Semiparametric Asymmetric Stochastic Volatility*

Mark J. Jensen
Federal Reserve Bank of Atlanta
Mark.Jensen@atl.frb.org

John M. Maheu
University of Toronto
RCEA
jmaheu@chass.utoronto.ca

Abstract. This paper extends the stochastic volatility with leverage model, where returns are correlated with volatility, by flexibly modeling the bivariate distribution of the return and volatility innovations nonparametrically. The novelty of the paper is in modeling the unknown distribution with an infinite ordered mixture of bivariate normals with mean zero, but whose mixture probabilities and covariance matrices are unknown and modeled with the Dirichlet Process prior. A Bayesian Markov chain Monte Carlo sampler is designed to fully characterize the parametric and distributional uncertainty. Cumulative marginal likelihoods and log predictive Bayes factors for the semiparametric and parametric asymmetric stochastic volatility models are compared. We find substantial empirical evidence in favor of the semiparametric leverage version of the stochastic volatility model.
Keywords: Bayesian nonparametrics, Dirichlet process mixture, leverage effect

*We would like to thank the seminar participants at the 2010 Remini Conference in Economics and Finance, the 10th Annual All-Georgia Financial Conference, and the Department of Economics at Instanbul Bilgi University. The views expressed here are ours and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

This paper proposes a semiparametric, asymmetric, stochastic volatility model of continuously compounded returns where the instantaneous volatility of returns, the volatility-of-volatility, and the correlation between returns and volatility, in other words the leverage effect, are random second-order effects drawn from an unspecified distribution.[1] In the model, log-volatility is a parametric, stationary, first-order autoregressive process, but the rest of the model is nonparametric in the sense that no assumptions are made about the underlying joint distribution of the return and volatility innovations. Instead, a flexible, bivariate, Dirichlet process mixture prior is assumed for the unknown distribution. This nonparametric prior of Lo (1984) models the unknown distribution as a infinitely ordered mixture of bivariate, normal distributions with mean zero and the unknown covariance mixture matrices. Each daily return and latent volatility is assigned a mixture covariance according to the Dirichlet process, in other words, a random second-order effect.

By relaxing the distributional assumptions of the asymmetrical, stochastic volatility model, the paper's stochastic volatility model has a more flexible innovation distribution. Since the stochastic volatility model is unable to produce the level of skewness and kurtosis observed empirically except under implausible parameter values (see Das and Sundaram (1999)), the nonparametric joint distribution is designed to capture such non-Gaussian type behavior.

The paper provides a Markov chain, Monte Carlo posterior sampler for the nonparametric, asymmetric, stochastic volatility model. The sampler extends the univariate algorithm of the semiparametric stochastic volatility model by Jensen and Maheu (2010). A restricted version of the algorithm can also be applied to the parametric, asymmetric, stochastic volatility model. Parameter, volatility, and distributional uncertainty are integrated out via the unrestricted posterior sampler. Draws of the unknowns are used to generate one-day-ahead predictive densities of returns, which result in the cumulative log predictive Bayes factor when evaluated at the observed returns.

The paper's empirical analysis departs from the assumption that one of the models is the actual data generating process, and, hence, we do not carry out model specification tests. Instead, we believe all models are, sooner or later, misspecified in some form or another. We evaluate the nonparametric and parametric models predictive performance to market conditions by calculating their cumulative log Bayes factor over periods of market normalcy and market stress. Positive and negative market shocks are well known to have different

[1]See Harvey et al. (1994), Yu (2005) and Omori et al. (2007) for the parametric version of the asymmetric stochastic volatility model.

impacts on overall market volatility (see Chen and Ghysels (2011). Volatility increases more following a large drop in the market than volatility does after an increase in the market of similar magnitude. The cumulative log-Bayes factor will provide a dynamic measure of the role the Gaussian distribution and a constant leverage effect plays in the forecasts of returns and volatility during periods of market stress.

The paper is organized as follows. In the next section, the asymmetric, stochastic volatility model with a nonparametric Dirichlet process mixture prior for the unknown distribution is constructed under discrete and continuous time. Section 3 spells out a Markov chain, Monte Carlo sampler of all the unknowns in the nonparametric model. In Section 4 we apply the semiparametric and parametric version of the asymmetric, stochastic volatility model to 28 years of daily market returns as measured by the value-weighted market portfolio from the Center of Research in Security Prices. Section 4 also contains analysis of the parametric and nonparametric versions of the model by computing the log predictive Bayes factor over the last two years of returns. A summary and conclusions are contained in Section 5.

2 Asymmetric Stochastic Volatility with DPM

We model asset returns with the following semiparametric, asymmetric, stochastic, log-volatility model

$$y_t = \mu + \exp\{h_t/2\}\epsilon_t \tag{1}$$

$$h_{t+1} = \delta h_t + \eta_t \tag{2}$$

where y_t is the continuously compounded daily return at time periods $t = 1, \ldots, n$, and h_{t+1} is the value of the latent, log-volatility, one-day-ahead. The absolute value of the autoregressive parameter, δ, is constrained to the unit interval, ensuring the log-volatility process in Eq. (2) is stationary.[2]

We relax all assumptions concerning the joint distribution of ϵ_t and η_t, and, instead, allow their distribution to be completely unknown and random as if the distribution were an additional unknown to the parameters and latent volatilities of the SV model. Being unknown and random, the joint innovation distribution requires a prior, which can then be used to obtain the random distributions posterior once data has been collected.

[2]Because the mean of h_{t+1} can be subsumed into the variance of ϵ_t, identification requires the mean of log-volatility to be zero; i.e., the intercept term of h_{t+1} must be set equal to zero.

We choose the following Lo (1984) type Dirichlet process mixture prior (DPM)

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N(\mathbf{0}, \Lambda_t), \tag{3}$$

$$\Lambda_t \sim G, \tag{4}$$

$$G \sim DP(\alpha, G_0), \tag{5}$$

to model the unknown distribution. In Eq. (3)-(5), the tth innovations are distributed as a bivariate normal with a mean zero vector but whose covariance matrix Λ_t is random and distributed as G, which is also unknown and distributed as a Dirichlet process.

The DPM builds on the well known property that a flexible distribution can be found by mixing together a number of known distributions. It extends this concept by mixing together an infinite number of distributions. In its simplest and most basic form the Dirichlet process mixture models the innovation vector, $(\epsilon_t, \eta_t)'$ as independent realizations from the same, unknown, distribution which is modeled as a mixture of distributions

$$\int F_N(\theta)G(d\theta), \tag{6}$$

where F_N is a normal distribution function with mean zero and covariance matrix Λ, and G is a weighted mixture of the Λs.

Eq. (1)-(5) constitute the semiparametric, asymmetric, stochastic volatility model with DPM prior model (ASV-DPM). At first glance, the ASV-DPM model, with its mean zero, bivariate, normal distribution function, might seem to lack the capacity to fit the non-Gaussian behavior of returns and log-volatility. This, however, is incorrect. Fixing the mean of F to zero only limits the DPM prior to the class of distributions having one mode. This is hardly a limitation since asset returns are not known to have distributions with more than one mode.

The DPM prior for the ASV model can also be viewed in terms of a random, second-order, effects model, where Λ_t is the random effect, but with a slight twist. Unlike a random effects model, where Λ_t is typically assumed to follow a parametric, Inverse-Wishart distribution, in the ASV-DPM model the distribution G is unknown and is modeled nonparametrically. As a random distribution modeled nonparametrically, G allows the Λ_ts to take on dramatically different values, and to be distributed with "multimodality", and more "skewness" and "kurtosis" than is possible with a parametric distribution. However, because G is nonparametric the second-order, random effects, matrices do not have any financial or economic meaning. They are simply building blocks in fitting the unknown distribution of (ϵ_t, η_t).

Employing Sethuraman (1994) representation of $DP(\alpha, G_0)$, G will almost surely be equal to the discrete distribution

$$G(d\Lambda) = \sum_{j=1}^{\infty} \pi_j \mathcal{I}_{\Sigma_j}(d\Lambda), \tag{7}$$

where $\mathcal{I}_{\Sigma_j}(\cdot)$ is a degenerative distribution on the covariance matrix

$$\Sigma_j = \begin{pmatrix} \sigma^2_{y,j} & \sigma_{yh,j} \\ \sigma_{yh,j} & \sigma^2_{h,j} \end{pmatrix}. \tag{8}$$

Each Σ_j is a covariance matrix randomly drawn from the DP prior's base distribution, G_0. To ensure conjugacy, we let G_0 be the Inverse-Wishart distribution with scale matrix S_0 and v_0 degrees of freedom, i.e.,

$$G_0 \equiv \text{Inv-Wish}(S_0, v_0). \tag{9}$$

The probability of Λ_t being equal to a particular Σ_j is π_j where $\pi_1 = V_1, \pi_j = V_j \prod_{j'<j}(1 - V_{j'})$, and $V_j \sim \text{Beta}(1, \alpha)$, for $\alpha > 0$.

Eq. (7) is referred to as the "stick-breaking" representation of the DP prior since the π_js can be viewed as having been formed by breaking a stick of unit-length into pieces of length π_j and assigning Σ_j to Λ at that probability level. In (7), G_0 is our "best" guess at the distribution of the Λ_ts. Because the π_js are dependent on the V_js being drawn from the $\text{Beta}(1, \alpha)$ distribution, whose expected value is $1/(1 + \alpha)$, for relatively large values of α, the DP prior for G converges to G_0; i.e., each π_j is close to zero with a unique Σ_j drawn from G_0. Hence, as α gets larger, it follows that the uncertainty drops about Λ_t being distributed according to the parametric distribution G_0. On the other hand, for values of α close to zero, the prior for G will consist of a discrete distribution whose support is located on only a few covariances Σ_j. The DP precision parameter, α can, thus, be understood as controlling the number of random, second-order, effects.

2.1 Parsimony of the DP

Parsimony of the ASV-DPM model, in other words, clustering in the covariances, Λ_t, can be seen in the discrete nature of G. The almost sure discreteness of G ensures there will ties among the Λ_ts. To be explicit, the marginal joint distribution of the covariances $\pi(\Lambda_1, \Lambda_2, \ldots, \Lambda_n) = \pi(\Lambda_1)\pi(\Lambda_2|\Lambda_1)...\pi(\Lambda|\Lambda_1, \ldots, \Lambda_{n-1})$ is defined sequentially by $\pi(\Lambda_1) \equiv G_0$ and

$$\Lambda_t | G, \Lambda_1, \ldots, \Lambda_{t-1} \sim G, \tag{10}$$

$$G|\Lambda_1, \ldots, \Lambda_{t-1} \sim DP\left(\frac{\alpha}{\alpha + t - 1} G_0 + \sum_{t'=1}^{t-1} \frac{1}{\alpha + t - 1} \mathcal{I}_{\Lambda_{t'}}(dG), \alpha\right), \tag{11}$$

for $t = 2, \ldots, n$ (see Blackwell and MacQueen (1973)). Integrating out G from each of the conditional distributions in Eq. (10), we obtain what can be described as the "strong getting stronger" form of the conditional distribution for Λ_t

$$\Lambda_t | \Lambda_1, \ldots, \Lambda_{t-1} \sim \begin{cases} G_0 & \text{with probability } \frac{\alpha}{\alpha+t-1}, \\ \Sigma_j & \text{with probability } \frac{n_j}{\alpha+t-1}, \quad j = 1, \ldots, k_t, \end{cases} \tag{12}$$

where Σ_j, $j = 1, \ldots, k_t$ are the unique covariance matrices among the $\Lambda_{t'}$, $t' = 1, \ldots, t$, and k_t is the number of unique covariances.

Simply put, Eq. (12) says the more $\Lambda_{t'}$s belonging to the covariance cluster Σ_j, the larger n_j will be and the greater the probability of Λ_t belonging to the jth cluster. On the other hand, if only a few covariances have been assigned to a cluster, the selection of that particular covariance cluster is less likely. Also note that there is a non-trivial chance, proportional to α, of a new covariance cluster being selected from G_0. If $\alpha \to \infty$ no clustering occurs and every observations mixture covariances matrix is assigned its very own unique Σ_j, $j = 1, \ldots, n$. In this case, the ASV-DPM model's joint, return, log-volatility distribution is a multivariate Student-t distribution. Extreme clustering occurs when the precision parameter $\alpha \to 0$. The ASV-DPM model is a parametric model in the same sense as the ASV model of Yu (2005), Jacquier et al. (2004), and Omori et al. (2007) whose innovations to returns and log-volatility are conditionally normal with an unknown covariance that is modeled with an Inverse-Wishart prior.

2.2 Orthogonal Representation

We can write the ASV-DPM model in terms of orthogonal innovations by first defining the latent assignment variable $s_t = j$ when Λ_t equals the jth unique covariance Σ_j; i.e., when $\Lambda_t = \Sigma_j$ then $s_t = j$. Under the DP prior s_t is distributed

$$s_t \sim \sum_{j=1}^{\infty} \pi_j \mathcal{I}_j,$$

where the probability weights, π_j, $j = 1, \ldots,$ are the same as those defined in Eq. (7).

Incorporating s_t into the definition of the ASV-DPM model, we arrive at

$$y_t = \mu + \sigma_{y,s_t} \exp\{h_t/2\} u_t, \tag{13}$$

$$h_{t+1} = \delta h_t + \sigma_{h,s_t} v_t, \tag{14}$$

$$\left. \begin{pmatrix} u_t \\ v_t \end{pmatrix} \right| s_t, \Upsilon_{s_t} \sim N(\mathbf{0}, \Upsilon_{s_t}), \tag{15}$$

$$s_t \sim \sum_{j=1}^{\infty} \pi_j \mathcal{I}_j, \tag{16}$$

$$\Sigma_{s_t} \sim G_0 \tag{17}$$

where the covariance matrix $\Upsilon_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix}$, with $\rho_j = \sigma_{hy,j}/(\sigma_{h,j}\sigma_{y,j})$.

Letting $\Upsilon_{s_t}^{1/2}$ represent the Cholesky decomposition, $\Upsilon_{s_t} \equiv \Upsilon_{s_t}^{1/2}\Upsilon_{s_t}^{1/2\prime}$, we pre-multiply $(u_t, v_t)'$ by the inverse of $\Upsilon_{s_t}^{1/2}$ to obtain the uncorrelated innovation vector

$$\begin{pmatrix} w_t \\ v_t \end{pmatrix} \equiv \left(\Upsilon_{s_t}^{1/2}\right)^{-1} \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} (u_t - v_t\rho_{s_t})/\sqrt{1-\rho_{s_t}^2} \\ v_t \end{pmatrix}.$$

Solving for u_t in terms of w_t and substituting this into Eq. (13), the ASV-DPM model in terms of the orthogonal shocks $(w_t, v_t)'$ equals:

$$y_t = \mu + \sigma_{y,s_t}\exp\{h_t/2\}\rho_{s_t}v_t + \sigma_{y,s_t}\exp\{h_t/2\}\sqrt{1-\rho_{s_t}^2}w_t \tag{18}$$

$$h_{t+1} = \delta h_t + \sigma_{h,s_t}v_t. \tag{19}$$

2.3 Continuous Representation

Conditional on s_t and Σ_{s_t}, Eq. (18)-(19) is the Euler approximation to the stochastic differential equations

$$\begin{pmatrix} dS_t \\ dh_t \end{pmatrix} = \begin{pmatrix} \mu \\ (\delta-1)h_t \end{pmatrix} dt + \begin{pmatrix} \sigma_{y,s_t}\exp\{h_t/2\} & 0 \\ \sigma_{h,s_t}\rho_{s_t} & \sigma_{h,s_t}\sqrt{1-\rho_{s_t}^2} \end{pmatrix} dW_t, v \tag{20}$$

where S_t is the log price of the asset. The innovation vector, $W_t = (W_{S,t}, W_{h,t})'$, is standard Brownian motion in $\mathbb{R}^2$ with $\text{Corr}[dW_{S,t}, dW_{h,t}] = 0$.

Unlike the traditional parametric asymmetric stochastic volatility model, the ASV-DPM model's stochastic differential equation for the log-price process, $dS_t = \mu dt + \sigma_{y,s_t}\exp\{h_t/2\}dW_{S,t}$, is comprised of two sources of time-varying conditional volatility. The first is $\exp\{h_t/2\}$ - the stochastic volatility model's original source of time-varying conditional variance whose persistent causes the impact of a volatility shock to slowly die out. The other source of time-varying behavior is the second-order, random effect, associated with the variance of the price innovations, σ_{y,s_t}^2. This instantaneous variance inherits the properties of the DP prior through Σ_{s_t} and s_t.

The impact this second source of time-varying variation has on asset returns can also be understood by rewriting the differential equation for the log-price process in terms of the leverage effect

$$dS_t = \mu dt + \exp\{h_t/2\}\left(\sqrt{1-\rho_{s_t}^2}\sigma_{y,s_t}dW_{S,t} + \rho_{s_t}dW_{h,t}\right)$$

where the innovation to volatility, $dW_{h,t}$, impacts returns through the correlation measure ρ_{s_t}. If ρ_{s_t} is negative the leverage effect is present, causing positive shocks to volatility to skew dS_t to the left - a desirable property given the empirical regularity of large negative market returns occurring during volatile periods in equity markets. If, however, ρ_{s_t} is the same magnitude as the leverage effect, but instead of being negative, is positive, dS_t will be skewed to the right. As for kurtosis, the larger the magnitude of ρ_{s_t}, the greater is the excess kurtosis in the continuously compounded return process.

Because the DP prior leads to clustering, it is unlikely σ_{y,s_t} and ρ_{s_t} take on a large number of different values. It is more likely σ_{y,s_t} and ρ_{s_t} will take on a few unique combinations. Those Λ_ts corresponding in time to similar market behavior may be assigned the same unique covariance matrix, Σ_j. During periods of large negative return shocks, it could be imagined that their unique cluster has values of $\sigma^2_{y,j}$ and $\sigma^2_{h,j}$ that are larger, and a ρ_j that is more negative, than the random effects matrix associated with periods of market tranquility. With such a cluster returns would be more skewed to the left during a market crash. Occurrences of large positive market returns may have their own unique second-order, random effect, covariance matrix. One possibly exhibiting a positive ρ_j that is smaller in magnitude so as to capture slowly increasing levels of future volatility following a positive market return shock.

It is also possible that a large positive return shock will increase the level of uncertainty just as a large negative shock is expected too. Both large positive and negative types of shocks are deviation from the current state and cause traders to reevaluate their understanding of the market. If this the case, negative correlation between volatility and returns will dampen the effect a large positive shock has on volatility - diminishing volatility. The flexibility of the ASV-DPM model allows for a unique positive correlation in such circumstances, enabling volatility to increase after a large return shock.

3 Estimating the ASV-DPM Model

In this section we provide a likelihood-based approach to making parameter inference and model comparison with the ASV-DPM model by using a Markov chain Monte Carlo (MCMC) sampler. The MCMC sampler has a number of advantages. Along with providing parameter estimates, the MCMC sampler also estimates the latent volatilities and integrates out the uncertainty of the latent mixture variables from the DPM prior.

Let $y = (y_1, \ldots, y_n)'$ be the observed asset returns and $h = (h_1, \ldots, h_n)'$ the vector of its unobserved log-volatilities. The ASV-DPM posterior distribution

$$\pi(\mu, \delta, h, \{\Lambda_t\}, \alpha|y) \propto f(y|\mu, \delta, h, \{\Lambda_t\}, \alpha)\pi(\mu)\pi(h|\delta)\pi(\delta)\pi(\{\Lambda_t\}|\alpha)\pi(\alpha)$$

does not have a closed form. As a result, we strategically group the unknown parameters, latent volatilities, and mixture order, identities and assignments into manageable blocks where the selected blocks conditional posterior distributions are either known or have a tractable form. A Markov chain is then constructed by iteratively sampling through each block's posterior distribution conditioning on the value of the other parameters and latent variables drawn earlier.

The blocks of conditional distributions are:

- $\pi(\{\Lambda_t\}|y, h, \mu, \delta, \alpha)$
- $\pi(h|y, \mu, \delta, \{\Lambda_t\})$
- $\pi(\delta|y, h, \mu, \{\Lambda_t\})$
- $\pi(\mu|y, h, \{\Lambda_t\})$
- $\pi(\alpha|\{\Lambda_t\})$

3.1 Λ_t sampler

Inference of the ASV-DPM models $\pi(\{\Lambda_t\}|y, h, \mu, \delta, \alpha)$ can be carried out by employing a Polya urn type Gibbs sampler of Escobar (1994). The Polya urn approach sequentially samples Λ_t, for $t = 1, \ldots, n$, from the conditional posterior distribution. More formally, let $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$, where $\boldsymbol{z}_t = ((y_t - \mu)\exp\{-h_t/2\}, h_{t+1} - \delta h_t)'$, and draw Λ_t from the distribution:

$$\begin{aligned}\Lambda_t|\{\Lambda_{t'} : t' \neq t\}, \boldsymbol{z}_t, \alpha \quad \sim \quad & \frac{\alpha}{\alpha + n - 1} g(\boldsymbol{z}_t) G(d\Lambda|\boldsymbol{z}_t) \\ & + \frac{1}{\alpha + n - 1} \sum_{t' \neq t} f_N(\boldsymbol{z}_t|\Lambda_{t'}) \mathcal{I}_{\Lambda_{t'}}(d\Lambda), \end{aligned} \tag{21}$$

where $g(\boldsymbol{z}_t) \equiv \int f_N(\boldsymbol{z}_t|\mathbf{0}, \Lambda)\ G_0(\Lambda) d\Lambda$, and by the law of conditional probability, $G(d\Lambda|\boldsymbol{z}_t) \propto f_N(\boldsymbol{z}_t|\mathbf{0}, \Lambda) G_0(d\Lambda)$. Applying the prior information of Eq. (3) and (9) it follows that the density function of $G(d\Lambda|\boldsymbol{z}_t)$:

$$\begin{aligned} g(\Lambda|\boldsymbol{z}_t) \quad & \propto \quad |\Lambda|^{-1/2} \exp\left\{-\frac{1}{2}\mathrm{tr} \boldsymbol{z}_t \boldsymbol{z}_t' \Lambda^{-1}\right\} \frac{|S_0|^{v_0/2}}{|\Lambda|^{(v_0+3)/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\Lambda^{-1} S_0\right\}, \\ & = \quad \frac{|S_0|^{v_0/2}}{|\Lambda|^{(v_0+4)/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(S_0 + \boldsymbol{z}_t \boldsymbol{z}_t')\Lambda^{-1}\right\}, \end{aligned} \tag{22}$$

is the kernel to a Inverse-Wishart distribution; i.e., $G(d\Lambda|\boldsymbol{z}_t) \equiv \text{Inv-Wish}(S_0 + \boldsymbol{z}_t \boldsymbol{z}_t', v_0 + 1)$ (see Zellner, 1987, p.395). Integrating out the Λ from the Inverse-Wishart distribution is the marginal likelihood $g(\boldsymbol{z}_t)$, which equals the density of a bivariate Student-t distribution:

$$g(\boldsymbol{z}_t) \quad = \quad f_{MSt}(\boldsymbol{z}_t|\mathbf{0}, (S_0/(v_0 - 1))^{-1}, v_0 - 1), \tag{23}$$

with $v_0 - 1$ degrees of freedom, mean-zero vector, and covariance, $S_0/(v_0-3)$ (see Zellner 1987, Eq. (B.20), p. 383 for the exact formula of f_{MSt}).

The more efficient Poly urn approach of West et al. (1994) and MacEachern and Müller (1998) can be applied to the ASV-DPM model to generate draws from a distribution equivalent to $\pi(\{\Lambda_t\}|y,h,\mu,\delta,\alpha)$. This distribution is $\pi(\Sigma_1,\ldots,\Sigma_k,s|y,h,\mu,\delta,\alpha)$ where the Σ_j, $j=1,\ldots,k$, $k\le n$, are the distinct Λ_t, $t=1,\ldots,n$, and $s=(s_1,\ldots,s_n)'$ is the vector consisting of the assignment variables, s_t; i.e., $s_t=j$ when $\Lambda_t=\Sigma_j$.

Define n_j to be the number of observations where $s_t=j$, $k^{(t)}$ to be the distinct number of Σ_j in $\{\Lambda_{t'}:t'\neq t\}$ and $n_j^{(t)}$ the number of observations where $s_{t'}=j$, $t'\neq t$. For a given h, μ, δ and α, draws from the posterior $\pi(\{\Sigma_j\},s|\boldsymbol{z},\alpha)\equiv\pi(\{\Sigma_j\},s|y,h,\mu,\delta,\alpha)$ are made with the following 2-step algorithm where:

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

$$s_t|\{\Lambda_{t'}:t'\neq t\},z_t,\alpha \quad \sim \quad \begin{cases} \frac{\alpha}{\alpha+n-1} g(\boldsymbol{z}_t)\,\mathcal{I}_0(ds_t) \\ \frac{1}{\alpha+n-1}\sum_{j=1}^{k^{(t)}} n_j^{(t)} f_N(\boldsymbol{z}_t|\mathbf{0},\Sigma_j)\,\mathcal{I}_j(ds_t). \end{cases} \tag{24}$$

   If $s_t=0$, we then draw a new Σ_j from the Inverse-Wishart distribution in Eq. (22), increase k by one and set s_t equal to the new k. Otherwise, we set s_t equal to the randomly drawn j and leave k unchanged.

2. Discard the Σ_js from Step 1 and use the s and k to iteratively draw new Σ_j, for $j=1,\ldots,k$, from:

$$\pi(\Sigma_j|\boldsymbol{z},s,k) \propto \prod_{t:s_t=j} f_N(\boldsymbol{z}_t|\mathbf{0},\Sigma_j)G_0(d\Sigma) \tag{25}$$

$$\propto \prod_{t:s_t=j} |\Sigma_j|^{-1/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\boldsymbol{z}_t\boldsymbol{z}_t'\Sigma_j^{-1}\right\} \times\frac{|S_0|^{v_0/2}}{|\Sigma_j|^{(v_0+3)/2}}\exp\left\{-\frac{1}{2}\mathrm{tr}S_0\Sigma_j^{-1}\right\} \tag{26}$$

$$= \frac{|S_0|^{v_0/2}}{|\Sigma_j|^{(v_0+n_j+3)/2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\sum_{t:s_t=j}\boldsymbol{z}_t\boldsymbol{z}_t'+S_0\right)\Sigma_j^{-1}\right\} \tag{27}$$

$$\sim \text{Inv-Wish}\left(S_0+\sum_{t:s_t=j}\boldsymbol{z}_t\boldsymbol{z}_t', v_0+n_j\right). \tag{28}$$

Breaking up the draws of the assignment variables s_t from the draws of the identities of the random effects, Σ_j, reduces the inherent dependency that exists when sampling the

DPM covariances from the existing set of covariances. This helps the randomly drawn covariances to span the entire support of the posterior distribution.

3.2 Latent volatility sampler

Given the heightened correlation that exists between the log-volatilities, we propose a efficient tailored, Metropolis-Hasting sampler of randomly drawn blocks of h. Volatility draws are made by forming random partitions of h where the length of each subvector in the partition is equal to a random draw from a Poisson distribution. Randomly lengthed blocks promote mixing from sweep to sweep by ensuring that volatilities are not drawn conditionally on time adjacent volatilities where their time position is fixed.

Given a particular partition of h one sequentially draws each volatility block conditional on the value of the other volatilities. The conditional distribution of a volatility block $h_{(t',\tau)} = (h_{t'}, h_{t'+1}, \ldots, h_\tau)'$, where $1 \leq t' \leq \tau < n$, and the vector length $l_{t'} = \tau - t' + 1$ is equal to the random draw from $l_{t'} \sim \text{Pois}(\lambda_h)$, is:

$$\begin{aligned} \pi(h_{(t',\tau)}|y_{(t',\tau)}, h_{-(t',\tau)}) &\propto \pi(h_{(t',\tau)}|h_{\tau+1}, h_{t'-1}) f(y_{(t',\tau)}|h_{(t',\tau)}, h_{\tau+1}) \\ &= \prod_{t=t'}^{\tau} \pi(h_t|h_{t+1}, h_{t-1}) f(y_t|h_t, h_{t+1}) \\ &= \prod_{t=t'}^{\tau} \left[\exp\left\{ -\frac{1}{2} \left[\frac{(h_{t+1} - \delta h_t)^2}{\sigma^2_{h,s_t}} + \frac{(h_t - \delta h_{t-1})^2}{\sigma^2_{h,s_{t-1}}} + h_t \right] \right\} \right. \\ &\quad \left. \times \exp\left\{ -\frac{\left(y_t - \mu - e^{h_t/2}\sigma_{y,s_t}\rho_{s_t}(h_{t+1} - \delta h_t)/\sigma_{h,s_t}\right)^2}{2(1-\rho^2_{s_t})\exp\{h_t\}\sigma^2_{y,s_t}} \right\} \right] \end{aligned} \tag{29}$$

When the conditional distribution is for the first block, $\pi(h_{(1,\tau)}|h_{\tau+1}, h_0)$ depends on h_0. We choose to model h_0 with the prior $\pi(h_0) \sim N(0, \sigma^2_{h,0}/(1-\delta^2)$ where $\sigma^2_{h,0} \equiv E[G_0(d\Lambda)]_{2,2}$ is the expected variance of log-volatility from the the DPM base distribution. Drawing h_0 from $\pi(h_0)$, we numerically integrate out h_0 from the draws of $h_{(1,\tau)}$.

If the draw of $l_{t'}$ were to cause τ to be greater than or equal to n, the volatility blocks conditional distribution is the same as above except $\tau = n$. For this last block in the partition on h we integrate out the one period ahead, out of sample volatility, h_{n+1}, by replacing it with a random draw from:

$$h_{n+1}|y_n, h_n, \Sigma_{s_n} \sim N\left(\delta h_n + \sigma_{h,s_n}\rho_{s_n} y_n \exp\{-h_n/2\}/\sigma_{y,s_n}, \sigma^2_{h,s_n}(1-\rho^2_{s_n})\right).$$

where the value of h_n is from the previous sweep of the sampler.

Since the conditional distributions of Eq. (29) are nonstandard we use a Metropolis-Hasting sampler (see Chib and Greenberg (1995)). Candidate draws of $h_{(t,\tau)}$ are made with a l_t-variate Student-t distribution with mean vector $\boldsymbol{m}$, covariance matrix, S, and ξ degrees of freedom where $\boldsymbol{m}$ is the argument maximizing $\pi(h_{(t',\tau)}|h_{\tau+1}, h_{t'-1})f(y_{(t',\tau)}|h_{(t',\tau)}, h_{\tau+1})$ and S is the negative inverted Hessian evaluated at $\boldsymbol{m}$.

3.3 Sampler of δ

We assume a prior for δ equal to the truncated normal distribution, $\pi(\delta) \propto N(\mu_\delta, \sigma_\delta^2)I_{|\delta|<1}$. Under this prior, we sample from $\pi(\delta|y, h, \{\Lambda_t\})$ by carrying out a Metropolis-Hasting draw where the candidate draw δ' is made from the $N(\widehat{\delta}, \widehat{\sigma}_\delta^2)$ distribution where:

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

and is accepted with probability $\alpha(\delta', \delta) = \min\left\{ \frac{g(\delta')}{g(\delta)} \frac{f_N(\delta|\widehat{\delta}, \widehat{\sigma}_\delta^2)}{f_N(\delta'|\widehat{\delta}, \widehat{\sigma}_\delta^2)}, 1 \right\}$ where

$$\begin{aligned} g(\delta) &\propto f(y|\delta, h, \{\Sigma_j\}, s, h_{n+1})\pi(h|\delta, \{\Sigma_j\}, s)\pi(h_{n+1}|h_n, \delta)\pi(h_0|\delta)\pi(\delta), \\ &= \prod_{t=1}^{n} \exp\left\{ -\frac{1}{2} \left[\frac{\left(y_t - \mu - \rho_t e^{h_t/2}\sigma_{y,s_t}(h_{t+1} - \delta h_t)/\sigma_{h,s_t}\right)^2}{(1-\rho_{s_t}^2)\sigma_{y,s_t}^2 e^{h_t}} + \frac{(h_{t+1} - \delta h_t)^2}{\sigma_{h,s_t}} \right] \right\} \\ &\quad \times \exp\left\{ -\frac{1}{2} \left[\frac{(h_1 - \delta h_0)^2}{\sigma_{h,0}^2} + \frac{h_0^2}{\sigma_{h,0}^2/(1-\delta^2)} \right] \right\} f_N(\delta|\mu_\delta, \sigma_\delta^2) I_{|\delta|<1}. \end{aligned}$$

For those draws when δ' is rejected, the previous sweep's draw of δ is kept as the current sweeps draw from $\pi(\delta|y, h, \{\Lambda_t\})$.

3.4 Sampler of μ

To perform draws from $\pi(\mu|y, h, \{\Lambda_t\})$ we let $\pi(\mu) \sim N(m, \tau)$. Since $\pi(\mu)$ is a conjugate prior, draws of μ are made from $N(\hat{\mu}, \hat{\tau})$ where:

$$\hat{\mu} = \hat{\tau} \left(\frac{m}{\tau} + \sum_{t=1}^{n} \frac{\tilde{y}_t}{\tilde{\sigma}_t^2} \right), \qquad \hat{\tau} = \left(\frac{1}{\tau} + \sum_{t=1}^{n} \frac{1}{\tilde{\sigma}_t^2} \right)^{-1},$$

and $\tilde{y} = y_t - \rho_t e^{h_t/2}\sigma_{y,s_t}(h_{t+1} - \delta h_t)/\sigma_{h,s_t}$ and $\tilde{\sigma}_t^2 = (1-\rho_{s_t}^2)\sigma_{y,s_t}^2 \exp\{h_t\}$.

3.5 Sampler of α

The two step algorithm of Escobar and West (1995) is used to sample the ASV-DPM model's precision parameter α. When the mixture order, k, identifying vector, s, and locations $\{\Sigma_j\}$, are all known, the posterior of α will only depend on k. Assuming a gamma prior, $\Gamma(a, b)$, where $a > 0$ and $b > 0$, for α, draws from $\pi(\alpha|k)$ can be made by

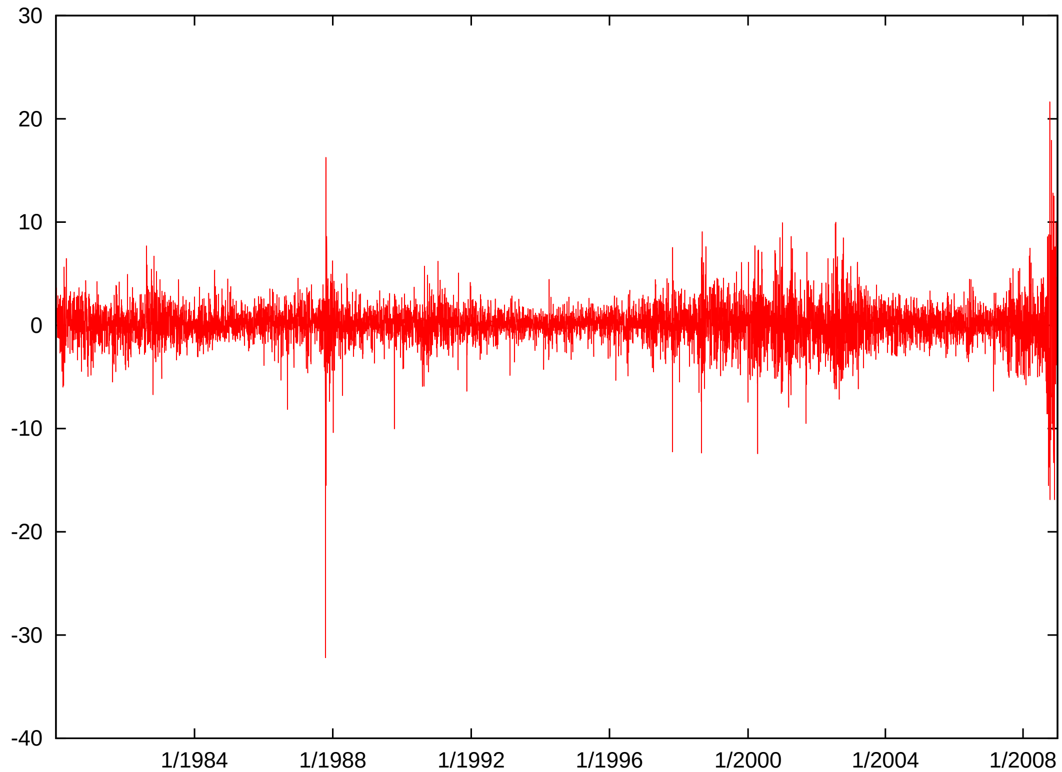
1. Sampling the random variable ξ from $\pi(\xi|\alpha, k) \sim Beta(\alpha + 1, n)$

2. Sampling α from the mixture $\pi(\alpha|\xi, k) \sim \pi_\xi \Gamma(a+k, b-\ln\xi)+(1-\pi_\xi)\Gamma(a+k-1, b-\ln\xi)$, where $\pi_\xi/(1-\pi_\xi) = (a+k-1)/[n(b-\ln\xi)]$.

4 Empirical Results

We estimate the ASV-DPM model using 7319 daily returns (multiplied by a 100) over the period of January 2, 1980 to December 31, 2008 from the Center of Research in Security Prices (CRSP) value-weighted portfolio index. In Figure 1 we plot the value-weighted portfolio returns. This time period is ideal since returns exhibit a number of different dynamics, for example, the pre- and post-1987 market crash periods, the tech bubble of the late 90s, and the financial crisis of 2008. Over the entire sample returns average 0.045 and have a variance of 1.12. Returns appear to be asymmetrically distributed with fat-tails as is evident in their negative skewness of -0.757, and a excess kurtosis measure of 19.296.

Figure 1: CRSP value-weighted portfolio returns from January 2, 1980 to December 31, 2008.



To benchmark the results for the ASV-DPM model we also apply the parametric asymmetric, stochastic volatility model (ASV) of Harvey et al. (1994) to the CRSP value-weighted portfolio returns. The ASV model equals

$$\begin{aligned} y_t &= \mu + \exp\{h_t/2\}\epsilon_t && (30) \\ h_{t+1} &= \delta h_t + \eta_t && (31) \\ \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} &\sim N(\mathbf{0}, \Sigma). && (32) \end{aligned}$$

Except for the distribution of the innovations $(\epsilon_t, \eta_t)'$, which is normally distributed with a constant covariance Σ, the ASV model has the same form as the ASV-DPM model in Eq. (1)-(2).

The priors for the ASV-DPM model are $\pi(\mu) \equiv N(0, 0.1)$ and $\pi(\delta) \equiv N(0, 100)I_{|\delta|<1}$. For the DPM, we let the base distribution, $G_0 \equiv \text{Inv-Wish}(S_0, v_0)$ with $S_0 = \boldsymbol{I}_2$ and $v_0 = 10$. The prior for the DPM precision parameter is $\pi(\alpha) = \text{Gamma}(2, 8)$ so that $E[\alpha] = 1/4$ and $\text{Var}[\alpha] = 1/32$. For the ASV model, the priors will be the same as those for the ASV-DPM. This means the prior for Σ is also G_0. Estimation of the ASV model will follow the sampling algorithm spelled out in Section 3 for the ASV-DPM model, but with k and s_t, $t = 1, \ldots, n$, set equal to 1.

To reduce the influence the starting values will have on the MCMC sampler, we first perform 1000 sweeps over the log-volatilities using the step-by-step volatility sampler of Kim et al. (1998) while holding the other parameters constant. We then let the entire sampler of Section 3 iterate 40,000 times and keep only the last 30,000 draws from the two models for inference purposes.

In Table 1 we report the posterior mean, standard deviation, and 95% probability interval for the structural parameters of the ASV-DPM and ASV models. For the parameters common to both models, the posterior mean in the ASV-DPM model of the unconditional mean of return, μ, is slightly larger at 0.063 than the ASV models 0.048. But the posterior standard deviations of μ are nearly the same at 0.008 for the two models. This similarity in the posterior distribution of μ is also evident in the two models 95% posterior probability intervals. The difference in the 2.5% and 97.5% percentiles for μ in the two models are similar to the difference in the two models posterior mean of μ. Hence, the posterior distribution of μ has the same shape for the two models but is shifted slightly to the left for the ASV model.

Persistence in the two model's log-volatilities, as captured by the posterior distribution of the autoregressive parameter, δ, is essentially the same. For the ASV-DPM model, the

Table 1: Posterior estimates of the ASV-DPM and ASV models for daily compounded CRSP value-weighted portfolio returns from January 2, 1980 to December 29, 2008 (7319 observations, 41,000 draws with the first 1000 draws of log-volatility followed by the next 10,000 draws of all the unknowns being discarded).

	ASV-DPM			ASV		
	mean	stdev	95% prob interval	mean	stdev	95% prob interval
α	0.3729	0.1820	(0.1050, 0.7997)			
k	5.3158	1.6914	(3, 9)			
δ	0.9799	0.0035	(0.9728, 0.9865)	0.9793	0.0033	(0.9725, 0.9854)
μ	0.0630	0.0083	(0.0468, 0.0792)	0.0482	0.0085	(0.0315, 0.0648)
σ_y^2				0.7096	0.0644	(0.5931, 0.8459)
σ_h^2				0.0332	0.0037	(0.0268, 0.0412)
ρ				-0.5243	0.0324	(-0.5850,-0.4585)

posterior mean of δ is 0.9799, whereas, in the ASV model it equals 0.9793. The standard deviations of δ's posterior distribution are also nearly identical with the ASV-DPM model's equaling 0.0035 and the ASV being 0.0033. This suggests that regardless of the distribution for the ASV model, log-volatility continues to cluster together with large fluctuations in volatility following large fluctuations and small fluctuations following small ones. The magnitude of δ is also evidence of a strong level of persistence in volatility where the impact of a volatility shock lives on for a very long time.

We can infer from the ASV-DPM posterior mean of k that there are close to six different second-order effects. Because of the nature of the DPM sampler in Section 3.1, it is not possible to identify where these occur in the return series since the elements of indicator vector s from sweep to sweep may be different. For example, suppose in one sweep of the MCMC sampler a particular effect is labeled j. There is no restriction that this effect will be labeled as the j'-cluster, where $j \neq j$ in another sweep. However, the posterior estimate of k is evidence in favor of a non-Gaussian distribution for the innovation vector of returns and log-volatility.

4.1 Model Comparison & Predictive Power

To compare our ASV-DPM model with the ASV model, and more generally other stochastic volatility models, we need to compute each models marginal likelihood. In addition to the marginal likelihood integrating out parameter uncertainty, a stochastic volatility model's marginal likelihood requires integrating out the uncertainty associated with the latent volatilities. Particle filters have been applied in this regards by Chib et al. (2002). The marginal likelihood for the ASV-DPM model also requires integrating out the latent DP covariance matrices. Basu and Chib (2003) have a method for this but only for a DPM type model, not a DPM model with stochastic volatility. The ASV-DPM model would require a particle filter for the integrating out the latent volatilities.

Because of the additional computation costs involved with integrating out the volatilities and DPM covariances, and also because of the increased availability of parallel computing, clusters, and quad-core processors, we do not use a particle filter approach. Instead we compute a SV model's marginal likelihood sequentially with its one-step-ahead predictive likelihoods. Given the low cost of multi-thread computing and availability of multiple processors, a large number of individual and independent MCMC draws can be conducted on unique filters of the return series. For the ASV-DPM and ASV models we carry out 50 separate and unique MCMC samplers simultaneous on 5 servers each possessing two quad-core processors. As a result the marginal likelihood is computed in less than a day.

Our approach is as follows. Let the vectors $y^{t-1} = (y_1, \ldots, y_{t-1})'$, where $t = 2, \ldots, n$, denote the histories of returns up to time period $t-1$. By the law of conditional probability, the marginal likelihood can be expressed in terms of the one-step-ahead predictive likelihoods

$$m(y|M) = \prod_{t=1}^{n} f(y_t|y^{t-1}, M), \tag{33}$$

where M = ASV-DPM, ASV denotes the particular model. In practice, the product of one-step-ahead predictive likelihood in Eq. (33) does not begin at $t = 1$, since the model's priors will overly influence the predictive likelihoods having a short history. Instead, in our empirical calculations we use $m_L(y|M) = \prod_{t=L}^{n} f(y_t|y^{t-1}, M)$, where $L = 100$.

The marginal likelihood in Eq. (33) is a recursive method similar to the cumulative sum of the recursive residuals of Brown et al. (1975), except where the instability associated with the parameters is integrated out. Each one-step-ahead predictive likelihood is approximated by averaging out the randomly sampled draws of the unknown parameters, μ and δ, volatilities, h^{t-1}, and the DPM order, indicator vector, locations, and precision from their posterior conditional on the data history y^{t-1}. For the ASV-DPM model the predictive likelihood equals

$$\begin{aligned} f\left(y_t|y^{t-1}, \text{ASV-DPM}\right) &= \int f(y_t|h_t, \mu, \delta, \{\Lambda\}, \alpha)\pi(h_t, \mu, \delta, \{\Lambda\}, \alpha|y^{t-1})\, d(h_t, \mu, \delta, \{\Lambda\}, \alpha), \\ &\approx R^{-1} \sum_{l=1}^{R} f\left(y_t \,\middle|\, \mu^{(l)}, h_t^{(l)}, \left\{\sigma_{y,j}^{2(l)}, n_j^{(l)}\right\}_{j=1}^{k^{(l)}}, \alpha^{(l)}\right), \end{aligned} \tag{34}$$

where

$$f\left(y_t\left|\mu^{(l)}, h_t^{(l)}, \left\{\sigma_{y,j}^{2(l)}, n_j^{(l)}\right\}_{j=1}^{k^{(l)}}, \alpha^{(l)}\right.\right) = \frac{\alpha^{(l)}}{\alpha^{(l)}+t-1} f_{St}\left(y_t\left|\mu^{(l)}, \left(\frac{s_{11}e^{h_t^{(l)}}}{v_0-1}\right)^{-1}, v_0-1\right.\right)$$
$$+\frac{1}{\alpha^{(l)}+t-1}\sum_{j=1}^{k^{(l)}} n_j^{(l)} f_N\left(y_t\left|\mu^{(l)}, e^{h_t^{(l)}}\sigma_{y,j}^{2(l)}\right.\right), \quad (35)$$

with the $h_t^{(l)}$s being random draws from the normal distribution, $N(\widehat{h}_{t|t-1}, \widehat{\sigma}^2_{t|t-1})$, where

$$\widehat{h}_{t|t-1} = \delta^{(l)}h_{t-1}^{(l)} + \sigma_{h,s_{t-1}}^{(l)}\rho_{s_{t-1}}^{(l)}\left(y_{t-1}-\mu^{(l)}\right)\exp\left\{-h_{t-1}^{(l)}/2\right\}\Big/\sigma_{y,s_{t-1}}^{(l)}, \quad (36)$$

$$\widehat{\sigma}_{t|t-1} = \sigma_{h,s_{t-1}}^{2(l)}\left(1-\rho_{s_{t-1}}^{2(l)}\right), \quad (37)$$

and $\mu^{(l)}, h_t^{(l)}, \sigma_{y,s_{t-1}}^{2(l)}, \sigma_{h,s_{t-1}}^{2(l)}, \rho_{s_{t-1}}^{2(l)}$ are lth posterior draw from the MCMC sampler of Section 3.

Each of the $n-1$ MCMC posterior conditional samplers is independent from the others. Given this independence we need only to supply a particular sampler with one of the $n-1$ histories y^{t-1} before letting it run. First, we farm out as many unique histories as there are processors available. On a quad-core, multi-threaded, computer this generally equals 10 potential MCMC samplers with unique histories. When a processor's task of sampling R draws from the posterior distribution conditioned on that history is complete, the processor computes the one-step-ahead predictive likelihood in Eq. (35) from these draws and returns it for later use. If there are predictive likelihoods for histories still needing to be computed, the processor will request another uniquely lengthed history and sample from its posterior. Once all the predictive likelihoods have been computed the marginal likelihood is calculated.

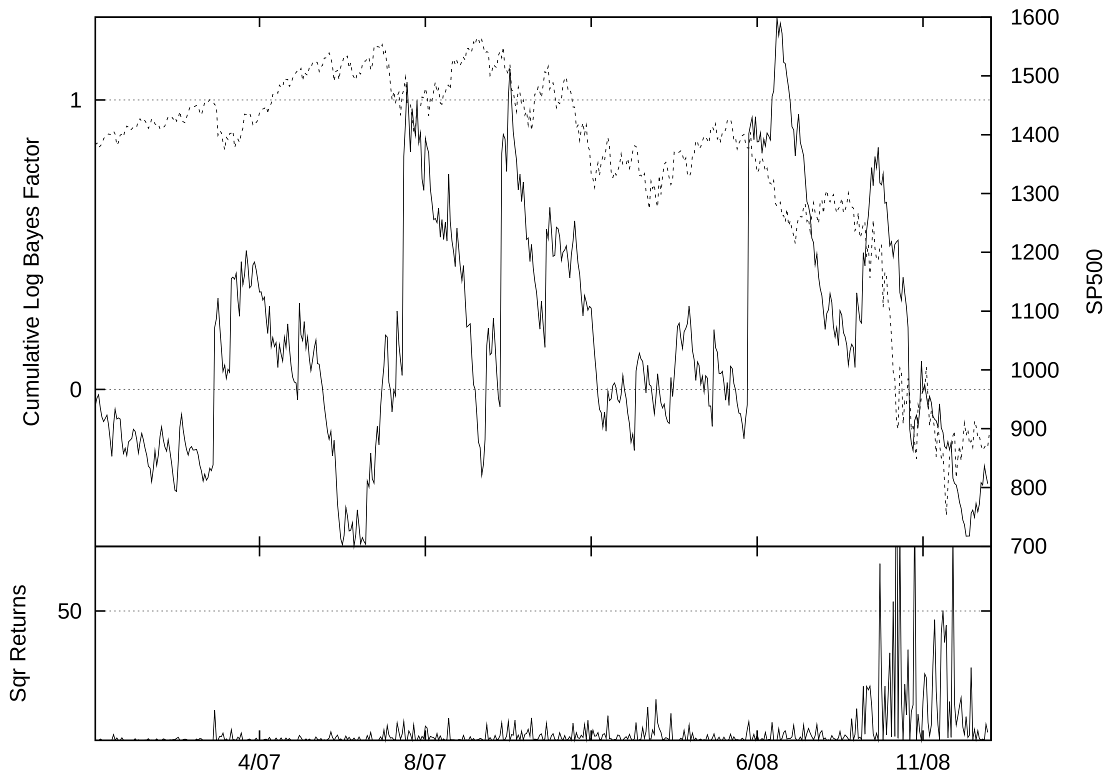
Given $f(y_t|y^{t-1}, \text{ASV-DPM})$ and $f(y_t|y^{t-1}, \text{ASV})$, for $t = L, \ldots, n$, we can find the cumulative Bayes factors

$$\frac{m_L(y^\tau|\text{ASV-DPM})}{m_L(y^\tau|\text{ASV})} = \prod_{t=L}^{\tau} \frac{f(y_t|y^{t-1}, \text{ASV-DPM})}{f(y_t|y^{t-1}, \text{ASV})}, \qquad \tau = L, \ldots, n. \quad (38)$$

By plotting the cumulative Bayes factors in Eq. (38) we can identify those days where one model out preforms the other in predicting the next days return. The Bayes factor, $m_L(y|\text{ASV-DPM})/m_L(y|\text{ASV})$, will thus favor the model possessing better predictive power.

We investigate the cumulative Bayes factors of the ASV-DPM and ASV models for the time period from January 3, 2006 to December 31, 2008, using the CRSP value-weighted portfolio returns over the above 755 trading days. For the two SV models we individually compute the 540, one-day-ahead predictive likelihoods, $f(y_t|y^{t-1}, M)$, $t =$

Figure 2: Cumulative log Bayes factor of the ASV-DPM model relative to the ASV model (solid line) for the period of November 8, 2006 to December 31, 2008 (using returns starting on January 3, 2006) and the S&P 500 (dotted line) plotted above the squared value-weighted portfolio return.



216 (Nov. 8, 2006), . . . , 755 (Dec 31, 2008), by sampling the model's unknowns 11,000 times and discarding the first 1,000 draws.

In the top panel of Figure 2 we plot the cumulative log Bayes (CLB) factors from November 8, 2006 to December 31, 2008, along with the S&P 500 market index from November 8, 2006 to December 31, 2008. In the bottom panel we provide a proxy for the level of volatility by plotting the squared CRSP value-weighted portfolio returns.

The end point of the cumulative log Bayes factor on the right hand side of the figure equals the log Bayes factor for the ASV-DPM model versus the ASV model for the time period of January 3, 2006 to December 31, 2008. Since $\log[m(y|ASV - DPM)/m(y|ASV]$ at the endpoint equals -0.33, its Bayes factor of 1.4 in favor of the ASV over the ASV-DPM model is in Jefferies (1961) words "barely" worth mentioning.

There are sub-periods over the sample where the cumulative log Bayes factor is positive and in support of the ASV-DPM model. During times like August and October of 2007, and September 2008, the value of the cumulative log Bayes factor is slightly larger than one. These values for the CLB leads us to infer that up to these points in time there is, again in the words of Jefferies, "substantial" evidence in favor of the ASV-DPM model. In the spring of 2007 the cumulative log Bayes factor favors the ASV model. However, the accumulation seen in the predictive Bayes factor when favoring the ASV model (downward movement in the CLB line) is small relative to the sizable upward accumulations favoring the predictive likelihood of the ASV-DPM model.

Each period of upward accumulation in the CLB demonstrates a sharp, sizable increase with the largest occuring on the sixth of June, 2008. The other points of rapid accumulations that we focus on are February 27, 2007, August 9, 2007, and November 1, 2007.[3] In three of the four episodes, this increase in the predictive power of the ASV-DPM model corresponds to a increase in the S&P 500 index followed the next day by a sizable decline. For the episode on June 6, 2008, the market increased 2.1% the day before and then fell 2.7% on the sixth. The exception to this pattern is in the first occurrence on February 27, 2007. In that instance there is no initial increase in the market only a sizable decline of 3.4%.

These episodes of rapid accumulation in the CLB for the ASV-DPM model also line up with large increases in volatility. Until the rapid run up in the CLB, volatility is small but in each of the above instances volatility suddenly increases at the same instance as the jump in the CLB.

[3]There are days when the CLB rapidly declines in favor of the ASV model's log predictive likelihood, but nothing in size nor quickness as the ASV-DPM model.

5 Conclusion

In this paper we extended the asymmetric, stochastic, log-volatility model whose innovations are correlated and normally distributed by modeling the uncertainty in their joint distribution with a nonparametric, bi-variate Dirichlet process mixture prior. We provide a sampling algorithm to integrate out the parameter, volatility and distributional uncertainty of our semiparametric, asymmetric, stochastic volatility model. Our algorithm is also used to compute the log Bayes predictive forecast of the semiparametric model relative to the parametric version. These log Bayes predictions are used to evaluate and compare the forecasting abilities of the two models.

The nonparametric prior increases the flexibility of the asymmetric stochastic volatility model by allowing the correlation between its innovations to take on infinite number of values, but to also remain managable and parsimonious by taking on only a finite number of correlations values over a finite lengthed dataset. This flexibility is important when forecasting market returns, especially when the market has experienced a positive return shock, followed by a large negative one. In the empirical case study, forecasts from our semiparametric asymmetric stochastic volatility model capture the changes in the level of correlation occurring when returns experience a large positive shock, whereas the parametric model does not. This leads to the daily predictive Bayes factors favoring the nonparametric asymmetric stochastic volatility model more often than the parametric version.

References

Basu, S. and Chib, S.: 2003, Marginal likelihood and bayes factors for dirichlet process mixture models, *Journal of the American Statistical Association* **98**(461), 224–235.

Blackwell, D. and MacQueen, J.: 1973, Ferguson distributions via polya urn schemes, *The Annals of Statistics* **1**, 353–355.

Brown, R. L., Durbin, J. and Evans, J. M.: 1975, Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society. Series B (Methodological)* **37**(2), 149–192.

Chen, X. and Ghysels: 2011, News - good or bad - and its impact on volatility predictions over multiple horizons, *RFS* **24**, 46–81.

Chib, S. and Greenberg, E.: 1995, Understanding the metropolis-hastings algorithm, *American Statistician* **49**, 327–335.

Chib, S., Nardari, F. and Shephard, N.: 2002, Markov chain monte carlo methods for stochastic volatility models, *Journal of Econometrics* **108**, 281–316.

Das, S. R. and Sundaram, R. K.: 1999, Of smiles and smirks: A term structure perspective, *Journal of Financial and Quantitave Analysis* **34**(2), 211–239.

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. and West, M.: 1995, Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* **90**(430), 577–588.

Harvey, A., Ruiz, E. and Shephard, N.: 1994, Multivariate stochastic variance models, *The Review of Economic Studies* **61**(2), 247–264.

Jacquier, E., Polson, N. G. and Rossi, P. E.: 2004, Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* **122**, 185–212.

Jefferies, H.: 1961, *Theory of Probability*, Claredon Press.

Jensen, M. J. and Maheu, J. M.: 2010, Bayesian semiparametric stochastic volatility modeling, *Journal of Econometrics* **157**(2), 306 – 316.

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

Lo, A. Y.: 1984, On a class of bayesian nonparametric estimates. i. density estimates, *The Annals of Statistics* **12**, 351–357.

MacEachern, S. N. and Müller, P.: 1998, Estimating mixture of Dirichlet process models, *Journal of Computational and Graphical Statistics* **7**(2), 223–238.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J.: 2007, Stochastic volatility with leverage: Fast and efficient likelihood inference, *Journal of Econometrics* **140**(2), 425–449.

Sethuraman, J.: 1994, A constructive definition of dirichlet priors, *Statistica Sinica* **4**, 639–650.

West, M., Muller, P. and Escobar, M.: 1994, Hierarchical priors and mixture models with applications in regression and density estimation, *in* P. R. Freeman and 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.