Beyond Conditional Second Moments: Does Nonparametric Density Modelling Matter to Portfolio Allocation?

Li, Chenxing and Maheu, John M

Center for Economics, Finance and Management Studies, Hunan University, DeGroote School of Business, McMaster University

13 September 2023

Beyond Conditional Second Moments: Does Nonparametric Density Modelling Matter to Portfolio Allocation? *

Chenxing Li† John M. Maheu‡

December 2024

Abstract

This paper investigates the economic importance of nonparametrically/semiparametrically modelling the shape and the change in the unknown distribution of returns in portfolio allocation problems from a Bayesian perspective. Besides parametric multivariate GARCH (MGARCH) benchmark models for returns, we consider an MGARCH with innovations following a Dirichlet process mixture and an infinite hidden Markov model (IHMM). We introduce a new Bayesian semiparametric model that combines the MGARCH component with the IHMM for innovations. This new model nonparametrically approximates both the shape and evolution through time of the unknown distribution of returns beyond that captured by the MGARCH part. The results show that the Bayesian nonparametric/semiparametric models lead to improved statistical forecast accuracy and economic gains for a quadratic utility and CRRA utility investor. The new model makes the greatest gains. Portfolio choice is improved by modelling beyond the conditional second moments.

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

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

*Li acknowledges the financial support from the Philosophy and Social Science Foundation of Hunan Province (project number: 23YBA031), Maheu acknowledges financial support from SSHRC of Canada.
†Center for Economics, Finance and Management Studies, Hunan University, China, lichenxing@hnu.edu.cn
‡DeGroote School of Business, McMaster University, maheujm@mcmaster.ca

1 Introduction

It is well known that the accurate modelling of conditional covariances improves portfolio outcomes (Fleming et al., 2001; Guidolin and Timmermann, 2007; Pettenuzzo and Timmermann, 2011; Bollerslev et al., 2018, and many others). However, it is not clear whether modelling the unknown shape of the distribution and allowing the shape to change over time is important. This goes beyond second moments and can impact asymmetry and various tail shapes. Is it worth the additional costs of model complexity and estimation to use more sophisticated specifications? The first contribution of this paper is to investigate the economic importance of time variation in the multivariate return distribution over and above second moments through an extensive application to portfolio optimization. The second contribution is to propose a new multivariate Bayesian semiparametric model that approximates the shape and various dynamic features of the unknown conditional distribution.

Dynamics in conditional second moments are well acknowledged in multivariate financial time series. Multivariate GARCH (MGARCH) models (Bollerslev et al., 1988; Engle and Kroner, 1995; Bollerslev, 1990; Engle, 2002) and their leptokurtic extensions (see Pesaran and Pesaran, 2010; Kawakatsu, 2006; Bonato, 2012, and many others) have been extensively developed to capture the gradual changes in conditional covariance matrices.

To further capture the unknown shape of the return innovation distribution, in addition to the highly persistent dynamics of conditional covariances, Jensen and Maheu (2013) and Maheu and Shamsi Zamenjani (2021) add a Dirichlet process mixture (DPM) model to the multivariate GARCH model (MGARCH-DPM). This Bayesian semiparametric model approximates the unknown distribution of conditional returns by mixing an infinite number of multivariate normal kernels with a Dirichlet process prior (Antoniak, 1974). Mixtures of normals can capture distributional features such as skewness, kurtosis, and asymmetric tails for any continuous distribution. In the MGARCH-DPM model, the DPM component relaxes the parametric assumption that the innovation distribution is known, but the mixture weights remain constant, and any time variation in the return distribution can only come from the GARCH part of the model.

The new Bayesian semiparametric model introduced in this paper extends the MGARCH-DPM model directly by replacing the DPM with an infinite hidden Markov model (IHMM). The IHMM proposed by Beal et al. (2002) has Markovian mixing weights that are constructed from a hierarchical Dirichlet process (HDP) prior formalized by Teh et al. (2006). The shape of the unknown distribution can change over time. The DPM version is also nested as a special case. The IHMM can also be seen as a Bayesian nonparametric extension of the Markov switching model as it generalizes the predefined finite number of states into an infinite number of states. Conditional return 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 to economic and financial time series since econometricians introduced it from computer science. Univariate examples include Song (2014) and Jochmann (2015) on inflation rates, Maheu and Yang (2016) on short-term interest rates, and Jin et al. (2022) on macroeconomic and financial forecasting. Multivariate applications in Jin and Maheu (2016); Jin et al. (2019); Hou (2017) focus on estimation and forecast improvements from the IHMM but do not consider whether it is beneficial for portfolio decisions.

Although the IHMM extends the Markov switching models into an unbounded state space, a regime-switching model may not be able to capture the smooth changes in conditional second moments (Rydén et al., 1998). The infinite hidden Markov multivariate GARCH model (MGARCH-IHMM) in this paper uses the MGARCH component to capture the strong persistence in covariance dynamics and allows the IHMM component to focus on approximating the changes in the unknown shape of the conditional distribution of returns. The IHMM component is also able to capture the potential state recurrence in addition to abrupt changes in variance levels.

Our MGARCH-IHMM is related to Dufays (2016) who applies the IHMM with a univariate GARCH model for demeaned data where skewness is not allowed. This paper extends his work to a flexible multivariate setting that allows for time-varying covariance, skewness, and kurtosis at the same time. We parameterize the MGARCH component to avoid any path dependence issue that can arise in switching GARCH models (Bauwens et al., 2010), making estimation tractable with MCMC methods for IHMM. Although we document the forecast performance of this new model and others, our focus is on the economic value of the IHMM component for portfolio choice.

The portfolio problem involves maximizing expected utility subject to a wealth constraint and transaction costs. To make the problem realistic, we impose no-short-selling restriction and a separate case with no short-selling and no leverage-trading at the same time. From a Bayesian perspective, the expected utility is taken with respect to the predictive distribution. In the nonparametric/semiparametric models, this will involve integrating out parameter uncertainty as well as distributional uncertainty. Following MCMC estimation, the draws from the predictive density are constructed to empirically estimate an investor's expected utility, which is maximized to solve for the optimal portfolio weights.

Model estimation and portfolio choice are applied to two monthly return datasets, Fama-French 5 industry portfolios and Fama-French 5 factors portfolios. Several benchmark parametric MGARCH models, along with the nonparametric IHMM and semiparametric MGARCH-DPM and MGARCH-IHMM, are used for portfolio choice. Posterior results show that, in addition to the smooth volatility changes captured by the MGARCH component, significant parameter changes are present from the IHMM portion of the specifications. Posterior estimates indicate many active components in mixture models, and multivariate measures of skewness and kurtosis indicate departures from normality.

In terms of out-of-sample forecasts, the MGARCH-IHMM model performs significantly better than all benchmark models through predictive likelihoods. It also has accurate covariance forecasts as measured against monthly realized covariance, although this depends on the dataset. Do these forecast improvements lead to better portfolio outcomes?

We compute an ex-post break-even performance fee that equates the utility performance of two portfolios optimized from different models. This fee represents the annualized return the investor would pay to switch from one model to another to perform portfolio optimization while accounting for transaction costs. Overall, with a quadratic or CRRA utility, the investor would pay a positive fee against all alternative econometric models to obtain forecasts from the MGARCH-IHMM. The fees are substantial. For example, a CRRA utility investor would pay an annual fee around $1 - 3\%$ to move from one of the parametric MGARCH models to the proposed semiparametric alternative under the no-short-selling restriction. Against simple alternatives like an equally-weighted (EW) portfolio or value-weighted (VW) portfolio the fees can be in excess of 5% in some settings. The fee generally increases with transaction costs but declines with risk aversion. For high risk aversion and high transaction costs or no leverage, the simple VW portfolio becomes the dominant choice.

Although our portfolios are not optimized for the Sharpe ratio, we find that the MGARCH-IHMM model is often preferred when there are no transaction costs but otherwise the EW portfolio is best. DeMiguel et al. (2007) point out that the EW buy-and-hold portfolio is robust to parameter uncertainty and difficult to beat. We show that although the EW portfolio is competitive, the MGARCH-IHMM model strictly dominates it based on utility outcomes for low to medium risk aversion levels.

We conclude that dynamic density modelling using the more sophisticated but flexible semiparametric (MGARCH-IHMM) specification does matter to portfolio choice. This semiparametric model, to the extent that it improves the predictive covariances, leads to better portfolio outcomes for a quadratic utility investor. For the CRRA investor who will be sensitive to the whole distribution of wealth, the models with better density forecasts tend to provide better portfolio results than other econometric alternatives. The fees a CRRA investor would pay for the MGARCH-IHMM forecasts against alternatives is generally larger than the quadratic utility investor. A robust and useful alternative model is the VW portfolio. It performs well when there is no leverage and is otherwise good when risk aversion and/or transaction costs are high.

This paper is organized as follows. Section 2 presents the portfolio allocation problem of the investor for a given utility function and discusses the transaction cost and optimization. The next section focuses on the new MGARCH-IHMM, including estimation and forecasting in addition to detailing several benchmark models. Section 4 discusses the returns series used to estimate models and perform portfolio allocation. Section 5 presents the posterior estimations of the models and compares their out-of-sample forecast performance. Section 6 compares the portfolio allocation performance of different econometric models under different risk-aversion levels, transaction costs, and trading restriction settings. Section 7 concludes, while the appendices contain additional details on models and estimation steps.

2 Utility-Based Portfolio Optimization

We make use of the following notation. Let $\boldsymbol{R}_t$ denote the simple return vector from N risky assets and $R_{f,t}$ be the return from the risk-free asset.[1] Let the information set be $\boldsymbol{R}_{1:t-1} = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_{t-1}\}$ and $\boldsymbol{\iota}$ be a vector of 1.

2.1 Dynamic Optimal Portfolio Weights

Consider a rational and risk-averse investor whose utility function is $U(W)$, where W denotes their wealth. Each period, they distribute their wealth into N risky assets and the risk-free asset. Without loss of generality, assume that their wealth is set as 1 at the beginning of each period. The investor rebalances their portfolio given the information set $\boldsymbol{R}_{1:t-1}$ for period t by maximizing their conditional expected utility:

$$\max_{\boldsymbol{w}_t} \mathrm{E}\left[U(W_t)|\boldsymbol{R}_{1:t-1}\right] \tag{1a}$$

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

$\boldsymbol{w}_t$ represents the vector of the portfolio weights for the risky assets, and $C(\cdot)$ is the transaction cost incurred when rebalancing at the end of period t and is a function of $\boldsymbol{w}_t$. We focus on two realistic cases that investors are likely to face. The first is the no-short-sale constraint ($w_{i,t} \geq 0$ for each asset i). Jagannathan and Ma (2003) show this can lead to less extreme weights and better portfolio outcomes. The second case is the no-short-sales and no-leverage-trading constraint ($w_{i,t} \geq 0$ for each asset i and $\sum_{i=1}^{N} w_{i,t} \leq 1$) which means the investor cannot borrow to invest in risky assets.

[1]For each element in $\boldsymbol{R}_t$, $R_{i,t} = \exp(r_{i,t})-1$, where $r_{i,t}$ is the log return for asset i predicted from the econometric model.

In selecting the optimal portfolio at time t, the investor is subject to the transaction fee,

$$C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}) = c \sum_{i=1}^{N} |w_{i,t} - (1 + R_{i,t-1})w_{i,t-1}| + c \left|(1 - \boldsymbol{w}_t'\boldsymbol{\iota}) - (1 + R_{f,t-1})(1 - \boldsymbol{w}_{t-1}'\boldsymbol{\iota})\right|. \quad (2)$$

Here, c denotes the flexible fee as a certain percentage of the wealth. The transaction costs reflect the change in weights on risky and riskless assets but also the cost to rebalance when weights are kept constant from $t-1$ to t given the wealth is normalized to 1 at the beginning of each period. Note that in contrast to others, such as Bollerslev et al. (2018), in which turnover and transaction costs are ex-post, our investor optimizes their portfolio subject to these costs.

The analytical solution for this problem is generally unavailable, but a numerical solution can be found. In practice, we approximate expected utility with draws from the predictive distribution of the econometric model $\mathcal{M}_i$ for $\boldsymbol{R}_t$ given $\boldsymbol{R}_{1:t-1}$ through,

$$\mathrm{E}\left[U(W_t)|\boldsymbol{R}_{1:t-1}, \mathcal{M}_i\right] \approx \frac{1}{M} \sum_{m=1}^{M} U(W_{t,\mathcal{M}_i}^{(m)}), \quad (3a)$$

$$W_{t,\mathcal{M}_i}^{(m)} = 1 + \boldsymbol{w}_t'\boldsymbol{R}_{t,\mathcal{M}_i}^{(m)} + (1 - \boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t} - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}), \quad (3b)$$

where $\boldsymbol{R}_{t,\mathcal{M}_i}^{(m)}$ is simulated from the predictive distribution for $\boldsymbol{R}_t$ by model $\mathcal{M}_i$, and M is the total number of draws. As discussed below, the draws from the predictive distribution will integrate out parameter uncertainty and, in the nonparametric/semiparametric models, distributional uncertainty.

2.2 Break-even Management Fees

Since the notion of "risk" can go beyond the second moment of a return distribution which traditional measures like the Sharpe ratio use, we consider the break-even performance fee an investor is willing to pay to switch from one econometric specification of returns to another (e.g., Fleming et al., 2001) to compare models. The fee embodies the risk-return trade-off that the investor has based on their utility function.

Given an out-of-sample period from $t+1$ to T, a utility function $U(W)$ and two models $\mathcal{M}_A$ and $\mathcal{M}_B$, we compute the optimal weights for each model at each time following the discussion above.[2] Given the optimal weights, the realized wealth for each model can be computed, $\{W_{i,l}\}_{l=t+1}^{T}$ for $i = A, B$. The performance fee, Δ, that an investor would pay to move from model $\mathcal{M}_A$ to $\mathcal{M}_B$ is found by equating the ex-post utility as

$$\sum_{l=t+1}^{T} U(W_{A,l} - \Delta|\mathcal{M}_A) = \sum_{l=t+1}^{T} U(W_{B,l}|\mathcal{M}_B), \quad (4)$$

and solving for Δ. This fee automatically takes into account transaction costs incurred in each out-of-sample period.

[2] Note that each model $\mathcal{M}_A$ and $\mathcal{M}_B$ are re-estimated at each point in the out-of-sample period to obtain draws from the predictive density.

3 Models

In this section, we present the various multivariate models of returns used to solve the portfolio optimization problem. We begin with the new model that nonparametrically allows for time dependence in the distribution coupled with a multivariate GARCH specification. As this is a new model, we present more details on the model than the other benchmarks, including estimation and simulation from the predictive density. Following this, we briefly detail existing nonparametric models from the literature with some extensions. These are the MGARCH-DPM of Jensen and Maheu (2013); Maheu and Shamsi Zamenjani (2021) and a version of the IHMM from Maheu and Yang (2016) without a multivariate GARCH component. Finally, the benchmark multivariate GARCH models with parametric return distributions are considered. These existing models, except for IHMM, allow the conditional covariance to change, but the innovation distribution is constant through time.

3.1 MGARCH-IHMM

To investigate the importance of general distributional shapes and changes over time beyond second moments, this paper proposes a sophisticated model that extends an MGARCH structure to capture general nonparametric patterns. The proposed Bayesian semiparametric model consists of an MGARCH component and an IHMM component. Let $\boldsymbol{r}_t$ be an $N \times 1$ vector of log-returns, and $\boldsymbol{r}_{1:T} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_T\}$. Define $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots\}$ as the set of state-dependent parameters, where $\boldsymbol{\theta}_j = \{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}$. Let $\boldsymbol{\Theta}$ be the joint set of all parameters. The stick-breaking representation (Sethuraman, 1994; Teh et al., 2006) of the model is as follows:

$$\boldsymbol{\Gamma}|\beta_0 \sim GEM(\beta_0), \quad \boldsymbol{\Pi}_j|\alpha_0, \boldsymbol{\Gamma} \sim DP(\alpha_0, \boldsymbol{\Gamma}) \tag{5a}$$

$$s_t|s_{t-1}, \boldsymbol{\Pi} \sim categorical(\boldsymbol{\Pi}_{s_{t-1}}) \tag{5b}$$

$$p(\boldsymbol{r}_t|\boldsymbol{\Theta}, s_{t-1}) = \sum_{k=1}^{\infty} \pi_{s_{t-1}k} N\left(\boldsymbol{r}_t \middle| \boldsymbol{\mu}_k, \boldsymbol{H}_t^{1/2} \boldsymbol{\Sigma}_k \boldsymbol{H}_t^{1/2'}\right) \tag{5c}$$

$$\boldsymbol{\mu}_k \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0), \quad \boldsymbol{\Sigma}_k \sim IW(\boldsymbol{\Sigma}_0, \nu + N) \tag{5d}$$

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\boldsymbol{r}_{t-1} - \boldsymbol{\eta})(\boldsymbol{r}_{t-1} - \boldsymbol{\eta})' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1} \tag{5e}$$

where $\boldsymbol{\Gamma} = (\gamma_1, \gamma_2, \dots)'$, $\boldsymbol{\Pi}_j = (\pi_{j1}, \pi_{j2}, \dots)$. To be more specific,

$$\gamma_k = \tilde{\gamma}_k \prod_{l=1}^{k-1} (1 - \tilde{\gamma}_l), \qquad \tilde{\gamma}_k \sim Beta(1, \beta_0), \tag{6a}$$

$$\pi_{jk} = \tilde{\pi}_{jk} \prod_{l=1}^{k-1} (1 - \tilde{\pi}_{jl}), \qquad \tilde{\pi}_{jk} \sim Beta\left(\alpha_0\gamma_k, \alpha_0\left(1 - \sum_{l=1}^{k} \gamma_l\right)\right). \tag{6b}$$

$GEM(\beta_0)$[3] is a general stick-breaking process of (6a) with a concentration parameter β_0. As shown in equation (5c), the conditional distribution of the returns is a mixture of an infinite number of Gaussian kernels with a vector of weights $\boldsymbol{\Pi}_j$. $\boldsymbol{\Pi}_j$ is the jth row of the infinite dimension squared transition matrix $\boldsymbol{\Pi}$ and a draw from a particular Dirichlet process. 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 parameterized

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

around an identity matrix. The general level and long-run dynamics of conditional volatility are captured by the GARCH component, and $\mathbf{\Sigma}_{s_t}$ serves as an amplifier to either boost or shrink the conditional covariance from $\boldsymbol{H}_t$. Clearly, when $\boldsymbol{\mu}_{s_t} = \mathbf{0}$ and $\mathbf{\Sigma}_{s_t} = \boldsymbol{I}$, 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). (5a) represents this HDP structure. This component is essentially an infinite dimension Markov-switching model. It is designed to capture the sudden changes in the conditional distribution through a regime-switching scheme, and all of the states can recur with certain probabilities. To ensure this recurrence, another Dirichlet process is required to 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-IHMM is more flexible and nests the MGARCH-DPM model of Jensen and Maheu (2013) as a special case. Because the MGARCH-IHMM is a Bayesian semiparametric model, where one does not need to impose any distributional assumption on the conditional return distributions, 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 period-by-period evolution of the conditional distributions.

The MGARCH component (5e) 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 $\boldsymbol{H}_t$.[4] Unlike traditional MGARCH models, the effect of the lagged return shock is centred around the additional parameter $\boldsymbol{\eta}$. This specification, on one hand, avoids the path dependence problem in estimation since $\boldsymbol{H}_t$ is not a function of the states. On the other hand, it allows $\boldsymbol{H}_t$ to capture the potential asymmetric volatility feedback effect. When $\boldsymbol{r}_{t-1} > \boldsymbol{\mu}_{s_{t-1}} > \boldsymbol{\eta}$, the distance between $\boldsymbol{r}_{t-1}$ and $\boldsymbol{\eta}$ is greater than that between $\boldsymbol{r}_{t-1}$ and $\boldsymbol{\mu}_{s_{t-1}}$, and 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 any feedback is instead learned from data. The posterior estimates of Fama-French 5 industry portfolio returns show that η_i is generally greater than $\mu_{i,s_{t-1}}$ for each asset i, indicating a negative feedback in volatility empirically.

The state-dependent parameters $\boldsymbol{\mu}_k$ and $\mathbf{\Sigma}_k$ allow us to identify potential regime switches. Furthermore, mixing over $\boldsymbol{\mu}_k$ also generates possible skewness in conditional return distributions, and mixing over $\mathbf{\Sigma}_k$ 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 of both the MGARCH model and the IHMM. It approximates both the shape and the period-by-period evolution of the unknown conditional distributions semiparametrically.

3.2 Hierarchical Priors

Prior settings are important when estimating Bayesian nonparametric/semiparametric models where the number of active states is estimated jointly with other parameters. To allow learning

[4]This will be true for a prior expectation of $\mathrm{E}(\mathbf{\Sigma}_k) = \boldsymbol{I}$ but in general need not be true for other posterior parameter values.

on new state-dependent parameters, we employ hierarchical priors. 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 and can improve out-of-sample forecasts and portfolio outcomes (Maheu and Yang, 2016).

The following set of hierarchical priors motivated by Song (2014) are used:

$$\boldsymbol{b}_0 \sim N\left(\boldsymbol{h}_0, \boldsymbol{H}_0\right), \quad \boldsymbol{B}_0 \sim IW\left(\boldsymbol{A}_0, a_0\right), \quad \boldsymbol{\Sigma}_0 \sim W\left(\boldsymbol{C}_0, d_0\right), \quad \nu \sim Exp\left(g_0\right). \tag{7}$$

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

3.3 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. 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 in the main MCMC procedure, reduces the complexity of posterior sampling. Let $\bar{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{r}_t$ be the sample mean 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 sample covariance. Assuming $\mathrm{E}\left(\boldsymbol{H}_t\right) = \bar{\boldsymbol{H}}$ for all $t = 1, \ldots, T$, then in the MGARCH-IHMM, $\boldsymbol{C}\boldsymbol{C}'$ can be replaced as

$$\boldsymbol{C}\boldsymbol{C}' = \bar{\boldsymbol{H}} \odot \left[\mathbf{1} - \boldsymbol{\alpha}\boldsymbol{\alpha}' - \boldsymbol{\beta}\boldsymbol{\beta}'\right] - \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)\left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)', \tag{8}$$

where $\mathbf{1}$ is an $N \times N$ matrix with all the elements being 1. This result only holds for a restricted version of the model. See Appendix A for details. Note that any draw of $\boldsymbol{\theta}_H$ from the posterior that results in non-positive definite $\boldsymbol{C}\boldsymbol{C}'$ is rejected.

3.4 Sampling Algorithm

The MGARCH-IHMM is estimated through an MCMC algorithm. For the nonparametric component (IHMM), we 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 *active* states with assigned observations and an additional *remaining* state where no observation is allocated. This allows for standard finite Markov switching posterior simulation methods (Chib, 1996) to be employed. Conditional on the states, the MGARCH parameters can be sampled with a multivariate random-walk proposal. Detailed steps of posterior simulations are found in Appendix B.

Collecting M MCMC samples after dropping a suitable amount of burn-in samples, the posterior moments of some function $g(\cdot)$ can be computed by

$$\mathrm{E}\left[g(\theta)|\boldsymbol{r}_{1:T}\right] \approx \frac{1}{M}\sum_{m=1}^{M} g\left(\theta^{(m)}\right),$$

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

3.5 Predictive Density

A key input into model comparison, forecasting and hence portfolio choice is the predictive density of a model. For a general model $\mathcal{M}_A$, this is defined as

$$p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \mathcal{M}_A) \quad = \quad \int p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}, \mathcal{M}_A) p(\boldsymbol{\Theta}|\boldsymbol{r}_{1:t}, \mathcal{M}_A) d\boldsymbol{\Theta}, \tag{9}$$

where $p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}, \mathcal{M}_A)$ is the data density for $\boldsymbol{r}_{t+1}$ given data $\boldsymbol{r}_{1:t}$ and parameter vector $\boldsymbol{\Theta}$ and $p(\boldsymbol{\Theta}|\boldsymbol{r}_{1:t}, \mathcal{M}_A)$ is the posterior density of $\boldsymbol{\Theta}$ given $\boldsymbol{r}_{1:t}$.

Computing the expected utility for portfolio choice will require simulating from the predictive density. For each MCMC draw $\boldsymbol{\Theta}^{(m)}$ from the posterior, we simulate a $\boldsymbol{r}_{t+1}^{(m)}$ using the data density. In practice, for the MGARCH-IHMM this will require simulating a future state which may be a new state that requires a draw from the hierarchical prior.

The predictive likelihood evaluates the predictive density at the realized $\boldsymbol{r}_{t+1}$ and is the basis of model comparison. For $\mathcal{M}_A$, the predictive likelihood can be estimated as $\frac{1}{M}\sum_{m=1}^{M} p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}^{(m)}, \mathcal{M}_A)$ with $\boldsymbol{\Theta}^{(m)}$ draws from $p(\boldsymbol{\Theta}|\boldsymbol{r}_{1:t}, \mathcal{M}_A)$.

Although our main focus is portfolio choice, we also document the econometric gains in density forecasts that the models provide by comparing predictive likelihoods. This is computed out-of-sample recursively as detailed in Appendix C. The log-predictive likelihood over the out-of-sample period $t+1, \ldots, T$ for model $\mathcal{M}_A$ is

$$\log PL_A = \log p(\boldsymbol{r}_{t+1:T}|\boldsymbol{r}_{1:t}, \mathcal{M}_A) = \sum_{l=t}^{T-1} \log p(\boldsymbol{r}_{l+1}|\boldsymbol{r}_{1:l}, \mathcal{M}_A). \tag{10}$$

Models with larger $\log PL$ are more consistent with the observed data and the log-Bayes factor in favour of $\mathcal{M}_A$ vs $\mathcal{M}_B$, is $\log BF_{AB} = \log PL_A - \log PL_B$. A log-Bayes factor greater than 5 is usually considered strong evidence for model $\mathcal{M}_A$.

3.6 Other Models

We include the following semiparametric and fully nonparametric models, as well as purely parametric models.

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

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

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, as discussed in Section 3. $\boldsymbol{C}\boldsymbol{C}'$ is targeted as in equation (8).

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

$$\begin{aligned}
\boldsymbol{\Gamma}|\beta_0 \sim GEM(\beta_0), &\quad \boldsymbol{\Pi}_j|\alpha_0, \boldsymbol{\Gamma} \sim DP(\alpha_0, \boldsymbol{\Gamma}) \\
s_t|s_{t-1}, \boldsymbol{\Pi} \sim categorical(\boldsymbol{\Pi}_{s_{t-1}}), &\quad \boldsymbol{r}_t|s_t, \boldsymbol{\theta} \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{\Sigma}_{s_t}\right) \\
\boldsymbol{\mu}_s \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0), &\quad \boldsymbol{\Sigma}_s \sim IW(\boldsymbol{V}, \nu + N).
\end{aligned}$$

This model is nested within the MGARCH-IHMM when $\boldsymbol{H}_t = \boldsymbol{I}$ for all t. The same set of hierarchical priors in (7) is also employed in this model. The only way to capture conditional heteroskedasticity is through changes in $\boldsymbol{\mu}_{s_t}$ and $\boldsymbol{\Sigma}_{s_t}$.

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

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

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

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

MGARCH-A A fully parametric asymmetric multivariate GARCH model with normal innovations 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 (8). Unlike the MGARCH-N model, this model can capture the potential shock asymmetry in volatility feedback through $(\boldsymbol{r}_{t-1} - \boldsymbol{\eta})$.

4 Data

We consider two datasets in our work. The first includes monthly holding period returns of the Fama-French 5 industry portfolios (hereinafter, industry portfolios), consisting of Consumer, Manufacture, High Tech, Health and Other portfolios, while the second is the Fama-French 5 factor portfolios (hereinafter, factor portfolios), and the 1-month Treasury bill rate from the website of Kenneth French.[5] The industry portfolio returns range from July 1926 to December 2020 at a monthly frequency (1,134 observations), and the factor portfolio returns range from July 1963 to December 2023 (726 observations). The model estimation uses the associated continuous compounded returns scaled by 100. In addition, 5-dimensional monthly realized covariances are computed as the summation of the outer product of the daily return vector of the 5 assets of each data set. This is used to compare predictive covariances from models.

Table 1 illustrates some descriptive statistics for both portfolio returns. All of the industries are negatively skewed and leptokurtic and highly correlated. For the factor portfolios, only the market and the RMW portfolio are strongly negatively skewed, and only the RMW portfolio is heavily leptokurtic. The factor portfolios are in general less correlated than the industry portfolios, as expected.

5 Model Estimates and Forecasts

Before analyzing the economic gains in portfolio allocation, we report model estimates and comparisons of out-of-sample forecasts. We focus on the new semiparametric MGARCH-IHMM model.

[5] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The 1-month T-bill rate serves as the risk-free investment in the portfolio optimization.

5.1 Hyper parameters

The hyperparameters for the priors, the hyperpriors 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 hyperpriors for the HDP concentration parameters β_0 and α_0 are assumed as

$$\beta_0 \sim Gamma(2, 8), \qquad \alpha_0 \sim Gamma(2, 8), \tag{11}$$

where $\mathrm{E}(\beta_0) = \mathrm{E}(\alpha_0) = 0.25$. During estimation, these hyperpriors strongly favour less active states for better state identification and faster computation. The hyperprior for the concentration parameter of the DPM is also assumed to be distributed as $Gamma(2, 8)$. The hierarchical priors discussed in Section 3.2 are assumed as

$$\boldsymbol{b}_0 \sim N\left(\mathbf{0}, \boldsymbol{I}\right), \quad \boldsymbol{B}_0 \sim IW\left(\boldsymbol{I}, N+2\right), \quad \boldsymbol{\Sigma}_0 \sim W\left(\frac{\boldsymbol{I}}{N+2}, N+2\right), \quad \nu \sim Exp\left(\frac{1}{N+2}\right), \tag{12}$$

and ensure $E[\boldsymbol{\Sigma}_0] = \boldsymbol{I}$, $E[\nu] = N+2$ and using these values centers the hierarchical prior for $\boldsymbol{\Sigma}_k$ on $\boldsymbol{I}$. For the MGARCH parameters $\boldsymbol{\theta}_H = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta})'$, assume a truncated multivariate normal prior

$$\boldsymbol{\theta}_H \sim N(\mathbf{0}, \boldsymbol{I})\mathbb{1}\{\alpha_i > 0, \beta_i > 0, \alpha_i^2 + \beta_i^2 < 1, \text{ for all } i\}, \tag{13}$$

where $\alpha_i > 0$ and $\beta_i > 0$ are imposed for parameter identification and $\alpha_i^2 + \beta_i^2 < 1$ for covariance targeting.

5.2 In-Sample Estimation

In all models, the first 20,000 iterations were discarded as burn-in, and the next 20,000 MCMC samples were collected for posterior inference. This section summarizes the results of the full sample estimates for both datasets.

Tables 2 and 3 list 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, the MGARCH-DPM, the MGARCH-N and the MGARCH-A models all have a high $\boldsymbol{\beta}$ that is above 0.89, and low $\boldsymbol{\alpha}$ that is below 0.37 in general. As mentioned in Section 3, $\boldsymbol{\eta}$ helps to capture an asymmetric volatility feedback. The $\boldsymbol{\eta}$ estimates are often considerably larger than the sample mean of returns in Table 1 as well as the estimates of $\boldsymbol{\mu}$ in the parametric MGARCH models. This indicates an asymmetric volatility response to return shocks for both portfolio returns. Put another way, bad news (negative shocks) has a larger impact on future $\boldsymbol{H}_t$ than good news (positive shocks). The MGARCH persistence measure $(\alpha_i^2 + \beta_i^2)$ is very similar in all models, although each of the individual α_i and β_i values are a bit different.

Comparing the semiparametric MGARCH-IHMM and the nonparametric IHMM, the MGARCH-IHMM has higher concentration parameters (α_0 and β_0) than the IHMM and a greater number of active states (K) for each dataset. The MGARCH-DPM model has the lowest concentration parameter ($\beta_0 = 0.387$ and 0.458) and the least number of active states ($K \approx 5$).

Figure 1 plots the posterior mean of several time-varying parameters for the MGARCH-IHMM based on the industry portfolios. The top plot shows the posterior mean of the average over the five portfolios of $\boldsymbol{\mu}_{s_t}$. The other three plots show the log-determinants of the posterior mean for the time-varying second-moment parameters. As discussed earlier, $\boldsymbol{\Sigma}_{s_t}$ has a prior centred around the identity matrix whose log-determinant is 0. To the extent that $\boldsymbol{\Sigma}_{s_t}$ differs from the identity, this indicates that $\boldsymbol{H}_t$ is scaled or shrunk as it enters the covariance of the normal kernels of the mixture. The figure shows this to be a regular occurrence and highlights deviations from the MGARCH-A model.

While many of the active states are there to approximate the tails and shapes of the conditional return distribution, there are three visually identifiable approximate active states: a *bull* state with high expected returns and low state-dependent volatility (green background), a *bear* state with low expected returns and high state-dependent volatility (red background) and a *correction* state with both high expected returns and low state-dependent volatility (white background). Note that these are states that have been identified in addition to the GARCH effect, indicating that the regular MGARCH-N model is insufficient in capturing the full dynamics of the conditional distribution.

The last graph of Figure 1 plots the log-determinant of the posterior mean of the overall conditional covariance from each model, namely $\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2\prime}$ for the MGARCH-IHMM and MGARCH-DPM, $\boldsymbol{\Sigma}_{s_t}$ for the IHMM, and $\boldsymbol{H}_t$ for the MGARCH-N and MGARCH-A. All the models roughly track the same trend. The two middle plots show that the MGARCH-IHMM decomposes the strong persistent component into $\boldsymbol{H}_t$ and abrupt discrete changes to $\boldsymbol{\Sigma}_{s_t}$.

A heat map of the three main states discussed above is found in Figure 2 for the MGARCH-IHMM. This displays the empirical probability that two periods share the same state. The redder the colour, the higher the probability of sharing states. 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. The correction state mainly corresponds to the pre- and post-WW2 market corrections from the bull markets.

5.3 Out-of-Sample Forecasts

A recursive prediction is performed for the last 360 out-of-sample periods (January 1999 to December 2023) for the industry portfolio and the last 300 out-of-sample periods (January 1991 to December 2020) for the shorter factor portfolio dataset. In each case, we re-estimated each model, at each out-of-sample period t, and computed the one period ahead log-predictive likelihood, predictive mean, and predictive covariance. Point forecasts are based on 20,000 simulations for a model's predictive density.

Table 4 compares the performance of those recursive forecasts. The MGARCH-IHMM produces large improvements against all models, with the second most competitive model being the MGARCH-DPM. The statistical gains from going dynamically semiparametric (MGARCH-IHMM) are very large with log-Bayes factors in excess of 32 for both datasets against parametric MGARCH alternatives. The importance of the MGARCH dynamics is clear from the MGARCH-A and MGARCH-N models having a larger log-predictive likelihood value than the IHMM, which omits that feature.

There is little to differentiate the predictive mean forecasts among the models.[6] Turning to the predictive covariance forecasts measured against monthly realized covariance, the best point

[6] The forecast error is computed from the difference between the realized $\boldsymbol{r}_t$ and the predicted mean of $\boldsymbol{r}_t$ from a model.

forecasts are from the MGARCH-IHMM while for the factor portfolio dataset, the MGARCH-IHMM is only marginally better. Similar to the log-predictive likelihood rankings, the MGARCH component is important to competitive forecasts with the IHMM performing the worst.

Finally, dynamic infinite mixture models such as the MGARCH-IHMM can capture various tail shapes and asymmetries. To see if these are present, we plot the Mardia (1970) multivariate skewness and kurtosis measures for the predictive density over the out-of-sample periods. The time series is shown in Figures 3 and 4 along with the Mardia measure for the normal distribution. There is a clear deviation from the normal distribution as well as evidence that the Mardia multivariate measures change dramatically over time.

Do the improved forecasts given by the MGARCH-IHMM and to a lesser extent the MGARCH-DPM, transfer over to better portfolio choice outcomes? This is the question we turn to in the next section.

6 Empirical Portfolio Choice

6.1 Quadratic Utility

The first set of results is for portfolio optimization with quadratic utility,

$$U(W) = W - \frac{a}{2(1+a)}W^2,$$

where a is the risk aversion parameter, the absolute risk aversion is $\frac{a}{1+a-aW}$ and the relative risk aversion is $\frac{aW}{1+a-aW}$. The portfolio is optimized when $a \in \{2, 4, 6\}$ for all five econometric models with transaction costs $c \in \{0\%, 1\%, 2\%\}$, respectively. We use $M = 20,000$ from the predictive distribution of a model to estimate (3a). Due to the possible corner solution, a simplex method is used for optimization with 100 different initial values being used to avoid a local optimum. The Brent-Dekker method is used to find the scalar root in equation (4). Each econometric model is recursively estimated to compute the draws from the respective predictive density of the returns to estimate (3a) at each point in the out-of-sample period for each dataset.

In addition to the benchmark models listed in Section 3.6, we consider two simple weighted portfolios. The first is a buy-and-hold strategy of an equally-weighted (EW) portfolio that consists of the 5 assets in a portfolio, where each component has a weight of 1/5. The buy-and-hold EW portfolio is recommended by DeMiguel et al. (2007) as very competitive primarily because it has no parameter uncertainty. The second option is the value-weighted (VW) portfolio of the 5 assets.

Table 5 and 6 reports the annualized fee, after transaction costs, that an investor is willing to pay for switching from the econometric model in the first column to the MGARCH-IHMM for both datasets. This fee is based on the ex-post utility using the same out-of-sample period as the statistical forecast evaluations. In practice, short-selling can be very expensive or unavailable for many investors, so we impose a restriction where short-selling is not permitted for any risky asset (referred as “No Short-Selling”, $w_{i,t} \geq 0$ for all $i = 1, \ldots, N$).

In Panel A of the tables, a quadratic utility investor, with a few exceptions, is always willing to pay a positive annualized fee to switch from any benchmark model to the MGARCH-IHMM, regardless of the risk-averse parameter or the transaction costs. Most of the fees are in the $1-3$ percent range but several are larger. For example, for $a = 2$ and $c = 0$ the fee is 5.10% and 5.35% to move from the EW and VW to the MGARCH-IHMM in Table 5. The fee is smaller, about 2.25% to move from the MGARCH-N and MGARCH-A to the MGARCH-IHMM. This can only come from better predictive mean and predictive covariance forecasts. Table 4 shows it is mostly the latter that provides improvements. In general, the fee declines as risk aversion a increases while the fee increases with transaction costs c. The one notable exception to these results is the VW portfolio. An investor is willing to pay a positive fee to move from this portfolio to the MGARCH-IHMM for low risk aversion $a = 2$, but would not be willing to switch for higher risk aversion parameters and transaction costs. Thus the advantage of the MGARCH-IHMM specification for the quadratic utility investor is in the low risk aversion, low transaction cost setting.

As expected, imposing the additional constraint of no leverage-trading leads to lower performance fees in Panel B of the tables. The fees are often below 1% level for $a = 2$ but increase for larger risk aversion levels for many cases. For low risk aversion and transaction fees the VW portfolio is preferred. For high risk aversion or high transaction costs there are a few cases in which the MGARCH-IHMM is preferred but overall the VW is the top model. Nevertheless, the MGARCH-IHMM is always a better choice over other econometric models.

The utility-based performance fees provide an internally coherent method to compare models in that the utility function is used to optimize the portfolio weights as well as rank the models, all while accounting for transaction costs. We have not optimized Sharpe ratios, but we present ex-post Sharpe ratios for selected optimal portfolios after transaction costs for the no short-selling case in Table 7.[7] The no transaction cost case $c = 0$ favours the MGARCH-IHMM, but this disappears once transaction costs are positive. In fact, none of the time-series models are preferred while the EW portfolio has the best overall performance.

In summary, the improved point forecasts of the MGARCH-IHMM often translate into positive performance fees that the investor would pay to obtain these forecasts and resulting positive portfolio outcomes compared to many benchmarks. The fee greatly depends on the risk aversion parameter as well as the transaction costs and other constraints imposed. The MGARCH-IHMM consistently provides the largest Sharpe ratios with zero transaction costs but becomes dominated by the EW portfolio for positive transaction costs.

6.2 CRRA Utility

Next, we turn to the investor with constant relative risk-averse (CRRA) utility,

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

where $a \geq 0$. Relative risk aversion is a and absolute risk aversion is $\frac{a}{W}$. Unlike quadratic utility, expected utility will be sensitive to the shape of the distribution of wealth through returns. As such, the quality of density forecasts as measured by the log-predictive likelihoods may be a better rank of model inputs to portfolio choice.

As before, Tables 8 and 9 report the annualized fee that a CRRA investor will pay to move from the model in the first column to the MGARCH-IHMM over the out-of-sample period for both datasets. For almost all cases in Panel A of the tables, the investor is willing to pay a positive fee to obtain forecasts from the MGARCH-IHMM to make portfolio decisions. Compared to the quadratic utility case, the fees are generally larger. This may indicate the importance of the density beyond the summary moments of predictive mean and covariance in portfolio choice.

[7] Although the Sharpe ratio and expected quadratic utility are both functions of the mean and variance of a portfolio, they represent different trade-offs between risk and return. In addition, measures like the Sharpe ratio in the presence of non-Guassian distributions, which we have, can be inconsistent in ranking investments (Smetters and Zhang, 2013).

The fee to move from the nonparametric IHMM is larger than the parametric MGARCH models, indicating that capturing the strong persistence in the conditional second moments is vital to a good model and portfolio choice. We also see that the semiparametric MGARCH-DPM improves on the parametric MGARCH models, but an investor would still pay 0.62% to 2.25% for industry portfolios (Table 8) and 0.84% to 2.35% for factor portfolios (Table 9) to obtain forecasts from the MGARCH-IHMM. The simple EW and VW portfolios are generally dominated by the MGARCH-IHMM except for high transaction costs and higher risk aversion. When there are no transaction costs and $a = 2$, the fee can be very large, for example, 5.55% and 5.30% for the industry portfolios (Table 8) and 7.95% and 4.85% for the factor portfolios (Table 9) to move away from these simple portfolios.

Further imposing no leverage trading (Panel B) generally reduces all fees except for the high risk aversion case $a = 6$. The proposed MGARCH-IHMM dominates all other models except when risk aversion and transaction costs are both low, where the VW portfolio can be preferred but the MGARCH-IHMM is still better than other econometric models. In the factor portfolio application, VW is the dominating model while the MGARCH-IHMM performs the second best among other models, including the passive EW portfolio.

Table 10 reports Sharpe ratios for selected model portfolios for the no short-selling scenario. As in the quadratic utility case, the MGARCH-IHMM does well with no transaction costs, but with transaction costs the EW portfolio is preferred.[8]

With only no short-selling restriction, the optimal portfolio calls for active margin-buying for the CRRA investor. Figures 5 and 6 display the total weight allocated to risky assets for the MGARCH-IHMM in various settings. Leverage trading is a number above one and indicates borrowing at the risk-free rate to invest in risky assets. From the left side of these subfigures, we see the investor will borrow to invest most of the time, and that leverage can be high during some periods. For example, at the beginning of 1991 when $a = 2$, the investor borrows over 200% of their initial wealth to invest. This strategy can be risky when transaction costs are involved.

The right side of these subfigures shows the impact of the additional constraint of no leverage-trading on the risky weights. When there are no transaction costs, the investor can adjust their portfolio freely and it is optimal to only hold risky assets, without holding the risk-free asset. The exception to this is a bear market, namely during the Dotcom Bubble, the 2008 Financial Crisis, the US-China Trade War or the recent COVID-19 Crash. When transaction costs are imposed, instead of pursuing market timing aggressively, it is better to gradually change positions to lower transaction costs which we see for $c = 1$ and $c = 2$. 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.

7 Conclusions

This paper investigates the economic importance of nonparametrically/semiparametrically modelling the shape and the change of the unknown distribution of returns in portfolio allocation problems from a Bayesian perspective. Besides parametric multivariate GARCH (MGARCH) benchmark models for returns, we consider an infinite hidden Markov model (IHMM) and an MGARCH with innovations following a Dirichlet process mixture. We introduce a new Bayesian semiparametric model that combines the MGARCH component with the IHMM for innovations.

[8] See footnote 7 for problems with using the Sharpe ratio to rank investments with non-Gaussian distributions.

This new model nonparametrically approximates both the shape and evolution through time of the unknown distribution of returns beyond that captured by the MGARCH part.

The MGARCH-IHMM shows a clear advantage against the benchmark MGARCH-DPM, IHMM, MGARCH-N, and MGARCH-A models in terms of density forecasts as well as point forecasts for realized covariance. The MGARCH-IHMM captures distributional features such as changing conditional skewness and kurtosis.

Empirical results from a utility-based portfolio optimization show that a risk-averse investor, whose utility function is quadratic or CRRA, is willing to pay a positive fee that is economically significant for switching from the benchmark econometric models, including the passive EW portfolio, to the more sophisticated Bayesian semiparametric model MGARCH-IHMM in almost all tested scenarios, except for high risk aversion and high transaction cost when leverage is permitted. Without leverage, the MGARCH-IHMM still dominates alternative econometric models but the VW portfolio is preferred in many cases.

We conclude that Bayesian nonparametric/semiparametric models can lead to improved statistical forecast accuracy and economic gains for an investor. Portfolio choice is improved by modelling beyond the conditional second moments.

References

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

Bauwens, L., Preminger, A., and Rombouts, J. V. K. (2010). Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13(2):218–244.

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. (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. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75(1):79–97.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

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.

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

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

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

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*, 34(5):641–660.

Jin, X., Maheu, J. M., and Yang, Q. (2022). Infinite Markov pooling of predictive distributions. *Journal of Econometrics*, 228(2):302–321.

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

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.

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.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530.

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.

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.

Smetters, K. and Zhang, X. (2013). A sharper ratio: A general measure for correctly ranking non-normal investment risks. NBER Working Paper No. w19500, SSRN: https://ssrn.com/abstract=2336365.

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.

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 Portfolio Returns

A: Industry Portfolios

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

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

B: Factor Portfolios

Factor	Mean	Median	StDev	Skewness	Ex.Kurtosis	Min	Max
MKT	0.8261	1.2571	4.5080	-0.7309	2.4547	-25.6700	15.3665
SMB	0.1689	0.0850	3.0179	0.1192	2.8072	-16.6291	16.7885
HML	0.2467	0.2247	2.9889	-0.1227	2.4310	-14.9312	12.0003
RMW	0.2580	0.2447	2.2332	-0.7397	13.5542	-20.6409	12.2837
CMA	0.2510	0.0900	2.0673	0.1601	1.4160	-7.4939	8.6820

Correlations	MKT	SMB	HML	RMW	CMA
MKT	1.0000	0.2867	-0.1961	-0.1873	-0.3568
SMB	0.2867	1.0000	-0.0094	-0.3553	-0.0915
HML	-0.1961	-0.0094	1.0000	0.0826	0.6812
RMW	-0.1873	-0.3553	0.0826	1.0000	-0.0192
CMA	-0.3568	-0.0915	0.6812	-0.0192	1.0000

1. Source: Kenneth French's Data Library.
2. Fama-French 5 industry portfolios, log-returns in percentage, July 1926 to December 2020, (1,134 observations).
3. Fama-French 5 factor portfolios, log-returns in percentage, July 1963 to December 2023 (726 observations).

Table 2: Posterior Estimates for Industry Portfolios

MGARCH-IHMM:

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

MGARCH-DPM:

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

IHMM:

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

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:

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

	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 3: Posterior Estimates for Factor Portfolios

MGARCH-IHMM:

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

MGARCH-DPM:

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

IHMM:

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

MGARCH-N:

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

MGARCH-A:

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

	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							1.0791	(0.7875,1.3692)	0.8868	(0.5946,1.1759)
μ_2							0.1803	(0.0063,0.3685)	0.1757	(-0.0217,0.3753)
μ_3							0.1155	(-0.0515,0.2600)	0.1229	(-0.0451,0.2931)
μ_4							0.2208	(0.1123,0.3310)	0.1919	(0.0781,0.3035)
μ_5							0.1327	(0.0036,0.2536)	0.1680	(0.0433,0.2936)
α_1	0.2624	(0.2064, 0.3176)	0.2837	(0.2296,0.3348)			0.2836	(0.2353,0.3254)	0.2658	(0.2232,0.3067)
α_2	0.1684	(0.1179, 0.2271)	0.2228	(0.1546,0.2814)			0.2576	(0.2144,0.3154)	0.2923	(0.2227,0.3444)
α_3	0.2770	(0.2474, 0.3101)	0.2899	(0.2618,0.3283)			0.3138	(0.2806,0.3538)	0.3036	(0.2848,0.3234)
α_4	0.3465	(0.3169, 0.3691)	0.4010	(0.3730,0.4185)			0.3673	(0.3442,0.3977)	0.3500	(0.3131,0.3825)
α_5	0.3140	(0.2903, 0.3477)	0.3380	(0.3089,0.3631)			0.3099	(0.2860,0.3346)	0.3037	(0.2786,0.3350)
β_1	0.9063	(0.8486, 0.9419)	0.8982	(0.8454,0.9384)			0.9176	(0.8827,0.9447)	0.9185	(0.8792,0.9457)
β_2	0.9540	(0.8885, 0.9818)	0.9305	(0.8694,0.9691)			0.9211	(0.8768,0.9521)	0.8997	(0.8510,0.9436)
β_3	0.9338	(0.9059, 0.9522)	0.9300	(0.9050,0.9480)			0.9312	(0.9094,0.9476)	0.9365	(0.9240,0.9467)
β_4	0.9155	(0.8985, 0.9303)	0.8924	(0.8744,0.9077)			0.9141	(0.8960,0.9265)	0.9223	(0.9049,0.9399)
β_5	0.9162	(0.8934, 0.9360)	0.9060	(0.8810,0.9252)			0.9261	(0.9097,0.9401)	0.9268	(0.9078,0.9422)
η_1	3.0921	(1.9440, 4.1389)	1.7665	(1.7665,4.3057)					2.0661	(1.2499,2.8341)
η_2	1.8429	(0.9898, 2.5481)	0.3854	(0.3854,2.2283)					0.4190	(-0.0757,0.8550)
η_3	0.1893	(-0.4190, 0.6982)	-0.0983	(-0.0983,0.7245)					0.3110	(0.0312,0.6278)
η_4	0.2483	(-0.0980, 0.5901)	0.0443	(0.0443,0.5702)					0.3423	(0.0867,0.5779)
η_5	-0.1238	(-0.5701, 0.2141)	-0.3739	(-0.3739,0.2424)					-0.0261	(-0.3114,0.2281)
α_0	1.4730	(0.8003, 2.4805)			0.6277	(0.3496, 0.9934)				
β_0	0.8972	(0.3878, 1.6222)	0.4580	(0.1346,0.9423)	0.7942	(0.3414, 1.4409)				
K	9.3923	(7.0000,13.0000)	5.6217	(3.0000,9.0000)	7.8127	(7.0000,10.0000)				

Table 4: Out-of-Sample Forecasts

Panel A: Industry portfolios

Model	log PL	log BF	RMSFE	
			Pred Mean	Pred Cov
MGARCH-IHMM	**-4478.602**	—	4.1480	**93.537**
MGARCH-DPM	-4492.075	13.4731	4.1487	99.603
IHMM	-4550.889	72.2869	**4.1340**	144.611
MGARCH-A	-4521.526	42.9239	4.1493	100.984
MGARCH-N	-4523.410	44.8084	4.1508	101.391

Panel B: Factor portfolios

Model	log PL	log BF	RMSFE	
			Pred Mean	Pred Cov
MGARCH-IHMM	**-3523.930**	—	6.4685	**36.4118**
MGARCH-DPM	-3538.222	14.2917	6.4680	36.6132
IHMM	-3617.323	93.3930	6.4801	50.1767
MGARCH-A	-3556.786	32.8565	6.4686	37.4624
MGARCH-N	-3561.274	37.3438	**6.4649**	37.7417

1. Fama-French 5 industry portfolio returns start in July 1926, the out-of-sample period is from January 1991 to December 2020.
2. Fama-French 5 factor portfolio returns start in July 1963, the out-of-sample period is from January 1999 to December 2023.
3. Log PL is the log-predictive likelihood, Log BF is the log-Bayes factor for the MGARCH-IHMM against the model in the first column.
4. RMSFE is the mean of the L^2 norm of the vector forecast error for the predictive mean of $\boldsymbol{r}_t$, while it is the mean of the spectral norm of the matrix difference between the predictive covariance and the monthly realized covariance.

Table 5: Annualized Fees a Quadratic Investor Is Willing to Pay for the MGARCH-IHMM, Industry Portfolios

$$U(W) = W - \frac{a}{2(1+a)}W^2,$$
$$W_t = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t + (1-\boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t} - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}),$$
$$C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}) = c\sum_{i=1}^{N}|w_{i,t} - (1+R_{i,t-1})w_{i,t-1}| + c\left|(1-\boldsymbol{w}_t'\boldsymbol{\iota}) - (1+R_{f,t-1})(1-\boldsymbol{w}_{t-1}'\boldsymbol{\iota})\right|$$

a	$a=2$			$a=4$			$a=6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Panel A: No Short-Selling									
MGARCH-DPM	1.31%	1.75%	2.01%	0.98%	1.03%	1.39%	0.63%	0.71%	0.87%
IHMM	3.18%	3.28%	3.35%	2.04%	2.46%	2.72%	1.11%	1.59%	1.81%
MGARCH-A	2.24%	2.57%	2.87%	1.72%	1.98%	2.23%	0.68%	1.15%	1.63%
MGARCH-N	2.27%	2.59%	2.88%	1.74%	2.04%	2.25%	0.73%	1.17%	1.67%
EW	5.35%	2.22%	1.61%	1.14%	0.40%	0.08%	1.10%	0.19%	0.06%
VW	5.10%	2.01%	1.44%	0.79%	0.08%	-0.20%	0.63%	-0.24%	-0.33%
Panel B: No Short-Selling and No Leverage-Trading									
MGARCH-DPM	0.00%	0.27%	0.79%	0.06%	1.74%	2.66%	0.10%	2.30%	2.67%
IHMM	0.18%	0.73%	1.22%	0.34%	2.37%	2.68%	0.49%	2.68%	3.03%
MGARCH-A	0.02%	0.35%	0.84%	0.15%	1.85%	2.48%	0.38%	2.35%	2.70%
MGARCH-N	0.03%	0.43%	0.86%	0.18%	1.92%	2.55%	0.42%	2.39%	2.76%
EW	-0.43%	-0.10%	0.12%	0.04%	0.19%	0.45%	0.30%	0.61%	0.85%
VW	-0.66%	-0.30%	-0.04%	-0.31%	-0.12%	0.17%	-0.16%	0.19%	0.46%

1. 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 Leverage-Trading indicates that a constraint of $\sum_{i=1}^{N} w_{i,t} \leq 1$ is 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 model in column one is better.
4. EW indicates a buy-and-hold strategy of an equally-weighted portfolio that consists of the Fama-French 5 industry portfolios. VW indicates a buy-and-hold strategy of a value-weighted portfolio.

Table 6: Annualized Fee a Quadratic Investor Is Willing to Pay for the MGARCH-IHMM, Factor Portfolios

$$U(W) = W - \frac{a}{2(1+a)}W^2,$$
$$W_t = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t + (1 - \boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t} - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}),$$
$$C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}) = c\sum_{i=1}^{N} |w_{i,t} - (1 + R_{i,t-1})w_{i,t-1}| + c\left|(1 - \boldsymbol{w}_t'\boldsymbol{\iota}) - (1 + R_{f,t-1})(1 - \boldsymbol{w}_{t-1}'\boldsymbol{\iota})\right|.$$

a	$a=2$			$a=4$			$a=6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Panel A: No Short-Selling									
MGARCH-DPM	1.32%	1.90%	2.63%	1.10%	1.54%	1.72%	0.66%	0.85%	1.15%
IHMM	3.41%	3.80%	4.15%	2.51%	2.95%	3.49%	1.71%	2.10%	2.42%
MGARCH-A	2.48%	3.23%	3.75%	1.86%	2.50%	2.72%	1.21%	1.75%	2.09%
MGARCH-N	2.49%	3.28%	3.78%	1.92%	2.53%	2.75%	1.24%	1.77%	2.10%
EW	7.77%	3.41%	2.89%	3.18%	2.04%	1.55%	2.08%	1.06%	0.40%
VW	4.57%	0.49%	0.07%	0.92%	-0.09%	-0.51%	0.62%	-0.30%	-0.86%
Panel B: No Short-Selling and No Leverage-Trading									
MGARCH-DPM	0.08%	0.31%	0.47%	0.22%	0.61%	1.45%	0.39%	1.23%	1.83%
IHMM	0.46%	0.75%	2.11%	1.00%	1.80%	2.44%	1.06%	2.12%	2.95%
MGARCH-A	0.20%	0.47%	0.80%	0.52%	0.80%	1.99%	0.84%	1.25%	2.08%
MGARCH-N	0.22%	0.51%	0.85%	0.53%	0.81%	2.01%	0.86%	1.28%	2.09%
EW	1.00%	1.45%	1.49%	0.97%	1.15%	1.26%	0.50%	0.68%	1.35%
VW	-1.94%	-1.42%	-1.29%	-1.22%	-0.96%	-0.77%	-0.92%	-0.66%	0.07%

1. 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 Leverage-Trading indicates that a constraint of $\sum_{i=1}^{N} w_{i,t} \leq 1$ is 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 model in column one is better.
4. EW indicates a buy-and-hold strategy of an equally-weighted portfolio that consists of the Fama-French 5 factor portfolios. VW indicates a buy-and-hold strategy of a value-weighted portfolio.

Table 7: Ex-Post Sharpe Ratios of a Quadratic Investor's Optimal Portfolio after Transaction Costs, No Short-Selling

$$U(W) = W - \frac{a}{2(1+a)}W^2,$$
$$W_t = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t + (1 - \boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t} - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}),$$
$$C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}) = c\sum_{i=1}^{N} |w_{i,t} - (1 + R_{i,t-1})w_{i,t-1}| + c\left|(1 - \boldsymbol{w}_t'\boldsymbol{\iota}) - (1 + R_{f,t-1})(1 - \boldsymbol{w}_{t-1}'\boldsymbol{\iota})\right|.$$

a	$a = 2$			$a = 4$			$a = 6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Industry Portfolios									
MGARCH-IHMM	**0.2123**	-0.1383	-0.2028	**0.2155**	-0.1164	-1.0968	**0.2162**	-0.7675	-0.3887
MGARCH-DPM	0.2022	-0.5813	-0.3589	0.1995	-0.1154	-0.2458	0.2006	-0.2286	-0.4861
IHMM	0.1862	-0.1488	-0.3766	0.1822	-0.1574	-0.2383	0.1920	-0.1314	-0.1383
MGARCH-A	0.1963	-0.0858	-0.3711	0.1896	-0.3087	-0.4522	0.1992	-0.1312	-0.5571
MGARCH-N	0.1959	-0.4114	-0.4391	0.1891	-0.2060	-0.4555	0.1978	-0.1379	-0.5261
EW	0.1980	**0.1482**	**0.0985**	0.1980	**0.1482**	**0.0985**	0.1980	**0.1482**	**0.0985**
VW	0.2061	-0.1599	-0.4861	0.2061	-0.1599	-0.4861	0.2061	-0.1599	-0.4861
Factor Portfolios									
MGARCH-IHMM	**0.1882**	-0.7012	-0.5339	**0.1807**	-0.7668	-0.9574	**0.1883**	-0.6929	-0.7952
MGARCH-DPM	0.1651	-0.6566	-0.3600	0.1762	-0.9096	-1.1132	0.1788	-0.6764	-0.2492
IHMM	0.1464	-0.3517	-0.3805	0.1426	-0.3420	-0.1605	0.1793	-0.2201	-0.4421
MGARCH-A	0.1534	-0.8095	-0.7049	0.1586	-0.7137	-0.6996	0.1578	-0.5362	-0.4382
MGARCH-N	0.1740	-0.5589	-0.3231	0.1634	-0.9338	-0.4815	0.1602	-0.6906	-0.9790
EW	0.1223	**0.0314**	**-0.0568**	0.1223	**0.0314**	**-0.0568**	0.1223	**0.0314**	**-0.0568**
VW	0.1704	-0.3614	-0.7596	0.1704	-0.3614	-0.7596	0.1704	-0.3614	-0.7596

1. No Short-Selling indicates a constraint of $w_{i,t} \geq 0$ for all $i = 1, \ldots, N$ is imposed in the optimization problem.
2. The number indicates the ex-post Sharpe ratio as the result of the realized returns from the optimized weights for a model. Bold entries are the largest Sharpe ratio in a column.
3. EW indicates a buy-and-hold strategy of an equally-weighted portfolio that consists of the 5 assets. VW indicates a buy-and-hold strategy of a value-weighted portfolio.

Table 8: Annualized Fees a CRRA Investor Is Willing to Pay for the MGARCH-IHMM, Industry Portfolios

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

a	$a = 2$			$a = 4$			$a = 6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Panel A: No Short-Selling									
MGARCH-DPM	1.58%	2.02%	2.25%	1.06%	1.38%	1.50%	0.62%	0.85%	0.91%
IHMM	3.15%	3.30%	3.68%	2.27%	2.58%	2.90%	1.24%	1.77%	1.92%
MGARCH-A	2.56%	2.90%	3.14%	1.80%	2.10%	2.44%	0.87%	1.28%	1.74%
MGARCH-N	2.59%	2.93%	3.15%	1.84%	2.15%	2.48%	0.90%	1.32%	1.75%
EW	5.55%	2.64%	1.88%	1.49%	0.58%	0.22%	1.37%	0.62%	0.16%
VW	5.30%	2.42%	1.70%	1.10%	0.23%	-0.10%	0.82%	0.11%	-0.31%
Panel B: No Short-Selling and No Leverage-Trading									
MGARCH-DPM	0.12%	0.27%	0.54%	0.26%	1.12%	1.59%	0.39%	1.37%	1.82%
IHMM	0.47%	1.26%	1.97%	1.04%	1.44%	2.72%	1.74%	2.50%	2.88%
MGARCH-A	0.30%	1.02%	1.38%	0.85%	1.33%	2.56%	1.22%	2.37%	2.64%
MGARCH-N	0.31%	1.02%	1.41%	0.86%	1.37%	2.58%	1.28%	2.38%	2.65%
EW	-0.26%	0.08%	0.25%	0.50%	0.60%	0.63%	0.70%	0.81%	1.08%
VW	-0.50%	-0.13%	0.07%	0.12%	0.25%	0.32%	0.16%	0.30%	0.60%

1. 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 Leverage-Trading indicates that a constraint of $\sum_{i=1}^{N} w_{i,t} \leq 1$ is 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 model in column one is better.
4. EW indicates a buy-and-hold strategy of an equally-weighted portfolio that consists of the Fama-French 5 industry portfolios. VW indicates a buy-and-hold strategy of a value-weighted portfolio.

Table 9: Annualized Fee a CRRA Investor Is Willing to Pay for the MGARCH-IHMM, Factor portfolios

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

a	$a = 2$			$a = 4$			$a = 6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Panel A: No Short-Selling									
MGARCH-DPM	1.51%	1.98%	2.35%	1.38%	1.73%	1.86%	0.84%	1.12%	1.33%
IHMM	3.70%	3.99%	4.33%	2.71%	3.10%	3.69%	1.93%	2.37%	2.81%
MGARCH-A	2.72%	3.37%	3.93%	2.18%	2.65%	2.94%	1.49%	1.94%	2.26%
MGARCH-N	2.74%	3.37%	3.97%	2.22%	2.71%	2.96%	1.53%	1.99%	2.29%
EW	7.95%	3.67%	3.06%	3.22%	2.16%	1.74%	2.37%	1.12%	0.60%
VW	4.85%	0.76%	0.24%	1.00%	0.04%	-0.29%	0.96%	-0.20%	-0.63%
Panel B: No Short-Selling and No Leverage-Trading									
MGARCH-DPM	0.18%	0.43%	0.71%	0.32%	0.86%	1.61%	0.55%	1.36%	1.95%
IHMM	0.56%	1.06%	2.22%	1.14%	1.93%	2.56%	1.21%	2.28%	3.13%
MGARCH-A	0.28%	0.62%	1.08%	0.67%	1.02%	2.13%	1.00%	1.41%	2.36%
MGARCH-N	0.30%	0.67%	1.09%	0.71%	1.06%	2.18%	1.03%	1.46%	2.39%
EW	1.31%	1.48%	1.52%	1.09%	1.22%	1.38%	0.67%	1.23%	1.52%
VW	-1.63%	-1.38%	-1.26%	-1.10%	-0.88%	-0.65%	-0.73%	-0.10%	0.28%

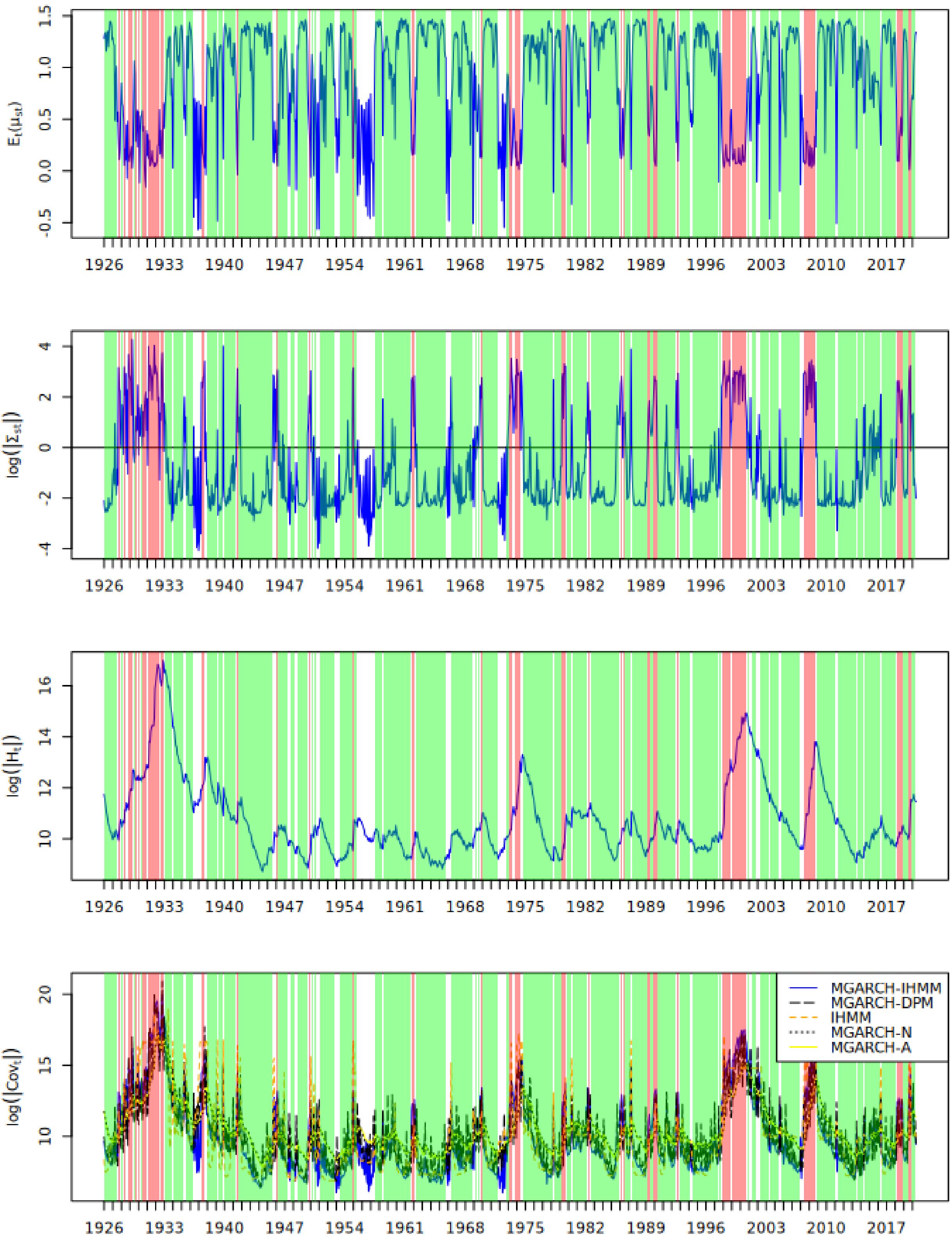
1. No Short-Selling indicates a constraint of $w_{i,t} \geq 0$ for all $i = 1, \ldots, N$ is imposed in the optimization problem; and No Leverage-Trading indicates that a constraint of $\sum_{i=1}^{N} w_{i,t} \leq 1$ is imposed in the optimization problem.
2. IHMM means an investor switches from IHMM to MGARCH-IHMM.
3. A positive fee means the MGARCH-IHMM is better, and a negative fee means the corresponding model in column one is better.
4. EW indicates a buy-and-hold strategy of an equally-weighted portfolio that consists of the Fama-French 5 factor portfolios. VW indicates a buy-and-hold strategy of a value-weighted portfolio.

Table 10: Ex-Post Sharpe Ratios of a CRRA Investor's Optimal Portfolio after Transaction Costs, No Short-Selling

$$U(W) = W - \frac{a}{2(1+a)}W^2,$$
$$W_t = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t + (1 - \boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t} - C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}),$$
$$C(\boldsymbol{w}_t, \boldsymbol{w}_{t-1}) = c\sum_{i=1}^{N} |w_{i,t} - (1 + R_{i,t-1})w_{i,t-1}| + c\left|(1 - \boldsymbol{w}_t'\boldsymbol{\iota}) - (1 + R_{f,t-1})(1 - \boldsymbol{w}_{t-1}'\boldsymbol{\iota})\right|.$$

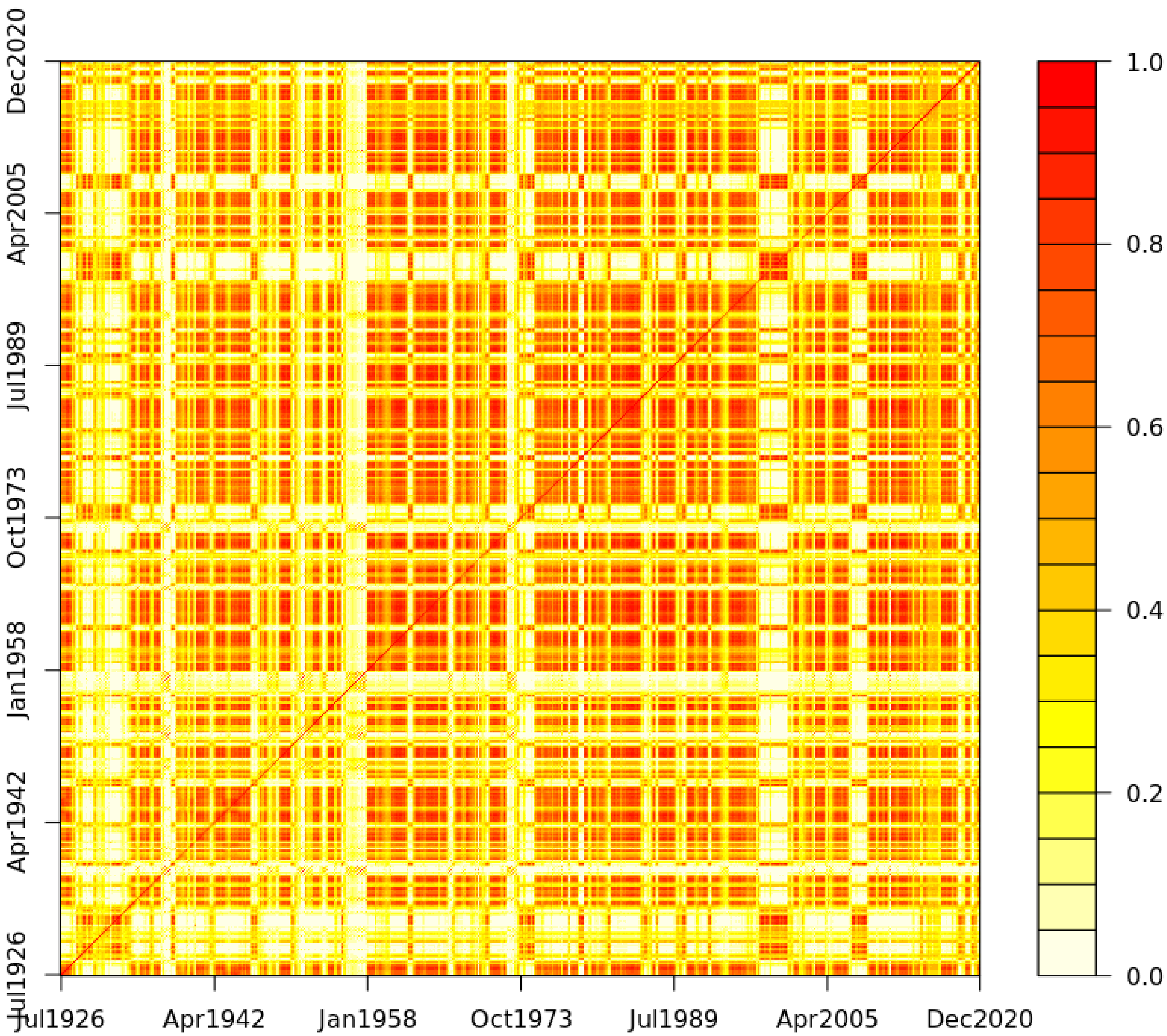
a	$a=2$			$a=4$			$a=6$		
c	0%	1%	2%	0%	1%	2%	0%	1%	2%
Industry Portfolios									
MGARCH-IHMM	**0.2151**	-0.2831	-0.3643	**0.2197**	-0.1776	-0.3474	**0.2142**	-0.1186	-0.8108
MGARCH-DPM	0.2043	-0.2568	-0.3506	0.2041	-0.3378	-0.3617	0.1999	-0.7956	-0.7721
IHMM	0.1889	-0.2019	-0.3613	0.1824	-0.2275	-0.4770	0.1840	-0.6579	-0.4975
MGARCH-A	0.1950	-0.1408	-0.1769	0.1926	-0.2452	-0.4214	0.1961	-0.0867	-0.5941
MGARCH-N	0.1947	-0.0541	-0.3872	0.1922	-0.8305	-0.4851	0.1951	-0.0852	-0.5997
EW	0.1980	**0.1482**	**0.0985**	0.1980	**0.1482**	**0.0985**	0.1980	**0.1482**	**0.0985**
VW	0.2061	-0.1599	-0.4861	0.2061	-0.1599	-0.4861	0.2061	-0.1599	-0.4861
Factor Portfolios									
MGARCH-IHMM	**0.1806**	-0.7437	-0.6260	**0.1843**	-0.5142	-0.8693	**0.1981**	-0.6885	-0.9685
MGARCH-DPM	0.1793	-0.8911	-0.6069	0.1636	-0.5412	-0.7679	0.1843	-0.1135	-0.6647
IHMM	0.1659	-0.9040	-0.9695	0.1505	-0.8335	-1.0381	0.1930	-0.4782	-1.0340
MGARCH-A	0.1581	-0.3220	-0.4846	0.1498	-0.6525	-0.4867	0.1751	-0.9042	-0.9935
MGARCH-N	0.1583	-0.3327	-0.5262	0.1422	-0.8489	-0.4749	0.1706	-0.7004	-0.9842
EW	0.1223	**0.0314**	**-0.0568**	0.1223	**0.0314**	**-0.0568**	0.1223	**0.0314**	**-0.0568**
VW	0.1704	-0.3614	-0.7596	0.1704	-0.3614	-0.7596	0.1704	-0.3614	-0.7596

1. No Short-Selling indicates a constraint of $w_{i,t} \geq 0$ for all $i = 1, \ldots, N$ is imposed in the optimization problem.
2. The number indicates the ex-post Sharpe ratio as the result of the realized returns from the optimized weights for a model. Bold entries are the largest Sharpe ratio in a column.
3. EW indicates a buy-and-hold strategy of an equally-weighted portfolio that consists of the 5 assets. VW indicates a buy-and-hold strategy of a value-weighted portfolio.



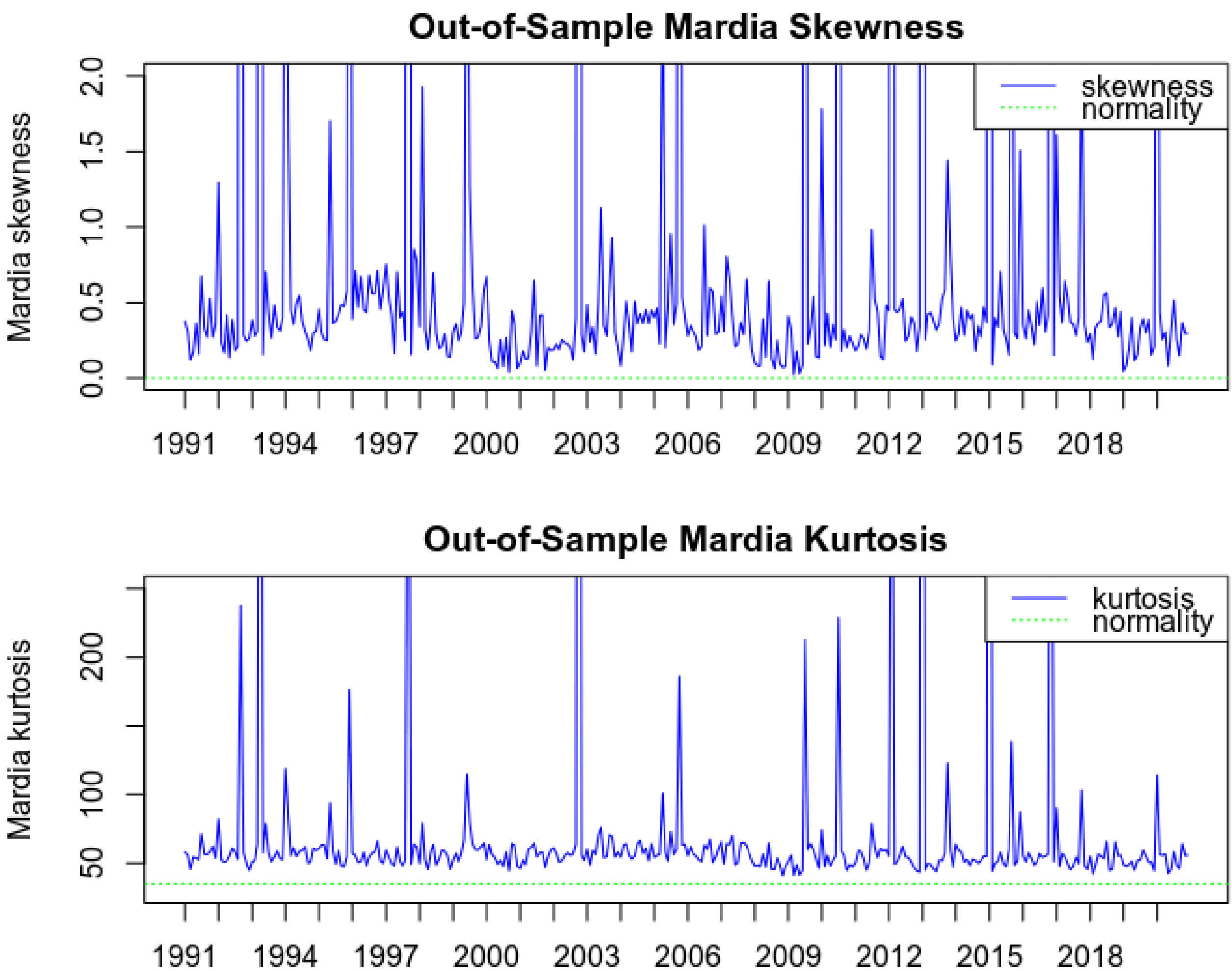
$$\text{MGARCH-IHMM, MGARCH-DPM: } \mathrm{Cov}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2'},$$
$$\text{IHMM: } \mathrm{Cov}_t = \boldsymbol{\Sigma}_{s_t}, \quad \text{MGARCH-N, MGARCH-A: } \mathrm{Cov}_t = \boldsymbol{H}_t$$

Figure 1: Posterior Means of the Time-Varying Parameters over Time, Industry Portfolios



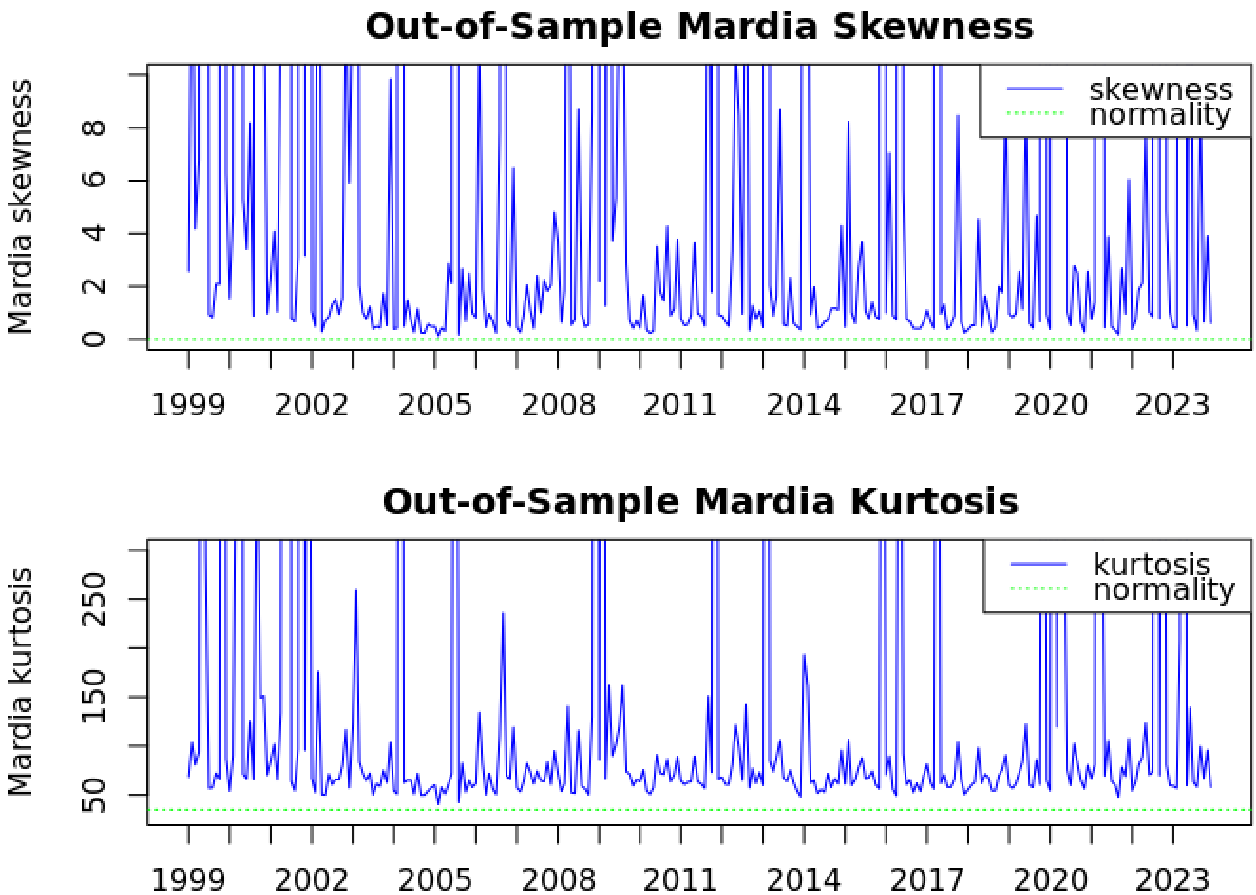
The redder the colour, the higher probability that the periods on the horizontal and vertical axes share the same state.

Figure 2: Heat Map of States for MGARCH-IHMM, Industry Portfolios



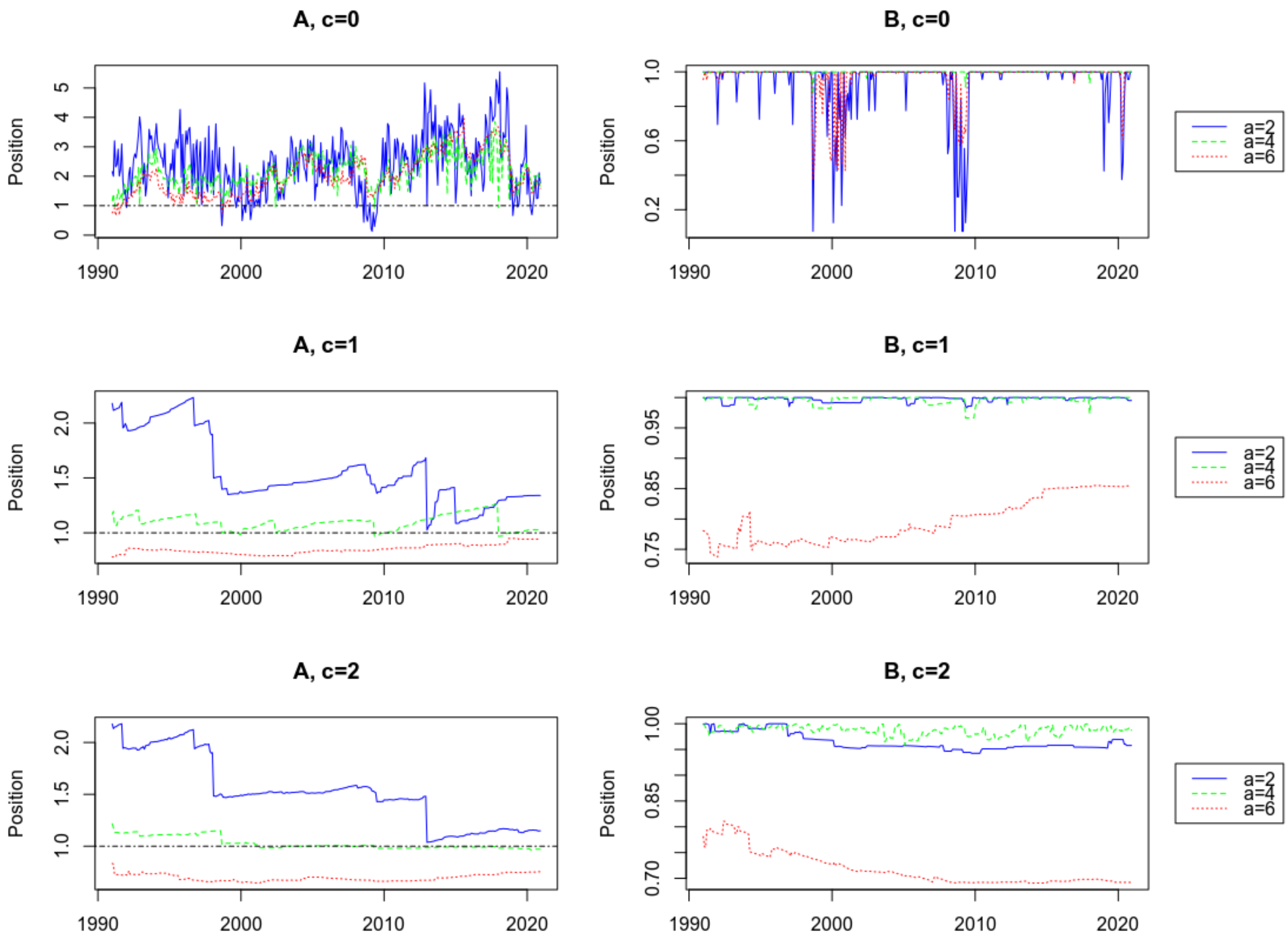
The Mardia skewness and kurtosis measures are defined as $skew = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[(x_i - \bar{x})V^{-1}(x_j - \bar{x})\right]^3$ and $kurt = \frac{1}{N}\sum_{i=1}^{N}\left[(x_i - \bar{x})V^{-1}(x_i - \bar{x})\right]^2$ where $V = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})'$ for N observations of the $p \times 1$ vector x_i.

Figure 3: Mardia Skewness and Kurtosis, Industry Portfolios



The Mardia skewness and kurtosis measures are defined as $skew = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[(x_i - \bar{x})V^{-1}(x_j - \bar{x})\right]^3$ and $kurt = \frac{1}{N}\sum_{i=1}^{N}\left[(x_i - \bar{x})V^{-1}(x_i - \bar{x})\right]^2$ where $V = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})'$ for N observations of the $p \times 1$ vector x_i.

Figure 4: Mardia Skewness and Kurtosis, Factor Portfolios



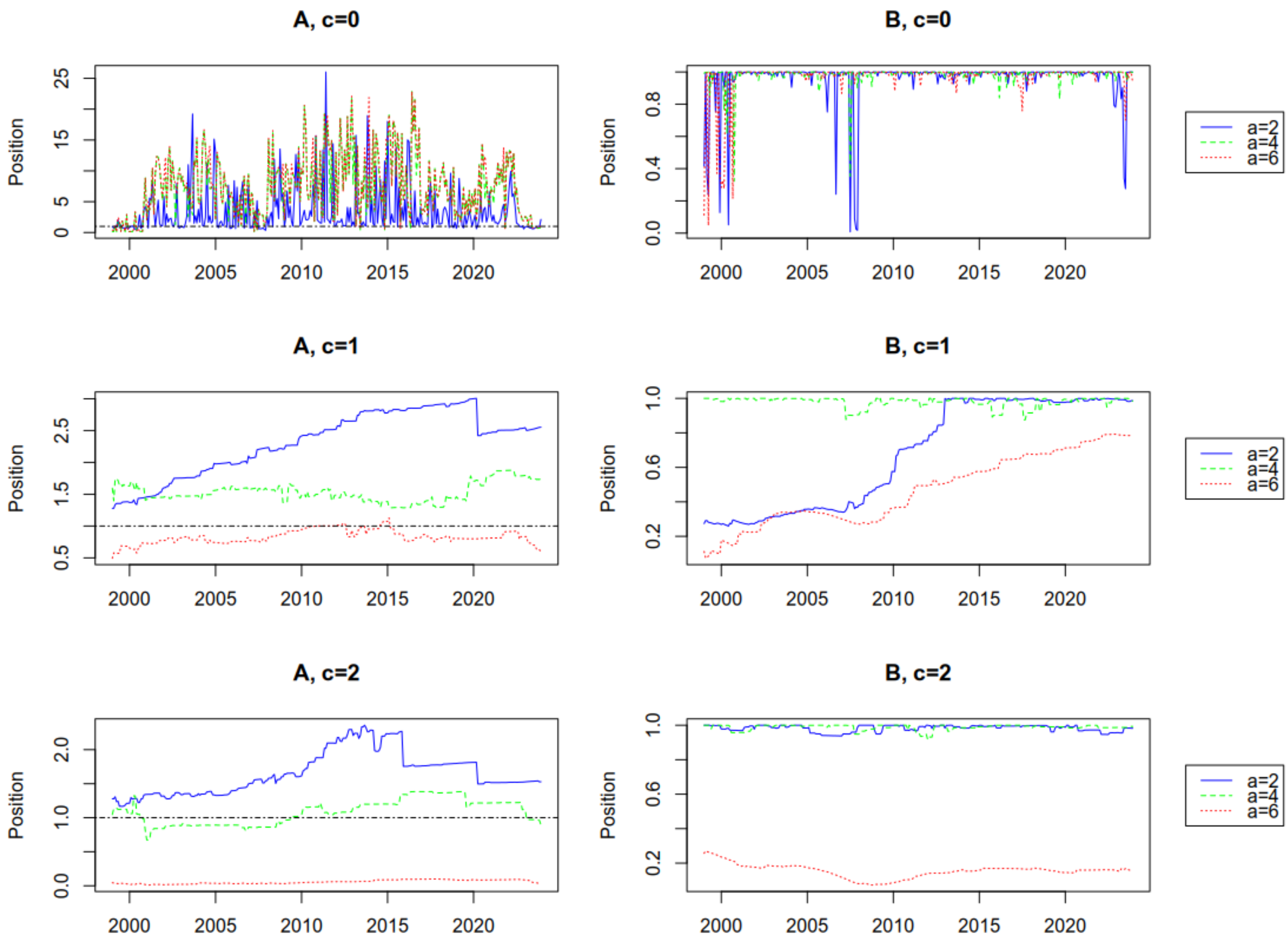
CRRA Utility

A: No Short-Selling

B: No Short-Selling and No Leverage-Trading

Net risky position indicates the summation of the weights of the risky assets in a portfolio ($\sum_{i=1}^{N} w_{t,i}$)

Figure 5: Net Risky Positions in the Portfolio Optimized with MGARCH-IHMM over Time, Industry Portfolios



CRRA Utility

A: No Short-Selling

B: No Short-Selling and No Leverage-Trading

Net risky position indicates the summation of the weights of the risky assets in a portfolio ($\sum_{i=1}^{N} w_{t,i}$)

Figure 6: Net Risky Positions in the Portfolio Optimized with MGARCH-IHMM over Time, Factor Portfolios

Appendix A Proof of Covariance Targeting of $\boldsymbol{CC}'$

In a reduced form of the MGARCH-IHMM, where $\boldsymbol{\Sigma}_t = \boldsymbol{I}$ for all t, referring to equation (5e), the unconditional expectation of $\boldsymbol{H}_t$ is

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

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

$$\boldsymbol{CC}' = \bar{\boldsymbol{H}} \odot \left[\mathbf{1} - \boldsymbol{\alpha\alpha}' - \boldsymbol{\beta\beta}'\right] - \boldsymbol{\alpha\alpha}' \odot \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)\left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)',$$

where $\mathbf{1}$ is an $N \times N$ matrix with all the elements being 1.

Appendix B Posterior Sampling Details

Recall that $\boldsymbol{\Gamma} = \left(\gamma_1, \ldots, \gamma_K, \gamma_R\right)'$ and $\boldsymbol{\Pi}_j = \left(\pi_{j1}, \ldots, \pi_{jK}, \pi_{jR}\right)$, 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 $u_{1:T} | \boldsymbol{\Gamma}, \boldsymbol{\Pi}$. The auxiliary slice variable $U = \{u_t\}_{t=1}^{T}$ 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
$$\min\{u_t\}_{t=1}^{T} > \max\{\pi_{jR}\}_{j=1}^{K} \tag{14}$$
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, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Pi}$ also need to be updated accordingly:

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

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

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

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

   Repeat the above steps until inequality (14) holds.

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

   (a) The prediction step for the initial state s_1 is as follows:
$$p(s_1 = k | u_1, \boldsymbol{\Gamma}) \propto \mathbb{1}\left(u_1 < \gamma_k\right), \qquad k = 1, \ldots, K \tag{15}$$
   for the following states $s_{2:T}$:
$$p(s_t = k | \boldsymbol{r}_{1:t-1}, u_{1:t}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t-1}) \propto \sum_{j=1}^{K} \mathbb{1}\left(u_t < \pi_{jk}\right) p\left(s_{t-1} = j | \boldsymbol{r}_{1:t-1}, u_{1:t-1}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t-1}\right) \tag{16}$$

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

$$p\left(s_t = k | \boldsymbol{r}_{1:t}, u_{1:t}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \propto p\left(\boldsymbol{r}_t | \boldsymbol{r}_{t-1}, \boldsymbol{\Theta}_k, \boldsymbol{H}_t\right) p\left(s_t = k | \boldsymbol{r}_{1:t-1}, u_{1:t}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t-1}\right) \tag{17}$$

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,

$$p\left(s_t = k | s_{t+1} = j, \boldsymbol{r}_{1:t}, u_{1:t+1}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \propto \mathbb{1}\left(u_{t+1} < \pi_{kj}\right) p\left(s_t = k | \boldsymbol{r}_{1:t}, u_{1:t}, \boldsymbol{\Pi}, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \tag{18}$$

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 of different colours in the "oracle" urn. Fox et al. (2011) propose simulating c_k from the hierarchical 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 towards 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 the Dirichlet process, the conjugate posterior is

$$\boldsymbol{\Gamma} | c_{1:K}, \beta_0 \sim Dir\left(c_1, \ldots, c_K, \beta_0\right) \tag{19}$$

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

$$\boldsymbol{\Pi}_j | n_{j,1:K}, \boldsymbol{\Gamma}, \alpha \sim Dir\left(\alpha_0 \gamma_1 + n_{j1}, \ldots, \alpha_0 \gamma_K + n_{jK}, \alpha_0 \gamma_R\right) \tag{20}$$

10. Sample $\boldsymbol{\Theta}|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{H}_{1:T}$. Assume conjugate priors $\boldsymbol{\mu} \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0)$ and $\boldsymbol{\Sigma} \sim IW(\boldsymbol{\Sigma}_0, \nu + N)$. 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

$$\boldsymbol{Y}_k = \boldsymbol{X}_k \boldsymbol{\mu}_k + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_k) \tag{21}$$

The posteriors are

$$p(\boldsymbol{\mu}_k | \boldsymbol{Y}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_{1:T}) \sim \prod_{t:s_t=k} p(\boldsymbol{Y}_t | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_t)\, p(\boldsymbol{\mu}_k) \tag{22}$$

$$\sim N(\boldsymbol{M}_\mu, \boldsymbol{V}_\mu) \tag{23}$$

where

$$\boldsymbol{M}_\mu = \boldsymbol{V}_\mu \left(\sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{H}_t^{-1/2} \boldsymbol{r}_t + \boldsymbol{B}_0^{-1} \boldsymbol{b}_0\right) \tag{24}$$

$$\boldsymbol{V}_\mu = \left(\sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{H}_t^{-1/2} + \boldsymbol{B}_0^{-1}\right)^{-1} \tag{25}$$

and

$$p(\boldsymbol{\Sigma}_k | \boldsymbol{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{H}_{1:T}) \propto \prod_{t:s_t=k} p(\boldsymbol{r}_t | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_t)\, p(\boldsymbol{\Sigma}_k) \tag{26}$$

$$\sim IW\left(\bar{\boldsymbol{\Sigma}}, \bar{\nu} + N\right) \tag{27}$$

where

$$\bar{\nu} = T_k + \nu = \sum_{t=1}^{T} \mathbb{1}(s_t = k) + \nu \tag{28}$$

$$\bar{\boldsymbol{\Sigma}} = \sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2} (\boldsymbol{r}_t - \boldsymbol{\mu}_k)(\boldsymbol{r}_t - \boldsymbol{\mu}_k)' \boldsymbol{H}_t^{-1/2'} + \boldsymbol{\Sigma}_0 \tag{29}$$

11. Sample hierarchical priors.

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

$$\boldsymbol{\mu}_b = \boldsymbol{\Sigma}_b \left(\boldsymbol{B}_0^{-1} \sum_{k=1}^{K} \boldsymbol{\mu}_k + \boldsymbol{H}_0^{-1} \boldsymbol{h}_0\right) \tag{30}$$

$$\boldsymbol{\Sigma}_b = \left(K \boldsymbol{B}_0^{-1} + \boldsymbol{H}_0^{-1}\right)^{-1} \tag{31}$$

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

$$\omega_b = K + a_0 \tag{32}$$

$$\boldsymbol{\Omega}_B = \sum_{k=1}^{K} (\boldsymbol{\mu}_k - \boldsymbol{b}_0)(\boldsymbol{\mu}_k - \boldsymbol{b}_0)' + \boldsymbol{A}_0 \tag{33}$$

(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):

$$\nu'|\nu \sim Gamma\left(\tau, \frac{\tau}{\nu}\right) \tag{34}$$

and the acceptance rate is

$$\min\left\{1, \frac{p\left(\nu'|\boldsymbol{\Sigma}_{1:K}, s_0, g_0\right)/q\left(\nu'|\nu\right)}{p\left(\nu|\boldsymbol{\Sigma}_{1:K}, s_0, g_0\right)/q\left(\nu|\nu'\right)}\right\} \tag{35}$$

(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

$$\boldsymbol{C}_s = \left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1} + \boldsymbol{C}_0^{-1}\right)^{-1} \tag{36}$$

$$d_s = K\left(\nu + N\right) + d_0 \tag{37}$$

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

$$p\left(\boldsymbol{\theta}_H|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{\Theta}\right) \sim \prod_{t=1}^{T} p\left(\boldsymbol{r}_t|\boldsymbol{\Theta}, \boldsymbol{H}_t\right) p\left(\boldsymbol{\theta}_H\right) \tag{38}$$

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

Appendix C Out-of-Sample Forecast Algorithms

The predictive likelihood for an out-of-sample period $t+1$ can be computed in the following steps:

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 3.4 and Appendix B.

2. Simulate the state indicator from equation (5b):

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

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

4. Propagate $\boldsymbol{H}_{t+1}^{(i)}$ from equation (5e):

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

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

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

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

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

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

To simulate from the predictive distribution, replace the above Step 5 with the following step:

- Generate predictive draw $\boldsymbol{r}^{(i)}_{t+1}$ from $\boldsymbol{r}^{(i)}_{t+1} \sim N\left(\boldsymbol{\mu}^{(i)}_{s^{(i)}_{t+1}}, \boldsymbol{H}^{(i)1/2}_{t+1}\boldsymbol{\Sigma}^{(i)}_{s^{(i)}_{t+1}}\boldsymbol{H}^{(i)1/2'}_{t+1}\right)$ for every MCMC sample (i).