Improving Markov Switching Models using Realized Variance*

Jia Liu † John M. Maheu ‡

This draft June 2017

Abstract

This paper proposes a class of models that jointly model returns and ex-post variance measures under a Markov switching framework. Both univariate and multivariate return versions of the model are introduced. Estimation can be conducted under a fixed dimension state space or an infinite one. The proposed models can be seen as nonlinear common factor models subject to Markov switching and are able to exploit the information content in both returns and ex-post volatility measures. Applications to equity returns compare the proposed models to existing alternatives. The empirical results show that the joint models improve density forecasts for returns and point predictions of return variance. Using the information in ex-post volatility measures can increase the precision of parameter estimates, sharpen the inference on the latent state variable and improve portfolio decisions.

Keywords: Infinite hidden Markov model, Realized covariance, Density forecast, MCMC, Portfolio choice

*We thank the editor Andrew Patton and two anonymous referees for many helpful comments that resulted in a significant revision of the paper. We thank Qiao Yang and seminar participants at the CEA annual meetings 2015, CESG 2015 and MEG 2015 for comments. Maheu is grateful to the SSHRC (grant 435-2014-2015) for financial support.

†DeGroote School of Business, McMaster University, Canada, liuj46@mcmaster.ca

‡DeGroote School of Business, McMaster University, Canada and RCEA, Italy, maheujm@mcmaster.ca

1 Introduction

This paper proposes a new way of jointly modelling return and ex-post volatility measures under a Markov switching framework. Parametric and nonparametric versions of the models are introduced in both univariate and multivariate settings. The models are able to exploit the information content in both return and ex-post volatility series. Compared to existing models, the joint models improve density forecasts of returns, point predictions of realized variance and portfolio decisions.

Since the pioneering work by Hamilton (1989) the Markov switching model has became one of the standard econometric tools in studying various financial and economic data series. The basic model postulates a discrete latent variable governed by a first-order Markov chain that directs an observable data series. This modelling approach has been fruitfully employed in many applications. For instance, Markov switching models have been used to identify bull and bear markets in aggregate stock returns (Maheu and McCurdy, 2000; Lunde and Timmermann, 2004; Guidolin and Timmermann, 2006; Maheu et al., 2012), to capture the risk and return relationship (Pastor and Stambaugh, 2001; Kim et al., 2004), for portfolio choice (Guidolin and Timmermann, 2008), for interest rates (Ang and Bekaert, 2002; Guidolin and Timmermann, 2009) and foreign exchange rates (Engel and Hamilton, 1990; Dueker and Neely, 2007). Recent work has extended the Markov switching model to an infinite dimension. The infinite hidden Markov model (IHMM), which is a Bayesian nonparametric model, allows for a very flexible conditional distribution that can change over time. Applications of IHMM include Jochmann (2015), Dufays (2016), Song (2014), Carpantier and Dufays (2014) and Maheu and Yang (2016).

Realized variance (RV), constructed from intraperiod returns, is an accurate measure of ex-post volatility. Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) formalized the idea of using higher frequency data to measure the volatility of lower frequency data and show RV is a consistent estimate of quadratic variation under ideal conditions. Barndorff-Nielsen and Shephard (2004b) generalized the idea of RV and introduced a set of variance estimators called realized power variations. Furthermore, RV has been extended to realized covariance (RCOV), which is an ex-post nonparametric measure of the covariance of multivariate returns, by Barndorff-Nielsen and Shephard (2004a). A survey of RV and related volatility proxies is Andersen and Benzoni (2009).

This paper is the first to exploit the information content of ex-post volatility measures in a Markov switching context to improve estimation, forecasting and portfolio decisions.[1] This

[1] Other papers that have used high frequency data to improve estimation and forecasting include Alizadeh et al. (2002), Blair et al. (2001), Shephard and Sheppard (2010), Noureldin et al. (2012), Hansen et al. (2012) Hansen et al. (2014), Jin and Maheu (2013, 2016) and Maheu and McCurdy (2011).

is done for finite and infinite state models. We assume that regime changes in returns and realized variance are directed by one common unobserved state variable. Closely related to our approach is Takahashi et al. (2009) who propose a stochastic volatility model in which unobserved log-volatility affects both RV and the variance of returns. They find improved fixed parameter and latent volatility estimates but do not investigate forecast performance or portfolio choice.

There is no reason to confine attention to RV, and therefore we investigate the use of other volatility measures and in the multivariate setting realized covariance. Four versions of the univariate return models are proposed. We consider RV, log(RV), realized absolute variation (RAV), or log(RAV) as ex-post volatility measures coupled with returns to construct joint models. We then extend the MS-RV specification to its multivariate version with RCOV.

It is more flexible to drop the finite state assumption and let the data determine the number of states needed to fit the data. Using Bayesian nonparametric techniques, we extend the finite state joint MS models to nonparametric versions. These models allow the conditional distribution to change more flexibly and accommodate any nonparametric relationship between returns and ex-post volatility.

Markov switching models have been particularly useful at monthly and quarterly frequencies as the Markov chain dynamics are a dominate feature of the data.[2] Therefore, the proposed joint MS and joint IHMM models are compared to existing models in empirical applications to monthly U.S. stock market returns including forecasting and portfolio applications.

Based on the log-predictive Bayes factors, the proposed joint models strongly dominate the models that only use returns. Moreover, we find the gains from joint modelling are particularly large during high volatility episodes. The empirical results also show that the joint models reduce the error in predicting realized variance. With the help of additional information offered by RV, RAV and RCOV, posterior density intervals for parameters are tighter and the inference on the unobservable state variables is improved. In general, adding RV to a model improves the forecast performance of any MS model, finite or infinite, but the best performing models are the joint infinite hidden Markov models.

Exploiting measures of ex-post volatility also lead to better portfolio choice outcomes. Several portfolio exercises that use the models are considered over two sample periods. A robust result is larger Sharpe ratios for models that incorporate realized variance or realized covariance. In addition, investors are always willing to pay a positive performance fee to obtain forecasts from these models for their investment decisions.

[2] Our models could be applied to higher frequency data but in this case additional structure would be needed to capture the strong persistence in high frequency volatility Rydén et al. (1998).

This paper is organized as follows. In Section 2, we show how to incorporate ex-post measures of volatility into Markov switching models. The joint MS models are extended to the nonparametric versions in Section 3. Benchmark models used for comparison are found in Section 4. Section 5 illustrates the Bayesian estimation steps and model comparison. A univariate return application to the market index is found in Section 6 while a multivariate return application to equity follows in Section 7. The next section concludes followed by an appendix that gives detailed steps of posterior simulation.

2 Joint Markov Switching Models

In this section, we will focus on simple specifications of the conditional mean but dynamic models with lags of the dependent variables could be used. We will first discuss the four versions of univariate return joint models, then introduce the multivariate version.

Higher frequency data is used to construct ex-post volatility measures. Let $r_{t,i}$ denote the i^{th} intraperiod continuously compounded return in period t, $i = 1, \dots, n_t$, where n_t is the number of intraperiod returns. Then the return and realized variance from $t-1$ to t is

$$r_t = \sum_{i=1}^{n_t} r_{t,i}, \tag{1}$$

$$RV_t = \sum_{i=1}^{n_t} r_{t,i}^2. \tag{2}$$

Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) formalized the idea of using higher frequency data to measure the volatility of r_t. They show that RV_t is a consistent estimate of quadratic variation under ideal conditions.[3] Similarly, for multivariate returns $R_{t,i}$ is the i^{th} intraperiod $d \times 1$ return vector at time t and the time t return is $R_t = \sum_{i=1}^{n_t} R_{t,i}$. $RCOV_t$ denotes the associated realized covariance (RCOV) matrix which is computed as follows,

$$RCOV_t = \sum_{i=1}^{n_t} R_{t,i} R_{t,i}^{'}. \tag{3}$$

[3]We have not made adjustments for market microstructure dynamics since our high-frequency data consists of daily returns and are relatively clean. Nevertheless, any of the existing approaches that correct for microstructure dynamics in computing ex-post volatility measures could be used.

2.1 MS-RV Model

We first use RV as the proxy for ex-post volatility to build a joint MS-RV model. The proposed K-state MS-RV model is given as follows.

$$\begin{aligned} r_t \big| s_t &\sim \mathrm{N}(\mu_{s_t}, \sigma^2_{s_t}), & (4)\\ RV_t \big| s_t &\sim \mathrm{IG}(\nu+1, \nu\sigma^2_{s_t}), & (5)\\ P_{i,j} &= p(s_{t+1}=j|s_t=i), & (6) \end{aligned}$$

where $s_t \in \{1,\ldots,K\}$. Conditional on state s_t, RV_t is assumed to follow an inverse-gamma distribution[4] $\mathrm{IG}(\nu+1, \nu\sigma^2_{s_t})$, where $\nu+1$ is the shape parameter and $\nu\sigma^2_{s_t}$ is the scale parameter.

The basic assumption of this model is that RV_t is subject to the same regime changes as r_t and share the same parameter $\sigma^2_{s_t}$.[5] Note, that RV_t and the other volatility measures used in this paper, are assumed to be a noisy measure of the state dependent variance $\sigma^2_{s_t}$. Conditional on the latent state, the mean and variance of RV_t are

$$\begin{aligned} E(RV_t|s_t) &= \frac{\nu\sigma^2_{s_t}}{(\nu+1)-1} = \sigma^2_{s_t}, & (7)\\ \mathrm{Var}(RV_t|s_t) &= \frac{(\sigma^2_{s_t})^2}{(\nu-1)}. & (8) \end{aligned}$$

Therefore RV_t is centered around $\sigma^2_{s_t}$, but in general, not equal to it. The variance of the distribution of RV_t is positively correlated with the realized variance itself. During high volatility periods, the movements of realized variances are more volatile. Both the return process and realized variance process are governed by the same underlying Markov chain with transition matrix P.

Since $\sigma^2_{s_t}$ influences both the return process and RV_t process, the model can be seen as a nonlinear common factor model.[6] Exploiting the information content of RV_t for $\sigma^2_{s_t}$ can lead to more precise estimates of model parameters, state variables and forecasts.

[4]If $x \sim \mathrm{IG}(\alpha,\beta)$, $\alpha>0$, $\beta>0$ then it has density function:

$$g(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right).$$

The mean of x is $E(x) = \frac{\beta}{\alpha-1}$ for $\alpha>1$ and $\mathrm{Var}(x) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha>2$.

[5]Formally, the high frequency data generating process is assumed to be $r_{t,i} = \mu_{s_t}/n_t + (\sigma_{s_t}/\sqrt{n_t})z_{t,i}$, with $z_{t,i} \sim \mathrm{NID}(0,1)$. Then $E\left[\sum_{i=1}^{n_t} r^2_{t,i}|s_t\right] = \mu^2_{s_t}/n_t + \sigma^2_{s_t} \approx \sigma^2_{s_t}$ when the term $\mu^2_{s_t}/n_t$ is small due to n_t being large.

[6]This is in contrast to linear dynamic factor models such as Forni and Reichlin (1998), Kose et al. (2003) and Stock and Watson (2010).

2.2 MS-logRV Model

Another possibility is to model the logarithm of RV as normally distributed. The MS-logRV model is shown as follows,

$$
\begin{aligned}
r_t \big| s_t &\sim \mathrm{N}\left(\mu_{s_t}, \exp(\zeta_{s_t})\right), &(9)\\
\log(RV_t) \big| s_t &\sim \mathrm{N}\left(\zeta_{s_t} - \frac{1}{2}\delta_{s_t}^2, \delta_{s_t}^2\right), &(10)\\
P_{i,j} &= p(s_{t+1} = j | s_t = i), &(11)
\end{aligned}
$$

where $s_t \in \{1, \ldots, K\}$. In this model there are three state-dependent parameters: μ_{s_t}, ζ_{s_t} and $\delta_{s_t}^2$, which enable both the mean and variance of returns and $\log(RV_t)$ to be state-dependent. $\zeta_{s_t} - \frac{1}{2}\delta_{s_t}^2$ is the mean of $\log(RV_t)$ and $\exp(\zeta_{s_t})$ is the variance of returns. Since RV_t is log-normal, $E[RV_t|s_t] = \exp(\zeta_{s_t})$ which is assumed to be the variance of returns.

2.3 MS-RAV Model

Now we consider using realized absolute variation (RAV), instead of RV in the joint MS model. Calculated using the absolute values of intraperiod returns, RAV is robust to jumps and may be less sensitive to outliers Barndorff-Nielsen and Shephard (2004b). RAV_t is computed using intraperiod returns as

$$
RAV_t = \sqrt{\frac{\pi}{2}} \sqrt{\frac{1}{n_t}} \sum_{i=1}^{n_t} |r_{t,i}|, \tag{12}
$$

where $r_{t,i}$ denotes the i^{th} intraperiod log-return in period t, $i = 1, \ldots, n_t$. It can be shown that RAV_t provides an estimate of the standard deviation of r_t.[7]

Consistent with the inverse-gamma distribution to model the variance or its proxy, we assume RAV follows a square-root inverse-gamma distribution (sqrt-IG). The density function of sqrt-IG(α, β) is given by

$$
f(x) = \frac{2\beta^\alpha}{\Gamma(\alpha)} x^{-2\alpha-1} \exp\left(-\frac{\beta}{x^2}\right), \; x > 0, \alpha > 0, \beta > 0, \tag{13}
$$

[7] As before, if $r_{t,i} = \mu_{s_t}/n_t + (\sigma_{s_t}/\sqrt{n_t}) z_{t,i}$, with $z_{t,i} \sim \mathrm{NID}(0,1)$ and μ_{s_t}/n_t is small, then we have $E\left[\sqrt{\frac{\pi}{2}}\sqrt{\frac{1}{n_t}} \sum\limits_{i=1}^{n_t} |r_{t,i}| \big| s_t\right] \approx \sqrt{\frac{\pi}{2n_t}} \frac{1}{\sqrt{n_t}} \sum_{i=1}^{n_t} E\left[|\sigma_{s_t} z_{t,i}| \big| s_t\right] = \sqrt{\frac{\pi}{2}} \frac{\sigma_{s_t}}{n_t} \sum_{i=1}^{n_t} E\left[|z_{t,i}| \big| s_t\right] = \sigma_{s_t}$, since for a standard normal variate x, $E\left[|x|\right] = \sqrt{\frac{2}{\pi}}$.

and the first and second moments of sqrt-IG(α, β) are given as follows

$$E[x] = \sqrt{\beta} \cdot \frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma(\alpha)} \text{ and } E[x^2] = \frac{\beta}{\alpha - 1}, \text{ for } \alpha > 1. \tag{14}$$

These results can be found in Zellner (1971).

We define the joint MS model of return and RAV as,

$$\begin{aligned} r_t \big| s_t &\sim \mathrm{N}(\mu_{s_t}, \sigma^2_{s_t}), & (15) \\ RAV_t \big| s_t &\sim \text{sqrt-IG}\left(\nu, \sigma^2_{s_t}\left[\frac{\Gamma(\nu)}{\Gamma(\nu - \frac{1}{2})}\right]^2\right), & (16) \\ P_{i,j} &= p(s_{t+1} = j | s_t = i), & (17) \end{aligned}$$

where $s_t \in \{1, \ldots, K\}$ and $\nu > 1$. As in the MS-RV model, the mean and variance of both r_t and RAV_t are state-dependent. In each state, the return follows a normal distribution with mean μ_{s_t} and variance $\sigma^2_{s_t}$. The mean and variance of RAV_t conditional on state s_t are given as follows.

$$\begin{aligned} E(RAV_t \big| s_t) &= \sigma_{s_t}, & (18) \\ \mathrm{Var}(RAV_t \big| s_t) &= \frac{\sigma^2_{s_t}}{\nu - 1}\left[\frac{\Gamma(\nu)}{\Gamma(\nu - \frac{1}{2})}\right]^2 - \sigma^2_{s_t}. & (19) \end{aligned}$$

2.4 MS-logRAV Model

Similar to the MS-logRV model discussed in Section 2.2, the logarithm of RAV can be modelled as opposed to RAV. The MS-logRAV specification is

$$\begin{aligned} r_t \big| s_t &\sim \mathrm{N}\left(\mu_{s_t}, \exp(2\zeta_{s_t})\right), & (20) \\ \log(RAV_t) \big| s_t &\sim \mathrm{N}\left(\zeta_{s_t} - \frac{1}{2}\delta^2_{s_t}, \delta^2_{s_t}\right), & (21) \\ P_{i,j} &= p(s_{t+1} = j | s_t = i), & (22) \end{aligned}$$

where $s_t \in \{1, \ldots, K\}$. The model is close to the MS-logRV parametrization, but now $\zeta_{s_t} - \frac{1}{2}\delta^2_{s_t}$ is the mean of $\log(RAV_t)$ and $\exp(2\zeta_{s_t})$ is the state-dependent variance of returns. Since RAV_t is log-normal, $E(RAV_t | s_t) = \exp(\zeta_{s_t})$ which is the standard deviation of returns.

2.5 Multivariate MS-RCOV Model

The univariate return models can be extended to the multivariate setting by including realized covariance matrices. The multivariate MS-RCOV model we consider is

$$R_t | s_t \sim \mathrm{N}(M_{s_t}, \Sigma_{s_t}), \tag{23}$$

$$RCOV_t | s_t \sim \mathrm{IW}(\Sigma_{s_t}(\kappa - d - 1), \kappa), \kappa > d + 1, \tag{24}$$

$$P_{i,j} = p(s_{t+1} = j | s_t = i). \tag{25}$$

where $s_t \in \{1, \ldots, K\}$. M_{s_t} is a $d \times 1$ state-dependent mean vector and Σ_{s_t} is the $d \times d$ covariance matrix. $RCOV_t$ is assumed to follow an inverse-Wishart distribution $\mathrm{IW}(\Sigma_{s_t}(\kappa - d - 1),\ \kappa)$, where $\Sigma_{s_t}(\kappa - d - 1)$ is the scale matrix and κ is the degree of freedom.

Σ_{s_t} is the covariance of returns as well as the mean of $RCOV_t$ since

$$E[RCOV_t | s_t] = \frac{1}{\kappa - d - 1} \Sigma_{s_t}(\kappa - d - 1) = \Sigma_{s_t}, \tag{26}$$

assuming $\kappa > d + 1$. The parameter κ controls the variation of the inverse-Wishart distribution and the smaller κ is, the larger spread the distribution has. Both R_t and $RCOV_t$ are governed by the same Markov chain with transition matrix P.

3 Joint Infinite Hidden Markov Model

3.1 Dirichlet Process and Hierarchical Dirichlet Process

All of the Markov switching models we have discussed require the econometrician to set the number of states. An alternative is to incorporate the state dimension into estimation. The Bayesian nonparametric version of the Markov switching model is the infinite hidden Markov model, which can be seen as a Markov switching model with infinitely many states. Given a finite dataset, the model selects a finite number of states for the system. Since the number of states is no longer a fixed value, the Dirichlet process, an infinite dimensional version of Dirichlet distribution, is used as a prior for the transition probabilities.

There are several benefits to using an infinite Markov chain. First, this flexible structure can accommodate both recurring regimes as well as structural changes. As new data arrives if a new state of the market occurs the model can introduce a new state and associated parameter to account for this. Since the model is applied to a finite dataset a finite set of states will be used and this is estimated along with the rest of the parameters. Since as many states can be used as needed the model is nonparametric in nature and able to capture an unknown continuous density that changes over time.

The Dirichlet process DP(α, H), was formally introduced by Ferguson (1973) and is a distribution of distributions. A draw from a DP(α, H) is a distribution and is almost surely discrete and centered around the base distribution H. $\alpha > 0$ is the concentration parameter that governs how close the draw is to H.

We follow Teh et al. (2006) and build an infinite hidden Markov model (IHMM) using a hierarchical Dirichlet process (HDP). This consists of two linked Dirichlet processes. A single draw of a distribution is taken from the top level Dirichlet processes with base measure H and precision parameter η. Subsequent to this, each row of the transition matrix is distributed according to a Dirichlet processes with base measure taken from the top level draw. This ensures that each row of the transition matrix governs the moves among a common set of model parameters. In addition, each row of the transition matrix is centered around the top level draw but any particular draw will differ. If Γ denotes the top level draw and P_j the j^{th} row of the transition matrix P then the previous discussion can be summarized as

$$\Gamma \big| \eta \sim \text{DP}(\eta, H), \tag{27}$$

$$P_j \big| \alpha, \Gamma \overset{iid}{\sim} \text{DP}(\alpha, \Gamma),\ j = 1, 2, ... \tag{28}$$

This formulation provides a prior over the natural numbers (states) such that each P_j has $E[P_j] = \Gamma$.[8] For more details and examples of the HDP used in the infinite hidden Markov model see Maheu and Yang (2016).

Combining the HDP, (27)-(28), with the state indicator s_t and the data density, forms the infinite hidden Markov model,

$$s_t \big| s_{t-1}, P \sim P_{s_{t-1}}, \tag{29}$$

$$\theta_j \overset{iid}{\sim} H,\ j = 1, 2, ..., \tag{30}$$

$$y_t \big| s_t, \theta \sim F(y_t \big| \theta_{s_t}), \tag{31}$$

where $\theta = \{\theta_1, \theta_2, ...\}$ and $F(\cdot|\cdot)$ is the data distribution. The two concentration parameters η and α control the number of active states in the model. Larger values favour more states while small values promote a parsimonious state space. Rather than set these hyperparameters they can be treated as parameters and estimated from the data. In this case, the hierarchical

[8] To make this concrete, consider the following simple example. Abstracting from an infinite chain to a three dimensional one, suppose $\Gamma = (0.51, 0.32, 0.17)$. Each row of the transition matrix is obtained as $P_j \sim \text{Dir}(\Gamma)$ where $\text{Dir}(\Gamma)$ is the Dirichlet distribution with parameter Γ and $E[P_j] = \Gamma$. Then, for instance, the rows sampled could be $P_1 = (0.55, 0.2, 0.25)$, $P_2 = (0.42, 0.41, 0.17)$ and $P_3 = (0.49, 0.30, 0.21)$.

prior for η and α are

$$\eta \sim \mathrm{G}(a_\eta, b_\eta), \tag{32}$$
$$\alpha \sim \mathrm{G}(a_\alpha, b_\alpha), \tag{33}$$

where $\mathrm{G}(a, b)$ stands for the gamma distribution with shape parameter a and rate parameter b.

The models can be estimated with MCMC methods. We discuss the specific details below for each model.

3.2 IHMM with RV and RAV

RV or RAV can be jointly modelled in the IHMM model as we did in the finite Markov switching models. The joint IHMM is constructed by replacing the Dirichlet distributed prior of the MS model by a hierarchical Dirichlet process. Hierarchical priors are used for concentration parameter α and η and allow the data to influence the state dimension. For example, the IHMM-RV model is given as (27)-(28), (32)-(33) with

$$s_t \big| s_{t-1}, P \sim P_{s_{t-1}}, \tag{34}$$
$$\theta_j = \{\mu_j, \sigma_j^2\} \overset{iid}{\sim} H,\ j = 1, 2, ..., \tag{35}$$
$$r_t \big| s_t, \theta \sim \mathrm{N}(\mu_{s_t}, \sigma_{s_t}^2), \tag{36}$$
$$RV_t \big| s_t \sim \mathrm{IG}(\nu + 1, \nu\sigma_{s_t}^2). \tag{37}$$

The base distribution is $H(\mu) \equiv \mathrm{N}(m, v^2), H(\sigma^2) \equiv \mathrm{IG}(v_0, s_0)$. The parameter $\sigma_{s_t}^2$ is common to the distribution of r_t and RV_t.

The IHMM-logRV and IHMM-logRAV models are formed similarly by replacing the fixed dimension transition matrix with infinite dimensional versions with a HDP prior. For instance, the IHMM-logRV specification replaces (35)-(37) with

$$\theta_j = \{\mu_j, \zeta_j, \delta_j^2\} \overset{iid}{\sim} H,\ j = 1, 2, ..., \tag{38}$$
$$r_t \big| s_t, \theta \sim \mathrm{N}\left(\mu_{s_t}, \exp(\zeta_{s_t})\right), \tag{39}$$
$$\log(RV_t) \big| s_t \sim \mathrm{N}\left(\zeta_{s_t} - \frac{1}{2}\delta_{s_t}^2, \delta_{s_t}^2\right). \tag{40}$$

The base distribution is $H(\mu) \equiv \mathrm{N}(m_\mu, v_\mu^2), H(\zeta) \equiv \mathrm{N}(m_\zeta, v_\zeta^2)$ and $H(\delta) \sim \mathrm{IG}(v_0, s_0)$. The parameter ζ_{s_t} is common to the distribution of r_t and $\log(RV_t)$. Similarly, the IHMM-logRAV model, replaces (35)-(37) with

$$\theta_j = \{\mu_j, \zeta_j, \delta_j^2\} \overset{iid}{\sim} H, j = 1, 2, ..., \quad (41)$$

$$r_t \big| s_t, \theta \sim \mathrm{N}\left(\mu_{s_t}, \exp(2\zeta_{s_t})\right), \quad (42)$$

$$\log(RAV_t)\big| s_t \sim \mathrm{N}\left(\zeta_{s_t} - \frac{1}{2}\delta_{s_t}^2, \delta_{s_t}^2\right). \quad (43)$$

The base distribution is $H(\mu) \equiv \mathrm{N}(m_\mu, v_\mu^2), H(\zeta) \equiv \mathrm{N}(m_\zeta, v_\zeta^2)$ and $H(\delta) \sim \mathrm{IG}(v_0, s_0)$. Now, ζ_{s_t} affects both r_t and $\log(RAV_t)$.

3.3 Multivariate IHMM with RCOV

The multivariate MS-RCOV model can be extended to its nonparametric version, labelled IHMM-RCOV as follows. In this model (27)-(28), (32)-(33) is combined with

$$s_t \big| s_{t-1}, P \sim P_{s_t-1}, \quad (44)$$

$$\theta_j = \{M_j, \Sigma_j\} \overset{iid}{\sim} H,\ j = 1, 2, ..., \quad (45)$$

$$R_t \big| s_t, \theta \sim \mathrm{N}(M_{s_t}, \Sigma_{s_t}), \quad (46)$$

$$RCOV_t \big| s_t \sim \mathrm{IW}(\Sigma_{s_t}(\kappa - d - 1), \kappa). \quad (47)$$

The base distribution is $H(M) \equiv \mathrm{N}(m, V), H(\Sigma) \equiv \mathrm{W}(\Psi, \tau)$, where $\mathrm{W}(\Psi, \tau)$ denotes Wishart distribution, Ψ are $d \times d$ positive definite matrices and $\kappa > d+1$ being the degree of freedom. Σ_{s_t} is a common parameter affecting the distributions of R_t and $RCOV_t$.

4 Benchmark Models

Each of the new models are compared to benchmark models that do not use ex-post variance measures. The benchmark specifications are essentially the same model with RV_t or RAV_t omitted. For example, in the univariate application we compare to the following MS specification.

$$r_t \big| s_t \sim \mathrm{N}(\mu_{s_t}, \sigma_{s_t}^2), \quad (48)$$

$$P_{i,j} = p(s_{t+1} = j | s_t = i), \quad (49)$$

where $s_t \in \{1, \ldots, K\}$. The IHMM comparison model is given with (27)-(33) with (30) and (31) specified as

$$\theta_j = \{\mu_j, \sigma_j^2\} \overset{iid}{\sim} H, \; j = 1, 2, ..., \tag{50}$$

$$r_t \big| s_t, \theta \sim \mathrm{N}(\mu_{s_t}, \sigma_{s_t}^2). \tag{51}$$

The benchmark models for the multivariate application are similarly derived by omitting $RCOV_t$.

5 Estimation and Model Comparison

For notation, let $r_{1:t} = \{r_1, \ldots, r_t\}$, $RV_{1:t} = \{RV_1, \ldots, RV_t\}$, $y_{1:t} = \{y_1, \ldots, y_t\}$ where $y_t = \{r_t, RV_t\}$. We further define $R_{1:T} = \{R_1, \ldots, R_T\}$, $RCOV_{1:T} = \{RCOV_1, \ldots, RCOV_T\}$ and $Y_{1:t} = \{Y_1, \ldots, Y_t\}$ where $Y_t = \{R_t, RCOV_t\}$.

5.1 Estimation of Joint Finite MS Models

The joint finite MS models are estimated using Bayesian inference. Taking the MS-RV model as an example, model parameters include $\theta = \{\mu_j, \sigma_j^2\}_{j=1}^K$, $\phi = \{\nu\}$ and transition matrix P. By augmenting the latent state variable $s_{1:T} = \{s_1, s_2, \cdots, s_T\}$, MCMC methods are used to simulate from the conditional posterior distributions. The prior distributions are listed in Table 1. One MCMC iteration contains the following steps.

1. $s_{1:T} \big| y_{1:T}, \theta$
2. $\theta_j \big| y_{1:T}, s_{1:T}, \phi$, for $j = 1, 2, \ldots, K$
3. $\phi \big| y_{1:T}, s_{1:T}, \theta$
4. $P \big| s_{1:T}$

The first MCMC step is to sample the latent state variable $s_{1:T}$ from the conditional posterior distribution $s_{1:T} \big| y_{1:T}, \theta, P$. We follow Chib (1996) and use the forward filter backward sampler. In the second step, μ_j is sampled using the Gibbs sampler for the linear regression model. The conditional posterior of σ_j^2 is of unknown form and a Metropolis-Hasting step is used. The proposal density follows a gamma distribution formed by combining the likelihood for $RV_{1:T}$ and the prior. $\nu \big| y_{1:T}, \{\sigma_j^2\}_{j=1}^K$ is sampled using the Metropolis-Hasting algorithm with a random walk proposal. Finally, each row of P is drawn from its conditional posterior, which is a Dirichlet distribution. Additional details of posterior sampling are collected in the appendix.

After an initial burn-in of iterations are discarded we collect N additional MCMC iterations for posterior inference. Simulation consistent estimates of posterior quantities can be formed.[9] For example, the posterior mean of θ_j is estimated as,

$$E[\theta_j|y_{1:T}] \approx \frac{1}{N}\sum_{i=1}^{N}\theta_j^{(i)}, \tag{52}$$

where $\theta_j^{(i)}$ is the i^{th} iteration from posterior sampling of parameter θ_j. The smoothed probability of s_t can be estimated as follows.

$$p(s_t = k|y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}(s_t^{(i)} = k), \tag{53}$$

where $\mathbb{1}(A) = 1$ if A is true and otherwise 0.

The estimation of MS-RAV, MS-logRV, MS-logRAV and MS-RCOV models are done in a similar fashion. Detailed estimation steps of the models are in the appendix.

5.2 Estimation of Joint IHMM Models

In the IHMM-RV model, estimation of the unknown parameters $\theta = \{\mu_j, \sigma_j^2\}_{j=1}^{\infty}$, $\phi = \{\nu, \alpha, \eta\}$, P, Γ and $s_{1:T}$ are different given the unbounded nature of the state space. The beam sampler, introduced by Gael et al. (2008), is an extension of the slice sampler by Walker (2007), and is an elegant solution to estimation challenges that an infinite parameter model presents. An auxiliary variable $u_{1:T} = \{u_1, u_2, \cdots u_T\}$ is introduced that randomly truncates the state space to a finite one at each MCMC iteration. Conditional on $u_{1:T}$ the number of states is finite and the forward filter backward sampler previously discussed can be used to sample $s_{1:T}$.

The key idea behind the beam sampling is to introduce the auxiliary variable u_t that preserves the target distributions, and has the following conditional density

$$p(u_t|s_{t-1}, s_t, P) = \frac{\mathbb{1}(0 < u_t < P_{s_{t-1},s_t})}{P_{s_{t-1},s_t}}, \tag{54}$$

[9] For a good textbook treatment of MCMC estimation see Greenberg (2014).

where $P_{i,j}$ denotes element (i, j) of P. The forward filtering step becomes

$$p(s_t|y_{1:t}, u_{1:t}, P) \propto p(y_t|y_{1:t-1}, s_t) \sum_{s_{t-1}=1}^{\infty} \mathbb{1}(0 < u_t < P_{s_{t-1},s_t}) p(s_{t-1}|y_{1:t-1}, u_{1:t-1}, P) \quad (55)$$

$$\propto p(y_t|y_{1:t-1}, s_t) \sum_{s_{t-1}:u_t<P_{s_{t-1},s_t}} p(s_{t-1}|y_{1:t-1}, u_{1:t-1}, P), \quad (56)$$

which renders an infinite summation into a finite one. Conditional on u_t, only states satisfying $u_t < P_{s_{t-1},s_t}$ are considered and the number of states become a finite number, say K. The same considerations hold for the backward sampling step.

Each MCMC iteration loop contains the following steps.

1. $u_{1:T}\big|s_{1:T}, P, \Gamma$
2. $s_{1:T}\big|y_{1:T}, u_{1:T}, \theta, \phi, P, \Gamma$
3. $\Gamma\big|s_{1:T}, \eta, \alpha$
4. $P\big|s_{1:T}, \Gamma, \alpha$
5. $\theta_j\big|y_{1:T}, s_{1:T}, \phi$ for $j = 1, 2, \ldots,$
6. $\phi\big|y_{1:T}, s_{1:T}, \theta$

$u_{1:T}$ is sampled from its conditional densities $u_t\big|s_{1:T}, P \sim \text{Uniform}(0, P_{s_{t-1},s_t})$ for $t = 1, \cdots, T$. Following the discussion above, conditional on $u_{1:T}$ the effective state space is finite of dimension K and $s_{1:T}$ is sampled using the forward filter backward sampler. Γ and each row of transition matrix follow a Dirichlet distribution after additional latent variables are introduced. The sampling of μ_j, σ_j^2 and ν are the same as in the joint finite MS models. Posterior sampling of the IHMM-logRV, IHMM-logRAV and IHMM-RCOV models can be done following similar steps. The appendix provides the detailed steps.

Given N MCMC iterations collected after a burn-in period are discarded, posterior statistics can be estimated as usual. The estimation of state-dependent parameters suffer from a label-switching problem. This means that the states and the associated parameters can be permuted over different MCMC iterations while maintaining the same likelihood value. Thus, it is not possible to track state j over the MCMC iterations as the definition of state j can change. Therefore we focus on label invariant quantities. For example, the posterior mean of θ_{s_t} is computed as

$$E[\theta_{s_t}|y_{1:T}] \approx \frac{1}{N} \sum_{i=1}^{N} \theta_{s_t^{(i)}}^{(i)}. \quad (57)$$

5.3 Density Forecasts

The predictive density is the distribution governing a future observation given a model $\mathcal{M}$, prior and data. It is computed by integrating out parameter uncertainty. The predictive likelihood is the key quantity used in model comparison and is the predictive density evaluated at next period's return

$$p(r_{t+1}|y_{1:t}, \mathcal{M}) = \int p(r_{t+1}|y_{1:t}, \Lambda, \mathcal{M})p(\Lambda|y_{1:t}, \mathcal{M}) \, \mathrm{d}\Lambda, \tag{58}$$

where $p(r_{t+1}|y_{1:t}, \Lambda, \mathcal{M})$ is the data density given $y_{1:t}$ and parameter Λ and $p(\Lambda|y_{1:t}, \mathcal{M})$ is the posterior distribution of Λ.

To focus on model performance and comparison it is convenient to consider the log-predictive likelihood and use the sum of log-predictive likelihoods from time $t+1$ to $t+s$ given as

$$\sum_{l=t+1}^{t+s} \log p(r_l|y_{1:l-1}, \mathcal{M}). \tag{59}$$

The log-predictive Bayes factor between $\mathcal{M}_1$ and model $\mathcal{M}_2$ is defined as

$$\sum_{l=t+1}^{t+s} \log p(r_l|y_{1:l-1}, \mathcal{M}_1) - \sum_{l=t+1}^{t+s} \log p(r_l|y_{1:l-1}, \mathcal{M}_2). \tag{60}$$

A log-predictive Bayes factor greater than 5 provides strong support for $\mathcal{M}_1$.

5.3.1 Predictive Likelihood of MS Models

Both parameter uncertainty and state uncertainty need to be integrated out in order to calculate the predictive likelihood. The predictive likelihood of a K-state joint MS model can be estimated as follows

$$p(r_{t+1}|y_{1:t}) \approx \frac{1}{N} \sum_{i=1}^{N} \sum_{s_{t+1}=1}^{K} \mathrm{N}(r_{t+1}|\mu_{s_{t+1}}^{(i)}, \sigma_{s_{t+1}}^{2(i)}) P^{(i)}_{s_{t+1}, s_t^{(i)}}, \tag{61}$$

where $\mu_{s_{t+1}}^{(i)}$ and $\sigma_{s_{t+1}}^{2(i)}$ are the i^{th} draw of $\mu_{s_{t+1}}$ and $\sigma_{s_{t+1}}^2$ respectively, and $P_{j1,j2}^{(i)}$ denotes element $(j1, j2)$ of $P^{(i)}$ all based on the posterior distribution given data $y_{1:t}$.

The calculation of the predictive likelihood for the multivariate MS models follows the same method,

$$p(R_{t+1}|Y_{1:t}) \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{s_{t+1}=1}^{K} \mathrm{N}(R_{t+1}|M_{s_{t+1}}^{(i)}, \Sigma_{s_{t+1}}^{(i)}) P^{(i)}_{s_{t+1}, s_t^{(i)}}. \tag{62}$$

5.3.2 Predictive Likelihood of IHMM

For the IHMM models the state next period may be a recurring one or it may be new, the calculation of predictive likelihood is sightly different and is estimated as follows,

$$p(r_{t+1}|y_{1:t}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathrm{N}(r_{t+1}|\mu_i, \sigma_i^2) \tag{63}$$

where the parameter values μ_i and σ_i^2 are determined using the following steps. Given $s_t^{(i)}$, draw $s_{t+1} \sim \text{Multinomial}(P^{(i)}_{s_t^{(i)}}, K^{(i)}+1)$.

1. If $s_{t+1} <= K^{(i)}$, set $\mu_i = \mu^{(i)}_{s_{t+1}}$, $\sigma_i^2 = \sigma^{2(i)}_{s_{t+1}}$.
2. If $s_{t+1} = K^{(i)} + 1$, draw a new set of parameter values from the prior: $\mu_i \sim \mathrm{N}(m, v^2)$ and $\sigma_i^2 \sim \mathrm{IG}(v_0, s_0)$.

In multivariate IHMM models, the predictive likelihood is calculated exactly the same way except that the base measure draw is from a multivariate normal and an inverse-Wishart distribution.

5.4 Point Predictions for Returns and Volatility

In addition to density forecasts we evaluate the predictive mean of returns and the predictive variance (covariance) of returns. For the finite state MS models, conditional on the MCMC output, the predictive mean for r_{t+1} is estimated as

$$E[r_{t+1}|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} \mu_j^{(i)} P^{(i)}_{s_t^{(i)},j}. \tag{64}$$

The second moment of the predictive distribution can be estimated as follows.

$$E[r_{t+1}^2|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} (\mu_j^{(i)2} + \sigma_j^{2(i)}) P^{(i)}_{s_t^{(i)},j}, \tag{65}$$

so that the variance can be estimated from

$$\text{Var}(r_{t+1}|y_{1:t}) = E[r_{t+1}^2|y_{1:t}] - (E[r_{t+1}|y_{1:t}])^2. \tag{66}$$

For the multivariate model we have

$$E[R_{t+1}|Y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} M_j^{(i)} P_{s_t^{(i)},j}^{(i)}, \tag{67}$$

$$E[R_{t+1}R_{t+1}'|Y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}(\Sigma_j^{(i)} + M_j^{(i)} M_j^{(i)\prime}) P_{s_t^{(i)},j}^{(i)} \tag{68}$$

which can be used to estimate

$$\text{Cov}(R_{t+1}|Y_{1:t}) = E[R_{t+1}R_{t+1}'|Y_{1:t}] - E[R_{t+1}|Y_{1:t}]E[R_{t+1}|Y_{1:t}]'. \tag{69}$$

For the IHMM models, the predictive mean and variance are derived from the following.

$$E[r_{t+1}|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\mu_i, \tag{70}$$

$$E[r_{t+1}^2|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}(\sigma_i^2 + \mu_i^2). \tag{71}$$

The parameters μ_i, σ_i^2 are selected following the steps in Section 5.3.2. Similar results hold for the multivariate versions.

Finally, estimation and forecasting for the benchmark model follow along the same lines as discussed for the joint models with minor simplifications.

6 Univariate Return Application

Four versions of joint MS models (MS-RV, MS-logRV, MS-RAV and MS-logRAV), joint IHMM-RV and the benchmark alternatives are applied to model monthly U.S. stock market returns from March 1885 to December 2013. The data from March 1885 to December 1925 are the daily capital gain returns provided by Bill Schwert, see Schwert (1990). The rest of returns are from the value-weighted S&P 500 index excluding dividends from CRSP, for a total of 1542 observations. The daily simple returns are converted to continuous compounded returns and are scaled by 12. The monthly return r_t is the sum of the daily returns. RV_t and other ex-post volatility measures are computed according to the definitions previously stated. Table 2 reports the summary statistics of monthly returns along with the summary statistics of monthly RV_t, $\log(RV_t)$, RAV_t and $\log(RAV_t)$.

Table 1 lists the priors for the various models. The priors provide a wide range of empirically realistic parameter values. The benchmark models have the same prior. Results are based on 5000 MCMC iterations after dropping the first 5000 draws.

6.1 Out-of-Sample Forecasts

Table 3 reports the sum of log-predictive likelihoods of 1 month ahead returns for the out-of-sample period from January 1951 to December 2013 (756 observations). Recursive out-of-sample forecasts are computed beginning at the start of the out-of-sample period. As new data arrives the models are re-estimated and forecasts computed one period ahead. This is repeated until the end of the sample is reached.

In each case of a finite state assumption, the joint MS specifications outperform the benchmark model that do not use ex-post volatility measures. The log-predictive Bayes factors between the best finite joint MS model and the benchmark model are greater than 15, which provide strong evidence that exploiting higher frequency data leads to more accurate density forecasts. Using ex-post volatility data offer little to no gains in forecasting the mean of returns, however, it does lead to better variance forecasts as measured against realized variance.

The lower panel of Table 3 report the same results for the infinite hidden Markov models with and without higher frequency data. The overall best model according to the predictive likelihood is the IHMM-RV specification. This model has a log-predictive Bayes factor of 20.8 against the best finite state model that does not use high-frequency data. It has about the same log-predictive Bayes factor against the IHHM. The IHMM-RV has the lowest RMSE for RV_t forecasts. It is 10.6% lower than the best MS model that uses returns only.

All of the joint models that include some form of ex-post volatility lead to improved forecasts but generally the best performance comes from using RV_t.

Table 4 provides a check on these results over the shorter sample period from January 1984 to December 2013 (360 observations). The general results are the same as the longer sample, high-frequency data offer substantial improvements to density forecasts of returns and gains on forecasts of RV_t.

Figures 1 gives a breakdown of the period-by-period difference in the predictive likelihood values. Positive values are in favour of the model with RV_t. The overall sum of the log-predictive likelihoods is not due to a few outliers or any one period but represent ongoing improvements in accuracy. The joint models do a better job in forecasting return densities when the market is in a high volatility period, such as the period of 1973-1974 crash, the period before and after the internet bubble and the 2008 financial crisis. Table 5 reports the log-predictive Bayes factors using data from four major bear regimes. The data windows are small, nevertheless, the log-predictive Bayes factors are all positive and support the joint models which exploit RV.

6.2 Parameter Estimates and State Inference

Table 6 reports the posterior summary of parameters of the 2 state MS, MS-RV and MS-RAV models based on the full sample. To avoid label switching issues, we use informative priors $\mu_1 \sim$ N(-1, 1), $\mu_2 \sim$ N(1, 1), $P_1 \sim$ Dir(4, 1), $P_2 \sim$ Dir(1, 4) and $\sigma_j^2 \sim$ IG(5, 5) for $j = 1, 2$ and restrict $\mu_1 < 0$ and $\mu_2 > 0$. The results show that all three models are able to sort stock returns into two regimes. One regime has a negative mean and high volatility, the other regime has positive mean return mean along with lower variance. This is consistent with the results of Maheu et al. (2012) and several other studies.

Compared with the benchmark MS model, the joint models specify the return distribution more precisely in each state, as can be seen the smaller estimated values $\sigma_{s_t}^2$. For instance, in the first state the innovation variance is 1.6127 for the MS model, while the estimates of variance are 0.5633 and 0.6581 in the MS-RV and MS-RAV models, respectively. The variance estimates in the positive mean regimes drop from 0.2187 to 0.1556 and 0.1460 after joint modelling RV and RAV. We would expect this reduction in the innovation variance to result in better forecasts which is what we found in the previous section.

Another interesting result is that MS-RV and MS-RAV models provide more precise estimates of all the model parameters. As shown in Table 6, most parameters have smaller posterior standard deviations and shorter 0.95 density intervals. For example, the length of the density interval of μ_1 from the benchmark model is 0.437, while the values are 0.187 and 0.167 from the MS-RV and MS-RAV models, respectively.

Figure 2 plots $E\left[\sigma_{s_t}^2 \middle| y_{1:T}\right]$ for the IHMM-RV and IHMM models. Volatility estimates vary over a larger range from the IHMM-RV model. For example, it appears that the IHMM overestimates the return variance during calm market periods and underestimates the return variance in several high volatile periods, such as the October 1987 crash and the financial crisis in 2008. In contrast, the return variance from the IHMM-RV model is closer to RV during these times. The differences between the models is due to the additional information from ex-post volatility.

Figure 3 plots the smoothed probability of the high return state from the 2 state MS, MS-RV and MS-RAV models. The benchmark MS model does a fairly good job in identifying the primary downward market trends, such as the big crash of 1929, 1973-1974 bear market and the 2008 market crash, but it ignores a series of panic periods before and after 1900, the internet bubble crash and several other relatively smaller downward periods. The joint MS-RV and MS-RAV models not only identify the primary market trends but also are able to capture a number of short lived market drops. The main difference is that the joint model appears to have more frequent state switches and state identification is more precise. One obvious example is the joint models identify the dot-com collapse from 2000 to 2002 and the market crash of 2008-2009.

In summary, the joint models lead to better density forecasts, better forecasts of realized variance, improved parameter precision and minor differences in latent state estimates.

6.3 Market Timing Portfolio

As shown in Section 6.1, the joint MS model improves density forecast of returns. We further investigate if the better forecasts lead to actual economic gain. The proposed joint models (MS-RV and IHMM-RV) are compared with benchmark models from a portfolio allocation perspective.

Suppose an investor uses a market timing strategy to manage her portfolio. Let r_t^m, r_t^p and r_t^f denotes the monthly simple return of the market (index), a market-timing portfolio and the risk-free asset, respectively. The trading strategy is designed as follows. Each model $\mathcal{M}$ is used to forecast the direction of the market next month. That is, if $P(r_{t+1}^m > r_t^f | \mathcal{M}, r_{1:t}, RV_{1:t}) > 0.5$, one dollar is invested in the market and held for one month to receive that month's return. In this case, the portfolio return $r_{t+1}^{\mathcal{M}} = r_{t+1}^m$. Otherwise, the risk-free asset is held and the monthly portfolio return is $r_{t+1}^{\mathcal{M}} = r_{t+1}^f$. This is repeated for several different models $\mathcal{M}_1, \dots, \mathcal{M}_p$ and each produces its own portfolio of returns, $r_t^{\mathcal{M}_1}, \dots, r_t^{\mathcal{M}_p}$, over time.

The evaluation of portfolios (models) is based on risk-return tradeoff and utility-based approach. Two types of utility functions are used. The quadratic utility function used in Fleming et al. (2001) has the following form

$$U_q(r_t^{\mathcal{M}}) = (1 + r_t^{\mathcal{M}}) - \frac{\gamma}{2(1+\gamma)}(1 + r_t^{\mathcal{M}})^2, \tag{72}$$

where γ denotes the risk aversion coefficient and is set to be $\gamma = 5$.

Following Skouras (2007) and Clements and Silvennoinen (2013), the second utility function is exponential, given as follows

$$U_e(r_t^{\mathcal{M}}) = -\exp(-\gamma(1 + r_t^{\mathcal{M}})). \tag{73}$$

The performance fee Δ that an investor would pay to switch from one portfolio to another is used to evaluate the competing portfolios. Δ is a constant equalizing the ex-post utility from models $\mathcal{M}_1$ and $\mathcal{M}_2$ in

$$\sum_{t=1}^{T} U(r_t^{\mathcal{M}_1}) = \sum_{t=1}^{T} U(r_t^{\mathcal{M}_2} - \Delta). \tag{74}$$

Table 7 reports the summary statistics of the portfolio returns. The out-of-sample market timing is conducted from 1951-2013 and for a shorter sample 1984-2013. The portfolios based on the joint MS or IHMM models yield higher average returns, lower risks and result in higher Sharpe ratios in all the cases, compared with the ones based on models without RV. It is only possible to beat the Sharpe ratio from the buy and hold strategy if a model uses RV information. For instance, in all cases except one, the Sharpe ratio from the joint models is higher than from a buy and hold strategy. The last two columns of Table 7 record the annualized basis point performance fees that an investor would be willing to pay to switch from the portfolio based on the 2-state MS benchmark model to the one based on another model. It shows incorporating RV in MS models always improves the utility level of an investor with either quadratic or exponential utility. Independent of the number of states or the sample period an investor is always willing to move from the simple MS two state model to a specification that exploits RV.

7 Multivariate Return Application

We also evaluate the multivariate joint models through multivariate applications to a vector of equity returns. The daily prices of three equities (stock symbol: IBM, XOM and GE) listed in NYSE are obtained from CRSP. These firms were chosen since they have been actively traded over the full sample period, January 1926 to December 2013 (1056 observations). The continuous compounded returns are constructed and the monthly RCOV is computed using daily values following equation (3). The summary statistics of monthly returns R_t and $RCOV_t$ are found in Table 9. The prior specification is found in Table 8.

7.1 Out-of-Sample Forecasts

Table 10 reports the results of density forecasts and the root mean squared error of predictions based on 756 out-of-sample observations. We found the larger finite state models are the most competitive and therefore do not include results for small dimension models. The 8 and 12 state models that exploit RCOV are all superior to the models that do not according to log-predictive values. The improvement in the log-predictive likelihood is 30 or more. Further improvements are found on moving to the Bayesian nonparametric models. The IHMM-RCOV model is the best over the alternative models.

As for point predictions of return and realized covariance, the results is similar to the univariate return applications. The proposed joint models improve predictions of $RCOV_t$ but offer no gains for return predictions.

Figure 4 displays the posterior average of active states in both IHMM and IHMM-RCOV models at each point in the out-of-sample period. It shows that more states are used in the joint return-RCOV model in order to better capture the dynamics of returns and volatility.

7.2 Portfolio Performance

Beyond the forecasts of return density and covariance, we also evaluate the out-of-sample performance of models in portfolio allocation. Suppose an investor forms her portfolio using three equities (IBM, XOM and GE) and applies the modern portfolio theory by Markowitz (1952) to select a portfolio on the efficient frontier. The weight of a minimum variance portfolio given required portfolio return μ_p can be calculated by solving the problem below.

$$\min_{w_{t+1}} \quad w'_{t+1}\mathrm{Cov}_{t+1|t}w_{t+1} \quad \text{s.t.} \quad w'_{t+1}\boldsymbol{\mu} + (1 - w'_{t+1}\mathbf{1})r_t^f = \mu_p, \tag{75}$$

where w_{t+1} is the portfolio weight, $\mathrm{Cov}_{t+1|t}$ denotes the predictive covariance from a model using data up to time t, $\mathbf{1}$ is a vector of ones and $\boldsymbol{\mu}$ is the return mean vector set to the sample average and is the same across models. The solution is

$$w_{t+1} = \frac{(\mu_p - r_t^f)\mathrm{Cov}_{t+1|t}^{-1}(\boldsymbol{\mu} - r_t^f\mathbf{1})}{(\boldsymbol{\mu} - r_t^f\mathbf{1})'\mathrm{Cov}_{t+1|t}^{-1}(\boldsymbol{\mu} - r_t^f\mathbf{1})}. \tag{76}$$

The portfolio return for model $\mathcal{M}$ is $r_{t+1}^{\mathcal{M}} = w'_{t+1}R_{t+1} + (1 - w'_{t+1}\mathbf{1})r_{t+1}^f$. R_{t+1} and r_{t+1}^f are all expressed as simple returns. $\boldsymbol{\mu}$ is set to be the sample average of returns and is common to all models.[10] $\mathrm{Cov}_{t+1|t}$ is computed following equation (69). The utility functions and performance fee calculations are the same as in the univariate portfolio application.

A second portfolio application compares the models on their ability to produce a global minimum variance portfolio Engle and Colacito (2006). The global minimum variance port-

[10]This is done to focus on the differences in the the predictive covariance. When both the predictive mean and predictive covariance are used from each model the IHMM-RCOV specification is strongly favored compared to other models.

folio can be determined by solving the following minimization problem.

$$\min_{w_{t+1}} \quad w'_{t+1}\text{Cov}_{t+1|t}w_{t+1} \quad \text{s. t.} \quad \text{and} \quad w'_{t+1}\mathbf{1} = 1. \tag{77}$$

The weight of global minimum-variance portfolio is given by

$$w_{t+1} = \frac{\text{Cov}_{t+1|t}^{-1}\mathbf{1}}{\mathbf{1}'\text{Cov}_{t+1|t}^{-1}\mathbf{1}}. \tag{78}$$

The model parameters are re-estimated each month after new data arrives and the out-of-sample period is from 1951–2013 and 1984–2013. Tables 11 and 12 summarizes the performance of minimum-variance portfolios for two sample periods based on the benchmark model and proposed joint models. Panel A and B show the results for required annual portfolio returns of $\mu_p = 10\%$ and $\mu_p = 20\%$, respectively.[11]

The benefit of jointly modelling both return and RCOV is clear. In all the cases but one using the predictive covariance from a model incorporating RCOV increases the Sharpe ratio relative to the model without RCOV. The investor is always willing to pay to move to a model that exploits RCOV information. For instance the investor using an 8-state MS model would pay 7.08 basis points to use the forecasts from an 8-state MS-RCOV model. In Table 12 performance fees are considerably larger for the more recent sample period.

Finally, Table 13 reports the variances of global minimum-variance portfolios based on competing models over the two sample periods. The specifications that jointly model returns and RCOV always lead to portfolios with smaller variances, compared to the model without RCOV, no matter the number of states.

In summary, the joint modelling of returns and RCOV leads to better out-of-sample forecasts and improved portfolio decisions.

8 Conclusion

This paper shows how to incorporate ex-post measures of volatility with returns to improve forecasts, parameter and state estimation under a Markov switching assumption. We show how to build and estimate joint nonlinear factor models. Markov switching can be specified as fixed and finite or countably infinite. In empirical applications the new models give dramatic improvements in density forecasts for returns, forecasts of realized variance and lead to improved portfolio decisions.

[11] Note that the annualized mean return for IBM, XOM and GE are 13.4%, 11.6% and 10.3%, respectively.

9 Appendix

9.1 MS-RV Model

(1) $s_{1:T} \big| y_{1:T}, \theta, \phi, P$

The latent state variable $s_{1:T}$ is sampled using the forward filter backward sampler (FFBS) in Chib (1996). The forward filter part contains the following steps.

i. Set the initial value of filter $p(s_1 = j \big| y_1, \theta, \phi, P) = \pi_j$, for $j = 1, \ldots, K$, where π is the stationary distribution, which can be computed by solving $\pi = P^{\intercal}\pi$.

ii. Prediction step: $p(s_t \big| y_{1:t-1}, \theta, \phi, P) \propto \sum_{j=1}^{K} P_{j,s_t} \cdot p(s_{t-1} = j \big| y_{1:t-1}, \theta, \phi, P)$.

iii. Update step:

$$p(s_t \big| y_{1:t}, \theta, \phi, P) \propto f(r_t \big| \mu_{s_t}, \sigma^2_{s_t}) \cdot g(RV_t \big| \nu + 1, \nu\sigma^2_{s_t}) \cdot p(s_t \big| y_{1:t-1}, \theta, \phi, P),$$

where function $f(\cdot)$ and $g(\cdot)$ denote normal density and inverse-gamma density, respectively.

The underlying states are drawn using backward sampler as follows.

i. For $t = T$, draw s_T from $p(s_T \big| y_{1:T}, \theta, \phi, P)$.

ii. For $t = T-1, \ldots, 1$, draw s_t from $P_{s_t, s_{t+1}} \cdot p(s_t \big| y_{1:t}, \theta, \phi, P)$.

Let $n_j = \sum_{t=1}^{T} \mathbb{1}(s_t = j)$ denotes the number of observations belong to state j.

(2) $\mu_j \big| r_{1:T}, s_{1:T}, \sigma^2_j$ for $j = 1, \ldots, K$

μ_j is sampled using the Gibbs sampling for linear regression model. Given prior $\mu_j \sim \mathrm{N}(m_j, v_j^2)$, μ_j is sampled from conditional posterior $\mathrm{N}(\overline{m_j},\ \overline{v_j}^2)$, where

$$\overline{m_j} = \frac{v_j^2 \sum_{s_t=j} r_t + m_j\sigma_j^2}{\sigma_j^2 + n_j v_j^2}, \quad \text{and} \quad \overline{v_j}^2 = \frac{\sigma_j^2 v_j^2}{\sigma_j^2 + n_j v_j^2}.$$

(3) $\sigma^2_j \big| y_{1:T}, \mu_j, \nu, s_{1:T}$ for $j = 1, \ldots, K$

The prior of σ_j^2 is assumed to be $\sigma_j^2 \sim \mathrm{G}(v_0,\ s_0)$. The conditional posterior of σ_j^2 is given as follows,

$$p(\sigma_j^2 \big| y_{1:T}, \mu_j, \nu, s_{1:T}) \propto \prod_{s_t=j} \left\{ \frac{1}{\sigma_j} \exp\left[-\frac{(r_t - \mu_j)^2}{2\sigma_j^2} \right] \cdot (\nu\sigma_j^2)^{(\nu+1)} \exp\left(-\frac{\nu\sigma_j^2}{RV_t} \right) \right\} \cdot (\sigma_j^2)^{v_0 - 1} \exp\left(-s_0\sigma_j^2\right).$$

The conditional posterior of σ_j^2 is not of any known form, therefore Metropolis-Hasting algorithm is applied to sample σ_j^2. Combining RV likelihood function and prior provides a good proposal density $q(\cdot)$ for sampling σ_j^2, which follows a gamma distribution and is derived as follows,

$$\begin{aligned} p(\sigma_j^2 | y_{1:T}, \mu_j, \nu, s_{1:T}) &< \prod_{s_t=j} \left\{ (\nu\sigma_j^2)^{(\nu+1)} \exp\left(-\frac{\nu\sigma_j^2}{RV_t} \right) \right\} \cdot (\sigma_j^2)^{v_0-1} \exp\left(-s_0\sigma_j^2\right) \\ &\sim \mathrm{G}\left(n_j(\nu+1) + v_0, \nu \sum_{s_t=j} \frac{1}{RV_t} + s_0 \right) \equiv q(\sigma_j^2). \end{aligned}$$

(4) $\nu | y_{1:T}, \{\sigma_j^2\}_{j=1}^K, s_{1:T}$

The prior of ν is assumed to be $\nu \sim \mathrm{IG}(a, b)$. The posterior of ν is given as follows.

$$p(\nu | y_{1:T}, \sigma_{s_t}^2, s_{1:T}) \propto \prod_{t=1}^{T} \left\{ \frac{(\nu\sigma_{s_t}^2)^{(\nu+1)}}{\Gamma(\nu+1)} RV_t^{-\nu-2} \exp\left(-\frac{\nu\sigma_{s_t}^2}{RV_t} \right) \right\} \cdot (\nu)^{-a-1} \exp\left(-\frac{b}{\nu} \right)$$

ν is drawn from a random walk proposal and negative draws will be dropped.

(5) $P | s_{1:T}$

Using conjugate prior for rows of the transition matrix P: $P_j \sim \mathrm{Dir}(\alpha_{j1}, \cdots, \alpha_{jK})$, the posterior is given by $\mathrm{Dir}(\alpha_{j1} + n_{j1}, \cdots, \alpha_{jK} + n_{jK})$, where vector $(n_{j1}, n_{j2}, \cdots, n_{jK})$ records the numbers of switches from state j to the other states.

9.2 MS-logRV Model

Forward filter backward sampler is used to sample $s_{1:T}$. The sampling of P is same as step (5) in MS-RV model estimation. The sampling of μ_j is same as step (2) in MS-RV model except replacing σ_j^2 with $\exp(\zeta_j)$.

Let $n_j = \sum_{t=1}^{T} \mathbb{1}(s_t = j)$ denotes the number of observations belong to state j. Assuming the prior of ζ_j is $\zeta_j \sim N(m_{\zeta,j}, v_{\zeta,j}^2)$, the conditional posterior of ζ_j is given as follows,

$$\begin{aligned} p(\zeta_j | y_{1:T}, \mu_j, \delta_j^2, s_{1:T}) \quad \propto \quad & \prod_{s_t=j} \left\{ \exp\left[-\frac{\zeta_j}{2} - \frac{(r_t - \mu_j)^2}{2\exp(\zeta_j)} \right] \cdot \exp\left[-\frac{(\log RV_t - \zeta_j + \frac{1}{2}\delta_j^2)^2}{2\delta_j^2} \right] \right\} \\ & \cdot \exp\left[-\frac{(\zeta_j - m_{\zeta j})^2}{2v_{\zeta,j}^2} \right] \end{aligned}$$

Metropolis-Hasting algorithm is applied to sample ζ_j. The proposal density is formed as follows,

$$
\begin{aligned}
p(\zeta_j \big| y_{1:T}, \mu_j, \delta_j^2, s_{1:T}) &\propto \prod_{s_t=j} \left\{ \exp\left[-\frac{\zeta_j}{2} - \frac{(r_t-\mu_j)^2}{2\exp(\zeta_j)} \right] \right\} \cdot \exp\left[-\frac{(\zeta_j-\mu^*)^2}{2\sigma^{2*}} \right] \\
&< \exp\left[-\frac{n_j\zeta_j}{2} - \frac{\sum_{s_t=j}(r_t-\mu_j)^2 \exp(-\mu^*)(-\zeta_j)}{2} - \frac{(\zeta_j-\mu^*)^2}{2\sigma^{2*}} \right] \\
&\sim \mathrm{N}(\mu_j^{**}, \sigma_j^{*2}) \equiv q(\zeta_j), \quad \text{where}
\end{aligned}
$$

$$
\begin{aligned}
\mu_j^* &= \frac{v_{\zeta,j}^2 \sum_{s_t=j} \log(RV_t) + \frac{1}{2} n_j v_\zeta^2 \delta_j^2 + \delta_j^2 m_{\zeta,j}}{n_j v_{\zeta,j}^2 + \delta_j^2}, \qquad \sigma_j^{*2} = \frac{\delta_j^2 v_{\zeta,j}^2}{n_j v_{\zeta,j}^2 + \delta_j^2}, \\
\mu_j^{**} &= \mu^* + \frac{1}{2}\sigma^{*2} \left[\sum_{s_t=j} (r_t-\mu_j)^2 \exp(-\mu^*) - n_j \right].
\end{aligned}
$$

Using conjugate prior $\delta_j^2 \sim \mathrm{IG}(v_0, s_0)$, the posterior density of δ_j^2 is given by

$$
p(\delta_j^2 \big| y_{1:T}, \mu_j, \sigma_j^2) \propto \prod_{s_t=j} \left\{ \frac{1}{\delta_j} \exp\left[-\frac{(\log RV_t - \zeta_j + \frac{1}{2}\delta_j^2)^2}{2\delta_j^2} \right] \right\} \cdot \ \delta_j^{-v_0-1} \exp\left(-\frac{s_0}{\delta_j^2} \right)
$$

Metropolis-Hasting is used to sample δ_j with the following proposal

$$
q(\delta_j^2) \equiv \mathrm{IG}\left(\frac{n_j}{2} + v_0, \frac{\sum_{s_t=j} (\log RV_t - \zeta_j)^2}{2} + s_0 \right).
$$

9.3 MS-RAV Model

The sampling step of $s_{1:T}$, $\{\mu_j\}_{j=1}^K$ and P are same as in MS-RV model estimation. Let $y'_{1:t} = \{y'_1, \ldots, y'_t\}$, where $y'_t = \{r_t, RAV_t\}$. $\{\sigma_j\}_{j=1}^K$ and ν are sampled as follows.

The prior of σ_j^2 is assumed to be $\sigma_j^2 \sim \mathrm{G}(v_0, s_0)$. The conditional posterior of σ_j^2 is given as follows,

$$
\begin{aligned}
p(\sigma_j^2 \big| y'_{1:T}, \mu_j, \nu, s_{1:T}) &\propto \prod_{s_t=j} \left\{ \frac{1}{\sigma_j} \exp\left[-\frac{(r_t-\mu_j)^2}{2\sigma_j^2} \right] \cdot \sigma_j^{2\nu} \exp\left[-\left(\frac{\sigma_j \Gamma(\nu)}{\Gamma(\nu-\frac{1}{2})} \right)^2 \frac{1}{RAV_t^2} \right] \right\} \\
&\quad \cdot (\sigma_j^2)^{v_0-1} \exp\left(-s_0 \sigma_j^2 \right).
\end{aligned}
$$

$\sigma_j^{2\prime}$ is drawn from random walk proposal and negative draws discarded.

The prior for ν is assumed to be $\nu \sim \mathrm{IG}(a,b)$. The posterior of ν is given as follows,

$$\begin{aligned} p(\nu \big| RAV_{1:T}, \{\sigma_j^2\}_{j=1}^K, s_{1:T}) \;\; \propto \;\; & \prod_{t=1}^{T} \left\{ \left[\frac{\sigma_j \Gamma(\nu)}{\Gamma(\nu - \frac{1}{2})} \right]^{2\nu} \frac{RAV_t^{-2\nu-1}}{\Gamma(\nu)} \exp\left[- \left(\frac{\sigma_j \Gamma(\nu)}{\Gamma(\nu-\frac{1}{2})} \right)^2 \frac{1}{RAV_t^2} \right] \right\} \\ & \cdot \nu^{-a-1} \exp\left(\frac{b}{\nu} \right). \end{aligned}$$

Random walk proposal is used to sample ν and negative values discarded.

9.4 MS-logRAV Model

The estimation of MS-logRAV model is very similar to that of MS-logRV model except changing the return variance $\exp(\zeta_{s_t})$ to $\exp(2\zeta_{s_t})$.

9.5 MS-RCOV

See step (1) and (5) in Appendix 9.1 for the estimation of $s_{1:T}$ and P. Let $n_j = \sum_{t=1}^{T} \mathbb{1}(s_t = j)$ denotes the number of observations belong to state j.

Given conjugate prior $M_j \sim \mathrm{N}(G_j, V_j)$, the posterior density of M_j is given by

$$M_j \big| R_{1:T}, s_{1:T}, \Sigma_j \sim \mathrm{N}(\overline{M}, \overline{V}), \quad \text{where}$$

$$\overline{V} = \left(\Sigma_j^{-1} n_j + V_j^{-1} \right)^{-1}, \qquad \overline{M} = \overline{V} \left(\Sigma_j^{-1} \sum_{s_t = j} R_t + G_j V_j^{-1} \right).$$

The prior of Σ_j is assumed to be $\Sigma_j \sim \mathrm{W}(\Psi, \tau)$. The conditional posterior of Σ_j is given as follows,

$$\begin{aligned} p(\Sigma_j \big| Y_{1:T}, M_j, \kappa, s_{1:T}) \;\; \propto \;\; & \prod_{s_t=j} \left\{ |\Sigma_j|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (R_t - M_j)^\intercal \Sigma_j^{-1} (R_t - M_j) \right] \right\} \\ & \cdot \prod_{s_t=j} \left\{ |\Sigma_j|^{\frac{\kappa}{2}} |RCOV_t|^{-\frac{\kappa+d+1}{2}} \exp\left[-\frac{1}{2} \mathrm{tr}(\Sigma_j RCOV_t^{-1}) \right] \right\} \\ & \cdot |\Sigma_j|^{\frac{\tau-d-1}{2}} \exp\left[-\frac{1}{2} \mathrm{tr}(\Psi^{-1} \Sigma_j) \right]. \end{aligned}$$

Metropolis-Hasting algorithm is applied to sample Σ_j. The proposal density $q_j(\cdot)$ is formed as follows,

$$
\begin{aligned}
p(\Sigma_j \big| Y_{1:T}, M_j, \kappa, s_{1:T}) \quad &< \quad \prod_{s_t=j} \left\{ |\Sigma_j|^{\frac{\kappa}{2}} \exp\left[-\frac{1}{2} \mathrm{tr}(\Sigma_j RCOV_t^{-1}) \right] \right\} \cdot |\Sigma_j|^{\frac{\tau-d-1}{2}} \exp\left[-\frac{1}{2} \mathrm{tr}(\Psi^{-1}\Sigma_j) \right] \\
&\sim \quad \mathrm{W}\left(\left[(\kappa-1-d) \sum_{s_t=j} RCOV_t^{-1} + \Psi^{-1} \right]^{-1},\ n_j\kappa + \tau \right) \equiv q_j(\cdot).
\end{aligned}
$$

Assuming the prior of κ is $\kappa \sim \mathrm{G}(a, b)$, the posterior density of κ is given as follows,

$$
\begin{aligned}
p(\kappa \big| Y_{1:T}, M_{s_t}, \Sigma_{s_t}, S) \propto \prod_{t=1}^{T} \Bigg\{ & \frac{|\Sigma_{s_t}(\kappa-d-1)|^{\frac{\kappa}{2}}}{2^{\frac{\kappa d}{2}} \Gamma(\frac{\kappa}{2})} |RCOV_t|^{-\frac{\kappa+d+1}{2}} \\
& \cdot \exp\left[-\frac{1}{2} \mathrm{tr}\left((\kappa-d-1)\Sigma_{s_t} RCOV_t^{-1} \right) \right] \Bigg\} \cdot \kappa^{a-1} \exp(-b\kappa).
\end{aligned}
$$

Metropolis-Hasting algorithm with random walk proposal is used to sample κ.

9.6 Univariate Joint IHMM-RV

Define vector $C = \{c_1, \cdots, c_K\}$ and another $K \times K$ matrix A, which will be used in sampling Γ and η. The MCMC steps are illustrated as follows. Several estimation steps are based on Maheu and Yang (2016) and Song (2014).

(1) $u_{1:T} \big| s_{1:T}, \Gamma, P$

Draw $u_1 \sim \mathrm{Uniform}(0, \gamma_{s_1})$ and draw $u_t \sim \mathrm{Uniform}(0, P_{s_{t-1}, s_t})$ for $t = 2, \ldots, T$.

(2) Adjust the Number of States K

i. Check if $\max\{P^r_{1,K+1}, \cdots, P^r_{K,K+1}\} > \min\{u_{1:T}\}$. If yes, expand the number of clusters by making the following adjustments (ii) - (vi), otherwise, move to step (3).

ii. Set $K = K + 1$.

iii. Draw $u_\beta \sim \mathrm{Beta}(1, \eta)$, set $\gamma_K = u_\beta \gamma^r_K$ and the new residual probability equals to $\gamma^r_{K+1} = (1 - u_\beta)\gamma^r_K$.

iv. For $j = 1, \cdots, K$, draw $u_\beta \sim \mathrm{Beta}(\gamma_K, \gamma_{K+1})$, set $P_{j,K} = u_\beta P^r_{j,K}$ and $P^r_{j,K+1} = (1 - u_\beta)P^r_{j,K}$. Also, add an additional row to transition matrix P. $P_{K+1} \sim \mathrm{Dir}(\alpha\gamma_1, \cdots, \alpha\gamma_K)$.

v. Expand the parameter size by 1 by drawing $\mu_{K+1} \sim \mathrm{N}(m, v^2)$ and $\sigma^2_{K+1} \sim \mathrm{IG}(v_0, s_0)$.

vi. Go back to step(i.).

(3) $s_{1:T}|y_{1:T}, u_{1:T}, \theta, \phi, P, \Gamma$

In this step, the latent state variable is sampled using the forward filter backward sampler Chib (1996). The forward filter part contains the following steps:

i. Set the initial value of filter $p(s_1 = j \big| y_1, u_1, \theta, \phi, P) = \mathbb{1}(u_0 < \gamma_j)$ and normalize it.

ii. Prediction step:

$$p(s_t \big| y_{1:t-1}, u_{1:t-1}, \theta, \phi, P) \propto \sum_{j=1}^{K} \mathbb{1}(u_t < P_{j,s_t}) \cdot p(s_{t-1} = j \big| y_{1:t-1}, u_{1:t-1}, \theta, \phi, P).$$

iii. Update step:

$$p(s_t \big| y_{1:t}, u_{1:t}, \theta, \phi, P) \propto f(r_t \big| \mu_j, \sigma_j^2) \cdot g(RV_t \big| \nu + 1, \nu\sigma_j^2) \cdot p(s_t \big| y_{1:t-1}, u_{1:t-1}, \theta, \phi, P).$$

The underlying states are drawn using backward sampler as follows.

i. For $t = T$, draw s_T from $p(s_T \big| y_{1:T}, u_{1:T}, \theta, \phi, P)$.

ii. For $t = T-1, \cdots, 1$, draw s_t from $\mathbb{1}(u_t < P_{j,s_{t+1}}) \cdot p(s_t \big| y_{1:t}, u_{1:t}, P, \theta, \phi)$.

Then we count the number of active clusters and removing inactive states by making following adjustments.

i. Calculate the number of active states (states with at least one observation assigned to it) denoted by L. If $L < K$, remove the inactive states by adjusting the value of states.

ii. Adjust the order of state-dependent parameters μ, σ^2 and Γ according to the adjusted state $s_{1:T}$.

iii. Set $K = L$. Recalculate the residual probabilities of Γ_{K+1}^r for $j = 1, \cdots, K$. Then set the values of parameter μ_j, σ_j^2 and γ_j, to be zero for $j > K$.

(4) $\Gamma \big| s_{1:T}, \eta, \alpha$

i. Let $n_{j,i}$ denotes the number of state moves from state j to i. Calculate $n_{j,i}$ for $i = 1, \cdots, K$ and $j = 1, \cdots, K$.

ii. For $i = 1, \cdots, K$ and $j = 1, \cdots, K$, if $n_{j,i} > 0$, then for $l = 1, \cdots, n_{j,i}$, draw $x_l \sim$ Bernoulli$(\frac{\alpha\gamma_i}{l-1+\alpha\gamma_i})$. If $x_l = 1$, set $A_{j,i} = A_{j,i} + 1$.

iii. Draw $\Gamma \sim \text{Dir}(c_1, \ldots, c_K, \eta)$, where $c_i = \sum_{j=1}^{K} A_{ji}$.

(5) $P \big| s_{1:T}, \Gamma, \alpha$

For $j = 1, \cdots K$, draw $P_j \sim \text{Dir}(\alpha\gamma_1 + n_{j,1}, \cdots, \alpha\gamma_k + n_{j,K}, \alpha\gamma_{K+1}^{r})$.

(6) $\theta \big| y_{1:T}, s_{1:T}, \nu$

See the step (2) and step (3) in Appendix 9.1. for the estimation of state-dependent parameters μ_j, σ_j^2, for $j = 1, \ldots, K$.

(7) $\nu \big| y_{1:T}, s_{1:T}, \{\sigma_j^2\}_{j=1}^{K}, \nu$

Same as the step (4) in Appendix 9.1.

(8) $\eta \big| s_{1:T}, \Gamma, \alpha$

Recompute C vector again as in step (4) and define ν and λ, where $\nu \sim \text{Bernoulli}(\frac{\sum_{i=1}^{K} c_i}{\sum_{i=1}^{K} c_i + \eta})$ and $\lambda \sim \text{Beta}(\eta + 1, \sum_{i=1}^{K} c_i)$. Then draw a new value of $\eta \sim \text{G}(a_1 + K - \nu, b_1 - \log(\lambda))$.

(9) $\alpha \big| s_{1:T}, C$

Define ν_j', λ_j', for $j = 1, \cdots, K$, where $\nu_j' \sim \text{Bernoulli}(\frac{\sum_{i=1}^{K} n_{j,i}}{\sum_{i=1}^{K} n_{j,i} + \alpha})$ and $\lambda_j' \sim \text{Beta}(\alpha + 1, \sum_{i=1}^{K} n_{j,i})$. Then draw $\alpha \sim \text{G}(a_2 + \sum_{j=1}^{K} c_j - \sum_{j=1}^{K} \nu_j', b_2 - \sum_{j=1}^{K} \log(\lambda_j'))$.

9.7 Univariate Joint IHMM-logRV and Joint IHMM-logRAV

See step (1) - (5), (8) and (9) in Appendix 9.6 for the estimation of auxiliary variable $u_{1:T}$, latent state variable $s_{1:T}$, Γ, transition matrix P, DP concentration parameter η and α. The estimation of $\theta = \{\mu_j, \zeta_j, \delta_j^2\}_{j=1}^{\infty}$ in IHMM-logRV are same as the MS-logRV model, see Appendix 9.3. The parameter estimation of IHMM-logRAV can be done similarly.

9.8 Joint IHMM-RCOV

See step (1) - (5), (8) and (9) in Appendix 9.6 for the estimation of $u_{1:T}$, $s_{1:T}$, Γ, P, η and α. The estimation of $\theta = \{M_j, \Sigma_j\}_{j=1}^{\infty}$ and κ are same as the estimation of MS-RCOV model, see Appendix 9.5.

References

Alizadeh S, Brandt MW, Diebold F. 2002. Range-based estimation of stochastic volatility models. *The Journal of Finance* **57**: 1047–1091.

Andersen TG, Benzoni L. 2009. Realized volatility. In *Handbook of Financial Time Series*. Springer, 555–576.

Andersen TG, Bollerslev T, Diebold FX, Ebens H. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* **61**: 43–76.

Ang A, Bekaert G. 2002. Regime switches in interest rates. *Journal of Business and Economic Statistics* **20**: 163–182.

Barndorff-Nielsen OE, Shephard N. 2002. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* **17**: 457–477.

Barndorff-Nielsen OE, Shephard N. 2004a. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72**: 885–925.

Barndorff-Nielsen OE, Shephard N. 2004b. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* **2**: 1–48.

Blair BJ, Poon SH, Taylor SJ. 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics* **105**: 5–26.

Carpantier JF, Dufays A. 2014. Specific Markov-switching behaviour for arma parameters. CORE Discussion Papers 2014014, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Chib S. 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* **75**: 79–97.

Clements A, Silvennoinen A. 2013. Volatility timing: How best to forecast portfolio exposures. *Journal of Empirical Finance* **24**: 108–115.

Dueker M, Neely CJ. 2007. Can Markov switching models predict excess foreign exchange returns? *Journal of Banking and Finance* **31**: 279–296.

Dufays A. 2016. Infinite-state Markov-switching for dynamic volatility. *Journal of Financial Econometrics* **14**: 418–460.

Engel C, Hamilton JD. 1990. Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review* **80**: 689–713.

Engle R, Colacito R. 2006. Testing and valuing dynamic correlations for asset allocation. *Journal of Business and Economic Statistics* **24**: 238–253.

Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**: 209–230.

Fleming J, Kirby C, Ostdiek B. 2001. The economic value of volatility timing. *The Journal of Finance* **56**: 329–352.

Forni M, Reichlin L. 1998. Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *The Review of Economic Studies* **65**: 453–473.

Gael JV, Saatci Y, Teh YW, Ghahramani Z. 2008. Beam sampling for the infinite hidden Markov model. In *In Proceedings of the 25th International Conference on Machine Learning*. 1088–1095.

Greenberg E. 2014. *Introduction to Bayesian Econometrics*. Cambridge University Press.

Guidolin M, Timmermann A. 2006. An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics* **21**: 1–22.

Guidolin M, Timmermann A. 2008. International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies* **21**: 889–935.

Guidolin M, Timmermann A. 2009. Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics* **150**: 297–311.

Hamilton JD. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**: 357–384.

Hansen P, Huang Z, Shek HH. 2012. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* **27**: 877–906.

Hansen PR, Lunde A, Voev V. 2014. Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics* **29**: 774–799.

Jin X, Maheu JM. 2013. Modeling realized covariances and returns. *Journal of Financial Econometrics* **11**: 335–369.

Jin X, Maheu JM. 2016. Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics* **192**: 19–39.

Jochmann M. 2015. Modeling U.S. inflation dynamics: A Bayesian nonparametric approach. *Econometric Reviews* **34**: 537–558.

Kim CJ, Morley JC, Nelson CR. 2004. Is there a positive relationship between stock market volatility and the equity premium? *Journal of Money, Credit and Banking* **36**: 339–360.

Kose MA, Otrok C, Whiteman CH. 2003. International business cycles: World, region, and country-specific factors. *American Economic Review* **93**: 1216–1239.

Lunde A, Timmermann AG. 2004. Duration dependence in stock prices: An analysis of bull and bear markets. *Journal of Business and Economic Statistics* **22**: 253–273.

Maheu JM, McCurdy TH. 2000. Identifying bull and bear markets in stock returns. *Journal of Business and Economic Statistics* **18**: 100–112.

Maheu JM, McCurdy TH. 2011. Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics* **160**: 69–76.

Maheu JM, McCurdy TH, Song Y. 2012. Components of bull and bear markets: Bull corrections and bear rallies. *Journal of Business and Economic Statistics* **30**: 391–403.

Maheu JM, Yang Q. 2016. An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance* **38**: 202–220.

Markowitz H. 1952. Mean-variance analysis in portfolio choice and financial markets. *The Journal of Finance* **7**: 77–91.

Noureldin D, Shephard N, Sheppard K. 2012. Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* **27**: 907–933.

Pastor L, Stambaugh RF. 2001. The equity premium and structural breaks. *The Journal of Finance* **56**: 1207–1239.

Rydén T, Teräsvirta T, Åsbrink S. 1998. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics* **13**: 217–244.

Schwert GW. 1990. Indexes of U.S. stock prices from 1802 to 1987. *Journal of Business* **63**: 399–426.

Shephard N, Sheppard K. 2010. Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* **25**: 197–231.

Skouras S. 2007. Decisionmetrics: A decision-based approach to econometric modelling. *Journal of Econometrics* **137**: 414–440.

Song Y. 2014. Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics* **29**: 825–842.

Stock J, Watson M. 2010. *Dynamic Factor Models*. Oxford: Oxford University Press.

Takahashi M, Omori Y, Watanabe T. 2009. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics and Data Analysis* **53**: 2404–2426.

Teh YW, Jordan MI, Beal MJ, Blei DM. 2006. Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101**: 1566–1581.

Walker SG. 2007. Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation* **36**: 45–54.

Zellner A. 1971. *An Introduction to Bayesian Inference in Econometrics*. John Wikey and Sons.

Table 1: Prior Specifications of Univariate Return Models

Panel A: Priors for MS and Joint MS Models					
Model	μ_{s_t}	$\sigma^2_{s_t}$	ν	$\delta^2_{s_t}$	P_j
MS	N$(0,1)$	IG$(2, \widehat{\text{var}}(r_t))$	-		Dir$(1,\ldots,1)$
MS-RV	N$(0,1)$	G$(\overline{RV_t}, 1)$	IG$(2,1)$		Dir$(1,\ldots,1)$
MS-RAV	N$(0,1)$	G$(\overline{RV_t}, 1)$	IG$(2,1)$		Dir$(1,\ldots,1)$
MS-logRV	N$(0,1)$	N$(\overline{\log(RV_t)}, 5)$	-	IG$(2,0.5)$	Dir$(1,\ldots,1)$
MS-logRAV	N$(0,1)$	N$(\overline{\log(RAV_t)}, 5)$	-	IG$(2,0.5)$	Dir$(1,\ldots,1)$

Panel B: Priors for IHMM and Joint IHMM Models						
Model	μ_{s_t}	$\sigma^2_{s_t}$	ν	$\delta^2_{s_t}$	η	α
IHMM	N$(0,1)$	IG$(2, \widehat{\text{var}}(r_t))$	-	-	G$(1,4)$	G$(1,4)$
IHMM-RV	N$(0,1)$	G$(\overline{RV_t}, 1)$	IG$(2,1)$	-	G$(1,4)$	G$(1,4)$
IHMM-logRV	N$(0,1)$	N$(\overline{\log(RV_t)}, 5)$	-	IG$(2,0.5)$	G$(1,4)$	G$(1,4)$
IHMM-logRAV	N$(0,1)$	N$(\overline{\log(RAV_t)}, 5)$	-	IG$(2,0.5)$	G$(1,4)$	G$(1,4)$

$\widehat{\text{var}}(r_t)$ is the sample variance, $\overline{RV_t}$, $\overline{\log(RV_t)}$ and $\overline{\log(RAV_t)}$ are the sample means. All are computed using in-sample data.

Table 2: Summary Statistics for Monthly Equity Returns and Volatility Measures

Data	Mean	Median	Stdev	Skewness	Kurtosis	Min	Max
r_t	0.047	0.097	0.612	-0.539	9.123	-4.154	3.884
RV_t	0.328	0.156	0.621	6.853	68.499	0.010	8.580
RAV_t	0.470	0.394	0.287	2.807	14.358	0.103	2.747
$\log(RV_t)$	-1.720	-1.856	0.964	0.714	3.992	-4.608	2.149
$\log(RAV_t)$	-0.882	-0.931	0.476	0.682	3.869	-2.274	1.010

This table reports the summary statistics for monthly returns and various ex-post proxies of volatility. See the text for definitions. The sample period is from March 1885 to December 2013 and the number of observations is 1542. (Note: Market closed between July 1914 and December 1914 due to World War I).

Table 3: Equity Forecasts: Jan. 1951 - Dec. 2013

No. of States	Models	Log-predictive Likelihoods	RMSE[r_{t+1}]	RMSE[RV_{t+1}]
2 States	MS	-548.409	0.5268	0.5285
	MS-RV	-535.003	0.5242*	0.5338
	MS-logRV	-534.914	0.5276	0.5229
	MS-RAV	-533.370*	0.5263	0.5263
	MS-logRAV	-534.256	0.5269	0.5199*
3 States	MS	-538.437	0.5244*	0.5240
	MS-RV	-523.000*	0.5290	0.5070
	MS-logRV	-524.754	0.5286	0.5087
	MS-RAV	-523.171	0.5276	0.5032*
	MS-logRAV	-525.353	0.5283	0.5048
4 States	MS	-535.454	0.5232*	0.5193
	MS-RV	-520.363*	0.5273	0.5029
	MS-logRV	-528.631	0.5284	0.4902*
	MS-RAV	-527.708	0.5277	0.4976
	MS-logRAV	-530.697	0.5290	0.4920
-	IHMM	-535.165	0.5229	0.5348
	IHMM-RV	**-514.662**	**0.5216**	0.4724
	IHMM-logRV	-516.643	0.5228	**0.4647**
	IHMM-logRAV	-517.148	0.5244	0.4775

This table reports the sum of 1-period ahead log-predictive likelihoods of return $\sum_{j=t+1}^{T} \log(p(r_j|y_{1:j-1}, \text{Model}))$, root mean squared error for return and realized variance predictions over period from Jan 1951 to Dec 2013 (756 observations). The symbol $*$ identifies the largest log-predictive likelihood and the smallest RMSE given a fixed number of states. Bold text indicates the largest log-predictive likelihood and the smallest RMSE in the entire column.

Table 4: Equity Forecasts: Jan. 1984 - Dec. 2013

No. of States	Models	Log-Predictive Likelihoods	RMSE$[r_{t+1}]$	RMSE$[RV_{t+1}]$
2 States	MS	-300.019	0.5542	0.6863
	MS-RV	-293.311	0.5512*	0.7130
	MS-logRV	-291.794	0.5563	0.6938*
	MS-RAV	-290.353*	0.5543	0.7050
	MS-logRAV	-290.709	0.5553	0.6968
3 States	MS	-294.914	0.5522*	0.6817
	MS-RV	-283.877	0.5570	0.6794
	MS-logRV	-284.135	0.5568	0.6781
	MS-RAV	-281.126*	0.5556	0.6764*
	MS-logRAV	-282.150	0.5561	0.6772
4 States	MS	-292.397	0.5506*	0.6755
	MS-RV	-281.211*	0.5553	0.6749
	MS-logRV	-285.383	0.5570	0.6564*
	MS-RAV	-282.523	0.5559	0.6696
	MS-logRAV	-284.563	0.5580	0.6614
-	IHMM	-291.091	0.5529	0.7002
	IHMM-RV	**-279.504**	**0.5475**	0.6344
	IHMM-logRV	-281.344	0.5503	**0.6209**
	IHMM-logRAV	-280.019	0.5529	0.6434

This table reports the sum of 1-period ahead log-predictive likelihoods of return $\sum_{j=t+1}^{T} \log(p(r_j|y_{1:j-1}, \text{Model}))$, root mean squared error for return and realized variance predictions over period from Jan 1984 to Dec 2013 (360 observations). The symbol $*$ identifies the largest log-predictive likelihood and the smallest RMSE given a fixed number of states. Bold text indicates the largest log-predictive likelihood and the smallest RMSE in the entire column.

Table 5: Log-Predictive Bayes Factors for Market Declines

Market Declines	Period	$\sum \log \frac{p(r_{t+1}\|r_{1:t}, \text{IHMM-RV})}{p(r_{t+1}\|r_{1:t}, \text{IHMM})}$
1973-74 stock market crash	Feb. 1973 - Dec. 1974	0.2878
Black Monday	Oct. 1987 - Dec. 1987	1.4915
Dot-com bubble	Jan. 2000 - Dec. 2002	2.3788
Financial crisis of 2007-08	Jul. 2007 - Dec. 2008	1.5615

This table reports log-predictive Bayes factors for the IHMM-RV model versus the IHMM over several sample periods.

Table 6: Estimates for Stock Market Returns

	MS		MS-RV		MS-RAV	
Parameter	Mean	Stdev	Mean	Stdev	Mean	Stdev
μ_1	-0.2995 (-0.528, -0.089)	0.1104	-0.0840 (-0.159, -0.020)	0.0354	-0.1372 (-0.229, -0.050)	0.0445
μ_2	0.0875 (0.060, 0.116)	0.0140	0.1224 (0.096, 0.149)	0.0139	0.1130 (0.089, 0.138)	0.0125
σ_1^2	1.6127 (1.191, 2.217)	0.2630	0.5633 (0.481, 0.678)	0.0486	0.6581 (0.597, 0.748)	0.0394
σ_2^2	0.2187 (0.196, 0.241)	0.0115	0.1556 (0.143, 0.171)	0.0049	0.1460 (0.138, 0.154)	0.0044
ν	-	-	1.3431 (1.183, 1.501)	0.0824	2.1529 (2.009, 2.316)	0.0758
$P_{1,1}$	0.8775 (0.793, 0.943)	0.0396	0.9023 (0.862, 0.937)	0.0193	0.8716 (0.828, 0.910)	0.0210
$P_{2,2}$	0.9849 (0.972, 0.994)	0.0058	0.9442 (0.924, 0.962)	0.0099	0.9538 (0.937, 0.969)	0.0083

This table reports the posterior mean, standard deviation and 0.95 density intervals (values in brackets) of parameters of selected 2 state models. The prior restriction $\mu_1 < 0$ and $\mu_2 > 0$ is imposed. The sample period is from March 1885 to December 2013 (1542 observations).

Table 7: Performance of Market Timing Portfolios

Number of states	Model	Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
Panel A: Jan. 1951 - Dec. 2013						
Buy-Hold		0.0826	0.5148	0.1604	-	-
2 States	MS	0.0663	0.4198	0.1578	-	-
	MS-RV	0.0720	0.3929	**0.1831**	**92.256**	**96.192**
4 States	MS	0.0690	0.4427	0.1560	-7.188	-4.512
	MS-RV	0.0698	0.4040	**0.1729**	**55.860**	**60.372**
-	IHMM	0.0657	0.4414	0.1487	-37.608	-37.884
	IHMM-RV	0.0770	0.4348	**0.1770**	**82.392**	**87.888**
Panel B: Jan. 1984 - Dec. 2013						
Number of states	Model	Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
Buy-Hold		0.0911	0.5408	0.1684	-	-
2 States	MS	0.0564	0.4755	0.1187	-	-
	MS-RV	0.0772	0.3909	**0.1975**	**325.536**	**337.728**
4 States	MS	0.0684	0.4845	0.1412	100.476	108.540
	MS-RV	0.0645	0.4291	**0.1502**	**145.476**	**156.852**
-	IHMM	0.0563	0.4975	0.1132	-37.212	-37.692
	IHMM-RV	0.0794	0.4590	**0.1729**	**247.956**	**261.276**

The summary statistics are based on annualized returns. The values in the last two columns are annualized basis point performance fees that an investor is willing to pay to switch away from the 2-state Markov switching model to the one in the respective row. Δ_q is annualized basis point performance fee based on quadratic utility and Δ_e is the performance fee for investor with exponential utility. The risk aversion coefficient is $\gamma = 5$. Bold numbers indicate the largest Sharpe ratio and the largest performance fee for a class of models.

Table 8: Prior Specification of Multivariate Models

Model	M_{s_t}	Σ_{s_t}	κ	P_j
Panel A: Priors for Multivariate MS Models				
MS	$N(\mathbf{0}, 5\mathbf{I})$	$IW(\widehat{\text{Cov}}(R_t), 5)$	-	$\text{Dir}(1, \ldots, 1)$
MS-RCOV	$N(\mathbf{0}, 5\mathbf{I})$	$W(\frac{1}{3}\overline{RCOV_t}, 3)$	$G(20, 1)\mathbb{1}_{\kappa>4}$	$\text{Dir}(1, \ldots, 1)$

Model	M_{s_t}	Σ_{s_t}	κ	η	α
Panel B: Priors for Multivariate IHMM Models					
IHMM	$N(\mathbf{0}, 5\mathbf{I})$	$IW(\widehat{\text{Cov}}(R_t), 5)$	-	$G(1, 4)$	$G(1, 4)$
IHMM-RCOV	$N(\mathbf{0}, 5\mathbf{I})$	$W(\frac{1}{3}\overline{RCOV_t}, 3)$	$G(20, 1)\mathbb{1}_{\kappa>4}$	$G(1, 4)$	$G(1, 4)$

$\mathbf{0}$ denotes zero vector, $\mathbf{I}$ is the identity matrix. $\widehat{\text{Cov}}(R_t)$, and $\overline{RCOV_t}$ are computed using in sample data.

Table 9: Summary Statistics of Returns (IBM, XOM, GE)

Panel A: Summary of Returns

Data	Mean	Median	St. Dev	Skewness	Kurtosis	Min	Max
IBM	0.134	0.135	0.824	-0.192	5.169	-3.644	3.635
XOM	0.116	0.096	0.707	-0.152	6.942	-3.930	3.773
GE	0.103	0.088	0.940	-0.324	7.755	-5.265	5.336

Panel B: Return Covariance and RCOV Mean

	Covariance of Return			Average of RCOV		
Data	IBM	XOM	GE	IBM	XOM	GE
IBM	0.678	0.236	0.411	0.742	0.250	0.370
XOM	0.236	0.500	0.338	0.250	0.641	0.394
GE	0.411	0.338	0.882	0.370	0.394	0.970

The panel A of above table reports the summary statistics of the monthly return of IBM, XOM and GE. The reported data are annualized values after scaling the raw returns by 12. The panel B reports the covariance matrix calculated from monthly return vectors and the averaged RCOV matrix, which are calculated using daily returns. The sample period is from Jan 1926 to Dec 2013 (1056 observations).

Table 10: Multivariate Equity Forecasts

No. of States	Models	Log Predictive Likelihoods	$\|\|\text{RMSE}(R_{t+1})\|\|$	$\|\|\text{RMSE}(RCOV_{t+1})\|\|$
8 States	MS	-2294.571	**1.2843**	2.4230
	MS-RCOV	-2264.490*	1.2857	2.2049*
12 States	MS	-2315.345	1.2849*	2.4979
	MS-RCOV	-2270.969*	1.2856	2.2320*
-	IHMM	-2274.063	1.2877	2.3651
	IHMM-RCOV	**-2262.383**	1.2873*	**2.1956**

This table summarizes the sum of 1 month log predictive likelihoods of return, $\sum_{j=t+1}^{T} \log(p(R_j|y_{1:j-1}, \text{Model}))$, root mean squared errors of mean and covariance prediction over Jan 1951 to Dec 2013 (totally 756 predictions), when the models are applied to analyze IBM, XOM, GE jointly. The root mean squared errors provided in this table are matrix norms. $||A|| = \sqrt{\sum_i \sum_j a_{ij}^2}$. The symbol $*$ identifies the largest log-predictive likelihood and the smallest RMSE given a fixed number of states. Bold text indicates the largest log-predictive likelihood and the smallest RMSE in the entire column.

Table 11: Performance of Minimum Variance Portfolios (Jan. 1951 - Dec. 2013)

Panel A: Required Annualized Portfolio Return: 10%						
		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.1103	0.3296	0.3345	-	-
	MS-RCOV	0.1090	0.3155	**0.3455**	**7.080**	**3.888**
12 States	MS	0.1103	0.3369	0.3273	-10.704	-10.188
	MS-RCOV	0.1084	0.3137	**0.3456**	**3.744**	**0.900**
-	IHMM	0.1081	0.3257	0.3320	-15.924	-16.224
	IHMM-RCOV	0.1089	0.3170	**0.3436**	**4.188**	**0.600**
Panel B: Required Annualized Portfolio Return: 20%						
		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.1956	0.8916	0.2194	-	-
	MS-RCOV	0.1962	0.8660	**0.2265**	**106.752**	**83.784**
12 States	MS	0.1934	0.9161	0.2111	-123.468	-116.364
	MS-RCOV	0.1970	0.8626	**0.2284**	**128.772**	**103.104**
-	IHMM	0.1933	0.8715	0.2218	56.940	45.876
	IHMM-RCOV	0.1985	0.8695	**0.2283**	**116.520**	**90.996**

The summary statistics are based on annualized return (scaled by 12). Δ_q is annualized basis point performance fees that an investor with quadratic utility is willing to pay to switch from the 8-state MS model. Δ_e is the fee for investor with exponential utility. The risk aversion coefficients are both utility function are $\gamma = 5$. Panel A and B shows the results of portfolio with 10% and 20% required return, respectively. Out of sample period: Jan 1951 - Dec 2013, totally 756 months. Bold numbers indicate the largest Sharpe ratio and the largest performance fee for a class of models.

Table 12: Performance of Minimum Variance Portfolios (Jan. 1984 - Dec. 2013)

Panel A: Required Annualized Portfolio Return: 10%

		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.0907	0.3370	0.2692	-	-
	MS-RCOV	0.0925	0.3239	**0.2856**	**36.612**	**33.060**
12 States	MS	0.0889	0.3491	0.2545	-36.708	-36.336
	MS-RCOV	0.0938	0.3222	**0.2912**	**51.936**	**48.348**
-	IHMM	0.0936	0.3354	0.2791	30.972	**33.420**
	IHMM-RCOV	0.0925	0.3232	**0.2860**	**36.948**	32.880

Panel B: Required Annualized Portfolio Return: 20%

		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.1788	0.9140	0.1956	-	-
	MS-RCOV	0.1833	0.8934	**0.2052**	**128.520**	**99.360**
12 States	MS	0.1728	0.9542	0.1811	-229.368	-217.260
	MS-RCOV	0.1869	0.8857	**0.2110**	**195.588**	**161.340**
-	IHMM	0.1871	0.8966	**0.2087**	**153.588**	**154.812**
	IHMM-RCOV	0.1831	0.8967	0.2042	113.532	75.228

The summary statistics are based on annualized return (scaled by 12). Δ_q is annualized basis point performance fees that an investor with quadratic utility is willing to pay to switch from the 8-state MS model. Δ_e is the fee for investor with exponential utility. The risk aversion coefficients for both utility function are $\gamma = 5$. Panel A and B shows the results of portfolio with 10% and 20% required return, respectively. Out of sample period: Jan 1984 - Dec 2013, totally 360 months. Bold numbers indicate the largest Sharpe ratio and the largest performance fee for a class of models.

Table 13: Return Variances of Global Minimum Variance Portfolios

Panel A: Jan. 1951 - Dec. 2013		
No. of States	Model	Variance
8 States	MS	0.31123
	MS-RCOV	**0.29601**
12 States	MS	0.32271
	MS-RCOV	**0.29191**
-	IHMM	0.30325
	IHMM-RCOV	**0.29675**
Panel B: Jan. 1984 - Dec. 2013		
No. of States	Model	Variance
8 States	MS	0.31498
	MS-RCOV	**0.31112**
12 States	MS	0.33399
	MS-RCOV	**0.30378**
-	IHMM	0.31098
	IHMM-RCOV	**0.31047**

This table reports variances of annualized return of global minimum variance portfolios based on competing models. Bold numbers indicate the smallest portfolio variance for a class of models.

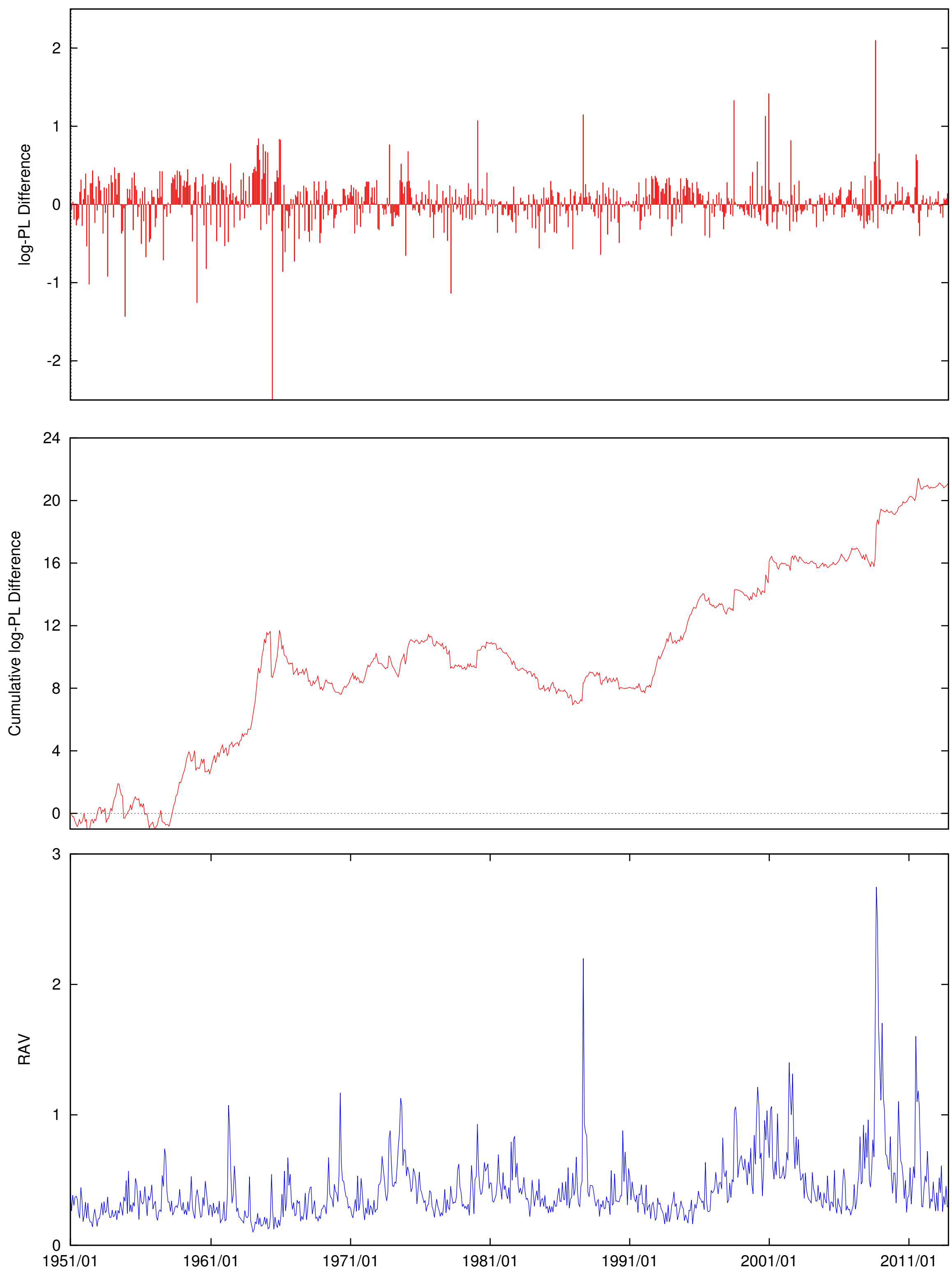

Figure 1: The top panel shows $\log\left(p(r_{t+1}|y_{1:t}, \text{IHMM-RV})\right) - \log\left(p(r_{t+1}|y_{1:t}, \text{IHMM})\right)$ over January 1984 to December 2013. The second panel plots the cumulative log-predictive likelihood difference. The final panel is the time series plot of realized absolute variation.

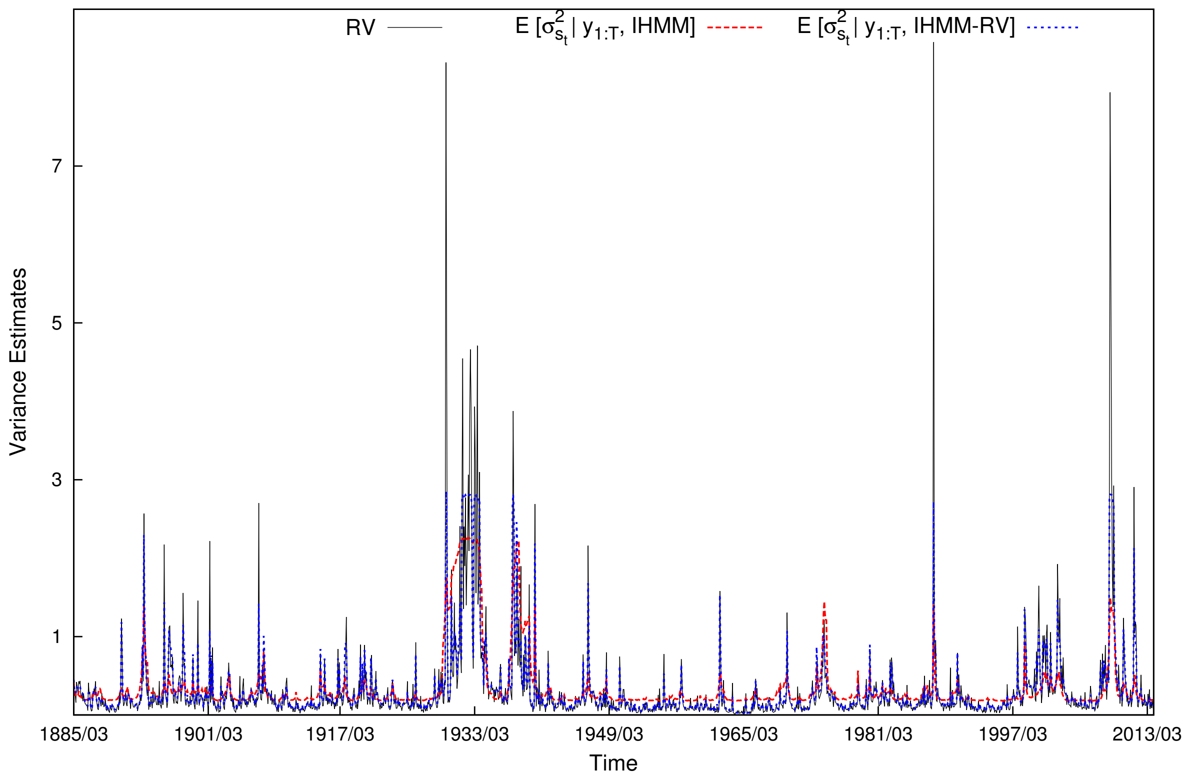

Figure 2: $E\left[\sigma^2_{s_t}\middle|y_{1:T},\text{IHMM}\right]$, $E\left[\sigma^2_{s_t}\middle|y_{1:T},\text{IHMM-RV}\right]$ and Realized Variance

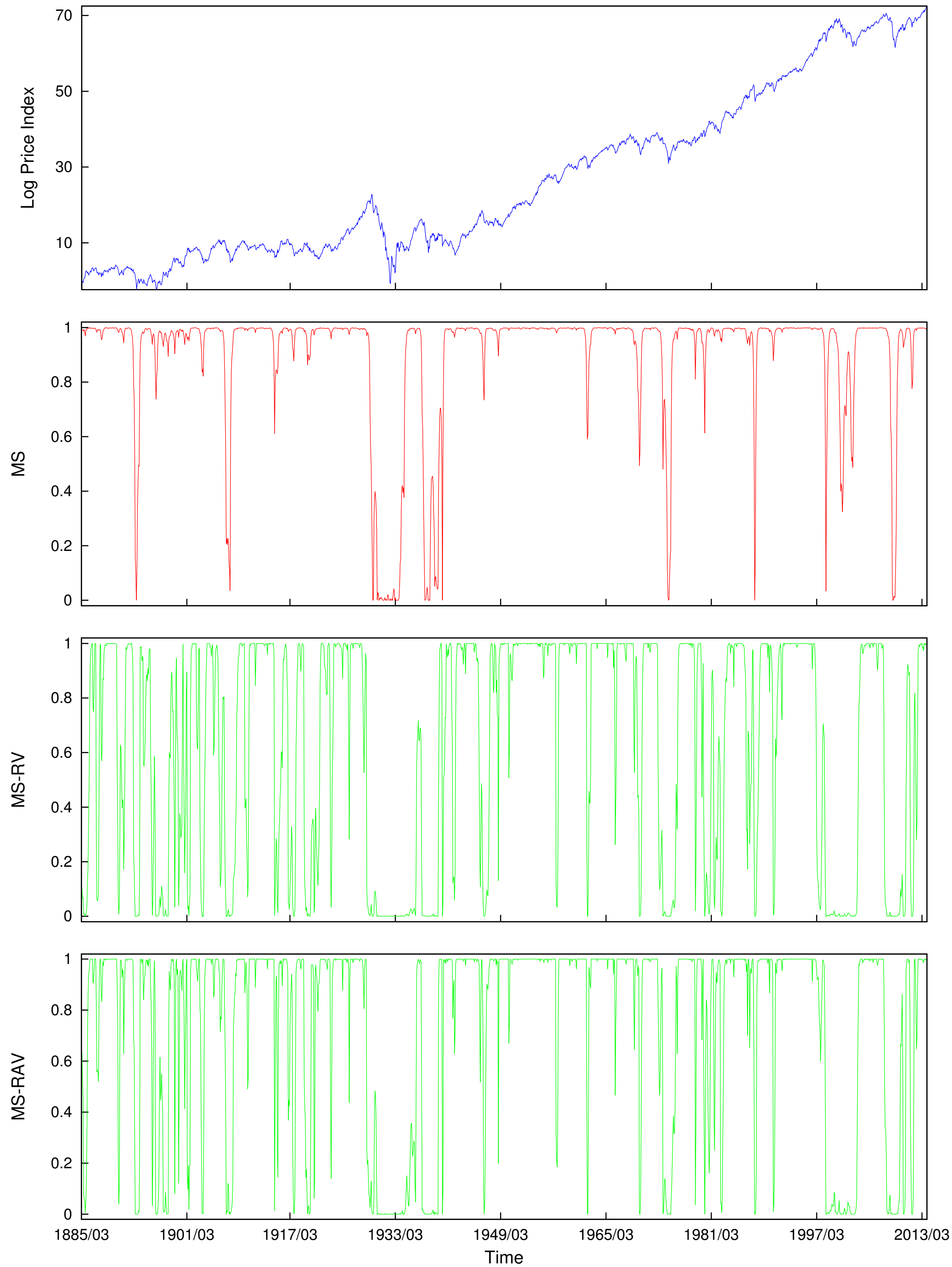

Figure 3: Smoothed Probability of High Return State

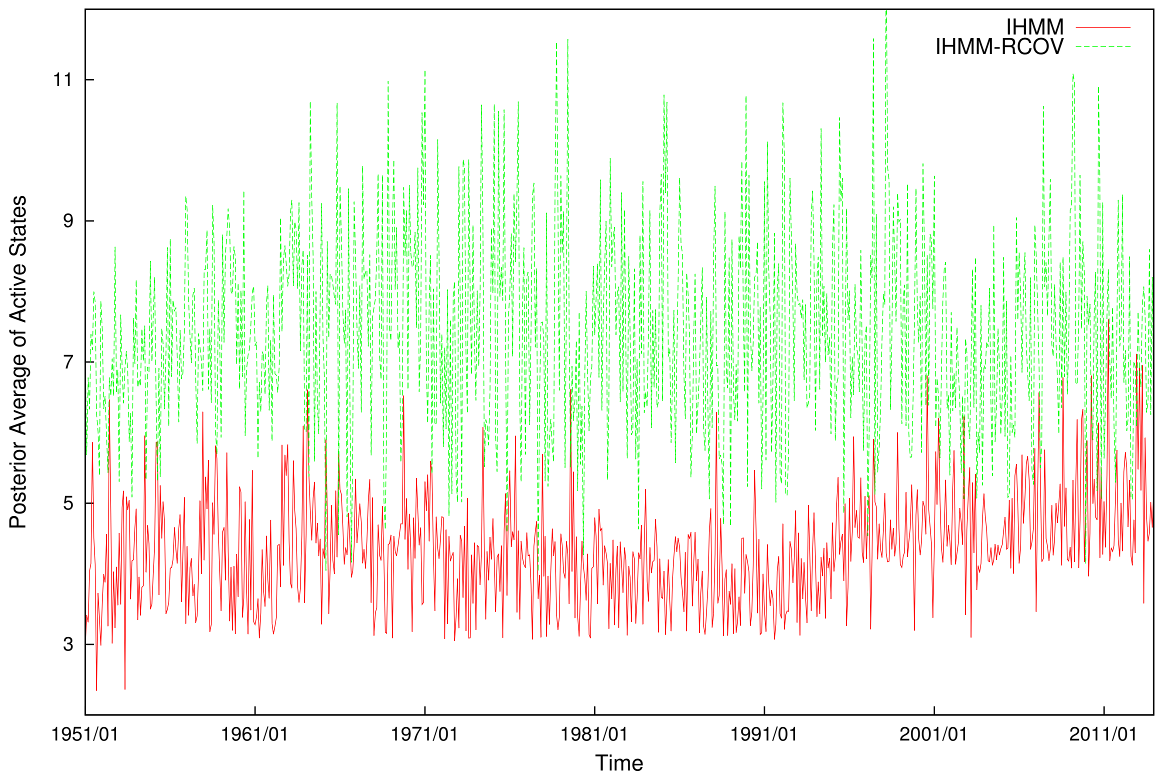

Figure 4: Number of Active Clusters: IHMM and IHMM-RCOV Models