Bayesian semiparametric stochastic volatility modeling

Mark J. Jensen* and John M. Maheu†

preliminary
September 2007

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 innovations, nonparametric Bayesian methods are used to flexibly model the distribution's skewness and kurtosis while volatility dynamics follow a parametric structure. Our Bayesian approach provides a full characterization of uncertainty. A Markov chain Monte Carlo sampling approach to estimation is presented with theoretical and computational issues for simulation from the posterior and predictive distributions. The new models are assessed based on simulation evidence, an empirical example, and comparison to parametric models.

*Federal Reserve Bank of Atlanta, Mark.Jensen@atl.frb.org
†Department of Economics, University of Toronto, and RCEA, jmaheu@chass.utoronto.ca.

1 Introduction

In this paper we propose a semiparametric stochastic volatility (SV) model that draws from the existing financial econometric literature on stochastic volatility models of returns with time-varying conditional variance, along with the recent developments in nonparametric Bayesian modeling and sampling. By bringing together these different approaches to modeling the distribution of returns, we produce an estimable stochastic volatility model whose return innovation distribution is modeled nonparametrically. The nonparametric distribution consists of a mixture of normals with parameters distributed according to a particularly flexible nonparametric Bayesian prior. The predictive density of our model is capable of producing the high levels of kurtosis and negative skewness observed in returns but not captured by current parametric SV models.

There exists a long history in financial economics where observed returns are modeled as realization from a mixture of normals (see Press (1967); Praetz (1972); Clark (1973); Gonedes (1974); Kon (1984)). The general makeup of these mixture models consists of an infinite order of mean zero normal distributions with variances that are independently and identically distributed (*iid*) over some pre-specified distribution. It is well known that mixture models produce a wide variety of fat-tailed behavior, in other words, levels of kurtosis in excess of normality. Given the leptokurtotic behavior of asset returns the fat-tailed behavior of a mixture model makes it an appealing specification for returns. However, mixture models alone are not capable of capturing the strong persistence in the conditional variance of returns.

Stochastic volatility models were designed with the time-varying behavior of returns in mind (see Taylor (1986); Harvey et al. (1994)). Like its mixture predecessors, stochastic volatility is a continuous mixture of normals. However, unlike earlier mixture models the variances of the mixture components are not distributed *iid*, instead, they follow a dynamic stochastic process. This allows the SV model to produce both the higher levels of kurtosis in the return density and persistence in time-varying conditional variances. Unfortunately, these properties do not enable the SV model to capture the asymmetries and leptokutotic behavior present in return data. The current class of SV models do not produce sufficient kurtosis and asymmetry (see Gallant et al. (1997); Mahieu & Schotman (1998), and Durham (2006)).

SV models with a parametric return distribution are thus restrictive and limited in their ability to characterize the distribution of returns. Even as a fat-tailed, time-varying mixture model, the parametric SV model is not flexible enough to model the skewness in

returns. For instance, modeling the return innovation of the SV model with a standard Gaussian distribution results in a return distribution that is symmetrical and centered at zero. Features of the return distribution such as asymmetry and different tail thickness are important in derivative pricing, risk measurement and portfolio choice. Therefore, it is important to have a flexible nonparametric distribution for return innovations within the class of SV models.

The Dirichlet process prior of Ferguson (1973) applied to an infinite order mixture model is called a Dirichlet process mixture (DPM) model and is the basis of Bayesian nonparametrics for continuous distributions (Escobar & West (1995)). The DPM possesses a number of attractive features as a Bayesian nonparametric estimator; 1) as a prior to a infinite order mixture model the DPM is more flexible and realistic than a nonparametric mixture model with a predetermined number of components, 2) as a conjugate prior the DPM is easy to use and facilitates Gibbs sampling, 3) with the DPM model the data determines the correct number of mixture clusters, 4) we can impose parsimony through the prior and 5) it works well in practice. Examples of the DPM model in economics include Chib et al. (1988), Chib & Hamilton (2002), Griffin & Steel (2004), Hirano (2002), Kacperczyk et al. (2005), and Jensen (2004)[1].

The idea set forth in this paper is to create a semiparametric stochastic volatility model by combining a DPM nonparametric model for the distribution of the return innovations with a latent autoregressive process governing the return's conditional variance. We construct a Markov chain Monte Carlo (MCMC) sampler to estimate the unknown parameters of our Bayesian semiparametric SV model. Most of the convenient Gibbs sampling for the basic DPM carry over to our model. Our MCMC approach draws on existing Bayesian nonparametric methods for *iid* data (West et al. (1994) MacEachern & Müller (1998)) and parametric stochastic volatility estimation (Eraker et al. (2003), Jacquier et al. (1994, 2004), Kim et al. (1998), Chib et al. (2002)). Due to the independence between the volatility process and the Dirichlet process mixture model of the return innovations, we provide a tractable efficient method of posterior simulation. Our estimation approach provides smoothed estimates of the volatility process, and the predictive density for returns fully accounts for uncertainty from the unobserved volatility process as well as the unknown return innovation density.

We evaluate our semiparametric SV model against a stochastic volatility specification with normal return innovations and one with Student-t return innovations, both

[1]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.

standard specifications in the literature. 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 are misspecified while the semiparametric model consistently performs well. In an empirical application with daily CRSP return data the predictive distribution from the semiparametric model is very different from the parametric SV models. Our semiparametric SV model's predictive density displays asymmetry and fatter tails whereas neither the SV model with normal nor Student-t innovations do.

The paper is organized as follows. The next section introduces basic concepts in 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 concludes and Section 9 contains a brief account of some posterior simulation methods used in this paper.

2 Bayesian Nonparametric

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 $(\Phi, \mathcal{F})$ with the unknown probability distribution function F. As a random function, F represents the unknown "parameter" in a nonparametric model of the z's distribution. As with all Bayesian estimators, a Bayesian approach to nonparametrically estimating F requires placing a prior distribution on F. 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 a prior that will lead to an analytically manageable posterior distribution, Ferguson (1973) derived the Dirchlet 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 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 above

definition for the $DP(G_0, \alpha)$ prior, $F(\Phi_0)$ is distributed as a $Beta(\alpha G_0(\Phi_0), \alpha G_0(\Phi_1))$ distribution (the Beta distribution is a special case of the Dirichlet distribution when $J = 2$). It follows from F being distributed as a Beta distribution and G_0 being a distribution that $E[F(\Phi_0)] = G_0(\Phi_0)$ and $\text{Var}[F(\Phi_0)] = G_0(\Phi_0)G_0(\Phi_1)/(\alpha + 1) = G_0(\Phi_0)(1 - G_0(\Phi_0))/(\alpha + 1)$. The first moment of F reveals that the $DP(G_0, \alpha)$ prior is well centered around G_0. From the variance of F, the precision parameter, α, can be understood as a measure of one's belief as to how well G_0 represents F. Since α is in the denominator of $\text{Var}[F]$, a larger α leads to a prior with a smaller variance.

In our example, the conjugacy property of the Beta distribution with the binomial likelihood function 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 $z = (z_1, \ldots, z_n)'$, and $\delta_{z_i}(\cdot)$ 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 equals:

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

$E[F(\Phi_0)|z]$ 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. From 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, 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.

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 $F|z \sim DP(G_0^*, \alpha^*)$, where $G_0^* = \frac{\alpha}{\alpha^*} G_0 + \frac{1}{\alpha^*} \sum_{i=1}^{n} \delta_{z_i}$ and $\alpha^* = \alpha + n$. As a conjugate prior 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 also 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 help show why the Dirichlet process 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 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 in Section 2.1 (Lo (1984)). With the DPM 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 number of clusters).

Suppose $f(\cdot|\phi_j)$ is the normal density function with $\phi_j = (\eta_j, \lambda_j^{-2})$, where 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 mixture model's infinite number of unknowns, $\{V_j, \phi_j\}_{j=1,\ldots}$, are not identified given the finite number of observations in the dataset 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 V_j and ϕ_j. Since the distribution F is modeled as $z_i|\phi_i \overset{iid}{\sim} f(\cdot|\phi_i)$ 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 can be applied to ϕ. The probability of ϕ_i conditional on the fixed values $\phi_1, \ldots, \phi_{i-1}$ equals:

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

¿From the construct of Equation (3) $\phi_i|\phi_1, \ldots, \phi_{i-1}$ also follows a Polya urn scheme. Also notice 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. At the other extreme, when $\alpha \to \infty$, F is modeled by a mixture prior consisting of distinct clusters where each cluster's distribution has a unique parameter equal to a realization from $\phi_j \sim G_0$.

Except for some pathological cases analytical expressions of the ϕ's posterior expectation are not possible. Fortunately, a Markov chain of the ϕ_i's conditional posteriors exist that 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 distributions with their likelihood $f(z_i|\phi_i)$ produces the posterior distribution:

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

Using the definition of $\pi(\phi_i|\phi_{i-1}, \ldots, \phi_1)$ in Equation (3) the marginal posteriors equal:

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

where $h_i(z_i) = \int f(z_i|\phi) G_0(d\phi)$ 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 a proportional constant that ensures the probabilities sum to one.

Equation (4) shows how the DPM nests the Praetz (1972) mixture model of the distribution of returns. In the Praetz model, the DPM's $f(z_i|\phi_i)$ equals a normal density function with mean zero and random variances, σ_i^2; i.e., $\phi_i = (0, \sigma_i^2)$. The hyperparameters to the DP-prior are $\alpha \to \infty$ and the base distribution equals a Inv-$\Gamma(m+2, \sigma_0^2(m-1))$ conjugate prior distribution for σ_i^2.[2] Under this DP-prior, the first right-hand term in

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

Equation (4) is well defined with $h_i(z_i)$ equal to the Student-t probability density function with $2m$ degrees of freedom and the scaling factor $\sqrt{2m/(2(m-1))}$. Given the conjugate nature of G_0, $G(d\phi|z_i)$ equals the inverse-Gamma distribution with shape, $m+3$, and scale, $\sigma_0^2(m-1)+z_i^2/2$. Since $\alpha \to \infty$, the second term on the right-hand side of Equation (4) converges to zero. As a result $\sigma_i^2|z_i$ will be distributed as the inverse-Gamma distribution $G(d\phi|z_i)$ and not as point masses at the existing variances, σ_j^2, $j = 1, \ldots, i-1$.

The lognormal-normal mixture model of Clark (1973) has a similar DPM representation as the Praetz model. However, because Clark assumes σ_i^2 is log-normally distributed (a non-conjugate G_0 prior to the normal distribution, $f(z_i|\phi_i)$), neither h_i nor $G(\phi|z_i)$ will have a analytical distribution.

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

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

where $\phi^{(i)}$ is the vector containing the elements ϕ_j, $j \neq i$. Draws from the posterior can thus be obtained by sequentially sampling from Equation (5) for $i = 1, \ldots, n$. When G_0 is a conjugate base distribution to the likelihood $f(\cdot|\phi_j)$ this is relatively straight forward, otherwise, a more taxing approach is required (see MacEachern & Müller (1998) and Neal (2000) on how to handle the non-conjugate case).

Unfortunately, this sampler can produce highly correlated draws from $\phi_1, \ldots, \phi_n|z$. High levels of correlation between the sampled ϕ_is means the algorithm needs to run for a long time in order to generate realizations from over the entire support of the distribution. This inefficiency comes from the finite nature of the DP prior. Under the DP prior the elements of ϕ will often equal one another to produce a group of ϕ_is with the same value. If the group of observations having the same valued ϕ_i is large, this step-by-step sampler of ϕ_i will continually produce realizations equal in value to the existing groups. As a result, the algorithm often gets stuck sampling the same set of ϕ_j and not generating 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 $\phi|z$. 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 $i' \neq i$. With this notation Equation (5) can be rewritten as:

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

Draws from $\phi | z$ are again made from the conditionals, however, each sweep of the sampler now consists of 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 first sampler, except now instead of recording the drawn ϕ_is, we discard them and only keep track of the state vector, s, θ and the number of clusters, k. In the context of sampling s_i, if $s_i = 0$, a new θ_{k+1} is sampled from $G(\phi | z_i)$ and k is increased by 1. Likewise, if $n_j^{(i)} = 0$, in other words, θ_j is only observed at the ith observation and it is not resampled, θ_j is dropped from θ. A new value for s_i is then sampled from one of the other existing orders and k is decreased by 1.

In Step 2, the ϕ_i's associated with the jth-cluster are all updated simultaneously by sampling from the posterior of θ_j conditional on the observations associated with the jth cluster. Thus, instead of sampling the ϕ_is step-by-step as in the previous sampler, a more efficient block sampler is employed in drawing ϕ. This ensures that the realizations of ϕ_i will be uncorrelated and representative of a nice mix of draws from the posterior distribution.

After collecting a large number of draws from the posterior by iterating on steps 1 and 2 we obtain $\{\theta^{(s)}\}_{r=1}^{R}$. Note that for each drawn $\theta = \{\theta_1, ..., \theta_k\}$, there is an associated state vector s, and number of observations in each cluster $\{n_1, ..., n_k\}$. k will vary over draws, so that the size of θ will change. The Bayesian estimate, or predictive

density, for a given α is obtained by integrating out these random variables as in

$$\pi(z^*|z) = \int \pi(z^*|\theta)\pi(\theta|z)d\theta \tag{7}$$

$$\approx \frac{1}{R}\sum_{r=1}^{R}\pi(z^*|\theta^{(r)}) \tag{8}$$

where

$$\pi(z^*|\theta) = \frac{\alpha}{\alpha+n}h(z^*) + \sum_{i=1}^{k}\frac{n_i}{\alpha+n}f(z^*|\theta_i), \tag{9}$$

and $h(z^*) = \int f(z^*|\phi)G_0(d\phi)$. The parameter α can also be sampled, which would add a third step to the above procedure. The model in the next section allows for this.

3 Stochastic Volatility and DPM Innovations

We now extend the DPM model of the unconditional distribution to time-varying conditional distributions. Let the stochastic volatility-Dirichlet process mixture model (SV-DPM) be defined as:

$$y_t = \exp(h_t/2)z_t, \tag{10}$$

$$h_t = \delta h_{t-1} + \sigma_v v_t, \quad v_t \overset{iid}{\sim} N(0,1), \tag{11}$$

$$z_t|\eta_t,\lambda_t^2 \overset{iid}{\sim} N(\eta_t, 1/\lambda_t^2) \tag{12}$$

$$\left.\begin{pmatrix}\eta_t\\ \lambda_t^2\end{pmatrix}\right| G \overset{iid}{\sim} G \tag{13}$$

$$G|G_0,\alpha \sim \mathrm{DP}(G_0,\alpha) \tag{14}$$

$$G_0(\eta_t,\lambda_t^2) \equiv N(m,(\tau\lambda_t^2)^{-1}) - \Gamma(v_0/2, s_0/2) \tag{15}$$

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 a first-order autoregressive (AR) process with the AR-parameter δ. For identification of the model the intercept in returns and log-volatility are removed and their effect is subsumed into the DPM component. To ensure the stationarity of y_t, δ is restricted to the interval $(-1,1)$. This guarantees that the mean and variance of h_t will be finite. The innovation to volatility, v_t, is independently and identically distributed standard normal. As currently constituted, our SV-DPM model does not address the issue of leverage effects (Jacquier et al. (2004);

Yu (2005); Omori et al. (2007)); i.e., the possible correlation that exists between y_t and h_t. Instead, we assume the innovations z_t and v_t are independent.[3]

Equations (10)-(11) are the typical representation for the stochastic volatility model. It is Equations (12)-(15), where a hierarchical representation of the DPM model is defined for the z_t's unknown distribution, F, that distinguishes our semiparametric SV model from a parametric stochastic volatility model. In this particular DPM, the kernel function from Equation (1) is defined by Equation (12) to be a normal density function with the parameter vector $\phi_t = (\eta_t, \lambda_t^2)$, where η_t is the mean and $1/\lambda_t^2$ the variance. The distribution G in Equations (13)-(14) represents the distribution of the parameter vector ϕ_t and the mixing probabilities V_t. A DP-prior with a conjugate normal-gamma base distribution, G_0, for η_t and λ_t^2 is assigned in Equations (14)-(15) to G. Like the DPM representation of the Praetz model, the $h_t(z_t)$ for the SV-DPM model is a Student-t distribution. However, since the variance and mean of $f(\cdot|\phi_t)$ are both unknown, the marginal posterior distribution, $G(d\phi|z_t)$, equals a conditional normal-gamma distribution.

In addition to this hierarchical definition, the SV-DPM model has the concise mixture representation:

$$y_t|f_N, h_t \sim f_N(\cdot|h_t, G) = \int f_N(\cdot|\eta_t, \lambda_t^{-2}\exp(h_t))\, dG \tag{16}$$

where G is again defined by Equation (14)-(15) and $f_N(\cdot|\eta, \sigma^2)$ is the density function of a normal distribution with mean η and variance σ^2. The flexibility of the semiparametric SV-DPM model is also seen in its continuous nonparametric conditional distribution form:

$$y_t|h_t \overset{\perp}{\sim} \sum_{j=1}^{\infty} V_j f_N\left(\cdot|\eta_j, \lambda_j^{-2}\exp(h_t)\right), \tag{17}$$

where $V_1 = W_1$, and $V_j = W_j \prod_{s=1}^{j-1}(1 - W_s)$ with $W_j \sim Beta(1, \alpha)$, and $\overset{\perp}{\sim}$ denotes a sequence of random variables that are independently distributed.

As defined the semiparametric SV-DPM model avoids the distributional assumptions associated with the class of parametric SV models. Instead, in the terminology of Müller & Quintana (2004), the SV-DPM model "robustifies" the class of parametric SV models. By modeling the innovations with a DPM model diagnostics and sensitivity analysis can now be conducted by nesting parametric SV models within the SV-DPM model. This

[3]Extending the SV-DPM to include leverage effect can be done but the DPM portion of the model becomes much more computationally challenging. As a result, leverage effect are a topic for a future research.

can be seen in the continuous mixture representation of Equation (17) where a SV model with *iid* normal errors is found with the semiparametric SV model when $V_1 = 1$, $V_j = 0$, for $j > 1$ and $\phi_t = (\eta, \lambda^2)'$, for $t = 1, \ldots, n$. The Student-t SV model of Harvey et al. (1994) is also nested within the SV-DPM model when $\alpha \to \infty$, $\phi_t = (0, \lambda_t^2)$, and $\lambda_t^2 \sim \Gamma(\nu/2, \nu/2)$, where ν is the degrees of freedom to the Student-t distribution.

4 Bayesian Inference of the SV-DPM

The inherent difficulty with every stochastic volatility model, regardless of its innovations being modeled parametrically or nonparametrically, is the intractability of the SV's likelihood function. Since the return process, y_t, is comprised of the two innovations, z_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 (MCMC) 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 naturally 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 distributions into tractable blocks of conditional posterior distribution and design a stylized MCMC sampler for each blocked conditional distribution. The accuracy of the sampler and its computational costs are highly 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 interation through these conditional distributions denotes a sweep with the MCMC 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 regions of δ's parameter space, and $\pi(\sigma_v^2) = \text{Inv-}\Gamma(v_0/2, s_0/2)$. 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 $\sigma_v^2|h,\delta \sim \text{Inv-}\Gamma((n-1+v_0)/2, (s_0 + \sum_{t=2}^n (h_t - \delta h_{t-1})^2)/2)$. If the 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 of δ is obtained.

Drawing the latent volatility vector, h, from $\pi(h|y,\psi,\phi)$ presents a difficult problem that has attracted much attention (see Jacquier et al. (1994); Pitt & Shephard (1997); Kim et al. (1998); and Chib et al. (2002)). One approach is applying Jacquier et al. (1994) element-by-element algorithm to the SV-DPM model and sampling h by sequentially drawing h_t, $t = 1, \ldots, n$, from:

$$\begin{aligned} \pi(h_t|y_t, h_{t-1}, h_{t+1}, \phi_t, \psi) \quad \propto \quad & \frac{1}{\exp\{h_t/2\}} \exp\left\{ -\frac{(y_t - \eta_t)^2}{2\exp\{h_t\}\lambda_t^{-2}} \right\} \\ & \times \exp\left\{ -\frac{(h_t - \delta(h_{t+1} + h_{t-1})/(1+\delta^2))^2}{2\sigma_v^2/(1+\delta^2)} \right\}. \end{aligned} \tag{18}$$

Conditional on the elements of the mixture vector equaling $\eta_t = 0$ and $\lambda_t^{-2} = 1$ for all t, this volatility sampler for the SV-DPM model is exactly the same as the JPR sampler of the volatilities in their SV model. When the mixture parameters η_t and λ_t^{-2} are not 0 and 1, but are still conditioned on some fixed value, the first term in (18) is

$$\frac{1}{\exp\{h_t/2\}} \exp\left\{ -\frac{\tilde{y}_t^2}{2\exp\{h_t\}} \right\}$$

where $\tilde{y}_t^2 = (y_t - \eta_t)^2/\lambda_t^{-2}$, and so the entire suite of existing element-by-element samplers by Geweke (1994), Pitt & Shephard (1997), Kim et al. (1998), and Jacquier et al. (2004) are directly applicable by reparameterizing to $\tilde{y}_t^2$.

Since each draw of h_t is conditional on the previous draw of 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 with the algorithm 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 the entire h from $h|y, \psi, \phi$ in one single draw (see Kim et al. (1998); and Chib et al. (2002))). Drawing h in such a manner eliminates this correlation, but it requires taking the log-squared transformation of y. In the context of the SV-DPM model the tangible nature of the DPM model of z's unknown distribution is lost under a log-square tranformation of y. Thus, sampling the entire h all at once is not feasible for the SV-DPM model. Fortunately, draws of the volatilities that are less correlated than the single-state draws can be obtained by applying the basic idea of the entire block sampler to random length blocks of the latent volatilities (see Elerian et al. (2001)).

The random length block sampler divides h into blocks of subvectors $\{\boldsymbol{h}_j\}$, where $\boldsymbol{h}_j = (h_j, h_{j+1}, \ldots, h_{j+l_j})'$, $1 \leq j < n$ and the length of the subvector l_j is randomly drawn from a Poisson distribution with hyperparameter $\lambda_h = 5$. By letting the length of the blocks be random we ensure that with each sweep different subblocks of h are sampled. This helps reduce the degree of dependency that would exist in the draws of the $\boldsymbol{h}_j$ if the l_js were fixed. By lowering the level of correlation in the draws of the $\boldsymbol{h}_j$s, we reduce the number of sweeps needed to produce reliable estimates of the model parameters.

Because $\boldsymbol{h}_j|y, h_{j-1}, h_{j+l_j+1}, \psi, \phi$ is not a standard distribution, draws are made with the tailored Metropolis-Hastings sampler (see Chib & Greenberg (1995,1998) for detailed description). A tailored-MH sampler produces draws from the target posterior distribution $\boldsymbol{h}_j|y, h_{j-1}, h_{j+l_j+1}, \psi, \phi$ by generating candidate draws from a l_j-variate Student-t distribution with v_h degrees of freedom, a mean of:

$$\boldsymbol{h}_j^* = \arg\max_{\boldsymbol{h}_j} \log f(\boldsymbol{y}_j|\boldsymbol{\phi}_j, \boldsymbol{h}_j)\ \pi(\boldsymbol{h}_j|h_{j-1}, h_{j+l_j+1}, \psi)$$

and a covariance, Σ_{h_j}, equal to the negative inverse Hessian of the maximized function evaluated at $\boldsymbol{h}_j^*$.[4] The candidate draw, $\boldsymbol{h}_j'$, is accepted as a realization with Metropolis-

[4]The conditional posterior depends on the unknown h_0 and h_{n+1} when $j = 1$ and $j + l_j = n$, respectively. We integrate out these unknowns by sampling $h_0 \sim N(0, \sigma_v^2/(1-\delta^2))$ and $h_{n+1}|h_n \sim N(\delta h_n, \sigma_v^2)$.

Hastings probability:

$$\min\left\{\frac{f(\boldsymbol{y}_j|\boldsymbol{\phi}_j,\boldsymbol{h}'_j)\;\pi(\boldsymbol{h}'_j|h_{j-1},h_{j+l_j+1},\psi)}{f(\boldsymbol{y}_j|\boldsymbol{\phi}_j,\boldsymbol{h}_j)\;\pi(\boldsymbol{h}_j|h_{j-1},h_{j+l_j+1},\psi)}\frac{St(\boldsymbol{h}_j|\boldsymbol{h}^*_j,\Sigma_{h_j},v_h)}{St(\boldsymbol{h}'_j|\boldsymbol{h}^*_j,\Sigma_{h_j},v_h)},1\right\},$$

where $f(\boldsymbol{y}_j|\boldsymbol{\phi}_j,\boldsymbol{h}'_j)=\prod_{t=j}^{j+l_j}f_N(y_t|\eta_t,\lambda_t^{-2}\exp\{h_t\})$,

$$\pi(\boldsymbol{h}_j|h_{j-1},h_{j+l_j+1},\psi)=\prod_{t=j}^{j+l_j+1}\exp\left\{-\frac{(h_t-\delta h_{t-1})^2}{2\sigma_v^2}\right\},$$

and $St(\boldsymbol{h}_j|\boldsymbol{h}^*_j,\Sigma,v)\propto[1+(\boldsymbol{h}_j-\boldsymbol{h}^*_j)'\Sigma^{-1}(\boldsymbol{h}-\boldsymbol{h}^*_j)/v]^{-(l_j+v)/2}$ denotes the l_j-variate Student-t probability density function evaluated at $\boldsymbol{h}_j$.[5]

For a given ψ and h, draws from the posterior distribution $\phi|y,h$ can be made with a variant of Section 2.2 Gibb sampler of $\pi(\phi|z)$. To improve the mixing behavior of the sampler we again use the equivalent distribution $\pi(\theta,s|y,h)$ and sample from it by:

1. Drawing s and k by sampling s_t for $t=1,\ldots,n$ from:

$$\begin{aligned}\phi_t|y,h,\phi^{(t)} \;\sim\; & c\frac{\alpha}{\alpha+n-1}g(y_t,h_t)G(d\phi|y_t,h_t)\\ & +\frac{c}{\alpha+n-1}\sum_{j=1}^{k}n_j^{(t)}f(y_t|\theta_j,h_t)\delta_{\theta_j}(\phi_t),\end{aligned}\tag{19}$$

where $f(y_t|\theta_j,h_t)=f_N(y_t|\eta_j,\lambda_j^{-2}\exp\{h_t\})$ and the probability of a new draw is proportional to:

$$\begin{aligned}g(y_t,h_t) &= \int f_N(y_t|\eta_t,\lambda_t^{-2}\exp\{h_t\})\;G_0(d\phi_t)\\ &= \int\frac{1}{\sqrt{2\pi\lambda_t^{-2}\exp(h_t)}}\exp\left(-\frac{(y_t-\eta_t)^2}{2\lambda_t^{-2}\exp(h_t)}\right)G_0(d\phi_t)\\ &= \exp(-h_t/2)\int\frac{1}{\sqrt{2\pi\lambda_t^{-2}}}\exp\left(-\frac{(\xi_t-\eta_t x_t)^2}{2\lambda_t^{-2}}\right)G_0(d\phi_t)\\ &= \exp(-h_t/2)St(\xi_t|mx_t,(x_t^2+\tau)s_0/(\tau v_0),v_0)\\ &= St(y_t|m,(1+\tau\exp(h_t))s_0/(\tau v_0),v_0)\end{aligned}\tag{20}$$

with $\xi_t=y_t\exp\{-h_t/2\}$, and $x_t=\exp\{-h_t/2\}$. When $s_t=0$ is drawn, a new value for ϕ_t is sampled from:

$$G(d\phi_t|y_t,h_t)=f_N(y_t|\eta_t,\lambda_t^{-2}\exp\{h_t\})\;G_0(d\phi_t).$$

[5]We set $v_h=10$ to ensure that the tails of the candidate Student-t distribution are fat enough to envelope the unknown target distribution.

By the conjugate results of Geweke (2005) this posterior distribution equals the normal-gamma distribution:

$$\begin{aligned}
\lambda_t^2 &\sim G(\bar{v}/2, \bar{s}/2) && (21)\\
\eta_t|\lambda_t^2 &\sim N(\bar{\mu}, (\bar{\tau}\lambda_t^2)^{-1}) && (22)\\
\bar{\tau} &= \tau + x_t^2 && (23)\\
\bar{\mu} &= \bar{\tau}^{-1}(\tau m + x_t\xi_t) && (24)\\
\bar{v} &= v_0 + 1 && (25)\\
\bar{s} &= s_0 + (\bar{\mu} - \xi_t)^2 x_t^2 + (\bar{\mu} - m)^2\tau && (26)
\end{aligned}$$

2. Given s and k, θ is sampled by drawing θ_j, $j = 1, \ldots, k$ from:

$$\theta_j|y, h, s, k \propto \prod_{t:s_t=j} f_N\left(y_t|\eta_j, \lambda_j^{-2}\exp\{h_t\}\right) G_0(d\theta_j) \quad (27)$$

which is simply the linear model $\xi_t = x_t\eta_j + u_t$ over the observations $\{t : s_t = j\}$. Conjugacy again produces the normal-gamma posterior but now over the observations associated with the state j. Thus, the posterior equals:

$$\begin{aligned}
\lambda_j^2 &\sim G(\bar{v}/2, \bar{s}/2) && (28)\\
\eta_j|\lambda_j^2 &\sim N(\bar{\mu}, (\bar{\tau}\lambda_j^2)^{-1}) && (29)\\
\bar{\tau} &= \tau + \sum_{t:s_t=j} x_t^2 && (30)\\
\bar{\mu} &= \bar{\tau}^{-1}\left(\tau m + b\sum_{t:s_t=j} x_t^2\right) && (31)\\
\bar{v} &= v_0 + n_j && (32)\\
\bar{s} &= s_0 + SSE_j + (\bar{\mu} - b)^2\sum_{t:s_t=j} x_t^2 + (\bar{\mu} - m)^2\tau && (33)
\end{aligned}$$

where b is the OLS estimate from regressing ξ_t on x_t over the observations $\{t : s_t = j\}$, and SSE_j is the sum of squares errors from this regression.

Finally, the DPM precision parameter α can be sampled from $\alpha|k$ with the algorithm of Escobar & West (1995).

5 Features of the Model

The above MCMC approach, after dropping a suitable number of initial iterations, will produce a set of draws $\{k^{(r)}, \theta^{(r)}, \alpha^{(r)}, \eta^{(r)}, \lambda^{(r)}, \delta^{(r)}, \sigma_v^{2(r)}, h^{(r)}\}_{r=1}^R$, from the posterior

density. Associated with each draw $\theta = \{\theta_1, ..., \theta_k\}$, with a typical element $\theta_i = (\eta_i, \lambda_i^{-2})$, there is the state vector s, and number of observations in each cluster $\{n_1, ..., n_k\}$. Recall that k will vary over draws, so that the size of θ will change.

Given these draws we can produce simulation consistent estimates of posterior quantities. For example, the posterior mean of δ is $E[\delta|y] \approx \frac{1}{R}\sum_{r=1}^{R}\delta^{(r)}$, and this estimate can be made more accurate by increasing R. For a full treatment on MCMC methods see Robert & Casella (1999). In a similar way various quantities of the predictive density can be estimated.

5.1 Predictive Density

In density estimation the key quantity of interest is the predictive density. Gelfand & Mukhopadhyay (1995) discuss this and more generally the estimation of linear functionals for DPM models. Drawing on their results, the in-sample predictive density at observation t is

$$\begin{aligned} \pi(y_t^*|y) &= \int \pi(y_t^*|\theta, h_t)\pi(\theta, h_t|y)d\theta dh_t \qquad (34) \\ &\approx \frac{1}{R}\sum_{r=1}^{R}\pi(y_t^*|\theta^{(r)}, h_t^{(r)}) \qquad (35) \end{aligned}$$

where $\theta^{(r)}$, and $h_t^{(r)}$ are draws from the posterior simulator.[6] The conditional density in (35) is

$$\begin{aligned} \pi(y_t^*|\theta, h_t) = \frac{\alpha}{\alpha+n}St(y_t^*|m, (1+\tau\exp(h_t))s_0/(\tau v_0), v_0) \\ + \sum_{i=1}^{k}\frac{n_i}{\alpha+n}f_N(y_t^*|\eta_i, \lambda_i^{-2}\exp(h_t)). \qquad (36) \end{aligned}$$

The predictive density for the SV-DPM model is analogous to the predictive density in Equation (8) of Escobar & West (1995) for the simple DPM model except that we have the added structure of the stochastic volatility process. The predictive one step out-of-sample density is computed in a similar way as

$$\begin{aligned} \pi(y_{n+1}^*|y) &= \int \pi(y_{n+1}^*|\theta, h_{n+1})\pi(\theta, h_{n+1}|y)d\theta dh_{n+1} \qquad (37) \\ &\approx \frac{1}{R}\sum_{r=1}^{R}\pi(y_{n+1}^*|\theta^{(r)}, h_{n+1}^{(r)}) \qquad (38) \end{aligned}$$

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

where

$$\pi(y^*_{n+1}|\theta, h_{n+1}) = \frac{\alpha}{\alpha+n} St(y^*_{n+1}|m, (1+\tau\exp(h_{n+1}))s_0/(\tau v_0), v_0) + \sum_{i=1}^{k} \frac{n_i}{\alpha+n} f_N(y^*_{n+1}|\eta_i, \lambda_i^{-2}\exp(h_{n+1})) \quad (39)$$

and for each posterior draw $h_n^{(r)}$ we simulate $h_{n+1}^{(r)} = \delta h_n^{(r)} + \sigma_v v_{n+1}$, $v_{n+1} \sim N(0,1)$.

The predictive likelihood is calculated by evaluating $\pi(y^*_{n+1}|y)$ at the observed data value $y^*_{n+1} = y_{n+1}$. The SV-DPM model's predictive likelihood for observations $y_{n+1}, ..., y_{n+\tau}$, conditional on y is defined as

$$PL_{SV-DPM} = \pi(y_{n+1}, ..., y_{n+\tau}|y) = \prod_{i=1}^{\tau} \pi(y_{n+i}|y_1, ..., y_{n+i-1}) \quad (40)$$

and each of the terms of the right-hand side can be approximated using (38). The predictive likelihood records the predictive record of a model on a set of data, and provides a natural quantity for model comparison (Geweke & Whiteman (2006)). Given the predictive likelihood for another specification, such as a stochastic volatility model with normal return innovations, $PL_{SV-N} = \sum_{i=1}^{\tau} \pi(y_{n+i}|y_1, ..., y_{n+i-1})$, we can compare the two specifications with a log-predictive Bayes factor defined as $\log(PL_{SV-DPM}/PL_{SV-N})$. A value greater than 0 is evidence in favor of the SV-DPM model while values less than 0 are evidence in favor of the SV-N model.

5.2 Conditional Moments

Using Equation (34) the in-sample moments can be computed. For instance, the first and second moments of the model can be approximated as

$$E[y_t|y] \approx \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\alpha^{(r)}}{\alpha^{(r)}+n}m + \sum_{r=1}^{k^{(r)}}\frac{n_i^{(r)}}{\alpha^{(r)}+n}\eta_i^{(r)}\right) \quad (41)$$

$$E[y_t^2|y] \approx \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\alpha^{(r)}}{\alpha^{(r)}+n}\left[\frac{(1+\tau\exp(h_t^{(r)}))s_0}{(\tau(v_0-2))} + m^2\right] + \sum_{i=1}^{k^{(r)}}\frac{n_i^{(r)}}{\alpha^{(r)}+n}\left[\eta_i^{(r)2} + \lambda_i^{(r)-2}\exp(h_t^{(r)})\right]\right) \quad (42)$$

and the variance as $E[y_t^2|y] - E[y_t|y]^2$.

6 Simulation Examples

In this section we consider two examples from simulated 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),

$$\begin{aligned} y_t &= \mu + \exp(h_t/2)z_t, \quad z_t \sim N(0,1) \\ h_t &= \gamma + \delta h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0,1) \end{aligned} \tag{43}$$

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),

$$\begin{aligned} y_t &= \mu + \exp(h_t/2)z_t, \quad z_t \sim St(0, (\nu-2)/\nu, \nu) \\ h_t &= \gamma + \delta h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0,1). \end{aligned} \tag{44}$$

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 $\nu \sim U(2, 100)$. Details of the estimation of these two models is briefly listed in the Appendix.

The priors for the SV-DPM model are chosen to match the parametric models $\delta \sim N(0, 100)I_{|\delta|<1}, \sigma_v^2 \sim \text{Inv} - \Gamma(10/2, 0.5/2), m = 0, \tau = 10, v_0 = 10, s_0 = 10$, and $\alpha \sim \Gamma(2, 8)$.

6.1 Example 1

In the first example we simulate from the SV-t specification (44) with the parameters $\mu = 0, \gamma = -0.01025, \delta = 0.95, \sigma_v^2 = 0.04$, and $\nu = 6$. The first 100 simulated draws from this 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. The SV-N is misspecificed and the estimates of δ and σ_v^2 are poor. The only way the SV-N can approximate the fat-tails in z_t is to increase the variance of log-volatility, σ_v^2. On the other hand, the SV-DPM does a better job in estimating the parameters and is generally very close to the correctly specificied SV-t. Note that the SV-DPM is using, on average, a mixture of 3 normals to approximate the density of return innovations. The last row in the table displays the root mean squared error (RMSE) for the smoothed variance estimates measured against the true variance.[7] The RMSE for the SV-DPM is similar to the SV-t model and worse for the SV-N. Figure 1 displays a time-series plot of the estimated variances for these models. The log-predictive densities, one period out-of-sample, are displayed in Figure 2. The tails of the SV-N are too thin relative the the SV-t while the SV-DPM is very close to the true model.

[7]This is computed as $\sqrt{\frac{1}{1500}\sum_{t=1}^{1500}(\text{Var}_T(r_t) - \exp(h_t))^2}$ where $\exp(h_t)$ denotes the true variance from the DGP and $\text{Var}_T(r_t)$ is a full sample model estimate computed from Section 5.2.

6.2 Example 2

In this example the DGP is the same as in (44) except that z_t is replaced by a mixture of two normals,

$$z_t \sim \begin{cases} N(\mu_1, \sigma_1^2) \text{ with probability } p \\ N(\mu_2, \sigma_2^2) \text{ with probability } 1-p. \end{cases} \tag{45}$$

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 0.0, 1.0, -1.3056, and 5.2042, respectively. The density is shown in Figure 3 and has a fat left tail and a thin right tail. In this example both the SV-N and SV-t are misspecified and cannot accommodate the asymmetry in z_t.

Parameter estimates are found in Table 2. The SV-N estimates are strongly effected by the misspecification while the SV-t is somewhat more robust. The SV-t uses the degrees of freedom parameter to approximate the asymmetry, however, this implies very fat tails for both negative and positive returns and is inconsistent with the density of z_t. The SV-DPM produces reasonable estimates and is the only model that accurately estimates σ_v^2. The RMSE for the variance further illustrates the problems for the SV-N and the better performance of the SV-t and SV-DPM. The predictive densities, one period out-of-sample, in Figure 4 show the SV-DPM is able to capture the asymmetry in the return distribution unlike the parametric alternatives.

7 Empirical Example

In this section we report the results for the SV-DPM as it is applied to daily stock return data. More specifically, we apply the SV-DPM model's MCMC sampler of Section 4 to 4000 daily returns from January 2, 1990 to November 8, 2005 of the Center for Research in Security Prices value-weighted portfolio index. The data scaled by 100 are displayed in Figure 5, and have a mean 0.0446, variance 0.9636, skewness -0.1163, and kurtosis 7.0571. We also fit the SV-N and SV-t models to this daily series. The priors for the three SV models are the same as those used in the previous simulation examples. The SV-DPM sampler is first burned in for 7000 draws with the following 10,000 draws used to make inferences.

In Table 3 we report the means and standard deviations of each SV model's parameter draws. As in the simulation examples, the SV-DPM model's estimate of variance of volatility parameter, σ_v^2, is the smallest of the three SV models. The dynamics of

volatility as captured by AR-parameter, δ, are close to being indistinguishable from one another. The SV-DPM has the highest degree of persistence in volatility with $\delta = 0.9887$. This is only slightly larger than the SV-t model's $\delta = 0.9877$ and the SV-N model's $\delta = 0.9848$. The variances of these draws for each of the three SV models are tightly distributed around the mean with the largest standard deviation equaling 0.0037 for the SV-N model.

The average mixture order for the daily portfolio return is sizeably larger at $k = 5.93$ than the estimate of $k = 3.32$ found for the simulated SV-t model data series.[8] Such a large k for the SV-DPM model suggests that not only are the daily returns leptokurtotic, but also possibly skewed. SV models that fail to account for this non-Gaussian behavior compensate for it by having increased levels of volatility during high volatility periods. This can be seen in the smooth volatility plots of Figure 6. During high volatility periods, the SV-t model's smoothed volatility are always larger than that of the SV-DPM model. Although in Example 1 there were differences in the two models estimate of smoothed volatility during high volatility events, the difference found in Figure 1 is small relative to Figure 6 differences between the SV-DPM and SV-t.

Figure 7 plots the predictive densities for the models at the end of the sample. The SV-DPM displays clear differences compared to the SV-N and SV-t. For instance, the SV-DPM is more peaked around 0, and shows asymmetry. In addition to these differences, Figure 8 which plots the log-predictive densities, show the SV-DPM to produce fatter tails than the other models.

8 Conclusion

This paper proposes a new Bayesian semiparametric stochastic volatility model. We provide a MCMC approach to posterior simulation and discuss how various features of the model can be estimated. Compared to existing fully parametric stochastic volatility models, the semiparametric version performs well based on both simulation examples and an application to daily stock returns.

9 Appendix

to be completed

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

References

Blackwell, D. & MacQueen, J. (1973), 'Ferguson distributions via polya urn schemes', *The Annals of Statistics* **1**, 353–355.

Chib, S. & Greenberg, E. (1995), 'Understanding the metropolis-hastings algorithm', *American Statistician* **49**, 327–335.

Chib, S. & Greenberg, E. (1998), 'Analysis of multivariate probit models', **85**, 347–361.

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.

Chib, S., Tiwari, R. & Jammalamadaka, S. (1988), 'Bayes prediction density and regression estimation: A semi parametric approach', *Empirical Economics* **13**, 209–222.

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.

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. (2005), *Contemporary Bayesian Econometrics and Statistics*, Wiley.

Geweke, J. & Whiteman, C. (2006), Bayesian forecasting, *in* G. Elliot, C. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting', Elsevier, Amsterdam.

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 semiparametric models with application to option pricing. manuscript University of British Columbia.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with arch models', *Review of Economic Studies* **65**, 361–393.

Kon, S. (1984), 'Models of stock returns – a comparison', *Journal of Finance* **39**, 147–165.

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.

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.

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.

Table 1: Parameter Estimates for Simulation Example 1

	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	10.08094		
α		0.2789	0.17342				
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: Parameter Estimates for Simulation Example 2

	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: Estimates for Value-weighted CRSP Index Returns

	SV-DPM		SV-t		SV-N	
	mean	stdev	mean	stdev	mean	stdev
μ			0.0768	0.0108	0.0763	0.0108
γ			-0.0072	0.0028	-0.0071	0.0031
δ	0.9887	0.0032	0.9877	0.0033	0.9848	0.0037
σ_v^2	0.0128	0.0033	0.0196	0.0036	0.0245	0.0044
ν			18.6129	9.5954		
α	0.4243	0.2356				
k	5.9250	2.7575				

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

Figure 1: Variance Estimates

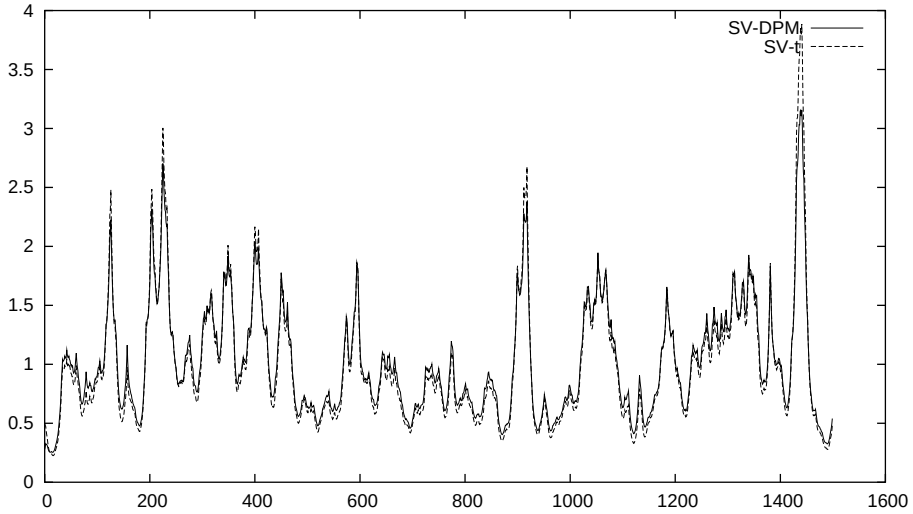

Figure 2: Log-predictive Densities

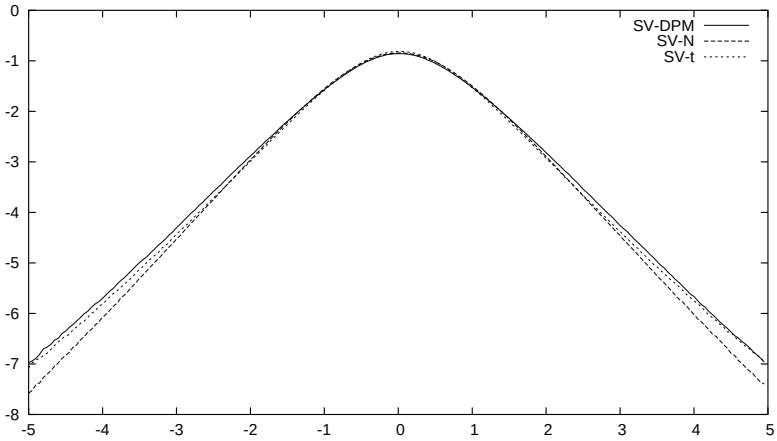

Figure 3: Density of z_t from a Mixture of Normals

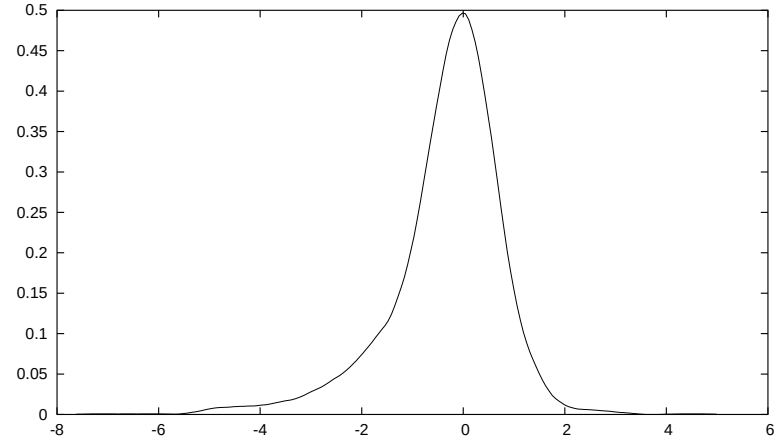

Figure 4: Predictive Densities

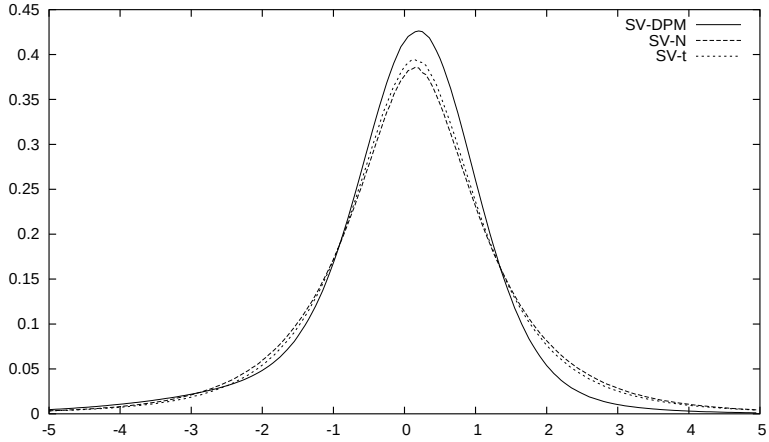

Figure 5: Value-weighted CRSP Index Returns

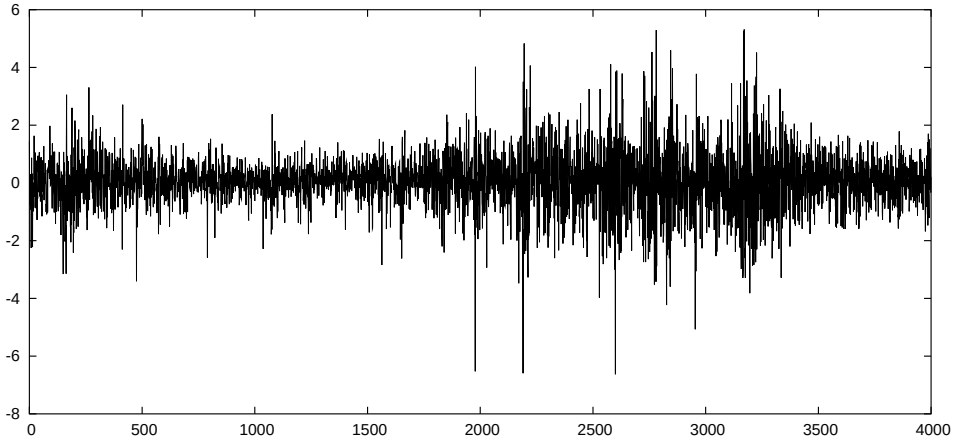

Figure 6: Variance Estimates, Value-weighted CRSP Index Returns

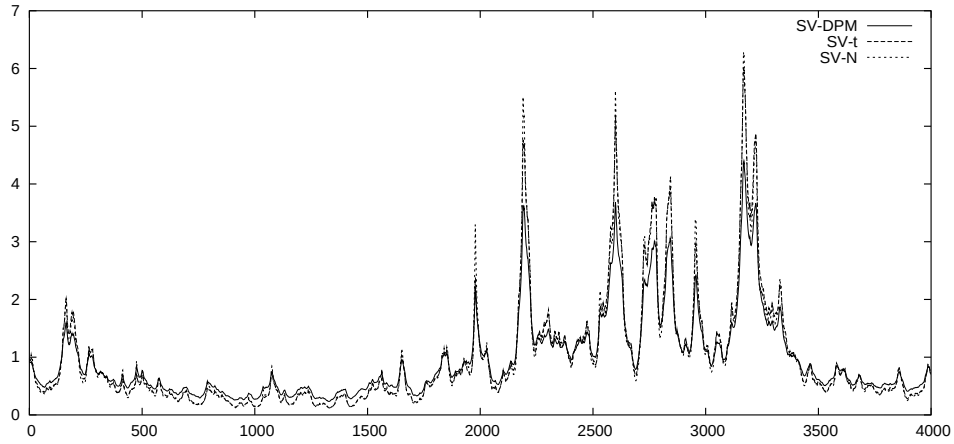

Figure 7: Predictive Densities, Value-weighted CRSP Index Returns

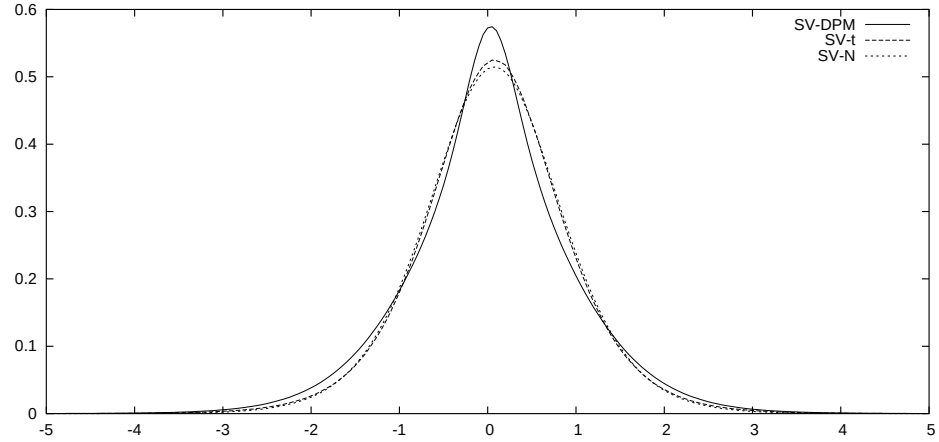

Figure 8: Log-predictive Densities, Value-weighted CRSP Index Returns