Assessing the Performance of Symmetric and Asymmetric Implied Volatility Functions

Panayiotis C. Andreou,[a,*] Chris Charalambous,[b] Spiros H. Martzoukos[c]

[a] *Lecturer in Finance at the Dept. of Commerce, Finance and Shipping, Cyprus University of Technology, and Research Fellow at Durham Business School, Durham University*
[b] *Professor in Management Science at the Dept. of Public and Business Administration, University of Cyprus*
[c] *Associate Professor in Finance at the Dept. of Public and Business Administration, University of Cyprus*

This version: November 2012

Abstract

This study examines several alternative symmetric and asymmetric model specifications of regression-based deterministic volatility models to identify the one that best characterizes the implied volatility functions of S&P 500 Index options in the period 1996–2009. We find that estimating the models with nonlinear least squares, instead of ordinary least squares, always results in lower pricing errors in both in- and out-of-sample comparisons. In-sample, asymmetric models of the moneyness ratio estimated separately on calls and puts provide the overall best performance. However, separating calls from puts violates the put-call-parity and leads to severe model mis-specification problems. Out-of-sample, symmetric models that use the logarithmic transformation of the strike price are the overall best ones. The lowest out-of-sample pricing errors are observed when implied volatility models are estimated consistently to the put-call-parity using the joint data set of out-of-the-money options. The out-of-sample pricing performance of the overall best model is shown to be resilient to extreme market conditions and compares quite favorably with continuous-time option pricing models that admit stochastic volatility and random jump risk factors.

JEL classification: G13, G14.

Keywords: option pricing, deterministic volatility functions, implied volatility forecasting, model selection, stochastic volatility.

***** Corresponding author: Panayiotis C. Andreou. Cyprus University of Technology, Tel.: +357 25 002286; Fax: +357 25 002816. P.O. Box 50329, 3603, Lemesos, Cyprus
E-mail addresses: benz@pandreou.com (P.C. Andreou), bachris@ucy.ac.cy (C. Charalambous), baspiros@ucy.ac.cy (S.H. Martzoukos)

1. Introduction

The Black and Scholes (1973) option pricing model (henceforth BS) has shown considerable time endurance and is still widely used by practitioners since it generates reasonable prices for a wide spectrum of European financial options. Scholars and practitioners have often accounted for deviations from its assumptions by employing a regression-based Deterministic Volatility Function (DVF) approach to generate volatility values that depend on the option moneyness level and maturity.[1] As discussed in Christoffersen and Jacobs (2004), Berkowitz (2010), and Christoffersen et al. (2009), although the regression-based DVF approach cannot be considered a replacement for structural (continuous-time) option pricing models, it is widely used as a benchmark because it is an effective way to mitigate the BS volatility smile/smirk anomaly, and is quite simple and easy to apply in practice.

This study seeks to identify the overall best regression-based estimation approach for modeling the implied volatility functions of S&P 500 Index (SPX) options. Our analysis is carried out on daily data over the period January 1996 to October 2009 using the Option-Metrics database. The best models are identified by gauging their performance both in- and out-of-sample. Although the focus of this study is on one-day-ahead out-of-sample pricing tests, we reach similar conclusions when assessing the model's performance on one-week and two-week out-of-sample pricing comparisons. As a starting point, we consider the three *symmetric* DVF specifications introduced by Dumas et al. (1998), which are widely used as benchmark models in the literature (see, for instance, Andreou et al. 2010; Christoffersen et al. 2009; Linaras and Skiadopoulos 2005; Christoffersen and Jacobs 2004; Brandt and Wu 2002). In addition, we consider input-variable transformations that can differentiate between a *relative* and an *absolute* volatility smile, as well as *asymmetric* DVF specifications that allow for different shapes for in- and out-of-the-money options.[2] We estimate all regression-based DVF models with both ordinary least squares (OLS) and nonlinear least squares

[1] DVF is an interpolative (smile-consistent) regression-based approach in which implied volatilities are smoothed across moneyness and maturities, and as such, it is effective in relaxing the BS assumption of having a single volatility per day. This approach is also commonly known as the practitioner BS model (see Christoffersen and Jacobs 2004; Christoffersen et al. 2009) or as the ad hoc BS model (see Linaras and Skiadopoulos 2005; Kim 2009; Berkowitz, 2010).

[2] The moneyness ratio K (i.e., $K = S/X$, stock index value, S, over the option's strike price, X) is used in the relative smile approach to model implied volatility as a fixed function of moneyness, while in the absolute smile approach, the strike price is used to model implied volatility as a fixed function of the strike price (see discussions in Kim 2009).

(NLS). This rationale is indicated by Christoffersen and Jacobs (2004) who suggest that when calculating implied parameters, optimization should be in respect to the option pricing function. We also seek to identify which data set allows for the best model estimation, by comparing models' pricing performance estimated using either call options only or put options only, or using joint data sets both of out-of-the-money call and put options. The search for the most suitable estimation data set is motivated by Bollen and Whaley (2004) whose findings suggest that information from puts has explanatory power on the implied volatilities of calls (see also Ahoniemi and Lenne 2009).

We also endeavor to shed more light on the performance robustness of alternative specifications that regression-based DVF models could employ. In this vein, we stress test the models' resistance to cases of significant mispricing, while, to assess their accuracy in rapidly changing market conditions, we consider pricing performance on days with large jumps in index levels and on days of high market volatility. We also compare the best-performing DVF models with Heston's (1993) stochastic volatility model and Bates' (1996) stochastic volatility and random jump model. Furthermore, we consider two models that have received significantly less attention in empirical applications using the SPX contracts - namely Bates' (1991) single-jump model, and Merton's (1976) two-jump variant model discussed in Jones (1984).

To the best of our knowledge, this is the first study to simultaneously compare and contrast all the aforementioned model specification combinations. While, no doubt, the search to identify the overall best regression-based DVF model may seem endless, we provide a comprehensive and systematic investigation of the models and methods that have been widely employed by many researchers thus far.[3] Overall, this study seeks to contribute to options pricing research by attempting to resolve the *ambiguity* in the literature about the use of regression-based DVF models.

Our findings pose a number of important implications for other studies in this area. First, in the spirit of Dumas et al. (1998), it is essential for both practitioners and scholars to be able to gauge the size of prediction (pricing) error which should be considered as 'large'. We find that the overall best regression-based DVF specification is not the one introduced by Dumas et al. (1998), even

[3] For a review of smile/smirk consistent deterministic volatility models, see Skiadopoulos (2001); for applications that involve the use of DVF models, see Ncube (1996), Dumas et al. (1998), Brandt and Wu (2002), Christoffersen and Jacobs (2004), Linaras and Skiadopoulos (2005), Kim (2009), and Andreou et al. (2010), among others.

though it has since been used as a benchmark model in many studies.[4] Second, there are studies that rely on regression-based DVF models to investigate time-series economic determinants affecting the shape of implied volatility functions (see, for instance, Peña et al. 1999, 2001; Chang et al. 2009). In the same vein, many other studies need to model stock market uncertainty - as measured by implied volatility - accurately, so that researchers, analysts and observers can gauge the effects of macroeconomic news announcements and monetary policy decisions (e.g., Vähämaa and Äijö 2011). Hence, in all of these cases it is imperative for the researchers to rely on an 'overall best' implied volatility model, otherwise they will never know whether their results are spurious (i.e., driven by the BS mis-specification that could have been otherwise mitigated by using an alternative DVF specification). Third, it is well established in the literature that the shape of the DVF specification is directly linked to the shape of the risk-neutral density implied by option market prices.[5] All-in-all, it is reasonable to infer that relying on an overall best DVF model is essential to permit the derivation of risk-neutral densities that can be used in a credible manner. Fourth, many researchers model implied volatility functions using only out-of-the-money put and call options. Although these studies make valid contributions to the literature, they are less informative on the performance of the option pricing model when relying, instead, on other definitions of options data sets. So it is both valid and interesting to examine the relative pricing merits of DVF models estimated on different option data sets. Finally, it is useful to observe whether the DVF approach performs better than structural (continuous-time) models, such as those that admit stochastic volatility and random jump risk factors. As Tompkins (2001) discusses, even if the process of pricing an underlying asset is correctly identified - for instance as being a stochastic volatility and jump process - it is not certain whether market option prices would conform to this process when non-traded sources of risk (for example, unknown risk-premium dynamics) are involved. In contrast, the DVF approach is not impeded by the

[4] Relying on the overall best DVF model is important for other studies as well, such as that of Brandt and Wu (2002), where option parameters are estimated from liquid European options and are then applied to price illiquid, exotic or other derivatives (see also, Linaras and Skiadopoulos 2005; Berkowitz 2010; Chang et al. 2012). Moreover, several studies using nonparametric option pricing methods such as kernel regression or neural networks also have a need for a proper benchmark model (see Andreou et al. 2008).

[5] Shimko (1993) was the first to show how to recover the risk-neutral probability density function by fitting a smooth curve to the implied volatility smirk. Moreover, Bakshi et al. (2003) show that the shape of the implied volatility function is directly linked to the risk-neutral skewness and kurtosis of the implied distribution. Panigirtzoglou and Skiadopoulos (2004) show how to model the dynamics of implied distributions for (smile-consistent) option pricing purposes.

existence of such factors. In cases where DVF models perform better than structural option pricing models, we suggest that they should be seriously considered for valuation and risk-management purposes.

We proceed as follows: section 2 describes the alternative DVF modeling approaches we use; section 3 discusses the data and research methodology; sections 4 and 5 presents our empirical findings and discuss various robustness tests; and section 6 sets out our conclusions.

2. The Black and Scholes Deterministic Volatility Functions Approach

The BS model is widely referenced as a benchmark. The formulae for European options, modified for a dividend-paying underlying asset are:

$$c^{BS} = Se^{-d_y T} N(d) - Xe^{-rT} N\left(d - \sigma\sqrt{T}\right), \qquad (1.1)$$

$$p^{BS} = Xe^{-rT} N\left(-d + \sigma\sqrt{T}\right) - Se^{-d_y T} N(-d), \qquad (1.2)$$

with

$$d = \frac{ln(S/X) + (r - d_y)T + (\sigma\sqrt{T})^2/2}{\sigma\sqrt{\mathrm{T}}}, \qquad (1.3)$$

where c^{BS} (p^{BS}) is the price of the European call (put) option; S is the spot price of the underlying asset; X is the strike price of the option; r is the continuously compounded risk-free interest rate; d_y is the continuous dividend yield paid by the underlying asset; T is the time left until the option maturity date; σ is the yearly volatility of the rate of return for the underlying asset, and $N(.)$ represents the standard normal cumulative distribution. The regression-based DVF approach is used to smooth the implied volatility surface across the option's moneyness and time-to-maturity; in this manner, the resulting volatility function produces contract-specific values for σ that can be subsequently used with the model to price options.

The main purpose of this study is to examine several alternative specifications of regression-based DVF models so as to determine which best characterizes the daily implied volatility functions of the S&P 500 Index options: to the best of our knowledge, there is no clear consensus regarding the best model for this purpose. Since this problem can be best addressed with empirical examination, we consider the following symmetric DVF models as a starting point:

$$\sigma_X^s = \max\left(0.01, a_0 + a_1 X + a_2 X^2 + a_3 T + a_4 XT + a_5 T^2\right), \tag{2}$$

$$\sigma_{LnX}^s = \max\left(0.01, a_0 + a_1 LnX + a_2 (LnX)^2 + a_3 T + a_4 (LnX) T + a_5 T^2\right), \tag{3}$$

$$\sigma_K^s = \max\left(0.01, a_0 + a_1 K + a_2 K^2 + a_3 T + a_4 KT + a_5 T^2\right), \tag{4}$$

$$\sigma_{LnK}^s = \max\left(0.01, a_0 + a_1 LnK + a_2 (LnK)^2 + a_3 T + a_4 (LnK) T + a_5 T^2\right), \tag{5}$$

where, σ_j^s represents the volatility of the underlying asset, which is obtained by fitting each of the regression-based DVF models to option market prices per day *t*; further, superscript symbol '*s'* is used to denote a '*symmetric'* volatility specification, while subscripts $j = X, LnX, K, LnK$, denote the underlying asset price variable used to measure the options' moneyness (where $K = S / X$ determines the option's moneyness ratio and *LnK* determines the natural logarithm transformation of *K*). Following Dumas et al. (1998), a minimum value of 1% is imposed when estimating the regression functions to prevent negative volatility values.

The symmetric DVF specifications define implied volatilities as quadratic polynomial functions of moneyness, $j = X, LnX, K, LnK$, and time-to-maturity, T; hence, all models are able to capture the (empirical) curvature of implied volatility surfaces with respect to moneyness, as well as the empirical presence of curvature (even humps) in their term structures.[6] The cross product of moneyness and time-to-maturity is important, since it allows capturing changes in the shape of the implied volatility functions over different maturities.

From a computational point of view, we expect σ_{LnX}^s to perform better than σ_X^s because of its better scaling of the regressors, which is important in the NLS estimation of the models. Under a proper scaling scheme, such as using the natural logarithm of the strike price to measure option moneyness, we expect greater precision of the estimated parameters, faster convergence of the

[6] There is abundance of empirical evidence of such patterns in the time structure of SPX implied volatilities in previous studies, such as Bakshi et al. (1997), Christoffersen et al. (2009), and Andreou et al. (2010).

optimization process, and significant reductions in computation time.[7] In contrast, we do not expect σ^s_{LnK} to perform better than σ^s_K, because the natural logarithm of the moneyness ratio in the latter does not offer a significant scaling benefit in this case. One advantage of σ^s_{LnK} over σ^s_K, however, is that the implied at-the-money volatility value is explicitly defined by the intercept coefficient of the regression specification in Eq. (5), since the natural logarithm of an option that is exactly at-the-money equals zero.

Despite the focus of most prior studies on the σ^s_X specification and the utilization of shorter time periods, empirical evidence supports the use of symmetric DVF specifications for SPX contracts. Thus it is not surprising to find the symmetric DVF specifications discussed above to perform well. Dumas et al. (1998), for example, consider the period June 1988 - December 1993 to find that σ^s_X performs quite well (see also Brandt and Wu, 2002 and Linaras and Skiadopoulos, 2005 for applications of implied volatility trees).[8] Christoffersen and Jacobs (2004), consider the June 1988 to May 1991 period to demonstrate that when σ^s_X is estimated using NLS, it outperforms Heston's (1993) stochastic volatility model. Moreover, Christoffersen et al. (2009) compare σ^s_X against a two-factor stochastic volatility model, and report that it underperforms the structural option pricing model in the early part of their sample period, and slightly outperforms the two-factor stochastic volatility model in the latter sample part. However, prior empirical evidence relevant to the performance of the other symmetric models σ^s_{LnX}, σ^s_K, and σ^s_{LnX} is rather limited, so it is interesting to study their pricing performance for SPX contracts. Our study contributes in this respect by filing this significant empirical void.

Tompkins (2001) reports that implied volatility surfaces produced by futures options display

---

[7] Nonlinear optimization algorithms employed for estimating the nonlinear least squares versions of the DVF models might be affected by large differences in the levels of the observed variables. As suggested by Nocedal and Wright (1999, pg. 27), the performance of an algorithm may depend critically on how the problem is formulated. In addition, the use of very high or low numbers, which can occur when the (untransformed) strike price is used, can cause underflow or overflow problems.

[8] Dumas et al. (1998) use σ^s_X in most of their analysis, but in a few cases where option cross-sections have fewer than three expiration dates available, they use a reduced version that ignores some of the time-terms. In our case, there was no such problem, since there were no days with less than three expiration dates.

much greater asymmetry than what can be produced by stochastic volatility models, suggesting that asymmetric DVF specifications can be useful in cases where the implied volatility functions diverge significantly from the well-known skew-shaped smirk pattern. The asymmetric DVF models we consider for analysis are:

$$\sigma_X^a = \max\left(0.01,\, a_0 + a_1 D^u X + a_2 X^2 + a_3 D^d X + a_4 T + a_5 XT + a_6 T^2\right), \tag{6}$$

$$\sigma_{LnX}^a = \max\left(0.01,\, a_0 + a_1 D^u LnX + a_2 (LnX)^2 + a_3 D^d LnX + a_4 T + a_5 (LnX)T + a_6 T^2\right), \tag{7}$$

$$\sigma_K^a = \max\left(0.01,\, a_0 + a_1 D^u K + a_2 K^2 + a_3 D^d K + a_4 T + a_5 KT + a_6 T^2\right), \tag{8}$$

$$\sigma_{LnK}^a = \max\left(0.01,\, a_0 + a_1 D^u LnK + a_2 (LnK)^2 + a_3 D^d LnK + a_4 T + a_5 (LnK)T + a_6 T^2\right), \tag{9}$$

with D^u and D^d defined as:

$$D_i^u = \begin{cases} 1 \text{ if } & S/X_i > 1 \\ 0 \text{ if } & S/X_i \le 1 \end{cases}, \qquad \text{and} \qquad D_j^d = \begin{cases} 1 \text{ if } & S/X_i < 1 \\ 0 \text{ if } & S/X_i \ge 1 \end{cases}, \tag{10}$$

where subscript *i* tracks the different strike prices throughout a given day; D^u takes the value 1 for in-the-money (out-of-the-money) call (put) options, and 0 otherwise. D^d takes the value 1 for out-of-the-money (in-the-money) call (put) options, and 0 otherwise.[9] In the notation σ_j^a, superscript '*a'* is used to denote an '*asymmetric'* volatility specification, while (as before) subscripts $j = X, LnX, K, LnK$ denote the underlying asset price variable used to measure option moneyness. The asymmetric DVF models define implied volatility as a second degree relation of moneyness, $j = X, LnX, K, LnK$, and time-to-maturity, T, and allow different coefficients for the left and right wings of the specifications' linear element. Therefore asymmetric DVF specifications offer even greater flexibility in modeling implied volatility than their symmetric counterparts.

Prior literature makes only very limited use of asymmetric DVF specifications: for instance, Peña et al. (1999) employ some reduced versions (ignoring the time dimension) of the asymmetric models σ_K^a and σ_{LnK}^a to investigate the behavior of Spanish implied volatility functions, and

[9] In a previous version of this paper, which used SPX data for the period 1998 to 2004, we also estimated all asymmetric DVF models with the threshold value for the dummy variable to be either K = 1.05 or K = 1.10. These specifications do not work better than those where the threshold is set equal to 1 (K = 1).

Engström (2002) uses a reduced version of σ^{a}_{LnK} to investigate Swedish stock options. Although asymmetric DVF specifications might be more suitable for implied volatility functions characterizing emerging option markets, there is neither theoretical nor empirical evidence to preclude their use in investigating developed options markets. To the best of our knowledge, this study is the first to report evidence for the pricing performance of asymmetric DVF specifications for SPX contracts.

3. Data and Methodology

3.1. Options and other data

Our data set covers the period January 1996 to October 2009, a total of 3,480 trading days. Daily option contract transaction prices and dividend yields are from Option-Metrics, and daily risk-free rates from U.S. Federal Reserve Bank Statistical Releases. We use the midpoint of the closing option bid–ask spread, since relying on such midpoints rather than trading prices reduces noise in the cross-sectional estimation of implied parameters (Dumas et al. 1998). The midpoint of each day's call (put) option bid–ask spread at the market's close , c^{mrk} (p^{mrk}), is matched against the closing value of the S&P 500 Index.[10] In our analysis, time-to-maturity, T , is computed assuming 252 trading days per year. We also use nonlinear cubic spline interpolation to match each option contract against a continuous interest rate, r , that corresponds to the option's maturity. For this purpose, 1-, 3-, 6-, and 12-month constant maturity T-bills rates are considered.

We rely on the following filtering rules (Bakshi et al. 1997; see also Andreou et al. 2010) to create our final data set. First, we eliminate all observations that have zero trading volume, since these do not represent actual trades. Second, we eliminate observations that violate either the lower or the upper arbitrage pricing bounds and, similarly, we exclude observations with price quotes of less than 1.0 index points, with either implied volatility lower than 5% or higher than 70%, or with midpoint option price lower than the bid–ask spread difference. Third, we eliminate all observations with either

[10] Data synchronicity is a minor issue for this highly active market. Among others, Constantinides et al. (2009), Christoffersen et al. (2006), Christoffersen and Jacobs (2004), and Chernov and Ghysels (2000) use daily closing prices of European options written on the S&P 500 Index. Other related studies that use daily closing prices include Ncube (1996), Peña et al. (1999), Engström (2002), Brandt and Wu (2002), Chen et al. (2009), Andreou et al. (2010) and Mozumber et al. (2012).

less than 6 or more than 253 trading days to maturity or moneyness ratio that is either less than 0.75 or higher than 1.25; these eliminations are adopted to avoid any illiquid option observations. The final data set has a total of 502,630 observations - 231,215 from call options and 271,415 from put options.[11]

Sample characteristics for the whole data set are reported in Table 1, from which it is evident that volatility anomalies are present and are stronger for short-term options. It is also clear that out-of-the-money calls and puts involve higher levels of volume trading and lower bid–ask spreads compared to those that are in-the-money.

[Table 1 here]

3.2. Methodology for estimating and evaluating the models

Studies in this area rely mostly on OLS estimations (e.g., Ncube 1996; Peña et al. 1999; Engström 2002; Chang et al. 2009; Kim, 2009). Estimating the DVF parameters with OLS is straightforward - for any given day, we back out the contract-specific BS volatility and regress on the explanatory variables via OLS. However, Christoffersen and Jacobs (2004) demonstrate that OLS estimates of the DVF parameters yield biased option pricing results, and suggest that NLS pricing loss functions should be used instead. The methodology to obtain the NLS coefficients is more complex: specifically, it is similar to that employed in previous studies (Bakshi et al. 1997; Dumas et al. 1998; Christoffersen and Jacobs 2004; Christoffersen et al. 2006; Chen et al. 2009; Andreou et al. 2010; Kuo 2011; Mozumder et al. 2012) and relies on a simultaneous equation procedure to minimize a price deviation function with respect to the unknown parameters. Market option prices, o^{mrk}, are assumed to be the corresponding model prices, o^{m}, plus a random additive disturbance term, ε:

$$o^{mrk} = o^{m} + \varepsilon . \tag{11}$$

[11] We use the following moneyness ratio classes: deep out-of-the-money (DOTM) for $0.75 \leq K < 0.90$, out-of-the-money (OTM) for $0.90 \leq K < 0.95$, just out-of-the-money (JOTM) for $0.95 \leq K < 0.99$, at-the-money (ATM) for $0.99 \leq K < 1.01$, just in-the-money (JITM) for $1.01 \leq K < 1.05$, in-the-money (ITM) for $1.05 \leq K < 1.10$, and deep in-the-money (DITM) for $1.10 \leq K \leq 1.25$. In terms of maturity, an option contract is classified as being of short-term maturity (where its maturity is ≤ 60 calendar days), as medium-term (where its maturity is between 61 and 180 calendar days) and as long-term (where its maturity is > 180 calendar days).

To find the implied parameter values per model, the Sum of Squared Errors (*SSE*) optimization problem is solved, as follows:

$$SSE(t) = \min_{\xi^m} \sum_{j=1}^{n_t} (\varepsilon_j)^2 , \qquad (12)$$

where n_t refers to the number of different option observations available in day *t*, and ξ^m refers to the unknown parameters associated with a particular DVF specification. Model option prices (o^m) are computed using the BS model shown in Eqs. (1.1) and (1.2), with contract-specific volatility values defined in each case by some of the DVF specifications in Eqs. (2) through (9). The *SSE* function is minimized using the Levenberg-Marquardt method, with a line search based on a mixed quadratic and cubic polynomial interpolation and the extrapolation method offered by Matlab. Berkowitz (2010) demonstrates theoretically that the DVF constitutes a reduced-form approximation to an unknown structural model which, if estimated frequently (e.g., daily), can yield exceptional option pricing performance; so the unconstrained optimization problem in Eq. (12) is solved on a daily basis for all DVF models considered.[12] For sake of brevity, our results only report pricing performance of the models based on NLS estimations (since in this study OLS results are always found to be inferior to the NLS ones).

As Christoffersen and Jacobs (2004) point out, the estimation and evaluation of a model should be based on the same error measure; they also suggest that, among different loss functions, Root Mean Squared Error (RMSE) estimates may perform best (see also Chang et al. 2012). Bates (2000) also points out that the RMSE is also a relatively intuitive error measure and very useful when comparing the empirical performance of different option pricing models: so we only report statistics based on this error measure.[13]

[12] Daily recalibration of the implied coefficients is also adopted by Christoffersen and Jacobs (2004); see also similar treatments and discussions in Ncube (1996), Peña et al. (1999), Engström (2002), Kim (2009), and Andreou et al. (2010).

[13] Patton (2010) reports that among many alternative loss functions Mean Squared Error is robust to the noise in the volatility proxy and yields correct rankings of models used to forecast volatility. Despite the overwhelming (theoretical and empirical) evidence in favor of the Mean Squared Error metric, we also compute the Mean Absolute Error, the Median Absolute Error, and the 5th and 95th Percentile of Absolute Error (results are available upon request—see also our robustness analysis in Section 5.1). Overall, we observe that all of our findings remain unchanged when using any of the alternative loss functions.

3.3. Estimation and evaluation data sets

A novel feature of our analysis is that all DVF specifications are *estimated separately* using three different data sets. The first data set is composed of all daily call options; the second that comprises all daily put options; the third takes in the joint data set of all available out-of-the-money call ($K < 1$) and put ($K \geq 1$) options. As Table 1 shows, out-of-the-money options, which are exclusively included in the third dataset, preserve the highest levels of liquidity in terms of trading volume and the smallest bid–ask spreads (compared with the rest in-the-money options). Out-of-the-money options also contain vital pieces of information regarding the demand for portfolio insurance - especially the out-of-the-money puts - which allows the model to capture market volatility more effectively. We would therefore expect the third set to be more informative about market volatility.

Searching for the best estimation data set that could allow the option model to achieve the highest out-of-sample pricing accuracy has received little attention in the literature - in particular, no other study has employed an approach similar to the one proposed in this study.

The majority of prior research (e.g., Christoffersen and Jacobs 2004; Christoffersen et al. 2009; Andreou et al. 2010) reports option pricing performance using the same estimation and evaluation data set (i.e., models are estimated and their performance is assessed using, for example, only call options). However, in this study we also seek to identify the DVF model that performs best across the entire cross-section of options contacts, so we evaluate the models' pricing performance, separately, on data sets of call options only, of put options only, and of the entire cross-section of options (i.e., all available call and put options together). This thorough approach makes our analysis more informative than previous studies in this regard as well.

4. Discussion of Results

4.1. Pricing performance of the option models

Table 2 summarizes the pricing performances of the different DVF models, in which their in-sample, and one-day-ahead, out-of-sample, pricing accuracy are measured in terms of RMSE. Table 2 is divided into three panels: Panel A reports the models' RMSE performance when used to price the daily cross-section of call options only; Panel B reports model performance when used to price the daily cross-section of put options only, while Panel C reports RMSE values when the models are used to price daily the entire cross-section of options (all call and put options together). Each panel is further divided into three parts - the upper parts reporting pricing results when the DVF models are estimated using call options (i.e., first estimation data set), the middle parts reporting the results when they are estimated using put options (i.e., second estimation data set), and the lower parts reporting the results when the DVF models are estimated using the joint data set of out-of-the-money call and put options (i.e., third estimation data set where the moneyness ratio K is used to identify out-of-the-money options).

[Table 2 here]

4.1.1. In-sample pricing performance

As in previous studies (e.g., Christoffersen and Jacobs 2004; Andreou et al. 2010), we observe that the DVF models estimated with NLS exhibit significantly smaller pricing RMSEs compared to those estimated with OLS - in most cases, the improvements in pricing accuracy are over 40% in terms of RMSE (for brevity reasons, OLS results are not reported but are available on request). This is to be expected since - as Christoffersen and Jacobs (2004) point out - estimating DVF models with OLS produces biased estimates for the observed market option prices.

We begin by discussing the models' in-sample pricing performance, from which we infer the following. First, the best in-sample fitting accuracy for call (put) options is obtained when the DVF models are estimated using call (put) options only, and not in any other way. For instance, the in-sample fitting accuracy for call options ranges from 0.772 to 0.815 when the DVF models are estimated using call options, but from 1.445 to 1.567 when the models are estimated using the joint data set of out-of-the-money call and put options, which is the second-best approach. The fitting performance of the models on the call data set when the models are estimated with put options is the poorest, with RMSE values ranging from 2.783 to 3.153. Similar observations can be made for the models when fitting put options (first best case is when the DVF models are estimated using puts, second best is when they are estimated with the joint data set of out-of-the-money options, and third best when estimated with calls).

Second, we observe that, for both symmetric and asymmetric model classes, better in-sample results can be obtained when using the moneyness ratio K as an input (i.e., using a relative smile approach). Looking, for instance, into the performance of alternative asymmetric models when are estimated using call (put) options and evaluated in-sample on call (put) options, we observe that σ_K^a RMSE equals 0.772 (0.821) against the RMSE for σ_X^a, which equals 0.806 (0.843), while σ_{LnK}^a RMSE equals 0.779 (0.817) against the RMSE of σ_{LnX}^a, which equals 0.783 (0.821). This result coincides with Kim's (2009) findings, according to which the relative smile DVF models that rely on the moneyness ratio K exhibited the best in-sample pricing performance for KOSPI 200 Index options.

Third, in comparing the symmetric DVF models against their counterpart asymmetric ones, we find that the asymmetric models perform better in-sample. This happens since extra terms in their regression specifications allow for different in- and out-of-the-money slopes for the volatility function, and thus offer greater modeling flexibility when fitting the daily implied volatilities of the SPX contracts. These findings hold true only for cases in which models are estimated and evaluated with the same data set.

Our in-sample results have important implications for the literature that attempts to model implied volatility functions by (usually) relying on a single DVF specification without taking into account the type of options data set (i.e., call vs. put options) - in particular, Peña et al. (1999), Engström (2002), Brandt and Wu (2002), Christoffersen and Jacobs (2004), Christoffersen et al. (2009). Our results suggest both that DVF models should be fitted separately to call and put options for in-sample applications, and that different DVF model specifications are needed to fit properly the daily shapes of call and put implied volatility functions.

4.1.2. Put-call parity violations

Before moving into the out-of-sample comparisons, there is one tacit implication that relates to put-call-parity violations when DVF models are fitted to call and put options separately, and which has been widely ignored in the literature. Theoretically speaking, put-call parity does not break down and precludes arbitrage opportunities as long as the relevant calls and puts exhibit the same levels of volatility. When a single DVF model is estimated by using the joint data set of out-of-the-money call and put options, there is no resulting violation and put-call parity holds true for any paired call and put option combination. But this does not hold true when (different) DVF models are used to model implied volatilities of calls and puts separately - in such case, it is highly likely to observe different volatilities for paired call and put option combinations. Effectively, this translates to a breakdown of put-call parity when (theoretical) BS values of the options are used to validate that parity.[14]

We examine in-sample violations related to put-call parity when call options are valued using σ_{LnX}^{s} estimated with call options (i.e., the upper part of Panel A of Table 2) and separately estimated with put options (i.e., the middle part of Panel B of Table 2).[15] From unreported statistics (available from the authors on request), we observe that price violations in terms of RMSE vary across the seven moneyness classes we consider in our analysis between 1.41 and 2.45 for short-term options (category average is 2.08), between 2.35 and 3.51 (average 2.82) for medium-term options, and between 3.47 and 6.34 (average 4.13) for long-term options These price violations are economically significant when compared to the average bid–ask spreads of the options involved in the computation of the put–call parity values, and therefore reveal severe mis-specification issues in adopting such an approach.[16] As shown above, the approach of estimating DVF models using the joint data set of out-of-the-money call and put options leads to suboptimal in-sample fitting of the models. Nevertheless, model estimation in this respect is well specified in theoretical terms and should be preferred (at least) when trying to assess model pricing accuracy out-of-sample (as shown below).

4.1.3. Out-of-sample pricing performance

[14] Refer to Ahoniemi and Lanne (2009) for reasons that might cause the empirical inequality of paired put and call implied volatilities in index options (see also Bollen and Whaley, 2004).

[15] We use σ_{LnX}^{s} in BS because we find it to be superior in terms of out-of-sample pricing accuracy; yet, similar violations are obtained with all other DVF specifications.

[16] Average bid–ask spreads for the seven moneyness classes vary between 1.57 and 2.86 for short-term options (category average is 1.64), between 1.83 and 2.73 for medium-term options (average is 2.04), and between 2.14 and 2.67 for long-term options (average is 2.29).

Table 2 reports the models' out-of-sample pricing performance and Table 3 reports the *t*-statistics from comparing the respective mean squared pricing residuals performance (models' performance in the columns against the performance of those in the rows). All *t*-statistics regarding the one-day-ahead out-of-sample pricing errors are adjusted for heteroskedasticity using White's (1980) robust standard errors.[17]

First, as in the in-sample comparison cases, we can observe that DVF models estimated with OLS are significantly inferior to those estimated with NLS when pricing options out-of-sample (for brevity sake, OLS estimation results are not reported but are available on request). In most cases, the pricing accuracy improvement exceeds 20–30% in terms of RMSE (the one-week and two-week-ahead comparisons show similar improvements). Our findings point to potential problems in identifying the correct model performance ranking when DVF models are estimated with OLS instead of NLS. In this respect, we can explain the puzzling performance of DVF models in relation to other widely referenced models, such as Heston's (1993) stochastic volatility model. When DVF models are estimated with OLS, they underperform stochastic volatility models (see for instance, Christoffersen and Jacobs 2004; Kim 2009), while when they are estimated with NLS, they are found to outperform such models (see results in Christoffersen and Jacobs 2004; Andreou et al. 2010).

Second, in direct comparisons of the best-performing models, we find that symmetric DVF models perform much better than asymmetric ones in all cases. The superiority of the less-parameterized symmetric models is not surprising, since over-parameterized models entail the risk of over-fitting the options data resulting into poor out-of-sample pricing performance (Bates 2000; Chen et al. 2009). Our findings indicate that, relative to asymmetric DVF models, symmetric ones are significantly more resilient to the choice of the data set used in the estimation, making them more trustworthy tools for out-of-sample pricing purposes.

Our third observation regards the out-of-sample performance of symmetric models when estimated with a particular type of data set and evaluated on the entire cross-section of option

[17] For brevity, in Table 3 we report *t*-statistics for a subset of models that exhibit the overall best out-of-sample pricing accuracy. For all model comparisons, we have also computed Student's *t*-statistics, as well as the Johnson (1978) modified *t*-statistics for non-normal distributions, which account for the presence of skewness in the model residuals. All results are qualitatively similar to those reported in Table 3 (the same holds true for the *t*-statistics in Table 6).

contracts (i.e., all available call and put options together). For instance, we observe that their RMSE values when estimated with call options are between 2.498 and 2.981 (upper part of Panel C, Table 2), while when estimated with put options, they are between 2.368 and 2.783 (middle part of Panel C,). We can also see that symmetric DVF models estimated with the joint data set of out-of-the-money call and put options (lower part of Panel C) perform significantly better than the two previous models, with RMSEs between 1.922 and 2.470. From our previous observations, we may conjecture that the overall best strategy is to estimate the regression-based DVF models using the joint data set of all available out-of-the-money call and put options; the same holds true for asymmetric DVF models (although the former should be preferred due to their better out-of-sample pricing performance).

Fourth, by comparing models within each class (symmetric and asymmetric) when they are estimated using the joint data set of out-of-the-money call and put options, the best pricing performance is obtained when the logarithmic transformation of the strike price (*LnX*) is used as the input. Specifically, the smallest RMSE within the symmetric class is observed for σ^{s}_{LnX} (equal to 1.922), while the smallest RMSE within the asymmetric class is observed for σ^{a}_{LnX} (equal to 2.046). Combining all the evidence, it is obvious that σ^{s}_{LnX} estimated with the joint data set of out-of-the-money call and put options is the *overall best model* (i.e., the one that outperforms all other models when pricing performance is assessed on the entire cross-section of option observations). By investigating the *t*-statistics in Panel C of Table 3, we observe that this model is also superior to all others in statistical terms; for instance, comparing σ^{s}_{LnX} to the second-best performing model σ^{a}_{LnX} (when both are estimated with the joint options data set) results in a *t*-statistic equal to -45.846. In addition, comparing σ^{s}_{LnX} when estimated using call (put) options only, to σ^{s}_{LnX} when estimated with the joint data set, results in *t*-statistics of 54.821 (34.340); note that *t*-statistics in these cases are positive since we compare models with higher RMSE to those with a lower RMSE.

[Table 3 here]

Fifth, another interesting observation is that σ^{s}_{X} and σ^{a}_{X} are the second-best overall models within each class. By assessing the out-of-sample RMSE values reported in the lower part of Panel C in Table 2, we observe that the absolute smile DVF models (those that use the strike price, X, to measure option moneyness) significantly outperform the relative smile models (those that use the moneyness ratio, K, to measure option moneyness). In fact, the differences in pricing accuracy are huge - for instance, we observe that the out-of-sample RMSE performance of absolute smile symmetric (asymmetric) models is between 1.922 and 1.938 (2.046 and 2.058), while the RMSE for relative smile models is between 2.465 and 2.470 (2.474 and 2.550). The superiority of the absolute smile models in out-of-sample pricing accuracy appears to contradict the fact that better in-sample results are obtained when using relative smile models instead. Our results are in the same line of reasoning as those of Kim (2009), according to whom the relative smile DVF models perform better in-sample and the absolute smile DVF models better out-of-sample. Kim (2009) conjectures that this result is due to the fact that the absolute model treats the smile as a fixed function of the strike price, and so predicts a smaller implied volatility than the relative smile model when there is an increase in asset price. We further believe that the superiority of the absolute smile over the relative smile model indicates that there are significant idiosyncratic factors associated with implied volatility functions. In that respect, traders seem to be concerned with specific contracts (defined by their strike price, *X*), not simply the percentage by which a contract is in- or out-of-the-money (as defined by their moneyness ratio, *K*). Such an explanation is in the spirit of the empirical evidence reported by Bollen and Whaley (2004), who find that the shape of the SPX-implied volatility functions is attributable to the buying pressure of specific options series and arbitrageurs' limited ability to bring prices back into alignment.

Table 4 provides further evidence that the best approach is to estimate the DVF models using the joint data set of out-of-the-money call and put options. Out-of-sample pricing results in this table report the RMSE values of σ^{s}_{LnX} in terms of 21 moneyness ratio and time-to-maturity classes. The bottom panel also reports RMSE per moneyness class (aggregating time-to-maturity), and the last column RMSE per time-to-maturity (aggregating moneyness). We report results for six different testing schemes: C$\rightarrow$C **(**P$\rightarrow$C**)** indicates that the model is estimated daily with call (put) options and evaluated out-of-sample solely on call options, while CP$\rightarrow$C indicates that the model is estimated daily with the joint data set of out-of-the-money call and put options and evaluated out-of-sample solely on call options. Similarly, in the notations C→P, P→P, CP→P, the first (left-hand) capital letter indicates the data set used to estimate the model, while the second (right-hand) capital letter indicates the evaluation data set (in this case all models are evaluated with put options).

[Table 4 here]

The overall conclusion from Table 4 confirms again that the most robust strategy is to estimate the DVF models by using the joint data set of available out-of-the-money call and put options. Specifically, one would expect the C→C approach to be the most accurate when used to price call options out-of-sample, but, in fact, we observe that the CP→C approach outperforms C→C in almost all cases (underperforming only in the DOTM case). The P→C approach is, in general, noticeably inferior (even to the C→C approach) in pricing call options out-of-sample. In the same vein, one would expect the P→P approach to be most accurate when used to price put options out-of-sample, but we can observe that it underperforms CP→P in many cases. For example, in the case of short-term options, CP→P (with an overall RMSE of 1.648) proves superior to P→P (RMSE 1.736) - while C→P (RMSE 2.051) is even less accurate. This forms a strong evidence against the P→P approach, since short-term options comprise the largest part of our data, are the most frequently traded in the market, are responsible for the highest levels of trading volumes and present the most pronounced implied volatility smile anomaly (see Table 1). Further, we find that CP→P also dominates the P→P approach in many of the moneyness classes of medium- and long-term options: for instance, it performs better than P→P for all JOTM, ATM, and JITM cases – as Table 1 shows, these options involve the highest daily volumes in medium- and long-term options.

As a final test, we compare the aggregate pricing performance of: (i) σ^{s}_{LnX} when estimated with the joint data set of out-of-the-money call and put options and evaluated out-of-sample on the entire cross section of call and put options (i.e., using CP→C and CP→P results together), against the combined performance of: (ii) σ^{s}_{LnX} when estimated with call options and used to price out-of-sample call options only (i.e., the pricing performance of C→C) combined with σ^{s}_{LnX} when estimated with put options and used to price out-of-sample put options only (i.e., the pricing performance of P→P). In case (i) the one-day-ahead out-of-sample RMSE performance is 1.922 (lower part of Panel C, Table 2), while (from unreported statistics) RMSE performance in case (ii) is 1.954: this performance difference (1.954 vs. 1.922) is statistically significant at the 1% level (*t*-statistic equals 3.43). These results again indicate that employing σ_{LnX}^{s} with the combined CP→C and CP→P approach (consistent with put-call parity during in-sample estimation) is superior to the mis-specified case where σ_{LnX}^{s} is employed separately with the C→C and P→P approach (which is inconsistent with put-call-parity during in-sample estimation).

We offer three justifications to explain why estimating the DVF models using the joint data set of out-of-the-money call and put options is found to be superior to any other approach. First, Bollen and Whaley (2004) find that the slope of the daily SPX-implied volatility functions is erratic across time. The options contracts included in the joint data set bear the highest sensitivity with respect to volatility (options with high vega values), and thus are more informative as to the likely future behavior of SPX-implied volatilities. Second (as Table 1 shows) out-of-the-money options also involve the highest trading volumes and the lowest bid–ask spreads; as a result, this estimation data set minimizes the possibility of using options that are affected by illiquidity and measurement errors. Finally, a third explanation is associated with the optimization approach of Eq. (12). As Bakshi et al. (1997) discuss (see also Bates 1996, 2000; Dumas et al. 1998; Kuo 2011), the loss function used in Eq. (12) may force the estimation to assign more weight to relatively expensive options (which are usually those that are in-the-money with long maturities), and less to relatively cheap options, which are usually short-term, at-the-money and out-of-the-money options. In contrast, using only out-of-the-money calls and out-of-the-money puts minimizes this inherent limitation of the estimation methodology, since all options included in the estimation data set are relatively cheap and have comparable market values.

4.2. Hedging analysis

Many recent empirical studies - such as those of Kim (2009) and Andreou et al. (2010) - assess the volatility forecasting power of option pricing models by including single-instrument hedging strategies in their analyses, such as that employed in Bakshi et al. (1997). However, we abstain from employing such a hedging strategy for our best-performing DVF models in this study for two reasons. First, there is abundant empirical evidence to suggest that the models' hedging performance of different models is virtually indistinguishable, and also that their hedging-based rankings are in sharp contrast to their out-of-sample pricing performance (see discussions in Dumas et al. 1998; Chernov and Ghysels 2000; Kim 2009; Kuo 2011; Andreou et al. 2008, 2010), making it more difficult for us to draw sensible conclusions. Second, Christoffersen and Jacobs (2004) suggest that the aforementioned ambiguity may be due to inappropriate choice of the loss function used to calibrate the models for hedging purposes, and suggest the best possible parameter estimates for a hedging exercise are likely to be obtained using a hedging based loss function. Andreou et al. (2010) work with this intuition and calibrate their option pricing models based on a hedging criterion to observe significant variations in their hedging performance, thus making it easier to rank the models according to their hedging outcomes. Nevertheless, these authors also find that this calibration approach significantly deteriorates the models' out-of-sample pricing accuracy. Their results suggest that option pricing models intended to be used for pricing purposes should be calibrated based on a pricing loss function, while those intended to be used for risk-management purposes (i.e., hedging) should be calibrated based on a hedging loss function. Such estimation treatments are beyond the scope of the current study and we leave it for future research.[18]

5. Robustness Analyses

5.1. Comparisons with reduced-version DVF models

Bates (2000) notes that over-parameterized models entail the risk of over-fitting the options data and so exhibit poor performance when used to price options out-of-sample (see also Brandt and Wu 2002;

[18] Mozumder et al. (2012) document that more complex hedging strategies such those that rely on delta and delta-gamma approaches yield inaccurate approximations of option portfolio values, especially in the face of large swings in the price of the underlying asset. Yet, the more sophisticated delta-gamma approximation is not significantly more accurate than the easier to implement single-instrument delta-hedging approximation. These authors also report that even the most sophisticated option pricing models they employ in their empirical investigation deliver poor risk-management effectiveness during times of financial market turbulence.

Linaras and Skiadopoulos 2005). In this respect, a more general/flexible DVF specification may not necessarily improve the pricing performance of the BS model over a volatility specification that includes fewer free parameters. To examine this issue even further, we repeated our entire analysis, carrying out all the aforementioned estimations on a broad set of reduced-version DVF specifications. Specifically, in addition to the eight DVF models detailed in Eqs. (2) through (9), we considered 44 more reduced-version DVF specifications. The least complex DVF specifications included only linear and squared terms of moneyness measures (ignoring time-to-maturity terms), while other specifications also included time-to-maturity terms and/or cross products of moneyness and time-to-maturity. We summarize all these estimation in an Appendix which includes all relevant tables accompanied by a brief discussion of the results available on request from the authors.[19]

In terms of pricing accuracy, the DVF specifications considered for our main analysis in this study (Eqs. (2) through (9)) still delivered the overall best in- and out-of-sample performance when compared with these reduced-version DVF alternatives.[20] Moreover, following careful inspection of all additional results, we find that the overall conclusions relevant to all aspects of the comparisons carried out above remained unaltered.

5.2. Comparisons with other BS volatility measures and structural option pricing models

Among others, Linaras and Skiadopoulos (2005) note that the performance of smile-consistent models vs. the more complex stochastic-volatility-implied models demands further investigation. As part of our robustness analysis, we compared the best-performing regression-based DVF models with more advanced structural (continuous-time) parametric option pricing models. In particular, we considered Heston's (1993) Stochastic Volatility (SV) model, which has been extensively used in

---

[19] As in the main analysis of this study, all additional (reduced-version) DVF specifications are estimated daily based on three different data sets: *i)* all call options, *ii)* all put options, and *iii)* the joint data set of out-of-the-money call and put options. As discussed in Section 5.3., to validate the robustness of the estimation data set, we also estimated all DVF specifications using the joint data set of out-of-the-money call and put options, whereas moneyness was based on the options' delta value. We estimate daily the additional 44 DVF models separately on each data set using both OLS and NLS. In this Appendix, we tabulate in- and out-of-sample pricing results (one-day, one-week, and two-weeks ahead) of all models considered in this study on all three evaluation data sets as discussed in the main analysis, using a variety of error metrics (i.e., Root Mean Squared Error, Mean Absolute Error, Median Absolute Error, and 5th and 95th Percentile of Absolute Error).

[20] We find, though, a few asymmetric specification cases where DVF models that exclude the time-term T^2 perform slightly better (this is more pronounced when the estimation and evaluation data sets are different).

previous studies (e.g., Christoffersen and Jacobs 2004; Chen et al. 2009; Kim 2009), and for a more complete analysis, we also considered Bates' (1996) Stochastic Volatility and random Jump (SVJ) model which is known to perform better for the case of SPX contracts than the SV (Andreou et al. 2010). We also considered two further models that have received significantly less attention in empirical applications of SPX contracts - namely Bates' (1991) single-jump model (1Jump), and Merton's (1976) two-jump model (2Jump) as discussed in Jones (1984). For sake of brevity, we refrain from displaying the closed-form solutions of these option pricing models and discussing their diffusion properties here since they can be found in the original published papers.

Moreover, we also estimated Black and Scholes' (1973) model (BS) using both a daily overall-average implied volatility (BS-OV), and the one-day lagged value of the CBOE VIX index (BS-VIX). All these models (except BS-VIX) are estimated via Eq. (12) and, for comparability with our previous analysis, estimated and evaluated with all possible data sets (as discussed in the main analysis of this study). Table 5 presents the in-sample and one-day-ahead out-of-sample pricing performance of these models (using a structure similar to that of Table 2).

[Table 5 here]

From results in Table 5, it is clear that both BS-OV and BS-VIX perform poorly, both in- and out-of-sample regardless of the data set used. Of the two, BS-OV performs better since its implied volatility is more informative, as it is computed using all available expiration dates in the options cross-section (while VIX is always computed using the proximate and the second-proximate options with at least eight days left to expiration).[21]

We observe that models with more structural parameters produce smaller in-sample pricing RMSE values; in particular 2Jump and SVJ models outperform 1Jump and SV models, respectively, for almost all evaluation data sets. This can be expected, since as in the case of asymmetric DVF models the presence of extra parameters allows these models to have more flexible distributions: that is, 2Jump and SVJ models are more flexible and thereby able to internalize negative skewness and leptokurtosis to better fit daily options.

[21] The underperformance of the VIX is not surprising. In a financial econometrics context, Becker and Clements (2008) find that the VIX index produces volatility forecasts that are inferior to model-based forecasts of realized volatility.

Regarding the out-of-sample comparisons between the SV and 1Jump models with the BS model when estimated with the joint data set of out-of-the-money call and put options (lower part of Panel C in Table 5), the former is able to lower BS-OV RMSE value from 5.166 to 3.184, while the latter can only lower its value to 3.486. This makes the presence of stochastic volatility more important in capturing the salient features of asset returns. It is noteworthy that out-of-sample, 2Jump is inferior to SV and only slightly better than 1Jump. As in Bakshi et al. (1997), we find that the presence of a stochastic volatility risk factor achieves a first-order pricing improvement over the BS model. However, as Tompkins (2001) notes, stochastic volatility on its own is not sufficient to reconcile theoretical/model-based volatility smiles with those observed in reality. Thus, the addition of a jump risk factor should help to improve out-of-sample pricing performance even further, since it allows both long and short-maturity smiles to be matched within a single model (Bates 1996; see also discussions in Andreou et al. 2010). As expected, the combination of stochastic volatility and jump risk factors in the SVJ model help to reduce BS-OV RMSE from 5.166 to 2.465, making SVJ the best-performing structural model in pricing the entire cross-section of options in the out-of-sample data set. It is notable that, in pricing this entire options cross-section out-of-sample - as in the case with DVF models - our results show that the smallest RMSE is obtained when each structural model is estimated using the joint data set of out-of-the-money call and put options.

Finally, we observe that SVJ does not perform better than the overall best-performing DVF model (σ^{s}_{LnX}) when the models are estimated with the joint data set of out-of-the-money call and put options and evaluated out-of-sample on the entire cross-section of option prices.[22] In that respect, from the standpoint of both the academic researcher and the risk-management practitioner, the performance of σ^{s}_{LnX} (when estimated to be consistent with put-call parity) is critical in gauging the size of pricing error that should be considered as 'large'.

[22] SV and SVJ sometimes result in implausible implied parameters to rationalize the observed option prices. Bakshi et al. (1997) report that both are clearly mis-specified (see also Bates 1996 and discussions in Skiadopoulos 2001). Moreover, while it can take only a few seconds to fit a regression-based DVF model to daily option prices, it sometimes can take up to a few minutes to do the same using the SV, and even longer for the SVJ model. For these reasons, traders in practice usually prefer the regression-based DVF approach and rely less on mathematically advanced continuous-time option pricing models such as SV or SVJ (see also discussions in Dumas et al. 1998; Brandt and Wu 2002; Chen et al. 2009; Kim, 2009).

5.3. Alternative definition of the estimation data set

In a further robustness test of the nature of the options data set, we check whether our results change with a different definition of the estimation data set. In that respect, we employed a fourth set comprised of the joint data of all available delta-based out-of-the-money call and out-of-the-money put options, where delta is defined according to Bollen and Whaley (2004).[23] In the spirit of Tables 2 and 5, all model specifications are re-estimated and re-evaluated using the delta-based out-of-the-money calls and puts. Overall, estimating the models using the delta-based out-of-the-money options definition of moneyness does not improve their in- or out-of-sample performance compared to estimating the models with the joint data set of out-of-the-money call and put options (using the ratio *K* to measure options moneyness). Consequently, all previous conclusions drawn are robust in terms of how we define the (estimation) data set of out-of-the-money options.

5.4. Stress-testing model out-of-sample performance

We stress tested the performance of the three overall best symmetric models (σ_X^s, σ_{LnX}^s, and σ_K^s), and the three overall best asymmetric models (σ_X^a, σ_{LnX}^a, and σ_K^a) - results for the other DVF models are available on request. The analysis concerns the out-of-sample pricing accuracy of the models on the entire cross-section of options estimated using the joint data set of out-of-the-money call and put options (using the ratio *K* to measure options moneyness). We employed three alternative stress tests to allow us to assess model performance under various market conditions. Table 6 presents the results. In Panel A, we test model resistance to large mispricing by only considering the tails of the daily pricing residual distribution. In particular, the RMSE values are computed by considering only the five largest squared pricing residuals per day. This test can reveal the models which are

---

[23] Out-of-the-money calls are those with $\Delta_c < 0.50$, where Δ_c is the delta value for a call option computed as: $\partial c / \partial S = e^{-d_y T}\,\mathrm{N}(d)$, while out-of-the-money puts are those with $\Delta_p \geq -0.50$, where Δ_p is the delta value for a put option computed as: $\partial p / \partial S = e^{-d_y T}\left(\mathrm{N}(d) - 1\right)$, whereas d is given by Eq. (1.3). Bollen and Whaley (2004) explain that it is preferable to split the data in delta moneyness categories, since delta is sensitive to the volatility of the underlying asset, as well as to the option's time-to-maturity. As in their case, the proxy for the volatility rate is the realized return volatility of the S&P 500 Index over the most recent 60 trading days.

mostly prone to large mispricing, which is significantly essential for risk-management purposes. In Panel B, we test the robustness of their pricing accuracy in rapidly-changing market conditions by taking into account only the days on which there is more than a 1% change (in absolute terms) in the level of the index compared to the preceding day. Days with large jumps induce (extra) skewness in asset returns and there is greater likelihood of observing asymmetric patterns of the implied volatility functions, because positive (negative) price jumps make out-of-the-money calls (puts) more expensive. Such cases are tough to handle, since implied volatility functions are expected to differ significantly compared to the preceding day. In the same way, in Panel C we stress test model performance by only computing RMSE values for days when VIX values are over 30% (so all options become more expensive).

In equal-terms comparisons, we find that symmetric models still perform better than their asymmetric counterparts. Moreover, there was considerable consensus among the test results supporting our contention that the σ_{LnX}^{s} model still appears to be the best-performing one.

[Table 6 here]

6. Conclusions

This paper seeks to identify the overall best approach to model implied volatility functions for pricing S&P 500 Index options in- and out-of-sample by examining the performance of several DVF models over the period 1996-2009. We estimate all DVF specifications, daily, and separately on three different data sets – specifically, the models are estimated either using all available calls, or all available puts or the joint data set of all out-of-the-money call and put options. In our analysis, we report the in- and out-of-sample pricing performance of each DVF model on different evaluation options data sets – specifically, the models are evaluated either using the entire cross-section of call options only, or the entire cross-section of put options only, or the entire cross-section of call and put options. By doing so, we identify which DVF specifications exhibit the overall best in- and out-of-sample pricing performance. The data suggest that DVF specifications that model implied volatility as a function of the moneyness ratio (relative smile approach) work better in-sample, while specifications that model implied volatility as a function of the logarithm of the strike price (absolute smile approach) work better out-of-sample. Moreover, symmetric DVF specifications perform better out-of-sample, although asymmetric models are superior in-sample. We provide strong evidence that DVF models estimated with NLS outperform those estimated with OLS. Finally, we find that the overall best strategy is to estimate the regression-based DVF models using the joint data set of all available out-of-the-money call and put options.

Our results indicate that σ_{LnX}^{s} , when estimated consistently with put-call parity, is the best model overall, since its performance is better than any other DVF model, and compares quite favorably with advanced structural continuous-time option pricing models.

Acknowledgements

This article has benefited from the useful comments of Hans Bystrom (discussant) and the participants of the 16th Annual Conference of the Multinational Finance Society (Crete, June 2009). Comments by Christodoulos Louca (Cyprus University of Technology), George Skiadopoulos (University of Piraeus and University of Warwick), Panayiotis Theodossiou (Cyprus University of Technology), and Christos Savva (Cyprus University of Technology) are also acknowledged. The authors are also grateful to Isabella Karasamani (Cyprus University of Technology and North Central University) for proofreading the paper and providing insightful and helpful comments.

References

Andreou PC, Charalambous C, Martzoukos SH (2008) Pricing and trading European options by combining artificial neural networks and parametric models with implied parameters. Eur J Op Res 185(3): 1415 – 1433

Andreou PC, Charalambous C, Martzoukos SH (2010) Generalized parameter functions for option pricing. J Bank Financ 34(3): 633-646

Ahoniemi K, Lanne M, (2009) Joint modeling of call and put implied volatility. Int J Forecast 25(2): 239–258

Bakshi G, Cao C, Chen Z (1997) Empirical performance of alternative options pricing models. J Financ 52(5): 2003-2049

Bakshi G, Kapadia N, Madan D (2003) Stock return characteristics, skew laws, and the differential pricing of individual equity options. Rev Financ Stud 16(1): 101-143

Bates DS (1991) The crash of ’87: was it expected? The evidence from options markets. J Financ 46(3): 1009–1044

Bates DS (1996) Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. Rev Financ Stud 9(1):69-107

Bates DS (2000) Post-'87 crash fears in the S&P 500 futures option market. J Econom 94(1-2): 181-238

Becker R, Clements AE (2008) Are combination forecasts of S&P 500 volatility statistically superior? Int J Forecast 24(1): 122–133

Berkowitz J (2010) On justifications for the ad hoc Black-Scholes method of option pricing. Stud Nonlinear Dyn Econom 14(1): Article 4

Black F, Scholes M (1973) The pricing of options and corporate liabilities. J Polit Econ 81(3): 637-654

Bollen NPB, Whaley RE (2004) Does net buying pressure affect the shape of implied volatility functions? J Financ 59(2): 711-753

Brandt MW, Wu T (2002) Cross-sectional tests of deterministic volatility functions. J Bank Financ 9(5): 525-550

Chen RR, Lee CF, Lee HH (2009) Empirical performance of the constant elasticity variance option pricing model. Rev Pac Basin Financ Mark Policies 12(2): 177-217

Chang CC, Lin JB, Tsai WC, Wang YH (2012) Using Richardson extrapolation techniques to price American options with alternative stochastic processes. Rev Quant Financ Account 39(3): 383-406

Chang EC, Ren J, Shi Q (2009) Effects of the volatility smile on exchange settlement practices: the Hong Kong case. J Bank Financ 33(1): 98-112

Chernov M, Ghysels E (2000) Towards a unified approach to the joint estimation objective and risk

neutral measures for the purpose of option valuation. J Financ Econ 56(3): 407-458
Christoffersen P, Heston S, Jacobs K (2006) Option valuation with conditional skewness. J Econom 131(1-2): 253-284
Christoffersen P, Heston SL, Jacobs K (2009) The shape and term structure of the index option smirk: why multifactor stochastic volatility models work so well. Manag Sci 55(12): 1914-1932
Christoffersen P, Jacobs K (2004) The importance of the loss function in option valuation. J Financ Econ 72(2): 291-318
Constantinides GM, Jackwerth JC, Perrakis S (2009). Mispricing of S&P 500 index options. Rev Financ Stud 22(3): 1247-1277
Dumas B, Fleming J, Whaley RE (1998) Implied volatility functions: empirical tests. J Financ 53(6): 2059-2106
Engström M (2002) Do Swedes smile? On implied volatility functions. J Multinatl Financ Manag 12(4-5): 285-304
Heston SL (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev Financ Stud 6(2): 327-343
Johnson NJ (1978) Modified *t* test and confidence intervals for asymmetrical populations. J Am Stat Assoc 73(363): 536-544
Jones EP (1984) Option arbitrage and strategy with large price changes. J Financ Econ 13(1): 91-113
Kim S (2009) The performance of traders' rules in option market. J Futures Mark 29(11): 999-1020
Kuo ID (2011) Pricing and hedging volatility smile under multifactor interest rate models. Rev Quant Financ Account 36(1): 83-104
Linaras CE, Skiadopoulos G (2005) Implied volatility trees and pricing performance: evidence from the S&P 100 options. Int J Theor Appl Financ 8(8): 1085–1106
Merton RC (1976) Option pricing when underlying stock return are discontinuous. J Financ Econ 3(1-2): 125–144
Mozumder S, Sorwar G, Dowd K (2012) Option pricing under non-normality: a comparative analysis. Rev Quant Financ Account (forthcoming)
Ncube M (1996) Modelling implied volatility with OLS and panel data models. J Bank Financ 20(1): 71-84
Nocedal J, Wright JS (1999) Numerical Optimization. Springer-Verlag, New York
Panigirtzoglou N, Skiadopoulos G (2004) A new approach to modeling the dynamics of implied distributions: theory and evidence from the S&P 500 options. J Bank Financ 28(7): 1499–1520
Patton AJ (2011) Volatility forecast comparison using imperfect volatility proxies. J Econom 160(1): 246-256
Peña I, Rubio G, Serna G (1999) Why do we smile? On the determinants of the implied volatility function. J Bank Financ 23(8): 1151-1179
Peña I, Rubio G, Serna G (2001) Smiles, bid-ask spreads and option pricing. Eur Financ Manag 7(3):

351-374

Shimko DC (1993) Bounds of probability. RISK 6(4): 33-37

Skiadopoulo G (2001) Volatility smile consistent option models: a survey. Int J Theor Appl Financ 4(3): 403-437

Tompkins R (2001) Stock index futures markets: stochastic volatility models and smiles. J Futures Mark 21(1): 43-78

Vähämaa S, Äijö J (2011) The Fed's policy decisions and implied volatility. J Futures Mark 31(10): 995-1010

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48(4): 817-838

Tables

Moneyness class	DOTM	OTM	JOTM	ATM	JITM	ITM	DITM
Moneyness range	**[0.75, 0.90)**	**[0.90, 0.95)**	**[0.95, 0.99)**	**[0.99, 1.01)**	**[1.01, 1.05)**	**[1.05, 1.10)**	**[1.10, 1.25]**
				Short Term Options: < 60 calendar days			
Implied volatility	0.31 (0.38)	0.21 (0.26)	0.18 (0.19)	0.18 (0.19)	0.21 (0.21)	0.26 (0.25)	0.33 (0.32)
Spread	1.32 (3.45)	1.04 (2.61)	1.17 (2.04)	1.62 (1.58)	1.97 (1.21)	2.13 (0.98)	2.25 (0.91)
Volume	1,115 (303)	1,261 (338)	1,836 (821)	2,348 (2,730)	704 (2,369)	353 (1,971)	238 (1,637)
Obs.	5,680 (2,942)	17,593 (7,709)	36,131 (25,214)	20,962 (21,413)	25,319 (35,782)	12,353 (28,231)	6,336 (23,640)
				Medium Term Options: 60-180 calendar days			
Implied volatility	0.21 (0.27)	0.18 (0.22)	0.18 (0.20)	0.20 (0.20)	0.21 (0.21)	0.23 (0.23)	0.27 (0.26)
Spread	1.34 (3.35)	1.49 (2.52)	1.83 (2.13)	2.04 (1.96)	2.14 (1.79)	2.19 (1.57)	2.27 (1.29)
Volume	553 (252)	716 (292)	894 (732)	1,702 (1,822)	661 (1,248)	261 (1,091)	185 (877)
Obs.	11,623 (2,323)	16,782 (5,220)	17,478 (12,425)	9,872 (11,395)	9,038 (17,454)	4,794 (16,666)	3,451 (20,767)
				Long Term Options: ≥ 180 calendar days			
Implied volatility	0.18 (0.23)	0.18 (0.21)	0.19 (0.20)	0.20 (0.20)	0.21 (0.21)	0.22 (0.22)	0.24 (0.24)
Spread	1.76 (2.85)	2.12 (2.35)	2.26 (2.21)	2.30 (2.16)	2.33 (2.09)	2.32 (2.03)	2.31 (1.90)
Volume	447 (243)	514 (202)	491 (446)	660 (815)	397 (708)	270 (590)	149 (580)
Obs.	8,984 (1,534)	7,753 (3,250)	6,418 (5,628)	3,757 (4,546)	3,455 (7,837)	1,911 (7,621)	1,525 (9,818)

Table 1: Sample descriptive statistics per moneyness and time-to-maturity.
Numbers in parentheses refer to put options while numbers above refer to call options. In each block of numbers, the first row reports daily average values of Black and Scholes (1973) implied volatilities, the second row reports the daily average values of options bid-ask spread, the third row reports the daily average option volume, while the last row reports the number of observations per moneyness/time-to-maturity. Descriptive statistics are tabulated based on the following moneyness cases ($K = S/X$, stock index value, S, over the option's strike price, X): deep out-of-the-money (DOTM) for $0.75 \leq K < 0.90$, out-of-the-money (OTM) for $0.90 \leq K < 0.95$, just out-of-the-money (JOTM) for $0.95 \leq K < 0.99$, at-the-money (ATM) for $0.99 \leq K < 1.01$, just in-the-money (JITM) for $1.01 \leq K < 1.05$, in-the-money (ITM) for $1.05 \leq K < 1.10$, and deep in-the-money (DITM) for $1.10 \leq K \leq 1.25$.

DVF Models	σ^s_X	σ^s_{LnX}	σ^s_K	σ^s_{LnK}	σ^a_X	σ^a_{LnX}	σ^a_K	σ^a_{LnK}
	Symmetric DVF Models				**Asymmetric DVF Models**			
Panel A. Model performance with call options								
Data used for model estimation: Call options								
IN	0.815	0.793	0.782	0.793	0.806	0.783	0.772	0.779
OUT	2.082	2.066	2.625	2.629	2.098	2.082	2.634	2.643
Data used for model estimation: Put options								
IN	3.020	2.833	2.783	2.833	3.153	2.866	2.807	2.853
OUT	3.068	2.856	3.176	3.230	3.193	2.899	3.200	3.247
Data used for model estimation: Joint dataset of out-of-the-money call and put options								
IN	1.445	1.462	1.465	1.462	1.567	1.563	1.558	1.462
OUT	1.940	1.930	2.510	2.516	2.066	2.046	2.583	2.520
Panel B. Model performance with put options								
Data used for model estimation: Call options								
IN	2.842	2.857	2.897	2.857	2.860	2.869	2.911	2.918
OUT	2.805	2.823	3.254	3.219	2.833	2.840	3.268	3.270
Data used for model estimation: Put options								
IN	0.851	0.831	0.832	0.831	0.843	0.821	0.821	0.817
OUT	1.885	1.854	2.330	2.336	1.897	1.863	2.330	2.339
Data used for model estimation: Joint dataset of out-of-the-money call and put options								
IN	1.688	1.650	1.641	1.650	1.811	1.801	1.798	1.656
OUT	1.936	1.915	2.427	2.431	2.051	2.045	2.521	2.434
Panel C. Model performance with entire cross-section of options (calls and puts together)								
Data used for model estimation: Call options								
IN	2.160	2.167	2.194	2.167	2.172	2.174	2.202	2.208
OUT	2.498	2.504	2.981	2.962	2.522	2.520	2.993	2.998
Data used for model estimation: Put options								
IN	2.142	2.016	1.984	2.016	2.226	2.035	1.997	2.026
OUT	2.500	2.368	2.752	2.783	2.576	2.396	2.765	2.794
Data used for model estimation: Joint dataset of out-of-the-money call and put options								
IN	1.581	1.567	1.562	1.567	1.703	1.696	1.692	1.569
OUT	1.938	1.922	2.465	2.470	2.058	2.046	2.550	2.474

Table 2: In-sample and one-day ahead out-of-sample pricing performance of DVF models.
In- and out-of-sample RMSE pricing performance of the symmetric and asymmetric DVF models estimated daily with nonlinear least squares during the period from January 1996 to October 2009. Panel A reports in-sample (IN) and one-day ahead out-of-sample (OUT) pricing performance when the models are evaluated with the daily cross-section of call options. Panel B reports pricing performance when the models are evaluated with the daily cross-section of puts options, while Panel C reports pricing performance when the models are evaluated with the entire cross-section of options (call and put options together). The upper part of each panel reports pricing performance when the models are estimated with call options, the middle part reports pricing performance when the models are estimated with put options, while the lower part of each panel reports pricing performance when the models are estimated with the joint dataset of out-of-the-money call and put options.

DVF Models			σ^{s}_{LnX}	σ^{s}_{LnX}	σ^{a}_{LnX}	σ^{a}_{LnX}	σ^{a}_{LnX}
	Dataset used for estimation ↓	→	Puts	Out-of-the-money calls and puts	Calls	Puts	Out-of-t money c and pu
Panel A: Options dataset used for one-day ahead out-of-sample pricing evaluation: Call options							
		RMSE	2.856	1.930	2.082	2.899	2.046
σ^{s}_{LnX}	Calls	2.066	-31.177	8.809	-15.576	-32.343	1.281
σ^{s}_{LnX}	Puts	2.856		40.179	30.612	-16.350	35.536
σ^{s}_{LnX}	Out-of-the-money calls and puts	1.930			-9.854	-40.534	-29.351
σ^{a}_{LnX}	Calls	2.082				-31.795	2.327
σ^{a}_{LnX}	Puts	2.899					36.112
Panel B: Options dataset used for one-day ahead out-of-sample pricing evaluation: Put options							
		RMSE	1.854	1.915	2.840	1.863	2.045
σ^{s}_{LnX}	Calls	2.823	55.218	62.277	-13.356	55.114	54.840
σ^{s}_{LnX}	Puts	1.854		-5.476	-56.003	-5.772	-16.625
σ^{s}_{LnX}	Out-of-the-money calls and puts	1.915			-62.962	4.743	-35.289
σ^{a}_{LnX}	Calls	2.840				55.913	55.712
σ^{a}_{LnX}	Puts	1.863					-16.114
Panel C: Options dataset used for one-day ahead out-of-sample pricing evaluation: Entire cross-section of options (calls and puts together)							
		RMSE	2.368	1.922	2.520	2.396	2.046
σ^{s}_{LnX}	Calls	2.504	9.072	54.821	-18.961	7.158	43.882
σ^{s}_{LnX}	Puts	2.368		34.340	-10.177	-17.271	24.989
σ^{s}_{LnX}	Out-of-the-money calls and puts	1.922			-56.099	-35.409	-45.846
σ^{a}_{LnX}	Calls	2.520				8.255	45.310
σ^{a}_{LnX}	Puts	2.396					26.392

Table 3: Pricing performance and *t*-statistics for the one-day ahead out-of-sample pricing difference of the best performing models.

The table reports *t*-statistics that result from the comparison of the means of squared pricing residuals between models in the vertical heading versus models in the horizontal heading. All *t*-statistics regard the one-day ahead out-of-sample pricing errors of the models and are adjusted for heteroskedasticity using the White (1980) robust standard errors. Each model is estimated daily using three different datasets: *i)* call options only, *ii)* put options only, and, *iii)* the joint dataset of out-of-the-money call and put options. The data set used to estimate each model is designated below and on the right of the column entitled as: “Dataset used for estimation”. Panel A shows *t*-statistics that result by comparing model performance against call options, Panel B shows *t*-statistics by comparing model performance against put options, while Panel C shows *t*-statistics by comparing models performance against the entire cross-section of options (calls and puts together). For convenience, the models’ RMSE performance is also reported in each case (taken accordingly from Table 2).

Moneyness class	DOTM	OTM	JOTM	ATM	JITM	ITM	DITM	Maturity Classes (aggregating moneyness)
Moneyness range	**[0.75, 0.90)**	**[0.90, 0.95)**	**[0.95, 0.99)**	**[0.99, 1.01)**	**[1.01, 1.05)**	**[1.05, 1.10)**	**[1.10, 1.25]**	
				Short term options: ≤ 60 days				
C→C	1.544	1.427	1.613	1.909	2.025	2.199	2.217	1.826
P→C	2.322	2.177	2.118	2.131	1.987	1.997	2.085	2.099
CP→C	1.613	1.454	1.566	1.721	1.776	1.969	2.135	1.700
P→P	3.180	2.865	2.295	2.059	1.911	1.825	1.702	2.051
C→P	3.099	2.789	2.324	1.885	1.422	1.178	1.027	1.736
CP→P	3.212	2.806	2.089	1.667	1.325	1.213	1.106	1.648
				Medium term options: 60-180 days				
C→C	1.397	1.552	1.875	2.083	2.300	2.657	3.031	1.960
P→C	2.798	2.744	2.625	2.454	2.409	2.471	2.579	2.622
CP→C	1.658	1.659	1.730	1.814	2.017	2.433	2.832	1.871
P→P	4.044	3.243	2.680	2.378	2.488	2.537	2.778	2.685
C→P	3.327	3.020	2.364	1.835	1.552	1.231	1.097	1.780
CP→P	4.301	3.213	2.210	1.646	1.465	1.283	1.347	1.837
				Long term options: ≥ 180 days				
C→C	2.839	2.830	2.673	2.888	2.966	3.373	4.263	2.937
P→C	6.393	4.906	4.324	3.798	3.975	3.874	4.007	4.972
CP→C	2.683	2.272	2.508	2.431	2.916	3.414	4.225	2.698
P→P	6.421	4.894	4.510	4.637	4.453	4.817	5.008	4.812
C→P	4.291	3.652	2.550	2.242	1.888	1.793	1.990	2.351
CP→P	6.524	4.246	3.185	2.689	2.045	1.838	1.973	2.778
				Moneyness classes (aggregating maturity)				
C→C	2.033	1.812	1.831	2.086	2.194	2.460	2.830	
P→C	4.312	3.067	2.590	2.457	2.340	2.372	2.574	
CP→C	2.059	1.711	1.738	1.837	1.965	2.274	2.721	
P→P	4.389	3.481	2.789	2.603	2.540	2.683	2.959	
C→P	3.477	3.054	2.366	1.917	1.526	1.302	1.279	
CP→P	4.515	3.272	2.295	1.817	1.476	1.343	1.390	

Table 4: Moneyness and time-to-maturity tabulation of the one-day ahead out-of sample pricing performance of the DVF model: σ^{s}_{LnX} .

One-day ahead out-of-sample RMSE pricing performance of the DVF model: $\sigma = \max\left(0.01,\ a_0 + a_1 LnX + a_2 (LnX)^2 + a_3 T + a_4 (LnX) T + a_5 T^2\right)$. Estimation is done daily using nonlinear least squares for the period from January 1996 to October 2009. The bottom panel reports RMSE per moneyness class (aggregating time-to-maturity) while the last column on the right reports RMSE per time-to-maturity (aggregating moneyness). Six different estimation approaches are tabulated in each panel. C→C indicates model estimation with call options and out-of-sample evaluation on call options; P→C indicates model estimation with put options and out-of-sample evaluation on call options; CP→C indicates model estimation with the joint data set of out-of-the-money call and put options and out-of-sample evaluation on call options; P→P indicates model estimation with put options and out-of-sample evaluation on put options; C→P indicates model estimation with call options and out-of-sample evaluation on put options; CP→P indicates model estimation with the joint data set of out-of-the-money call and put options and out-of-sample evaluation on put options.

	BS-OV	BS-VIX	1Jump	2Jump	SV	SVJ
Panel A. Model performance with call options						
Data used for model estimation: Call options						
IN	4.285	---	2.654	2.596	0.881	0.595
OUT	4.736	9.588	3.426	3.387	3.810	2.519
Data used for model estimation: Put options						
IN	6.291	---	4.303	4.335	3.793	3.289
OUT	6.397	9.588	4.494	4.530	4.364	3.556
Data used for model estimation: Joint dataset of out-of-the-money call and put options						
IN	4.832	---	3.090	3.051	1.552	1.395
OUT	5.128	9.588	3.622	3.591	3.477	2.544
Panel B. Model performance with put options						
Data used for model estimation: Call options						
IN	6.136	---	3.871	3.795	3.497	2.670
OUT	6.238	7.411	4.067	4.000	3.930	3.038
Data used for model estimation: Put options						
IN	4.614	---	2.622	2.556	1.591	1.350
OUT	4.910	7.411	3.213	3.174	3.292	2.530
Data used for model estimation: Joint dataset of out-of-the-money call and put options						
IN	4.995	---	2.986	2.923	1.820	1.599
OUT	5.199	7.411	3.366	3.317	2.911	2.396
Panel C. Model performance with entire cross-section of options (calls and puts together)						
Data used for model estimation: Call options						
IN	5.364	---	3.366	3.298	2.638	2.003
OUT	5.597	8.482	3.786	3.730	3.875	2.811
Data used for model estimation: Put options						
IN	5.450	---	3.497	3.489	2.826	2.441
OUT	5.643	8.482	3.856	3.857	3.823	3.045
Data used for model estimation: Joint dataset of out-of-the-money call and put options						
IN	4.920	---	3.034	2.983	1.702	1.508
OUT	5.166	8.482	3.486	3.446	3.184	2.465

Table 5: In-sample and one-day ahead out-of-sample pricing performance of structural option pricing models.

The structure presentation of this table is similar to that of Table 2. The models considered in this table are as follows: BS-OV is the Black and Scholes (1973) model performance employed with a daily overall-average implied volatility, BS-VIX is the Black and Scholes (1973) model performance when employed daily with the one-day lagged closing value of the VIX index, 1Jump is the single-jump model of Bates (1991), 2Jump is a two-jump variant of the Merton (1976) model as discussed in Jones (1984), SV is the stochastic volatility of Heston (1993) and SVJ is the stochastic volatility and jump model of Bates (1996).

DVF Models	Dataset used for estimation ↓ →		σ^{s}_{LnX}	σ^{s}_{K}	σ^{a}_{X}	σ^{a}_{LnX}	σ^{a}_{K}
			Out-of-the-money call and put options				
Panel A: One-day ahead out-of-sample pricing tests on the cross-section of call and put options: 5-largest pricing errors							
		RMSE	5.001	5.514	5.408	5.317	5.854
σ^{s}_{X}	Out-of-the-money call and put options	5.086	5.845	-15.598	-20.417	-13.789	-26.049
σ^{s}_{LnX}		5.001		-27.539	-17.001	-20.533	-34.808
σ^{s}_{K}		5.514			3.234	8.641	-23.324
σ^{a}_{X}		5.408				5.968	-15.168
σ^{a}_{LnX}		5.317					-27.524
Panel B: One-day ahead out-of-sample pricing tests on the cross-section of call and put options: Index jumps over |1%|							
		RMSE	2.384	3.399	2.524	2.515	3.495
σ^{s}_{X}	Out-of-the-money call and put options	2.388	2.529	-83.571	-19.160	-17.581	-76.736
σ^{s}_{LnX}		2.384		-85.909	-19.811	-18.799	-78.009
σ^{s}_{K}		3.399			68.545	69.579	-14.604
σ^{a}_{X}		2.524					-77.550
σ^{a}_{LnX}		2.515					-78.795
Panel C: One-day ahead out-of-sample pricing tests on the cross-section of call and put options: High VIX (over 30%)							
		RMSE	2.851	3.993	2.998	2.995	4.081
σ^{s}_{X}	Out-of-the-money call and put options	2.865	10.629	-85.199	-22.913	-22.126	-81.280
σ^{s}_{LnX}		2.851		-86.585	-23.921	-23.716	-81.911
σ^{s}_{K}		3.993			73.160	73.675	-14.977
σ^{a}_{X}		2.998				3.195	-79.852
σ^{a}_{LnX}		2.995					-80.562

Table 6: Pricing performance and *t*-statistics for the one-day ahead out-of-sample pricing difference of the best performing models under different market conditions.

The structure presentation of this table is similar to that of Table 3. All DVF models are estimated with nonlinear least squares using daily the joint data set of out-of-the-money call and put options during the period from January 1996 to October 2009. One-day ahead out-of-sample evaluation is always done with the entire cross-section of options (calls and puts together). Panel A reports pricing performance with *t*-statistics for the best performing models when RMSE is computed by considering daily the 5-largest squared pricing residuals. Panel B reports pricing statistics for the best performing models only for those days with a price change in the S&P 500 value over 1% (in absolute terms) compared to the day before. Panel C reports pricing statistics by considering only those days with a VIX value greater than 30%.