Bayesian Quadrature Approaches to State-Price Density Estimation

Huei-Wen Teng\textsuperscript{a}, John Liechty\textsuperscript{b}

\textsuperscript{a}Department of Statistics, Pennsylvania State University
\textsuperscript{b}Department of Marketing and Statistics, Pennsylvania State University

Abstract

The Risk Neutral or State Price Density (SPD) is used for a variety of important activities in finance, including providing an arbitrage-free tool for pricing complex and less liquid securities and making inference for volatility that is consistent with observed option prices. The importance of understanding this density with respect to asset pricing and risk management has led to a competing number of approaches for estimating the SPD. We propose a finite-dimensional model for the SPD in a Bayesian framework. This modeling approach can be viewed as a Bayesian Quadrature model, where the locations and weights of support points in the finite-dimensional representation of the SPD are random variables. This modeling approach allows a 'prior' reference distribution which can be a parametric distribution (e.g. the lognormal density) or which can be uniform and completely non-informative, and it also provides a posterior distribution of SPD that is consistent with the observed option prices. We assess the performance of the proposed model using simulation studies based on synthetic data and then by contrasting the method with a number of competing methods using S&P 500 index option data.
1. Introduction

Under the usual conditions, markets are arbitrage free and the price of a contingent contract is the discounted value of the expected payoff function, under the State Price Density (SPD) or Risk Neutral Density (RND). The connection between the SPD and the RND are well established, see Harrison and Kreps (1979) and Cox and Ross (1976) for a description of the fundamental asset pricing theory and the theory of asset pricing in complete markets. The SPD is critical for risk management and for pricing assets, for example it can be used to price illiquid or complicated securities. As a result there have been a number of competing approaches for estimating the SPD based on observed European option prices. These include parametric models and non-parametric models. A survey of some of the SPD estimation approaches is given in Jackwerth (1999) and a discussion of the advantage of parametric and nonparametric approaches are given in Engle and Gonzalez-Rivera (1991) and Ait-Sahalia and Lo (1998).

All of the proposed methods, find the best density with respect to the ability for the resulting estimate of the SPD to fit a set of empirical option prices (typically European options). These solutions are attractive as they typically result in closed-form formulas which result in rapid calculations for pricing other assets or calculating risk positions. The natural initial parametric SPD to consider is the lognormal density which follows from the Black-Scholes model, but many empirical studies have shown that this
assumption does not hold (Longstaff, 1995). There have been some efforts that use mixtures of lognormal distributions, such as Melick and Thomas (1997) and Söderlind and Svensson (1997), and others that have used more general distributions, like Abadir and Rockinger (2003) where they develop density functionals based on confluent hypergeometric functions, which cover several popular distributions, such as normal and gammas distributions, and mixtures of these distributions. In a related effort, Giacomini et al. (2008) use mixtures of scaled and shifted $t$-distributions.

A related approach follows from assuming a more general underlying process than the lognormal diffusion and using the resulting parametric distribution for the SPD. For example, Bates (1996a), Bates (1996b), and Malz (1996), present SPD’s that are based on jump-diffusion diffusions. In practice this assumption has empirical weaknesses as Das and Sundaram (1999) shows that option pricing models based on jump process fail to reproduce observed patterns of skewness and kurtosis at long horizons and in general stochastic volatility models fail to match option prices with short maturities. As a result, many diffusion based approaches to pricing use stochastic volatility models with jumps, e.g. Nandi (1998), Bates (2000), Duffie et al. (2000), Pan (2002), Eraker et al. (2003), Maheu and McCurdy (2004).

Several nonparametric methods have been developed for estimating the SPD. Curve fitting methods have been presented by Rubinstein (1994) and Jackwerth and Rubinstein (1996). Ait-Sahalia and Lo (1998) propose the kernel method. In practice the kernel methods rely on a long time series of data to achieve reasonable convergence properties and while the kernel estimation models offer much better predictions for longer maturity contracts,
it is lacking in terms of pricing options across different moneyness levels and may result in poor estimates of the tails of the distribution.

We propose a finite-dimensional model for the SPD in a Bayesian framework. This modeling approach can be viewed as a Bayesian Quadrature model, where the locations and weights of the support points in the finite-dimensional representation of the SPD are random variables. Our method is a Quadrature approach, which uses a finite number of support points to approximate the unknown SPD. From a modeling perspective, our method can be seen as a special case of mixture models, where the component densities are point measures. One advantage of our method is that our resulting pricing formula is linear, allowing for quick calculation. In addition, our method can provide a guide for the complexity, in terms of the number of underlying support points of the reduced form function that is needed to effectively replicate the SPD.

In contrast to standard parametric methods, our approach avoids the problem of model misspecification. It also provides an option pricing model that is parsimonious and can be easily calculated without extra algebraic efforts and very little numerical efforts. In contrast to other non-parametric approaches, given the derivation of the full-conditional densities and detailed description of the inference algorithm provided in this manuscript, our approach is straightforward and simple. In addition, our Bayesian Quadrature approach can be seen as an extension of the curve fitting method proposed by Rubinstein (1994), in the sense that both the locations and the weights of the support points are random variables, whereas in Rubinstein (1994) the researcher needs to specify and fix the locations of the support points. As we
demonstrate, in practice, our approach requires only a comparatively small number of support points to obtain good model fit.

Our paper is organized as follows. In Section 2, we describe our method in more detail and presents a slice sampling algorithm for inference in the Markov chain Monte Carlo (MCMC) inference framework. In Section 3, we demonstrate the performance of the method with a number of simulation studies and consider model selection using cross-validation. We apply our method to S&P 500 data in Section 4, where our method performs better than many existing methods and illustrates the need for both call and put data in order to effectively estimate both tails of the SPD using our Bayesian Quadrature approach. We conclude in Section 5 and discuss potential areas of future research.

2. The Bayesian Quadrature Method

Options or derivatives are financial contracts that give the buyer a cash flow contingent on a payoff function and the behavior of the underlying asset (the payoff function just only depends on the value of the assets at the time of the maturity of the contract for European options). Typically the underlying asset is single stock or an index. In this paper we restrict our attention to options on single assets, but this work could be extended to consider options on multiple assets. To be more specific, if we let $x$ denote the underlying asset’s price at maturity, and $K$ denote the strike price, a put option has the payoff function $\max(K - x, 0)$ and a call option has the payoff function...
max(x − K, 0). Let \( \varphi_{ij}(x) \) denote the payoff function of an option,

\[
\varphi_{ij}(x) = \left((-1)^i(x - c_{ij})\right)^+,
\]

where \((\cdot)^+\) is the positive part of a function; the option is a put option for \(i = 1\), and call option for \(i = 2\). The strike price for each option is denoted by \(c_{ij}\), where \(j = 1, \ldots, N_i\) indexes possible strike prices.

Standard no-arbitrage, asset pricing theory gives the theoretical option price as the discounted value of the expected payoff function under the risk-neutral measure,

\[
e^{-rT}E[\varphi_{ij}(X)],
\]

where \(X\) denotes the stock price at maturity, and the expectation operator \(E\) is taken under the risk-neutral measure. When the SPD \(f(x)\) exists, the theoretical price is

\[
e^{-rT} \int \varphi_{ij}(x) f(x) dx. \tag{1}
\]

If the SPD is discrete, the integral in Eq. (1) is replaced with the summation.

In practice we assume that the option price \(y_{ijk}\) is a perturbation of the theoretical option price and follows

\[
y_{ijk} = e^{-rT}E[\varphi_{ij}(X)] e^{\varepsilon_{ijk}}, \tag{2}
\]

for \(i = 1, 2, j = 1, \ldots, N_i, k = 1, \ldots, N_{ij}\), where the error \(\varepsilon_{ijk}\) is normally distributed with mean zero and variance \(\sigma_{\varepsilon}^2\) and the variance is small enough to be negligible with respect to potential bias.
We define $\delta_{x}(\cdot)$ as a unit point measure at the location $x$ by

$$
\delta_{x}(x) = \begin{cases} 
1 & \text{for } x = x, \\
0 & \text{otherwise}.
\end{cases}
$$

and our proposed Bayesian Quadrature method, models the SPD $f(x)$ with a finite dimensional approximation or

$$
Q(x|w, \theta) = w_{1}\delta_{\theta_{1}}(x) + \cdots + w_{M}\delta_{\theta_{M}}(x),
$$

with the locations $\theta = (\theta_{1}, \ldots, \theta_{M})$ and the weights $w = (w_{1}, \ldots, w_{M})$ of the support points in the finite-dimensional representation are random variables. The locations $\theta$ are constrained to be non-negative and the weights $w$ are constrained to be nonnegative quantities that sum to one. From a modeling perspective, $Q(x|w, \theta)$ is a finite mixture distribution with the point measure as the component densities. The theoretical price under $Q(x|w, \theta)$ comes from the discrete integral or sum,

$$
G_{ij}(w, \theta) = e^{-r_{T}^T \sum_{m=1}^{M} w_{m}G_{ij}^{\theta_{m}}(\theta_{m})},
$$

and the likelihood is given by

$$
L(y|w, \theta, \sigma^{2}) = \prod_{i=1}^{2} \prod_{j=1}^{N_{i}} \prod_{k=1}^{N_{ij}} \frac{1}{\sqrt{2\pi\sigma^{2}_{x}}} e^{-\frac{(\log y_{ijk} - \log G_{ij}(\theta, \theta))^{2}}{2\sigma^{2}_{x}}}, \quad (3)
$$

From an inference perspective, the parameters of interest are $w$, $\theta$ and $\sigma^{2}_{x}$.

For simplicity, we assume a priori that the distribution of $\sigma^{2}_{x}$ is an inverse-
gamma distribution with shape parameter $\alpha_\varepsilon$ and scale parameter $\beta_\varepsilon$, denoted by $\sigma^2_\varepsilon \sim IG(\alpha_\varepsilon, \beta_\varepsilon)$. We assume a vague prior for the weights $w_m$ of the support points or a priori they are uniformly distributed over the unit simplex $\mathcal{W}$,

$$\mathcal{W} = \{ w \in \mathbb{R}^M : 0 \leq w_m, w_1 + \cdots + w_M = 1 \}. \quad (4)$$

If we let $1_A(\cdot)$ be an indicator function on the support set $A$, the prior distribution on $w$ is $p(w) \propto 1_{\mathcal{W}}(w)$. We let $c_{\min}$ and $c_{\max}$ denote the minimum and the maximum of the observed strike prices $c_{ij}$ for $i = 1, 2$ and $j = 1, \ldots, N_i$, respectively. To avoid zero option prices, we assume a priori that the distribution of the locations $\theta$ of the support points are uniformly distributed over the support set $\Theta$,

$$\Theta = \{ \theta \in \mathbb{R}^M : 0 < \theta_m, \min(\theta) < c_{\min}, \max(\theta) > c_{\max} \}. \quad (5)$$

The prior distribution on $\theta$ is $p(\theta) \propto 1_{\Theta}(\theta)$. It is important to note that all three of these assumptions for the prior distributions can be changed in cases where appropriate subjective information is available. Inferences for the parameters of interest is based on the posterior distribution of $w$, $\theta$, and $\sigma^2_\varepsilon$, or

$$p(w, \theta, \sigma^2_\varepsilon|y, \alpha_\varepsilon, \beta_\varepsilon) \propto L(y|w, \theta, \sigma^2_\varepsilon)1_{\mathcal{W}}(w)1_{\Theta}(\theta)p(\sigma^2_\varepsilon|\alpha_\varepsilon, \beta_\varepsilon). \quad (6)$$

Before presenting the main inference algorithm, the Markov chain Monte Carlo (MCMC) algorithm, we describe the idea behind the slice sampling, which we use in part of larger MCMC algorithm. This description builds on
the discussion in (Damien et al., 1999; Liu, 2001). If we wish to sample \( x \) and it has a density proportional to a product of \( k \) nonnegative functions \( f_i(x) \),

\[
p(x) \propto \prod_{k=1}^{K} f_k(x),
\]

we can introduce \( K \) auxiliary variables, \( u_1, \ldots, u_K \), as described in Edwards and Sokal (1988) and define a joint distribution for \((x, u_1, \ldots, u_k)\), which is uniform over the region in which \( 0 < u_k < f_k(x) \) for \( k = 1, \ldots, K \). It is straightforward to show that the marginal distribution of \( x \) is given in Eq. (7). We call the set \( \{ x : 0 < u_k < f_k(x) \} \) the sub-slice, and the interaction of \( \bigcap_{k=1,\ldots,K} \{ x : 0 < u_k < f_k(x) \} \) the slice. Samples from the augmented joint distribution can be generated by generating samples from each sub-slice, conditional on the current values of the remaining auxiliary variables and the variable of interest; then samples from the variable of interest \( x \) can be generated by sampling from the slice, conditional on the current value of the auxiliary variables. Sampling from conditional densities, which are created from a joint density that includes auxiliary variables is known as the slice sampling. To be more explicit, let \( x^0 \) denote the current value of \( x \) at each iteration, and let \( U(A) \) denote a uniform distribution over the set \( A \). The algorithm to sample \( x \) having distribution in Eq. (7) is implemented as follows.

1. Start \( x \) randomly.

2. Repeat the following procedures until convergence.
   (a) Sample \( u_k \sim U([0, f_k(x^0)]) \) for \( k = 1, \ldots, K \).
   (b) Find sub-slices \( S_k = \{ x : f_k(x) > u_k \} \) and the slice \( S = \bigcap_{k=1}^{K} S_k \).
(c) Sample $x \sim U(S)$.

To avoid possible floating-point underflow problems, it is typically better to calculate $g_k(x) = \log(f_k(x))$ instead of $f_k(x)$. When a log transformation is used, the auxiliary variable becomes $z_k = g_k(x^0) - e$, where $e$ is independently and identically distributed as an Exponential distribution with mean 1 (denoted by $e \sim \text{Exp}(1)$) for $k = 1, \ldots, K$. The algorithm to sample $x$ having distribution in Eq. (7) is implemented as follows.

1. Start $x$ randomly.

2. Repeat the following procedures until convergence.
   
   (a) Sample $z_k = g_k(x^0) - e$, where $e \sim \text{Exp}(1)$, for $k = 1, \ldots, K$.

   (b) Find sub-slices $S_k = \{x : g_k(x) > z_k\}$ and the slice $S = \bigcap_{k=1}^{K} S_k$.

   (c) Sample $x \sim U(S)$.

We use the slice sampler in the main MCMC algorithm, which is used to generate samples of $w$, $\theta$ and $\sigma^2_\varepsilon$ that will have a distribution that converges to the posterior distribution in Eq. (6). A summary of the MCMC algorithm is given below and a detail derivation of the proposed slice samplers are given in the Appendix.

1. Start $w$, $\theta$ and $\sigma^2_\varepsilon$ at random values, with in the support of the joint distribution.

2. At each iteration, repeat the following procedures until the distribution of the samples from $w$, $\theta$ and $\sigma^2_\varepsilon$ converges.

   (a) Sample $w_m \sim U(T_m)$, using the slice sampler described in the Appendix, and update $w_M$ as $1 - w_1 - \cdots - w_{M-1}$, where the set $T_m$ is an open interval generated by Eq. (16), for $m = 1, \ldots, M-1$. 

10
(b) Sample $\theta_m \sim U(S_m)$, using the slice sampler described in the Appendix, where the set $S_m$ is an open interval generated by Eq. (22), for $m = 1, \ldots, M$.

c) Sample

$$\sigma^2 \sim IG \left( \alpha + \sum_{i=1}^{2} \sum_{j=1}^{N_{ij}} \frac{v_{ijk}}{2}, \beta + \sum_{i=1}^{2} \sum_{j=1}^{N_{ij}} \sum_{k=1}^{N_{ij}} v_{ijk} (\log y_{ijk} - \log G_{ij}(w, \theta))^2 / 2 \right).$$

3. Simulation Studies

In order to understand how the proposed Bayesian Quadrature model performs, we consider two simulation studies. In the first study we considered a finite, discrete distribution, and in the second study we considered three different continuous distributions, including a lognormal distribution, a mixture of $t$-distributions with two components, and a mixture of $t$-distributions with three components. The number of components, $M$, is assumed to be known in the first simulation, and it is selected using a 10-fold cross-validation for the second simulation. The trace plot, autocorrelation function (ACF) plot, and the kernel density estimation (KDE) of the log-likelihood (LL) are used to assess the convergence of the MCMC algorithm. When the sampling distribution appears to have converged to joint distribution of the simulation is satisfied, we assess model fits using residuals plots and summarize the posterior densities of parameters of interest.

3.1. Simulation Study One

In this simulation study, we consider a finite, discrete distribution with four support points as described in the second column of Table 1, and assume
that the number of support points is known or that $M = 4$. We let $K$ denote the number of different strike prices used to generate the synthetic data sets. To investigate how the number of different number of strike prices affects the inference, we generate four data sets having 4, 8, 16, and 32 different strike prices. The strike prices are equally distributed in the interval between 5 and 30. We then calculate theoretical call and put option prices at these strike prices, and generate additional 500 option prices using Eq. (2) with $\sigma^2$ at 0.05. Scatterplots of these four data sets are in Figure 1.

**Insert Figure 1 About Here**

For each MCMC analysis, we begin with random starting values for the parameters, discard the first 100,000 burn-in samples, and make inference using the the following 1000 samples. Figure 2 shows the trace plots, ACF plots, and KDEs of LL of these four data sets. These plots show that the sampling density appears to have converged.

**Insert Figure 2 About Here**

Figure 3 depicts the 90% credible regions of residuals plots for put and call options. These residuals have zero mean and constant variance across strike prices, indicating that the proposed method produces a good model fit.

**Insert Figure 3 About Here**

Figure 4 shows the trace plots, ACF plots, and KDEs of $\sigma^2$ of these four data sets, and these plots show that the sampling density of $\sigma^2$ appears to have converged.
Figure 5 shows the trace plots of $w_m$ and $\theta_m$ for $m = 1, \ldots, 4$, and Table 1 summarizes the posterior means and standard deviations of these parameters.

In summary, Figure 5 and Table 1 show that the posterior means of these parameters are closer to the true values and the posterior standard deviations decrease as the number of strike prices increases.

3.2. Simulation Study Two

To understand how the proposed method works for a more realistic SPD, we consider the following three distributions,

$$LN(6.7319, 0.0031),$$  \hspace{1cm} (8)

$$0.4 \, t_5(800, 100^2) + 0.6 \, t_5(1100, 60^2),$$  \hspace{1cm} (9)

$$0.3 \, t_5(650, 60^2) + 0.5 \, t_5(850, 100^2) + 0.2 \, t_5(1150, 70^2).$$  \hspace{1cm} (10)

For each density, we uniformly select 50 strike prices from a proper interval, calculate the theoretical call and put option prices, and generate additional 50 option prices using Eq. (2) with $\sigma^2$ at 0.05. Figure 6 gives the KDE and scatter plots of the option prices against strike prices for these data sets.
Cross-validation suggests an optimal model in terms of correctness of out-sample forecasting using the in-sample data set. A 10-fold cross-validation splits the data set randomly into ten sub-data sets and for each sub-data set, we use the remaining nine data sets to calibrate the model and predict the remaining data set. For each MCMC analysis, we begin with random starting values for the parameters, discard the first 10,000 burn-in samples, and make inference using the the following 1000 samples. Figure 7 shows the trace plots, ACF plots, and KDEs of LL of these four data sets. These plots show that the sampling density appears to have converged.

**Insert Figure 7 About Here**

Figure 8 shows the boxplots of the prediction errors against the number of support points $M$ for the three densities (8)–(10), and Table 2 reports the median, the 5-th, and the 95-th quantile of prediction errors.

**Insert Figure 8 About Here**

**Insert Table 2 About Here**

Based on these the criteria that the prediction error has stabilized and that we want as parsimonious model as possible, we selected $M = 14$ for the Log Normal data, $M = 15$ for the mixture of two t-distributions data, and $M = 17$ for the mixture of three t-distribution data. Figure 9 gives the residuals plots, indicating that the method is able to recover the prices across a range of strick prices.

**Insert Figure 9 About Here**
Figure 10 gives the posterior estimate of the empirical density of the SPDs, showing that the method is able to recover the essential structure of the densities used to generate the option prices.

**Insert Figure 10 About Here**

4. Empirical Study

In order to understand how the proposed model would work in practice, we apply it to the four data sets used in Ait-Sahalia and Lo (1998). There are a number of competing methods for estimating the SPD, using this data, making it ideally suited for comparison. We begin with random starting values for the parameters, discard the first 50,000 burn-in samples, and make inference using the the following 1000 samples for each MCMC analysis. Table 3 provides a summary of the performance of the proposed Bayesian Quadrature method with a number of competing methods, using an $R^2$ measure of fit.

**Insert Table 3 About Here**

We considered a range of different support points $M = 5$ up to $M = 10$ and two different ways of incorporating the input data. Because of the put-call parity and the assumed symmetry that this implies with regards to input data, existing methods only use call data as input. We found that for our procedure, as we are explicitly estimating support points based on a payoff function, that using only call data and hence call payoff functions resulted in considerable uncertainty about the lower tail of the SPD, see Figure 11.

**Insert Figure 11 About Here**
Even though the lower tail is not as well defined, the model still fits very well and has an extremely tight predictive range as shown in Figures 11 to 12.

**Insert Figure 12 About Here**

We found that the ability to estimate the tail can be improved, along with the ability to recover the option prices, by using the put-call parity to create both call and put input data. This is done by creating a set of put data based on the call data and then calibrating the model using both sets of data. See numerical results in Tables 4.

**Insert Table 4 About Here**

Using this extended input data, the lower tail behaves in a fashion consistent with the upper tail, see Figure 13 and the fit actually improves, see Figures 13 to 14.

**Insert Figure 13 About Here**

**Insert Figure 14 About Here**

Not only was the fit of the proposed method improved using additional put data, the model turns out to be a strong competitor to the existing methods in the literature. Using the $R^2$ as a measure of this, as is used in earlier studies, the Bayesian Quadrature method clearly dominates the lognormal density assumption and it can improve on all of the competing methods depending on the wether the just the call data or call and put data are used and on the number of support points. One advantage of the Bayesian Quadrature method is that it gives a posterior distribution of the goodness of fit $R^2$, or it
naturally provides a range of reasonable values. We report the 5-th and 95-th quantile of the posterior distribution of $R^2$ values in Tables 3 and 4. One interesting point is that the performance of the Bayesian Quadrature method improves as the number of support points increase. In the best case that we reported, $M = 10$ and using both call and put data, the lower 5-th quantile for the $R^2$ value is better than the $R^2$ value for all of the competing methods for all of the data sets considered, expect in one case. When there are just 15 observations, the “Abadir-Rockinger” density function approach presented in Abadir and Rockinger (2003) has an $R^2$ that is slightly higher than the 5-th quantile of best performing Bayesian Quadrature model. In general, even the Bayesian Quadrature model with just five support points $M = 5$ performs at a comparable level with the two strongest performing methods, the “Abadir-Rockinger” density function approach and the mixtures of $t$ approach as reported in Giacomini et al. (2008) and it dominates the remaining approaches, the “Hermite” approach used in Jondeau and Rockinger (2001), the “Jumps” approach or the Malz-type jump-diffusion model used in Malz (1996), the “Mixtures of lognormal” which uses a mixture of lognormal distributions, and the ”lognormal” which uses a lognormal distributions. The parsimony of this model, just 10 parameters, suggests that there is a limited amount of information in the original option data and it points to the fact that, in practice, the Bayesian Quadrature model may be a very computationally efficient tool for pricing options.
5. Conclusions and Future Work

We propose a Bayesian Quadrature method as a nonparametric approach for the SPD estimation, and provide an efficient MCMC algorithm using slice sampling, for making inference about the resulting finite dimensional approximation to the SPD. In two simulation studies we demonstrate that the proposed inference method can correctly recover the option prices and the underlying SPD used to generate these options price, for a variety of different distributions. We provide an empirical studies that demonstrates both the ability to recover the underlying option prices and the parsimony of the proposed Bayesian Quadrature method. The proposed method performs at a comparable level to the best competing methods with as few as five support points in the Bayesian Quadrature and dominate all of the competing methods when the number of support points are increased to ten. One finding from our empirical study is that the Bayesian Quadrature method does better, when it uses both put and call data (were the put data is generated using the Put-call parity and using call data). The Bayesian modeling framework, reports a distribution for a range of different summaries of interest. It allows for a predictive distribution (conditional on the observed data) for option prices, it results in a posterior distribution of model fit criteria (e.g. $R^2$) and a posterior distribution of the SPD. It also allows the researcher to include different prior distributions, different from the uninformative distributions that we used in our model description, if the researcher has subjective information that merits more informative priors.

The success of this Bayesian Quadrature approach for making inference about the SPD based on European options, suggests that the method could
be extended to accommodate American options (including path dependent, exotic options). One natural way to extend the model is to make the locations multidimensional or to make them realizations of diffusion processes. Another, alternate way would be to build recursive binomial trees, which are consistent with the American and European prices as is discussed in (Rubinstein, 1994). Both of these extensions, with applications to American options and Interest rate options (e.g. Credit Default Spreads) offer fruitful areas for future study.
Table 1: Numerical results for the first simulation. We generate four data, each with 4, 8, 16, and 32 strike prices, respectively. We reported the true parameter values, and the posterior mean and the standard deviation in parenthesis of the parameter. The values of standard deviations below for $\sigma^2$ and $w$ are a multiple of 100 and 10, respectively.

<table>
<thead>
<tr>
<th>Num. Strikes</th>
<th>True Value</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_Y$</td>
<td>0.05</td>
<td>0.05 (0.11)</td>
<td>0.05 (0.08)</td>
<td>0.06 (0.06)</td>
<td>0.05 (0.04)</td>
</tr>
<tr>
<td>$w_1$</td>
<td>0.1</td>
<td>0.08 (0.06)</td>
<td>0.10 (0.01)</td>
<td>0.10 (0.00)</td>
<td>0.10 (0.01)</td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.2</td>
<td>0.23 (0.08)</td>
<td>0.28 (0.03)</td>
<td>0.34 (0.01)</td>
<td>0.20 (0.02)</td>
</tr>
<tr>
<td>$w_3$</td>
<td>0.3</td>
<td>0.30 (0.27)</td>
<td>0.38 (0.02)</td>
<td>0.45 (0.01)</td>
<td>0.30 (0.03)</td>
</tr>
<tr>
<td>$w_4$</td>
<td>0.4</td>
<td>0.39 (0.27)</td>
<td>0.24 (0.03)</td>
<td>0.11 (0.02)</td>
<td>0.40 (0.02)</td>
</tr>
<tr>
<td>$x_1$</td>
<td>5</td>
<td>3.68 (0.60)</td>
<td>4.90 (0.05)</td>
<td>5.07 (0.02)</td>
<td>4.99 (0.01)</td>
</tr>
<tr>
<td>$x_2$</td>
<td>15</td>
<td>14.76 (0.15)</td>
<td>16.35 (0.06)</td>
<td>16.89 (0.02)</td>
<td>15.01 (0.08)</td>
</tr>
<tr>
<td>$x_3$</td>
<td>25</td>
<td>25.55 (0.35)</td>
<td>29.31 (0.07)</td>
<td>32.19 (0.02)</td>
<td>25.02 (0.08)</td>
</tr>
<tr>
<td>$x_4$</td>
<td>35</td>
<td>35.11 (0.44)</td>
<td>37.12 (0.10)</td>
<td>39.70 (0.09)</td>
<td>34.98 (0.01)</td>
</tr>
</tbody>
</table>
Table 2: Numerical results for the second simulation study. The 5-th quantile ($Q_5$), median ($Q_{50}$), and 95-th quantile ($Q_{95}$) of the prediction errors in a 10-fold cross-validation for data generated from a lognormal distribution (LN), a mixture of $t$-distributions with two components ($t_2$) and a mixture of $t$-distributions with three components ($t_3$) are reported with respect to the number of support points, which go from $M = 5$ to $M = 20$. The results for optimal number of support points are underlined, where the optimal number is based on the criteria that the prediction error has stabilized and that we want as parsimonious model as possible.

<table>
<thead>
<tr>
<th>True Dist.</th>
<th></th>
<th>LN</th>
<th>t2</th>
<th>t3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td></td>
<td>$Q_5$</td>
<td>$Q_{50}$</td>
<td>$Q_{95}$</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1.77</td>
<td>1.98</td>
<td>2.46</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1.06</td>
<td>1.25</td>
<td>1.41</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.80</td>
<td>0.91</td>
<td>2.79</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.68</td>
<td>0.76</td>
<td>1.13</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0.54</td>
<td>0.79</td>
<td>2.45</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.59</td>
<td>0.69</td>
<td>2.56</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>0.52</td>
<td>0.59</td>
<td>1.14</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>0.56</td>
<td>0.63</td>
<td>0.72</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>0.47</td>
<td>0.56</td>
<td>0.70</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>0.48</td>
<td>0.57</td>
<td>0.66</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>0.46</td>
<td>0.54</td>
<td>0.75</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>0.46</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>0.46</td>
<td>0.52</td>
<td>0.59</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>0.45</td>
<td>0.53</td>
<td>0.57</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>0.45</td>
<td>0.53</td>
<td>0.57</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>0.45</td>
<td>0.52</td>
<td>0.56</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>0.44</td>
<td>0.53</td>
<td>0.87</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>0.44</td>
<td>0.51</td>
<td>0.56</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>0.44</td>
<td>0.52</td>
<td>0.56</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>0.43</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>0.43</td>
<td>0.51</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Table 3: Numerical results in the empirical studies. We report the 5-th quantile ($Q_5$) and the 95-th quantile ($Q_{95}$) of the posterior distribution of $R^2$ produced by our Bayesian Quadrature approach with different number of support points $M = 5$ up to $M = 10$ for these four data sets used in Ait-Sahalia and Lo (1998), where these data sets vary with respect to the time to maturity in days (T), trading date (Date), and number of observations (#). In addition, we report the $R^2$ values for the same data sets, as summarized by Giacomini et al. (2008), where ”Mixtures of t” refers to the method used by Giacomini et al. (2008), “Abadir-Rockinger” refers to the functional densities used in Abadir and Rockinger (2003), “Hermite” refers to the method used by Jondeau and Rockinger (2001), “Jumps” refers to the Malz-type jump-diffusion model in Malz (1996), “Mixtures of lognormal” uses the mixtures of lognormal distributions, and “lognormal” uses a lognormal density.

<table>
<thead>
<tr>
<th>T</th>
<th>Date</th>
<th>#</th>
<th>Bayesian Quadrature</th>
<th>$Q_5$</th>
<th>$Q_{95}$</th>
<th>$Q_5$</th>
<th>$Q_{95}$</th>
<th>$Q_5$</th>
<th>$Q_{95}$</th>
<th>$Q_5$</th>
<th>$Q_{95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.996008</td>
<td>0.997358</td>
<td>0.999718</td>
<td>0.999820</td>
<td>0.999048</td>
<td>0.999303</td>
<td>0.999159</td>
<td>0.999414</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>0.999136</td>
<td>0.999339</td>
<td>0.999774</td>
<td>0.999858</td>
<td>0.999693</td>
<td>0.999757</td>
<td>0.999809</td>
<td>0.999865</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>0.999363</td>
<td>0.999542</td>
<td>0.999748</td>
<td>0.999812</td>
<td>0.999465</td>
<td>0.999592</td>
<td>0.999905</td>
<td>0.999933</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>0.999844</td>
<td>0.999821</td>
<td>0.999869</td>
<td>0.999887</td>
<td>0.999645</td>
<td>0.999755</td>
<td>0.999653</td>
<td>0.999700</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>0.999607</td>
<td>0.999701</td>
<td>0.999931</td>
<td>0.999958</td>
<td>0.997406</td>
<td>0.998663</td>
<td>0.999668</td>
<td>0.999724</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.999705</td>
<td>0.999792</td>
<td>0.999960</td>
<td>0.999964</td>
<td>0.999974</td>
<td>0.999989</td>
<td>0.999769</td>
<td>0.999821</td>
</tr>
<tr>
<td></td>
<td>Mixtures of t</td>
<td></td>
<td></td>
<td>0.99968</td>
<td>0.99982</td>
<td>0.99965</td>
<td>0.99969</td>
<td>0.99967</td>
<td>0.99989</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixtures of t</td>
<td>0.99922</td>
<td>0.99911</td>
<td>0.99966</td>
<td>0.997214</td>
<td>0.984918</td>
<td>0.993403</td>
<td>0.997926</td>
<td>0.991013</td>
<td>0.995244</td>
<td>0.998267</td>
<td>0.990682</td>
</tr>
<tr>
<td>Abadir-Rockinger</td>
<td>0.99661</td>
<td>0.99700</td>
<td>0.99982</td>
<td>0.98508</td>
<td>0.92857</td>
<td>0.98067</td>
<td>0.951508</td>
<td>0.92857</td>
<td>0.98067</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: We report the 5-th quantile ($Q_5$) and the 95-th quantile ($Q_{95}$) of the posterior distribution of $R^2$ produced by our Bayesian Quadrature approach, using the original Call data and Put data built from the original Call data using the put-call parity, with different number of support points $M = 5$ up to $M = 10$ for these four data sets used in Ait-Sahalia and Lo (1998), where these data sets vary with respect to the time to maturity in days (T), trading date (Date), and number of observations (#).

<table>
<thead>
<tr>
<th>T</th>
<th>Date</th>
<th>#</th>
<th>Bayesian Quadrature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>46</td>
<td>30</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
<tr>
<td>930503</td>
<td>58</td>
<td>58</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
<tr>
<td>931020</td>
<td>98</td>
<td>98</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
<tr>
<td>930611</td>
<td>176</td>
<td>176</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
<tr>
<td>930923</td>
<td>40</td>
<td>40</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
<tr>
<td>48</td>
<td>34</td>
<td>34</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
<tr>
<td>930923</td>
<td>48</td>
<td>48</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
<tr>
<td>48</td>
<td>34</td>
<td>34</td>
<td>$Q_5$  $Q_{95}$</td>
</tr>
</tbody>
</table>
Figure 1: Scatters plots, for the first simulation, of put and call data against strike prices in the left and right panels, respectively, for four data, each with 4, 8, 16, and 32 strike prices, respectively.
Figure 2: The trace plot, autocorrelation function (ACF) plot, and kernel density estimate (KDE) of the log likelihood for the first simulation data sets, each with 4, 8, 16 and 32 strike prices, respectively.
Figure 3: The 90% credible region of residuals and plots of the residuals against strike prices for put and call in the left and right panels, respectively, for the first simulation data sets, each with 4, 8, 16 and 32 strike prices, respectively.
Figure 4: The trace plot, autocorrelation function (ACF) plot, and kernel density estimate (KDE) of $\sigma_\epsilon^2$ for data for the first simulation data sets, each with 4, 8, 16 and 32 strike prices, respectively.
Figure 5: Trace plots of the locations, $\theta_m$, and weights, $w_m$, for $m = 1, \ldots, 4$ for the first simulation data sets, each with 4, 8, 16 and 32 strike prices, respectively.
Figure 6: The true density, scatter plots of put and call data against strike prices for data used in the second simulation study, data was generated from a lognormal distribution (LN), a mixture of $t$-distributions with two components ($t_2$) and a mixture of $t$-distributions with three components ($t_3$).
Figure 7: The trace plot, autocorrelation function (ACF) plot, and kernel density estimate (KDE) of the log likelihood for data for the second simulation, where data was generated from a lognormal distribution (LN), a mixture of $t$-distributions with two components ($t_2$) and a mixture of $t$-distributions with three components ($t_3$).
Figure 8: A summary of the 10-fold cross-validation prediction errors using boxplots against the number of support points, from $M = 5$ up to $M = 25$, for data for the second simulation, where data was generated from a lognormal distribution (LN), a mixture of $t$-distributions with two components ($t_2$) and a mixture of $t$-distributions with three components ($t_3$).
Figure 9: The 90% credible region of residuals and plots of the residuals against strike prices for put and call in the left and right panels, respectively, for the second simulation, where data was generated from a lognormal distribution (LN), a mixture of $t$-distributions with two components ($t_2$) and a mixture of $t$-distributions with three components ($t_3$).
Figure 10: The true density (in bold) and a 90% credible region of the estimated State Price Density (SPD) using the Bayesian Quadrature approach for the second simulation, where data was generated from a lognormal distribution (LN), a mixture of $t$-distributions with two components (t2) and a mixture of $t$-distributions with three components (t3). Strike prices from the original data are represented as crosses on the horizontal axis. Kernel density estimates (KDE) of the posterior draws of the discrete SPD were used to provide visual summaries of the estimated SPD.
Figure 11: The 90% credible region of the estimated State Price Density (SPD) using the Bayesian Quadrature approach for the four data sets used in Ait-Sahalia and Lo (1998). Strike prices from the original data are represented as crosses on the horizontal axis. Kernel density estimates (KDE) of the posterior draws of the discrete SPD were used to provide visual summaries of the estimated SPD.
Figure 12: The scatter plots of the actual options prices (in diamond) and the 90% credible region of the fitted option prices, using ten support points, for the data sets used in Ait-Sahalia and Lo (1998).
Figure 13: The 90% credible region of the estimated State Price Density (SPD) using the Bayesian Quadrature approach for the four data sets used in Ait-Sahalia and Lo (1998), using the original call data and put data built from the original call data using the put-call parity. Strike prices from the original data are represented as crosses on the horizontal axis. Kernel density estimates (KDE) of the posterior draws of the discrete SPD were used to provide visual summaries of the estimated SPD.
Figure 14: The scatter plots of the actual options prices (in diamond) and the 90% credible region of the fitted option prices, using ten support points, for the data sets used in Ait-Sahalia and Lo (1998), using the original call data and put data built from the original call data using the put-call parity.
Acknowledgements

We thank Christian Haefke for providing us data set and codes used in Giacomini et al. (2008).
References


A. Appendix: Slice Sampling

To begin with, we define

\[ g_{ij}(w, \theta) = -\sum_{k=1}^{N_{ij}} v_{ijk} (\log y_{ijk} - \log G_{ij}(w, \theta))^2 / 2\sigma_\varepsilon^2. \]  

(11)

The logarithm of the posterior distribution in Eq. (6) can be written as

\[ \left( \sum_{i=1}^{2} \sum_{j=1}^{N_i} g_{ij}(w, \theta) + \log(p(\sigma_\varepsilon^2|\alpha_\varepsilon, \beta_\varepsilon)) \right) \mathbf{1}_W(x) \mathbf{1}_\Theta(\theta) + J, \]

for some constant \( J \). Given an arbitrary value \( z \), we define the set

\[ B_{ij}(z) = \{(w, \theta) \in \mathbb{R}^{M \times M} : g_{ij}(w, \theta) > z \}. \]

To simplify the set \( B_{ij}(z) \), we calculate

\[
\begin{align*}
a &= \sum_{k=1}^{N_{ij}} v_{ijk}, \\
b &= -2 \sum_{k=1}^{N_{ij}} v_{ijk} \log y_{ijk}, \\
c &= \sum_{k=1}^{N_{ij}} v_{ijk} (\log y_{ijk})^2 + 2\sigma_\varepsilon^2 z.
\end{align*}
\]

Note that \( a \) and \( b \) only depend on the observed options prices and volume, but \( c \) also depends on \( z \). Now, \( B_{ij}(z) \) equals
\[
\{ (w, \theta) \in \mathbb{R}^{M \times M} : a(\log G_{ij}(w, \theta))^2 + b \log G_{ij}(w, \theta) + c < 0 \} \\
= \{ (w, \theta) \in \mathbb{R}^{M \times M} : G_{ij}(w, \theta) \in \left( e^{-\frac{b+\sqrt{b^2-4ac}}{2a}}, e^{-\frac{b-\sqrt{b^2-4ac}}{2a}} \right) \} \\
= \{ (w, \theta) \in \mathbb{R}^{M \times M} : \sum_{m=1}^{M} w_m \phi_{ij}(\theta_m) \in (l_{ij}, u_{ij}) \} ,
\]

where

\[ l_{ij} = e^{r_T + \frac{b-\sqrt{b^2-4ac}}{2a}} , \quad \quad (12) \]
\[ u_{ij} = e^{r_T + \frac{b+\sqrt{b^2-4ac}}{2a}} . \quad \quad (13) \]

The slice \( T_m \) to update \( w_m \) is the intersections of sub-slices \( T_{m0} \) and \( T_{mij} \) defined as follows. Because \( w \) has support set given in Eq. (4), let \( T_{m0} \) be an open interval

\[ T_{m0} = (0, p_{-M}) , \quad (14) \]

where

\[ p_{-M} = 1 - w_1 - \cdots - w_{m-1} - w_{m+1} - \cdots - w_{M-1} . \]

Let \( w^0 = (w_1^0, \ldots, w_M^0) \) and \( \theta^0 = (\theta_1^0, \ldots, \theta_M^0) \) being the current values of \( w \) and \( \theta \) at each iteration in the MCMC algorithm, respectively. We sample the auxiliary variable \( z_{ij} = g_{ij}(w^0, \theta^0) - e \) with \( e \sim \text{Exp}(1) \), and define the sub-slice \( T_{mij} \) as

\[ T_{mij} = \{ w_m \in \mathbb{R} : g_{ij}(w, \theta) > z_{ij} \} . \]
Standard algebra gives

\[ e^{rT} G_{ij}(w, \theta) = \alpha_{ij} w_m + \beta_{ij}, \]

where

\[ \alpha_{ij} = \varphi_{ij}(\theta_m) - \varphi_{ij}(\theta_M), \]
\[ \beta_{ij} = \gamma_{ij-mM} + p_{-MM} \varphi_{ij}(\theta_M), \]
\[ \gamma_{ij-mM} = \sum_{m' = 1, \ldots, M \atop m' \neq m, M} w_{m'} \varphi_{ij}(\theta_m). \]

Recall \( l_{ij} \) and \( u_{ij} \) are given in Eqs. (12)-(13) given a constant \( z \). Now, \( T_{mij} \) equals

\[ T_{mij} = \{ w_m : g_{ij}(w, \theta) > z_{mij} \} \]
\[ = \{ w_m : \sum_{m=1}^{M} w_m \varphi_{ij}(\theta_m) \in (l_{ij}, u_{ij}) \} \]
\[ = \{ w_m : \alpha_{ij} w_m + \beta_{ij} \in (l_{ij}, u_{ij}) \}, \]

and is determined by \( \alpha_{ij} \),

\[ T_{mij} = \begin{cases} 
((l_{ij} - \beta_{ij})/\alpha_{ij}, (u_{ij} - \beta_{ij})/\alpha_{ij}) & \text{for } \alpha_{ij} > 0, \\
((u_{ij} - \beta_{ij})/\alpha_{ij}, (l_{ij} - \beta_{ij})/\alpha_{ij}) & \text{for } \alpha_{ij} < 0, \\
(-\infty, \infty) & \text{for } \alpha_{ij} = 0.
\end{cases} \]  \hspace{1cm} (15)

Because \( T_{m0} \) and each \( T_{mij} \) are open intervals containing \( w_m^0 \), the slice \( T_m \) is
again an interval containing \( u^0_m \). In summary, we have

\[
T_m = T_{m0} \cap_{j=1}^{i=1,2} T_{mij},
\]

where \( T_{m0} \) and \( T_{mij} \) are in Eqs. (14) and (15).

Similarly, the slice \( S_m \) for \( \theta_m \) is the intersection of the sub-slices \( S_{m0} \) and \( S_{mij} \) are defined as follows. Note the support set for \( \theta \) in Eq. (5) is to ensure that the theoretical option price is non-zero. To ensure each stock prices non-negative, we first define

\[
S_{m01} = \{ \theta_m \in \mathbb{R}^M : \theta_m > 0 \}.
\]

Let \( \theta_{-m} \) denote the vector \( \theta \) with \( \theta_m \) removed, \( \theta_{-m} = \{ \theta_1, \ldots, \theta_{m-1}, \theta_{m+1}, \ldots, \theta_M \} \).

To ensure that the minimum of \( \theta \) is less than \( c_{\min} \), we define

\[
S_{m02} = \begin{cases} 
(0, \infty) & \text{for } \min(\theta_{-m}) < c_{\min}, \\
(0, c_{\min}) & \text{for } \min(\theta_{-m}) > c_{\min}.
\end{cases}
\]

Similarly, to ensure that the maximum of \( \theta \) is larger than \( c_{\max} \), we define

\[
S_{m03} = \begin{cases} 
(0, \infty) & \text{for } \max(\theta_{-m}) > c_{\max}, \\
(c_{\max}, \infty) & \text{for } \max(\theta_{-m}) < c_{\max}.
\end{cases}
\]

We define

\[
S_{m0} = S_{m01} \cap S_{m02} \cap S_{m03},
\]

where the sets \( S_{m01}, S_{m02} \) and \( S_{m03} \) are given in Eqs. (17)-(19). Furthermore,
we sample the auxiliary variable \( z_{ij} = g_{ij}(w^0, \theta^0) - e \) with \( e \sim \text{Exp}(1) \), and define the sub-slice \( S_{mij} \) as

\[
S_{mij} = \{ \theta_m \in \mathbb{R} : g_{ij}(w, \theta) > z_{ij} \}.
\]

Recall the definitions of \( l_{ij} \) and \( u_{ij} \) in Eqs. (12)-(13) given a constant \( z \). \( S_{mij} \) equals

\[
\left\{ \theta_m \in \mathbb{R} : \sum_{m=1}^{M} w_m \wp_{ij}(\theta_m) \in (l_{ij}, u_{ij}) \right\} = \{ \theta_m \in \mathbb{R} : w_m \wp_{ij}(\theta_m) \in (l_{ij} - \gamma_{ij-m}, u_{ij} - \gamma_{ij-m}) \},
\]

where

\[
\gamma_{ij-m} = \sum_{m'=1,\ldots,M \neq m} w_{m'} \wp_{ij}(\theta_{m'}).
\]

Because \( w_m \wp_{ij}(\theta_m) \) is non-negative by definition, \( S_{mij} \) can be simplified as

\[
S_{mij} = \begin{cases} 
(c_j - \bar{u}_{m1j}/w_m, c_j - \bar{l}_{m1j}/w_m) & \text{for } i = 1, \\
(c_j + \bar{l}_{m2j}/w_m, c_j + \bar{u}_{m2j}/w_m) & \text{for } i = 2.
\end{cases}
\]

where

\[
\bar{l}_{mij} = \begin{cases} 
l_{ij} - \gamma_{ij-m} & \text{for } l_{ij} - \gamma_{ij-m} \geq 0, \\
-\infty & \text{for } l_{ij} - \gamma_{ij-m} < 0,
\end{cases}
\]

\[
\bar{u}_{mij} = u_{ij} - \gamma_{ij-m}.
\]

Because \( S_{m0} \) and \( S_{mij} \) are open intervals containing \( \theta^0_m \), \( S_m \) is again an open

46
interval containing $\theta_m^0$. In summary,

$$S_m = S_{m0} \bigcap_{i=1,2} \bigcap_{j=1,\ldots,N_i} S_{mij}, \quad (22)$$

where $S_{m0}$ and $S_{mij}$ are given in Eqs. (20)-(21).