



Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **09-26**

Bayesian Option Pricing Using Mixed Normal Heteroskedasticity Models

Jeroen V.K. Rombouts

Lars Stentoft

Août/August 2009

Rombouts: Institute of Applied Economics at HEC Montréal, CIRANO, CIRPÉE, Université Catholique de Louvain (CORE), 3000 Côte Sainte-Catherine, Montréal (QC) Canada H3T 2A7; phone : +1 514 340-6466; fax : +1 514 340-6469

jeroen.rombouts@hec.ca

Stentoft: Department of Finance at HEC Montréal, CIRANO, CREATES, CREF

The authors thank Eric Renault, Kris Jacobs, as well as seminar participants at Center (Tilburg University), CORE (Catholic University of Louvain), CREATES (University of Aarhus), and the Economics Department at Boston University for comments. The first author thanks IFM2 for financial support. The second author thanks CREATES (Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation) for financial support.

Abstract:

While stochastic volatility models improve on the option pricing error when compared to the Black-Scholes-Merton model, mispricings remain. This paper uses mixed normal heteroskedasticity models to price options. Our model allows for significant negative skewness and time varying higher order moments of the risk neutral distribution. Parameter inference using Gibbs sampling is explained and we detail how to compute risk neutral predictive densities taking into account parameter uncertainty. When forecasting out-of-sample options on the S&P 500 index, substantial improvements are found compared to a benchmark model in terms of dollar losses and the ability to explain the smirk in implied volatilities.

Keywords: Bayesian inference, option pricing, finite mixture models, out-of-sample prediction, GARCH models

JEL Classification: C11, C15, C22, G13

While the Black-Scholes-Merton (BSM) model of Black and Scholes (1973) and Merton (1973) remains a standard tool for practitioners, 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 put 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 non Gaussian returns, or a combination of the two. Thus, extensions to the BSM model have been developed focusing on relaxing these assumptions. This paper uses mixed normal heteroskedasticity models which allow for time varying volatility and can generate skewness and excess kurtosis in the distribution. Using this model, we find substantial improvements compared to a benchmark model in terms of dollar losses and the ability to explain the smirk in implied volatilities.

Several studies have examined models which allow for more flexible specifications of the volatility process compared to the BSM model. In particular, the stochastic volatility (SV) models have successfully been applied, see Hull and White (1987), Johnson and Shanno (1987), Scott (1987), Stein and Stein (1991), Wiggins (1987), Amin and Ng (1993), and Heston (1993). When comparing these models to the BSM model, empirically support is found for the stochastic volatility specification. This is documented in Bakshi, Cao, and Chen (1997), Bates (2000), and Nandi (1996) when considering options on the S&P 500 index or the index futures. In addition to the SV models, the generalized autoregressive conditional heteroskedasticity (GARCH) framework has been used for option pricing using the model of Duan (1995). This model has also been used with success for empirical option pricing by, among others, Christoffersen and Jacobs (2004), Heston and Nandi (2000), and Hsieh and Ritchken (2005). A recent contribution is Christoffersen, Jacobs, Ornathanalai, and Wang (2008) where the volatility is allowed to have both short and long run components. These studies have all analyzed options on the S&P 500 index and found that the GARCH models diminish the mispricings found when using the constant volatility BSM model. Stentoft (2005) documents the same using a sample individual stock options and options on the S&P

100 index.

However, while the improvements in option pricing performance of SV and GARCH models are important, mispricings still exist when comparing these models to actual option data as documented by Nandi (1996). This has led to the development of option pricing models which rely on alternative conditional distributions in addition to having non constant volatility. In the SV literature, jumps have been introduced in the return and volatility processes. Classical references are, among others, Bakshi, Cao, and Chen (1997), Bakshi, Cao, and Chen (2000), Bates (1991), Bates (2000), Pan (2002), and Eraker (2004). When examining the empirical performance of these models, most of the above papers find support for the existence of jumps again when looking at S&P 500 index options. One exception is Bakshi, Cao, and Chen (2000) which, however, analyzes long term options and finds only small gains of allowing for jumps. In addition to the jump-diffusion processes, models based on infinite activity Lévy processes have been proposed in Barndorff-Nielsen and Shephard (2001) and Carr, Geman, Madan, and Yor (2003), and on time changed Lévy processes in Carr and Wu (2004) with applications to currency options in Carr and Wu (2007) and Bakshi, Carr, and Wu (2008).

Within the GARCH literature extensions include the use of distributions which are either skewed or leptokurtic, or both, as is done in Duan (1999) and Christoffersen, Elkamhi, Feunou, and Jacobs (2008). For empirical applications of this framework, see Christoffersen, Heston, and Jacobs (2006), Christoffersen, Jacobs, Dorion, and Wang (2008) and Stentoft (2008). Although Christoffersen, Jacobs, Dorion, and Wang (2008) find little improvement for the non-normal models Christoffersen, Heston, and Jacobs (2006) observe that allowing for non-normal innovations is important when pricing out of the money put options on the S&P 500 index. Moreover, Stentoft (2008) documents improvements for both call and put options in terms of fitting the smile across moneyness for a sample of individual stock options as well as for options on the S&P 100 index. In addition to models with non Gaussian innovations, GARCH models with jumps have been developed by Duan, Ritchken, and Sun (2006) and Christoffersen, Jacobs, and Ornathanalai (2008). The latter of these papers also examines the empirical performance and shows that jumps are important empirically when pricing S&P 500 index options.

The discussion above documents that important advances have been made in the empirical option pricing literature when it comes to extending the BSM model. However, it is also clear that there is still much room for improvements as discussed in the reviews by Bates (2003) and Garcia, Gysels, and Renault (2009). 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. However, 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 requires substantial negative skewness, more than is often found in the underlying process, whereas the latter indicates that moments of higher order are time varying.

In the present paper, we price options using mixed normal heteroskedasticity models, and we show that our proposed model can address the above points. In particular, 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. We suggest a feasible way for option pricing within this general framework and we derive the appropriate risk neutral dynamics. In our application, we find pronounced differences between the risk neutral distribution and the conditional distribution of the underlying asset given the significant risk premium. We show that our model allows for significant negative skewness and time varying higher order moments.

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. In this framework, each distribution in the mixture can have its own mean and conditional variance process. Moreover, if required by the data, some conditional variance processes may even be weakly nonstationary, for example to capture turbulent periods, while the overall conditional variance remains weakly stationary. Finite mixture textbooks are for example McLachlan and Peel (2000) and Frühwirth-Schnatter

(2006). Early applications are Kon (1982) and Kim and Kon (1994) who investigate the statistical properties of stock returns using mixture models. Boothe and Glassman (1987), Tucker and Pond (1988), and Pan, Chan, and Fok (1995) use mixtures of normals to model exchange rates. Recent examples are Wong and Li (2000) and Wong and Li (2001) who model the conditional return distribution, extended by Haas, Mittnik, and Paoletta (2004) with an application of value at risk prediction, and Bauwens and Rombouts (2007a) for the clustering of financial time series. Durham (2007) investigates the power of finite mixtures of normal densities with stochastic volatility to model the return of the S&P 500 index. Using statistical criteria such as QQ-plots, goodness-of-fit tests and information criteria he finds that the finite mixture does a good job of capturing the salient features of the data. We extend his work by analyzing the performance of finite mixtures in out-of-sample option pricing. Hence, we focus mostly on the financial properties of the finite mixture instead of its statistical properties. In doing so, we examine both the physical and risk neutral measures, which are shown to be very different.

The main advantages of the Bayesian approach are twofold. Firstly, we avoid the maximization of the involved likelihood function of the mixture model. Instead, 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. Secondly, the predictive price densities, we compute for evaluating the option prices, are easily obtained as a by-product of the Bayesian sampler and take into account parameter uncertainty, because we integrate over the entire parameter space. This is unlike classical inference that almost always conditions on maximum likelihood estimates. Another interesting advantage of the Bayesian approach is that our procedure can be applied directly to the raw returns instead of percentage returns without running into problems of numerical stability. This is particularly convenient when pricing options since the riskneutralization procedure involves nonlinear transformations of the model parameters and the data. We note that Bayesian inference combines data information and prior information. However, the priors we use in this paper are diffusive, so that we give most of the weight to the data to learn about the model parameters.

Our results show that the added flexibility of the finite mixture model in terms of both skewness and excess kurtosis provides important improvements in terms of predicted option prices. We use our model to forecast out-of-sample prices of 8,637 call and put options on the S&P 500 index in 2006 and compute dollar losses and implied standard deviation losses. Using dollar losses the results are quite similar for call options, but for put options the losses are approximately 16% or \$0.34 smaller. Moreover, improvements are found for most categories of moneyness and maturity. The largest improvement in dollar terms is \$1.92 for the deep out of the money call options with very long maturity, and \$1.71 for the deep in the money put options also with very long maturity. In percentage of the mean option prices, the improvement is largest for the shorter term options which are out of the money. For the call options it is as high as 37% for options which are deep out of the money with short maturity. Finally, using implied volatility losses, we also show that the mixed normal heteroskedasticity model performs significantly better for options far away from being at the money. Moreover, when the losses are considered across moneyness for various maturity categories, it is observed that our model does a much better job than the GARCH model in explaining the smirk found in our sample of options. The improvements are particularly pronounced for the short term options.

It should be noted that we use only historical data on the underlying asset to obtain the parameters to be used for option valuation. However, it is very likely that historical option prices themselves contain important information on the model parameters. Therefore, an alternative approach is to infer these parameters either from historical option data alone or by using both returns and options data, by calibrating the option pricing model to existing option data, as detailed in Chernov and Ghysels (2000). This is for example the procedure used in Jacquier and Jarrow (2000), who incorporates parameter uncertainty and model error in the BSM model, and in Eraker (2004), who develops an option pricing model allowing for stochastic volatility with jumps in both the return and the volatility processes. In both of these papers, option pricing models within the affine class are considered, a choice which is motivated by the existence of closed or semiclosed form pricing formulas. However, this class of models could be considered as restrictive since many of the well known specifications within for example the GARCH framework are non-affine. An important exception is Heston

and Nandi (2000). Moreover, research has shown that in terms of option pricing, non-affine models often perform better than affine models, see for example Hsieh and Ritchken (2005) and Christoffersen, Jacobs, Dorion, and Wang (2008). In the framework of the mixed normal heteroskedasticity model considered here, closed or semiclosed form solutions for option prices do not exist. Thus, using option data for estimation purposes along the lines above becomes extremely cumbersome if at all possible.

The rest of the paper is organized as follows: Section 1 presents the mixed normal heteroskedasticity model. In Section 2 we discuss the risk neutral dynamics, and in Section 3 we explain how options can be priced using the mixed normal heteroskedasticity model. Section 4, explains how to conduct Bayesian inference and compute predictive price densities taking into account parameter uncertainty. Section 5, reports the results on the empirical application to options on the S&P 500 index. Finally, Section 6 contains the conclusion.

1 Mixed normal heteroskedasticity

Building on the finite mixtures with autoregressive means and variances of Wong and Li (2000) and Wong and Li (2001), Haas, Mittnik, and Paoletta (2004) develop a mixture of normals coupled with the GARCH specification to capture, for example, conditional kurtosis and skewness. They define a mixture model on a demeaned series $\varepsilon_t = y_t - E(y_t|\mathcal{F}_{t-1})$ where \mathcal{F}_{t-1} is the information set up to time t and the conditional mean does not depend on the mixture. The conditional distribution of ε_t is a combination of K densities

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

where

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

and $\Phi(\cdot)$ is the standard Gaussian distribution. We denote this the mixed normal heteroskedasticity (MNGARCH) model. For each t in this finite mixture framework, ε_t is drawn from one of the K conditional distributions with probabilities π_1, \dots, π_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 $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$. The zero mean assumption on ε_t is ensured by the restriction

$$\mu_K = - \sum_{k=1}^{K-1} \frac{\pi_k \mu_k}{\pi_K}. \quad (3)$$

Note that this zero mean restriction does not imply a symmetric distribution. The latter would only happen if all μ_K 's are zero, something that can be tested. Indeed, the fact that our model is able to generate substantial skewness is an important advantage as documented also in Durham (2007).

Conditional moments of ε_t for the MNGARCH model are combinations of the K distribution moments in (1). The conditional variance of ε_t is given by

$$\sigma_t^2 = \sum_{k=1}^K (\pi_k \mu_k^2 + \pi_k \sigma_{k,t}^2). \quad (4)$$

The conditional third moment is given by

$$E_{t-1}(\varepsilon_t^3) = \sum_{k=1}^K (\pi_k \mu_k^3 + 3\pi_k \sigma_{k,t}^2 \mu_k), \quad (5)$$

and the conditional fourth moment is given by

$$E_{t-1}(\varepsilon_t^4) = \sum_{k=1}^K (\pi_k \mu_k^4 + 6\pi_k \mu_k^2 \sigma_{k,t}^2 + 3\pi_k \sigma_{k,t}^4), \quad (6)$$

where $E_{t-1}(\cdot)$ means the expectation conditional on \mathcal{F}_{t-1} . These formulas illustrate that we can have flexible dynamics compared to the classical GARCH model which arises when $K = 1$. For example, the skewness of the conditional distribution of returns would be forced to zero in the latter case.

The MNGARCH model defined here is substantially different from the component volatility models introduced by Ding and Granger (1996) and Engle and Lee (1999). The component volatility model incorporates in a convenient way long range dependence in volatility. Each component allows the variance innovations to decay at a different rate, however there is only one conditional variance process. It can easily be shown that the component volatility model is a GARCH model with more lags in the squared innovations and the conditional variances than the conventional GARCH(1,1) model. For example, in Engle and Lee (1999), one component captures the long run movements in volatility while a second component

accounts for the short run volatility movements. Skewness in the conditional return distribution is only possible when the innovation distribution is skewed. However, it is possible to have skewness in the unconditional return distribution if there is a time varying conditional mean in the model, see He, Silvennoinen, and Terasvirta (2008) for more details.

What makes the mixture model attractive from an econometric point of view is the common innovation term ε_t that feeds in the K conditional variance equations. Therefore, evaluation of the likelihood function is possible since there is no path dependence problem as would be the case in a Markov switching GARCH model. Note that to have an overall variance process that is weakly stationary, only one of the conditional variance processes is required to be weakly stationary. The other $K - 1$ conditional variance processes are allowed to be explosive ($\alpha_k + \beta_k > 1$) as long as their combined probability is not too high. More formally, the weak stationarity condition for the model is

$$\left[\sum_{k=1}^K \frac{\pi_k}{(1 - \beta_k)} \left(1 - \alpha_k - \beta_k \right) \right] \prod_{k=1}^K (1 - \beta_k) > 0. \quad (7)$$

More details on the unconditional moments of the mixed normal heteroskedasticity model can be found in Haas, Mittnik, and Paolella (2004).

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 give a financial interpretation of the parameters, like in this paper, we add an identification restriction like $\pi_1 \geq \pi_2 \geq \dots > \pi_K$. Other restrictions, for example on the mean of the distributions, are possible as explained in Hamilton, Zha, and Waggoner (2007).

Finally, we note that the mixture model can incorporate extreme events by having a distribution with very low probability and with a large mean and a small constant variance for example. Moreover, the model can be made even more general by considering other conditional variance models than the GARCH model of Bollerslev (1986) used here. Examples include, but are not restricted to, the NGARCH model of Engle and Ng (1993), the GJR-GARCH model of Glosten, Jagannathan, and Runkle (1993), or the EGARCH model of Nelson (1991) all of which allow for asymmetries in the volatility process. Furthermore, other

distributions than the normal can be considered in the finite mixture like the exponential power distribution proposed by Bouaddi and Rombouts (2009).

1.1 Asset return dynamics with mixed normal heteroskedasticity

In this paper, we use the dynamics above as the driving innovation for asset returns. To be specific, we assume that the underlying return process $R_t \equiv \ln(S_t/S_{t-1})$ can be characterized by

$$R_t = \mu_t - \Psi_t(-1) + \varepsilon_t, \quad (8)$$

where S_t is the index level on day t and where ε_t follows (1). In (8), the term $\Psi_t(\cdot)$ denotes the conditional cumulant generating function which corresponds to the logarithm of the conditional moment generating function. The conditional moment generating function (also called the conditional Laplace transform) of ε_t in (1) is given by

$$E_{t-1} \exp(-u\varepsilon_t) = \sum_{k=1}^K \pi_k \exp\left(-u\mu_k + \frac{u^2\sigma_{k,t}^2}{2}\right). \quad (9)$$

In fact, this is just a convex combination of Gaussian moment generating functions and thus very easy to calculate. For option pricing purposes, the logarithm of this function will be used extensively, and the fact that it may easily be calculated is therefore convenient.

Using (9), the conditional cumulant generating function evaluated at $u = -1$ is now simply given by

$$\ln(E_{t-1}[\exp(\varepsilon_t)]) = \Psi_t(-1). \quad (10)$$

Substituting this into (8), we see that

$$\begin{aligned} E_{t-1}[S_t/S_{t-1}] &= E_{t-1}[\exp(\mu_t - \Psi_t(-1) + \varepsilon_t)] \\ &= \exp(\mu_t). \end{aligned} \quad (11)$$

It makes indeed sense to define the return process as above since with this particular specification, since μ_t can be interpreted as the expected gross rate of return.

2 Risk neutral dynamics

Once the conditional distribution of R_t is specified, the proposed system can be taken to the data. However, when the ultimate goal is that of option pricing some further work is needed. In particular, for option pricing to proceed an equivalent martingale measure (EMM) is needed. While technically complicated, intuitively what we are looking for is a "transformation" to the distribution of ε_t . This transformation, however, has to be special in the sense that the two measures, the original and the transformed one, have the same null sets, i.e. they are equivalent. Furthermore, under the transformed measure, the expected gross rate of returns should equal the risk free rate, i.e. discounted returns are martingales. The new conditional distribution is distinguished from the original density F by adding a superscript Q . We will refer to this as either the distribution under Q , the distribution under the risk neutral measure, or simply as the risk neutral distribution.

In this paper, we follow the approach of Christoffersen, Elkamhi, Feunou, and Jacobs (2008) which involves the specification of a candidate EMM through a specification of a Radon-Nikodym derivative. A similar method, which would provide an equal set of conditions, specifies a candidate stochastic discount factor directly as is done in Gourieroux and Monfort (2007). For a discussion of the relationship between the two probability measures F and Q and the corresponding stochastic discount factor see Bertholon, Monfort, and Pegoraro (2008). Note that an alternative method is to 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 and an actual application of the method computationally complex (see e.g. Stentoft (2008)). Thus, this approach appears to be more restrictive.

Once a candidate EMM is obtained, European options may be priced as the expected value, under the EMM, of future cash flows discounted using the risk free interest rate. For example, the price of a European call option at time T with maturity T^* and strike price K can be computed as

$$C_T(S, T^*, K) = e^{-r(T^*-T)} \int_0^\infty \max(S_{T^*} - K, 0) f^Q(S_{T^*}) dS_{T^*} \quad (12)$$

where $f^Q(S_{T^*})$ is the density of the underlying asset price at expiration under the EMM, i.e.

the risk neutral density.

2.1 Specification of a candidate EMM

We specify a candidate EMM from the following Radon-Nikodym derivative

$$\frac{dQ}{dF} \Big|_{\mathcal{F}_t} = \exp \left(- \sum_{i=1}^t (\nu_i \varepsilon_i + \Psi_i(\nu_i)) \right), \quad (13)$$

where $\Psi_t(u)$ is the conditional cumulant generating function specified above. It is immediately observed that by using a Radon-Nikodym of this type we are guaranteed that the two measures are equivalent as they have the same null sets. To ensure that discounted asset prices are martingales under the risk neutral measure it can be shown that the sequence ν_t has to satisfy the following equation

$$0 = \Psi_t(\nu_t - 1) - \Psi_t(\nu_t) - \Psi_t(-1) + \mu_t - r_t. \quad (14)$$

Moreover, we have that under the risk neutral measure the conditional cumulant generating function of ε_t is given by

$$\Psi_t^Q(u) = \Psi_t(\nu_t + u) - \Psi_t(\nu_t). \quad (15)$$

Using the Inversion Theorem (see for example Billingsley (1995, Theorem 26.2) or Davidson (1997, Theorem 11.12)) this can be used to obtain (15) the distribution under Q .

The two above equations completely characterize the risk neutral process and hence this is, in fact, all that is needed for option pricing purposes. In particular, equation (15) characterizes the risk neutral distribution in terms of the sequence ν and equation (14) provides the link between this sequence and the properties under the original measure F . In order to apply the method for pricing, however, all that is left is to derive these dynamics explicitly. To do this, we simply use (14) and (15) in the mean equation in (8). Doing so, we obtain the following specification of the risk neutral dynamics

$$R_t^Q = r_t - \Psi_t^Q(-1) + \varepsilon_t^Q, \quad (16)$$

where the superscript Q indicates that the variables are considered under the risk neutral distribution. Using this specification, we can calculate the gross rate of return on assets

under Q as

$$\begin{aligned} E_{t-1}^Q [S_t/S_{t-1}] &= E_{t-1}^Q \left[\exp \left(r_t - \Psi_t^Q(-1) + \varepsilon_t \right) \right] \\ &= \exp(r_t). \end{aligned} \tag{17}$$

Thus, it equals the risk free interest rate as required.

As it was noted in Christoffersen, Elkamhi, Feunou, and Jacobs (2008), the above method is only one of several ways to derive the risk neutral dynamics to be used for option pricing. In fact, in the present setting markets are incomplete and hence in general no unique EMM exists as is the case in for example the constant volatility BSM model. However, in this setup the obtained option prices are in fact unique conditional on the choice of Radon-Nikodym derivative.

2.2 A strategy for riskneutralization

Based on the above, the strategy to follow for option pricing is clear and consists of the following three steps:

- First, a sequence ν_t which satisfy (14) has to be specified. If possible, this can be done directly by inverting the conditional cumulant generating function. Below, we propose an alternative and somewhat indirect way of specifying this sequence which can be used in general.
- Secondly, given the sequence for ν_t the conditional cumulant generating function for ε_t under the risk neutral density can be obtained from (15). With this the distribution of ε_t under the Q measure can be obtained.
- Finally, the return process under the risk neutral measure can be obtained from (16) and this then specifies the dynamics to be used for option pricing purposes. Note that if the conditional variance depends on the innovations, as it is the case in for example the GARCH specification, it is the innovations from this distribution which should be used.

It is immediately seen, that the procedure we use riskneutralizes by changing the distribution of the innovations. This is in line with the procedure used in the BSM model and in most of the SV literature. It also corresponds to the method used in Duan (1999), although the obtained distribution is given somewhat implicitly in this framework as mentioned above. However, the method differs from the procedure used in Barone-Adesi, Engle, and Mancini (2008), where it is the actual parameters of the mean and variance process, that are changed under Q .

2.2.1 The Gaussian special case

We now consider the Gaussian special case for illustrative purposes. In this case, the conditional cumulant generating function of ε_t is given by

$$\Psi_t(u) = \frac{1}{2}\sigma_t^2 u^2. \quad (18)$$

We now proceed with the three steps outlined above for risk neutralization:

- Substituting the conditional cumulant generating function into (14), we can rewrite the EMM restriction as

$$\begin{aligned} 0 &= \frac{1}{2}\sigma_t^2 (\nu_t - 1)^2 - \frac{1}{2}\sigma_t^2 (\nu_t)^2 - \frac{1}{2}\sigma_t^2 (-1)^2 + \mu_t - r_t \\ &= -\sigma_t^2 \nu_t + \mu_t - r_t. \end{aligned} \quad (19)$$

Thus, in this case, there is an exact and analytically tractable mapping between μ_t and the required ν_t . In particular, for any choice of μ_t , the corresponding ν_t process is given by

$$\nu_t = \frac{\mu_t - r_t}{\sigma_t^2}. \quad (20)$$

- Substituting (20) into (15) and using the specification of the cumulant generating function, we obtain

$$\begin{aligned} \Psi_t^Q(u) &= \frac{1}{2}\sigma_t^2 \left(\frac{\mu_t - r_t}{\sigma_t^2} + u \right)^2 - \frac{1}{2}\sigma_t^2 \left(\frac{\mu_t - r_t}{\sigma_t^2} \right)^2 \\ &= (\mu_t - r_t)u + \frac{1}{2}\sigma_t^2 u^2. \end{aligned} \quad (21)$$

This, however, is recognized as the cumulant generating function of a Gaussian variable with mean equal to $-(\mu_t - r_t)$ and variance equal to σ_t^2 . Thus, it follows that ε_t is distributed as $N(r_t - \mu_t, \sigma_t^2)$ under Q .

- The risk neutral dynamics may now be obtained as

$$\begin{aligned} R_t^Q &= r_t - \Psi_t^Q(-1) + \varepsilon_t^Q \\ &= \mu_t - \frac{1}{2}\sigma_t^2 + \varepsilon_t^Q. \end{aligned} \tag{22}$$

To be even more specific, we take as an example the particular choice of mean specification used in Duan (1995) where $\mu_t = r_t + \lambda\sigma_t$. Substituting this into (22), the risk neutral dynamics are obtained as

$$R_t^Q = r_t + \lambda\sigma_t - \frac{1}{2}\sigma_t^2 + \varepsilon_t^Q. \tag{23}$$

Moreover, assume that the conditional variance is of the GARCH type given by

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{24}$$

Then, under the risk neutral dynamics the conditional variance is given by

$$\sigma_t^2 = \omega + \alpha(\varepsilon_{t-1}^Q)^2 + \beta\sigma_{t-1}^2. \tag{25}$$

3 Option pricing under general assumptions

In the previous section, we outlined a strategy for riskneutralization and we provided an example on how this can be implemented in the Gaussian special case. In this section, we show how this can be done under more general assumptions and we interpret the effect of riskneutralization on the model parameters. We first provide a general method for obtaining a solution to the restriction in equation (14) used in the first step. We then derive the risk neutral distribution in the mixed normal heteroskedasticity model from the conditional cumulant generating function obtained in the second step from (15).

3.1 Feasible option pricing

As mentioned above, we effectively choose the appropriate EMM by solving (14) for ν_t given μ_t and the assumed distribution of ε_t . However, this is potentially complicated due to the nonlinearity of this relationship and an analytical expression for ν_t may therefore not be available in general. In particular, this is the case with the mixed normal heteroskedasticity model used here, and it may thus seem to be impossible to derive the EMM given μ_t for this model.

However, (14) may equally well be solved for μ_t given ν_t and the assumed distribution of ε_t as

$$\mu_t = r_t - \Psi_t(\nu_t - 1) + \Psi_t(\nu_t) + \Psi_t(-1). \quad (26)$$

From this, we note that for any choice of ν_t a closed form expression exist for μ_t given that the cumulant generating function exists. Substituting this into the return equation in (8) we obtain

$$R_t = r_t - \Psi_t(\nu_t - 1) + \Psi_t(\nu_t) + \varepsilon_t, \quad (27)$$

which can be used for estimation directly. This way of implying μ_t , given a particular specification of ν_t , is used in Stentoft (2008) in the Normal Inverse Gaussian framework using the option pricing model of Duan (1999). However, it is equally applicable here.

It should be noted that, depending on the specification of ν_t and the assumed distribution of ε_t , we may interpret the specification in (27) differently. We next illustrate this for the Gaussian case and then for the general case.

3.1.1 Interpreting ν_t in the Gaussian special case

With the Gaussian distribution, we obtain the following in (27) above

$$\begin{aligned} R_t &= r_t - \Psi_t(\nu_t - 1) + \Psi_t(\nu_t) + \varepsilon_t \\ &= r_t - \frac{1}{2}\sigma_t^2(\nu_t - 1)^2 + \frac{1}{2}\sigma_t^2(\nu_t)^2 + \varepsilon_t \\ &= r_t - \frac{1}{2}\sigma_t^2 + \nu_t\sigma_t^2 + \varepsilon_t. \end{aligned} \quad (28)$$

Hence, we see that ν_t is related to the assumption of the unit risk premium. In particular, if we were to specify $\nu_t = \nu$, that is as a constant, the implied mean specification corresponds to assuming a unit risk premium proportional to the level of the variance. Alternatively, if $\nu_t = \nu/\sigma_t$, the unit risk premium becomes proportional to the level of the standard deviation, and finally with $\nu_t = \nu/\sigma_t^2$ a constant unit risk premium independent of the level of the variance is obtained. Thus, while it may appear that we, by implying the mean, are constraining the potential mean specification in an unreasonably way from an econometric point of view, this is in fact not the case.

3.1.2 Interpreting ν_t in the general case

In the Gaussian case, the sequence ν_t was immediately interpreted as determining the unit risk premium. In the mixed normal heteroskedasticity model this is less obvious. However, we may analyze the effect of ν_t using a Taylor series expansion in (27). To do this, we first note that the two terms involving the log moment generating functions may be approximated by

$$\begin{aligned} \Psi_t(\nu_t - 1) &\approx \Psi_t(0) + \Psi_t'(0)(\nu_t - 1) + \frac{1}{2}\Psi_t''(0)(\nu_t - 1)^2 \\ &\quad + \frac{1}{6}\Psi_t'''(0)(\nu_t - 1)^3 + \frac{1}{24}\Psi_t''''(0)(\nu_t - 1)^4 \end{aligned} \quad (29)$$

and

$$\Psi_t(\nu_t) \approx \Psi_t(0) + \Psi_t'(0)\nu_t + \frac{1}{2}\Psi_t''(0)\nu_t^2 + \frac{1}{6}\Psi_t'''(0)\nu_t^3 + \frac{1}{24}\Psi_t''''(0)\nu_t^4, \quad (30)$$

respectively. Furthermore, by the definition of ε_t as a zero mean random variable with conditional variance σ_t^2 we have that $\Psi_t'(0) = 0$ and $\Psi_t''(0) = \sigma_t^2$. Moreover, by the definition of $\Psi_t(u)$, we have that $\Psi_t'''(0) = -skew_t\sigma_t^3$ and $\Psi_t''''(0) = exkurt_t\sigma_t^4$, where $skew_t$ and $exkurt_t$ denotes the conditional skewness and excess kurtosis at time t respectively.

With the above expressions, (27) may be rewritten as

$$R_t \approx r_t - \Psi_t(-1) + \nu_t\sigma_t^2 + \frac{(3\nu_t^2 - 3\nu_t)}{6}(-skew_t)\sigma_t^3 + \frac{(4\nu_t^3 - 6\nu_t^2 + 4\nu_t)}{24}exkurt_t\sigma_t^4 + \varepsilon_t, \quad (31)$$

where we have collected all the terms which do not involve ν_t in $\Psi_t(-1)$. Equation (31) shows that even in the more general setting it makes sense to interpret ν_t as the unit risk

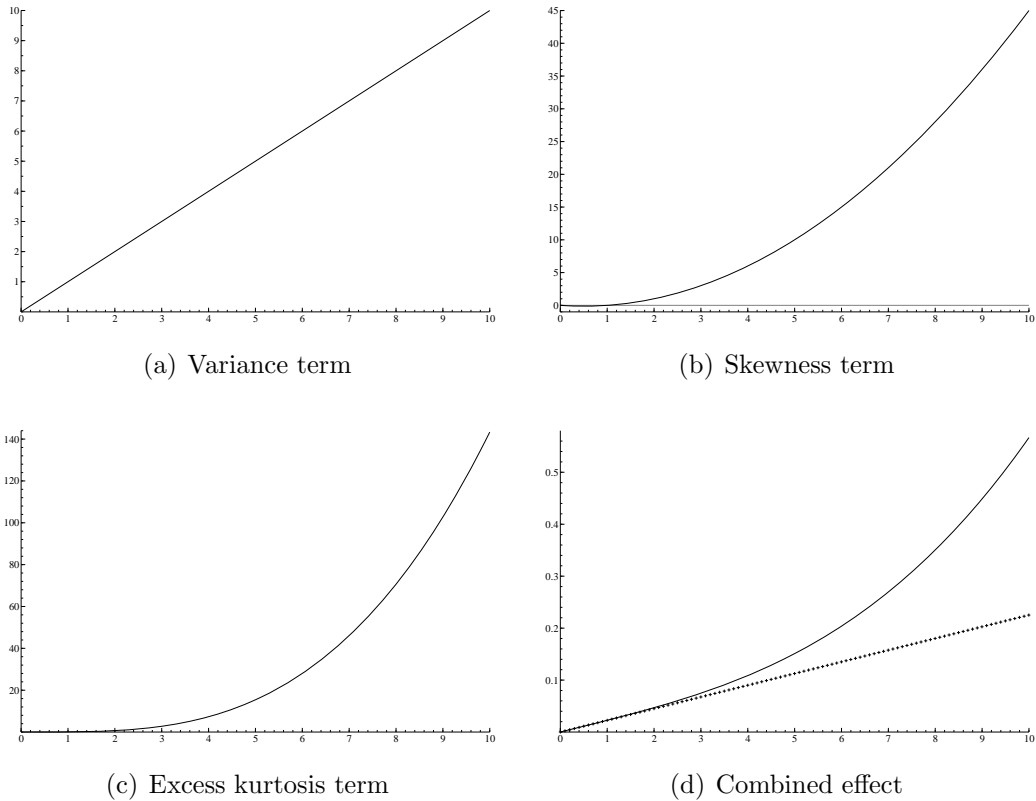


Figure 1: Unit risk premium effects

This figure plots the coefficients of the variance term, the skewness term, and the excess kurtosis term in (31) as a function of ν_t in panels (a) through (c). Panel (d) plots the combined effect.

premium. In particular, it may be observed that for reasonable values of ν_t , say larger than 1, the coefficients above are all positive. In Figure 3.1.2, we plot in panels (a) through (c) the coefficients of the variance term, the skewness term, and the excess kurtosis term as a function of ν_t . Thus, in a general setting, the premium is increasing in the variance, decreasing in the skewness, and increasing in the excess kurtosis.

In the case of our model for the S&P 500 index, for which we obtain negative skewness and significantly excess kurtosis, equation (31) shows that the higher the value of ν_t the larger the premium required by investors for holding this particular risky asset. Panel (d) of Figure 3.1.2 plots the combined effect of the three terms with a solid line. Note also, that if we were to neglect these higher order moment properties of the return process and assume them equal to zero, then to compensate for this a higher value of ν_t would be required to generate the same overall level of compensation for risk. This is clear from the dotted line

in Panel (d) of Figure 3.1.2 which shows the effect without the higher order terms.

3.2 Risk neutral distribution in the mixed normal heteroskedasticity model

Once a sequence ν_t has been obtained, in the second step in the riskneutralization strategy, the risk neutral distribution are obtained from (15). For this, we need the conditional cumulant generating function for the mixed normal heteroskedastic model which from equation (9), becomes

$$\Psi_t(u) = \ln \left(\sum_{k=1}^K \pi_i \exp \left(-u\mu_k + \frac{u^2\sigma_{k,t}^2}{2} \right) \right). \quad (32)$$

Using this, the conditional cumulant generation function of ε_t under the risk neutral measure Q is easily obtained as

$$\begin{aligned} \Psi_t^Q(u) &= \Psi_t(\nu_t + u) - \Psi_t(\nu_t) \\ &= \ln \left(\frac{\sum_{k=1}^K \pi_i \exp \left(-(\nu_t + u)\mu_k + \frac{(\nu_t + u)^2\sigma_{k,t}^2}{2} \right)}{\sum_{k=1}^K \pi_i \exp \left(-\nu_t\mu_k + \frac{\nu_t^2\sigma_{k,t}^2}{2} \right)} \right) \\ &= \ln \left(\sum_{k=1}^K \pi_i^* \exp \left(-u\mu_k^* + \frac{u^2\sigma_{k,t}^2}{2} \right) \right), \end{aligned} \quad (33)$$

where

$$\mu_{k,t}^* = \mu_k - \nu_t\sigma_{k,t}^2, \quad (34)$$

and

$$\pi_{k,t}^* = \frac{\pi_k \exp \left(-\nu_t\mu_k + \frac{\nu_t^2\sigma_{k,t}^2}{2} \right)}{\sum_{k=1}^K \pi_k \exp \left(-\nu_t\mu_k + \frac{\nu_t^2\sigma_{k,t}^2}{2} \right)}, \quad (35)$$

for $k = 1, \dots, K$. Thus, the risk neutral distribution of ε_t remains within the family of normal mixtures. In fact, in the special case with no unit risk premium, the above equations show that the risk neutral distribution correspond to the original distribution. In general, however, under Q the distribution of ε_t will have changed means and probabilities.

We remark that for the risk neutral distribution, the weak stationarity condition in (7) is not appropriate anymore, since the finite mixture under Q has time varying distribution probabilities π_t^* . Therefore, it can in principle occur, that the physical distribution is weakly stationary, but the risk neutral distribution is not.

3.2.1 Interpreting the impact of riskneutralization

For simplicity, we now consider the special case where $K = 2$. With respect to the risk neutral means from (34), it is immediately seen that the correction is very similar to what is obtained with the Gaussian model, where the mean of ε_t under Q is equal to $r_t - \mu_t = -\nu_t \sigma_t^2$. The intuition behind this is the following: If volatility risk carries a positive premium, then in the risk neutral world the mean of the innovations is shifted downwards to compensate for this.

With respect to the risk neutralized probabilities, the relationship is somewhat less straightforward. However, when $K = 2$ (35) simplifies to

$$\pi_t^* = \frac{\pi \exp\left(-\nu_t \mu_1 + \frac{\nu_t^2 \sigma_{1,t}^2}{2}\right)}{\pi \exp\left(-\nu_t \mu_1 + \frac{\nu_t^2 \sigma_{1,t}^2}{2}\right) + (1 - \pi) \exp\left(-\nu_t \mu_2 + \frac{\nu_t^2 \sigma_{2,t}^2}{2}\right)}. \quad (36)$$

Moreover, by dividing through with $\exp\left(-\nu_t \mu_1 + \frac{\nu_t^2 \sigma_{1,t}^2}{2}\right)$, it is seen that

$$\pi_t^* \begin{matrix} \geq \\ \leq \end{matrix} \pi \text{ if } \exp\left(-\nu_t \mu_1 + \frac{\nu_t^2 \sigma_{1,t}^2}{2}\right) \begin{matrix} \geq \\ \leq \end{matrix} \exp\left(-\nu_t \mu_2 + \frac{\nu_t^2 \sigma_{2,t}^2}{2}\right). \quad (37)$$

Thus, if we further assume that the variance term $\frac{\nu_t^2 \sigma_{k,t}^2}{2}$ is negligible compared to the mean term $-\nu_t \mu_k$, for $k = 1, 2$, this restriction simplifies to

$$\pi_t^* \begin{matrix} \geq \\ \leq \end{matrix} \pi \text{ if } \nu_t \mu_2 \begin{matrix} \geq \\ \leq \end{matrix} \nu_t \mu_1. \quad (38)$$

However, by construction $\mu_2 < \mu_1$ and hence the effect on the probabilities depends entirely on the sign of the unit risk premium. Since we expect this premium to be positive, it follows that we will in general have that $\pi_t^* < \pi$. The intuition behind this result is the following: If volatility risk carries a positive premium then the probability attributed to the explosive part, that is $1 - \pi_t^*$, is increased in the risk neutral world to compensate appropriately for this.

4 Bayesian Inference

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. Polson and Johannes (2009) explain how to estimate equity price models, driven for example by stochastic volatility and jumps, using Markov Chain Monte Carlo (MCMC) methods.

In order to learn about the model parameters of the mixed normal heteroskedasticity 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.

In the next section, we explain the Gibbs sampler and its blocks to perform inference on the parameters of the mixed normal heteroskedasticity model. Then we explain how to compute predictive densities and to price options.

4.1 Gibbs sampling

The mixture model is defined in (1) for the innovations ε_t only to leave the specification of the conditional mean for option pricing. As required by (27), the model for the returns is

defined as

$$\begin{aligned} R_t &= r_t - \Psi_t(\nu_t - 1) + \Psi_t(\nu_t) + \varepsilon_t \\ &= \rho_t(\nu) + \varepsilon_t. \end{aligned} \tag{39}$$

That is, we consider ν_t as constant although this can be specified otherwise if desired by the econometrician. Note that a constant ν does not imply a constant risk premia as (31) shows. The likelihood of the model for T observations is given by

$$\mathcal{L}(\xi | R) = \prod_{t=1}^T \sum_{k=1}^K \pi_k \phi(R_t | \mu_k + \rho_t(\nu), \theta_k), \tag{40}$$

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

A direct evaluation of the likelihood function is difficult 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 $G_t \in \{1, 2, \dots, K\}$ 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

$$\varphi(G^T | \pi) = \prod_{t=1}^T \varphi(G_t | \pi) = \prod_{t=1}^T \pi_{G_t}, \tag{41}$$

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

$$\mathcal{L}(\xi | G^T, R) = \prod_{t=1}^T \pi_{G_t} \phi(R_t | \mu_{G_t} + \rho_t(\nu), \theta_{G_t}), \tag{42}$$

which is easier to evaluate than (40) 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, Polson, and Rossi (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 $G^T = (G_1, \dots, G_T)'$, $\pi = (\pi_1, \pi_2, \dots, \pi_K)$, $\mu = (\mu_1, \mu_2, \dots, \mu_K)$, $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ and the parameter ν . The joint posterior distribution is given by

$$\varphi(G^T, \nu, \mu, \theta, \pi | R) \propto \varphi(\nu) \varphi(\mu) \varphi(\theta) \varphi(\pi) \prod_{t=1}^T \pi_{G_t} \phi(R_t | \mu_{G_t} + \rho_t(\nu), \theta_{G_t}), \quad (43)$$

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

- $\varphi(G^T | \nu, \mu, \theta, \pi, R)$

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

$$\varphi(G^T | \nu, \mu, \theta, \pi, R) = \prod_{t=1}^T \varphi(G_t | \nu, \mu, \theta, \pi, R). \quad (44)$$

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

$$P(G_t = k | \nu, \mu, \theta, \pi, R) = \frac{\pi_k \phi(R_t | \mu_k + \rho_t(\nu), \theta_k)}{\sum_{j=1}^K \pi_j \phi(R_t | \mu_j + \rho_t(\nu), \theta_j)}, \quad (k = 1, \dots, K). \quad (45)$$

To sample G_t , we draw one observation from a uniform distribution on $[0, 1]$ and decide which group k to take according to (45).

- $\varphi(\pi | G^T, \nu, \mu, \theta, R)$

The full conditional posterior density of π depends only on G^T and is given by

$$\varphi(\pi | G^T) \propto \varphi(\pi) \prod_{k=1}^K \pi_k^{x_k}, \quad (46)$$

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 $a_k = a_{k0} + x_k$, $k = 1, 2, \dots, K$.

- $\varphi(\mu|G^T, \nu, \pi, \theta, R)$

The conditional distribution of $\tilde{\mu} = (\mu_1, \mu_2, \dots, \mu_{K-1})'$ is Gaussian with mean $-A^{-1}b$ and covariance matrix A^{-1} where

$$A = \text{diag} \left(\sum_{(1)} \frac{1}{\sigma_{1,t}^2}, \dots, \sum_{(K-1)} \frac{1}{\sigma_{K-1,t}^2} \right) + \frac{\tilde{\pi} \tilde{\pi}'}{\pi_K^2} \sum_{(K)} \frac{1}{\sigma_{K,t}^2}, \quad (47)$$

and

$$b = \begin{bmatrix} \frac{\pi_1}{\pi_K} \sum_{(K)} \frac{\varepsilon_t}{\sigma_{K,t}^2} - \sum_{(1)} \frac{\varepsilon_t}{\sigma_{1,t}^2} \\ \vdots \\ \frac{\pi_{K-1}}{\pi_K} \sum_{(K)} \frac{\varepsilon_t}{\sigma_{K,t}^2} - \sum_{(K-1)} \frac{\varepsilon_t}{\sigma_{K-1,t}^2} \end{bmatrix}, \quad (48)$$

where $\tilde{\pi} = (\pi_1, \dots, \pi_{K-1})$ 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 (3).

- $\varphi(\theta|G^T, \nu, \mu, \pi, R)$

By assuming prior independence between the θ_k 's, that is $\varphi(\theta) = \prod_{k=1}^K \varphi(\theta_k)$, it follows that

$$\begin{aligned} \varphi(\theta|G^T, \nu, \mu, \pi, R) &= \varphi(\theta|G^T, \nu, \mu, R) \\ &= \varphi(\theta_1|\nu, \mu_1, \tilde{R}^1) \varphi(\theta_2|\nu, \mu_2, \tilde{R}^2) \cdots \varphi(\theta_K|\nu, \mu_K, \tilde{R}^K) \end{aligned} \quad (49)$$

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

$$\varphi(\theta_k|\nu, \mu_k, \tilde{R}^k) \propto \varphi(\theta_k) \prod_{t \in G_t=k} \phi(R_t|\mu_k + \rho_t(\nu), \theta_k). \quad (50)$$

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, Lubrano, and Richard (1999)). We take bounded uniform supports for ω_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.

- $\varphi(\nu|G^T, \mu, \pi, \theta, R)$

The conditional posterior distribution for this block does not belong to a known family of distributions. Since ν is a scalar, 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 . We use a diffuse prior for ν .

Convergence of the Gibbs sampler is checked with CUMSUM plots of the parameter draws as explained for example in Bauwens, Lubrano, and Richard (1999). To ensure a good precision of posterior moments in this paper, we take $N = 20,000$ and a warm up of 5,000 draws to ensure convergence to the target joint posterior distribution and to eliminate the impact of the starting values on the final results. All of the posterior results and the option prices in this paper are obtained within hours on a standard desktop computer.

4.2 Predictive densities and option prices

To compute option prices, we first need to have predictive return densities under the Q measure. The predictive density of R_{T+1} under the risk neutral measure Q is given by

$$f^Q(R_{T+1} | R) = \int f^Q(R_{T+1} | \xi, R) \varphi(\xi | R) d\xi \quad (51)$$

where $f^Q(R_{T+1} | \xi, R) = \sum_{k=1}^K \pi_k^* \phi(R_{T+1} | \mu_k^* + \rho_{T+1}(\nu), \theta_k)$ as implied by the finite mixture distribution defined in (1). Unlike prediction in the classical framework, note that predictive densities take into account parameter uncertainty by construction while integrating the predictive likelihood over the parameter space. An analytical solution to (51) is unavailable but extending the algorithm of Geweke (1989b), it can be approximated by

$$\widehat{f}^Q(R_{T+1} | R) = \frac{1}{N} \sum_{j=1}^N \left(\sum_{k=1}^K \pi_k^{*(j)} \phi \left(R_{T+1} | \mu_k^{*(j)} + \rho_{T+1}(\nu), \theta_k^{(j)}, R \right) \right) \quad (52)$$

where the superscript (j) 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 $\xi^{(j)} \sim \varphi(\xi | R)$, we simulate $R_{T+1}^{(j)} \sim f^Q(R_{T+1} | \xi^{(j)}, R)$. 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

$$\begin{aligned}
f^Q(R_{T+s} | R) &= \int \left[\int \int \dots \int f^Q(R_{T+s} | R_{T+s-1}, \dots, R_{T+1}, R, \xi) \times \right. \\
&\quad f^Q(R_{T+s-1} | R_{T+s-2}, \dots, R_{T+1}, R, \xi) \times \\
&\quad \dots \times \\
&\quad \left. f^Q(R_{T+1} | R, \xi) dR_{T+s-1} dR_{T+s-2} dR_{T+1} \right] \varphi(\xi | R) d\xi, \quad (53)
\end{aligned}$$

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. These draws 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 , which ignores the estimation uncertainty.

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 option example in (12), we obtain

$$C_T(S, T^*, K) \approx e^{-r(T^*-T)} \frac{1}{N} \sum_{j=1}^N \max \left(S_T \exp \left(\sum_{i=T}^{T^*} R_i^{(j)} \right) - K, 0 \right), \quad (54)$$

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.

5 Pricing S&P 500 options

We now take our model to the data and evaluate its out-of-sample forecast performance by pricing a large sample of options. We compare the performance to a benchmark Gaussian GARCH model, which is a special case of our MNGARCH model with $K = 1$. This benchmark model has been used extensively in the literature, and it has been shown to improve significantly on the performance of the constant volatility BSM model for example. Note

that the constant volatility mixture model, which allows for non zero skewness and excess kurtosis, also performs worse than this benchmark. We therefore refrain from reporting results on these two unconditional models.

Until now, we have considered the general formulation of the MNGARCH model without fixing K , the number of distributions in the finite mixture. However, when applying the model empirically a choice has to be made about K , and this is the first issue we investigate. It turns out that for our sample, there is no evidence for adding a third distribution. In fact, we find that the probability to be in the last distribution is found to be indistinguishable from zero for $K = 3$. Therefore, $K = 2$ is considered as optimal, and in the following we report results for this case only.

In the next section, we describe the option data as well as the data we use for inference purposes in detail. We then proceed to report detailed results for the parameter estimates for the first week for which option data is available. The last section contains the overall pricing results using the full sample of 8,637 option contracts on the S&P 500 index in 2006.

5.1 Data

In the present paper, we use option data from the CBOE provided by Market Data Express, return and dividend data for the S&P 500 index from Datastream, and interest data from the H.15 Federal Reserve Statistical Release, all of which we now describe in detail.

5.1.1 Option data

The option data set contains one daily observation for each option contract on the S&P 500 index traded at the CBOE. We work with option data for 2006 for which approximately 180,000 observations are available. Thus, for this to be manageable we impose the following restrictions. First, we consider weekly data and choose the options traded on Wednesdays. 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 those contracts which had a daily volume of trades larger than or equal to 5. Thirdly, we exclude options which have a midquote

Table I: Properties of the S&P 500 index options data set (Calls)

Mean price in USD, mean ISD, and number of contracts in the cells of this table. The maturity (columns) and moneyness (lines) are defined in the text.

	Time to maturity in days					
	VST	ST	MT	LT	VLT	ALL
DOTM	-	\$ 0.86	\$ 1.76	\$ 5.20	\$ 12.33	\$ 4.54
	-	0.0988	0.0946	0.1021	0.1090	0.1001
	0	60	135	103	68	366
OTM	\$ 1.32	\$ 2.84	\$ 6.69	\$ 17.86	\$ 29.81	\$ 7.50
	0.0898	0.0910	0.0969	0.1103	0.1181	0.0967
	107	350	229	93	68	847
ATM	\$ 11.67	\$ 17.45	\$ 25.34	\$ 39.30	\$ 51.13	\$ 21.70
	0.0969	0.1041	0.1120	0.1223	0.1281	0.1072
	363	450	360	127	68	1368
ITM	\$ 42.87	\$ 45.81	\$ 52.21	\$ 63.11	\$ 76.54	\$ 52.40
	0.1225	0.1231	0.1253	0.1307	0.1370	0.1261
	148	144	109	65	70	536
DITM	\$ 117.41	\$ 117.46	\$ 135.02	\$ 125.57	\$ 122.61	\$ 123.65
	0.1342	0.1448	0.1561	0.1534	0.1485	0.1475
	81	136	115	65	80	477
ALL	\$ 28.95	\$ 27.60	\$ 33.87	\$ 42.94	\$ 60.76	\$ 24.95
	0.1056	0.1070	0.1128	0.1209	0.1289	0.1122
	699	1140	948	453	354	3594

below half a dollar, and we use the midquote as the price. 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 8,637 options of which 3,594 are calls and 5,043 are puts.

In Table I, we provide descriptive statistics for the call option in terms of the number of options, the average prices as well as average implied standard deviations (ISD) from the BSM model. We tabulate data for various categories of maturity measured in trading days, T^* , and moneyness measured as $M = S / (K \exp(-rT^*))$, where S is the value of the

underlying, K is the strike price, and r is the risk free interest rate. The maturity categories 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 $T^* \geq 169$. For the call options, the moneyness categories 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 $M \geq 1.05$. Table II contains similar descriptive statistics for these categories for the put options. For the put options, the moneyness categories are inverted, for example the DOTM put options have $M \leq 1.05$.

Besides the overrepresentation of DOTM put options, Tables I and II show that our sample contains a diverse sample of traded options. First of all, in terms of the number of contracts even the VLT category contains a large number and so does the DITM category. Naturally, most of the options in our sample are ATM options with ST maturity. Next, in terms of option prices, the two tables show that the mean of these vary from \$0.86 to \$135.02 for the call options and from \$1.20 to \$142.04 for the put options thus spanning a very large interval. Finally and most importantly, we observe the well known volatility smirk which is present for both call and put options. To be specific, Table I shows that the mean ISD is higher for DITM options than for DOTM options and Table II shows that the mean ISD is higher for DOTM options than for DITM options.

5.1.2 Return, dividend and interest rate data

As return data, we use a total return index (Datastream data type RI), which is calculated under the assumption that dividends are re-invested. Figure 2 displays the sample path, the normal quantile plot, and the autocorrelation functions for returns and for squared returns. Panel (a) of this figure shows the well known pattern of time varying volatility with periods of high volatility levels followed by periods of low levels of volatility, which is often observed in financial data. Panel (b) clearly shows that returns are far from being Gaussian and instead have fat tails. Finally, Panel (c) provides evidence of the lack of persistency in raw returns, i.e. the first moment of the series, whereas Panel (d) provides evidence of the strong persistency in squared returns, i.e. a proxy of the second moment of the series.

Table II: Properties of the S&P 500 index options data set (Puts)

Mean price in USD, mean ISD and number of contracts in the cells of this table. The maturity (columns) and moneyness (lines) are defined in the text.

	Time to maturity in days					
	VST	ST	MT	LT	VLT	ALL
DITM	\$ 99.12 0.0999 24	\$ 123.69 0.0936 24	\$ 140.04 0.0839 32	\$ 132.87 0.0955 20	\$ 142.04 0.1020 16	\$ 127.23 0.0937 116
ITM	\$ 43.37 0.0746 87	\$ 49.33 0.0966 82	\$ 57.75 0.1054 45	\$ 74.83 0.1104 31	\$ 94.64 0.1111 12	\$ 53.98 0.0931 257
ATM	\$ 13.13 0.0985 352	\$ 20.21 0.1090 398	\$ 28.20 0.1170 306	\$ 49.07 0.1256 103	\$ 68.49 0.1282 39	\$ 24.22 0.1100 1198
OTM	\$ 2.73 0.1280 251	\$ 8.03 0.1285 315	\$ 15.84 0.1295 229	\$ 34.41 0.1338 120	\$ 50.37 0.1357 58	\$ 14.28 0.1297 973
DOTM	\$ 1.20 0.1775 267	\$ 2.52 0.1711 715	\$ 5.20 0.1713 637	\$ 11.42 0.1745 521	\$ 21.85 0.1663 359	\$ 7.69 0.1718 2499
ALL	\$ 12.00 0.1255 981	\$ 12.63 0.1410 1534	\$ 18.13 0.1457 1249	\$ 25.29 0.1575 795	\$ 34.80 0.1561 484	\$ 18.00 0.1432 5043

Table III provides the standard descriptive statistics for the S&P 500 index return series. The numbers in this table show that the return data is slightly negatively skewed and very leptokurtic. The classical Jarque-Bera test rejects the null of normality at any level of significance due to the high level of kurtosis. However, the individual test for skewness results in a p-value of 1.15%.

For pricing purposes, we need an estimate of future dividend payments since options are written on the actual index value and not on the total return index. For this, we choose to use the Dividend Yield (Datastream data type DY). This series is calculated as a rolling

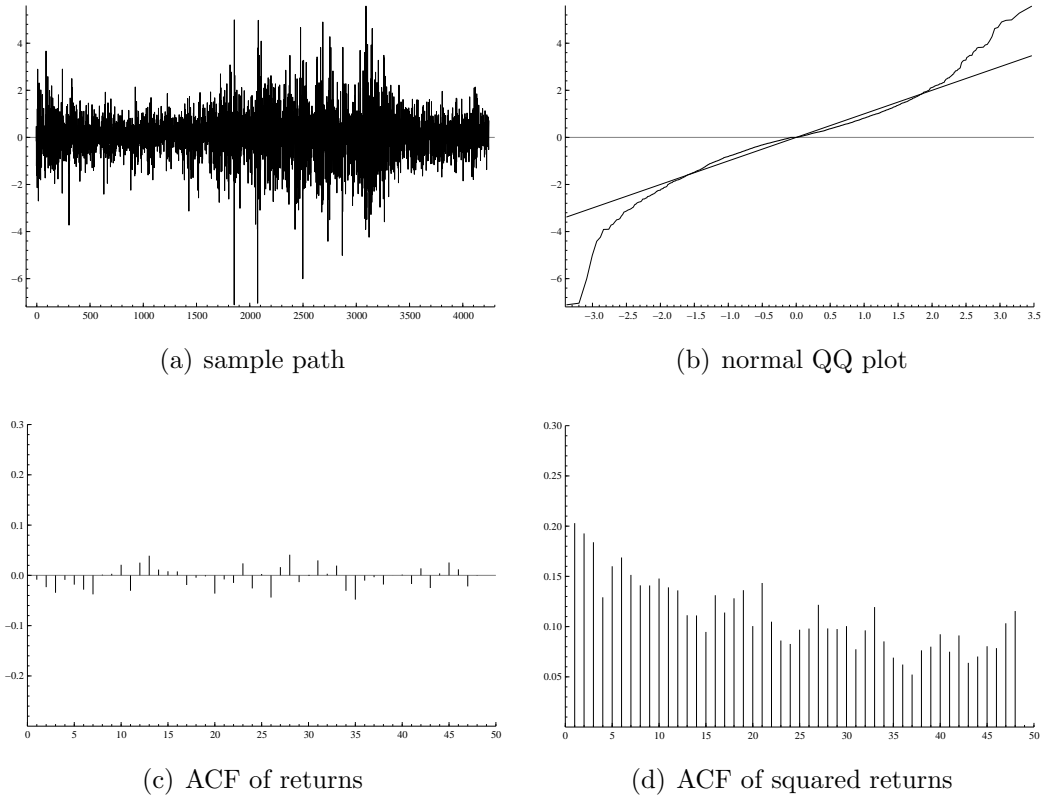


Figure 2: Properties of the S&P 500 returns

This figure plots the sample path, QQ plot, ACF's of returns and squared returns using a sample period from September 11, 1990 to December 27, 2006 (4,252 observations).

average of the daily yields for the previous year. Panel (a) of Figure 3 displays the sample paths for this series. From this, it is seen that while the dividend yield changes through time, for the last part of the period which is used for pricing, it is quite stable. Thus, in the pricing part, we simply set the future dividend equal to the historical dividend yield at the day of pricing.

Finally, in order to perform inference and forecast option prices, we need a series of interest rates. We use the series of one month Eurodollar deposit rates provided by the H.15 Federal Reserve Statistical Release. Panel (b) of Figure 3 displays the sample paths for the daily interest rates. While the plot clearly shows that this series changes over time, it may also be observed that during the last part of the sample, it remains relatively constant. Since this is the period over which we will be pricing options, we simply assume that the interest rate remains constant over the life of the option, and set it equal to the prevailing rate at

Table III: Descriptive statistics for S&P 500 index percentage returns

Sample period: September 11, 1990 to December 27, 2006 (4,252 observations)

Mean	0.042963	Maximum	5.5754
Standard deviation	0.97980	Minimum	-7.1130
Skewness	-0.094923	Kurtosis	7.2501

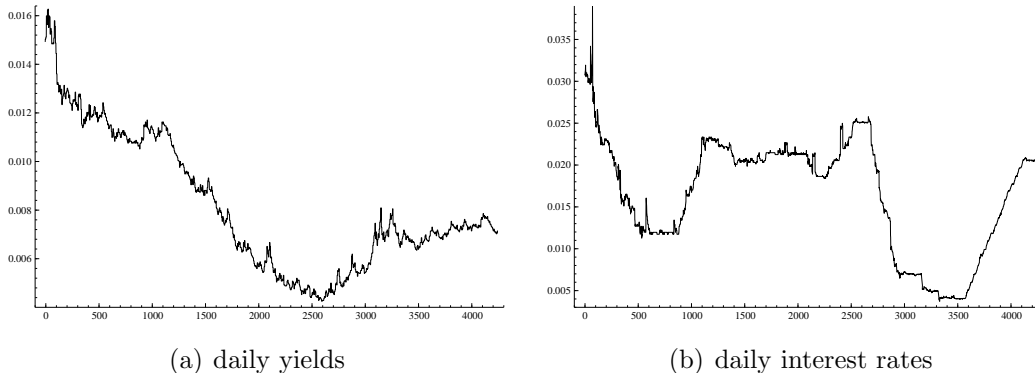


Figure 3: Sample paths of the dividend yield and interest rate

This figure plots the sample paths of the dividend yields and interest rates using a sample period from September 11, 1990 to December 27, 2006 (4,252 observations).

the day of pricing.

5.2 Posterior results

Table IV displays posterior moments for the GARCH and MNGARCH models. The results are based on the data used for pricing of week one in the option pricing data. As it is often found in the literature, the posterior means of the GARCH model imply a highly persistent variance process since $\alpha_1 + \beta_1 = 0.995$. For the MNGARCH model, we see that the second distribution has an explosive variance process since $0.237 + 0.869 > 1$. Given that the probability to be in this state is small, i.e. $1 - \pi = 0.246$, the second order stationarity condition stated in equation (7) is still met. It is thanks to this distribution that we can accommodate for example the high kurtosis in the returns of the underlying.

Table V informally compares the moments of the data with the moments implied by the posterior mean parameters of the GARCH and MNGARCH models. From this table, we

Table IV: Posterior means and standard deviations

Week 1. September 5, 1990 to January 3, 2006 (4,000 observations). A - symbol means that the parameter is set to 0. Posterior moments based on 20,000 draws from the Gibbs sampler defined in Section 4.

	GARCH		MNGARCH	
	mean	standard deviation	mean	standard deviation
ν	6.139	(1.081)	4.534	(1.348)
μ	-	-	0.001	(0.0001)
π	-	-	0.754	(0.082)
ω_1	5.1E-7	(9.9E-7)	6.8E-8	(5.8E-8)
α_1	0.054	(0.007)	0.018	(0.006)
β_1	0.941	(0.008)	0.970	(0.008)
ω_2	-	-	4.3E-6	(2.0E-6)
α_2	-	-	0.237	(0.071)
β_2	-	-	0.869	(0.031)

see that the GARCH model can match the variance but clearly not the kurtosis of the data. The MNGARCH model, on the other hand, matches both the standard deviation and the kurtosis nicely. With respect to the skewness, it should be noted that, given the size of the estimates, these are both insignificantly different from zero.

With respect to the unit risk premium parameter ν , Table IV shows that the posterior mean of ν is centered around 6.139 for the GARCH model and 4.534 for the MNGARCH model. Figure 4 plots the posterior marginals for ν and shows that they are symmetrically

Table V: Empirical and implied moments of the returns (week 1)

Simulated moments using 50,000 draws for the GARCH and MNGARCH models.

	Standard deviation	Skewness	Kurtosis
Empirical	0.9984	-0.09517	7.1077
GARCH	1.0316	0.06911	4.5840
MNGARCH	1.0143	0.10518	7.4169

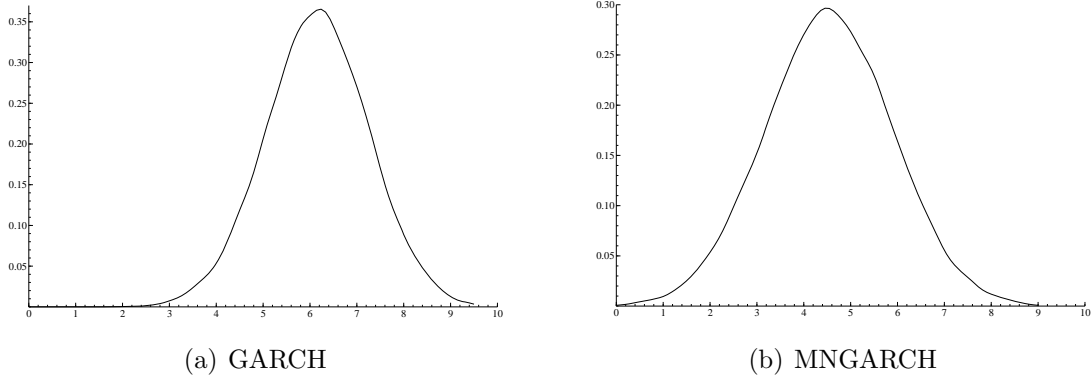


Figure 4: Posterior marginal densities of ν

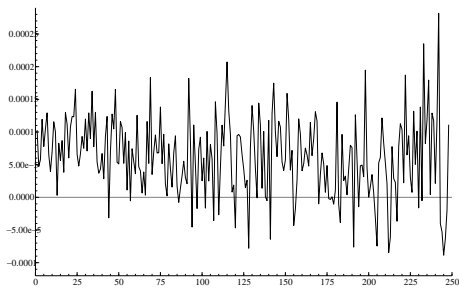
This figure plots the posterior marginal densities of the unit risk premium ν using 20,000 draws from the Gibbs sampler.

distributed away from zero. Comparing the size of the estimates, we note that while it may seem that the unit risk premium is larger for the GARCH model than for the MNGARCH model this, in fact, is to be expected as this model neglects the impact of higher order conditional moments. In particular, as it was mentioned above the higher order terms in (31) varies negatively with conditional skewness and positively with conditional excess kurtosis for a given level of ν . Hence, when these terms are neglected, a higher ν is required to compensate investors appropriately for the risk of the asset.

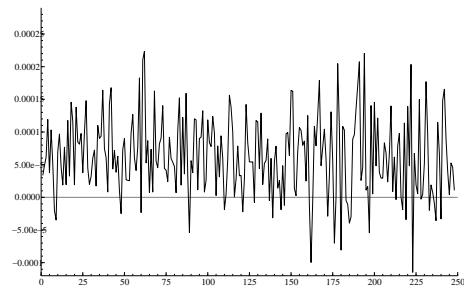
5.3 Option pricing results

For option pricing to proceed, we need the risk neutral dynamics, i.e. the dynamics under Q . As mentioned in the introduction, our method allows us to derive these using only historical data on returns and hence no calibration to historical option prices is needed. Figure 5 displays the term structure of the mean, variance, skewness and kurtosis of the predicted returns under Q until the highest maturity, i.e. 248 trading days, for week 1 for the GARCH and MNGARCH models.

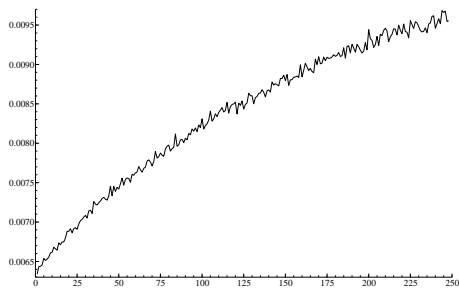
While the first two moments are similar for the two models, this is clearly not the case for the skewness and the kurtosis. In particular, for the GARCH model skewness is zero as expected, whereas the risk neutral dynamics of the MNGARCH model has pronounced negative skewness. Likewise, for the kurtosis large differences are also observed as the level



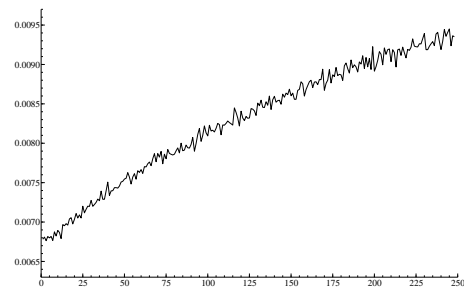
(a) mean GARCH



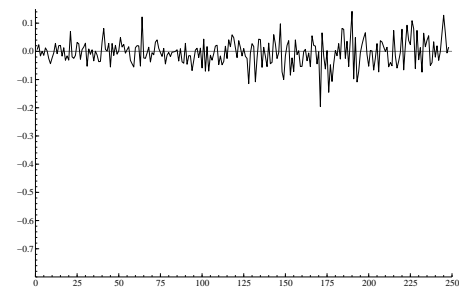
(b) mean MNGARCH



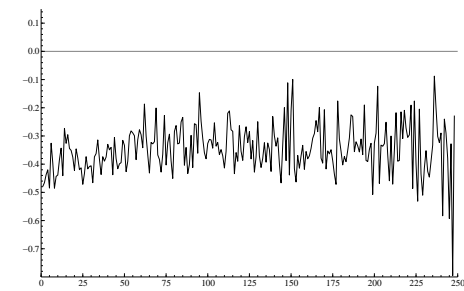
(c) variance GARCH



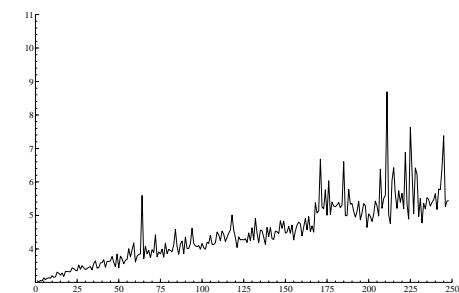
(d) variance MNGARCH



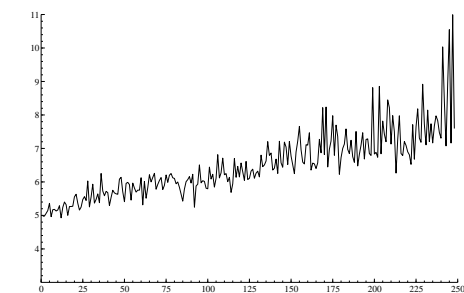
(e) skewness GARCH



(f) skewness MNGARCH



(g) kurtosis GARCH



(h) kurtosis MNGARCH

Figure 5: Week 1 term structure under Q

This figure plots the term structure of the first four moments of the predictive return densities under Q using 20,000 draws from the Gibbs sampler.

is much higher for the MNGARCH model than for the GARCH model. These features are related to the explosive volatility process in the second distribution of the finite mixture which in the risk neutral dynamics becomes even more important as $\pi_t^* < \pi$.

It is well known that skewness and excess kurtosis are important for option pricing purposes, and will potentially affect different categories of options. In particular, excess kurtosis increases the predicted value of both DITM and DOTM options and lowers predicted prices of ATM options. The large negative skewness will, everything else equal, contribute positively to OTM and DOTM put options and ITM and DITM call options. Table I and II show that both of these effects are relevant for our sample of options.

5.3.1 Dollar losses

We now proceed to compare out-of-sample our MNGARCH model to the classical GARCH model in terms of actual prices. Since we price options for 52 weeks in 2006, we update the return data on a weekly basis to compute the predictive price densities. We could also update the posterior densities weekly, but given that our inference is based on the last 4,000 observations neither the parameters nor the predictive densities would change substantially when doing this. Table VI provides dollar losses for the calls, calculated as the observed price minus the predicted price, whereas Table VII provides dollar losses for the puts. The tables also report the root mean squared error (RMSE). In both tables, the top panels provide the results for the GARCH model whereas the results for the MNGARCH model are found in the bottom panels. In all tables and panels, we categorize the losses according to the maturity and moneyness as defined in Section 5.1.2.

Examining first the overall results, we see that there are pronounced differences between the performance of the two models. Using the dollar losses, it appears that the GARCH models performs the best for call options whereas the MNGARCH models perform the best for put options. For the call options, Table VI shows a clear pattern across both moneyness and maturity. In particular, when comparing the two panels, it is only for the ATM category that the GARCH model has the smallest losses. For the DOTM options with VLT maturity, the MNGARCH model improves on the GARCH model with an average \$1.92. In percentage of the mean option prices, the improvement is as high as 37% for the DOTM options with

Table VI: Dollar losses for Calls

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 5.1.2.

		GARCH					
		VST	ST	MT	LT	VLT	ALL
DOTM	-	-1.532	-2.477	-4.772	-6.836	-3.778	
	-	1.866	2.839	5.205	7.251	4.576	
OTM	-1.443	-1.977	-2.580	-3.430	-4.520	-2.436	
	1.768	2.443	3.185	4.306	5.320	3.148	
ATM	-0.756	-0.454	0.421	0.355	-0.593	-0.236	
	1.632	2.157	2.756	3.145	3.525	2.402	
ITM	0.933	1.868	3.241	3.703	4.191	2.415	
	1.556	2.730	4.123	4.707	5.311	3.535	
DITM	0.477	1.428	3.467	5.803	6.913	3.274	
	1.385	2.228	4.325	6.539	7.580	4.660	
ALL	-0.361	-0.461	-0.023	-0.325	0.095	-0.092	
	1.612	2.321	3.272	4.696	6.041	3.432	
		MNGARCH					
		VST	ST	MT	LT	VLT	ALL
DOTM	-	-1.211	-2.015	-3.540	-4.915	-2.851	
	-	1.453	2.269	3.860	5.259	3.402	
OTM	-1.260	-1.794	-2.261	-2.519	-2.896	-2.021	
	1.523	2.166	2.771	3.495	3.806	2.610	
ATM	-0.860	-0.639	0.235	0.806	0.780	-0.263	
	1.639	2.087	2.668	3.158	3.311	2.343	
ITM	0.621	1.361	2.495	3.402	4.731	2.075	
	1.493	2.361	3.365	4.299	5.551	3.268	
DITM	0.417	1.200	2.971	5.481	6.748	3.008	
	1.368	2.150	3.882	6.245	7.216	4.395	
ALL	-0.460	-0.552	-0.096	0.179	1.110	-0.158	
	1.562	2.128	2.902	4.116	5.297	3.001	

Table VII: Dollar losses for Puts

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 5.1.2.

		GARCH					
		VST	ST	MT	LT	VLT	ALL
DITM	-0.208	-0.142	-1.624	-4.340	-8.151	-2.393	
	1.262	1.789	2.517	4.899	8.601	4.133	
ITM	-1.029	-1.612	-2.217	-2.744	-6.188	-1.871	
	1.803	2.708	3.288	3.454	6.627	2.970	
ATM	-0.562	0.122	1.126	0.274	-0.962	0.155	
	1.584	2.138	2.852	3.247	3.806	2.360	
OTM	1.036	2.400	3.645	4.026	3.553	2.610	
	1.405	2.870	4.248	4.794	4.695	3.402	
DOTM	0.995	1.829	3.223	5.074	6.418	3.431	
	1.196	2.189	3.771	5.752	7.202	4.416	
ALL	0.238	1.289	2.466	3.752	4.686	2.174	
	1.456	2.357	3.620	5.252	6.768	3.831	
		MNGARCH					
		VST	ST	MT	LT	VLT	ALL
DITM	-0.318	-0.166	-1.117	-2.797	-6.441	-1.778	
	1.228	1.929	2.131	3.646	6.972	3.366	
ITM	-0.925	-1.308	-1.729	-1.830	-4.590	-1.468	
	1.703	2.549	2.974	3.067	4.885	2.617	
ATM	-0.699	-0.118	-0.916	1.137	0.589	0.106	
	1.563	2.051	2.731	3.081	3.402	2.283	
OTM	0.724	1.828	2.951	3.973	4.055	2.205	
	1.203	2.441	3.628	4.601	4.853	3.069	
DOTM	0.868	1.554	2.728	4.508	5.901	3.021	
	1.084	1.938	3.271	5.184	6.565	3.965	
ALL	0.081	0.997	2.066	3.559	4.584	1.831	
	1.364	2.114	3.183	4.768	6.158	3.384	

ST maturity.

For the put options, the MNGARCH model outperforms the GARCH model and as Table VII shows the losses are approximately 16% or \$0.34 smaller. This holds true along both dimensions of the data indicating clearly that the flexible dynamics accommodated by this model are important for the pricing of put options. The largest average improvement in dollar losses is \$1.71 for the DITM options with VLT maturity. In percentage of the mean option prices, the improvement is largest for the shorter term options which are out of the money. For several of these categories, this improvement exceeds 10% to 15%.

While the ranking of the models in terms of dollar losses are ambiguous for the call options, the ranking of the models in terms of RMSE is clear. In particular, using this metric the MNGARCH model is the best for both calls and puts. This holds not only overall but also along both moneyness and maturity dimensions. Moreover, the improvements of the MNGARCH model are in several cases of a significant size. Consider, as an example, the DOTM category where the GARCH RMSE equals 4.576 and the MNGARCH RMSE equals 3.402. This corresponds to a reduction of approximately 25%.

5.3.2 Implied standard deviation losses

Whereas the section above considered actual dollar losses, another often used metric for comparing the performance of alternative option pricing models is losses in ISD's. Tables VIII and IX provide these losses for the call and put options, respectively. In both tables, the losses are calculated as the difference between the ISD of the market price and the ISD of the predicted price. Thus, when comparing these numbers across all the options we may firstly gauge whether or not the models generate a sufficiently high level of volatility under the risk neutral measure Q in general. Secondly, if a particular model can explain the smile or smirk, which is documented in Tables I and II for our option sample, losses should be constant across moneyness for the various maturity.

Although the overall losses are not exactly equal to zero, they are in fact quite close to zero. This is so especially for the call options. For the put options, the overall losses are slightly positive and somewhat larger for the GARCH model. This indicates that this model fails to produce a high enough level of volatility under Q , a feature which is well known in

Table VIII: Implied volatility losses for Calls

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 5.1.2.

		GARCH					
		VST	ST	MT	LT	VLT	ALL
DOTM	-	-0.024	-0.026	-0.027	-0.023	-0.025	
	-	0.027	0.028	0.030	0.024	0.028	
OTM	-0.023	-0.020	-0.016	-0.011	-0.011	-0.017	
	0.026	0.023	0.020	0.014	0.012	0.021	
ATM	-0.007	-0.003	0.002	0.001	-0.001	-0.003	
	0.016	0.013	0.012	0.009	0.008	0.013	
ITM	0.017	0.015	0.016	0.010	0.009	0.014	
	0.036	0.022	0.020	0.014	0.012	0.028	
DITM	0.018	0.034	0.048	0.029	0.021	0.032	
	0.085	0.076	0.073	0.035	0.031	0.067	
ALL	-0.002	-0.003	0.001	-0.003	0.001	-0.001	
	0.037	0.032	0.031	0.022	0.020	0.031	
		MNGARCH					
		VST	ST	MT	LT	VLT	ALL
DOTM	-	-0.020	-0.022	-0.021	-0.017	-0.020	
	-	0.022	0.024	0.023	0.018	0.022	
OTM	-0.020	-0.018	-0.014	-0.009	-0.007	-0.015	
	0.023	0.019	0.017	0.012	0.009	0.019	
ATM	-0.008	-0.004	0.001	0.002	0.001	-0.003	
	0.015	0.013	0.012	0.009	0.007	0.013	
ITM	0.010	0.010	0.012	0.009	0.010	0.010	
	0.034	0.018	0.016	0.012	0.012	0.022	
DITM	0.012	0.029	0.039	0.026	0.018	0.027	
	0.090	0.059	0.073	0.036	0.021	0.069	
ALL	-0.004	-0.003	0.000	-0.001	0.002	-0.002	
	0.037	0.033	0.018	0.019	0.015	0.030	

Table IX: Implied volatility losses for Puts

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 5.1.2.

		GARCH					
		VST	ST	MT	LT	VLT	ALL
DITM		0.011	-0.026	-0.041	-0.034	-0.028	-0.025
		0.115	0.076	0.069	0.049	0.031	0.076
ITM		-0.037	-0.018	-0.013	-0.009	-0.015	-0.022
		0.059	0.031	0.019	0.011	0.016	0.040
ATM		-0.006	0.001	0.005	0.001	-0.002	-0.000
		0.017	0.013	0.012	0.009	0.008	0.0137
OTM		0.021	0.020	0.018	0.012	0.008	0.018
		0.025	0.023	0.021	0.014	0.011	0.022
DOTM		0.057	0.053	0.052	0.043	0.031	0.048
		0.063	0.058	0.058	0.048	0.035	0.054
ALL		0.016	0.028	0.029	0.029	0.022	0.025
		0.045	0.043	0.044	0.040	0.031	0.042
		MNGARCH					
		VST	ST	MT	LT	VLT	ALL
DITM		-0.007	-0.032	-0.037	-0.027	-0.023	-0.026
		0.121	0.090	0.061	0.043	0.025	0.078
ITM		-0.036	-0.016	-0.010	-0.006	-0.011	-0.020
		0.059	0.031	0.014	0.010	0.011	0.040
ATM		-0.007	-0.001	0.004	0.003	0.001	-0.001
		0.016	0.013	0.011	0.009	0.007	0.013
OTM		0.013	0.014	0.015	0.011	0.009	0.014
		0.019	0.019	0.018	0.013	0.011	0.018
DOTM		0.041	0.038	0.038	0.034	0.026	0.036
		0.047	0.043	0.042	0.037	0.028	0.040
ALL		0.009	0.019	0.021	0.024	0.019	0.019
		0.038	0.034	0.033	0.032	0.025	0.033

empirical option pricing as highlighted in the introduction. The MNGARCH model, on the other hand, is able to generate a higher volatility under Q than the GARCH model, which reduces the ISD loss with 24%. The conclusions are the same when considering the RMSE's which are 21% smaller for the MNGARCH model than for the GARCH model for the put options.

Inspecting the individual cells in Tables VIII and IX, we see that in most cases the losses are smaller for the MNGARCH model than for the GARCH model. In fact, this is so for 41 of the 49 categories for which options are available. Moreover, this is so for all the DOTM categories and for all but one of the DITM categories (the ST put options). The improvement is larger than 28% for DOTM put options with less than MT maturity. Taken together, the evidence in the two tables indicates that the MNGARCH model is capable of explaining a larger fraction of the smile in ISD's than the GARCH model.

6 Conclusion

In this paper, we perform option pricing using mixed normal heteroscedasticity models. We provide details on how to obtain the appropriate risk neutral dynamics and we suggest a feasible way for option pricing within this general framework. Moreover, we perform inference in a Bayesian framework which allows us to compute easily predictive price densities that take into account parameter uncertainty. We compare our option pricing model to a benchmark model and find substantial improvements in terms of both dollar losses and implied standard deviation losses for a large sample of options on the S&P 500 index.

In terms of the risk neutralization, we show that the risk neutral dynamics stay within the class of mixed normal heteroscedasticity models, although the parameters of the distribution are changed. These risk neutral parameters are easily interpreted as providing investors with compensation for specific distributional features like for example the distribution with an explosive variance process found in our data. Moreover, when comparing the properties of the risk neutral distribution these differ in a pronounced way from the properties under the original measure used for inference.

We document significant difference between our model and the benchmark in terms of

the risk neutral dynamics used for option pricing. In particular, our model is capable of generating negative skewness and significant amounts of excess kurtosis. In terms of pricing performance, our results confirm the importance of both of these features for our sample of index options. Our model performs best for options which are deep out of the money and deep in the money. In terms of explaining the smirk in implied volatilities, improvements are found for all maturities, and this is particularly pronounced when considering the shorter maturities.

There are several interesting extensions for further research. The most obvious will be to extend our data to include the recent period of financial turmoil as data becomes available. We conjecture that this may require additional distributions in the mixture, i.e. increasing K , which is easily accommodated in our framework. With this new data it would be interesting to consider alternative benchmarks which could include discrete jumps in returns and volatility as suggested by Christoffersen, Jacobs, and Ornathanalai (2008). Another extension would be to consider Markov switching models in which returns can have a high or low mean and variance, and switches between these states are determined by a Markov process. Within this framework, it is possible to allow for state dependent unit risk premiums that further drive a wedge between the physical and risk neutral dynamics. Lastly, this paper does not use option prices for inference on the model parameters. We explained in the introduction how complicated this is for our model which is non-affine. However, it would nevertheless be interesting to investigate how the information contained in past option prices can be included in the inference procedure of this paper.

References

- ALBERT, J., AND S. CHIB (1993): “Bayesian Analysis of Binary and Polychotomous Response Data,” *Journal of the American Statistical Association*, 88, 669–679.
- AMIN, K. I., AND V. K. NG (1993): “Option Valuation with Systematic Stochastic Volatility,” *Journal of Finance*, 48(3), 881–910.
- BAKSHI, G., C. CAO, AND Z. CHEN (1997): “Empirical Performance of Alternative Option Pricing Models,” *Journal of Finance*, 52, 2003–2049.

- (2000): “Pricing and Hedging Long-Term Options,” *Journal of Econometrics*, 94, 277–318.
- BAKSHI, G., P. CARR, AND L. WU (2008): “Stochastic Risk Premiums, Stochastic Skewness in Currency Options, and Stochastic Discount Factors in International Economies,” *Journal of Financial Economics*, 87, 132–156.
- BARNDORFF-NIELSEN, O., AND N. SHEPHARD (2001): “Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics,” *Journal of the Royal Statistical Society-Series B*, 63, 167–241.
- BARONE-ADESI, G., R. F. ENGLE, AND L. MANCINI (2008): “A GARCH Option Pricing Model with Filtered Historical Simulation,” *Review of Financial Studies*, 21(3), 1223–1258.
- BATES, D. (1991): “The Crash of ’87: Was it expected? The Evidence from Option Markets,” *Journal of Finance*, 46, 1009–1044.
- BATES, D. (2000): “Post-’87 Crash Fears in the S&P 500 Futures Option Market,” *Journal of Econometrics*, 94, 181–238.
- BATES, D. (2003): “Empirical Option Pricing: a Retrospection,” *Journal of Econometrics*, 116, 387–404.
- BAUWENS, L., AND M. LUBRANO (2002): “Bayesian Option Pricing using Asymmetric GARCH Models,” *Journal of Empirical Finance*, 9, 321–342.
- BAUWENS, L., M. LUBRANO, AND J. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.
- BAUWENS, L., AND J. ROMBOUTS (2007a): “Bayesian Clustering of Many GARCH Models,” *Econometric Reviews*, 26, 365–386.
- (2007b): “Bayesian Inference for the Mixed Conditional Heteroskedasticity Model,” *Econometrics Journal*, 10, 408–425.
- BERTHOLON, H., A. MONFORT, AND F. PEGORARO (2008): “Econometric Asset Pricing Modelling,” *Journal of Financial Econometrics*, 6(4), 407–458.
- BILLINGSLEY, P. (1995): *Probability and Measure*. Wiley-Interscience, New York.
- BLACK, F., AND M. SCHOLES (1973): “The Pricing of Options and Corporate Liabilities,” *Journal of Political Economy*, 81, 637–654.
- BOLLERSLEV, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.

- BOOTHE, P., AND D. GLASSMAN (1987): “The Statistical Distribution of Exchange Rates,” *Journal of International Economics*, 22, 297–319.
- BOUADDI, M., AND J. ROMBOUTS (2009): “Mixed Exponential Power Asymmetric Conditional Heteroskedasticity,” *forthcoming in Studies on Nonlinear Dynamics and Econometrics*.
- CARR, P., H. GEMAN, D. MADAN, AND M. YOR (2003): “Stochastic volatility for Lévy processes,” *Mathematical Finance*, 13, 345–382.
- CARR, P., AND L. WU (2004): “Time-changed Lévy processes and option pricing,” *Journal of Financial Economics*, 71, 113–141.
- CARR, P., AND L. WU (2007): “Stochastic Skew in Currency Options,” *Journal of Financial Economics*, 86, 213–247.
- CHERNOV, M., AND E. GHYSELS (2000): “Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Pricing Evaluation,” *Journal of Financial Economics*, 56, 407–458.
- CHRISTOFFERSEN, P., R. ELKAMHI, B. FEUNOU, AND K. JACOBS (2008): “Option Valuation with Conditional Heteroskedasticity and Non-Normality,” *Working Paper, McGill University*.
- CHRISTOFFERSEN, P., S. HESTON, AND K. JACOBS (2006): “Option valuation with Conditional skewness,” *Journal of Econometrics*, 131, 253–284.
- CHRISTOFFERSEN, P., AND K. JACOBS (2004): “Which GARCH Model for Option Valuation?,” *Management Science*, 50(9), 1204–1221.
- CHRISTOFFERSEN, P., K. JACOBS, C. DORION, AND Y. WANG (2008): “Volatility Components, Affine Restrictions and Non-Normal Innovations,” *Creates Working Paper 2008-10*.
- CHRISTOFFERSEN, P., K. JACOBS, AND C. ORNTHANALAI (2008): “Exploring Time-Varying Jump Intensities: Evidence from S&P500 Returns and Options,” *Manuscript*.
- CHRISTOFFERSEN, P., K. JACOBS, C. ORNTHANALAI, AND Y. WANG (2008): “Option Valuation with Long-run and Short-run Volatility Components,” *Journal of Financial Economics*, 90, 272–297.
- DAVIDSON, J. (1997): *Stochastic Limit Theory*. Oxford University Press, Oxford.
- DING, Z., AND C. GRANGER (1996): “Modeling Volatility Persistence of Speculative Returns: A New Approach,” *Journal of Econometrics*, 73, 185–215.
- DUAN, J. (1995): “The GARCH Option Pricing Model,” *Mathematical Finance*, 5, 13–32.

- (1999): “Conditionally Fat-Tailed Distributions and the Volatility Smile in Options,” *Mimeo, Hong Kong University of Science and Technology*.
- DUAN, J.-C., P. RITCHKEN, AND Z. SUN (2006): “Approximating GARCH-Jump Models, Jump-Diffusion Process, and Option Pricing,” *Mathematical Finance*, 16(1), 21–52.
- DURHAM, G. (2007): “SV mixture models with application to S&P 500 index returns,” *Journal of Financial Economics*, 85, 822–856.
- ENGLE, R., AND G. LEE (1999): *A Permanent and Transitory Component Model of Stock Return Volatility*, pp. 475–497, *Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger*. R.F. Engle and H. White (eds), Oxford University Press.
- ENGLE, R., AND G. NG (1993): “Measuring and Testing the Impact of News on Volatility,” *Journal of Finance*, 48, 1749–1778.
- ERAKER, B. (2004): “Do Stock Prices and Volatility jump? Reconciling Evidence from spot and option Prices,” *Journal of Finance*, 59, 1367–1403.
- FRÜHWIRTH-SCHNATTER, S. (2006): *Finite Mixture and Markov Switching Models*. Springer, New York.
- GARCIA, R., E. GYSELS, AND E. RENAULT (2009): “The Econometrics of Option Pricing,” forthcoming in Ait-Sahalia and Hansen (eds.), *Handbook of Financial Econometrics*. Elsevier.
- GELFAND, A., AND A. SMITH (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.
- (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.
- GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1993): “On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks,” *Journal of Finance*, 48(5), 1779–1801.
- GOURIEROUX, C., AND A. MONFORT (2007): “Econometric Specification of Stochastic Discount Factor Models,” *Journal of Econometrics*, 136, 509–530.
- HAAS, M., S. MITTNIK, AND M. PAOLELLA (2004): “Mixed Normal Conditional Heteroskedasticity,” *Journal of Financial Econometrics*, 2, 211–250.

- HAMILTON, J., T. ZHA, AND D. WAGGONER (2007): “Normalization in Econometrics,” *Econometric Reviews*, 26, 221–252.
- HE, C., A. SILVENNOINEN, AND T. TERASVIRTA (2008): “Parameterizing Unconditional Skewness in Models for Financial Time Series,” *Journal of Financial Econometrics*, 6, 208–230.
- HESTON, S. L. (1993): “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options,” *Review of Financial Studies*, 6(2), 327–343.
- HESTON, S. L., AND S. NANDI (2000): “A Closed-Form GARCH Option Valuation Model,” *Review of Financial Studies*, 13(3), 585–625.
- HSIEH, K. C., AND P. RITCHKEN (2005): “An Empirical Comparison of GARCH Option Pricing Models,” *Review of Derivatives Research*, 8, 129–150.
- HULL, J., AND A. WHITE (1987): “The Pricing of Options on Assets with Stochastic Volatilities,” *Journal of Finance*, 42(2), 281–300.
- JACQUIER, E., AND R. JARROW (2000): “Bayesian Analysis of Contingent Claim Model Error,” *Journal of Econometrics*, 94, 145–180.
- JACQUIER, E., N. POLSON, AND P. ROSSI (1994): “Bayesian analysis of stochastic volatility models (with discussion),” *Journal of Business and Economic Statistics*, 12, 371–417.
- JOHNSON, H., AND D. SHANNO (1987): “Option Pricing When the Variance is Changing,” *Journal of Financial and Quantitative Analysis*, 22(2), 143–151.
- KIM, D., AND S. KON (1994): “Alternative Models for the Conditional Heteroscedasticity of Stock Returns,” *Journal of Business*, 67, 563–598.
- KON, S. (1982): “Models of Stock Returns, a Comparison,” *Journal of Finance*, 39, 147–165.
- KOOP, G. (2003): *Bayesian Econometrics*. Wiley-Interscience, New York.
- MCLACHLAN, G., AND D. PEEL (2000): *Finite Mixture Models*. Wiley Interscience, New York.
- MERTON, R. C. (1973): “Theory of Rational Option Pricing,” *Bell Journal of Economics and Management Science*, 4, 141–183.
- NANDI, S. (1996): “Pricing and Hedging Index Options under Stochastic Volatility: An Empirical Examination,” *Federal Reserve Bank of Atlanta Working Paper 96-9*.
- NELSON, D. (1991): “Conditional Heteroskedasticity In Asset Returns: A New Approach,” *Econometrica*, 59(2), 347–370.

- PAN, J. (2002): “The Jump-risk Premia implicit in Options: Evidence from an Integrated Time-series Study,” *Journal of Financial Economics*, 63, 3–50.
- PAN, M., K. CHAN, AND C. FOK (1995): “Currency Futures Price Changes: A Two-Piece Mixture of Normals Approach,” *International Review of Economics and Finance*, 4, 69–78.
- POLSON, N., AND M. JOHANNES (2009): “MCMC Methods for Financial Econometrics,” forthcoming in Ait-Sahalia and Hansen (eds.), *Handbook of Financial Econometrics*. Elsevier.
- ROBERT, C., AND G. CASELLA (2004): *Monte Carlo Statistical Methods*. Springer-Verlag, New York.
- SCOTT, L. O. (1987): “Option Pricing When the Variance Changes Randomly: Theory, Estimation, and an Application,” *Journal of Financial and Quantitative Analysis*, 22(4), 419–438.
- STEIN, E. M., AND J. C. STEIN (1991): “Stock Price Distributions with Stochastic Volatility: An Analytical Approach,” *Review of Financial Studies*, 4(4), 727–752.
- STENTOFT, L. (2005): “Pricing American Options When the Underlying Asset follows GARCH Processes,” *Journal of Empirical Finance*, 12(4), 576–611.
- (2008): “American Option Pricing using GARCH models and the Normal Inverse Gaussian distribution,” *Journal of Financial Econometrics*, 6(4), 540–582.
- TANNER, M., AND W. WONG (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.
- TUCKER, A., AND L. POND (1988): “The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes,” *The Review of Economics and Statistics*, 70, 638–647.
- VAN DIJK, H., AND T. KLOEK (1978): “Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo,” *Econometrica*, 46, 1–19.
- WIGGINS, J. B. (1987): “Option Values Under Stochastic Volatility: Theory and Empirical Estimates,” *Journal of Financial Economics*, 19, 351–372.
- WONG, C., AND W. LI (2000): “On a Mixture Autoregressive Model,” *Journal of the Royal Statistical Society, Series B*, 62, 95–115.
- WONG, C., AND W. LI (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.