Computational Statistics and Data Analysis 76 (2014) 588–605

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bayesian option pricing using mixed normal heteroskedasticity models

Jeroen V.K. Rombouts[a,*], Lars Stentoft[b,c,d]

[a] ESSEC Business School, Avenue B. Hirsch, 95021 Cergy-Pontoise Cedex, France
[b] Department of Finance at CBS, Denmark
[c] CIRANO and CIRPÉE, Canada
[d] CREATES, Denmark

ARTICLE INFO

Article history:
Received 9 February 2012
Received in revised form 28 June 2013
Accepted 28 June 2013
Available online 10 July 2013

Keywords:
Bayesian inference
Option pricing
Finite mixture models

ABSTRACT

Option pricing using mixed normal heteroscedasticity models is considered. It is explained how to perform inference and price options in a Bayesian framework. The approach allows to easily compute risk neutral predictive price densities which take into account parameter uncertainty. In an application to the S&P 500 index, classical and Bayesian inference is performed on the mixture model using the available return data. Comparing the ML estimates and posterior moments small differences are found. When pricing a rich sample of options on the index, both methods yield similar pricing errors measured in dollar and implied standard deviation losses, and it turns out that the impact of parameter uncertainty is minor. Therefore, when it comes to option pricing where large amounts of data are available, the choice of the inference method is unimportant. The results are robust to different specifications of the variance dynamics but show however that there might be scope for using Bayesian methods when considerably less data is available for inference.

1. Introduction

The Black–Scholes–Merton (BSM) model of Black and Scholes (1973) and Merton (1973) remains a standard tool for practitioners. However, when it comes to option pricing many attempts have been made on extending the model to obtain a better fit to actually observed prices. In particular, several studies have shown that the BSM model, which assumes constant volatility and Gaussian returns, severely underprices out of the money options, particularly those with short maturity. In terms of implied volatilities, this leads to the well-known smile or smirk across moneyness, which is found to be particularly pronounced for index options. Intuitively, such findings can be the result of either non constant volatility or of non-Gaussian returns, or a combination of the two.

In the existing literature extensions to the BSM model have been developed within the continuous time stochastic volatility (SV) framework and within the discrete time generalized autoregressive conditional heteroskedasticity (GARCH) framework. While important advances have been made there is still much room for improvements as discussed in the reviews by Bates (2003) and Garcia et al. (2010). In particular, the existing research has shown that there are large differences between the conditional distribution of the underlying asset and the distribution implied from option pricing. One difference is that the volatility implied by at the money options is significantly different from that observed over the life of the option. Much more importantly is the finding that the implied volatility curve, that is the implied volatility plotted against moneyness, is not only asymmetric but also changes through time. The first of these findings implies substantial negative

* Corresponding author. Tel.: +33 134433000; fax: +33 134433001.
E-mail address: rombouts@essec.edu (J.V.K. Rombouts).

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2013.06.023

skewness, more so than is often found in the underlying process, whereas the latter indicates that moments of higher order are time varying.

In this paper, we price options using asymmetric heteroskedastic normal mixture models. The type of finite mixture model we use is flexible enough to approximate arbitrarily well any kind of conditional distribution, for example highly skewed and leptokurtic, and to allow for stochastic volatility of the returns on the underlying asset of the option contract. Finite mixture models, which are convex combinations of densities, are becoming a standard tool in financial econometrics. They are attractive because of the parsimonious flexibility they provide in the specification of the distribution of the underlying random variable, which gives them a semiparametric flavor. Recent examples are Wong and Li (2000) and Wong and Li (2001) who model the conditional return distribution, extended by Haas et al. (2004) with an application of value at risk prediction, and Bauwens and Rombouts (2007a) for the clustering of financial time series (see also Ausin and Galeano, 2007 and Giannikis et al., 2008 for alternative models). Rombouts and Stentoft (2010) use an asymmetric heteroskedastic normal mixture model to price S&P 500 index options and document substantial improvements compared to several benchmark models in terms of dollar losses and the ability to explain the smirk in implied volatilities.

The are two main contributions of this paper. First of all, we explain how to perform inference and price options in a Bayesian framework. We compute posterior moments of the model parameters by sampling from the posterior density. This is possible thanks to data augmentation, a technique that includes latent variables in the parameter space, so that they also can be drawn using the Bayesian sampler. Option prices are computed using the risk neutral predictive price densities that are easily obtained as a by-product of the Bayesian sampler. This method takes into account parameter uncertainty, because the sampler integrates over the entire parameter space. This is unlike classical inference that almost always conditions on maximum likelihood estimates. In fact, for large data sets this corresponds to using only the mode of the posterior density.

As a second contribution, we compare Bayesian inference with classical inference in terms of parameter estimation and in terms of option pricing performance. Our application is on the S&P 500 index and index options, which are by far the most studied in the literature. In fact, this index is studied by, among others, Christoffersen and Jacobs (2004), Heston and Nandi (2000), Hsieh and Ritchken (2005), and Christoffersen et al. (2008) using Gaussian GARCH models, by Christoffersen et al. (2010b) and Christoffersen et al. (2006) using non-Gaussian models, and by Christoffersen et al. (2012) using GARCH models with jumps. We first of all compare the ML estimates and the posterior moments when fitting the model on S&P 500 index returns available from 1962 to 2011. Given the large amount of data, as expected we find small differences. Next, we price a rich sample of options between 1996 and 2011 using the simulated price densities based on the ML estimates and the predictive densities resulting from the posterior sampler. In terms of dollar losses, the two methods lead to similar overall pricing errors, 1.071 and 0.942 for the bias and 7.491 and 7.467 for the root mean squared error, respectively. Comparing both methods using implied standard deviation losses yields virtually equal performance. Thus, it turns out that the impact of parameter uncertainty for option pricing is minor even in this rather complex model with ten parameters.

As a robustness check we first of all consider two alternative specifications for the conditional volatility. Here our results show that our conclusions hold true for these alternative choices of volatility dynamics both in terms of inference and in terms of option pricing performance. Moreover, we find that no single choice of conditional volatility model produces the smallest losses across both maturity and moneyness. Secondly, we consider the robustness of our results to changing sample size and consider two shorter samples. Here our results show that the two methods differ only for the shortest sample for which the Bayesian framework yields very different losses. However, though the losses are smaller for this sample of options when using the Bayesian framework, the overall losses are still smaller when a larger sample is used. Thus, our results indicate that parameter uncertainty could potentially play a role when pricing options on underlyings for which considerably less data is available.

Note that the comparison between Bayesian inference and classical inference is possible since the likelihood of the mixture model in this paper can still be evaluated without simulation because there is no path dependence problem. In fact, a mixture model can be thought of as a model with latent but independent states, which makes it easy to integrate them out of the likelihood. For Markov switching volatility models, that is models with latent but dependent states, maximum likelihood becomes infeasible for most financial time series. In this situation, Bayesian inference is the only alternative, see for example the Markov switching GARCH model of Bauwens et al. (2010) and Bauwens et al. (2011). A Markov switching GARCH model is a flexible volatility model while the finite mixture model used here is a flexible distribution that for example explicitly can handle skewness. Furthermore, by adding components with constant variance to the mixture, one is able to replicate jump like behavior sometimes observed in returns data. Although Markov switching GARCH models are promising when fitted to financial return data, their application to option pricing remains quite limited. The reason is that in the general model the state variable is unobserved and proper risk neutralization difficult unless additional assumptions are made. For example, in Elliot et al. (2006) this difficulty is circumvented by using an observable Markov chain, in Satoyoshi and Mitsui (2011) agents are assumed to be risk neutral, and in Chen and Hung (2010) the model is calibrated to options directly. Furthermore, the latter two papers consider very restricted Markov switching type models. Nevertheless, an empirical comparison of both types of models is an interesting topic for further research.

The rest of the paper is organized as follows: Section 2 presents the asymmetric heteroskedastic normal mixture model and derives the risk neutral dynamics needed for option pricing. Section 3, explains how to conduct Bayesian inference, compute predictive price densities taking into account parameter uncertainty, and compares ML estimates to posterior moments for the S&P 500 index returns. Section 4, reports the results on the empirical application to options on the S&P 500 index and in Section 5 we perform several robustness checks. Finally, Section 6 concludes.

2. The framework

Letting $\mathcal{F}_t$ denote the information set up to time t, we assume that the underlying return process $R_t \equiv \ln(S_t/S_{t-1})$, where S_t is the index level on day t, can be characterized by

$$R_t = r_t - \Psi_t(\nu_t - 1) + \Psi_t(\nu_t) + \varepsilon_t, \tag{1}$$

where r_t is the risk free rate, ν_t can be interpreted as the unit risk premium, and Ψ_t is the conditional cumulant generating function of ε_t. Its conditional distribution is given by a combination of K distributions

$$P(\varepsilon_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \pi_k \Phi\left(\frac{\varepsilon_t - \mu_k}{\sigma_{k,t}}\right), \tag{2}$$

where

$$\sigma_{k,t}^2 = \omega_k + \alpha_k(\varepsilon_{t-1} + \gamma_k\sigma_{k,t-1})^2 + \beta_k\sigma_{k,t-1}^2, \tag{3}$$

with $\omega_k > 0$, $\alpha_k \geq 0$ and $\beta_k \geq 0$, and $\Phi(\cdot)$ is the standard Gaussian distribution. For this model, the conditional cumulant generating function $\Psi_t(u)$ is given by

$$\Psi_t(u) = \ln\left(\sum_{k=1}^{K} \pi_i \exp\left(-u\mu_k + \frac{u^2\sigma_{k,t}^2}{2}\right)\right), \tag{4}$$

which is just the logarithm of a convex combination of Gaussian moment generating functions. We denote this the asymmetric heteroskedastic normal mixture (MN-NGARCH) model, since the conditional variance specification in (3) corresponds to the well known NGARCH model of Engle and Ng (1993). The MN-GARCH model developed by Haas et al. (2004) corresponds to the case when $\gamma = 0$ in (3). We note that from an econometric point of view the mixture model is attractive since a common innovation term ε_t feeds in the K conditional variance equations, which circumvents the well known path dependence problem.

In the MN-NGARCH model at each t the innovation ε_t is drawn from one of the K conditional distributions with probabilities $\pi_1, \ldots, \pi_K$. Consequently, the parameter π_k is restricted to be positive for all k and $\sum_{k=1}^{K} \pi_k = 1$, which is imposed by setting $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$. Since ε_t is an innovation its zero mean restriction is easily ensured by

$$\mu_K = -\sum_{k=1}^{K-1} \frac{\pi_k\mu_k}{\pi_K}. \tag{5}$$

Note that this zero mean restriction does not imply a symmetric distribution as this would only happen if all μ_k's are zero.

It should be noted that the parameters of the mixture model are not identified as such because of the label switching problem which leaves the model likelihood unchanged when we change the order of the distributions in the finite mixture. This is not a problem if the objects of interest are label invariant, an example would be the predictive density of future returns. However, if we want to interpret the parameters, like in this paper, we add an identification restriction like $\pi_1 \geq \pi_2 \geq \cdots > \pi_K$.

In the MN-NGARCH model markets are incomplete and hence there is no unique way to derive the equivalent martingale measure (EMM) needed for option pricing. In this paper, we follow the approach of Christoffersen et al. (2010a) in which a candidate EMM is specified from the following Radon–Nikodym derivative

$$\left.\frac{dQ}{dP}\right| \mathcal{F}_t = \exp\left(-\sum_{i=1}^{t} (\nu_i\varepsilon_i + \Psi_i(\nu_i))\right), \tag{6}$$

where $\Psi_t(u)$ is the conditional cumulant generating function from above. It can be shown that under the risk neutral measure the conditional cumulant generating function of ε_t is given by

$$\Psi_t^Q(u) = \Psi_t(\nu_t + u) - \Psi_t(\nu_t). \tag{7}$$

For the asymmetric heteroskedastic normal mixture model this becomes

$$\Psi_t^Q(u) = \ln\left(\sum_{k=1}^{K} \pi_i^* \exp\left(-u\mu_k^* + \frac{u^2\sigma_{k,t}^2}{2}\right)\right), \tag{8}$$

where

$$\mu_{k,t}^* = \mu_k - \nu_t\sigma_{k,t}^2, \tag{9}$$

and

$$\pi_{k,t}^{*} = \frac{\pi_k \exp\left(-\nu_t \mu_k + \frac{\nu_t^2 \sigma_{k,t}^2}{2}\right)}{\sum_{k=1}^{K} \pi_k \exp\left(-\nu_t \mu_k + \frac{\nu_t^2 \sigma_{k,t}^2}{2}\right)}, \tag{10}$$

for $k = 1, \ldots, K$. Thus, comparing (4) and (8) the risk neutral distribution of ε_t remains within the family of normal mixtures. Note that the method above assumes that traders know the true parameter values. Therefore, while we assume that traders know the parameter values, econometricians do not know them in Bayesian inference.

In general, the distribution of ε_t will have changed means and probabilities under Q. In particular, with respect to the risk neutral means, it is immediately seen from (9) that the correction is very similar to what is obtained with a Gaussian model, where the mean of ε_t under Q is equal to $-\nu_t \sigma_t^2$. The intuition behind this is that if volatility risk carries a positive premium, then in the risk neutral world the mean of the innovations is shifted downwards to compensate. With respect to the risk neutral probabilities the relationship is less straightforward. However, for the special case where $K = 2$ it can be shown that in general $\pi_t^* < \pi$. Empirically it is often found that the least probable state is driven by an explosive volatility process. Thus, if volatility risk carries a positive premium then the probability attributed to this state, that is $1 - \pi_t^*$, is increased in the risk neutral world to compensate appropriately.

Using the stochastic discount factor principle of Gourieroux and Monfort (2007) similar results are derived by Bertholon et al. (2006) in the particular restricted case with $K = 2$ and for the GARCH specification only. For a discussion of the relationship between the two probability measures P and Q and the corresponding stochastic discount factor see Bertholon et al. (2008). One can also work within a general equilibrium setup as is done in Duan (1999). While this method also yields the dynamics to be used for option pricing, the specification is generally less explicit. In particular, an actual application of the method is computationally complex, see e.g. Stentoft (2008), and this approach therefore appears to be more restrictive.

3. Bayesian inference and option pricing

Bayesian inference is an approach to statistics that describes the model parameters as well as the data by probability distributions. It has become of widespread use in economics since Zellner (1971), van Dijk and Kloek (1978) and Geweke (1989a). Recent introductions to Bayesian inference are Koop (2003) and Geweke (2005), whereas the last chapter of Tsay (2005) treats inference for some particular models often used in financial econometrics. Johannes and Polson (2010) explain how to estimate equity price models, driven for example by stochastic volatility and jumps, using Markov Chain Monte Carlo (MCMC) methods.

More recently, Bayesian inference techniques have been used together with option data. For example, Jacquier and Jarrow (2000) incorporate parameter uncertainty and model error into the implementation of contingent claim models and apply their framework to the BSM model, and Martin et al. (2005) conduct Bayesian inference implicitly via observed option prices and apply their method to S&P 500 option data from 1995. In addition to this, Jones (2003) implements a SV model for both the underlying and the option market data and applies it to the S&P 100 index for which the Chicago Board Options Exchange Market Volatility Index is available. With respect to inference our paper is closest to that of Bauwens and Lubrano (2002) in which artificial option prices are calculated from a Bayesian viewpoint using Gaussian GARCH models.

In the next section, we explain the MCMC method we use to perform inference on the parameters of the asymmetric heteroskedastic normal mixture model. We then explain how to compute risk neutral predictive price densities and how these can be used to price options. In the final section, we report parameter inference results using data on the S&P 500 index. As a comparison we also perform maximum likelihood estimation.

3.1. Gibbs sampling

In order to learn about the model parameters of the asymmetric heteroskedastic mixture model and to forecast out-of-sample option prices, we need to draw from the posterior density. Unfortunately, the posterior density is too involved to sample from directly because it is nonstandard. We implement a MCMC procedure known as Gibbs sampling which is an iterative procedure to sample sequentially from the posterior distribution, see Gelfand and Smith (1990). Each iteration in the Gibbs sampler produces a draw from a Markov chain. Under regularity conditions, see for example Robert and Casella (2004), the simulated distribution converges to the posterior distribution. The Markov chain is generated by drawing iteratively from lower dimensional distributions, called blocks or complete conditional distributions, of this joint target distribution. These complete conditional distributions are easier to sample from because either they are known in closed form or approximated by a lower dimensional additional sampler.

The MN-NGARCH model for the returns considered is given by

$$\begin{aligned} R_t &= r_t - \Psi_t(\nu_t - 1) + \Psi_t(\nu_t) + \varepsilon_t \\ &= \rho_t(\nu_t) + \varepsilon_t. \end{aligned} \tag{11}$$

The likelihood of the model for T observations is given by

$$\mathcal{L}(\xi \mid R) = \prod_{t=1}^{T} \sum_{k=1}^{K} \pi_k \phi(R_t | \mu_k + \rho_t(\nu_t), \theta_k), \tag{12}$$

where ξ is the vector regrouping the parameters ν_t and π_k, μ_k, and θ_k for $k = 1, \ldots, K$, R denotes the vector of returns i.e. $R = (R_1, R_2, \ldots, R_T)'$, and $\phi(\cdot|\mu_k + \rho_t(\nu_t), \theta_k)$ denotes a normal density with mean $\mu_k + \rho_t(\nu_t)$ and variance $\sigma^2_{k,t}$ that depends on $\theta_k = (\omega_k, \gamma_k, \alpha_k, \beta_k)$.

A direct evaluation of the likelihood function is difficult for large K because it consists of a product of sums. It is this function that is maximized in the classical inference framework. To alleviate this evaluation, in the Bayesian framework, we introduce for each observation a state variable $G_t \in \{1, 2, \ldots, K\}$ that takes the value k if the observation R_t belongs to distribution k. The vector G^T contains the state variables for the T observations. We assume that the state variables are independent given the distribution probabilities. Then, the probability that G_t is equal to k is equal to π_k which can be written as

$$\varphi(G^T|\pi) = \prod_{t=1}^{T} \varphi(G_t|\pi) = \prod_{t=1}^{T} \pi_{G_t}, \tag{13}$$

where $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$. Given G^T and R, the likelihood function is

$$\mathcal{L}(\xi \mid G^T, R) = \prod_{t=1}^{T} \pi_{G_t} \phi(R_t | \mu_{G_t} + \rho_t(\nu_t), \theta_{G_t}), \tag{14}$$

which is easier to evaluate than (12) since the sum has disappeared. Since G^T is not observed, we treat it as an extra parameter of the model. This technique is called data augmentation, see Tanner and Wong (1987) and Albert and Chib (1993) for more details, and Jacquier et al. (1994) for a well known application in stochastic volatility modeling. Although the augmented model contains more parameters, the initial parameters plus the states, inference becomes easier by making use of MCMC methods as we will see next.

Since the posterior density of the mixed normal heteroskedasticity model is too involved to sample from directly, we implement a hybrid Gibbs sampling algorithm that allows to sample from the posterior distribution by sampling from its conditional posterior densities, see also Bauwens and Rombouts (2007b) for more details. The blocks of the Gibbs sampler, and the prior densities are explained next using the parameter vectors $G^T = (G_1, \ldots, G_T)'$, $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$, $\mu = (\mu_1, \mu_2, \ldots, \mu_K)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$ and the parameter ν_t. The joint posterior distribution is given by

$$\varphi(G^T, \nu_t, \mu, \theta, \pi | R) \propto \varphi(\nu_t)\, \varphi(\mu)\, \varphi(\theta)\, \varphi(\pi) \prod_{t=1}^{T} \pi_{G_t} \phi(R_t | \mu_{G_t} + \rho_t(\nu_t), \theta_{G_t}), \tag{15}$$

where $\varphi(\nu_t), \varphi(\mu), \varphi(\theta), \varphi(\pi)$ are the corresponding prior densities. Thus, we assume prior independence between ν_t, π, μ, and θ. This makes the construction of the Gibbs sampler easier. It does, however, not imply posterior independence between the parameters.

- $\varphi(G^T|\nu_t, \mu, \theta, \pi, R)$
  Given ν_t, μ, θ, π and y, the posterior density of G^T is proportional to $\mathcal{L}(\xi \mid G^T, R)$. Since the G_t's are mutually independent, we can write the relevant conditional posterior density as

  $$\varphi(G^T|\nu_t, \mu, \theta, \pi, R) = \prod_{t=1}^{T} \varphi(G_t|\nu_t, \mu, \theta, \pi, R). \tag{16}$$

  As the sequence $\{G_t\}_{t=1}^{T}$ is equivalent to a multinomial process, we simply have to sample from a discrete distribution where the K probabilities are given by

  $$P(G_t = k|\nu_t, \mu, \theta, \pi, R) = \frac{\pi_k \phi(R_t|\mu_k + \rho_t(\nu_t), \theta_k)}{\sum_{j=1}^{K} \pi_j \phi(R_t|\mu_j + \rho_t(\nu_t), \theta_j)}, \quad k = 1, \ldots, K. \tag{17}$$

  To sample G_t, we draw one observation from a uniform distribution on [0, 1] and decide which group k to take according to (17).
- $\varphi(\pi|G^T, \nu_t, \mu, \theta, R)$
  The full conditional posterior density of π depends only on G^T and is given by

  $$\varphi(\pi|G^T) \propto \varphi(\pi) \prod_{k=1}^{K} \pi_k^{x_k}, \tag{18}$$

where x_k is the number of times that $G_t = k$. The prior $\varphi(\pi)$ is chosen to be a Dirichlet distribution, $Di(a_{10}, a_{20} \cdots a_{K0})$ with parameter vector $a_0 = (a_{10}, a_{20} \cdots a_{K0})'$. As a consequence, $\varphi(\pi|G^T)$ is also a Dirichlet distribution, $Di(a_1, a_2 \cdots a_K)$ with $a_k = a_{k0} + x_k,\ k = 1, 2, \ldots, K$.

- $\varphi(\mu|G^T, \nu_t, \pi, \theta, R)$
  The conditional distribution of $\tilde{\mu} = (\mu_1, \mu_2, \ldots, \mu_{K-1})'$ is Gaussian with mean $-A^{-1}b$ and covariance matrix A^{-1} where

$$A = \text{diag}\left(\sum_{(1)} \frac{1}{\sigma_{1,t}^2}, \ldots, \sum_{(K-1)} \frac{1}{\sigma_{K-1,t}^2}\right) + \frac{\tilde{\pi}\tilde{\pi}'}{\pi_K^2} \sum_{(K)} \frac{1}{\sigma_{K,t}^2}, \tag{19}$$

and

$$b = \begin{bmatrix} \dfrac{\pi_1}{\pi_K} \displaystyle\sum_{(K)} \frac{\varepsilon_t}{\sigma_{K,t}^2} - \sum_{(1)} \frac{\varepsilon_t}{\sigma_{1,t}^2} \\ \vdots \\ \dfrac{\pi_{K-1}}{\pi_K} \displaystyle\sum_{(K)} \frac{\varepsilon_t}{\sigma_{K,t}^2} - \sum_{(K-1)} \frac{\varepsilon_t}{\sigma_{K-1,t}^2} \end{bmatrix}, \tag{20}$$

where $\tilde{\pi} = (\pi_1, \ldots, \pi_{K-1})$ and $\sum_{(k)}$ means summation over all t for which $G_t = k$. Once $\tilde{\mu}$ has been drawn, the last mean μ_K is obtained from (5).

- $\varphi(\theta|G^T, \nu_t, \mu, \pi, R)$
  By assuming prior independence between the θ_k's, that is $\varphi(\theta) = \prod_{k=1}^K \varphi(\theta_k)$, it follows that

$$\begin{aligned}\varphi(\theta|G^T, \nu_t, \mu, \pi, R) &= \varphi(\theta|G^T, \nu_t, \mu, R) \\ &= \varphi(\theta_1|\nu_t, \mu_1, \widetilde{R}^1) \times \cdots \times \varphi(\theta_K|\nu_t, \mu_K, \widetilde{R}^K),\end{aligned} \tag{21}$$

where $\widetilde{R}^k = \{R_t|G_t = k\}$ and

$$\varphi(\theta_k|\nu_t, \mu_k, \widetilde{R}^k) \propto \varphi(\theta_k) \prod_{t \in G_t = k} \phi(R_t|\mu_k + \rho_t(\nu_t), \theta_k). \tag{22}$$

Since we condition on the state variables, we can simulate each block θ_k separately. We do this with the griddy-Gibbs sampler (for further details, see Bauwens et al., 1999). We take bounded uniform supports for ω_k, γ_k, α_k, and β_k. The choice of these bounds are finely tuned in order to cover the relevant posterior parameter support. Doing so, we only impose diffuse priors. Note that instead of the griddy-Gibbs sampler one could easily implement the Metropolis Hastings sampler. Bauwens et al. (2010) compare both samplers for a GARCH type volatility process and find similar results.

- $\varphi(\nu_t|G^T, \mu, \pi, \theta, R)$
  The conditional posterior distribution for this block does not belong to a known family of distributions. If ν_t is constant, as is often the case, we can sample directly numerically by drawing one observation from a uniform distribution on [0, 1] and finding the corresponding quantile of the conditional posterior distribution of ν_t. If, on the other hand, ν_t has a specific parametric form, an additional griddy-Gibbs or Metropolis Hastings sampler can be used.

Convergence of the Gibbs sampler is checked with CUMSUM plots of the parameter draws as explained for example in Bauwens et al. (1999). To ensure a good precision of posterior moments in this paper, we take $N = 20{,}000$ and a warm up of 5000 draws to ensure convergence to the target joint posterior distribution and to eliminate the impact of the starting values on the final results.

3.2. Predictive densities and option pricing

In our application, we price European call options and we now explain how this can be done in a Bayesian framework. The theoretical value of this option at time T with maturity T^* and strike price K is given by

$$C_T(S, T^*, K) = e^{-r(T^*-T)} \int_0^\infty \max(S_{T^*} - K, 0) f^Q(S_{T^*}) dS_{T^*}, \tag{23}$$

where $f^Q(S_{T^*})$ is the density of the underlying asset price at expiration under the risk neutral density. The major problem with (23) is that, with the potential exception of models with constant volatility, no closed form expression exists for this price in general. Thus, instead a numerical method will have to be used. Our proposed method consists of two steps. The first involves sampling asset returns under the risk neutral measure. In the second step, these returns are aggregated to obtain the predicted price density and the expectation in (23) is approximated by the sample average.

The predictive density of R_{T+1} under the risk neutral measure Q is given by

$$f^Q(R_{T+1} \mid R) = \int f^Q(R_{T+1} \mid \xi, R)\, \varphi(\xi \mid R)\, \mathrm{d}\xi, \tag{24}$$

Table 1
Descriptive statistics for S&P 500 index percentage returns.

Minimum	−22.900	Standard deviation	1.034
Mean	0.025	Skewness	−1.040
Maximum	10.957	Kurtosis	30.590

This table reports descriptive statistics for S&P 500 index percentage returns using the sample period from July 2, 1962, to December 28, 2011, for a total of 12,459 observations. The autocorrelation coefficient is 0.0205 with a standard error of 0.0090 and thus not statistically significant at the one percent level.

where $f^Q(R_{T+1} \mid \xi, R) = \sum_{k=1}^{K} \pi_k^* \phi(R_{T+1}|\mu_k^* + \rho_{T+1}(\nu_t), \theta_k)$ as implied by the finite mixture distribution defined in (2), and π_k^* and μ_k^* are given by (9) and (10), respectively. Note that the predictive density is invariant to label switching. Unlike prediction in the classical framework, the predictive densities take into account parameter uncertainty by construction while integrating the predictive likelihood over the parameter space. An analytical solution to (24) is unavailable but extending the algorithm of Geweke (1989b), it can be approximated by

$$\widehat{f^Q}(R_{T+1} \mid R) = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{K}\pi_k^{*(n)}\phi\left(R_{T+1}|\mu_k^{*(n)} + \rho_{T+1}(\nu_t), \theta_k^{(n)}, R\right)\right), \tag{25}$$

where the superscript n indexes the draws generated with the Gibbs sampler and N is the number of draws. Therefore, simultaneously with the Gibbs sampler where we simulate N times $\xi^{(n)} \sim \varphi(\xi \mid R)$, we simulate $R_{T+1}^{(n)} \sim f^Q(R_{T+1} \mid \xi^{(n)}, R)$. A similar algorithm is used by Bauwens and Lubrano (2002). Extending the idea used for R_{T+1}, the predictive density for R_{T+s} under the Q measure may be written as

$$\begin{aligned} f^Q(R_{T+s} \mid R) = \int \Bigg[\iint \ldots \int & f^Q(R_{T+s} \mid R_{T+s-1}, \ldots, R_{T+1}, R, \xi) \\ & \times f^Q(R_{T+s-1} \mid R_{T+s-2}, \ldots, R_{T+1}, R, \xi) \ldots \\ & \times f^Q(R_{T+1} \mid R, \xi) \mathrm{d}R_{T+s-1}\mathrm{d}R_{T+s-2}\mathrm{d}R_{T+1}\Bigg] \varphi(\xi \mid R)\, \mathrm{d}\xi, \end{aligned} \tag{26}$$

for which draws can be obtained by extending the above algorithm to a $(s+1)$-step algorithm. The draw of R_{T+1} serves as conditioning information to draw R_{T+2}, both realizations serve to draw R_{T+3}, etc.

The predictive densities described until here are return densities. To obtain predictive option prices, we need price densities at the maturity date of the option contract. This predictive price density is obtained by aggregating the predicted returns until maturity for each of the N draws. For the European call option example in (23), we obtain

$$C_T(S, T^*, K) \approx e^{-r(T^*-T)}\frac{1}{N}\sum_{n=1}^{N}\max\left(S_T \exp\left(\sum_{i=T}^{T^*} R_i^{(n)}\right) - K, 0\right), \tag{27}$$

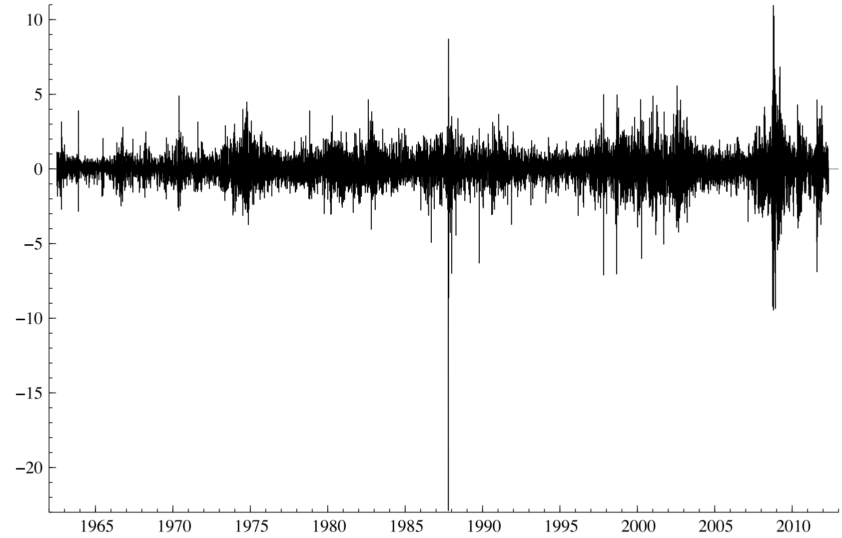
which can be evaluated rapidly. Using the Bayesian inference approach, we have the complete predictive price density at maturity. In fact, once we have the N draws from the predictive densities until maturity T^*, we can price any contract defined on the underlying S until that maturity.

In our application, the draws used for pricing purposes are generated from the finite mixture of normal densities with adjusted parameters to sample under the Q measure. A non Bayesian procedure typically proceeds by conditioning on a point estimate of ξ, and it will therefore ignore the estimation uncertainty. Thus, an extra benefit of the Bayesian framework is that it allows us to consider the issue of parameter uncertainty by comparing our results to those obtained when e.g. the means are used instead of the full set of draws.

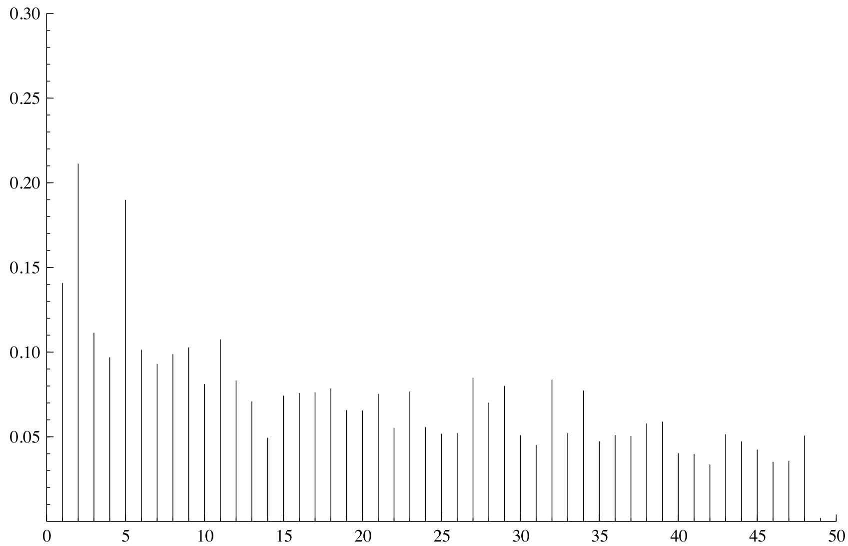
Note that instead of simple Monte Carlo integration, variance reductions techniques such as antithetic acceleration exist. However, in our application we find only small differences between these methods. A more exhaustive comparison of different variance reduction techniques is therefore left for future research.

3.3. Application to the S&P 500 index

As has become common practice in the option pricing literature, we use the S&P 500 index returns from July 2, 1962, which is the first day data from CRSP is available, to December 28, 2011, which corresponds to the last week for which we have option data, for a total of 12,459 observations. Table 1 provides the standard descriptive statistics for the S&P 500 index return series. The numbers in this table show that the return data is negatively skewed and very leptokurtic. Fig. 1(a) displays the sample path and shows the well known pattern of time varying volatility with periods of high volatility levels followed by periods of low levels of volatility. Finally, Fig. 1(b) provides evidence of the strong persistence in squared returns, a proxy of the second moment of the series. Considering the evidence of Table 1 and Fig. 1, it is clear that a model which allows for conditional heteroskedasticity and non-Gaussian features is appropriate for this type of data.



(a) Sample path.



(b) ACF of squared returns.

Fig. 1. Properties of the S&P 500 returns. This figure plots the sample path and the ACF of squared returns using the sample period from July 2, 1962, to December 28, 2011, for a total of 12,459 observations.

Until now, we have considered the general formulation of the MN-NGARCH model without fixing K, the number of distributions in the finite mixture. However, when applying the model empirically a choice has to made about K. Using one, two, and three distributions in the mixture Rombouts and Stentoft (2010) show that $K = 2$ is optimal in terms of option pricing performance. Furthermore, they show that for this data the asymmetry parameter γ_k is significant, but the risk premium parameter ν_t is insignificant. Therefore, in Table 2 we report results only for the asymmetric heteroskedastic normal mixture model with $K = 2$ and for $\nu_t = 0$. Note that fixing $\nu_t = 0$ simplifies the inference as there is one less parameter and as Eq. (7) shows in this case the risk neutral distribution is the same as the physical distribution. However, even when $\nu_t = 0$ restrictions are needed on the mean processes to ensure the appropriate risk neutral dynamics as shown in (1) (see also the discussion in Rombouts and Stentoft, 2010). Though the evaluation of the log likelihood function is potentially complicated for large K because of the "product of sums", following Haas et al. (2004) ML estimation can be performed straightforwardly for the two component mixture. Finally, to avoid the label switching problem, and hence the multi-modality of the log likelihood function, we use the parameter restriction that $\pi > 0.5$.

The first two columns of Table 2 report ML estimation results. The mixture probability π amounts to 94.7% for the mixture distribution with conditional variance process that is highly persistent, equal to $\alpha_1(1+\gamma_1^2)+\beta_1 = 0.991$, but nonexplosive. However, the second distribution in the mixture is driven by an explosive variance process, with persistence equal to 1.308.

Table 2
Results for the MN2NGARCH model.

	ML estimates		Posterior moments		Prior support
	Estimate	Std error	Mean	Std deviation	
μ	0.025	0.005	0.019	0.004	–
π	0.947	0.014	0.976	0.008	–
ω_1	0.004	0.001	0.005	0.006	[0.02;0.09]
γ_1	−0.813	0.060	−0.804	0.051	[−1.00;0.63]
α_1	0.050	0.004	0.052	0.003	[0.038;0.0625]
β_1	0.908	0.006	0.907	0.005	[0.89;0.93]
ω_2	0.167	0.084	0.297	0.013	[0.005;0.90]
γ_2	−0.206	0.163	−0.128	0.100	[−0.40;0.10]
α_2	0.500	0.164	0.794	0.212	[0.20;1.25]
β_2	0.787	0.057	0.745	0.056	[0.50;1.00]

This table reports ML estimation results and posterior moments based on 20,000 draws from the Gibbs sampler defined in Section 3.1. The data used is from July 2, 1962, to December 28, 2011, for a total of 12,459 observations. The parameters ω and μ have been multiplied by 10,000 and 100 respectively. The last column contains the θ prior supports of the griddy-Gibbs sampler. A uniform Dirichlet prior is used for π, and for the mixture mean μ the Gaussian prior is centered around zero.

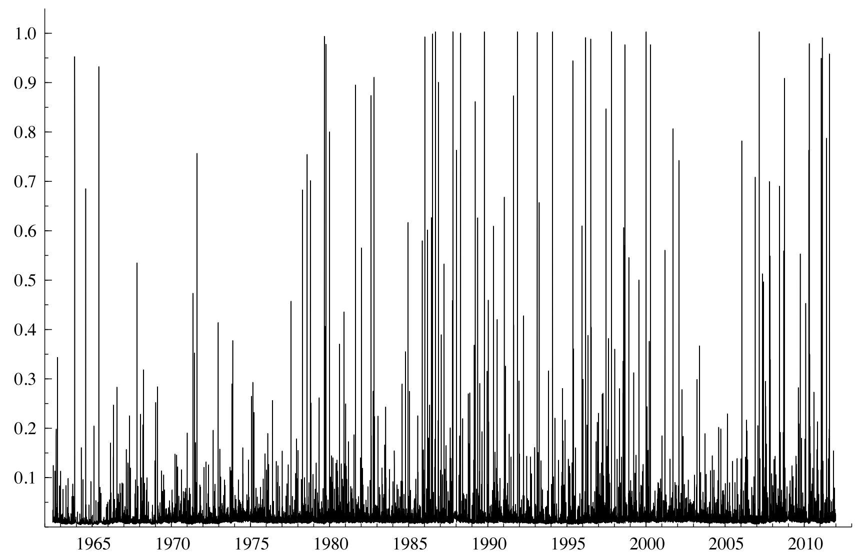


Fig. 2. Posterior probabilities to be in the second mixture distribution. This figure plots posterior probabilities to be in the second mixture distribution using the sample period from July 2, 1962, to December 28, 2011, for a total of 12,459 observations.

Though the probability to draw from this distribution is only to 5.3%, it is thanks to this volatility process that we can accommodate for example the high kurtosis in the returns of the underlying.

Columns three and four of Table 2 report the posterior moments based on 20,000 draws from the Gibbs sampler defined in Section 3.1. As mentioned above, the choice of the bounds for the variance processes, shown in the last column, are finely tuned in order to cover the relevant posterior parameter support. The posterior mean of the probability of the first mixture distribution is slightly higher, 97.6%, than the ML estimate. Though, the posterior mean of the individual parameters are similar to the ML estimates, with posterior standard deviations slightly smaller than the standard errors as expected. The posterior means for the volatility process in the second mixture distribution are somewhat different from the ML estimates except for β_2. For example, the posterior mean of the asymmetry parameter γ_2 increases to -0.128 compared to the ML estimate of -0.206, and α_2 increases to 0.794 compared to 0.500. Overall, the persistence of this volatility process increases to 1.552, though the probability to draw from this distribution is only 2.4%.

To understand better which observations are likely to be drawn from the second mixture distribution, Fig. 2 plots the corresponding posterior probabilities as estimated for each observation as $\sum_{n=1}^{N}(G_t^{(n)}-1)/N$ for $t=1,\ldots,T$. As expected, we see that small returns are less likely to be in the second mixture distribution. Also, although the latent states G_t are independent by construction, we can see episodes where it is quite less likely to be in the second mixture distribution.

To conclude, the ML and Bayesian inference results for the volatility process in the first mixture distribution are very similar. For the second mixture distribution, from which it is more likely to draw with the ML results, small differences are observed. In the next section, we analyze the implications of these differences for option pricing.

Table 3
Properties of the S&P 500 index options data set.

	VST	ST	MT	LT	VLT	ALL
DOTM	2.43	4.27	8.30	15.86	24.85	9.62
	22.47	19.79	18.46	17.79	17.00	19.07
	1375	3098	3123	2080	1159	10 835
OTM	4.69	9.67	19.47	35.22	52.79	14.80
	16.14	15.35	16.28	16.68	16.68	15.98
	2406	2609	1775	733	380	7903
ATM	15.96	25.53	39.35	57.03	76.19	31.46
	15.85	16.72	17.94	17.77	17.77	16.90
	3801	3448	2934	1119	618	11 920
ITM	42.07	49.10	62.03	78.34	93.21	58.76
	18.58	18.77	20.09	17.84	16.88	18.64
	869	722	611	405	335	2942
DITM	114.77	126.02	162.01	145.11	149.94	136.98
	21.02	19.36	18.98	18.44	16.76	19.20
	524	589	467	343	280	2203
ALL	19.16	22.56	32.49	43.62	61.02	29.91
	17.51	17.58	18.00	17.66	17.09	17.64
	8975	10 466	8910	4680	2772	35 803

Mean price in USD, mean ISD in percentage, and number of contracts in the cells of this table. The maturity categories, in columns, are divided into very short term (VST), with $T^* < 22$, short term (ST), with $22 \leq T^* < 43$, medium term (MT), with $43 \leq T^* < 85$, long term (LT), with $85 \leq T^* < 169$, and very long term (VLT), with $T^* \geq 169$. The moneyness categories, in rows, are divided into deep out of the money (DOTM), with $M < 0.95$, out of the money (OTM), with $0.95 \leq M < 0.98$, at the money (ATM), with $0.98 \leq M < 1.02$, in the money (ITM), with $1.02 \leq M < 1.05$, and deep in the money (DITM), with $M \geq 1.05$.

4. Pricing S&P 500 options

We now assess if there are differences in pricing performance for a large sample of options using either the simulated price densities based on the ML estimates in Table 2 or the predictive densities resulting from the posterior sampler. In the next section, we describe the option data in detail. We then report results on the overall pricing performance using two well known metrics, the dollar losses and the implied standard deviation losses.

4.1. Data

In this paper, we use data on call options on the S&P 500 index. The use of call options follows the literature and this approach is in line with most of the papers mentioned in the Introduction. Our data covers the period from 1996 through 2011 which is the entire period for which data is provided by OptionMetrics. We impose the following standard restrictions on our sample: Firstly, we consider weekly data only and choose the options traded on Wednesdays. If Wednesday is not a trading day we pick the date closest to it. This choice is made to balance the tradeoff between having a long time period against the computational complexity. We choose Wednesdays as these options are least affected by weekend effects. Secondly, we choose to work only with contracts which had a daily traded volume of at least 100 contracts. Thirdly, we exclude options which have an ask price below 50 cents. Fourthly, we exclude options with less than 7 or more than 252 calendar days to maturity. Finally, we eliminate options in the LEAPS series as the contract specifications for these options do not correspond to that of the standard options. In total, we end up with a sample of 35,803 call options.

Table 3 provides descriptive statistics for the options in terms of the average prices, the average implied standard deviations (ISD) from the BSM model, and the number of options. We tabulate data for various categories of maturity measured in trading days, T^*, and moneyness measured as $M = S/\left(K \exp\left(-rT^*\right)\right)$, where S is the value of the underlying, K is the strike price, and r is the risk free interest rate, which we take to be given by the available LIBOR rate. The table shows that our data contains a diverse sample of traded options. First of all, in terms of the average prices these range from 2.43 for the DOTM–VST options to 162.01 for the DITM–MT category. Next, while the number of contracts varies even the smallest category is as large as 280. Finally and most importantly, we observe the well known volatility smile, i.e. the mean ISD is higher for the DOTM and DITM options than for the ATM options, as shown in e.g. the last column of Table 3. Note that, while the volatility smile becomes less pronounced as the maturity increases, it is present for all maturities.

4.2. Pricing performance measured by dollar and ISD losses

We now proceed to compare the predicted prices from the ML and Bayesian methods. To keep the computation time reasonable, in both cases the prices are calculated using Monte Carlo simulation using the values in Table 2, and on each day of pricing the spot variance is filtered out using the available historical data and the actual value of the LIBOR rate. Thus, the option price at day T is calculated using the returns up to day T but with parameters estimated with the full sample. Note

Table 4
Dollar losses.

	VST	ST	MT	LT	VLT	ALL
	ML estimation					
DOTM	2.291	2.668	2.346	1.903	0.609	2.161
	4.381	5.954	8.108	10.894	12.021	8.352
OTM	1.265	1.269	0.547	0.465	1.830	1.058
	3.153	4.828	7.257	11.085	14.649	6.653
ATM	0.721	0.267	−0.321	−0.525	0.191	0.189
	3.058	4.939	7.888	11.064	14.396	6.898
ITM	0.326	−0.423	−0.656	1.290	4.423	0.537
	3.131	4.710	8.046	10.661	15.858	8.128
DITM	0.328	0.174	0.290	1.566	6.361	1.238
	2.912	3.921	6.049	10.587	16.134	8.026
ALL	1.046	1.175	0.796	1.020	1.725	1.071
	3.318	5.171	7.770	10.923	13.882	7.491
	Bayesian inference					
DOTM	2.273	2.607	2.220	1.673	0.222	2.019
	4.393	5.898	8.014	10.905	12.057	8.324
OTM	1.237	1.183	0.372	0.131	1.315	0.926
	3.152	4.803	7.216	11.104	14.645	6.640
ATM	0.693	0.187	−0.476	−0.831	−0.282	0.065
	3.062	4.933	7.860	11.068	14.381	6.889
ITM	0.314	−0.475	−0.760	1.000	3.975	0.409
	3.150	4.710	8.045	10.577	15.639	8.067
DITM	0.320	0.145	0.231	1.419	6.084	1.158
	2.936	3.896	6.005	10.554	15.964	7.968
ALL	1.022	1.104	0.655	0.756	1.305	0.942
	3.325	5.143	7.717	10.923	13.842	7.467

In the cells of this table, the first number is the mean error and the second number is the root mean squared error both measured in USD. The maturity (columns) and moneyness (lines) are defined in Section 4.1.

that every week contains options with short maturities, but also maturities that reach up to one year. In fact, forecasting at long horizons is an inherent part of option pricing, whether the model parameters are re-estimated or not.

Table 4 shows the mean error (BIAS), calculated as the difference between the model and the market prices, and root mean squared error (RMSE) in dollar terms for our sample of options. Overall, the two methods lead to similar pricing errors, 1.071 and 0.942 for the BIAS and 7.491 and 7.467 for the RMSE, respectively. In general, the BIAS and RMSE increase with maturity for all moneyness categories, though slightly less so for the Bayesian prices. In fact, for the DITM–VLT options the BIAS is 6.361 and 6.084 and the RMSE 16.134 and 15.964, respectively. Compared to the mean price of the DITM–VLT options the differences are only 0.18% and 0.11% though.

Another often used metric for comparing the performance of option pricing models is losses in ISD's. The option ISD is a nonlinear transformation of the dollar price, which corrects for maturity and moneyness effects. Table 5 provides these losses for our sample of options and shows that in general the results are in line with the dollar losses in Table 4. More precisely, the pricing errors are 0.972% and 0.895% for the BIAS and 4.854% and 4.846% for the RMSE for the ML and Bayesian methods respectively yielding virtually equal performance. The differences observed across maturity for the dollar losses are much less pronounced when using ISD losses. For example, the RMSE for the VLT options is now 4.064% and 4.066% respectively. Thus, it turns out that in terms of option pricing performance, especially when measured in ISD losses, parameter uncertainty is of minor importance.

Finally, note that compared to the constant volatility BSM model, which uses the volatility over the full sample as input, or even the Gaussian NGARCH model, both of which are special cases of our mixture model, the mixture model's performance is striking. When pricing all options using the two restricted models the increase in dollar loss as measured by the RMSE is substantial, up to 46% and 15% respectively compared to the mixture model. Note that this large advantage is there irrespective of the inference method used and that Bayesian and ML inference yield virtually identical results for these simpler models.

5. Robustness checks

Though the results presented above showed that using Bayesian inference does not improve on option performance, this was for one particular choice of conditional variance dynamics and for one particular sample size. In this section, we analyze the robustness of the results along these two dimensions. To be specific, we first of all compare the results to what is obtained with two alternative specifications for the dynamics of the conditional variance. Secondly, since parameter uncertainty could play a role when pricing options on underlyings for which considerably less data is available we consider two different and shorter sample sizes.

Table 5
Implied volatility losses.

	VST	ST	MT	LT	VLT	ALL
	ML estimation					
DOTM	4.050	2.805	1.670	0.937	0.179	1.996
	7.859	6.946	6.023	5.087	3.801	6.213
OTM	1.592	0.943	0.263	0.123	0.403	0.886
	4.023	3.754	3.900	3.758	3.602	3.864
ATM	0.736	0.121	0.149	0.217	0.046	0.210
	3.256	3.376	3.924	3.589	3.415	3.504
ITM	0.545	0.332	0.318	0.314	0.950	0.165
	4.607	3.772	4.073	3.367	3.793	4.050
DITM	0.729	1.235	1.554	1.753	2.706	1.450
	8.153	6.971	7.038	6.397	6.571	7.153
ALL	1.454	1.152	0.648	0.540	0.508	0.972
	4.922	5.026	4.955	4.558	4.064	4.854
	Bayesian inference					
DOTM	3.994	2.713	1.540	0.815	0.047	1.888
	7.869	6.898	5.971	5.091	3.822	6.187
OTM	1.541	0.864	0.170	0.015	0.286	0.808
	4.023	3.739	3.879	3.771	3.619	3.856
ATM	0.706	0.069	−0.218	−0.309	−0.150	0.155
	3.263	3.375	3.908	3.597	3.423	3.503
ITM	0.521	−0.374	−0.367	0.225	0.849	0.114
	4.630	3.772	4.082	3.348	3.749	4.053
DITM	0.620	1.133	1.489	1.684	2.662	1.367
	8.306	7.050	6.997	6.354	6.538	7.197
ALL	1.411	1.079	0.555	0.434	0.397	0.895
	4.943	5.010	4.923	4.557	4.066	4.846

In the cells of this table, the first number is the mean error and the second number is the root mean squared error both measured in ISD's. The maturity (columns) and moneyness (lines) are defined in Section 4.1.

5.1. Alternative volatility specifications

The option pricing framework of Christoffersen et al. (2010a) is applicable to any choice of conditional variance dynamics and not just to dynamics of the NGARCH type given in (3). Hence, as a first robustness check we consider two alternative specifications for the dynamics of the conditional variance in this section. The first of these is the GJRGARCH model of Glosten et al. (1993), in which the component variances are given by

$$\sigma_{k,t}^2 = \omega_k + \alpha_k\varepsilon_{t-1}^2 + \gamma_k\varepsilon_{t-1}^2 I(\varepsilon_{t-1} > 0) + \beta_k\sigma_{k,t-1}^2, \tag{28}$$

with $\omega_k > 0, \alpha_k \geq 0, \gamma_k \geq -\alpha_k$ and $\beta_k \geq 0$, and where $I(\cdot)$ denotes the indicator function. Note that for an easier comparison with the NGARCH and AGARCH models we use $I(\varepsilon_{t-1} > 0)$ instead of $I(\varepsilon_{t-1} < 0)$, which yields negative values for the asymmetry parameter. The second is the AGARCH model of Engle (1990), which we specify as

$$\sigma_{k,t}^2 = \omega_k + \alpha_k\varepsilon_{t-1}^2 + \gamma_k\varepsilon_{t-1} + \beta_k\sigma_{k,t-1}^2, \tag{29}$$

with $\omega_k > 0, \alpha_k \geq 0, \omega_k > -\gamma_k^2/(4\alpha_k)$ and $\beta_k \geq 0$. Like the NGARCH model in (3) both of these models allow for asymmetries through the parameter γ, something which has been shown to be of importance for the data we consider.

In Table 6, we provide results for these two alternative models. Compared to the results in Table 2 we see that although some of the actual parameter values are different the three models are very similar. In particular, the mixture probabilities are very close and so is the persistence, calculated as $\alpha + 0.5\gamma + \beta$ for the GJRGARCH model and as $\alpha + \beta$ for the AGARCH model, of the first component. This holds for the ML estimates as well as for the posterior moments. For the second component the difference in persistence is somewhat larger, though these are within 6% of that of the NGARCH model for the ML estimates and within 12% for the posterior moments. When comparing the ML estimates to the posterior moments for the two alternative models we see, as was the case for the NGARCH model, the mixture probability is slightly higher and the second mixture distribution has slightly higher values for γ_2, α_2, as well as for the persistence. Thus, in conclusion we obtain qualitatively the same results when comparing ML estimates to posterior moments for these alternative specifications of the conditional volatility dynamics.

In Tables 7 and 8, we compare the dollar and implied volatility losses using the three different specifications for the conditional volatility across maturity and moneyness and when using ML estimation as well as Bayesian inference. The first thing to notice from the table is that generally both the dollar and implied volatility losses are very similar across models, and this holds both when ML estimation and Bayesian inference is used. When using dollar losses, the AGARCH model does perform somewhat better than the two other specifications, and the reason for this is that the AGARCH model performs much better for LT and VLT options. However, the tables also show that no single choice of conditional volatility model produces the smallest errors across both moneyness and maturity.

Table 6
Results with different volatility dynamics.

	ML estimates		Posterior moments		Prior support
	Estimate	Std error	Mean	Std deviation	
Panel A: Using GJRGARCH volatility dynamics					
μ	0.026	0.006	0.021	0.005	–
π	0.934	0.019	0.968	0.010	–
ω_1	0.003	0.001	0.004	0.001	[0.002;0.006]
γ_1	−0.069	0.007	−0.068	0.003	[−0.09;−0.02]
α_1	0.087	0.007	0.088	0.003	[0.075;0.10]
β_1	0.937	0.005	0.937	0.002	[0.92;0.95]
ω_2	0.183	0.078	0.265	0.088	[0.10;0.60]
γ_2	−0.580	0.187	−0.446	0.117	[−0.70;−0.20]
α_2	0.727	0.185	0.797	0.075	[0.50;1.10]
β_2	0.793	0.050	0.783	0.031	[0.60;0.90]
Panel B: Using AGARCH volatility dynamics					
μ	0.023	0.005	0.018	0.004	–
π	0.952	0.014	0.978	0.006	–
ω_1	0.006	0.001	0.007	0.001	[0.005;0.009]
γ_1	−0.047	0.004	−0.049	0.003	[−0.06;−0.034]
α_1	0.051	0.004	0.054	0.003	[0.035;0.060]
β_1	0.935	0.005	0.934	0.004	[0.91;0.95]
ω_2	0.190	0.100	0.320	0.120	[0.10;0.60]
γ_2	−0.259	0.233	−0.145	0.081	[−0.28;0.00]
α_2	0.600	0.227	0.837	0.097	[0.20;0.95]
β_2	0.787	0.065	0.762	0.049	[0.50;0.90]

This table reports ML estimation results and posterior moments based on 20,000 draws from the Gibbs sampler defined in Section 3.1 using different volatility dynamics. The data used is from July 2, 1962, to December 28, 2011, for a total of 12,459 observations. The parameters ω and μ have been multiplied by 10,000 and 100 respectively. In addition, for the AGARCH model the parameters γ have been multiplied by 100.

Table 7
Dollar losses with different volatility specifications.

	ML estimation					
	VST	ST	MT	LT	VLT	ALL
NGARCH	1.046	1.175	0.796	1.020	1.725	1.071
	3.318	5.171	7.770	10.923	13.882	7.491
GJRGARCH	1.212	1.405	0.937	0.853	0.801	1.121
	3.492	5.259	7.905	10.803	13.776	7.526
AGARCH	1.037	1.171	0.574	0.496	0.133	0.820
	3.340	4.987	7.230	10.447	13.264	7.139
	DOTM	OTM	ATM	ITM	DITM	ALL
NGARCH	2.161	1.058	0.189	0.537	1.238	1.071
	8.352	6.653	6.898	8.128	8.026	7.491
GJRGARCH	2.459	1.285	0.146	0.108	0.586	1.121
	8.467	6.678	6.945	8.104	7.700	7.526
AGARCH	2.519	0.998	−0.302	−0.528	−0.301	0.820
	8.047	6.236	6.598	7.789	7.325	7.139
	Bayesian inference					
	VST	ST	MT	LT	VLT	ALL
NGARCH	1.022	1.104	0.655	0.756	1.305	0.942
	3.325	5.143	7.717	10.923	13.842	7.467
GJRGARCH	1.154	1.290	0.685	0.281	−0.364	0.846
	3.488	5.232	7.863	10.852	13.715	7.510
AGARCH	1.031	1.139	0.493	0.330	−0.202	0.742
	3.367	4.995	7.248	10.499	13.257	7.157
	DOTM	OTM	ATM	ITM	DITM	ALL
NGARCH	2.019	0.926	0.065	0.409	1.158	0.942
	8.324	6.640	6.889	8.067	7.968	7.467
GJRGARCH	2.342	1.042	−0.214	−0.384	0.156	0.846
	8.493	6.634	6.948	7.993	7.578	7.510
AGARCH	2.450	0.917	−0.385	−0.629	−0.362	0.742
	8.087	6.252	6.617	7.761	7.295	7.157

In the cells of this table, the first number is the mean error and the second number is the root mean squared error both measured in USD. The maturity and moneyness are defined in Section 4.1.

Table 8
Implied volatility losses with different volatility specifications.

	ML estimation					
	VST	ST	MT	LT	VLT	ALL
NGARCH	1.454	1.152	0.648	0.540	0.508	0.972
	4.922	5.026	4.955	4.558	4.064	4.854
GJRGARCH	1.680	1.383	0.753	0.516	0.284	1.102
	5.118	5.054	4.935	4.434	3.978	4.887
AGARCH	1.468	1.235	0.591	0.450	0.147	0.946
	4.957	4.908	4.647	4.397	3.790	4.713
	DOTM	OTM	ATM	TM	DITM	ALL
NGARCH	1.996	0.886	0.210	0.165	1.450	0.972
	6.213	3.864	3.504	4.050	7.153	4.854
GJRGARCH	2.313	1.125	0.251	0.034	1.102	1.102
	6.310	3.932	3.507	4.001	7.001	4.887
AGARCH	2.439	0.944	0.007	−0.297	0.356	0.946
	6.188	3.668	3.276	3.855	6.821	4.713
	Bayesian inference					
	VST	ST	MT	LT	VLT	ALL
NGARCH	1.411	1.079	0.555	0.434	0.397	0.895
	4.943	5.010	4.923	4.557	4.066	4.846
GJRGARCH	1.607	1.294	0.621	0.324	−0.006	0.977
	5.144	5.050	4.910	4.466	3.972	4.891
AGARCH	1.452	1.204	0.537	0.383	0.056	0.904
	4.986	4.904	4.645	4.410	3.790	4.721
	DOTM	OTM	ATM	ITM	DITM	ALL
NGARCH	1.888	0.808	0.155	0.114	1.367	0.895
	6.187	3.856	3.503	4.053	7.197	4.846
GJRGARCH	2.279	1.011	0.094	−0.174	0.769	0.977
	6.338	3.907	3.496	3.993	6.996	4.891
AGARCH	2.388	0.899	−0.026	−0.333	0.308	0.904
	6.194	3.680	3.288	3.864	6.817	4.721

In the cells of this table, the first number is the mean error and the second number is the root mean squared error both measured in ISD's. The maturity and moneyness are defined in Section 4.1.

When comparing the results with ML estimation and Bayesian inference for the individual models some differences do occur, and this in particular so when dollar losses are used. For example, for the GJRGARCH model the overall BIAS is 32.5% larger when using ML estimation than when using Bayesian inference. For the NGARCH and AGARCH models the differences are only 13.7% and 10.5%, respectively. When looking across moneyness and maturity it is seen that this is due to a better performance for the LT and VLT options and for options that are DITM. However, for the implied volatility BIAS and for both the RMSEs the losses are much more similar.

5.2. Shorter estimation periods

The results presented above were based on the full sample from 1962 through 2011 for a total of 12,459 observations. With this amount of data and considering the time period spanned one might worry about structural breaks and the effect of these on the option pricing performance of the models. To examine this further, in this section we consider two alternative and shorter sample sizes. Since the option data we consider spans the period from 1996 through 2011 for a total of 16 years, or roughly 4032 trading days using the convention of 252 trading days per year, our shortest possible period is set to 4032 observations. We also consider twice this as a sample size, i.e. 8064 trading days or roughly 32 years. These two sample sizes correspond to roughly one third and two thirds of the sample size of 12,459 used previously.

In Table 9, we provide results for the MN2NGARCH model for the two alternative sample periods. Compared to the results in Table 2, we see that the mixture probability decreases when shorter samples are used and is down to 73.4% when using 4032 observations and ML estimation. The same holds for the persistence of the second component. In fact, for the shortest sample neither of the two components is explosive. When comparing the ML estimates to the posterior moments for the three sample sizes we see that although the properties of the first component are very close, the mixture probabilities are somewhat higher when using Bayesian inference. Moreover, the means of the posterior distributions are different from the ML estimates for the parameters of the second component. The largest differences are seen for γ_2. Finally, it is noteworthy that the persistence of the second component is much higher when using Bayesian inference than with ML estimates for the longer samples. When using 4032 observations only, these two methods lead to virtually identical values.

In Tables 10 and 11, we compare the dollar and implied volatility losses using the three different sample sizes, the full sample of 12,459 observations and the two shorter samples with respectively 8064 and 4032 observations across maturity

Table 9
Results with different sample sizes.

	ML estimates		Posterior moments		Prior support
	Estimate	Std error	Mean	Std deviation	
Panel A: Using 8064 observations					
μ	0.046	0.010	0.029	0.007	–
π	0.891	0.036	0.973	0.014	–
ω_1	0.006	0.002	0.010	0.002	[0.003;0.16]
γ_1	−0.905	0.109	−0.921	0.089	[−1.15;−0.70]
α_1	0.040	0.005	0.048	0.006	[0.024;0.06]
β_1	0.915	0.010	0.898	0.014	[0.86;0.94]
ω_2	0.248	0.077	0.289	0.105	[0.017;0.47]
γ_2	−0.807	0.263	−0.442	0.103	[−0.60;−0.20]
α_2	−0.285	0.079	0.566	0.095	[0.04;0.70]
β_2	0.630	0.096	0.703	0.077	[0.40;0.86]
Panel B: Using 4032 observations					
μ	0.121	0.024	0.119	0.018	–
π	0.734	0.046	0.787	0.044	–
ω_1	0.001	0.002	0.002	0.002	[0.01;0.009]
γ_1	−1.215	0.203	−1.202	0.084	[−1.50;−0.60]
α_1	0.041	0.006	0.041	0.005	[0.02;0.06]
β_1	0.896	0.018	0.894	0.005	[0.85;0.94]
ω_2	0.167	0.032	0.173	0.010	[0.03;0.27]
γ_2	−2.349	0.378	−1.920	0.075	[−2.2;−1.20]
α_2	0.126	0.025	0.131	0.012	[0.05;0.20]
β_2	0.154	0.134	0.349	0.040	[0.20;0.65]

This table reports ML estimation results and posterior moments based on 20,000 draws from the Gibbs sampler defined in Section 3.1 using different sample sizes. The parameters ω and μ have been multiplied by 10,000 and 100 respectively.

Table 10
Dollar losses with different sample sizes.

	ML estimation					
	VST	ST	MT	LT	VLT	ALL
Full	1.046	1.175	0.796	1.020	1.725	1.071
	3.318	5.171	7.770	10.923	13.882	7.491
8064	0.913	0.989	0.579	1.036	2.922	1.024
	2.663	4.086	6.100	9.313	14.063	6.525
4032	1.114	1.400	1.841	4.302	9.769	2.465
	2.687	4.156	6.494	10.750	17.945	7.576
	DOTM	OTM	ATM	ITM	DITM	ALL
Full	2.161	1.058	0.189	0.537	1.238	1.071
	8.352	6.653	6.898	8.128	8.026	7.491
8064	1.126	1.264	0.737	1.026	1.212	1.024
	6.040	5.962	6.513	8.358	7.878	6.525
4032	1.620	2.221	2.662	4.052	4.313	2.465
	6.319	6.797	7.590	10.641	10.376	7.576
	Bayesian inference					
	VST	ST	MT	LT	VLT	ALL
Full	1.022	1.104	0.655	0.756	1.305	0.942
	3.325	5.143	7.717	10.923	13.842	7.467
8064	0.980	1.058	0.644	1.123	3.109	1.103
	2.865	4.387	6.516	9.611	14.260	6.787
4032	1.039	1.215	1.287	2.877	6.886	1.845
	2.675	4.125	6.295	9.975	15.968	7.027
	DOTM	OTM	ATM	ITM	DITM	ALL
Full	2.019	0.926	0.065	0.409	1.158	0.942
	8.324	6.640	6.889	8.067	7.968	7.467
8064	1.264	1.292	0.763	1.133	1.428	1.103
	6.503	6.186	6.680	8.511	8.077	6.787
4032	1.303	1.775	1.897	2.880	3.098	1.845
	6.124	6.354	7.020	9.546	9.211	7.027

In the cells of this table, the first number is the mean error and the second number is the root mean squared error both measured in USD. The maturity and moneyness are defined in Section 4.1.

Table 11
Implied volatility losses with different sample sizes.

	ML estimation					
	VST	ST	MT	LT	VLT	ALL
Full	1.454	1.152	0.648	0.540	0.508	0.972
	4.922	5.026	4.955	4.558	4.064	4.854
8064	1.254	0.918	0.424	0.400	0.697	0.794
	4.088	4.082	3.917	3.710	3.961	3.987
4032	1.456	1.148	1.000	1.465	2.435	1.330
	4.056	4.152	4.150	4.125	4.888	4.186
	DOTM	OTM	ATM	ITM	DITM	ALL
Full	1.996	0.886	0.210	0.165	1.450	0.972
	6.213	3.864	3.504	4.050	7.153	4.854
8064	1.048	0.932	0.438	0.350	1.580	0.794
	4.591	3.188	3.041	3.718	7.047	3.987
4032	0.945	1.185	1.255	1.664	3.698	1.330
	4.501	3.322	3.323	4.170	7.911	4.186
	Bayesian inference					
	VST	ST	MT	LT	VLT	ALL
Full	1.411	1.079	0.555	0.434	0.397	0.895
	4.943	5.010	4.923	4.557	4.066	4.846
8064	1.319	0.969	0.455	0.436	0.762	0.843
	4.325	4.362	4.190	3.865	4.059	4.225
4032	1.365	1.022	0.722	0.976	1.700	1.080
	4.081	4.125	4.050	3.916	4.448	4.095
	DOTM	OTM	ATM	ITM	DITM	ALL
Full	1.888	0.808	0.155	0.114	1.367	0.895
	6.187	3.856	3.503	4.053	7.197	4.846
8064	1.165	0.948	0.449	0.385	1.629	0.843
	5.001	3.392	3.187	3.832	7.157	4.225
4032	0.880	1.039	0.936	1.165	2.873	1.080
	4.545	3.264	3.190	3.950	7.537	4.095

In the cells of this table, the first number is the mean error and the second number is the root mean squared error both measured in ISD's. The maturity and moneyness are defined in Section 4.1.

and across moneyness and when using ML estimation as well as Bayesian inference. The first thing to notice from the table is that dollar and implied volatility BIASs vary significantly with the sample size. For example, the shortest sample size leads to losses that are 130.2% and 36.8% larger than with the full sample, respectively, when using the ML estimates. In terms of RMSE losses, however, there is much less difference. In fact, in terms of RMSE using a smaller sample size leads to smaller losses in most cases.

When comparing the results with ML estimation and Bayesian inference for the different sample sizes, on the other hand, large differences do occur, and this is particular so for the smallest sample and when considering dollar losses. For example, when using the ML estimates the dollar BIAS is 33.6% larger than when using Bayesian inference. The reason for this is that the Bayesian method performs much better for LT and VLT options and for options that are ITM and DITM. For the implied volatility BIAS, this difference is down to 23.1% however, and it is much smaller also when considering the RMSEs.

6. Conclusion

We consider option pricing using mixed normal heteroscedasticity models. We explain how to perform inference and price options in a Bayesian framework. We compute posterior moments of the model parameters by sampling from the posterior density. Option prices are computed using the risk neutral predictive price densities that are easily obtained as a by-product of the Bayesian sampler. This method takes into account parameter uncertainty, because the sampler integrates over the entire parameter space. This is unlike classical inference that almost always conditions on maximum likelihood estimates.

An application using real data on the S&P 500 index is provided. We perform classical and Bayesian inference on a two component asymmetric normal mixture model using the available index return data. Comparing the ML estimates and posterior moments small differences are found. When pricing a rich sample of options on the index, both methods yield similar pricing errors measured in dollar and implied standard deviation losses, and it turns out that the impact of parameter uncertainty is minor. Therefore, when it comes to option pricing where large amounts of data are available, the choice of the inference method is unimportant.

As a robustness check we first of all consider two alternative specifications for the conditional volatility. Here our results show that our conclusions hold true for these alternative choices of volatility dynamics. Secondly, we consider the robustness of our results to changing sample size and consider two shorter samples. Here our results show that the two methods yield different results only for the shortest sample for which the Bayesian framework yields very different and smaller losses. Thus, our results indicate that parameter uncertainty could potentially play a role when pricing options on underlyings for which considerably less data is available.

A definite strength of the option pricing approach used here is that it can be implemented using only historical returns data. However, an obvious extension is to investigate how past option prices can be included in the inference procedure of this paper. In this case, the impact of parameter uncertainty could play a more important role. Another extension, where Bayesian inference is of particular interest, is Markov switching models in which returns can have a state dependent mean and variance, and recurrent or non-recurrent switches are determined by a Markov process. Within this framework of Markov switching and change-point GARCH models, it is possible to allow for state dependent unit risk premiums that drive a wedge between the physical and risk neutral dynamics.

Acknowledgments

The authors thank Eric Renault, Kris Jacobs, as well as seminar participants at Center (Tilburg University), CORE (Catholic University of Louvain), CREATES (Aarhus University), and the Economics Department at Boston University for comments. They also thank participants and various discussants at the 2009 FMA, NFA, and CIRPEE annual meetings as well as the Conference on Periodicity, Non-stationarity, and Forecasting of Economic and Financial Time Series and the 2009 CIREQ Financial Econometrics Conference. The second author thanks CREATES (Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation) for financial support.

References

Albert, J., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669–679.
Ausin, M.C., Galeano, P., 2007. Bayesian estimation of the Gaussian mixture Garch model. Computational Statistics & Data Analysis 51, 2636–2652.
Bates, D., 2003. Empirical option pricing: a retrospection. Journal of Econometrics 116, 387–404.
Bauwens, L., Dufays, A., Rombouts, J., 2011. Marginal likelihood for Markov-switching and change-point Garch models. CORE DP 2011/13.
Bauwens, L., Lubrano, M., 2002. Bayesian option pricing using asymmetric Garch models. Journal of Empirical Finance 9, 321–342.
Bauwens, L., Lubrano, M., Richard, J., 1999. Bayesian Inference in Dynamic Econometric Models. Oxford University Press, Oxford.
Bauwens, L., Preminger, A., Rombouts, J., 2010. Theory and inference for a Markov switching Garch model. Econometrics Journal 13, 218–244.
Bauwens, L., Rombouts, J., 2007a. Bayesian clustering of many Garch models. Econometric Reviews 26, 365–386.
Bauwens, L., Rombouts, J., 2007b. Bayesian inference for the mixed conditional heteroskedasticity model. Econometrics Journal 10, 408–425.
Bertholon, H., Monfort, A., Pegoraro, F., 2006. Pricing and inference with mixtures of conditional normal processes. CREST Working Paper 28.
Bertholon, H., Monfort, A., Pegoraro, F., 2008. Econometric asset pricing modelling. Journal of Financial Econometrics 6, 407–458.
Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.
Chen, C., Hung, M., 2010. Option pricing under Markov-switching GARCH processes. Journal of Futures Markets 30, 444–464.
Christoffersen, P., Elkamhi, R., Feunou, B., Jacobs, K., 2010a. Option valuation with conditional heteroskedasticity and non-normality. Review of Financial Studies 23, 2139–2183.
Christoffersen, P., Heston, S., Jacobs, K., 2006. Option valuation with conditional skewness. Journal of Econometrics 131, 253–284.
Christoffersen, P., Jacobs, K., 2004. Which GARCH model for option valuation? Management Science 50, 1204–1221.
Christoffersen, P., Jacobs, K., Dorion, C., Wang, Y., 2010b. Volatility components, affine restrictions and non-normal innovations. Journal of Business and Economic Statistics 28, 483–502.
Christoffersen, P., Jacobs, K., Ornthanalai, C., 2012. Dynamic jump intensities and risk premiums: evidence from S&P 500 returns and options. Journal of Financial Economics 106, 447–472.
Christoffersen, P., Jacobs, K., Ornthanalai, C., Wang, Y., 2008. Option valuation with long-run and short-run volatility components. Journal of Financial Economics 90, 272–297.
Duan, J., 1999. Conditionally fat-tailed distributions and the volatility smile in options. Mimeo, Hong Kong University of Science and Technology.
Elliot, R., Siu, T., Chan, L., 2006. Option pricing for Garch models with Markov switching. International Journal of Theoretical and Applied Finance 9, 825–841.
Engle, R., 1990. Discussion: stock market volatility and the crash of '87. Review of Financial Studies 3, 103–106.
Engle, R., Ng, G., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.
Garcia, R., Ghysels, E., Renault, E., 2010. The econometrics of option pricing. In: Aït-Sahalia, Hansen (Eds.), Handbook of Financial Econometrics, Vol 1. Elsevier, pp. 479–552.
Gelfand, A., Smith, A., 1990. Sampling based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398–409.
Geweke, J., 1989a. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.
Geweke, J., 1989b. Exact predictive densities in linear models with arch disturbances. Journal of Econometrics 40, 63–86.
Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics. Wiley-Interscience, New York.
Giannikis, D., Vrontos, I., Dellaportas, P., 2008. Modelling nonlinearities and heavy tails via threshold normal mixture Garch models. Computational Statistics & Data Analysis 52, 1549–1571.
Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.
Gourieroux, C., Monfort, A., 2007. Econometric specification of stochastic discount factor models. Journal of Econometrics 136, 509–530.
Haas, M., Mittnik, S., Paolella, M., 2004. Mixed normal conditional heteroskedasticity. Journal of Financial Econometrics 2, 211–250.
Heston, S.L., Nandi, S., 2000. A closed-form GARCH option valuation model. Review of Financial Studies 13, 585–625.
Hsieh, K.C., Ritchken, P., 2005. An empirical comparison of GARCH option pricing models. Review of Derivatives Research 8, 129–150.
Jacquier, E., Jarrow, R., 2000. Bayesian analysis of contingent claim model error. Journal of Econometrics 94, 145–180.
Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–417 (with discussion).
Johannes, M., Polson, N., 2010. MCMC methods for continuous-time financial econometrics. In: Aït-Sahalia, Hansen (Eds.), Handbook of Financial Econometrics, Vol 2. Elsevier, pp. 1–72.
Jones, C., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. Journal of Econometrics 116, 181–223.
Koop, G., 2003. Bayesian Econometrics. Wiley-Interscience, New York.
Martin, G., Forbes, C., Martin, V., 2005. Implicit Bayesian inference using option prices. Journal of Time Series Analysis 26, 437–462.
Merton, R.C., 1973. Theory of rational option pricing. Bell Journal of Economics and Management Science 4, 141–183.

Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods. Springer-Verlag, New York.
Rombouts, J., Stentoft, L., 2010. Option pricing with asymmetric heteroskedastic normal mixture models. CREATES Working Paper 44.
Satoyoshi, K., Mitsui, H., 2011. Empirical study of nikkei 225 options with the Markov switching Garch model. Asia-Pacific Financial Markets 18, 55–68.
Stentoft, L., 2008. American option pricing using GARCH models and the normal inverse Gaussian distribution. Journal of Financial Econometrics 6, 540–582.
Tanner, M., Wong, W., 1987. The calculation of Posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–540.
Tsay, R., 2005. Analysis of Financial Time Series. Wiley-Interscience, New York.
van Dijk, H., Kloek, T., 1978. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. Econometrica 46, 1–19.
Wong, C., Li, W., 2000. On a mixture autoregressive model. Journal of the Royal Statistical Society: Series B 62, 95–115.
Wong, C., Li, W., 2001. On a mixture autoregressive conditional heteroscedastic model. Journal of the American Statistical Association 96, 982–995.
Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.