Robust Estimation of Shape Constrained State Price Density Surfaces*

Markus Ludwig§

First Version: August 30, 2012 This Version: March 25, 2013

Abstract

In order to better capture empirical phenomena, research on option price and implied volatility modeling increasingly advocates the use of nonparametric methods over simple functional forms. This, however, comes at a price, since they require dense observations to yield sensible results. Calibration is therefore typically performed using aggregate data. Ironically, the use of time-series data in turn limits the accuracy with which current observations can be modeled. We propose a novel approach that enables the use of flexible functional forms using daily data alone. The resulting estimators yield excellent fits and generalize well beyond available data, all the while respecting theory imposed shape constraints. We demonstrate the numerical stability and the pricing performance of our method by approximating arbitrage-free implied volatility, price and state price density surfaces from S&P 500 options over a period of 12 years.

Keywords: implied volatility, state price density, shape constraints, neural networks

JEL classification: C14, C58, G13

Research on derivatives traditionally revolves around developing and refining methods for pricing and hedging. However, recently there has been increased interest in approaches that take market prices as given, and focus on the information content priced in by means of supply and demand, cf. Xing et al. (2010), Conrad et al. (2013) and Ross (2013). A popular method to analyze the sentiment of market participants is to use the seminal option pricing formula proposed by Black and Scholes (1973) and Merton (1973) to map observations from the space of prices to the space of implied volatilities. Implied volatilities correspond to the volatility inputs that equate model and market prices, and contrary to the assumptions underlying the model used to derive them, vary both with respect to strike prices and times to maturity. The well known smile and term structure patterns highlight that the market does not consider asset prices to be log-normally distributed.

Another approach builds on the fact that, analogous to the distinction between implied and historic volatility, there also exists an implied density that inherently differs from the historic density of the underlying price process. Breeden and Litzenberger (1978) show that the state price density can be obtained by differentiating the option price function twice with respect to strike. This density reflects the expectations of market participants regarding the evolution of the underlying asset as well as their risk preferences and - unlike implied volatility - is model-free.

Approaches to approximate option price functions can be distinguished into parametric and nonparametric. Parametric methods are structured techniques and rely on specific assumptions about the process generating the observable data. While stringent assumptions allow for easier calibration and facilitate both extrapolation and the incorporation of shape constraints, they pose the risk of misspecification, i.e., the estimator may fail to capture salient properties of the data. Nonparametric methods, in contrast, are data-driven and do not rely on strong assumptions about the underlying process. Their disadvantage is that they are typically not very effective on small samples and beyond the support of given data.

Recovering well-behaved state price densities from a set of option prices is a non-trivial exercise and poses several challenges: (i) smooth interpolation despite noisy and sparse data, (ii) extension of the density into the tails, i.e., beyond the range of observations, (iii) compliance with shape constraints following from no-arbitrage arguments, and (iv) accurate fits, both in-sample and out-of-sample.

*I would like to gratefully acknowledge the valuable comments by Markus Leippold, Francesco Audrino, Stephen Figlewski, Jens Carsten Jackwerth, Manfred Gilli and Robert Huijema.

§Department of Banking and Finance, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland
These issues are further exacerbated by the fact that the state price density corresponds to the second derivative of the function giving rise to observable prices. The quality of a derivative is typically significantly worse than that of its corresponding primitive, since differentiation amplifies even minor irregularities in the estimated function. This issue is usually referred to as the curse of differentiation, cf. Aït-Sahalia and Lo (1998).

This paper proposes a novel approach that produces robust and adaptive implied volatility, price and state price density surfaces that satisfy theory imposed shape constraints, such that both strike and calendar arbitrage are ruled out. Potential applications range from pricing of non-traded or illiquid products to risk management, asset allocation and market timing, see e.g., Rosenberg (1998), Aït-Sahalia and Lo (2000), Kostakis et al. (2011). Option implied densities also enable the analysis of how market participants respond to new information, and how risk perceptions evolve, see e.g., Birru and Figlewski (2012).

Our work contributes to the existing literature along several dimensions. We propose a flexible approach that allows the incorporation of shape constraints in the model selection process of neural network modeling, show that excellent generalization can be achieved without resorting to time-series data, and obtain perfectly smooth surfaces over a fixed domain as opposed to maturity-wise strings.

The stable observational grid provided by robust and adaptive surfaces over a fixed domain in turn enables the analysis of how option implied measures evolve over time.

The remainder of this paper is organized as follows. Section 1 provides an introduction to state price densities and discusses some key challenges. Section 2 introduces neural networks as a tool for function approximation and presents the main ideas behind our shape constrained networks. Section 3 presents the results of an empirical study based on S&P 500 options data and examines them in the context of existing literature. Section 4 concludes.

1 Option Prices and Implied Densities

Options are derivative instruments whose payoff depends on future states of an underlying asset. Option markets serve to integrate the expectations of market participants, as well as their perceptions regarding risk and ambiguity.

A widely used measure of market sentiment, calculated from observable prices of plain vanilla options, is implied volatility (IV). A related concept is the state price density (SPD), which captures the risk neutral probabilities that the market assigns to the various possible states of the underlying asset upon expiry of the option.

Ross (1976) and Cox and Ross (1976) demonstrate that in a dynamically complete arbitrage-free market, the price of an option is given by the expected present value of the payoff computed under the SPD. Building on these results Banz and Miller (1978) and Breeden and Litzenberger (1978) show that an explicit expression for the SPD can be obtained as the second derivative of the option price function with respect to strike prices. Harrison and Kreps (1979) prove the existence of a probability density for the underlying process such that the value of a call option $C(S_t, K, \tau, r, \delta)$ is given by

$$C(\cdot) = e^{-r\tau} \int_0^{+\infty} \max(S_T - K, 0)q(S_T)dS_T, \quad (1)$$

where $S_t$ is the price of the asset at date $t$, $K$ the strike price, $q$ the state price density, $\tau$ the time to maturity, $T = t + \tau$ the expiration date, $r$ the deterministic risk-free interest rate for that maturity, and $\delta$ the corresponding dividend yield of the asset. Following from (1) the state price density can be expressed as

$$q(S_T) = e^{-r\tau} \frac{\partial^2 C(\cdot)}{\partial K^2}. \quad (2)$$

As noted by Cont (1997), the SPD should be viewed as a way of characterizing the prices of options on an asset as opposed to a mathematical property of the underlying asset’s stochastic process.

Related Work

Extracting well-behaved densities from a discrete set of option prices is a non-trivial exercise and poses multiple challenges, as evidenced by the considerable quantity of alternative approaches put forward in the literature. The field subsumes both research on implied volatility and option price modeling, but has additional and more stringent requirements such as smoothness of the second derivative, support beyond available data and absence of static arbitrage.

Existing methods can be distinguished by whether they model the SPD directly, or estimate the implied volatility or price function first, and then derive the corresponding density via the result of Breeden and Litzenberger (1978).


Another way to categorize the literature is according to whether the techniques are parametric or nonparametric. Parametric approaches aim to provide tractable models using only a parsimonious number of parameters. While functional forms facilitate the estimation procedure and allow for easy extrapolation, they typically rely on strong assumptions regarding the data generating process and often lack the flexibility to fit the observable data.

Nonparametric methods, such as kernel smoothing and maximum entropy, however, are data-driven and thus less restricted. Their main drawback pertains to the risk of overfitting the observable data, instead of capturing the salient features of the underlying functional relationship. Overfitting in price or implied volatility space will lead to sharp spikes in the corresponding state price density. This risk is further amplified on small and sparse data sets. For a lucid discussion of small sample, as opposed to asymptotic, properties in the context of shape constraints, see e.g., Aït-Sahalia and Duarte (2003).

Detailed reviews of the literature can be found in Bliss and Panigirtzoglou (2002), Jackwerth (2004) and Figlewski (2010), who concludes that none of the existing techniques is clearly superior. Expansion methods may give rise to negative tails, mixtures of lognormals tend to be unstable and exhibit tails that are too thin, maximum entropy distributions can be multimodal, while kernel methods suffer from slow convergence, cf. Cont (1997). By contrast, curve fitting methods, especially in implied volatility space, have been shown to yield stable densities that exhibit good fits, cf. Bliss and Panigirtzoglou (2002).

Fitting the implied volatility smile with a quadratic polynomial was originally proposed by Shimko (1993) and has gained wide acceptance among practitioners. In the context of option pricing this approach is also known as practitioners or ad hoc Black-Scholes.

Modeling Challenges

One of the key challenges in modeling option data arises from the highly irregular data design. At any given time, there are only a limited number of maturities with a discrete set of strikes available. Options appear in strings that are not evenly distributed and advance along the maturity dimension over time. The data is furthermore noisy which poses a significant challenge in the context of estimating SPDs via curve fitting, since derivatives exacerbate noise and irregularities, cf. Rebonato (2004).

For small samples, asymptotic properties offer little to no guidance about the actual performance of an estimator cf. Pritsker (1998). In order to mitigate the problems arising from finite observations, nonparametric methods often resort to aggregating data over time. For example, the kernel smoothing estimator of the call price surface in Aït-Sahalia and Lo (1998) is based on one year of options. Fan and Mancini (2009) aggregate data over one week to fit a two-dimensional local linear estimator. While data aggregation alleviates the problems related to small samples, it opens the door to nonstationarity and regime shift issues, cf. Aït-Sahalia and Duarte (2003). Furthermore, a surface estimator based on time-series data will not result in a real average, but an amalgamate that captures daily fluctuations along its term structure.

Kernel smoothing methods are also highly sensitive to the chosen bandwidth parameter and furthermore exhibit severe biases near the support boundary of observations, as well as in interior regions if data spacing is irregular. While local linear methods are more robust, they tend to be biased in regions of curvature, a phenomenon known as trimming the hills and fitting the valleys. Local quadratic methods, in turn, are generally able to yield better fits, but re-introduce the erratic behavior near the boundaries, cf. Hastie et al. (2009). This sensitivity to small perturbations in the data also affects other flexible techniques based on splines and high-degree polynomials.

Sensible extrapolation beyond available data poses a particular challenge, since the range of observable strikes is typically not sufficient to recover the tails of the density. While numerous authors, such as Aït-Sahalia and Duarte (2003) and Fan and Mancini (2009) neglect to address this issue, others such as Shimko (1993) and Bliss and Panigirtzoglou (2004) assume implied volatility to remain constant outside the range of observable strikes. This is equivalent to pasting normal tails onto the implied density and is not only questionable since asset returns exhibit fat tails, but also because it gives rise to individual strings that imply globally inconsistent shapes. Figlewski (2010) proposes a combination of a fourth-degree spline interpolation and tails from generalized extreme value distributions, but only yields two-dimensional estimators.

The vast majority of the literature on SPD extraction is confined to single option series. Due to the special data design of options moving towards expiration, the use of such estimators leads to either maturity jumping, or the construction of a new estimator by means of interpolating between neighboring strings. While the former precludes an analysis of the SPD at fixed maturities, the latter is prone to be afflicted by calendar arbitrage. As we will show, the surfaces we obtain, do not only provide a stable observational grid, but are also considerably more robust.
Another critical aspect is smoothness, both in the original estimator and in the resulting SPD. Rebonato (2004) notes that a smooth price density is important because there is a link between the unconditional (marginal) price densities obtained from the quoted prices today, and the conditional densities that will prevail in the future. The smoothness of the estimator also determines the accuracy with which it can fit the data. The classical bias-variance trade-off relates to the problem that, while more complex models provide better fits to available observations, they are prone to generalize poorly to previously unseen data.

A final issue that ties back to all the other challenges is arbitrage. Ait-Sahalia and Duarte (2003) were the first to consider shape constrained SPD estimation. Arbitrage constraints for entire surfaces are discussed in Carr and Madan (2005) and Fengler (2012). Roper (2010) remarks that a call price surface is free of static arbitrage if there can be no arbitrage opportunities trading in the surface. The following conditions must hold to guarantee that a surface is free of static arbitrage.

General price bounds:

\[ Se^{-\delta \tau} \geq C \geq \max \{0, Se^{-\delta \tau} - Ke^{-\tau r}\}, \quad (3) \]

Constraints on strike and butterfly spreads:

\[-e^{-\tau r} \leq \frac{\partial C}{\partial K} \leq 0 \quad \text{and} \quad \frac{\partial^2 C}{\partial K^2} \geq 0. \quad (4)\]

To avoid calendar arbitrage, implied total variance must be non-decreasing in forward-moneyness \( m \equiv K/F_T \).

Defining total variance \( \nu^2(m, T) \equiv \sigma^2(m, T)T \), we have

\[ \nu^2(m, T_2) > \nu^2(m, T_1) \quad \text{given} \quad T_2 > T_1. \quad (5) \]

Arbitrage poses an issue, since (i) the data we observe is noisy and often contains recording errors, (ii) methods that lack smoothness are prone to yield negative regions in the corresponding SPD (iii) extrapolating the slope at \( K_{\min} \) and \( K_{\max} \) may lead to negative tails, and (iv) two-dimensional fits might suffer from calendar spread arbitrage between maturities.

Against the backdrop of these challenges we developed a novel approach that we present in the following section.

## 2 Neural Networks

Our approach revolves around the Darwinian principle of random variation and natural selection. In contrast to the goals of classical optimization, we actively encourage a large variety of different solutions, as this allows us to check for properties beyond deviations from given prices.

The method builds on a specific class of neural networks, namely multilayer perceptrons, which perform function approximation via superpositions of sigmoid functions. Given observable data pairs \( \{x_1, y_1\}, \ldots, \{x_N, y_N\} \) with \( x_i \in \mathbb{R}^p \) and \( y_i \in \mathbb{R} \), where \( N \) denotes the number of observations and \( p \) indexes the dimension of the input space, function approximation aims to identify a mapping \( f(x) \) via a model such as \( y_i = f(x_i) + \epsilon_i \), where the error \( \epsilon \) is assumed to be iid noise. Neural networks achieve their flexibility through the layered use of primitive functions, each performing a nonlinear transformation of linearly combined inputs. Given a set of parameters \( \theta = \{\beta, w\} \) and omitting the intercepts \( \beta_0 \) and \( w_{0j} \), a network with one nonlinear layer can be written as

\[ f_\theta(x) = \sum_{j=1}^{M} \beta_j \cdot h(w^T_j x) + \epsilon, \quad (6) \]

where \( M \) denotes the number of nonlinear expansions, \( \beta \) corresponds to the coefficients in a linear model, and \( h(z) = 1/(1 + e^{-z}) \) specifies a family of log-sigmoid basis functions parametrized by \( w \). Since the parameters of the basis functions are learned from the data, such a network can be thought of as an adaptive basis function method.

Multilayer perceptrons have been shown to be universal approximators, i.e., given a sufficient number of hidden nodes \( M \), they can approximate any continuous function on a compact input domain up to an arbitrary degree of accuracy, cf. Hornik et al. (1989).

The parameters are typically estimated by minimizing the residual sum-of-squares

\[ \text{RSS}(\theta) = \sum_{i=1}^{N} (y_i - f_\theta(x_i))^2. \quad (7) \]

Since the basis functions have hidden parameters \( w \), the optimization has no closed-form solution and needs to be solved by means of iterative numerical methods, typically gradient descent. Probably the most common algorithm to minimize (7) is Levenberg-Marquardt, which combines gradient descent with a Gauss-Newton algorithm in the vicinity of a minimum. It achieves very fast convergence, since it does not require the computation of second order derivatives, cf. Hagan et al. (1996). However, like other gradient-based methods it can’t guarantee convergence to a global optimum. Furthermore, in neural networks, permutations of the parameter values can yield the same functional input output mapping, cf. Chen et al. (1993).

Due to these symmetries in the loss function, the number of local minima is high. This causes solutions to be very sensitive to the initial starting values of the optimization.
**Model Complexity and Generalization**

At first glance, the existence of local minima may seem as a serious drawback. However, as for all flexible nonlinear methods, the question of interest is not to which degree of accuracy we can match the training data, but whether or not the resulting estimator is predictive for novel data.

The practical usefulness of an estimator depends on its ability to successfully generalize beyond observable data. The real challenge is thus to find a model that is flexible enough to capture the relationships implicit in a set of finite and noisy observations, without memorizing them. In statistical learning theory this problem is known as the bias-variance trade-off. Overly simplistic models, while robust to variations in the data, typically exhibit a lack of fit due to a high bias. Overly complex models on the other hand will start to fit the idiosyncratic noise in the data. This variance phenomenon is known as overfitting.

The optimal degrees of freedom can however not solely be determined from the error on the training data, since it generally decays with model complexity. One method to estimate generalization performance is to partition the available data into a training and a validation set. The error on the validation data typically only decays up to a point with increasing complexity, and then rises again.

In neural networks, model complexity can be controlled both through the choice of $M$ and through regularization. Regularization modifies the loss function such that large weights $w$ are being penalized. If they are close to zero, the operative part of the sigmoid is roughly linear, and the network collapses into an approximately linear model.

A similar effect can be achieved through early stopping. Since the parameters are typically initialized at random starting values near zero, the model starts out nearly linear, cf. Nguyen and Widrow (1990). During training the weights are updated to introduce nonlinearities where needed. Stopping the optimization after a few iterations keeps them close to their highly regularized initial values.

This in turn raises the issue of how to determine the optimal penalty function, or number of training epochs. Since regularization methods express our belief that the function we are looking for exhibits some kind of smooth behavior, they can be cast in a Bayesian framework. MacKay (1992a,b) presents a method that determines the optimal penalty parameters for neural network training in an automated fashion based on Bayes’ theorem. One of the biggest advantages of this approach is that it needs no validation set, i.e., all the available training data can be used for parameter estimation and model comparison.

Despite their parametric functional form, the effective complexity of neural networks is thus data-driven. They can be viewed as parametric models with nonparametric interpretation. While they provide the same flexibility as local methods they have superior small sample properties and provide infinitely differentiable closed-form solutions.

**Random Variation and Model Selection**

Given the bias-variance trade-off, selecting the optimal model complexity is an issue of prime importance, be it the degree of a polynomial estimator, the bandwidth parameter for a local method or the number of knotpoints for splines. Going back to Malliaris and Salchenberger (1993) and Hutchinson et al. (1994), research on modeling option prices with neural networks uses cross-validation, often employing the first half of a year to train several architectures, the third quarter for validation, and the fourth to test the predictive quality of the chosen model, see e.g., Garcia and Gençay (2000) or Dugas et al. (2009).

The number of iterations for parameter estimation is typically determined by the convergence of the objective function, with a training process consisting of hundreds if not thousands of steps. In order to mitigate the influence of local minima, the final estimator is usually an average over several runs, initialized at different starting values.

For our approach, we adopt the perspective of heuristic optimization and perform both parameter estimation and model selection by perturbing random solutions, and then evaluating their properties. Initializing a neural network with small random weights corresponds to a Monte Carlo search of the parameter space. The local neighborhoods of these initial solutions are then explored by running Levenberg-Marquardt with Bayesian regularization for a few steps, stopping the process long before convergence.

The resulting population of estimators is then checked for the no-arbitrage conditions discussed in the previous section. Due to the considerable amount of extrapolation involved in obtaining surfaces that extend into the tails of the SPD, these checks allow us to eliminate solutions that exhibit inferior generalization performance, without resorting to cross-validation. Within the subset of valid networks, solutions can then be chosen based on the $RSS$.

In order to offset the reduced flexibility of this highly regularized arrangement, we use significantly more basis functions than comparable approaches. We furthermore stochastically perturb $M$, which allows us to incorporate model selection, and further increases variability among the solutions. While employing the same building blocks as traditional neural network modeling, we essentially sample smooth manifolds and harness their diversity to ensure that the final estimator conforms to the desired constraints. Our approach is thus more akin to stochastic search than gradient-based optimization.

The resulting estimators are both highly adaptive and robust. They provide excellent results over a large variety of market conditions, and do not sensitively depend on the number of iterations, $M$, or the quantity of available training data. The fact that we are able to perform both parameter estimation and model selection using only a sparse cross-section of current observations, allows us to yield better fits than methods relying on aggregate data.
3 Empirical Analysis

This section demonstrates both the pricing performance and robustness of our shape constrained network (SCN) approach by fitting arbitrage-free implied volatility, price and state price density surfaces over a fixed domain, with \( m \in [0.5, 1.5] \) and \( \tau \in [20, 365] \). We contrast the results with two versions of the widely used ad hoc Black-Scholes model, which, due to its simple functional form, can also be fitted on small samples.

Data

We use daily closing prices of out-of-the-money (OTM) call and put options on the S&P 500 for each Wednesday between January 5, 2000 and December 28, 2011. The choice of only working with OTM options is motivated by the fact that they are more liquid. In case a particular Wednesday was a holiday, we use the preceding trading day. Option data and interest rates were obtained from OptionMetrics. We take the mean of bid and ask prices during the day. Option data and interest rates were obtained from OptionMetrics. We take the mean of bid and ask prices during the day.

Following Aït-Sahalia and Lo (1998) we use (8) to derive the implied forward from close to at-the-money (ATM) call and put pairs. Given the implied forward, we can translate OTM puts into in-the-money (ITM) calls, and examined for arbitrage. We check (3) and (4) in their respective spaces, use Black-Scholes to map from implied volatility to price space, and obtain the derivatives via numerical differentiation.

Last but not least, they increase the overall quality of the data, for which we assume the put-call parity to hold

\[
C + Ke^{-r\tau} = P + Fe^{-r\tau},
\]

where \( F \) denotes the forward, and \( P \) is the price of a put option with the same strike price and time to maturity. Following Aït-Sahalia and Lo (1998) we use (8) to derive the implied forward from close to at-the-money (ATM) call and put pairs. Given the implied forward, we can translate OTM puts into in-the-money (ITM) calls, and back out the unobservable implied dividend yield via the spot-forward parity

\[
F = Se^{(r-\delta)\tau}. 
\]

Table 1 summarizes the resulting data set, which contains a total of 121'510 call options, along the dimensions of moneyness and time to maturity.

Benchmark Models

Going back to Dumas et al. (1998), ad hoc Black-Scholes models have been documented to be a tough benchmark. The most common specification models implied volatility along moneyness and maturity via quadratic polynomials

\[
\sigma = \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 \tau + \beta_4 \tau^2 + \beta_5 m \tau. \quad (10)
\]

Ad hoc models (AHG) have been shown to perform better than the stochastic volatility model proposed by Heston (1993), and even two-factor extensions, cf. Christoffersen and Jacobs (2004) and Christoffersen et al. (2009). In the context of extracting option implied densities, the use of quadratic forms to fit implied volatilities along strikes, has been proposed by Shimko (1993)

\[
\sigma = \beta_0 + \beta_1 m + \beta_2 m^2. \quad (11)
\]

The string-wise specification (AHS) poses a challenging benchmark due to its additional flexibility. We fit (11) to all observable maturities and then linearly interpolate the coefficients to recover the entire implied volatility surface for the out-of-sample test.

Shape Constrained Networks

After having introduced our approach to neural network training, we will now focus on the implementation. We model implied total variance as a function of moneyness and maturity, which allows us to directly check the resulting solutions for calendar arbitrage

\[
\nu = \beta_0 + \sum_{j=1}^{M} \beta_j \cdot h \left( w_{0j} + w_{1j} m + w_{2j} \sqrt{\tau} \right). \quad (12)
\]

During modeling, up to 500 different solutions are created and examined for arbitrage. We check (3) and (4) in their respective spaces, use Black-Scholes to map from implied volatility to price space, and obtain the derivatives via numerical differentiation.\(^3\) The networks vary both with respect to their initial parameter values, which we choose following Nguyen and Widrow (1990), and the number of nonlinear basis expansions: \( M = B + [s], \ s \sim N(0, 2) \).

We set \( B = 20 \) and train each network for 10 iterations using the method proposed by Foresee and Hagan (1997). This procedure stops either once 25 valid solutions have been obtained, or the global maximum has been reached. The final estimator is then a simple average over the three solutions with the lowest in-sample errors. Since we work with small samples and limit training to a few steps, we can evaluate hundreds of solutions in a matter of seconds.

\(^3\)Since our method yields an analytic solution, a direct mapping from implied volatilities to state price densities would also be possible, see e.g., Jackwerth (2000).
Since options close to ATM are more liquid, we weight the training errors with $\omega = \mathcal{N}(m, 1, 0.2) + \mathcal{N}(m, 1, 0.1)$. In order to provide some guidance at the boundaries of the model domain, we keep the first string with $\tau > 365$ days, and repeat the first string with $\tau \geq 20$ from $m \in [0.9, 1.1]$ at $\tau = 10$ days. Furthermore, we augment the training data with artificial observations at the average ATM IV from $m \in [1.2, 1.5]$ and $\tau \in [10, 20]$ days. This procedure, which we refer to as anchoring, curbs calendar arbitrage.

As elucidated in Figure 1, both string augmentation and anchoring take place outside of our model domain. While they are not an integral part of the SCN approach, they increase the likelihood of obtaining valid solutions. This is especially the case for anchoring, which provides both guidance in the short term OTM region and counteracts the notoriously steep IV slopes caused by discrete quotes.

The figure also illustrates that the moneyness range of observable strings varies significantly between maturities.

**Table 1. Sample characteristics**

<table>
<thead>
<tr>
<th>$K/F$</th>
<th>DITM</th>
<th>ITM</th>
<th>ATM</th>
<th>OTM</th>
<th>DOTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt; 0.90$</td>
<td>36.19</td>
<td>22.11</td>
<td>17.79</td>
<td>17.63</td>
<td>26.84</td>
</tr>
<tr>
<td>$0.90 - 0.98$</td>
<td>22.11</td>
<td>75.34</td>
<td>25.46</td>
<td>8.38</td>
<td>3.32</td>
</tr>
<tr>
<td>$0.99 - 1.01$</td>
<td>7562</td>
<td>8935</td>
<td>3847</td>
<td>7554</td>
<td>1484</td>
</tr>
<tr>
<td>$1.02 - 1.10$</td>
<td>9739</td>
<td>45.61</td>
<td>20.84</td>
<td>5.73</td>
<td></td>
</tr>
<tr>
<td>$&gt; 1.10$</td>
<td>17204</td>
<td>3871</td>
<td>9637</td>
<td>7131</td>
<td></td>
</tr>
</tbody>
</table>

**Short-term options < 60 days**

- IV (%): 36.19, 22.11, 17.79, 17.63, 26.84
- Call Price: 208.97, 75.34, 25.46, 8.38, 3.32
- # Observations: 7562, 8935, 3847, 7554, 1484

**Medium-term options 60 – 180 days**

- IV (%): 33.34, 22.72, 19.62, 18.48, 20.88
- Call Price: 270.10, 93.72, 45.61, 20.84, 5.73
- # Observations: 17204, 9739, 3871, 9637, 7131

**Long-term options > 180 days**

- IV (%): 29.42, 22.14, 20.57, 19.21, 19.26
- Call Price: 315.48, 124.12, 79.09, 47.98, 13.71
- # Observations: 17179, 6959, 2382, 6654, 11372

Notes. For each Wednesday between January 5, 2000 and December 28, 2011 we combine out-of-the-money calls and puts on the S&P 500 index to create a data set of call options with moneyness $m \in [0.5, 1.5]$ and maturity $\tau \in [20, 365]$.

**Results**

Table 2 shows the results for both the benchmark models and our SCN. For the out-of-sample test we compare the implied volatilities observed at $t + 7$ with $\hat{\sigma}(m_{t+7}, \tau_{t+7})$.

As expected, the global specification AHG exhibits the highest in-sample errors, followed by the more flexible AHS, and our shape constrained networks. Interestingly, the ordering changes for the out-of-sample analysis, with the AHS falling behind the AHG, hinting at overfitting. Despite having the highest degree of freedom, our model also exhibits the smallest errors in the out-of-sample test. Over the entire period, we have to initialize 35 networks on average, to obtain a set of 25 arbitrage-free solutions.

**Table 2. Model comparison**

<table>
<thead>
<tr>
<th>Model</th>
<th>SCN</th>
<th>AHS</th>
<th>AHG</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000 – 2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV Bias</td>
<td>-0.0030</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IV RMSE</td>
<td>0.1661</td>
<td>0.3420</td>
<td>1.0561</td>
</tr>
<tr>
<td>Out-of-Sample IV $R^2$</td>
<td>0.9374</td>
<td>0.7876</td>
<td>0.9043</td>
</tr>
<tr>
<td>2004 – 2007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV Bias</td>
<td>-0.0029</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IV RMSE</td>
<td>0.1428</td>
<td>0.2582</td>
<td>0.7452</td>
</tr>
<tr>
<td>Out-of-Sample IV $R^2$</td>
<td>0.9497</td>
<td>0.8894</td>
<td>0.9309</td>
</tr>
<tr>
<td>2008 – 2011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV Bias</td>
<td>-0.0085</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IV RMSE</td>
<td>0.2103</td>
<td>0.3263</td>
<td>1.1585</td>
</tr>
<tr>
<td>Out-of-Sample IV $R^2$</td>
<td>0.9419</td>
<td>0.9161</td>
<td>0.9258</td>
</tr>
</tbody>
</table>

Notes. In-sample, we report the bias and the root-mean-square error (RMSE). Out-of-sample, the coefficient of determination $R^2$. 

**Figure 1.** Error weighting and augmentation of training data

Notes. The schematic shows the model domain (black rectangle), Gaussian weighting of errors, string augmentation, and anchoring.
Figure 2. Model comparison

Notes. Index level (light) and evolution of daily errors for the AHG (grey), AHS (dotted), and SCN model (dark). The in-sample plots cover each Wednesday between January 5, 2000 and December 28, 2011, the out-of-sample plots span January 12, 2000 to January 4, 2012.

Figure 2 contrasts the evolution of errors with that of the S&P 500. It is evident that our SCN is almost completely unfazed by crisis periods. We can also see that the AHS model generalizes extremely poorly and is highly volatile out-of-sample. Table 3 provides a more detailed look at the behavior of our method. The low ATM errors show the effect of error weighting, the short-term DOTM levels stem from shape constraints impeding calendar arbitrage.

Table 3. SCN performance

<table>
<thead>
<tr>
<th>K/F</th>
<th>DITM</th>
<th>ITM 0.90 – 0.98</th>
<th>ATM 0.99 – 1.01</th>
<th>OTM 1.02 – 1.10</th>
<th>DOTM &gt; 1.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-term options &lt; 60 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV Bias</td>
<td>0.0132</td>
<td>0.0333</td>
<td>−0.0504</td>
<td>−0.0984</td>
<td>−0.2582</td>
</tr>
<tr>
<td>IV RMSE</td>
<td>0.3433</td>
<td>0.1741</td>
<td>0.1826</td>
<td>0.2990</td>
<td>0.7046</td>
</tr>
<tr>
<td>Out-of-Sample IV $R^2$</td>
<td>0.9294</td>
<td>0.8978</td>
<td>0.9089</td>
<td>0.8755</td>
<td>0.7729</td>
</tr>
<tr>
<td>Medium-term options 60 – 180 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV Bias</td>
<td>−0.0022</td>
<td>0.0132</td>
<td>0.0067</td>
<td>0.0019</td>
<td>0.0122</td>
</tr>
<tr>
<td>IV RMSE</td>
<td>0.1981</td>
<td>0.0865</td>
<td>0.0794</td>
<td>0.1074</td>
<td>0.1895</td>
</tr>
<tr>
<td>Out-of-Sample IV $R^2$</td>
<td>0.9537</td>
<td>0.9241</td>
<td>0.9286</td>
<td>0.9113</td>
<td>0.8708</td>
</tr>
<tr>
<td>Long-term options &gt; 180 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV Bias</td>
<td>0.0010</td>
<td>−0.0003</td>
<td>−0.0052</td>
<td>−0.0005</td>
<td>−0.0007</td>
</tr>
<tr>
<td>IV RMSE</td>
<td>0.1049</td>
<td>0.0462</td>
<td>0.0427</td>
<td>0.0520</td>
<td>0.0977</td>
</tr>
<tr>
<td>Out-of-Sample IV $R^2$</td>
<td>0.9704</td>
<td>0.9558</td>
<td>0.9530</td>
<td>0.9524</td>
<td>0.9294</td>
</tr>
</tbody>
</table>

Notes. In-sample and out-of-sample errors for the SCN, partitioned along moneyness and maturity (DITM stands for deep-in-the-money).
Figures 3 and 4 illustrate the quality of the SCN surfaces. The implied volatility surface provides both an excellent fit to the given data and beautifully extrapolates beyond.

**Figure 3.** Implied volatility surface


Figure 3 also shows how anchoring effectively modulates the extrapolation of the DOTM wing for short maturities.

**Figure 4.** State price density surface


Despite the impeccable smoothness of the corresponding SPD surface, the SCN is adaptive enough to fit the convex as well as the concave IV regions on September 5, 2007. In order to fully appreciate these results, they have to be considered in the context of the current literature.

Most parametric models, including specialized forms, like the one proposed by Gatheral and Jacquier (2012), assume IV to be convex in moneyness, and can thus not achieve a comparable fit. Local polynomial models would not be capable to achieve the same smoothness, given the limited amount data. They would have to either use very large bandwidth parameters or work with aggregate data, both of which severely reduces the capacity to accurately reproduce current observations. Another issue with local methods is extrapolation, which is crucial to obtain tails.

The methods proposed in Benko et al. (2007), Glaser and Heider (2010), and Fengler and Hin (2011) are limited to the range of observable strikes and do typically not yield true probability densities. Rescaling to unity has questionable implications regarding the null space of the resulting density and affects both the expected value and higher order moments. Dennis and Mayhew (2002) show that asymmetries in the domain of integration also distort the implied skewness put forward in Bakshi et al. (2003).

Figure 6 highlights another key advantage of our global estimator, namely its robustness regarding variations in the training data. The second derivatives of the jackknife estimates (orange) are virtually identical to our solution.
4 Conclusion

The exigency for accurate state price density surfaces has recently been highlighted by Ross (2013). In this paper, we propose a novel neural network-based approach to approximate arbitrage-free implied volatility, price and state price density surfaces from a sparse cross-section of observations. We demonstrate that our method is robust enough to carry out both model selection and parameter estimation using daily data alone, and obtain excellent in-sample and out-of-sample fits over a period of 12 years. The corresponding state price density surfaces provide a comprehensive snapshot of the current market sentiment. Unlike maturity-wise estimators they enable us to trace the evolution of expectations and risk perceptions along a continuum of future spot trajectories and time horizons.

A natural extension would be to research whether or not the superior quality of our estimators translates into new insights regarding the information content of implied densities. We will investigate this question in the future.

References


