A multivariate GARCH model with an infinite hidden Markov mixture

Li, Chenxing

Center for Economics, Finance and Management Studies, Hunan University

16 March 2022

A Multivariate GARCH Model with an Infinite Hidden Markov Mixture*

Chenxing Li†

This draft March 2022

*I would like to thank John M. Maheu, Ronald Balvers, Dean Mountain and Narat Charupat for their valuable suggestions for this paper. I also thank Sebastien Blais, Marco Del Negro, Arnaud Dufays, Hedibert Lopes, Davide Pettenuzzo, Neil Shephard, Herman van Dijk for their helpful comments and suggestions, and also for the comments from the conference participants of the 2019 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics (SBIES), the 13th RCEA Bayesian Econometrics Workshop and 10th European Seminar on Bayesian Econometrics (ESOBE).

†Center for Economics, Finance and Management Studies, Hunan University, China, lichenxing@hnu.edu.cn

Abstract

This paper proposes a new Bayesian semiparametric model that combines a multivariate GARCH (MGARCH) component and an infinite hidden Markov model. The new model nonparametrically approximates both the shape of unknown returns distributions and their short-term evolution. It also captures the smooth trend of the second moment with the MGARCH component and the potential skewness, kurtosis, and volatility roughness with the Bayesian nonparametric component. The results show that this more-sophisticated econometric model not only has better out-of-sample density forecasts than benchmark models, but also provides positive economic gains for a CRRA investor at different risk-aversion levels when transaction costs are assumed. After considering the transaction costs, the proposed model dominates all benchmark models/portfolios when No Short-Selling or No Margin-Trading restriction is imposed.

Keywords: Multivariate GARCH, IHMM, Bayesian nonparametric, Portfolio allocation, Transaction costs

JEL codes: C53; C58; C14; C32; C11; C34

1 Introduction

It is well known that conditional returns distributions exhibit fat tails and potential skewness, and the shape of those distributions are constantly changing over time. Providing better density forecasts of conditional distributions requires an econometric model that is flexible enough to capture all these shapes and dynamics. This paper proposes a new multivariate Bayesian semiparametric model that allows the shape of the conditional distributions to change over time nonparametrically. The new model produces better density forecasts and portfolio allocations than all of the benchmark models.

Changing volatility is among the first distributional features brought to econometricians' attention. One branch in the multivariate setting deals with extensions of the univariate GARCH models (Bollerslev, 1986). Popular examples include VEC (Bollerslev et al., 1988), BEKK (Engle and Kroner, 1995), constant conditional correlation (CCC, Bollerslev, 1990), dynamic conditional correlation (DCC, Tse and Tsui, 2002; Engle, 2002) and many others. All of these models assume that the returns innovation distributions are i.i.d normal. The innovation can be easily replaced with some fat-tailed distributions to incorporate leptokurtic returns (see Pesaran and Pesaran, 2010; Kawakatsu, 2006; Bonato, 2012; Peng and Kim, 2020, and many others), but the conditional distribution is still assumed to be symmetric and constant.

Mixture modelling is an alternative approach to modelling non-normal innovation distributions. Examples include the structural breaking model (Chib, 1998; Pástor and Stambaugh, 2001; Pesaran et al., 2006; Pettenuzzo and Timmermann, 2011) and the Markov-switching (MS) model (Hamilton, 1989; Rydén et al., 1998; Maheu and McCurdy, 2000; Guidolin and Timmermann, 2007, 2008). Both models can generate skewness and kurtosis but they also assume the returns distribution is known and mixed from a fixed number of states.

Instead of employing a finite number of states, the Dirichlet process mixture (DPM) model approximates the unknown conditional returns distributions by mixing an infinite number of normal kernels over a Dirichlet process (Antoniak, 1974). This model can fruitfully mimic distributional features, such as skewness, kurtosis and asymmetric tails for any continuous distribution. Jensen and Maheu (2013) and Maheu and Shamsi Zamenjani (2021) combine the DPM with a multivariate GARCH model (MGARCH-DPM) to capture the unknown shape of returns innovation distributions in addition to the smooth dynamics of conditional covariances. While the DPM relaxes the parametric assumption that the data-generating process of the innovation distributions is known, the mixture weights remain constant, and the time-variation of the returns distributions is only contributed from the GARCH part of the model.

The infinite hidden Markov model (IHMM) extends the DPM by using time-varying mixing weights (Beal et al., 2002). Its Markovian mixing weights are constructed from a hierarchical Dirichlet process (HDP) prior formalized by Teh et al. (2006), which consists of two layers of nested Dirichlet processes, so the shape of the approximated distribution is allowed to change over time. The IHMM can also be seen as a Bayesian nonparametric extension of the MS model as it generalizes the predefined finite number of states into an infinite number of states. Conditional returns distributions are approximated by mixing an infinite number of normal kernels and, for each period, the mixture weights depend on which state the previous period was in.

The IHMM has been applied in many economic and financial time-series since econometricians introduced it from computer science. Examples include Song (2014) and Jochmann (2015) on inflation rates, Maheu and Yang (2016) on short-term interest rates, Jin and Maheu (2016) and Jin et al. (2019) on realized covariance modelling, and Hou (2017) and Jin et al. (2021) on macroeconomic forecasting.

Although the IHMM can mimic smooth volatility transitions, empirically it would require many probably non-recurring states, which would greatly reduce the estimation efficiency. Dufays (2016) combines an IHMM with a univariate GARCH model to solve this problem, but a relevant multivariate model is still lacking.

In this paper, I propose a new multivariate Bayesian semiparametric model that combines the IHMM with a multivariate GARCH (MGARCH) model. The MGARCH component captures the strong dependence and smooth changes in the conditional second moments. The IHMM component not only nests a finite-state MS model and captures the short-run changes in the conditional distribution, but also nests the DPM model so that one does not need to assume the shape (tails, skewness, etc.) of the underlying data-generating process for the unknown returns distribution. Further, both the unknown conditional distribution and its short-term evolution are approximated nonparametrically through an infinite hidden Markov structure in addition to the GARCH dynamics. To compare these models' performance, this paper selects four benchmark models: an MGARCH model with normal innovations, an MGARCH model with asymmetric volatility feedback, an IHMM without GARCH effects and an MGARCH-DPM model. The proposed MGARCH-IHMM exhibits a clear advantage in monthly density forecasts over all of the benchmark models.

To test the model's economic performance in portfolio allocation, I follow Guidolin and Timmermann (2007), Pettenuzzo and Timmermann (2011) and Pettenuzzo and Ravazzolo (2016) and maximize an investor's expected utility by employing the full information acquired from predictive distributions when integrating out all of the parameters and the possible distributional uncertainties, in contrast to the traditional mean-variance optimization. Without assuming that the predictive returns distributions are of a known type, this method is more appropriate as it is sensitive to the shape of the predictive density of the returns. The optimized portfolios (models) are then evaluated based on an ex-post performance fee motivated by Fleming et al. (2001) and Bollerslev et al. (2018). The empirical results show that a risk-averse investor is always willing to pay at least 44 bps annually (and often more) to switch from any benchmark model/portfolio, including a buy-and-hold equally weighted one, to the MGARCH-IHMM when no transaction costs or trading restrictions are assumed. After considering the transaction costs, the proposed MGARCH-IHMM dominates all of the benchmarks when No Short-Selling or No Margin-Trading restriction is imposed. The investor is willing to pay at least 17 bps annually for switching to my new model. These results are robust to different risk-aversion levels.

This paper is organized as follows. Section 2 illustrates the specifications and the benefits of the proposed MGARCH-IHMM, along with the sampling and forecasting algorithms and the benchmark models. Section 3 details the returns series uses to test this model. Section 4 presents the posterior estimations of the MGARCH-IHMM and its out-of-sample forecasts. Section 5 compares the portfolio-allocation performance of different econometric models under different risk-aversion levels, transaction costs and trading restriction settings. Section 6 concludes.

2 MGARCH-IHMM Model

This paper proposes a new Bayesian semiparametric model with an MGARCH component and an infinite hidden Markov model (IHMM) component. Jensen and Maheu (2013) propose an MGARCH-DPM model that models the returns innovations nonparametrically through a static infinite mixture. The new model proposed in this paper extends this model by replacing the DPM component with an infinite hidden Markov structure that allows the mixture to change over time. The proposed MGARCH-IHMM is more flexible and nests the MGARCH-DPM model as a special case. Let $\boldsymbol{r}_t$ be an $N \times 1$ vector of returns, and $\boldsymbol{r}_{1:T} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_T\}$. Define $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots\}$ as the set of state-dependent parameters, where $\boldsymbol{\Theta}_j = \left\{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right\}$. The hierarchical representation of the model is

$$G_0|\beta_0, \Xi \sim DP\left(\beta_0, \Xi\right) \tag{1a}$$

$$G_j|\alpha_0, G_0 \sim DP\left(\alpha_0, G_0\right), \quad j = 1, 2, \ldots \tag{1b}$$

$$\boldsymbol{\Theta}_j|G_j \overset{\text{iid}}{\sim} G_j \tag{1c}$$

$$\boldsymbol{r}_t|s_t, \boldsymbol{\Theta}, \mathcal{F}_{t-1} \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2'}\right) \tag{1d}$$

$$\Xi = N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right) \times IW\left(\boldsymbol{\Sigma}_0, \nu + N\right), \quad \nu > 0 \tag{1e}$$

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1} \tag{1f}$$

where $\mathcal{F}_{t-1}$ is the information set for $t-1$, $DP(\beta_0, \Xi)$ is a Dirichlet process with concentration parameter β_0 and base measure Ξ, where the concentration parameter represents how dispersed the draw is away from the base measure, and Ξ contains a normal prior $N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right)$ for $\boldsymbol{\mu}_{s_t}$ and an inverse Wishart prior $IW\left(\boldsymbol{\Sigma}_0, \nu + N\right)$ for $\boldsymbol{\Sigma}_{s_t}$. The Bayesian nonparametric component IHMM and the parametric component MGARCH of this model are linked through (1d). s_t denotes the state/cluster for time t. s_t, $\boldsymbol{\mu}_{s_t}$ and $\boldsymbol{\Sigma}_{s_t}$ are determined by the IHMM component, and $\boldsymbol{H}_t$ is determined by the MGARCH component. $\boldsymbol{H}_t^{1/2}$ is the Cholesky decomposition of $\boldsymbol{H}_t$. $\boldsymbol{\Sigma}_{s_t}$ is parametrized around an identity matrix. The general level and long-run dynamics of conditional volatility are captured by the GARCH component, and $\boldsymbol{\Sigma}_{s_t}$ serves as an amplifier to either boost or shrink the conditional covariance from $\boldsymbol{H}_t$. Clearly, when $\boldsymbol{\mu}_{s_t} = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_{s_t} = \boldsymbol{I}$, the MGARCH-IHMM reduces to a parametric MGARCH model.

The IHMM component of this model is a Bayesian nonparametric one that employs a hierarchical Dirichlet process (HDP). (1a) and (1b) represents this HDP structure. There are an infinite number of states/clusters in the model, and the state-dependent parameter set $\boldsymbol{\Theta}_j$ is a random draw from their corresponding probability measure G_j. Accordingly, there are also an infinite number of G_js, and each one of them is drawn from a separate bottom layer Dirichlet process with precision parameter α_0 and base measure G_0. Further, G_0 is a draw from the top layer Dirichlet process with precision parameter β_0 and base measure Ξ. Ξ can be also seen as the prior of the state-dependent parameters $\boldsymbol{\Theta}$, a multivariate normal prior for $\boldsymbol{\mu}_{s_t}$ and an inverse Wishart prior for $\boldsymbol{\Sigma}_{s_t}$. The IHMM component is essentially an infinitely dimensional Markov-switching (MS) model. It is designed to capture the sudden changes in the conditional distribution through a regime-switching scheme. This regime-switching behaviour is shown more clearly later in the stick-breaking representation. Because this is a Bayesian nonparametric model, where one does not need to impose any distributional assumption to the conditional returns distributions, so it can capture features such as asymmetries and fat tails. Moreover, unlike the DPM, which is another Bayesian nonparametric model, but where the conditional distribution is static, the IHMM also nonparametrically approximates the unknown short-term evolution of the conditional distributions.

The MGARCH component takes a variant of the diagonal BEKK-GARCH representation (Engle and Kroner, 1995). $\boldsymbol{C}$ is an $N \times N$ lower triangular matrix, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are $N \times 1$ vectors, and $\odot$ is the Hadamard operator representing element-by-element multiplication. The parameter restriction of $\alpha_i^2 + \beta_i^2 < 1$ for all $i = 1, \ldots, N$ is imposed for stationarity in $\boldsymbol{H}_t$. Unlike traditional MGARCH models, where the source of the $\boldsymbol{H}_t$ variation is the returns shock in the last period (the difference between the last period returns and the conditional mean), the $\boldsymbol{H}_t$ dynamics in this model are determined by the change in the returns from an additional parameter $\boldsymbol{\eta}$. This allows the model to capture the potential asymmetric volatility feedback effect. When $\boldsymbol{\eta} > \boldsymbol{\mu}_{s_{t-1}}$, for a $\boldsymbol{r}_{t-1}$ less than $\boldsymbol{\mu}_{s_{t-1}}$, $\|\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\| > \|\boldsymbol{r}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}\|$, this would increase $\boldsymbol{H}_t$ as in an asymmetric dynamic covariance (ADC) model (Kroner and Ng, 1998). However, the MGARCH-IHMM does not enforce an asymmetric volatility feedback or the sign of this asymmetry but the feedback is instead learned from data. The posterior estimates of Fama-French 5 industry portfolios returns show that η_i is generally greater than $\mu_{i,s_{t-1}}$ for each asset i.

This model nests many models as special cases. For example, when $\alpha_0 \to 0$, $G_j \to \delta_x$, where δ_x is a Dirac measure centred at x and x is G_0 distributed. Then the MGARCH-

IHMM reduces to the MGARCH-DPM model from Jensen and Maheu (2013). Additionally, if $\beta_0 \to 0$, the MGARCH-IHMM reduces to a parametric MGARCH model with a normal innovation, and if $\beta_0 \to \infty$, the MGARCH-IHMM reduces to a parametric MGARCH model with Student-t innovation.

From a mixture perspective, this model can be rewritten as a stick-breaking representation (Sethuraman, 1994; Teh et al., 2006):

$$\mathbf{\Gamma}|\beta_0 \sim GEM\left(\beta_0\right) \tag{2a}$$

$$\mathbf{\Pi}_j|\alpha_0, \mathbf{\Gamma} \sim DP\left(\alpha_0, \mathbf{\Gamma}\right) \tag{2b}$$

$$s_t|s_{t-1}, \mathbf{\Pi} \sim Catagorical(\mathbf{\Pi}_{s_{t-1}}) \tag{2c}$$

$$p\left(\boldsymbol{r}_t|\mathbf{\Theta}, \mathbf{\Pi}, s_{t-1}\right) = \sum_{k=1}^{\infty} \pi_{s_{t-1}k} N\left(\boldsymbol{r}_t; \boldsymbol{\mu}_k, \boldsymbol{H}_t^{1/2} \mathbf{\Sigma}_k \boldsymbol{H}_t^{1/2'}\right) \tag{2d}$$

$$\boldsymbol{\mu}_k \sim N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right), \quad \mathbf{\Sigma}_k \sim IW\left(\mathbf{\Sigma}_0, \nu + N\right) \tag{2e}$$

where $\mathbf{\Gamma} = (\gamma_1, \gamma_2, \dots)'$, $\mathbf{\Pi}_j = (\pi_{j1}, \pi_{j2}, \dots)$, and $\boldsymbol{H}_t$ is defined as in (1f). To be more specific,

$$\gamma_k = \hat{\gamma}_k \prod_{l=1}^{k-1} \left(1 - \hat{\gamma}_l\right), \qquad \hat{\gamma}_k \sim Beta\left(1, \beta_0\right), \tag{3a}$$

$$\pi_{jk} = \hat{\pi}_{jk} \prod_{l=1}^{k-1} \left(1 - \hat{\pi}_{jl}\right), \qquad \hat{\pi}_{jk} \sim Beta\left(\alpha_0 \gamma_k, \alpha_0 \left(1 - \sum_{l=1}^{k} \gamma_l\right)\right). \tag{3b}$$

$GEM(\beta_0)$[1] is a general stick-breaking process with a precision parameter β_0. As shown in equation (2d), the conditional distribution of the returns is a mixture of an infinite number of Gaussian kernels with a vector of weights $\mathbf{\Pi}_j$. $\mathbf{\Pi}_j$ is the jth row of the infinitely dimensional squared transition matrix $\mathbf{\Pi}$ and a draw from a particular Dirichlet process. Because the IHMM is also a Bayesian nonparametric extension of the MS model, all of the states can recur with certain probabilities. To ensure this recurrence, another Dirichlet process is required to

[1]GEM stands for Griffiths, Engen, and McCloskey. See Pitman (2002) as an example.

share the atoms among all the bottom-layer Dirichlet processes. π_{jk} indicates the probability of switching from state j to state k. Note that in the MGARCH-DPM model, $\pi_{s_{t-1}k} = \pi_k$ for all $t = 1, \ldots, T$, so the MGARCH-DPM model is nested in the MGARCH-IHMM. This scenario will occur when $\alpha_0 \to 0$. Please note that (3b) is not applicable when α_0 approaches to 0, as it is derived from the formal definition of the Dirichlet process, where the precision parameter is strictly positive.

The state-dependent parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ allow us to identify potential regime switches. Furthermore, mixing over $\boldsymbol{\mu}_k$ also generates possible skewness in the conditional returns distributions, and mixing over $\boldsymbol{\Sigma}_k$ generates kurtosis. Since the mixture weights are Markovian, the unknown conditional distribution is allowed to change over time in an unknown pattern. In summary, the proposed MGARCH-IHMM retains all the advantages from both the MGARCH model and the IHMM. The new model is designed to capture the highly persistent long-run volatility dynamics and any potential drastic regime switches. Additionally, it approximates both the shape and the short-term evolution of the unknown conditional distributions nonparametrically.

2.1 Hierarchical Priors

Prior settings are very important in Bayesian inferences, especially when estimating Bayesian nonparametric models where the number of useful states is estimated jointly with other parameters. Each state-dependent parameter is estimated based on the subsamples assigned to its corresponding state. For states with fewer observations, the parameter posteriors can be heavily influenced by the priors, compared to states with more observations where the posteriors are dominated by the likelihoods. On the other hand, whenever a new state is introduced, the parameters are drawn directly from the priors. When the priors are far away from the support of the data, the introduction of the new state is easily wasted.

This problem can be successfully solved by introducing hierarchical priors, which are the priors of the base measure parameters. The base measure parameters are now estimated from the state-dependent parameters instead of being preset as constants. This allows the base measure to learn from the data so the state-dependent parameters, especially those of a new state, can always land within a reasonable region and then greatly improve both the in-sample fit and the out-of-sample forecasts.

Consider the following set of hierarchical priors motivated by Song (2014) and Maheu and Yang (2016):

$$\boldsymbol{b}_0 \sim N\left(\boldsymbol{h}_0, \boldsymbol{H}_0\right), \quad \boldsymbol{B}_0 \sim IW\left(\boldsymbol{A}_0, a_0\right), \quad \boldsymbol{\Sigma}_0 \sim W\left(\boldsymbol{C}_0, d_0\right), \quad \nu \sim Exp\left(g_0\right). \tag{4}$$

Then $\boldsymbol{b}_0$, $\boldsymbol{B}_0$, $\boldsymbol{\Sigma}_0$ and ν are drawn conditional on both the hierarchical priors and the corresponding parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$).

2.2 Covariance Targeting

In the MGARCH component, $\boldsymbol{C}$ has $N(N+1)/2$ parameters to estimate, while $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ all have N parameters, respectively. Obviously, the number of parameters grows quadratically in $\boldsymbol{C}$ and linearly in $\boldsymbol{\theta}_H = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}\}$. Hence, by targeting the symmetric $\boldsymbol{C}\boldsymbol{C}'$ matrix instead of estimating it, one can greatly reduce the total number of parameters in the estimation. Let $\bar{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{r}_t$ be the unconditional expectation and $\bar{\boldsymbol{H}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_t - \bar{\boldsymbol{\mu}})(\boldsymbol{r}_t - \bar{\boldsymbol{\mu}})'$ be the unconditional covariance matrix. In a reduced form of the MGARCH-IHMM, where $\boldsymbol{\Sigma}_t = \boldsymbol{I}$ for all t, referring to equation (1f), the unconditional expectation of $\boldsymbol{H}_t$ is

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{H}_t\right) &= \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \mathrm{E}\left[\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)'\right] + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \mathrm{E}\left(\boldsymbol{H}_{t-1}\right) \\
&= \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \mathrm{E}\left[\left(\boldsymbol{r}_{t-1} - \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)'\right] + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \mathrm{E}\left(\boldsymbol{H}_{t-1}\right) \\
&= \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \bar{\boldsymbol{H}} + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)\left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \mathrm{E}\left(\boldsymbol{H}_{t-1}\right).
\end{aligned}$$

Further assuming $\mathrm{E}(\boldsymbol{H}_t) = \bar{\boldsymbol{H}}$ for all $t = 1, \ldots, T$, we have

$$\boldsymbol{C}\boldsymbol{C}' = \bar{\boldsymbol{H}} \odot [\mathbf{1} - \boldsymbol{\alpha}\boldsymbol{\alpha}' - \boldsymbol{\beta}\boldsymbol{\beta}'] - \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\bar{\boldsymbol{\mu}} - \boldsymbol{\eta})(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta})', \tag{5}$$

where $\mathbf{1}$ is an $N \times N$ matrix with all the elements being 1. Note that any draw of $\boldsymbol{\theta}_H$ from the posterior that results in non-positive definite $\boldsymbol{C}\boldsymbol{C}'$ is rejected.

2.3 Sampling Algorithm

The MGARCH-IHMM is estimated through an MCMC algorithm. For the Bayesian non-parametric component (IHMM), I employ the beam sampler introduced by Van Gael et al. (2008) (see also Fox et al., 2011; Maheu and Yang, 2016). Similar to the slice sampler for the DPM model, the beam sampler partitions the infinite number of states in the IHMM into a finite set of "major" states with assigned observations and an additional "remaining" state where no observation is assigned.. Whenever a new state is introduced, the corresponding data density parameters are drawn directly from the hierarchical priors since no data have yet been assigned to that state. Let state R be the "remaining" state, then $\boldsymbol{\Gamma} = (\gamma_1, \ldots, \gamma_K, \gamma_R)'$ and $\boldsymbol{\Pi}_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jR})$, where $\gamma_R = \sum_{k=K+1}^{\infty} \gamma_k = 1 - \sum_{k=1}^{K} \gamma_k$ and $\pi_{jR} = \sum_{k=K+1}^{\infty} \pi_{jk} = 1 - \sum_{k=1}^{K} \pi_{jk}$. The sampling steps are as follows:

1. Sample the auxiliary slice variable $u_{1:T}|\boldsymbol{\Gamma}, \boldsymbol{\Pi}$.

2. Update K. If a new "major" state is introduced, then draw the corresponding parameters and the transition probabilities from the prior. The transition matrix now has an additional column and row.

3. Forward filter, backward sampler (FFBS) for the state variable $s_{1:T}|\boldsymbol{r}_{1:T}, u_{1:T}, \boldsymbol{\Gamma}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:T}$.

4. Simulate the coloured ball counts in the "oracle" urn[2] $c_{1:K}|s_{1:T}, \boldsymbol{\Gamma}, \alpha_0$.

[2]An "oracle" urn is a special urn in the hierarchical Pòlya urn scheme introduced by Beal et al. (2002). Please refer to Appendix A for details.

5. Sample β_0 and α_0 following Fox et al. (2011).

6. Sample $\boldsymbol{\Gamma}|c_{1:K}, \beta_0$.

7. Sample $\boldsymbol{\Pi}|n_{1:K,1:K}, \boldsymbol{\Gamma}, \alpha_0$.

8. Sample the state-dependent parameters $\boldsymbol{\Theta}|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{H}_{1:T}$.

9. Sample hierarchical priors.

   (a) Sample $\boldsymbol{b}_0|\boldsymbol{\mu}_{1:K}, \boldsymbol{B}_0, \boldsymbol{h}_0, \boldsymbol{H}_0$.

   (b) Sample $\boldsymbol{B}_0|\boldsymbol{\mu}_{1:K}, \boldsymbol{b}_0, a_0, \boldsymbol{A}_0$.

   (c) Sample $\nu|\sigma^2_{1:K}, s_0, g_0$.

   (d) Sample $\boldsymbol{\Sigma}_0|\boldsymbol{\Sigma}_{1:K}, v_0, \boldsymbol{C}_0, d_0$.

10. sample GARCH parameters $\boldsymbol{\theta}_H = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta})' \, |\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{\Theta}$. Apply a block-move random-walk Metropolis-Hastings algorithm to sample $\boldsymbol{\theta}_H$.

Repeat the above steps and discard the first burn-in samples. Details of all the sampling steps can be found in Appendix B. After simulating all of the M MCMC samples, the posterior mean of each parameter and latent variable can be computed by

$$\mathrm{E}(\theta|\boldsymbol{r}_{1:T}) \approx \frac{1}{M}\sum_{i=1}^{M}\theta^{(i)},$$

where $\theta^{(i)}$ is the ith MCMC draw of the given parameter θ.

2.4 Predictive Likelihood

I define the predictive likelihood for a particular out-of-sample period as $p\left(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}\right)$. This predictive likelihood is computed as follows:

1. Estimate the model for $\boldsymbol{r}_{1:t}$ and collect M posterior samples for $\left\{\boldsymbol{\Theta}^{(i)}, \boldsymbol{\theta}_H^{(i)}, \boldsymbol{\Pi}^{(i)}, s_{1:t}^{(i)}, K^{(i)}\right\}_{i=1}^{M}$ as described in Section 2.3.

2. Simulate the state indicator from equation (2c):

$$s_{t+1}^{(i)}|s_t^{(i)}, \mathbf{\Pi}^{(i)} \sim Catagorical(\mathbf{\Pi}_{s_t^{(i)}}^{(i)}).$$

3. If $s_{t+1}^{(i)} \leq K^{(i)}$, then the posterior of $\mathbf{\Theta}_{s_{t+1}^{(i)}}^{(i)}$ is already drawn. Otherwise state $s_{t+1}^{(i)}$ is inactive and without any assigned observation, so the posterior of $\mathbf{\Theta}_{s_{t+1}^{(i)}}^{(i)}$ is essentially the prior. Draw $\mathbf{\Theta}_{s_{t+1}^{(i)}}^{(i)}$ from the base measure $\boldsymbol{\mu}_{s_{t+1}^{(i)}}^{(i)} \sim N(\boldsymbol{b}_0^{(i)}, \boldsymbol{B}_0^{(i)})$ and $\mathbf{\Sigma}_{s_{t+1}^{(i)}}^{(i)} \sim IW(\mathbf{\Sigma}_0^{(i)}, \nu^{(i)} + N)$.

4. Propagate $\boldsymbol{H}_{t+1}^{(i)}$ from equation (1f):

$$\boldsymbol{H}_{t+1}^{(i)} = \boldsymbol{C}^{(i)}\boldsymbol{C}^{(i)'} + \boldsymbol{\alpha}^{(i)}\boldsymbol{\alpha}^{(i)'} \odot \left(\boldsymbol{r}_t - \boldsymbol{\eta}^{(i)}\right)\left(\boldsymbol{r}_t - \boldsymbol{\eta}^{(i)}\right)' + \boldsymbol{\beta}^{(i)}\boldsymbol{\beta}^{(i)'} \odot \boldsymbol{H}_t^{(i)}.$$

5. Evaluate the predictive likelihood for the realized return $\boldsymbol{r}_{t+1}$ conditional on every MCMC sample

$$p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \mathbf{\Theta}_{s_{t+1}^{(i)}}^{(i)}, \boldsymbol{\theta}_H^{(i)}) = N\left(\boldsymbol{r}_{t+1}\middle|\boldsymbol{\mu}_{s_{t+1}^{(i)}}^{(i)}, \boldsymbol{H}_{t+1}^{(i)1/2}\mathbf{\Sigma}_{s_{t+1}^{(i)}}^{(i)}\boldsymbol{H}_{t+1}^{(i)1/2'}\right),$$

where $N(\boldsymbol{x}|\boldsymbol{\mu}, \mathbf{\Sigma})$ is a multivariate normal density with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$ evaluated at $\boldsymbol{x}$.

6. Average out the conditional predictive likelihoods with respect to the MCMC draws

$$p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}) \approx \frac{1}{M}\sum_{i=1}^{M} p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \mathbf{\Theta}_{s_t^{(i)}}^{(i)}, \boldsymbol{\theta}_H^{(i)}).$$

The log-predictive likelihood over the whole out-of-sample period $t+1, \ldots, T$ for model A is

$$\log PL_A = \log p(\boldsymbol{r}_{t+1:T}|\boldsymbol{r}_{1:t}, \mathcal{M}_A) = \sum_{l=t}^{T-1} \log p(\boldsymbol{r}_{l+1}|\boldsymbol{r}_{1:l}, \mathcal{M}_A). \quad (6)$$

The log-predictive likelihood essentially evaluates the predictive density at the actual returns realization point. This can serve as a metric when comparing the out-of-sample performance among the different models. It also measures how likely the actual out-of-sample returns can be realized, given the predictive densities foretasted by the selected model. To compare, one can compute the difference of the log-predictive likelihoods between the two models, which is also called the log-Bayes factor:

$$\log BF_{AB} = \log PL_A - \log PL_B.$$

A log-Bayes factor greater than 5 is usually considered strong evidence that supports one model over the other. The log score differential (LSD) is another measure used to compare model density forecasts. The LSD is simply the ratio between the log-Bayes factor and the benchmark log-predictive likelihood.

$$LSD_{AB} = \frac{\log BF_{AB}}{\log PL_B}. \tag{7}$$

2.5 Benchmark Models

MGARCH-DPM A semiparametric multivariate GARCH model where the innovation is a constant mixture can be written as follows:

$$\begin{gathered}
\mathbf{\Gamma}|\beta_0 \sim GEM\left(\beta_0\right), \quad s_t|\mathbf{\Gamma} \sim catagorical(\mathbf{\Gamma}) \\
\boldsymbol{r}_t|\mathbf{\Theta}, \boldsymbol{H}_t, \mathbf{\Pi}, s_t \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{H}_t^{1/2}\mathbf{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2'}\right) \\
\boldsymbol{\mu}_s \sim N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right), \quad \mathbf{\Sigma}_s \sim IW\left(\boldsymbol{V}, \nu + N\right) \\
\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}.
\end{gathered}$$

This model replaces the IHMM component of the MGARCH-IHMM with a DPM component. The DPM model is a special case of the IHMM, where the mixture is static instead of

Markovian, so the MGARCH-DPM model is nested within the MGARCH-IHMM when $\alpha_0 \to 0$, as discussed in Section 2. $\boldsymbol{CC}'$ is targeted as in equation (5).

IHMM A fully nonparametric multivariate model with regime switching is specified as

$$\begin{aligned} \boldsymbol{\Gamma}|\beta_0 \sim GEM\,(\beta_0), &\quad \boldsymbol{\Pi}_j|\alpha_0,\boldsymbol{\Gamma} \sim DP\,(\alpha_0,\boldsymbol{\Gamma}) \\ s_t|s_{t-1},\boldsymbol{\Pi} \sim \boldsymbol{\Pi}_{s_{t-1}}, &\quad \boldsymbol{r}_t|s_t,\boldsymbol{\Theta},\mathcal{F}_{t-1} \sim N\left(\boldsymbol{\mu}_{s_t},\boldsymbol{\Sigma}_{s_t}\right) \\ \boldsymbol{\mu}_s \sim N\,(\boldsymbol{b}_0,\boldsymbol{B}_0), &\quad \boldsymbol{\Sigma}_s \sim IW\,(\boldsymbol{V},\nu+N) \\ \boldsymbol{b}_0 \sim N\,(\boldsymbol{h}_0,\boldsymbol{H}_0)\,, \quad \boldsymbol{B}_0 \sim IW\,(\boldsymbol{A}_0,a_0), &\quad \boldsymbol{\Sigma}_0 \sim W\,(\boldsymbol{C}_0,d_0)\,, \quad \nu \sim Exp\,(g_0)\,. \end{aligned}$$

This model is similar to the MGARCH-IHMM but without the MGARCH component. So this model is nested within the MGARCH-IHMM when $\boldsymbol{H}_t = \boldsymbol{I}$ for all t. The same set of hierarchical priors is also employed in this model.

MGARCH-N A fully parametric multivariate GARCH model with normal innovation is specified as

$$\begin{aligned} \boldsymbol{r}_t &= \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t, \quad \boldsymbol{z}_t \overset{\text{iid}}{\sim} N(\boldsymbol{0},\boldsymbol{I}) \\ \boldsymbol{H}_t &= \boldsymbol{CC}' + \boldsymbol{\alpha\alpha}' \odot (\boldsymbol{r}_{t-1}-\boldsymbol{\mu})\,(\boldsymbol{r}_{t-1}-\boldsymbol{\mu})' + \boldsymbol{\beta\beta}' \odot \boldsymbol{H}_{t-1}, \end{aligned}$$

where, through covariance targeting, $\boldsymbol{CC}'$ is defined as

$$\boldsymbol{CC}' = \bar{\boldsymbol{H}} \odot (\boldsymbol{1} - \boldsymbol{\alpha\alpha}' - \boldsymbol{\beta\beta}')\,.$$

MGARCH-A A fully parametric asymmetric multivariate GARCH model with normal innovation is specified as

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t, \quad \boldsymbol{z}_t \overset{\text{iid}}{\sim} N(\boldsymbol{0}, \boldsymbol{I})$$

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\boldsymbol{r}_{t-1} - \boldsymbol{\eta})(\boldsymbol{r}_{t-1} - \boldsymbol{\eta})' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1},$$

where $\boldsymbol{C}\boldsymbol{C}'$ is targeted as in equation (5). Unlike the MGARCH-N model, this model can also capture the potential asymmetry in the volatility feedback.

2.6 Hyper parameters

The hyper parameters for the priors, the hyper priors and the hierarchical priors are common across models and between posterior estimations and out-of-sample forecasts for comparability in the results. Following Fox et al. (2011), the hyper priors for the HDP concentration parameters β_0 and α_0 are assumed as

$$\beta_0 \sim Gamma(2, 8), \qquad \alpha_0 \sim Gamma(2, 8), \tag{8}$$

where $\mathrm{E}(\beta_0) = \mathrm{E}(\alpha_0) = 0.25$. During estimation, these hyper priors strongly favour less-active states for better state identification and faster computation speed. The hyper prior for the concentration parameter of the DPM is also assumed as $Gamma(2, 8)$. The hierarchical priors discussed in Section 2.1 are assumed as

$$\boldsymbol{b}_0 \sim N(\boldsymbol{0}, \boldsymbol{I}), \quad \boldsymbol{B}_0 \sim IW(\boldsymbol{I}, N+2), \quad \boldsymbol{\Sigma}_0 \sim W\left(\frac{\boldsymbol{I}}{N+2}, N+2\right), \quad \nu \sim Exp(2). \tag{9}$$

For the MGARCH parameters $\boldsymbol{\theta}_H = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta})'$, assume a truncated multivariate normal prior

$$\boldsymbol{\theta}_H \sim N(\boldsymbol{0}, \boldsymbol{I})\mathbb{1}\{\alpha_i > 0, \beta_i > 0, \alpha_i^2 + \beta_i^2 < 1, \text{ for all } i\}, \tag{10}$$

where $\alpha_i > 0$ and $\beta_i > 0$ are imposed for parameter identification and $\alpha_i^2 + \beta_i^2 < 1$ is imposed for the stationary MGARCH process.

3 Data

The data used in this paper include monthly returns of the Fama-French 5 industry portfolios, consisting of Consumer, Manufacture, High Tech, Health and Other portfolios, and the 1-month Treasury bill rate from the website of Kenneth French. All of the returns range from July 1926 to December 2020 at a monthly frequencies, for a total of 1,134 observations. Furthermore, all of the returns are converted into log-returns for continuous compounding and scaled by 100 for percentage values.

Table 1 is approximately here.

Panel A of Table 1 illustrates some univariate descriptive statistics of the five industry portfolio log-returns. The expected returns are from around 0.71% to 0.93%, and the standard deviation from around 5.27% to 6.34%. All of the industries are negatively skewed and leptokurtic. Panel B shows that all five industries are highly but not perfectly correlated, so the diversification benefit still exists. Moreover, as discussed later in Section 5, if an investor's utility is not quadratic, then the diversification benefits are not solely determined by these industries' correlations.[3]

4 Model Performance

4.1 In-Sample Estimation

In all three models, the first 20,000 iterations were discarded as burn-ins, and 20,000 MCMC samples were collected for posterior inference. This section summarizes the results of the

[3]Also see the simulation experiment in Jondeau and Rockinger (2006) where the sensitivities to skewness and kurtosis are tested for portfolio allocations.

full sample estimates, consisting of 1,106 observations.

Table 2 is approximately here.

Panel A of Table 2 lists the posterior estimates of the non-state-dependent parameters for the five models, and the assets in $\boldsymbol{r}_t$ are indexed in the same order as in Table 1. The MGARCH-IHMM, and the MGARCH-DPM, the MGARCH-N and the MGARCH-A models all have high $\boldsymbol{\beta}$s that are above 0.92, and low $\boldsymbol{\alpha}$s that are below 0.31 in general. As mentioned in Section 2, $\boldsymbol{\eta}$ helps to capture the asymmetric volatility feedback. Comparing the $\boldsymbol{\eta}$ estimates and the posterior means of $\boldsymbol{\mu}$ over time in Panel B, η_is are in general greater than the corresponding μ_i. This shows that the Fama-French 5 industry portfolio returns usually exhibit negative feedback to their volatility (a return that is lower than its expectation leads to higher volatility next period). This is also the case for the MGARCH-A model, where η_is are greater than the corresponding μ_i except for those in the third portfolio (High Tech). The MGARCH persistence ($\alpha_i^2 + \beta_i^2$) in the MGARCH-IHMM are almost the same as those in the MGARCH-DPM, the MGARCH-N and the MGARCH-A models, although each of the α_is and β_is are a little different. However, the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ estimates are almost identical between the MGARCH-N and the MGARCH-A, so the difference in the MGARCH-IHMM is unlikely to be due to the asymmetric volatility feedback but rather to the Bayesian nonparametric IHMM component. On the other hand, the 0.95 density intervals for the GARCH parameters are slightly wider in the MGARCH-IHMM compared to those in the MGARCH-N and the MGARCH-A models. For example, the 0.95 density interval for β_1 is $(0.9385, 0.9556)$ in the MGARCH-IHMM, $(0.9391, 0.9510)$ in the MGARCH-N model and $(0.9404, 0.9527)$ in the MGARCH-A model. The posterior for the GARCH parameters in the MGARCH-IHMM are generally more dispersed, probably due to the complexity of the model.

Comparing the semiparametric MGARCH-IHMM and the nonparametric IHMM, the MGARCH-IHMM has higher concentration parameters ($\alpha_0 = 1.6190$ and $\beta_0 = 0.9140$, on average) than the IHMM ($\alpha_0 = 0.9919$ and $\beta_0 = 0.7723$, on average) and a greater number of active states ($K = 9.7826$, compared to $K = 7.8118$, on average). The MGARCH-DPM model has the lowest concentration parameter ($\beta_0 = 0.3872$) and the least number of active states ($K = 4.8338$).

Comparing the posterior means of the state-dependent mean parameter $\boldsymbol{\mu}_{s_t}$ over time between the MGARCH-IHMM and the MGARCH-DPM model, both models share similar medians and means, but the unconditional distribution of the posterior means of $\boldsymbol{\mu}_{s_t}$ are more dispersed for the MGARCH-IHMM than for the MGARCH-DPM model. This shows that the MGARCH-IHMM allows for much richer dynamics for the state-dependent parameters than the MGARCH-DPM model does for its Markovian mixtures.

Figure 1 is approximately here.

In addition to the GARCH dynamics, the Fama-French 5 industry portfolios still exhibit clear regime-switching behaviour, implying that both long-run smooth changes and short-run regime switches are important. Figure 1 plots the heat map generated by the MGARCH-IHMM. This heat map shows the empirical probability that two periods share the same state. The redder the colour, the higher the probability of sharing states. There are two dominating states: a "bull" state with high expected returns and low state-dependent volatility, and a "bear" state with low expected returns and high state-dependent volatility. Most periods fall within the "bull" state, but there are several exceptions. The "bear" state is shared by multiple eras, including but not limited to the Great Depression, from March 1928 to September 1933; the Stagflation, from September 1973 to March 1975; the Savings and Loan Crisis, from November 1989 to October 1990; the Dotcom Bubble, from May 1998 to September 2001; the 2008 Financial Crisis, from November 2007 to April 2009; the US-China Trade War, from October 2018 to May 2019; and the recent Coronavirus Crash, from January to August 2020. Note that these are states that have been identified *in addition to* the GARCH effect, indicating that the regular MGARCH-N model is insufficient especially during crises. The rest of the active states approximate the tails and shapes of the conditional returns distributions.

Figure 2 is approximately here.

Figure 2 plots the posterior means of some of the time-varying parameters, with the previously mentioned major "bear" periods shaded in red. To be more specific, the top plot shows the average over five industry portfolios of the posterior means for the state-dependent mean parameter $\boldsymbol{\mu}_{s_t}$, and the other three plots show the log determinants of the posterior means for the time-varying second-moment parameters. As discussed earlier, $\boldsymbol{\Sigma}_{s_t}$ is parameterized around an identity matrix whose log determinant is 0. If $\boldsymbol{\Sigma}_{s_t}$ equals to identity, the conditional covariance dynamics of the MGARCH-IHMM reduces to those of the MGARCH-A model. From the graph, the "bear" state (red) is accompanied by low expected returns and high state-dependent volatility (log determinant greater than 0), which represents a slowdown in the growth of investors' wealth and high investment uncertainty, and the "bull" state is the opposite. Because the state switching captures the dynamics in addition to the GARCH recursion, if the MGARCH component is sufficient for the conditional covariance dynamics, then the state-dependent covariance $\boldsymbol{\Sigma}_t$ will be an identity matrix, which clearly is not the case here. Therefore, the "bear" state, where the log determinant of the state-dependent variance is considerably greater than zero (the horizontal dashed line in the second graph), compensates for the extra volatility that the MGARCH component fails to capture, and the "bull" state, where the state-dependent variance is lower than one, causes the overall conditional variance of the returns to shrink.. As for the $\boldsymbol{H}_t$ dynamics, the "bear" state usually occurs when the log determinant of $\boldsymbol{H}_t$ is increasing, and the "bull" state usually occurs when the log determinant of $\boldsymbol{H}_t$ is decreasing. This implies that these states are identified because $\boldsymbol{H}_t$ is not climbing or declining fast enough, and an additional multiplier $\boldsymbol{\Sigma}_t$ is required to boost the conditional covariances.

The last graph of Figure 2 plots the log determinant of the posterior means of the overall conditional variances that are estimated from each model, namely $\mathrm{E}\left(\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_k\boldsymbol{H}_t^{1/2'}\right)$ for the MGARCH-IHMM and MGARCH-DPM, $\mathrm{E}\left(\boldsymbol{\Sigma}_{s_t}\right)$ for the IHMM, and $\mathrm{E}\left(\boldsymbol{H}_t\right)$ for the MGARCH-N and MGARCH-A. The conditional variance that is estimated from the

MGARCH-IHMM is the most-flexible one among all of the selected models in terms of both its strong persistent and rough volatility dynamics.[4] For the IHMM, the log determinant of the covariances tries to mimic the smooth volatility trend but still changes drastically and switches among the different levels, while for the MGARCH-N and MGARCH-A models, the paths of the log determinants are approximately the same. They can change gradually, but the range of the change is confined and it cannot capture the roughness in the volatility. Both the MGARCH-IHMM and the MGARCH-DPM models allow for highly persistent and rough volatility dynamics, but the volatility path of the MGARCH-IHMM appears to be rougher than that of the MGARCH-DPM model.

4.2 Out-of-Sample Forecasts

Table 3 is approximately here.

In Bayesian econometrics, one usually compares forecasts from different models by using the log predictive likelihood, the log Bayes factor or the log score differential (LSD). The empirical results show that the MGARCH-IHMM outperforms the benchmark MGARCH-DPM, IHMM, MGARCH-N and MGARCH-A models in terms of the density forecasts. A recursive prediction is performed for 360 out-of-sample periods from January 1991 to December 2020 by re-estimating each model for each period. Table 3 compares the performance of these recursive forecasts among the four models. The difference in the log-predictive likelihoods between two models represents a log Bayes factor, and a number greater than 6 is usually considered strong evidence that one model predicts better than the other. The MGARCH-IHMM produces a log-predictive likelihood of -4478.602, which is 13.4731 higher than the log-predictive likelihood of the MGARCH-DPM (-4492.075), 72.2869 higher than that of the IHMM (-4550.889), 42.9239 higher than that of the MGARCH-A (-4521.526), and 44.8084 higher than that of the MGARCH-N (-4523.410). The average log score LSDs

[4]For the roughness of volatility and its importance, see Andersen et al. (2007); Bayer et al. (2016); Gatheral et al. (2018); Glasserman and He (2020); Livieri et al. (2018); Shi et al. (2021).

give the same model rankings. The proposed MGARCH-IHMM provides a prediction that is 0.30% better than the MGARCH-DPM, 1.59% better than the IHMM, 0.95% better than the MGARCH-A and 0.99% better than the MGARCH-N. However, the MGARCH-IHMM does not show an advantage in terms of the point forecast. All five models yield similar root mean squared forecast errors (RMSFE)[5], with 4.1480 for the MGARCH-IHMM, 4.1487 for the MGARCH-DPM model, 4.1340 for the IHMM, 4.1493 for the MGARCH-N model and 4.1508 for the MGARCH-A model.

5 Utility-Based Portfolio Optimization

Clearly, the proposed MGARCH-IHMM is more flexible and can provide better density forecasts than all of the benchmark models, but does it translate into actual economic gains in portfolio allocation? As discussed at the beginning of this paper, in most research, a portfolio is optimized on a mean-variance space, which means that a risk-averse investor wants to maximize the expected returns while minimizing the variance of her portfolio. Here, the "risk" is defined as the variance or standard deviation of the asset returns, and it ignores any higher moments or the whole returns distribution. This implies either that the investor's utility function is quadratic or approximately quadratic with respect to her wealth, or that the asset returns are multivariate elliptically distributed.

On one hand, stock returns are better approximated by a mixture of normals, as discussed in Section 4.1, so ellipticity is not guaranteed. Moreover, after being converted into simple returns, all of the predictive distributions will be more skewed than those of the log returns. On the other hand, a quadratic utility is usually considered as unrealistic due to its increasing absolute risk-aversion. The constant relative risk-aversion (CRRA) utility, whose absolute risk-aversion is decreasing when wealth increases, is relatively more realistic. To incorporate the CRRA utility along with general returns distributions, we need to consider a more-general

[5]The forecast error is computed from the average of the difference between the predicted mean of $\boldsymbol{r}_t$ and the actual realized returns across five industry portfolios.

portfolio-optimization problem.

5.1 Dynamic Optimal Portfolio Weights

Consider a rational, risk-averse investor whose utility function is $U(W)$, where W denotes their wealth. She distributes her wealth into N risky assets and one risk-free asset.[6] Without loss of generality, assume her wealth set as 1 at the beginning of each period and her wealth at the end of each period t is

$$W_t = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t + (1 - \boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t} - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}), \tag{11}$$

where $\boldsymbol{w}_t$ represents the vector of the portfolio weights of the risky assets, $\boldsymbol{R}_t$ is a vector of the simple returns of the risky assets,[7] $\boldsymbol{\iota}$ is a vector of 1, and $R_{f,t}$ is the simple return of the risk-free asset. $C(\cdot)$ is the transaction cost incurred when rebalancing at the end of period t, which is a function of $\boldsymbol{w}_t$.

Suppose the investor rebalances her portfolio given the information set $\mathcal{F}_{t-1}$. For a particular period t, she maximizes her conditional expected utility:

$$\begin{aligned} &\max_{\boldsymbol{w}_t} \mathrm{E}\left[U(W_t)|\mathcal{F}_{t-1}\right] \qquad (12)\\ &\approx \frac{1}{M}\sum_{m=1}^{M} U(W_t^{(m)}),\\ \text{where} \qquad & W_t^{(m)} = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t^{(m)} + (1 - \boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t} - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}). \end{aligned}$$

The superscript (m) corresponds to the mth draw of the simulated predictive returns for time t. No constraint on $\boldsymbol{w}_t$ is required since the rest of the wealth is distributed to the risk-free asset. The No short-sale constraint ($w_{i,t} \geq 0$ for each asset i) or no leverage trading constraint ($w_{i,t} \geq 0$ for each asset i and $\sum_{i=1}^{N} w_{i,t} \leq 1$) can also be applied.

[6] $N = 5$. Consumer, Manufacture, High Tech, Health and Other portfolio are from the Fama-French 5 industry portfolios.

[7] For each element in $\boldsymbol{R}_t$, $R_{i,t} = \exp(r_{i,t}/100) - 1$.

At the end of period t, the investor is subject to transaction fee $C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1})$ on every transactions she made to rebalances her portfolio.

$$C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}) = c \sum_{i=1}^{N} |w_{i,t} - w_{i,t-1}| + c \left| \sum_{i=1}^{N} w_{i,t} - \sum_{i=1}^{N} w_{i,t-1} \right|. \tag{13}$$

where c denotes the flexible fee as a certain percentage of the wealth. The first term represents the portfolio turnovers due to the balance change in each asset in the risky portfolio. The second term represents the amount of the risk-free asset the investor needs to trade when rebalancing.

The analytical solution for this problem is in general unavailable, so it needs to be found numerically. Note that if the utility function $U(W_t^{(i)})$ is strictly concave with respect to $W_t^{(i)}$, then it is also strictly concave with respect to $\boldsymbol{w}_t$, and so is its empirical expectation. This ensures the existence of a unique solution to the above optimization problem.

5.2 Break-even Management Fees

Because the notion of "risk" is not measured by the second moment of returns distribution but by the uncertainty of the returns, the Sharpe ratio, which measures risk by its standard deviation, may not be a desirable criteria to compare portfolio-allocation performance. Instead, this paper implements a more-direct method, by computing the break-even management fee an investor is willing to pay for switching from one model to another (e.g., Fleming et al., 2001; Bollerslev et al., 2018) to compare the economic performance of different portfolios.

At the end of period $t-1$, the investor rebalances her portfolio to the optimal weights $\boldsymbol{w}_t$ found above, and carries this portfolio into the next period when the actual returns $\boldsymbol{R}_t$ are realized. One can easily compute her realized wealth W_t and the realized utility $U(W_t)$ for each out-of-sample period. The break-even management fee the investor is willing to pay,

Δ, for switching from model $\mathcal{M}_2$ to model $\mathcal{M}_1$ is given by equating the ex-post utilities as

$$\sum_{l=t+1}^{T} U(W_l - \Delta | \mathcal{M}_1) = \sum_{l=t+1}^{T} U(W_l | \mathcal{M}_2). \tag{14}$$

Note that the investor's initial wealth is always 1 for each period, so Δ can be seen either as a dollar fee or a returns fee.

5.3 Empirical Results

I optimize the portfolio under a constant relative risk-averse (CRRA) utility. The utility function is specified as

$$U(W) = \frac{W^{1-a}}{1-a},$$

where a is the risk-aversion parameter, which indicates the relative risk aversion as a and the absolute risk aversion as $\frac{a}{W}$.

The portfolio is optimized when $a = \{2, 4, 6\}$ for all five econometric models with transaction costs $c = \{0\%, 1\%, 2\%\}$, respectively. A simplex method is used for each optimization, with 200 different initial values being used to avoid the local optimum problem. The Brent-Dekker method is used to find the scalar root in equation (14). In addition to the benchmark models listed in Section 2.5, I consider two equally weighted portfolios. The first mixes a risky, equally weighted portfolio with risk-free asset (EW + RF). The second is a buy-and-hold strategy of an equally weighted portfolio that consists of the Fama-French 5 industry portfolios and a risk-free asset (EW), where each component has a weight of 1/6.

Table 4 is approximately here.

Table 4 shows the annualized fee that an investor is willing to pay for switching from one econometric model to another over the out-of-sample period from January 1991 to December 2020. When no trading restriction is imposed, the investor is always willing to pay a positive fee to switch from all of the benchmark models to the proposed MGARCH-IHMM model when there are no transaction costs. Among all of the benchmark models/portfolios, the Bayesian semiparametric model MGARCH-DPM is the most competitive, but the investor is still willing to pay for an annualized fee from 0.44% to 0.82%, depending on the relative risk-aversion level when switching to the MGARCH-IHMM model. This shows that the distributional change over time, other than volatility dynamics, still matters economically. The investor is willing to pay less for switching from the two parametric GARCH models than for switching from the Bayesian nonparametric IHMM model, a finding that is consistent with the density forecast results. This result emphasizes the importance of smooth changes in the second moment for stock returns, both statistically and economically. When transaction costs are not involved with unlimited margin trading, a risk-averse investor always prefers a more-sophisticated MGARCH-IHMM than a less-sophisticated benchmark.

After considering the transaction costs but still having no trading restrictions, the buy-and-hold strategy for an equally weighted portfolio (EW) becomes very competitive. Since the weights are constant over time, there are no transaction costs. The advantage is even greater when the transaction costs are high (2%). This is because when there are no transaction costs or trading restrictions, the optimal weights from the MGARCH-IHMM change over time to such as extent as to achieve the highest possible expected utility, as shown in Figure 3a. After introducing the transaction costs, the change in the optimal weights over time is penalized. The paths of positions over time are significantly stabilized in Figure 3b and 3c, but still change too much (both in longs and shorts), which causes greater transaction costs than the benefits gained from the better density forecasts.

Figure 3 is approximately here.

Table 5 lists the average positions over time for the risky portfolio that is optimized using different models. When there are no trading restrictions (Panel A), all of the econometric models suggest taking large margin-buy and short-sale positions. After considering the transaction costs, the MGARCH-IHMM still uses the largest margin-buy and short-sale positions among all of the models. The net risky positions are relatively low since it is offset by the large long and short positions.

In practice, short-selling can be very expensive or even not available for some investors. When short-selling is not allowed for any risky asset ($w_{i,t} \geq 0$ for all $i = 1, \ldots, N$), the investor is always willing to pay for a positive annualized fee to switch from any benchmark model to my new model (from 0.13% to 5.97%, depending on the model, the risk-aversion parameters and transaction cost settings), regardless of the risk-averse parameter or the transaction costs. Among all the benchmark models, when there are no transaction costs, the Bayesian nonparametric and semiparametric models are worse than the most-flexible MGARCH-IHMM (a fee no more than 0.81% for switching to the MGARCH-IHMM) but much better than the parametric MGARCH models and equally weighted portfolios (a fee no less than 0.98%). This suggests that the economic value of a smooth volatility structure is better realized by short-selling in the out-of-sample period. After considering the transaction costs, the buy-and-hold equally weighted portfolio (EW) becomes one of the best models after the MGARCH-IHMM, especially when the investor is more risk-averse. When $a = 4$ or 6, an investor is willing to pay no more than 0.65% annually for switching from the EW to the MGARCH-IHMM optimized portfolio, while she is always willing to pay at least 10 bps more for switching from any other benchmark model to the proposed MGARCH-IHMM, given the same risk-averse parameter and transaction costs. Compared to the scenario without trading restrictions, the net risky positions are approximately the same when short-selling is not allowed (Table 5), but the range of the weight change is greatly reduced since the entire space of the negative weights is excluded. Although the solution found in the optimization problem may be sub-optimal to the solution without trading restrictions, the realized results are better than those for all of the benchmarks.

Under the "No Short-Selling" restriction, the optimal portfolio calls for active margin-buying. From the bottom-left plot in each of Figures 3a, 3b and 3c, the investor will take margin-buying positions most of the time, and the leverage can be high during some periods, for example, in the beginning of 1991 when $a = 2$. This strategy can be risky when transaction costs are involved. When an investor holds a margin-buying position with high leverage, it is much harder to liquidate her position in time if the market is frictional.

As a result, the "No Margin-Trading" restriction is further considered, where $w_{i,t} \geq 0$ for all $i = 1, \ldots, N$ and $\sum_{i=1}^{N} w_{i,t} \leq 1$. Similar to the scenario without short-selling, the MGARCH-IHMM model dominates all the benchmarks, regardless of risk aversion levels and transaction cost settings, with an annualized fee between 0.07% and 2.72% (Table 4). The other Bayesian semiparametric model MGARCH-DPM, along with the Bayesian non-parametric model IHMM and the buy-and-hold equally weighted portfolio (EW), are the second-best candidates among all of the models. For the risky holdings of the MGARCH-IHMM portfolio, it shows that the investor, on average, should hold the risky portfolio in as a greater portion as possible for the best outcome (Figure 3). When there are no transaction costs, the investor can adjust her portfolio freely, and it is optimal to hold the risky part of the portfolio, without having the risk-free asset except when there is a bear market, namely during the Dotcom Bubble, the 2008 Financial Crisis, the US-China Trade War or the recent Coronavirus Crash. When transaction costs are relevant, instead of pursuing market timing, it is better not to react so drastically but to maintain risky holdings that are relatively stable. But this does not necessarily advocate a buy-and-hold strategy because the relative weights for each individual risky asset can still change even when the net risky holding is stable.

6 Conclusions

This paper investigates the relationship between accurate returns density forecasts and whether they lead to economic gains in portfolio choice. The results show that under a relatively realistic CRRA utility, positive economic gains can be generated from a more-sophisticated model that produces the highest out-of-sample log-predictive likelihood.

A new multivariate Bayesian semiparametric model (MGARCH-IHMM) is proposed in this paper. The additional IHMM component successfully captures the regime-switching phenomenon around GARCH dynamics. Multiple occurrences of the "bear" state are identified apart from the "bull" state. The rest of the estimated states are there to approximate the unknown shape of the conditional distributions. The state-dependent covariance parameter serves as a multiplier to the GARCH covariance, helping to boost the covariance when it is not increasing fast enough and to shrink it when it is not decreasing fast enough. Compared to the parametric models, the Bayesian semiparametric models, especially the MGARCH-IHMM, can capture the roughness in the volatility dynamics well.

The MGARCH-IHMM shows a clear advantage against the benchmark MGARCH-DPM, IHMM, MGARCH-N and MGARCH-A models in terms of density forecast. The MGARCH-IHMM allows for distributional change, such as skewness, kurtosis and tails, in addition to smooth volatility changes and the approximated unknown conditional distributions captured by the MGARCH-DPM. By adding the IHMM component, the MGARCH-IHMM is capable of predicting skewed and leptokurtic returns distributions, compared to the MGARCH-N and MGARCH-A models. Compared to the IHMM, the MGARCH-IHMM is able to predict smooth changes in the volatility dynamics.

Since the commonly used Markowitz Portfolio Theory implies that the investor only cares about the first two moments of the conditional returns distribution, the rest of the information from the returns distribution is discarded. Therefore, a more-general portfolio-optimization problem, which maximizes the expected utility, is implemented instead.

Empirical results show that, when there are no transaction costs, a risk-averse investor is always willing to pay a positive fee for switching to the more-sophisticated Bayesian semiparametric model MGARCH-IHMM with better density forecasts, regardless the risk-aversion level. However, the MGARCH-IHMM optimized portfolio involves heavy short-selling and margin-buying, both of which may result in high transaction costs. After considering the transaction costs, the benefits of frequently rebalancing portfolios to their optimum are negated by the associated transaction costs since the optimal weights change too much in the absense of trading restrictions.

Margin trading, including margin-buying and short-selling, can be expensive or not accessible for some investors. When No Short-Selling or No Margin-Trading constraints are imposed, the MGARCH-IHMM dominates all the benchmarks, including a buy-and-hold equally weighted portfolio that has zero transaction costs, regardless of the risk-aversion level or the transaction cost settings.

References

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, pages 1152–1174.

Bayer, C., Friz, P., and Gatheral, J. (2016). Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904.

Beal, M. J., Ghahramani, Z., and Rasmussen, C. E. (2002). The infinite hidden Markov model. In *Advances in Neural Information Processing Systems*, pages 577–584.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics*, pages 498–505.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207(1):71–91.

Bonato, M. (2012). Modeling fat tails in stock returns: a multivariate stable-GARCH approach. *Computational Statistics*, 27(3):499–521.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of econometrics*, 86(2):221–241.

Dufays, A. (2016). Infinite-state Markov-switching for dynamic volatility. *Journal of Financial Econometrics*, 14(2):418–460.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, pages 1020–1056.

Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.

Glasserman, P. and He, P. (2020). Buy rough, sell smooth. *Quantitative Finance*, 20(3):363–378.

Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *The Review of Financial Studies*, 21(2):889–935.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, pages 357–384.

Hou, C. (2017). Infinite hidden Markov switching VARs with application to macroeconomic forecast. *International Journal of Forecasting*, 33(4):1025–1043.

Jensen, M. J. and Maheu, J. M. (2013). Bayesian semiparametric multivariate GARCH modeling. *Journal of Econometrics*, 176(1):3–17.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics*, 192(1):19–39.

Jin, X., Maheu, J. M., and Yang, Q. (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *Journal of Applied Econometrics*.

Jin, X., Maheu, J. M., and Yang, Q. (2021). Infinite Markov pooling of predictive distributions. *Journal of Econometrics*, forthcoming.

Jochmann, M. (2015). Modeling U.S. inflation dynamics: A Bayesian nonparametric approach. *Econometric Reviews*, 34(5):537–558.

Jondeau, E. and Rockinger, M. (2006). Optimal portfolio allocation under higher moments. *European Financial Management*, 12(1):29–55.

Kawakatsu, H. (2006). Matrix exponential GARCH. *Journal of Econometrics*, 134(1):95–128.

Kroner, K. F. and Ng, V. K. (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies*, 11(4):817–844.

Livieri, G., Mouti, S., Pallavicini, A., and Rosenbaum, M. (2018). Rough volatility: evidence from option prices. *IISE transactions*, 50(9):767–776.

Maheu, J. M. and McCurdy, T. H. (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics*, 18(1):100–112.

Maheu, J. M. and Shamsi Zamenjani, A. (2021). Nonparametric dynamic conditional beta. *Journal of Financial Econometrics*, 19(4):583–613.

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

Pástor, L. and Stambaugh, R. F. (2001). The equity premium and structural breaks. *The Journal of Finance*, 56(4):1207–1239.

Peng, C. and Kim, Y. S. (2020). Portfolio optimization on multivariate regime switching GARCH model with normal tempered stable innovation. *arXiv preprint arXiv:2009.11367*.

Pesaran, B. and Pesaran, M. H. (2010). Conditional volatility and correlations of weekly returns and the var analysis of 2008 stock market crash. *Economic Modelling*, 27(6):1398–1416.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73(4):1057–1084.

Pettenuzzo, D. and Ravazzolo, F. (2016). Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics*, 31(7):1312–1332.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78.

Pitman, J. (2002). Poisson-Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. *Combinatorics, Probability and Computing*, 11(5):501–514.

Rydén, T., Teräsvirta, T., and Åsbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13(3):217–244.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, pages 639–650.

Shi, S., Liu, X., and Yu, J. (2021). Fractional stochastic volatility model. *Available at SSRN 3724733*.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics*, 29(5):825–842.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3):351–362.

Van Gael, J., Saatci, Y., Teh, Y. W., and Ghahramani, Z. (2008). Beam sampling for the infinite hidden Markov model. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1088–1095. ACM.

Table 1: Descriptive Statistics of the Industry Portfolio Returns

Panel A: Univariate statistics

Industry	Mean	Median	StDev	Skewness	Ex.Kurtosis	Min	Max
Consumer	0.8818	1.2324	5.2714	-0.5883	6.7366	-33.6592	36.2349
Manufacture	0.7977	1.2669	5.5026	-0.4705	7.4474	-36.9037	36.1374
High Tech	0.8373	1.2472	5.5998	-0.6536	3.9009	-31.1702	29.1475
Health	0.9314	1.0989	5.5460	-0.6361	7.4813	-41.6728	31.5759
Other	0.7118	1.2768	6.3401	-0.2871	8.1140	-35.7104	46.2160

Panel B: Correlations

	Consumer	Manufacture	High Tech	Health	Other
Consumer	1.0000	0.8749	0.8175	0.7828	0.8832
Manufacture	0.8749	1.0000	0.8101	0.7448	0.8935
High Tech	0.8175	0.8101	1.0000	0.7098	0.8031
Health	0.7828	0.7448	0.7098	1.0000	0.7416
Other	0.8832	0.8935	0.8031	0.7416	1.0000

[1.] Source: Kenneth French's Data Library.

[2.] From July 1926 to December 2020, 1,134 observations.

Table 2: Posterior Estimates

MGARCH-IHMM:

$$\mathbf{\Gamma}|\beta_0 \sim GEM\left(\beta_0\right), \quad \mathbf{\Pi}_j|\alpha_0, \mathbf{\Gamma} \sim DP\left(\alpha_0, \mathbf{\Gamma}\right)$$
$$s_t|s_{t-1}, \mathbf{\Pi} \sim catagorical(\mathbf{\Pi}_{s_{t-1}})$$
$$\boldsymbol{r}_t|\mathbf{\Theta}, \boldsymbol{H}_t, \mathbf{\Pi}, s_t \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{H}_t^{1/2}\mathbf{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2'}\right)$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

MGARCH-DPM:

$$\mathbf{\Gamma}|\beta_0 \sim GEM\left(\beta_0\right), \quad s_t|\mathbf{\Gamma} \sim catagorical(\mathbf{\Gamma})$$
$$\boldsymbol{r}_t|\mathbf{\Theta}, \boldsymbol{H}_t, \mathbf{\Pi}, s_t \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{H}_t^{1/2}\mathbf{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2'}\right)$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

IHMM:

$$\mathbf{\Gamma}|\beta_0 \sim GEM\left(\beta_0\right), \quad \mathbf{\Pi}_j|\alpha_0, \mathbf{\Gamma} \sim DP\left(\alpha_0, \mathbf{\Gamma}\right)$$
$$s_t|s_{t-1}, \mathbf{\Pi} \sim catagorical(\mathbf{\Pi}_{s_{t-1}})$$
$$\boldsymbol{r}_t|s_t, \mathbf{\Theta}, \mathcal{F}_{t-1} \sim N\left(\boldsymbol{\mu}_{s_t}, \mathbf{\Sigma}_{s_t}\right)$$

MGARCH-N:

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\boldsymbol{r}_{t-1} - \boldsymbol{\mu}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\mu}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

MGARCH-A:

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

Panel A: non-state-dependent parameters

	MGARCH-IHMM		MGARCH-DPM		IHMM		MGARCH-N		MGARCH-A	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
μ_1							0.7586	(0.5143,1.0043)	0.7944	(0.5488,1.0373)
μ_2							0.7141	(0.4703,0.9591)	0.7443	(0.5034,0.9872)
μ_3							0.7869	(0.5403,1.0327)	0.7851	(0.5359,1.0305)
μ_4							0.8317	(0.5620,1.0937)	0.8368	(0.5651,1.1050)
μ_5							0.5930	(0.3095,0.8787)	0.6652	(0.3803,0.9486)
α_1	0.2404	(0.2278, 0.2525)	0.2513	(0.2426,0.2604)			0.2644	(0.2541,0.2732)	0.2575	(0.2466,0.2678)
α_2	0.2554	(0.2373, 0.2691)	0.2621	(0.2521,0.2723)			0.2738	(0.2624,0.2882)	0.2820	(0.2681,0.2944)
α_3	0.2589	(0.2443, 0.2696)	0.2891	(0.2798,0.3000)			0.3013	(0.2787,0.3208)	0.3014	(0.2894,0.3168)
α_4	0.2524	(0.2328, 0.2747)	0.2743	(0.2510,0.3069)			0.2807	(0.2594,0.2999)	0.2643	(0.2491,0.2869)
α_5	0.2556	(0.2425, 0.2698)	0.2618	(0.2484,0.2749)			0.2896	(0.2730,0.3049)	0.2897	(0.2768,0.3040)
β_1	0.9477	(0.9385, 0.9556)	0.9422	(0.9327,0.9514)			0.9452	(0.9391,0.9510)	0.9471	(0.9404,0.9527)
β_2	0.9485	(0.9402, 0.9573)	0.9446	(0.9364,0.9516)			0.9455	(0.9384,0.9517)	0.9419	(0.9344,0.9489)
β_3	0.9425	(0.9334, 0.9512)	0.9328	(0.9224,0.9411)			0.9328	(0.9222,0.9443)	0.9319	(0.9229,0.9395)
β_4	0.9361	(0.9188, 0.9490)	0.9249	(0.9049,0.9428)			0.9336	(0.9208,0.9450)	0.9390	(0.9276,0.9483)
β_5	0.9470	(0.9396, 0.9534)	0.9426	(0.9337,0.9503)			0.9359	(0.9266,0.9454)	0.9348	(0.9253,0.9426)
η_1	2.1068	(1.5072, 2.6198)	2.3519	(1.6730,2.8626)					1.3033	(0.6774,1.7108)
η_2	1.7347	(1.2468, 2.2381)	1.9518	(1.2454,2.5057)					1.0098	(0.4078,1.3683)
η_3	1.3616	(0.8971, 1.9095)	1.5638	(0.9139,2.1334)					0.6037	(0.0804,1.0255)
η_4	1.8997	(1.2746, 2.5574)	1.9541	(1.2088,2.5770)					0.9182	(0.2964,1.4279)
η_5	2.3398	(1.6549, 3.0103)	2.6564	(1.7991,3.3445)					1.5565	(0.9303,2.0199)
α_0	1.6190	(0.9509, 2.5345)			0.9919	(0.5945, 1.4988)				
β_0	0.9140	(0.3807, 1.6428)	0.3872	(0.1115,0.8235)	0.7723	(0.3303, 1.4078)				
K	9.7826	(6.0000,13.0000)	4.8338	(3.0000,8.0000)	7.8118	(7.0000,10.0000)				

Table 2: Posterior Estimates (cont.)

Panel B: Summary of Posterior means of $\boldsymbol{\mu}_{s_t}$ over time

MGARCH-IHMM	Min	25%	Median	Mean	75%	Max
μ_1	-1.5289	0.4779	1.1288	0.9243	1.4756	1.6909
μ_2	-0.2949	0.6587	1.0097	0.8995	1.1963	1.3202
μ_3	-0.1667	0.7207	1.0791	0.9552	1.2517	1.3592
μ_4	0.2566	0.7789	1.1245	1.0241	1.3033	1.4447
μ_5	-0.9827	0.2400	1.0149	0.8024	1.4083	1.6510

MGARCH-DPM	Min	25%	Median	Mean	75%	Max
μ_1	0.4415	0.9007	1.0628	0.9761	1.1222	1.1634
μ_2	0.5529	0.8484	0.9694	0.9109	1.0128	1.0417
μ_3	0.3734	0.9071	1.0098	0.9546	1.0468	1.0713
μ_4	0.4632	0.9841	1.0797	1.0294	1.1143	1.1355
μ_5	0.2759	0.7731	0.9417	0.8559	1.0018	1.0406

Table 3: Log-Predictive Likelihoods and Log-Bayes Factors

Model	log PL	log BF	LSD	RMSFE
MGARCH-IHMM	**-4478.602**	—	—	4.1480
MGARCH-DPM	-4492.075	13.4731	0.30%	4.1487
IHMM	-4550.889	72.2869	1.59%	**4.1340**
MGARCH-A	-4521.526	42.9239	0.95%	4.1493
MGARCH-N	-4523.410	44.8084	0.99%	4.1508

Training samples are from July 1926 to December 1990, and out-of-sample periods are from January 1991 to December 2020.

Table 4: Annualized Fees a Risk-Averse Investor Is Willing to Pay

$$U(W) = \frac{W^{1-a}}{1-a},$$
$$W_t = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t + (1 - \boldsymbol{w}_t'\boldsymbol{\iota}) - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}),$$
$$C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}) = c\sum_{i=1}^{N}|w_{i,t} - w_{i,t-1}| + c\left|\sum_{i=1}^{N} w_{i,t} - \sum_{i=1}^{N} w_{i,t-1}\right|.$$

a	$a=2$			$a=4$			$a=6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Panel A: No Trading Restriction									
MGARCH-DPM	0.82%	-0.96%	-4.17%	0.55%	0.47%	-1.00%	0.44%	0.71%	-0.67%
IHMM	6.39%	1.16%	-5.08%	3.43%	-0.21%	-0.34%	2.31%	0.85%	-0.01%
MGARCH-A	2.20%	-2.41%	-6.96%	1.31%	0.84%	-0.42%	0.91%	1.22%	-0.13%
MGARCH-N	1.37%	-2.45%	-6.14%	0.88%	0.28%	-0.89%	0.62%	0.92%	-0.37%
EW+RF	3.55%	-0.14%	-5.71%	1.94%	1.74%	0.27%	1.32%	-0.28%	-0.11%
EW	5.41%	-0.81%	-6.30%	1.04%	-0.47%	-1.91%	0.53%	-0.00%	-1.36%
Panel B: No Short-Selling									
MGARCH-DPM	0.81%	2.51%	2.20%	0.78%	0.99%	1.50%	0.65%	0.80%	1.31%
IHMM	0.35%	1.51%	1.20%	0.13%	0.66%	0.75%	0.25%	0.59%	0.87%
MGARCH-A	3.17%	3.09%	3.55%	1.85%	1.82%	2.47%	1.36%	1.84%	1.99%
MGARCH-N	1.93%	2.42%	2.67%	1.26%	2.03%	2.48%	0.98%	1.81%	2.08%
EW+RF	4.09%	4.11%	3.75%	2.24%	3.03%	2.73%	1.62%	2.10%	2.28%
EW	5.97%	3.42%	3.06%	1.34%	0.22%	0.65%	0.83%	0.40%	0.59%
Panel C: No Margin-Trading									
MGARCH-DPM	0.07%	0.33%	0.47%	0.82%	1.11%	1.18%	0.63%	0.78%	1.16%
IHMM	1.44%	0.51%	0.93%	1.14%	0.95%	0.39%	0.71%	0.68%	0.77%
MGARCH-A	0.20%	0.77%	0.91%	1.35%	2.09%	2.47%	1.03%	1.45%	1.68%
MGARCH-N	0.15%	0.38%	0.90%	1.52%	1.80%	2.19%	1.22%	1.60%	1.92%
EW+RF	0.11%	1.51%	1.56%	1.88%	2.61%	2.72%	1.72%	1.86%	2.13%
EW	0.50%	0.83%	0.88%	0.98%	0.44%	0.36%	0.92%	0.17%	0.43%

1. No Trading Restriction indicates no restriction on $\boldsymbol{w}_t$ during expected utility optimization; No Short-Selling indicates a constraint of $w_{i,t} \geq 0$ for all $i = 1, \ldots, N$ is imposed in the optimization problem; and No Margin-Trading indicates that in addition to the No Short-Selling constraint, $\sum_{i=1}^{N} w_{i,t} \leq 1$ is also imposed in the optimization problem.

2. IHMM means an investor switches from IHMM to MGARCH-IHMM.

3. A positive fee means the MGARCH-IHMM is better, and a negative fee means the corresponding benchmark model is better.

4. EW+RF indicates mixing an equally weighted portfolio with the risk-free asset, assuming iid; and EW indicates an equally weighted portfolio of the Fama-French 5 industry portfolios and a risk-free asset (each has a weight of 1/6).

Table 5: Risky Positions Averaged over Time

a	$a=2$			$a=4$			$a=6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Panel A: No Trading Restriction									
Long:									
MGARCH-IHMM	4.3408	2.6729	2.4037	2.4173	1.8978	1.7783	1.6499	1.2949	1.3901
MGARCH-DPM	3.6612	1.7090	1.7372	2.0483	1.1633	1.2402	1.3988	0.7789	0.8577
IHMM	4.6610	1.7763	1.9958	2.5441	1.2783	1.1557	1.7265	0.9052	0.9494
MGARCH-A	4.0941	1.5820	1.4847	2.0573	0.8012	0.7603	1.3722	0.5027	0.5099
MGARCH-N	4.5350	2.0060	1.7757	2.2802	1.0503	1.0330	1.5210	0.6702	0.6655
EW+RF	1.1331	0.7300	0.7299	0.5797	0.3676	0.3676	0.3888	0.2455	0.2455
EW	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333
Short:									
MGARCH-IHMM	-1.8967	-1.6987	-1.7893	-1.0116	-0.9702	-0.8749	-0.6838	-0.6489	-0.6507
MGARCH-DPM	-1.3671	-0.5988	-0.6841	-0.7511	-0.3395	-0.4043	-0.5114	-0.2381	-0.2570
IHMM	-2.3146	-0.5282	-0.7440	-1.2233	-0.4435	-0.3216	-0.8243	-0.3134	-0.2841
MGARCH-A	-1.8056	-0.5583	-0.5356	-0.9072	-0.3075	-0.3065	-0.6051	-0.1997	-0.2079
MGARCH-N	-2.2006	-0.8794	-0.8665	-1.1065	-0.4734	-0.4852	-0.7381	-0.3224	-0.3186
EW+RF	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EW	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Total:									
MGARCH-IHMM	2.4441	0.9742	0.6143	1.4057	0.9276	0.9034	0.9662	0.6460	0.7394
MGARCH-DPM	2.2941	1.1102	1.0530	1.2972	0.8239	0.8359	0.8874	0.5408	0.6007
IHMM	2.3464	1.2482	1.2518	1.3208	0.8348	0.8341	0.9022	0.5918	0.6653
MGARCH-A	2.2885	1.0237	0.9492	1.1500	0.4936	0.4538	0.7671	0.3030	0.3020
MGARCH-N	2.3344	1.1266	0.9092	1.1737	0.5769	0.5478	0.7829	0.3478	0.3469
EW+RF	1.1331	0.7300	0.7299	0.5797	0.3676	0.3676	0.3888	0.2455	0.2455
EW	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333
Panel B: No Short-Selling									
MGARCH-IHMM	2.4021	1.4434	1.4614	1.3703	0.9411	1.0755	0.9393	0.6369	0.7638
MGARCH-DPM	2.1723	1.0154	1.0345	1.2188	0.6417	0.6139	0.8315	0.4214	0.4222
IHMM	2.2498	1.0489	1.0619	1.2500	0.6243	0.7111	0.8520	0.4100	0.4672
MGARCH-A	2.0384	0.9041	0.7852	1.0229	0.4592	0.4285	0.6824	0.2866	0.2996
MGARCH-N	2.0194	0.9512	0.9077	1.0138	0.4326	0.4354	0.6759	0.2882	0.2799
EW+RF	1.1331	0.7300	0.7299	0.5797	0.3676	0.3676	0.3888	0.2455	0.2455
EW	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333
Panel C: No Margin-Trading									
MGARCH-IHMM	0.9730	0.9866	0.9895	0.8883	0.8795	0.9564	0.7784	0.6551	0.7665
MGARCH-DPM	0.9979	0.9655	0.9801	0.9567	0.5774	0.5791	0.7886	0.4215	0.4056
IHMM	0.9687	0.9805	0.9795	0.8992	0.6446	0.7145	0.7625	0.4210	0.4650
MGARCH-A	0.9950	0.9436	0.7840	0.8597	0.4078	0.4060	0.6617	0.2969	0.3258
MGARCH-N	0.0022	0.8713	0.7904	0.8603	0.4328	0.4327	0.6603	0.3029	0.2949
EW+RF	0.9344	0.7300	0.7299	0.5797	0.3676	0.3676	0.3888	0.2456	0.2455
EW	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333

1. The table shows the positions of the risky portfolio optimized with each model and averaged over time. To be more specific, Total = $\mathrm{E}(\sum_{i=1}^{N} w_i)$, Long = $\mathrm{E}(\sum_{i=1}^{N} w_i \mathbb{1}_{\{w_i>0\}})$, and Short = $\mathrm{E}(\sum_{i=1}^{N} w_i \mathbb{1}_{\{w_i<0\}})$. When No Short-Selling or No Margin-Trading is imposed, the short position is always 0 and Long = Total.
2. The meaning of No Trading Restriction, No Short-Selling and No Margin-Trading are the same as in Table 4.
3. The meaning of EW+RF and EW are the same as in Table 4.

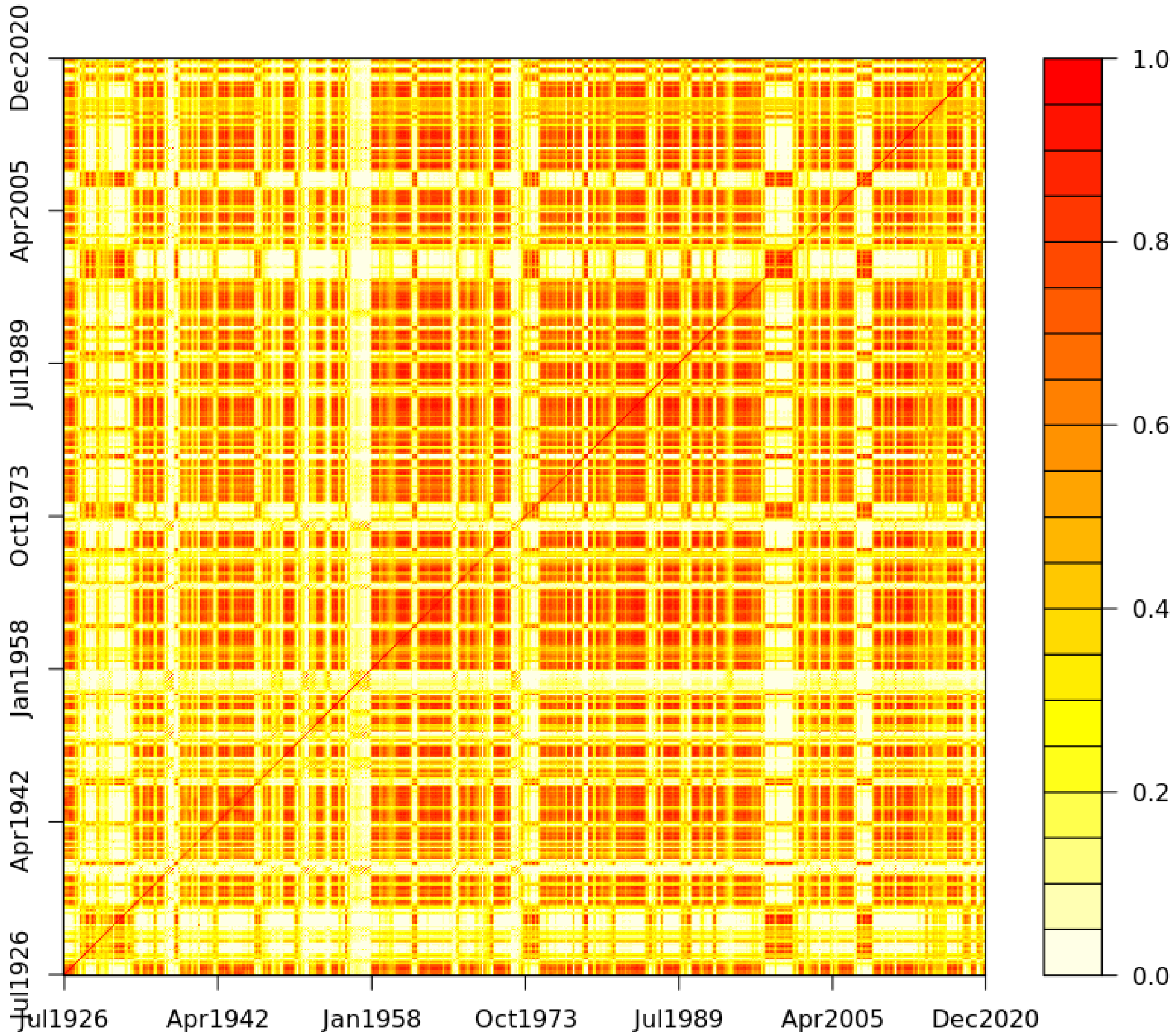
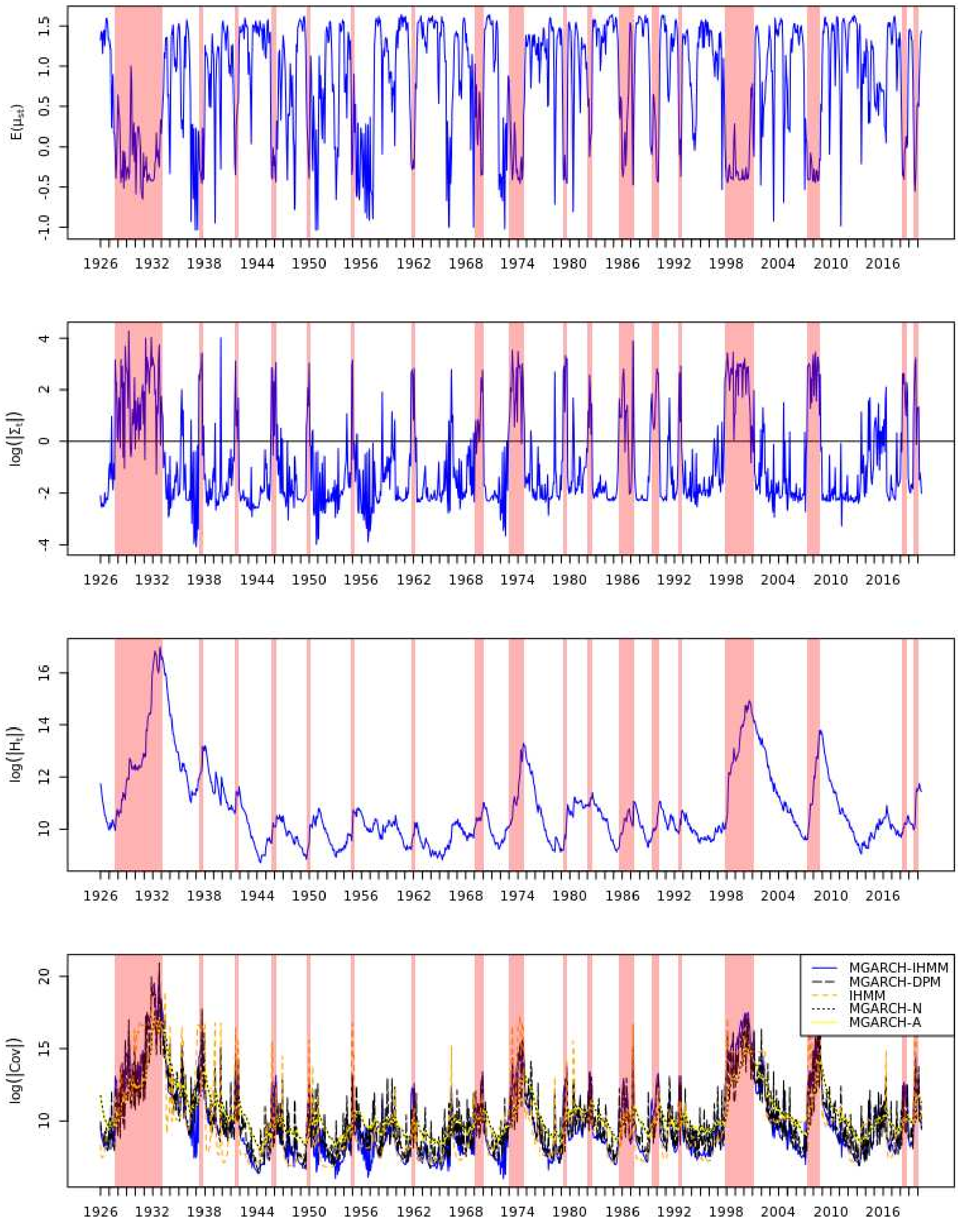
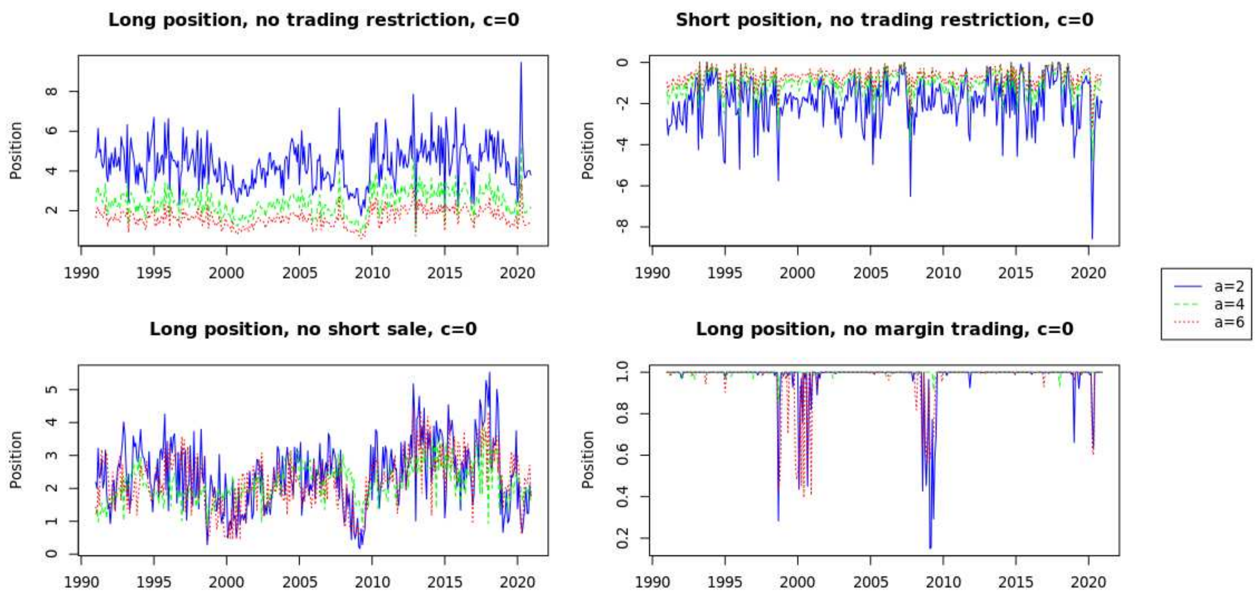


Figure 1: Heat Map of States Estimated by the MGARCH-IHMM

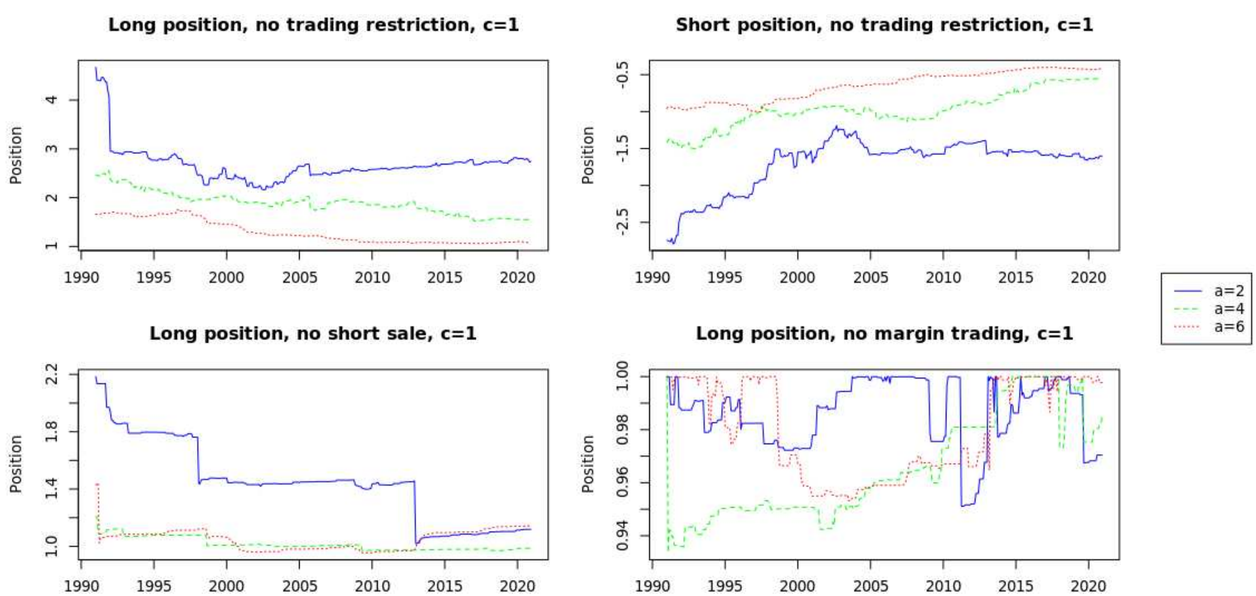


MGARCH-IHMM, MGARCH-DPM: $\mathrm{Cov} = \mathrm{E}(\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_k\boldsymbol{H}_t^{1/2'})$,
IHMM: $\mathrm{Cov} = \mathrm{E}(\boldsymbol{\Sigma}_{s_t})$, MGARCH-N, MGARCH-A: $\mathrm{Cov} = \mathrm{E}(\boldsymbol{H}_t)$

Figure 2: (Log Determinants of) the Posterior Means of the Time-Varying Parameters over Time

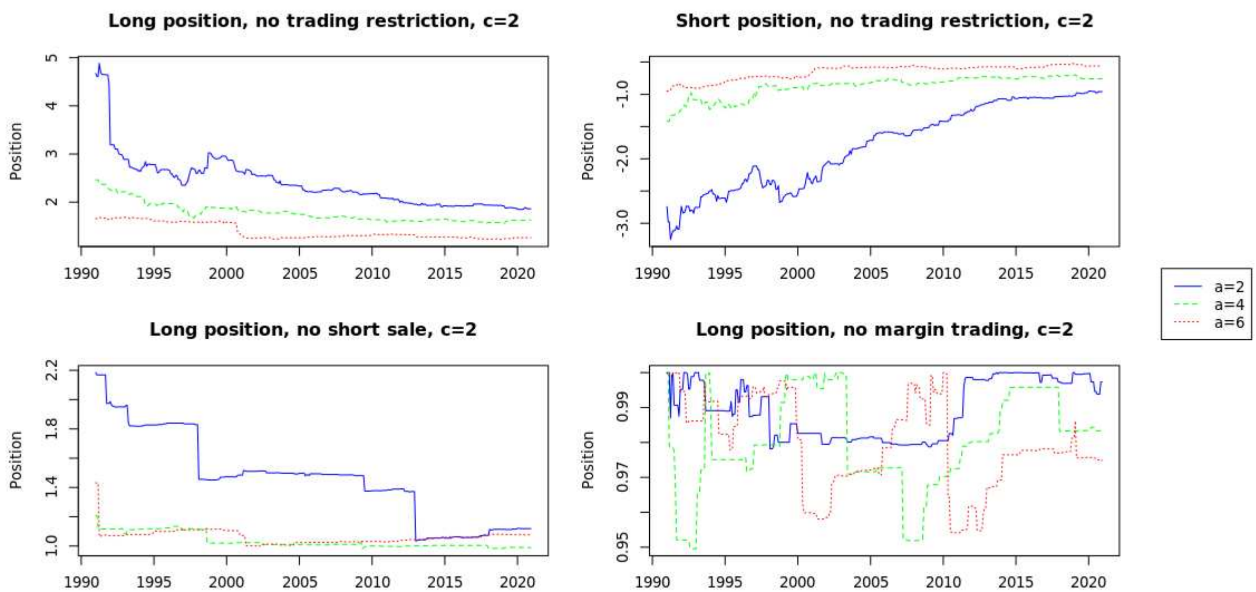


(a) $c = 0\%$



(b) $c = 1\%$

Figure 3: Risky Positions in the Portfolio Optimized with MGARCH-IHMM over Time



(c) $c = 2\%$

Figure 3: Risky Positions in the Portfolio Optimized with MGARCH-IHMM over Time (cont.)

Appendix A Pólya Urn Scheme for HDP

To estimate the IHMM, a hierarchical Pólya urn scheme introduced by Beal et al. (2002) is more convenient. This representation consists of a set of regular urns with an additional "oracle" urn. Suppose there are currently K existing "major" states, then firstly draw from the regular urn as follows:

$$p\left(\tilde{s}=s|s_t=j,s_{1:t-1}\right)=\sum_{k=1}^{K}\frac{n_{jk}}{\alpha_0+\sum_{j=1}^{K}n_{jk}}\delta\left(k,s\right)+\frac{\alpha_0}{\alpha_0+\sum_{j=1}^{K}n_{jk}}\delta\left(K+1,s\right) \tag{15}$$

where n_{jk} counts the number of balls in colour k from urn j; $\delta\left(a,b\right)$ is the Kronecker delta function that equals 1 if $a=b$, otherwise 0. If $\tilde{s}\leqslant K$, then $s_{t+1}=\tilde{s}$; if a new state is drawn, an additional sampling step from the "oracle" urn is involved:

$$p\left(s_{t+1}=s|s_t=j,s_{1:t-1}\right)=\sum_{k=1}^{K}\frac{c_k}{\beta_0+\sum_{j=1}^{K}c_k}\delta\left(k,s\right)+\frac{\beta_0}{\beta_0+\sum_{j=1}^{K}c_k}\delta\left(K+1,s\right) \tag{16}$$

where c_k counts the number of balls in colour k from the "oracle" urn. If a new state is also drawn in the "oracle" urn, the number of existing "major" states will become $K+1$.

Appendix B Sampling Details

Recall that $\mathbf{\Gamma}=\left(\gamma_1,\ldots,\gamma_K,\gamma_R\right)'$ and $\mathbf{\Pi}_j=\left(\pi_{j1},\ldots,\pi_{jK},\pi_{jR}\right)$, where $\gamma_R=\sum_{k=K+1}^{\infty}\gamma_k=1-\sum_{k=1}^{K}\gamma_k$ and $\pi_{jR}=\sum_{k=K+1}^{\infty}\pi_{jk}=1-\sum_{k=1}^{K}\pi_{jk}$. The sampling steps are:

1. Sample $u_{1:T}|\mathbf{\Gamma},\mathbf{\Pi}$. The auxiliary slice variable $U=\left\{u_t\right\}_{t=1}^{T}$ is drawn by $u_1\sim U\left(0,\gamma_{s_1}\right)$ and $u_t\sim U\left(0,\pi_{s_{t-1}s_t}\right)$.

2. Update K. Similar to the DPM model, if K does not meet the condition

$$\min\left\{u_t\right\}_{t=1}^{T}>\max\left\{\pi_{jR}\right\}_{j=1}^{K} \tag{17}$$

then K needs to be increased by 1 ($K' = K+1$) and all of the corresponding parameters need to be drawn from the base measure. In addition, since a new "major" state is introduced, $\mathbf{\Gamma}$ and $\mathbf{\Pi}$ also need to be updated accordingly:

(a) $\mathbf{\Theta}_{K'} \sim H$;

(b) Draw $v \sim Beta\,(1, \beta_0)$, then update $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_K, \gamma_{K'}, \gamma_R)'$, where $\gamma_{K'} = v\gamma_R$ and $\gamma_R = (1-v)\,\gamma_R$;

(c) Draw $v_j \sim Beta\,(\alpha_0\gamma_{K'}, \alpha_0\gamma_R)$, update $\mathbf{\Pi}_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jK'}, \pi_{jR})$ for $j = 1, \ldots, K$, where $\pi_{jK'} = v\pi_{jR}$ and $\pi_{jR} = (1-v)\,\pi_{jR}$;

(d) Draw the K'th row of $\mathbf{\Pi}$, $\mathbf{\Pi}_{K'}$, by $\mathbf{\Pi}_{K'} \sim Dir\,(\alpha_0\gamma_1, \ldots, \alpha_0\gamma_K, \alpha_0\gamma_{K'}, \alpha_0\gamma_R)$.

Repeat the above steps until inequality (17) holds.

3. The forward filter for $s_{1:T}|\boldsymbol{r}_{1:T}, u_{1:T}, \mathbf{\Gamma}, \mathbf{\Pi}, \mathbf{\Theta}, \boldsymbol{H}_{1:T}$. Iterating the following steps forward from 1 to T:

(a) The prediction step for initial state s_1 is as follows:

$$p(s_1 = k|u_1, \mathbf{\Gamma}) \propto \mathbb{1}\,(u_1 < \gamma_k)\,, \qquad k = 1, \ldots, K \tag{18}$$

for the following states $s_{2:T}$:

$$p(s_t = k|\boldsymbol{r}_{1:t-1}, u_{1:t}, \mathbf{\Pi}, \mathbf{\Theta}, \boldsymbol{H}_{1:t-1}) \propto \sum_{j=1}^{K} \mathbb{1}\,(u_t < \pi_{jk})\, p\,(s_{t-1} = j|\boldsymbol{r}_{1:t-1}, u_{1:t-1}, \mathbf{\Pi}, \mathbf{\Theta}, \boldsymbol{H}_{1:t-1}) \tag{19}$$

(b) The update step for $s_{1:T}$:

$$p\,(s_t = k|\boldsymbol{r}_{1:t}, u_{1:t}, \mathbf{\Pi}, \mathbf{\Theta}, \boldsymbol{H}_{1:t}) \propto p\,(\boldsymbol{r}_t|\boldsymbol{r}_{t-1}, \mathbf{\Theta}_k, \boldsymbol{H}_t)\, p\,(s_t = k|\boldsymbol{r}_{1:t-1}, u_{1:t}, \mathbf{\Pi}, \mathbf{\Theta}, \boldsymbol{H}_{1:t-1}) \tag{20}$$

4. The backward sampler for $s_{1:T}|\boldsymbol{r}_{1:T}, u_{1:T}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:T}$. Sample the states $s_{1:T}$ using the previously filtered values backward from T to 1:

   (a) for the terminal state s_T directly from $p\left(s_T|\boldsymbol{r}_{1:T}, u_{1:T}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:T}\right)$

   (b) for the rest states,

$$p\left(s_t = k|s_{t+1} = j, \boldsymbol{r}_{1:t}, u_{1:t+1}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \propto \mathbb{1}\left(u_{t+1} < \pi_{kj}\right) p\left(s_t = k|\boldsymbol{r}_{1:t}, u_{1:t}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \tag{21}$$

5. Sample $c_{1:K}|s_{1:T}, \boldsymbol{\Gamma}, \alpha_0$. $c_{1:K}$ is essential for sampling $\boldsymbol{\Gamma}$, and it counts balls in different colours in the "oracle" urn. However, recall that the Pòlya urn scheme in section A, the "oracle" urn is only involved when a new state is drawn from the regular urn, and it would be difficult to directly sample c_k. Fox et al. (2011) propose simulating c_k from the original Pòlya urn scheme instead of sampling it.

   (a) Count the number of each transition type, n_{jk}, for the number of times switching from state j to state k.

   (b) Simulate an auxiliary trail variable $x_i \sim Bernoulli\left(\frac{\alpha_0\gamma_k}{i-1+\alpha_0\gamma_k}\right)$, for $i = 1, \ldots, n_{jk}$. If the trial is successful, an "oracle" urn step is involved at the ith step toward n_{jk} and we increase the corresponding "oracle" counts, o_{jk}, by one.

   (c) $c_k = \sum_{j=1}^{K} o_{jk}$.

6. Sample β_0. Following Fox et al. (2011); Maheu and Yang (2016), assume a Gamma prior $\beta_0 \sim Gamma\left(a_1, b_1\right)$, and let $c = \sum_{j=1}^{K} c_j$,

   (a) $\nu \sim Bernoulli\left(\frac{c}{c+\beta_0}\right)$

   (b) $\lambda \sim Beta\left(\beta_0 + 1, c\right)$

   (c) $\beta_0 \sim Gamma\left(a_1 + K - \nu, b_1 - \log\lambda\right)$

7. Sample α_0. Following Fox et al. (2011), assume a Gamma prior $\alpha_0 \sim Gamma\left(a_2, b_2\right)$, and let $n_j = \sum_{k=1}^{K} n_{jk}$,

(a) $\nu_j \sim Bernoulli\left(\frac{n_j}{n_j+\alpha_0}\right)$

(b) $\lambda_j \sim Beta\left(\alpha_0+1, n_j\right)$

(c) $\alpha_0 \sim Gamma\left(a_2+c-\sum_{j=1}^{K}\nu_j, b_2-\sum_{j=1}^{K}\log\left(\lambda_j\right)\right)$

8. Sample $\boldsymbol{\Gamma}|c_{1:K},\beta_0$. Given the "oracle" urn counts $c_{1:K}$ and the property of Dirichlet process, the conjugate posterior is

$$\boldsymbol{\Gamma}|c_{1:K},\beta_0 \sim Dir\left(c_1,\ldots,c_K,\beta_0\right) \tag{22}$$

9. Sample $\boldsymbol{\Pi}|n_{1:K,1:K},\boldsymbol{\Gamma},\alpha_0$. Similarly, the conjugate posterior of $\boldsymbol{\Pi}_j$ is

$$\boldsymbol{\Pi}_j|n_{j,1:K},\boldsymbol{\Gamma},\alpha \sim Dir\left(\alpha_0\gamma_1+n_{j1},\ldots,\alpha_0\gamma_K+n_{jK},\alpha_0\gamma_R\right) \tag{23}$$

10. Sample $\boldsymbol{\Theta}|\boldsymbol{r}_{1:T},s_{1:T},\boldsymbol{H}_{1:T}$. Assume conjugate priors $\boldsymbol{\mu} \sim N\left(\boldsymbol{b}_0,\boldsymbol{B}_0\right)$ and $\boldsymbol{\Sigma} \sim IW\left(\boldsymbol{\Sigma}_0,\nu+N\right)$. Define $\boldsymbol{Y}_k \equiv \left\{\boldsymbol{H}_t^{-1/2}\boldsymbol{r}_t|s_t=k\right\}_{t=2}^{T}$ and $\boldsymbol{X}_k \equiv \left\{\boldsymbol{H}_t^{-1/2}|s_t=k\right\}_{t=2}^{T}$. The linear model is now

$$\boldsymbol{Y}_k = \boldsymbol{X}_k\boldsymbol{\mu}_k+\boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim N\left(\mathbf{0},\boldsymbol{\Sigma}_k\right) \tag{24}$$

The posteriors are

$$p\left(\boldsymbol{\mu}_k|\boldsymbol{Y}_k,\boldsymbol{\Sigma}_k,\boldsymbol{H}_{1:T}\right) \sim \prod_{t:s_t=k} p\left(\boldsymbol{Y}_t|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k,\boldsymbol{H}_t\right)p\left(\boldsymbol{\mu}_k\right) \tag{25}$$

$$\sim N\left(\boldsymbol{M}_\mu,\boldsymbol{V}_\mu\right) \tag{26}$$

where

$$\boldsymbol{M}_{\mu} = \boldsymbol{V}_{\mu} \left(\sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{H}_t^{-1/2} \boldsymbol{r}_t + \boldsymbol{B}_0^{-1} \boldsymbol{b}_0 \right) \tag{27}$$

$$\boldsymbol{V}_{\mu} = \left(\sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{H}_t^{-1/2} + \boldsymbol{B}_0^{-1} \right)^{-1} \tag{28}$$

and

$$p\left(\boldsymbol{\Sigma}_k | \boldsymbol{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{H}_{1:T}\right) \propto \prod_{t:s_t=k} p\left(\boldsymbol{r}_t | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_t\right) p\left(\boldsymbol{\Sigma}_k\right) \tag{29}$$

$$\sim IW\left(\bar{\boldsymbol{\Sigma}}, \bar{\nu} + N\right) \tag{30}$$

where

$$\bar{\nu} = T_k + \nu = \sum_{t=1}^{T} \mathbb{1}\left(s_t = k\right) + \nu \tag{31}$$

$$\bar{\boldsymbol{\Sigma}} = \sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2} \left(\boldsymbol{r}_t - \boldsymbol{\mu}_k\right) \left(\boldsymbol{r}_t - \boldsymbol{\mu}_k\right)' \boldsymbol{H}_t^{-1/2'} + \boldsymbol{\Sigma}_0 \tag{32}$$

11. Sample hierarchical priors.

    (a) Sample $\boldsymbol{b}_0 | \boldsymbol{\mu}_{1:K}, \boldsymbol{B}_0, \boldsymbol{h}_0, \boldsymbol{H}_0 \sim N\left(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b\right)$, where

$$\boldsymbol{\mu}_b = \boldsymbol{\Sigma}_b \left(\boldsymbol{B}_0^{-1} \sum_{k=1}^{K} \boldsymbol{\mu}_k + \boldsymbol{H}_0^{-1} \boldsymbol{h}_0 \right) \tag{33}$$

$$\boldsymbol{\Sigma}_b = \left(K \boldsymbol{B}_0^{-1} + \boldsymbol{H}_0^{-1}\right)^{-1} \tag{34}$$

    (b) Sample $\boldsymbol{B}_0 | \boldsymbol{\mu}_{1:K}, \boldsymbol{b}_0, a_0, \boldsymbol{A}_0 \sim IW\left(\boldsymbol{\Omega}_B, \omega_b\right)$, where

$$\omega_b = K + a_0 \tag{35}$$

$$\boldsymbol{\Omega}_B = \sum_{k=1}^{K} \left(\boldsymbol{\mu}_k - \boldsymbol{b}_0\right) \left(\boldsymbol{\mu}_k - \boldsymbol{b}_0\right)' + \boldsymbol{A}_0 \tag{36}$$

(c) Sample $\nu|\sigma^2_{1:K}, s_0, g_0$. There is no easily applicable conjugate prior for ν so a Metropolis-Hastings step needs to be applied. Implement a Gamma proposal following Maheu and Yang (2016):

$$\nu'|\nu \sim Gamma\left(\tau, \frac{\tau}{\nu}\right) \tag{37}$$

and the acceptance rate is

$$\min\left\{1, \frac{p\left(\nu'|\boldsymbol{\Sigma}_{1:K}, s_0, g_0\right)/q\left(\nu'|\nu\right)}{p\left(\nu|\boldsymbol{\Sigma}_{1:K}, s_0, g_0\right)/q\left(\nu|\nu'\right)}\right\} \tag{38}$$

(d) Sample $\boldsymbol{\Sigma}_0|\boldsymbol{\Sigma}_{1:K}, v_0, \boldsymbol{C}_0, d_0 \sim W\left(\boldsymbol{C}_s, d_s\right)$, where

$$\boldsymbol{C}_s = \left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1} + \boldsymbol{C}_0^{-1}\right)^{-1} \tag{39}$$

$$d_s = K\left(\nu + N\right) + d_0 \tag{40}$$

12. Sample the GARCH parameters $\boldsymbol{\theta}_H = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}\}|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{\Theta}$. With normal prior $\boldsymbol{\theta}_H \sim N\left(\boldsymbol{0}, \boldsymbol{I}\right)$, the posterior is

$$p\left(\boldsymbol{\theta}_H|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{\Theta}\right) \sim \prod_{t=1}^{T} p\left(\boldsymbol{r}_t|\boldsymbol{\Theta}, \boldsymbol{H}_t\right) p\left(\boldsymbol{\theta}_H\right) \tag{41}$$

Apply a random-walk Metropolis-Hastings algorithm to sample $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. $\boldsymbol{C}$ is jointly targeted.