

Venturing into uncharted territory: An extensible implied volatility surface model

Pascal François¹  | Rémi Galarneau-Vincent³ | Geneviève Gauthier^{4,5}  | Frédéric Godin² 

¹Department of Finance, HEC Montréal, Montréal, Canada

²Department of Mathematics and Statistics, Concordia University, Montréal, Canada

³Department of Decision Science, HEC Montréal, Montréal, Canada

⁴GERAD and Department of Decision Science, HEC Montréal, Canada

⁵Associate member, Oxford-Man Institute of Quantitative Finance, Oxford, UK

Correspondence

Pascal François, Department of Finance,
 HEC Montréal, 3000 Chemin de la Côte-
 Ste-Catherine, Montréal (QC) H3T 2A7,
 Canada.

Email: pascal.francois@hec.ca

Funding information

TMX Group; Canadian Network for
 Research and Innovation in Machining
 Technology, Natural Sciences and
 Engineering Research Council of Canada,
 Grant/Award Numbers: RGPIN-2019-
 04029, RGPIN-2017-06837; Hydro-Québec;
 Institut de Valorisation des Données;
 Letko-Brosseau; HEC Montreal,
 Grant/Award Number: Professorship

Abstract

A new factor-based representation of implied volatility (IV) surfaces is proposed. The factors adequately capture the moneyness and maturity slopes, the smile attenuation, and the smirk. Furthermore, the IV specification is twice continuously differentiable and well-behaved asymptotically, allowing for clean interpolation and extrapolation over a wide range of moneyness and maturity. Fitting performance on Standard and Poor's 500 options compares favorably with existing benchmarks. The benefits of a smoothed IV surface are illustrated through the valuation of illiquid index derivatives, the extraction of the risk-neutral density and risk-neutral moments, and the calculation of option price sensitivities.

KEY WORDS

derivatives pricing, factor models, Greeks, implied volatility surfaces

1 | INTRODUCTION

Nearly 50 years after their introduction on trading exchanges, derivatives have become a central piece in modern asset pricing theory. Beyond their fundamental role as risk management tools, options are contingent claims whose market prices convey all the information needed to determine state prices (Cox & Ross, 1976). In conjunction with the information embedded in the underlying asset returns, these state prices can further be disentangled into natural probabilities and the pricing kernel (Jackwerth, 2000; Ross, 2015). Breeden and Litzenberger (1978) show that the risk-neutral density of the underlying asset price can be retrieved from the continuum of options across strikes. From this seminal property, a rich set of extended results has emerged in the financial economics literature, including but not limited to: the spanning of a terminal payoff with a portfolio of discount bonds and options (Bakshi & Madan, 2000; Carr & Madan, 2001), the inference of risk-neutral moments (Ammann & Feser, 2019; Bakshi & Kapadia, 2003; Conrad et al., 2013; Neumann & Skiadopoulos, 2013), the construction of the risk-neutral density (Birru & Figlewski, 2012; Figlewski, 2018), the static hedging of options (Carr & Wu, 2014), and the calculation of risk metrics, such as the Volatility Index (VIX; Neuberger, 1994), the SVIX (I. Martin, 2017; I. W. Martin & Wagner, 2019), or the rare disaster index (RIX; Gao et al., 2018, 2019).

The implementation of these applications is hampered, however, by the limited availability of traded options. Given the one-to-one correspondence between the European option premium and the implied volatility (IV) established by Black and Scholes (1973), the continuum of options across strikes and maturities can be represented by the IV surface. In practice, observable IVs form a cloud of points only, and the construction of a smoothed IV surface is an empirical challenge for academics and a mandatory daily exercise for industry practitioners who trade options. Observed implied volatilities are particularly scarce far from the money, which is precisely where information about the tails of the risk-neutral distribution can be retrieved. They are also scarce for medium and long maturities. Thus, an accurate extrapolation of the IV surface can help better estimate the high-order risk-neutral moments and how they aggregate over the time horizon.

This paper introduces a new functional form for the IV with three major benefits. First, it is designed to accommodate the well-documented, stylized features of the IV surface for Standard and Poor's 500 (S&P 500) index options. Chalamandaris and Tsekrekos (2011; hereafter CT) explicitly introduce the level, slope, and curvature factors, in the spirit of how Nelson and Siegel (1987) and Diebold and Li (2006) model the yield curve. The IV specification presented in this paper takes a step further by assigning precise roles to the factors. Specifically, the factors are explicitly designed to capture the slopes in the moneyness and the maturity dimensions, the smile attenuation (i.e., the fact that the smile convexity decreases with the maturity), and the smirk (i.e., the fact that the IVs for short-term, deep out-of-the-money (OTM) calls are higher than those at-the-money [ATM]). In particular, the specific account of the slope attenuation and the smirk greatly enhances the calibration performance on the SPX option IV surface over the January 1996–December 2019 period compared with the Heston (1993), the Goncalves and Guidolin (2006; hereafter GG), and the Chalamandaris and Tsekrekos (2011) benchmarks.

Second, the factors are tailored to admit a stable asymptotic behavior, which makes it possible to extrapolate beyond quoted moneyness levels and maturities. This feature proves to be particularly valuable when it comes to valuing long-term, illiquid, index-option-related contracts (such as structured notes). Third, the factors are constructed so that the surface is twice continuously differentiable, which produces a well-behaved risk-neutral density function for each time horizon. These last two features represent a significant improvement over factor-based benchmarks as extrapolating the IV surface has remained an empirical challenge in the literature.¹

Several smoothing methods have been proposed in the literature. They can be broadly classified into parametric and nonparametric approaches.²

Parametric smoothing allows seamless interpolation, parsimony, interpretability, and limited computational requirements for option pricing. However, the parametric specifications proposed in earlier works are chosen for their simplicity and convenience. Dumas et al. (1998) examine the IV as a polynomial function of strike and maturity. Goncalves and Guidolin (2006) apply a similar specification to the log-IV replacing the strike with a moneyness factor. Jackwerth and Rubinstein (1996) and Bliss and Panigirtzoglou (2002) use cubic splines.

The paper adds to four strands of the literature. First, it is related to the mark-to-market of thinly traded securities (Célérier & Vallée, 2017; Henderson & Pearson, 2011; Skiadopoulos, 2001). A wide variety of derivatives are not publicly traded on exchanges, for example, over-the-counter (OTC) derivatives, structured products, and options embedded in corporate securities. Nevertheless, they can be priced in a consistent manner when the valuation models are anchored to the IV surface of the vanilla options with the same underlying asset (Bayraktar & Yang, 2011; Daglish et al., 2007). The pricing consistency requires an arbitrage-free smoothed IV surface (Fengler, 2009). The IV surface specification presented in this paper successfully passes the arbitrage detection assessment using butterfly and calendar spreads (Davis & Hobson, 2007) which represents sufficient conditions for the absence of static arbitrage. A numerical application shows how a complete, smoothed IV surface streamlines the mark-to-market of an equity-index-linked note, making it less reliant on the entry and exit of available strikes and maturities. For the seller of the note, this, in turn, reduces unnecessary liquidity stress on the management of the position.

Second, the paper contributes to the extraction of the risk-neutral density and risk-neutral moments. As argued by Jackwerth (2004), the challenge in constructing the risk-neutral density does not reside in the center of the distribution but in the tails, where few options are observable. Parametric smoothing methods therefore complete the shape of the density either with composite distributions or with mixture models (Figlewski, 2018). The IV specification of this paper

¹Our numerical experiments show that CT and GG models often induce anomalous risk-neutral densities when our method does not.

²Among nonparametric smoothing methods are the Gaussian kernel (Cont & Da Fonseca, 2002), the principal component analysis (PCA; Israelov & Kelly, 2017), and the neural networks (Ackerer et al., 2019). One major limitation of nonparametric approaches is that the extrapolation of the smoothed IV surface is a nontrivial issue.

circumvents this issue by using factors that are twice continuously differentiable and well-behaved asymptotically. The generated risk-neutral densities are shown to be smooth and regular, and the mass of probability over the full spectrum of moneyness levels adds up to exactly one. Furthermore, numerical experiments show the substantial correction in the computation of risk-neutral skewness and kurtosis when the Carr and Madan (2001) formula is discretized.

Third, the specification for the IV surface proposed in this paper can be applied to the dynamic hedging of options. Delta and gamma are derived analytically and are shown to be smile-implied Greeks (Alexander & Nogueira, 2007; Bates, 2005; François & Stentoft, 2021). That is, these Greeks are consistent with the observed shape of the volatility smile and they do not depend on any assumption regarding the underlying asset dynamics. Furthermore, the specification for the IV surface allows for the enhanced management of volatility risk, not only through the traditional definition of the option vega, but also through the sensitivity of the option value with respect to the long-term volatility level and to the IV maturity slope.

The paper is structured as follows. Section 2 presents the data. Section 3 introduces the IV specification, compares its fitting performance with selected benchmarks, and checks for the presence of arbitrage opportunities in the smoothed IV surface. Section 4 details the applications of the IV surface model to derivatives pricing and risk management. Section 5 concludes.

2 | DATA

The data set extracted from the OptionMetrics database consists of European call and put options on the S&P 500 index (SPX options) quoted daily on the CBOE from January 4, 1996, to December 31, 2019.³ For each option quote, the data includes bid and ask prices, from which midprices are calculated to serve as option prices. The data also include forward prices of the S&P 500 index associated with the same maturity date for each option. For each quote, the associated option moneyness is defined as

$$M = \frac{1}{\sqrt{\tau}} \log\left(\frac{F_{t,\tau}}{K}\right), \quad (1)$$

where τ is the annualized time-to-maturity⁴ of the option, $F_{t,\tau}$ is the day- t forward price of maturity τ given by OptionMetrics, and K is the strike price. Thus, $M = 0$ corresponds to ATM options, $M < 0$ to OTM calls, and $M > 0$ to OTM puts. As time-to-maturity increases, the range of traded strike prices also widens due to the scaling property of volatility. Scaling the moneyness by $\frac{1}{\sqrt{\tau}}$ generates a comparable moneyness measure across time-to-maturity.

Data exclusion procedures are in line with the filters of Bakshi et al. (1997). Options with the following characteristics are removed: (i) a time-to-maturity shorter than 6 days, (ii) a price lower than \$3/8, (iii) a zero bid price, and (iv) options with a bid–ask spread larger than 175% of the option's midprice.⁵ Furthermore, all in-the-money options are excluded (i.e., puts with $M < 0$ and calls with $M \geq 0$). The resulting sample contains 3,468,515 option quotes spanning over 6039 days.⁶

For each option in the sample, its IV is computed with the Black (1976) formula. Table 1 presents descriptive statistics about option implied volatilities (hereafter IV) associated with all quotes from the data set. The IV surfaces from the various days covered by the data typically display asymmetry and often the well-known smirk (i.e., calls with a moneyness $M < -0.2$ having higher IV than calls with a moneyness $-0.2 < M \leq 0$). Moreover, as seen in Table 1, there are slightly fewer options with short maturities (less than or equal to 90 days) than there are options with long maturities (more than 90 days) options.

Figure 1 shows the set of all option IV considered on four selected days. These days are representative of stylized features of the IV surface. As observed when comparing the first day of the sample (Panel A) to others, the number of

³Only spot options are considered, that is, options clearing on the spot price of the S&P 500 index.

⁴For clarity of exposition, τ is reported in days in tables and figures.

⁵This filter is inspired by Azzone and Baviera (2022). Options with a large ratio of bid–ask spread over price yield inaccurate midprices, which then induce anomalies in the IV surfaces and artificial spikes in the factors' time series. Options excluded by this filter represent a minuscule proportion of the sample (0.3%).

⁶The October 9, 2006 IV surface is removed from the data set, since the IV surface displays a broken shape which most likely indicates that the data are erroneous for that specific day.

TABLE 1 Descriptive statistics of the SPX options data

| | Calls | | Puts | | | All |
|---------------------------|----------------|---------------------|----------------------|-----------------------|-----------------|-----------|
| | $M \leq -0.2$ | $-0.2 < M \leq 0$ | $0 < M \leq 0.2$ | $0.2 < M \leq 0.8$ | $M \geq 0.8$ | |
| Average IV (%) | 17.90 | 15.23 | 19.95 | 28.43 | 44.12 | 23.94 |
| Standard deviation IV (%) | 7.19 | 5.46 | 5.53 | 6.40 | 9.60 | 10.30 |
| Number of contracts | 274,252 | 774,321 | 772,735 | 1,303,428 | 279,447 | 3,404,183 |
| | $\tau \leq 30$ | $30 < \tau \leq 90$ | $90 < \tau \leq 180$ | $180 < \tau \leq 365$ | $\tau \geq 365$ | All |
| Average IV (%) | 24.68 | 23.93 | 23.88 | 24.18 | 23.55 | 23.94 |
| Standard deviation IV (%) | 13.08 | 11.27 | 10.32 | 9.50 | 8.42 | 10.30 |
| Number of contracts | 277,521 | 1,008,508 | 610,650 | 703,008 | 804,496 | 3,404,183 |

Note: Descriptive statistics of the S&P 500 index (SPX) options implied volatility (IV) daily data from January 4, 1996, to June 26, 2019 across multiple times-to-maturity and moneyness buckets. M represents the moneyness defined in Equation (1) and τ is the time-to-maturity of days.

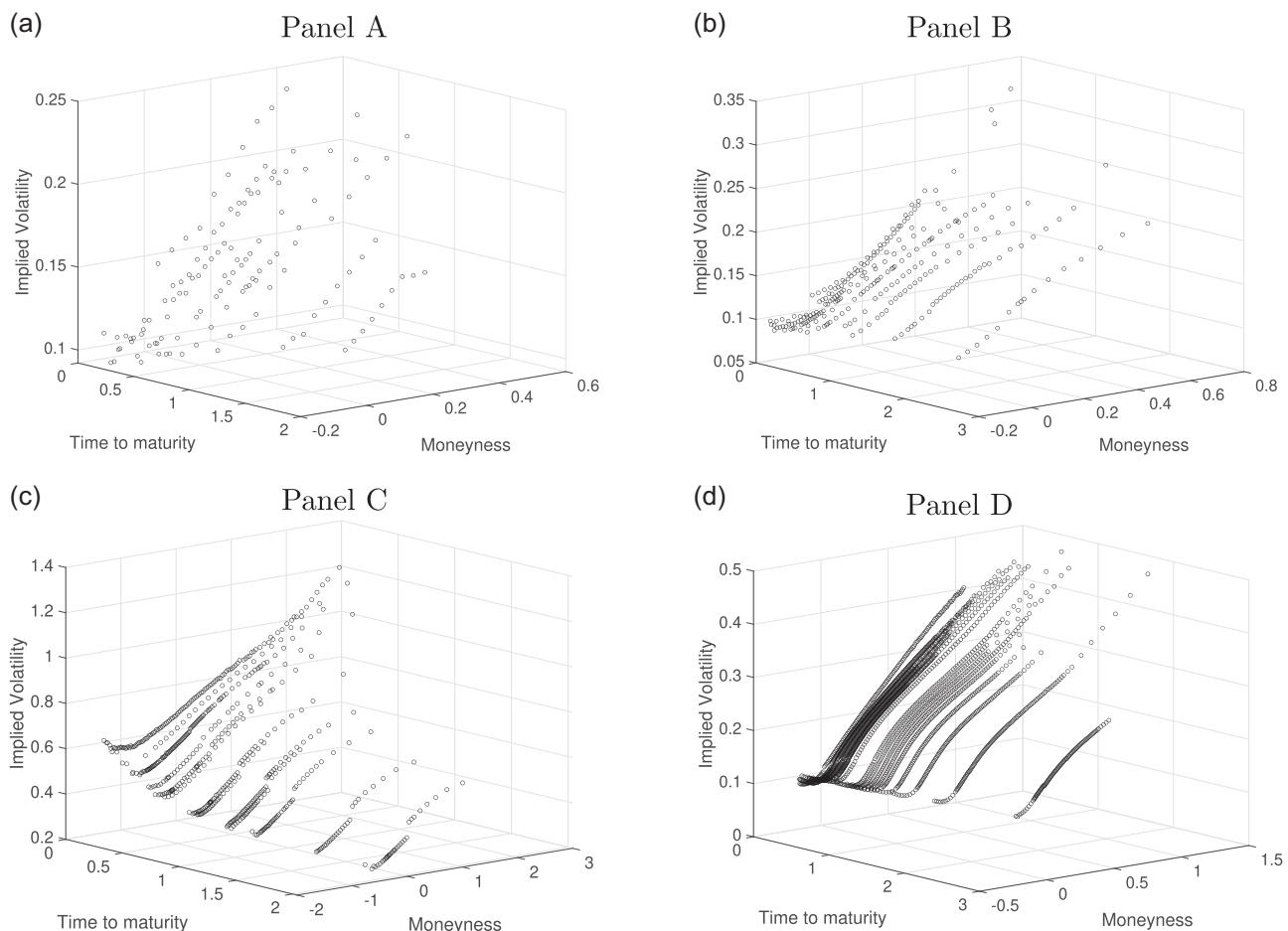


FIGURE 1 Observed IV surfaces on four selected dates: (a) Panel A, January 4, 1996; (b) Panel B, May 8, 2006; (c) Panel C, December 1, 2008; (d) Panel D, December 31, 2019. The various panels display the set of implied volatilities associated with retained option quotes from the OptionMetrics data set on four selected days. January 4, 1996, is the first day in the sample. May 8, 2006, is a low volatility day. December 1, 2008, represents the peak of the 2008 financial crisis. December 31, 2019, is the last day of the sample. The moneyness is defined in Equation (1). IV, implied volatility.

quoted options has significantly gone up with time. The increased number of quoted options mainly results in much lower strike intervals between the quotes of a given maturity, but also in additionally traded maturities. Panel B clearly displays the smile attenuation, where the moneyness slope flattens as the time-to-maturity increases. The IV surface from Panel C observed during the subprime crisis exhibits a time-to-maturity slope, that is, a downward trend of the IV with respect to the option's time-to-maturity. The volatility smirk is particularly visible in Panel D, OTM calls having a higher IV than ATM options. Such characteristics of IV surfaces are well known in the literature and are mentioned, for instance, in Cont and Da Fonseca (2002) and Rebonato (2005). Furthermore, all four panels highlight the positive impact of scaling the moneyness by $\frac{1}{\sqrt{\tau}}$ where the range of moneyness remains similar for all time-to-maturities and the IV levels becomes comparable for options with similar moneyness across time-to-maturity.

3 | MODEL SPECIFICATION AND PERFORMANCE

We opt for a static representation of the IV surface based on five factors capturing the main empirical characteristics of IV surfaces. The following three requirements are considered: (i) factor interpretability, (ii) twice differentiability, and (iii) extrapolation ability.

First, our factors are carefully selected functions of the strike and time-to-maturity. Their specification has been designed to match patterns commonly observed on IV surfaces, such as the smile attenuation or the volatility smirk. Assigning a specific role to each factor allows for a clear interpretation. It also contributes to an accurate calibration of the observed IV surfaces, as shown below. Second, the functional forms for our factors are twice continuously differentiable, which ensures the existence of a continuous risk-neutral density function for the price of the underlying asset (Breeden & Litzenberger, 1978). This, in turn, limits the presence of arbitrage opportunities. Third, our factors are asymptotically stable and can therefore be easily extrapolated beyond the observed moneyness levels and maturities. As shown in Section 4, this property is helpful for various applications, including option pricing and risk management.

In the next subsections, we introduce and discuss our IV model. We then report how the calibration of our specification strongly outperforms that of the parametric model of Heston (1993), the polynomial model of Goncalves and Guidolin (2006), and the factor model of Chalamandaris and Tsekrekos (2011). We then proceed to a formal screening procedure detecting the presence of theoretical arbitrage opportunities on both the raw data and the set of model-calibrated IV surfaces.⁷ For an agent trading at the bounds of the bid–ask spread, arbitrage opportunities turn out to be rare in the data, and even more so in the model-calibrated IV surfaces.

3.1 | A factor-based IV specification

The IV $\sigma(M, \tau)$ observed on a given day for an option with moneyness M defined in Equation (1) and time-to-maturity τ is modeled as

$$\begin{aligned} \sigma(M, \tau) = & \underbrace{\beta_1}_{\text{long-term ATM level}} + \underbrace{\beta_2 \exp(-\sqrt{\tau/T_{\text{conv}}})}_{\text{Time-to-maturity slope}} + \underbrace{\beta_3 \left(M \mathbb{1}_{\{M \geq 0\}} + \frac{e^{2M} - 1}{e^{2M} + 1} \mathbb{1}_{\{M < 0\}} \right)}_{\text{Moneyness slope}} \\ & + \underbrace{\beta_4 (1 - \exp(-M^2)) \log(\tau/T_{\text{max}})}_{\text{Smile attenuation}} + \underbrace{\beta_5 (1 - \exp(-(3M)^3)) \log(\tau/T_{\text{max}}) \mathbb{1}_{\{M < 0\}}}_{\text{Smirk}}. \end{aligned} \quad (2)$$

Fixed values T_{max} and T_{conv} are the coefficients selected based on empirical observation. T_{max} is the maximal maturity represented by the model. Although the longest time-to-maturity for options in the sample is 3 years, the value $T_{\text{max}} = 5$ years is considered herein to allow for extrapolation beyond the longest time-to-maturity. T_{conv} represents the location of a fast convexity change in the IV term structure with respect to time-to-maturity. It is set to $T_{\text{conv}} = 0.25$ for the reasons explained below.

⁷The method detects static arbitrage opportunities via calendar and butterfly spreads (Davis & Hobson, 2007; Fengler, 2009).

The five-factor representation in Equation (2) is parsimoniously designed to capture the stylized facts of the IV surface. The first factor is constant and its coefficient β_1 is a proxy for the long-term ATM IV. Since $\lim_{\tau \rightarrow 0} \lim_{M \rightarrow 0} \sigma(M, \tau) = \beta_1 + \beta_2$, the coefficient β_2 measures the time-to-maturity slope of the ATM IV. Given that the observed IV curvature over time-to-maturity is more pronounced for short-term options, the convexity correction is increased for horizons less than 3 months by relating β_2 with a nonlinear function of $\tau/0.25$. The term $\tau/0.25$ creates substantially more convexity for the short-term part (under 3 months) of the IV surface.

The third factor picks up the moneyness slope separately for put and call options. In the data, put IV increases almost linearly as the moneyness increases. Conversely, while call IV decreases at first when the moneyness decreases, it then either stabilizes or starts increasing for deeper OTM options as shown in Panels B–D of Figure 1. The hyperbolic tangent function in the third term generates a shape reminiscent of the blade of a hockey stick. It is strictly decreasing with a negative second derivative equal to 0 at $M = 0$ creating this desirable shape for call IV and generating a smooth moneyness slope function. The fourth factor accounts for the smile attenuation, that is, it ensures that the smile gets flatter as time-to-maturity increases. The coefficient β_4 also controls for the smile convexity. The observed IV surface sometimes exhibits some local concavity with respect to the moneyness. This phenomenon can be seen on December 31, 2019 but also appears on several other dates during our sample. First, it should be noted that this local concavity is not inconsistent with the absence of arbitrage. Simple calculus rules show that the call price convexity in the strike (i.e., $\frac{d^2C}{dK^2} > 0$) is equivalent to a lower bound for $\frac{d^2IV}{dK^2}$ which is not necessarily negative (see Reiswich, 2010, for a derivation). Second, recent research (Alexiou et al., 2021; Baker et al., 2018) argues that this local concavity reflects a bimodal risk-neutral distribution for the underlying and is therefore a manifestation of event risk. Lee and Sokolinskiy (2015) show that standard stochastic volatility models, like, Heston (1993) cannot generate this local concavity. This is not the case for our IV specification. Indeed, for $M > 0$, $\frac{d^2}{dM^2}\sigma(M, \tau) = -\beta_4(4M^2 - 2)\exp(-M^2)\log\left(\frac{\tau}{T_{\max}}\right)$, which is not necessarily positive. Finally, the fifth factor captures the tilt in the smile for deep-OTM calls, referred to as the IV smirk. The smirk factor is designed to fade away as time-to-maturity increases toward the T_{\max} bound.

Figure 2 compares the five factors of Equation (2) with those extracted from a PCA decomposition of IV surfaces.⁸ The use of PCA for the representation of volatility surfaces is motivated by the seminal paper of Cont and Da Fonseca (2002) proposing a Karhunen–Loève decomposition of the log-surface dynamics generating orthogonal factors. Directly applying PCA to the option sample is not possible since option numbers and characteristics (moneyness and time-to-maturity) vary from day to day. Because PCA requires a stable sample every day, a grid with respect to moneyness and time-to-maturity is constructed over the densely populated regions of the surface. When applied to our data set, the PCA is fitted to a subsurface inside which only 41% of observed options lie. There are similarities between our five factors (left panels) and the first five PCA factors (right panels). Given that the PCA generates by construction the linear factors fitting the IV surface best in terms of root-mean-square error (RMSE), such similarity entails that Model (2) has the potential to adequately capture the IV surface patterns without requiring the construction of a subgrid of moneyness/time-to-maturity and discarding an important part of the observations as in the PCA approach.

3.2 | Daily calibration

As described in Appendix B, a daily re-estimation of the set of coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ is obtained by minimizing the sum of the squared fitting errors while integrating Bayesian information for regularization purposes. Calibration may be performed on option pricing errors, such as in Azzone and Baviera (2022) or Schoutens et al. (2003). Alternatively, we choose to minimize the distance between model and observed IVs because the latter is of the same order of magnitude across moneyness and maturity. Figure 3 plots the time series of estimated coefficients. The level of the ATM IV (panel A) remains low during quiet market conditions. It spikes during financial turmoil, in particular around the LTCM crisis in September 1998 and, most notably, after the Lehman Brothers collapse by the end of 2008. The time-to-maturity slope of the IV surface (panel B) is usually slightly positive (i.e., negative β_2), but the 2008 financial crisis is associated with a strong and short-lived slope inversion. The smile asymmetry appears to be more stable in the second half of the sample with a less volatile coefficient β_3 (panel C). The smile attenuation β_4 varies

⁸The PCA approach is described in Appendix A.

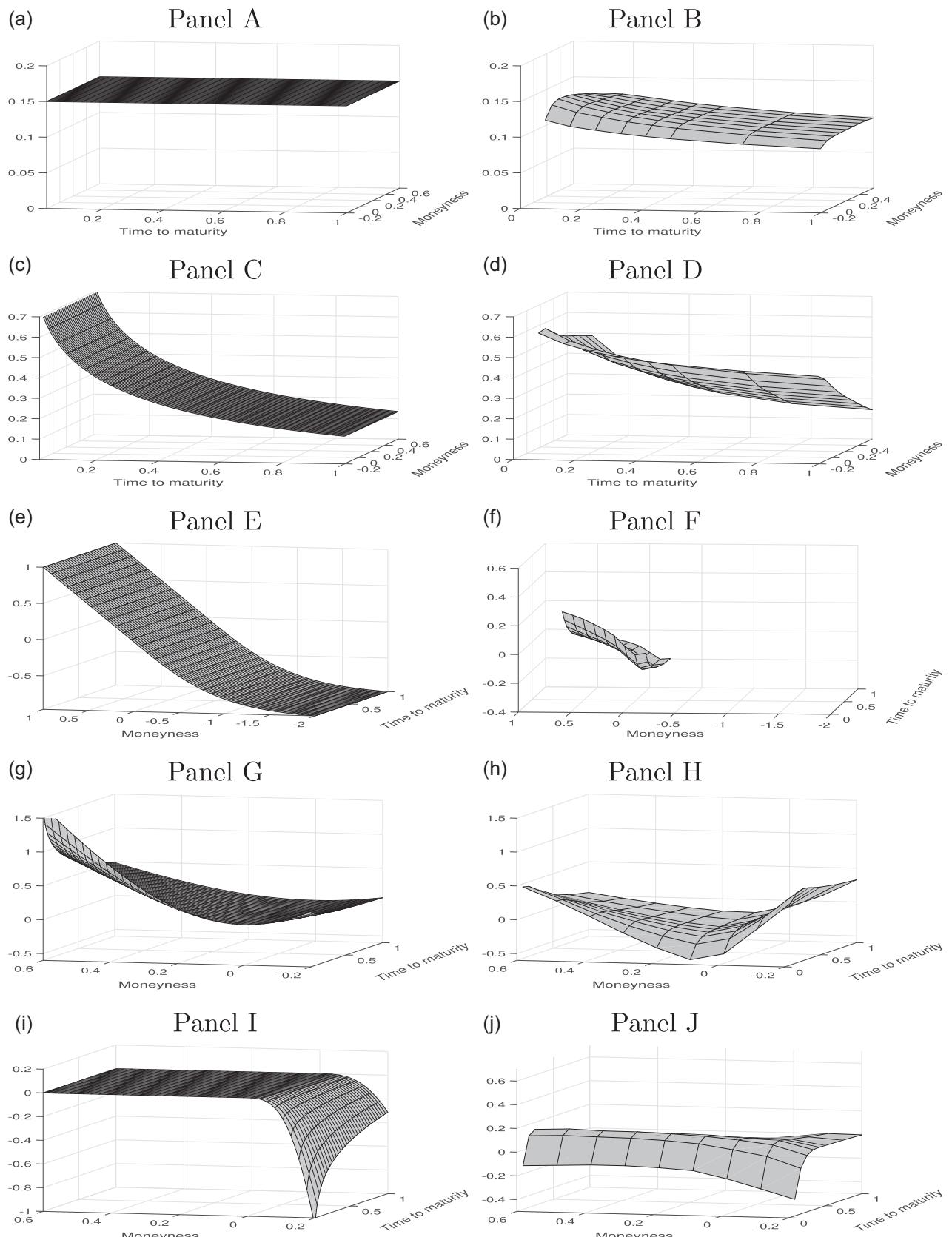


FIGURE 2 (See caption on next page)

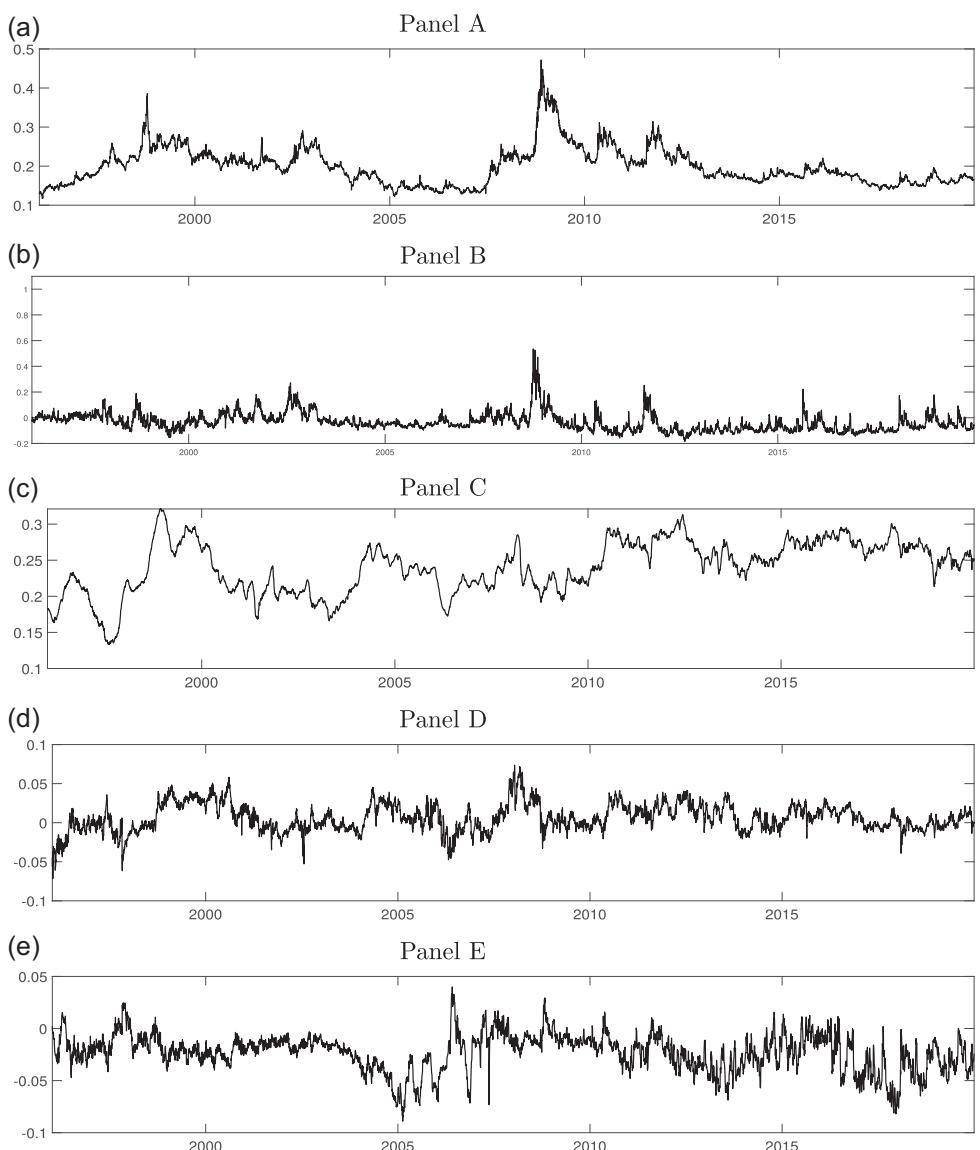


FIGURE 3 Daily coefficient estimates for the IV surface model: (a) Panel A, long-term ATM level; (b) Panel B, time-to-maturity slope; (c) Panel C, moneyness slope; (d) Panel D, smile attenuation; (e) Panel E, moneyness smirk. Daily evolution of estimates of coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ of Model (2), respectively, capturing the long-term ATM level, the time-to-maturity slope, the moneyness slope, the smile attenuation, and the moneyness smirk, from January 4, 1996, to June 26, 2019. The estimates are obtained by minimizing the sum of the squared fitting errors while integrating Bayesian information (Appendix B). ATM, at-the-money; IV, implied volatility.

FIGURE 2 Model factors versus PCA factors: (a) Panel A, long-term ATM level; (b) Panel B, factor 1; (c) Panel C, time-to-maturity slope; (d) Panel D, factor 2; (e) Panel E, moneyness slope; (f) Panel F, factor 3; (g) Panel G, smile attenuation; (h) Panel H, factor 4; (i) Panel I, moneyness smirk; (j) Panel J, factor 5. The five factors of Equation (2) are represented on the left panels. The right column panels present the first five factors obtained through a PCA applied on the option data as described in Appendix A. The proportions of variation explained by each factor are 94.84%, 3.32%, 1.45%, 0.22%, and 0.16%. As explained in Appendix A, these proportions are calculated on a reduced sample that only includes 41% of the option data, and therefore must be interpreted as such. The top four panels are displayed using an angle highlighting the factor variation with respect to the time-to-maturity, while the remaining six bottom panels show the variation with respect to the moneyness. Panels E and F have a larger moneyness axis than the other panels to show the entire shape of the third factor. As mentioned in Appendix A, the PCA approach requires IV values at all points of a two-dimensional lattice: the moneyness M ranges between -0.2 and 0.6 by increments of 0.1 and the maturities are 30, 60, 91, 122, 152, 182, 270, and 365 days. ATM, at-the-money; IV, implied volatility; PCA, principal component analysis.

more at the beginning of the sample and during the financial crisis (panel D). Finally, the smirk effect seems more pronounced but also more volatile over the most recent period (panel E). The dynamics of the coefficients is complex: in addition to the obvious dependence between them, each β coefficient is strongly correlated with its lagged value as well as with those of other coefficients. The variability of the coefficients changes over time. Their distribution is asymmetrical, the level and the time-to-maturity slope exhibiting more high positive shocks than negative ones.

3.3 | Benchmarking

The calibration performance of Model (2) is compared with that of three benchmarks: the polynomial model of Goncalves and Guidolin (2006), the factor model of Chalamandaris and Tsekrekos (2011), and the Heston (1993) parametric model.

Goncalves and Guidolin (2006), hereafter GG, regress the daily log IV surface on five factors:

$$\log \sigma(\tilde{M}, \tau) = \delta_1 + \delta_2 \tilde{M} + \delta_3 \tilde{M}^2 + \delta_4 \tau + \delta_5 (\tilde{M}\tau),$$

where their moneyness is defined as $\tilde{M} = \frac{1}{\sqrt{\tau}} \ln \left[\frac{K}{S \exp(r\tau)} \right]$, with S being the underlying asset price and r being the annualized risk-free rate. Working with log-volatilities ensures a positive IV surface. The various factors clearly represent moneyness and maturity slopes along with their interaction, on top of a convexity factor for the moneyness. The daily coefficients are obtained by minimizing the squared difference between the model and observed IVs.

Chalamandaris and Tsekrekos (2011), hereafter CT, fit the daily IV surface of foreign exchange options using seven factors:

$$\begin{aligned} \sigma(m, \tau) = & \theta_1 + \theta_2 1_{m>0} m^2 + \theta_3 1_{m<0} m^2 + \theta_4 \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \\ & + \theta_5 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \theta_6 1_{m>0} m\tau + \theta_7 1_{m<0} m\tau, \end{aligned} \quad (3)$$

where $m = (\Delta - 0.5) \times 100$ and Δ represents the Black–Scholes option delta.⁹

The first term of Equation (3) is the level of the surface. The second and third terms account for the right and left smiles. The curvature of the IV surface at the short and at the medium maturities is picked up by the fourth and fifth terms, respectively. The sixth and seventh terms capture the right and left smile attenuation.

Another approach to characterize IV surfaces is the family of stochastic volatility models. Heston (1993) is one of the widely used representatives. The risk-neutral dynamics is

$$\begin{aligned} dS_t &= (r - q)S_t dt + \sqrt{v_t} S_t dW_t, \\ dv_t &= \kappa(\theta - v_t)dt + \xi \sqrt{v_t} dB_t, \end{aligned}$$

where the correlated Brownian motions W and B capture part of the dependence between the underlying asset price S_t and instantaneous variance v_t . In the Heston (1993) framework, the values of European options are available in quasianalytical form, and are transformed into IVs by inverting the Black and Scholes (1973) formula.

The Heston (1993) model parameters as well as the instantaneous variance are calibrated daily by minimizing the squared fitting errors between the model's generated IVs and the observed IVs. This calibration procedure is consistent

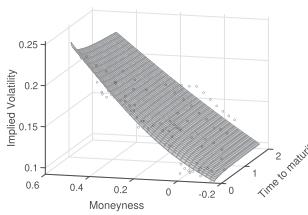
⁹The CT study uses the Garman and Kohlhagen (1983) delta defined in the context of the foreign exchange market. The call option Δ is computed as follows:

$$\begin{aligned} \Delta &= \frac{\partial c(K, \tau)}{\partial S} = e^{(d-r)\tau} \phi(d_1(K)), \\ d_1(K) &= \frac{1}{\sigma_{K,\tau}^{\text{obs}} \sqrt{\tau}} \log \frac{F_{0,\tau}}{K} + \sigma_{K,\tau}^{\text{obs}} \sqrt{\tau}. \end{aligned}$$

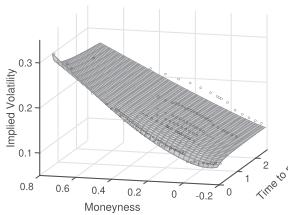
where d is the foreign risk-free rate. The $\sigma_{K,\tau}^{\text{obs}}$ used when fitting the model to the observed IV surfaces corresponds to the observed option IV.

Model (2)

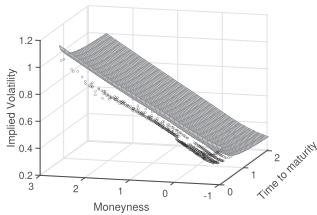
January 4, 1996



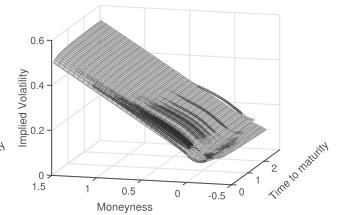
May 8, 2006



December 1, 2008

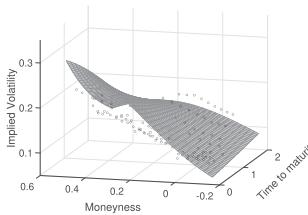


December 31, 2019

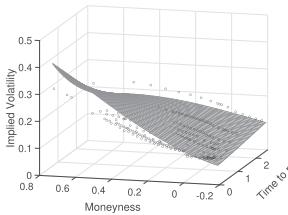


GG model

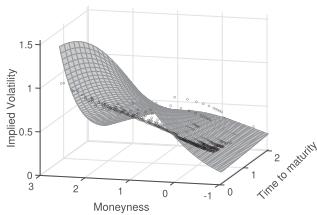
January 4, 1996



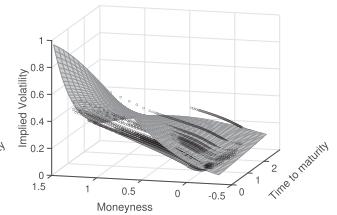
May 8, 2006



December 1, 2008

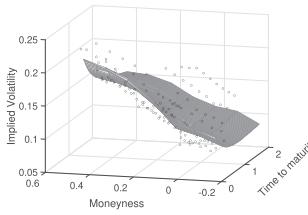


December 31, 2019

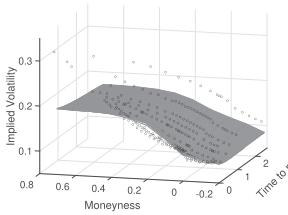


CT model

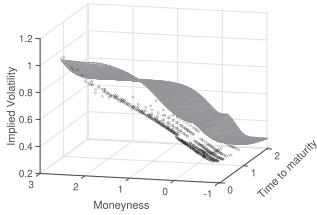
January 4, 1996



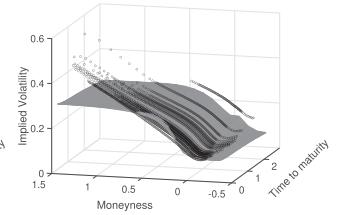
May 8, 2006



December 1, 2008

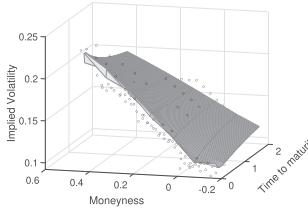


December 31, 2019

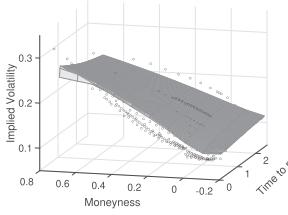


Heston model

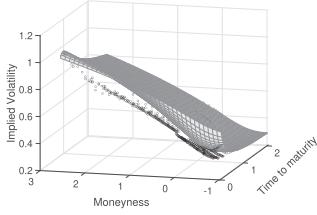
January 4, 1996



May 8, 2006



December 1, 2008



December 31, 2019

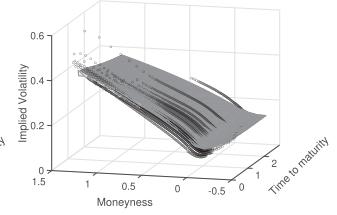


FIGURE 4 Model (2) fitted surfaces compared with the benchmark surfaces. Observed IVs are plotted against the fitted IV surfaces derived from Model (2)—first row—and the three benchmark models in rows 2–4. GG refers to Goncalves and Guidolin (2006) and CT stands for Chalamandaris and Tsekrekos (2011). The Heston model parameters have been calibrated on a daily basis. January 4, 1996 is the first day in the sample. May 8, 2006 is a low volatility day. December 1, 2008 represents the peak of the 2008 financial crisis. December 31, 2019 is the last day of the sample. IV, implied volatility.

with that used for the other benchmarks. It allows for a fair comparison of the fitting performance across the benchmarks using the RMSE as the relevant metric.

3.4 | Calibration performance

Figure 4 illustrates how Model (2) and the CT, GG, and Heston benchmarks fit the IV data points for the four selected dates of Figure 1.

Inspection of Figure 4 confirms the well-behaved extrapolation of the IV surface using Model (2). In particular, the shape of the IV surface remains consistent with the few implied volatilities observed in extreme regions of time-to-

maturity and moneyness. By contrast, extrapolation of the IV surface using the GG benchmark induces a twist in the maturity slope that does not fit with the data. Likewise, extrapolation of the IV surface using the CT benchmark often induces a cap on far from the money implied volatilities, which, again, is not in line with market observations.

The calibration performance of each model over the entire sample is assessed by computing the daily RMSE:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\sigma(M_i, \tau_i) - \sigma_i^{Obs})^2},$$

where σ_i^{Obs} is the observed IV for the i th quote available on that day, and M_i and τ_i are its associated moneyness and time-to-maturity, respectively. We also compute the average relative percentage error (ARPE):

$$ARPE = \frac{1}{N} \sum_{i=1}^N \frac{|O(M_i, \tau_i) - O_i^{Obs}|}{O_i^{Obs}},$$

where O_i^{Obs} is the observed option midprice for the i th quote available and $O(M_i, \tau_i)$ is the model option price described in Equations (4) and (5).

Figure 5 shows that the daily RMSE and ARPE of Model (2) are lower than those of the GG, CT, and Heston models for most of the sample period. Despite the increase in the number of quoted option contracts in more recent periods, the fitting performance of Model (2) does not deteriorate over time. This is in sharp contrast with the GG, CT, and Heston benchmarks. Note that since the GG model is estimated on the log IV surfaces and the measurement error is

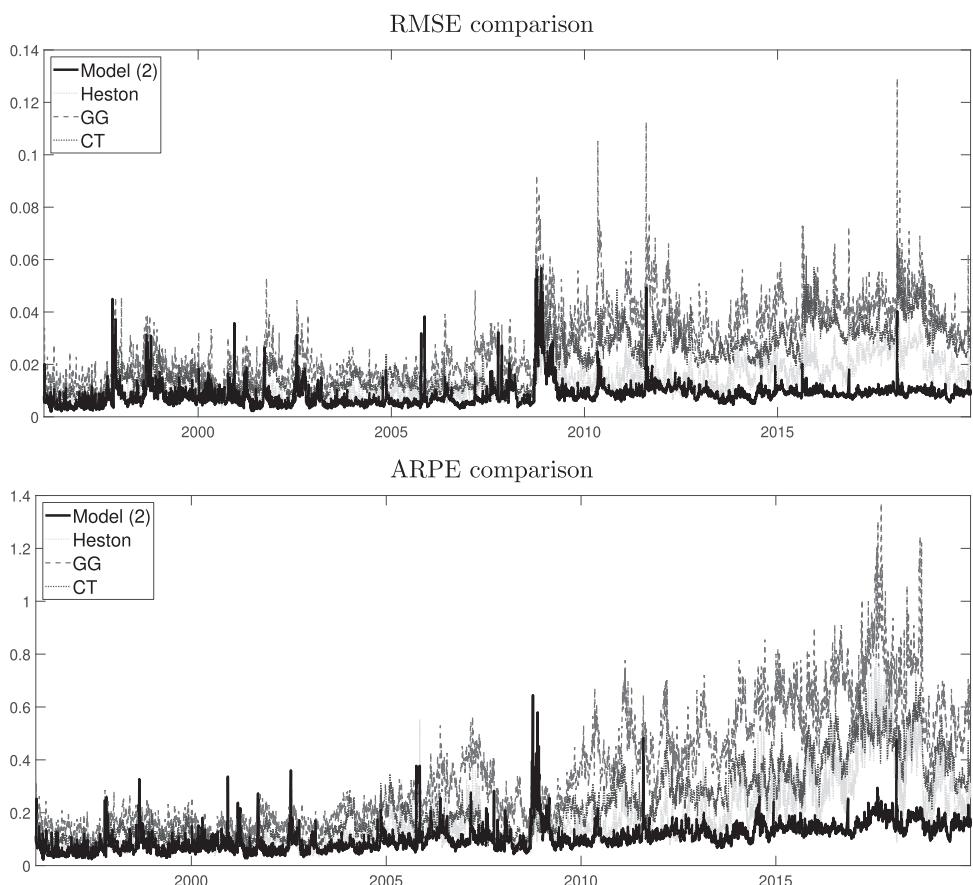


FIGURE 5 RMSE and ARPE across time. The daily RMSE and ARPE are reported for Model (2), and the CT, the GG, and the Heston (1993) benchmarks. ARPE, average relative percentage error; CT, Chalamandaris and Tsekrekos; GG, Goncalves and Guidolin; RMSE, root-mean-square error.

TABLE 2 Average RMSE and ARPE over time from IV surface estimation

| Panel A: RMSE | | | | |
|----------------------|--|---|--------------------------------------|------------|
| Model | $M \leq -0.1$ (call) | $-0.1 < M \leq 0.1$ | $M \geq 0.1$ (put) | All |
| Model (2) | 0.0126 | 0.0083 | 0.0099 | 0.0101 |
| CT | 0.0114 | 0.0151 | 0.0366 | 0.0291 |
| GG | 0.0203 | 0.0141 | 0.0482 | 0.0381 |
| Heston | 0.0140 | 0.0106 | 0.0217 | 0.0182 |
| Number of contracts | 621,839 | 838,754 | 1,943,590 | 3,404,183 |
| Model | $\tau \leq 60$ | $60 < \tau \leq 180$ | $\tau \geq 180$ | All |
| Model (2) | 0.0104 | 0.0081 | 0.0112 | 0.0101 |
| CT | 0.0347 | 0.0301 | 0.0246 | 0.0291 |
| GG | 0.0505 | 0.0241 | 0.0385 | 0.0381 |
| Heston | 0.0303 | 0.0133 | 0.0111 | 0.0182 |
| Number of contracts | 813,778 | 1,082,901 | 1,507,504 | 3,404,183 |
| Panel B: ARPE | | | | |
| Model | $M \leq -0.1$ (call) | $-0.1 < M \leq 0.1$ | $M \geq 0.1$ (put) | All |
| Model (2) | 0.2476 | 0.0669 | 0.1155 | 0.1276 |
| CT | 0.2573 | 0.1081 | 0.4002 | 0.3021 |
| GG | 0.6084 | 0.1339 | 0.5609 | 0.4643 |
| Heston | 0.5175 | 0.0933 | 0.2208 | 0.2436 |
| Number of contracts | 621,839 | 838,754 | 1,943,590 | 3,404,183 |
| Model | $\tau \leq 60$ | $60 < \tau \leq 180$ | $\tau \geq 180$ | All |
| Model (2) | 0.1277 | 0.1202 | 0.1329 | 0.1276 |
| CT | 0.4275 | 0.3167 | 0.2239 | 0.3021 |
| GG | 0.5958 | 0.3880 | 0.4482 | 0.4643 |
| Heston | 0.4030 | 0.2347 | 0.1639 | 0.2436 |
| Number of contracts | 813,778 | 1,082,901 | 1,507,504 | 3,404,183 |

Note: Panel A reports the average RMSE and Panel B the average ARPE, for each bucket of moneyness (M) and days to maturity (τ). The sample period is January 4, 1996–June 26, 2019.

Abbreviations: ARPE, average relative percentage error; CT, Chalamandaris and Tsekrekos; GG, Goncalves and Guidolin; IV, implied volatility; RMSE, root-mean-square error.

based on the IV level, the RMSE does not only capture the variability of the error term, but also the convexity bias caused by Jensen's inequality. The drop in performance of the CT, GT, and Heston models following the 2008 financial crisis is linked to the wider range of moneyness levels traded after the crisis. Due to the asymptotic behavior of the models, capturing the surface shape with deeper OTM options becomes an issue. The ARPE leads to the same conclusions. This measure is heavily influenced by deep-OTM options due to the denominator being very small in their case, which explains the magnitude of ARPEs reported in the second panel of Figure 5. The gap between Model (2) and all other benchmarks is even wider for the second half of the sample in part because there are more quoted deep-OTM options.

Panel A of Table 2 displays the average RMSE across time over different subregions of the IV surface. The average RMSE across the entire surface for the GG and CT benchmarks is 3–4 times larger than that of Model (2) and twice as large in the case of the Heston model. The quality of the fit obtained for the CT and GG benchmarks is strongly sensitive to moneyness and time-to-maturity. In particular, the two models poorly match the OTM put implied

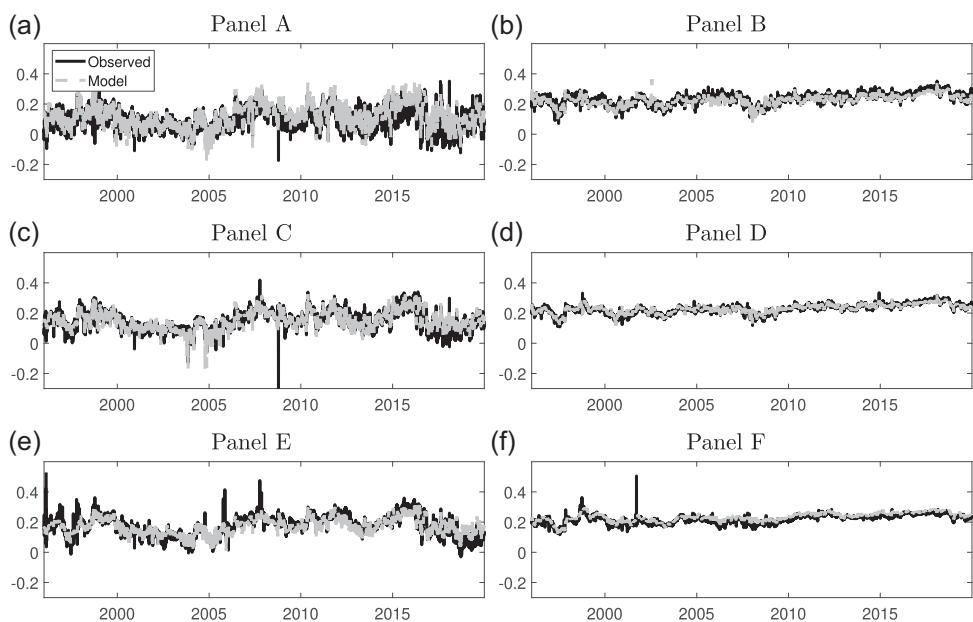


FIGURE 6 Implied volatility skew for various buckets of time-to-maturity: (a) Panel A, IV skew calls, $\tau \leq 60$; (b) Panel B, IV skew puts, $\tau \leq 60$; (c) Panel C, IV skew calls, $60 < \tau \leq 180$; (d) Panel D, IV skew puts, $60 < \tau \leq 180$; (e) Panel E, IV skew calls, $180 < \tau \leq 365$; (f) Panel F, IV skew puts, $180 < \tau \leq 365$. The daily time series of estimated IV skew for OTM calls and puts is reported over the sample period (1996–2019). The black line represents the IV skew estimated using the observed IV surfaces, while the gray line shows the IV skew estimated with Model (2) fitted IV surfaces. IV, implied volatility; OTM, out-of-the-money.

volatilities. Interestingly, the average RMSEs associated with the Model (2) specification are of similar magnitude across all moneyness and time-to-maturity buckets. Moreover, Panel B shows that Model (2) also outperforms the benchmarks in terms of average ARPE. Furthermore, relative pricing errors are larger for deep-OTM options. The calibration performance is excellent when the short-term volatility is low. The errors (RMSE and ARPE) are about twice as large when the short-term volatility reaches very high levels, which is the case during market turmoils (short-term deep-OTM calls are the most affected).

Figure 6 illustrates the estimated IV skews for both calls and puts at various time-to-maturities. To estimate the IV skew, the IV slopes between an ATM put (call) option and the options with the 20% lowest (highest) strikes are computed for each available time-to-maturity. The IV slopes are then averaged by buckets of time-to-maturity to produce the reported IV skews.¹⁰ Results indicate that the time series of IV skews estimated using the daily observed IVs and the Model (2)'s fitted surfaces are very similar for both calls and puts across every bucket of time-to-maturity.

3.5 | Arbitrage opportunities

To verify if Model (2) generates prices consistent with no-arbitrage principles, a screening procedure inspired by the work of Davis and Hobson (2007) is applied. The detection of prices violating no-arbitrage restrictions is performed on both the calibrated surfaces and sample observations. The comparison of the number of such violations in both data sets serves as a sanity check to assess the propensity of Model (2) to either (i) generate prices incompatible with the absence of arbitrage, or (ii) smooth out arbitrage opportunities found in the data.

¹⁰The slope between an ATM and OTM option is computed as follows:

$$\text{slope} = \frac{\sigma(\tau, M_{\text{OTM}}) - \sigma(\tau, M_{\text{ATM}})}{M_{\text{OTM}} - M_{\text{ATM}}},$$

where $\sigma(\tau, M)$ is the option IV. Due to the considered definition of the moneyness M (i.e., M increases when the strike price decreases), the IV skew is positive. If the slope was computed using the option strike price instead, the sign of the slope would be negative.

Davis and Hobson (2007) study instances of the so-called static arbitrage opportunities, which is a convenient relaxation of the general no-arbitrage theory outlined, for instance, in Delbaen and Schachermayer (1994). The distinction between both types of arbitrages lies in the difference between information sets (see Carr & Madan, 2005 for a more thorough discussion). Davis and Hobson (2007) provide sufficient conditions precluding the presence of static arbitrage opportunities within a point-in-time set of European call option prices for several strikes and maturities on a single underlying asset. They show that the absence of no-arbitrage violations within prices of a set of butterfly spread and calendar spread portfolios ensures that the entire set of option prices is arbitrage-free. The approach for the construction of such spread portfolios is described in Appendix D.

When applying the screening on observed option quotes, boundaries of the bid–ask interval are used instead of midprices. More precisely, whenever a call option is purchased (sold) in the construction of the spread portfolio, the ask (bid) price is considered. Indeed, using midprices instead would have led to flagging spurious arbitrage opportunities in the data which would have been impossible to realize due to the limited ability to trade within the bid–ask range. Conversely, when detecting no-arbitrage violations among calibrated surfaces, prices generated by Model (2) are used without any correction for illiquidity considerations. Such discrepancy in arbitrage detection is a conservative choice as it puts more stringent requirements on Model (2) for its prices to satisfy no-arbitrage constraints during the screening. Moreover, the Davis and Hobson (2007) methodology is based on call option prices, whereas the current sample contains quotes for both call and put options. Thus, to screen for arbitrage opportunities, put option bid and ask quotes in the data are transformed into call ask and bid prices using the put–call parity.

Table 3 displays the number of arbitrage opportunities detected in the entire data sample and on fitted surfaces. Numbers provided are aggregates across all dates of the data sample. For each date, one butterfly spread arbitrage check is performed for each option in the data set. Moreover, we investigate arbitrage regarding one calendar spread per option in the data set, with the exception of options for which the construction of the calendar spread is impossible due to the lack of traded options that would have been needed for inclusion in the spread portfolio. For instance, options whose maturity is the last one available on a given day are not checked for calendar spread arbitrage. The numbers provided in Table 3 thus include both the butterfly and calendar arbitrage counts. Results show that the calibrated surfaces exhibit fewer arbitrage opportunities than do quotes from the data sample in all buckets of moneyness and time-to-maturity. This result provides reassurance about the suitability of factors designed in Model (2). Indeed, the model tends to correct for the arbitrage opportunities found in the data while it generally avoids producing prices violating no-arbitrage constraints.

TABLE 3 Detected static arbitrage opportunities

| | Observed prices | | Fitted surfaces | | Number of checks |
|----------------------|--------------------|----------------------|--------------------|----------------------|------------------|
| | Arbitrage detected | % Arbitrage detected | Arbitrage detected | % Arbitrage detected | |
| $\tau \leq 60$ | | | | | |
| $M \leq 0$ | 603 | 0.260 | 363 | 0.002 | 231,839 |
| $0 < M \leq 0.3$ | 714 | 0.307 | 0 | 0 | 232,572 |
| $M > 0.3$ | 134 | 0.038 | 0 | 0 | 349,470 |
| $60 < \tau \leq 180$ | | | | | |
| $M \leq 0$ | 637 | 0.197 | 22 | 0 | 322,562 |
| $0 < M \leq 0.3$ | 1263 | 0.296 | 0 | 0 | 426,227 |
| $M > 0.3$ | 1966 | 0.451 | 0 | 0 | 435,698 |
| $\tau > 180$ | | | | | |
| $M \leq 0$ | 1117 | 0.226 | 0 | 0 | 494,172 |
| $0 < M \leq 0.3$ | 1954 | 0.378 | 0 | 0 | 516,397 |
| $M > 0.3$ | 11,333 | 2.332 | 0 | 0 | 486,033 |

Note: Summary statistics on violations per moneyness and time-to-maturity buckets of no-arbitrage constraints on butterfly spreads and calendar spreads designed as per the methodology outlined in Section 3.5 inspired by Davis and Hobson (2007). The numbers and proportions of violations, which are aggregated across all dates of the sample, are reported for both the data sample and fitted surfaces obtained with Model (2).

4 | APPLICATIONS OF THE VOLATILITY SURFACE MODEL

Having a complete surface with implied volatilities available on a large range of moneyness and maturity has several practical applications, some of which are presented in this section. Two main categories of applications are outlined: derivatives pricing and risk management.

4.1 | Derivatives pricing applications

The most direct application of the volatility surface model developed herein is the pricing of financial derivatives. The model can be applied in conjunction with three main approaches each described subsequently: direct interpolation or extrapolation of the IV, the Carr and Madan (2001) formula, and the extraction of the underlying asset price risk-neutral density. The three techniques are tailored to different classes of derivatives, which explains why all three are necessary.

Within the Model (2) framework, option prices are obtained by substituting the IV $\sigma(M, \tau)$ into the Black–Scholes formula. More precisely, the Black (1976) formula using the forward price instead of the underlying asset price is considered. Indeed, both the forward-based and underlying-based pricing formulas are equivalent in theory, but the former takes advantage of the OptionMetrics data set which provides forward prices rather than underlying prices, and it allows circumventing the cumbersome task of performing a daily extraction of implied dividend rates for the various option maturities. Because the moneyness (1) can be inverted to retrieve the strike price through $K = F_{0,\tau} e^{-\sqrt{\tau}M}$, the call¹¹ and put prices are

$$c(M, \tau) = e^{-r\tau} F_{0,\tau} [\Phi(\delta_1(M)) - e^{-\sqrt{\tau}M} \Phi(\delta_2(M))], \quad (4)$$

$$p(M, \tau) = e^{-r\tau} F_{0,\tau} [-\Phi(-\delta_1(M)) + e^{-\sqrt{\tau}M} \Phi(-\delta_2(M))], \quad (5)$$

where Φ is the standard normal cumulative distribution function,

$$\delta_1(M) = \frac{M}{\sigma(M, \tau)} + \frac{1}{2}\sigma(M, \tau)\sqrt{\tau} \quad \text{and} \quad \delta_2(M) = \frac{M}{\sigma(M, \tau)} - \frac{1}{2}\sigma(M, \tau)\sqrt{\tau}.$$

4.1.1 | Interpolation and extrapolation of the IV

Derivatives traded OTC often lack liquidity. Their pricing through mark-to-market procedures can therefore be a challenging exercise. Determining prices of illiquid derivatives is needed for several reasons: balance sheet assessment and corresponding risk metrics calculation, financial statement reporting, or margin calls determination in the presence of compensation by a clearinghouse.

Figure 1 shows that only a few maturities are actively traded on any day. A vanilla option whose strike or maturity is not quoted publicly on an exchange is considered illiquid. For such a contract, the pricing is made completely seamless by Model (2) as the IV can be directly obtained by substituting the option moneyness and maturity in the

¹¹Depending on the application, the call option prices are sometimes expressed as a function of the strike price instead of the moneyness (1). In the Model (2) framework, the call price becomes

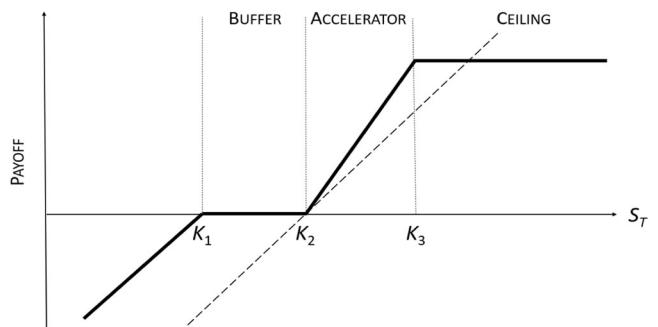
$$C(K, \tau) = \exp(-r\tau)(F_{0,\tau} \Phi(d_1(K)) + K \Phi(d_2(K)))$$

with

$$d_1(K) = \frac{1}{\sigma \left(\frac{1}{\sqrt{\tau}} \log \frac{F_{0,\tau}}{K}, \tau \right) \sqrt{\tau}} \log \frac{F_{0,\tau}}{K} + \sigma \left(\frac{1}{\sqrt{\tau}} \log \frac{F_{0,\tau}}{K}, \tau \right) \sqrt{\tau},$$

$$d_2(K) = d_1(K) - \sigma \left(\frac{1}{\sqrt{\tau}} \log \frac{F_{0,\tau}}{K}, \tau \right) \sqrt{\tau}.$$

FIGURE 7 Typical index-linked Standard and Poor's 500 note payoff function



latter formula. The computational effort required is close to nil, which makes the approach extremely convenient for the quick valuation of a large portfolio of derivatives.

Consider, for instance, the mark-to-market of an index-linked S&P 500 note. A typical terminal payoff is displayed in Figure 7. There are three thresholds: K_1 for the “buffer” region, K_2 for the “accelerator” region, and K_3 for the “ceiling” region. The terminal payoff X_τ can be replicated with

$$X_\tau = S_\tau - C(K_1, 0) + \alpha C(K_2, 0) - \alpha C(K_3, 0), \quad (6)$$

where S_τ is the time- τ underlying asset price and α is the return enhancement factor. The price of calls whose payoff appears in (6) are often illiquid and thus can seldom be traded directly, which complicates the valuation of the contract. Nevertheless, the IV surface model (2) allows for continued, accurate mark-to-market of the note.

The left panels of Figure 8 present the estimated note price using either the quoted options or Model (2) IV surfaces, whereas the right panels contain the difference between the two approaches. The rows correspond to the pre-2008 crisis period (first row), the financial turmoil (second and third row), and its aftermath (last row). Because only a limited number of maturities and strikes are quoted, it is often not possible to observe quotes for options whose characteristics match exactly these of options embedded in the note, an issue for the model-free approach. To remedy this issue, two alternative notes are priced for the two closest (smaller and larger) maturities using the closest strikes available for the buffer, the accelerator, and the ceiling. The desired note price is then computed by linear interpolation. Due to liquidity issues, some strike prices appear and disappear during the note's life, sometimes generating undesired price variation which can be witnessed in Panels A, C, E, and G of Figure 8. This phenomenon does not seem to be related to the economic cycle and can appear at any moment during the note's life.

These undesired price variations can result in unnecessary, large, and erroneous mark-to-market movements. For example, in the two first rows of Figure 8, the note price computed with the model-free approach spikes on two occasions at the end of 2003 and 2007. Panels A and G also display long periods where note prices estimated with quoted options are either overvalued or undervalued. However, using Model (2)'s implied surfaces generate more stable note prices during the note's life span.

4.1.2 | The Carr and Madan formula and its discretization

For the pricing of several other contingent claims which are not call and put options, the interpolation and extrapolation approach serves as a building block within the Carr and Madan (2001) methodology. Such a method relies on options prices for a continuum of strike prices, which is provided by Model (2).

Carr and Madan (2001) show that any twice differentiable payoff function can be evaluated using infinitely many OTM put and call option prices with the same time-to-maturity as the payoff horizon. In practice, the valuation of such a payoff is applied in a model-free fashion using a discrete set of traded options. The volatility surface of Model (2) improves the numerical implementation in two ways. First, the integrals involving OTM options can be truncated at levels of moneyness beyond those taken from the data. Second, these same integrals do not need to be discretized with respect to the strike price dimension.

The main result of Carr and Madan (2001) is adapted herein to provide an integration with respect to the moneyness (1) rather than the strike price. In what follows, the cash-and-carry relationship $F_{t,\tau} = S_t \exp((r - q)\tau)$ is assumed to hold, where r is the risk-free rate and q is the continuously compounded dividend yield.

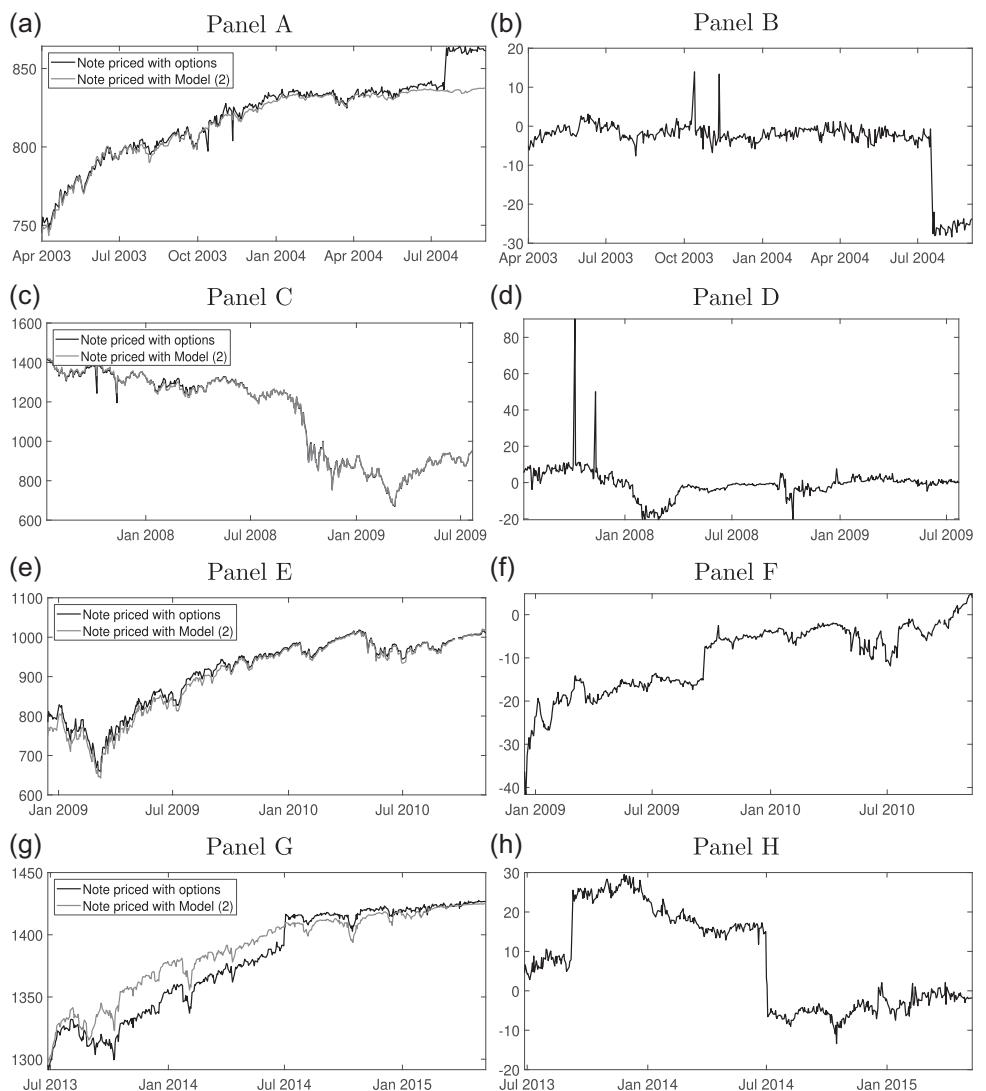


FIGURE 8 Mark-to-market of structured notes. Left panels display the price of a note estimated using (i) the closest observed quoted option prices in conjunction with linear interpolation and (ii) Model (2). Right panels display the price difference between the two methods. The first row is a note with a maturity of 550 days and strikes $K_1 = 775$, $K_2 = 850$, and $K_3 = 975$. The second row is a note with a maturity of 800 days and strikes $K_1 = 1425$, $K_2 = 1550$, and $K_3 = 1675$. The third row is a note with a maturity of 730 days and strikes $K_1 = 1000$, $K_2 = 1200$, and $K_3 = 1400$. The last row is a note with a maturity of 730 days and strikes $K_1 = 1350$, $K_2 = 1550$, and $K_3 = 1700$.

For a twice differentiable payoff function f with second derivative f'' , the Carr and Madan (2001) formula becomes

$$\begin{aligned} e^{-r\tau} E^Q[f(S_\tau)] &= e^{-r\tau} f(F_{0,\tau}) + \sqrt{\tau} F_{0,\tau} \int_{-\infty}^0 f''(F_{0,\tau} e^{-\sqrt{\tau} M}) c(M, \tau) e^{-\sqrt{\tau} M} dM \\ &\quad + \sqrt{\tau} F_{0,\tau} \int_0^\infty f''(F_{0,\tau} e^{-\sqrt{\tau} M}) p(M, \tau) e^{-\sqrt{\tau} M} dM, \end{aligned} \tag{7}$$

where Q is the risk-neutral probability measure. The proof is in Appendix E. A numerical implementation of Equation (7) requires a truncation of the tails of the integral. The lower and upper bounds \underline{m} and \bar{m} are set such that

$$\begin{aligned} e^{-r\tau} E^Q[f(S_\tau)] &\cong e^{-r\tau} f(F_{0,\tau}) + \sqrt{\tau} F_{0,\tau} \int_{\underline{m}}^0 f''(F_{0,\tau} e^{-\sqrt{\tau} M}) c(M, \tau) e^{-\sqrt{\tau} M} dM \\ &\quad + \sqrt{\tau} F_{0,\tau} \int_0^{\bar{m}} f''(F_{0,\tau} e^{-\sqrt{\tau} M}) p(M, \tau) e^{-\sqrt{\tau} M} dM. \end{aligned} \tag{8}$$

TABLE 4 S&P 500 return risk-neutral moments

| 12–14 | December 1, 2008 | | | | | | December 2, 2019 | | | | | |
|-------------|------------------|-------|-------|--------------|-------|-------|------------------|-------|-------|--------------|-------|-------|
| | $\tau = 45$ | | | $\tau = 199$ | | | $\tau = 45$ | | | $\tau = 199$ | | |
| | A | B | C | A | B | C | A | B | C | A | B | C |
| VIX | 66.72 | 66.70 | 66.70 | 54.95 | 57.68 | 58.80 | 15.48 | 16.10 | 16.22 | 18.27 | 18.11 | 18.17 |
| Skewness | -1.36 | -1.41 | -1.41 | -1.35 | -1.24 | -1.63 | -2.40 | -2.07 | -2.31 | -2.43 | -2.52 | -2.65 |
| Kurtosis | 5.98 | 6.21 | 6.22 | 5.08 | 4.69 | 7.15 | 11.84 | 10.39 | 14.10 | 11.70 | 13.89 | 15.72 |
| Num options | 100 | | | 47 | | | 218 | | | 99 | | |

Note: Estimation of the VIX and risk-neutral moments of the forward contract log-return $\log(F_{\tau,0}/F_{0,\tau})$ using either the quoted options and a discretization of Equation (8) (column A) or Equation (8) on fitted surfaces. In the latter case, the moneyness range is the one provided by quoted options of the same maturity (column B) or it is extrapolated (column C). The extrapolated moneyness range is the largest and smallest moneyness (M) observed for all maturities for a given day.

Abbreviations: S&P, Standard Poor's; VIX, Volatility Index.

The bounds selection is determined by the extrapolation capacity of the IV surface model. Because Model (2) provides a closed-form solution for option prices for a continuum of moneyness, the two integrals in Equation (8) can be computed without being adversely impacted by the discretization bias stemming from the availability of only a finite number of strike prices.

Among all potential applications, a particularly interesting one consists in recovering risk-neutral moments of the underlying asset price. The continuum of OTM options across moneyness levels is also involved in the calculation of the VIX and the valuation of variance swaps (Carr & Madan, 2001; Neuberger, 1994; Schneider & Trojani, 2015). As a matter of fact, the VIX is the expected log-return of the S&P 500 forward contract. These economic quantities can be assessed with increased accuracy when the IV surface is smoothed and not restricted to available market data points.

Table 4 presents the VIX and risk-neutral moments of the forward contract log-return $\log(F_{\tau,0}/F_{0,\tau})$ using either the quoted options and a discretization of Equation (8) (column A) or Equation (8) on fitted surfaces. In the latter case, the moneyness range is the one provided by the quoted options with interpolation only¹² (column B), or it is extrapolated (column C). For short maturities, the number of quoted options spans a wider range resulting in similar results for the three approaches. However, for the longest two maturities, there are important disparities across the three methods due to the lack of observations.

As an additional verification of Model (2)'s extrapolation ability, we work with simulated instead of real data (details reported in the Supporting Information). We choose to simulate the Andersen et al. (2015) underlying asset dynamics model with various starting values for the latent variables to create different scenarios. Then we use Model (2) to extrapolate the IV surface from simulated option prices. We find a high-quality fit between the extrapolated and the simulated surfaces. Furthermore, we confirm a very accurate estimation of the risk-neutral moments using the extrapolated IV surfaces.

4.1.3 | Extraction of the risk-neutral density

A particularly relevant application of option pricing on a continuum of strikes is the extraction of the risk-neutral density of the underlying asset. Breeden and Litzenberger (1978) initially noted that such distribution is related to the convexity of option prices with respect to the strike, and that it could in theory be retrieved from a continuous IV surface. Although the risk-neutral underlying price distribution is interesting in itself, its extraction provides incremental benefits in terms of derivatives pricing over the two other aforementioned methods. Indeed, some derivatives are not options and do not have a twice differentiable payoff function, rendering the interpolation and the Carr and Madan (2001) methods inapplicable. Digital options with binary payoffs are a particular example. For such derivatives, obtaining the risk-neutral distribution is a necessary endeavor for pricing.

For a given day, the index spot price risk-neutral density function g_τ in τ years can be calculated through

¹²The integration is applied on the same moneyness range as that provided by quoted options of the same maturity.

$$g_\tau(K) = e^{r\tau} \frac{\partial^2 C(K, \tau)}{\partial K^2}, \quad K > 0.$$

Within the Model (2) framework, the risk-neutral density function¹³ is

$$e^{r\tau} \frac{\partial^2 c}{\partial K^2} = \frac{F_{0,\tau}}{K^2} \varphi(\delta_1) \left(\frac{1}{\sqrt{\tau}} \frac{\partial \delta_1}{\partial M} - \frac{1}{\sqrt{\tau}} \delta_1 \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + \frac{1}{\sqrt{\tau}} \frac{\partial^2 \sigma}{\partial M^2} \right), \quad (9)$$

where $\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$ is the density function of a standard normal random variable,

$$\begin{aligned} \frac{\partial \delta_1}{\partial M} &= \frac{1}{\sigma} - \left(\frac{M}{\sigma^2} - \frac{1}{2}\sqrt{\tau} \right) \frac{\partial \sigma}{\partial M}, \\ \frac{\partial \sigma}{\partial M} &= \beta_3 \mathbb{1}_{M \geq 0} + \beta_3 \left(1 - \left(\frac{e^{2M} - 1}{e^{2M} + 1} \right)^2 \right) \mathbb{1}_{M < 0} + \beta_4 2M e^{-M^2} \log \frac{T}{T_{\max}} \\ &\quad - \beta_5 81M^2 e^{27M^3} \log \frac{T}{T_{\max}} \mathbb{1}_{M < 0}, \\ \frac{\partial^2 \sigma}{\partial M^2} &= -\beta_3 8e^{2M} \frac{e^{2M} - 1}{(e^{2M} + 1)^3} \mathbb{1}_{M < 0} + \beta_4 2(1 - 2M^2) e^{-M^2} \log \frac{T}{T_{\max}} \\ &\quad - \beta_5 (162 + 6561M^3) M e^{27M^3} \log \frac{T}{T_{\max}} \mathbb{1}_{M < 0}. \end{aligned}$$

The proof is in Appendix F.1. Appendix F.1 shows that $\int_0^\infty g_\tau(K) dK = 1$ which is one of the fundamental properties of a density function.

For the same 4 days that were selected in Figure 1, a set of risk-neutral log-price densities $h_\tau(y) = \exp(y)g_\tau(\exp(y))$, each corresponding to a different maturity τ , is displayed in Figure 9. In all cases, volatility increases with time-to-maturity. Obviously, the financial crisis shows a greater volatility for all maturities. For all four dates, Model (2) implied densities exhibit negative skewness which is particularly apparent in 2008. Appendix G presents the analog of Figure 9 for both the CT and GG models. The polynomial structure of GG model which misbehaves for moneyness levels lying outside the observed range and the discontinuities in the CT model produce risk-neutral density functions that may take negative values and have irregular patterns, especially in the tails of the distribution.

4.2 | Risk management for options

The smooth IV surface of Model (2) provides the additional benefit, on top of derivatives pricing, to allow for the computation of “Greek” parameter sensitivities of option prices that are essential for replicating option payoffs or, more broadly speaking, managing positions on option portfolios. As shown in Appendix F.2, the call option delta and gamma¹⁴ are

$$\Delta = e^{-q\tau} \left(\Phi(\delta_1) + \varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right), \quad (10)$$

$$\Gamma = \frac{e^{-q\tau}}{\sqrt{\tau} S_0} \varphi(\delta_1) \left(\frac{\partial \delta_1}{\partial M} - \delta_1 \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + \frac{\partial^2 \sigma}{\partial M^2} \right), \quad (11)$$

¹³The arguments of the functions are omitted to simplify the notation.

¹⁴The functions’ arguments are omitted to simplify the notation.

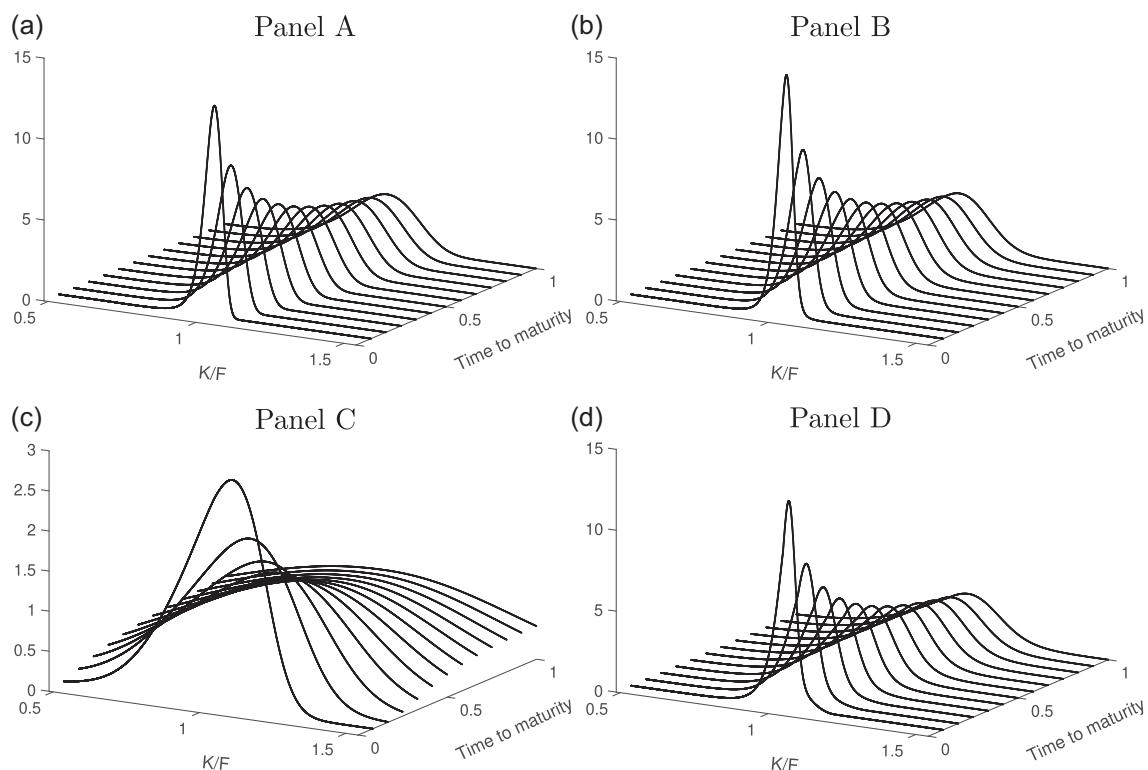


FIGURE 9 Log-price risk-neutral densities implied by Model (2): (a) Panel A, January 4, 1996; (b) Panel B, May 8, 2006; (c) Panel C, December 1, 2008; and (d) Panel D, December 31, 2019. Risk-neutral density functions $h_\tau(y) = \exp(y)g_\tau(\exp(y))$ of the log-prices derived from Model (2) call prices. January 4, 1996 is the first day in the sample. May 8, 2006 is a low volatility day. December 1, 2008 represents the peak of the 2008 financial crisis. December 31, 2019 is the last day of the sample.

where $\partial d_1 / \partial M$, $\partial \sigma / \partial M$, and $\partial^2 \sigma / \partial M^2$ are defined in Section 4.1.3. If the IV surface is flat, then $\partial \sigma / \partial M = \partial^2 \sigma / \partial M^2 = 0$ and the above formulas simplify into the Black-Scholes Greeks. Therefore, the extra terms are measuring the sensitivity of the IV to the variation of the underlying asset price through the moneyness variation.

Using the factor specification of IV, the computation of $\partial \sigma / \partial M$ and $\partial^2 \sigma / \partial M^2$ is immediate and its accuracy is not undermined by the limited availability of traded strikes and maturities. This clearly represents a significant advantage for option risk management purposes.

It should also be noted that the delta and the gamma defined above do not depend on any assumption regarding the dynamics of the underlying asset. Rather, they are consistent with the observed shape of the volatility smile and, as such, they comply with Bates's (2005) definition of smile-implied Greeks.¹⁵

$$\Delta = \frac{1}{S_0} \left(c - K \frac{\partial c}{\partial K} \right), \quad \Gamma = \frac{K^2}{S_0^2} \frac{\partial^2 c}{\partial K^2}.$$

This is particularly useful for illiquid options, for example, OTC transactions, having a moneyness or a maturity that is quite remote from those of publicly quoted ones; a model-free assessment of associated partial derivatives relying purely on finite differences would most likely prove unstable due to the paucity of related observations.

Another Greek parameter of high importance is the vega, that is, the sensitivity of the option price with respect to the IV. As shown in Appendix F.2, the call option vega is

$$\vartheta = \frac{\partial c}{\partial \sigma} = e^{-r\tau} F_{0,\tau} \varphi(\delta_1) \sqrt{\tau}.$$

¹⁵Bates (2005) formulas for the delta and the gamma rely on the scale invariance of option prices, a property verified by Model (2). The IV is a function of the moneyness $M = \frac{1}{\sqrt{\tau}} \ln \frac{F_{0,\tau}}{K}$ which is not affected when both the strike price and the underlying price are multiplied by a constant.

The call price sensitivity to the long term volatility level is $\frac{\partial \sigma}{\partial \beta_1} \frac{\partial c}{\partial \sigma}$, and its sensitivity to the maturity slope is $\frac{\partial \sigma}{\partial \beta_2} \frac{\partial c}{\partial \sigma}$.

5 | CONCLUSION

This study develops a factor model for the representation of volatility surfaces. The design of the model makes it very parsimonious, easy to interpret, seamless to estimate and quick to compute. Factors underlying the representation possess very intuitive meaning, that is, long-term level, time-to-maturity slope, moneyness slope, smile attenuation, and smirk. They are designed in accordance with the most salient empirical features of volatility surfaces as evidenced by the commonalities between loading vectors obtained from a PCA analysis on historical surfaces and the designed model factors.

The construction of the model factors leads to volatility-surface-implied underlying asset densities that are well-behaved and smooth. In particular, the convenient asymptotic behavior of the surface imposed by the model for large maturities and moneyness levels allows for extrapolation beyond ranges observed in the data. This key characteristic of the model is paramount in the applications discussed in this paper which crucially rely upon extrapolation ability. Moreover, performing such extrapolation is a necessary endeavor for multiple market participants requiring frequent marking-to-market of illiquid options. Furthermore, our model produces IV surfaces that are twice continuously differentiable with respect to the moneyness level, thereby limiting instances of model prices inconsistent with no-arbitrage principles.

The fitting performance is assessed on historical S&P 500 option prices obtained from the OptionMetrics database. Benchmarking against alternative literature models provides evidence of strong calibration outperformance versus the competing Goncalves and Guidolin (2006), Heston (1993), and Chalamandaris and Tsekrekos (2011) models, especially in recent periods. Indeed, the specification of the two latter factor-based benchmark exhibits explosive asymptotic behavior, leading to poor fitting performance for deep-OTM puts. By contrast, our model is much better suited for extrapolation to extreme moneyness levels.

Moreover, when restricting to a liquid subarea of the volatility surface, our model exhibits a fitting performance that is very close to that of the nonparametric PCA-based Israelov and Kelly (2017) model. Such result is very encouraging as it indicates our approach leads to very little loss of performance against purely data-driven approaches, such as PCA, while allowing circumventing shortfalls of the latter approach, such as (i) the inability to incorporate information provided by options in illiquid areas of the surface and (ii) its inadequacy for extrapolation beyond observed ranges for moneyness and time-to-maturity. A screening procedure based on the Davis and Hobson (2007) methodology highlights that option prices generated by the model developed herein very infrequently violate no-arbitrage restrictions, and that the model tends to smooth out theoretical arbitrage opportunities found in the data.

Several applications of the model are related to derivatives pricing and risk management. Three different derivatives pricing methods, namely (i) pure interpolation or extrapolation, (ii) the Carr–Madan formula, and (iii) risk-neutral density extraction through the Breeden and Litzenberger (1978) methodology, are shown to be applicable in conjunction with the model, using fitted surfaces stemming from the latter as their main input. The presented pricing methods enable the valuation of different classes of illiquid derivatives for which the mark-to-market is not straightforward. Closed-form expressions for European option Greek letters (e.g., the delta and the gamma) in the context of the IV surface model are presented. The straightforward computation of option sensitivities in this setting can facilitate the implementation of hedging procedures for options, especially for those whose strike or maturity is not quoted on public exchanges.

Further work expanding on the present study could include modeling the time dynamics of coefficients associated with each factor of the model, designing hedging strategies consistent with the volatility surface factor representation, and assessing the adequacy of the model for options on individual stocks or alternative underlying assets instead of equity indices.

ACKNOWLEDGMENTS

We thank an anonymous referee for helpful comments that significantly improved the paper. François is supported by a professorship funded by HEC Montréal. Galarneau-Vincent is supported by the TMX Group, Hydro-Québec, Letko-Brosseau, IVADO, MITACS, and Manulife. Gauthier is supported by NSERC (RGPIN-2019-04029), by a professorship funded by HEC Montréal and the HEC Montréal Foundation, and by IVADO. Godin is supported by NSERC (RGPIN-2017-06837).

DATA AVAILABILITY STATEMENT

The option data that support the findings of this study are available from OptionMetrics. Restrictions apply to the availability of these data, which were used under license for this study. The option data set used is available with the permission of the above data vendor.

ORCID

Pascal François  <http://orcid.org/0000-0003-0164-0467>

Geneviève Gauthier  <http://orcid.org/0000-0001-6098-0790>

Frédéric Godin  <http://orcid.org/0000-0001-5097-5269>

REFERENCES

- Ackerer, D., Tagasovska, N., & Vatter, T. (2019). Deep smoothing of the implied volatility surface. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*. Available at SSRN 3402942.
- Alexander, C., & Nogueira, L. M. (2007). Model-free price hedge ratios for homogeneous claims on tradable assets. *Quantitative Finance*, 7(5), 473–479.
- Alexiou, L., Goyal, A., Kostakis, A., & Rompolis, L. (2021). Pricing event risk: Evidence from concave implied volatility curves. Available at SSRN 3840081.
- Ammann, M., & Feser, A. (2019). Robust estimation of risk-neutral moments. *Journal of Futures Markets*, 39(9), 1137–1166.
- Andersen, T. G., Bondarenko, O., & Gonzalez-Perez, M. T. (2015). Exploring return dynamics via corridor implied volatility. *The Review of Financial Studies*, 28(10), 2902–2945.
- Azzone, M., & Baviera, R. (2022). Additive normal tempered stable processes for equity derivatives and power-law scaling. *Quantitative Finance*, 22(3), 501–518.
- Baker, M., Gillberg, T., & Thomas, S. (2018). Trading events: Smiles, frowns and moustaches—the many faces of the options market. Available at SSRN 3263368.
- Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5), 2003–2049.
- Bakshi, G., & Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *The Review of Financial Studies*, 16(2), 527–566.
- Bakshi, G., & Madan, D. (2000). Spanning and derivative-security valuation. *Journal of Financial Economics*, 55(2), 205–238.
- Bates, D. S. (2005). Hedging the smirk. *Finance Research Letters*, 2(4), 195–200.
- Bayraktar, E., & Yang, B. (2011). A unified framework for pricing credit and equity derivatives. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 21(3), 493–517.
- Birru, J., & Figlewski, S. (2012). Anatomy of a meltdown: The risk neutral density for the S&P 500 in the fall of 2008. *Journal of Financial Markets*, 15(2), 151–180.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.
- Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1–2), 167–179.
- Bliss, R. R., & Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions. *Journal of Banking & Finance*, 26(2–3), 381–422.
- Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4), 621–651.
- Carr, P., & Madan, D. (2001). Towards a theory of volatility trading. *Option Pricing, Interest Rates and Risk Management, Handbooks in Mathematical Finance*, 22(7), 458–476.
- Carr, P., & Madan, D. B. (2005). A note on sufficient conditions for no arbitrage. *Finance Research Letters*, 2(3), 125–130.
- Carr, P., & Wu, L. (2014). Static hedging of standard options. *Journal of Financial Econometrics*, 12(1), 3–46.
- Célerier, C., & Vallée, B. (2017). Catering to investors through security design: Headline rate and complexity. *The Quarterly Journal of Economics*, 132(3), 1469–1508.
- Chalamandaris, G., & Tsekrekos, A. E. (2011). How important is the term structure in implied volatility surface modeling? Evidence from foreign exchange options. *Journal of International Money and Finance*, 30(4), 623–640.
- Conrad, J., Dittmar, R. F., & Ghysels, E. (2013). Ex ante skewness and expected stock returns. *The Journal of Finance*, 68(1), 85–124.
- Cont, R., & DaFonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1), 45–60.
- Cox, J. C., & Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(1–2), 145–166.
- Daglish, T., Hull, J., & Suo, W. (2007). Volatility surfaces: Theory, rules of thumb, and empirical evidence. *Quantitative Finance*, 7(5), 507–524.
- Davis, M. H., & Hobson, D. G. (2007). The range of traded option prices. *Mathematical Finance*, 17(1), 1–14.
- Delbaen, F., & Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(1), 463–520.
- Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2), 337–364.
- Dumas, B., Fleming, J., & Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *The Journal of Finance*, 53(6), 2059–2106.
- Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, 9(4), 417–428.

- Figlewski, S. (2018). Risk-neutral densities: A review. *Annual Review of Financial Economics*, 10, 329–359.
- François, P., & Stentoft, L. (2021). Smile-implied hedging with volatility risk. *Journal of Futures Markets*, 41(8), 1220–1240.
- Gao, G. P., Gao, P., & Song, Z. (2018). Do hedge funds exploit rare disaster concerns? *The Review of Financial Studies*, 31(7), 2650–2692.
- Gao, G. P., Lu, X., & Song, Z. (2019). Tail risk concerns everywhere. *Management Science*, 65(7), 3111–3130.
- Garman, M. B., & Kohlhagen, S. W. (1983). Foreign currency option values. *Journal of International Money and Finance*, 2(3), 231–237.
- Gauthier, G., & Simonato, J.-G. (2012). Linearized Nelson-Siegel and Svensson models for the estimation of spot interest rates. *European Journal of Operational Research*, 219(2), 442–451.
- Goncalves, S., & Guidolin, M. (2006). Predictable dynamics in the S&P 500 index options implied volatility surface. *Journal of Business*, 79(3), 1591–1635.
- Henderson, B. J., & Pearson, N. D. (2011). The dark side of financial innovation: A case study of the pricing of a retail financial product. *Journal of Financial Economics*, 100(2), 227–247.
- 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.
- Israelov, R., & Kelly, B. T. (2017). Forecasting the distribution of option returns. Available at SSRN 3033242.
- 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. (2004). *Option-implied risk-neutral distributions and risk aversion*. Research Foundation of AIMR.
- Jackwerth, J. C., & Rubinstein, M. (1996). Recovering probability distributions from option prices. *The Journal of Finance*, 51(5), 1611–1631.
- Lee, C.-F., & Sokolinskiy, O. (2015). R-2GAM stochastic volatility model: Flexibility and calibration. *Review of Quantitative Finance and Accounting*, 45(3), 463–483.
- Martin, I. (2017). What is the expected return on the market? *The Quarterly Journal of Economics*, 132(1), 367–433.
- Martin, I. W., & Wagner, C. (2019). What is the expected return on a stock? *The Journal of Finance*, 74(4), 1887–1929.
- Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60(4), 473–489.
- Neuberger, A. (1994). The log contract. *Journal of Portfolio Management*, 20, 74–80.
- Neumann, M., & Skiadopoulos, G. (2013). Predictable dynamics in higher-order risk-neutral moments: Evidence from the S&P 500 options. *Journal of Financial and Quantitative Analysis*, 48(3), 947–977.
- Rebonato, R. (2005). *Volatility and correlation: The perfect hedger and the fox*. John Wiley & Sons.
- Reiswich, D. (2010). A note on standard no-arbitrage conditions. Available at SSRN 1619844.
- Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2), 615–648.
- Schneider, P., & Trojani, F. (2015). *Fear trading* [Swiss Finance Institute Research Paper, 15-03].
- Schoutens, W., Simons, E., & Tistaert, J. (2006). A perfect calibration! now what? In *The Best of Wilmott* (p. 281). John Wiley & Sons.
- Skiadopoulos, G. (2001). Volatility smile consistent option models: A survey. *International Journal of Theoretical and Applied Finance*, 4(3), 403–437.

How to cite this article: François, P., Galarneau-Vincent, R., Gauthier, G., & Godin, F. (2022). Venturing into uncharted territory: An extensible implied volatility surface model. *Journal of Futures Markets*, 42, 1912–1940.
<https://doi.org/10.1002/fut.22364>

APPENDIX A: PRINCIPAL COMPONENT ANALYSIS

Directly applying PCA to the option sample is not possible since option numbers and characteristics (moneyness and time-to-maturity) vary from day to day; PCA requires a stable sample every day. To circumvent this issue, a grid with respect to moneyness and time-to-maturity is constructed. For each point of the grid the IV can be interpolated using quoted options IV. The interpolation scheme can be achieved through a variety of methods. As in Israelov and Kelly (2017), a spline interpolation scheme is implemented using the MATLAB fit function with “thinplateinterp” fit type.

Because quoted moneyness levels vary greatly from day to day depending on the market conditions, the grid covers a smaller surface than the one spanned by the quoted moneyness and maturities. Two grids are used in this paper. In Figure 2 the moneyness M varies between -0.2 and 0.6 by increments of 0.1 and the days-to-maturity are 30, 60, 91, 122, 152, 182, 270, and 365 days. The other figures and tables of Appendix A are based on the moneyness definition and the grid of Israelov and Kelly (2017), that is, the moneyness M^* is between -2 and 1 with increments of 0.25 and the days-to-maturity are 30, 60, 91, 122, 152, 182, 270, and 365 days-to-maturity. Because M^* depends on the VIX value, the two grids do not include the same option subsample, especially during financial turmoil.

The PCA is constructed from the sample covariance matrix of the interpolated IV time series. The right panels of Figure 2 display the five factors with the greatest explanatory power. Factor 1 can be interpreted as the level, factor 2 is the time-to-maturity slope, factor 3 corresponds to the moneyness slope, factor 4 is a curvature factor, and factor 5 is the smirk. The first factor explains 94.84% of the surface variations.

APPENDIX B: BAYESIAN REGRESSION

Model (2) is a linear function of the factors, which suggests that the ordinary least square (OLS) estimation approach is straightforward. However, as documented in Gauthier and Simonato (2012) in the alternative context of zero-coupon yields, there can be several sets of parameters which produce very similar surfaces. To preserve the economic interpretation of each factor, the least-squares method is coupled with a Bayesian approach for regularization purposes, thereby avoiding avoid erratic behavior in parameter time series.

A typical linear model can be expressed as $Y = X\beta + \epsilon$. A linear system incorporates prior information as follows:

$$\begin{bmatrix} Y \\ \beta_{\text{prior}} \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \epsilon \\ \delta \end{bmatrix}, \quad (\text{B1})$$

where Y represents the observations, β_{prior} the prior's expectation, X the factors, R the matrix linking the parameters to the priors, and (ϵ, δ) the vector of errors which follows a multivariate normal distribution with a diagonal covariance matrix

$$\Omega = \begin{bmatrix} \Sigma_\epsilon & 0 \\ 0 & \Sigma_\delta \end{bmatrix}.$$

The generalized least-squares estimator of β with prior information is

$$\hat{\beta} = \left(\begin{bmatrix} X \\ R \end{bmatrix}^\top \Omega^{-1} \begin{bmatrix} X \\ R \end{bmatrix} \right)^{-1} \begin{bmatrix} X \\ R \end{bmatrix}^\top \Omega^{-1} \begin{bmatrix} Y \\ \beta_{\text{prior}} \end{bmatrix}.$$

The value of Σ_ϵ can be estimated using the OLS estimator without any priors and the matrix Σ_δ is a hyperparameter that controls the prior distribution. The smaller the values on the diagonal of Σ_δ are, the closer to the prior expected value the estimated β is.

The observed 1-year ATM IV ($\text{ATM}_{1y,t}$) serves as prior mean for $\beta_{1,t}$. The slope prior is constructed from the 1-month ATM IV ($\text{ATM}_{1m,t}$):

$$\text{Slope}_t = \frac{\text{ATM}_{1y,t} - \text{ATM}_{1m,t}}{\exp(-\sqrt{4/12})}.$$

The priors for $\beta_{3,t}$ and $\beta_{5,t}$ are the previous day estimates $\beta_{3,t-1}$ and $\beta_{5,t-1}$, respectively. Due to its interconnectedness with the other parameters, no prior is assigned to $\beta_{4,t}$. This entails setting

$$\beta_{\text{prior}} = \begin{bmatrix} \text{ATM}_{1y,t} \\ \text{Slope}_t \\ \beta_{3,t-1} \\ \beta_{5,t-1} \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \text{and } \Sigma_\delta = \begin{bmatrix} 0.38 & 0 & 0 & 0 & 0 \\ 0 & 5.60 & 0 & 0 & 0 \\ 0 & 0 & 0.73 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times 10^{-4}.$$

The prior variances need to be set. For the first two priors associated with the long-term level and the slope, the sample variance of the observable ($\text{ATM}_{1y,t}$ and Slope_t) is considered. The β_3 prior's variance is the sample variance of the proxy variable for the moneyness slope, namely, the 1-month ATM IV minus the 1-month IV with moneyness $M = 0.4$. Finally, for the last prior of β_5 , based on judgmental consideration, the standard deviation is set to half the parameter's level, resulting in a prior variance of 1×10^{-4} . Because the prior variances are large, the estimation procedure has extensive leeway to match the option data, while keeping the economic interpretation of the coefficients due to regularization removing large erratic movements in time series of parameter estimates.

APPENDIX C: BENCHMARKING WITH A NONPARAMETRIC PCA APPROACH

In Section 3.3, the calibration performance of Model (2) is compared with that of two parametric benchmarks. In this section, the nonparametric PCA approach of Israelov and Kelly (2017), hereafter IK, is instead considered for benchmarking. As mentioned by the authors, their methodology is best suited to represent densely populated regions of

the surface. When applied to our data set, their model is fitted to a subsurface inside which only 41% of observed options lie, leaving aside useful information about extreme movements expectations contained in the deep-OTM options. Furthermore, PCA methods cannot extrapolate beyond the restricted grid to generate prices for out-of-sample deep-OTM options. The main conclusion of this section is that on this restricted sample for which the PCA approach is optimal, Model (2) does almost as well, while allowing for IV interpolation and extrapolation on the wider IV surface.

Israelov and Kelly (2017) rely on an alternative definition of moneyness that is proportional to the 30-day VIX level at day t ,

$$M^* = \frac{\log(K/S)}{VIX_t \sqrt{(\tau)}}.$$

Such specification allows the model to adjust the grid relied upon by the PCA to the prevalent market state, that is, to include more OTM options in periods of market turmoil. Indeed, to perform a PCA over daily IV surfaces, Israelov and Kelly (2017) construct a synthetic IV surface on a predefined grid covering time-to-maturity values of 30, 60, 91, 122, 152, 182, 273, and 365 days, and moneyness M^* values varying between -2 and 1 with increments of 0.25 . They interpolate the grid points from available options IV with a thin plate spline. The factors extracted from the PCA represent the most efficient linear decomposition to minimize the squared fitting errors.

The fitting performance is assessed by computing the daily RMSE between observed IV and the corresponding fitted values.

Figure C1 shows that the daily RMSE is very similar between the PCA approach and the specification (2) over the restricted sample. Such an outcome was expected due to the similarity of Model (2) and PCA factors highlighted in Figure 2.

Table C1 confirms that the ARMSEs of the PCA approach and Model (2) are very close within each bucket of moneyness and time-to-maturity. As expected, the PCA approach slightly outperforms Model (2) for all buckets

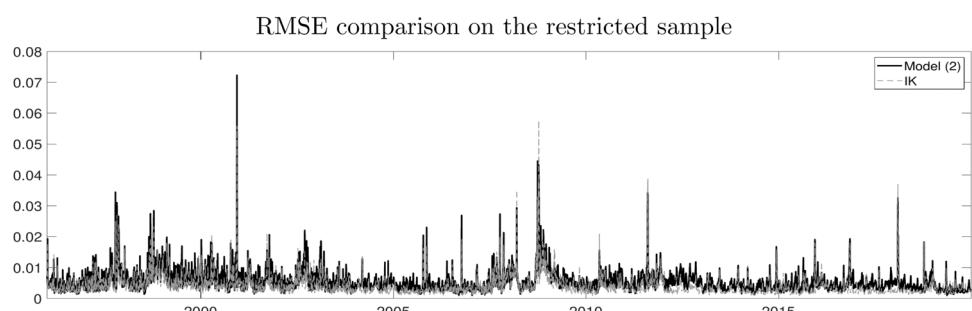


FIGURE C1 RMSE by bucket over the restricted sample. Model (2) is estimated on a restricted sample corresponding to options with a time-to-maturity of between 30 and 365 days, and an M^* moneyness of between -2 to 1 . IK, Israelov and Kelly; RMSE, root-mean-square error.

TABLE C1 Average RMSE over time from IV surface estimation over the restricted sample

| Model | $M \leq -0.1$ | $-0.1 < M \leq 0.1$ | $M > 0.1$ | All |
|-------------------|----------------|----------------------|--------------|-----------|
| Model (2) | 0.0080 | 0.0059 | 0.0057 | 0.0063 |
| IK | 0.0060 | 0.0045 | 0.0054 | 0.0052 |
| Number of options | 231,356 | 606,327 | 587,042 | 1,424,725 |
| Model | $\tau \leq 60$ | $60 < \tau \leq 180$ | $\tau > 180$ | All |
| Model (2) | 0.0089 | 0.0036 | 0.0058 | 0.0063 |
| IK | 0.0082 | 0.0030 | 0.0031 | 0.0052 |
| Number of options | 435,642 | 585,176 | 403,907 | 1,424,725 |

Note: The average RMSE over time is reported for each bucket of moneyness (M) and days to maturity (τ). The sample period is January 4, 1996–June 26, 2019. The restricted sample corresponds to options with a time-to-maturity of between 30 and 365 days, and an M^* moneyness of between -2 to 1 . Model (2) is re-estimated on the restricted sample in Panel B.

Abbreviations: IK, Israelov and Kelly; RMSE, root-mean-square error.

because by construction, the PCA designs factors so as to minimize mean squared discrepancies between data and fitted values. It is reassuring to see that Model (2) has a very similar performance than the PCA, while being able to fit over a much larger IV surface and even extrapolate beyond the observable IV.

APPENDIX D: BUTTERFLY AND CALENDAR SPREADS

On a given trading day t , let $D_\tau = \exp(-r_\tau \tau)$ be the risk-free discount factor where r_τ is the OptionMetrics zero-coupon interest rate associated with the maturity τ . For any traded maturity τ , denoted by $K_0^{(\tau)} < K_1^{(\tau)} < \dots < K_{n(\tau)}^{(\tau)}$ the set of all strikes for which quotes are provided in the data set. Thus, $n(\tau)$ represents the number of available options with time-to-maturity τ . $C(K, \tau)$ denotes the call price for time-to-maturity τ and strike price K .

As outlined in Davis and Hobson (2007), the butterfly spread value $BS_i^{(\tau)}$, $i = 1, \dots, n(\tau) - 1$, defined as

$$\begin{aligned} BS\left(K_i^{(\tau)}, \tau\right) = & \frac{C\left(K_{i+1}^{(\tau)}, \tau\right)}{\left(K_{i+1}^{(\tau)} - K_i^{(\tau)}\right)D_\tau} + \frac{C\left(K_{i-1}^{(\tau)}, \tau\right)}{\left(K_i^{(\tau)} - K_{i-1}^{(\tau)}\right)D_\tau} \\ & - \frac{C\left(K_i^{(\tau)}, \tau\right)}{D_\tau} \left(\frac{1}{K_i^{(\tau)} - K_{i-1}^{(\tau)}} + \frac{1}{K_{i+1}^{(\tau)} - K_i^{(\tau)}} \right) \end{aligned}$$

needs to be positive, as otherwise a butterfly spread arbitrage opportunity would arise.¹⁶

The calendar spread value $CS_i^{(\tau)}$ is defined as

$$\begin{aligned} CS_i^{(\tau)}(\tau_1, \tau_2, i_1, i_2) = & C\left(K_i^{(\tau)}, \tau\right) - \left(\frac{K_{i_2}^{(\tau_2)}/F_{0,\tau_2} - K_i^{(\tau)}/F_{0,\tau}}{K_{i_2}^{(\tau_2)}/F_{0,\tau_2} - K_{i_1}^{(\tau_1)}/F_{0,\tau_1}} \right) \frac{D_\tau F_{0,\tau}}{D_{\tau_1} F_{0,\tau_1}} C\left(K_{i_1}^{(\tau_1)}, \tau_1\right) \\ & - \left(1 - \left(\frac{K_{i_2}^{(\tau_2)}/F_{0,\tau_2} - K_i^{(\tau)}/F_{0,\tau}}{K_{i_2}^{(\tau_2)}/F_{0,\tau_2} - K_{i_1}^{(\tau_1)}/F_{0,\tau_1}} \right) \right) \frac{D_\tau F_{0,\tau}}{D_{\tau_2} F_{0,\tau_2}} C\left(K_{i_1}^{(\tau_2)}, \tau_2\right). \end{aligned}$$

where $\tau_1, \tau_2 > \tau$ and i_1, i_2 are such that the strike prices satisfy $\frac{K_{i_1}^{(\tau_1)}}{F_{0,\tau_1}} \leq \frac{K_i^{(\tau)}}{F_{0,\tau}} \leq \frac{K_{i_2}^{(\tau_2)}}{F_{0,\tau_2}}$. If $CS_i^{(\tau)}(\tau_1, \tau_2, i_1, i_2) \leq 0$, then there is a calendar spread arbitrage opportunity in the data.

There are many combinations of maturities and strike prices that satisfy the calendar spread restrictions. In Section 3.5, only one calendar spread per option in the sample is tested: for $C\left(K_i^{(\tau)}, \tau\right)$, τ_1 and τ_2 are the smallest maturities greater than τ and the ratios $\frac{K_{i_1}^{(\tau_1)}}{F_{0,\tau_1}}$ and $\frac{K_{i_2}^{(\tau_2)}}{F_{0,\tau_2}}$ are the closest to $\frac{K_i^{(\tau)}}{F_{0,\tau}}$. When no such strike and maturity combinations are available, no calendar spread test is performed for that particular option.

APPENDIX E: CARR-MADAN FORMULA

This appendix contains the proof of Equation (7). The Carr and Madan (2001) formula states that for a twice differentiable payoff function f ,

$$\begin{aligned} E^Q[e^{-\tau}f(S_\tau)] = & f(k)e^{-\tau} + f'(k)[C(k) - P(k)] \\ & + \int_0^k f''(K)P(K) dK + \int_k^\infty f''(K)C(K) dK, \end{aligned} \tag{E1}$$

¹⁶The butterfly spreads are computed differently at extremities of the strike price set:

$$BS\left(K_0^{(\tau)}, \tau\right) = 1 - \frac{C\left(K_0^{(\tau)}, \tau\right) - C\left(K_1^{(\tau)}, \tau\right)}{\left(K_1^{(\tau)} - K_0^{(\tau)}\right)D_\tau} \quad \text{and} \quad BS\left(K_{n(\tau)}^{(\tau)}, \tau\right) = \frac{C\left(K_{n(\tau)-1}^{(\tau)}, \tau\right) - C\left(K_{n(\tau)}^{(\tau)}, \tau\right)}{\left(K_{n(\tau)}^{(\tau)} - K_{n(\tau)-1}^{(\tau)}\right)D_\tau}.$$

where C and P denote the put and call prices written on the underlying index S_τ with maturity τ and strike price K . Due to the put-call parity, the second term vanishes if $k = F_{0,\tau}$. In our framework, the moneyness $M = (\log F_{0,\tau} - \log K)/\sqrt{\tau}$ can be inverted to retrieve the strike price $K = F_{0,\tau}e^{-\sqrt{\tau}M}$. Since $dK = -\sqrt{\tau}F_{0,\tau}e^{-\sqrt{\tau}M}dM$, applying a change of variable in Equation (E1), assuming that $k = F_{0,\tau}$, leads to Equation (7). \square

APPENDIX F: GREEKS AND OTHER PARTIAL DERIVATIVES

This appendix establishes Equations (9)–(11). The functions' arguments are omitted to simplify the notation. Recall that the call price is

$$c = e^{-r\tau}F_{0,\tau}(\Phi(\delta_1) - e^{-M\sqrt{\tau}}\Phi(\delta_2))$$

with $F_{0,\tau} = S_0 \exp((r - q)\tau)$, $M = \frac{1}{\sqrt{\tau}} \ln \frac{F_{0,\tau}}{K}$, $\delta_1 = \frac{M}{\sigma} + \frac{1}{2}\sigma\sqrt{\tau}$, $\delta_2 = \delta_1 - \sigma\sqrt{\tau}$, and σ is the IV from Model (2).

F1 | Risk-neutral density function

Equation (9) is derived in this appendix. Note that $\frac{\partial F}{\partial S} = e^{(r-q)\tau}$, $\frac{\partial M}{\partial F} = \frac{1}{\sqrt{\tau}} \frac{1}{F_{0,\tau}}$,

$$\frac{\partial \delta_1}{\partial M} = \frac{1}{\sigma} - \left(\frac{M}{\sigma^2} - \frac{1}{2}\sqrt{\tau} \right) \frac{\partial \sigma}{\partial M} \quad \text{and} \quad \frac{\partial \delta_2}{\partial M} = \frac{1}{\sigma} - \left(\frac{M}{\sigma^2} + \frac{1}{2}\sqrt{\tau} \right) \frac{\partial \sigma}{\partial M} = \frac{\partial \delta_1}{\partial M} - \sqrt{\tau} \frac{\partial \sigma}{\partial M}.$$

Moreover,

$$\varphi(\delta_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\delta_1 - \sigma\sqrt{\tau})^2} = \varphi(\delta_1)e^{\sigma\sqrt{\tau}\delta_1 - \frac{1}{2}\sigma^2\tau} = \varphi(\delta_1)e^{\sigma\sqrt{\tau}\left(\frac{M}{\sigma} + \frac{1}{2}\sigma\sqrt{\tau}\right) - \frac{1}{2}\sigma^2\tau} = \varphi(\delta_1)e^{\sqrt{\tau}M}.$$

Because $\frac{\partial M}{\partial K} = -\frac{1}{\sqrt{\tau}K}$,

$$\begin{aligned} \frac{\partial c}{\partial K} &= e^{-r\tau}F_{0,\tau}(\Phi(\delta_1) - e^{-M\sqrt{\tau}}\Phi(\delta_2)) \\ &= e^{-r\tau}F_{0,\tau} \frac{\partial M}{\partial K} \left(\varphi(\delta_1) \frac{\partial \delta_1}{\partial M} + \sqrt{\tau}e^{-M\sqrt{\tau}}\Phi(\delta_2) - e^{-M\sqrt{\tau}}\varphi(\delta_2) \frac{\partial \delta_2}{\partial M} \right) \\ &= e^{-r\tau}F_{0,\tau} \frac{\partial M}{\partial K} \left(\varphi(\delta_1) \frac{\partial \delta_1}{\partial M} + \sqrt{\tau}e^{-M\sqrt{\tau}}\Phi(\delta_2) - \varphi(\delta_1) \left(\frac{\partial \delta_1}{\partial M} - \sqrt{\tau} \frac{\partial \sigma}{\partial M} \right) \right) \\ &= -e^{-r\tau} \frac{F_{0,\tau}}{K} \left(e^{-M\sqrt{\tau}}\Phi(\delta_2) + \varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right) \\ &= -e^{-r\tau} \left(\Phi(\delta_2) + e^{M\sqrt{\tau}}\varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right). \end{aligned}$$

The Gaussian density function satisfies $\frac{\partial \varphi}{\partial z}(z) = -z\varphi(z)$. Therefore, the risk-neutral density function is

$$\begin{aligned} e^{r\tau} \frac{\partial^2 c}{\partial K^2} &= \frac{\partial M}{\partial K} \frac{\partial}{\partial M} \left(\Phi(\delta_2) + e^{M\sqrt{\tau}}\varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right) \\ &= \frac{\partial M}{\partial K} \left(\varphi(\delta_2) \frac{\partial \delta_2}{\partial M} + \sqrt{\tau}e^{M\sqrt{\tau}}\varphi(\delta_1) \frac{\partial \sigma}{\partial M} - e^{M\sqrt{\tau}}\delta_1\varphi(\delta_1) \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + e^{M\sqrt{\tau}}\varphi(\delta_1) \frac{\partial^2 \sigma}{\partial M^2} \right) \\ &= \frac{\partial M}{\partial K} \left(\varphi(\delta_1)e^{\sqrt{\tau}M} \left(\frac{\partial \delta_1}{\partial M} - \sqrt{\tau} \frac{\partial \sigma}{\partial M} \right) \right. \\ &\quad \left. + \sqrt{\tau}e^{M\sqrt{\tau}}\varphi(\delta_1) \frac{\partial \sigma}{\partial M} - e^{M\sqrt{\tau}}\delta_1\varphi(\delta_1) \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + e^{M\sqrt{\tau}}\varphi(\delta_1) \frac{\partial^2 \sigma}{\partial M^2} \right) \\ &= \frac{\partial M}{\partial K} e^{\sqrt{\tau}M} \varphi(\delta_1) \left(\frac{\partial \delta_1}{\partial M} - \delta_1 \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + \frac{\partial^2 \sigma}{\partial M^2} \right) \\ &= \frac{F_{0,\tau}}{K^2} \frac{\varphi(\delta_1)}{\sqrt{\tau}} \left(\frac{\partial \delta_1}{\partial M} - \delta_1 \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + \frac{\partial^2 \sigma}{\partial M^2} \right). \end{aligned}$$

The density integrates to one since

$$\begin{aligned}
 \int_0^\infty e^{-rt} \frac{\partial^2 c}{\partial K^2} dK &= e^{-rt} \left(\lim_{K \rightarrow \infty} \frac{\partial c}{\partial K} - \lim_{K \rightarrow 0} \frac{\partial c}{\partial K} \right) \\
 &= - \lim_{M \rightarrow -\infty} \left(\Phi(\delta_2) + e^{M\sqrt{r}} \varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right) + \lim_{M \rightarrow \infty} \left(\Phi(\delta_2) + e^{M\sqrt{r}} \varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right) \\
 &= \lim_{M \rightarrow \infty} \Phi(\delta_2) + \lim_{M \rightarrow \infty} e^{M\sqrt{r}} \varphi(\delta_1) \frac{\partial \sigma}{\partial M} = 1.
 \end{aligned}$$

F2 | Greeks computation

Equations (10) and (11) are derived in this appendix. Because $\frac{\partial \delta_2}{\partial M} = \frac{\partial \delta_1}{\partial M} - \sqrt{r} \frac{\partial \sigma}{\partial M}$ and $\varphi(\delta_2) = \varphi(\delta_1) e^{\sqrt{r}M}$,

$$\begin{aligned}
 \Delta &= \frac{\partial c}{\partial S} = \frac{\partial F}{\partial S} \frac{\partial c}{\partial F} \\
 &= e^{-rt} \frac{\partial F}{\partial S} ((\Phi(\delta_1) - e^{-M\sqrt{r}} \Phi(\delta_2)) \\
 &\quad + F_{0,r} \left(\varphi(\delta_1) \frac{\partial M}{\partial F} \frac{\partial \delta_1}{\partial M} + e^{-M\sqrt{r}} \sqrt{r} \frac{\partial M}{\partial F} \Phi(\delta_2) - e^{-M\sqrt{r}} \varphi(\delta_2) \frac{\partial M}{\partial F} \frac{\partial \delta_2}{\partial M} \right)) \\
 &= e^{-rt} \frac{\partial F}{\partial S} ((\Phi(\delta_1) - e^{-M\sqrt{r}} \Phi(\delta_2)) \\
 &\quad + F_{0,r} \frac{\partial M}{\partial F} \left(\varphi(\delta_1) \frac{\partial \delta_1}{\partial M} + e^{-M\sqrt{r}} \Phi(\delta_2) \sqrt{r} - \varphi(\delta_1) \left(\frac{\partial \delta_1}{\partial M} - \sqrt{r} \frac{\partial \sigma}{\partial M} \right) \right)) \\
 &= e^{-qt} \left(\Phi(\delta_1) + \varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right).
 \end{aligned}$$

Recall that $\frac{\partial \varphi}{\partial z}(z) = -z\varphi(z)$. Therefore,

$$\begin{aligned}
 \Gamma &= \frac{\partial^2 c}{\partial S^2} = \frac{\partial F}{\partial S} \frac{\partial}{\partial F} \left(\frac{\partial c}{\partial S} \right) = e^{-qt} \frac{\partial F}{\partial S} \frac{\partial}{\partial F} \left(\Phi(\delta_1) + \varphi(\delta_1) \frac{\partial \sigma}{\partial M} \right) \\
 &= e^{-qt} \frac{\partial F}{\partial S} \left(\varphi(\delta_1) \frac{\partial M}{\partial F} \frac{\partial \delta_1}{\partial M} - \delta_1 \varphi(\delta_1) \frac{\partial M}{\partial F} \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + \varphi(\delta_1) \frac{\partial M}{\partial F} \frac{\partial^2 \sigma}{\partial M^2} \right) \\
 &= \frac{e^{-qt} e^{(r-q)\tau}}{\sqrt{r} F_{0,r}} \varphi(\delta_1) \left(\frac{\partial \delta_1}{\partial M} - \delta_1 \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + \frac{\partial^2 \sigma}{\partial M^2} \right) \\
 &= \frac{e^{-qt}}{\sqrt{r} S_0} \varphi(\delta_1) \left(\frac{\partial \delta_1}{\partial M} - \delta_1 \frac{\partial \delta_1}{\partial M} \frac{\partial \sigma}{\partial M} + \frac{\partial^2 \sigma}{\partial M^2} \right).
 \end{aligned}$$

APPENDIX G: RISK-NEUTRAL DENSITIES WITHIN CT AND GG FRAMEWORKS

This section illustrates the importance of the smoothness of the IV surface model as well as its asymptotic behavior. For the same 4 days that were selected in Figure 1, Figures G1 and G2 present risk-neutral densities function obtained from the GG and CT benchmark models. We observe negative values and irregular tail behavior.

As explained in Section 3, the Chalamandaris and Tsekrekos (2011) moneyness measure consists of a linear transformation of the Black–Scholes Δ . To compute the risk-neutral density from the Chalamandaris and Tsekrekos (2011) IV surface, the moneyness measure is transformed back to the displayed moneyness measure K/F . To transform back the Chalamandaris and Tsekrekos (2011) moneyness measure to K/F , the option IV is required. A two-dimensional interpolation method using the observed options IV with their respective characteristics is implemented to obtain implied volatilities with the desired characteristics (strike price and time-to-maturity). To extrapolate from the surface, the options with the largest (smallest) moneyness are considered.

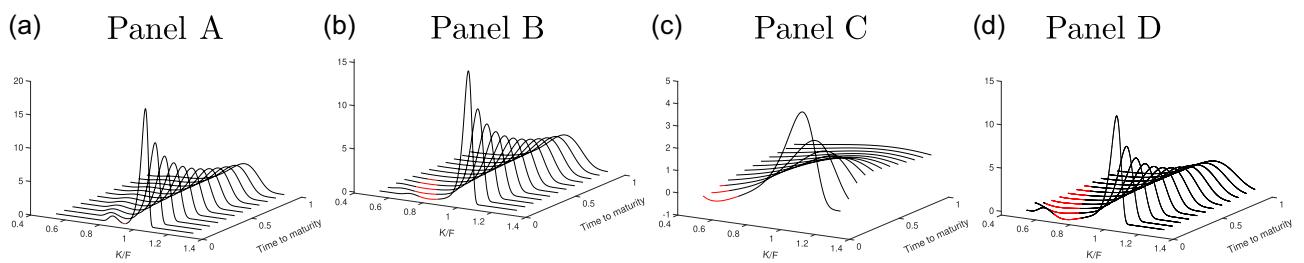


FIGURE G1 Log-price risk-neutral densities implied by the GG model: (a) Panel A, January 4, 1996; (b) Panel B, May 8, 2006; (c) Panel C, December 1, 2008; and (d) Panel D, December 31, 2019. Risk-neutral density functions $h_\tau(y) = \exp(y)g_\tau(\exp(y))$ of the log-prices derived from GG model call prices. January 4, 1996 is the first day in the sample. May 8, 2006 is a low volatility day. December 1, 2008 represents the peak of the 2008 financial crisis. December 31, 2019 is the last day of the sample. The red line highlights regions where the density is negative. GG, Goncalves and Guidolin.

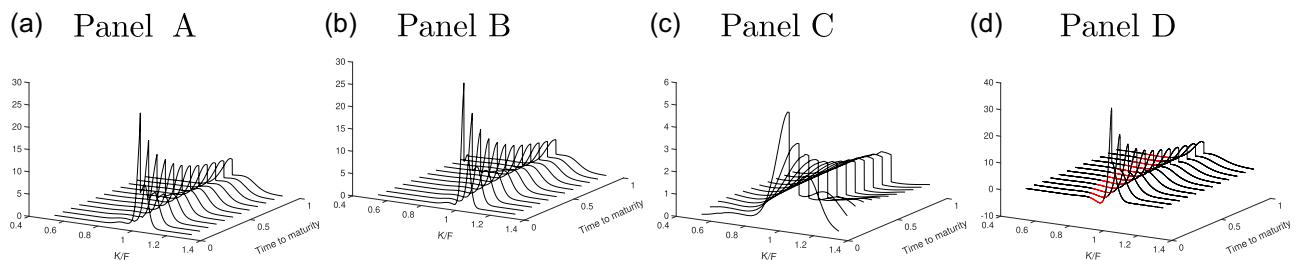


FIGURE G2 Log-price risk-neutral densities implied by CT model: (a) Panel A, January 4, 1996; (b) Panel B, May 8, 2006; (c) Panel C, December 1, 2008; and (d) Panel D, December 31, 2019. Risk-neutral density functions $h_\tau(y) = \exp(y)g_\tau(\exp(y))$ of the log-prices derived from CT model call prices. January 4, 1996 is the first day in the sample. May 8, 2006 is a low volatility day. December 1, 2008 represents the peak of the 2008 financial crisis. December 31, 2019 is the last day of the sample. The red line highlights regions where the density is negative. CT, Chalamandaris and Tsekrekos.

APPENDIX H: ABNORMAL IV SURFACE ON OCTOBER 9, 2006

The observed implied volatilities from October 9, 2006 are provided in Figure H1.

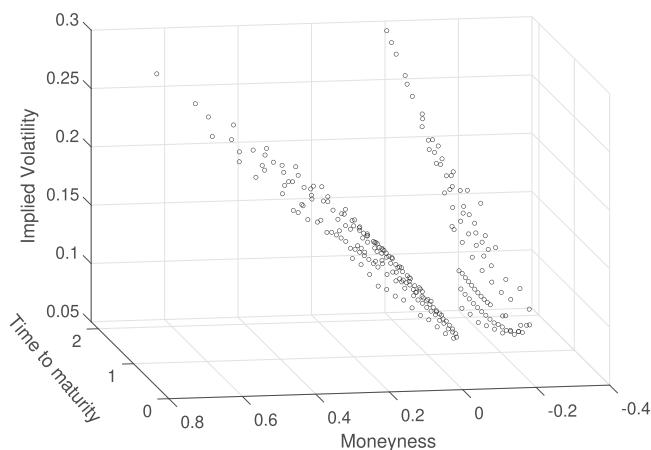


FIGURE H1 IV surface on October 9, 2006. IV, implied volatility.