Estimating a Semiparametric Asymmetric Stochastic Volatility Model with a Dirichlet Process Mixture[*]

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. In this paper we extend the parametric, asymmetric, stochastic volatility model (ASV), where returns are correlated with volatility, by flexibly modeling the bivariate distribution of the return and volatility innovations nonparametrically. Its novelty is in modeling the joint, conditional, return-volatility, distribution with a infinite mixture of bivariate Normal distributions with mean zero vectors, but having unknown mixture weights and covariance matrices. This semiparametric ASV model nests stochastic volatility models whose innovations are distributed as either Normal or Student-t distributions, plus the response in volatility to unexpected return shocks is more general than the fixed asymmetric response with the ASV model. The unknown mixture parameters are modeled with a Dirichlet Process prior. This prior ensures a parsimonious, finite, posterior, mixture that bests represents the distribution of the innovations and a straightforward sampler of the conditional posteriors. We develop a Bayesian Markov chain Monte Carlo sampler to fully characterize the parametric and distributional uncertainty. Nested model comparisons and out-of-sample predictions with the cumulative marginal-likelihoods, and one-day-ahead, predictive log-Bayes factors between the semiparametric and parametric versions of the ASV model shows the semiparametric model projecting more accurate empirical market returns. A major reason is how volatility responds to an unexpected market movement. When the market is tranquil, expected volatility reacts to a negative (positive) price shock by rising (initially declining, but then rising when the positive shock is large). However, when the market is volatile, the degree of asymmetry and the size of the response in expected volatility is muted. In other words, when times are good, no news is good news, but when times are bad, neither good nor bad news matters with regards to volatility.

Keywords: Bayesian nonparametrics, cumulative Bayes factor, Dirichlet process mixture, infinite mixture model, leverage effect, marginal likelihood, MCMC, non-normal, stochastic volatility, volatility-return relationship

1 Introduction

In this paper we extend the parametric, asymmetric, stochastic volatility model (ASV), where returns are correlated with volatility, by flexibly modeling the bivariate distribution of the return and volatility innovations nonparametrically.[1] In the model, log-volatility belongs to the parametric, stationary, first-order autoregressive class of stochastic volatility made popular by Jacquier et al. (1994) and Kim et al. (1998). The rest of the model is nonparametric in the sense that no assumptions are made about the underlying joint distribution of returns and volatility. Instead, the flexible Dirichlet process mixture (DPM) class of priors by Lo (1984) along with return data is used to estimate the unknown distribution.

The version of the DPM used is an infinite mixture of bivariate, normal distributions with mean zero, but unknown covariance matrices and mixture probabilities. This is used to fit the return and log-volatility distribution. As a mixture, each observations covariance matrix is distributed according to a Dirichlet process prior – a nonparametric prior over the value of the covariance matrix and the probability of its occurrence.

Others have nonparametrically modeled the return distribution of a stochastic volatility model, but the joint return, log-volatility distribution has not. Jensen (2004), Griffin and Steel (2006), Jensen and Maheu (2010), Griffin and Steel (2011), and Delatola and Griffin (2011a), each apply a Dirichlet process mixture type prior to the return distribution. For the asymmetric stochastic volatility model, Delatola and Griffin (2011b) include a constant leverage effect and model the distribution of the log-squared return innovation nonparametrically with the infinite mixture, Constant Component Variance model of Griffin (2011). Yu (2011) directly models the leverage effect nonparametrically by modeling the correlation between returns and volatility with a fixed ordered linear spline, and Durham (2007) models the return distribution with a finite mixture model where its order is fixed a prior by the econometrician.

By relaxing the parametric distribution of the asymmetric stochastic volatility model, the approach taken in this paper is more flexible and better positioned to model non-Gaussian behavior. As pointed out by Das and Sundaram (1999), the parametric stochastic volatility model can only produce the level of skewness and kurtosis observed in the data when it takes on implausible parameter values. By design a nonparametric joint distribution allows the stochastic volatility model to capture these types of characteristics in the data without sacrificing the time dependent nature of the stochastic volatility model.

A Markov chain, Monte Carlo posterior algorithm for sampling the nonparametric and

---

[1]See Harvey et al. (1994), Yu (2005) and Omori et al. (2007) or Eq. (30)-(32) for the parametric version of the asymmetric stochastic volatility model.

parametric portions of the model is presented. Our semiparametric sampler extends the univariate algorithm of the semiparametric stochastic volatility model by Jensen and Maheu (2010). A restricted version of the algorithm is also applied to a fully parametric, asymmetric, stochastic volatility model. Parameter, volatility, and distributional uncertainty are integrated out with the sampler of the posterior. These draws will be used to generate both the one-day-ahead predictive joint density for daily market returns and log-volatility, and the marginal density for one-day-ahead returns.

Stochastic volatility models and econometric models in general are chosen based on their predictability (see Geweke (2001), and Geweke and Whiteman (2006)). This is understandable given the important role predictions play in valuing stocks and options, constructing portfolios, and creating hedging strategies. Strong evidence in favor of our semiparametric model relative to parametric models is provided by the sequential predictive likelihoods for returns. In particular, the DPM estimate of the unknown predictive distribution for returns is found to be robust over low and high volatility periods and to large return shocks.

Great emphasis is also placed on the ability of a model to forecast volatility.[2] Asymmetry, where an unexpected decline in price leads to higher volatility, whereas an increase in price causes volatility to decline, is common in volatility models (see Bekaert and Wu (2000), Chen and Ghysels (2011)). In the stochastic volatility model this asymmetry comes from negatively correlated returns and log-volatility. Since the covariance matrix of the nonparametric return and log-volatility innovations distribution follows a random second-order effect, this paper's semiparametric volatility model exhibits asymmetry. However, the random second-order effect for the covariance also allows this correlation to change depending on market conditions. During a regular market the asymmetry is like that found by Chen and Ghysels (2011), moderate increases in stock prices reduces expected volatility, whereas any decline in stock prices or a large unexpected increase in prices leads to higher expected volatility. However, during highly volatile times, expected volatility barely increases following an unexpected shock to market returns. In other words, the asymmetric return and volatility relationship is not a high volatility phenomenon.

The paper is organized as follows. In the Section 2, the asymmetric, stochastic volatility model with a nonparametric Dirichlet process mixture prior for the unknown distribution is constructed. Section 3 spells out the nonparametric model's Markov chain, Monte Carlo sampler of all the unknown parameters and latent variables. In Section 4 we apply our semiparametric and an existing parametric asymmetric, stochastic volatility models to 28 years worth of daily market returns as measured by the value-weighted market portfolio

---

[2]See Poon and Granger (2003) for an extensive overview of the volatility forecasting literature

from the Center of Research in Security Prices. In Section 5 and 6, we compare our empirical results by first evaluating the Bayes factor in favor of a nested parametric versions of the nonparametric model with the Savage-Dickey density ratio. We then compare the forecasting performance of the models over the last two years of the data with their cumulative log predictive Bayes factors. In Section 7 we estimate volatility's response to unexpected changes in stock prices. A summary and conclusions are contained in Section 8.

2   Model

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

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

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

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, φ, is constrained to the unit interval, ensuring the log-volatility process in Eq. (2) is stationary.[3]

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 ASV 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. 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 observation's innovations are distributed as a bivariate Normal with a mean zero vector but with a random covariance matrix, $\Lambda_t = \begin{pmatrix} \sigma_{y,t}^2 & \sigma_{yh,t} \\ \sigma_{yh,t} & \sigma_{h,t}^2 \end{pmatrix}$, that is distributed as G. The distribution G is also unknown and its prior is the Ferguson (1973) Dirichlet process distribution, $DP(\alpha, G_0)$, where the

---

[3]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.

nonzero scalar α is the precision parameter and G_0 is the base distribution. We will provide a concrete definition for the DP prior, shortly.

The DPM builds on the well known property that a flexible distribution can be found by mixing together a finite 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 we model as a mixture of distributions

$$\int F_N(0, \Lambda) G(d\Lambda), \tag{6}$$

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 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 G is unknown and is modeled nonparametrically. As a unknown random distribution function, G enables the Λ_ts 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, Λ_ts, 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

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

where $\delta_{\Sigma_j}(\cdot)$ is a degenerative distribution on the covariance matrix atom

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

where $\sigma_{yh,j} = \rho_j \sigma_{y,j} \sigma_{h,j}$. Each Σ_j is a unique 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 $\alpha > 0$.

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 about Λ_t distribution declines and the Λ_ts will be distributed as G_0. On the other hand, if α is close to zero, the prior for G will consist of a discrete distribution whose support is located on only a few covariance matrices, Σ_j. The DP precision parameter, α, can thus, be understood as controlling the complexity of the random, second-order, effects.

2.1 Parsimony with the DP

Parsimony, in other words, clustering or uniqueness in the covariances, Λ_t, of the ASV-DPM model is provided by the almost sure discreteness of Eq. (7). By modeling G as a DP prior there is guaranteed to be ties among the Λ_ts. To be explicit, the joint distribution of the covariances can be defined sequentially as $\pi(\Lambda_1, \Lambda_2, \ldots, \Lambda_n) \equiv \pi(\Lambda_1)\pi(\Lambda_2|\Lambda_1)...\pi(\Lambda|\Lambda_1, \ldots, \Lambda_{n-1})$ where $\pi(\Lambda_1) \equiv G_0$ and

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

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

for $t = 2, \ldots, n$ (see Blackwell and MacQueen (1973)). Integrating out G from each of the conditional distributions in Eq. (10), we obtain 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.

Eq. (12) shows the self-reinforcing property of the DP where previously sampled values are more likely to be resampled in the future. The more $\Lambda_{t'}$s belonging to the cluster whose covariance matrix is Σ_j, the larger n_j will be and the greater the probability Λ_t will be

assigned Σ_j. On the other hand, if only a few $\Lambda_{t'}$ have been assigned Σ_j, both n_j and the likelihood of Λ_t being assigned Σ_j will be smaller.

Notice in Eq. (12) that there is also a non-trivial chance, proportional to α, of a new covariance cluster being selected from G_0. Extreme clustering occurs when the precision parameter $\alpha \to 0$. On the other hand, if $\alpha \to \infty$, no clustering occurs and every observation 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 but the model will typically not have finite unconditional moments (see Nelson (1991)).

2.2 Orthogonal representation

The ASV-DPM model can be written 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 for G, s_t will be distributed

$$s_t \sim \sum_{j=1}^{\infty} \pi_j \delta_j,$$

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

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

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

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

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

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

where the correlation matrix $\Upsilon_j = \left(\begin{array}{cc} 1 & \rho_j \\ \rho_j & 1 \end{array} \right)$, with $\rho_j = \sigma_{hy,j}/(\sigma_{h,j}\sigma_{y,j})$.

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

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

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:

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

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

where $(w_t, v_t)' \overset{iid}{\sim} N(0, \mathbf{I_2})$. This form of the ASV-DPM model will be shown to be convenient for the posterior sampling of log-volatilities and φ.

3 Estimation

In this section we provide a likelihood-based approach to parameter inference, distributional uncertainty, 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

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

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:

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

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

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

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

- $\pi(\alpha | \{\Lambda_t\})$

3.1 Λ_t sampler

Sampling $\{\Lambda_t\} | y, h, \mu, \varphi, \alpha$ 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

$z_t = ((y_t - \mu) \exp\{-h_t/2\}, h_{t+1} - \varphi h_t)'$, and draw Λ_t from the distribution:

$$\Lambda_t | \{\Lambda_{t'} : t' \neq t\}, z_t, \alpha \quad \sim \quad \frac{\alpha}{\alpha + n - 1} g(z_t) G(d\Lambda | z_t)$$
$$+ \frac{1}{\alpha + n - 1} \sum_{t' \neq t} f_N(z_t | \Lambda_{t'}) \delta_{\Lambda_{t'}}(d\Lambda), \tag{20}$$

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

$$g(\Lambda | z_t) \quad \propto \quad |\Lambda|^{-1/2} \exp\left\{-\frac{1}{2} \text{tr} z_t z_t' \Lambda^{-1}\right\} \frac{|S_0|^{v_0/2}}{|\Lambda|^{(v_0+3)/2}} \exp\left\{-\frac{1}{2} \text{tr} \Lambda^{-1} S_0\right\},$$
$$= \quad \frac{|S_0|^{v_0/2}}{|\Lambda|^{(v_0+4)/2}} \exp\left\{-\frac{1}{2} \text{tr}(S_0 + z_t z_t') \Lambda^{-1}\right\}, \tag{21}$$

is the kernel to a Inverse-Wishart distribution; i.e., $G(d\Lambda | z_t) \equiv \text{Inv-Wish}(S_0 + z_t z_t', v_0 + 1)$ (see Zellner (1971), p.395). Integrating out the Λ from the Inverse-Wishart distribution is the marginal likelihood $g(z_t)$, which equals the density of a bivariate Student-t distribution:

$$g(z_t) \quad = \quad f_{MSt}(z_t | \mathbf{0}, (S_0/(v_0 - 1))^{-1}, v_0 - 1), \tag{22}$$

with $v_0 - 1$ degrees of freedom, mean-zero vector, and covariance, $S_0/(v_0 - 3)$ (see Zellner (1971), 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 $\pi(\{\Lambda_t\} | y, h, \mu, \varphi, \alpha)$. This distribution is $\pi(\Sigma_1, \ldots, \Sigma_k, s | y, h, \mu, \varphi, \alpha)$ where the Σ_j, $j = 1, \ldots, k$, $k \leq n$, 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, φ and α, draws from the posterior $\pi(\{\Sigma_j\}, s | z, \alpha) \equiv \pi(\{\Sigma_j\}, s | y, h, \mu, \varphi, \alpha)$ are made with the following 2-step algorithm

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

$$s_t | \{\Lambda_{t'} : t' \neq t\}, z_t, \alpha \quad \sim \quad \begin{cases} \frac{\alpha}{\alpha + n - 1} g(z_t) \, \delta_0(ds_t) \\ \frac{1}{\alpha + n - 1} \sum_{j=1}^{k^{(t)}} n_j^{(t)} f_N(z_t | \mathbf{0}, \Sigma_j) \, \delta_j(ds_t). \end{cases} \tag{23}$$

If a zero is drawn for s_t, we then draw a new Σ_j from the Inverse-Wishart distribution in Eq. (21), 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:

$$\pi(\Sigma_j|\mathbf{z}, s, k) \quad \propto \quad \prod_{t:s_t=j} f_N(\mathbf{z}_t|\mathbf{0}, \Sigma_j)G_0(d\Sigma) \tag{24}$$

$$\propto \quad \prod_{t:s_t=j} |\Sigma_j|^{-1/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\mathbf{z}_t\mathbf{z}_t'\Sigma_j^{-1}\right\}$$

$$\times \frac{|S_0|^{v_0/2}}{|\Sigma_j|^{(v_0+3)/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}S_0\Sigma_j^{-1}\right\} \tag{25}$$

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

$$\sim \quad \mathrm{Inv\text{-}Wish}\left(S_0 + \sum_{t:s_t=j} \mathbf{z}_t\mathbf{z}_t', v_0 + n_j\right). \tag{27}$$

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. Random length 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 $l_{t'} \sim \mathrm{Pois}(\lambda_h)$, is:

$$\pi(h_{(t',\tau)}|y_{(t',\tau)}, h_{-(t',\tau)}) \quad \propto \quad \pi(h_{(t',\tau)}|h_{\tau+1}, h_{t'-1})f(y_{(t',\tau)}|h_{(t',\tau)}, h_{\tau+1})$$

$$= \exp\left\{-\frac{1}{2}\frac{(h_{t'} - \varphi h_{t'-1})^2}{\sigma_{h,s_{t'}}^2}\right\} \prod_{t=t'}^{\tau} \exp\left\{-\frac{1}{2}\left[\frac{(h_{t+1} - \varphi h_t)^2}{\sigma_{h,s_t}^2} + h_t\right]\right\}$$

$$\times \prod_{t=t'}^{\tau} \exp\left\{-\frac{1}{2}\frac{\left(y_t - \mu - e^{h_t/2}\sigma_{y,s_t}\rho_{s_t}(h_{t+1} - \varphi h_t)/\sigma_{h,s_t}\right)^2}{(1 - \rho_{s_t}^2)\exp\{h_t\}\sigma_{y,s_t}^2}\right\} \tag{28}$$

10

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 $\pi(h_0) \sim N(0, \sigma_{h,0}^2/(1-\varphi^2))$ where $\sigma_{h,0}^2 \equiv E[G_0(d\Lambda)]_{2,2}$ is the expected variance of log-volatility from the the DPM base distribution. Drawing h_0 from $\pi(h_0|h_1) \equiv N(\varphi h_1, \sigma_{h,0}^2)$, 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 of h we integrate out the one period ahead, out of sample volatility, h_{n+1}, by replacing it with a random draw from:

$$\pi(h_{n+1}|y, h_n) \sim N(\bar{h}_{n+1}, \bar{\sigma}_{h_{n+1}}^2), \tag{29}$$

where

$$\bar{h}_{n+1} = \varphi h_n + \frac{(y_n - \mu)\rho_{s_n}\sigma_{h,s_n}e^{-h_n/2}}{\sigma_{y,s_n}}, \qquad \bar{\sigma}_{h_{n+1}}^2 = \left(\frac{\rho_{s_n}^2}{(1-\rho_{s_n}^2)\sigma_{h,s_n}^2} + \frac{1}{\sigma_{h,s_n}^2}\right)^{-1},$$

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

Since the conditional distributions of Eq. (28) 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 φ

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

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

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

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

When the candidate φ' is rejected, the prior sweep's draw of φ is kept as the draw from $\pi(\varphi|y, h, \{\Lambda_t\})$.

11

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 $\tilde{y} = y_t - \rho_t e^{h_t/2} \sigma_{y,s_t} (h_{t+1} - \varphi h_t)/\sigma_{h,s_t}$ 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 $\pi_\xi/(1 - \pi_\xi) = (a + k - 1)/[n(b - \ln \xi)]$.

3.6 Results with quasi-return data

To benchmark the ASV-DPM sampler, we apply it to 1,000 returns generated with Harvey et al. (1994) parametric, asymmetric, stochastic volatility model

$$y_t = \mu + \exp\{h_t/2\}\epsilon_t \tag{30}$$

$$h_{t+1} = \varphi h_t + \eta_t \tag{31}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N(\mathbf{0}, \Sigma). \tag{32}$$

where $\Sigma = \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_h \\ \rho\sigma_y\sigma_h & \sigma_h^2 \end{pmatrix}$ and $\rho \equiv \text{Corr}(\epsilon_t, \eta_t)$. 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 structural form as the ASV-DPM model.

The ASV parameters are set equal to the estimates reported in Table 2 of Section 4. These parameter estimates come from estimating the ASV model with 7,319 daily returns (multiplied by a 100) of the Center of Research in Security Prices (CRSP) value-weighted portfolio index over the January 2, 1980 to December 31, 2008 time period. We use these parameter values in the ASV model to generate 1,000 quasi-returns.

We fit both the ASV-DPM and ASV model to the simulated returns. The priors for the ASV-DPM model are $\pi(\mu) \equiv N(0, 0.1)$ and $\pi(\varphi) \equiv N(0, 100)I_{|\varphi|<1}$. For the DPM, we choose the base distribution $G_0 \equiv$ Inv-Wish(S_0, v_0) where $S_0 = \boldsymbol{I}_2$ and $v_0 = 10$. The prior for the DPM precision parameter is $\pi(\alpha) =$ Gamma$(2, 8)$ so that $E[\alpha] = 1/4$ and Var$[\alpha] = 1/32$. For the ASV model, the priors are the same as those for the ASV-DPM, meaning the prior for Σ is G_0. The sampler of the ASV model is the same as the ASV-DPM model's except with k and s_t, $t = 1, \ldots, n$, fixed and set equal to 1.

Using the initial starting parameter values, we throw away the first 1,000 draws of log-volatility and then the following 10,000 draws of both the volatilities and parameters, before keeping the last 30,000 parameter draws. Table 1 reports the posterior mean, standard deviation, and 95% probability interval for the ASV-DPM and ASV parameters. There is nothing out of the ordinary in these posterior estimates. Each model's parameter estimates are reasonably close to the true values and the ASV-DPM model finds on average one mixture covariance matrix.

Table 1: The posterior estimates for the ASV-DPM and ASV models as applied to 1,000 simulated ASV returns using the parameter values from Table 2. Given the initial starting parameter values, the first 1,000 draws of log-volatility are discarded and then the next 10,000 draws of both the volatilities and parameters are thrown away, before keeping the last 30,000 draws.

	ASV-DPM			ASV		
	mean	stdev	95% prob interval	mean	stdev	95% prob interval
α	0.1690	0.1414	(0.0207, 0.4727)			
k	1.0	0.8454	(1, 4)			
φ	0.9634	0.0118	(0.9386, 0.9849)	0.9897	0.0122	(0.9545, 0.9995)
μ	0.0501	0.0237	(0.0032, 0.0962)	0.0653	0.0237	(0.0195, 0.1116)
σ_y^2				0.1633	0.1644	(0.0427, 0.6495)
σ_h^2				0.0480	0.0117	(0.0299, 0.0770
ρ				-0.3350	0.1059	(-0.5377,-0.1248)

Using data generated from the ASV model, $y_t = 0.06 + \exp\{h_t/2\}\epsilon_t$ and $h_{t+1} = 0.97h_t + \eta_t$ where

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.61 & -0.46 * \sqrt{0.61 * 0.04} \\ -0.46 * \sqrt{0.61 * 0.04} & 0.04 \end{pmatrix} \right).$$

We also compare the joint predictive densities, $f((y_{n+1}, h_{n+2})'|y, M)$, $M =$ ASV-DPM, ASV, by numerically calculating them. The predictive density for the ASV-DPM is approximately

$$f\left(\begin{pmatrix} y_{n+1} \\ h_{n+2} \end{pmatrix} \Big| y, \text{ASV-DPM} \right) \approx$$

$$R^{-1} \sum_{l=1}^{R} f\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right) \bigg| \mu^{(l)}, \varphi^{(l)}, h^{(l)}, h_{n+1}^{(l)}, \left\{\Sigma_j^{(l)}\right\}, s^{(l)}, \alpha^{(l)}\right) \quad (33)$$

where $R = 30,000$, and $\mu^{(l)}, \varphi^{(l)}, h^{(l)}, \left\{\Sigma_j^{(l)}\right\}, s^{(l)}, \alpha^{(l)}$ are the lth draw from the posterior and

$$f\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right) \bigg| \mu, \varphi, h, h_{n+1}, \{\Sigma_j\}, s, \alpha\right) =$$

$$\frac{\alpha}{\alpha + n} f_{MSt}\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right) \bigg| \left(\begin{array}{c} \mu \\ \varphi h_{n+1} \end{array}\right), \left(\frac{H_{n+1} S_0 H_{n+1}}{v_0 - 1}\right)^{-1}, v_0 - 1\right)$$

$$+ \sum_{j=1}^{k} \frac{n_j}{\alpha + n} f_N\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right) \bigg| \left(\begin{array}{c} \mu \\ \varphi h_{n+1} \end{array}\right), H_{n+1} \Sigma_j H_{n+1}\right) \quad (34)$$

with $H_{n+1} = \left(\begin{array}{cc} e^{h_{n+1}/2} & 0 \\ 0 & 1 \end{array}\right)$ and h_{n+1} being a draw from the posterior distribution of Eq. (29). By averaging over the weighted draws of the parameters and the unknown volatilities, the predictive density integrates out both parameter and log-volatility uncertainty leaving a distribution dependent on only the return series, y. In the ASV model, the predictive density integrates out μ, φ, h, Σ, from the sampling distribution,

$$f_N\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right) \bigg| \left(\begin{array}{c} \mu \\ \varphi h_{n+1} \end{array}\right), H_{n+1} \Sigma H_{n+1}\right). \quad (35)$$

In Figure 1 we plot the joint predictive densities of both models. Since the densities are three-dimensional, we plot each model's density from two vantage points – the left side figures plot the joint densities from the y_{n+1}-axis vantage point, and the righthand figures plot the densities, but from the h_{n+2}-axis perspective. The two model's densities are nearly identical in their shape and location. Both densities are centered at $y_{n+1} = 0$ and $h_{n+2} = 1.25$, and both show a slight upward skewness in the h_{n+2} dimension. If there is a difference to be found it is in their height. The ASV-DPM model's predictive density reaches a maximum density value of 0.2 that is slightly larger than the ASV, indicating the predictive distribution of the ASV-DPM model is leptokurtotic relative to the predictive distribution of the ASV model. This does not come as a surprise. From the results in Table 1, there are sweeps where the DPM sampler drew a mixture representation with two or more clusters. Furthermore, in Eq. (34) we see that the ASV-DPM predictive density includes the Student-t distribution with $v_0 - 1$ degrees of freedom – a known leptokurtotic distribution.

14

Figure 1: Joint predictive densities of the ASV-DPM and ASV model from the vantage points of the y_{n+1}-axis (lefthand side plots) and h_{n+2}-axis (righthand side plots) as applied to return data simulated from the ASV model.
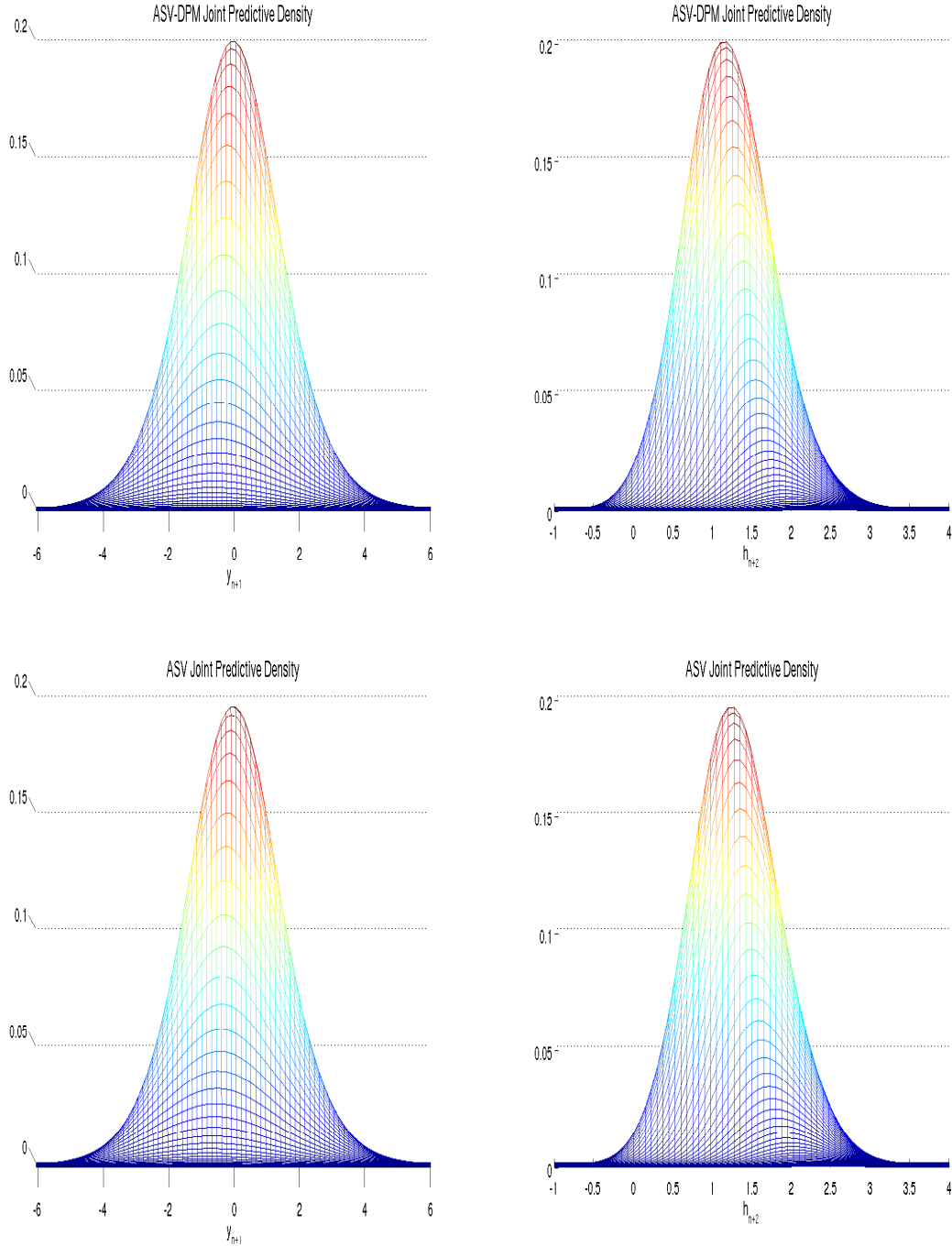
4   Empirical Application

We analyze the ASV-DPM and ASV models by applying them to 7,319 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 2, we plot the value-weighted portfolio returns. The chosen time period is ideal since market 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. Daily market returns appear to be asymmetrically distributed with fat-tails as is evident in a negative skewness of -0.757 and an excess kurtosis measure of 19.296.

Figure 2: CRSP value-weighted portfolio daily compounded returns from January 2, 1980 to December 31, 2008 (in percentages).
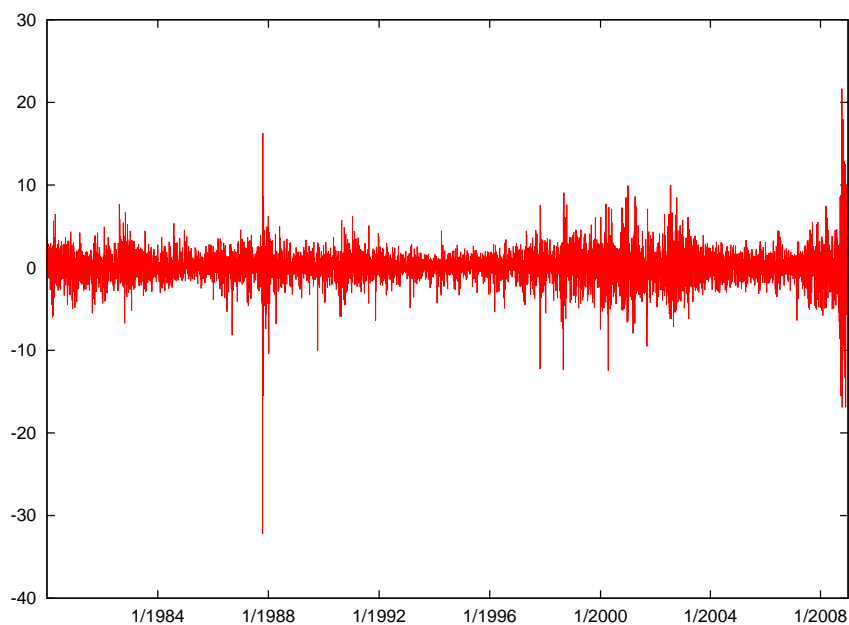
The priors applied to the ASV-DPM and ASV models are the same as were used in Section 3.6. They are $\pi(\mu) \equiv N(0, 0.1)$, $\pi(\varphi) \equiv N(0, 100)I_{|\varphi|<1}$, $G_0 \equiv$ Inv-Wish$(\boldsymbol{I}_2, 10)$, and $\pi(\alpha) =$ Gamma$(2, 8)$. For the ASV model, the prior for Σ is the Inv-Wish$(\boldsymbol{I}_2, 10)$ distribution and k and s_t, $t = 1, \ldots, n$, are set equal to 1.

To reduce the influence of the starting values, 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

Table 2: 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	0.3188	0.1609	(0.0862, 0.7007)			
k	4.4220	1.3120	(3, 8)			
φ	0.9799	0.0030	(0.9739, 0.9855)	0.9740	0.0037	(0.9662, 0.9809)
μ	0.0665	0.0082	(0.0506, 0.0826)	0.0611	0.0083	(0.0448, 0.0774)
σ_y^2				0.6100	0.0514	(0.5112, 0.7169)
σ_h^2				0.0391	0.0046	(0.0308, 0.0490)
ρ				-0.4682	0.0371	(-0.5402,-0.3944)

times, keeping only the last 30,000 draws from the two models for inference purposes.

Table 2 reports the posterior mean, standard deviation, and 95% posterior, probability interval for the parameters of the ASV-DPM and ASV models. The posterior mean for the ASV-DPM model's unconditional mean of return, μ, at 0.067 is slightly larger than the ASV model's estimate of 0.061. But their posterior standard deviations have the same value of 0.008. Hence, the model's posterior distributions, $\pi(\mu|y)$, are similar in shape but the ASV-DPM is shifted slightly to the right. Although the difference in the estimates of μ are small, it can still have an effect on one's expected median wealth say 20 years into the future. For example the median expected wealth per unit of investment over twenty years $(e^{\mu/100*365*20})$ is 128.32 for the ASV-DPM model as opposed to 86.51 for the ASV.

Persistence in volatility, as captured by the posterior distribution of the autoregressive parameter, φ, is close to being the same for the two models. In the ASV-DPM model, the posterior mean of φ is 0.98, whereas, in the ASV model it is 0.97. The standard deviations are also similar with the ASV-DPM model's equaling 0.003 and the ASV slightly larger at 0.004. Values of φ so close to one is evidence of a strongly persistent volatility process where a shock to volatility impacts its future values and lives on for a very long time in either model. Even under a nonparametric distribution, the φ for ASV-DPM finds clustering in volatility where large and small fluctuations follow similar type of behavior.

In Table 2, the posterior average number of mixture clusters is $k = 4.42$. Because of label switching in the sweep to sweep draws of the s_ts,[4] we are unable to identify the posterior distribution of the assignment vector, s. However, we are able to calculate the

---

[4]See Richardson and Green (1997) for a extensive discussion on label switching in mixture models.

average posterior value for each observation's covariance matrix, Λ_t.

Figure 3: Time plots of the posterior average of $\sigma^2_{y,t}$, $\sigma^2_{h,t}$, ρ_t, t=1,...,7319
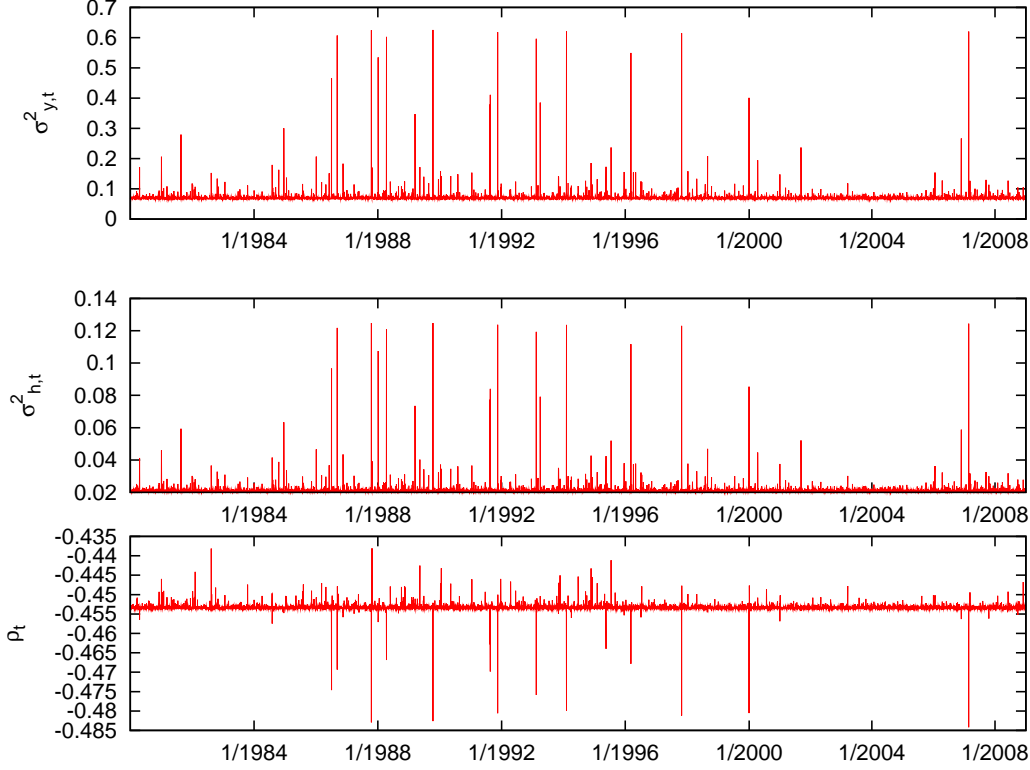
In Figure 3, we plot the estimates of Λ_t by graphing the average posterior draw of $\sigma^2_{y,t}$, $\sigma^2_{h,t}$, ρ_t, for $t = 1, \ldots, 7319$. In the figure, the movement over time in $\sigma^2_{y,t}$ is nearly indistinguishable from $\sigma^2_{h,t}$. Both move at exactly the same instances in exactly the same direction. Such dynamics suggests markets requiring a large $\sigma^2_{y,t}$ also need $\sigma^2_{h,t}$ to be large. Furthermore, those instances where the value of $\sigma^2_{y,t}$ and $\sigma^2_{h,t}$ are large also identifies a mixture cluster whose probability is relatively low but is still important. For instance, the triplet ($\sigma^2_{y,t} = 0.6$, $\sigma^2_{h,t} = 0.12$, $\rho_t = -0.48$) only occurs a dozen or so times over the sample period, but each occurrence corresponds to a sizable market decline.

This contrasts with the covariance cluster where the spikes in $\sigma^2_{y,t}$ and $\sigma^2_{h,t}$ emerge from and whose value is approximately ($\sigma^2_{y,t} = 0.07$, $\sigma^2_{h,t} = 0.02$, $\rho_t = -0.455$). This cluster can be viewed as the market's typical covariance behavior. It consists of most of the sample, and, hence, has the highest probability of occurring.

Figure 3 also identifies three clear mixture clusters. Two of the three are those mentioned in the previous paragraph; the one cluster representing the typical market, and the other

for when there is a large market decline. The third cluster consists of those days where both $\sigma_{y,t}^2$ and $\sigma_{h,t}^2$ are elevated, but the level of correlation remains unchanged at $\rho_t = -0.455$. For instance, the third and fourth episode of $\sigma_{y,t}^2$ and $\sigma_{h,t}^2$ spiking in Figure 3 are occurrences of this cluster.

4.1 Fit

A point of possible contention with the ASV-DPM model is that it nonparametrically fits an unknown joint distribution where one of the random variables, h, is not observed. This latency could cause the parameters of the DPM to be unidentified. Since the DP parameters suffer from label switching we are unable to analyze the sampled distribution parameters to determine if they are identified. Alternatively, we can compare the nonparametric model's draws of h with those from the ASV model. If the ASV-DPM model is unidentified, we would expect to find the draws from the smooth distribution $\pi(h|y)$ to be less precise and have a very different posterior mean from the parametric model. We would also expect the location and spread of the ASV-DPM model's joint predictive density to not match up with the ASV model's predictive distribution.

In Figure 4 we plot two graphs. In Figure 4(a) we plot the difference between the ASV and ASV-DPM sampler's standard deviations of $\pi(h|y)$ over the trading days, January 2, 1980 to December 29, 2008, and in Figure 4(b) the sample means of the two model's draws of h. Except for a few trading days, the standard deviation of the ASV draws are slightly larger than the ASV-DPM models. Because of the large difference between the ASV-DPM and ASV standard deviation on those days where the ASV-DPM standard deviations is larger, on average, the ASV-DPM standard deviations are 0.0026 larger than the ASV. On these days market volatility was generally low.

In Figure 4(b), the two model's smoothed volatilities, $E[h|y]$, are similar in their level and pattern. However, there are differences, such as when volatility reaches a local peak. In these instances, the ASV model's smooth volatility is larger than the ASV-DPM model. Because the instantaneous variance of returns, σ_y^2, in the ASV model is constant, large changes in the return process correspond to large changes in volatility. This contrasts with the ASV-DPM model where σ_{y,s_t}^2 is flexible and, as we saw in Figure 3, changes value when it needs to adapt to a market decline. As a result the ASV-DPM model's volatilities are less volatile and do not increase by as much as the ASV model.

Figure 4: (a) difference between the ASV and ASV-DPM standard deviation from the MCMC draws of $\pi(h_t|y)$, and (b) the average draw of $h_t|y$ from the ASV and ASV-DPM. Both figures are for the period of January 2, 1980 to December 31, 2008.
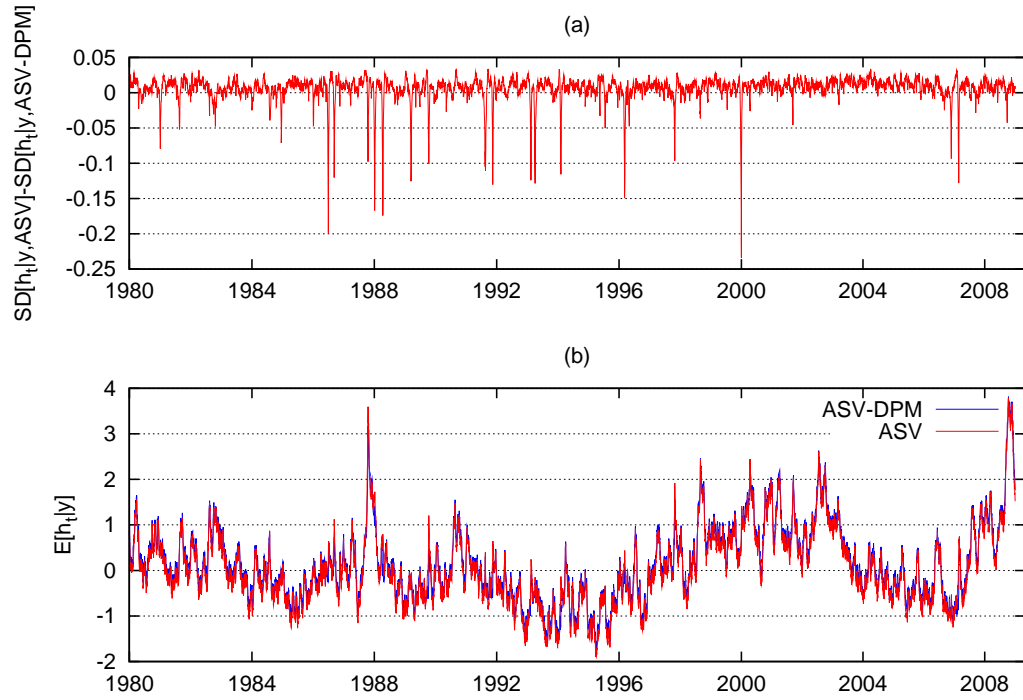
5 Nested Model Comparison

The DP precision parameter α can be understood as being a tuning parameter to the number of unique Λ_ts; i.e., the degree of clustering imposed by the DP prior. Under the $DP(\alpha, G_0)$ prior, a data set of length n is expected to have $E[k_n] = \sum_{i=1}^{n} \alpha/(\alpha + i - 1)$ clusters. By increasing or decreasing the value of α, the ASV-DPM model is more or less likely to add new clusters. As α approaches zero, the probability of a second unique covariance matrix, $\alpha/(1 + \alpha)$, approaches zero, as does the probability of there being a second, third, fourth, etc. cluster. It follows that the ASV-DPM model is equivalent to the parametric ASV model when α is equal to zero.

At the other end of the spectrum, the prior probability of drawing a new cluster for Λ_t goes to one as $\alpha \to \infty$. Because the DP prior for G is no longer discrete, but is instead equal to the distribution G_0, when $\alpha \to \infty$, the prior for Λ_t, $t = 1, \ldots, n$, is G_0. Hence, as $\alpha \to \infty$, the ASV-DPM model is the ASV-t model – a parametric ASV model whose innovations are distributed as a bivariate, Student-t with mean-vector zero, covariance matrix, $S_0/(v_0 - 3)$ and $v_0 - 1$ degrees of freedom (see Eq. (22)).

Because the ASV and ASV-t models are nested versions of the ASV-DPM that depend on the value of α, Bayes factors in favor of the nested models can be computed using the Savage-Dickey density ratio test (see Dickey (1971)). In general, the Savage-Dickey density ratio favors the nested model $M' : \alpha = \alpha_0$, where, in our case, $\alpha_0 = \{0, \infty\}$, over the general model $M : \alpha$, where $\alpha \geq 0$, when Bayes-factor

$$BF(\alpha = \alpha_0) \equiv \frac{m(y|M')}{m(y|M)}, \tag{36}$$

$$= \frac{\pi(\alpha = \alpha_0|y)}{\pi(\alpha = \alpha_0)}, \tag{37}$$

where m is the marginal likelihood, is large. In our case, $M = $ ASV-DPM and $M' = $ ASV, ASV-t, respectively.

The limit $\alpha_0 \to \infty$ does not lend itself easily to the Savage-Dickey density ratio, so we transform α into the random variable $u \equiv \alpha/(\alpha + 1)$ and assume u is distributed as the maximum entropy prior

$$\pi(u|\lambda) = \frac{e^{\lambda u}\lambda}{e^\lambda - 1}, \qquad \lambda \in \mathbb{R}, \tag{38}$$

over the unit interval, $u \in [0, 1]$. The corresponding maximum entropy prior for α is

$$\pi(\alpha|\lambda) = \frac{\lambda \exp\{\alpha\lambda/(1 + \alpha)\}}{(e^\lambda - 1)(1 + \alpha)^2}. \tag{39}$$

Under this transformation, $u \to 0$, as $\alpha \to 0$, and $u \to 1$, as $\alpha \to \infty$. The random variable u is thus, the prior probability of there being a second mixture cluster. If $u = 1$, then Λ_2 will be drawn from G_0. Whereas, if $u = 0$, the probability of drawing a second cluster is zero and $\Lambda_2 = \Lambda_1$.

Let the nested ASV and ASV-t models $M' : u = u_0$, be, respectively, $u_0 = \{0, 1\}$ and the unrestricted model $M : u \in [0, 1]$. In terms of u, the Savage-Dickey density ratio in favor of the nested model M' is

$$BF(u = u_0) \equiv \frac{\pi(u = u_0|y)}{\pi(u = u_0)}.$$

Unlike the Gamma prior we used earlier for α, the maximum entropy prior does not lend itself to a standard distribution for the conditional posterior of α, so, we modify the Escobar and West (1995) sampler of α. Since drawing either α or u requires a Metropolis sampler, we could choose to sample either one. However, because α is defined on the positive real line, whereas u is constrained to the unit interval, we choose to draw α and use a random walk proposal with unit variance to generate the candidate draws.

Denote the candidate draw by α'. It will be accepted as a draw from $\pi(\alpha|y)$ with probability

$$\frac{\pi(k|\alpha', n)\pi(\alpha'|\lambda)}{\pi(k|\alpha, n)\pi(\alpha|\lambda)} = \frac{\alpha'^k \Gamma(\alpha')/\Gamma(\alpha' + n)\pi(\alpha'|\lambda)}{\alpha^k \Gamma(\alpha)/\Gamma(\alpha + n)\pi(\alpha|\lambda)}, \tag{40}$$

where α is the draw from the previous sweep and k is the number of clusters from the current sweep (see Escobar and West (1995) for the formula of the likelihood function, $\pi(k|\alpha, n)$). For each draw of α we compute the corresponding draw of u and evaluate u's empirical posterior distribution $\pi(u|y)$ at zero and one. If the Savage-Dickey ratio at these points is greater than one then there is evidence in favor of the nested model, M'.

We choose the maximum entropy prior for u where $\lambda = 0$ when testing the nested ASV and ASV-t models against the ASV-DPM. When $\lambda = 0$, the maximum entropy prior is a uniform prior over the unit interval. Except for the prior of u, the MCMC sampler is the same as Section 4.

In Figure 5, we plot the uniform prior and empirical posterior density of u. The Savage-Dickey density ratio in favor of the ASV model ($u_0 = 0$) has a 1-in-10 chance of occurring and indicates that there is some likelihood for the parametric ASV model but not much. Since $\pi(u = 1|y) = 0$, there is virtually no evidence supporting the ASV-t model ($u_0 = 1$).

Though there is little empirical evidence supporting the nested ASV and ASV-t models, in Figure 5, there is still a range of values for u_0 where a sharp hypothesis is supported. If we restrict u_0 to values where $\pi(u = u_0|y)$ is greater than $\pi(u_0)$, we find a number of

Figure 5: The prior, $\pi(u|\lambda = 0) \equiv \mathrm{Unif}(0,1)$, and the empirical posterior density $\pi(u|y)$.
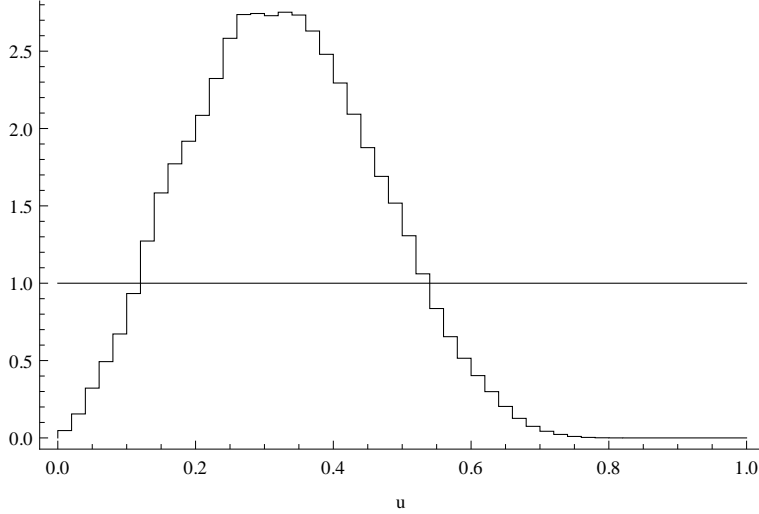
Bayes-factors in favor of the restricted ASV-DPM model. From Figure 5, the data supports a ASV-DPM model where u_0 is between 0.125 and 0.525. In other words, the data supports ASV-DPM models having on average 1 to 9 clusters.

5.1 Prior sensitivity

Because M' is a sharp hypothesis, the prior for u under M' is the Dirac delta function, $\pi(u|M') = \delta_{u_o}(u)$. For the unrestricted model, M, the prior is the maximum entropy distribution of Eq. (38) with hyperparameter $\lambda \in \mathbb{R}$. As $\lambda \to -\infty$, the prior $\pi(u|\lambda, M) \to \delta_0(u)$; i.e., the prior of the unrestricted model converges to that of the sharp ASV model hypothesis. As $\lambda \to \infty$, $\pi(u|\lambda, M) \to \delta_1(u)$ and the prior is that of the ASV-t model.

The marginal likelihood for M is

$$m(y|\lambda, M) = \int l(y|u, \lambda, M)\pi(u|\lambda, M) \, du, \tag{41}$$

and the marginal likelihood for the nested model M' is

$$
\begin{aligned}
m(y|M') &= \int l(y|u, \lambda, M)\delta_{u_0}(u) \, du \tag{42} \\
&= l(y|u = u_0, M). \tag{43}
\end{aligned}
$$

The Bayes factor in favor of the restricted model M' written in terms of these marginal likelihoods is

$$BF(u = u_0|y, \lambda, M) = \frac{l(y|u = u_0, M)}{m(y|\lambda, M)}.$$

Under this prior for u, there exists limits where the unrestricted model equals the restricted. The two restricted models are $u_0 = 0$ (ASV) and $u_0 = 1$ (ASV-t). To obtain the ASV model with the unrestricted model, the prior

$$\lim_{\lambda \to -\infty} \pi(u|\lambda, M) = \delta_0(u),$$

and for the ASV-t model

$$\lim_{\lambda \to \infty} \pi(u|\lambda, M) = \delta_1(u).$$

Thus, the Bayes factor in favor of the ASV model can be written in terms of M's marginal likelihood function as

$$BF(u = 0|y, \lambda, M) = \frac{m(y|\lambda \to -\infty, M)}{m(y|\lambda, M)} \tag{44}$$

and for the ASV-t the Bayes factor can be expressed as

$$BF(u = 1|y, \lambda, M) = \frac{m(y|\lambda \to \infty, M)}{m(y|\lambda, M)}. \tag{45}$$

Eq. (44) illustrates how the Bayes factor of a sharp hypothesis is influenced by the prior even when the posterior is robust to the prior. The situation occurs when the marginal likelihood of the unrestricted model is sensitive to the prior. For example, the maximum entropy prior causes the Bayes factor in favor of the ASV model to get closer and closer to one for more and more negative values of λ.

We use the prior $\pi(u|\lambda = -10, M)$ to compute the Bayes factor favoring the ASV model and the prior $\pi(u|\lambda = 10, M)$ for the Bayes factor favoring the ASV-t. These priors give the benefit of doubt to the restricted ASV-DPM model. In Figure 6 we graph in subplot (a) the prior $\pi(u|\lambda = 10, M)$ and in subplot (b) the prior $\pi(u|\lambda = -10, M)$. Each plot also contains the empirical distribution of u using the respective prior.

Using these two maximum entropy priors, the Savage-Dickey density ratio for the two sharp hypothesis reinforces the findings of Section 5. In the case where the prior lends support to a ASV model ($\lambda = -10$), the density ratio at $u_0 = 0$ in Figure 6(b) again shows there to be a 1-in-10 chance of a ASV model. The Savage-Dickey density ratio goes to zero in Figure 6(a) when the prior heavily weights ASV-DPM models having a large number of clusters. Neither prior changes the likelihood of the data coming from a ASV-t model. In each case $\pi(u = 1|y, M)$ is zero.

Figure 6 also shows how the expected number of posterior clusters is robust to the prior. As was the case with the uniform prior, the data supports a u between 0.125 and 0.5 (0.75) when $\lambda = -10$ ($\lambda = 10$). Again these are sharp hypothesis that the data can support.

24

Figure 6: The the empirical posterior density of u under the maximum entropy priors (a) $\pi(u|\lambda = 10)$ and (b) $\pi(u|\lambda = -10)$.
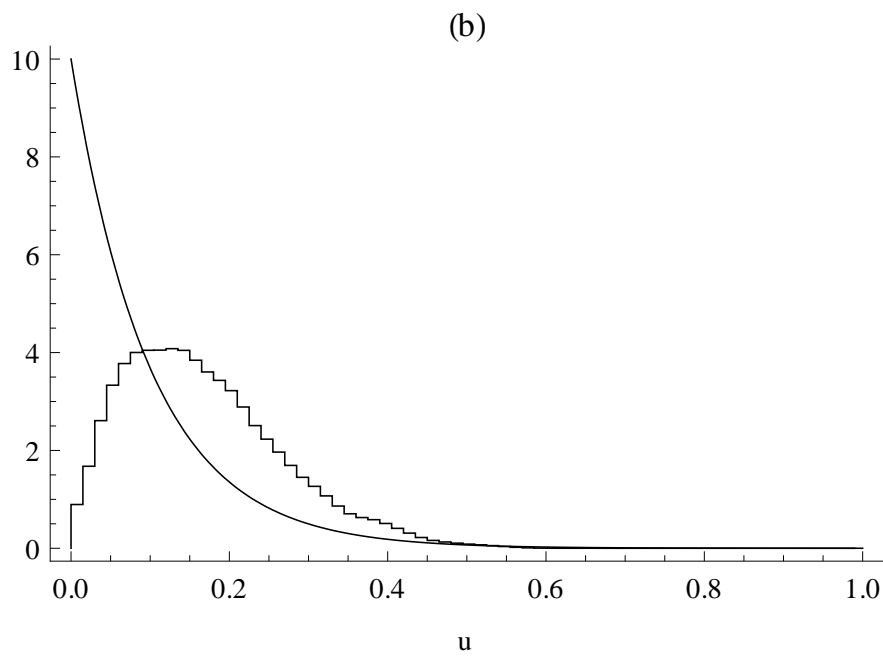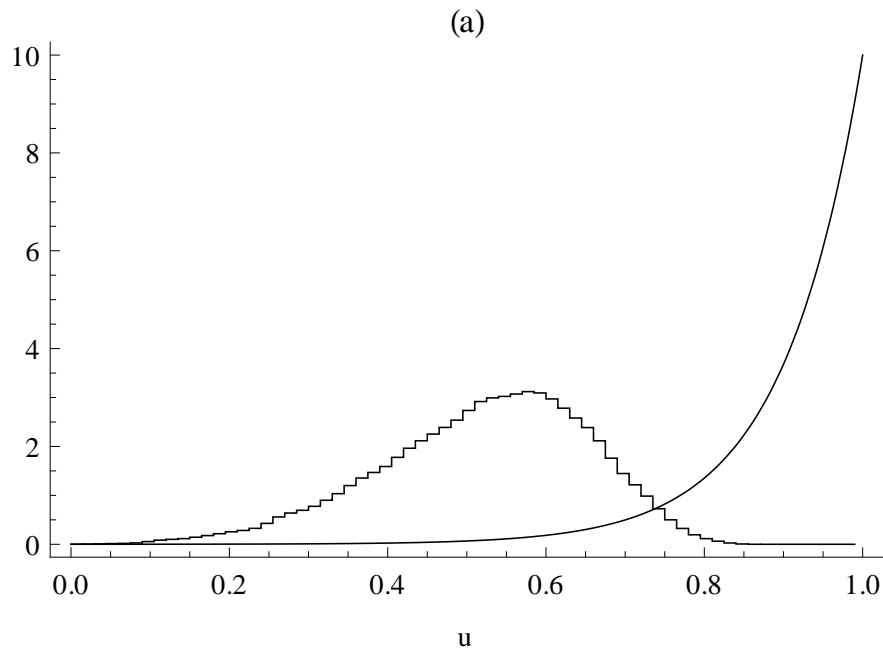
6 Predictability

To compare the ASV-DPM model with the ASV model, and more generally other stochastic volatility models, we compute each model's marginal likelihood using the product of its one-step-ahead predictions. As we have mentioned, in addition to integrating out parameter uncertainty, the marginal likelihood of stochastic volatility models also integrates out the uncertainty associated with the latent volatilities. In the past, particle filters have been applied to stochastic volatility models to integrate out volatility (see Chib et al. (2002)). However, the marginal likelihood for the ASV-DPM model also requires integrating out the latent DP covariance matrices. Basu and Chib (2003) have a way of doing this but only for a DPM type model, not a DPM model with stochastic volatility. The ASV-DPM model requires a particle filter for integrating out the latent volatilities and DP parameters, making the Basu and Chib approach infeasible.[5]

Because of the additional computation costs involved in integrating out the volatilities and DPM covariances, and also because of the increased availability of parallel computing, Beowulf clusters, and quad-core processors, we choose to compute the ASV model's marginal likelihood sequentially with 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 the histories 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

$$m(y|M) = \prod_{t=1}^{n} f(y_t|y^{t-1}, M), \tag{46}$$

where $M = $ ASV-DPM, ASV denotes the particular model. The selected priors overly influence the predictive likelihoods for y^t with small t. So, in practice the product in Eq. (46) does not begin at $t = 1$. In our empirical calculations we use $m_L(y|M) = \prod_{t=L}^{n} f(y_t|y^{t-1}, M)$, where $L = 216$.

The marginal likelihood in Eq. (46) is a recursive computation in which all parameter uncertainty is integrated out. Each one-step-ahead predictive likelihood is approximated by

---

[5]An appealing alternative to the approach we take here is the sequential Monte Carlo method to estimating and filtering DPM type model of Carvalho et al. (2010).

averaging out the randomly sampled draws of the unknown parameters, μ and φ, 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

$$
\begin{aligned}
f\left(y_t|y^{t-1}, \text{ASV-DPM}\right) &= \int f(y_t|h_t, \mu, \varphi, \{\Lambda\}, \alpha)\pi(h_t, \mu, \varphi, \{\Lambda\}, \alpha|y^{t-1}) \\
&\quad \times d(h_t, \mu, \varphi, \{\Lambda\}, \alpha), \\
&\approx R^{-1}\sum_{l=1}^{R} f\left(y_t \left| \mu^{(l)}, h_t^{(l)}, \left\{\sigma_{y,j}^{2(l)}, n_j^{(l)}\right\}_{j=1}^{k^{(l)}}, \alpha^{(l)}\right.\right),
\end{aligned}
\tag{47}
$$

where

$$
\begin{aligned}
f\left(y_t \left| \mu^{(l)}, h_t^{(l)}, \left\{\sigma_{y,j}^{2(l)}, n_j^{(l)}\right\}_{j=1}^{k^{(l)}}, \alpha^{(l)}\right.\right) &= \\
\frac{\alpha^{(l)}}{\alpha^{(l)} + t - 1} f_{St}\left(y_t \left| \mu^{(l)}, \left(\frac{s_{11}e^{h_t^{(l)}}}{v_0 - 1}\right)^{-1}, v_0 - 1\right.\right) & \\
+ \frac{1}{\alpha^{(l)} + t - 1}\sum_{j=1}^{k^{(l)}} n_j^{(l)} f_N\left(y_t \left| \mu^{(l)}, e^{h_t^{(l)}}\sigma_{y,j}^{2(l)}\right.\right), &
\end{aligned}
\tag{48}
$$

with the $h_t^{(l)}$s being random draws from $\pi\left(h_t|y^{t-1}, h_{t-1}^{(l)}\right) \equiv N\left(\bar{h}_t^{(l)}, \bar{\sigma}_{h_t}^{(l)2}\right)$ where

$$
\begin{aligned}
\bar{h}_t^{(l)} &= \varphi^{(l)}h_{t-1}^{(l)} + \frac{(y_{t-1} - \mu^{(l)})\rho_{s_{t-1}}^{(l)}\sigma_{h,s_{t-1}}^{(l)}e^{-h_{t-1}^{(l)}/2}}{\sigma_{y,s_{t-1}}^{(l)}}, \\
\bar{\sigma}_{h_t}^{(l)2} &= \left(\frac{\rho_{s_{t-1}}^{(l)2}}{(1 - \rho_{s_{t-1}}^{(l)2})\sigma_{h,s_{t-1}}^{(l)2}} + \frac{1}{\sigma_{h,s_{t-1}}^{(l)2}}\right)^{-1},
\end{aligned}
$$

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

Each of the $n - L$ MCMC posterior conditional samplers is independent from the others. Given this independence we only need to supply a particular sampler with one of the $n - L$ histories y^{t-1} before letting it run. First, we farm out as many histories as there are processors available. On a quad-core, multi-threaded, computer this generally equals 10 potential MCMC samplers. When a processor's task of sampling R draws from the posterior distribution conditioned on that particular history is completed, the processor computes the one-step-ahead predictive likelihood in Eq. (48) and returns it for later use. If predictive likelihoods for other histories still need to be computed, the processor will request another history and sample from its posterior. Once all $n - L$ predictive likelihoods have been computed the marginal likelihood is calculated.

6.1 Cumulative Bayes-factor

From the one-step-ahead predictive likelihoods, $f(y_t|y^{t-1}, \text{ASV-DPM})$ and $f(y_t|y^{t-1}, \text{ASV})$, $t = L, \ldots, n$, the cumulative log-Bayes factor (CLBF) for the two models is equal to

$$\log \left(\frac{m_L(y^\tau|\text{ASV-DPM})}{m_L(y^\tau|\text{ASV})} \right) = \sum_{t=L}^{\tau} \log \left(\frac{f(y_t|y^{t-1}, \text{ASV-DPM})}{f(y_t|y^{t-1}, \text{ASV})} \right), \quad \tau = L, \ldots, n. \quad (49)$$

Each point on the CLBF represents the log-Bayes factor comparing the two models for the data up to y_τ. By plotting the CLBF over τ we are able to identify those instances where one model out predicts the other. Its day-to-day change depicting how well the models perform relative to one another in forecasting the next day's return as illustrated by the quantities $\log f(y_t|y^{t-1}, \text{ASV-DPM}) - \log f(y_t|y^{t-1}, \text{ASV})$.

Because of the enormous computational costs involved in computing the predictive likelihoods over the lengthy return series of Section 4, we investigate the cumulative log-Bayes factors for the ASV-DPM and ASV models using value-weighted CRSP portfolio returns from January 3, 2006 to December 31, 2008 (755 trading days). For the two models we compute 540, one-day-ahead predictive likelihoods, $f(y_t|y^{t-1}, M)$, from $t = 216$ (Nov. 8, 2006) to 755 (Dec 31, 2008), by sampling the model's unknowns 11,000 times and discarding the first 1,000 draws.

The cumulative log-Bayes factors for $\tau = 216, \ldots, 755$, are plotted in the top panel of Figure 7. In the bottom panel of the figure, we plot the daily return for the CRSP value-weighted portfolio over the corresponding period. From the initial CLBF value of zero, the ASV-DPM and ASV models are equivalent in their prediction of y_{215}. The CLBF then climbs steadily in favor of the ASV-DPM model while experiencing a few temporary drops, until it briefly enjoys a period during the summer of 2008 where it exceeds five before declining during the financial crisis. This general upward trend in the CLBF reflects a ASV-DPM model producing better day-ahead predictions of market returns than the ASV model. According to Jefferies (1961), the final CLBF value of 4.1 is "strong" evidence in favor of the ASV-DPM model over the ASV model.

There are a few days and time periods where the CLBF either quickly jumps, or rapidly climbs upwards – the most notable being the jump on Feb. 27, 2007. On this day the market portfolio declined 3.4 percent. By comparing the CLBF to the plotted returns in the bottom panel of Figure 7, we find that those periods of rapid increase in the CLBF often correspond with a shift from a low to high volatility state. For instance, in the middle of the year 2007 the CLBF begins to climb from a local low just as returns become more volatile. However, it is also the steady strength of the ASV-DPM model's daily prediction

Figure 7: Cumulative log-Bayes factor of the ASV-DPM model relative to the ASV model (red line) using return data back to January 3, 2006, for the period of November 8, 2006 to December 31, 2008 and CRSP value-weighted portfolio returns (blue line).
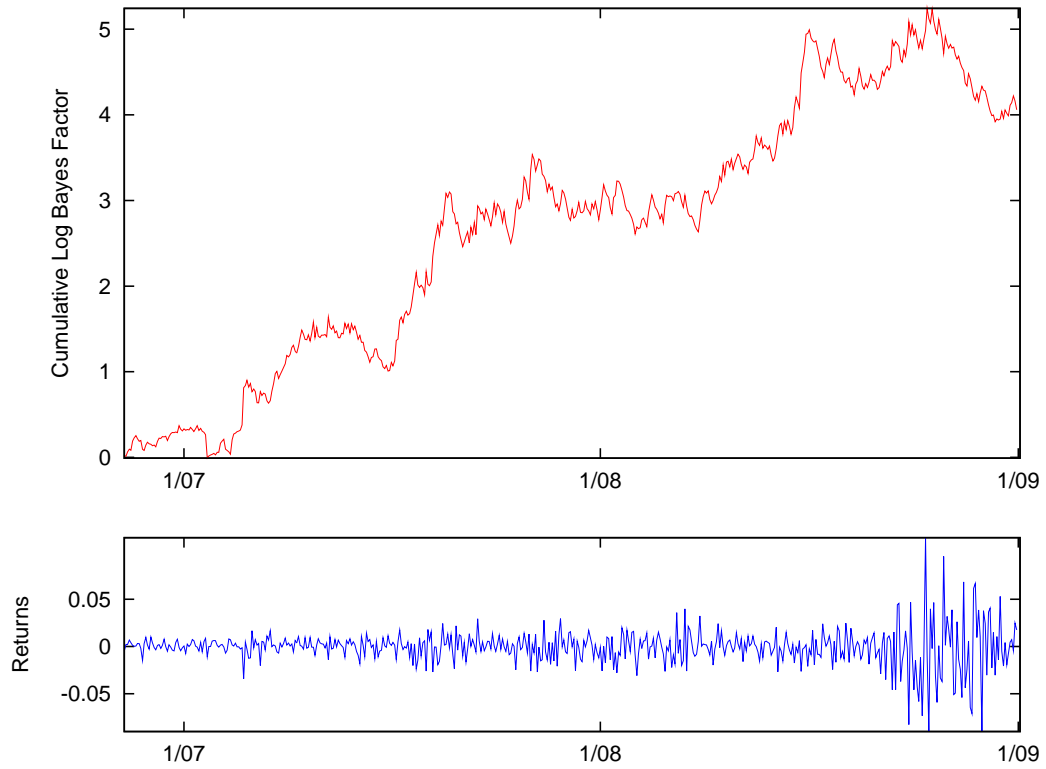
over the entire period that results in it being the "better" model.

7 Volatility Response

Ever since Black (1976) proposed the leverage hypothesis and French et al. (1987) the volatility feedback effect, many have studied how volatility reacts to changes in market returns.[6] Yu (2005) and Asai and McAleer (2009) both establish the theoretical relationship for stochastic volatility and returns when leverage is present. Yu (2005) derives the volatility-return relationship for the asymmetric stochastic volatility model and Asai and McAleer (2009) for the multivariate stochastic volatility model, but neither empirically investigates the relationship.

Volatility's response in the ASV-DPM to a change in market returns is found in Eq. (33) – the joint posterior predictive density, $f((y_{n+1}, h_{n+2})'|y^n)$. This predictive density dispenses with parameter and latent volatility uncertainty by integrating them over their posterior.[7] We compute $f((y_{n+1}, h_{n+2})'|y^n)$ over a 200×200 grid of equally spaced values of $(y_{n+1}, h_{n+2})'$ centered around the market return, y_{n+1}.

The ASV-DPM's predictive joint density contains a healthy amount of information concerning the return-volatility relationship. However, since the predictive posterior density is bivariate and depends on today's return, y_n, it is difficult to visualize this information in a density plot. Instead, we choose to summarize some of the return-volatility relationship by computing and plotting the conditional expectation of the next period's log-volatility, $E[h_{n+2}|y^n, y_{n+1}]$, over a range of values for y_{n+1}.

We compute the ASV-DPM model's joint predictive density using the MCMC draws from Section 6 and the CRSP portfolio returns from January 3, 2006 to December 31, 2008. In Figure 8, we plot fifteen of the contour lines from the predictive density (solid lines) along with $E[h_{n+2}|y^n, y_{n+1}]$ (dashed line). In the figure, the contour lines bow up and out when tomorrow's return is negative, while the contours are nearly linear over small values of h_{n+2}; i.e., conditional on returns being negative, the predictive distribution for h_{n+1} is not symmetrical but is skewed upward. Skewness is also present when $y_{n+1} > 0$. However, these contour lines are less (more) bowed out for large (small) values of h_{n+1}. When market returns are positive, the conditional distribution of tomorrow's log-volatility is skewed upward, but less so relative to when y_{n+1} is negative. Thus, tomorrow's volatility

---

[6]See Bekaert and Wu (2000) for a review of the research prior to 2000 and Chen and Ghysels (2011) for more recent work on the subject.

[7]The volatility-return relationship of Yu (2005) and Asai and McAleer (2009) do not integrate out this uncertainty and depend on the value of the estimated parameters.

is likely to be higher following either a market decline or increase, but because of the difference in the degree of skewness, volatility's expected response is asymmetric.

Figure 8: Contour lines (solid lines) of the ASV-DPM model's joint predictive density, $f((y_{n+1}, h_{n+2})'|y^n)$, and the conditional expected value of log-volatility given tomorrow's return, $E[h_{n+2}|y^n, y_{n+1}]$, (dashed line) where y^n contains 755 daily CRSP value-weighted portfolio returns from January 3, 2006, to December 31, 2008.
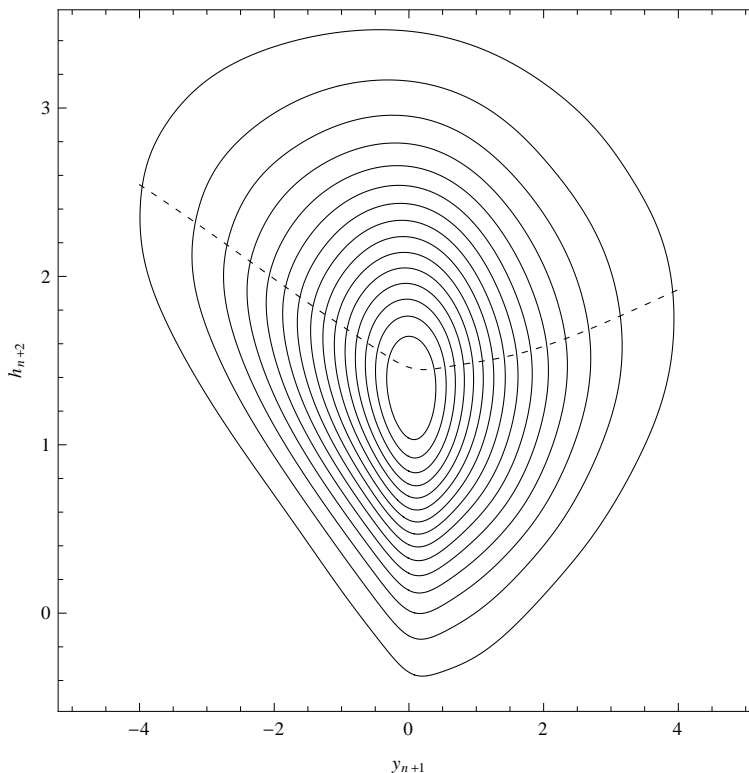

This asymmetric response in volatility is seen in the dashed line of $E[h_{n+2}|y^n, y_{n+1}]$. When the market does not move, tomorrow's expected log-volatility is 1.46. With each percentage point decline in the daily market return, expected log-volatility increases at approximately the rate of 0.3. This contrasts with a market increase, where expected log-volatility responds in a nonlinear manner. Expected log-volatility ever so slightly declines for market gains smaller than 0.2%. increasing at a rate of 0.1. But as returns get even larger expected log-volatility rises at a faster rate of 0.2. Even though this rate of increase in expected log-volatility is smaller than had the market dropped, tomorrow's expected volatility will be larger after any sizable move in the market, be it negative or positive. Thus, in the words of Campbell and Hentschel (1992) and Chen and Ghysels (2011),"No

news is good news," when talking about market returns and volatility.

7.1 Time-varying volatility response

Like volatility, the volatility-return relationship can vary. This begs the question, does the dashed line in Figure 8 represent the typical joint predictive distribution of tomorrow's market return and log-volatility, or is it an abnormality? To answer this question we delete the last 540 returns from the return history and sequentially estimate the volatility-return relationship beginning with the return history ending on November 8, 2006 and adding one return at a time until we reach December 31, 2008.

Figure 9: The ASV-DPM conditional expected log-volatility, $E[h_{n+2}|y^n, y_{n+1}]$, plotted against future return, y_{n+1}, for the return series, $y^n = (y_1, \ldots, y_n)$, $n = 216, \ldots, 755$.
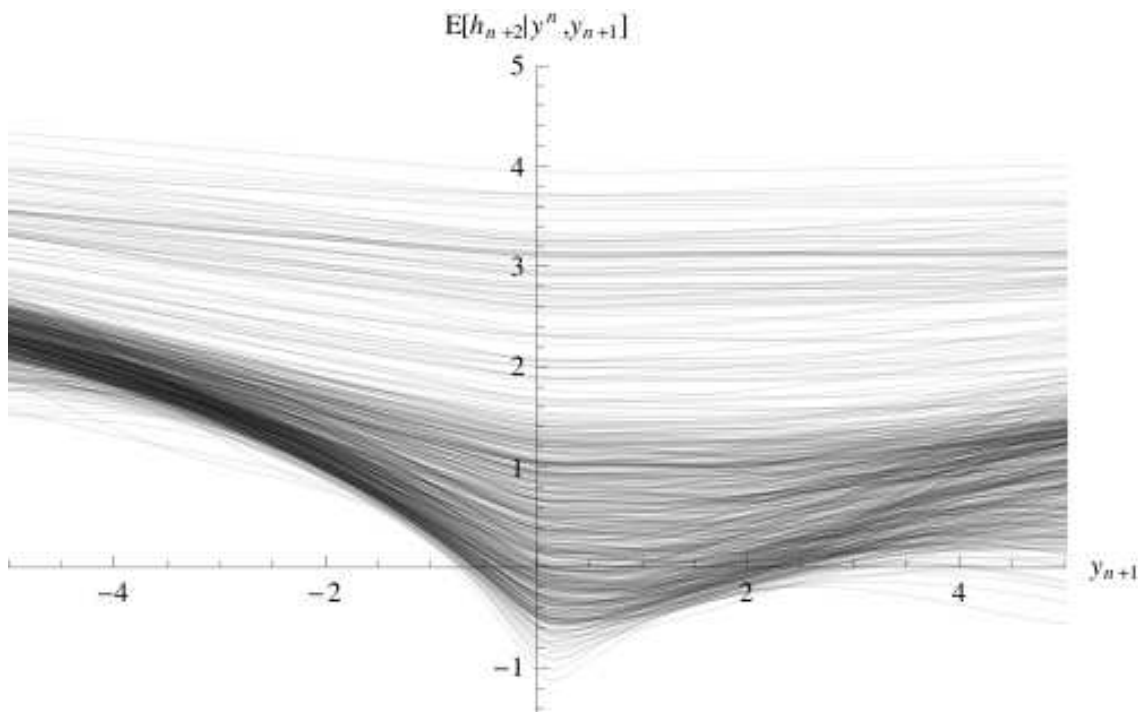
In Figure 9, we plot $E[h_{n+2}|y^n, y_{n+1}]$ for each of the 540 return histories, y^n, $n = 216, \ldots, 755$. This figure illustrates how the volatility-return relationship depends on the return history. The value of $E[h_{n+2}|y^n, y_{n+1} = 0]$ ranges from -1 to 4, with the highest concentration falling between -0.5 to 1.5. Histories associated with the largest $E[h_{n+2}|y^n, y_{n+1} = 0]$ end during the most volatile period of September, 2008. The $E[h_{n+2}|y^n, y_{n+1} = 0]$ near the origin have y^n that end during calmer more normal market periods. Current

volatility explains 89 percent of the variation in $E[h_{n+2}|y^n, y_{n+1} = 0]$ when regressing it on $E[h_{n+1}|y^n]$. Hence, as one would expect tomorrow's expected volatility is strongly correlated with today's volatility.

Many studies of the leverage effect assume that debt is unaffected by changes in the stock market, and is hence, riskless; e.g. see Schwert (1989) and Figlewski and Wang (2000). Christie (1982) shows that by allowing the value of debt to change in the same direction as market capital, the leverage effect is weakened. Under risky debt, a negative market return leads to an increase in volatility, but the size of the impact declines as leverage grows. By most practitioners, leverage was viewed as being high during 2008. The return histories ending during this period have the highest values for $E[h_{n+2}|y^n, y_{n+1}]$, but also the flattest.

The degree of asymmetry in $E[h_{n+2}|y^n, y_{n+1}]$ is also affected. Asymmetry is greatest when $E[h_{n+2}|y^n, y_{n+1} = 0]$ is negative and today's volatility is low. At the other end of the spectrum, volatility's response to the market is nearly symmetrical when $E[h_{n+2}|y^n, y_{n+1} = 0]$ is greater than 2 and the return history ends during volatile bear markets. Volatility response is not only symmetrical for these histories, there is almost no volatility response at all.

The degree of shading in Figure 9 provides a rough density measure of the different volatility-return responses. Dark bands over the negative returns show that for many of the 540 volatility-return responses a drop in the market today corresponds to an increase in expected log-volatility. A similar but not quite as dark a band can also be seen for positive daily returns.[8] The lighter lines are the volatility-return relationships whose histories end during rare, but very volatile, times. As mentioned above, these responses are symmetrical and nearly flat. Thus, we see that during "normal" times, no news is good news for volatility, which is even more true during tranquil times. During the most turbulent of markets, news of any sort just does not matter for volatility. This is a notable feature of the semiparametric ASV model. When markets are highly volatile, a return shock must be large in order to affect the market's expectations about tomorrow's volatility. Whereas, on a typical day only a little bit of news is required to cause expected volatility to increase.

8 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

---

[8]The dark band in Figure 9 has a shape very similar to Chen and Ghysels (2011) nonparametric news impact curve on p. 49.

distribution with a nonparametric, bivariate 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, while being manageable and parsimonious by taking on a finite number of correlations for a finite length data set. This flexibility is important when forecasting market returns, especially when the market transitions from a low to high volatility state or the market suddenly declines. In the empirical case study, forecasts from our semiparametric asymmetric stochastic volatility model capture these types of episodes, 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.

The flexibility of having more than one value to model the correlation between volatility and return is also important for modeling the volatility-return relationship and the response in expected volatility to a unexpected change in market returns. In particular, the size and and degree of asymmetry in the response of volatility to an unexpected change in market prices can vary. Using the semiparametric model of the asymmetric stochastic volatility, the response in the expected value of volatility to a decline in stock prices versus an increase is highly asymmetric when volatility is currently low and the market is calm. However, if volatility is high and the market irregular, the asymmetry nearly disappears and the response in expected volatility becomes muted. In other words, during normal times, just a little bit of news affects volatility, whereas, during a turbulent market, it takes a significant amount of news to impact volatility.

34

References

Asai, M. and McAleer, M.: 2009, Multivariate stochastic volatility, leverage and news impact surfaces, *Econometric Journal* **12**, 292–309.

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.

Bekaert, G. and Wu, G.: 2000, Asymmetric volatility and risk in equity markets, *Review of Financial Studies* **13**, 1–42.

Black, F.: 1976, Studies in stock price volatility changes, *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section of the American Statistical Association* pp. 177–181.

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

Campbell, J. Y. and Hentschel, L.: 1992, No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* **31**, 281–318.

Carvalho, C. M., Lopes, H. F., Polson, N. G. and Taddy, M. A.: 2010, Particle learning for general mixtures, *Bayesian Analysis* **5**(4), 709–740.

Chen, X. and Ghysels: 2011, News - good or bad - and its impact on volatility predictions over multiple horizons, *Review of Financial Studies* **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.

Christie, A. A.: 1982, The stochastic behavior of common stock variances: Value, leverage and interest rate effects, *Journal of Financial Economics* **10**, 407–432.

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.

Delatola, E.-I. and Griffin, J. E.: 2011a, Bayesian nonparametric modelling of the return distribution with stochastic volatilty, *Bayesian Analysis* **6**, 1–26.

Delatola, E.-I. and Griffin, J. E.: 2011b, A Bayesian semiparametric model for volatility with a leverage effect, *Technical report*, University of Kent.

Dickey, J. D.: 1971, The weighted likelihood ratio, linear hypotheses on normal location parameters, *Annals of Mathematical Statistics* **42**, 204–223.

Durham, G. B.: 2007, SV mixture models with application to S&P 500 index return, *Journal of Financial Economics* **85**, 822–856.

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.

Ferguson, T.: 1973, A Bayesian analysis of some nonparametric problems, *The Annals of Statistics* **1**(2), 209–230.

Figlewski, S. and Wang, X.: 2000, Is the "leverage effect" a leverage effect?, *Technical report*, NYU Stern School of Business.

French, K. R., Schwert, G. W. and Stambaugh, R. F.: 1987, Expected stock returns and volatility, *Journal of Financial Economics* **19**, 3–29.

Geweke, J.: 2001, Bayesian econometric and forecasting, *Journal of Econometrics* **100**, 11–15.

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

Griffin, J. E.: 2011, Default priors for density estimation with mixture models, *Bayesian Analysis* **5**, 45–64.

Griffin, J. E. and Steel, M. F. J.: 2006, Ordered-based dependent Dirichlet processes, *Journal of the American Statistical Association* **101**, 179–194.

Griffin, J. E. and Steel, M. F. J.: 2011, Stick-breaking autoregressive processes, *Journal of Econometrics* **162**, 383–396.

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.: 1994, Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics* **12**, 371–417.

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

Jensen, M. J.: 2004, Semiparametric Bayesian inference of long-memory stochastic volatility models, *Journal of Time Series Analysis* **25**(6), 895–922.

Jensen, M. J. 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.

Nelson, D. B.: 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**, 347–370.

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.

Poon, S. and Granger, C. W.: 2003, Forecasting volatility in financial markets: A review, *Journal of Economics Literature* **41**, 478–539.

Richardson, S. and 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.

Schwert, G.: 1989, Why does stock market volatility change over time?, *Journal of Finance* **44**, 1115–1154.

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.

Yu, J.: 2011, A semiparametric stochastic volatility model, *Journal of Econometrics* **176**, 473–482.

Zellner, A.: 1971, *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York, New York.