

Ulrich, Maxim; Walther, Simon

**Article — Published Version**

## Option-implied information: What's the vol surface got to do with it?

Review of Derivatives Research

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Ulrich, Maxim; Walther, Simon (2020) : Option-implied information: What's the vol surface got to do with it?, Review of Derivatives Research, ISSN 1573-7144, Springer US, New York, NY, Vol. 23, Iss. 3, pp. 323-355,  
<https://doi.org/10.1007/s11147-020-09166-0>

This Version is available at:

<https://hdl.handle.net/10419/288886>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Option-implied information: What's the vol surface got to do with it?

Maxim Ulrich<sup>1</sup> · Simon Walther<sup>2</sup> 

Published online: 7 May 2020  
© The Author(s) 2020

## Abstract

We find that option-implied information such as forward-looking variance, skewness and the variance risk premium are sensitive to the way the volatility surface is constructed. For some state-of-the-art volatility surfaces, the differences are economically surprisingly large and lead to systematic biases, especially for out-of-the-money put options. Estimates for risk-neutral variance differ across volatility surfaces by more than 10% on average, leading to variance risk premium estimates that differ by 60% on average. The variations are even larger for risk-neutral skewness. To overcome this problem, we propose a volatility surface that is built with a one-dimensional kernel regression. We assess its statistical accuracy relative to existing state-of-the-art parametric, semi- and non-parametric volatility surfaces by means of leave-one-out cross-validation, including the volatility surface of OptionMetrics. Based on 14 years of end-of-day and intraday S&P 500 and Euro Stoxx 50 option data we conclude that the proposed one-dimensional kernel regression represents option market information more accurately than existing approaches of the literature.

**Keywords** Option-implied · Risk-neutral variance · Risk-neutral density · Tail risk · Option standardization · Interpolation

**JEL Classification** G13 · G17 · C14

---

We are thankful for comments from Marliese Uhrig-Homburg, Jonathan Wright, Melanie Schienle and Philipp Schuster, seminar participants at the KIT and two anonymous referees. Special thanks to Thorsten Luedecke from the Karlsruhe Capital Market Database and to Deutsche Boerse for providing access to the Euro Stoxx 50 options. Simon Walther gratefully acknowledges financial support by the Konrad-Adenauer-Stiftung.

---

✉ Simon Walther  
simon.walther@kit.edu  
Maxim Ulrich  
maxim.ulrich@kit.edu

<sup>1</sup> Karlsruhe Institute of Technology (KIT), Bluecherstr. 17, E008, 76185 Karlsruhe, Germany

<sup>2</sup> Karlsruhe Institute of Technology (KIT), Bluecherstr. 17, E004, 76185 Karlsruhe, Germany

## 1 Introduction

Many popular option-implied metrics such as risk-neutral variance, skewness and the variance risk premium are calculated based on an estimate of the option-implied volatility surface. We document in this paper that the method for constructing the volatility surface affects these standard option-implied quantities. Our findings hold more generally for any quantity that is extracted from the aggregation of option prices along the strike range.

State-of-the-art methodologies such as the semi-parametric spline interpolation<sup>1</sup> of Figlewski (2008) or the three-dimensional kernel regression of OptionMetrics (2016) produce surprisingly large differences in standard option-implied quantities. In our sample for S&P 500 options (2004–2017), Bakshi et al. (2003) risk-neutral variance for medium-term maturities, computed with exactly the same procedure, but based on the volatility surface from the interpolation scheme of Figlewski (2008) or OptionMetrics (2016), differs in relative terms by more than 10% on average. The 1-month-ahead variance risk premium varies across both volatility surfaces by a relative margin of on average 60%. Differences are even more troublesome for risk-neutral skewness, where we document relative differences in the order of 200% and more.

The key question to ask is which volatility surface represents market information most accurately? As information is extracted from option prices by means of deterministic manipulations of the observed portions of the volatility surface, it is natural that it is the most accurate volatility surface that also reprices options most accurately. We therefore perform a detailed empirical investigation to understand which volatility surface captures market information most accurately. Our test incorporates the semi-parametric spline interpolation (Figlewski 2008), a three-dimensional non-parametric kernel regression (OptionMetrics 2016), and the parametric Gram–Charlier expansion (Beber and Brandt 2006). In addition, we also propose an one-dimensional non-parametric kernel regression. We compare the statistical accuracy of these four volatility surfaces by means of leave-one-out cross-validation root mean squared errors (RMSEs) and mean absolute errors (MAE). Calculating the average integrated squared second derivative of the respective implied volatility smiles allows us to identify differences in smoothness. Our tests expand across two dimensions: (i) options on the S&P 500 and on the Euro Stoxx 50 and (ii) with an end-of-day and intraday frequency. The time span of the analysis is 2004–2017 for US data and 2002–2017 for European data.

Our main findings are as follows: First, the one-dimensional kernel regression generates the most accurate volatility surface for S&P 500 options by means of the lowest leave-one-out cross-validation RMSE and MAE. This holds for intraday and for end-of-day data.

Second, for the end-of-day analysis of S&P 500 options, the spline-based volatility surface turns out to be the second best, with a RMSE (MAE) that is on average 128% (91%) higher than the RMSE of the one-dimensional kernel regression surface. The

---

<sup>1</sup> The spline interpolation represents an implied volatility smile parametrically. However, there is no explicit parametric form of the risk-neutral density (RND), that the volatility smile implies. For this reason, we consider this methodology, and more broadly the class of parametric implied volatility models that do not allow to pin down the RND parametrically, as semi-parametric.

three-dimensional kernel regression produces a RMSE that is more than 5 times larger than the RMSE of the one-dimensional kernel regression and an 11 times higher MAE. The Gram–Charlier volatility surface produces the largest RMSEs (MAEs), on average over 9 (23) times larger than the RMSE (MAE) of the best performing volatility surface. The results for Euro Stoxx 50 options confirm our general findings, though the one-dimensional kernel regression and the spline interpolation appear to perform roughly at par here.

Third, state-of-the-art volatility surfaces turn out to be less accurate than the one-dimensional kernel regression volatility surface because these surfaces do not accurately capture market information in the thinly traded tails of the volatility surface. The volatility surface based on spline interpolation shows weakness in capturing the left tail of short-term options. The three-dimensional kernel regression shows severe shortcomings in capturing the left tail of short-, medium- and long-term options. The method of the Gram–Charlier expansion partially captures the at-the-money region of the volatility smiles well, but shows otherwise shortcomings in capturing the left and right tail of options across all maturities.

Fourth, for the intraday analysis we find that the most accurate volatility surface is again the one that is generated by the one-dimensional kernel regression, if data is very scarce. However, the indication about the most accurate method depends on the number of trades that are used to construct the implied volatility surface and diverges between the error measures. While the one-dimensional kernel regression continues to lead in most set-ups with respect to the RMSE, the spline interpolation quickly shows lower MAEs than all other methods as the number of trades grows. This finding can be rationalized by occasional over-fitting in the spline method, which occurs less often if more trades are available.

Our analysis concludes that option-implied information can differ substantially across volatility surfaces. The three-dimensional kernel regression of OptionMetrics (2016) underpredicts option-implied tail risk at the end-of-day data frequency, which translates into systematic biases in risk-neutral skewness and variance. The Gram–Charlier expansion is not flexible enough to closely track observed option prices. The one-dimensional kernel regression appears to produce the most accurate volatility surface for the end-of-day use-case and if data is very scarce. Yet, using the spline interpolation might be beneficial if an intermediate amount of implied volatility observations is available. In that case, the over-fitting tendency in the spline interpolation is already dampened enough to produce the lowest MAEs, although the performance spread to the one-dimensional kernel regression is not large.

Our research study adds to the growing empirical finance literature that exploits option-implied information. By now, this literature is too vast to be reviewed here in detail. Hence, we cannot give credit to all studies and have to leave out important contributions. However, we discuss a selection of recent studies and focus on how these have constructed the option-implied volatility surface.

A large and diverse amount of research studies work with the volatility surface of OptionMetrics (2016): For example, Buss and Vilkov (2012) estimate option-implied correlations and CAPM betas and find that higher option-implied betas go along with higher average returns. Chang et al. (2012) predict a stock's beta based on Bakshi et al. (2003) implied volatility and skewness estimates. Martin and Wagner (2019) extract

a measure for predicting the risk premium for each stock from its associated implied volatility surface. Christoffersen et al. (2017) propose a parametric option pricing model where illiquidity is a driver of the jump intensity of the underlying and can thus explain time variation in the implied volatility surface. Du and Kapadia (2012) compare the information content of the VIX and the Bakshi et al. (2003) implied variance measure and construct a tail risk index from its difference. Hofmann and Uhrig-Homburg (2018) use the wedge between implied volatility observations and respective OptionMetrics (2016) estimates to construct a limits of arbitrage measure.

Driessen et al. (2009) compare portfolios of single stock options with options on the index and find evidence for a substantial option-implied correlation risk premium. Bollerslev and Todorov (2011) introduce various risk measures for realized and expected continuous risk and jump risk under the empirical and risk-neutral probability measure and find that a significant portion of the equity risk premium is compensation for jump risk. The option-implied volatility surface in these innovative studies is constructed based on end-of-day closing prices and based on a version of the spline interpolation methodology that we use in this paper.

Martin (2017) shows that options contain information about the lower bound of the underlying's expected return. Schneider and Trojani (2015) construct tradable option-implied strategies for higher moments. These studies do not interpolate the option-implied volatility surface, but work with observed option-implied volatilities.

Wright (2016) adopts the multi-dimensional kernel regression of Aït-Sahalia and Lo (1998) to construct a monthly option-implied volatility surface of real interest rates. The author pools all end-of-day volatilities (roughly 25 per day and maturity) within 1 month to stabilize the procedure and obtain a sufficiently smooth volatility surface. Swanson (2016) highlights that pooling prevents that study from working at a higher frequency. Moreover, Swanson (2016) suggests to use the spline methodology to construct daily option-implied volatility surfaces. As evidence in Bliss and Panigirtzoglou (2002) suggests, that method works well for 10 or more option prices per day and maturity. In order to stay in a non-parametric framework while still obtaining smoothness, Jackwerth and Rubinstein (1996) and Jackwerth (2000) propose a method that directly fits the volatility surface to observed data by minimizing the squared fitting error while at the same time maximizing the smoothness of the surface. Our research study contributes to this discussion, as we formally compare the statistical accuracy and smoothness of several state-of-the-art volatility surfaces across different frequencies and different currency zones. For our tests, we select two kernel regression approaches as representatives for the non-parametric class of volatility surfaces. These methods do not explicitly optimize for smoothness, thus enabling us to perform a fair evaluation between methods, since neither the kernel regression of OptionMetrics (2016) nor the spline method of Figlewski (2008) or the Gram–Charlier expansion features an explicit smoothness treatment. As we document, especially the one-dimensional kernel regression is still well capable of producing smooth volatility surfaces.

A large body of the literature studies parametric representations of the implied volatility surface. An early example is Schönbucher (1999), who assumes a diffusion

process for the evolution of implied volatility.<sup>2</sup> He derives restrictions on the parameters to ensure arbitrage-freeness and to prevent the potential model-implied emergence of bubbles in implied volatility as the maturity decreases. The latter leads to a constraint that can essentially serve as a model for the volatility smile at a given maturity. Based on a two-dimensional diffusion for the forward price of an asset and its volatility, Hagan et al. (2002) derive a closed-form solution for a parametric volatility smile in their *SABR* model. The model is completely specified by four parameters, which essentially describe the level, skew and curvedness of the volatility smile. However, as Gatheral (2006) points out, the lack of a mean reversion component in the volatility diffusion makes it only applicable to short-maturity options. In his *SVI* model, Gatheral (2004) assumes a parametric function for the volatility smile that is similar to the representation of Schönbucher (1999), but features an additional parameter to locate the volatility smile across the strike range. By construction, the *SVI* model assumes that the implied volatility smile becomes approximately linear in the tails. This assumption is not always fulfilled in the data, which has triggered the development of generalized versions of the *SVI* model that allow concavity, for example Zhao and Hodges (2013) or Damghani and Kos (2013). Damghani (2015) further elaborates on this work in the context of the FX option market. His *IVP* model features an explicit treatment of the bid-ask spreads in option prices and allows to incorporate liquidity factors. By adding a maturity interpolation scheme, he is able to capture the whole volatility surface with a dramatic reduction in parameters that need to be estimated. Figlewski (2008) follows a different approach: He argues to use a 4th-order smoothing spline with one knot point at-the-money to model the volatility smile at each maturity separately. The higher amount of free parameters is accepted to reach a higher accuracy while still smoothing out noise. In our study, we use the method of Figlewski (2008) as representative of the class of parametric models of the implied volatility surface.

Survey and methodology papers that are closest to ours are Jackwerth (1999), Jondeau and Rockinger (2000), Bliss and Panigirtzoglou (2002) and Bahaludin and Abdullah (2017). Relative to these studies, our empirical assessment covers a much longer time span (14 years), US and EU equity option markets and distinguishes between characteristics of the input option data sets.

## 2 Data

We use data for two of the most actively traded equity option contracts: options on the S&P 500 and options on the Euro Stoxx 50. The former is traded at the CBOE while the latter is traded at the Eurex. The S&P 500 and the Euro Stoxx 50 do both stand for the Blue Chip stocks of their respective currency zone. Both option contracts are of European exercise style.

Historical data for S&P 500 options come from the CBOE Livevol data shop. We collect both end-of-day bid-ask prices as well as intraday transaction prices. For the

<sup>2</sup> Assuming a diffusion process for implied volatility is different from a diffusion process for the volatility of the underlying, as in Heston (1993). In the model of Schönbucher (1999), movements in implied volatility are correlated with movements in the underlying, though explicitly feature components that are independent of the underlying asset.

**Table 1** Data summary

	EOD		Trades	
	S&P 500	Euro Stoxx 50	S&P 500	Euro Stoxx 50
Total number of price observations	9,246,926	6,877,495	43,749,042	16,921,911
Avg number of maturities per day	17.21	16.58	16.96	12.65
Avg number of strikes per maturity	61.27	45.28	16.97	5.93
Avg number of price observations per day and maturity	157.58	104.22	740.43	356.84
Total number of short-term ATM prices	771,000	73,000	24,627,422	5,866,507
Total number of short-term left tail prices	1,326,413	274,227	4,449,056	1,956,662
Total number of short-term right tail prices	293,823	198,558	1,330,209	903,600
Total number of medium-term ATM prices	1,186,655	293,028	7,348,889	3,148,740
Total number of medium-term left tail prices	3,536,880	1,218,912	3,996,122	3,016,860
Total number of medium-term right tail prices	1,048,097	982,266	1,628,093	1,586,504
Total number of long-term ATM prices	120,389	340,796	100,016	168,532
Total number of long-term left tail prices	656,001	1,768,878	179,571	170,934
Total number of long-term right tail prices	307,668	1,727,830	89,664	103,572

This table provides some aggregated summary statistics of our option panel. 'EOD' stands for 'end-of-day' (data-rich environment), whereas the 'Trades' columns summarize the aggregated data for our intraday analysis (data-poor environment). Options with a remaining maturity of less than 30 days are considered short-term, options with 30–365 days to maturity are considered medium-term and longer maturity options are considered long-term. At-the-money (ATM) options have a moneyness between 0.95 and 1.05. Options with a moneyness below 0.95 are considered as 'left tail' and options with a moneyness above 1.05 are considered as 'right tail'

end-of-day data set, we use the mid price for our calculations, whereas we use actual trade prices for the intraday data set. The end-of-day data is from January 2004 to July 2017. The intraday data spans the period January 2004 to October 2017. The data comes with the matched bid and ask prices of the underlying at the point of time of the record.

We obtain end-of-day settlement prices and intraday transaction prices for the Euro Stoxx 50 options and the underlying Euro Stoxx 50 index from the *Karlsruher Kapitalmarktdatenbank* (KKMDB), which is hosted at the Finance Institute of the Karlsruhe Institute of Technology. The KKMDB receives its data directly from the Eurex and Stoxx. The index data comes in 15 second intervals, while the intraday option data is time stamped at the point of time of the trade. We match the option trade with the index price directly at or prior to the trade's time stamp. The end-of-day data spans January 2002 to September 2017, whereas the intraday data covers the period January 2003 to September 2017. Table 1 summarizes our option data sets in more detail.

For the risk-free rate, we use the US-Dollar and Euro OIS curves from Bloomberg, which come in discrete maturities between 1 day and 20 years and which we obtain at the daily frequency. We match each option record with the risk-free rate for its respective date. We interpolate the OIS rate linearly along the maturity dimension to match the maturity of the respective option.

We apply a number of filtering steps to ensure that only valid option prices enter the standardization. We only consider options with a price that fulfills basic no-arbitrage



considerations. More precisely, we keep a Call option price  $C$ , if  $C \geq Se^{-q\tau} - Ke^{-r\tau}$ , and a Put option price  $P$ , if  $P \geq Ke^{-r\tau} - Se^{-q\tau}$ , with the current price of the underlying  $S$ , a strike of  $K$ , maturity of  $\tau$  years, the riskfree rate  $r$  and the continuous dividend yield  $q$ . In addition, as most methods work in implied volatility space, we remove options, for which the calculation of the Black-Scholes implied volatility did not converge. Inline with Bliss and Panigirtzoglou (2002), we only keep out-of-the-money options and recover prices for in-the-money options (where needed) through put-call parity. The latter is not done for the three-dimensional kernel regression of OptionMetrics (2016), as that method explicitly works with both Call and Put prices.

Both, the S&P 500 and the Euro Stoxx 50 index, are price indices and are thus quoted ex-dividends. Filtering the raw option prices and the calculation of implied volatilities for maturity  $\tau$  requires an estimate of the continuous  $\tau$ -period ahead dividend yield. For each day, we determine the maturity-specific dividend yield as the median (across strikes) of the put-call parity implied dividend yields, using only options with the desired maturity. This estimate is used for all quote filters and implied volatility calculations of a respective day and maturity  $\tau$ . More details on the dividend yield estimation and on matching Call and Put prices for put-call parity can be found in "Appendix A".

### 3 Constructing a volatility surface

We start this section with providing a quick overview on the most common parametric, semi-parametric and non-parametric approaches for constructing an option-implied volatility surface.

One popular representative for a parametric approach is the Gram–Charlier expansion. It approximates an unknown density by starting with a Gaussian density and, similar to a Taylor expansion, iteratively adding higher-order terms to reduce deviations between the true unknown density and its previous approximation. This parametric approach captures volatility, skewness and kurtosis of the unknown risk-neutral density and allows for closed-form prices of Calls and Puts (Backus et al. 2004), and hence a semi-closed form expression for the option-implied volatility surface. A detailed explanation of this method can be found in "Appendix B.1".

Interpolating observed option-implied volatilities with polynomial splines is certainly the most popular semi-parametric approach for constructing the option-implied volatility surface. Several variations of the spline interpolation approach exist, mainly differing in the degree of the spline or in the dimension over which the interpolation runs (see, for example, Bliss and Panigirtzoglou 2002; Jiang and Tian 2005; Chang et al. 2012). We follow Figlewski (2008) and Fengler (2009) as their procedure nests many other approaches of the literature. A detailed description of that method is in "Appendix B.2".

An often applied non-parametric approach is the kernel regression. A prominent volatility surface, that is often used in financial economic research, is the kernel regression specification of OptionMetrics (2016). That data provider approximates missing values of the option-implied volatility surface (across strikes and maturity) with a three-dimensional kernel regression. The corresponding three bandwidth parameters



are fixed in OptionMetrics (2016). To keep our paper self-contained, we provide in “Appendix B.3” a detailed description of that non-parametric approach.

We now continue to present our own kernel regression specification, that combines advantages of the above mentioned methods. First, it is a non-parametric approach, which ensures flexibility. Second, it is essentially only one-dimensional, which makes it more robust and less data intensive than existing multi-dimensional kernel regression approaches. Third, our approach ensures the volatility surface to be arbitrage-free, borrowing a technique from Fengler (2009). Fourth, conditional on having sufficient data, our approach takes special care of capturing market information that is hidden in the thinly traded tails.

### 3.1 One-dimensional kernel regression with tail extrapolation

We continue here to present in detail our suggested variation of the kernel regression method that was popularized by Aït-Sahalia and Lo (1998). Aït-Sahalia and Lo (1998) propose to approximate unobserved option-implied volatilities by the following kernel regression set-up:

$$\hat{\sigma}(F_j, K_j, \tau_j) = \sum_{i=1}^N \frac{k(F_j - F_i) k(K_j - K_i) k(\tau_j - \tau_i)}{\sum_{i=1}^N k(F_j - F_i) k(K_j - K_i) k(\tau_j - \tau_i)} \times \sigma(F_i, K_i, \tau_i), \quad (1)$$

where  $N$  is the number of observed option prices that enter the kernel regression as inputs,  $\sigma(\cdot)$  is the observed Black-Scholes implied volatility and  $\hat{\sigma}$  is the interpolated Black-Scholes implied volatility for a desired tuple of strike  $K_j$ , forward price of the underlying  $F_j$  and maturity  $\tau_j$ . The kernel function  $k(x)$  has a Gaussian shape with its own bandwidth parameter  $h_x$  for each dimension, i.e.,

$$k(x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2h_x}\right)}. \quad (2)$$

We are now going to present refinements that distinguish our approach from the tested application in Aït-Sahalia and Lo (1998). First, we reduce the input dimension by one unit, as we combine the underlying forward price with the strike level. More precisely, we define the observed moneyness measure as  $m_i = \frac{K_i}{F_i}$  and apply the kernel regression to  $m_j - m_i$  instead of  $F_j - F_i$  and  $K_j - K_i$ .<sup>3</sup> In most cases, this step can be regarded as a mere technical rescaling of the strike axis, which does not affect the interpolation accuracy of the technique. In our end-of-day set-up, we only use the price observations of a single day to construct the respective day’s implied volatility surface, which all have the same underlying price as they are observed at the same point of time. Similarly, in the intraday set-ups, we construct the implied volatility surfaces based on subsequent option trade observations. In the vast majority of cases, the change in the underlying price between two option trades is very small,

<sup>3</sup> Although this adjustment has already been proposed by Aït-Sahalia and Lo (1998), their subsequent empirical analysis works with  $K$  and  $F$ , separately.

such that the underlying forward price can be considered approximately constant for the employed set of option trade observations.

The second refinement of the Aït-Sahalia and Lo (1998) approach is that we remove the time to maturity dimension,  $\tau$ , from the kernel regression. This means that the kernel regression is purely used to approximate the option-implied volatility surface in the moneyness dimension.<sup>4</sup> Mathematically, this means we use the following kernel weighting for the moneyness dimension:

$$\hat{\sigma}(m_j, \tau_i) = \sum_{i=1}^{N^*} \frac{k(m_j - m_i) \sigma(m_i, \tau_i)}{\sum_{i=1}^{N^*} k(m_j - m_i)}, \quad m_i = \frac{K_i}{F_i}, \quad (4)$$

where  $N^*$  denotes the size of the set of observed option prices with maturity  $\tau_i$ . We set the bandwidth parameter for the moneyness dimension by the following rule-of-thumb. We first compute the average distance between two neighboring strikes

$$\overline{\Delta K}(\tau_i) = \frac{1}{N^* - 1} (K^{\max}(\tau_i) - K^{\min}(\tau_i)),$$

where  $K^{\max}(\tau_i)$  ( $K^{\min}(\tau_i)$ ) represents the maximum (minimum) observed strike at maturity  $\tau_i$ . We set the moneyness bandwidth to  $h_m(\tau_i) = 0.75 \overline{\Delta K}(\tau_i)$ . While the coefficient of 0.75 may appear ad-hoc, our tests reveal that the results barely change if it is set at any value within a range of [0.6, 1].

Third, far out-of-the-money options are usually not observed, but very important ingredients for capturing the tails of the option-implied volatility surface. For the analysis at the end-of-day frequency, we use a linear extrapolation scheme to increase the number of 'observed' option-implied volatilities in both tails, before applying the kernel smoother. The extrapolation works as follows: We start with a set of Black-Scholes implied volatilities  $\{\sigma_i\}_{i \in [1, N]}$  for some point of time. We split this set into subsets  $\{\sigma_i^\tau\}_{i \in [1, N]}$  with all elements in a subset having the same time to maturity  $\tau$ . We assume that the elements of  $\{\sigma_i^\tau\}_{i \in [1, N]}$  are sorted ascendingly by their moneyness  $m_i$ .

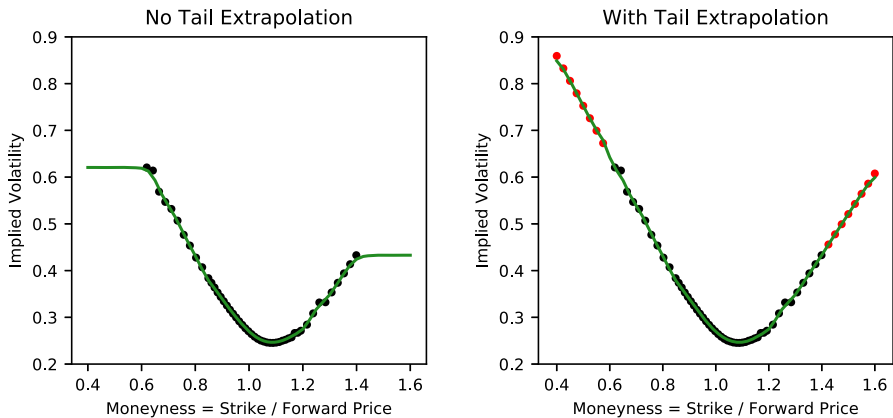
For the left tail extrapolation, we use the first 5 implied volatilities of each subset. The number of 5 observations is a trade-off between stability on the one hand and including only the most extreme tail observations on the other hand. Unreported results show that our findings are robust to variations in that considered number of observations. On these 5 observations, we estimate the following linear regression using OLS:

$$\sigma_i^\tau = \alpha + \beta m_i + \epsilon_i, \quad \epsilon_i \sim i.i.d.$$

<sup>4</sup> If needed, we recommend a linear interpolation along the log maturity axis between two observed maturities:

$$\hat{\sigma}(m_j, \tau_j) = \hat{\sigma}(m_j, \tau_l) + \frac{\log \tau_j - \log \tau_l}{\log \tau_h - \log \tau_l} (\hat{\sigma}(m_j, \tau_h) - \hat{\sigma}(m_j, \tau_l)), \quad (3)$$

where  $\tau_h$  ( $\tau_l$ ) represents the next longer (shorter) observed option maturity, relative to the targeted time to maturity  $\tau_j$ .



**Fig. 1** Effect of tail extrapolation on kernel regression. This figure visualizes for a particular day of the sample, May 22nd, 2012, the effect of the linear tail extrapolation on the option-implied volatility smile for Euro Stoxx 50 options with 24 days to maturity. The left panel shows the kernel regression without tail extrapolation, the right panel includes the tail extrapolation. The dots mark observed (and artificial) implied volatility observations, the line depicts the interpolation

**Table 2** Cross-validation errors for refinement steps of the kernel regression

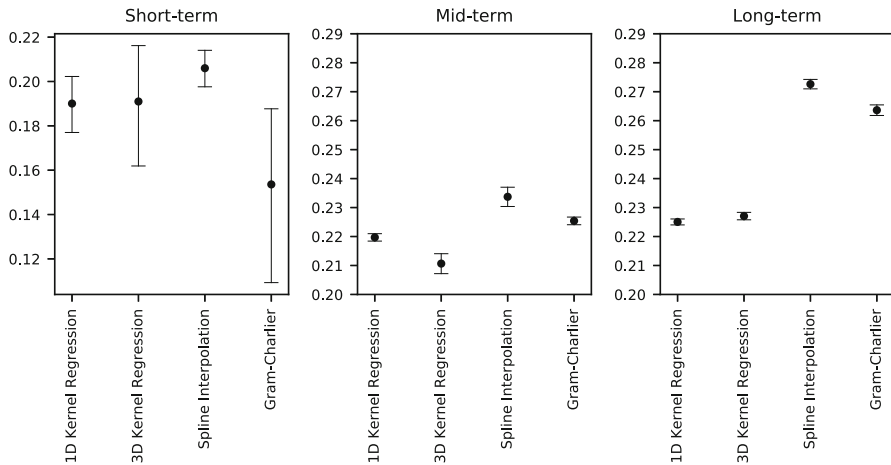
Step	S&P 500		Euro Stoxx 50	
	RMSE	MAE	RMSE	MAE
Initial	0.0479	0.0139	0.1964	0.0518
No maturity dimension	0.0109	0.0023	0.0156	0.0024
Tail extrapolation	0.0091	0.0019	0.0101	0.0021
No-arbitrage enforcement	0.0092	0.0019	0.100	0.0019

This table shows the leave-one-out cross-validation root mean squared error (RMSE) and mean absolute error (MAE) for our refinements steps of the kernel regression method of Aït-Sahalia and Lo (1998). All numbers refer to end-of-day data. *Initial* represents the original method, applied day by day. We first remove the maturity dimension (*No maturity dimension*) and then add the linear tail extrapolation in implied volatility space (*Tail extrapolation*). Finally, we additionally enforce no-arbitrage on the volatility surface in the *No-arbitrage enforcement* step

Starting from the lowest observed moneyness, we proceed in steps of  $\overline{\Delta K}(\tau_i)$  and calculate the extrapolated implied volatility as  $\tilde{\sigma}(m) = \alpha + \beta m$  until reaching a moneyness of  $m = 0.4$ . We proceed similarly for the right tail, using the last 5 observations in  $\{\sigma_i^\tau\}_{i \in [1, N]}$  and extrapolating from the largest observed moneyness in steps of  $\overline{\Delta K}(\tau_i)$  until a final moneyness of 1.6. We use the union of observed implied volatilities and thus artificially created implied volatilities as inputs for the kernel regression. Figure 1 presents the tail extrapolation visually for a sample day and maturity.

The last refinement that we apply to the Aït-Sahalia and Lo (1998) methodology is to guarantee that the resulting option-implied volatility surface is consistent with an arbitrage-free asset market by applying the algorithm of Fengler (2009).

Table 2 displays the performance gains that we achieve with each adjustment of the Aït-Sahalia and Lo (1998) method. Clearly, dropping the maturity dimension from the



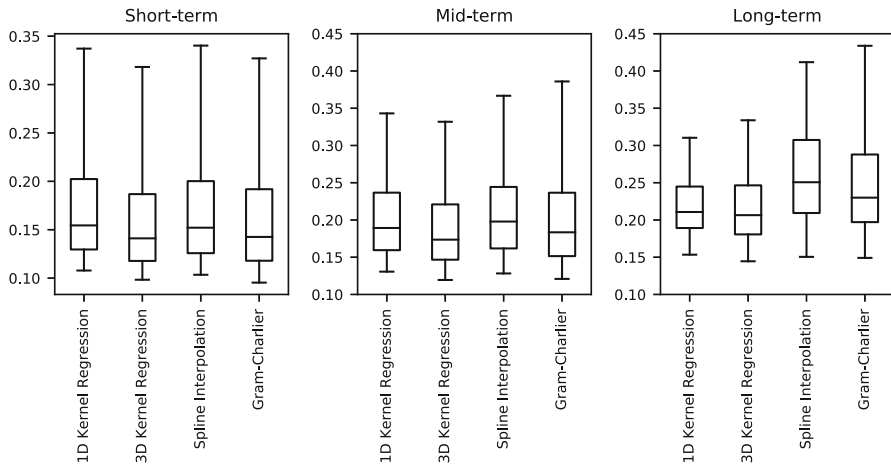
**Fig. 2** Unconditional mean of model-free implied volatility by methodology. This figure presents the sample mean of the Bakshi et al. (2003) model-free implied volatility for S&P 500 options, as a function of maturity and interpolation methodology, together with 95% confidence intervals. The short-term category refers to maturities below 30 days, the mid-term category groups maturities between 30 and 365 days, the long-term category summarizes maturities that are longer than 365 days

kernel regression leads to the strongest improvement, lowering the RMSE by 77% for the S&P 500 and 92% for the Euro Stoxx 50. Further adding the tail extrapolation leads to an additional 16% reduction in the RMSE of the S&P 500 implied volatility surface (minus 35% for the Euro Stoxx 50). Finally, enforcing no-arbitrage does generally not lead to a higher accuracy of the constructed implied volatility surface. However, as the error figures are not rising, we can reach an arbitrage-free implied volatility surface without having to sacrifice accuracy.

#### 4 On the uniqueness of option-implied information

Option-implied risk and return measures, such as forward-looking risk-neutral variance, skewness or the variance risk premium do have a unique mathematical representation. Yet, the numerical implementation of such measures does often require the integration of option prices across the whole strike spectrum, even though only a subset of these are actually traded in the market. Conditional on a correct volatility surface, Jiang and Tian (2005) show how to implement such an integration numerically. But in real-life applications the problem is that different state-of-the-art approaches for constructing the volatility surface in high resolution across the strike dimension do result in different volatility surfaces. It is the goal of our paper to assess the severity of these differences for a subset of methodologies and to perform a thorough statistical analysis to identify the methodology that comes closest to the unobserved true volatility surface.

Figure 2 summarizes the sample mean of the Bakshi et al. (2003) implied volatilities for S&P 500 options with a remaining maturity of (i) less than 30 days (short-term),



**Fig. 3** Distribution characteristics of model-free implied volatility by methodology. This figure presents box plots for the estimates of the Bakshi et al. (2003) model-free implied volatility for S&P 500 options, as a function of maturity and interpolation methodology. The short-term category refers to maturities below 30 days, the mid-term category groups maturities between 30 and 365 days, the long-term category summarizes maturities that are longer than 365 days. The rectangular boxes mark the quartiles of the implied model-free volatility distribution. The solid lines outside of the boxes expand to the 5% and 95% percentiles

(ii) 30–365 days (medium-term) and (iii) more than 365 days (long-term), respectively. Each sample mean is based on one of the four volatility surfaces that we have discussed above. We highlight two insights: First, different volatility surfaces imply different values for a standard risk measure like Bakshi et al. (2003) implied volatility. The average medium-term Bakshi et al. (2003) implied volatility has been 21.0% when using the three-dimensional kernel regression volatility surface and 23.5% when working with the spline-induced volatility surface. That 12% relative spread is statistically significant and obviously large from an economic point of view. Second, while the spline interpolation method and the kernel regression methods tend to produce Bakshi et al. (2003) volatilities that are close to each other for short-term options, the differences build up for maturities larger than 30 days. Noteworthy, the kernel regression methods produce average term structures of Bakshi et al. (2003) implied volatilities that are less steep in the maturity dimension, compared to the other two interpolation methodologies. This results in comparatively low average medium- and long-term estimates for Bakshi et al. (2003) implied volatility.

In contrast to the disperse, yet precisely estimated, sample means of the Bakshi et al. (2003) volatilities, the box plots in Fig. 3 describe the respective full distribution of an interpolation method's Bakshi et al. (2003) implied volatility. The distributions largely overlap for short- and medium-term maturities. For medium-term maturities and even more strongly for long-term maturities, the downward bias of the kernel regression methods' estimates becomes visible, which is even more pronounced in the right tail of the distribution.

Differences in model-free implied variance estimates induce different variance risk premium (VRP) estimates. The details of our VRP calculation are summarized in

**Table 3** Average annualized monthly variance risk premium and Sharpe ratio

	S&P 500		Euro Stoxx 50	
	Average VRP	Sharpe ratio	Average VRP	Sharpe ratio
3D kernel regression	0.015 (0.0006)	0.44	0.019 (0.0006)	0.5
1D kernel regression	0.020 (0.0006)	0.54	0.029 (0.0007)	0.68
Spline interpolation	0.024 (0.0008)	0.57	0.028 (0.0007)	0.65
Gram–Charlier expansion	0.016 (0.0010)	0.26	0.024 (0.0007)	0.56

This table shows average annualized 1-month variance risk premia and VRP Sharpe Ratios. Standard errors are given in parenthesis. The integration scheme and model for the physical variance expectations are the same for all estimates, such that the only difference is the method for constructing the implied volatility surface

“Appendix C”. Table 3 summarizes the average annualized VRP for S&P 500 and Euro Stoxx 50 options for different volatility surfaces, respectively. The average annualized VRP for S&P 500 (Euro Stoxx 50) options has been estimated to be between 1.5 and 2.4% (1.9% and 2.9%), depending on the volatility surface. In relative terms, the spline-based volatility surface results on average in a 60% higher S&P 500 VRP estimate relative to the same estimate for the three-dimensional kernel regression. The differences between the average VRP estimates are for nearly all pairs of methods statistically strongly significant. Monthly Sharpe ratios for the VRP range from 0.26 to 0.57 for the S&P 500 and from 0.5 to 0.68 for the Euro Stoxx 50. We highlight that these economically large differences are a direct result of the choice of the inter- and extrapolation method that builds the basis for a volatility surface; the input data and the integration scheme is the same across all methods.

Table 4 states the sample mean of model-free option-implied skewness for maturities between 15 and 91 days. Option-implied skewness is calculated as in Bakshi et al. (2003) and reported for different sub-samples.<sup>5</sup> We split the US data set into a pre-financial-crisis, a crisis and a post-financial-crisis period. We apply the same cuts for the European data set, but further split the post-financial-crisis period into a between-crisis-, an euro-crisis-, and a post-euro-crisis period due to the high impact of the European sovereign debt crisis for European stock markets. The risk-neutral skewness estimates are nearly all significantly different across different volatility surfaces. The estimates with the three-dimensional kernel regression are roughly only half the size when compared to our proposed one-dimensional kernel regression or the spline method. Further, the changes in the average risk-neutral skewness between one sub-period and the next also differ among the volatility surfaces, sometimes even disagreeing on the sign of the change. For example, risk-neutral skewness estimates based on a spline interpolation get less negative on average after the Euro Crisis in Europe, while the estimates based on all other volatility surfaces get more negative.

In summary, information extracted from option markets is supposed to be unique. But our analysis has documented that this information is sensitive with regard to the volatility surface that a researcher uses. We have shown that risk-neutral model-free estimates for variance and skewness differ by a large margin across different state-of-

<sup>5</sup> The results do not change qualitatively for different maturity intervals.

**Table 4** Average Bakshi et al. (2003) skewness by time period and method

Period	Start date	End date	3D kernel regression	1D kernel regression	Spline interpolation	Gram–Charlier expansion
(a) S&P 500						
Pre-crisis	2004-01-01	2007-12-31	− 1.256 (0.005)	− 2.347 (0.022)	− 2.583 (0.042)	− 1.332 (0.007)
Financial crisis	2008-01-01	2009-03-31	− 0.881 (0.005)	− 1.307 (0.01)	− 1.366 (0.035)	− 1.210 (0.01)
Post-crisis	2009-04-01	2017-07-21	− 1.316 (0.002)	− 2.791 (0.007)	− 3.703 (0.025)	− 1.492 (0.002)
(b) Euro Stoxx 50						
Pre-crisis	2002-01-01	2007-12-31	− 0.826 (0.009)	− 1.858 (0.009)	− 1.612 (0.01)	− 1.347 (0.015)
Financial crisis	2008-01-01	2009-03-31	− 0.708 (0.015)	− 1.324 (0.010)	− 1.335 (0.013)	− 1.242 (0.01)
Between-crisis	2009-04-01	2010-03-31	− 0.657 (0.016)	− 1.622 (0.011)	− 1.721 (0.015)	− 1.135 (0.009)
Euro crisis	2010-04-01	2013-12-31	− 0.699 (0.009)	− 1.855 (0.01)	− 1.921 (0.012)	− 1.277 (0.007)
Post-crisis	2014-01-01	2017-09-30	− 0.805 (0.008)	− 1.918 (0.009)	− 1.826 (0.015)	− 1.287 (0.007)

This table presents sample averages of the end-of-day Bakshi et al. (2003) model-free option-implied skewness, for different time periods and different interpolation schemes. All numbers in panel (a) are calculated for S&P 500 options, whereas numbers in panel (b) refer to Euro Stoxx 50 options. Only options with maturities between 15 and 91 days are considered. Standard errors are given in parenthesis



the-art methods. While we have used the Bakshi et al. (2003) moments for explanatory purpose, our findings hold more generally for any quantity that is extracted from the aggregation of option prices along the strike range.

## 5 Assessing the accuracy of a volatility surface

Here, we assess the relative advantages and disadvantages of different state-of-the-art volatility surfaces. Our previous findings have documented that standard option-implied measures differ across volatility surfaces. As these measures are deterministic functions of the volatility surface, it is natural that the most accurate volatility surface does imply the most accurate option-implied risk measures.

At the same time, accuracy is only one of multiple practical considerations when constructing an implied volatility surface. It is, for example, well thinkable to accept a lower accuracy in favor of a more informative representation of the volatility surface. Especially parametric models can provide such representations. Depending on the model, single parameters can be interpreted directly and serve as measures that express the situation at the option market in lower dimension. For example, the parameters of the SVI model of Gatheral (2004) inform about the level of the implied volatility smile, its rotation and how wide it is. Parametrization also allows to easily share an implied volatility surface, as it can be fully reconstructed from the relatively few parameters. The model for the volatility smile allows to extrapolate beyond observed option prices in a manner that is consistent with the central part of the modeled volatility smile. No-arbitrage constraints can be incorporated into the construction of the implied volatility surface at the parameter estimation phase already (Damghani and Kos 2013; Damghani 2015). These arguments speak in favor of a parametric representation of the implied volatility surface. However in this study, we explore how accurate different volatility surfaces capture option market information. The result of our analysis can then serve as an important input, next to the previously mentioned concerns, for the choice of a construction method in practice.

One might also be willing to deliberately sacrifice some accuracy in the representation of the observed implied volatilities in favor of a smoother volatility surface (Jackwerth 2000; Jackwerth and Rubinstein 1996). This is especially true, if the noise in the observations is expected to be large enough to not be fully canceled out by the interpolation and smoothing method. In that case, the constructed volatility surface would show spikes or bumps, that would decrease its smoothness. On the other hand, a very smooth volatility surface might plane out important features of the observations and thus introduce biases.<sup>6</sup> Therefore, we are going to implement a thorough investigation on the statistical accuracy of popular volatility surfaces and compare them with respect to their smoothness.

We follow a rich machine learning literature that assesses the accuracy of a model based on leave-one-out cross-validation (Geisser 1993; Kohavi 1995). The advantage of this cross-validation approach for our study is that every method that we use to

---

<sup>6</sup> The introduction of biases by very smooth volatility surfaces can easily be seen by the fact that the smoothest volatility surface is an uncurved plane.

construct an option-implied volatility surface is evaluated with data that was not used to construct the surface. This allows us to detect over- and under-fitting.

We evaluate the statistical quality of each volatility surface with end-of-day and intraday data for S&P 500 and Euro Stoxx 50 options. End-of-day data is characterized by a rich panel of option prices for different strikes and maturities, that all refer to the same point of time. We call this to be the *data-rich* environment. In contrast, intraday data is characterized by a limited amount of observed trade prices in a given time interval. We therefore call the intraday application to be a *data-poor* environment. The intraday set-up becomes increasingly data-poor as the considered time interval for pooling trade observations shrinks and thus the time resolution increases.<sup>7</sup>

The highest possible time resolution of a method is bound by the minimum amount of option trades that a method requires for constructing the option-implied volatility surface. Technically, the Gram–Charlier expansion requires only 3 observed option prices for distinct strikes at the same maturity. Price observations at 3 different strikes are also the theoretical minimum for capturing the key characteristics level, slope and curvature of the implied volatility smile. We undertake four independent intraday analyses, namely relying on 3, 4, 5 and 10 observations per maturity.

For the end-of-day data set, the cross-validation works as follows: Given price observations of a day, we in turn leave out a single observation and calculate the methodology-implied estimate for that observation. We repeat this procedure for each observed option price. The data-poor set-up with 3, 4, 5 and 10 price observations per maturity is treated similarly: For each new transaction price, we use the corresponding 3, 4, 5 or 10 preceding transaction prices with differing strike prices for the same maturity in order to create an estimate for the next transaction price. This approach is basically assessing out-of-sample how well a volatility surface predicts future option prices.

Given a set of  $N$  evaluation samples with Black-Scholes implied volatility observations  $\sigma_i^{BS}$ ,  $i \in \{1, \dots, N\}$ , our primary evaluation measure is the root mean squared error (RMSE), that arises when comparing the respective method's Black-Scholes-implied volatility estimator  $\hat{\sigma}_i^{BS}$  with the left out observation  $\sigma_i^{BS}$ :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\sigma_i^{BS} - \hat{\sigma}_i^{BS})^2}. \quad (5)$$

A detailed discussion of the relationship between the (root) mean squared error as evaluation criterion and the desire to obtain a smooth volatility surface can be found in “Appendix D”. By squaring the errors, the RMSE penalizes large deviations of the constructed volatility surface from the observed implied volatilities more strongly than small deviations. However, as discussed above, one might be willing to accept occasional large errors in favor of a smoother volatility surface. This is especially true, if one expects such large errors to be due to outliers, that are not representative of the

<sup>7</sup> End-of-day data does not necessarily create a data-rich environment. For some options, only a handful of strikes are traded, thus effectively constituting a data-poor environment in end-of-day data. On the other hand, if the time interval for pooling intraday trade prices becomes large enough, there will likely be enough observations for distinct strikes to constitute a data-rich environment.

true unobserved volatility surface. For this reason, we also calculate the mean absolute error (MAE),

$$MAE = \frac{1}{N} \sum_{i=1}^N \left| \sigma_i^{BS} - \hat{\sigma}_i^{BS} \right|, \quad (6)$$

which is less responsive to such occasional large errors and is still comparably low, if these errors only occur seldom and the remaining fit is good. For both error measures, we exclude observations with an implied volatility above 10.<sup>8</sup>

We also compare the end-of-day volatility surfaces with respect to their smoothness directly. Following Jackwerth (2000), we measure smoothness as the sum of squared second derivatives of the implied volatility surface along the moneyness dimension. More precisely, for each day  $t$  and maturity  $\tau$ , we consider a moneyness range of  $[1 - k s \sqrt{\tau}, 1 + k s \sqrt{\tau}]$ , where  $s$  is an estimate for the unconditional annual volatility of the underlying and  $k$  is a fix multiple. We discretize this range in steps of  $\Delta = \frac{2ks\sqrt{\tau}}{99}$  and construct the volatility smile  $\{\hat{\sigma}_{j,t,\tau}^{BS}\}_{j \in [0,100]}$  for the grid points of that day and maturity with each method. Our measure for the smoothness of the smile is then calculated as

$$S_{t,\tau} = \sum_{j=1}^{99} \left( \frac{\hat{\sigma}_{j-1,t,\tau}^{BS} - 2\hat{\sigma}_{j,t,\tau}^{BS} + \hat{\sigma}_{j+1,t,\tau}^{BS}}{\Delta^2} \right). \quad (7)$$

Finally, we compute the mean of that smoothness measure across all days and maturities and take the square root for better readability. In that, we exclude the 1% of the smiles with the highest and the 1% with the lowest smoothness measure to mitigate the impact of the tails of the distribution of smoothness figures on the mean estimates.

## 6 Findings

This section summarizes key findings of our empirical assessment of the statistical quality of different state-of-the-art volatility surfaces. We start with the end-of-day analysis for S&P 500 and Euro Stoxx 50 options. This section ends with the findings for the intraday analysis.

### 6.1 End-of-day: data-rich environment

The aggregated RMSEs for the end-of-day, data-rich, environment are summarized in Table 5, the respective MAE error figures can be found in Table 9 in “Appendix E”. For S&P 500 options we find that the volatility surface from the one-dimensional kernel regression produces the lowest RMSE (0.0092) and MAE (0.0019). The spline interpolation induced volatility surface ranks second with an overall RMSE (MAE) that

<sup>8</sup> In rare cases, observed option prices translate into unreasonably high implied volatilities. Our threshold of 10 leads to the exclusion of 0.012% or less of the data points, depending on the data set.

**Table 5** End-of-day cross-validation RMSE

Maturity	Moneyneess	3D kernel regression	1D kernel regression	Spline interpolation	Gram–Charlier expansion
(a) S&P 500					
All	All	578%	100% (0.0092)	228%	927%
Short	Center	397%	100% (0.0082)	373%	362%
Short	Left	1036%	100% (0.0088)	513%	1643%
Short	Right	201%	101%	100% (0.0109)	221%
Medium	Center	235%	131%	210%	100% (0.0084)
Medium	Left	622%	100% (0.0102)	138%	1039%
Medium	Right	192%	100% (0.0059)	110%	232%
Long	Center	249%	100% (0.0022)	150%	164%
Long	Left	515%	116%	100% (0.0066)	1073%
Long	Right	185%	100% (0.0052)	120%	209%
(b) Euro Stoxx 50					
All	All	1397%	100% (0.0100)	137%	1220%
Short	Center	371%	100% (0.0129)	398%	155%
Short	Left	1340%	100% (0.0395)	181%	506%
Short	Right	3369%	113%	100% (0.0120)	5204%
Medium	Center	125%	100% (0.0010)	460%	328%
Medium	Left	2521%	219%	100% (0.0059)	1892%
Medium	Right	4021%	152%	100% (0.0022)	4337%
Long	Center	164%	109%	100% (0.0020)	226%
Long	Left	1302%	216%	100% (0.0015)	2268%
Long	Right	755%	107%	100% (0.0013)	1169%

This table summarizes the cross-validation RMSE for the data-rich (end-of-day) set-up. The short maturity section groups options with remaining maturities of less than 30 days, medium maturity refers to a remaining maturity of 30–365 days and long maturity represents options with a remaining maturity of more than 365 days. The moneyneess  $m = \frac{K}{F}$  axis is divided into a left, center and right section, with the cutting lines being located at 0.95 and 1.05. Error figures are given in percent of the respective best-performing method's RMSE for each section of the implied volatility surface. For each section's best performing method, we report its RMSE in parenthesis

is 118% (91%) higher. The three-dimensional kernel regression ranks third place, with an RMSE (MAE) that is roughly 6 (12) times larger, relative to the one-dimensional kernel regression. The volatility surface of the Gram–Charlier expansion turns out to be the least accurate, producing a RMSE (MAE) that is more than 9 (23) times larger than the winning volatility surface.

Looking at the RMSE results for the Euro Stoxx 50 confirms that the volatility surface that is generated by the one-dimensional kernel regression is the most accurate one, followed by the spline-based volatility surface, whose overall RMSE is 37% higher. However, the ranking is reversed with respect to the MAE, where the spline method performs best and the kernel regression shows a 29% higher error. This reversal

**Table 6** Smoothness of the volatility surface by method

Method	S&P 500	Euro Stoxx 50
3D kernel regression	1657 (115)	3315 (213)
1D kernel regression	600 (55)	303 (24)
Spline interpolation	1402 (407)	196 (46)
Gram–Charlier expansion	623 (46)	1096 (73)

This table shows the average smoothness of the constructed implied volatility surface for each method. For each day and maturity, we sum up the squared second derivative along the moneyness dimension, aggregate across all days and maturities by taking the mean and report the root of this figure. The second derivative is calculated for a moneyness interval between  $-3$  and  $3$  times the unconditional volatility of the underlying, de-annualized to the respective maturity. Due to the existence of rare outliers, we remove the 1% largest and smallest single smoothness values from the calculation. Standard errors are given in parenthesis

is the result of a downward bias of the one-dimensional kernel regression in deep out-of-the-money medium-term options with a moneyness below 0.5. If these options were to be excluded from the error calculation, the one-dimensional kernel regression and the spline method would produce basically the same MAE for Euro Stoxx 50 data. Finally, the three-dimensional kernel regression and the Gram–Charlier expansion produce RMSEs (MAEs) that are 14 (25) and 12 (24) times higher, respectively.

We now continue to report the RMSE for different regions of the option-implied volatility surface. We split the option-implied volatility surface along the maturity and moneyness dimension. Options with a maturity of less than 30 calendar days are considered short-term, options with a maturity between 30 and 365 calendar days are considered medium-term, while options with a maturity of more than 1 year are called long-term. Along the moneyness axis, we label a moneyness of 0.95 to 1.05 as at-the-money (ATM), whereas the ‘left tail’ (‘right tail’) is characterized by a moneyness of below (above) 0.95 (1.05). In combination, these splits yield nine different regions of the option-implied volatility surface, for which we calculate the RMSE and MAE of each method.

Looking at panels (a) and (b) of Tables 5 and 9 highlights that across all volatility surfaces, short-term options and left tail options produce the highest errors. For the S&P 500, the one-dimensional kernel regression performs best in most sections of the implied volatility surface with respect to the RMSE and in all sections with respect to the MAE. Spline interpolation tends to outperform the one-dimensional kernel regression for medium- and long-term options for the Euro Stoxx 50, though, which is more pronounced in the RMSE results than in the MAE results. The three-dimensional kernel regression and the Gram–Charlier expansion do both show severe difficulties in capturing the left and the right tail of the surface. The problem of the Gram–Charlier volatility surface is that the parametric risk-neutral density approximation turns out to be insufficient for capturing market information in the tails. We identify that the relative weakness of the three-dimensional kernel regression is that it only considers options with a delta of 0.2–0.8, which ignores market information about the tails.

We proceed by comparing the smoothness of the volatility surfaces. Table 6 displays our smoothness measures, as defined in Eq. 7. A smooth volatility surface has low second derivatives and thus a low smoothness measure. Conversely, a high smoothness measure is an indicator for a more curved volatility surface. The volatility surface of the three-dimensional kernel regression is the least smooth in our tests. On the other end, the one-dimensional kernel regression appears to be smoother than most of the alternatives. For the spline interpolation, there is an interesting divergence in the smoothness measure between its comparably rough S&P 500 volatility surface and its very smooth Euro Stoxx 50 surface. This goes hand in hand with the low errors of the spline interpolation in constructing the Euro Stoxx 50 surface. It seems like the spline interpolation is not able to cancel out all noise in the S&P 500 data, which produces a more curved volatility surface and higher fitting errors. At the same time, a smooth volatility surface does not necessarily imply low errors, as can be seen from the S&P 500 results for the Gram–Charlier expansion. That method might produce a smooth volatility surface, though it does not accurately reflect the information in observed option prices.

In short summary, the cross-validation of the data-rich environment recommends to use a volatility surface that was constructed based on the one-dimensional kernel regression. The spline interpolation is a good alternative for Euro Stoxx 50 data, though appears to not be able to cancel out all noise in the S&P 500 data. The volatility surface of the Gram–Charlier expansion should only be applied if one is interested in option-implied information from the at-the-money region and should be avoided when inferring conclusions about the tails. The three-dimensional kernel regression produces RMSEs that, relative to the one-dimensional kernel regression, are roughly 6 times larger for S&P 500 options and roughly 14 times larger for Euro Stoxx 50 options.

## 6.2 Intraday: data-poor environment

Here, we summarize key findings for the intraday analysis. We start with the highest time resolution (3 trades), followed by interpolations based on 4, 5 and 10 trades for distinct strikes.

The findings from Table 7 (see Table 10 in “Appendix E” for MAE results) highlight that the one-dimensional kernel regression produces the most accurate overall volatility surface. The three-dimensional kernel regression is the second most accurate surface, though the performance differential is not as large as for the end-of-day set-up. In our sample, we find that the spline-based volatility surface suffers from frequent outliers, which increases the RMSE enormously. This is not surprising, because the spline interpolation is prone to over-fitting when it is applied to very few observations. Figure 4 visualizes the potential problem of over-fitting when constructing the spline-based volatility surface with 5 or less data points. As the MAE is less responsive to occasional large errors, which are a result of over-fitting, the spline interpolation performs better with respect to the MAE than with respect to the RMSE. It even shows a slightly lower MAE than the one-dimensional kernel regression for the Euro Stoxx 50 intraday set-up with 3 trades. The volatility surface of the Gram–Charlier expansion

**Table 7** Intraday cross-validation RMSE

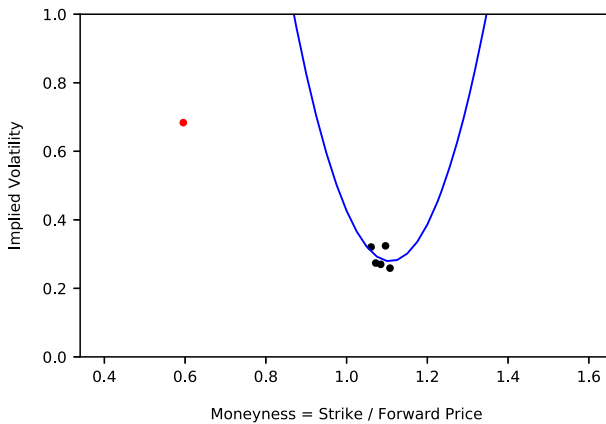
Maturity	Moneyness	3D kernel regression	1D kernel regression	Spline interpolation	Gram–Charlier expansion)
(a) S&P 500					
All	All	116%	100% (0.0705)	381%	127%
Short	Center	140%	100% (0.0371)	437%	174%
Short	Left	105%	100% (0.1445)	346%	106%
Short	Right	100% (0.05)	141%	264%	245%
Medium	Center	120%	100% (0.0315)	421%	165%
Medium	Left	110%	100% (0.0878)	471%	121%
Medium	Right	100% (0.0459)	111%	244%	195%
Long	Center	128%	100% (0.046)	301%	225
Long	Left	130%	100% (0.0747)	459%	139%
Long	Right	130%	100% (0.0635)	324%	228%
(b) Euro Stoxx 50					
All	All	111%	100% (0.0492)	290%	152%
Short	Center	145%	100% (0.0246)	166%	192%
Short	Left	102%	100% (0.0838)	195%	109%
Short	Right	100% (0.0438)	109%	146%	181%
Medium	Center	126%	100% (0.0306)	296%	228%
Medium	Left	108%	100% (0.0614)	423%	142%
Medium	Right	107%	100% (0.0458)	212%	203%
Long	Center	158%	100% (0.0425)	441%	308%
Long	Left	131%	100% (0.064)	617%	152%
Long	Right	126%	100% (0.0682)	275%	236%

This table summarizes the out-of-sample RMSE for the data-poor (intraday) set-up. The short maturity section groups options with remaining maturities of less than 30 days, medium maturity refers to a remaining maturity of 30–365 days and long maturity represents options with a remaining maturity of more than 365 days. The moneyness  $m = \frac{K}{F}$  axis is divided into a left, center and right section, with the cutting lines being located at 0.95 and 1.05. Evaluations are performed based on 3 prior trades on different strikes and the same maturity. Error figures are given in percent of the respective best-performing method's RMSE for each section of the implied volatility surface. For each section's best performing method, we report its RMSE in parenthesis

ranks always in the back, though with a much smaller error differential than in the end-of-day set-up. There does not appear to occur any over-fitting in the Gram–Charlier expansion, however in all of our tests, the general fit seems to be worse than for the kernel regression methods.

Comparing different sections of the implied volatility surface reveals that it is primarily the moneyness axis, and more precisely the left tail of shorter-term options that causes problems to all volatility surfaces. The volatility surface from the one-dimensional kernel regression generally copes best with these difficulties. With respect to both error measures, the volatility surface of the one-dimensional kernel regression implies the highest statistical accuracy across most regions of the volatility surface.





**Fig. 4** Sample for over-fitting in the spline interpolation. This figure visualizes the severity of the over-fitting in the spline interpolation methodology for a particular data point of the Euro Stoxx 50 sample, namely 30-day options on May 16th, 2012 (01:35 p.m.). The dots with a moneyiness between 1.0 and 1.2 represent the 5 previous trade observations based on which the spline interpolation constructs the spline, which is depicted by the solid line. The dot with a moneyiness around 0.6 marks the next trade observation

Relatively speaking, that volatility surface is especially precise for longer maturity options and for left tail options of all maturities.

We now report the main findings for the intraday analysis that builds the volatility surface with 4, 5 and 10 price observations per maturity, respectively. Table 8 summarizes the results. As one would expect, adding more observations as input for the interpolations, improves the statistical accuracy for all methods. Clearly, the spline-based volatility surface benefits most, since the higher amount of observations reduces the tendency of over-fitting. Its RMSE differential towards the best method is falling rapidly as the volatility surface is based on more price observations. With the decreasing occurrence of over-fitting, spline interpolation quickly becomes the leading method with respect to the MAE and even outperforms the other approaches with respect to the RMSE in Euro Stoxx 50 options, when using 10 trades for the volatility surface. We conclude that the spline interpolation is a good alternative for the one-dimensional kernel regression in intraday set-ups, if an intermediate amount of price observations is used and if the occasional occurrence of an over-fitted volatility smile is acceptable.

## 7 Conclusion

Our paper makes two contributions to the literature. First, we are the first to document that option-implied information, such as well-known forward-looking measures for variance, skewness and the variance risk premium, are sensitive to the choice of the underlying volatility surface. These findings hold more generally for any quantity that is extracted from the aggregation of option prices along the strike range.

Second, we implement a thorough statistical assessment of the accuracy of common parametric, semi-parametric and non-parametric approaches that the literature has entertained for constructing volatility surfaces at the end-of-day and intraday

**Table 8** Intraday cross-validation errors for varying amount of trades used

Interpolation base	3D kernel regression	1D kernel regression	Spline interpolation	Gram–Charlier expansion
(a) S&P 500				
RMSE				
3 trades	116%	100% (0.0705)	381%	127%
4 trades	120%	100% (0.0624)	257%	132%
5 trades	126%	100% (0.0565)	196%	139%
10 trades	149%	100% (0.0421)	119%	166%
MAE				
3 trades	123%	100% (0.0309)	139%	137%
4 trades	134%	101%	100% (0.025)	150%
5 trades	169%	116%	100% (0.0184)	189%
10 trades	216%	115%	100% (0.0113)	266%
(b) Euro Stoxx 50				
RMSE				
3 trades	111	100% (0.0492)	290%	152%
4 trades	120	100% (0.0435)	158%	167%
5 trades	128	100% (0.0399)	115%	182%
10 trades	198	123%	100% (0.026)	275%
MAE				
3 trades	108%	101%	100% (0.0199)	139%
4 trades	160%	133%	100% (0.0127)	210%
5 trades	181%	140%	100% (0.0107)	249%
10 trades	197%	117%	100% (0.0093)	293%

This table presents the out-of-sample RMSE and MAE for the implied volatility surfaces by amount of trades used in the construction. Error figures are given in percent of the lowest respective error for each amount of trades. The lowest error for each measure and amount of trades is given in parenthesis.

frequency. The methods under consideration are the Gram–Charlier expansion, the spline interpolation of Figlewski (2008), the three-dimensional kernel regression of OptionMetrics (2016) and the one-dimensional kernel regression with a linear tail extrapolation for the end-of-day setting. We have recorded the root mean squared error (RMSE) and mean absolute error (MAE) based on a leave-one-out cross-validation for each method and have compared the smoothness of the constructed end-of-day volatility surfaces directly. The test assets are S&P 500 and Euro Stoxx 50 options at the daily and intraday frequency over the past 14 years.

The result of our analysis is that the volatility surface that is constructed with the one-dimensional kernel regression is generally the most accurate one for end-of-day and intraday options on the S&P 500 and the Euro Stoxx 50. We recommend to use that volatility surface for extracting option-implied information at the daily and high-resolution intraday frequency. However, if an intermediate amount of observations is available, spline interpolation might be a preferable alternative, despite the occasional occurrence of over-fitting. The parametric Gram–Charlier volatility surface is

in many cases too restrictive to approximate the true risk-neutral distribution and thus the true volatility surface accurately. The three-dimensional kernel regression of OptionMetrics (2016) performs comparatively well in the intraday analysis, but shows shortcomings in capturing the left tail of the option-implied volatility surface, especially at the end-of-day frequency.

**Acknowledgements** Open Access funding provided by Projekt DEAL. Simon Walther gratefully acknowledges financial support by the Konrad-Adenauer-Stiftung.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Calculation of dividend yield estimates

We make use of the put-call parity to obtain daily model-free estimates of the dividend yield for an option's underlying. More precisely, let  $C$  be the Call price and  $P$  be the Put price of 2 options with maturity  $\tau$  on the same underlying with spot price  $S$  and forward price  $F$ . Both options have the same strike  $K$ . Let the risk-free rate be  $r$  and the dividend yield be  $q$ . Following Hull (2018), we can express the put-call parity in the following equation:

$$C - P = e^{-r\tau}(F - K) = e^{-q\tau}S - e^{-r\tau}K$$

Solving this equation for the dividend yield  $q$  yields

$$q = \frac{1}{\tau} [\log(S) - \log(C - P + e^{-r\tau}K)]$$

In order to estimate the dividend yield with this equation, one needs to find a Put and a Call price with the same strike and maturity at the same point of time. In our end-of-day data set, settlement or bid/ask prices for (nearly) all strikes and maturities of both Puts and Calls are available. As all settlement prices refer to the same point of time, the price of the underlying coincides for all price quotes. For each pair of options at a specific date, strike and maturity, we calculate the implied dividend yield and take the median over the strike range to arrive at a single estimate of the dividend yield per date and maturity.

The situation is more complex for intraday transaction data, since reported trades do not occur simultaneously. For this reason, there will always be a time differential when matching a Put and a Call with the same strike and maturity for calculating the dividend yield via the put-call parity. However, the time differential might be large enough for the price of the underlying to change significantly. In that case, the put-call parity does not hold any more, even if the dividend yield remains constant. When

matching Put and Call prices in the intraday set-up, we therefore impose the constraints that the price quotes are from the same day and that the price of the underlying has changed less than 0.01%. If multiple pairs of a Put and a Call fulfill these constraints, we choose the ones with the smallest time differentials. This may result in multiple Put-Call pairs for the same strike and maturity on a single day. Again, we take the median over the strike range to arrive at a single dividend yield estimate per date and maturity.

## Appendix B: Volatility surfaces

Let  $\{O_{i,t}\}_{i \in [1, \dots, N]}$ , be a panel of option prices with  $N$  being the number of observed option prices at time  $t$ .  $S_t$  denotes the spot price of the option's underlying and  $F_t$  the corresponding forward price of the underlying with the same maturity as the option. As all methodologies apply per point of time, we will drop the time index to save notation. Furthermore, each option price  $O_i$  is associated with a strike  $K_i$ , a remaining time to maturity  $\tau_i$ , an option delta  $\Delta_i$  and an indicator  $I_i$ , which is 1 for Call options and 0 for Put options. In the following subsections, we may identify an option as a function of its parameters, e.g.,  $O_i = O(K_i, \tau_i, I_i)$ .

### B.1 Gram-Charlier expansion

The Gram-Charlier expansion applies a fourth-order approximation of the risk-neutral density  $dp^Q(x, \tau)$ . Mathematically, this means

$$dp^Q(x, \tau) = \phi(x) - \frac{\gamma_{1,\tau}}{3!} \frac{\partial^3 \phi}{\partial x^3}(x) + \frac{\gamma_{2,\tau}}{4!} \frac{\partial^4 \phi}{\partial x^4}(x) + H.O.T.(x^5), \quad \forall x, \tau, \quad (8)$$

where  $\phi(x)$  is the Gaussian density function with volatility  $\sigma_\tau$ , evaluated at the point  $x$ ;  $\gamma_{1,\tau}$  and  $\gamma_{2,\tau}$  are maturity-specific free parameters that account for the degree of skewness and excess kurtosis in  $dp^Q(x, \tau)$ , and  $H.O.T.$  stands for higher order error terms.

We follow Beber and Brandt (2006) and fit the free parameters  $(\sigma_\tau, \gamma_{1,\tau}, \gamma_{2,\tau})$  by minimizing the squared pricing errors between observed option prices and Backus et al. (2004) implied option prices, i.e.,

$$\min_{\sigma_\tau, \gamma_{1,\tau}, \gamma_{2,\tau}} \sum_{i=1}^N \left( O(K_i, \tau_i, I_i) - \hat{O}(K_i, \tau_i, I_i; \sigma_\tau, \gamma_{1,\tau}, \gamma_{2,\tau}) \right)^2 \quad (9)$$

where the Backus et al. (2004) implied Call price coincides with

$$\begin{aligned} \hat{C}(K, \tau; \sigma_\tau, \gamma_{1,\tau}, \gamma_{2,\tau}) &= e^{-r\tau} (F\phi(d) - K\phi(d - \sigma_\tau)) \\ &\quad + Fe^{-r\tau} \phi(d) \sigma_\tau \left[ \frac{\gamma_{1,\tau}}{3!} (2\sigma_\tau - d) \right. \\ &\quad \left. - \frac{\gamma_{2,\tau}}{4!} (1 - d^2 + 3d\sigma_\tau - 3\sigma_\tau^2) \right] \end{aligned}$$

$$d = \frac{\ln(F/K) + \sigma_\tau^2/2}{\sigma_\tau}.$$

Put prices can then be obtained via the put-call parity.

We follow the approach of Jondeau and Rockinger (2000) and fit the three free parameters separately for each observed maturity  $\tau$ . We then feed Backus et al. (2004) implied option prices to the Black-Scholes formula to recover the full option-implied volatility surface.

## B.2 Spline interpolation and extrapolation

We follow Figlewski (2008) and Fengler (2009) as their procedure nests many other approaches of the literature. Their variation consists of 4 steps: (i) data pre-processing, (ii) spline interpolation, (iii) tail extrapolation and (iv) data post-processing. We now sketch each step in more detail.

The pre-processing of the observed ATM option-implied volatilities addresses a well-known practical problem of the data: Most approaches construct the option-implied volatility surface with only out-of-the-money (OTM) options. They do so, because OTM options are especially liquid and hence, their prices are especially informative about the current market environment. However, the switch from OTM Puts to OTM Calls at a moneyness of  $m = 1$  results in many occasions in a small jump in implied volatility, as ATM Puts often imply a slightly different volatility than ATM Calls. This volatility jump may deteriorate the quality of the fitted spline and may even lead to inconsistent smoothed prices, which in the end induces arbitrage.

The pre-processing step mitigates this problem, by considering—separately for each maturity—a small region around the ATM level, which contains options that fulfill  $0.98 \leq m_i \leq 1.02$ . Let  $K_{high}$  be the highest and  $K_{low}$  be the lowest traded strike within that moneyness region. By linearly combining the observed Put- and Call-implied volatilities, we smooth the hard volatility jump in the region of ATM options and replace the observed option-implied volatilities with their smoothed counterparts (Figlewski 2008):

$$\sigma^*(K_i) = w \sigma_{Put}(K_i) + (1 - w) \sigma_{Call}(K_i), \quad (10)$$

$$w = \frac{K_{high} - K_i}{K_{high} - K_{low}}, \quad \forall K_i \in [0.98 F_i, 1.02 F_i]. \quad (11)$$

$\sigma_{Put}(K_i)$  and  $\sigma_{Call}(K_i)$  are the observed Put- and Call-implied volatilities in the ATM region  $0.98 \leq m_i \leq 1.02$ , whereas  $\sigma^*(K_i)$  is the smoothed option-implied volatility for the strike level  $K_i$ .

With regard to the second step, we consider each maturity separately and fit a spline to the observed (and pre-processed) option-implied volatilities. Figlewski (2008) proposes to use a 4th order polynomial with one knot at a moneyness of  $m = 1$ . Using a 4th order polynomial guarantees a continuous 3rd derivative of the spline, which translates into a smooth RND without any edges. In addition, it is not too restrictive on the shape of the implied volatility curve, while still smoothing out market noise.

If enough option observations to determine all parameters of the spline are available, we use this setting for interpolating the moneyness range. In settings where we do not have a sufficient amount of option-implied volatility data available, we simply drop the knot point and reduce the polynomial's order, thus effectively reducing the number of parameters.

The third step concerns the approximation of the RND tails. Depending on the degree of moneyness of the traded options, it might occur that these do not reach far enough into the tails of the RND, which would make it difficult to recover these tails. We follow Figlewski (2008) and fit a generalized extreme value density (GEV) to both tails of the interpolated risk-neutral density and extrapolate option prices based on that GEV density. Said differently, we first convert interpolated option-implied volatilities back to Call option prices  $C(K_i, \tau_i)$  via the Black-Scholes formula and numerically approximate the second derivative of the interpolated Call prices along the strike dimension. This provides us with an estimate for the central part of the RND, i.e.,

$$dp^Q(K_i, \tau_i) \approx e^{r\tau_i} \frac{C(K_{i+1}, \tau_i) - 2C(K_i, \tau_i) + C(K_{i-1}, \tau_i)}{(K_{i+1} - K_i)(K_i - K_{i-1})}. \quad (12)$$

We fit the GEV distribution to both ends of the supported RND interval as described in Figlewski (2008).

The fourth step of the interpolation method follows the algorithm of Fengler (2009) to post-process the option-implied volatility surface in order to exclude arbitrage across strikes and maturities.

### B.3 Three-dimensional kernel regression

We follow the methodology from OptionMetrics' IvyDB US File and Data Reference Manual, Version 3.1, Rev. 8/19/2016. The option-implied volatility estimate for an option with a given delta  $\Delta_j$ , a maturity  $\tau_j$  and an Put/Call flag  $I_j \in \{0, 1\}$  is given by

$$\hat{\sigma}(\Delta_j, \tau_j, I_j) = \sum_{i=1}^N \frac{v_i \times k(\Delta_j - \Delta_i, \log(\tau_j) - \log(\tau_i), I_j - I_i)}{\sum_{i=1}^N v_i \times k(\Delta_j - \Delta_i, \log(\tau_j) - \log(\tau_i), I_j - I_i)} \sigma(\Delta_i, \tau_i, I_i), \quad (13)$$

where the left-hand side stands for the approximation of the unobserved option-implied volatility, while the right-hand side is a weighted average of all  $N$  observed option-implied volatilities. The weight that each observed  $\sigma(\Delta_i, \tau_i, I_i)$  obtains depends on that option's vega  $v_i$  and on that option's proximity to the target location on the volatility surface, where the degree of proximity is measured by a Gaussian kernel function

$$k(x, y, z) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2h_1} + \frac{y^2}{2h_2} + \frac{z^2}{2h_3}\right)}. \quad (14)$$

The IvyDB US Reference Manual of OptionMetrics LLC specifies the bandwidth parameters to coincide with

$$h_1 = 0.05, \quad h_2 = 0.005, \quad h_3 = 0.001.$$

OptionMetrics LLC limits the kernel regression to the delta interval of [0.2, 0.8].

The volatility surface of OptionMetrics (2016) is not necessarily arbitrage-free. In order to compare apples with apples, we need all considered volatility surfaces to be arbitrage-free, so we apply a post-processing step as in Fengler (2009) on the original OptionMetrics (2016) volatility surface. All of our results are qualitatively and quantitatively robust to whether or not one applies this post-processing step. In unreported results we have found that ensuring no-arbitrage improves the statistical accuracy of the OptionMetrics (2016) volatility surface.

## Appendix C: Calculating the variance risk premium

In line with Drechsler and Yaron (2011), we define the variance risk premium (VRP) as

$$\begin{aligned} VRP_t &= E_t^Q[\text{Total Return Variation}(t, t + 21)] \\ &\quad - E_t[\text{Total Return Variation}(t, t + 21)] \end{aligned}$$

where  $t + 21$  stands for a 1-month (or 21-trading-day) increment.

We use Bakshi et al. (2003) model-free implied variance as the measure for the variance expectation under  $Q$ . For those days, during which the 1-month maturity option is not traded, we linearly interpolate the implied variance measures for surrounding traded maturities along the maturity dimension to obtain a 1-month-ahead expectation.

We use an adjusted form of the model of Corsi (2009) to obtain variance expectations under the statistical measure:

$$\begin{aligned} RV_t &= r_t^2 = \left(\log \frac{S_t}{S_{t-1}}\right)^2, \quad RV_t^{(w)} = \frac{1}{5} \sum_{i=t-4}^t r_i^2, \quad RV_t^{(m)} = \frac{1}{21} \sum_{i=t-20}^t r_i^2 \\ \log(RV_{t+21}^{(m)}) &= \alpha + \beta_1 \log(RV_t) + \beta_2 \log(RV_t^{(w)}) + \beta_3 \log(RV_t^{(m)}) + \epsilon_{t+21}, \end{aligned}$$

with a random error term  $\epsilon_{t+21}$ .

## Appendix D: The use of the MSE for model evaluation and smoothness

The following discussion shows that the MSE is indeed a valid measure for assessing the quality of a standardization methodology, especially considering the desire



to obtain a smooth volatility surface that still captures all characteristics of the true unobserved volatility surface. In the end-of-day set-up, all price recordings for a given day refer to the same point of time, but may be measured with noise. We can split the error of a standardization methodology,  $\sigma_i^{BS} - \hat{\sigma}_i^{BS}$ , into a noise component  $\epsilon_i^{noise}$  and a methodology-specific bias component  $\epsilon_i^{bias}$ . The noise component is the difference between the implied volatility observation and the true unobserved implied volatility and is thus the same across all methodologies. All methodologies are not purely interpolating between observed option prices, but attempt to smooth the implied volatility surface in order to reduce the noise error. However, smoothing comes at the cost that a potential over-smoothing introduces a bias into the implied volatility estimator, which is independent of the noise component. The methodology-specific residual can thus be rewritten as  $\epsilon_i^{method} = b\epsilon_i^{noise} + \epsilon_i^{bias}$ , where we would expect  $b$  to be positive but smaller than 1. If a method just replicates the exact implied volatility observations without any smoothing, its volatility surface would not be biased ( $\forall i : \epsilon_i^{bias} = 0$ ), but exposed to the full noise error component ( $b = 1$ ). Substituting these considerations into the MSE equation yields for the expected MSE

$$E[MSE] = E[(b\epsilon_i^{noise} + \epsilon_i^{bias})^2] = b^2 E[(\epsilon_i^{noise})^2] + E[(\epsilon_i^{bias})^2]. \quad (15)$$

The expected MSE consists of two components: the noise component and the bias component. If a method works particularly well in smoothing out price measurement noise, its  $b$  parameter is close to 0 and the expected MSE reduces. At the same time, the bias error directly increases the MSE if a method starts to over-smooth the implied volatility surface. Splitting the estimation error into its 2 components would be tricky, as the noise component is not observed directly. However, a perfect standardization methodology would have just the right smoothness to cancel out noise without introducing any biases. Both objectives lead to a lower MSE.

The situation becomes a bit more complex for the intraday set-up. Here, transaction prices are recorded at different points of time, leading to a situation, in which there is a small time gap between the option prices that are used to construct the volatility surface and the option price that is used to evaluate the volatility surface. During this time, market expectations and fundamentals may have changed, which may result in a change in the true implied volatility surface. When calculating the mean squared error, the time difference introduces a third error component,  $\epsilon_i^{time}$ , into the estimation error. We assume this error component to be independent from the noise and the bias components in the estimation error. For the intraday set-up, the expected MSE can hence be written as

$$E[MSE] = b^2 E[(\epsilon_i^{noise})^2] + E[(\epsilon_i^{bias})^2] + E[(\epsilon_i^{time})^2]. \quad (16)$$

Since all methods are evaluated on the same data set,  $E[(\epsilon_i^{time})^2]$  should be the same for all considered methodologies. While the time component biases the measurement of the full method-specific MSE upwards, a comparison of standardization methodologies is still possible for the intraday set-up, as methodologies with more

favorable characteristics (better noise cancellation, no biases) still possess a lower expected MSE.

## Appendix E: MAE results

See Tables 9 and 10.

**Table 9** End-of-day cross-validation MAE

Maturity	Moneyess	3D kernel regression	1D kernel regression	Spline interpolation	Gram–Charlier expansio
(a) S&P 500					
All	All	1174%	100% (0.0019)	191%	2370%
Short	Center	506%	100% (0.0013)	483%	655%
Short	Left	1844%	100% (0.0032)	379%	3535%
Short	Right	377%	100% (0.0037)	112%	375%
Medium	Center	299%	100% (0.0005)	370%	561%
Medium	Left	1487%	100% (0.0021)	121%	3275%
Medium	Right	249%	100% (0.0018)	114%	432%
Long	Center	252%	100% (0.0007)	195%	303%
Long	Left	706%	100% (0.0024)	128%	1402%
Long	Right	349%	100% (0.0014)	131%	490%
(b) Euro Stoxx 50					
All	All	2536%	129%	100% (0.0015)	2435%
Short	Center	309%	100% (0.006)	155%	104%
Short	Left	2618%	100% (0.0123)	104%	1184%
Short	Right	7037%	100% (0.0035)	122%	9905%
Medium	Center	127%	100% (0.0005)	217%	422%
Medium	Left	4284%	265%	100% (0.0014)	4570%
Medium	Right	3262%	112%	100% (0.0012)	3419%
Long	Center	165%	100% (0.0010)	106%	284%
Long	Left	1401%	119%	100% (0.0008)	2066%
Long	Right	640%	100% (0.0006)	116%	1388%

This table summarizes the cross-validation mean absolute error (MAE) for the data-rich (end-of-day) set-up. The short maturity section groups options with remaining maturities of less than 30 days, medium maturity refers to a remaining maturity of 30–365 days and long maturity represents options with a remaining maturity of more than 365 days. The moneyess  $m = \frac{K}{F}$  axis is divided into a left, center and right section, with the cutting lines being located at 0.95 and 1.05. Error figures are given in percent of the respective best-performing method's MAE for each section of the implied volatility surface. For each section's best performing method, we report its MAE in parenthesis

**Table 10** Intraday cross-validation MAE

Maturity	Moneyness	3D kernel regression	1D kernel regression	Spline interpolation	Gram–Charlier expansion
(a) S&P 500					
All	All	123%	100% (0.0309)	139%	137%
Short	Center	121%	100% (0.0191)	117%	156%
Short	Left	121%	100% (0.0813)	155%	116%
Short	Right	100% (0.0254)	117%	163%	239%
Medium	Center	103%	100% (0.0164)	107%	135%
Medium	Left	117%	100% (0.0437)	171%	131%
Medium	Right	100% (0.179)	132%	166%	231%
Long	Center	115%	100% (0.0229)	147%	208%
Long	Left	143%	100% (0.0372)	158%	154%
Long	Right	101%	100% (0.0313)	166%	213%
(b) Euro Stoxx 50					
All	All	108%	101%	100% (0.0199)	139%
Short	Center	164%	156%	100% (0.0078)	191%
Short	Left	153%	129%	100% (0.0269)	144%
Short	Right	100% (0.0114)	122%	136%	206%
Medium	Center	108%	108%	100% (0.0145)	170%
Medium	Left	113%	100% (0.0299)	130%	132%
Medium	Right	100% (0.0149)	129%	155%	241%
Long	Center	105%	100% (0.0239)	132%	246%
Long	Left	119%	100% (0.0282)	249%	173%
Long	Right	108%	100% (0.033)	139%	230%

This table summarizes the out-of-sample MAE for the data-poor (intraday) set-up. The short maturity section groups options with remaining maturities of less than 30 days, medium maturity refers to a remaining maturity of 30–365 days and long maturity represents options with a remaining maturity of more than 365 days. The moneyness  $m = \frac{K}{F}$  axis is divided into a left, center and right section, with the cutting lines being located at 0.95 and 1.05. Evaluations are performed based on 3 prior trades on different strikes and the same maturity. Error figures are given in percent of the respective best-performing method's MAE for each section of the implied volatility surface. For each section's best performing method, we report its MAE in parenthesis

## References

- Aït-Sahalia, Y., & Lo, A. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2), 499–547.
- Backus, D., Foresi, S., & Wu, L. (2004). *Accounting for biases in Black–Scholes*. Available at SSRN, <https://ssrn.com/abstract=585623>. Accessed 29 Apr 2020.
- Bahaludin, H., & Abdullah, M. H. (2017). Empirical performance of interpolation techniques in risk-neutral density (RND) estimation. *Journal of Physics: Conference Series*, 819(1), 012026.
- Bakshi, G., Kapadia, N., & Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16(1), 101–143.
- Beber, A., & Brandt, M. W. (2006). The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. *Journal of Monetary Economics*, 53(8), 1997–2039.

- Bliss, R. R., & Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions. *Journal of Banking & Finance*, 26(2–3), 381–422.
- Bollerslev, T., & Todorov, V. (2011). Tails, fears, and risk premia. *The Journal of Finance*, 66(6), 2165–2211.
- Buss, A., & Vilkov, G. (2012). Measuring equity risk with option-implied correlations. *The Review of Financial Studies*, 25(10), 3113–3140.
- Chang, B. Y., Christoffersen, P., Jacobs, K., & Vainberg, G. (2012). Option-implied measures of equity risk. *Review of Finance*, 16(2), 385–428.
- Christoffersen, P., Feunou, B., Jeon, Y., & Ornthanalai, C. (2017). *Time-varying crash risk: The role of stock market liquidity*. Rotman School of Management Working Paper No. 2797308. Available at SSRN, <https://ssrn.com/abstract=2797308>.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196.
- Damghani, B. M. (2015). Introducing the implied volatility surface parametrization (IVP): Application to the FX market. *Wilmott*, 2015(77), 68–81.
- Damghani, B. M., & Kos, A. (2013). De-arbitraging with a weak smile: Application to skew risk. *Wilmott*, 64, 40–49.
- Drechsler, I., & Yaron, A. (2011). What's vol got to do with it. *The Review of Financial Studies*, 24(1), 1–45.
- Driessen, J., Maenhout, P. J., & Vilkov, G. (2009). The price of correlation risk: Evidence from equity options. *The Journal of Finance*, 64(3), 1377–1406.
- Du, J., & Kapadia, N. (2012). *Tail and volatility indices from option prices*. [https://people.umass.edu/nkapadia/docs/Du\\_Kapadia\\_August2012.pdf](https://people.umass.edu/nkapadia/docs/Du_Kapadia_August2012.pdf). Accessed 29 Apr 2020.
- Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, 9(4), 417–428.
- Figlewski, S. (2008). Estimating the implied risk neutral density. In M. W. Tim Bollerslev & J. R. Russell (Eds.), *Volatility and time series econometrics: Essays in honor of Robert F. Engle*. Oxford: Oxford University Press.
- Gatheral, J. (2004). *A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives*. In *Presentation at Global Derivatives & Risk Management, Madrid*. <http://faculty.baruch.cuny.edu/jgatheral/madrid2004.pdf>. Accessed 29 Apr 2020.
- Gatheral, J. (2006). *The volatility surface: A practitioner's guide*. Hoboken, New Jersey: Wiley.
- Geisser, S. (1993). *Predictive inference: An introduction*. Monographs on statistics and applied probability (1st ed., p. 55). New York: Chapman & Hall.
- Hagan, P. S., Kumar, D., Lesniewski, A. S., & Woodward, D. E. (2002). Managing smile risk. *The Best of Wilmott*, 1, 249–296.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2), 327–343.
- Hofmann, M., & Uhrig-Homburg, M. (2018). *Volatility noise*. Available at SSRN, <https://ssrn.com/abstract=3130045>. Accessed 29 Apr 2020.
- Hull, J. (2018). *Options, futures, and other derivatives* (10th ed.). New York, NY: Pearson.
- Jackwerth, J. C. (1999). Option-implied risk-neutral distributions and implied binomial trees: A literature review. *The Journal of Derivatives*, 7(2), 66–82.
- Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *The Review of Financial Studies*, 13(2), 433–451.
- Jackwerth, J. C., & Rubinstein, M. (1996). Recovering probability distributions from contemporaneous security prices. *The Journal of Finance*, 51(5), 1611–1631.
- Jiang, G. J., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, 18(4), 1305–1342.
- Jondeau, E., & Rockinger, M. (2000). Reading the smile: The message conveyed by methods which infer risk neutral densities. *Journal of International Money and Finance*, 19(6), 885–915.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 14(2), 1137–1145.
- Martin, I.W.R., (2017). What is the expected return on the market? *The Quarterly Journal of Economics*, 132(1), 367–433.
- Martin, I., & Wagner, C. (2019). What is the expected return on a stock? *The Journal of Finance*, 74(4), 1887–1929.

- OptionMetrics (2016). IvyDB US File and Data Reference Manual, Version 3.1, Rev. 8/19/2016. 1776 Broadway, Suite 1800, New York, NY 10019.
- Schneider, P., & Trojani, F. (2015). Fear trading Swiss Finance Institute Research Paper No. 15-03. Available at SSRN, <https://ssrn.com/abstract=1994454>.
- Schönbucher, P. J. (1999). A market model for stochastic implied volatility. *Philosophical Transactions of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, 357, 2071–2092.
- Swanson, E. T. (2016). Discussion of options-implied probability density functions for real interest rates. [http://www.ijcb.org/journal/ijcb16q3a3\\_disc.pdf](http://www.ijcb.org/journal/ijcb16q3a3_disc.pdf).
- Wright, J. H. (2016). Options-implied probability density functions for real interest rates. *International Journal of Central Banking*, 12(3), 129–149.
- Zhao, B., & Hodges, S. D. (2013). Parametric modeling of implied smile functions: A generalized SVI model. *Review of Derivatives Research*, 16(1), 53–77.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.