FEDERAL RESERVE BANK *of* ATLANTA

Bayesian Semiparametric Stochastic Volatility Modeling

Mark J. Jensen and John M. Maheu

Working Paper 2008-15
June 2008

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA | WORKING PAPER SERIES

Bayesian Semiparametric Stochastic Volatility Modeling

Mark J. Jensen and John M. Maheu

Working Paper 2008-15
June 2008

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, we use nonparametric Bayesian methods to flexibly model the skewness and kurtosis of the distribution while continuing to model the dynamics of volatility with a parametric structure. Our semiparametric Bayesian approach provides a full characterization of parametric and distributional uncertainty. We present a Markov chain Monte Carlo sampling approach to estimation with theoretical and computational issues for simulation from the posterior predictive distributions. The new model is assessed based on simulation evidence, an empirical example, and comparison to parametric models.

JEL classification: C11, C14, C53

Key words: Bayesian nonparametrics, Dirichlet process mixture prior, Markov chain Monte Carlo, mixture models, stochastic volatility

---

The authors thank the seminar participants at the Canadian Econometric Study Group conference held in Montreal, the All-Georgia Conference held at the Federal Reserve Bank of Atlanta, the Rimini Center for Economic Analysis's 2007 Conference on Econometrics, held in Rimini, Italy, the annual symposium of the Society for Nonlinear Dynamics and Econometrics held at the Federal Reserve Bank of San Francisco, and the Department of Economics at Oregon State University. They also appreciate the comments and suggestions of Thanasis Stengos and George Tauchen. Maheu is grateful to the Social Sciences and Humanities Research Council for financial support. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Mark J. Jensen, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA, 30309-4470, 404-498-8019, mark.jensen@atl.frb.org, or John M. Maheu, Associate Professor, Department of Economics, University of Toronto, 150 St. George Street, Toronto, Canada, M5S 3G7, 416-978-1495, jmaheu@chass.utoronto.ca.

1 Introduction

This paper proposes a model of asset returns that draws from the existing literature on autoregressive stochastic volatility models and the recent advances made in Bayesian nonparametric models and their sampling to create a semiparametric stochastic volatility model of returns. By applying both parametric and nonparametric features to the return process, an estimable stochastic volatility model with a flexible nonparametric innovation distribution is provided. The nonparametric portion of the model consists of an infinitely ordered mixture of normal distributions whose mixture probabilities, means, variances, and most importantly, number of mixtures, are distributed according to a particular Bayesian prior. With this nonparametric representation of the conditional distribution of returns, the predictive density from the model is able to fit both the high level of kurtosis and negative skewness not currently captured with parametric stochastic volatility models. Our approach is likelihood based and provides exact finite sample inference, 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)). The general makeup of these models consist of an infinite mixture of normal distributions with their means fixed to zero and their variances independently and identically distributed (*iid*) over some pre-specified distribution. It is well known that mixture models produce fat-tailed behavior, in other words, levels of kurtosis in excess of normality. However, mixture models alone do not capture the strong level of empirical persistence observed in the conditional variance of returns.

Stochastic volatility models (SV) are designed to fit this time-varying behavior in the conditional variance of returns (see Taylor (1986); Harvey et al. (1994)). Like its nonparametric mixture predecessors, stochastic volatility models are a continuous mixture of normals, however, their variance follows a dynamic stochastic process. This stochastic behavior enables the SV model to produce both the high levels of kurtosis and the persistence found in the conditional variances of returns. Unfortunately, parametric SV models have not fully captured the asymmetries and leptokurtotic behavior present in return data (see Gallant et al. (1997); Mahieu & Schotman (1998), and Durham (2006)). For example, a SV model with a standard normal distribution (SV-N) cannot fit the skewness in returns since its distribution is symmetrical. Furthermore, by its parametric nature the SV-N model is restricted in the level of kurtosis it can produce (see Liesenfeld & Jung (2000) and Meddahi (2001)). Skewness and kurtosis are important distributional features that play an important role in the pricing of derivatives, the measuring of risk, and the selection of a portfolio. A flexible nonparametric version of the SV model will be useful to risk managers and analysts.

The Dirichlet process mixture (DPM) prior consists of modeling the clusters and probabilities of an infinite 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 is a basis function spanning the entire class of continuous distributions (Escobar & West (1995) and Ghosal et al. (1999)), *ii*) as a prior to a infinite ordered mixture model, the DPM is more flexible and realistic than a mixture model with a predetermined number of components, *iii*) with the DPM the data determines the number of mixture clusters that best fit the data, *iv*) parsimony can be imposed through the DPM prior's hyperparameters, *v*) as a conjugate prior the DPM is easy to use and facilitates Gibbs sampling, and *vi*) it works well in practice.[1]

The goal of this paper is to create a flexible semiparametric stochastic volatility model by combining a nonparametric *iid* DPM model of innovations scaled by an autoregressive model of the return's latent conditional variance process. As a semiparametric model the DPM version of the stochastic volatility model nests parametric volatility models commonly used in finance within it. The paper's semiparametric SV model is also capable of modeling skewness, multimodality, and kurtotic type behavior. Because the Dirichlet process prior is a discrete distribution with proba-

[1] Examples of the DPM model in economics include Chib & Hamilton (2002), Griffin & Steel (2004), Hirano (2002), Jensen (2004), Kacperczyk et al. (2005), and Tiwari et al. (1988). Jensen (2004) used 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.

bility one, the DPM places a high probability on mixture models with a manageable number of mixture clusters. The posterior distribution of the return innovations will thus consist of a finite mixture whose exact number of clusters and parameter values will be determined by the return data. In other words, by using the DPM we are able to impose parsimony through the prior, which is important for producing good forecasts. Something that is difficult in the classical setting.

A Markov chain Monte Carlo (MCMC) sampler is constructed to estimate the unknown parameters of our Bayesian semiparametric SV model. The paper's MCMC algorithm extends the DPM samplers of West et al. (1994) and MacEachern & Müller (1998) to the time-varying structure of the stochastic volatility model. Due to the independence between the volatility process and the Dirichlet process mixture model, a tractable efficient posterior sampler is possible. Conditional on the value of the other, one block of the sampler consists of drawing the parameters associated with the DPM, whereas in the other blocks the parametric parameters and latent volatility associated with the stochastic volatility model are drawn (see Chib et al. (2002), Eraker et al. (2003), Jacquier et al. (1994, 2004), and Kim et al. (1998)). In addition to providing smoothed estimates of the latent volatility process, the sampler generates a predictive density for returns that fully accounts for the uncertainty in the volatility process as well as the unknown return distribution.

A second contribution found in the paper is a simple random block sampler of latent volatility. We extend Fleming & Kirby (2003) block sampler of volatility by including the return data in the sampler's proposal distribution. This results in better candidate draws to the Metropolis-Hasting sampler resulting in lower correlation among draws and fewer sweeps. The sampler can also be used for all the SV models discussed in the paper.

We evaluate our semiparametric SV model against standard SV models found in the literature; the SV-N model and the SV model with Student-t innovations (SV-t). In simulation studies, we find that the semiparametric model accurately captures the return distribution and volatility clustering. The parametric models display severe parameter bias when they misspecify the conditional distribution while the semiparametric model performs well for each simulated SV model. In an empirical application with daily CRSP return data, the predictive distribution for the semiparametric model is very different from the parametric SV models. The semiparametric SV model's predictive density displays negative skewness and kurtosis whereas neither the SV-N nor SV-t do. 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 *iid* DPM component of returns.

The paper is organized as follows. The next section introduces basic concepts concerning Bayesian nonparametrics including the Dirichlet process prior and the Dirichlet process mixture model. The semiparametric stochastic volatility model with DPM return innovations is discussed in Section 3. Section 4 present Bayesian inference for the model and Section 5 discusses features of the model. Simulation examples comparing existing parametric models with our semiparametric model are presented in Section 6 while an application to daily return data is found in Section 7. Section 8 contain our conclusions and suggestions for possible future extensions for our Bayesian semiparametric SV model.

2 Bayesian Nonparametric Modeling

2.1 Dirichlet Process Prior

Let $z_1, z_2, \ldots, z_n$ be a sequence of independently and identically distributed random variables defined on some measurable space $(\Phi, \mathcal{F})$ whose probability distribution function F is unknown. Being unknown, F represents the "parameter" in a nonparametric model of the z's distribution. As with all Bayesian estimators, estimating F requires placing a prior distribution on it. In an effort to produce a prior for F whose support is not only large enough to span the space of probability distribution functions, but also a prior that will lead to an analytically manageable posterior distribution, Ferguson (1973) derived the Dirichlet process prior. A Dirichlet process prior, denoted by $F \sim DP(G_0, \alpha)$, with base distribution G_0 and scalar precision parameter, $\alpha > 0$, generates the random probability distribution F if for all finite measurable partitions, $\{\Phi_i\}_{i=1}^{J}$, of the sample space, Φ, the distribution of the random vector, $(F(\Phi_1), \ldots, F(\Phi_J))$, is the Dirichlet distribution with parameters $(\alpha G_0(\Phi_1), \ldots, \alpha G_0(\Phi_J))$.

To better understand and appreciate the flexibility of the Dirichlet process as a prior for F, let $\{\Phi_0, \Phi_1\}$ form a simple partition of Φ; i.e., $\Phi_0 \cup \Phi_1 = \Phi$ and $\Phi_0 \cap \Phi_1 = \{\}$. Using the definition of $DP(G_0, \alpha)$, the random function $F(\Phi_0)$ will be distributed as a $Beta(\alpha G_0(\Phi_0), \alpha G_0(\Phi_1))$ distribution (when $J = 2$, the Beta distribution is a special case of the Dirichlet distribution). It follows from the properties of the Beta distribution and G_0 being a probability measure that the prior for F has a mean distribution of $E[F(\Phi_0)] = G_0(\Phi_0)$ with variance $\text{Var}[F(\Phi_0)] = G_0(\Phi_0)G_0(\Phi_1)/(\alpha + 1) = G_0(\Phi_0)(1 - G_0(\Phi_0))/(\alpha + 1)$. In other words, the $DP(G_0, \alpha)$ prior for F centers F around G_0. Because α is found in the denominator of $\text{Var}[F]$, larger values of α lead to the prior of F having a smaller variance. Hence, α can be viewed as measuring one's belief as to how well G_0 represents F.

In our example, the conjugacy property of the Beta distribution with the binomial likelihood function for $z = (z_1, \ldots, z_n)'$ leads to the posterior distribution:

$$F(\Phi_0)|z \sim Beta\left(\alpha G_0(\Phi_0) + \sum_{i=1}^{n} \delta_{z_i}(\Phi_0),\ \alpha G_0(\Phi_1) + \sum_{i=1}^{n} \delta_{z_i}(\Phi_1)\right),$$

where $\delta_{z_i}(\cdot)$ is the Dirac function such that $\delta_{z_i}(\Phi_j) = 1$ if $z_i \in \Phi_j$ and zero otherwise. It follows from the properties of the Beta distribution that the posterior mean and Bayesian estimate of $P[Z \in \Phi_0]$ equals:

$$E[F(\Phi_0)|z] = \frac{\alpha}{\alpha + n} G_0(\Phi_0) + \frac{n}{\alpha + n} \sum_{i=1}^{n} \delta_{z_i}(\Phi_0)/n.$$

$E[F|z]$ is equivalent to a Polya urn scheme (see Blackwell & MacQueen (1973)). A Polya urn scheme involves sequentially drawing from an urn filled with colored balls whose colors are distributed according to the distribution G_0. Upon observing the color of the sampled ball another ball of exactly the same color is added to the urn along with the sampled ball. With this interpretation of the DP-prior, z_1 is distributed as the base distribution G_0 (assuming $\alpha \neq 0$) since there are no other observations. The distribution of subsequent z_is is either the empirical distribution of the observed $z_1, \ldots, z_{i-1}$, or like z_1, distributed as G_0. Notice also that as more and more z_is are observed, in other words, as $n \to \infty$, the distribution of z_n will tend to the empirical distribution of the observed z_is; i.e., $E[F|z] \to \sum_{i=1}^n \delta_{z_i}/n$ as $n \to \infty$.

The Dirichlet process's posterior properties for the partition, $\{\Phi_0, \Phi_1\}$, apply in a general manner to all partitions of Φ. Thus, the Dirichlet process prior, $z_i|F \sim DP(G_0, \alpha)$, $i = 1, \ldots, n$, produces the Dirichlet process posterior distribution $F|z \sim DP(G_0^*, \alpha + n)$, where $G_0^* = \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^n \delta_{z_i}$. As a conjugate prior to multinomial outcomes, the DP is thus both manageable and intuitive, leading to a posterior distribution equal to a weighted average of the prior, G_0, and the empirical distribution, $n^{-1} \sum_{i=1}^n \delta_{z_i}$.

The DP-prior for F can be concisely written in terms of an infinite mixture of point mass functions:

$$F = \sum_{j=1}^{\infty} V_j \delta_{Z_j},$$

where the probabilities are defined by $V_1 = W_1$, and $V_j = W_j \prod_{s=1}^{j-1}(1 - W_s)$ with $W_j \sim Beta(1, \alpha)$, and $Z_j \sim G_0$, $j = 1, 2, \ldots$ (Sethuraman (1994)). This mixture representation of the DP-prior helps illustrate why it is referred to as a stick-breaking prior. At each stage j a stick initially of unit length is independently and randomly broken into length V_j by breaking off W_j percent of the remaining stick. This stick-breaking representation of F, however, also reveals one of the DP-prior's shortcomings. Although the DP-prior spans the space of all discrete probability distributions it does so with probability one. As a result the class of continuous distributions lies outside the scope of the DP-prior.

2.2 Dirichlet Process Mixture

A prior that does span the entire set of continuous probability distributions with probability one is the Dirichlet process mixture (DPM) model:

$$z_i \overset{iid}{\sim} \sum_{j=1}^{\infty} V_j f(\cdot|\phi_j), \tag{1}$$

where f is a continuous, nonnegative valued kernel and V_j and ϕ_j, $j = 1, \ldots,$ are defined in the same stick-breaking manner as Section 2.1.[2] In addition to this Setheruman type representation, the DPM also has the hierarchical form:

$$\begin{aligned} z &\sim f(\cdot|\phi), \\ \phi|G &\overset{iid}{\sim} G, \\ G|G_0, \alpha &\sim DP(G_0, \alpha). \end{aligned}$$

Under the DPM prior the unknown distribution F is modeled as a mixture of mixtures with a countably infinite number of clusters. With an infinite number of clusters the DPM is more flexible than a finite ordered mixture model. It also eliminates the trouble of having to choose the "best" number of clusters (see Richardson & Green (1997) for a Bayesian approach to inferring the correct number of clusters for a finite mixture model).

Suppose $f(\cdot|\phi_j)$ is the normal density function where $\phi_j = (\eta_j, \lambda_j^{-2})$, η_j is the mean, and λ_j^{-2} the variance. If we make no distributional assumptions concerning V_j or ϕ_j, estimating F cannot be carried out since the model's infinite number of unknowns, $\{V_j, \phi_j\}_{j=1,\ldots}$, are not identified by a finite length vector z. Fortunately, the discrete nature of the Dirichlet process that earlier posed a problem as a prior for F becomes useful as a prior for ϕ_j. Since F's prior models $z_i|\phi_i \overset{iid}{\sim} f(\cdot|\phi_i)$ with $\phi_i|G \overset{iid}{\sim} G$, we can write:

$$z_1, \ldots, z_n|F \overset{iid}{\sim} \int f(\cdot|G)\, G(d\phi), \tag{2}$$

where $G \sim DP(G_0, \alpha)$; i.e., $G = \sum_{j=1}^{\infty} V_j \delta_{\phi_j}$.

Because $\phi_i|G \sim G$ and $G \sim DP(G_0, \alpha)$, our example of the DP-prior in Section 2.1 applies to ϕ. The probability of ϕ_i conditional on the values of $\phi_1, \ldots, \phi_{i-1}$ equals:

$$\begin{aligned} P(\phi_i \in \Phi_0|\phi_1, \ldots, \phi_{i-1}) &= E[G(\Phi_0)|\phi_1, \ldots, \phi_{i-1}] \\ &= \frac{\alpha}{\alpha + i - 1} G_0(\Phi_0) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\phi_j}(\Phi_0). \end{aligned} \tag{3}$$

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

¿From the construct of Equation (3), $\phi_i|\phi_1,\ldots,\phi_{i-1}$ follows a Polya urn scheme. Notice also that since the probability of drawing a new ϕ_i approaches zero as $\alpha \to 0$, a smaller α causes Equation (2) to have fewer clusters and parameters. At the other extreme, as $\alpha \to \infty$, F will be a heavily parameterized mixture model consisting of a large number of clusters where each clusters parameter ϕ_i is a unique realization from G_0.

Combining $\pi(\phi_i|\phi_1,\ldots,\phi_{i-1})$ with the likelihood $f(z_i|\phi_i)$ produces the posterior:

$$\phi_i|\phi_1,\ldots,\phi_{i-1},z_i \sim c\frac{\alpha}{\alpha+i-1}g(z_i)G(d\phi|z_i) + \frac{c}{\alpha+i-1}\sum_{j=1}^{i-1} f(z_i|\phi_j)\delta_{\phi_j} \qquad (4)$$

where $g(z_i) = \int f(z_i|\phi)G_0(d\phi)d\phi$ is the normalizing constant to the posterior distribution $G(d\phi|z_i) \propto f(z_i|\phi)G_0(d\phi)$, and c is the proportional constant ensuring the probabilities in Equation (4) sum to one.

Suppose the kernel for the DPM is the normal density function with a fixed mean of zero but a random variance, σ_j^2; i.e., $f(\cdot|\phi_j) \equiv f_N(\cdot|0,\sigma_j^2)$. By letting Inv-$\Gamma(m+2,\sigma_0^2(m-1))$ be the base distribution[3] to the DP prior of σ_j^2 and allowing $\alpha \to \infty$, the DPM is equivalent to the scaled t-distribution return model of Praetz (1972). The prior on σ_j^2 as described by Praetz represents the changing expectations of investors concerned with moving interest rates, random earnings, varying levels of risk, altering states of the economy, etc. Under the DPM the first term in Equation (4) is well defined and equal to the product of a Student-t density function, with $2m$ degrees of freedom and the scaling factor $\sqrt{2m/(2(m-1))}$, and a inverse Gamma density function, with shape $m+3$ and scale $\sigma_0^2(m-1)+z_i^2/2$. Given the Polya urn interpretation of the DPM prior, as $\alpha \to \infty$ there is zero probability σ_i^2 will be drawn from one of the existing $\sigma_1^2,\ldots,\sigma_{i-1}^2$. Instead, at every observation, z_i, a new σ_i^2 will be sampled from the inverse Gamma distribution, Inv-$\Gamma(m+3,\sigma_0^2(m-1)+z_i^2/2)$.

The lognormal-normal mixture model of returns by Clark (1973) has a similar DPM representation. However, Clark assumes σ_j^2 is distributed as a log-normal; i.e., in the DPM representation $G_0 \equiv \ln N$. Since the log-normal distribution is not a

[3] In the following we use $\Gamma(a,b)$ to denote a Gamma distribution and Inv-$\Gamma(c,d)$ to denote an inverse Gamma distribution.

conjugate prior to $f_N(\cdot|0, \sigma_j^2)$, the posterior predictive density, $g(z_i)$, and distribution, $G(d\phi|z_i)$, do not have an analytical form like in Praetz's model.

2.3 DPM Gibbs Sampler

Except for some pathological cases analytical expressions of ϕ's posterior expectation are not possible. Fortunately, a Markov chain of the ϕ_i's conditional posteriors can be formed and shown to converge in the limit to the posterior distribution, $\pi(\phi_1, \ldots, \phi_n|z)$. Applying the law of total probability, the prior for the ϕ_is can be written as $\pi(\phi_1, \ldots, \phi_n) = \pi(\phi_1)\pi(\phi_2|\phi_1)\ldots\pi(\phi_n|\phi_{n-1}, \ldots, \phi_1)$. Combining these conditional priors with their likelihood $f(z_i|\phi_i)$ produces the posterior distribution:

$$\pi(\phi_1, \ldots, \phi_n|z) = \pi(\phi_1)f(z_1|\phi_1)\prod_{i=2}^{n}\pi(\phi_i|\phi_{i-1}, \ldots, \phi_1)f(z_i|\phi_i).$$

Equation (4) is helpful in designing a sampler of the conditional posteriors, but a Markov chain requires the draws of the ϕ_is to be conditional on all the other ϕ_j, $j \neq i$. Fortunately, Escobar (1994) proves that since the ϕ_i's are exchangeable, in other words, because their joint probability distribution is invariant to permutation, the ϕ_i and z_i can always be treated as if they were the last observation. Applying the exchangeability property to Equation (4) and $\pi(\phi_i|\phi^{(i)}, z_i)$, where $\phi^{(i)}$ is the vector containing the elements ϕ_j, $j \neq i$, the conditional posterior distribution equals:

$$\phi_i|\phi^{(i)}, z_i \sim c\frac{\alpha}{\alpha + n - 1}g(z_i)G(d\phi|z_i) + \frac{c}{\alpha + n - 1}\sum_{j \neq i} f(z_i|\phi_j)\delta_{\phi_j}. \tag{5}$$

Draws from the posterior can then be obtained by sequentially sampling from Equation (5) for $i = 1, \ldots, n$. When G_0 is a conjugate based distribution to the likelihood $f(\cdot|\phi_j)$ sampling from Equation (5) is relatively straight forward (see Escobar & West (1995)). Otherwise, a more taxing approach is required (see MacEachern & Müller (1998) and Neal (2000) on how to handle the non-conjugate case).

Unfortunately, sampling from $\phi_i|\phi^{(i)}, z_i$ produces highly correlated draws of the ϕs. High levels of correlation in the realizations require a large number of sweeps in order to generate realizations from the entire support of the posterior distribution,

$\phi_1,\ldots,\phi_n|z$. This inefficiency in the sampler comes from the finite nature of the DP-prior. Under the DP-prior the elements $\{\phi_i\}$ will often equal one another and produce a group of ϕ_is having the same value. If the size of the group of ϕ_is having the same value is large, the element-by-element sampler of $\phi_1,\ldots,\phi_n|z$ will continually produce realizations equal in value to the existing draws. As a result, the algorithm often gets stuck sampling from the same set of ϕ_j and does not generate any new unique realizations of ϕ_i.

West et al. (1994) and MacEachern & Müller (1998) overcome this inefficiency by designing a sampling algorithm that draws from an equivalent distribution to $\phi_1,\ldots,\phi_n|z$. Let $\theta = (\theta_1,\ldots,\theta_k)'$ denote the set of distinct ϕ_i's, where $k \leq n$. Define the state vector $s = (s_1,\ldots,s_n)'$ to be configured such that $s_i = j$, when $\phi_i = \theta_j$, where $i = 1,\ldots,n$, and $j = 1,\ldots,k$. Let n_j be the number of $s_i = j$ for $i = 1,\ldots,n$. Also define $k^{(i)}$ to be the number of distinct θ_j in $\phi^{(i)}$, and $n_j^{(i)}$ to be the number of observations where $s_{i'} = j$, for $i' \neq i$. Using this notation, Equation (5) can be rewritten as:

$$\phi_i|\phi^{(i)}, z_i \sim c\alpha g(z_i)G(d\phi|z_i) + c\sum_{j=1}^{k^{(i)}} n_j^{(i)} f(z_i|\theta_j)\delta_{\theta_j}. \tag{6}$$

Draws from $\phi_1,\ldots,\phi_n|z$ are again made from the conditional distribution (either Equation (5) or (6)), however, each sweep of the sampler now consists of the following two steps:

Step 1. Draw s and k by drawing s_i for $i = 1,\ldots,n$, from Equation (6).

Step 2. Given s and k, sample θ_j, $j = 1,\ldots,k$ from:

$$\theta_j|z, s, k \propto \left[\prod_{i:s_i=j} f(z_i|\theta_j)\right] G_0(d\theta_j).$$

Step 1 is the same as in the previous DPM sampler except instead of retaining the drawn ϕ_is they are now discarded after Step 1 and only the state vector, s, and the number of clusters, k, are used in Step 2. In the context of sampling s_i, if a new

θ is sampled from $G(d\phi|z_i)$, k is increased by 1, and s_i is set equal to the new value of k. Likewise, if $n_j^{(i)} = 0$, in other words, θ_j is only observed at the ith observation, θ_j is dropped from θ and is not resampled. Instead, a new value for s_i is drawn either from one of the existing clusters, in which case k would decrease by 1, or θ_j is sampled from $G(d\phi|z_i)$ and s_i continues to equal j.

In Step 2, the ϕ_i's associated with the jth-cluster are block updated by sampling from the posterior of θ_j conditional on the observations associated with the jth cluster. Thus, instead of sampling from $\phi_1, \ldots, \phi_n|z$ element-by-element as in the sampler of Escobar & West (1995), a more efficient block sampler of drawing from $\theta_j|\theta^{(j)}, z, s, k$ is employed. This ensures that the realizations of ϕ_i will be uncorrelated and representative of a nice mixture of draws from the posterior distribution. The parameter α can also be sampled, which will add a third step to the above procedure. We allow for this in the stochastic volatility model of the next section.

After iterating on Steps 1 and 2 a number times we obtain a large collection of draws denoted as $\{\theta^{(r)}\}_{r=1}^{R}$ from the posterior. Note that for each drawn $\theta = \{\theta_1, ..., \theta_k\}$, there is an associated state vector s, and number of observations in each cluster $\{n_1, ..., n_k\}$, such that $\sum_{j=1}^{k} n_j = n$. The number of clusters k will vary from sweep to sweep, so that the size of θ will change and hence, the number of mixture orders will too. The Bayesian estimate of the DPM model's predictive density is obtained by integrating out these unknowns as in:

$$\begin{aligned} \pi(z_{n+1}|z) &= \int \pi(z_{n+1}|\theta, n_1, \ldots, n_k)\pi(\theta, n_1, \ldots, n_k, \alpha|z) d\theta dn_1 \ldots dn_k, \\ &\approx \frac{1}{R}\sum_{r=1}^{R} \pi(z_{n+1}|\theta^{(r)}, n_1^{(r)}, \ldots, n_{k^{(r)}}^{(r)}, \alpha^{(r)}), \end{aligned} \tag{7}$$

where:

$$\pi(z_{n+1}|\theta, n_1, \ldots, n_k) = \frac{\alpha}{\alpha + n} g(z_{n+1}) + \sum_{j=1}^{k} \frac{n_j}{\alpha + n} f(z_{n+1}|\theta_j), \tag{8}$$

and $g(z_{n+1}) = \int f(z_{n+1}|\phi) G_0(d\phi) d\phi$.

3 Stochastic Volatility and DPM Innovations

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

$$
\begin{aligned}
y_t &= \eta_t + \lambda_t^{-1} \exp(h_t/2)\epsilon_t, \qquad \epsilon_t \stackrel{iid}{\sim} N(0,1), && (9)\\
h_t &= \delta h_{t-1} + \sigma_v v_t, \qquad v_t \stackrel{iid}{\sim} N(0,1), \quad \epsilon_t \perp v_t, && (10)\\
\left(\begin{array}{c} \eta_t \\ \lambda_t^2 \end{array} \right) \Bigg| G &\stackrel{iid}{\sim} G, && (11)\\
G|G_0, \alpha &\sim \mathrm{DP}(G_0, \alpha), && (12)\\
G_0(\eta_t, \lambda_t^2) &\equiv N\left(m, (\tau\lambda_t^2)^{-1}\right) - \Gamma(v_0/2, s_0/2), && (13)
\end{aligned}
$$

where at time $t = 1, \ldots, n$ 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 Equation (10) with the AR-parameter δ. Identification of the SV-DPM model requires the intercepts of both y_t and h_t to equal zero with their effect subsumed into η_t and 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. By the notation $\epsilon_t \perp v_t$ found in Equation (10), we are assuming the innovations to the return and volatility process are independent of one another; i.e., there are no explicit leverage effects in the SV-DPM model (see Jacquier et al. (2004); Yu (2005); Omori et al. (2007)).[4]

In Equation (11)-(13), the SV-DPM unconditional distribution is modeled nonparametrically by an infinite ordered mixture of normals. Being a dense basis function to the entire class of continuous distributions, this mixture of normals with its different means η_t and variances λ_t^{-2} is fully flexible with regards to the type of distribution it is able to fit. Equation (11)-(12) assumes the mixture parameters η_t and λ_t^2 are distributed *a priori* as a Dirichlet process. The Dirichlet process prior for G is

[4]Including leverage effects can be done 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.

formally defined in Equation (12)-(13) by the conjugate conditional normal-gamma base distribution, G_0, and the nonnegative precision parameter α.

The SV-DPM model of Equation (9)-(13) can also be written in terms of its Sethurman representation:

$$y_t|h_t \overset{\perp}{\sim} \sum_{j=1}^{\infty} V_j f_N \left(\cdot|\eta_j, \lambda_j^{-2} \exp\{h_t\}\right), \tag{14}$$

where $f_N(\cdot|m, v^2)$ is a normal density with mean m and variance v^2, $\overset{\perp}{\sim}$ denotes a sequence of random variables that are independently distributed, $V_1 = W_1$, and $V_j = W_j \prod_{s=1}^{j-1}(1 - W_s)$ with $W_j \sim Beta(1, \alpha)$, and $(\eta_j, \lambda_j^2) \sim G_0$.

The SV-DPM is more flexible at modeling the distribution of y_t than are the existing class of parametric SV models. In the terminology of Müller & Quintana (2004), the SV-DPM model "robustifies" the class of parametric SV models. By modeling the distribution of y_t innovation 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 $j > 1$, and $\phi_t = (\eta, \lambda^2)$ for $t = 1, \ldots, n$, Equation (14) equals the 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 $\alpha \to \infty$, $\phi_t = (0, \lambda_t^2)$ and $G_0(\lambda_t^2) \equiv \Gamma(\nu/2, \nu/2)$. Geweke & 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.

4 Bayesian Inference of the SV-DPM

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. Since y_t is comprised of the two innovations, ϵ_t and

v_t, and because log-volatility h_t enters through the variance of y_t, the 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 $\psi = (\delta, \sigma_v)'$, $h = (h_1, \ldots, h_n)'$ and $y = (y_1, \ldots, y_n)'$ (see Jacquier et al. (1994); Kim et al. (1998); Chib et al. (2002); and Jensen (2004)).

In the context of the SV-DPM model for y_t and its unknown parameters $\phi = (\phi_1, \ldots, \phi_n)'$, Bayesian augmenting can be extended to include a MCMC sampler of the posterior $\pi(\psi, h, \phi|y)$. Since the likelihood function of the SV model is intractable and because we do not know the number of mixtures of the nonparametric distribution nor 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 we design for the SV-DPM consists of iteratively sampling through the following conditional distributions:

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

One full iteration through these conditional distributions denotes a sweep of the MCMC sampler.

4.1 Parameter sampler

Sampling from $\pi(\psi|y,h)$ is straight forward. We assume 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, $\delta|h, \sigma_v^2 \sim N(\widehat{\delta}, \widehat{\sigma}_v^2)I(|\delta| < 1)$, where:

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

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 $\delta|h, \sigma_v^2$ result in a realization outside the stationary set for δ, the draw is discarded and another draw is made until a value from within the parameter space is obtained.

4.2 Latent volatility sampler

Drawing latent volatilities is a difficult sampling problem that has attracted the attention of the profession (see Jacquier et al. (1994); Pitt & Shephard (1997); Kim et al. (1998); Chib et al. (2002), and Fleming & Kirby (2003)). One option for drawing the volatilities for the SV-DPM model is to apply the element-by-element volatility sampling algorithm of Jacquier et al. (1994) (JPR) and sequentially draw from $h_t|y_t, h_{t-1}, h_{t+1}, \phi_t, \psi$, $t = 1, \ldots, n$. Conditional on the mixture mean, $\eta_t = 0$, and variance, $\lambda_t^{-2} = 1$, for all t, the JPR volatility sampler for the SV-DPM model is exactly the same as the SV-N model. If η_t and λ_t^{-2} do not respectively equal 0 and 1 then the JPR volatility sampler is applied to the standardized return, $\widetilde{y}_t = (y_t - \eta_t)/\lambda_t^{-1}$, $t = 1, \ldots, n$. Given any value for η_t and λ_t^{-2}, the entire suite of existing element-by-element samplers by Geweke (1994), Pitt & Shephard (1997), Kim et al. (1998), and Jacquier et al. (2004) can be directly applied to $\widetilde{y}$.

Since each draw of h_t is conditional on the previous draw of h_{t-1} and h_{t+1}, element-by-element samplers are known to be very inefficient and require throwing away a large number of initial draws of h to ensure that the sampler is not dependent on its starting values. This dependency between the h_ts also leads to strong levels of correlation between their realizations. As a result, a larger number of sweeps must be carried out in order for the sampler to produce draws from across the support of $h|y, \psi, \phi$. This is very taxing for the SV-DPM model since each additional sweep requires sampling from $\phi|y, h$ which costs a number of computing cycles.

Ideally one would like to sample from $h|y, \psi, \phi$ in a single draw (see Kim et al. (1998); and Chib et al. (2002)). This eliminates the correlation between draws, but requires taking the log-squared transformation of y. In the context of the SV-DPM model the tangible nature of the DP prior for ϕ_t is lost under a log-square transformation of y. Thus, sampling the entire h in one draw is not feasible with the SV-DPM model. Fortunately, less correlated draws of the volatilities can be found by sampling random length blocks of volatilities instead of the entire vector (see Elerian et al. (2001) and Fleming & Kirby (2003)).

Our random length block sampler divides h into blocks of subvectors $\{h_{(t,\tau)}\}$, where $h_{(t,\tau)} = (h_t, h_{t+1}, \ldots, h_\tau)'$, $1 \leq t \leq \tau \leq n$, 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., $E[l_t] = 4$. By letting the length be random we ensure that with each sweep different subblocks of h are sampled. This helps 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:

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

does not come from a standard distribution, we design a Metropolis-Hastings (MH) sampler of the above target density where we extend the sampler of Fleming & Kirby (2003) to include the return data, y. Fleming & 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:

$$h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \sigma_v^2 \sim N\left(m_{(t,\tau)}, \Sigma_{(t,\tau)}\right), \tag{15}$$

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

$$m_{t+i} = \frac{(l_t - i)h_{t-1} + (i+1)h_{\tau+1}}{l_t + 1}, \quad i = 0, \ldots, l_t - 1, \tag{16}$$

$$\sigma^{(t)}_{i,j} = \sigma^2_v \frac{\min(i,j)(1 + l_t) - ij}{l_t + 1}, \quad i = 1, \ldots, l_t, \text{ and, } j = 1, \ldots, l_t. \tag{17}$$

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

$$\Sigma^{-1}_{(t,\tau)} = \begin{pmatrix} 2/\sigma^2_v & -1/\sigma^2_v & 0 & \ldots \\ -1/\sigma^2_v & 2/\sigma^2_v & -1/\sigma^2_v & \ddots \\ 0 & -1/\sigma^2_v & 2/\sigma^2_v & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix} \tag{18}$$

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

Since the proposal distribution in Equation (15) ignores the information found in the return vector, $y_{(t,\tau)} = (y_t, \ldots, y_\tau)'$, a better proposal distribution would be one that incorporates this 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:

$$\begin{aligned} \pi(h_{(t,\tau)}|y_{(t,\tau)}, h_{t-1}, h_{\tau+1}, \psi, \phi) &\propto f(y_{(t,\tau)}|h_{(t,\tau)}, \phi)\pi(h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi), \\ &\approx f(y_{(t,\tau)}|h_{(t,\tau)}, \phi_{(t,\tau)})\ f_N\left(h_{(t,\tau)}\left|m_{(t,\tau)}, \Sigma_{(t,\tau)}\right.\right), \end{aligned} \tag{19}$$

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

$$f(y_{(t,\tau)}|h_{(t,\tau)}, \phi_{(t,\tau)}) \propto \exp\left\{-0.5\left(\iota' h_{(t,\tau)} + \widetilde{y}^{2\prime}_{(t,\tau)} \exp\{-h_{(t,\tau)}\}\right)\right\}, \tag{20}$$

with ι being a $l_t \times 1$ vector of ones, $\widetilde{y}^2_{(t,\tau)} = (\widetilde{y}^2_t, \ldots, \widetilde{y}^2_\tau)'$, and $\exp\{-h_{(t,\tau)}\} = (\exp\{-h_t\}, \ldots, \exp\{-h_\tau\})'$. Replacing the $\exp\{-h_{(t,\tau)}\}$ vector in Equation (20) with its first-order, Taylor series approximation, $\exp\{-h_{(t,\tau)}\} \approx D_{(t,\tau)}(\iota + m_{(t,\tau)} - h_{(t,\tau)})$, where the $l_t \times l_t$ diagonal matrix $D_{(t,\tau)} = \text{diag}\{\exp(-m_{(t,\tau)})\}$, results in:

$$\exp\left\{-0.5\left(\iota' h_{(t,\tau)} + \widetilde{y}^{2\prime}_{(t,\tau)} \exp\{-h_{(t,\tau)}\}\right)\right\} \leq \exp\left\{-0.5\left(\iota' - \widetilde{y}^{2\prime}_{(t,\tau)} D_{(t,\tau)}\right) h_{(t,\tau)}\right\}. \quad (21)$$

Substituting the righthand side of Equation (21) for the $f(y_{(t,\tau)}|h_{(t,\tau)}, \phi_{(t,\tau)})$ term in Equation (19) and collecting terms in the quadratic form of $h_{(t,\tau)}$ leads to our MH sampler's fat-tailed proposal density:

$$f_{S_t}(h_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu) \propto \left[1 + (h_{(t,\tau)} - \zeta_{(t,\tau)})' \Sigma^{-1}_{(t,\tau)} (h_{(t,\tau)} - \zeta_{(t,\tau)})/\nu\right]^{-(l_t+\nu)/2} \quad (22)$$

where $f_{S_t}(h_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)$ is the density of a l_t-variate Student-t distribution with mean, $\zeta_{(t,\tau)} = m_{(t,\tau)} - 0.5\Sigma_{(t,\tau)}(\iota - D_{(t,\tau)}\widetilde{y}^2_{(t,\tau)})$, covariance, $\Sigma_{(t,\tau)}\nu/(\nu-2)$, and ν degrees of freedom (in the simulated and empirical examples of Sections 6 and 7 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 sweeps MCMC draw of $h_{(t,\tau)}$, the candidate draw, $\widehat{h}_{(t,\tau)} \sim St(\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)$, will be accepted as a realization from the target distribution with MH probability:

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

where $f(y_{(t,\tau)}|\phi_{(t,\tau)}, h_{(t,\tau)}) = \prod_{j=t}^{\tau} f_N(y_j|\eta_j, \lambda_j^{-2}\exp\{h_j\})$ and:

$$\pi(h_{(t,\tau)}|h_{t-1}, h_{\tau+1}, \psi) = \prod_{j=t}^{\tau+1} \exp\left\{-\frac{(h_j - \delta h_{j-1})^2}{2\sigma_v^2}\right\}.$$

4.3 DPM sampler

Conditional on a draw of ψ and h from $\pi(\psi, h|y, \phi)$, sampling from the posterior distribution $\phi|y, h$ is done through a variant of the sampler in Section 2.2. To describe the sampler of ϕ we rewrite Equation (9), the compound return equation, as:

$$y_t^* \;=\; \eta_t \exp\{-h_t/2\} + \lambda_t^{-1}\epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} N(0,1), \quad (23)$$

where $y_t^* \equiv y_t \exp\{-h_t/2\}$. Once again, to improve the mixing behavior of the sampled ϕs we appeal to the equivalent distribution $\theta, s|y^*$ and indirectly draw ϕ by sampling θ and s. Draws from $\theta, s|y^*$ are made with the two step procedure:

Step 1. Sample s and k by drawing sequentially ϕ_t and s_t for $t = 1, \ldots, n$ from:

$$\begin{aligned} \phi_t|y_t^*, \phi^{(t)} \quad \sim \quad & c\frac{\alpha}{\alpha+n-1} g(y_t^*)\ G(d\phi|y_t^*) \\ & + \frac{c}{\alpha+n-1} \sum_{j=1}^{k} n_j^{(t)} f(y_t^*|\theta_j)\ \delta_{\theta_j}(\phi_t). \end{aligned} \tag{24}$$

Step 2. Given the s and k from Step 1, sample θ_j, $j = 1, \ldots, k$ from:

$$\theta_j|y^*, s, k \propto \prod_{t:s_t=j} f_N\left(y_t^*|\eta_j \exp\{-h_t/2\}, \lambda_j^{-2}\right) G_0(d\theta_j). \tag{25}$$

In Step 1 the probability of s_t equaling the jth cluster is proportional to the number of other times the jth cluster occurs, $n_j^{(t)}$, times the likelihood of y_t^* belonging to the jth 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 a new cluster and k increasing by one is proportional to the predictive density:

$$\begin{aligned} g(y_t^*) \quad & \equiv \quad \int f(y_t^*|\phi)\ G_0(d\phi)\ d\phi, \\ & = \quad \int \frac{1}{\sqrt{2\pi \exp\{h_t\}\lambda^{-2}}} \exp\left\{-\frac{(y_t^* - \eta \exp\{-h_t/2\})^2}{2\lambda^{-2}}\right\} G_0(d\phi)\ d\phi, \\ & = \quad f_{St}\left(y_t^*|m \exp\{-h_t/2\}, (\exp\{h_t\}+\tau)s_0/(\tau v_0), v_0\right), \\ & = \quad f_{St}(y_t|m, (1+\tau \exp\{h_t\})s_0/(\tau v_0), v_0), \end{aligned} \tag{26}$$

where $f_{St}(.|m, s, v)$ 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, a new ϕ_t and, hence, a new θ_{k+1}, is sampled from the posterior distribution:

$$G(d\phi_t|y_t^*) \equiv f(y_t^*|\phi_t)\ G_0(d\phi_t).$$

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

$$\lambda_t^2|y_t^* \sim \Gamma(\overline{v}/2, \overline{s}_t/2), \tag{27}$$

$$\eta_t|y_t^*, \lambda_t^2 \sim N\left(\overline{\mu}_t, (\overline{\tau}_t\lambda_t^2)^{-1}\right), \tag{28}$$

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

Before moving to Step 2, the ϕ drawn in Step 1 is discarded. Step 2 then consists of generating a new draw of ϕ, conditional on the s sampled in Step 1, by sampling the unique mixture parameters, θ_j, $j = 1, \ldots, k$, from the linear regression model:

$$y_t^*|h_t, s_t, \eta_j, \lambda_j^2 \sim N(\eta_j\exp\{-h_t/2\}, \lambda_j^{-1}), \tag{29}$$

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 Equation (29) leads us to find the posterior distribution $\theta_j|y^*, h, s, k$ being the normal-gamma distribution:

$$\lambda_j^2|y^*, h, s, k \sim \Gamma(\overline{v}_j/2, \overline{s}_j/2), \tag{30}$$

$$\eta_j|y, h, s, k, \lambda_j^2 \sim N\left(\overline{\mu}_j, (\overline{\tau}_j\lambda_j^2)^{-1}\right), \tag{31}$$

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

The DPM precision parameter α is sampled using the two step algorithm of Escobar & 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., $\pi(\alpha|\phi, k, s) = \pi(\alpha|k) \propto \pi(\alpha)f(k|\alpha)$.

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 $\pi(\xi|\alpha,k) \sim Beta(\alpha+1,n)$, 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 $\pi_\xi/(1-\pi_\xi) = (a+k-1)/[n(b-\ln\xi)]$.

5 Features of the Model

After an initial burn-in phase, our MCMC algorithm for the SV-DPM model produces a set of draws, $\{\theta^{(r)}, s^{(r)}, \alpha^{(r)}, \delta^{(r)}, \sigma_v^{2(r)}, h^{(r)}\}_{r=1}^{R}$, 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.[5]

5.1 Predictive density

In a similar way various quantities of the predictive density can be estimated. The key quantity of interest in density estimation is the predictive density. Gelfand & 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:

$$f(Y_t|y) = \int f(Y_t|\theta, h_t, \alpha)\, \pi(\theta, h_t, \alpha|y)\, d\theta\, dh_t\, d\alpha, \tag{32}$$

$$\approx \frac{1}{R}\sum_{r=1}^{R} f\left(Y_t|\theta^{(r)}, h_t^{(r)}, \alpha^{(r)}\right), \tag{33}$$

[5]For a full treatment on MCMC methods see Robert & Casella (1999).

where Y_t is the random asset return at time $t = 1, \ldots, n$, $\theta^{(r)}$, $h_t^{(r)}$ and $\alpha^{(r)}$ are the rth draw from the posterior simulator[6], and the conditional posterior density is:

$$f\left(Y_t \middle| \theta^{(r)}, h_t^{(r)}, \alpha^{(r)}\right) = \frac{\alpha^{(r)}}{\alpha^{(r)}+n} f_{St}\left(Y_t \middle| m, \frac{\left(1+\tau \exp\{h_t^{(r)}\}\right) s_0}{\tau v_0}, v_0\right) + \sum_{j=1}^{k^{(r)}} \frac{n_j^{(r)}}{\alpha^{(r)}+n} f_N\left(Y_t \middle| \eta_j^{(r)}, \lambda_j^{-2(r)} \exp\{h_t^{(r)}\}\right). \quad (34)$$

From Equation (34) one can see how flexible the SV-DPM is as a semiparametric model of the returns predictive density. The SV-DPM models conditional predictive density consists of a weighted mixture of normals and Student-t densities. Thus, the predictive density is equipped to produce multiple modes, negative or positive skewness, and other non-Gaussian type behavior.

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 & West (1995), p. 580. The SV-DPM one-step-ahead predictive return density equals:

$$f(Y_{n+1}|y) = \int f(Y_{n+1}|\theta, h_{n+1}, \alpha)\, \pi(\theta, h_{n+1}, \alpha|y)\, d\theta\, dh_{n+1}\, d\alpha, \quad (35)$$

$$\approx \frac{1}{R}\sum_{r=1}^{R} f\left(Y_{n+1} \middle| \theta^{(r)}, h_{n+1}^{(r)}, \alpha^{(r)}\right), \quad (36)$$

where the conditional density:

$$f(Y_{n+1}|\theta^{(r)}, h_{n+1}^{(r)}, \alpha^{(r)}) = \frac{\alpha^{(r)}}{\alpha^{(r)}+n} f_{St}\left(Y_{n+1} \middle| m, \frac{\left(1+\tau \exp\{h_{n+1}^{(r)}\}\right) s_0}{\tau v_0}, v_0\right) + \sum_{i=1}^{k^{(r)}} \frac{n_i^{(r)}}{\alpha^{(r)}+n} f_N\left(Y_{n+1} \middle| \eta_i^{(r)}, \lambda_i^{-2(r)} \exp\{h_{n+1}^{(r)}\}\right) \quad (37)$$

has the same form as Equation (34) but with $h_{n+1}^{(r)}$ being sampled from $N\left(\delta^{(r)} h_n^{(r)}, \sigma_v^{2(r)}\right)$.

[6]To minimize notation we have omitted conditioning on $n_1, ..., n_k$ which is the number of observations in each cluster.

5.2 Conditional Moments

Using Equation (33) in-sample moments of the equity return can be computed. For instance, the first and second moments of the SV-DPM models return can be approximated as:

$$E[Y_t|y] \approx \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\alpha^{(r)}}{\alpha^{(r)}+n}m + \sum_{r=1}^{k^{(r)}}\frac{n_i^{(r)}}{\alpha^{(r)}+n}\eta_i^{(r)}\right), \tag{38}$$

$$\begin{aligned} E[Y_t^2|y] \approx\ & \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\alpha^{(r)}}{\alpha^{(r)}+n}\left[\frac{\left(1+\tau\exp\{h_t^{(r)}\}\right)s_0}{\tau(v_0-2)}+m^2\right]\right. \\ & \left. + \sum_{i=1}^{k^{(r)}}\frac{n_i^{(r)}}{\alpha^{(r)}+n}\left[\eta_i^{2(r)}+\lambda_i^{-2(r)}\exp\{h_t^{(r)}\}\right]\right), \end{aligned} \tag{39}$$

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

5.3 Label switching

Mixture models in general suffer from what is referred to as "label switching"; a short-coming where the mixture parameters are unidentified. In Equation (34), 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, $\{n_{g(j)}, \eta_{g(j)}, \lambda_{g(j)}\}_{j=1,\ldots,k}$, where $g(j)$ is the permutation function of k elements. As a result the mixture parameters of the jth 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 & 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 ϵ_t nonparametrically, label switching will not present a problem in making inferences concerning the parameters or forecasts of the stochastic volatility model.

6 Simulation Examples

In this section we consider two examples of simulated return data. In each case we estimate the SV-DPM model along with conventional parametric specifications. The first benchmark is a stochastic volatility model with normal innovations (SV-N):

$$\begin{aligned} y_t &= \mu + \exp(h_t/2) z_t, \quad z_t \sim N(0,1), \\ h_t &= \gamma + \delta h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0,1). \end{aligned} \tag{40}$$

Priors are $\mu \sim N(0, 0.1), \gamma \sim N(0, 100), \delta \sim N(0, 100) I_{|\delta|<1}$, and $\sigma_v^2 \sim \text{Inv-}\Gamma(10/2, 0.5/2)$.

The second specification is a stochastic volatility model with Student-t return innovations (SV-t):

$$\begin{aligned} y_t &= \mu + \exp(h_t/2) z_t, \quad z_t \sim St(0, (\nu-2)/\nu, \nu), \\ h_t &= \gamma + \delta h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0,1), \end{aligned} \tag{41}$$

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

The priors for the SV-DPM model 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 precision parameter prior, $\alpha \sim \Gamma(2, 8)$.

Estimation of the 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 4.2 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 1,000 sweeps of the step-by-step volatility sampler of Kim et al. (1998) is 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 model are then conducted of which we keep the last 10,000 draws for inference of the two models. Because of computation time involved in drawing the DPM parameters only 11,000 sweeps are made of the SV-DPM model with the first 1,000 draws being discarded.

6.1 Example 1

In the first example we simulate from the SV-t specification (41) with the parameters $\mu = 0, \gamma = -0.01025, \delta = 0.95, \sigma_v^2 = 0.04$, and $\nu = 6$. The first 100 simulated observations from this data generating process (DGP) are discarded and the next 1500 are collected for estimation.

Table 1 reports the posterior mean and standard deviation for the SV-N, SV-t and SV-DPM as applied to the simulated SV-t data. The SV-N is misspecified and as a result it underestimates the autoregressive parameter δ with a posterior mean of 0.9, while overestimating the volatility of volatility parameter σ_v^2 with a estimate of 0.11. The estimate for σ_v^2 is not all that surprising since the only way the SV-N can approximate the fat-tails in the distribution of z_t is to increase the variance of log-volatility, σ_v^2.

The SV-DPM, on the other hand, does a better job estimating the volatility parameters, producing estimates that are in general very close to the correctly specified SV-t. Of the three SV models, the SV-DPM estimates of δ and σ_v^2 are the closest to the truth at 0.93 and 0.05, respectively. The volatility of volatility estimates is better than the true SV-t model estimate of 0.06 for σ_v^2. Note also that the estimated k of 3.2343 reported in Table 1 suggests the SV-DPM is using, on average, a mixture of normals consisting of three clusters to approximate the return innovations Student-t distribution.

Because the SV-DPM model must set the mean of the latent volatility process to equal to zero, a comparison between the three SV models estimate of $E[h_t|y]$ cannot be done. What can be done, however, is to compute each models posterior conditional return variance, $\text{Var}[Y_t|y]$, and compare it with the DGP true conditional variance, $\exp\{h_t\}$. For each of the three SV models, we calculate the root mean squared error, $\text{RMSE} = \sqrt{(1/1500)\sum_{t=1}^{1500}(\text{Var}(Y_t|y) - \exp\{h_t\})^2}$, where $\text{Var}(Y_t|y)$ for the SV-DPM model is the full sample model estimate computed with Equation (38)-(39), and is equal to $\text{Var}(Y_t|y) = R^{-1}\sum_{r=1}^{R}\exp\{h_t^{(r)}\} - [R^{-1}\sum_{r=1}^{R}\mu^{(r)}]^2$ for the SV-N model and $\text{Var}(Y_t|y) = R^{-1}\sum_{r=1}^{R}\exp\{h_t^{(r)}\}\nu^{(r)}/(\nu^{(r)} - 2) - [R^{-1}\sum_{r=1}^{R}\mu^{(r)}]^2$ for the SV-t model.

The last row in the Table 1 displays the RMSE for each of the three models. The

RMSE of 0.56 for the SV-DPM is nearly indistinguishable from the SV-t model's RMSE of 0.57 and is much smaller than the SV-N model's RMSE of 0.64. Figure 1 displays a time-series plot of the $\text{Var}(Y_t|y)$ for the SV-DPM and SV-t model. Except for the couple of episodes where the DGP conditional variance dramatically increases and the SV-t models $\text{Var}(Y_t|y)$ noticeably exceeds the SV-DPM models, the two models estimates of $\exp\{h_t\}$ have very similar behavior. Although we only use the DPM to model the distribution of the return innovations, the RMSE show how important modeling the return distribution is in estimating volatility. Not having fat enough tails in the return process causes the SV-N model to try and compensate for it with more extreme levels of volatility.

The estimated log-predictive densities, one period out-of-sample, $\ln f(Y_{T+1}|y)$, for each of the three SV models are displayed in Figure 2. From this figure one is able to conclude that the tails for the SV-N model are too thin relative the SV-t while the SV-DPM tails are very close to those obtained from the true SV-t model.

6.2 Example 2

In this example the DGP for the SV model's volatility process is the same as in the previous example, but now the innovation distribution for the asset return, z_t, is replaced by a mixture of two normals:

$$z_t \sim \begin{cases} N(\mu_1, \sigma_1^2) \text{ with probability } p, \\ N(\mu_2, \sigma_2^2) \text{ with probability } 1-p. \end{cases} \tag{42}$$

Setting $\mu_1 = -1.3791, \mu_2 = 0.3448$, $\sigma_1^2 = 1.3112, \sigma_2^2 = 0.3278$, $p = 0.2$, implies a mean, variance, skewness and kurtosis of 0, 1, -1.3056, and 5.2042, respectively, for the return process. A plot of the mixture's density function is provided in Figure 3, which illustrate the negative skewness and the fat lefthand and thin righthand tails of the conditional return distribution. In this simulation example both the SV-N and SV-t are misspecified and cannot accommodate the asymmetric distribution of z_t.

Table 2 reports the posterior parameter estimates. The SV-N estimates are adversely affected by the misspecification while the SV-t estimates are somewhat more robust to the second order mixture of normals. The SV-N estimate of 0.8 for σ_v^2 is twenty times larger than its true value of 0.04. By assigning so much of the variability in volatility to its variance instead of the dynamics of volatility, the SV-N model fails to fit the highly persistent behavior of the simulated volatility data. As a result, the SV-N model severely underestimates the AR parameter with a point estimate of 0.5 and an imprecise posterior distribution whose standard deviation is 0.1.

The SV-t model utilizes its degrees of freedom parameter, ν, to approximate some of the asymmetry seen in the density of z. However, the SV-t smaller 3.996 estimate of ν implies very fat symmetrical tails. Such symmetry in the distribution of returns is inconsistent with the skewness of the true distribution.

The SV-DPM produces reasonable estimates of the autoregressive parameter and is the only model to accurately estimates σ_v^2. Volatility of volatility is found by the SV-DPM model to be equal to 0.05; a point estimate very similar to that reported in Example 1. Volatility's persistence, however, is slightly less than the true value of δ, equalling 0.89 as opposed to 0.95.

The RMSE for the estimated posterior variances further illustrates the problems faced with the SV-N and the better performance available with the SV-t and SV-DPM. Again the SV-N over compensates for its inability to fit the thick tails, and in this case skewness, of the return innovations distribution by attributing this behavior to large fluctuations in volatility. Whereas both the SV-t and SV-DPM models RMSE are approximately equal at 0.6, the RMSE of the SV-N model equals 0.9, nearly seventy percent larger than the other two models.

The three models predictive, one period out-of-sample, densities in Figure 4 shows how flexible the SV-DPM model is in capturing both the negative skewness and the fat-tail behavior of the return distribution. Neither the SV-N, nor SV-t, with their symmetric distribution, is able to simultaneously fit these two dominate characteristics of the distribution. As a result forecasts from either the SV-N and SV-t models produce too many positive returns, while not generating enough negative ones.

7 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 model and the MCMC sampler developed in Section 4 to 6815 compounded daily returns from the Center for Research in Security Prices (CRSP) value-weighted portfolio index over the trading days January 2, 1980 to December 29, 2006. Figure 5 plots the percentage returns (the return series multiplied by 100). CRSP portfolio returns average 0.0529 during this time period with a variance of 0.9225. Non-Gaussian behavior is seen in the return processes 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 the SV-N and SV-t models to them. Priors for the three SV models are the same as those used in Section 6 for the simulated data examples. Like the simulated return data, we first sweep over the latent volatilities 1000 times with the step-by-step algorithm of Kim et al. (1998) before applying the full MCMC sampler. In each of these initial 1000 sweeps the unknown volatilities are drawn conditional on a normal return distribution and a fixed parameter vector equal to its starting value.

We increase the efficiency of the SV-DPM sampler and reduce the samplers total computing time by respectively taking every tenth draw while running three independent chains (consisting of 110,000 sweeps each) of the SV-DPM model's sampler simultaneously. To reduce the samplers dependency on the starting parameters and the initialization of the volatilities, the first 1000 thinned draws of each chain are discarded, leaving a total of 30,000 thinned draws for inference (10,000 from each chain). Independence between the chains is ensured by using a different random number generator for each chain. The three random number generators are the maximally equidistributed combined Tausworthe generator by L'Ecuyer (1999), a variant of the twisted generalized feedback shift-register algorithm known as the Mersenne Twister generator by Matsumoto & Nishimura (1998), and a lagged-fibonacci generator by Ziff (1998). More over, a different set of starting values is used with each chain. One chain is initialized at $\delta = 0.9$, $\sigma_v^2 = 0.05$ and $h = 0$, another with

$\delta = 0.95$, $\sigma_v^2 = 0.02$ and $h = \ln y^2$, and lastly, $\delta = 0.1$, $\sigma_v^2 = 0.01$ and $h = 1/(1-\delta)$.

Table 3 reports the MCMC sample means and standard deviations for the parameters of the SV-DPM, SV-t, and SV-N models. Similar to the estimates found for the simulated return data in Section 6, 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 0.0103 with a standard deviation of 0.0018. 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 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 σ_v^2 in the SV-t model is smaller in value than in the SV-N model, but slightly larger than the SV-DPM, with a mean and standard deviation of 0.0154 and 0.0023. In Fig. 6 the posterior densities of σ_v^2 are consistent with these observations. Notice the upper tail of the SV-DPM model's density for σ_v^2 barely overlaps with the lower tail of the SV-N model's density, whereas there is considerable overlap with the lower tail of the SV-t model.

Dynamic behavior in volatility as captured by the AR-parameter δ is nearly indistinguishable between the three SV models. First-order dynamics in the volatility of the SV-DPM model is precisely estimated at 0.9887 with the tight posterior standard deviation of 0.0026. This estimate of δ is only slightly smaller than the SV-t estimate of 0.9878, but with the same posterior standard deviation. The volatility in the SV-N model reverts to its mean at a slightly faster pace with a posterior estimate of δ equal to 0.9795.

For the daily portfolio return the average SV-DPM mixture order is $k = 7.16$. This is noticeably larger than the average three clusters found for the simulated SV-t model data series.[7] A larger k suggests that the SV-DPM not only captures the daily stock returns leptokurtotic behavior, but its skewness too. Because of the SV-N models 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 volatile is highest.

This increase in the volatility of the SV-N and SV-t model relative to the SV-DPM model is apparent in Figure 7 where the SV-DPM posterior conditional variance of returns is plotted in Panel (a) and the SV-DPM models difference from the conditional variances of the SV-N model are graphed in Panel (b) and the SV-t model in Panel (c). During those periods where the SV-DPM models conditional daily variance is greater than 2, the SV-N conditional variance is on the order of 2 to 14 points larger. The conditional variances of the SV-t model, while still greater than the SV-DPM model, only range from approximately 1 to 4 points larger than the SV-DPM variances. Even though there were differences found in the conditional variances of Figure 1 for the three SV models estimate of the simulated data of Example 1, those differences are small compared with those of Figure 7.

As for the behavior of skewness, because of their symmetrical distribution neither the SV-N nor SV-t model is able to capture the skewness of daily returns. This is borne out in the one day ahead, out of sample, predictive density plots of Figure 8. 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. In addition, the log-predictive densities plots of Figure 9 shows the SV-DPM producing fatter tails than either of the SV-N or SV-t model.

7.1 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 $G_0 \equiv$

[7] Note that this comparison is affected by sample size as k is increasing in the number of observations (Antoniak (1974)).

$N(m,(\tau\lambda_t^2)^{-1})-\Gamma(v_0/2,s_0/2)$ to test the robustness of the posterior estimates of the SV-DPM model to different priors. Table 4 reports these robustness findings for the posterior estimates of the SV-DPM model for the different priors.

To determine the impact 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 (to review the complete list of of priors please see Section 6). These hyperparameter values cause the prior distribution for α to be more tightly distributed and centered closer to zero than did the original prior. As a result the posterior estimate of α is found to be closer to zero at 0.1217. 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 4.4465. 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. The only difference being the estimate of σ_v^2 is slightly larger at 0.0112 with a standard deviation of 0.0019.

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 tests for the robustness to changes in the prior of the mixture variance, λ^2. In the posterior results reported in Table 4 neither of the changes in the hyperparameters to η nor λ^2 base distribution 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 0.978, and the estimates of σ_v^2 are equal out to the second decimal place at

0.01. Subtle differences between the estimates of α can be found under the different priors, with the posterior estimates α ranging from 0.4730 under Prior 4 to 0.4881 for the original prior. Similar results are found for k, where Prior 4 produces an estimate of $k = 6.9221$, while $k = 7.1644$ for Prior 1.

7.2 Robustness to number of draws

Because the DPM sampler is a step-by-step algorithm, making 30,000 thinned draws from the SV-DPM model requires a considerable number of computing cycles. This is understandable given the level of inefficiency associated with the posterior draws of the SV-DPM model. It would, however, be preferable if a fewer number of draws could be used in making inference concerning the SV-DPM model. To determine if this is possible, the SV-DPM model for the CRSP portfolio return data is reestimated 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 5. The table also includes the results from Table 3 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.

8 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 prior* 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 simulated and real world daily return data. The semiparametric stochastic volatility model performed well on both the simulated and 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 models increased flexibility and robustness to non-Gaussian behavior and its superior forecasts makes it an appealing specification for risk and portfolio managers.

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 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.

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. (1994), 'Estimating normal means with a dirichlet process prior', *Journal of the American Statistical Association* **89**(425), 268–277.

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.

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 (1994), 'Comment on jacquier, polson and rossi's "bayesian analysis of stochastic volatility models"', *Journal of Business & Economic Statistics* **12**, 389–392.

Geweke, J. & Keane, M. (2007), 'Smoothly mixing regressions', *Journal of Econometrics* **138**, 252–290.

Ghosal, S., Ghosh, J. & Ramamoorth, 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. 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.

Kacperczyk, M., Damien, P. & Walker, S. (2005), A new class of Bayesian semi-parametric 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.

L'Ecuyer, P. (1999), 'Tables of maximally equidistributed combined lfsr generators', *Mathematics of Computation* **68**(225), 261269.

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.

Mahieu, R. J. & Schotman, P. C. (1998), 'An empirical application of stochastic volatility models', *Journal of Applied Econometrics* **13**(4), 333–360.

Matsumoto, M. & Nishimura, T. (1998), 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator', *ACM Transactions on Modeling and Computer Simulation* **8**(1), 330.

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.

Neal, R. (2000), 'Markov chain sampling methods for dirichlet process mixture models', *Journal of Computational and Graphical Statistics* **9**, 249–265.

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* **B,59**, 731–92.

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.

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* P. R. Freeman & A. F. Smith, eds, 'Aspects of Uncertainty', John Wiley.

Yu, J. (2005), 'On leverage in a stochastic volatility model', *Journal of Econometrics* **127**(2), 165–178.

Ziff, R. M. (1998), 'Four-tap shift-register-sequence random-number generators', *Computers in Physics* **12**(4), 385392.

Table 1: Posterior estimates using the simulated return data of Example 1 where the date is generated from a SV-t model ($n = 1500$).

	true	SV-DPM		SV-t		SV-N	
		mean	stdev	mean	stdev	mean	stdev
μ	0.0			0.0153	0.0211	0.0151	0.0213
γ	-0.01025			-0.0343	0.0123	-0.0310	0.0133
δ	0.95	0.9296	0.0217	0.9252	0.0206	0.9007	0.0262
σ_v^2	0.04	0.0548	0.0218	0.0648	0.0214	0.1108	0.0314
ν	6.0			13.0150	10.0809		
α		0.2789	0.1734				
k		3.3243	1.6373				
RMSE (Variance)		0.5607		0.5715		0.6364	

$$\begin{aligned}
\text{SV-DPM: } y_t|\phi_t, h_t &\sim N(\eta_t, \lambda_t^{-2}\exp(h_t)),\ \phi_t|G \sim G,\ G|\alpha, G_0 \sim DP(G_0, \alpha) \\
h_t &= \delta h_{t-1} + \sigma_v v_t,\ v_t \sim N(0,1) \\
\text{SV-t: } y_t &= \mu + \exp(h_t/2) z_t,\ h_t = \gamma + \delta h_{t-1} + \sigma_v v_t,\ \ z_t \sim t_\nu(0,1), v_t \sim N(0,1) \\
\text{SV-N: } y_t &= \mu + \exp(h_t/2) z_t,\ h_t = \gamma + \delta h_{t-1} + \sigma_v v_t,\ \ z_t \sim N(0,1), v_t \sim N(0,1)
\end{aligned}$$

Table 2: Posterior estimates using the simulated return data of Example 2 where the date is generated from a SV model whose distribution is a second ordered mixture of normals ($n = 1500$).

	true	SV-DPM		SV-t		SV-N	
		mean	stdev	mean	stdev	mean	stdev
μ	0.0			0.1493	0.0223	0.1462	0.0235
γ	-0.020			-0.0584	0.0208	-0.1756	0.0576
δ	0.95	0.8864	0.0345	0.9076	0.0284	0.5049	0.1016
σ_v^2	0.04	0.0455	0.0189	0.0744	0.0281	0.8131	0.1706
ν				3.9959	0.5764		
α		0.2789	0.1623				
k		3.2781	1.3300				
RMSE (Variance)		0.5745		0.5822		0.9064	

$$
\begin{aligned}
\text{SV-DPM: } y_t|\phi_t, h_t &\sim N(\eta_t, \lambda_t^{-2}\exp(h_t)),\ \phi_t|G \sim G,\ G|\alpha, G_0 \sim DP(G_0, \alpha) \\
h_t &= \delta h_{t-1} + \sigma_v v_t,\ v_t \sim N(0,1) \\
\text{SV-t: } y_t &= \mu + \exp(h_t/2) z_t,\ h_t = \gamma + \delta h_{t-1} + \sigma_v v_t,\ \ z_t \sim t_\nu(0,1), v_t \sim N(0,1) \\
\text{SV-N: } y_t &= \mu + \exp(h_t/2) z_t,\ h_t = \gamma + \delta h_{t-1} + \sigma_v v_t,\ \ z_t \sim N(0,1), v_t \sim N(0,1)
\end{aligned}
$$

Table 3: 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 from three independent chains of the SV-DPM sampling algorithm where every tenth draw is retained and the first 1,000 thinned draws from each chain are discarded).

	SV-DPM			SV-t		SV-N	
	mean	stdev	ineff	mean	stdev	mean	stdev
μ				0.0786	0.0084	0.0793	0.0086
γ				-0.0087	0.0023	-0.0106	0.0028
δ	0.9877	0.0026	10.625	0.9878	0.0026	0.9795	0.0037
σ_v^2	0.0103	0.0018	72.288	0.0154	0.0023	0.0276	0.0040
ν				9.9149	1.3035		
α	0.4881	0.2357	28.474				
k	7.1644	2.5996	57.765				

ineff is the inefficiency factor.

$$\begin{aligned}
\text{SV-DPM: } y_t|\phi_t, h_t &\sim N(\eta_t, \lambda_t^{-2}\exp(h_t)),\ \phi_t|G \sim G,\ G|\alpha, G_0 \sim DP(G_0,\alpha) \\
h_t &= \delta h_{t-1} + \sigma_v v_t,\ v_t \sim N(0,1) \\
\text{SV-t: } y_t &= \mu + \exp(h_t/2)z_t,\ h_t = \gamma + \delta h_{t-1} + \sigma_v v_t,\ \ z_t \sim t_\nu(0,1), v_t \sim N(0,1) \\
\text{SV-N: } y_t &= \mu + \exp(h_t/2)z_t,\ h_t = \gamma + \delta h_{t-1} + \sigma_v v_t,\ \ z_t \sim N(0,1), v_t \sim N(0,1)
\end{aligned}$$

Table 4: 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 from three independent chains of the SV-DPM sampling algorithm where every tenth draw is retained and the first 1,000 thinned draws from each chain are discarded).

	Prior 2	Prior 3	Prior 4	Prior 5	Prior 6
δ	0.9877	0.9879	0.9877	0.9878	0.9876
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0027)
σ_v^2	0.0112	0.0103	0.0104	0.0115	0.0100
	(0.0019)	(0.0017)	(0.0018)	(0.0019)	(0.0023)
α	0.1217	0.4733	0.4730	0.4827	0.4837
	(0.0080)	(0.2300)	(0.2278)	(0.2253)	(0.2490)
k	4.4465	6.9364	6.9221	7.0739	7.100
	(1.3456)	(2.4933)	(2.4716)	(2.3095)	(2.9155)

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

$$
\begin{aligned}
\text{SV-DPM: } y_t|\phi_t, h_t &\sim N(\eta_t, \lambda_t^{-2}\exp(h_t)),\ \phi_t|G \sim G,\ G|\alpha, G_0 \sim DP(G_0, \alpha) \\
h_t &= \delta h_{t-1} + \sigma_v v_t,\ v_t \sim N(0,1)
\end{aligned}
$$

Table 5: 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 1,000 thinned draws are discarded.

T	30,000			10,000		
	mean	stdev	ineff	mean	stdev	ineff
δ	0.9877	0.0026	10.625	0.9878	0.0026	15.538
σ_v^2	0.0103	0.0018	72.288	0.0102	0.0018	65.403
α	0.4881	0.2357	28.474	0.4961	0.2418	39.304
k	7.1644	2.5996	57.765	7.3002	2.7332	78.165

ineff is the inefficiency factor.

$$\begin{aligned}\text{SV-DPM: } y_t|\phi_t, h_t &\sim N(\eta_t, \lambda_t^{-2}\exp(h_t)),\ \phi_t|G \sim G,\ G|\alpha, G_0 \sim DP(G_0, \alpha)\\ h_t &= \delta h_{t-1} + \sigma_v v_t,\ v_t \sim N(0,1)\end{aligned}$$

Figure 1: Posterior variance of returns, $\mathrm{Var}[Y_t|y]$, for the simulated SV-t return data of Example 1.

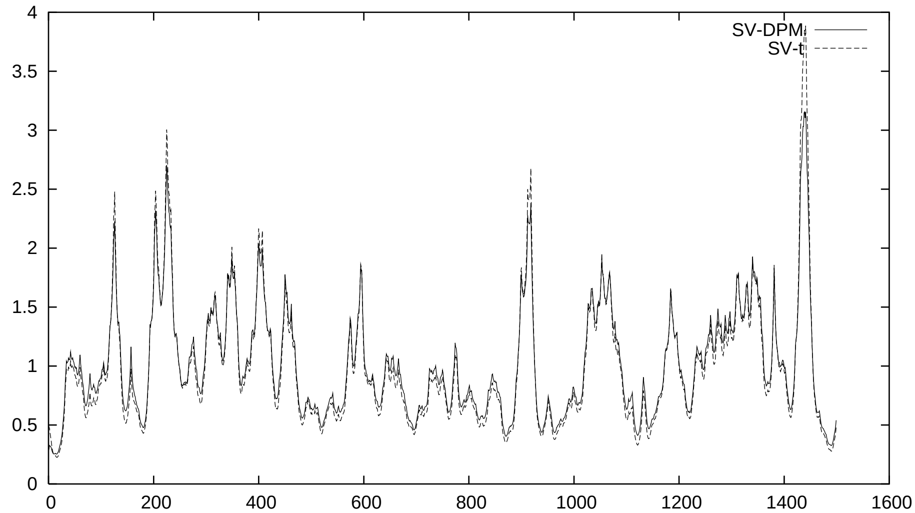

Figure 2: Log-predictive densities, $\ln f(Y_{n+1}|y)$, for simulated SV-t return data of Example 1.

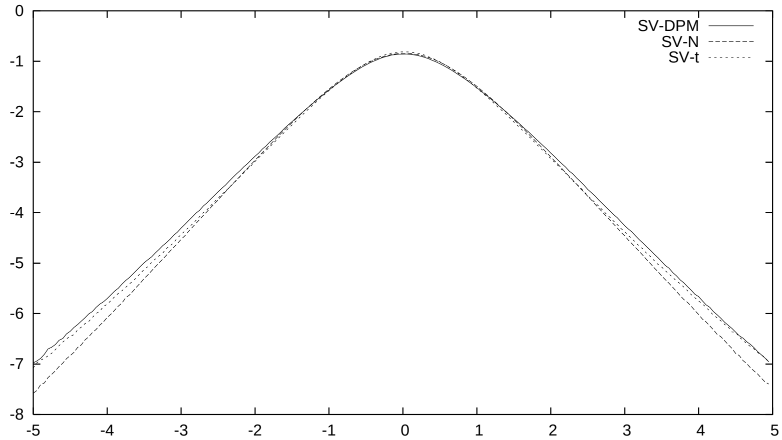

Figure 3: Density of the second order mixture of normal distributions, $f(z_t) \equiv 0.2f_N(z_t| -1.3791, 1.3112) + 0.8f_N(z_t|0.3448, 0.3278)$, used in simulating the return data of Example 2.

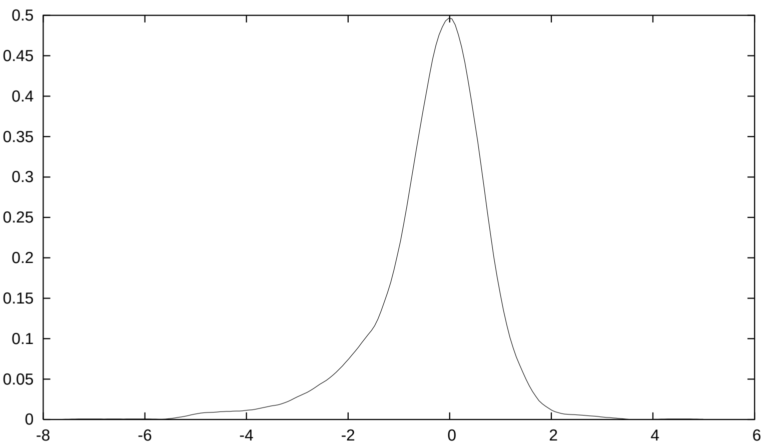

Figure 4: Predictive density, $f(Y_{n+1}|y)$, of the estimated SV-DPM, SV-N, and SV-t model for the simulated data of Example 2.

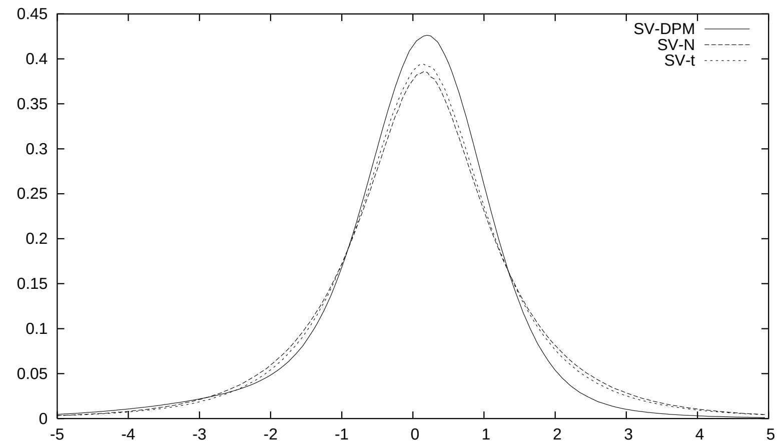

Figure 5: CRSP value-weighted portfolio index returns from Jan. 2, 1980 - Dec. 29, 2006 ($n = 6815$).

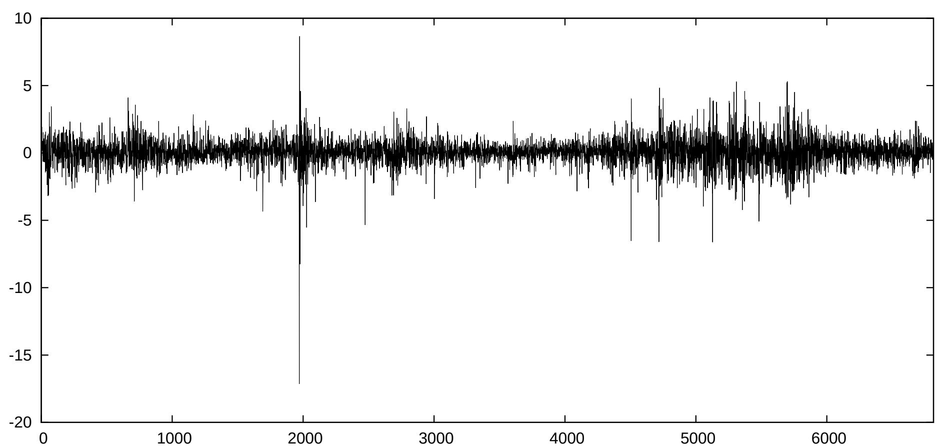

Figure 6: Posterior density of σ_v^2 for the SV-DPM (solid line), SV-t (dashed-dot line), and SV-N (dashed line) model as applied to the value-weighted CRSP portfolio daily return data.

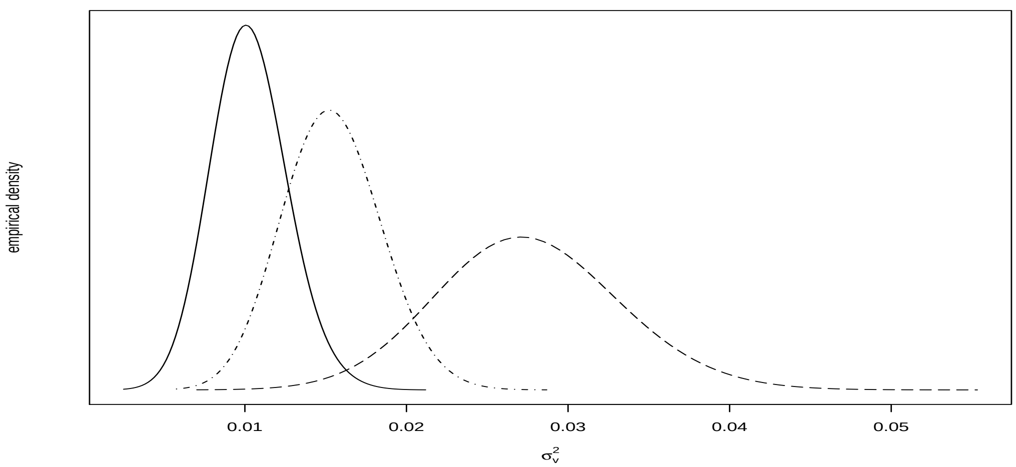

Figure 7: The SV-DPM posterior variance of returns, $\text{Var}[Y_t|y]$, for the value-weighted CRSP index returns (Panel a), and its difference from the SV-N (Panel b) and SV-t (Panel c) model.

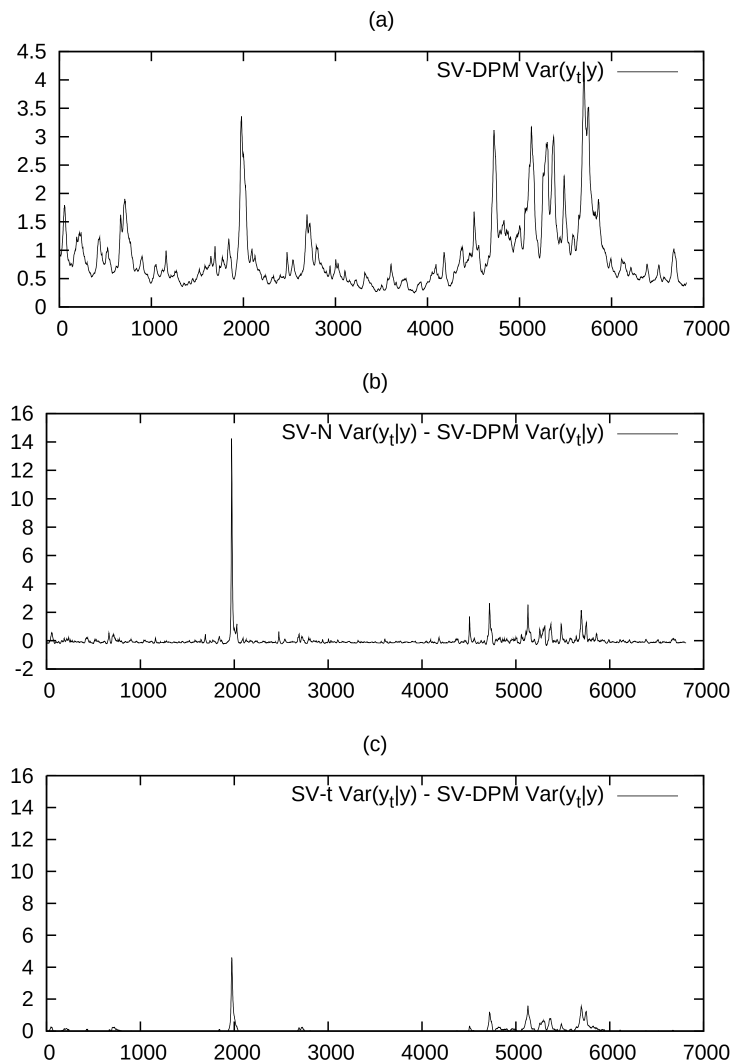

Figure 8: Predictive density, $f(Y_{n+1}|y)$, of the SV-DPM, SV-N, and SV-t model for the value-weighted CRSP portfolio daily return.

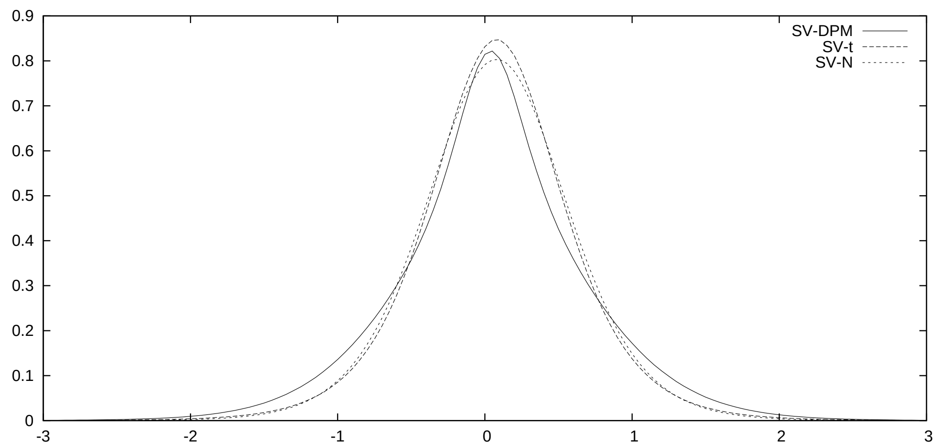

Figure 9: Log-predictive density, $\ln f(Y_{n+1}|y)$, of the SV-DPM, SV-N, and SV-t model for the value-weighted CRSP portfolio daily return.