional quantile curves yields average monthly returns of 21.3 % (calls) and 9.9 % (puts) with Sharpe ratios of 0.523 and 0.212, respectively. All trading strategies based on conditional quantiles yield returns that are both economically and statistically significant

[23]Delta-hedged option positions are formed by buying one option contract and buying (for puts) or short-selling (for calls) delta shares of the underlying stock. We follow the conservative approach of Goyal and Saretto (2009) and do not rebalance the portfolio during the holding period, see the discussion ibidem.

with t-statistics of at least 3.5. While for calls the Sharpe ratio for a strategy based on conditional quantile curves is higher than for a strategy based on the log-differences of RV and IV (0.523 vs. 0.434), the opposite is true for puts (0.212 vs. 0.298).[24] Except for delta-hedged call returns, the differences in the Sharpe ratios are statistically significant at the 5 % level when testing according to Ledoit and Wolf (2008). We attribute these differences between delta-hedged returns and raw option returns as well as between calls and puts to the systematic differences in RV when sorting options on the log-difference of RV and IV. While we control for the level of RV across the decile portfolios by using conditional quantile estimates, the long-short strategy based on the log-differences is long (short) in high (low) RV (see Table II). This differing influence of RV on the returns from various strategies highlights the importance of controlling for the influence of the level of RV when empirically analyzing the VRP.

Options of arbitrary moneyness

So far, we have analyzed our sample of ATM calls and puts. As ATM calls (puts) account for only 11.7 % (13.9 %) of all calls (puts) in our sample, we extend our analysis to options with moneyness (defined as the ratio K/S) between 0.5 and 1.5. In doing so, we can analyze if there is still a volatility risk premium in the cross-section of option returns when additionally including in the money and out of the money options. Furthermore, this allows us to increase the number of observations by a factor of about 8. Finally, by requiring ATM options to have a moneyness in the interval [0.975, 1.025] (see Section 3.1) we also exclude many underlyings from our analysis. This happens when there are no options with appropriate moneyness available at a specific date. Not restricting moneyness to the small interval [0.975, 1.025] therefore doubles the number of underlyings available in our sample on average over all months. In particular, the minimum number of different underlyings at a given day increases from 171 to 734 (for calls) and 105 to 387 (for puts). This facilitates

[24] These differences are robust to the choice of the method for estimating conditional quantiles, see Section 4.3.

controlling for further moments of the underlyings' return distribution, see Section 4.1.

Our approach of using conditional quantiles for deriving decile portfolios generalizes naturally to options of arbitrary moneyness. Therefore, we form portfolios based on extreme values of IV conditional on RV *and* options' moneyness.[25] We condition on options' moneyness to avoid results from such a trading strategy to be biased by systematic differences in options' moneyness (volatility skew, see, e.g., Toft and Prucyk (1997)). Summary statistics for monthly delta-hedged and raw option returns for both calls and puts are reported in Table V. The results are in line with our previous findings and confirm that the VRP is also priced in the cross-section of returns of options with arbitrary moneyness. For example, a delta-hedged strategy that is long (short) in options with a high (low) VRP yields an average monthly return of 2.4 % for calls and 2.5 % for puts with a monthly Sharpe ratio of 0.816 (calls) and 0.844 (puts), respectively. The corresponding long-short raw option strategy exhibits an average monthly return of 20.1 % for calls and 13.1 % for puts with a monthly Sharpe ratio of 0.390 and 0.217, respectively. All trading strategies yield returns that are both economically and statistically significant with t-statistics above 3.6.

– Insert Table V about here. –

4 Robustness checks

4.1 Controlling for further moments of the underlyings' return distribution

In our baseline analysis we sort options into decile portfolios based on their IV conditional on their RV. This allows us to identify large deviations between IV and RV to study the

[25] To put this into perspective: If one were to replicate this based on portfolio sorts, one would be required to apply a conditional triple-sort to the data. This is something that is impracticable if not infeasible against the background of only 826 put options at the beginning of our data sample. In Section 4.1 we additionally control for skewness and kurtosis in the underlyings' return distributions. This would be, at the latest, impossible to achieve with portfolio sorts.

VRP without potential biases arising from different levels in RV. However, Table II indicates that after controlling for RV (the second moment of the underlyings' return distribution) via conditional quantiles there remain differences in the third (skewness) and fourth (kurtosis) moment of the underlyings' return distribution between the long (decile portfolio 10) and the short portfolio (decile portfolio 1). These differences might potentially bias returns from a strategy exploiting the VRP.

Fortunately, our approach of using conditional quantiles allows us to additionally condition on the underlyings' skewness and kurtosis. This is impossible via conditional portfolio sorts as this would require a quadruple sort on IV, RV, skewness, and kurtosis in our ATM option sample.[26] Additionally controlling for moneyness in our sample of arbitrary moneyness would even require a quintuple sort.

Summary statistics for the monthly returns from the delta-hedged and raw option strategies for puts and calls are reported in Table VI for the ATM option sample as well as the sample of arbitrary moneyness. Skewness and kurtosis are calculated based on the realized returns from the option contracts' underlying over the most recent 12 months. The results are in line with our previous findings.

– Insert Table VI about here. –

4.2 Transaction costs

Transaction costs in option markets can be quite large and might in part explain some pricing anomalies such as put-call parity violations, see Goyal and Saretto (2009) and the references therein. For example, in our ATM option sample the average bid-ask spread relative to the mid-prices is 23.7 % for calls and 22.7 % for puts. We therefore study

[26] This becomes even more apparent when considering the number of only 105 puts at the beginning of our sample period in January 1996. This makes a conditional portfolio sort in 10 bins for each variable (10,000 bins in total) infeasible. Instead performing a portfolio sort based on only 5 bins for each variable would still be infeasible (625 bins). Apart from this, it would be a very imprecise approximation of the true conditional quantile curves (see the illustrative example in Figure 2). These and further issues are discussed in Section 2.2.

limitations of investors in exploiting profits from the long-short strategies based on the VRP.

So far, we assumed investors to trade options at their mid-point price. However, to account for transaction costs we recalculate returns from our strategies when incorporating bid-ask spreads. This not only concerns the option contracts but also the underlying stocks. For the raw option strategy transaction costs in the underlying stocks only occur at expiration as all equity options have to be delivered physically. In our delta-hedged strategy, stock related transaction costs additionally arise at the beginning of each monthly trading period.

As already mentioned, (quoted) bid-ask spreads can be quite high. However, although *effective* bid-asks spreads are usually still quite high in absolute terms, there is empirical evidence that they are small relative to the quoted spreads with ratios below 0.5 (see, e.g., Mayhew, 2002; de Fontnouvelle et al., 2003). We therefore recalculate returns for our raw and delta-hedged option strategies based on an effective spread of 50% relative to the quoted spreads. For example, for an option contract with quoted bid and ask equal to $2 and $3, respectively, we assume investors to buy at $2.75 and sell at $2.25, that is, at an effective spread of $0.5 instead of the quoted spread of $1. However, Battalio et al. (2004) find higher ratios between effective and quoted spreads when studying a small sample of large stocks between June 2000 and January 2002. We therefore follow Goyal and Saretto (2009) and additionally consider ratios of 75 % and 100 % between effective and quoted spreads.

Average returns for raw and delta-hedged long-short strategies for puts and calls, respectively, are reported in Table VII along with their t-statistics. As expected, average monthly returns decrease substantially when accounting for transaction costs. For example, the average monthly return from the ATM delta-hedged call strategy decreases from 2.0 % when trading at the mid-point prices to 0.3 % when considering an effective spread of 50 %. However, average delta-hedged returns remain positive for calls and puts both ATM as well as for arbitrary moneyness with t-statistics between 1.078 and 3.286. Nevertheless, while raw option returns are still positive after considering transaction costs (50 % ratio of effective to quoted spreads) for calls (both ATM and for arbitrary moneyness), raw option returns for puts are negative (both ATM and for arbitrary moneyness). When further increasing transaction costs to 75 % or even 100 % of effective to quoted spreads, returns deteriorate further and are negative for all reported strategies. This illustrates, that returns from our option strategies are strongly affected by transaction costs, especially in the case of raw option strategies.

– Insert Table VII about here. –

4.3 Other estimators of conditional quantiles

There are various methods for estimating conditional quantiles, see Section 2.3 for details on the estimation of conditional quantiles in general and specific estimators in particular. So far, we used the *leveraging estimator* to form decile portfolios and subsequently calculate returns from a long-short strategy. To make sure that our results are not driven by the specific choice of the estimator we repeat our analyses based on the *quantization estimator* due to Charlier et al. (2015b) and the *copula estimator* due to Kraus and Czado (2017). Results on the returns from long-short strategies based on the three different estimators are reported in Table VIII. For illustrative purposes, we also provide results obtained from a (conditional) double-sort (for ATM options) and triple-sort (for options of arbitrary moneyness) into 10 bins for each variable.[27]

– Insert Table VIII about here. –

The results are in line with our previous findings. In particular, all conclusions with regard to the comparison between our approach of calculating the VRP while controlling for

[27] As outlined in Section 2.2 portfolio sorts are inflexible with regard to the percentage of observations that are supposed to enter into the long and short portfolios, especially when the number of observations is low or when sorting on multiple variables. For example, in our put sample of arbitrary moneyness the percentage of observations entering into the long and short portfolios varies between 10.32 % and 18.62 % over all trading periods with a mean of 11.08 %. However, while conditional portfolio sorts might still be feasible in this case (with limitations), controlling for further characteristics as in Section 4.1 would be impossible.

RV and the approach based on the log-difference of RV and IV remain valid, see Section 3.4 for details. However, the variation between the different estimators is higher for the trading strategies involving options of arbitrary moneyness, which is mainly due to the quantization estimator.[28]

4.4 Including dividend-paying stocks

All equity options in our sample are American. In our baseline analysis we therefore follow Hu and Jacobs (2020) and exclude all options with an ex-dividend date during the remaining life of the contract. This is done to reduce the impact of early exercise, see Section 3.1. However, to show robustness of our results, we recalculate the returns of our option strategies when not excluding dividend-paying stocks. This increases our ATM option sample to 326,224 calls and 299,924 puts while the sample of arbitrary moneyness expands to 2,682,492 calls and 2,145,435 puts. Summary statistics on the monthly returns of delta-hedged and raw option strategies for calls and puts both ATM and for arbitrary moneyness are reported in Table IX. The results are in line with our previous findings.

– Insert Table IX about here. –

5 Conclusion

In this paper, we find new evidence that delta-hedged equity option returns include a volatility risk premium. We sort options on their implied volatility conditional on their realized volatility to proxy for the volatility risk premium. A strategy that is long (short) in high (low) deviations between realized and implied volatilities yields returns that are both economically and statistically significant. This result holds for call and put delta-hedged and raw option returns for both at the money (ATM) options and options of arbitrary moneyness.

[28]In an unreported simulation study we find the leveraging and the copula estimator to be most appropriate. However, we include the quantization estimator to provide a more comprehensive picture.

Changing the type of estimator of conditional quantiles as well as controlling for additional characteristics of the underlying does not affect our main finding.

The key to our main finding, and the difference to previous work, is our use of *conditional quantiles* in contrast to standard portfolio sorts or regression techniques. Using conditional quantiles estimated via nonparametric machine learning algorithms allows us to capture the non-linear relation between implied and realized volatility while at the same time controlling for characteristics that are known to affect stock volatility as well as the cross-section of expected option returns. As our main result, we find that previous work on the existence of a risk premium for volatility and volatility mispricing were correct, but considerably underestimated the size of the effect. Exploiting the estimated quantiles of implied volatility conditional on realized volatility and moneyness leads to returns that are higher than those reported in previous work on similar volatility strategies.

Our proposed use of conditional quantiles should be seen as a good compromise between standard portfolio sorts and nonparametric cross-sectional regressions. While the former easily fail to control for more than two covariates, the latter do not come with an easy interpretability in empirical asset pricing where one is interested in ready-to-use trading strategies involving a reasonably small number of assets. While our empirical study is concerned with the cross-section of option returns, our method is sufficiently general and can easily be applied to other assets, most importantly stocks.

References

Andreou, E. and E. Ghysels (2020): "Predicting the VIX and the volatility risk premium: The role of short-run funding spreads Volatility Factors," *Journal of Econometrics*, 220, 366–398.

Baele, L., J. Driessen, S. Ebert, J. M. Londono, and O. G. Spalt (2018): "Cumulative Prospect Theory, Option Returns, and the Variance Premium," *Review of Financial Studies*, 32, 3667–3723.

Banz, R. W. (1981): "The relationship between return and market value of common stocks," *Journal of Financial Economics*, 9, 3–18.

Barraclough, K. and R. E. Whaley (2012): "Early Exercise of Put Options on Stocks," *The Journal of Finance*, 67, 1423–1456.

Basu, S. (1977): "Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis," *The Journal of Finance*, 32, 663–682.

Battalio, R., B. Hatch, and R. Jennings (2004): "Toward a National Market System for U.S. Exchange-listed Equity Options," *The Journal of Finance*, 59, 933–962.

Bhattacharya, P. K. and A. K. Gangopadhyay (1990): "Kernel and nearest-neighbor estimation of a conditional quantile," *The Annals of Statistics*, 18, 1400–1415.

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

Boyer, B. H. and K. Vorkink (2014): "Stock Options as Lotteries," *The Journal of Finance*, 69, 1485–1527.

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

Cao, J., B. Han, Q. Tong, and X. Zhan (2021): "Option Return Predictability," *Review of Financial Studies, forthcoming.*

Cao, J., A. Vasquez, X. Xiao, and X. Zhan (2019): "Volatility Uncertainty and the Cross-Section of Option Returns," *SSRN Electronic Journal*.

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

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and E. Schaumburg (2020): "Characteristic-Sorted Portfolios: Estimation and Inference," *The Review of Economics and Statistics*, 102, 531–551.

Charlier, I., D. Paindaveine, and J. Saracco (2015a): "Conditional quantile estimation based on optimal quantization: From theory to practice," *Computational Statistics & Data Analysis*, 91, 20–39.

——— (2015b): "Conditional quantile estimation through optimal quantization," *Journal of Statistical Planning and Inference*, 156, 14–30.

Cochrane, J. H. (2011): "Presidential Address: Discount Rates," *The Journal of Finance*, 66, 1047–1108.

Connor, G., M. Hagmann, and O. Linton (2012): "Efficient Semiparametric Estimation of the Fama-French Model and Extensions," *Econometrica*, 80, 713–754.

Connor, G. and O. Linton (2007): "Semiparametric estimation of a characteristic-based factor model of common stock returns," *Journal of Empirical Finance*, 14, 694–717.

Cont, R. and J. da Fonseca (2002): "Dynamics of implied volatility surfaces," *Quantitative Finance*, 2, 45–60.

Coval, J. D. and T. Shumway (2001): "Expected Option Returns," *The Journal of Finance*, 56, 983–1009.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979): "Option pricing: A simplified approach," *Journal of Financial Economics*, 7, 229–263.

De Bondt, W. F. and R. Thaler (1985): "Does the stock market overreact?" *The Journal of Finance*, 40, 793–805.

de Fontnouvelle, P., R. P. H. Fishe, and J. H. Harris (2003): "The Behavior of Bid-Ask Spreads and Volume in Options Markets during the Competition for Listings in 1999," *The Journal of Finance*, 58, 2437–2463.

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

Duffy, N. and D. Helmbold (2002): "Boosting Methods for Regression," *Machine Learning*, 47, 153–200.

Eisdorfer, A., A. Goyal, and A. Zhdanov (2020): "Cheap Options Are Expensive," *SSRN Electronic Journal*.

Fama, E. F. and K. R. French (1992): "The cross-section of expected stock returns," *The Journal of Finance*, 47, 427–465.

Goyal, A. (2012): "Empirical cross-sectional asset pricing: a survey," *Financial Markets and Portfolio Management*, 26, 3–38.

Goyal, A. and A. Saretto (2009): "Cross-section of option returns and volatility," *Journal of Financial Economics*, 94, 310–326.

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical Asset Pricing via Machine Learning," *Review of Financial Studies*, 33, 2223–2273.

Hu, G. and K. Jacobs (2020): "Volatility and Expected Option Returns," *Journal of Financial and Quantitative Analysis*, 55, 1025–1060.

Huang, D., C. Schlag, I. Shaliastovich, and J. Thimme (2019): "Volatility-of-Volatility Risk," *Journal of Financial and Quantitative Analysis*, 54, 2423–2452.

Jegadeesh, N. (1990): "Evidence of predictable behavior of security returns," *The Journal of Finance*, 45, 881–898.

Jegadeesh, N. and S. Titman (1993): "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of Finance*, 48, 65–91.

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

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Kraus, D. and C. Czado (2017): "D-vine copula based quantile regression," *Computational Statistics & Data Analysis*, 110, 1–18.

Ledoit, O. and M. Wolf (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

Li, Q. and J. S. Racine (2008): "Nonparametric Estimation of Conditional CDF and Quantile Functions With Mixed Categorical and Continuous Data," *Journal of Business & Economic Statistics*, 26, 423–434.

Mayhew, S. (2002): "Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets," *The Journal of Finance*, 57, 931–958.

Meir, R. and G. Rätsch (2003): "An Introduction to Boosting and Leveraging," in *Advanced Lectures on Machine Learning*, ed. by S. Mendelson, Berlin, Heidelberg: Springer, Lecture Notes in Computer Science, 118–183.

Moritz, B. and T. Zimmermann (2016): "Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns," *SSRN Electronic Journal*.

Patton, A. and A. Timmermann (2010): "Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts," *Journal of Financial Economics*, 98, 605–625.

Stone, C. J. (1980): "Optimal Rates of Convergence for Nonparametric Estimators," *The Annals of Statistics*, 8, 1348–1360.

Toft, K. B. and B. Prucyk (1997): "Options on Leveraged Equity: Theory and Empirical Tests," *The Journal of Finance*, 52, 1151–1180.

Table I: Summary statistics for the option samples

This table presents summary statistics for implied (IV) and realized volatilites (RV) of ATM calls and puts as well as option contracts of arbitrary moneyness. All options are American and have a maturity of about one month. Our ATM sample consists of 267,147 calls and 244,892 puts while the sample of arbitrary moneyness is composed of 2,280,558 calls and 1,758,895 puts. The sample period is January 1996 to June 2019. IVs and option greeks (delta, gamma, vega) are retrieved from the OptionMetrics IvyDB US database and calculated based on a binomial tree model due to Cox et al. (1979). The volatilities are annualized. Option underlyings' returns are retrieved from CRSP. Skewness (skew) and kurtosis (kurt) are computed from realized returns over the most recent 12 months. The means are obtained by first taking the time-series average of IV and RV for each stock and then computing the cross-sectional average of these average volatilities. For the other statistics (median, mininum (min), maximum (max), standard deviation (sd), skewness (skew), and kurtosis (kurt)) we proceed analogously. Note that in the option samples of arbitrary moneyness, multiple IV values per stock and month can enter into the summary statistics.
The differences in RV between the ATM sample and the sample of arbitrary moneyness are due to the exclusion of option contracts and their corresponding underlyings with moneyness outside of the interval [0.975, 1.025] for ATM options. For more details on the construction of our sample we refer to Section 3.1.

			mean	**median**	**min**	**max**	**sd**	**skew**	**kurt**
ATM	**Calls**	IV	0.486	0.465	0.317	0.808	0.143	0.791	3.943
		RV	0.497	0.477	0.365	0.750	0.123	0.684	3.386
		delta	0.536	0.537	0.438	0.629	0.058	-0.094	2.360
		gamma	0.212	0.198	0.117	0.397	0.087	0.687	3.506
		vega	3.476	3.286	1.654	6.467	1.383	0.363	2.703
		skew	0.303	0.290	-1.223	1.909	0.931	0.028	3.513
		kurt	10.669	8.738	5.237	27.044	6.978	1.143	4.830
	Puts	IV	0.503	0.480	0.341	0.823	0.144	0.807	3.926
		RV	0.500	0.479	0.369	0.748	0.124	0.680	3.352
		delta	-0.465	-0.463	-0.557	-0.381	0.055	-0.137	2.274
		gamma	0.201	0.188	0.114	0.365	0.080	0.648	3.356
		vega	3.513	3.322	1.705	6.473	1.405	0.359	2.678
		skew	0.280	0.274	-1.213	1.836	0.936	0.021	3.401
		kurt	10.695	8.880	5.286	26.337	6.950	1.116	4.684
Arbitrary moneyness	**Calls**	IV	0.599	0.551	0.281	1.455	0.224	1.356	6.274
		RV	0.524	0.498	0.357	0.825	0.133	0.629	3.217
		delta	0.585	0.606	0.128	0.968	0.272	-0.148	1.763
		gamma	0.139	0.126	0.023	0.442	0.089	0.990	5.514
		vega	2.301	2.090	0.274	6.347	1.354	0.519	3.214
		skew	0.293	0.278	-1.582	2.406	0.977	0.119	4.261
		kurt	10.953	8.664	4.694	33.465	7.535	1.400	6.300
	Puts	IV	0.597	0.563	0.330	1.210	0.188	1.001	4.752
		RV	0.531	0.506	0.365	0.823	0.134	0.603	3.161
		delta	-0.405	-0.382	-0.808	-0.092	0.221	-0.222	1.921
		gamma	0.153	0.139	0.051	0.403	0.078	0.950	5.106
		vega	2.576	2.352	0.739	6.375	1.297	0.568	3.167
		skew	0.282	0.272	-1.529	2.233	0.958	0.066	4.030
		kurt	10.933	8.734	4.816	31.563	7.307	1.320	5.864

Table II: Decile portfolios for ATM options

This table provides information on the average values of various covariates within decile portfolios that are formed according to two approaches: This paper proposes sorting options (in descending order) on their IV conditional on their underlyings' RV into decile portfolios to proxy for the VRP. This yields portfolios where on average the differences RV-IV increase monotonically while RV remains nearly constant. We compare this to portfolios obtained by sorting options on the log-difference of RV and IV (according to Goyal and Saretto, 2009). This yields portfolios that are monotonically increasing in the average RV. This is due to an implicit linear assumption on the conditional quantile curves, see Figure 4 for an illustration. In opposition, we calculate conditional quantiles nonparametrically to mitigate this issue, see Figure 6.
Within the decile portfolios we provide means of IV, RV, option greeks (delta, gamma, vega) as well as skewness (skew), and kurtosis (kurt) of option underlyings' return distribution. The means are calculated by first computing averages for each portfolio and each month and then taking the time-series averages. Results are based on our ATM call and put sample January 1996 to June 2019, see Section 3.1 for details on the sample construction.

			Decile portfolios									
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Calls	**log-differences**	IV	0.438	0.423	0.417	0.413	0.410	0.410	0.411	0.413	0.414	0.418
		RV	0.423	0.421	0.421	0.421	0.421	0.428	0.431	0.439	0.447	0.467
		RV-IV	-0.015	-0.002	0.004	0.008	0.011	0.018	0.020	0.026	0.033	0.049
		delta	0.534	0.534	0.533	0.533	0.531	0.532	0.533	0.531	0.531	0.532
		gamma	0.151	0.150	0.149	0.152	0.150	0.152	0.151	0.153	0.155	0.164
		vega	4.857	4.941	5.037	4.983	4.986	4.921	4.956	4.952	4.861	4.753
		skew	0.160	0.169	0.165	0.170	0.172	0.166	0.156	0.175	0.179	0.242
		kurt	8.758	8.619	8.548	8.652	8.647	8.674	8.740	8.990	9.136	10.679
	cond. quantiles	IV	0.451	0.437	0.425	0.421	0.413	0.410	0.406	0.401	0.398	0.392
		RV	0.438	0.438	0.432	0.432	0.428	0.430	0.428	0.427	0.431	0.434
		RV-IV	-0.013	0.001	0.007	0.011	0.015	0.020	0.022	0.026	0.033	0.042
		delta	0.534	0.535	0.534	0.534	0.533	0.533	0.532	0.531	0.531	0.531
		gamma	0.153	0.151	0.150	0.152	0.151	0.151	0.151	0.154	0.154	0.163
		vega	4.770	4.840	4.957	4.865	4.947	4.981	4.991	4.923	5.094	4.987
		skew	0.171	0.175	0.171	0.181	0.175	0.161	0.159	0.171	0.177	0.218
		kurt	8.870	8.770	8.634	8.829	8.624	8.740	8.732	8.910	9.181	10.280
Puts	**log-differences**	IV	0.457	0.439	0.431	0.426	0.426	0.425	0.426	0.423	0.424	0.430
		RV	0.430	0.427	0.426	0.426	0.430	0.433	0.438	0.440	0.450	0.475
		RV-IV	-0.027	-0.012	-0.005	0.000	0.004	0.008	0.012	0.017	0.026	0.045
		delta	-0.469	-0.470	-0.470	-0.470	-0.473	-0.473	-0.472	-0.472	-0.473	-0.473
		gamma	0.146	0.143	0.145	0.145	0.145	0.145	0.146	0.148	0.148	0.154
		vega	4.818	4.990	5.002	5.036	5.032	5.005	5.001	4.989	5.001	4.866
		skew	0.170	0.161	0.163	0.163	0.183	0.167	0.147	0.161	0.169	0.228
		kurt	8.850	8.716	8.577	8.529	8.478	8.683	8.921	8.773	9.283	11.018
	cond. quantiles	IV	0.475	0.457	0.442	0.438	0.427	0.427	0.420	0.415	0.408	0.399
		RV	0.449	0.446	0.440	0.439	0.434	0.437	0.434	0.433	0.432	0.431
		RV-IV	-0.026	-0.011	-0.002	0.001	0.007	0.010	0.014	0.018	0.024	0.032
		delta	-0.470	-0.469	-0.469	-0.470	-0.471	-0.472	-0.471	-0.472	-0.474	-0.475
		gamma	0.147	0.145	0.145	0.145	0.145	0.146	0.145	0.147	0.148	0.153
		vega	4.702	4.853	4.901	4.920	4.973	5.014	5.091	5.011	5.112	5.150
		skew	0.191	0.175	0.177	0.178	0.165	0.170	0.151	0.160	0.153	0.196
		kurt	9.026	8.817	8.820	8.617	8.658	8.749	8.635	8.956	9.208	10.412

Table III: Delta-hedged returns of ATM options

This table provides summary statistics on monthly delta-hedged returns of decile portfolios for ATM call and put options. Decile portfolios are formed by sorting on the log-difference of RV and IV according to Goyal and Saretto (2009) and by sorting on options' IV *conditional* on their RV, respectively. We additionally provide returns from a long-short strategy, that is long in decile 10 and short in decile 1. The portfolios are equally weighted. Details on the trading strategy can be found in Section 3.4.
We report the mean, standard deviation (sd), minimum (min), maximum (max), and Sharpe ratio (SR) of the monthly returns. Since the long-short strategy is a zero investment strategy, the Sharpe ratio is simply calculated as the ratio between mean and standard deviation. We calculate the Sharpe ratios in the decile portfolios accordingly to make a comparison easier. The sample period is January 1996 to June 2019. Details on the sample construction can be found in Section 3.1.

Decile portfolios

			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**10-1**
Calls	**log-differences**	mean	-0.015	-0.007	-0.006	-0.005	-0.003	-0.003	-0.002	-0.001	0.000	0.005	0.020
		sd	0.027	0.026	0.026	0.028	0.029	0.028	0.029	0.029	0.030	0.032	0.025
		min	-0.078	-0.086	-0.073	-0.068	-0.069	-0.056	-0.081	-0.062	-0.069	-0.051	-0.045
		max	0.150	0.171	0.203	0.184	0.187	0.191	0.192	0.186	0.194	0.196	0.161
		SR	-0.571	-0.257	-0.243	-0.188	-0.106	-0.102	-0.068	-0.046	0.015	0.142	0.786
	cond. quant.	mean	-0.016	-0.008	-0.006	-0.005	-0.002	-0.003	-0.001	0.000	0.001	0.003	0.020
		sd	0.030	0.029	0.029	0.028	0.031	0.028	0.028	0.027	0.025	0.025	0.023
		min	-0.088	-0.076	-0.072	-0.070	-0.079	-0.071	-0.061	-0.060	-0.056	-0.038	-0.057
		max	0.161	0.199	0.183	0.180	0.230	0.185	0.191	0.184	0.179	0.169	0.113
		SR	-0.541	-0.273	-0.209	-0.178	-0.073	-0.103	-0.047	-0.018	0.047	0.129	0.842
Puts	**log-differences**	mean	-0.012	-0.006	-0.004	-0.003	-0.001	0.000	0.000	0.001	0.002	0.003	0.015
		sd	0.026	0.024	0.025	0.026	0.025	0.027	0.027	0.027	0.028	0.031	0.025
		min	-0.078	-0.065	-0.087	-0.078	-0.053	-0.064	-0.075	-0.088	-0.054	-0.057	-0.040
		max	0.132	0.169	0.145	0.177	0.171	0.161	0.203	0.160	0.170	0.202	0.209
		SR	-0.477	-0.256	-0.166	-0.112	-0.058	0.002	-0.001	0.026	0.086	0.104	0.612
	cond. quant.	mean	-0.014	-0.007	-0.003	-0.003	-0.001	0.000	0.001	0.001	0.002	0.003	0.017
		sd	0.028	0.027	0.027	0.026	0.026	0.026	0.026	0.025	0.025	0.025	0.021
		min	-0.078	-0.071	-0.075	-0.080	-0.066	-0.059	-0.061	-0.070	-0.053	-0.065	-0.058
		max	0.147	0.172	0.193	0.161	0.190	0.162	0.181	0.142	0.161	0.155	0.081
		SR	-0.492	-0.265	-0.120	-0.120	-0.026	-0.016	0.020	0.043	0.085	0.130	0.796

Table IV: Raw returns of ATM options

This table provides the same information on monthly returns from decile portfolios of ATM options as Table III but for a raw option strategy.

			Decile portfolios										
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**10-1**
Calls	**log-differences**	mean	0.006	0.098	0.084	0.104	0.112	0.129	0.120	0.129	0.159	0.185	0.179
		sd	0.512	0.591	0.566	0.622	0.646	0.648	0.657	0.659	0.681	0.687	0.413
		min	-0.919	-0.945	-0.998	-0.998	-0.987	-0.975	-1.000	-1.000	-0.990	-0.982	-1.307
		max	2.407	1.951	1.836	2.009	2.106	2.175	2.206	2.441	2.751	2.792	2.317
		SR	0.012	0.166	0.149	0.168	0.174	0.199	0.183	0.195	0.234	0.269	0.434
	cond. quant.	mean	-0.006	0.058	0.078	0.088	0.111	0.113	0.103	0.137	0.155	0.207	0.213
		sd	0.511	0.544	0.568	0.597	0.643	0.634	0.643	0.670	0.676	0.687	0.407
		min	-0.918	-0.982	-0.980	-0.992	-0.989	-1.000	-1.000	-1.000	-1.000	-0.975	-0.721
		max	2.375	1.674	1.737	1.786	2.322	2.093	2.294	2.819	2.157	2.475	1.885
		SR	-0.012	0.107	0.138	0.148	0.173	0.179	0.161	0.205	0.229	0.301	0.523
Puts	**log-differences**	mean	-0.187	-0.164	-0.143	-0.155	-0.103	-0.090	-0.076	-0.081	-0.062	-0.062	0.125
		sd	0.615	0.695	0.756	0.767	0.791	0.815	0.841	0.801	0.827	0.838	0.419
		min	-0.980	-0.978	-0.987	-1.000	-0.974	-0.990	-0.951	-0.976	-0.979	-0.985	-1.184
		max	3.825	4.728	4.338	4.980	4.971	4.575	5.665	4.699	4.665	4.827	1.635
		SR	-0.304	-0.235	-0.188	-0.202	-0.131	-0.111	-0.090	-0.101	-0.075	-0.074	0.298
	cond. quant.	mean	-0.176	-0.159	-0.122	-0.143	-0.105	-0.100	-0.087	-0.084	-0.077	-0.077	0.099
		sd	0.600	0.661	0.740	0.743	0.800	0.788	0.838	0.810	0.850	0.888	0.467
		min	-0.965	-0.982	-0.964	-0.992	-0.976	-0.947	-0.976	-0.990	-0.990	-0.958	-1.295
		max	3.794	4.017	5.054	4.481	5.425	4.738	5.136	4.439	5.013	5.054	2.741
		SR	-0.294	-0.240	-0.164	-0.192	-0.131	-0.127	-0.103	-0.103	-0.090	-0.087	0.212

Table V: Returns of trading strategies for options with arbitrary moneyness

This table provides similar information on monthly delta-hedged as well as raw option returns from decile portfolios to Tables III and IV but for options with arbitrary moneyness. Furthermore, decile portfolios are formed by sorting on options' IV conditional on their RV *and* their moneyness, see Section 3.4 for details.

			Decile portfolios										
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**10-1**
delta-hedged returns	**Calls**	mean	-0.023	-0.014	-0.010	-0.008	-0.007	-0.005	-0.004	-0.004	-0.002	0.001	0.024
		sd	0.032	0.030	0.029	0.029	0.029	0.029	0.030	0.030	0.029	0.032	0.030
		min	-0.117	-0.093	-0.105	-0.072	-0.067	-0.065	-0.065	-0.081	-0.061	-0.056	-0.078
		max	0.186	0.178	0.178	0.182	0.194	0.192	0.192	0.223	0.226	0.271	0.290
		SR	-0.726	-0.469	-0.333	-0.285	-0.233	-0.163	-0.143	-0.119	-0.071	0.029	0.816
	Puts	mean	-0.023	-0.013	-0.010	-0.008	-0.006	-0.005	-0.004	-0.003	-0.001	0.002	0.025
		sd	0.032	0.038	0.036	0.036	0.036	0.036	0.035	0.034	0.036	0.039	0.029
		min	-0.103	-0.074	-0.083	-0.085	-0.069	-0.062	-0.067	-0.071	-0.057	-0.049	-0.084
		max	0.245	0.308	0.341	0.308	0.326	0.299	0.284	0.286	0.334	0.356	0.250
		SR	-0.712	-0.333	-0.273	-0.213	-0.160	-0.134	-0.099	-0.087	-0.038	0.054	0.844
raw returns	**Calls**	mean	-0.059	-0.014	0.018	0.030	0.039	0.058	0.061	0.066	0.093	0.142	0.201
		sd	0.411	0.444	0.476	0.498	0.524	0.559	0.576	0.600	0.635	0.726	0.515
		min	-0.855	-0.889	-0.950	-0.949	-0.961	-0.959	-0.963	-0.948	-0.945	-0.935	-1.429
		max	1.384	1.477	1.723	2.047	2.321	2.847	3.562	4.246	4.242	5.699	4.871
		SR	-0.143	-0.032	0.037	0.060	0.075	0.104	0.105	0.110	0.147	0.196	0.390
	Puts	mean	-0.232	-0.202	-0.194	-0.179	-0.181	-0.176	-0.166	-0.152	-0.139	-0.101	0.131
		sd	0.596	0.750	0.795	0.835	0.862	0.885	0.906	0.888	0.957	1.030	0.603
		min	-0.919	-0.919	-0.959	-0.853	-0.929	-0.911	-0.942	-0.957	-0.941	-0.944	-1.634
		max	4.801	6.321	7.279	7.175	7.429	7.135	6.847	6.002	6.918	6.872	4.643
		SR	-0.389	-0.269	-0.244	-0.214	-0.210	-0.199	-0.184	-0.171	-0.145	-0.098	0.217

Table VI: Conditioning on further moments of the underlyings' return distribution

This table presents summary statistics on the monthly returns from long-short portfolios based on the conditional 10 % and 90 % quantiles of option IV. In our baseline analysis we conditioned on the second moment (RV) of the underlyings' return distribution (and additionally on moneyness for the sample of arbitrary moneyness). For robustness, we additionally condition on the third (skewness) and fourth (kurtosis) moment of the underlyings' return distribution. Summary statistics for the monthly returns (mean, standard deviation (sd), minimum (min), maximum (max), Sharpe ratio (SR)) are reported when conditioning on skewness (skew) and kurtosis (kurt) separately and jointly, respectively. We provide results for delta-hedged and raw option strategies for puts and calls for both ATM options as well as options with arbitrary moneyness. The sample period is January 1996 to June 2019. For more details on the sample construction and our trading strategy we refer to Sections 3.1 and 3.4.

				mean	**sd**	**min**	**max**	**SR**
ATM	**Delta-hedged**	**Calls**	skew	0.021	0.026	-0.088	0.104	0.802
			kurt	0.021	0.025	-0.058	0.148	0.855
			skew and kurt	0.021	0.025	-0.095	0.144	0.823
		Puts	skew	0.018	0.024	-0.093	0.124	0.752
			kurt	0.018	0.021	-0.046	0.091	0.850
			skew and kurt	0.018	0.023	-0.087	0.093	0.770
	Raw returns	**Calls**	skew	0.230	0.428	-0.982	1.954	0.537
			kurt	0.227	0.435	-0.864	2.302	0.522
			skew and kurt	0.232	0.421	-0.853	2.505	0.552
		Puts	skew	0.090	0.454	-0.751	2.812	0.197
			kurt	0.101	0.466	-1.166	2.826	0.216
			skew and kurt	0.081	0.456	-0.938	3.088	0.178
Arbitrary moneyness	**Delta-hedged**	**Calls**	skew	0.025	0.030	-0.104	0.291	0.840
			kurt	0.025	0.029	-0.090	0.275	0.864
			skew and kurt	0.025	0.029	-0.094	0.252	0.863
		Puts	skew	0.026	0.028	-0.103	0.174	0.919
			kurt	0.026	0.029	-0.101	0.221	0.902
			skew and kurt	0.026	0.027	-0.086	0.150	0.951
	Raw returns	**Calls**	skew	0.213	0.505	-1.122	4.872	0.421
			kurt	0.209	0.511	-1.171	4.565	0.408
			skew and kurt	0.206	0.489	-0.975	4.419	0.422
		Puts	skew	0.143	0.571	-1.421	4.664	0.251
			kurt	0.154	0.570	-1.280	4.144	0.270
			skew and kurt	0.149	0.549	-1.335	4.487	0.272

Table VII: Returns after accounting for transaction costs

This table reports means and t-statistics for monthly returns from our long-short trading strategies when considering transaction costs. In our baseline analysis we assume investors to buy and sell options at their mid-point price (MidP). This table reports results when considering ratios of effective to quoted spreads of 50 %, 75 %, and 100 %, see Section 4.2 for details. Note that we also take transaction costs for the underlying stocks into account as all equity options in our sample have to be delivered physically at option exercise.
Long-short portfolios are formed based on the lowest and highest deciles of options' IV *conditional* on their RV (for ATM options) or conditional on RV *and* moneyness (for options of arbitrary moneyness), see Section 3.3 for details. We report results from the delta-hedged and raw option strategies separately for calls and puts. Additionally, we consider the sample of ATM options as well as the sample of arbitrary moneyness, see Sections 3.1 and 3.4 for further information. The sample period is January 1996 to June 2019.

			MidP	**50 %**	**75 %**	**100 %**
ATM	**Delta-hedged**	Calls	0.020 (14.131)	0.003 (1.874)	-0.006 (-4.707)	-0.015 (-11.218)
		Puts	0.017 (13.364)	0.002 (1.772)	-0.005 (-3.815)	-0.013 (-9.350)
	Raw	Calls	0.213 (8.776)	0.021 (0.952)	-0.082 (-3.642)	-0.277 (-9.854)
		Puts	0.099 (3.565)	-0.032 (-1.296)	-0.109 (-4.509)	-0.252 (-8.264)
Arbitrary moneyness	**Delta-hedged**	Calls	0.024 (13.700)	0.002 (1.078)	-0.010 (-5.810)	-0.021 (-12.638)
		Puts	0.025 (14.178)	0.006 (3.286)	-0.004 (-2.136)	-0.013 (-7.678)
	Raw	Calls	0.201 (6.553)	0.026 (0.972)	-0.072 (-2.727)	-0.238 (-8.339)
		Puts	0.131 (3.644)	-0.010 (-0.348)	-0.091 (-3.345)	-0.225 (-7.918)

Table VIII: Returns from different estimators

This table provides information on the monthly returns from long-short portfolios formed based on the conditional 10 % and 90 % quantiles of options' IV conditional on RV (for ATM options) and conditional on RV *and* moneyness (for options of arbitrary moneyness). Details on the trading strategies are provided in Section 3.4.
Conditional quantiles are computed based on three different estimators, see Section 2.3 for details. We compare our baseline estimator (leveraging estimator) to the copula estimator by Kraus and Czado (2017) and the quantization estimator by Charlier et al. (2015b). For illustrative purposes we also include results from a conditional double-sort (ATM) and triple-sort (arbitrary moneyness) into 10 portfolios for each variable. We report the mean, standard deviation (sd), and Sharpe ratio (SR) of monthly returns for the delta-hedged and raw option strategy separately for calls and puts. The strategies are evaluated for ATM options and for options of arbitrary moneyness between January 1996 and June 2019, see Section 3.1 for information on the sample construction.

				Leveraging	**Copula**	**Quantization**	**Portfolio sort**
ATM	Delta-hedged returns	Calls	mean	0.020	0.021	0.019	0.018
			sd	0.023	0.025	0.022	0.022
			SR	0.842	0.836	0.874	0.835
		Puts	mean	0.017	0.018	0.016	0.018
			sd	0.021	0.022	0.020	0.022
			SR	0.796	0.813	0.808	0.813
	Raw returns	Calls	mean	0.213	0.223	0.208	0.201
			sd	0.407	0.447	0.408	0.414
			SR	0.523	0.499	0.510	0.487
		Puts	mean	0.099	0.113	0.094	0.113
			sd	0.467	0.483	0.446	0.483
			SR	0.212	0.235	0.210	0.235
Arbitrary moneyness	Delta-hedged returns	Calls	mean	0.024	0.026	0.023	0.023
			sd	0.030	0.032	0.030	0.028
			SR	0.816	0.796	0.786	0.803
		Puts	mean	0.025	0.023	0.028	0.023
			sd	0.029	0.032	0.030	0.032
			SR	0.844	0.730	0.955	0.730
	Raw returns	Calls	mean	0.201	0.221	0.190	0.197
			sd	0.515	0.556	0.510	0.493
			SR	0.390	0.398	0.372	0.399
		Puts	mean	0.131	0.147	0.198	0.147
			sd	0.603	0.611	0.636	0.611
			SR	0.217	0.241	0.312	0.241

Table IX: Including options on dividend-paying stocks

In our baseline analyses we excluded all options with an ex-dividend date during the remaining life of the contract. This was done to reduce the impact of early exercise. In this table we present summary statistics (mean, standard deviation (sd), minimum (min), maximum (max) and Sharpe ratio (SR)) of monthly returns that we obtain when we do *not* exclude options on dividend-paying stocks.
Long-short portfolios are formed based on the conditional 10 % and 90 % quantiles of options' IV conditional on RV (for ATM options) or conditional on RV *and* moneyness (for the option sample of arbitrary moneyness). We report returns from the delta-hedged and the raw option strategy separately for calls and puts. Details on the trading strategies can be found in Section 3.4. The ATM option sample consists of 326,224 calls and 299,924 puts while the sample of arbitrary moneyness comprises 2,682,492 calls and 2,145,435 puts. The sample period is January 1996 to June 2019.

			mean	**sd**	**min**	**max**	**SR**
ATM	**Delta-hedged**	Calls	0.019	0.022	-0.045	0.128	0.873
		Puts	0.016	0.019	-0.042	0.075	0.830
	Raw returns	Calls	0.205	0.419	-0.762	2.641	0.490
		Puts	0.110	0.439	-0.855	2.649	0.251
Arbitrary moneyness	**Delta-hedged**	Calls	0.023	0.028	-0.070	0.271	0.816
		Puts	0.023	0.027	-0.081	0.225	0.858
	Raw returns	Calls	0.193	0.487	-1.131	4.432	0.395
		Puts	0.139	0.572	-1.542	4.313	0.242

Figure 1: Conditional portfolio sort

This panel illustrates a conditional double-sort and derived conditional quantile curves based on 200 simulated observations according to $Y = \frac{1}{2} \cdot X + \frac{1}{10} \cdot X \cdot \epsilon$ where ϵ and X follow a standard normal and a $Beta(5, 5)$ distribution, respectively. We first sort observations into 5 bins based on their x-values (blue lines according to the 20 %, 40 %, 60 %, and 80 % (unconditional) quantile of all x-values). Subsequently, in each bin we further sort on the y-value. The dashed black and red lines mark the 20 % and 80 % quantile of y conditional on each of the 5 bins. The grey lines mark the 40 %, 60 %, and 80 % quantiles. Finally, we derive a long-short portfolio controlling for x by choosing observations in the lower quintile in each of the 5 bins (i.e., all observations below the black dashed line) for the long and observations in the upper quintile (i.e., all observations above the red dashed line) for the short portfolio. Observations that are not in the shaded areas do not enter into the long-short portfolio. For more details we refer to Section 2.2.
Connecting the red dashed lines inside each of the 5 bins yields a step function that approximates the true 80 % conditional quantile curve of Y conditional on X. The analogue is true for the black dashed lines, see Figure 2 for an illustration.

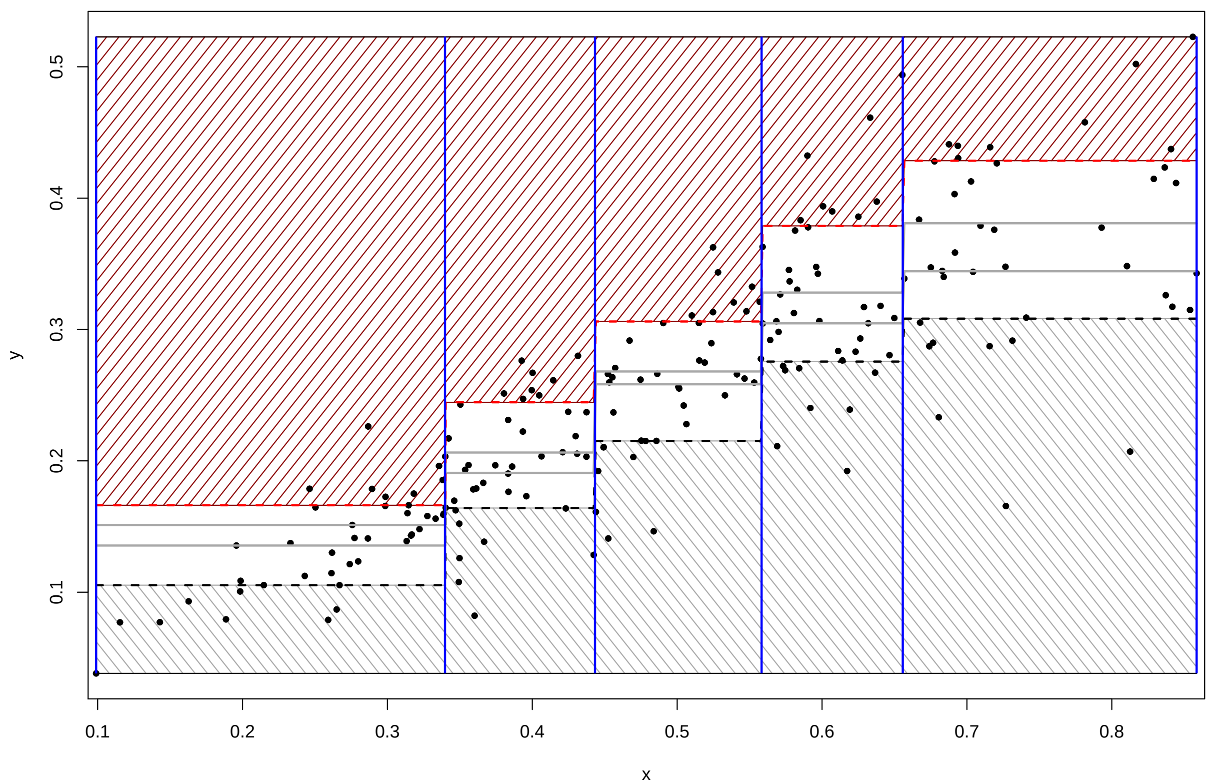


Figure 2: double-sort vs. conditional quantile curves

This figure compares the 20 % (black dashed line) and 80 % quantiles of y given x according to the double-sort from Figure 2 to the *true* conditional quantile curves of y given x. Note that the illustrative example is based on simulated data for which the true conditional quantile curves are known. For example, the 20 % conditional quantile function is linear and given by $q_{20\%}(x) = (0.5 + 0.1 \cdot q_{20\%}^{norm}) \cdot x$ where $q_{20\%}^{norm}$ denotes the (unconditional) 20 % quantile of the standard normal distribution. This figure illustrates that conditional portfolio sorts can be interpreted as a nonparametric estimator of conditional quantiles.
Based on the conditional portfolio sort, we want to form a long-short portfolio that is long (short) in low (high) values of y while controlling for x. However, as the step functions obtained via conditional portfolio sorts only provide a rough approximation to the true quantile curves, a long-short portfolio that is long (short) in securities corresponding to the observations below (above) the black (red) dashed line cannot properly control for x, see Figure 3.

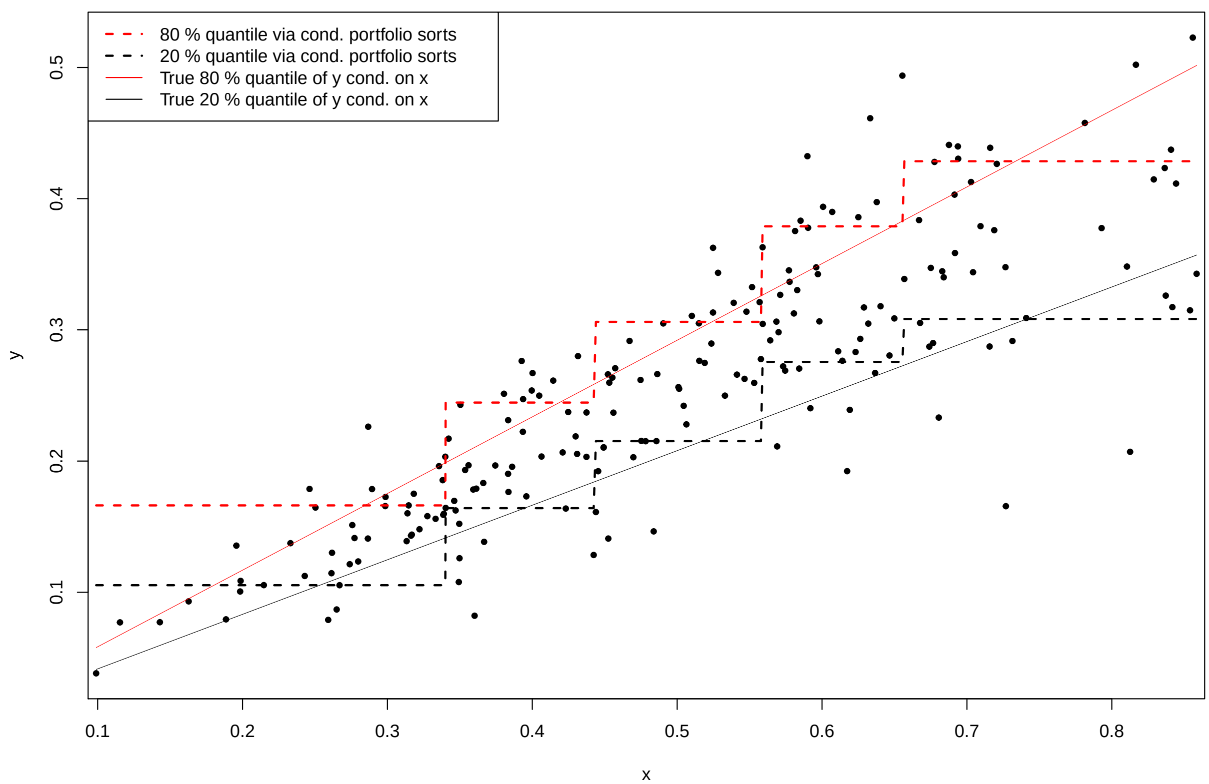


Figure 3: Long-short portfolio derived from a double-sort to control for x

This Figure illustrates limitations on forming a long-short portfolio via double-sorts that is long (short) in low (high) values of y while controlling for x. The black and red dashed lines are derived from a conditional portfolio sort (first on x, then on y), see Figure 1. Securities corresponding to the observations below (above) the black (red) dashed line enter into the long (short) portfolio. Since the step functions only provide a rough approximation to the true conditional quantile functions of y given x (see Figure 2) there are limitations in controlling for x. The black and red dashes on the x-axis correspond to the observations in the long and short portfolio, respectively, and illustrate that there remain systematic differences in x between the two portfolios. This is clear from mostly black dashes for low and red dashes for high x-values with clusters of black and red dashes in between. If instead portfolios were formed based on the true conditional quantile curves from Figure 2 the black and red dashes would be perfectly mixed along the x-axis.

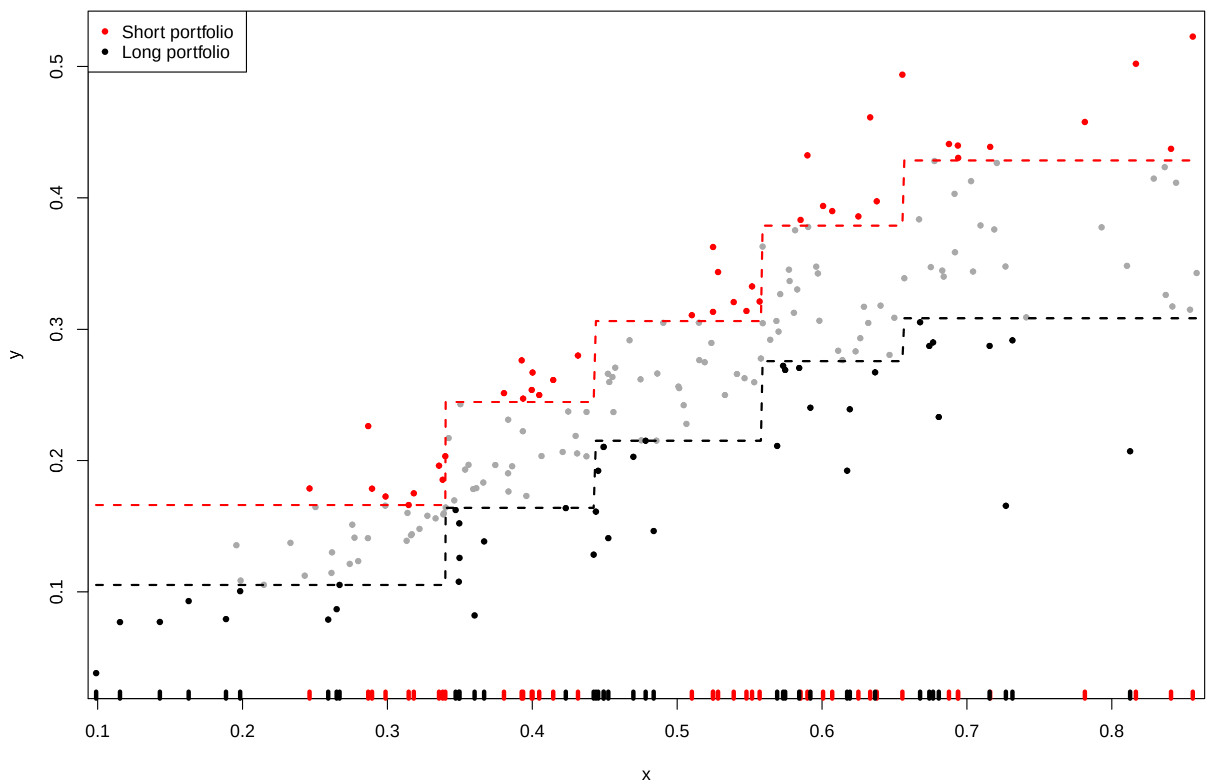


Figure 4: Sorting options on the *log* difference of RV and IV

This Figure shows the value pairs (RV, IV) of all ATM call options in our sample in January 2010. According to Goyal and Saretto (2009) options are sorted on the log-differences of RV and IV into decile portfolios. Options in decile 1 (red data points) constitute the short portfolio, while options in decile 10 enter into the long portfolio (black data points).

Sorting on the log-differences (unconditionally) and choosing the highest and lowest 10 % of the value pairs (RV, IV) is equivalent to requiring $IV_i \leq RV_i \cdot e^{-q_{90\%}}$ in the long portfolio and $IV_i \geq RV_i \cdot e^{-q_{10\%}}$ in the short portfolio, where $q_{10\%}$ and $q_{90\%}$ denote the (unconditional) empirical 10 % and 90 % quantiles of the log-differences of RV and IV, respectively. This translates into the black and red dash-dotted straight lines in the scatter plot.

The red and black dashes on the x-axis correspond to the value pairs (RV, IV) in the long and short portfolio and illustrate that there are systematic differences in RV between the two portfolios. The faint dashed lines correspond to the nonparametric 10 % (black) and 90 % (red) quantile curves of IV *conditional* on RV (see also Figure 6).

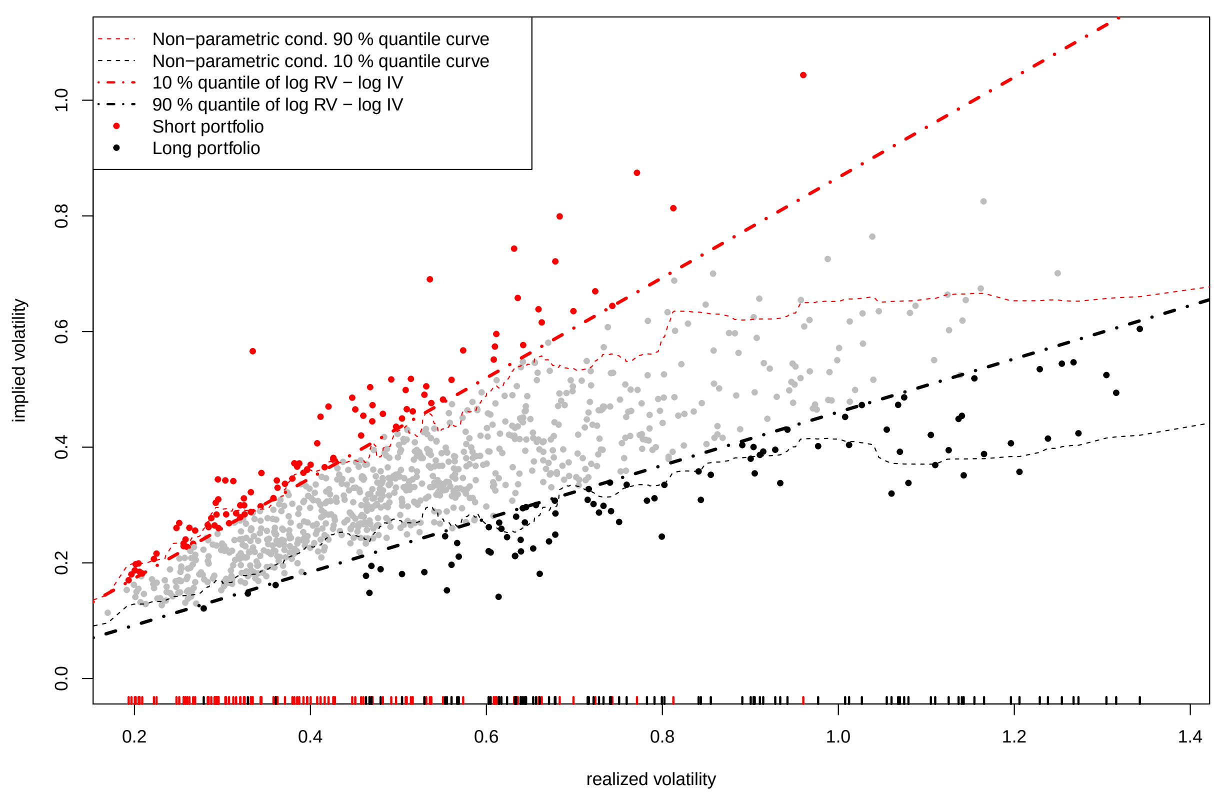


Figure 5: Sorting options on the difference of RV and IV

This Figure shows the value pairs (RV, IV) of all ATM call options in our sample in January 2010. Cao and Han (2013) define the VRP as the difference of RV and IV. We form a long-short portfolio where options in decile 1 according to the difference of RV and IV (red data points) constitute the short portfolio, while options in decile 10 enter into the long portfolio (black data points). Note that the only difference to Figure 4 is that we are sorting on the (simple) difference instead of the *log* difference of RV and IV.

Sorting on the difference (unconditionally) and choosing the highest and lowest 10 % of the value pairs is equivalent to requiring $IV_i \leq RV_i - q_{90\%}$ in the long portfolio and $IV_i \geq RV_i - q_{10\%}$ in the short portfolio, where $q_{10\%}$ and $q_{90\%}$ denote the (unconditional) empirical 10 % and 90 % quantiles of the differences of RV and IV, respectively. This translates into the black and red dash-dotted straight lines in the scatter plot.

The red and black dashes on the x-axis correspond to the value pairs (RV, IV) in the long and short portfolio and illustrate that there are systematic differences in RV between the two portfolios. The faint dashed lines correspond to the nonparametric 10 % (black) and 90 % (red) quantile curves of IV *conditional* on RV (see also Figure 6).

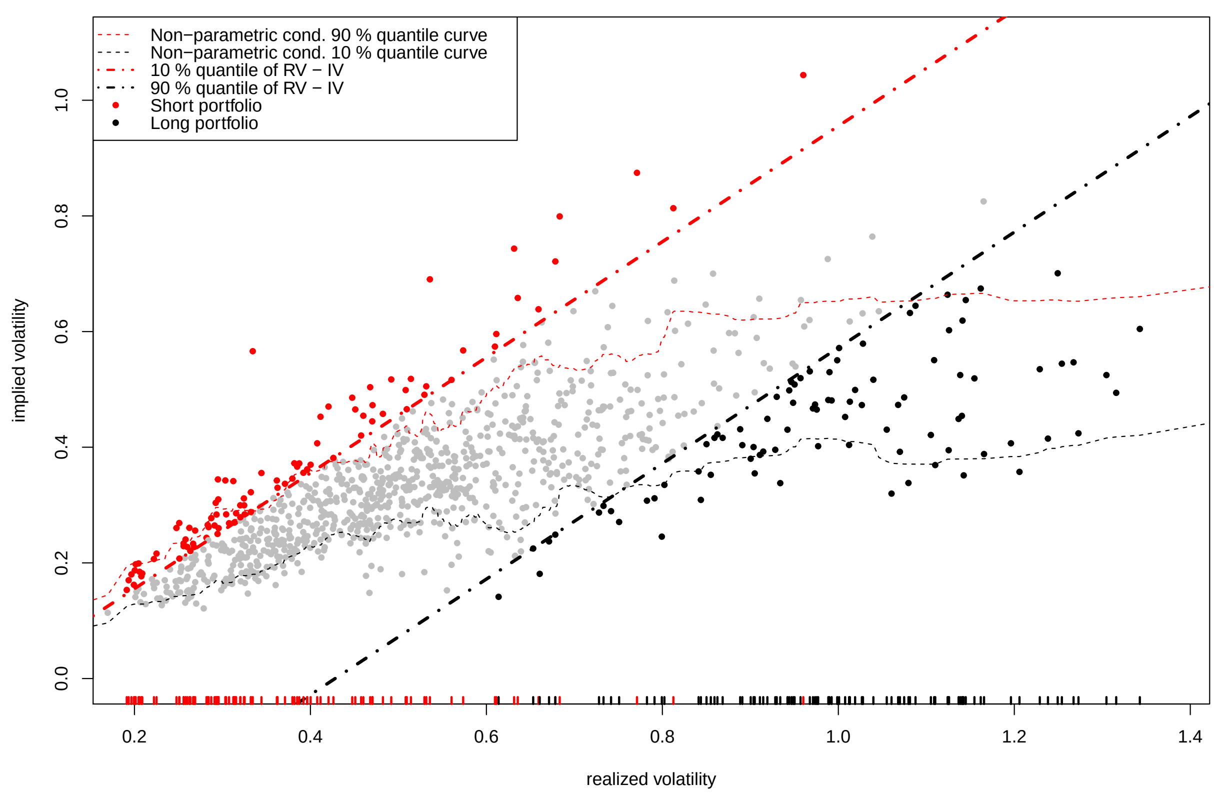
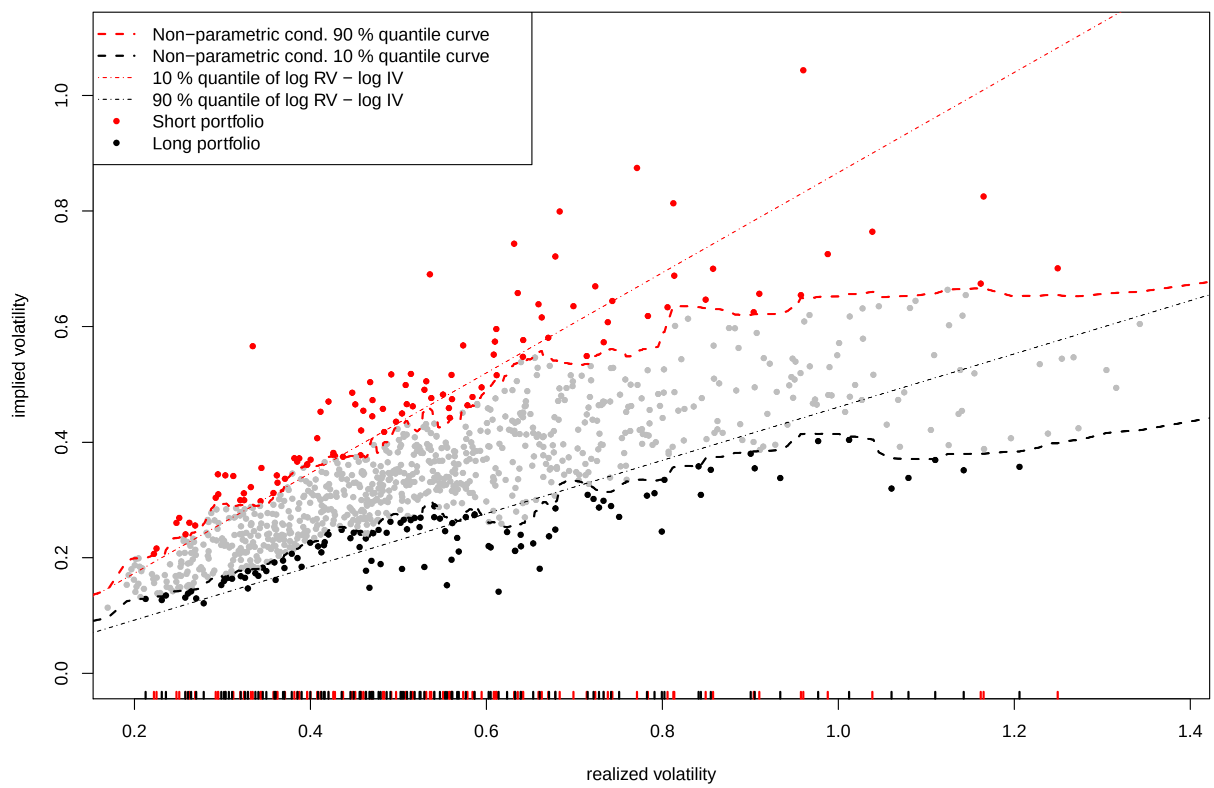


Figure 6: Sorting on IV conditional on RV

This Figure shows the value pairs (RV, IV) of all ATM call options in our sample in January 2010. Options are sorted according to IV *conditional* on RV into decile portfolios. Options in decile 10 (red data points) constitute the short portfolio, while options in decile 1 enter into the long portfolio (black data points). That is, the long (short) portfolio is constituted of the 10 % of the options with the highest (lowest) IV conditional on RV.
More exactly, for option i in the long portfolio we require $F(IV_i|RV_i) \leq 10$ % while options in the short portfolio have to fulfill $F(IV_i|RV_i) \geq 90$ %, where $F(\cdot|RV_i)$ denotes the conditional cumulative distribution function of implied volatility given a specific level of realized volatility (RV_i). This translates into the red and black dashed conditional quantile curves. The faint black and red dash-dotted lines correspond to the log-difference of RV and IV and are included for comparison.
The red and black dashes on the x-axis correspond to the value pairs (RV, IV) in the long and short portfolio and illustrate that there are on average no obvious differences in RV between the long and short portfolio as opposed to Figure 4.



The Centre for Responsible Banking & Finance
CRBF Working Paper Series
School of Management, University of St Andrews
The Gateway, North Haugh,
St Andrews, Fife,
KY16 9RJ.
Scotland, United Kingdom
http://www.st-andrews.ac.uk/business/rbf/

Recent CRBF Working papers published in this Series

Third Quarter | 2021

21-011 **Nhan Le, Duc Duy Nguyen, Vathunyoo Sila:** Does Shareholder Litigation Affect the Corporate Information Environment?

21-010 **Hiep N. Luu, Linh H. Nguyen, John O.S. Wilson:** Organizational Culture, Competition and Bank Loan Loss Provisions.

21-009 **Alessio Reghezza, Yener Altunbas, David Marques-Ibanez, Costanza Rodriguez d'Acri, Martina Spaggiari:** Do Banks Fuel Climate Change?

Second Quarter | 2021

21-008 **Dolores Gutiérrez-Mora and Daniel Oto-Peralías:** Gendered Cities: Studying Urban Gender Bias through Street Names.

21-007 **Linh H. Nguyen, John O.S. Wilson, Tuan Q. Le, Hiep N. Luu, Vinh X. Vo, Tram-Anh Nguyen:** Deposit Insurance and Credit Union Lending.

21-006 **Michael Koetter, Thomas Krause, Eleonora Sfrappini, Lena Tonzer:** Completing the European Banking Union: Capital Cost Consequences for Credit Providers and Corporate Borrowers.

21-005 **Marcel Lukas and Ray Charles "Chuck" Howard:** The Influence of Budgets on Consumer Spending.

First Quarter | 2021

21-004 **Ross Brown and Marc Cowling:** The Geographical Impact of the Covid-19 Crisis for Firms and Jobs: Evidence from the 100 Largest Cities and Towns of the UK.

21-003 **Javier Gomez-Biscarri, Germán López-Espinosa, Andrés Mesa Toro:** Drivers of Depositor Discipline in Credit Unions.