A new class of Bayesian semiparametric models with applications to option pricing[*]

Marcin Kacperczyk

University of Michigan Business School

email:mkacpe@bus.umich.edu

Paul Damien

University of Michigan Business School

email:pdamien@bus.umich.edu

Stephen Walker

University of Bath

email:S.G.Walker@bath.ac.uk

June 11, 2003

A new class of Bayesian semiparametric models with applications to option pricing

ABSTRACT

Several studies incorporating estimated volatilities into option pricing formulas have appeared in the literature. However, the models described in these studies tend to perform quite poorly in out-of-sample tests. In particular, significant departures from the observed prices can be seen for the deep out-of-the-money short-term call options where mispricing seems to be somewhat excessive. This paper develops a new family of semiparametric Bayesian models. A particular member from this family that includes a nonparametric component is used to model option prices with the aim of improving the out-of-sample predictions. The principal advantage of injecting a nonparametric component into the model is that wide ranges of kurtosis in the observed asset prices are allowed, leading to lower pricing errors in out-of-sample predictions; that is, significant departures from normality in the underlying distribution of the asset prices when modeled lead to reliable forecasts. A detailed comparative empirical analysis with recent approaches to this problem is made for European out-of-the-money call options for which maturity does not exceed 40 days; it is for this subset of options that the pricing errors from other approaches are significant. The results indicate that the semiparametric Bayesian approach does better in terms of out-of-sample valuation errors compared with other approaches to the problem. Also, consistent with evidence reported in recent literature, for the group of short-term options exhibiting similar moneyness, pricing errors tend to decrease with the time to maturity.

1

Numerous restrictions imposed in the seminal option pricing model by Black and Scholes (1973) have resulted in a plethora of studies aiming to relax distributional assumptions on asset returns with the goal of obtaining more accurate option pricing models.[1] However, when assessing their validity using out-of-sample forecasts, virtually all these parametric models predict prices that depart substantially from the observed prices. One of the common explanations for this mispricing is the inability of these pricing models to fit the volatility structure into the entire cross-section of options. This shortcoming may result either from an overfitting of the model, as is often the case for methods based on implied volatilities, or from the distributional restrictions imposed on the volatility structure as can be the case, for example, in many GARCH-type models. As Das and Sundaram (1999) argue, option pricing models based on jump processes fail to reproduce observed patterns of skewness and kurtosis at long horizons. In turn, stochastic volatility models fail in the same task at short maturities. Given the complementary merits of the above methods, a natural solution to predict all ranges of option prices effectively would be to consider a class of stochastic volatility models with a jump component. As noted by Das and Sundaram (1999), such models, typically, are not parsimonious and/or are difficult to implement routinely.

In this paper, we construct a Bayesian semiparametric, discrete-time stochastic volatility model in which we jointly model the variance and kurtosis of the underlying asset for each individual call option, with the objective of reducing mispricing of short-term options. We focus entirely on pricing of the short-term out-of-the-money (OTM) call options mainly because they tend to be mispriced the most. In particular, we analyze options on the S&P 500 index which constitute the second largest segment of the total options market, and are by far the most studied contracts in the literature. We will illustrate that for options with moneyness ≥ 1.05 and days to expiration less than 40 days, where the pricing errors are the most significant, the Bayesian approach does

[1]For a summary of the most influential papers in this context, the interested reader may consult Bates (1996) or Sundaresan (2000).

2

remarkably well. Also, in extensive empirical studies conducted on longer-term options with different levels of moneyness, our method does at least as well as other approaches. For brevity, we will not be reporting on these in this paper.

The existing literature related to the pricing of index options based on stochastic volatility approaches can be divided into two strands. The first relies on a continuous-time framework in which the volatility path is implied from the information contained in the cross-section of contracts. In a comprehensive summary of the methods falling into this category, Bakshi, Cao, and Chen (1997) show that the average pricing error present for the Black and Scholes model can be reduced by using models with stochastic volatility and stochastic interest rates (SVSI). Even though this improved framework alleviates the mispricing problem somewhat, it is highly parameter inefficient. Moreover, it relies on the existence of a liquid market for options. This last criticism is addressed by the second strand of literature. In this group, the volatility parameter is estimated in a discrete-time setting and is not implied using the cross-section of option data, but is solely dependent on the past information contained in the underlying asset. For example, in their option pricing paper, Heston and Nandi (2000) model the volatility path via the GARCH framework. The resulting prices, although considerably more accurate than those from the Black-Scholes model, still exhibit significant out-of-sample mispricing.[2] Similar magnitudes of mispricing are reported by Dumas, Fleming, and Whaley (1998), Chernov and Ghysels (2000), and Nandi (1998), among others.

The importance of kurtosis for modeling asset prices has been noted historically by Mandelbrot (1963), Fama (1965), while its importance in option pricing has been recognized by Backus, Foresi, Li, and Wu (1997), Drost, Nijman, and Werker (1998), and Das and Sundaram (1999), among others.

---

[2]The Root Mean Squared Errors, reported by Heston and Nandi, for options with maturity not longer than 40 days and moneyness above 1.05 exceed 87% in the best case scenario of the updated GARCH model.

In this study, we develop a new approach to modeling volatility and kurtosis based on the discrete-time setting that reduces the pricing errors of the short-term OTM options considerably. We introduce greater flexibility in modeling the distribution of returns using a semiparametric Bayesian model in the following ways. First, we explicitly model the volatility by conditioning it solely on the past available data. As a result, the method can be applied to pricing a single option without relying on the existence of a highly liquid option market, which is required in pricing options in the continuous-time framework. Second, we relax distributional assumptions; this is in sharp contrast to most of the popular parametric option pricing models where distributional restrictions have to be imposed.[3] In our model, the level of kurtosis is endogenously determined by the data, jointly with the variance process. This is key because, as we will later show, out-of-sample predictions improve dramatically in the absence of parametric assumptions imposed on the observed data. Third, in our framework current returns depend on past volatility; hence, we can capture the correlation between the asset return and the volatility parameter, similar to Nandi (1998), and Heston and Nandi (2000). Lastly, we provide a flexible Bayesian framework in which one can readily incorporate prior beliefs entertained by investors. In this respect, our paper is related to the papers by Bauwens and Lubrano (1998), Eraker, Johannes, and Polson (2003), and Guidolin and Timmermann (2003) who use a Bayesian approach to option pricing. All of the papers cited above use parametric models whereas we adopt a semiparametric method. Three papers of relevance in the present nonparametric context are Stutzer (1996), Aït-Sahalia and Lo (1998), and Duan (2002). Aït-Sahalia and Lo (1998) develop a density estimation (curve-fitting and smoothing) procedure to model the data. Duan builds on the nonparametric canonical valuation model of Stutzer. We differ from these other nonparametric approaches in two or more of the following aspects. Our approach: (a) provides a new class of Bayesian parametric and nonparametric models applicable to

---

[3]Duan (1995), Ritchken and Trevor (1999), Heston and Nandi (2000) assume normality of residuals in their models. Myers and Hanson (1993), Amin and Ng (1994), Bauwens and Lubrano (1998) employ t-distributions instead. Bollerslev, Chou, and Kroner (1992) and Bates (1996) provide summaries of the methods in a broad context of financial markets.

many financial applications; (b) does not employ the GARCH process as the description of the dynamic volatility structure; (c) does not require transforming asset returns to standard normal variates, bypassing the need to filter out the dynamic structure in the observed data: stated differently, imposing an i.i.d structure on the observed data is obviated; (d) is not curve-fitting; (e) is not data intensive; for instance, as noted by Aït-Sahalia and Lo, their classical nonparametric density estimation procedure requires several thousand datapoints to obtain even a reasonable level of accuracy; (f) allows the user to choose between parametric and nonparametric models based on the context, data availability and other practical considerations; (g) is remarkably easier to implement; and (h) provides the entire predictive distributions of the random variables of interest; in this paper, the out-of-sample risk-neutral predictive distributions of the underlying assets are readily obtained.

A semiparametric approach unifies two theoretical components to yield a practically desirable consequence: the parametric component provides structure around the volatility in the data via variance regression, while the nonparametric component makes minimal assumptions about the conditional distribution of the underlying asset. Constructing a family of models that allows parametric and nonparametric components to iteratively feed off each other is an appealing alternative to a purely parametric or purely nonparametric method because we would like to exploit the merits of both methods but avoid the pitfalls of either one. Having noted that, we show later, if a particular context warrants it, you could employ either a parametric or nonparametric approach using the family of models developed in this paper. This flexibility is a central feature of the models described in this study.

We analyze a sample of 253 short-term out-of-the-money S&P 500 index options listed on CBOE during the period 1991-1995. Our empirical results indicate a considerable improvement in the pricing accuracy compared to a standard Black and Scholes model, and the *ad hoc* Black and Scholes model of Dumas, Fleming, and Whaley (1998).

In our sample, the percentage Root Mean Squared Error (RMSE) is reduced to approximately 59%, compared to 119% for the best of the alternative models. This significant reduction in pricing errors suggests that jointly modeling the variance and kurtosis for each individual option better describes the risk-neutral predictive distribution of the underlying asset. Indirectly, our results can be viewed as evidence of the partial elimination of the "smile effect" in the volatility. Also, we compare pricing errors with respect to option maturity by collapsing our sample into four maturity groups. The empirical analysis from the Bayesian approach is contrasted with recent findings in this context.

The rest of the paper proceeds as follows. In Section I we introduce a new family of semiparametric Bayesian models. Section II discusses the construction and the properties of the data. Section III provides out-of-sample pricing results for the call index options written on the S&P 500. Conclusions are discussed in Section IV.

I. Methodology

In this section, we develop a new class of models. While the focus in this paper is on option pricing, it will become clear that the models could be used in other financial applications. Since one of the primary aims of the paper is to develop predictive distributions for the financial application under consideration, we now elaborate on point (h) in the Introduction to motivate the mathematics to be described later.

The predictive distribution of the variable Y (the log return on the S&P 500 index) at time j (r_j), which is subsequently used to price options, is the target variable of interest. Based on the predictive distribution of returns for each subsequent future period, we can obtain the predictive distribution of the S&P 500 index value at maturity (S_T) via the identity,

$$S_j = S_{j-1}e^{r_j}, \tag{1}$$

6

where $j = N, \ldots, T$; S_N is the observed value of the index at the time of prediction.

Now, given the predictive distribution of the terminal index value, we can evaluate the prices of call options (C) one period ahead using the identity

$$C_N = e^{-(r_f - q)\tau} E_N^Q (S_T - K)^+, \tag{2}$$

where E_N^Q stands for the expectation operator taken at time N under the risk-neutral measure. r_f is the annualized spot interest rate, q is the constant dividend rate, τ is the time to maturity in years, K is the option strike price, and $s^+ \equiv \max(s, 0)$. All empirical results are derived using this pricing formula.

A. Semiparametric Scale Mixture of Uniforms (SSMU)

There are four ideas that we link together to develop the class of models we call semi-parametric scale mixture of uniforms (SSMU). [1] A nonparametric family of prior distributions, namely, the Dirichlet Process; see, for example, Ferguson (1973) and Walker, Damien, Laud, and Smith (1999). [2] Scale mixture representation; see, for example, Feller (1971). [3] Variance regression; and [4] Gibbs sampling; see, for example, Smith and Roberts (1993).

Since the scale mixture of uniform representations is central to the construction of the SSMU family of models, we explain this in detail. The other features of the construction are then readily tagged on to the scale mixture approach. We start by considering the scale mixture of uniform representation given by Feller (1971). A word on notation: U denotes the uniform distribution, "$u \sim f$" should be read as "u has density f", and $[A|B]$ denotes the conditional distribution of A given B. Letting y denote the observed data with unknown mean μ, Feller's formulation is given by:

$$[y|u] \sim U(\mu - u, \mu + u),$$

7

$$u \sim F, \tag{3}$$

for some distribution function F with support on $(0, \infty)$. Later, we will generalize this representation which will be of particular relevance to finance.

Since we will be dealing with variance regression, there is no loss of generality by considering the scale mixture representation of a standard parametric linear regression model to describe the ideas; by describing the parametric approach at the outset, the nonparametric approach will be seen as a generalization that includes broad classes of parametric forms as special cases. Consider the following linear regression model.

$$y_i = \sum_{j=1}^{k} x_{ij} \beta_j + e_i, \ i = 1, \ldots, n, \tag{4}$$

where y_i is an observed dependent variable, the x_{ij} are known covariates, the β_j are unknown regression coefficients, and the e_i are independent and identically distributed (iid) error terms. We make the following assumptions about these errors:

A1. $E(e_i) = 0$ and the e_i are symmetric about 0.

A2. $\mathrm{var}(e_i) = \sigma^2$.

We can write the model in (2) in a different way. Introduce the latent variable $u = (u_1, \ldots, u_n)$, and consider the model

$$y_i | u_i = \sum_{j=1}^{k} x_{ij} \beta_j + \tau_i \sqrt{u_i}, \ i = 1, \ldots, n, \tag{5}$$

where the τ_i are iid from the uniform distribution on $(-1, +1)$, and the u_i are iid from some distribution G defined on $(0, \infty)$.

8

Proposition 1 *For the model given in (3), marginally for each i we have,*

(i) $E(y_i) = \sum_{j=1}^{k} x_{ij}\beta_j$ and y_i is symmetric about the mean.

(ii) If $E(u_i) = 3\sigma^2$ then $var(y_i) = \sigma^2$.

Proof See Appendix A.

Therefore we see that provided $E(u_i) = 3\sigma^2$, conditions A1 and A2 are satisfied. We now provide several insights into this approach. Each insight is labeled as a *Remark* because we will be cross-referencing them throughout the paper.

Remark 1: **The Normal Error Regression Model.** In the set up in equation (3), if G is the Gamma distribution with parameters $(3/2, \lambda/2)$, and mean value $3/\lambda$, where $\lambda = \sigma^{-2}$, then marginally each e_i is normally distributed with mean 0 and variance σ^2. This follows from the fact

$$\int_{u>y^2} \exp(-u)du = \exp(-y^2), \tag{6}$$

and noting that model (3) is equivalent to $y_i|u_i$ being uniformly distributed on the interval $(\mu_i - \sqrt{u_i}, \mu_i + \sqrt{u_i})$, where $\mu_i = \sum_{j=1}^{k} x_{ij}\beta_j$.

Remark 2: **The Student t Error Regression Model.** If G is defined via the mixture model

$$[u_i|\xi_i] \sim Gamma(3/2, \xi_i/2)$$

$$[\xi_i] \sim Gamma(v/2, v/2), \tag{7}$$

then marginally each e_i has a Student t distribution with mean 0, scale parameter σ^2, and with v degrees of freedom. This follows from the characterization of a Student t distribution as a chi-squared scale mixture of a normal distribution. Many stochastic volatility models in the literature assume a t-distribution, as in the case of a GARCH formulation; as examples, Bollerslev (1987), Baillie and Bollerslev (1989), Baillie and De-

Gennaro (1990), Bauwens and Lubrano (1998).[4] Also, for a discussion of continuous-time stochastic volatility models, see Ghysels, Harvey, and Renault (1996), Eraker, Johannes, and Polson (2003), and the references therein.

Remark 3: **Kurtosis and the Exponential Power Family (Box and Tiao (1973)).**
Probably one of the most widely used classes of models is the exponential power family of symmetric parametric models, which includes the Normal distribution as a special case. The family allows for different levels of kurtosis in the data. Box and Tiao (1973) note that modeling kurtosis is invaluable when the parent population is non-Normal. Also, Fama (1965) points out that stock returns tend to exhibit non-Normal unconditional sampling distributions in the form of skewness but more pronounced in the form of excess kurtosis.[5] Kurtosis is of two types. (A) Leptokurtic, where the parent population has less pronounced "shoulders" and heavier tails than the Normal. Thus, the Student t distribution is leptokurtic when the number of degrees of freedom are small. (B) Platykurtic, where the parent population has squarer shoulders and lighter tails as in the case of a rectangular distribution. It is easy to see that this parametric family is also characterized as a scale mixture of uniforms. Suppose y has an exponential power distribution. If $f(u)$ has density proportional to

$$f(u) \propto u^{1/\tau - 0.5} \exp(-u^{1/\tau}), \tag{8}$$

and

$$[y|u] \sim U(\mu - \sigma\sqrt{u}, \mu + \sigma\sqrt{u}), \tag{9}$$

then

$$f(y) \propto \exp\left(-\left|\frac{y-\mu}{\sigma}\right|^{2/\tau}\right), \tag{10}$$

where $\tau \in (0, 2]$.

---

[4]Other commonly used distributions include normal (Engle (1982), Bollerslev (1986)), power exponential (Baillie and Bollerslev (1989)), normal-Poisson mixture (Jorion (1988)), normal-lognormal mixture (Hsieh (1989)), generalized exponential(Nelson (1990)).

[5]A similar argument has been proposed by Bollerslev, Chou, and Kroner (1992).

10

Remark 4: **Scale Mixture of Normals.** Remarks 1, 2 and 3 point to the following result. Any scale mixture of normal representation also has a scale mixture of uniform representation; for if $y|\lambda \sim N(0, \lambda)$ and $\lambda \sim g$, then $y \sim U(-\sqrt{u}, +\sqrt{u})$ with $u \sim f$, and

$$f(u) \propto \sqrt{u} \int_{\lambda=0}^{\infty} \lambda^{3/2} \exp(-0.5\lambda u) g(\lambda) d\lambda. \qquad (11)$$

We now generalize the class of parametric models described in Remark 4.

Proposition 2 *Consider the following family of distributions:*

$$[y|u] \sim U(\mu - \sigma\sqrt{u}, \mu + \sigma\sqrt{u})$$

$$u \sim f. \qquad (12)$$

Then the following hold:

$$E(y) = \mu,$$

$$var(y) = \sigma^2 E(u/3)$$

and

$$\text{kurtosis}(y) = \frac{9E(u^2)}{5(E(u))^2} - 3.$$

Proof See Appendix A.

The importance of Proposition 2 is that the mean and variance of u determine the variance and kurtosis of y. To obtain $var(y) = \sigma^2$ and kurtosis $(y) = \tau$, we require $E(u) = 3$ and $var(u) = 5\tau + 6$. This implies τ has to be greater than $-6/5$, which is the kurtosis of the uniform density. A particular distribution that satisfies these requirements is given by

$$f(u) = Gamma(9\alpha, 3\alpha), \qquad (13)$$

where $\alpha = (5\tau + 6)^{-1}$. This new family of distributions has parameters (μ, σ^2, τ), where each element in the vector corresponds to the mean, variance and kurtosis, respectively. Clearly, the normal distribution is recovered when $\tau = 0$ ($\alpha = 1/6$).

Thus we have shown that the scale mixture of uniform distributions generalize a very large family of parametric models widely used in the finance literature, with the added advantage that one can impose wider ranges of kurtosis than is typically allowed in many parametric models such as, for example, a Student-t GARCH model. Since our aim is to introduce greater flexibility in the modeling process by relaxing parametric assumptions as much as possible, we require Propositions 3 and 4.

Proposition 3 *If f_Y is a unimodal, symmetric density about 0 and $f_Y'(y)$ exists for all y, then*

$$f_Y(y) = 1/2 \int_{u > y^2} f_{U(u)} du / \sqrt{u} \qquad (14)$$

where $f_{U(u)} = -f_Y'(\sqrt{u})$.

Proof This follows from a result of Feller (1971, page 155).

In words, Proposition 3 implies that the scale mixture of uniform family coincides with the class of unimodal, symmetric distributions.

The result in Proposition 3 is more general and simpler than the approach taken by Feller (equation (1)). Also note that the Feller approach only provides the unimodal, symmetric distribution for Y; besides, the first *four* moments of U are all required to specify the first four moments of Y. With our approach,

$$[y|u] \sim U(\mu - \sigma\sqrt{u}, \mu + \sigma\sqrt{u}),$$

$$u \sim F. \qquad (15)$$

12

Clearly, since we are explicitly modeling the variance of Y, only the first *two* moments of U are required to define the first four moments of Y.

Remark 5: **A nonparametric scale mixture model** is obtained by assigning F a stochastic process prior, say, the Dirichlet Process. $F \sim Dir(c, F_0)$ means, F is assigned a Dirichlet process prior with mean F_0 and scale parameter c. c is a measure of strength of belief in your prior guess at the mean. Note, as an example, you could center the location parameter, F_0, around any member of the exponential power family. This implies that our scale mixture of uniform representation encapsulates all ranges of kurtosis. We use the Dirichlet process for two reasons: (a) the theoretical properties of the process are very appealing; see Ferguson (1973) and Proposition 4 below; (b) implementing the overall model is highly simplified; see MacEachern (1998) and Appendix B. We note here that the scale mixture representation is such that we actually bypass simulating the Dirichlet process; that is, the computational burden is substantially reduced.

In the second stage of our hierarchical model, in equation (13), we investigate what happens if F is assigned a Dirichlet Process prior.

Proposition 4 *If $u_1, \ldots, u_t$ given F are independent and identically distributed from F and F has a Dirichlet process prior then the joint distribution of $(u_1 \ldots u_t)$ is symmetric, say $p(u_1 \ldots u_t)$, and such that U_i given U_{-i} is distributed from $f(\cdot)$ with probability proportional to c or is equal to U_j, $j \neq i$, with probability proportional to 1. Here $c > 0$ is the scale parameter of the Dirichlet process and f is the density function corresponding to the expected distribution of F.*

Proof See Escobar and West (1995).

Remark 6: **Kurtosis and Nonparametric Priors.** A central goal of this paper is to model the kurtosis in the observed asset prices using a nonparametric prior that can be centered around the exponential power family described under Remark 3. As is well-

known, the parent population of option prices is non-Normal. But at the same time, using parametric forms that deviate from normality to model option prices may also be unreliable because there is no *a priori* reason to endow the parent population with a predetermined degree of kurtosis. On the other hand, a nonparametric prior does not impose any degree of kurtosis on the parent population via a parametric component. Rather, conditioned on the observed data, the nonparametric component in the model encourages the predictive distributions to "gravitate" towards the underlying leptokurtic or platykurtic shape of the parent population from which the data arise. This is one reason why we would expect predictions from such a nonparametric model to be more accurate than predictions from parametric models. [6]

Now consider an observed sequence of financial time series $Y_1, Y_2, \ldots, Y_t$ in which the conditional density of the data based on its past values is unimodal and symmetric with the variance depending on the past. Without loss of generality, setting μ equal to 0, in canonical notation, consider the following development thus far; we have the following hierarchical modeling framework.

$$[Y_t | \mathcal{F}_{t-1}] \sim U\left(-\sigma_{t-1}\sqrt{u_t}, +\sigma_{t-1}\sqrt{u_t}\right),$$

and

$$u_1 \ldots u_t \sim p(u_1, \ldots, u_t), \tag{16}$$

where $\mathcal{F}_t = \sigma(Y_1, \ldots, Y_t)$. Since $\sigma > 0$ is the volatility parameter, equation (14) is essentially a scale mixture of uniforms representation of the observed data conditioned on past values of the volatility. We take $p(u_1, \ldots, u_t)$ to be based on a Dirichlet process prior.

We highlight several features of the discussion so far.

---

[6] Antoniak (1974) provides other theoretical reasons as well but they are not as critical in the present context.

- Not much importance should be attached to the uniform distribution. The important fact is that it is mixed with the latent variable u. Also, based on the development earlier, the u_is are iid from a random distribution function, F, which has a Dirichlet process prior. Hence, the distribution of Y_t is mixed via a random probability measure, which implies it could be any unimodal and symmetric distribution function; see, Proposition 4, Remarks 1,2, 3, 5 and 6.

- By introducing a latent variable, u, and rewriting the distribution of the observed data as a mixture model – the mixing taking place with respect to the distribution of u – we can characterize several well-known classes of parametric models. In particular, standard GARCH formulations and the traditional multiple linear regression models are seen to be special cases of this framework; see, Remarks 1, 2, 3, 5 and 6.

- By extending the scale mixture representation of Feller to include the variance of the observed data as part of the mixing measure, we introduce variance regression into the modeling framework. This is a central feature of our approach and is of direct importance to the financial application at hand. (Heston and Nandi (2000) note the importance of exploiting the correlation between asset returns and volatility.)

- The transition from parametric to nonparametric classes of models is readily obtained by merely taking the prior measure on the mixing variable to be infinite dimensional; that is, a stochastic process such as the Dirichlet. Indeed, as we showed earlier, the class of transition densities which we consider in this paper only have to be unimodal and symmetric. In particular, as described under Remarks 1, 2, 3, 5 and 6, our class of models includes all ranges of kurtosis.

- At the start of this section, we noted that there are four distinct ideas we would like to bring together. The first three have been developed so far. The fourth feature of the SSMU family of models is valuable, namely, the computational component.

It will be shown later that the computational model resulting from our approach leads to an easy-to-implement Gibbs sampling algorithm. This was one of the motivating factors underlying the scale mixture of uniform approach to modeling the S&P 500 options.

A.1. Prior Distributions

The following describes the various priors used in the empirical analysis. Where necessary, a conjugate hyper-prior is used; see, for example, Eraker, Johannes, and Polson (2003). Given recent advances in Bayesian computation, the practitioner can readily employ non-conjugate prior distributions if needed; for details, see MacEachern (1998), and Mira, Moller, and Roberts (2001).

The nonparametric component: The scale parameter of the Dirichlet process, c, is assigned a Gamma(a,b) hyper-prior distribution. The second parameter is the prior guess at F_0. In this paper, for illustrative purposes, we center the transition density on the normal distribution, and so we will take the location parameter of the Dirichlet process to be a Gamma($3/2, 1/2$) distribution; see also, Remark 1.

The parametric component: Consider the variance regression,

$$\sigma_t = \exp\left(\beta_0 + \sum_{k=1}^{M} \beta_k Z_{kt}\right), \tag{17}$$

where $\beta_0, \ldots, \beta_M$ are parameters to be estimated and the $\{Z_{kt}\}$ are observed information (independent variables) up to and including time t.

We assign a prior distribution to each of the β_k, which, without loss of generality, is assumed to be independent normal distributions with zero means and variances ψ_k^2.

A strength of the Bayesian approach is the ability to incorporate context-specific knowledge in the modeling process. However, for illustration purposes, all our prior settings were chosen to reflect vague prior knowledge; these are detailed in the section on estimation.

A.2. The Gibbs Sampler

Briefly, the Gibbs sampler is a Markov chain Monte Carlo method of sampling from conditional distributions, which, in the limit, induces samples from the required posterior marginal distributions of interest; see, Smith and Roberts (1993) and Mira, Moller, and Roberts (2001) for details. The key to implementing a Gibbs sampler therefore is to be able to obtain the conditional distributions, up to proportionality, of the random variables of interest. In our SSMU model, the following full conditional densities have to be sampled,

$$p(u_i \mid \text{everything else}), \quad i = 1, \ldots, N,$$

$$p(\beta_k \mid \text{everything else}), \quad k = 1, \ldots, K,$$

$$p(c \mid \text{everything else}).$$

N equals the number of observed returns in the sample and K is the number of independent variables in the variance regression. Appendix B provides the details of this Gibbs sampler.

The conditional structure of the time series in equation (16) provides us the intuition behind obtaining out-of-sample forecasts. Clearly, there is no closed-form description of the predictive distributions; rather, we approximate it using the sampled values from the Gibbs sampler. This procedure is detailed in Appendix C. Here we simply note that the procedure is remarkably easy to implement because the predictive distribution for Y is constructed by merely sampling the latent variable U; sampling U is straightforward and is described in Appendix C.

17

This ease in obtaining the risk-neutral predictive distributions of the underlying assets is a direct consequence of our scale mixture method, barring which obtaining the predictive distributions could be a formidable task.

B. Benchmarks for comparison

In order to evaluate the accuracy of our option pricing model, we need to specify appropriate benchmarks for comparison. Specifically, we select the following two models: the standard Black and Scholes (1973) model and the *ad hoc* BS model of Dumas, Fleming, and Whaley (1998). These two models have been used in this role by Heston and Nandi (2000) and Guidolin and Timmermann (2003). Both methods in their final form use the standard Black-Scholes formula to obtain option prices. As Heston and Nandi, and Dumas et al. note, the *ad hoc* Black-Scholes model performs better in an out-of-sample setting than many standard models such as the implied binomial tree or the deterministic volatility models of Derman and Kani (1994), Dupire (1994), and Rubinstein (1994). An important feature for this study though is that the *ad hoc* Black-Scholes model also gives similar valuation errors as the GARCH model of Heston and Nandi for the group of short-term (< 40 days) deep out-of-the-money (moneyness ≥ 1.05) options, the group with which we are most interested. Consequently, the *ad hoc* Black-Scholes model constitutes an appropriate benchmark when comparing our results to those obtained by Heston and Nandi.[7]

Differences occur, however, in the method of calculating the spot volatility that enters the Black-Scholes formula. To calculate the spot volatility under the Black-Scholes model, we follow the procedure of Engle and Mustafa (1992). We calculate the

---

[7]In the context of predicting short-term OTM options, the *ad hoc* BS model also does better than the Bayesian model of Guidolin and Timmermann (2003).

implied volatility using the information contained in all options present in our sample on a particular day by minimizing the following loss function.

$$\sigma_{t-1} = \text{argmin} \sum_i \epsilon^2{}_{i,t-1}, \tag{18}$$

where

$$\epsilon_{i,t-1} = BS_{i,t-1} - \hat{C}_{i,t-1}. \tag{19}$$

$\hat{C}_{i,t-1}$ is the observed price of option i at time $t-1$, and BS is the model price calculated as

$$BS = S_0 N(d1) - K e^{-r_f \tau} N(d2) \tag{20}$$

$$d1 = \frac{\ln(S_0/K) + (r_f + 0.5\sigma^2)\tau}{\sigma\sqrt{\tau}} \tag{21}$$

$$d2 = d1 - \sigma\sqrt{\tau}, \tag{22}$$

where S_0 is the spot index price; K is the strike price; r_f is the annualized risk-free rate; τ is the time to maturity, in years; σ is the constant volatility parameter. The volatility, σ_{t-1}, which minimizes the value of the loss function, is subsequently substituted into the Black-Scholes formula to calculate a desired option price at time t.

As pointed out by Heston and Nandi, the Black-Scholes model with a single implied volatility across all strikes and maturities is perhaps too restrictive. Therefore, as a second method of pricing, we use the *ad hoc* Black-Scholes model of Dumas et al. in which each option has its own implied volatility depending on its own strike price and the time to maturity. Following Heston and Nandi, we apply a particular functional form to estimate the value of implied volatility:

$$\sigma = \alpha_0 + \alpha_1 K + \alpha_2 K^2 + \alpha_3 \tau + \alpha_4 \tau^2 + \alpha_5 K \tau, \tag{23}$$

where σ is the implied volatility (using Black-Scholes) for an option of strike K and maturity τ. The coefficients of the *ad hoc* model are estimated on a particular day via

ordinary least squares, minimizing the squared errors between the Black-Scholes implied volatilities across different strikes (and maturities), and the model's functional form of the implied volatility. Finally, we substitute the estimated functional form of volatility in the standard Black-Scholes formula to calculate the price of the desired option.

II. The data

The empirical predictions of our model are derived for the intraday data of the European call options, written on the S&P 500 index, traded on Chicago Board Options Exchange (CBOE).[8] Our sample covers the period from January 1990 through December 1995. This is a slightly shorter period than in Chernov and Ghysels (2000), and longer than in Bakshi, Cao, and Chen (1997), Dumas, Fleming, and Whaley (1998), and Heston and Nandi (2000).[9] Bid-ask quotes for options are obtained from the Berkeley Option Database. For each day in the sample, only the first bid-ask quote of each option contract, reported between 9 a.m. (CST) and 9.30 a.m., is employed in the empirical tests. In particular, we take the mid-point price of the spread as an observed valid price. The minimum tick for series that trades below $3 is 1/16; for all other series the tick is 1/8. Strike prices are spaced 5 points apart for nearby 3 months and 25 points for other far away months. The options expire in the three near-term months in addition to the months from quarterly cycle of March, June, September, and December. The respective expiration day for each of the months is Saturday following the third Friday of the month. For each option price, corresponding index values are recorded to avoid the problem of nonsynchronous trading.[10] In order to make the observed S&P 500 index

---

[8]This example is mainly illustrative. The analysis for put options gives qualitatively similar results and thus we limit our exposition to call options only.

[9]Chernov and Ghysels (2000) analyze the period 1985-1994, Bakshi, Cao, and Chen (1997) consider the period from June 1988 until May 1991, the sample in Dumas, Fleming, and Whaley (1998) spans June 1988 through December 1993, while Heston and Nandi (2000) consider calendar years 1992-1994.

[10]For more details regarding this problem, see Stephan and Whaley (1990), Fleming, Ostdiek, and Whaley (1996).

and the simulated underlying fundamental data comparable, we adjust the S&P 500 index for dividends. However, instead of subtracting dividends directly from the index, we apply a constant, continuously compounded, dividend rate of 2%, consistent with the approach in Chernov and Ghysels (2000).[11]

We also apply certain exclusion criteria. First, we exclude options with fewer than six days to expiration as they may induce liquidity-related biases. Besides, shorter-term options have relatively small time premiums, hence the estimation of volatility is extremely sensitive to nonsynchronous option prices and other possible measurement errors. Second, to mitigate the impact of price discreteness on option valuation, price quotes lower than $3/8 are not included. Third, if the first quote of the index is missing or unreliable we take the next plausible observation. Fourth, we include only options for which prices form a non-increasing function of the strikes. Since the European call option price is calculated as the discounted value of the difference between the expected index level at maturity and its strike price, *ceteris paribus*, the higher the strike price the lower the call option price. Finally, each option price has to satisfy the following arbitrage conditions.

$$S_t \geq C(t, T) \geq \max(0, S_t - PVD - e^{-r_f(T-t)}K), \tag{24}$$

where $PVD = \sum_{i=1}^{n} D_i e^{-r_i(t_i - t)}$ denotes the present value of the cash dividends.

The focus of the main part of our analysis is mostly on the short-term deep out-of-the-money options. However, since the Black-Scholes and *ad hoc* Black-Scholes methods rely on the distributions of entire samples, we consider both the unrestricted and restricted samples.

---

[11]Guidolin and Timmermann (2003), for example, use the dividend yield of 3%. We have checked the sensitivity of our results with respect to such level of yield and the results remain qualitatively unchanged.

In Table I, Panel A, using the full sample, we report the summary statistics for the average bid-ask mid-point price, and the total number of observations for each moneyness-maturity category. The option is considered to be in the money (ITM) if its moneyness, defined as K/S, does not exceed 0.97, and at-the-money (ATM) if its moneyness falls between 0.97 and 1.02. All other contracts are out-of-the-money (OTM). The unrestricted sample includes a total of 84770 call option observations. In this group, in-the-money and at-the-money options make up approximately 37.5% percent and 36.7% percent of the total sample, respectively. The average call price ranges from $0.91 for short-term, deep out-of-the-money options to $57.83 for long-term, deep ITM contracts.

Panel B shows the properties of the restricted sample, which includes only the short-term options with maturity not longer than 40 days and moneyness greater than or equal 1.05. To illustrate the cross-sectional variation inside this class of contracts, we sort this restricted sample into four bins based on the time to maturity: $6 - 25$, $26 - 30$, $31 - 35$, and $36 - 40$ days. The largest bin includes calls with the longest maturity, while the smallest one includes calls with the shortest maturity.

Insert Table I about here

Since the main analysis in this paper requires the history of past stock returns, we use at least a minimum of one year of daily observations as an estimation period (here 1990), while the remaining years constitute our test period. We increase the estimation window as we move forward in time. Also, in order to compare the prices obtained from the model to the real-time prices, the maturity date of any option must not exceed December '95. These two constraints reduce the number of observations in our sample from 583 to 253. Each maturity bin, however, still includes at least 30 options, which seems to be sufficient when comparing the contracts across the time dimension. Note, from a statistical perspective, the sample size that we work with is substantially large.

22

Also from a statistical perspective, in a Bayesian framework, interest is seldom on large-sample properties of estimators.

III. Empirical Results

In Section I we outlined the procedure to obtain option prices. In practice, the implementation of this procedure involves the following three steps. First, we obtain the predictive distribution of the index at maturity for each analyzed option by substituting recursively simulated returns into equation (1). Second, using Girsanov's theorem, we transform this predictive distribution into a risk-neutral distribution. Assuming no arbitrage, under the risk-neutral density, the mean of the terminal index value is equal to the risk-free rate (r_f), and the variance of the underlying asset remains the same. This transformation is performed by first de-meaning the entire simulated distribution and then adding the future value of the index compounded at the risk-free rate. Third, to obtain option prices, we calculate the appropriate discounted expected value of the terminal *payoff* (Index - Strike) under previously derived risk-neutral measures.

The first step of our model involves the selection of the vector of covariates, Z. Since our model is entirely out-of-sample, we cannot use predetermined variables because they are unknown ex-ante. Instead, we must rely on a measure obtained using the lagged values of the predictor itself. In particular, we use the squared past log returns calculated as: $Z_t = \{\ln(S_t/S_{t-1})\}^2$, where S denotes the value of the S&P 500 index. We want to emphasize the fact that our results are not significantly dependent on a particular choice of the independent variable (or variables). For illustrative purposes, we consider just one predictor. Other contextually motivated predictors could be added to enrich the analysis.[12] This is yet another flexible feature of the SSMU model. Like we noted earlier,

---

[12]For example, we also use the squared past index normalized by the squared starting index value. The results remain qualitatively similar.

the SSMU framework can be used in other applications as well where, for example, a multiple variance regression might be warranted.

A. Estimation

The estimation problem proceeds as follows.[13] First, we obtain the posterior distributions of the parameters of the variance regression. Next, given the estimated distributions, we derive the predicted value of the variance. This value is subsequently used to derive the predicted value of the return, and hence the terminal index price. S_T – the terminal price – is calculated by iterating $S_{t+1} = S_t e^{r_{t+1}}$ for $t = 0, 1, \ldots, T - 1$; where r_{t+1} is the return from period t to period $t + 1$. Finally, we will use this value to predict the call price one period ahead.

Since our estimation relies on the Bayesian paradigm, it is essential to specify the prior setting for all the random quantities in the model. Our choice of prior values is designed to ensure a non-informative belief about the random variables, a so-called objective prior setting. Note that the Bayesian approach allows you to change or update these prior choices as more knowledge about the process is gained over time. Denoting π to be a prior distribution, $\pi(\beta_0) = N(0, 0.25)$, $\pi(\beta_1) = N(0, 0.01)$. We take $\pi(c)$ the scale parameter of the Dirichlet process, to be $Gamma(a, b)$ with $a = b = 0.01$. Since we are centering the transition density on the normal distribution, we take the location F_0 to be a $Gamma(3/2, 1/2)$ in the uniform scale mixture.

Armed with these prior choices, and using the Gibbs sampler, samples are drawn from the full conditional distributions detailed in Appendix B. Specifically, our Gibbs sampler involves one million Monte Carlo iterations. Using well-known convergence diagnostics (Smith and Roberts (1993)), having "burned-in" the first 500,000 iterates, approximate

---

[13]Given that in reality investors can observe the entire history of the past data, it is reasonable to assume a recursive approach in calculating the consecutive option prices; i.e., we will always use the entire available time-series. In our case, the shortest series equals one year, while the longest series is bounded by the last option data in our sample.

24

independence in the sampled variates is obtained by using every 1000th iterate from the chain.

B. Option Pricing

We only provide a detailed comparative empirical analysis for European out-of-the-money call options for which maturity does not exceed 40 days; it is for this subset of options that the pricing errors from other recent approaches are significant. The results indicate that the semiparametric Bayesian approach does better in terms of out-of-sample valuation errors compared, for example, to the *ad hoc* Black-Scholes model of Dumas, Fleming, and Whaley (1998), the GARCH model of Heston and Nandi (2000), and the Bayesian model of Guidolin and Timmermann (2003). We are not reporting extensive results for the other call options (or for that matter put options) because, in all other cases, we found that the SSMU method does at least as well as the parametric models that have appeared in the literature and which are relevant to this study. This is to be expected based on the theoretical discussions earlier. Here we merely recall that many of the parametric models that have been used in the present context are a member of the exponential power family, which as we showed earlier, is the family around which the nonparametric component is centered. Relaxing distributional assumptions therefore leads to more accurate predictions. Presenting the graphical results of the simulation for every option in our subset of 253 options is also not feasible. Rather, we show the important steps of this process for four particular contracts where each one is based on a different time to maturity. All other options are analyzed in exactly the same manner; we do present the pricing results for all 253 options in tables.

The contracts we analyze in detail have maturities equal to: 9, 29, 32, and 40 days, respectively. By construction, all the four contracts we consider exhibit similar moneyness.

Given the univariate nature of our variance regression, we have to estimate only two coefficients for each option. Figure 1 shows the posterior marginal density of the slope coefficient corresponding to the covariate $Z_t = \{\ln(S_t/S_{t-1})\}^2$ for each of the four contracts.

Insert Figure 1 about here

Using the sampled variates from the Gibbs sampler, the predictive terminal price distributions for the four contracts are constructed and presented in Figure 2. The technical details of the sampling process have been presented in Appendix C.

Insert Figure 2 about here

Recall Remark 6. As expected, the predictive terminal price distributions differ even among these four contracts chosen for illustration; specifically, note the differences in the degree of kurtosis in Figure 2.

The risk-neutral predictive terminal price distributions are subsequently used to price the corresponding options. The predicted (actual) price for the four contracts in increasing order of to time to maturity are as follows: 0.77 (0.51), 0.57 (0.69), 3.32 (2.13), 2.41 (1.63).

Next, to compare the efficiency of the SSMU method, we compute the dollar and percentage errors averaged across all 253 available options. In particular, for each option, we compare the accuracy of the SSMU pricing model to that of Black-Scholes and *ad hoc* Black-Scholes models. The metric for comparison is the root mean squared error (RMSE), defined as root of the mean squared deviations of the model price from the observed price. To calculate the percentage errors, the RMSEs are further scaled by the average price of the option in the sample.

Insert Table II about here

26

Table II indicates that the Root Mean Squared Error (RMSE) for SSMU equals 58.97%, which is substantially lower compared to 158.05% using Black-Scholes and 119.04% using the *ad hoc* Black-Scholes. This striking reduction in pricing errors can be mostly attributed to the fact that the nonparametric component allows for greater flexibility in capturing the uncertainty in the evolution of the volatilities of the underlying asset.

One could argue that the performance of our method may be sensitive to a particular choice of the sample period. We argue that this is not the case. Contrast our results above to the results reported by Heston and Nandi (2000). In particular, in their sample, the average percentage RMSE using the Black-Scholes and *ad hoc* Black-Scholes models equal 191.82% and 92.32%, respectively. Most importantly, Heston and Nandi do not exclude options with very low prices (lower than $0.375), so their average OTM option has a much lower price. This, in turn, implies much lower absolute pricing errors. Besides, the horizon of their data set is also shorter (1992-1994). Thus, in general, the results obtained using two different samples do not seem to be inconsistent. Hence, we can take comfort that our sample does not have any significant pricing bias.

Given that the Black-Scholes model underperforms significantly for the entire sample, the rest of our results will be compared to the *ad hoc* Black-Scholes model only, keeping in mind that this method actually constitutes a lower bound on the errors of the two pricing methods.

Many authors have noted that biases such as smiles and smirks are sensitive to term structure of options. In particular, biases in option prices driven by models' failure to incorporate excess kurtosis decrease with the time to maturity. In the remainder of this section, we study whether the pricing errors of the SSMU model are also related to the time to maturity. Recall that all options in our sample, by construction, are almost the same in terms of their moneyness so there is no additional confounding imposed by the moneyness characteristic.

27

As described in Section II, we divide our sample into four distinct maturity bins, and calculate the average pricing errors for options in each respective bin. The details of the analysis are presented in Table III.

Insert Table III about here

Table III indicates the following.

(1) The average mispricing is lower for each bin under the SSMU approach. Moreover, this reduction is larger for options with longer maturity, consistent with findings reported by other researchers. In some instances, the average mispricing is twice as large while using other models.

(2) The percentage pricing errors are lowest for the $36 - 40$ days bin, equaling 51.93%. They are the largest for the $6 - 25$ days bin, with an average of 78.27%. Thus, very short-term options pose a significant challenge for pricing models. This is not surprising given the importance of liquidity and price discreteness for this particular group of options.

IV. Concluding Remarks

Out-of-sample predictions are at the core of investors' decision-making process. In this paper, we develop a new class of Bayesian Semiparametric Scale Mixture of Uniforms (SSMU), which is subsequently used to price the short-term deep out-of-the-money index call options. Recent research in this area has documented significant pricing errors for these types of options for a variety of reasons; some of these errors are due to the inherent difficulty in modeling volatility, and some are due to the particular class of models used in the analysis. The empirical results in this paper show that mispricing can be reduced significantly by introducing a nonparametric component into the modeling process that accounts for all ranges of kurtosis. Subsequently, the pricing errors in many cases are at least 50% lower than those obtained using comparable alternatives. These results seem

to suggest that a significant part of the pricing errors using other models might be due to the somewhat restrictive assumptions imposed on the underlying asset in a purely parametric framework. An extension of the work presented in this paper would involve modeling skewness also as part of the scale mixture representation. Adding this feature leads to some new mathematical challenges, and will be reported in a subsequent paper.

AppendixA. Proofs

Proof of Proposition 1.

(i) follows from $E(\tau) = 0$ and the fact that τ is independent of u. The symmetry follows naturally from the properties of the uniform distribution.

(ii) $Var(y) = E(y^2) - [E(y)]^2 = E[(\sum_{j=1}^{k} x_{ij}\beta_j)^2 + 2\sum_{j=1}^{k} x_{ij}\beta_j\tau\sqrt{u} + \tau^2 u] - (\sum_{j=1}^{k} x_{ij}\beta_j)^2 =$

$= (\sum_{j=1}^{k} x_{ij}\beta_j))^2 + 2\sum_{j=1}^{k} x_{ij}\beta_j E(\tau\sqrt{u}) + E(\tau^2 u) - (\sum_{j=1}^{k} x_{ij}\beta_j)^2 = 2\sum_{j=1}^{k} x_{ij}\beta_j E(\tau)E(\sqrt{u}) +$

$+ E(\tau^2 u) = E(\tau^2)E(u) = \frac{1}{12}(1+1)^2 3 \ \sigma^2 = \sigma^2$ where the last line follows from the independence of τ and u, $E(\tau) = 0$, and the fact that for $y \sim U(a,b) \implies Var(y) = \frac{1}{12}(b-a)^2$.

Proof of Proposition 2.

The expected value follows from the fact that $y \sim U(a,b) \implies E(y) = \frac{1}{2}(a+b)$. Thus $E(y) = 2\mu/2 = \mu$. Variance follows from the formula provided in proof of Proposition 1, and the fact that $\tilde{u}$ is a random variable. Hence,

$$Var(y) = \frac{1}{12}(\mu + \sigma\sqrt{u} - \mu + \sigma\sqrt{u})^2 = \frac{1}{12}E(4\sigma^2 u) = \sigma^2 E(u/3).$$

Finally, kurtosis is calculated using the equation

$$\text{kurtosis}(y) = \frac{E(y - E(y))^4}{[Var(y)]^2} - 3.$$

$E[(y - E(y))^4] = E[(y - \mu)^4] = \int_{\mu - \sigma\sqrt{u}}^{\mu + \sigma\sqrt{u}} \frac{1}{(\mu + \sigma\sqrt{u} - \mu + \sigma\sqrt{u})} (y - \mu)^4 dy = \frac{1}{2\sigma\sqrt{u}}(\frac{1}{5}(y - \mu)^5)_{\mu - \sigma\sqrt{u}}^{\mu + \sigma\sqrt{u}} =$

$= \frac{1}{10\sigma\sqrt{u}}[(\mu + \sigma\sqrt{u} - \mu)^5 - (\mu - \sigma\sqrt{u} - \mu)^5] = \frac{2}{10\sigma\sqrt{u}}\sigma^5(\sqrt{u})^5 = \frac{1}{5}\sigma^4[E(u^2)]$. Also from the previous point,

$$[Var(y)]^2 = [\sigma^2(E(u)/3)]^2 = \frac{1}{9}\sigma^4[E(u)]^2.$$

Substituting into the equation for kurtosis and simplifying gives the result.

AppendixB. Semiparametric estimation

In this section, we describe the details of the estimation of the SSMU model. The Gibbs sampler successively samples from full conditional distributions. We now describe these conditional distributions.

    1 $p(u_i|\cdots)$. The full conditional here is given by

$$u_i^{-1/2}\, I(u_i > y_i^2/\sigma_{i-1}^2)\, p(u_i|u_{-i}).$$

    Now

$$p(u_i|u_{-i}) \propto c\, f(u_i) + \sum_{j\neq i}\delta_{u_j}(u_i),$$

where δ_u is the point mass 1 at u. Recall that $f(u) \propto u^{1/2}\exp(-u/2)$. Consequently, we either sample u_i from a truncated exponential distribution or take u_i to be u_j, for those $u_j > y_i^2/\sigma_{i-1}^2$, according to probabilities which are straightforward to compute. In fact,

$$u_i \begin{cases} \sim f^*(u) & \text{with probability } \propto \tau\exp(-a/2) \\ = u_j & \text{with probability } \propto 1/u_j \end{cases}$$

where $\tau = c/(4\sqrt{\pi})$, $a = y_i^2/\sigma_{i-1}^2$ and $f^*(u) = 0.5\exp\{(u-a)/2\}I(u>a)$.

    2 $p(\beta_k|\cdots)$. Define

$$\lambda_t = \frac{0.5\ln(Y_t^2)}{u_t} - \sum_{j\neq k}\beta_j Z_{jt} \tag{B1}$$

and $\pi(.)$ to be a prior distribution function for β, so

$$p(\beta_k|\ldots) \propto \pi(\beta_k)\ I\left(\beta_k \in \left[\max_{Z_{kt}>0}\{\lambda_t/Z_{kt}\}, \min_{Z_{kt}<0}\{\lambda_t/Z_{kt}\}\right]\right). \tag{B2}$$

If $Z_{kt} > 0$ for all t then

$$\max_{Z_{kt}>0}\{\lambda_t/Z_{kt}\} = -\infty \tag{B3}$$

31

and if $Z_{kt} < 0$ for all t then

$$\min_{Z_{kt}<0} \{\lambda_t/Z_{kt}\} = \infty. \tag{B4}$$

3 $p(c|\cdots)$. The sampling for c proceeds as follows. In the first step, we sample from the beta distribution for the new latent parameter $\eta \in (0,1)$;

$$(\eta \,|c, k) \sim \text{beta} \,(c+1, N). \tag{B5}$$

Then c is sampled from the mixture of gamma distributions, where the weights are defined as below,

$$(c \mid \eta, k) \sim \pi_\eta \, Gamma(k + a, b - \ln \eta) + (1 - \pi_\eta) \, Gamma(a + k - 1, b - \ln \eta). \tag{B6}$$

Here $Gamma(a, b)$ is the prior distribution for c with $a = b = 0.01$; and π_η is the solution of the equation

$$\pi_\eta/(1 - \pi_\eta) = (a + k - 1)/N(b - \ln(\eta)). \tag{B7}$$

AppendixC. Predictive distribution

In order to construct the predictive distribution of terminal index price we extend the Gibbs sampler outlined in Appendix B. In particular, we have to predict the value of the dependent variable Y until a fixed point of time T, which in our case depends on the option's maturity. In our case Y denotes the log return on the S&P 500 index. For the predictions for period $N + 1$, we would sample the following components

$$Y_{N+1} \sim \mathrm{Un}(-\sigma_N \sqrt{u_{N+1}}, +\sigma_N \sqrt{u_{N+1}}) \tag{C1}$$

where σ_N is defined as

$$\sigma_N = e^{\beta Z_N} \tag{C2}$$

and Z_N is the value of the covariate at the time of pricing the option while u_{N+1} is sampled as below,

$$u_{N+1} \begin{cases} \sim f(u) & \text{with probability} \propto c \\ = u_j & \text{with probability } \propto 1 \end{cases}$$

Here $f(u)$ is $Gamma(1.5, 0.5)$. A Y_{N+1} can be obtained from each iteration of the Gibbs sampler, using the current (c, β). The following algorithm can be extended in an obvious way until obtaining Y_T. The predicted value of Y_T is then obtained by taking a simple average of the $Y_T s$ from each iteration of the Gibbs sampler.

References

Aït-Sahalia, Yacine, and Andrew W. Lo, 1998, Nonparametric estimation of state-price densities implicit in financial assets, *Journal of Finance* 53, 499–547.

Amin, Kaushik I., and Victor K. Ng, 1994, A comparison of predictable volatility models using option data, Research Department Working Paper, International Monetary Fund.

Antoniak, Charles E., 1974, Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *Annals of Statistics* 2, 1152–1174.

Backus, David, Silverio Foresi, Kai Li, and Liuren Wu, 1997, Accounting for biases in Black-Scholes, unpublished working paper.

Baillie, Richard T., and Tim Bollerslev, 1989, The message in daily exchange rates: A conditional variance tale, *Journal of Business and Economic Statistics* 7, 297–305.

Baillie, Richard T., and Ramon P. DeGennaro, 1990, The impact of delivery terms on stock-return volatility, *Journal of Financial Services Research* 3, 55–76.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003–2049.

Bates, David S., 1996, *Testing option pricing models in: Handbook of Statistics Vol. 14 (eds: G.S. Maddala and C.R. Rao).* (Elsevier Science B.V. New York).

Bauwens, Luc, and Michel Lubrano, 1998, Bayesian inference on GARCH models using the Gibbs sampler, *Econometrics Journal* 1, 23–46.

Black, Fisher, and Myron S. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–659.

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.

Bollerslev, Tim, 1987, A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* 69, 542–547.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner, 1992, ARCH modeling in finance, *Journal of Econometrics* 52, 5–59.

Box, George, and George Tiao, 1973, *Bayesian Inference in Statistical Analysis*. (Addison-Wesley New York).

Chernov, Mikhail, and Eric Ghysels, 2000, A study toward a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation, *Journal of Financial Economics* 56, 407–458.

Das, Sanjiv R., and Rangarajan K. Sundaram, 1999, Of smiles and smirks: a term structure perspective, *Journal of Financial and Quantitative Studies* 34, 211–239.

Derman, Emanuel, and Iraj Kani, 1994, Riding on the smile, *Risk* 7, 32–39.

Drost, Feike C., Theo E. Nijman, and Burt J.M. Werker, 1998, Estimation and testing in models containing both jumps and conditional heteroskedasticity, *Journal of Business and Economic Statistics* 16, 237–243.

Duan, Jin Chuan, 1995, The GARCH option pricing model, *Mathematical Finance* 5, 13–32.

Duan, Jin Chuan, 2002, Nonparametric option pricing by transformation, unpublished working paper.

Dumas, Bernard, Jeff Fleming, and Robert E. Whaley, 1998, Implied volatility functions: empirical tests, *Journal of Finance* 53, 2059–2106.

Dupire, Bruno, 1994, Pricing with a smile, *Risk* 7, 18–20.

Engle, Robert F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987–1008.

Engle, Robert F., and Chowdhury Mustafa, 1992, Implied ARCH models from options prices, *Journal of Econometrics* 52, 289–311.

Eraker, Bjorn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *Journal of Finance* forthcoming.

Escobar, Michael D., and Mike West, 1995, Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* 90, 577–588.

Fama, Eugene F., 1965, The behavior of stock-market prices, *Journal of Business* 38, 34–105.

Feller, William, 1971, *An introduction to probability theory and its applications, Vol. II.* (John Wiley and Sons New York).

Ferguson, Thomas S., 1973, A Bayesian analysis of some nonparametric problems, *Annals of Statistics* 1, 209–230.

Fleming, Jeff, Barbara Ostdiek, and Robert E. Whaley, 1996, Trading costs and the relative rates of price discovery in the stock, futures, and option markets, *Journal of Futures Markets* 16, 353–387.

Ghysels, Eric, Andrew C. Harvey, and Eric Renault, 1996, *Stochastic volatility in: Handbook of Statistics Vol. 14 (eds: G.S. Maddala and C.R. Rao).* (Elsevier Science B.V. New York).

Guidolin, Massimo, and Allan Timmermann, 2003, Option prices under Bayesian learning: implied volatility dynamics and predictive densities, *Journal of Economic Dynamics and Statistics* forthcoming.

Heston, Steven L., and Saikat Nandi, 2000, A closed-form GARCH option valuation model, *Review of Financial Studies* 13, 585–625.

Hsieh, David A., 1989, Modeling heteroskedasticity in daily foreign exchange rate changes, *Journal of Business and Economic Statistics* 7, 307–317.

Jorion, Phillippe, 1988, On jump processes in the foreign exchange and stock markets, *Review of Financial Studies* 1, 427–445.

MacEachern, Steven N., 1998, *Computational methods for mixture of Dirichlet process models in: Practical Nonparametric and Semiparametric Bayesian Statistics (eds: Dipak Deys, Peter Muller and Debajyoti Sinha).* (Springer New York).

Mandelbrot, Benoit, 1963, The variation of certain speculative prices, *Journal of Business* 36, 394–419.

Mira, Antonietta, Jesper Moller, and Gareth O. Roberts, 2001, Perfect slice samplers, *Journal of the Royal Statistical Society, Series B* 63, 593–606.

Myers, Ron J., and Stuart D. Hanson, 1993, Pricing commodity options when the underlying futures price exhibits time-varying volatility, *American Journal of Agricultural Economics* 75, 121–130.

Nandi, Saikat M., 1998, How important is the correlation between returns and volatility in a stochastic volatility model? Empirical evidence from pricing and hedging in the S&P 500 index options market, *Journal of Banking and Finance* 22, 589–610.

Nelson, Daniel B., 1990, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.

Ritchken, Peter, and Rob Trevor, 1999, Pricing options under generalized GARCH and stochastic volatility processes, *Journal of Finance* 54, 377–402.

Rubinstein, Mark, 1994, Implied binomial trees, *Journal of Finance* 69, 771–818.

Smith, Adrian F.M., and Gareth O. Roberts, 1993, Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion), *Journal of the Royal Statistical Society, Series B* 55, 3–24.

Stephan, Jens A., and Robert E. Whaley, 1990, Intraday price change and trading volume relations in the stock and stock options markets, *Journal of Finance* 45, 191–220.

Stutzer, Michael, 1996, A simple nonparamteric approach to derivative security valuation, *Journal of Finance* 51, 1633–1652.

Sundaresan, Suresh M., 2000, Continuous-time methods in finance: a review and an assessment, *Journal of Finance* 55, 1569–1622.

Walker, Stephen G., Paul Damien, Purushottam W. Laud, and Adrian F.M. Smith, 1999, Bayesian nonparametric inference for random distributions and related functions (with discussion), *Journal of the Royal Statistical Society, Series B* 61, 485–527.
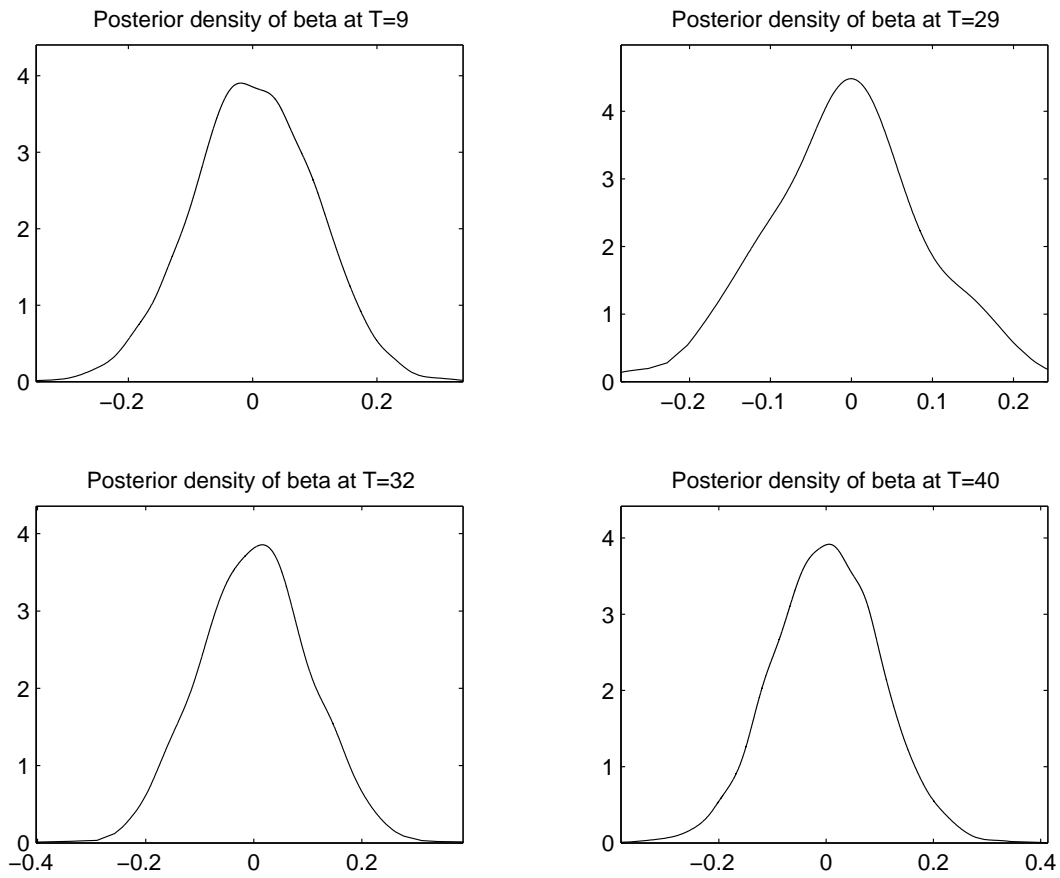
Figure 1. Posterior densities of the coefficient beta This figure depicts the posterior kernel densities of the coefficient beta for four different contracts in the sample with respective maturities of 9, 29, 32, and 40 days. Beta represents the coefficient beta in the variance regression for the option with appropriate spot index prices, and times to maturity.
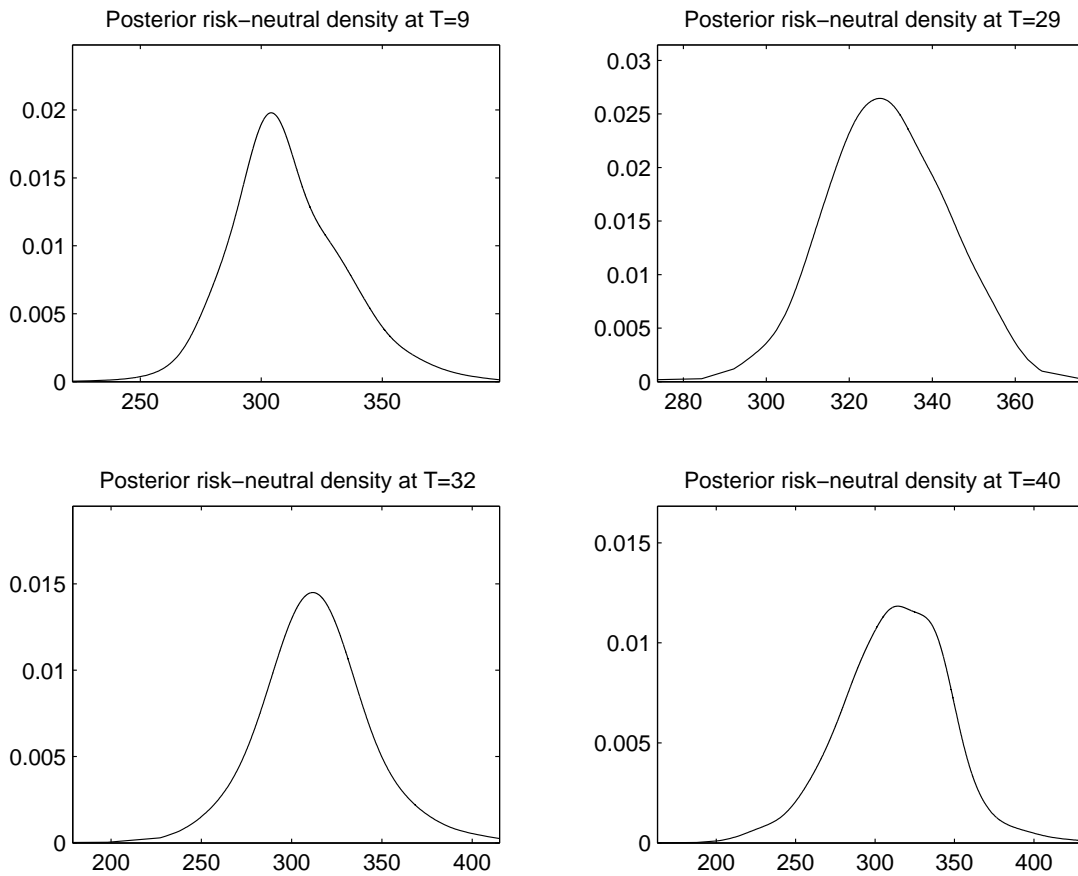
38

Figure 2. Posterior risk-neutral predictive densities of the terminal S&P 500 index
This figure depicts the posterior kernel risk-neutral densities of the *terminal* S&P 500 index. The predictions have been obtained for four different contracts with respective maturities of 9, 29, 32, and 40 days.

Table I
Sample properties of S&P 500 Index Options

This table reports the summary of the data used in the study. The cross-section of the call options has been divided into 18 categories: with respect to expiration date (≤ 40 days; $< 40, 180$) days, and > 180 days) and moneyness (out of the money (OTM); at the money (ATM), and in the money (ITM)). The sample covers the period of 01/01/1990-12/31/1995. Options with maturity less than 6 days, price lower than 0.375 and those violating arbitrage conditions have been excluded from the sample.

Panel A: Full Sample

Moneyness		Days-to-Expiration			
	K/S	≤ 40	41-180	> 180	Subtotal
ITM	< 0.94	$51.69	$47.53	$57.83	
		(4933)	(9086)	(3857)	(17876)
	0.94-0.97	$20.45	$26.55	$37.87	
		(3275)	(7841)	(2763)	(13879)
ATM	0.97-0.99	$10.24	$17.14	$29.79	
		(4220)	(8582)	(2791)	(15593)
	0.99-1.02	$3.16	$9.17	$20.58	
		(4151)	(8754)	(2592)	(15497)
OTM	1.02-1.05	$1.22	$4.68	$14.98	
		(1536)	(5563)	(1598)	(8697)
	≥ 1.05	$0.91	$2.82	$7.43	
		(583)	(7703)	(4942)	(13228)
Subtotal		(18698)	(47529)	(18543)	(84770)

Panel B: Short-term OTM Options (K/S$\geq$1.05)

	Maturity	Full Sample	Estimation Sample
	6-25	$0.67	$0.61
		(102)	(31)
	26-30	$0.75	$0.62
		(65)	(31)
	31-35	$0.96	$0.85
		(147)	(50)
	36-40	$1.01	$1.03
		(269)	(141)
	Total	(583)	(253)

Table II
Out-of-sample pricing errors of the short-term out-of-the-money S&P 500 Index Options

This table reports the average out-of-sample pricing errors for 253 out-of-the-money (OTM) short-term (≤ 40 days) call options with moneyness ≥ 1.05 during the period $1991 - 1995$. The mean absolute dollar errors (MAE), root mean squared errors (RMSE), and percentage errors have been calculated for the Black-Scholes (BS), *ad hoc* Black-Scholes, and Semiparametric Scale Mixture of Uniforms (SSMU)models defined in the paper. MAE has been calculated as a mean absolute difference between the model implied price and the observed price. RMSE has been calculated as a root of the average mean squared error, while the percentage price error further scales the RMSE by the average price of the option.

Model	Average Observed Price	MAE	RMSE	Percentage error
BS	$0.84	$1.39	$1.33	158.05%
Ad hoc BS	$0.84	$0.76	$1.00	119.04%
SSMU	$0.84	$0.39	$0.49	58.97%

Table III
Out-of-sample pricing errors with respect to time to maturity

This table reports the average out-of-sample pricing errors for 253 out-of-the-money (OTM) short-term ($\leq$40 days) call options with moneyness $\geq$ 1.05 during the period $1991 - 1995$. The options have been divided into four groups with respect to maturity time: 1) options with at most 25; 2) 26-30; 3) 31-35; and 4) 36-40 days to maturity. The mean absolute dollar errors (MAE), root mean squared errors (RMSE), and percentage errors have been calculated for *ad hoc* Black-Scholes (ABS), and Semiparametric Scale Mixture of Uniforms (SSMU) models as defined in the paper. MAE has been calculated as a mean absolute difference between the model implied price and the observed price in each maturity category. RMSE has been calculated as a root of the average mean squared error in each maturity category, while the percentage price error further scales the RMSE by the average price of the option in each maturity category.

Model	Maturity	Average Observed Price	MAE	RMSE	Percentage error
ABS	≤ 25	$0.61	$0.50	$0.62	101.56%
SSMU			$0.37	$0.48	78.27%
ABS	$< 26; 30 >$	$0.62	$0.92	$1.12	181.88%
SSMU			$0.29	$0.36	59.03%
ABS	$< 31; 35 >$	$0.85	$0.81	$1.06	124.51%
SSMU			$0.42	$0.54	63.36%
ABS	$< 36; 40 >$	$1.03	$0.79	$0.99	96.09%
SSMU			$0.44	$0.53	51.93%