



Contents lists available at ScienceDirect

Journal of Econometrics

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

Bayesian semiparametric stochastic volatility modeling[☆]

Mark J. Jensen^{a,*}, John M. Maheu^b

^a Federal Reserve Bank of Atlanta, United States

^b University of Toronto, Canada and RCEA

ARTICLE INFO

Article history:

Received 21 April 2008

Received in revised form

17 December 2009

Accepted 28 January 2010

Available online 12 February 2010

Keywords:

Bayesian nonparametrics

Dirichlet process mixture prior

Markov chain Monte Carlo

Mixture models

Stochastic volatility

ABSTRACT

This paper extends the existing fully parametric Bayesian literature on stochastic volatility to allow for more general return distributions. Instead of specifying a particular distribution for the return innovation, nonparametric Bayesian methods are used to flexibly model the skewness and kurtosis of the distribution while the dynamics of volatility continue to be modeled with a parametric structure. Our semiparametric Bayesian approach provides a full characterization of parametric and distributional uncertainty. A Markov chain Monte Carlo sampling approach to estimation is presented with theoretical and computational issues for simulation from the posterior predictive distributions. An empirical example compares the new model to standard parametric stochastic volatility models.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

This paper proposes a model of asset returns that draws from the existing literature on autoregressive stochastic volatility (SV) models and the advances made in Bayesian nonparametric modeling and sampling to create a semiparametric SV model. By applying both parametric and nonparametric features to the return process, an estimable SV model with a flexible nonparametric innovation distribution is provided. The nonparametric portion of the model consists of an infinitely ordered mixture of normals

[☆] We would like to thank the seminar participants at the 24th Canadian Econometric Study Group Conference held in Montreal, the 7th All-Georgia Conference held at the Federal Reserve Bank of Atlanta, the RCEA conference on Econometrics 2007 held in Rimini, Italy, the 15th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics held at the Federal Reserve Bank of San Francisco, the 2009 Seminar on Bayesian Inference in Econometrics and Statistics held at Washington University in Saint Louis, the 7th Workshop on Bayesian Nonparametric held at Collegio Carlo Alberto in Moncalieri, Italy, the 2010 North American Winter Meetings of the Econometric Society in Atlanta, and the Department of Economics at New York University, Oregon State University, and the J.M. Huntsman School of Business at Utah State University. In addition, we express our appreciation for the comments and suggestions of the anonymous referees, an associate editor, the editor John Geweke, and Mark Fisher, Thanasis Stengos, and George Tauchen. We also thank Jacob Smith for helping with the computer programming required by the paper. Maheu is grateful to the SSHRC for financial support. The views expressed here are ours and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

* Corresponding author. Tel.: +1 404 498 8019; fax: +1 404 498 8810.

E-mail addresses: mark.jensen@atl.frb.org (M.J. Jensen), jmaheu@chass.utoronto.ca (J.M. Maheu).

whose component probabilities and parameters are modeled with a particular Bayesian prior—the Dirichlet process mixture prior (DPM). Under the DPM representation of the returns conditional distribution, our model produces a more robust predictive density of returns than parametric SV models. The paper takes a likelihood based approach to model inference and provides exact finite sample properties, including a full characterization of parametric and distributional uncertainty.

There exists a long history of modeling asset returns with a mixture of normals (see Press, 1967; Praetz, 1972; Clark, 1973; Gonedes, 1974; Kon, 1984). These early mixture models produced fat-tailed behavior but could not capture the dynamic clustering observed in the conditional variance of returns. SV models were designed to fit this time-varying behavior (see Taylor, 1986; Harvey et al., 1994). They consist of a continuous mixture of normals where their variances follow a dynamic stochastic process. However, parametric SV models have not fully captured the asymmetries and leptokurtotic behavior present in return data (see Gallant et al., 1997; Mahieu and Schotman, 1998; Liesenfeld and Jung, 2000; Meddahi, 2001; Durham, 2006). These characteristics play an important role in the pricing of derivatives, the measuring and managing of risk, and in portfolio selection. A flexible nonparametric version of the SV model will be useful to risk and portfolio managers alike.

The DPM consists of modeling the probabilities and parameters of an infinitely ordered mixture model with the Dirichlet process prior of Ferguson (1973). As a Bayesian nonparametric estimator of a unknown distribution, the DPM offers a number of attractive features; (i) the DPM spans the class of continuous distributions

(Escobar and West, 1995; Ghosal et al., 1999), (ii) the DPM is more flexible and realistic than a mixture model with a predetermined number of components, (iii) the Dirichlet process prior helps determine the number of mixture clusters that best fits the data, (iv) as an almost surely discrete prior it is parsimonious, (v) as a conjugate prior it is easy to use and facilitates Gibbs sampling, and (vi) it works well in practice.¹

This paper provides a flexible semiparametric stochastic volatility, Dirichlet process mixture model (SV-DPM) by combining a nonparametric independently identically distributed DPM model of innovations scaled by an autoregressive model of the return's latent conditional variance process.² The SV-DPM will nest within its parametric versions of the SV model. We also show how to estimate a restricted version of the model (SV-DPM-P) in which only the precision parameters are governed by the Dirichlet process (DP) prior. A Markov chain Monte Carlo (MCMC) sampler is constructed to estimate the unknown parameters of the SV-DPM. Our MCMC algorithm extends the DPM samplers of West et al. (1994) and MacEachern and Müller (1998) to the time-varying structure of the SV model. Because volatility is independent from the DPM, a tractable efficient posterior sampler is possible. Conditional on the value of the other unknowns, one block of our sampler consists of drawing the parameters of the clusters, while the other blocks draw the parameters and volatilities for the SV model's latent volatility process (see Chib et al., 2002; Eraker et al., 2003; Jacquier et al., 1994, 2004; Kim et al., 1998). In addition to providing smoothed estimates of the latent volatility process, the sampler also generates the predictive density and likelihood of returns that fully accounts for the uncertainty in the latent volatility process as well as the unknown return distribution.

A second contribution of the paper is a simple random block sampler of latent volatility. We extend the Fleming and Kirby (2003) block sampler of volatility by including the return data in the proposal distribution. This results in better candidate draws to the Metropolis–Hastings sampler resulting in lower correlation, leading to fewer sweeps being required. Our simple random block sampler of volatility can be used for all the SV models discussed in the paper.

We evaluate our SV-DPM model against standard SV models found in the literature; the SV model with normal innovations (SV-N) and the SV model with Student-*t* innovations (SV-t). In an empirical application with daily CRSP return data over the period 1980–2006, the predictive distribution for the SV-DPM model is very different from those for the parametric SV models. The SV-DPM model's predictive density displays negative skewness and kurtosis whereas neither the SV-N nor the SV-t does. The estimate of the variance of log volatility is considerably smaller for the semiparametric model, indicating that some tail thickness in conditional returns is better captured by the DPM.

The results highlight important differences in the SV-DPM model's predictive density and parameter estimates relative to parametric alternatives in a large sample setting. Next we consider what the model can offer in a small sample analysis. We compare the relative quality of the density forecasts of the new models by pooling the log predictive score function (Geweke and Amisano, 2008) over a shorter sample of daily return data from 2006 to 2008.

The models in the pool are the SV-DPM, SV-N, SV-t, and SV-DPM-P models. This latter model displays the largest weight of 0.73 in the optimal pooling score function. Dropping this specification from the pool results in a decrease of 3 points in the log predictive score. We conclude that the SV-DPM models can provide improvements in both large and small samples.

The paper is organized as follows. The SV-DPM and SV-DPM-P models are constructed in Section 2. Section 3 presents Bayesian inference for the SV-DPM model and Section 4 discusses features of the model. An application to daily return data is found in Section 5. Section 6 contains our conclusions and suggestions for possible future extensions for our Bayesian semiparametric SV model. The working paper version Jensen and Maheu (2008) includes additional details and simulation results.

2. SV-DPM models

We model the return of an asset with a stochastic volatility model whose unconditional return distribution is modeled nonparametrically with the Dirichlet process mixture prior. The stochastic volatility, Dirichlet process mixture model (SV-DPM), is defined as

$$y_t | f_N, h_t, \eta_t, \lambda_t^2 \stackrel{\perp}{\sim} N(\eta_t, \lambda_t^{-2} \exp\{h_t\}), \tag{1}$$

$$h_t | h_{t-1}, \delta, \sigma_v^2 \sim N(\delta h_{t-1}, \sigma_v^2), \quad \text{and} \quad h_t \perp y_t, \tag{2}$$

$$\begin{pmatrix} \eta_t \\ \lambda_t^2 \end{pmatrix} \Big| G \stackrel{iid}{\sim} G, \tag{3}$$

$$G | G_0, \alpha \sim DP(G_0, \alpha), \tag{4}$$

$$G_0(\eta_t, \lambda_t^2) \equiv N(m, (\tau \lambda_t^2)^{-1}) - \Gamma(v_0/2, s_0/2), \tag{5}$$

where $\stackrel{\perp}{\sim}$ denotes independently distributed.

At time $t = 1, \dots, n$ the continuously compounded return from holding a financial asset equals y_t and the latent log volatility h_t follows the first-order autoregressive (AR) process defined by Eq. (2) with the AR parameter δ . Identification of the SV-DPM model requires the unconditional mean of h_t to be set to zero. The intercept is subsumed into λ_t^2 . Stationary returns are ensured by restricting δ to the interval $(-1, 1)$. This guarantees a finite mean and variance for the volatility process, h_t . In Eq. (2), $h_t \perp y_t$ assumes away any leverage effects (see Jacquier et al., 2004; Yu, 2005; Omori et al., 2007).³

Eqs. (3)–(5) places a nonparametric prior on the random unconditional return distribution. It consists of an infinite ordered mixture of normals, a basis that is dense over the entire class of continuous distributions.⁴ Eq. (3)–(4) assume the mixture's probabilities and parameters η_t and λ_t^2 follow the Dirichlet process prior (DP) of Ferguson (1973). The DP prior consists of the base distribution G_0 , defined in Eq. (5) as a conjugate conditional normal-gamma distribution, and a nonnegative precision parameter α .

Our SV-DPM model also has the Sethuraman (1994) representation

$$y_t | f_N, h_t \stackrel{\perp}{\sim} \sum_{j=1}^{\infty} V_j f_N(\cdot | c_j, d_j^{-2} \exp\{h_t\}), \tag{6}$$

where $f_N(\cdot | c_j, d_j^{-2} \exp\{h_t\})$ is a normal density with mean c_j and variance $d_j^{-2} \exp\{h_t\}$, with the mixture weights distributed as

¹ Examples of the DPM being used in economics include Chib and Hamilton (2002), Conley et al. (2008), Griffin and Steel (2004), Hirano (2002), Jensen (2004), Kacperczyk et al. (2005) and Tiwari et al. (1988). Jensen (2004) uses a DPM to model the distribution of additive noise of log-squared returns while in this paper we are concerned with the conditional distribution of returns.

² The Dirichlet process prior has been used in autoregressive time-series models (Lau and So, 2008; Muller et al., 1997) and in models with ARCH effects (Lau and Siu, 2008). A time-dependent Dirichlet process is introduced in Griffin and Steel (2006).

³ Leverage effects can be included but the DPM portion of the model becomes computationally challenging. As a result, we choose to focus on a SV model without leverage effects and leave this a topic for future research.

⁴ See Lo (1984), Ghosal et al. (1999) and Ghosal and van der Vaart (2007) for a discussion on the posterior consistency of the DPM model.

$V_1 = W_1$, and $V_j = W_j \prod_{s=1}^{j-1} (1 - W_s)$, where $W_j \sim \text{Beta}(1, \alpha)$. The mixture parameters (c_j, d_j^2) have the same prior as (η_t, λ_t^2) —the normal-gamma distribution of Eq. (5).

The discrete nature of Eq. (6) implies clustering in the mixture parameters c_j and d_j^2 . Except for some pathological cases analytical expressions for the DPM’s posterior expectations are not possible. Fortunately, there are Gibbs sampling techniques based on Escobar and West (1995) that exploit the Blackwell and MacQueen (1973) Polya urn representation of the DP prior to integrate out the mixture probabilities V_j and draw the k finite clusters $(c_j, d_j^2), j = 1, \dots, k$, where $k < n$, and cluster weights n_j/n , where n_j is the number of observations assigned to the j th cluster.

In another nonparametric DPM representation of the unconditional return distribution, we use a mixture of normals centered at zero with a DP prior placed only on the mixture probabilities and the mixture precision parameter λ_t^2 . We call this semiparametric stochastic volatility model the SV-DPM-P model. Its complete definition is

$$y_t | N, h_t, \mu, \lambda_t^2 \stackrel{\perp}{\sim} N(\mu, \lambda_t^{-2} \exp\{h_t\}),$$

$$h_t | h_{t-1}, \delta, \sigma_v^2 \sim N(\delta h_{t-1}, \sigma_v^2), \text{ and } h_t \perp y_t,$$

$$\lambda_t | G \stackrel{iid}{\sim} G,$$

$$G | G_0, \alpha \sim \text{DP}(G_0, \alpha),$$

$$G_0(\lambda_t^2) \equiv \Gamma(v_0/2, s_0/2).$$

Our SV-DPM class of volatility models are more flexible than the existing class of parametric SV models in modeling the distribution of y_t . In the terminology of Müller and Quintana (2004), SV-DPM models “robustifies” the class of parametric SV models. By modeling the innovation distribution of y_t with a Dirichlet process mixture, diagnostics and sensitivity analysis can be conducted by nesting parametric SV models within the SV-DPM model. For example, when $V_1 = 1, V_j = 0$ for $j > 1$, the SV-DPM model equals the autoregressive, stochastic volatility model of Jacquier et al. (1994). The SV-t model of Harvey et al. (1994) with ν degrees of freedom is also nested within the SV-DPM model by setting $\eta_t = 0$ for all t , letting $\alpha \rightarrow \infty$, and modeling the precision parameter prior as $\lambda_t^2 \sim G_0 \equiv \Gamma(\nu/2, \nu/2)$.

Geweke and Keane (2007) also model the return of an asset as a mixture with their smoothly mixing regression model. But unlike the infinite ordered mixture representation of the SV-DPM model, the smoothly mixing regression model sets the number of mixture clusters *a priori*. Probabilities of a particular cluster are then determined by a multinomial probit whose covariates are a nonlinear combination of lagged and absolute returns.

3. Bayesian inference

The inherent difficulty with all stochastic volatility models, regardless of the innovations being modeled parametrically or nonparametrically, is the intractability of the SV’s likelihood function. Because the log volatility process h_t enters through the variance of y_t , the SV model’s likelihood function does not have an analytical solution. Bayesian estimation of the SV model bridges this problem by augmenting the model’s unknown parameters with the latent volatilities and designing a hybrid Markov chain Monte Carlo algorithm (Tanner and Wong, 1987) to sample from the joint posterior distribution, $\pi(\psi, h|y)$, where $\psi = (\delta, \sigma_v)^{\prime}$, $h = (h_1, \dots, h_n)^{\prime}$ and $y = (y_1, \dots, y_n)^{\prime}$ (see Jacquier et al., 1994; Kim et al., 1998; Chib et al., 2002).

In the context of the SV-DPM models the additional unknown mixture parameters $\phi = (\phi_1, \dots, \phi_n)^{\prime}$, where $\phi_t = (\eta_t, \lambda_t^2)$ for the SV-DPM and $\phi_t = \lambda_t^2$ for SV-DPM-P, can be augmented with ψ and h and included in the MCMC sampler of the posterior $\pi(\psi, h, \phi|y)$. Since the likelihood function of SV models is intractable and because we do not know the number of mixtures of the nonparametric distribution or their values, we are precluded

from directly sampling from $\pi(\psi, h, \phi|y)$. Instead, we judiciously break up the augmented posterior distribution into tractable blocks of conditional posterior distributions and design a stylized MCMC sampler for each block. The accuracy of the sampler and its computational costs are dependent on how the blocks of the unknowns are selected, on the level of dependency between the conditional distributions and random variables, and on the type of sampling algorithm used.

The blocking scheme that we design for the SV-DPM models consists of iteratively sampling through the following conditional distributions:

1. $\pi(\psi|h)$,
2. $\pi(h|y, \psi, \phi)$,
3. $\pi(\phi|y, h)$,
4. $\pi(\alpha|\phi)$,
5. $\pi(\mu|y, h, \phi)$.

Step 5 is only required with the SV-DPM-P model. One full iteration through each conditional distributions denotes a sweep of the MCMC sampler.

3.1. The parameter sampler

Conditional on knowing the value of h , sampling from $\pi(\psi|h)$ in Step 1 is straightforward. Assume that the priors for δ and σ_v^2 are independent, in other words, $\pi(\psi) = \pi(\delta)\pi(\sigma_v^2)$, where the marginal prior distributions are $\pi(\delta) \propto N(\mu_\delta, \sigma_\delta^2)I_{|\delta|<1}$, a normal truncated to the stationary region of δ ’s parameter space, and $\pi(\sigma_v^2) \sim \text{Inv-}\Gamma(v_\sigma/2, s_\sigma/2)$. Under this prior for ψ , draws from $\delta, \sigma_v^2|h$ are made by sequentially sampling from the conditional marginal distributions, $\delta|h, \sigma_v^2 \sim N(\hat{\delta}, \hat{\sigma}_v^2)I(|\delta| < 1)$, where

$$\hat{\delta} = \hat{\sigma}_\delta^2 \left(\frac{\sum_{t=2}^n h_{t-1} h_t}{\sigma_v^2} + \frac{\mu_\delta}{\sigma_\delta^2} \right), \quad \hat{\sigma}_\delta^2 = \frac{\sigma_v^2 \sigma_\delta^2}{\sigma_\delta^2 \sum_{t=2}^n h_{t-1}^2 + \sigma_v^2},$$

and $\sigma_v^2|h, \delta \sim \text{Inv-}\Gamma((n-1+v_\sigma)/2, [s_\sigma + \sum_{t=2}^n (h_t - \delta h_{t-1})^2]/2)$. If a draw from $\delta|h, \sigma_v^2$ results in a realization outside the stationary set, the draw of δ is discarded and sampling continues until a value from within the parameter space is obtained.

To perform Step 5 for the SV-DPM-P model we assume $\pi(\mu) \sim N(m, \tau)$. Given values for ϕ and h , we can rewrite the return equation as

$$y_t \exp\{-h_t/2\} \lambda_t = \mu \exp\{-h_t/2\} \lambda_t + z_t, \quad z_t \sim \text{NID}(0, 1).$$

Given the conjugate nature of $\pi(\mu)$, draws of μ are made from $N(\bar{\mu}, \bar{\tau})$ where

$$\bar{\mu} = \frac{m/\tau + \sum_t y_t \exp\{-h_t\} \lambda_t^2}{1/\tau + \sum_t \exp\{-h_t\} \lambda_t^2},$$

$$\bar{\tau} = \left(1/\tau + \sum_t \exp\{-h_t\} \lambda_t^2 \right)^{-1}.$$

3.2. The latent volatility sampler

Drawing the latent volatilities is difficult and has attracted the attention of the profession (see Jacquier et al., 1994; Pitt and Shephard, 1997; Kim et al., 1998; Chib et al., 2002; Fleming and Kirby, 2003). One option for drawing the volatilities of the SV-DPM model is to apply an element-by-element volatility sampler. Conditional on ϕ , the entire suite of existing element-by-element

samplers given by Geweke (1994), Pitt and Shephard (1997), Kim et al. (1998), and Jacquier et al. (2004) can be directly applied to $\tilde{y}_t \equiv \lambda_t(y_t - \eta_t)$ for the SV-DPM model and $\tilde{y}_t \equiv \lambda_t(y_t - \mu)$ for SV-DPM-P.

Element-by-element samplers, however, are known to be very inefficient and require throwing away a large number of initial draws of h to reduce dependency on the starting values. Highly persistent h_t s also lead to a strong correlation between the sampled volatilities. As a result, a large number of sweeps must be carried out. This becomes very taxing for the SV-DPM models since each additional sweep also requires sampling from $\phi|y, h$.

Ideally one would like to sample from $h|y, \psi, \phi$ in a single draw (see Kim et al., 1998; Chib et al., 2002). This approach eliminates the correlation between the drawn h s, but requires approximating the log chi-squared distribution of $\log(y_t - \eta_t)^2 + \log \lambda_t^2$ with a finite order mixture of normals. While the approximating mixture's order, weights, means and variances are known *a priori*, each observation's cluster assignment is not. Because we are already modeling the unconditional return distribution nonparametrically we believe adding another layer of complexity with another mixture of normals takes away from the DPM prior flexibility to model the unconditional return distribution.

Fortunately, less correlated draws of the volatilities can be found by sampling random length blocks of volatilities instead of the entire vector (see Pitt and Shephard, 1997; Elerian et al., 2001; Fleming and Kirby, 2003). Our random length block sampler partitions h into blocks of subvectors $\{h_{(t,\tau)}\}$, where $h_{(t,\tau)} = (h_t, h_{t+1}, \dots, h_\tau)'$, $1 \leq t \leq \tau \leq n$, and the length of the subvector $l_t = \tau - t + 1$ is randomly drawn from a Poisson distribution with hyperparameter $\lambda_h = 3$; i.e., $E[l_t] = 4$.⁵ By letting the length be random we ensure that with each sweep a different partition of h is sampled, thus helping to reduce the degree of dependency that would exist if l_t were fixed. By lowering the level of correlation in the draws of the $h_{(t,\tau)}$, we reduce the number of sweeps needed to produce reliable estimates of the model parameters.

Because the desired density

$$\pi(h_{(t,\tau)}|y, h_{t-1}, h_{\tau+1}, \psi, \phi) \propto f(y|h_{(t,\tau)}, \phi, \psi) \pi(h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi),$$

does not come from a standard distribution, we design a Metropolis–Hastings (MH) sampler for the above target density where we extend the sampler of Fleming and Kirby (2003) to include the return data, y . Fleming and Kirby (2003) show that if the log volatility process is approximated by the random walk $h_t = h_{t-1} + \sigma_v v_t$ then a reasonable proposal for the target distribution is

$$h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \sigma_v^2 \sim N(m_{(t,\tau)}, \Sigma_{(t,\tau)}), \tag{7}$$

where the $l_t \times 1$ vector $m_{(t,\tau)} = (m_t, \dots, m_\tau)'$ and the $l_t \times l_t$ covariance matrix $\Sigma_{(t,\tau)} = \{\sigma_{ij}^{(t)}\}_{i,j=t,\dots,\tau}$ are defined by their elements:

$$m_{t+i} = \frac{(l_t - i)h_{t-1} + (i + 1)h_{\tau+1}}{l_t + 1}, \quad i = 0, \dots, l_t - 1,$$

$$\sigma_{ij}^{(t)} = \sigma_v^2 \frac{\min(i, j)(1 + l_t) - ij}{l_t + 1}, \quad i = 1, \dots, l_t,$$

and, $j = 1, \dots, l_t$.

The inverse of the covariance matrix for the proposal distribution has the convenient tridiagonal form

$$\Sigma_{(t,\tau)}^{-1} = \begin{pmatrix} 2/\sigma_v^2 & -1/\sigma_v^2 & 0 & \dots \\ -1/\sigma_v^2 & 2/\sigma_v^2 & -1/\sigma_v^2 & \ddots \\ 0 & -1/\sigma_v^2 & 2/\sigma_v^2 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

making evaluation of the proposal density's quadratic term $(h_{(t,\tau)} - m_{(t,\tau)})' \Sigma_{(t,\tau)}^{-1} (h_{(t,\tau)} - m_{(t,\tau)})$ quick and easy.

Since the proposal distribution in Eq. (7) ignores the information found in the return vector, $y_{(t,\tau)} = (y_t, \dots, y_\tau)'$, a better proposal distribution would be one that incorporates these data. Such a distribution would help the MH sampler converge more quickly and result in a better mixture of draws from the latent volatility's target distribution.

Once again the desired target density is

$$\begin{aligned} \pi(h_{(t,\tau)}|y_{(t,\tau)}, h_{t-1}, h_{\tau+1}, \psi, \phi) &\propto f(y_{(t,\tau)}|h_{(t,\tau)}, \phi) \pi(h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi), \\ &\approx f(y_{(t,\tau)}|h_{(t,\tau)}, \phi_{(t,\tau)}) f_N(h_{(t,\tau)}|m_{(t,\tau)}, \Sigma_{(t,\tau)}), \end{aligned} \tag{8}$$

where the random walk approximation of Fleming and Kirby (2003) has been applied to $\pi(h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi)$. The likelihood function

$$\begin{aligned} f(y_{(t,\tau)}|h_{(t,\tau)}, \phi_{(t,\tau)}) &\propto \exp \left\{ -0.5 \left(l' h_{(t,\tau)} + \tilde{y}_{(t,\tau)}' \exp\{-h_{(t,\tau)}\} \right) \right\}, \end{aligned} \tag{9}$$

with l being a $l_t \times 1$ vector of ones, $\tilde{y}_{(t,\tau)} = (\tilde{y}_t, \dots, \tilde{y}_\tau)'$, and $\exp\{-h_{(t,\tau)}\} = (\exp\{-h_t\}, \dots, \exp\{-h_\tau\})'$. Replacing the $\exp\{-h_{(t,\tau)}\}$ vector in Eq. (9) with its first-order, Taylor series approximation, $\exp\{-h_{(t,\tau)}\} \approx D_{(t,\tau)}(l + m_{(t,\tau)} - h_{(t,\tau)})$, where the $l_t \times l_t$ diagonal matrix $D_{(t,\tau)} = \text{diag}\{\exp(-m_{(t,\tau)})\}$, results in

$$\begin{aligned} \exp \left\{ -0.5 \left(l' h_{(t,\tau)} + \tilde{y}_{(t,\tau)}' \exp\{-h_{(t,\tau)}\} \right) \right\} &\leq \exp \left\{ -0.5 \left(l' - \tilde{y}_{(t,\tau)}' D_{(t,\tau)} \right) h_{(t,\tau)} \right\}. \end{aligned} \tag{10}$$

Substituting the right hand side of Eq. (10) for the $f(y_{(t,\tau)}|h_{(t,\tau)}, \phi_{(t,\tau)})$ term in Eq. (8) and collecting terms in the quadratic form of $h_{(t,\tau)}$ leads to our MH sampler's fat-tailed proposal density:

$$\begin{aligned} f_{St}(h_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu) &\propto \left[1 + (h_{(t,\tau)} - \zeta_{(t,\tau)})' \Sigma_{(t,\tau)}^{-1} (h_{(t,\tau)} - \zeta_{(t,\tau)})/\nu \right]^{-(l_t + \nu)/2} \end{aligned}$$

where $f_{St}(h_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)$ is the density of a l_t -variate Student- t distribution with mean, $\zeta_{(t,\tau)} = m_{(t,\tau)} - 0.5 \Sigma_{(t,\tau)}(l - D_{(t,\tau)} \tilde{y}_{(t,\tau)})$, covariance, $\Sigma_{(t,\tau)} \nu / (\nu - 2)$, and ν degrees of freedom (in the empirical example of Section 5 we set ν equal to 10). For the endpoints h_1 and h_n , we generate h_0 and h_{n+1} according to the volatility dynamics and use the same proposal density.

Given the previous MCMC draw of $h_{(t,\tau)}$, the candidate draw, $\hat{h}_{(t,\tau)} \sim St(\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)$, will be accepted as a realization from the target distribution with the MH probability as given in Box 1.

3.3. The DPM sampler

Although the SV-DPM model in (6) implies an infinite number of clusters, for a finite data set each sweep of the Gibbs sampler will divide the data into a finite set of clusters. Conditional on a draw of ψ and h , sampling from the posterior distribution $\phi|y, h$ is done through a variant of West et al. (1994) and MacEachern and Müller (1998) Gibbs samplers. To improve the efficiency of sampling from $\phi|y, h$, West et al. (1994) and MacEachern and Müller (1998) appeal to draws from the equivalent distribution $\theta, s|y, h$, where $\theta = (\theta_1, \dots, \theta_k)'$, $k \leq n$, contains the unique elements from the

⁵ λ_h was selected to minimize the numerical inefficiency values of the model parameters based on several trial runs.

$$\min \left\{ \frac{f(y_{(t,\tau)}|\phi_{(t,\tau)}, \widehat{h}_{(t,\tau)}) \pi(\widehat{h}_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi) f_{St}(h_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)}{f(y_{(t,\tau)}|\phi_{(t,\tau)}, h_{(t,\tau)}) \pi(h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi) f_{St}(\widehat{h}_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)}, 1 \right\}$$

where $f(y_{(t,\tau)}|\phi_{(t,\tau)}, h_{(t,\tau)}) = \prod_{j=1}^{\tau} f_N(y_j|\eta_j, \lambda_j^{-2} \exp\{h_j\})$ and $\pi(h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi) = \prod_{j=t}^{\tau+1} \exp\left\{-\frac{(h_j - \delta h_{j-1})^2}{2\sigma^2}\right\}$

Box I.

vector ϕ . The n -length vector s contains the indicator variables s_t , $t = 1, \dots, n$, where $s_t = j$ when $\phi_t = \theta_j$, $j = 1, \dots, k$. Together, θ and s completely identify ϕ . In the following, $\theta^{(t)}$ denotes the unique elements of ϕ when the element ϕ_t is deleted. The number of clusters in $\theta^{(t)}$ is indexed from $j = 1$ to $\mathcal{K}^{(t)}$.

To describe the sampler for $\theta, s|y, h$ we rewrite Eq. (1), the compound return equation, as

$$y_t^* = \eta_t \exp\{-h_t/2\} + \lambda_t^{-1} \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} N(0, 1),$$

where $y_t^* \equiv y_t \exp\{-h_t/2\}$. Draws are now made from $\theta, s|y^*$ with the following two-step procedure:

Step 1. Sample s and k by drawing $\phi_t = (\eta_t, \lambda_t^2)$ for $t = 1, \dots, n$ from

$$\begin{aligned} \phi_t|y_t^*, \theta^{(t)}, s^{(t)} &\sim c \frac{\alpha}{\alpha + n - 1} g(y_t^*) G(d\phi_t|y_t^*) \\ &+ \frac{c}{\alpha + n - 1} \sum_{j=1}^{\mathcal{K}^{(t)}} n_j^{(t)} f(y_t^*|\theta_j) \delta_{\theta_j}(d\phi_t), \end{aligned}$$

setting $s_t = j$ when $\phi_t = \theta_j$, or $s_t = k + 1$ and $k = k + 1$ when ϕ_t is drawn from $G(d\phi_t|y_t^*)$. c is the integrating constant.

Step 2. Given the s and k from Step 1, discard ϕ and sample $\theta_j = (\eta_j, \lambda_j^2)$, $j = 1, \dots, k$, from

$$\theta_j|y_t^* : s_t = j \propto \prod_{t:s_t=j} f_N(y_t^*|\eta_j \exp\{-h_t/2\}, \lambda_j^{-2}) G_0(d\theta_j).$$

In Step 1 the probability of s_t equaling the j th cluster is proportional to $n_j^{(t)}$, the number of other times that the j th cluster occurs after dropping ϕ_t , times the likelihood that y_t^* belongs to the j th cluster, $f(y_t^*|\theta_j) \equiv f_N(y_t^*|\eta_j \exp\{-h_t/2\}, \lambda_j^{-2})$. On the other hand, the probability of s_t being assigned to a new cluster is proportional to the predictive density:

$$\begin{aligned} g(y_t^*) &\equiv \int f(y_t^*|\phi_t) G_0(d\phi_t) d\phi_t, \\ &= \int \frac{1}{\sqrt{2\pi \exp\{h_t\} \lambda_t^{-2}}} \\ &\times \exp\left\{-\frac{(y_t^* - \eta_t \exp\{-h_t/2\})^2}{2\lambda_t^{-2}}\right\} G_0(d\phi_t) d\phi_t, \\ &= f_{St}(y_t^*|m \exp\{-h_t/2\}, (\exp\{h_t\} + \tau) s_0 / (\tau v_0), v_0), \\ &= f_{St}(y_t^*|m, (1 + \tau \exp\{h_t\}) s_0 / (\tau v_0), v_0), \end{aligned}$$

where $f_{St}(\cdot|m, s, v)$ denotes the probability density function of a Student- t distribution with mean m , variance $vs/(v - 2)$, and v degrees of freedom. If a new cluster is drawn, ϕ_t equals the new cluster parameter θ_{k+1} sampled from the posterior distribution:

$$G(d\phi_t|y_t^*) \equiv \frac{f(y_t^*|\phi_t) G_0(d\phi_t)}{g(y_t^*)}.$$

By the conjugate nature of the normal-gamma prior, G_0 , and the normality of the likelihood function, $f(y_t^*|\phi_t)$, $G(d\phi_t|y_t^*)$, equals the normal-gamma distribution:

$$\lambda_t^2|y_t^* \sim \Gamma(\bar{v}/2, \bar{s}_t/2),$$

$$\eta_t|y_t^*, \lambda_t^2 \sim N(\bar{\mu}_t, (\bar{\tau}_t \lambda_t^2)^{-1}),$$

where $\bar{v} = v_0 + 1$, $\bar{s}_t = s_0 + (\bar{\mu}_t - y_t^*)^2 \exp\{-h_t\} + (\bar{\mu}_t - m)^2 \tau$, with $\bar{\mu}_t = \bar{\tau}_t^{-1} (\tau m + y_t^* \exp\{-h_t/2\})$ and $\bar{\tau}_t = \tau + \exp\{-h_t\}$.

Step 2 consists of generating a new draw of ϕ , conditional on the s and k sampled in Step 1, by sampling the unique mixture parameters, θ_j , $j = 1, \dots, k$, from the linear regression model:

$$y_t^*|s_t, \eta_j, \lambda_j^2 \sim N(\eta_j \exp\{-h_t/2\}, \lambda_j^{-2}), \tag{11}$$

where $t \in \{t' : s_{t'} = j\}$, and the prior of η_j and λ_j^2 is distributed according to the base distribution, G_0 . Conjugacy between the normal-gamma base distribution, G_0 , and the likelihood function in Eq. (11) leads to the posterior distribution $\theta_j|y^*, s, k$ being the normal-gamma distribution

$$\lambda_j^2|y^*, s, k \sim \Gamma(\bar{v}_j/2, \bar{s}_j/2),$$

$$\eta_j|y^*, s, k, \lambda_j^2 \sim N(\bar{\mu}_j, (\bar{\tau}_j \lambda_j^2)^{-1}),$$

where $\bar{v}_j = v_0 + n_j$, $\bar{s}_j = s_0 + s_j + (\bar{\mu}_j - b_j)^2 \sum_{t:s_t=j} \exp\{-h_t\} + (\bar{\mu}_j - m)^2 \tau$, and $\bar{\mu}_j = \bar{\tau}_j^{-1} (\tau m + b_j \sum_{t:s_t=j} \exp\{-h_t\})$, with $\bar{\tau}_j = \tau + \sum_{t:s_t=j} \exp\{-h_t\}$, and b_j being the ordinary least square estimate from regressing y_t^* on $\exp\{-h_t/2\}$ over the set of observations $\{t : s_t = j\}$. Lastly, $s_j = \sum_{t:s_t=j} (y_t^* - b_j \exp\{-h_t/2\})^2$; i.e., the sum of squares of errors from the regression over the same set of observations where $s_t = j$.

3.4. The DPM-P sampler

For the SV-DPM-P model draws of ϕ are again made from $\theta, s|y$ but with $\theta = (\lambda_1^2, \dots, \lambda_k^2)$. The two-step DPM-P sampler involves:

Step 1. Sampling s and k by drawing λ_t^2 for $t = 1, \dots, n$ from

$$\begin{aligned} \lambda_t|y_t, \lambda^{(t)}, s^{(t)} &\sim c \frac{\alpha}{\alpha + n - 1} g(y_t) G(d\lambda_t^2|y_t) \\ &+ \frac{c}{\alpha + n - 1} \sum_{j=1}^{\mathcal{K}^{(t)}} n_j^{(t)} f(y_t|\mu, \exp\{h_t\} \lambda_j^{-2}) \delta_{\lambda_j^2}(d\lambda_t^2), \end{aligned}$$

where $g(y_t) = f_{St}(y_t|\mu, \exp\{h_t\} v_0/s_0, v_0)$, and $G(d\lambda_t|y_t)$ is the distribution $\Gamma(\bar{v}/2, \bar{s}_t/2)$ with $\bar{v} = v_0 + 1$ and $\bar{s}_t = s_0 + (y_t - \mu)^2 / \exp\{h_t\}$. c is the integrating constant.

Step 2. Given s and k from Step 1, sample λ_j^2 for $j = 1, \dots, k$, from

$$\lambda_j^2|y_t : s_t = j \propto \prod_{t:s_t=j} f_N(y_t|\mu, \exp\{h_t\} \lambda_j^{-2}) G_0(d\lambda_j)$$

which is the $\Gamma(\bar{v}_j/2, \bar{s}_j/2)$ distribution with $\bar{v}_j = v_0 + n_j$ and $\bar{s}_j = s_0 + \sum_{t:s_t=j} (y_t - \mu)^2 / \exp\{h_t\}$.

3.5. The α sampler

The DPM precision parameter α is sampled for both models with the two-step algorithm of Escobar and West (1995). Since y is conditionally independent of α when the mixture order, k , parameter vector, ϕ , and state indicator vector, s , are all known,

and because ϕ is also conditionally independent of α when both k and s are known, the posterior of α is only dependent on k ; i.e., $\pi(\alpha|\phi) = \pi(\alpha|k) \propto \pi(\alpha)f(k|\alpha)$. Assuming the gamma distribution, $\Gamma(a, b)$, where $a > 0$ and $b > 0$, is the prior for α , exact draws from $\pi(\alpha|k)$ are made by first sampling the random variable ξ from $\pi(\xi|\alpha, k) \sim \text{Beta}(\alpha + 1, n)$, and secondly, sampling α from the mixture $\pi(\alpha|\xi, k) \sim \pi_\xi \Gamma(a + k, b - \ln \xi) + (1 - \pi_\xi) \Gamma(a + k - 1, b - \ln \xi)$, where $\pi_\xi / (1 - \pi_\xi) = (a + k - 1) / [n(b - \ln \xi)]$.

4. Features of the SV-DPM model

After an initial burn-in phase, our MCMC algorithm for the SV-DPM model produces a set of draws, $\{\psi^{(r)}, h^{(r)}, \theta^{(r)}, s^{(r)}, \alpha^{(r)}\}_{r=1}^R$, from the desired posterior density, $\pi(\psi, h, \theta, s, \alpha|y)$. Given these draws we can produce simulation consistent estimates of posterior quantities. For example, the posterior mean of the AR parameter for volatility is $E[\delta|y] \approx R^{-1} \sum_{r=1}^R \delta^{(r)}$ where this approximation can be made more precise by increasing the number of draws, R .⁶ In a similar way various quantities of the predictive density and likelihood can be estimated.

4.1. Predictive density and likelihood

The key quantity of interest in density estimation is the predictive density. Gelfand and Mukhopadhyay (1995) discuss this and more generally the estimation of linear functionals for DPM models. Drawing on their findings, the in-sample predictive posterior density for the SV-DPM model equals

$$f(Y_t|y) = \int f(Y_t|\theta, h_t, \alpha) \pi(\theta, h_t, \alpha|y) d\theta dh_t d\alpha, \approx \frac{1}{R} \sum_{r=1}^R f(Y_t|\theta^{(r)}, h_t^{(r)}, \alpha^{(r)}), \tag{12}$$

where $Y_t, t = 1, \dots, n$, is the unobserved random return at time t , $\theta^{(r)}, h_t^{(r)}$ and $\alpha^{(r)}$ are the r th draws from the posterior simulator.⁷ The conditional posterior density in Eq. (12) equals

$$f(Y_t|\theta^{(r)}, h_t^{(r)}, \alpha^{(r)}) = \frac{\alpha^{(r)}}{\alpha^{(r)} + n} g(Y_t|h_t^{(r)}) + \sum_{j=1}^{k^{(r)}} \frac{n_j^{(r)}}{\alpha^{(r)} + n} f_N(Y_t|\theta_j^{(r)}, h_t^{(r)}). \tag{13}$$

For the SV-DPM model $g(Y_t|h_t^{(r)}) = f_{st}(Y_t|m, (1 + \tau \exp\{h_t^{(r)}\})s_0/(\tau v_0), v_0)$, and $f_N(Y_t|\theta_j^{(r)}, h_t^{(r)}) = f_N(Y_t|\eta_j^{(r)}, \lambda_j^{-2(r)} \exp\{h_t^{(r)}\})$. In the SV-DPM-P model $g(Y_t|h_t^{(r)}) = f_{st}(Y_t|\mu, \exp\{h_t^{(r)}\}v_0/s_0, v_0)$, and $f_N(Y_t|\theta_j^{(r)}, h_t^{(r)}) = f_N(Y_t|\mu^{(r)}, \lambda_j^{-2(r)} \exp\{h_t^{(r)}\})$.

Eq. (13) shows the flexibility of modeling the SV return innovation distribution with the nonparametric DPM prior. In our semiparametric SV model the conditional predictive density is a weighted mixture of normals and Student- t densities, enabling it to fit multi-modal distributions, negative or positive skewed distributions, and other non-Gaussian type behavior like fat tails.

Except for the additional structure of the stochastic volatility process, the one-step-ahead, out-of-sample predictive density for the SV-DPM model is the same as the predictive density of Escobar and West (1995), p. 580. The SV-DPM model's one-step-ahead predictive return density is

$$f(Y_{n+1}|y) = \int f(Y_{n+1}|\theta, h_{n+1}, \alpha) \pi(\theta, h_{n+1}, \alpha|y) d\theta dh_{n+1} d\alpha, \approx \frac{1}{R} \sum_{r=1}^R f(Y_{n+1}|\theta^{(r)}, h_{n+1}^{(r)}, \alpha^{(r)}), \tag{14}$$

where the conditional density

$$f(Y_{n+1}|\theta^{(r)}, h_{n+1}^{(r)}, \alpha^{(r)}) = \frac{\alpha^{(r)}}{\alpha^{(r)} + n} g(Y_{n+1}|h_{n+1}^{(r)}) + \sum_{j=1}^{k^{(r)}} \frac{n_j^{(r)}}{\alpha^{(r)} + n} f_N(Y_{n+1}|\theta_j^{(r)}, h_{n+1}^{(r)}),$$

has the same form as Eq. (13) but $h_{n+1}^{(r)}$ is a draw from $N(\delta^{(r)}h_n^{(r)}, \sigma_v^{2(r)})$.

The SV-DPM model time t one-step-ahead predictive likelihood equals Eq. (14) evaluated at the observed return y_t with $\{\theta^{(r)}, h_t^{(r)}, \alpha^{(r)}\}$ representing the draws from a full MCMC run on the posterior $\theta, h_t, \alpha|y_1, \dots, y_{t-1}$.

4.2. Conditional moments

Using Eq. (12), in-sample moments of the equity return can be computed. For instance, the first and second moments of the SV-DPM model's return distribution can be approximated as

$$E[Y_t|y] \approx \frac{1}{R} \sum_{r=1}^R \left(\frac{\alpha^{(r)}}{\alpha^{(r)} + n} m + \sum_{i=1}^{k^{(r)}} \frac{n_i^{(r)}}{\alpha^{(r)} + n} \eta_i^{(r)} \right), E[Y_t^2|y] \approx \frac{1}{R} \sum_{r=1}^R \left(\frac{\alpha^{(r)}}{\alpha^{(r)} + n} \left[\frac{(1 + \tau \exp\{h_t^{(r)}\})s_0}{\tau(v_0 - 2)} + m^2 \right] + \sum_{i=1}^{k^{(r)}} \frac{n_i^{(r)}}{\alpha^{(r)} + n} \left[\eta_i^{2(r)} + \lambda_i^{-2(r)} \exp\{h_t^{(r)}\} \right] \right),$$

and the posterior conditional variance equals $\text{Var}(Y_t|y) \equiv E[Y_t^2|y] - E[Y_t|y]^2$.

4.3. Label switching

Mixture models in general suffer from what is referred to as “label switching”; a shortcoming where the mixture parameters are unidentified. In Eq. (13), the conditional density is symmetrical over the k clusters, in other words, it will equal the same value regardless of the particular permutation of the mixture parameters, $\{n_{g(j)}, \eta_{g(j)}, \lambda_{g(j)}\}_{j=1, \dots, k}$, where $g(j)$ is the permutation function of k elements. As a result the mixture parameters of the j th cluster in one sweep of the sampler may be assigned a different cluster label, $g(j) \neq j$, during another sweep of the sampler (see Richardson and Green, 1997). The DPM clusters, therefore, cannot be used to identify time periods where markets are in a particular state such as an expansionary or recessionary economic state. Since our only purpose for using the DPM is to model the distribution of y_t^* nonparametrically, label switching will not present a problem in making inferences concerning the parameters or forecasts of the stochastic volatility model. For a more detailed discussion of this in the context of finite mixture models see Geweke (2007) and Frühwirth-Schnatter (2006).

5. An empirical example

In this section we report the results from applying the SV-DPM model to daily stock return data. More specifically, we apply the SV-DPM and SV-DPM-P models and the MCMC sampler developed in Section 3 to 6815 compounded daily returns from the Center for

⁶ For a full treatment of MCMC methods see Robert and Casella (1999).

⁷ To minimize notation we have omitted conditioning on n_1, \dots, n_k which is the number of observations in each cluster.

Table 1
 Posterior estimates for daily returns of the CRSP value-weighted portfolio from Jan. 2, 1980 to Dec. 29, 2006 (6815 observations, 30,000 thinned draws of the SV-DPM and SV-DPM-P sampling algorithm where every tenth draw is retained and the first 1000 thinned draws are discarded).

	SV-DPM			SV-DPM-P		SV-t		SV-N	
	Mean	Stdev	Ineff	Mean	Stdev	Mean	Stdev	Mean	Stdev
μ				0.0747	0.0115	0.0786	0.0084	0.0793	0.0086
γ						-0.0087	0.0023	-0.0106	0.0028
δ	0.9900	0.0022	13.510	0.9852	0.0033	0.9878	0.0026	0.9795	0.0037
σ_v^2	0.0149	0.0023	33.195	0.0205	0.0033	0.0154	0.0023	0.0276	0.0040
ν						9.9149	1.3035		
α	0.3323	0.1873	20.571	1.4034	0.5396				
k	4.6006	1.9579	49.939	21.589	6.9073				

Ineff is the inefficiency factor.

SV-DPM: $y_t | \eta_t, \lambda_t, h_t \sim N(\eta_t, \lambda_t^{-2} \exp(h_t))$, $(\eta_t, \lambda_t^2) | G \sim DP(G_0, \alpha)$, $G_0(\eta_t, \lambda_t^2) \equiv N(0, (10\lambda_t^2)^{-1}) - \Gamma(10/2, 10/2)$, $\alpha \sim \Gamma(2, 8)$

SV-DPM-P: $y_t | \lambda_t, h_t \sim N(\mu, \lambda_t^{-2} \exp(h_t))$, $\lambda_t | G \sim DP(G_0, \alpha)$, $G_0(\lambda_t^2) \equiv \Gamma(10/2, 10/2)$, $\alpha \sim \Gamma(2, 8)$

SV-t: $y_t = \mu + \exp(h_t/2)z_t$, $h_t = \gamma + \delta h_{t-1} + \sigma_v v_t$, $z_t \sim t_\nu(0, 1)$, $v_t \sim N(0, 1)$

SV-N: $y_t = \mu + \exp(h_t/2)z_t$, $h_t = \gamma + \delta h_{t-1} + \sigma_v v_t$, $z_t \sim N(0, 1)$, $v_t \sim N(0, 1)$.

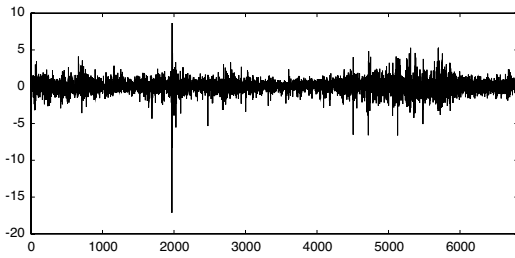


Fig. 1. CRSP value-weighted portfolio index returns from Jan. 2, 1980–Dec. 29, 2006 ($n = 6815$).

Research in Security Prices (CRSP) value-weighted portfolio index over the trading days January 2, 1980 to December 29, 2006. Fig. 1 plots the percentage returns (the return series multiplied by 100). The CRSP portfolio returns average 0.0529 during this time period with a variance of 0.9225. Non-Gaussian behavior is seen in the return’s significantly negative skewness of -0.9837 and highly elevated kurtosis measure of 22.9538.

In addition to modeling the CRSP returns with the SV-DPM, we also apply a stochastic volatility model with normal innovations (SV-N):

$$y_t = \mu + \exp(h_t/2)z_t, \quad z_t \sim N(0, 1),$$

$$h_t = \gamma + \delta h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0, 1).$$

Priors are $\mu \sim N(0, 0.1)$, $\gamma \sim N(0, 100)$, $\delta \sim N(0, 100)I_{|\delta| < 1}$, and $\sigma_v^2 \sim \text{Inv-}\Gamma(10/2, 0.5/2)$ and are similar to those of Jacquier et al. (1994). We also estimate a stochastic volatility model with Student- t return innovations (SV-t):

$$y_t = \mu + \exp(h_t/2)z_t, \quad z_t \sim St(0, (\nu - 2)/\nu, \nu),$$

$$h_t = \gamma + \delta h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0, 1),$$

where $St(0, (\nu - 2)/\nu, \nu)$ is a Student- t density standardized to have variance 1, and ν degrees of freedom. The priors are the same as in the SV-N model with $\nu \sim U(2, 100)$.

The priors for the SV-DPM and SV-DPM-P models are chosen to match the parametric SV models with $\delta \sim N(0, 100)I_{|\delta| < 1}$, $\sigma_v^2 \sim \text{Inv-}\Gamma(10/2, 0.5/2)$. The specific DPM prior is the base distribution, $G_0 \sim N(0, (10\lambda_t^2)^{-1}) - \Gamma(10/2, 10/2)$, and the precision parameter prior $\alpha \sim \Gamma(2, 8)$. This base distribution provides a wide range of empirically realistic values for the mean and precision. Section 5.1 considers other priors.

Estimation of the SV-N and SV-t models is carried out with the hybrid Gibbs, Metropolis–Hastings sampler of Jacquier et al. (2004) except that we use the random block sampler of Section 3.2 for h . Sampling of the degree of freedom parameter for the SV-t uses a tailored proposal density based on a quadratic approximation of the conditional posterior density at its mode.

To eliminate any dependencies on the initial volatilities, 1000 sweeps of the step-by-step volatility sampler of Kim et al. (1998) are carried out for each model while holding the initial parameter values constant. 30,000 sweeps of the sampler for the SV-N and SV-t models are then conducted of which we keep the last 10,000 draws for inference of the two models.

We increase the efficiency of the SV-DPM sampler and reduce the total computing time by respectively taking every tenth draw of the SV-DPM model’s sampler. To reduce the sampler dependency on the starting parameters and volatilities, the first 1000 thinned draws of each chain are discarded, leaving a total of 30,000 thinned draws for inference.

Table 1 reports the MCMC sample means and standard deviations for the parameters of the SV-DPM, SV-DPM-P, SV-t, and SV-N models. We report the observed serial correlation in the draws of the SV-DPM model parameters with the inefficiency measure:

$$1 + 2 \sum_{\tau=1}^L \frac{L - \tau}{L} \rho(\tau),$$

where $\rho(\cdot)$ is the sample autocorrelation function of the parameter draws, $L = 1000$ is the largest lag at which the autocorrelation function is computed. The inefficiency measure quantifies the loss associated with using correlated draws from the sampler, as opposed to truly independent draws, in computing the posterior mean. The numerical standard error equals the square root of the product of the inefficiency measure and the sample variance of the draws (Geweke, 1992).

The posterior estimate of the variance of volatility parameter, σ_v^2 , is the smallest with the SV-DPM model. The posterior estimate of σ_v^2 is 0.0149 with a standard deviation of 0.0023. This mean and standard deviation for σ_v^2 is substantially smaller than the SV-N models mean of 0.0276 and standard deviation of 0.004. For the SV-N model this is to be expected, given that the SV-N model requires a larger value of σ_v^2 in order to capture the excess kurtosis found in the return data.

Excess kurtosis is still, however, unaccounted for by the SV-N return process (Bakshi et al., 1997; Chib et al., 2002). A better characterization of the kurtosis is found in the SV-DPM, SV-DPM-P, and SV-t models where the distribution of the return process is fit by a fat-tailed mixture of normals. Mixture models assign volatile time periods to draws from the tail of the return distribution rather than to a more volatile volatility process. As a result the σ_v^2 in the SV-t and SV-DPM-P models are smaller in value than in the SV-N model, and closer to the SV-DPM ones, with the SV-t posterior mean and standard deviation of σ_v^2 equaling 0.0154 and 0.0023, respectively. For the SV-DPM-P model σ_v^2 is somewhat larger with a posterior mean of 0.0205 but still smaller than for the SV-N model. In Fig. 2 the posterior densities of σ_v^2 are consistent with these observations. Notice that the upper tail, for both the SV-DPM

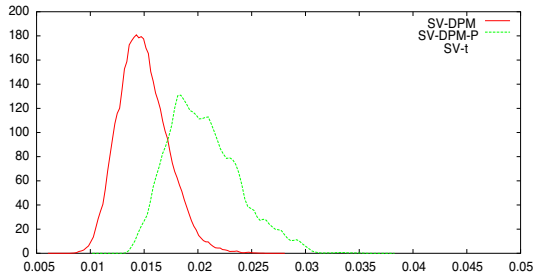


Fig. 2. Posterior density of σ_v^2 for the SV-DPM, SV-DPM-P, SV-t and SV-N models applied to the value-weighted CRSP portfolio daily return data.

and SV-t models' density, for σ_v^2 barely overlaps with the lower tail of the SV-N model's density.

Dynamic behavior in volatility as captured by the AR parameter δ is nearly indistinguishable between the four SV models. First-order dynamics in the volatility of the SV-DPM model is precisely estimated at 0.9900 with the tight posterior standard deviation of 0.0022. This estimate of δ is only slightly larger than the SV-DPM-P and SV-t estimates of 0.9852 and 0.9878 respectively. The volatility in the SV-N model reverts to its mean at a slightly slower pace with a posterior estimate of δ equal to 0.9795.

For the daily portfolio return the average SV-DPM mixture order is $k = 4.6006$ and this suggests that the SV-DPM not only captures the daily stock return leptokurtotic behavior, but its skewness too. The less flexible SV-DPM-P model requires significantly more clusters ($k = 21.589$) to fit the data. Because of the SV-N model's symmetrical Gaussian innovations, it is unable to account for this asymmetrical behavior. Instead, it compensates for this skewness behavior by increasing its level of volatility during those periods where volatility is highest.

This increase in the volatility of the SV-N model relative to the SV-DPM and SV-t models is apparent in Fig. 3 where the SV-DPM posterior conditional variance of returns is plotted in Panel (a) and the SV-DPM model's differences from the conditional variances of the SV-N model are graphed in Panel (b) and those from the SV-t model in Panel (c). There are differences in each of the models, with the largest disagreement being for the 1987 crash and the days following it. The crash period is captured as a spike in (a). Both fully parametric models have a larger variance during the crash than that of the SV-DPM model. For the SV-N the variance is more than 10 points larger while for the SV-t it is closer to 1 point larger than the SV-DPM.

As for the behavior of skewness, because of their symmetrical distributions, the SV-DPM-P, SV-N and SV-t models are all unable to capture the skewness of daily returns. This is borne out in the one-day-ahead, out-of-sample, predictive density plots of Fig. 4 and the log density in Fig. 5. The SV-DPM predictive density is clearly different from the SV-N or SV-t models. For example, the SV-DPM predictive density is more centered around 0 and exhibits the asymmetry associated with the negative skewness of returns. The log predictive density plots show the SV-DPM producing fatter tails than the other models. The tails of SV-DPM-P are fatter than the SV-N ones but thinner than the SV-t ones. Like for these parametric models, the SV-DPM-P tails are essentially symmetric.

5.1. Robustness to DP hyperparameters

Using the same empirical data set of CRSP portfolio returns we estimate the SV-DPM model under five different prior specifications of $\pi(\alpha) \equiv \Gamma(a, b)$ and $G_0 \equiv N(m, (\tau\lambda_t^2)^{-1}) - \Gamma(v_0/2, s_0/2)$ to test the robustness of the posterior estimates of the SV-DPM model to different priors. Table 2 reports these robustness findings for the posterior estimates of the SV-DPM model for the different priors.

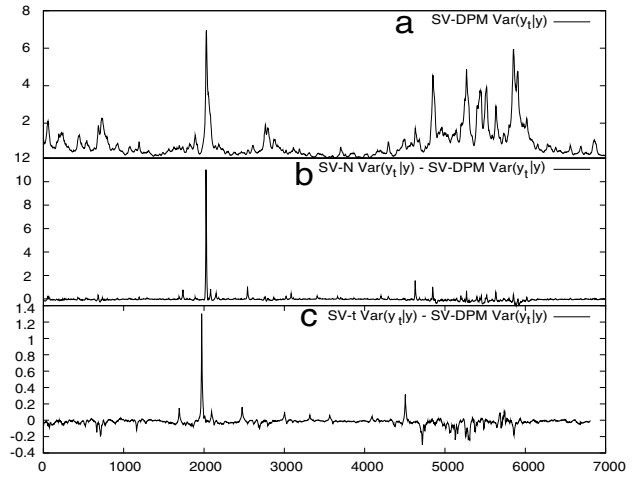


Fig. 3. The SV-DPM posterior variance of returns, $\text{Var}[Y_t|y]$, for the value-weighted CRSP index returns (Panel a), and its difference from the SV-N (Panel b) and SV-t (Panel c) models.

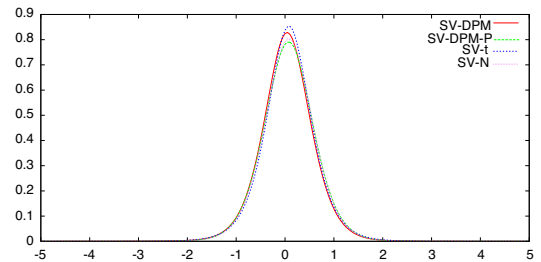


Fig. 4. Predictive density, $f(Y_{t+1}|y)$, of the SV-DPM, SV-N, and SV-t models for the value-weighted CRSP portfolio daily return.

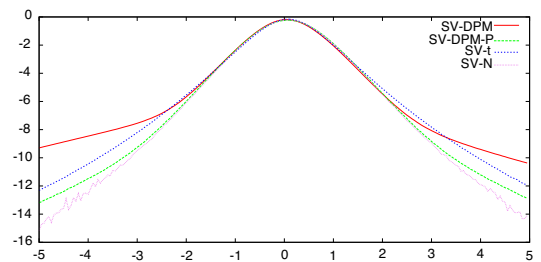


Fig. 5. Log predictive density, $\ln f(Y_{t+1}|y)$, of the SV-DPM, SV-N, and SV-t models for the value-weighted CRSP portfolio daily return.

Table 2

Robust sensitivity analysis of the SV-DPM to different precision parameter and base distribution priors for daily returns of the value-weighted CRSP portfolio from Jan. 2, 1980 to Dec. 29, 2006 (6815 observations, 30,000 thinned draws of the SV-DPM sampling algorithm where every tenth draw is retained and the first 1000 thinned draws are discarded).

	Prior 2	Prior 3	Prior 4	Prior 5	Prior 6
δ	0.9897 (0.0022)	0.9901 (0.0022)	0.9902 (0.0022)	0.9902 (0.0022)	0.9900 (0.0022)
σ_v^2	0.0154 (0.0023)	0.0147 (0.0023)	0.0147 (0.0022)	0.0145 (0.0022)	0.0149 (0.0023)
α	0.0604 (0.0553)	0.3371 (0.1853)	0.3330 (0.1846)	0.3313 (0.1798)	0.3451 (0.1945)
k	2.6603 (0.8952)	4.7302 (1.9750)	4.6341 (1.9555)	4.5886 (1.7908)	4.8409 (2.1363)

The posterior mean and standard deviation (in parenthesis) are reported.

To determine the impact that the prior of the precision parameter has on the estimates of the SV-DPM model we evaluate the model under the prior specification

- Prior 2: $\pi(\alpha) \sim \Gamma(0.1, 20)$,

where $E[\alpha] = 0.005$ and $\text{Var}[\alpha] = 0.00025$, and leave the other priors exactly as before. These hyperparameter values cause the prior distribution for α to be more tightly distributed and centered closer to zero than the original prior. As a result the posterior estimate of α is found to be closer to zero at 0.0604. Since a smaller value for α lowers the probability of selecting a new cluster from the Polya urn, under Prior 2 the estimate of k is smaller at 2.6603. Though the mixture representation for the distribution of returns now on average consists of fewer clusters, notice that the posterior estimates of the volatility parameters, δ and σ_v^2 , and their standard deviations are nearly the same as under the original prior.

In the other four priors we allow the DP prior's base distribution $N(m, (\tau\lambda_t^2)^{-1}) - \Gamma(v_0/2, s_0/2)$ to change in order to explore how sensitive the posterior estimates of the SV-DPM model are to prior's mean and spread. The four priors are

- Prior 3: $G_0 \equiv N(0, (5 * \lambda^2)^{-1}) - \Gamma(10/2, 10/2)$,
- Prior 4: $G_0 \equiv N(0, (15 * \lambda^2)^{-1}) - \Gamma(10/2, 10/2)$,
- Prior 5: $G_0 \equiv N(0, (10 * \lambda^2)^{-1}) - \Gamma(5/2, 5/2)$,
- Prior 6: $G_0 \equiv N(0, (10 * \lambda^2)^{-1}) - \Gamma(15/2, 15/2)$,

where Prior 3 & 4 change the variance of the mixture mean, η , and Prior 5 & 6 test for the robustness to changes in the prior of the mixture variance, λ^2 . In the posterior results reported in Table 2, the changes in the hyperparameters for η and λ^2 base distributions do not affect the posterior estimates of the SV-DPM model. Under each of the four priors the estimates of δ are the same up to the third decimal place at 0.990, and the estimates of σ_v^2 are equal out to the second decimal place at 0.01. Differences between the estimates of α are only found under Prior 2 while the remaining estimates have a posterior mean of approximately 0.33. Similar results are found for k for Prior 2 which produces the smallest estimate of $k = 2.6603$, while k is 4.6–4.8 for the other priors.

5.2. Robustness to number of draws

Because the DPM sampler is a step-by-step algorithm, making 30,000 thinned draws from the SV-DPM models is costly. This is understandable given the level of inefficiency associated with the posterior draws of the SV-DPM model. It would, however, be preferable if fewer draws could be used in making inferences concerning the SV-DPM model. To determine whether this is possible, the SV-DPM model for the CRSP portfolio return data is re-estimated with a MCMC sample of 10,000 thinned draws. The posterior results of the SV-DPM model from these 10,000 draws are reported in Table 3. The table also includes the results from Table 1 where 30,000 draws were made. Notice that there is little difference between the posterior means of the parameters. The volatility parameters, δ and σ_v^2 , have comparable posterior means and exactly the same standard deviations. The DP parameters α and k are also very similar.

5.3. Model comparison

The previous large sample analysis highlighted features of the predictive density that the standard parametric SV models could not account for. In this section we investigate the forecasting value of the predictive densities of the SV-DPM specifications in a small sample setting, using 755 daily CRSP returns over the period January 3, 2006 to December 31, 2008. Given the existing results on the good performance of the basic parametric SV models we focus on the relative value that the new models contribute to density forecasts. To do this we use the model pooling approach of Geweke and Amisano (2008). This approach recognizes that none of the models may be the true data generating process and advocates a

Table 3

Robust sensitivity analysis of the SV-DPM to the number of MCMC draws for daily returns of the value-weighted CRSP portfolio from Jan. 2, 1980 to Dec. 29, 2006 (6815 observations). T thinned MCMC draws where every tenth draw is retained and the first 1000 thinned draws are discarded.

T	30,000			10,000		
	Mean	Stdev	Ineff	Mean	Stdev	Ineff
δ	0.9900	0.0022	13.510	0.9901	0.0022	16.187
σ_v^2	0.0149	0.0023	33.195	0.0149	0.0022	39.031
α	0.3323	0.1873	20.571	0.3365	0.1911	29.518
k	4.6006	1.9579	49.939	4.6602	1.9640	69.544

linear prediction pool based on the log score function (predictive likelihood) from a set of models.

Given a set of predictive densities $\{f(y_t|y_1, \dots, y_{t-1}, M_i)\}_{i=1}^K$ from the set of models $\{M_i\}_{i=1}^K$ consider the combined predictive density of the form

$$\sum_{i=1}^K w_i f(y_t|y_1, \dots, y_{t-1}, M_i),$$

$$\sum_{i=1}^K w_i = 1, w_i \geq 0, i = 1, \dots, K.$$

Weights are chosen to maximize the log pooled, predictive score function:

$$\max_{w_i, i=1, \dots, K} \sum_{t=\tau_1}^{\tau_2} \log \left[\sum_{i=1}^K w_i f(y_t|y_1, \dots, y_{t-1}, M_i) \right], \tag{15}$$

where the predictive densities are evaluated at the realized data point y_t .

For each of the models we run a MCMC simulation consisting of 11,000 draws of which the first 1000 draws are thrown away to obtain 10,000 posterior draws conditional on the return data up to time period $t - 1$; i.e., y_1, \dots, y_{t-1} . These draws are then used to estimate the predictive likelihood $f(y_t|y_1, \dots, y_{t-1}, M_i)$.⁸ For the SV-DPM model the predictive likelihood is estimated using Eq. (14). MCMC draws of this size are carried out for each SV model and data set y_1, \dots, y_{t-1} where $t = \tau_1, \dots, \tau_2$. Given a history of predictive likelihood values for each model we can estimate the weights in Eq. (15).

The pool of models considered comprises: SV-DPM; SV-DPM-P; SV-t and SV-N; i.e., $K = 4$. Recall that in the SV-DPM-P model of Section 2 only the return precision parameter λ_t^2 is governed by the DP prior and the intercept is assumed to be the unknown constant μ . Conditional on return data back to January 3, 2006 ($t = 1$), we compute the log pooled predictive score function over the period of May 30, 2006 ($\tau_1 = 105$) to December 31, 2008 ($\tau_2 = 755$).⁹

Table 4 displays the optimal log score and the weights for the linear pool of models. Using all four models the log score is -886.38. The SV-DPM-P model dominates with a weight of 0.73 followed by the SV-t model with 0.27. Each of the subsequent table entries drop one of the models from the pool to assess the deleted model's relative importance for forecasting as measured by the model's contribution to the log score. As long as the SV-DPM-P model is in the pool, a similar log score is achieved, but once this model is dropped the log score declines just under 3

⁸ Because of the large number of predictive likelihoods that are required in the pooled predictive score function, the number of MCMC draws is smaller than the sampling performed in Section 5. For the largest series (745 observations) the SV-DPM sampler's compiled C-code takes just over 6 min on a 3 GHz Intel Xeon quad-core computer running Linux.

⁹ We decrease the computing time involved in calculating the pooled predictive score function by distributing the calculation of, for each model, 650 predictive likelihoods, $f(y_t|y_1, \dots, y_{t-1})$, $t = 105, \dots, 755$, to 25–30 separate processors each using the same initial values.

Table 4

The optimal pooled log predictive score function $\max_w f(w)$ and optimal weight vector $w^* = \arg \max_w f(w)$ where $f(w) \equiv \sum_t \log[\sum_i w_i f(y_t | y_1, \dots, y_{t-1}, M_i)]$, with t summing over the weighted combination of each model's one-day-ahead predictive likelihoods from May 30, 2006 ($t = 105$) to Dec. 31, 2008 ($t = 755$), conditional on return data back to Jan. 3, 2006 ($t = 1$). The x denotes a SV model being dropped from the predictive pool of models.

Log score	w_{DPM}^*	$w_{\text{DPM-P}}^*$	$w_{\text{msSV-t}}^*$	$w_{\text{SV-N}}^*$
−886.38	0	0.7280	0.2720	0
−886.38	0	0.7280	0.2720	x
−887.43	0	0.9469	x	0.0531
−889.14	0.3915	x	0.3881	0.2205
−886.38	x	0.7280	0.2720	0

points to −889.14. The SV-DPM-P nests both the SV-N and the SV-t model. The SV-t models a distinct precision parameter value for each observation, whereas the SV-DPM-P models prior leads to a clustering of distinct precision parameter values that are fewer in number than the sample size.¹⁰ The zero or near zero weight and lack of contribution to the pooled predictive likelihood function from the SV-DPM and SV-N models is likely due to the fact that to learn about asymmetry in the return distribution requires more observations than our data series of 755 returns affords.

6. Conclusion

This paper proposed a new Bayesian, semiparametric, autoregressive, stochastic volatility model where the conditional return distribution is modeled nonparametrically with an infinite ordered mixture of normal distributions. The unknown number of mixture clusters, their probability of occurrence, and their mean and variance are flexibly modeled *a priori* with a Dirichlet process prior. Conditional on a draw of the log volatilities, an efficient MCMC algorithm has been constructed to produce posterior draws of the unknown number of mixture clusters and the clusters' mean and variance. The sampler has been stress tested against existing parametric stochastic volatility models on real world daily return data. The semiparametric stochastic volatility model performed well on empirical return data, fitting both the negative skewness and leptokurtotic properties of returns, while still capturing the time-varying conditional heteroskedastic dynamics of returns. The semiparametric model's increased flexibility and robustness to non-Gaussian behavior and its superior forecasts makes it an appealing specification for risk and portfolio managers. The SV-DPM models can provide improvements in both large and small samples.

Important questions remain to be answered with the Bayesian semiparametric, stochastic volatility model. For instance, is it possible to attach structural meaning to the mixture parameters, such as a particular mixture cluster being identified with jumps in returns or to time periods where the economy is in a particular state of the business cycle? Placing such structural meaning on the mixture clusters is possible by assigning a prior rank ordering to the clusters within the Dirichlet process prior. Doing so overcomes the label switching problem discussed earlier.

Another area of potential research is that of leverage effects. Leverage effects have been used effectively with symmetrically distributed stochastic volatility models to produce negative skewness in returns. A natural question that one could ask is whether it is possible to introduce leverage effects into this paper's semiparametric, stochastic volatility model. If so, how do leverage effects affect the skewness of the mixture distribution? These and other interesting questions remain for future research.

References

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative options pricing models. *Journal of Finance* 52 (5), 2003–2049.

- Blackwell, D., MacQueen, J., 1973. Ferguson distributions via Polya urn schemes. *The Annals of Statistics* 1, 353–355.
- Chib, S., Hamilton, B., 2002. Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics* 110, 67–89.
- Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108 (2), 281–316.
- Clark, P.K., 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41 (1), 135–155.
- Conley, T., Hansen, C., McCulloch, R., Rossi, P., 2008. A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics* 144, 276–305.
- Durham, G. B., 2006. Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models. *Journal of Econometrics* 133 (1), 273–305.
- Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed non-linear diffusions. *Econometrica* 69, 959–993.
- Eraker, B., Johannes, M.S., Polson, N.G., 2003. The impact of jumps in volatility and returns. *Journal of Finance* 58 (3), 1269–1300.
- Escobar, M.D., West, M., 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90 (430), 577–588.
- Ferguson, T., 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1 (2), 209–230.
- Fleming, J., Kirby, C., 2003. A closer look at the relationship between GARCH and stochastic autoregressive volatility. *Journal of Financial Econometrics* 1, 365–419.
- Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. In: Springer Series in Statistics. New York/Berlin/Heidelberg.
- Gallant, A.R., Hsieh, D., Tauchen, G., 1997. Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81, 159–192.
- Gelfand, A.E., Mukhopadhyay, S., 1995. On nonparametric Bayesian inference for the distribution of a random sample. *The Canadian Journal of Statistics* 23 (4), 411–420.
- Geweke, J., 1994. Comment on Jacquier, Polson and Rossi's "Bayesian analysis of stochastic volatility models". *Journal of Business & Economic Statistics* 12, 389–392.
- Geweke, J., 1992. Evaluating the accuracy of sampling based approaches to the calculation of posterior moments. In: Bernardo, Berger, Dawid, Smith (Eds.), *Bayesian Statistics*, vol. 4. Clarendon Press, Oxford.
- Geweke, J., 2007. Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis* 51 (7), 3529–3550.
- Geweke, J., Amisano, G., 2008. Comparing and evaluating Bayesian predictive distributions of asset returns. *European Central Bank Working Paper*, vol. 969.
- Geweke, J., Keane, M., 2007. Smoothly mixing regressions. *Journal of Econometrics* 138, 252–290.
- Ghosal, S., Ghosh, J., Ramamoorthy, R., 1999. Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics* 27 (1), 143–158.
- Ghosal, S., van der Vaart, A., 2007. Posterior convergence rates of Dirichlet mixtures at smooth densities. *Annals of Statistics* 35 (2), 697–723.
- Gonedes, R.C.B.N.J., 1974. A comparison of the stable and student distributions as statistical models for stock prices. *The Journal of Business* 47 (2), 244–280.
- Griffin, J.E., Steel, M., 2006. Order-based dependent Dirichlet processes. *Journal of the American Statistical Association* 101 (473), 179–194.
- Griffin, J.E., Steel, M.F.J., 2004. Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics* 123 (1), 121–152.
- Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. *The Review of Economic Studies* 61 (2), 247–264.
- Hirano, K., 2002. Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica* 70, 781–799.
- Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* 12, 371–417.
- Jacquier, E., Polson, N.G., Rossi, P.E., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122, 185–212.
- Jensen, M.J., 2004. Semiparametric Bayesian inference of long-memory stochastic volatility models. *Journal of Time Series Analysis* 25 (6), 895–922.
- Jensen, M.J., Maheu, J.M., 2008. Bayesian semiparametric stochastic volatility modeling. Working Paper 314, Department of Economics, University of Toronto.
- Kacperczyk, M., Damien, P., Walker, S., 2005. A new class of Bayesian semiparametric models with application to option pricing. Manuscript University of British Columbia.
- Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Kon, S., 1984. Models of stock returns—A comparison. *Journal of Finance* 39, 147–165.
- Lau, J.W., Siu, T.K., 2008. Modelling long-term investment returns via Bayesian infinite mixture time series models. *Scandinavian Actuarial Journal* 108, 243–282.
- Lau, J.W., So, M.K., 2008. Bayesian mixture of autoregressive models. *Computational Statistics & Data Analysis* 53 (1), 38–60.
- Liesenfeld, R., Jung, R.C., 2000. Stochastic volatility models: Conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* 15, 137–160.
- Lo, A.Y., 1984. On a class of Bayesian nonparametric estimates. I. Density estimates. *The Annals of Statistics* 12, 351–357.
- MacEachern, S.N., Müller, P., 1998. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* 7 (2), 223–238.

¹⁰ The posterior mean of the number of clusters is 8.

- Mahieu, R.J., Schotman, P.C., 1998. An empirical application of stochastic volatility models. *Journal of Applied Econometrics* 13 (4), 333–360.
- Meddahi, N., 2001. An eigenfunction approach for volatility modeling. CRDE 29-2001.
- Müller, P., Quintana, F.A., 2004. Nonparametric Bayesian data analysis. *Statistical Science* 19 (1), 95–110.
- Muller, P., West, M., MacEachern, S., 1997. Bayesian models for non-linear autoregressions. *Journal of Time Series Analysis* 18, 593–614.
- Omori, Y., Chib, S., Shephard, N., Nakajima, J., 2007. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* 140 (2), 425–449.
- Pitt, M.K., Shephard, N., 1997. Likelihood analysis of non-gaussian measurement time series. *Biometrika* 84, 653–667.
- Praetz, P.D., 1972. The distribution of share price changes. *The Journal of Business* 45 (1), 49–55.
- Press, S.J., 1967. A compound events model for security prices. *Journal of Business* 40, 317–335.
- Richardson, S., Green, P.J., 1997. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B* 59, 731–792.
- Robert, C.P., Casella, G., 1999. *Monte Carlo Statistical Methods*. Springer, New York.
- Sethuraman, J., 1994. A constructive definition of Dirichlet priors. *Statistica Sinica* 4, 639–650.
- Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Society* 82, 528–550.
- Taylor, S.J., 1986. *Modeling Financial Time Series*. John Wiley.
- Tiwari, R., Jammalamadaka, S., Chib, S., 1988. Bayes prediction density and regression estimation: A semi parametric approach. *Empirical Economics* 13, 209–222.
- West, M., Muller, P., Escobar, M., 1994. Hierarchical priors and mixture models with applications in regression and density estimation. In: Freeman, P.R., Smith, A.F. (Eds.), *Aspects of Uncertainty*. John Wiley.
- Yu, J., 2005. On leverage in a stochastic volatility model. *Journal of Econometrics* 127 (2), 165–178.