



Munich Personal RePEc Archive

# Identification and Forecasting of Bull and Bear Markets using Multivariate Returns

Liu, Jia and Maheu, John M and Song, Yong

Saint Mary's University, McMaster University, University of Melbourne

2023

Online at <https://mpra.ub.uni-muenchen.de/119515/>  
MPRA Paper No. 119515, posted 30 Dec 2023 08:38 UTC

# Identification and Forecasting of Bull and Bear Markets using Multivariate Returns\*

Jia Liu<sup>1</sup>, John M. Maheu<sup>2</sup>, and Yong Song<sup>3</sup>

<sup>1</sup>Saint Mary's University

<sup>2</sup>McMaster University

<sup>3</sup>University of Melbourne

August 2023

## Abstract

Bull and bear market identification generally focuses on a broad index of returns through a univariate analysis. This paper proposes a new approach to identify and forecast bull and bear markets through multivariate returns. The model assumes all assets are directed by a common discrete state variable from a hierarchical Markov switching model. The hierarchical specification allows the cross-section of state specific means and variances to differ over bull and bear markets. We investigate several empirically realistic specifications that permit feasible estimation even with 100 assets. Our results show that the multivariate framework provides competitive bull and bear regime identification and improves portfolio performance and density prediction compared to several benchmark models including univariate Markov switching models.

Keywords: Markov switching, Multivariate analysis, Investment strategies, Market timing

---

\*We thank the editor Eric Ghysels and two anonymous referees for many constructive comments. We are grateful for feedback from 2021 NBER-NSF SBIES Conference, 4th Annual Workshop on Financial Econometrics, and 2022 CFE conference. We thank seminars attendants at the University of Nottingham. Maheu thanks SSHRC of Canada for financial support.

# 1 Introduction

There is a long tradition of sorting upward and downward movements of a broad market index as bull and bear markets, respectively. These terms date back to the 1700s and were popularized after the south sea bubble (Marriam-Webster 2022). Since then, bull and bear labels have become the de facto standard that the popular press uses to characterize the state of the stock market. Almost universally these descriptors have been applied to a single index of stock returns both in the financial press and in academic work. This paper moves in a new direction by identifying and forecasting bull and bear markets for an index by using the underlying multivariate stock returns that constitute the index. We propose new models to extract and exploit the information on bull and bear markets contained in disaggregated stock return data.

There are two main approaches to dating bull and bear markets. Historically market phases have been identified through some ex post sorting rule applied to historical index data. For instance, a bull market is an increase of 20% or more from the most recent low (trough) in the index. These rules have been formalized by Pagan & Sossounov (2003) and Lunde & Timmermann (2004) but are not directly applicable to prediction and online (real time) inference for investors. The other dominant approach to bull and bear market identification assumes the state of the market is unobserved and are model based. These methods feature some form of a Markov switching (MS) specification that estimates the state of the market and provides a probability law on the likelihood of future states. Examples of this approach include Turner et al. (1989), Maheu & McCurdy (2000) and Maheu et al. (2012, 2021). Kole & Van Dijk (2017) compare ex post sorting rules to model based approaches for bull and bear identification.

Other papers are concerned with predicting bull and bear markets and the role of additional non-market data. Chen (2009) and Haase et al. (2020) use the bull and bear regimes obtained from one of the methods discussed above as the dependent variable and investigate the predictability of regimes through various information sources. Candelon et al. (2008) investigate the synchronization of market phases obtained from an ex post filter of several Asian indices. Other papers that consider trends in stock returns include Ang & Bekaert (2002) and Chauvet & Potter (2000).

There are some examples of Markov switching for multivariate modeling of asset returns for the purpose of asset allocation. Four assets are included in Guidolin & Timmermann (2006) and Guidolin & Timmermann (2007) while Tu (2010) extends this to 28 variables (3 Fama-French factors and 25 portfolios).<sup>1</sup> This differs from our purpose which is to model a larger set of assets to better identify bull and bear markets that constitute an index and make asset allocation decision based on this inference.

The panel data literature on business cycle analysis is related to our approach as well. Part of that literature is concerned with inferring a common Markov switching structure in a panel setting (Kaufman 2010, Hamilton & Owyang 2012) or allowing for nonsynchronous cycles (Harding & Pagan 2006, Paap et al. 2009) among series. These approaches either assume independent series conditional on the state variable, or if the correlation is modeled, it is in a small dimension. Our approach assumes one or two common state variables direct

---

<sup>1</sup>Pelletier (2006) models exchange rates with Markov switching directing the correlation matrix.

a large dimension of returns in which the full correlation dynamics are modelled. How the correlation structure is modelled is critical to state identification. We characterize state-specific parameters of all stocks with a hierarchical prior.

At a firm level we show that the main distinguishing feature between bull and bear markets are differences in conditional mean and conditional variances with the latter being the most important. These observations lead us to use a hierarchical prior to learn and characterize the differences in the cross-section of firm returns' conditional moments. A hierarchical prior is an effective method of information pooling for many common parameters and can lead to improved forecasts (Pesaran et al. 2006, Song 2014). At the same time correlations do change but the changes are only loosely associated with market cycles. We explore several correlation assumptions from constant, perfectly coupled to the regime and loosely coupled to the regime. These empirical characteristics, only available from multivariate data, are absent in the MS models of bull and bear regimes based on a univariate index.

The largest application is to 100 stocks. In each of the multivariate applications the model provides accurate inference on the regimes and is close to that of a MS model applied to the univariate index, but tend to be more precise. The cross-section of returns over regimes is very different. Individual stocks' means are lower and more dispersed in the bear market as well as possessing larger variances in general. We show that our preferred models improve density forecasts of returns.

Our multivariate approach provides superior signals for market timing compared to the univariate model. Sharpe ratios are larger and investors are willing to pay a higher performance fee for the information from multivariate returns. Although our focus is on market timing for an index based on the underlying multivariate stock returns we also show the multivariate models lead to improved investment decisions for the S&P500. These benefits from multivariate data transfer over to a more general mean-variance portfolio selection.

The multivariate models are designed to be parsimonious but in larger applications computational challenges do arise. The main bottleneck is inversion of covariance and correlation matrices and lack of a simple posterior sampling method for correlation matrices. We adopt a geodesic Lagrangian Monte Carlo method, which is a special case of Hamiltonian Monte Carlo (HMC), from Holbrook et al. (2018) and show that this approach works well in our application.

Section 2 details the data and provides some motivation for the multivariate models. Section 3 delineates the modelling framework and Section 4 shows how to perform inference from these models along with prediction. The application to several datasets is in Section 5, which includes model comparison using out-of-sample forecasts, regime identification and portfolio investment strategies. Section 6 reports several robustness checks including for potential survival bias, duration dependence and sub-cycle dynamics. Section 7 concludes. The Appendix describes data and posterior sampling methods in detail including details on a geodesic Lagrangian Monte Carlo method to sample large dimension positive definite matrix from a non-standard distribution.

## 2 Data and Motivation

In this section, we introduce the data followed by estimates of bull and bear markets from a univariate MS model applied to the S&P 500 index. The resulting inference on historical market episodes is used to sort the firm level data and investigate their econometric properties that a multivariate model should capture.

### 2.1 Data

To analyze bull and bear markets a long calendar span of data is needed that includes as many market cycles as possible. However, going back further in history results in fewer firms that have survived to the present. As such, there is a tradeoff between long historical datasets with fewer firms and shorter historical datasets with more firms. Therefore, we consider three samples of monthly equity returns excluding dividends obtained from CRSP. The first sample contains data of 30 assets from January 1926 to December 2020 (1140 months). The second sample includes 60 stocks from January 1931 to December 2020 (1080 months). The last sample considers 100 assets from January 1951 to December 2020 (840 months). Table 14 in Appendix A.1 provide detailed information about the 100 assets, in which the first 30 and 60 assets from the first two samples. The monthly returns are converted to annualized continuously compounded returns.<sup>2</sup>

Similar to Tu (2010) we selected a monthly frequency. In our experience higher frequency data such as weekly suffers from asynchronous components unrelated to the broad market trends. Namely, the individual stocks have their own distinct sub-cycle behaviour. Modeling multivariate sub-cycle behaviour would require a significantly more complex model and restrict the dimension to a much smaller number of assets. Using monthly returns allows us to exploit the benefit of high dimensionality with a feasible cost of model complexity and computational burden.

At the aggregate level, we use the equally weighted (EW) index formed from the 30, 60 and 100 asset returns, which we name EW-30, EW-60 and EW-100, respectively. We also use the S&P500 index from CRSP to align with the literature. The S&P500 index data before 1957 is the 90-stock index from CRSP. The risk-free rate is obtained from the Kenneth French data library.

In the following, we use  $r_t$  to denote the index return,  $r_{t,i}$  for the return of firm  $i$  and  $R_t = (r_{t,1}, \dots, r_{t,N})'$  for the  $N \times 1$  vector of stock returns. The notation  $r_{1:t} = \{r_1, \dots, r_t\}$  and  $R_{1:t} = \{R_1, \dots, R_t\}$  are used for the index returns and stock returns up to time  $t$ , respectively.

### 2.2 Bull and Bear Markets from a Univariate Model

Consider the following 2-state univariate Markov switching model, denoted by UMS for the univariate S&P 500 portfolio return  $r_t$ .

$$r_t \mid s_t = k \sim N(\mu_k, \sigma_k^2), \quad (1)$$

$$P(s_{t+1} = k \mid s_t = j) = p_{jk}, \quad (2)$$

---

<sup>2</sup>That is, not excess returns.

where  $j, k = 1, 2, t = 1, \dots, T$ . The prior is

$$\mu_k \sim N(0, 1), \sigma_k^2 \sim IG(2, 2), k = 1, 2, \quad (3)$$

$$(p_{11}, p_{12}) \sim Dir(10, 1) \text{ and } (p_{22}, p_{21}) \sim Dir(10, 1). \quad (4)$$

An identification restriction is  $\mu_1 < \mu_2$ , that the bear mean is less than the bull mean. These priors are fairly uninformative but do favour persistent states. When  $s_t = 1$ , the period  $t$  is in the bear regime, while  $s_t = 2$  means the bull regime.  $IG()$  and  $Dir()$  denote an inverse gamma and Dirichlet distribution, respectively. The posterior inference is standard from Chib (1996).

Bayesian posterior simulation methods enable us to obtain the posterior probability of states as  $P(s_t = k | r_{1:T})$ . Our estimate of the bull state ( $s_t = 2$ ) probability is displayed in the top of Figure 2. Obviously,  $P(s_t = 1 | r_{1:T}) + P(s_t = 2 | r_{1:T}) = 1$  by model construction. If we call period  $t$  a bear regime when  $P(s_t = 1 | r_{1:T}) > 0.50$  and bull otherwise, the firm level data can be partitioned into two states using this rule. Collecting all data from state 1 and 2 gives us two samples, one of bear market returns and one of bull market returns.

Summary statistics for individual stocks in the bull and bear states are calculated for investigation using the 30 stock dataset from 1926–2020 for illustration. Figure 1 shows the scatter plot of the sample means and variances of individual stocks for different regimes. Blue dots are bear regime statistics while red ones are bull regime statistics. Clearly, individual stocks in the bear regime tend to have negative average returns and higher volatilities than in the bull regime. These results are consistent with the aggregate index data estimates that show negative mean, high variance for bear and high mean, low variance for bull markets (Turner et al. 1989, Maheu & McCurdy 2000, Pagan & Sossounov 2003, Lunde & Timmermann 2004). This similarity between firms and aggregate data suggests that we may improve bull and bear identification by using all firm level information.

One benefit from the multivariate analysis is that we can investigate the correlation structure among variables, which is infeasible for any univariate analysis. Figure 3 shows the sample correlations of all pairs of the first 30 stocks in bull and bear regimes. All but two points are below the 45 degree line, indicating a stronger correlation between stocks during a bear regime. In the bear market most correlations are between 0.2–0.8 while they are in 0.1–0.4 in the bull market. These observations are consistent with Tu (2010) but that analysis is restricted to the Fama-French 25 portfolios with a much shorter time span. The clear change in correlation structure over regimes indicates that correlations may provide a strong signal on the market phase.

This preliminary investigation reveals a potential benefit by using multiple stock returns to learn about bull and bear regimes. We find significant differences in the stock return moments and correlations of firm returns over bull and bear markets. There are differences in both the level and dispersion of these moments. However, a multivariate model can suffer from the curse of dimensionality and computational problems. Therefore, we pursue parsimonious parametrization that can exploit the empirical regularities we discuss in this section. To this end we design a hierarchical prior specification to consolidate information from individual stocks. Such a hierarchical approach explicitly estimates the uncertainty associated with regime-dependent means and variances of individual stocks. Our model allows counter-cyclical behaviour from any individual stocks.

The second new feature is that our models is able to decouple the correlation dynamic from the mean and variance process. This is motivated from Figure 2. Its top panel shows the posterior probability of being in the bull regime based on the S&P500 index. The middle and bottom panel plots the average monthly realized volatilities and pairwise correlation coefficients of 30 stock returns constructed at daily frequency, respectively. A visual check indicates that a bear (bull) regime is usually associated with higher (lower) volatility, as agreed by the literature and documented above. On the other hand, only mild evidence exists for regime dependent correlations. Hence, we consider several correlation structures from perfectly coupled, loosely coupled to independent with bull and bear regimes. These are discussed next.

### 3 Multivariate Models

We focus on four models in this paper. The first is a benchmark unrestricted multivariate Markov switching (MMS) model. The second model is multivariate Markov switching with constant correlation. This is more parsimonious, because only one set of the correlation coefficients is needed and it saves  $N(N - 1)/2$  parameters. The third specification is a multivariate Markov switching model with independent correlation dynamics. Here the mean and variance parameters follow a Markov chain while the correlation matrix follows its own independent Markov chain. The final model combines the two Markov chains to allow dependence between them. The regime governing the means and variances can extract information from the correlation dynamics but they are not restricted to switching regimes simultaneously.

#### 3.1 Benchmark Multivariate Markov Switching Model (MMS)

Our benchmark model is a standard multivariate Markov switching model with 2 states and multivariate normal data density denoted as MMS. In particular, the  $N \times 1$  vector  $R_t = (r_{t,1}, \dots, r_{t,N})'$  of stock returns follows,

$$R_t \mid s_t = k \sim N(M_k, \Sigma_k), \quad (5)$$

$$P(s_{t+1} = j \mid s_t = k) = p_{kj}, \quad (6)$$

where  $j, k = 1, 2$ . Without loss of generality, we name regime 1 as the bear market and 2 as the bull market. The  $N \times 1$  vector  $M_k$ , for  $k = 1, 2$ , is the mean vector of  $R_t$  in regime  $k$ . The  $N \times N$  matrix  $\Sigma_k$ , for  $k = 1, 2$ , is the covariance matrix of  $R_t$  in regime  $k$ .

For identification purposes, we assume that the equally-weighted portfolio constructed from  $R_t$  has a negative mean in the bear market and a positive mean in the bull market. We adopt an equally-weighted portfolio as our target *market* portfolio that emerges from the model. It is this portfolio that we are concerned about for tracking bull and bear phases. Define  $\iota_N$  as a  $N$  dimensional vector of ones. This restriction is equivalent to

$$\iota_N' M_1 < 0, \quad \iota_N' M_2 > 0. \quad (7)$$

Such a restriction is not binding at all in the applications and provides a great deal of flexibility for individual stock return means.<sup>3</sup>

Define the  $i$ th element in  $M_k$  as  $\mu_{ik}$ , hence  $M_k = (\mu_{1k}, \mu_{2k}, \dots, \mu_{Nk})'$  for  $k = 1, 2$ . The priors for the MMS model are

$$\Sigma_k \sim IW(\Psi, \tau), \text{ for } k = 1, 2, \quad (8)$$

$$\mu_{ik} \sim N(0, 1), \text{ for } i = 1, 2, \dots, N \text{ and } k = 1, 2, \quad (9)$$

$$(p_{11}, p_{12}) \sim Dir(10, 1) \text{ and } (p_{22}, p_{21}) \sim Dir(10, 1). \quad (10)$$

The  $IW$  means the invert Wishart distribution, and  $Dir$  means the Dirichlet distribution. We assume that  $\Psi = \widehat{\text{cov}}(R_t)(\tau - N - 1)$  and  $\tau = N + 2$ , so the prior mean of  $\Sigma$  is the sample covariance of  $R_t$ . The value  $\tau$  implies infinite variance, hence our prior covers a large range of reasonable values. The prior on the transition probabilities is informative and favours persistence of states which is consistent with past work.

### 3.2 Correlation Decomposition and Hierarchical Prior

The next set of models we consider all make use of the following covariance decomposition into a correlation matrix as

$$\Sigma_k = \Delta_k \Lambda_j \Delta_k, \quad (11)$$

$$\Lambda_j = \Gamma_j Q_j \Gamma_j, \quad \Gamma_j = \text{diag} \left( \frac{1}{\sqrt{Q_{j,11}}}, \dots, \frac{1}{\sqrt{Q_{j,NN}}} \right), \quad (12)$$

where the regime dependent diagonal matrix of standard deviations is  $\Delta_k = \text{diag}(\sigma_{1k}, \sigma_{2k}, \dots, \sigma_{Nk})'$  for  $k = 1, 2$  and  $\Lambda_j$  is a correlation matrix and can be constant or follow a Markov chain indexed by  $j$ . This specification allows for stock variances with significant regime differences and constant correlations as well as correlations displaying less regime dependence and possible independent dynamics.

The decomposition allows us to exploit regime differences and pool information among stocks' mean and variance parameters through the following hierarchical prior,

$$\mu_{ik} \sim N(m_k, v_k^2), \quad (13)$$

$$m_k \sim N(0, 0.5), \quad (14)$$

$$v_k^2 \sim IG(5, 0.25), \quad (15)$$

$$\log(\sigma_{ik}) \sim N(\zeta_k, b_k^2), \quad (16)$$

$$\zeta_k \sim N(0, 0.5), \quad (17)$$

$$b_k^2 \sim IG(5, 0.25), \quad (18)$$

$$Q_j \sim IW(\Psi, \nu) \quad (19)$$

$$(p_{11}, p_{12}) \sim Dir(10, 1) \text{ and } (p_{22}, p_{21}) \sim Dir(10, 1), \quad (20)$$

---

<sup>3</sup>We did not find a single violation of these restrictions in execution if such restriction is not imposed in a long sequence of MCMC iterations. We do not impose restrictions on the means of individual stocks. Namely, counter-cyclical dynamics for an individual stock is permitted.

for  $i = 1, 2, \dots, N$  and  $k = 1, 2, j = 1, 2$ . The first component implies all stock return means are drawn from a common normal distribution with mean  $m_k$  and variance  $v_k^2$  which are allowed to differ over bull and bear markets and characterize the cross-sectional differences in stock returns. The prior on both  $m_k$  and  $v_k^2$  allows for learning about these parameters from the data. A priori, the means of stock returns are centred around 0 with a standard deviation of around 2.2, which means a large dispersion of 220 percentage points. Hence, the prior of the means is very diffuse, so the posterior distribution would mainly reflect the data information.

For identification of regimes, we impose the restriction  $m_1 < m_2$  making the mean of the cross-section of stock return means less in the bear market than the bull market. No restrictions are imposed on variances or other parameters.

Similarly, the prior on  $\log(\sigma_{ik})$  indicates that all log-volatilities of stocks are drawn from a common normal distribution. The parameters governing this distribution,  $\zeta_k, b_k^2$ , have a prior and are estimated from the data and in general will differ in each regime. The centre of  $\sigma_{ik}$  is around 1 and a standard deviation of around 4, which is very large in the applications and makes the prior uninformative. The special case of  $v_k^2 \rightarrow 0$  or  $b_k^2 \rightarrow 0$  occur when all stocks share the same mean or the same log-volatility level. In general, these parameters will be positive and larger values indicate greater heterogeneity among stocks within a regime.

The prior on  $Q$  follows Barnard et al. (2000), which separates the volatility and correlation symmetrically, unlike some other approaches such as the Cholesky decomposition. We set  $\Psi = \widehat{\text{cov}}(R_t)(\nu - N - 1)$  and  $\nu = N + 2$  as in the MMS model.<sup>4</sup>

The prior of the transition probability is a standard setting for Markov switching models. It shows a weak belief that the regimes are persistent.

### 3.2.1 Constant Correlation

A constant correlation model is more parsimonious than the MMS because it only estimates one correlation matrix. However, only the mean and variance dynamics can be used for bull and bear identification. We denote this model as MMS-CC (CC as constant correlation) and is

$$R_t \mid s_t = k \sim N(M_k, \Delta_k \Lambda \Delta_k), \quad (21)$$

$$P(s_{t+1} = j \mid s_t = k) = p_{kj}, \quad (22)$$

where  $j, k = 1, 2$ . The MMS-CC model uses the hierarchical prior of Section 3.2 but simplifies the covariance decomposition with a constant correlation  $\Lambda$  across regimes. This model postulates that the key differences in regimes is the conditional mean and conditional variance which is partially consistent with Section 2 results.

### 3.2.2 Independent Correlation Process

The next model investigates if an independent correlation process may help bull and bear regime identification. It is motivated from Figure 2, as it does not clearly show if the

---

<sup>4</sup>We have also tried a uniformly distribution of the correlations by setting  $\Psi = I$  and  $\nu = N + 1$ . The results are qualitatively the same.

correlation synchronizes with bull and bear regimes. If there is more noise than signal in the correlation, it is probably better to let it follow an independent process. The idea is similar to MMS-CC model, but allows the correlation to change over time to give more flexibility. We denote this model as MMS-IC (IC as independent correlation).

In execution, we propose a second independent Markov process to govern the dynamics of the correlation matrix. This does not rule out synchronization of regimes and correlation states. If the data favours synchronization, then the posterior inference may indicate such, otherwise, modelling the correlation structure independently may improve inference on bull and bear cycles.

The model is,

$$R_t \mid s_t = k, w_k = j \sim N(M_k, \Delta_k \Lambda_j \Delta_k), \quad (23)$$

$$P(s_{t+1} = k \mid s_t = l) = p_{lk}^s, \quad (24)$$

$$P(w_{t+1} = j \mid w_t = n) = p_{nj}^w, \quad (25)$$

where  $j, k, l, n = 1, 2$  and the parameters have the hierarchical prior of Section 3.2. The state  $s_t$  governs the conditional mean vector  $M_k$  and conditional variance matrix  $\Delta_k$  while and the state variable  $w_t$  governs  $\Lambda_j$ . The transition probabilities prior for the correlation matrix is

$$(p_{11}^w, p_{12}^w) \sim Dir(10, 1) \quad \text{and} \quad (p_{22}^w, p_{21}^w) \sim Dir(10, 1). \quad (26)$$

This prior reflects state persistence in correlation dynamics. In the following we continue to refer to bull and bear regimes as  $s_t = 1, 2$  irrespective of the value of  $w_t$ .

### 3.2.3 Dependent Correlation

It is possible that the correlation structure depends on the regime process directing the mean and variances but does not fully co-move with them. In this case correlation dynamics can provide useful information on regime identification but the model extremes above will not accommodate this structure. The benchmark MMS imposes the restriction that the correlation matrix and the volatility must change at the same time. On the other hand, the MMS-CC and MMS-IC cut off any potential dependence between these two state processes.

As a middle ground, we propose a restricted 4-state model, denoted by MMS-DC (DC as dependent correlation), to allow stochastic dependence between the correlation structure and the regime process governing means and volatility parameters. We still preserve two sets of correlation matrices and means/volatilities as in MMS-IC. Instead of assuming two independent 2-state Markov chains, we allow for these two processes to have dependence, which is achieved by augmenting the state space to form one Markov process with 4 states.

The details are as follows,

$$R_t \mid s_t = k \sim N(M_k, \Delta_k \Lambda_k \Delta_k), \quad (27)$$

$$P(s_{t+1} = j \mid s_t = k) = p_{kj}, \quad (28)$$

where  $j, k = 1, 2, 3, 4$ . Note that  $\Lambda_k$  now shares the same state indicator  $k$  as  $\Delta_k$ . The restrictions are  $M_1 = M_2$ ,  $\Delta_1 = \Delta_2$ ,  $M_3 = M_4$ ,  $\Delta_3 = \Delta_4$ ,  $Q_1 = Q_3$  and  $Q_2 = Q_4$ . These

restrictions imply the parameters  $M_2, M_4, \Delta_2, \Delta_4, Q_3$  and  $Q_4$  are redundant while the free parameters  $M_1, M_3, \Delta_1, \Delta_3, Q_1, Q_2$  follow the hierarchical prior of Section 3.2.

The  $k$ th row of the transition matrix  $P$ , denoted by  $P_k$ , has a Dirichlet prior with more weight on the self-transition probability.

$$P_k \sim \text{Dir}(1 + 9 \times \mathbf{1}(k = 1), 1 + 9 \times \mathbf{1}(k = 2), 1 + 9 \times \mathbf{1}(k = 3), 1 + 9 \times \mathbf{1}(k = 4)). \quad (29)$$

This transition matrix differs from that in the MMS-IC model. In that model  $P = P_s \otimes P_w$ , where  $P_s$  is the transition matrix of the mean and volatility state and  $P_w$  is the transition matrix of the correlation state and is restricted to 4 parameters. In this model  $P$  is unrestricted with 12 parameters; and the bear regime occurs when  $s_t = 1$  or 2 while the bull regime is when  $s_t = 3$  or 4.

## 4 Inference and Prediction

### 4.1 Bayesian Inference

The models are estimated through standard Markov chain Monte Carlo (MCMC) techniques. Since Chib (1996) posterior sampling has become standard for conventional MS models so we do not elaborate on them here. All details are in the appendix.

The posterior simulation is carried out for  $G$  iterations after an initial burnin are discarded to remove potential initial value effects. These  $G$  iterations give a sample of parameters that are used to form simulation consistent estimators of model parameters. For instance, if one is interested in the probability of period  $t$  being in a bull regime  $P(s_t = 2 \mid R_{1:T})$  in the MMS model, where  $R_{1:T}$  represents the full sample data, we can use

$$P(s_t = 2 \mid R_{1:T}) = \frac{1}{G} \sum_{g=1}^G \mathbf{1}(s_t^{(g)} = 2)$$

as an estimate of the true value  $P(s_t = 2 \mid R_{1:T})$ . The superscript  $(g)$  is the iteration counter. The accuracy of this estimator increases as  $G$  increases. In the applications, we use 5000 posterior draws for inference after dropping 50000 burnin iterations.

The high dimensionality, however, creates a bottleneck from a simulation efficiency perspective for  $Q_k$  in MMS-IC and MMS-DC models and  $Q$  in the MMS-CC model. No simple Gibbs sampling method is available for these correlation matrices. One could use the Metropolis-Hastings algorithm on each individual parameter in  $Q_k$ , but it is extremely time consuming and practically infeasible as the dimension becomes large (up to 5050 parameters when  $N = 100$ ). A joint Metropolis-Hastings (M-H) sampler might be faster, however, in general it produces a highly persistent sample path and quickly becomes inefficient at a higher dimension than 30. A generic Hamiltonian Monte Carlo (HMC) method (see Neal et al. (2011) for an overview) works well when  $N = 30$  in our application but constantly fails to provide a positive definite proposal matrix when  $N$  becomes larger. Instead, we applied Holbrook et al. (2018)'s geodesic Lagrangian Monte Carlo (gLMC) method, which is a specially tailored HMC on manifolds (see Girolami & Calderhead (2011) and Byrne & Girolami (2013)), to sampling  $Q_k$  or  $Q$ .

We found the gLMC works well in our applications. The next section provides details on efficiency gains of gLMC while Appendix A.5 contains the technical details for estimation. The reader interested in the empirical results may skip Section 4.2.

## 4.2 Evaluating Geodesic Lagrangian Monte Carlo

To evaluate the gLMC’s usefulness on sampling positive definite matrices, we consider a very simple multivariate normal model as

$$y_t \sim N(0, \Sigma),$$

for  $t = 1, \dots, T$ . The vectors  $\{y_t\}_{t=1}^T$  have dimension  $N \times 1$  and are independent. There is only one parameter,  $\Sigma$ , which can have a large dimension as  $N(N + 1)/2$ . We assume a conjugate prior as

$$\Sigma \sim IW(A_0, a_0),$$

which returns a textbook analytic posterior distribution as

$$\Sigma | Y \sim IW(A_1, a_1),$$

where  $a_1 = a_0 + T$  and  $A_1 = A_0 + Y'Y$  with  $Y = [y_1, y_2, \dots, y_T]'$ . The analytic solution can verify the convergence of the M-H and gLMC methods and then help us understand their efficiency. *IW* means invert-Wishart.

We compare the gLMC to a joint random walk Metropolis-Hastings method. Table 1 compares these two sampling algorithms. We chose the dimensions and sample size to match the application in this paper. The true value  $\Sigma$  has 1 on the diagonal and 0.1 on the off-diagonals. In summary, for our application, the gLMC is more efficient for all dimensions (30, 60 and 100). However, the efficiency gain is diminishing as the dimension increases.

These calculations are performed in Matlab(2020b) and run on a Windows 10 operating system with Intel Core i7-4790 CPU @ 3.60GHz. We report the number of seconds, efficient sample size and their ratio (ESS per second). The last column of the table reports the efficiency gain if we use gLMC over the M-H method, defined as the ratio of ESS per sec of gLMC over that of the M-H method. For example, when  $N = 30$  and  $T = 1116$ , the gLMC is 11 times as efficient as the M-H method. Such an advantage diminishes when  $N$  grows larger. Nonetheless, when  $N = 100$ , the gLMC still produces more than double of the effective sample size that the M-H can offer.

For the M-H algorithm, we set the total number of iterations as 200,000 and discard the first half of the samples as burn-in. Then we save 1 out of each 100 iterations for inference. The inverse step size parameter  $v$  is set as 100k for  $N = 30$  and 60; 500k for  $N = 100$  (very small step sizes for a reasonable acceptance rate). For the gLMC, we set the total number of iterations as 1100 and discard the first 100 as burn-in. The leap size is set as 0.01, and the number of leaps is set as  $L = 10$  for  $N = 30$  and 60; and  $L = 5$  for  $N = 100$ . The posterior distributions are compared to the true posterior distribution, so the convergence is guaranteed.

### 4.3 Prediction

The posterior analysis provides insight into the models' in-sample performance. An investor could be more interested in a model's out-of-sample forecasts in order to make decisions to buy or sell. Because the models in this paper are all complete, we can evaluate their out-of-sample forecast precision as well as their economic values for an investor.

For statistical evidence, we compute the predictive likelihood  $p(R_{t+1} | R_{1:t}, \mathcal{M}_i)$ , where  $\mathcal{M}_i$  represent model  $i$  including MMS, MMS-CC, MMS-IC and MMS-DC. Using the MMS-IC model for example, define  $\Psi = \{\mu_{ik}, m_k, v_k^2, \sigma_{ik}, \zeta_k, b_k^2, Q_k, s_{1:t}, w_{1:t}, P^s, P^w\}_{k=1, i=1}^{2, N}$  as the collection of all model parameters and auxiliary variables. A one-period conditional predictive density  $p(R_{t+1} | R_{1:t}, \Psi)$  (ignore  $\mathcal{M}_i$  for simplicity) is given by

$$p(R_{t+1} | R_{1:t}, \Psi) = \sum_{k=1}^2 \sum_{j=1}^2 p(R_{t+1} | M_k, \Delta_k \Lambda_j \Delta_k) p_{s_{t,k}}^s p_{w_{t,j}}^w. \quad (30)$$

It is a mixture of multivariate normal distributions, where the weights are determined by the current state at time  $t$  and the transition probabilities.

Suppose that we have collected a sample of draws, denoted by  $\{\Psi^{(g)}\}_{g=1}^G$ , from the posterior distribution  $p(\Psi | R_{1:t})$ . The predictive density is

$$p(R_{t+1} | R_{1:t}) = \int p(R_{t+1} | R_{1:t}, \Psi) p(\Psi | R_{1:t}) d\Psi, \quad (31)$$

and can be consistently estimated from the posterior sample as

$$p(\widehat{R_{t+1}} | R_{1:t}) = \frac{1}{G} \sum_{g=1}^G p(R_{t+1} | R_{1:t}, \Psi^{(g)}), \quad (32)$$

where  $p(R_{t+1} | R_{1:t}, \Psi^{(g)})$  is available in (30). Evaluating the predictive density at the realized data  $R_{t+1}$  gives the predictive likelihood and forms the basis of model comparison amongst models. The predictive likelihood is forecast based and hence does not suffer from an overfitting problem because it penalizes more complex models that do not deliver improved predictions as measured by larger predictive likelihood values.

The data is divided into a training sample from period 1 to  $T_0$  and the rest is used to compute predictive likelihood values for all models. The predictive likelihood for data from  $T_0 + 1$  till the end of the sample  $T$  is expressed as the product of 1-period ahead predictive likelihoods as

$$p(R_{T_0+1:T} | R_{1:T_0}) = \prod_{t=T_0}^{T-1} p(R_{t+1} | R_{1:t}). \quad (33)$$

Each individual predictive likelihood,  $p(R_{t+1} | R_{1:t})$ , involves a separate estimation for each  $t$ . A higher predictive likelihood means more support for the data. Using predictive likelihood for model comparison is advocated by Geweke & Amisano (2010) and can be linked to the marginal likelihood comparison in Kass & Raftery (1995). A simple rule of thumb is that model  $\mathcal{M}_i$  is strongly supported by the data against model  $\mathcal{M}_j$  if the log predictive Bayes factor, defined as  $\log(B_{ij}) = \log\left(\frac{p(R_{T_0+1:T} | R_{1:T_0}, \mathcal{M}_i)}{p(R_{T_0+1:T} | R_{1:T_0}, \mathcal{M}_j)}\right)$ , is larger than 5.

To investigate the economic value of these models, we carry out investment strategies and evaluate their performance for different models. All these strategies are based on out-of-sample forecasting. Key inputs include the predictive probability of a state, predictive mean and covariance. The predictions are computed through simulations as a byproduct of the posterior simulations. Using MMS-IC again as an example, given a posterior sample  $(\Psi^{(g)})_{g=1}^G$  conditional on data  $R_{1:t}$ , for each  $\Psi^{(g)}$ , we simulate states  $s_{t+1}$  and  $w_{t+1}$  according to (24) and (25), followed by returns generated by (23). The simulated data, denoted by  $\{\tilde{R}_{t+1}^{(g)}\}_{g=1}^G$ , represents a sample from the predictive distribution of  $R_{t+1} | R_{1:t}$  from which simulation consistent statistics such as the predictive mean and covariance can be easily computed.

## 5 Application

### 5.1 Model Comparison

Table 2 reports log-predictive likelihoods (as the log value of (33)) for model comparison. The out-of-sample period is 1971/01-2020/12 (600 months) for 30 and 60 assets and 1991/01-2020/12 (360 months) for the 100 assets application. The first model UMS is a univariate 2-state MS model in Section 2.2 applied individually to each return series. It ignores any correlation or other dependencies between individual return series. The other models are delineated in Section 3. All models are recursively estimated by expanding samples, hence the model comparison is forecasting based.

From Table 2 we see significant improvements in density forecasts by moving from the simple UMS model to the benchmark MMS model. It shows that ignoring correlations could lead to erroneous prediction and likely to harm any associated investment strategy. Our models, however, provide huge additional gains in forecast accuracy. For example the log-Bayes factor for MMS-DC vs the MMS for 100 assets is an astonishing 5133. For each dataset the more sophisticated models with independent or dependent correlation are strongly preferred. As more assets are included, the flexible modeling of dependence in means/variance and correlations switching becomes important. Based on density forecasts, MMS-DC is our preferred model for capturing bull and bear market trends. This model comparison result is robust to various out-of-sample sizes. Additional results are available on request from the authors.

### 5.2 Full Sample Estimates

A contribution of this paper is the introduction of the hierarchical prior to characterize the cross-section of stock return moments in a bull and bear market. Table 3 reports the posterior means of the hierarchical prior parameters for the MMS-DC model for 30/60/100 assets cases. On average, stock returns are negative in the bear market and positive in the bull market. Average return standard deviation is larger in the bear market, which is about double the size of that in the bull market. Use the 60-asset case for illustration, the differences are displayed in Figure 4 and 5. This is the hierarchical distribution for stock means and volatilities (standard deviations) for bull and bear markets. The distribution

governing  $\mu_{i2}$  (mean of stock  $i$  in the bull market), as shown by the dashed line, is more concentrated and farther to the right. Nevertheless, a negative mean can occur in the bull market as well as a positive mean in the bear market. As such, this model can accommodate counter-cyclical stocks. The differences in the regimes are more pronounced for volatility as seen in Figure 5 and make volatility differences an important signal on regime inference.

The value of  $v_k^2$  for  $k = 1, 2$  decreases when the number of assets increases. This can be visualized in Figure 6, which plots the distribution of  $\mu_{ik}$  for  $k = 1, 2$  (bear) or  $3, 4$  (bull) as a function of 30/60/100 assets. The shape of the distributions becomes more concentrated as the number of assets increases and is more pronounced in the bull market. Because we do not impose sign restrictions on any individual stock return, such shape evolution shows that more assets are able to provide a better description of the *typical* stock in a bull and bear regime with a more focused meta-distribution over respective regimes. The same pattern of concentration also exists in the distribution of volatility in Figure 7, although such pattern is not as visible as in Figure 6. The same pattern appears in the MMS-CC and MMS-IC models as the number of assets increases and their results are available on request.

The posterior mean of the transition matrices for 30/60/100 assets from the MMS-DC model is shown in Table 4. Each top-left  $2 \times 2$  sub-matrix is the bear block and represents the bear market, and the bottom-right  $2 \times 2$  sub-matrix is the bull block and represents the bull market. The in-block transition probability ranges between 0.54 to 0.84 for the bear regime, and the in-block transition probability ranges between 0.53 to 0.94 for the bull regime. The bear and bull block are persistent, which is consistent with the existing literature.

Mixture models including Markov switching models might suffer from the label switching problem (Jasra et al. 2005). Because we impose restrictions to the mean vector for the multivariate models, the identification of bull and bear regimes is not a problem. On the other hand, the correlation matrices do not have any restriction and hence are not identified in theory. Geweke (2007) demonstrated that label switching is not an empirical issue as long as the posterior distribution of the components are well separated and convergence check is performed, which is confirmed in our applications. For example, Figure 8 shows the trace plot of the average of correlation coefficients in the 60 assets application and displays no evidence of label switching. We also checked the trace plots of a large spectrum of quantiles of the correlations' distribution in each regime for all three applications and find no label switching behaviour.

The absence of label switching allows us to interpret the correlation matrices. For illustration, Figure 9 shows the kernel density plot of the posterior means of correlation coefficients in each regime for the 60 assets application. The distributions of correlations over regimes are visually distinguished after we established that no label switching happens during the MCMC. Figure 9 indicates a *high* correlation regime (state 1 and 3) and a *low* (state 2 and 4) one.

The transition matrix  $\bar{P}_{60}$  from Table 4 shows that the bear block is more likely to be associated with the high correlation state. This can be seen from the bear block matrix  $\begin{pmatrix} 0.84 & 0.05 \\ 0.07 & 0.74 \end{pmatrix}$ . The high correlation state has a higher self-transition probability (0.84) than the low correlation state (0.74), and the low correlation state is more likely to switch to the high correlation state (0.07) than the other direction (0.05). With the same logic, we can see that the bull regime favours the low correlation state. This is consistent with Figure 3 and

shows that the correlation structure aids bull and bear regime identification. Meanwhile, the transition matrix also shows that low (high) correlation state is not perfectly coupled with bear (bull) regimes and points to the importance of the 4-state model which captures these dependencies. The unconditional probabilities associated with the posterior mean of  $\overline{P}_{60}$  are (0.139, 0.136, 0.039, 0.685) and the most time is spent in the last state, high mean return and low correlation.

### 5.3 State Identification

In this section, we consider state identification. Our focus is to compare univariate and multivariate models state identification for an EW index. The former only exploits the aggregate equally weighted return (EW-30, EW-60 and EW-100), while the latter use the full information from 30, 60 and 100 assets' returns. Since we never observe the real bull and bear regime in reality, the results here should be interpreted with caution.

Although the equally-weighted portfolio provides a consistent way to compare univariate and multivariate models we also include the S&P500 as another benchmark. However, the S&P500 is not directly comparable, since we do not model its underlying assets.

Figure 10, 11 and 12 display the results. The top panel is the cumulative log-return from the equally-weighted index and the S&P500. The second and third panel are the probability of the bull regime in the MMS-DC model and the benchmark MMS model, respectively. The bottom panel shows the probability of the bull regime from a UMS model applied to EW-30, EW-60 and EW-100, respectively, and the S&P500. The last panel of each figure reflects what is conventionally delivered if using an aggregate index.

All three models display close resemblance in capturing market states, but the multivariate models provide sharper regime identification. Focusing on 100 asset cases in Figure 12, the closest to a broad market index, we see that the multivariate models provide abrupt turning points with probability often shifting immediately from 1 to 0 or vice versa. The univariate model generates more uncertainty in regime inference. Recall that the multivariate models provide good out-of-sample density forecasts, which should translate into more clear-cut in-sample state identification. Next, we discuss whether such a difference in the regime inference improves investment decisions.

### 5.4 Investment Strategy

This section investigates the economic value of the multivariate models in comparison to the univariate model through various investment strategies. Model comparison is conducted on the same out-of-sample window that was used for the log-predictive likelihood calculations results in Section 5.1. Alternative sub-sample results have been carried out as robustness check and qualitatively similar.

#### 5.4.1 Market Timing Portfolio

We consider two strategies based on the predicted probability of market regimes. For the first, an investor invests his wealth in the equally-weighted index (EW-30, EW-60 or EW-100) at time  $t$  if  $p(s_t = \text{bull} \mid I_{t-1}, \mathcal{M}) > \tau$  for a pre-specified threshold  $\tau$ , otherwise,

he holds the risk-free asset. The symbol  $I_{t-1}$  means the information up to  $t - 1$  implied by the model. Namely, for multivariate models  $I_{t-1} = R_{1:t-1}$ , and for univariate models  $I_{t-1} = r_{1:t-1}$ . The  $s_t = \text{bull}$  is a shorthand meaning  $s_t = 2$  in the UMS models, MMS, MMS-CC and MMS-IC model; and  $s_t = 3$  or  $4$  in the MMS-DC model.

The second market timing strategy invests a proportion of wealth in an EW index using the predicted probability of the bull state to determine the proportion. A higher probability of a future bull state means a higher share of wealth being invested in the index. For example, if  $P(s_t = \text{bull} | I_{t-1}, \mathcal{M}) = 0.75$ , the investor allocates 75% of his wealth to buy the index and puts the remaining 25% in the risk-free asset.

The univariate model uses these indices directly for estimation and forecasting while multivariate models use all asset returns to forecast the future state. For all methods, univariate or multivariate, the investment strategy operates on the same financial assets: an EW index and the risk-free asset and the investor takes a position for day  $t$  and the return computed at the end of the day. This makes the univariate and multivariate model performance directly comparable.

We report the summary statistics associated with the return from the investment strategies. In addition, we also calculate the performance fee based on a quadratic utility function for an investor, which is

$$U(r_t^p | \mathcal{M}_A) = (1 + r_t^p) - \frac{\gamma}{2(1 + \gamma)}(1 + r_t^p)^2, \quad (34)$$

where  $\gamma$  denotes the risk aversion coefficient,  $r_t^p$  represents the portfolio return. The performance fee  $\Delta$  that an investor would pay to switch from the model  $\mathcal{M}_A$  to  $\mathcal{M}_B$  is inferred from

$$\sum_{t=1}^T U(r_t^{pA} | \mathcal{M}_A) = \sum_{t=1}^T U(r_t^{pB} - \Delta | \mathcal{M}_B). \quad (35)$$

Here  $\Delta$  is the fee that makes the investor indifferent in terms of ex post utility between using model  $\mathcal{M}_A$  and  $\mathcal{M}_B$ . A positive value of  $\Delta$  means  $\mathcal{M}_B$  is preferred to  $\mathcal{M}_A$ . The notation  $r_t^{pA}$  and  $r_t^{pB}$  are the return from the optimal decision based on the prediction from model  $\mathcal{M}_A$  and  $\mathcal{M}_B$ , respectively.

Table 5 – 7 report results for the 30, 60 and 100 assets application, respectively. Models are recursively estimated to produce a forecast of the bull regime at each point in the out-of-sample period, and make the investment decision. The top entry in the tables reports summary statistics and the Sharpe ratio of the buy-and-hold strategy of the EW indices. The buy-and-hold strategy of the EW index serves as the benchmark for model evaluation and performance fees calculation. We summarize some key information conveyed below.

Table 5 – 7, with few exceptions, show that the multivariate models on various strategies have much larger Sharpe ratios than the buy-and-hold strategy. For 30 assets application, simple multivariate models MMS or MMS-CC have the best Sharpe ratio. As the number of assets increase to 60 and 100, more flexible multivariate models as MMS-IC and MMS-DC provide more gains. As the dimension increases, the trade-off between signal and noise is more influential, a more careful treatment of the correlation structure may identify the state better.

Figure 13 displays the log-portfolio value for each model over time for investment strategy I with  $\tau = 0.5$ . MMS-IC not only has the largest terminal wealth but generally has the largest

value over the investment period. The buy and hold portfolio does a bit better for a few years after 2003.

The performance fee shows similar results as the Sharpe ratio. The multivariate approach is always better than the univariate approach (UMS model or the buy-and-hold approach). The best model changes with the number of assets. For 30 assets, the MMS model is often preferred; for 60 assets, the MMS-IC model performs the best most of the time; and for 100 assets, the MMS-DC model is superior to the others. These results confirm that a more flexible correlation structure becomes valuable as the dimension increases. This is consistent with the forecast results in Table 2 showing the MMS-DC model as the best at higher dimension.

#### 5.4.2 Mean-Variance Portfolio

In this section, we allow the weight to change in the portfolio choice problem. This is only feasible for the multivariate models since the univariate model does not utilize the individual stock data.

We assume a simple and classical CAPM world with focus on the mean-variance portfolio and consider two strategies. The first strategy is to construct the global minimum variance portfolio by solving the following optimization problem

$$\min_{w_t} w_t' \widehat{\Sigma}_t w_t \quad \text{s.t.} \quad w_t' \iota = 1, \quad (36)$$

where  $w_t$  represents the portfolio weights,  $\widehat{\Sigma}_t$  is the predicted covariance matrix based on information up to day  $t - 1$  using model  $\mathcal{M}$ . The solution to equation (36) is

$$w_t^{gmv} = \frac{\widehat{\Sigma}_t^{-1} \iota}{\iota' \widehat{\Sigma}_t^{-1} \iota}. \quad (37)$$

A better model should achieve a lower portfolio variance ex post.

The second strategy is to maximize the Sharpe ratio as follows

$$\max_{w_t} \frac{\widehat{\mu}_t' w_t}{w_t' \widehat{\Sigma}_t w_t} \quad \text{s.t.} \quad w_t' \iota = 1, \quad (38)$$

where  $\widehat{\mu}_t$  is the predictive mean of excess returns conditional on the information up to time  $t - 1$  using model  $\mathcal{M}$ . The optimal weight is

$$w_t^{sr} = \frac{\Sigma_t^{-1} \widehat{\mu}_t}{\iota' \widehat{\Sigma}_t^{-1} \widehat{\mu}_t}. \quad (39)$$

A larger ex post Sharpe ratio indicates a better model.

Table 8 summaries the performance of global minimum variance and maximum Sharpe ratio portfolios in 30, 60 and 100 assets application. The global minimum variance column shows that the standard deviation from the multivariate model is much lower (10–25%) than the corresponding EW index. Hence, an additional benefit from the multivariate models lies in its ability to allow the investor to choose weights among multiple assets. Notice that all outcomes are out-of-sample, so the conclusion does not suffer from over-fitting.

The maximum Sharpe ratio columns show that the multivariate models favor higher return (about double the return from EW indices) accompanied by higher risk. The extra return compensates for the additional risk as shown by a higher Sharpe ratio than the indices. Overall, the MMS-CC model does the best for global minimum variance portfolio and maximizing the Sharpe ratios.

The investment applications demonstrate the value of our multivariate Markov switching models for its efficient utilization of information, flexible accommodation of correlation structure and the ability to select investing weights on individual assets.

### 5.4.3 S&P500

This section reports results using the models to make market timing decisions using the S&P500. Note that our previous results used multivariate models to trade an equally-weighted index constructed from the underlying stocks. In this example, we use the same models and underlying stocks to make trading decisions for the S&P500. The underlying stocks do not make up the S&P500 index and we have made no attempt to match them.

Table 9 reports results for the 60 asset case. The UMS-S&P500 model is a univariate MS model that uses S&P500 returns. The top numeric entry is the statistics for the buy and hold strategy. The Sharpe ratio shows large improvements from using the multivariate models. The performance fee that an investor would pay to move from the buy and hold strategy for the S&P500, to the model decisions based on 60 assets is also substantial. The overall favoured model in this application is the MMS-IC specification. Similar results are found for the other cases with the MMS favoured in the 30 asset case and the MMS-DC preferred in the 100 asset case. These last two set of results are reported in the Appendix. Even though the models do not use the underlying stocks of the S&P500 this example shows the value of using the multivariate models to make investment decisions on a broad market index.

## 6 Robustness Checks

In this section, we consider robustness checks for data and models.

### 6.1 Volatility Timing?

Are the trading results detailed above only a function of volatility timing or does the conditional mean dynamics play a role? To investigate this Table 10 reports performance fees for three of the multivariate models against the EW portfolio for 60 assets. We include the full version of the MMS-CC, MMS-IC and MMS-DC models as well as a restricted version that has a constant conditional mean across states. This has the hierarchical prior only for the variance parameters. In 6 of the 9 cases the model with regime changes in the conditional mean has a larger Sharpe ratio and larger performance fee. We conclude that the conditional variance and conditional mean play a role in regime identification and can lead to improved market timing.

## 6.2 Survival Bias

To check if our results are impacted by potential survival bias from the data (individual stocks), we obtained 30 industrial portfolios' monthly returns between 1926/07 and 2022/10 from Kenneth French's website. We report the log predictive likelihoods of MMS, MMS-CC, MMS-IC and MMS-DC in Table 11. It shows that moving from the benchmark MMS model to the MMS-CC/MMS-IC/MMS-DC leads to significant improvements in density forecasts of return vectors. Table 12 summarizes the market timing performance. The order of model performance aligns with the results of the applications to individual stocks. Our proposed multivariate models on various strategies have larger Shape ratios and higher performance fees than the univariate UMS and the benchmark MMS model. However, the buy and hold EW-30 portfolio is now more competitive with a large Sharpe ratio and often being preferred to the other models in terms of performance fees. Nevertheless, the multivariate models are preferred over the univariate model, UMS. Figure 14, shows the smoothed probabilities of bull regime identified from the three multivariate models. Regime identification is similar to the results from individual stocks. Hence, our conclusion is not affected by potential survival bias from three perspectives,: density forecast, market timing performance and regime identification.

## 6.3 Duration Dependence

The state variables in our multivariate models have constant transition probabilities. In this extension, we consider time-varying transition probability (Maheu & McCurdy 2000) by explicitly modelling regime durations.

Let  $d_{j,t}$  represent the duration of state  $j$  in period  $t$ . If  $s_t = s_{t-1} = j$ ,  $d_{j,t} = d_{j,t-1} + 1$ . Suppose the foundation duration has a Gamma distribution  $d \sim G(\kappa, \theta)$ , where  $\kappa$  is the shape and  $\theta$  is the scale parameter. It implies that  $E(d) = \kappa/\theta$  and  $V(d) = \kappa/\theta^2$ . Let the bear market duration follows  $G(\kappa_1, \theta_1)$ . The implied hazard function is

$$P(d_{1,t} = d \mid d_{1,t} \geq d) = \frac{\int_{d-1}^d f(x)}{1 - F(d-1)} = \frac{\gamma(\kappa_1, \theta_1 d) - \gamma(\kappa_1, \theta_1 (d-1))}{\Gamma(\kappa_1) - \gamma(\kappa_1, \theta_1 (d-1))}, \quad (40)$$

where  $\gamma(\kappa_1, \theta_1 (d-1))$  is the lower incomplete Gamma function. Similarly, for bull regimes, the duration prior  $G(\kappa_2, \theta_2)$  implies

$$P(d_{2,t} = d \mid d_{2,t} \geq d) = \frac{\int_{d-1}^d f(x)}{1 - F(d-1)} = \frac{\gamma(\kappa_2, \theta_2 d) - \gamma(\kappa_2, \theta_2 (d-1))}{\Gamma(\kappa_2) - \gamma(\kappa_2, \theta_2 (d-1))} \quad (41)$$

These hazard functions are the time-varying transition probabilities.

We apply the duration dependence to the most robust MMS-CC model.

$$\begin{aligned} R_t \mid s_t = j &\sim N(M_j, \Delta_j \Lambda \Delta_j), \\ \Lambda &= \Gamma Q \Gamma, \quad \Gamma = \text{diag} \left( \frac{1}{\sqrt{Q_{11}}}, \dots, \frac{1}{\sqrt{Q_{NN}}} \right), \\ P(s_{t+1} = j \mid s_t = k, d_t) &= p(k, j, d_t), \end{aligned} \quad (42)$$

where  $p(k, j, d_t)$  is determined based on (40) and (41).

The hyper-parameters  $\kappa_j$  and  $\theta_j$  are calibrated such that the implied mean duration is centred between 20 and 30 months with various dispersion settings shown in Table 13. The implied bull and bear market durations under four priors are broadly in line with the constant transition probability models. If  $\kappa$  is set as degenerate, the implied duration is longer, however, they are still within a reasonable range from the results of the constant transition probability models.

Figure 15 in the Appendix evaluates the sensitivity of state identification results to duration distribution priors. Unless  $\kappa_j$  values are fixed, the regime identification results are almost identical to the constant transition probability models. Even when  $\kappa_j$ 's are fixed, the most different period is before the 1950's. Therefore, adding a simple regime duration dependence may not add much value to bull and bear regime identification.

## 6.4 Intra-regime Components

Bull and bear markets could contain sub-cycles such as bull corrections and bear rallies in Maheu et al. (2012) or require more than two states (Baele et al. 2019). We extend the multivariate model to a new 4-state version. Taking the MMS-CC model as an example, its 4-state version is given as

$$R_t \mid s_t = k \sim N(M_k, \Delta_k \Lambda \Delta_k), \quad (43)$$

$$\Lambda = \Gamma Q \Gamma, \quad \Gamma = \text{diag} \left( \frac{1}{\sqrt{Q_{11}}}, \dots, \frac{1}{\sqrt{Q_{NN}}} \right), \quad (44)$$

$$P(s_{t+1} = j \mid s_t = k) = p_{kj}, \quad j, k = 1, 2, 3, 4 \quad (45)$$

Let  $s_t = 1, 2, 3$  and 4 represent bear, bear rally, bull correction and bear states, respectively. Following Maheu et al. (2012), we set the transition matrix as follows

$$P = \begin{bmatrix} p_{11} & p_{12} & 0 & p_{14} \\ p_{21} & p_{22} & 0 & p_{24} \\ p_{31} & 0 & p_{33} & p_{34} \\ p_{41} & 0 & p_{43} & p_{44} \end{bmatrix} \quad (46)$$

The same identification restrictions are applied to the aggregate mean as below.

$$\iota'_N M_1 < 0, \quad \iota'_N M_2 > 0, \quad \iota'_N M_3 < 0, \quad \iota'_N M_4 > 0 \quad (47)$$

and

$$\begin{aligned} E(\iota'_N R_t \mid s_t = 1, 2) &= \frac{\pi_1}{\pi_1 + \pi_2} \iota'_N M_1 + \frac{\pi_2}{\pi_1 + \pi_2} \iota'_N M_2 < 0 \\ E(\iota'_N R_t \mid s_t = 3, 4) &= \frac{\pi_3}{\pi_3 + \pi_4} \iota'_N M_3 + \frac{\pi_4}{\pi_3 + \pi_4} \iota'_N M_4 > 0 \end{aligned} \quad (48)$$

where  $\iota = [1, 1, 1, 1]'$ ,  $\pi = [\pi_1, \pi_2, \pi_3, \pi_4] = (A'A)^{-1}A'e$  and  $A' = [P' - I, \iota]$  and  $e' = [0, 0, 0, 0, 1]$ . The state interpretation is the same as the model in Maheu et al. (2012). State 1 is the bear state, State 2 is the bear rally, State 3 is the bull correction and State 4

is the bull state. State 1 and 2 comprise the bear regime, while state 3 and 4 are the bull regime.

We found in multivariate applications the 4-state structure very difficult to estimate. Standard MCMC algorithms for the mean parameters in state 3 always rejects the parameter restrictions.<sup>5</sup> Therefore, we cannot correctly draw inferences on regime identification. This may be due to the heterogeneous behaviour of individual stocks and their asynchronous movements. To capture such complex dynamics, we need a more sophisticated structure than the simple 4-state model, which has been applied only to an index before. We leave this for future work.

## 7 Conclusion

We propose a high-dimensional multivariate regime switching model to exploit the information in individual stock returns to identify bull and bear regimes. A novel geodesic Lagrangian Monte Carlo method is applied to speed up posterior sampling of the high dimensional correlation matrix. The multivariate approach contains important information on regime changes that a univariate approach neglect or obscures. First, individual stocks show a clear distinction in the level of conditional mean and variance over bull and bear markets. In addition, the cross-section of stock return moments displays important distributional differences which a univariate approach ignores. We design a hierarchical prior to learn about these cross-sectional differences over regimes. Second, the correlation structure improves regime identification. Our multivariate models provide superior density forecasts and portfolio decisions out-of-sample compared to methods that use only a portfolio. The larger dimension multivariate models appear to identify regimes more sharply than the univariate model. We conclude that there are significant benefits from using multivariate data to identify and forecast bull and bear stock market regimes.

---

<sup>5</sup>We could apply the individual sampler to draw each stock's mean via a truncated normal distribution. However, we view such a strong rejection rate as an indication of model failure.

## References

- Ang, A. & Bekaert, G. (2002), ‘International asset allocation with regime shifts’, *Review of Financial Studies* **15**, 1137–1187.
- Baele, L., Bekaert, G., Inghelbrecht, K. & Wei, M. (2019), ‘Flights to Safety’, *The Review of Financial Studies* **33**(2), 689–746.
- Barnard, J., McCulloch, R. & Meng, X.-L. (2000), ‘Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage’, *Statistica Sinica* pp. 1281–1311.
- Byrne, S. & Girolami, M. (2013), ‘Geodesic monte carlo on embedded manifolds’, *Scandinavian Journal of Statistics* **40**(4), 825–845.
- Candelon, B., Piplack, J. & Straetmans, S. (2008), ‘On measuring synchronization of bulls and bears: The case of east asia’, *Journal of Banking & Finance* **32**(6), 1022–1035.
- Chauvet, M. & Potter, S. (2000), ‘Coincident and leading indicators of the stock market’, *Journal of Empirical Finance* **7**, 87–111.
- Chen, S.-S. (2009), ‘Predicting the bear stock market: Macroeconomic variables as leading indicators’, *Journal of Banking & Finance* **33**(2), 211–223.
- Chib, S. (1996), ‘Calculating posterior distributions and modal estimates in markov mixture models’, *Journal of Econometrics* **75**(1), 79–97.
- Geweke, J. (2007), ‘Interpretation and inference in mixture models: Simple mcmc works’, *Computational Statistics & Data Analysis* **51**(7), 3529–3550.
- Geweke, J. & Amisano, G. (2010), ‘Comparing and evaluating bayesian predictive distributions of asset returns’, *International Journal of Forecasting* **26**(2), 216–230.
- Girolami, M. & Calderhead, B. (2011), ‘Riemann manifold langevin and hamiltonian monte carlo methods’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(2), 123–214.
- Guidolin, M. & Timmermann, A. (2006), ‘An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns’, *Journal of Applied Econometrics* **21**(1), 1–22.
- Guidolin, M. & Timmermann, A. (2007), ‘Asset allocation under multivariate regime switching’, *Journal of Economic Dynamics and Control* **31**(11), 3503–3544.
- Haase, F., Neuenkirch, M. et al. (2020), Forecasting stock market recessions in the us: Predictive modeling using different identification approaches, Technical report, University of Trier, Department of Economics.
- Hamilton, J. D. & Owyang, M. T. (2012), ‘The propagation of regional recessions’, *The Review of Economics and Statistics* **94**(4), 935–947.

- Harding, D. & Pagan, A. (2006), ‘Synchronization of cycles’, *Journal of Econometrics* **132**(1), 59–79.
- Holbrook, A., Lan, S., Vandenberg-Rodes, A. & Shahbaba, B. (2018), ‘Geodesic lagrangian monte carlo over the space of positive definite matrices: with application to bayesian spectral density estimation’, *Journal of statistical computation and simulation* **88**(5), 982–1002.
- Jasra, A., Holmes, C. C. & Stephens, D. A. (2005), ‘Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling’, *Statistical Science* **20**(1), 50–67.
- Kass, R. E. & Raftery, A. E. (1995), ‘Bayes factors’, *Journal of the american statistical association* **90**(430), 773–795.
- Kaufman, S. (2010), ‘Dating and forecasting turning points by bayesian clustering with dynamic structure: a suggestion with an application to austrian data’, *Journal of Applied Econometrics* **25**, 309–344.
- Kim, C.-J., Nelson, C. R. et al. (1999), ‘State-space models with regime switching: classical and gibbs-sampling approaches with applications’, *MIT Press Books* **1**.
- Kole, E. & Van Dijk, D. (2017), ‘How to identify and forecast bull and bear markets?’, *Journal of Applied Econometrics* **32**(1), 120–139.
- Lunde, A. & Timmermann, A. (2004), ‘Duration dependence in stock prices: An analysis of bull and bear markets’, *Journal of Business & Economic Statistics* **22**(3), 253–273.
- Maheu, J. M. & McCurdy, T. H. (2000), ‘Identifying bull and bear markets in stock returns’, *Journal of Business & Economic Statistics* **18**(1), 100–112.
- Maheu, J. M., McCurdy, T. H. & Song, Y. (2012), ‘Components of bull and bear markets: bull corrections and bear rallies’, *Journal of Business & Economic Statistics* **30**(3), 391–403.
- Maheu, J. M., McCurdy, T. H. & Song, Y. (2021), ‘Bull and bear markets during the covid-19 pandemic’, *Finance Research Letters* **42**, 102091.
- Merriam-Webster (2022), ‘The history of ‘bull’ and ‘bear’ markets’, <https://www.merriam-webster.com/words-at-play/the-origins-of-the-bear-and-bull-in-the-stock-market>. Accessed: 2022-03-16.
- Neal, R. M. et al. (2011), ‘Mcmc using hamiltonian dynamics’, *Handbook of markov chain monte carlo* **2**(11), 2.
- Paap, R., Segers, R. & van Dijk, D. (2009), ‘Do leading indicators lead peaks more than troughs?’, *Journal of Business & Economic Statistics* **27**(4), 528–543.

- Pagan, A. R. & Sossounov, K. A. (2003), ‘A simple framework for analysing bull and bear markets’, *Journal of applied econometrics* **18**(1), 23–46.
- Pelletier, D. (2006), ‘Regime switching for dynamic correlations’, *Journal of Econometrics* **131**(1), 445–473.
- Pesaran, M. H., Pettenuzzo, D. & Timmermann, A. (2006), ‘Forecasting time series subject to multiple structural breaks’, *The Review of Economic Studies* **73**(4), 1057–1084.
- Song, Y. (2014), ‘Modelling regime switching and structural breaks with an infinite hidden markov model’, *Journal of Applied Econometrics* **29**(5), 825–842.
- Tu, J. (2010), ‘Is regime switching in stock returns important in portfolio decisions?’, *Management Science* **56**(7), 1198–1215.
- Turner, C., Startz, R. & Nelson, C. (1989), ‘A markov model of heteroskedasticity, risk, and learning in the stock market’, *Journal of Financial Economics* **25**, 3–22.

Table 1: Algorithm Comparison

	M-H	gLMC	Eff. gain gLMC/M-H
N=30, T= 1116			
Time/ESS/ESS per sec	124/212/1.71	43/809/18.8	11.0
N=60, T= 1068			
Time/ESS/ESS per sec	265/29/0.11	620/534/0.86	7.8
N=100, T= 816			
Time/ESS/ESS per sec	537/16/0.03	4707/382/0.08	2.7

The simulation is carried out in Matlab(2020b) on a Windows 10 operating system with Intel Core i7-4790 CPU @ 3.60GHz. For the M-H algorithm, the total number of iterations is 200,000, and the first half of the samples is discarded as burn-in. We apply the thinning method to select 1 out of each 100 iterations for inference. The step size parameter  $v$  is set as  $10^5$  for  $N = 30$  and  $60$ ;  $5 \times 10^5$  for  $N = 100$ . For the gLMC, we set the total number of iterations as 1100 and discard the first 100 as burn-in. The leap size is set as 0.01, and the number of leaps is set as  $L = 10$  for  $N = 30$  and  $60$ ; and  $L = 5$  for  $N = 100$ .

Table 2: Density Forecast of Return Vectors

	30 assets	60 assets	100 assets
UMS	-24106.55	-47711.06	-47392.79
MMS	-23032.29	-46329.33	-48049.96
MMS-CC	-22596.61	-43754.81	-43303.09
MMS-IC	<b>-22249.61</b>	-43420.93	-43020.77
MMS-DC	-22263.41	<b>-43190.61</b>	<b>-42917.07</b>

This table reports log predictive likelihoods of the models in this paper. The OOS periods for 30/60/100 assets are 1971/01-2020/12, 1971/01-2020/12 and 1991/01-2020/12, respectively.

Table 3: Posterior Means of Parameters in MMS-DC

Parameter	30 assets	60 assets	100 assets
$m_{bear}$ : mean of $\mu_i$ in bear market	-0.035	-0.017	-0.037
$m_{bul}$ : mean of $\mu_i$ in bull market	0.088	0.091	0.089
$v_{bear}^2$ : variance of $\mu_i$ in bear market	0.018	0.011	0.010
$v_{bull}^2$ : variance of $\mu_i$ in bull market	0.015	0.009	0.006
$\zeta_{bear}$ : mean of $\log(\sigma_i)$ in bear market	0.480	0.557	0.392
$\zeta_{bull}$ : mean of $\log(\sigma_i)$ in bull market	-0.113	-0.106	-0.153
$b_{bear}^2$ : variance of $\log(\sigma_i)$ in bear market	0.060	0.064	0.073
$b_{bull}^2$ : variance of $\log(\sigma_i)$ in bull market	0.081	0.062	0.070

This table reports posterior means of hyper-parameters for the MMS-DC model in 30, 60 and 100 assets applications.

Table 4: Posterior Means of Transition Matrix in MMS-DC

$$\bar{P}_{30} = \begin{pmatrix} 0.80 & 0.04 & 0.09 & 0.06 \\ 0.05 & 0.82 & 0.01 & 0.13 \\ 0.10 & 0.01 & 0.65 & 0.24 \\ 0.02 & 0.03 & 0.07 & 0.88 \end{pmatrix}, \bar{P}_{60} = \begin{pmatrix} 0.84 & 0.05 & 0.06 & 0.05 \\ 0.07 & 0.74 & 0.01 & 0.17 \\ 0.14 & 0.02 & 0.57 & 0.27 \\ 0.01 & 0.04 & 0.01 & 0.94 \end{pmatrix}, \bar{P}_{100} = \begin{pmatrix} 0.55 & 0.17 & 0.02 & 0.27 \\ 0.31 & 0.54 & 0.02 & 0.13 \\ 0.03 & 0.05 & 0.53 & 0.40 \\ 0.05 & 0.01 & 0.02 & 0.92 \end{pmatrix}$$

Table 5: Market Timing Portfolio - 30 Assets (Out-of-Sample: 1971/01-2020/12)

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
EW-30	0.0975	0.1668	0.3194		
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
UMS	0.0755	0.1550	0.2019	-134.94	-19.50
MMS	0.1056	0.1176	<b>0.5222</b>	<b>363.47</b>	<b>731.80</b>
MMS-CC	0.0943	0.1250	0.4007	218.45	547.50
MMS-IC	0.0987	0.1327	0.4105	220.88	496.84
MMS-DC	0.0971	0.1309	0.4044	215.68	504.61
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
UMS	0.0622	0.1513	0.1193	-242.94	-94.75
MMS	0.0992	0.1156	0.4760	308.80	689.53
MMS-CC	0.1013	0.1149	<b>0.4976</b>	<b>332.18</b>	<b>715.48</b>
MMS-IC	0.0945	0.1299	0.3876	194.88	490.87
MMS-DC	0.0986	0.1279	0.4254	245.65	554.42
<i>Panel C: Strategy II</i>					
UMS	0.0778	0.1471	0.2287	-62.99	116.09
MMS	0.0975	0.1100	<b>0.4846</b>	<b>319.09</b>	<b>733.68</b>
MMS-CC	0.0952	0.1129	0.4520	282.85	680.70
MMS-IC	0.0942	0.1248	0.4005	218.33	548.21
MMS-DC	0.0945	0.1223	0.4112	234.24	579.98

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index as the benchmark. The out of sample period is from 1971/01 to 2020/12. A bold number means the optimal value in the corresponding column and panel.

Table 6: Market Timing Portfolio - 60 Assets (Out-of-sample: 1971/01-2020/12)

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
EW-60	0.0983	0.1577	0.3431	-	-
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
UMS	0.0862	0.1489	0.2826	-59.70	22.88
MMS	0.0914	0.1358	0.3478	62.45	239.14
MMS-CC	0.0921	0.1255	0.3821	126.60	376.38
MMS-IC	0.0987	0.1296	<b>0.4206</b>	<b>172.67</b>	<b>396.62</b>
MMS-DC	0.0916	0.1287	0.3682	105.94	336.17
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
UMS	0.0725	0.1453	0.1947	-174.11	-59.70
MMS	0.0940	0.1349	0.3696	93.72	276.72
MMS-CC	0.0924	0.1241	0.3888	137.12	396.56
MMS-IC	0.0981	0.1279	<b>0.4214</b>	<b>175.76</b>	<b>411.48</b>
MMS-DC	0.0892	0.1285	0.3506	83.51	315.18
<i>Panel C: Strategy II</i>					
UMS	0.0849	0.1429	0.2852	-36.91	93.28
MMS	0.0930	0.1303	0.3747	108.78	324.85
MMS-CC	0.0915	0.1178	0.4013	158.08	<b>456.65</b>
MMS-IC	0.0966	0.1232	<b>0.4252</b>	<b>184.69</b>	451.18
MMS-DC	0.0909	0.1237	0.3780	125.30	388.45

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index as the benchmark. The out of sample period is from 1971/01 to 2020/12. A bold number means the optimal value in the corresponding column and panel.

Table 7: Market Timing Portfolio - 100 Assets (Out-of-sample: 1991/01-2020/12)

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
EW-100	0.1018	0.1521	0.5087	-	-
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
UMS	0.0907	0.1266	0.5233	-27.64	215.86
MMS	0.0946	0.1123	0.6254	135.23	411.31
MMS-CC	0.0731	0.0876	0.5554	23.20	430.73
MMS-IC	0.0688	0.0884	0.5022	-22.32	381.77
MMS-DC	0.0901	0.1001	<b>0.6563</b>	<b>144.50</b>	<b>489.98</b>
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
UMS	0.0834	0.1195	0.4935	-9.24	226.24
MMS	0.0946	0.1123	0.6254	<b>135.23</b>	411.31
MMS-CC	0.0698	0.0866	0.5240	-5.80	406.42
MMS-IC	0.0703	0.0879	0.5221	-5.56	400.92
MMS-DC	0.0865	0.0983	<b>0.6313</b>	116.03	<b>471.42</b>
<i>Panel C: Strategy II</i>					
UMS	0.0881	0.1149	0.5540	131.57	396.16
MMS	0.0956	0.1120	0.6357	252.16	532.18
MMS-CC	0.0726	0.0866	0.5557	202.87	616.99
MMS-IC	0.0754	0.0884	0.5763	207.35	613.34
MMS-DC	0.0882	0.0955	<b>0.6678</b>	<b>303.31</b>	<b>675.53</b>

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index as the benchmark. The out of sample period is from 1991/01 to 2020/12. A bold number means the optimal value in the corresponding column and panel.

Table 8: Summary of Mean-Variance Portfolio Performance

Portfolio	Global Minimum Variance		Max Sharpe Ratio		
		Stdev	Mean	St. Dev	Sharpe ratio
<i>Panel A: 30 Assets</i>					
EW-30		0.1668	0.0975	0.1668	0.3194
MMS		0.1501	0.1580	0.3271	0.4830
MMS-CC		0.1431	0.1566	0.3109	<b>0.5037</b>
MMS-IC		<b>0.1427</b>	0.1703	0.3558	0.4785
MMS-DC		0.1434	0.1691	0.3672	0.4605
<i>Panel B: 60 Assets</i>					
EW-60		0.1577	0.0983	0.1577	0.3431
MMS		0.1349	0.2835	0.4122	0.6371
MMS-CC		<b>0.1333</b>	0.2864	0.4426	<b>0.6472</b>
MMS-IC		0.1404	0.3182	0.6026	0.5280
MMS-DC		0.1345	0.3190	0.5489	0.5860
<i>Panel C: 100 Assets</i>					
EW-100		0.1521	0.1018	0.1521	0.5087
MMS		0.1290	0.2217	0.4568	0.4854
MMS-CC		<b>0.1112</b>	0.1959	0.2737	<b>0.7193</b>
MMS-IC		0.1167	0.2143	0.2982	0.7187
MMS-DC		0.1182	0.2055	0.3380	0.6080

This table provides the annual standard deviation of the global minimum variance portfolio and the mean, standard deviation and Sharpe ratio of the portfolio that maximizes the Sharpe ratio. The OOS periods for 30/60/100 assets are 1971/01-2020/12, 1971/01-2020/12 and 1991/01-2020/12, respectively. A bold number means the optimal value in the corresponding column and panel.

Table 9: Market Timing Portfolio using the S&amp;P500 Index - 60 Assets

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
S&P500	0.0742	0.1528	0.1963	-	-
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
UMS-S&P500	0.0576	0.1433	0.0936	-105.48	-25.16
MMS	0.0796	0.1317	0.2691	176.18	335.40
MMS-CC	0.0821	0.1225	0.3099	248.59	468.10
MMS-IC	0.0873	0.1264	<b>0.3412</b>	<b>281.70</b>	<b>477.14</b>
MMS-DC	0.0829	0.1254	0.3085	242.17	443.89
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
UMS-S&P500	0.0557	0.1420	0.0811	-117.00	-26.83
MMS	0.0828	0.1303	0.2963	214.85	383.00
MMS-CC	0.0812	0.1204	0.3080	250.88	484.45
MMS-IC	0.0884	0.1254	<b>0.3526</b>	<b>297.74</b>	<b>499.70</b>
MMS-DC	0.0796	0.1250	0.2833	211.43	416.00
<i>Panel C: Strategy II</i>					
UMS-S&P500	0.0614	0.1376	0.1253	-35.35	86.47
MMS	0.0785	0.1265	0.2711	191.54	385.54
MMS-CC	0.0787	0.1147	0.3015	251.93	<b>518.74</b>
MMS-IC	0.0844	0.1203	<b>0.3346</b>	<b>283.19</b>	517.12
MMS-DC	0.0810	0.1205	0.3058	247.86	480.55

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index as the benchmark. The out of sample period is from 1971/01 to 2020/12. A bold number means the optimal value in the corresponding column and panel.

Table 10: Market Timing Portfolio - State-dependent mean v.s. constant mean

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
MMS-CC	0.0921	0.1255	<b>0.3821</b>	<b>126.60</b>	<b>376.38</b>
MMS-CC (constant mean)	0.0896	0.1250	0.3637	104.16	357.40
MMS-IC	0.0987	0.1296	<b>0.4206</b>	<b>172.67</b>	<b>396.62</b>
MMS-IC (constant mean)	0.0931	0.1286	0.3807	122.33	353.46
MMS-DC	0.0916	0.1287	0.3682	105.94	336.17
MMS-DC (constant mean)	0.0936	0.1236	<b>0.4003</b>	<b>152.07</b>	<b>415.09</b>
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
MMS-CC	0.0924	0.1241	<b>0.3888</b>	<b>137.12</b>	<b>396.56</b>
MMS-CC (constant mean)	0.0906	0.1239	0.3748	120.00	380.94
MMS-IC	0.0981	0.1279	<b>0.4214</b>	<b>175.76</b>	<b>411.48</b>
MMS-IC (constant mean)	0.0913	0.1281	0.3676	106.16	340.46
MMS-DC	0.0892	0.1285	0.3506	83.51	315.18
MMS-DC (constant mean)	0.0905	0.1226	<b>0.3779</b>	<b>125.60</b>	<b>395.18</b>
<i>Panel C: Strategy II</i>					
MMS-CC	0.0915	0.1178	<b>0.4013</b>	<b>158.08</b>	<b>456.65</b>
MMS-CC (constant mean)	0.0905	0.1181	0.3920	146.93	444.26
MMS-IC	0.0966	0.1232	<b>0.4252</b>	<b>184.69</b>	<b>451.18</b>
MMS-IC (constant mean)	0.0918	0.1229	0.3879	138.40	406.79
MMS-DC	0.0909	0.1237	0.3780	125.30	388.45
MMS-DC (constant mean)	0.0921	0.1164	<b>0.4115</b>	<b>171.16</b>	<b>478.40</b>

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index as the benchmark. The results are based on 60 assets sample and the out of sample period is from 1971/01 to 2020/12.

Table 11: Density Forecast of Return Vectors - Industrial Portfolios

Models	Log predictive likelihoods
MMS	-12443.22
MMS-CC	-10835.84
MMS-IC	<b>-10451.14</b>
MMS-DC	-10589.17

This table provides log predictive likelihoods of four multivariate models applied to industrial portfolio returns. The out of sample period is from 1972/11 to 2022/10.

Table 12: Market Timing Performance - Industrial Portfolios

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
EW-30	0.1326	0.1689	0.5236		
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
UMS	0.1062	0.1568	0.3953	-173.24	-44.82
MMS	0.1019	0.1382	0.4176	-103.97	175.33
MMS-CC	0.1056	0.1370	0.4487	-59.93	227.63
MMS-IC	0.1110	0.1399	<b>0.4775</b>	<b>-24.28</b>	<b>239.63</b>
MMS-DC	0.1057	0.1374	0.4478	-61.46	223.13
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
UMS	0.0953	0.1485	0.3440	-230.51	-30.72
MMS	0.0999	0.1378	0.4041	-122.09	160.08
MMS-CC	0.1064	0.1350	0.4610	-40.60	261.83
MMS-IC	0.1080	0.1375	0.4640	-39.44	244.37
MMS-DC	0.1101	0.1293	<b>0.5101</b>	<b>27.05</b>	<b>368.80</b>
<i>Panel C: Strategy II</i>					
UMS	0.1066	0.1452	0.4303	-94.70	132.95
MMS	0.1065	0.1292	0.4824	-6.37	339.44
MMS-CC	0.1123	0.1289	0.5287	52.80	399.18
MMS-IC	0.1130	0.1325	0.5196	40.02	360.91
MMS-DC	0.1133	0.1278	<b>0.5413</b>	<b>68.56</b>	<b>422.00</b>

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index (buy and hold) as the benchmark. The UMS model is applied to the average returns of the 30 industrial portfolios. The out of sample period is from 1972/11 to 2022/10. A bold number means the optimal value in the corresponding column and panel.

Table 13: Posterior Means of Parameters in Duration Distribution

	$\kappa \sim G(4, 1)$ $\theta \sim G(0.1, 0.5)$	$\kappa \sim G(40, 10)$ $\theta \sim G(1, 5)$	$\kappa \sim G(60, 10)$ $\theta \sim G(1, 5)$	$\kappa_1 = \kappa_2 = 4$ $\theta \sim G(1, 5)$
$\kappa_1$	0.333	0.912	1.370	-
$\kappa_2$	0.389	0.802	1.144	-
$\theta_1$	0.102	0.257	0.349	0.605
$\theta_2$	0.032	0.062	0.079	0.165
Duration (bear)	3.26	3.55	3.92	6.61
Duration (bull)	12.13	12.94	14.43	24.27

This table reports the posterior averages of duration distribution parameters and implied bull and bear market durations under four priors.

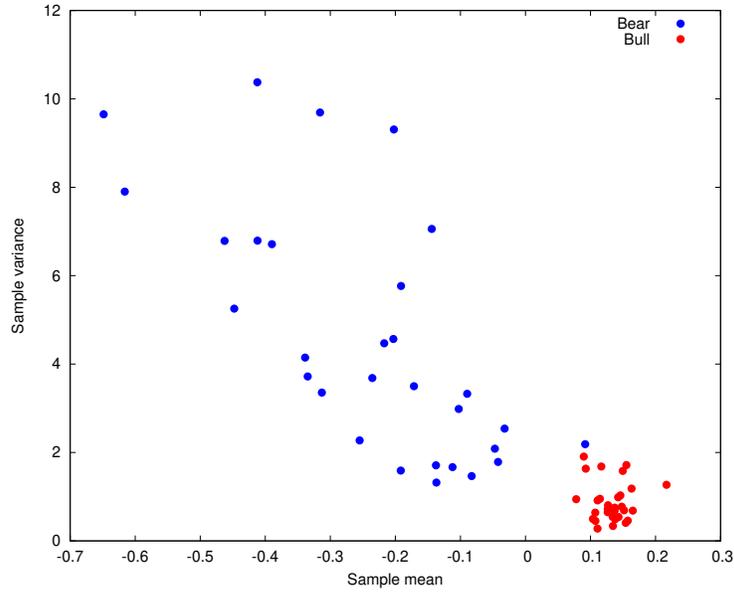


Figure 1: Scatter plot of sample means and variances of 30 assets in bull and bear regimes.

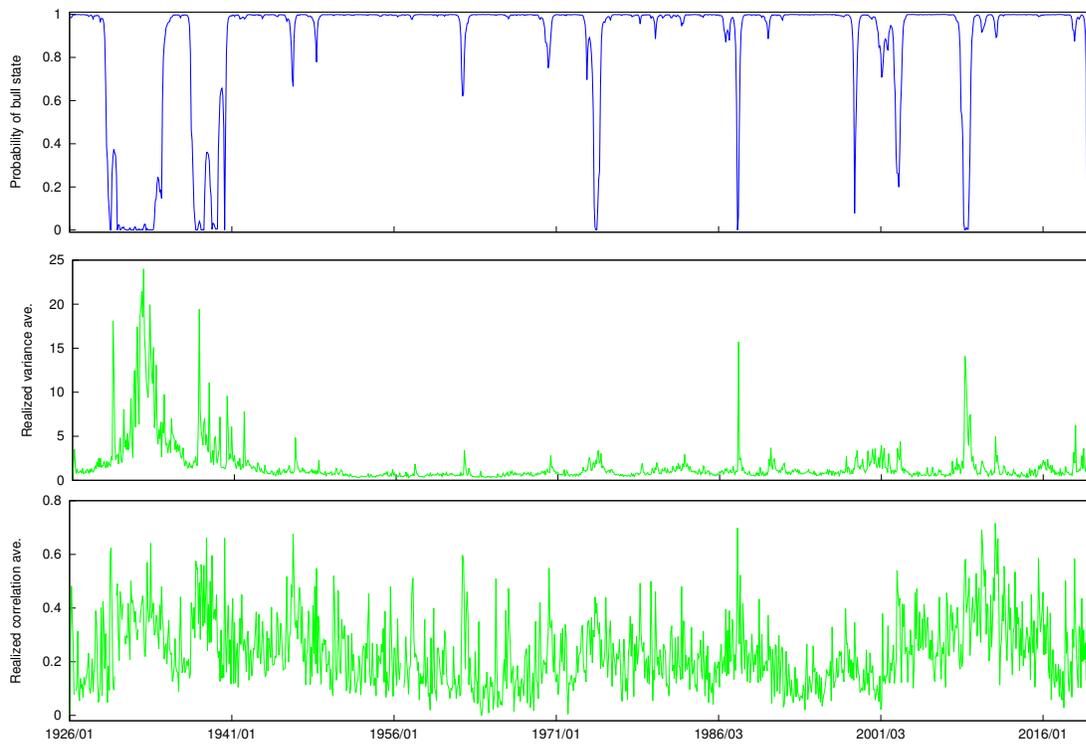


Figure 2: Bull and bear regimes based on UMS by using the S&P500 index.

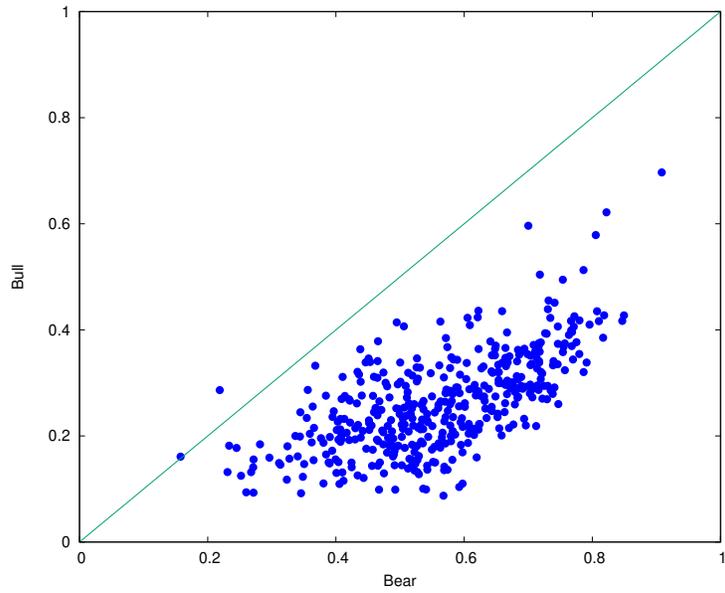


Figure 3: Sample correlation of stock pairs in bull and bear regimes (30 assets case)

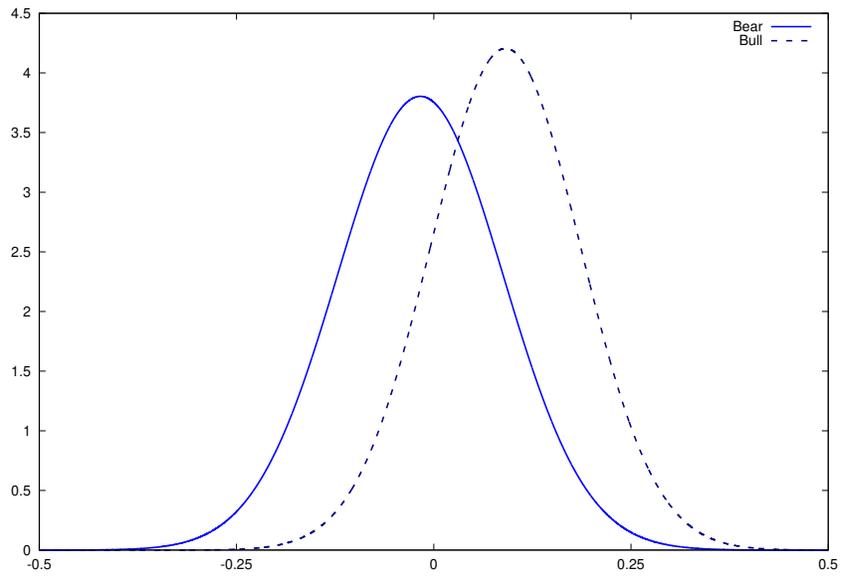


Figure 4: Distribution of  $\mu_{ibear}$  and  $\mu_{ibull}$  from MMS-DC model (60 assets case).

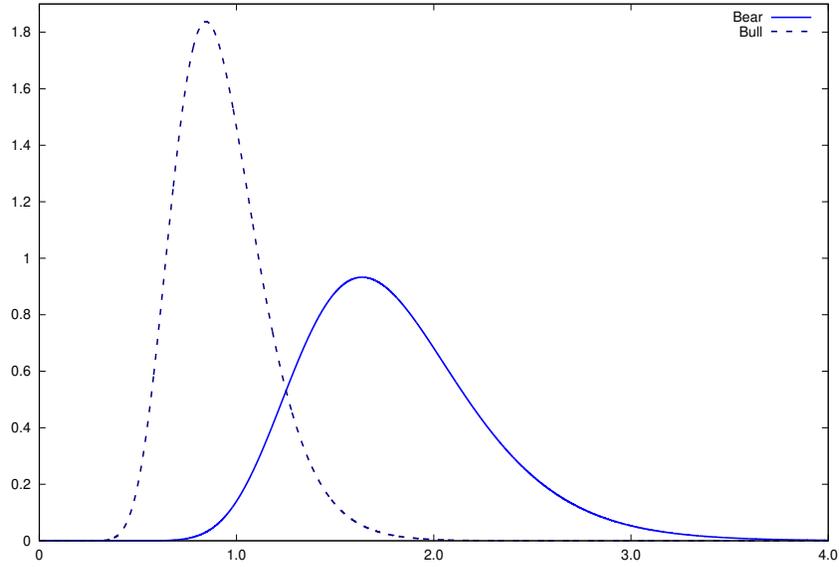


Figure 5: Distribution of  $\sigma_{ibear}$  and  $\sigma_{ibull}$  from MMS-DC model (60 assets case).

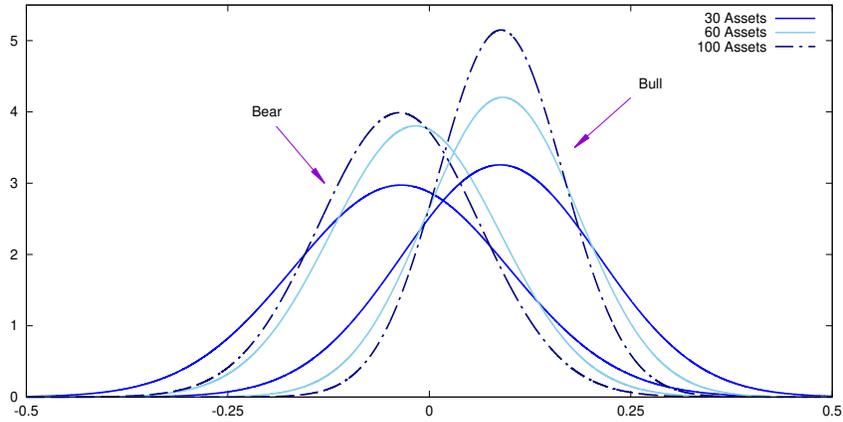


Figure 6: Distributions of  $\mu_{ibear}, \mu_{ibull}$  from MMS-DC model in 30, 60 and 100 assets cases.

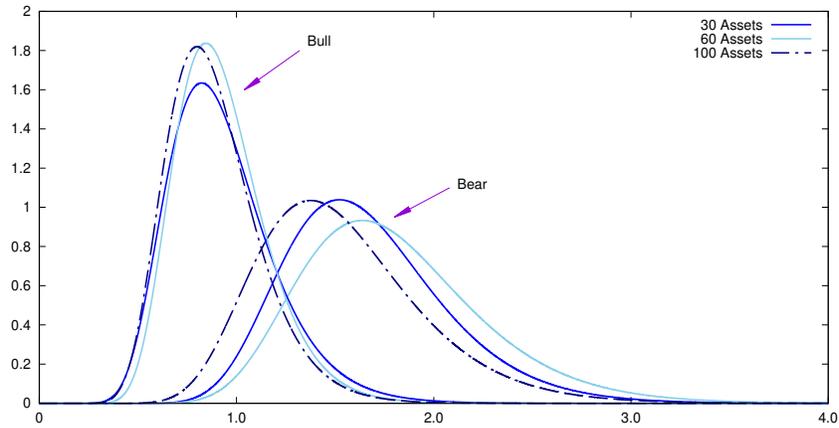


Figure 7: Distribution of  $\sigma_{ibear}, \sigma_{ibull}$  from MMS-DC model in 30, 60 and 100 assets cases.

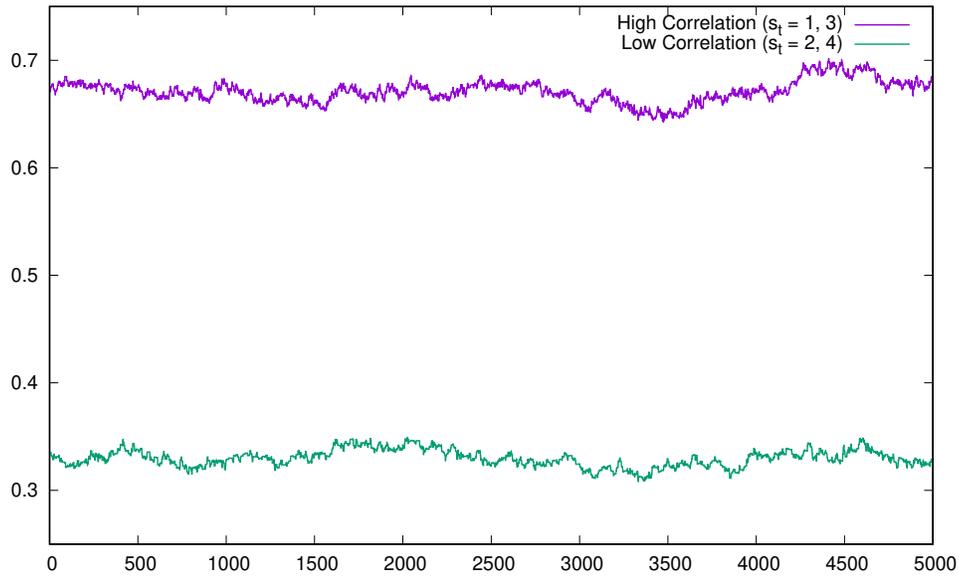


Figure 8: Trace plot of the average of correlations in each regimes (60 assets case).

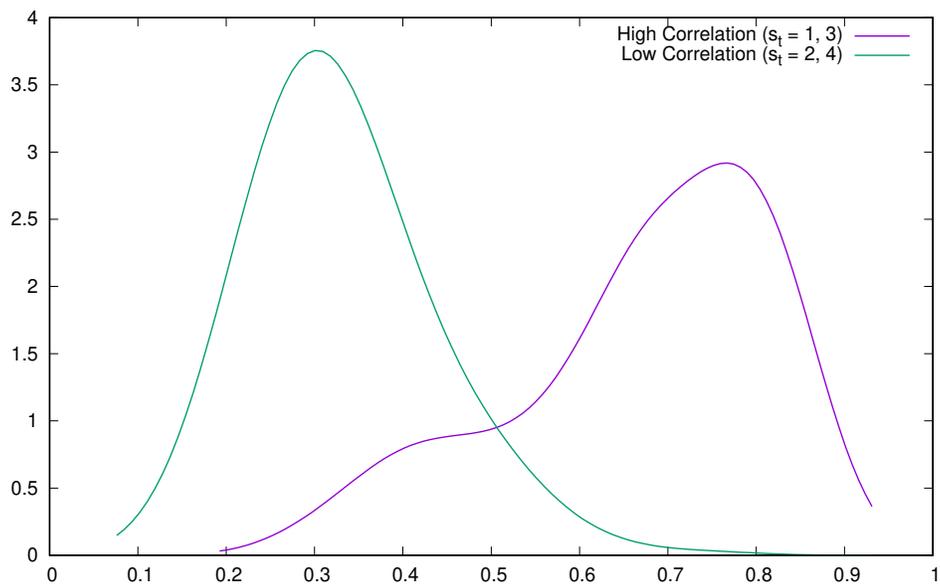


Figure 9: Density estimates of pairwise correlation coefficients in each regime (60 assets case).

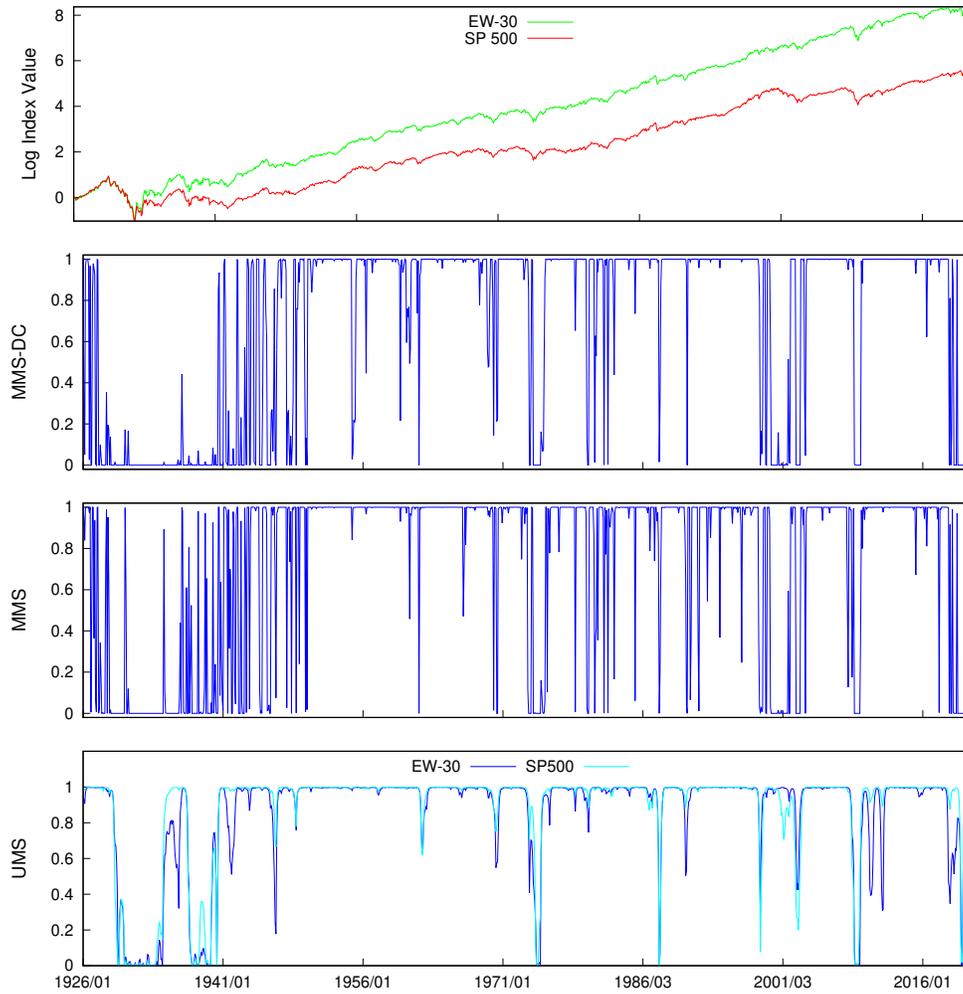


Figure 10: 30 asset case - smoothed probability of bull regime (from top to bottom: index cumulative values, MMS-DC, MMS and UMS).

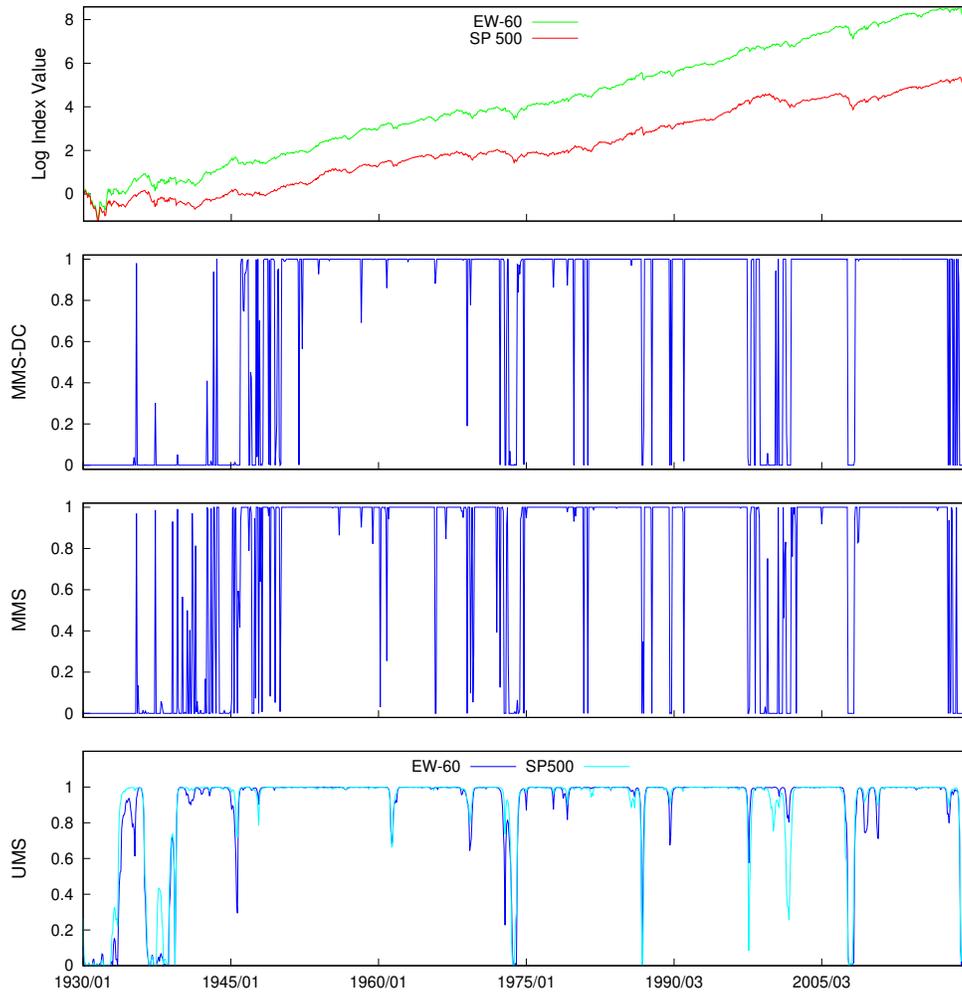


Figure 11: 60 assets case - smoothed probability of bull regime (from top to bottom: index cumulative values, MMS-DC, MMS and UMS).

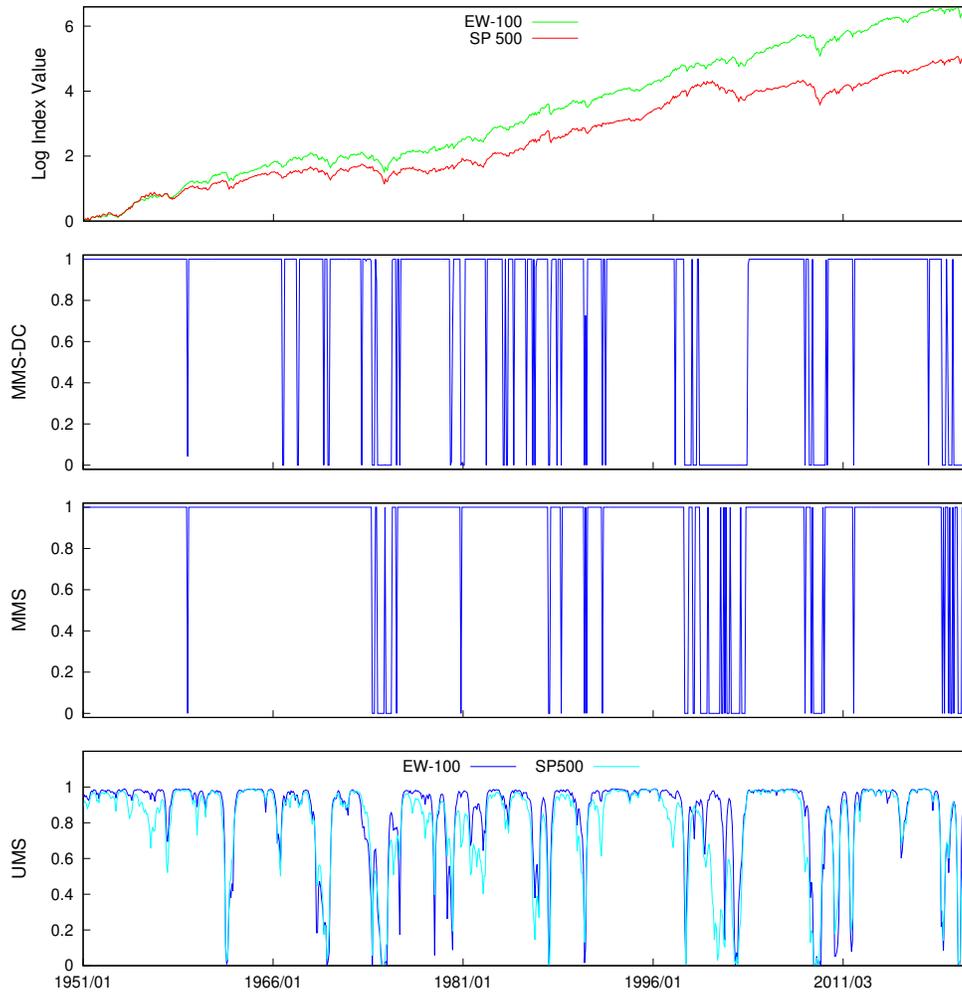


Figure 12: 100 assets case - smoothed probability of bull regime (from top to bottom: index cumulative values, MMS-DC, MMS and UMS).

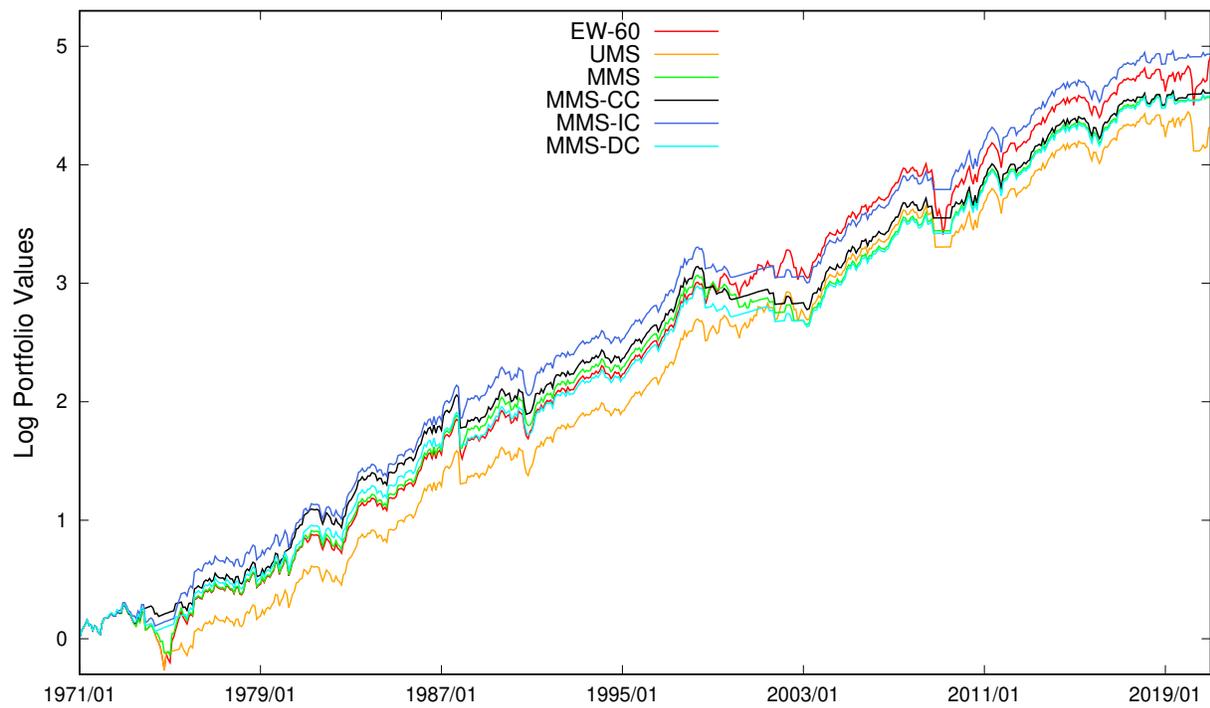


Figure 13: Log values of market timing portfolios (60 assets, investment strategy with  $\tau = 0.5$ )

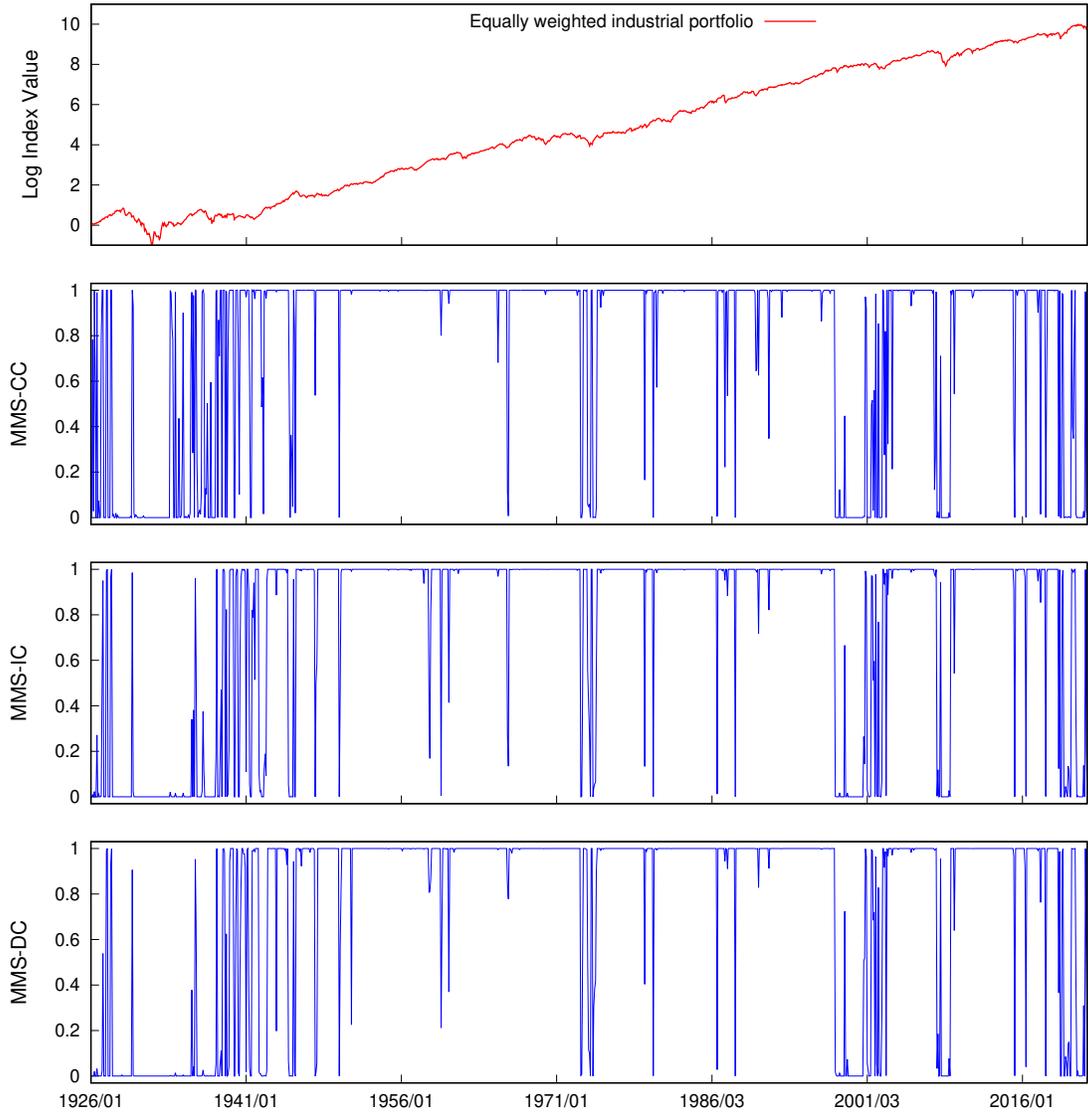


Figure 14: Smoothed probability of bull regime (Industrial Portfolios)

# A Appendix

## A.1 Data

The datasets of 30, 60 and 100 assets are formed from the first 30, the first 60 and the 100 assets below, respectively.

Table 14: Equity List

	<b>PERMNO</b>	<b>TICKER</b>	<b>COMPANY NAME</b>
1	10065	ADX	ADAMS EXPRESS CO
2	10145	HON	HONEYWELL INTERNATIONAL INC
3	10516	ADM	ARCHER DANIELS MIDLAND CO
4	10866	CAL	CALERES INC
5	10874	BC	BRUNSWICK CORP
6	10890	UIS	UNISYS CORP
7	11308	KO	COCA COLA CO
8	11404	ED	CONSOLIDATED EDISON INC
9	11674	DTE	D T E ENERGY CO
10	11762	ETN	EATON CORP PLC
11	11850	XOM	EXXON MOBIL CORP
12	12036	GATX	G A T X CORP
13	12052	GD	GENERAL DYNAMICS CORP
14	12060	GE	GENERAL ELECTRIC CO
15	12431	IR	INGERSOLL RAND PLC
16	12490	IBM	INTERNATIONAL BUSINESS MACHS COR
17	12503	NAV	NAVISTAR INTERNATIONAL CORP
18	12570	ITT	I T T INC
19	12650	KSU	KANSAS CITY SOUTHERN
20	12781	SR	SPIRE INC
21	13303	NL	N L INDUSTRIES INC
22	13610	OLN	OLIN CORP
23	13688	PCG	P G & E CORP
24	13856	PEP	PEPSICO INC
25	13901	MO	ALTRIA GROUP INC
26	13928	COP	CONOCOPHILLIPS
27	13936	PVH	P V H CORP
28	13987	PW	POWER REIT
29	14277	SLB	SCHLUMBERGER LTD
30	14541	CVX	CHEVRON CORP NEW
31	14752	TPL	TEXAS PACIFIC LAND TRUST
32	14795	TKR	TIMKEN COMPANY
33	14816	TR	TOOTSIE ROLL INDS INC

Continued on next page

Table 14 – continued from previous page

	<b>PERMNO</b>	<b>TICKER</b>	<b>COMPANY NAME</b>
34	15069	MRO	MARATHON OIL CORP
35	15202	VMC	VULCAN MATERIALS CO
36	15456	FL	FOOT LOCKER INC
37	15579	TXN	TEXAS INSTRUMENTS INC
38	15720	EIX	EDISON INTERNATIONAL
39	16432	GT	GOODYEAR TIRE & RUBBER CO
40	16555	UVV	UNIVERSAL CORPORATION
41	16600	HSY	HERSHEY CO
42	16678	KR	KROGER COMPANY
43	17005	CVS	C V S HEALTH CORP
44	17144	GIS	GENERAL MILLS INC
45	17478	SPGI	S & P GLOBAL INC
46	17523	SPA	SPARTON CORP
47	17726	CCK	CROWN HOLDINGS INC
48	17750	KMB	KIMBERLY CLARK CORP
49	17830	UTX	UNITED TECHNOLOGIES CORP
50	17929	UGI	U G I CORP NEW
51	17961	BGG	BRIGGS & STRATTON CORP
52	18075	AP	AMPCO PITTSBURGH CORP
53	18091	CW	CURTISS WRIGHT CORP
54	18163	PG	PROCTER & GAMBLE CO
55	18403	JCP	PENNEY J C CO INC
56	18411	SO	SOUTHERN CO
57	18438	SCX	STARRETT L S CO
58	18542	CAT	CATERPILLAR INC
59	18622	GAM	GENERAL AMERICAN INVESTORS INC
60	18649	BCO	BRINKS CO
61	18729	CL	COLGATE PALMOLIVE CO
62	18956	TY	TRI CONTINENTAL CORP
63	18964	PEO	ADAMS NATURAL RESOURCES FUND INC
64	19166	FMC	F M C CORP
65	19350	DE	DEERE & CO
66	19393	BMY	BRISTOL MYERS SQUIBB CO
67	19502	WBA	WALGREENS BOOTS ALLIANCE INC
68	19561	BA	BOEING CO
69	19721	VVI	VIAD CORP
70	19828	WEN	WENDYS CO
71	20204	CR	CRANE CO
72	20415	LUK	LEUCADIA NATIONAL CORP
73	20482	ABT	ABBOTT LABORATORIES
74	20618	CRS	CARPENTER TECHNOLOGY CORP
75	20626	DOW	DOW CHEMICAL CO

Continued on next page

Table 14 – continued from previous page

	<b>PERMNO</b>	<b>TICKER</b>	<b>COMPANY NAME</b>
76	21020	AAL	AMERICAN AIRLINES GROUP INC
77	21055	GCO	GENESCO INC
78	21135	FOE	FERRO CORP
79	21178	LMT	LOCKHEED MARTIN CORP
80	21186	WRK	WESTROCK CO
81	21207	NEM	NEWMONT MINING CORP
82	21776	EXC	EXELON CORP
83	21792	CNP	CENTERPOINT ENERGY INC
84	21928	IDA	IDACORP INC
85	21936	PFE	PFIZER INC
86	22103	EMR	EMERSON ELECTRIC CO
87	22111	JNJ	JOHNSON & JOHNSON
88	22293	GLW	CORNING INC
89	22509	PPG	P P G INDUSTRIES INC
90	22517	PPL	P P L CORP
91	22592	MMM	3M CO
92	22752	MRK	MERCK & CO INC NEW
93	22779	MSI	MOTOROLA SOLUTIONS INC
94	23026	FE	FIRSTENERGY CORP
95	23042	EDE	EMPIRE DISTRICT ELEC CO
96	23085	SCG	SCANA CORP NEW
97	23229	CMS	C M S ENERGY CORP
98	23536	WEC	W E C ENERGY GROUP INC
99	23579	TXT	TEXTRON INC
100	23712	PEG	PUBLIC SERVICE ENTERPRISE GP INC

## A.2 Market Timing Strategy Based on the S&P500 Index

Tables 15 and 16 report the performance of market timing strategy based on trading the S&P500 index. For example, in the 30-asset case, the univariate MS model is applied to the S&P500 index and the investor buys the S&P500 index if entering the market. These results compliment Table 15 in the main text.

Table 15: Market Timing Portfolio using the S&amp;P500 Index - 30 Assets

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
S&P500	0.0742	0.1528	0.1963		
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
UMS-S&P500	0.0594	0.1429	0.1068	-84.71	-1.43
MMS	0.0972	0.1092	<b>0.4858</b>	<b>459.58</b>	<b>753.73</b>
MMS-CC	0.0889	0.1162	0.3848	347.90	606.82
MMS-IC	0.0870	0.1230	0.3481	295.80	513.10
MMS-DC	0.0865	0.1218	0.3476	297.08	521.95
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
UMS-S&P500	0.0533	0.1399	0.0649	129.56	-23.76
MMS	0.0929	0.1068	0.4561	426.36	733.91
MMS-CC	0.0945	0.1060	<b>0.4751</b>	<b>445.03</b>	<b>755.17</b>
MMS-IC	0.0849	0.1200	0.3393	289.96	526.28
MMS-DC	0.0885	0.1186	0.3739	332.52	576.83
<i>Panel C: Strategy II</i>					
UMS-S&P500	0.0607	0.1357	0.1221	-32.04	102.83
MMS	0.0855	0.1020	<b>0.4046</b>	<b>372.94</b>	<b>706.49</b>
MMS-CC	0.0854	0.1047	0.3942	361.01	680.20
MMS-IC	0.0824	0.1156	0.3312	286.27	549.10
MMS-DC	0.0836	0.1134	0.3477	308.05	583.46

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index as the benchmark. The out of sample period is from 1971/01 to 2020/12. A bold number means the optimal value in the corresponding column and panel.

Table 16: Market Timing Portfolio using the S&amp;P500 Index - 100 Assets

Model	Mean	St. Dev.	Sharpe ratio	$\Delta (\eta = 5)$	$\Delta (\eta = 10)$
S&P500	0.0810	0.1464	0.3867	-	-
<i>Panel A: Strategy I with <math>\tau = 0.5</math></i>					
UMS-S&P500	0.0776	0.1445	0.3683	-39.49	89.58
MMS	0.0888	0.1072	0.6004	<b>274.79</b>	532.62
MMS-CC	0.0659	0.0822	0.5045	142.80	522.48
MMS-IC	0.0604	0.0821	0.4376	88.27	468.95
MMS-DC	0.0828	0.0954	<b>0.6123</b>	263.92	<b>583.19</b>
<i>Panel B: Strategy I with <math>\tau = 0.75</math></i>					
UMS-S&P500	0.0734	0.1394	0.3513	-70.44	119.94
MMS	0.0888	0.1072	<b>0.6004</b>	<b>274.79</b>	532.62
MMS-CC	0.0614	0.0801	0.4623	106.08	495.38
MMS-IC	0.0620	0.0816	0.4603	106.24	489.10
MMS-DC	0.0799	0.0946	0.5870	238.30	<b>561.67</b>
<i>Panel C: Strategy II</i>					
UMS-S&P500	0.0749	0.1300	0.3880	30.50	153.31
MMS	0.0868	0.1072	0.5817	<b>254.84</b>	513.01
MMS-CC	0.0634	0.0804	0.4848	124.55	512.20
MMS-IC	0.0612	0.0825	0.4453	95.23	474.65
MMS-DC	0.0791	0.0917	<b>0.5970</b>	241.96	<b>579.54</b>

This table provides annual sample mean, standard deviation and Sharpe ratio of market timing portfolio returns. The performance fees are annualized basis points and are calculated using the index as the benchmark. The out of sample period is from 1991/01 to 2020/12. A bold number means the optimal value in the corresponding column and panel.

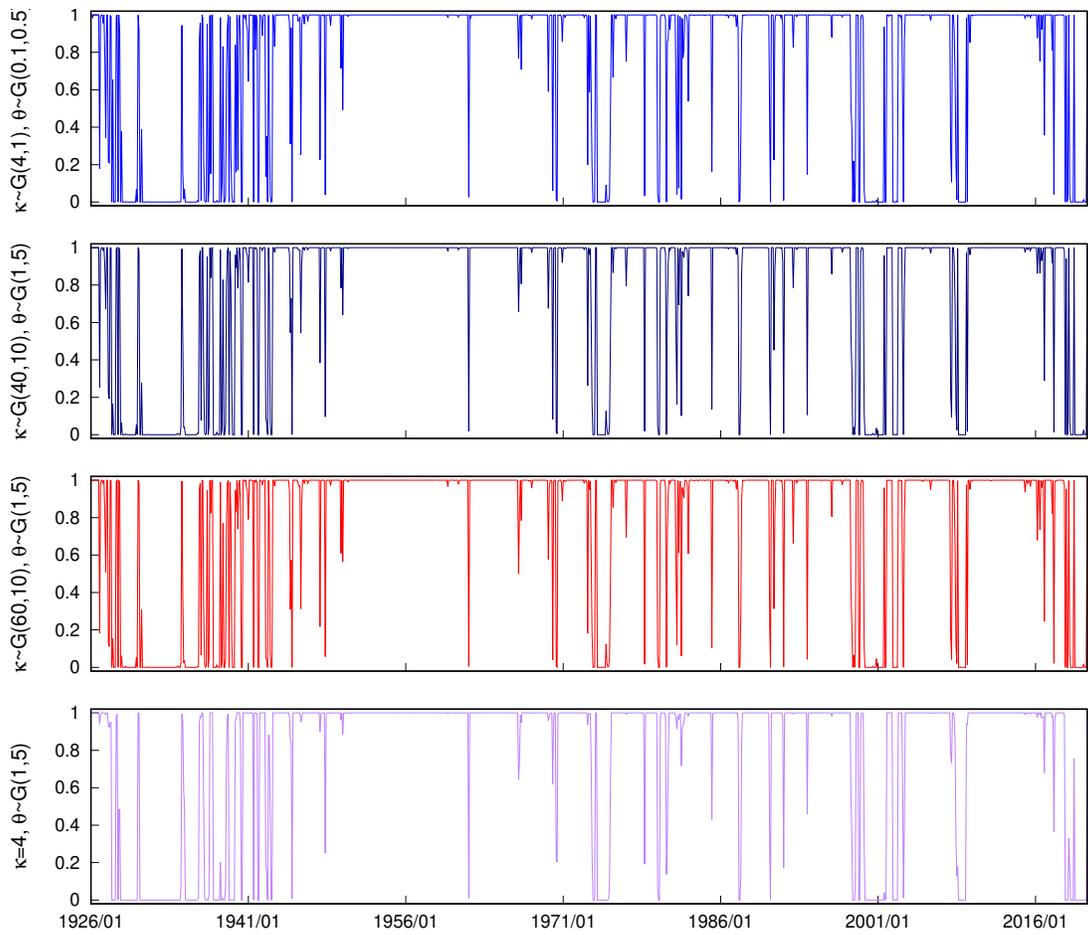


Figure 15: Smoothed probability of bull regime (MMS-CC-duration model under 4 priors)

### A.3 Inference from UMS

See Kim et al. (1999) Chapter 9.

### A.4 Inference from MMS

Inference from MMS is standard. We only sketch the algorithm here.

1. Draw  $M_k$  from a multivariate normal distribution. Keep the values that satisfy the bull and bear identification restriction.
2. Draw  $\Sigma_k$  from an inverse Wishart distribution.
3. Draw  $S$  by following Chib (1996).
4. Draw each row of  $P$  from a Dirichlet distribution.
5. Draw  $m_k$  from a normal distribution.
6. Draw  $v_k^2$  from an inverse Gamma distribution.
7. Draw  $\zeta_k$  from a normal distribution.
8. Draw  $b_k^2$  from an inverse Gamma distribution.

The details for the hierarchical parameter  $m_k, v_k^2, \zeta_k$  and  $b_k^2$  are identical as in Appendix A.6.

### A.5 Comparing gLMC and M-H methods

The theory of gLMC is in Holbrook et al. (2018), so we do not repeat it. This section provides evidence of the practicality of the gLMC method in comparison to the Metropolis-Hastings method, which is probably the easiest and most flexible (not without its deficiency) approach for posterior simulation. The comparison focuses on sampling large positive definite matrices which is the bottleneck of computation in this paper.

In order to illustrate the benefits we consider a very simple multivariate normal model as

$$y_t \sim N(0, \Sigma),$$

for  $t = 1, \dots, T$ . The vectors  $\{y_t\}_{t=1}^T$  have dimension  $N \times 1$  and are independent. There is only one parameter,  $\Sigma$ , which can have a large dimension as  $N(N + 1)/2$ . We assume a conjugate prior as

$$\Sigma \sim IW(A_0, a_0),$$

which returns a textbook analytic posterior distribution as

$$\Sigma \mid Y \sim IW(A_1, a_1),$$

where  $a_1 = a_0 + T$  and  $A_1 = A_0 + Y'Y$  with  $Y = [y_1, y_2, \dots, y_T]'$ . The analytic solution can verify the convergence of the M-H and gLMC method, and then help us understand their efficiency.

### A.5.1 Metropolis-Hastings Method

Denote the posterior kernel density as  $p(\Sigma \mid \cdot)$ , a random walk M-H method based on an inverse Wishart proposal is set up as follows. Given an initial value  $\Sigma^{(0)}$  and degree of freedom parameter  $v > 0$ , repeat the following.

1. Given value  $\Sigma^{(g)}$ , draw a new value  $\Sigma^*$  from an inverse Wishart distribution

$$\Sigma^* \sim IW(\Sigma^{(g)}(v - N - 1), v),$$

which assures that  $E(\Sigma^*) = \Sigma^{(g)}$ .

2. Accept the new value and set  $\Sigma^{(g+1)} = \Sigma^*$  with probability

$$\min \left\{ 1, \frac{p(\Sigma^* \mid \cdot)q(\Sigma^{(g)} \mid \Sigma^*)}{p(\Sigma \mid \cdot)q(\Sigma^* \mid \Sigma^{(g)})} \right\},$$

where  $q(\cdot \mid \cdot)$  is the proposal density in the previous step.

After discarding a burnin sample, we collect the remaining draws for posterior inference. To reduce dependence in the chain we apply thinning by selecting every 100th iteration. The thinning method is also helpful for saving the computer memory when  $N$  is large.

The degree of freedom parameter  $v$  plays an important role in the acceptance frequency. In large dimensions we find it necessary to set a very large value of  $v$  which reduces the variance of the IW draw. This means that the proposal  $\Sigma^*$  can be quite close to the existing value of the Markov chain  $\Sigma^{(g)}$ . This compromises sampling efficiency but still delivers the correct posterior distribution compared to the analytic posterior distribution in a manageable time.

### A.5.2 gLMC

We use Algorithm 2 in Holbrook et al. (2018) to randomly draw  $\Sigma$  from the posterior distribution. The detailed algorithm is in Appendix A.5.3 for interested readers. In this

subsection, we discuss the link between gLMC and the Hamiltonian Monte Carlo (HMC) method along with their benefit and cost.

The HMC relies on Hamiltonian dynamics to simulate values from the posterior distribution and is much faster in exploring the parameter space than a simple random-walk approach. See (Neal et al. 2011) for an excellent survey as well as a practitioner’s guide. The parameters are treated as the position of an object (imagine a surface if the number of parameters is 2) with its potential energy represented by the negative log posterior kernel density. Random shocks introduce fictitious “momentum” or kinetic energy. In other words, this object is repeatedly and randomly “kicked” to suddenly have a speed. In a frictionless world, it will move accordingly and its speed will change (magnitude and direction) during its travel along the curly surface. For example, if the object goes up the hill, its speed tends to decrease and its direction may revert eventually. Or, it may be diverted along a tilted hill. The random shocks (randomness) work together with the surface (negative log posterior kernel) would give a better chance for the object to reach the trough and its neighborhood while allowing its exploration into further area.

For illustration, a simple Hamiltonian function is

$$H(\theta, v) = U(\theta) + K(v), \tag{49}$$

where  $\theta$  is the parameter of interest and  $U(\theta) = -\log \pi(\theta | Y)$  for a Bayesian model (the posterior density  $\pi(\theta | Y)$  can be replaced by any posterior kernel density) is the potential energy. The  $v$  is the velocity and  $K(v) = v' M^{-1} v / 2$  is the kinetic energy with  $M$  being the “mass matrix”. As  $M$  is typically diagonal or even a scalar times identity matrix, denote  $m_i$  as the  $i$ th diagonal element, we can write  $K(v) = \sum_{i=1}^K \frac{v_i^2}{2m_i}$  with  $K$  as the dimension of  $\theta$  and  $v$ . The system moves according to the Hamiltonian’s equations

$$\frac{d\theta_i}{dt} = \frac{v_i}{m_i} \tag{50}$$

$$\frac{dv_i}{dt} = -\frac{\partial U}{\partial \theta_i} \tag{51}$$

The Hamiltonian dynamic system is continuously timed. In practice, it is executed through

a discrete approximation called the leap-frog algorithm with step size  $\varepsilon$ .

$$v_i(t + \varepsilon/2) = v_i(t) - (\varepsilon/2) \frac{\partial U}{\partial \theta_i}(\theta(t)) \quad (52)$$

$$\theta_i(t + \varepsilon) = \theta_i(t) + \varepsilon \frac{v_i(t + \varepsilon/2)}{m_i} \quad (53)$$

$$v_i(t + \varepsilon) = v_i(t + \varepsilon/2) - (\varepsilon/2) \frac{\partial U}{\partial \theta_i}(\theta(t + \varepsilon)) \quad (54)$$

After repeating Step (52)-(54) for  $L$  times, a Metropolis-Hastings adjustment is performed in the last step to safeguard any numeric error from approximation. The HMC is more efficient than a random walk, if ignoring its computational cost, because it explores the shape of the potential energy through the gradient as an direction.

The gLMC inherits HMC by using the Hamiltonian system, while being different in two perspectives. First, instead of using a constant mass function  $M$ , the second order information is exploited in each leap frog cycle ((52)-(54)) (see (Girolami & Calderhead 2011)). As a result, the mass function depends on the value of  $\theta$  (in our simulation,  $\Sigma$ ). Second, the gLMC exploits the geodesic flow to guarantee positive definiteness. Heuristically, for Step (53) in each iteration of the leap frog jump, the parameter space is cast to the subspace of the positive definite matrices so that no movement could violate such restriction. A generic HMC, however, may fail to provide a positive definite proposal without proper restrictions.

The cost from using the gLMC lies in its additional evaluation of a large dimensional gradient and large dimensional matrix operations associated with the varying mass function. Whether the M-H or gMLC provides more efficient samples for a given time period is an empirical issue.

### A.5.3 gLMC Algorithm in Simulation Study

1. The energy function is

$$\begin{aligned} E(\Sigma, V) &= -\log p(\Sigma | \cdot) - \frac{N+1}{2} \log |\Sigma| + \frac{1}{2} \text{vech}(V) G(\Sigma) \text{vech}(V) \\ &\doteq \frac{T}{2} \log |\Sigma| + \frac{1}{2} \text{tr}(Y'Y\Sigma^{-1}) + \frac{a_0 + N+1}{2} \log |\Sigma| + \frac{1}{2} \text{tr}(A_0\Sigma^{-1}) \\ &\quad - \frac{N+1}{2} \log |\Sigma| + \frac{1}{2} \text{vech}(V) G(\Sigma) \text{vech}(V) \\ &= \frac{T + a_0}{2} \log |\Sigma| + \frac{1}{2} \text{tr}([A_0 + Y'Y]\Sigma^{-1}) + \frac{1}{2} \text{vech}(V) G(\Sigma) \text{vech}(V), \end{aligned}$$

where  $G(\Sigma) = D'(\Sigma^{-1} \otimes \Sigma^{-1})D$  and  $G^{-1}(\Sigma) = D^+(\Sigma \otimes \Sigma)D^{+'}$ . Matrix  $D$  satisfies  $\text{vec}(V) = D\text{vech}(V)$  and  $D^+ = (D'D)^{-1}D'$ .

2. Choose step size  $\epsilon$ , number of steps  $L$  and an initial value  $\Sigma$ .
3. Carry out the iteration below.
  - (a) Save  $\Sigma_0 = \Sigma$  and draw

$$\text{vech}(V_0) \sim N(0, G^{-1}(\Sigma)).$$

To draw  $V_0$ , Cholesky decompose  $\Sigma$

$$\Sigma = LL'$$

and draw

$$X \sim N(0, \Sigma \otimes \Sigma)$$

with a Cholesky decomposition  $L \otimes L$ . A simple representation is

$$X = (L \otimes L) \times N(0, I_{N^2})$$

Then,

$$\text{vech}(V_0) = D^+X$$

Symmetrise  $V_0$ .

- (b) Set  $V = V_0$  and  $\Sigma = \Sigma_0$  and start the flow.
  - i.

$$\text{vech}(V) = \text{vech}(V) + \frac{\epsilon}{2}G^{-1}(\Sigma)\text{vech}\left(\frac{\partial}{\partial\Sigma}\left[\log p(\Sigma | \cdot) + \frac{N+1}{2}\log|\Sigma|\right]\right)$$

The derivative

$$\begin{aligned} & \frac{\partial}{\partial\Sigma}\left[\log p(\Sigma | \cdot) + \frac{N+1}{2}\log|\Sigma|\right] \\ &= -\frac{\partial}{\partial\Sigma}\left[\frac{T+a_0}{2}\log|\Sigma| + \frac{1}{2}\text{tr}([A_0 + Y'Y]\Sigma^{-1})\right] \\ &= -\frac{T+a_0}{2}(2\Sigma^{-1} - \text{diag}(\Sigma^{-1})) - \frac{1}{2}(-2\Sigma^{-1}[A_0 + Y'Y]\Sigma^{-1} + \text{diag}(\Sigma^{-1}[A_0 + Y'Y]\Sigma^{-1})) \\ &= -(T+a_0)\left(\Sigma^{-1} - \frac{1}{2}\text{diag}(\Sigma^{-1})\right) + \left(\Sigma^{-1}[A_0 + Y'Y]\Sigma^{-1} - \frac{1}{2}\text{diag}(\Sigma^{-1}[A_0 + Y'Y]\Sigma^{-1})\right) \end{aligned}$$

ii.

$$\Sigma = \Sigma^{1/2} \exp(\epsilon \Sigma^{-1/2} V \Sigma^{-1/2}) \Sigma^{1/2}$$

and

$$V = V \Sigma^{-1/2} \exp(\epsilon \Sigma^{-1/2} V \Sigma^{-1/2}) \Sigma^{1/2}.$$

The exponential is matrix exponential based on the spectral decomposition. Suppose

$$B = U D U',$$

then  $\exp(B) = U \exp(D) U'$  with  $\exp(D) = \text{diag}(e^{d_1}, \dots, e^{d_N})$ . The square root is defined as

$$B^{1/2} = U D^{1/2} U',$$

where  $D^{1/2} = (d_1^{1/2}, \dots, d_N^{1/2})$ .

Symmetrise  $V$ .

iii.

$$\text{vech}(V) = \text{vech}(V) + \frac{\epsilon}{2} G^{-1}(\Sigma) \text{vech} \left( \frac{\partial}{\partial \Sigma} \left[ \log p(\Sigma | \cdot) + \frac{N+1}{2} \log |\Sigma| \right] \right)$$

iv. Repeat for  $l = 1, \dots, L$  times.

(c) Accept new value  $\Sigma$  with probability

$$\min(1, \exp(E(\Sigma_0, V_0) - E(\Sigma, V))).$$

4. Repeat Step 3 many times and discard the initial burnin sample. The remaining  $G$  samples are used for inference. Because the gLMC explores the posterior surface rather efficiently, we do not use thinning and need significantly less simulations compared to the M-H method.

We need to fine-tune the step size  $\epsilon$  and number of leaps  $L$ . We monitor the posterior density with respect to the true distribution, trace plot, and autocorrelation. Acceptance rate is also monitored but not a key as long as it is not too low.

## A.6 Inference from MMS-CC

For completeness, we repeat the model and prior below.

$$\begin{aligned}
R_t \mid s_t = k &\sim N(M_k, \Delta_k \Lambda \Delta_k), \\
\Lambda &= \Gamma Q \Gamma, \quad \Gamma = \text{diag} \left( \frac{1}{\sqrt{Q_{11}}}, \dots, \frac{1}{\sqrt{Q_{NN}}} \right), \\
P(s_{t+1} = j \mid s_t = k) &= p_{kj}, \\
M_k &= (\mu_{1k}, \mu_{2k}, \dots, \mu_{Nk})', \\
\mu_{ik} &\sim N(m_k, v_k^2), \\
m_k &\sim N(0, 0.5), \\
v_k^2 &\sim IG(5, 5), \\
\Delta_k &= \text{diag}(\sigma_{1k}, \sigma_{2k}, \dots, \sigma_{Nk}), \\
\log(\sigma_{ik}) &\sim N(\zeta_k, b_k^2), \\
\zeta_k &\sim N(0, 0.5), \\
b_k^2 &\sim IG(5, 5), \\
Q &\sim IW(\Psi, \nu), \\
(p_{11}, p_{12}) \text{ and } (p_{22}, p_{21}) &\sim \text{Dir}(p_a, p_b),
\end{aligned}$$

where  $k, j = 1, 2$  and  $i = 1, \dots, N$ .

The parameter space  $\Theta$  includes  $M_k$ ,  $\Delta_k$ , matrix  $Q$ , the transition matrix  $P$  and the hyper-parameters  $m_k, v_k^2, \zeta_k, b_k^2$  for  $k = 1, 2$ . It also includes the regime indicator  $S = \{s_1, s_2, \dots, s_T\}$ . Let  $R_{1:T}$  denotes the data over the sample period. A posterior kernel density function is

$$\begin{aligned}
p(\Theta \mid R_{1:T}) &\propto p(P)p(Q) \prod_{k=1}^2 p(m_k)p(v_k^2)p(M_k \mid m_k, v_k^2)p(\zeta_k)p(b_k^2)p(\Delta_k \mid \zeta_k, b_k^2) \\
&\quad \times p(S \mid P)p(R_{1:T} \mid S, M, \Delta, Q),
\end{aligned}$$

where  $M = (M_1, M_2)$  and  $\Delta = (\Delta_1, \Delta_2)$ .

The detailed sampling steps are shown as follows.

1.  $M_k \mid R_{1:T}, \Delta_k, Q, S, m_k, v_k^2$  for  $k = 1, 2$ .

$M_k = (\mu_{1k}, \mu_{2k}, \dots, \mu_{Nk})'$  is sampled by using a Gibbs sampler. Given prior  $\mu_{ik} \sim$

$N(m_k, v_k^2)$ , the conditional posterior of  $M_k$  is

$$\begin{aligned}
p(M_k | \dots) &\propto \prod_{s_t=k} \exp \left\{ -\frac{1}{2} (R_t - M_k)' \Delta_k^{-1} \Lambda^{-1} \Delta_k^{-1} (R_t - M_k) \right\} \\
&\quad \cdot \exp \left\{ -\frac{1}{2} (M_k - m_{k\iota})' v_k^{-2} (M_k - m_{k\iota}) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ n_k M_k' \Sigma_k^{-1} M_k - 2 \sum_{t=1, s_t=k}^T R_t' \Sigma_k^{-1} M_k \right] - \frac{1}{2} \left[ v_k^{-2} M_k' M_k - 2 \frac{m_k}{v_k^2} \iota' M_k \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ M_k' (n_k \Sigma_k^{-1} + v_k^{-2} I) M_k - 2 \left( \sum_{t=1, s_t=k}^T R_t' \Sigma_k^{-1} + \frac{m_k}{v_k^2} \iota' \right) M_k \right] \right\}
\end{aligned}$$

Therefore,

$$M_k | \cdot \sim N(\bar{m}, \bar{H}),$$

where  $\bar{H} = n_k \Sigma_k^{-1} + v_k^{-2} I$  and  $\bar{m} = \bar{H}^{-1} \left( \frac{m_k}{v_k^2} \iota + \Sigma_k^{-1} \sum_{t=1, s_t=k}^T R_t \right)$ . And  $n_k$  is the number of periods that  $s_t = k$ .

2.  $\Delta_k | R_{1:T}, M_k, Q, S, \zeta_k, b_k^2$  for  $k = 1, 2$ .

Values  $\log \sigma_{ik}, i = 1, \dots, N$ , are sampled by using a Metropolis-Hasting algorithm with random walk proposal. A single-move sampler is used first and a joint sampling approach is then applied later to improve efficiency. The initial results obtained from the single-move sampler are used to construct the covariance matrix for the joint random walk proposal.

The conditional posterior of  $\log \sigma_{ik}$  is

$$p(\log \sigma_{ik} | \dots) \propto \prod_{t=1, s_t=k}^T |\Sigma_k|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (R_t - M_k)' \Sigma_k^{-1} (R_t - M_k) \right\} \exp \left\{ -\frac{(\log \sigma_{ik} - \zeta_k)^2}{2b_k^2} \right\}$$

Its log version is

$$\begin{aligned}
\log p(\log \sigma_{ik} | \dots) &\doteq -\frac{n_k}{2} \log |\Sigma_k| - \frac{1}{2} \sum_{t=1, s_t=j}^T (R_t - M_k)' \Sigma_k^{-1} (R_t - M_k) - \frac{(\log \sigma_{ik} - \zeta_k)^2}{2b_k^2} \\
&\doteq -\frac{n_k}{2} \log |\Delta_k \Lambda \Delta_k| - \frac{1}{2} \sum_{t=1, s_t=j}^T (R_t - M_k)' \Sigma_k^{-1} (R_t - M_k) - \frac{(\log \sigma_{ik} - \zeta_k)^2}{2b_k^2} \\
&\doteq -\frac{n_k}{2} \log |\Delta_k \Lambda \Delta_k| - \frac{1}{2} \sum_{t=1, s_t=j}^T (R_t - M_k)' \Delta_k^{-1} \Lambda^{-1} \Delta_k^{-1} (R_t - M_k) - \frac{(\log \sigma_{ik} - \zeta_k)^2}{2b_k^2} \\
&\doteq -n_k \log |\Delta_k| - \frac{1}{2} \sum_{t=1, s_t=j}^T (R_t - M_k)' \Delta_k^{-1} \Lambda^{-1} \Delta_k^{-1} (R_t - M_k) - \frac{(\log \sigma_{ik} - \zeta_k)^2}{2b_k^2}
\end{aligned}$$

Random walk proposal is used to simulate  $\log \dot{\sigma}_{ik}$  and accept it with probability  $\min\left(1, \frac{p(\log \dot{\sigma}_{ik} | \cdot)}{p(\log \sigma_{ik} | \cdot)}\right)$

### 3. $Q | R_{1:T}, M, \Delta, \nu$

The geodesic Lagrangian Monte Carlo introduced in Holbrook et al. (2018) is applied to sample  $Q$  in order to more efficiently explore the conditional posterior and ensure the positive definiteness.

The prior of  $Q \sim IW(\Psi, \nu)$ , the posterior of  $Q$  is

$$\begin{aligned}
p(Q | \dots) &\propto \prod_{t=1}^T |\Sigma_t|^{-1/2} \exp\left\{-\frac{1}{2}(R_t - M_{s_t})' \Sigma_t^{-1} (R_t - M_{s_t})\right\} |Q|^{-\frac{\nu+N+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\Psi Q^{-1})\right\} \\
&\propto \prod_{t=1}^T |\Delta_{s_t} \Gamma Q \Gamma \Delta_{s_t}|^{-1/2} \exp\left\{-\frac{1}{2}(R_t - M_{s_t})' \Delta_{s_t}^{-1} \Gamma^{-1} Q^{-1} \Gamma^{-1} \Delta_{s_t}^{-1} (R_t - M_{s_t})\right\} \\
&\quad |Q|^{-\frac{\nu+N+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\Psi Q^{-1})\right\}
\end{aligned}$$

The log density is

$$\begin{aligned}
\log p(Q | \cdot) &\doteq -\frac{1}{2} \sum_{t=1}^T \log |\Delta_{s_t} \Gamma Q \Gamma \Delta_{s_t}| - \frac{1}{2} \sum_{t=1}^T (R_t - M_{s_t})' \Delta_{s_t}^{-1} \Gamma^{-1} Q^{-1} \Gamma^{-1} \Delta_{s_t}^{-1} (R_t - M_{s_t}) \\
&\quad - \frac{\nu + N + 1}{2} \log |Q| - \frac{1}{2} \text{tr}(\Psi Q^{-1}) \\
&\doteq -\frac{T + \nu + N + 1}{2} \log |Q| - T \log |\Gamma| - \frac{1}{2} \text{tr}(\Psi Q^{-1}) \\
&\quad - \frac{1}{2} \sum_{t=1}^T (R_t - M_{s_t})' \Delta_{s_t}^{-1} \Gamma^{-1} Q^{-1} \Gamma^{-1} \Delta_{s_t}^{-1} (R_t - M_{s_t})
\end{aligned}$$

The Hamiltonian is

$$\begin{aligned}
E(Q, V) &= -\log(p(Q|\cdot)) - \frac{N+1}{2} \log |Q| + \frac{1}{2} \text{vech}(V)^T G(Q) \text{vech}(V) \\
&= \frac{T+\nu}{2} \log |Q| + T \log |\Gamma| + \frac{1}{2} \sum_{t=1}^T (R_t - M_{st})' \Delta_{st}^{-1} \Gamma^{-1} Q^{-1} \Gamma^{-1} \Delta_{st}^{-1} (R_t - M_{st}) \\
&\quad + \frac{1}{2} \text{tr}(\Psi Q^{-1}) + \frac{1}{2} \text{vech}(V)^T G(Q) \text{vech}(V),
\end{aligned}$$

where  $\text{vech}(V)$  is a  $N(N+1)/2 \times 1$  vector to represent the momentum of the free parameters of  $Q$ . Matrix  $G(Q) = D'(Q^{-1} \otimes Q^{-1})D$  and  $G^{-1}(Q) = D^+(Q \otimes Q)D^+$ , where  $D$  is the matrix satisfying  $\text{vec}(V) = D\text{vech}(V)$  and  $D^+ = (D'D)^{-1}D'$ . Matrix  $G(Q)$  is the Fisher information matrix for the Riemann Manifold Hamiltonian Monte Carlo (RMHMC) in Girolami & Calderhead (2011). Holbrook et al. (2018) adapted it to the covariance matrix sampling to safeguard positive definiteness.

The derivatives for each part are given as follows.

(a)

$$\frac{\partial}{\partial Q} \left( \frac{T+\nu}{2} \log |Q| \right) = \frac{T+\nu}{2} (2Q^{-1} - \text{diag}(Q^{-1}))$$

(b)

$$\begin{aligned}
\frac{\partial}{\partial Q} (T \log |\Gamma|) &= \frac{\partial}{\partial Q} \left( T \log(Q_{11}^{-1/2} \cdot Q_{22}^{-1/2} \cdots Q_{NN}^{-1/2}) \right) \\
&= -\frac{T}{2} \frac{\partial}{\partial Q} (\log Q_{11} + \log Q_{22} + \cdots + \log Q_{NN}) \\
&= -\frac{T}{2} \text{diag} \left( \frac{1}{Q_{11}}, \frac{1}{Q_{22}}, \dots, \frac{1}{Q_{NN}} \right)
\end{aligned}$$

(c)

$$\frac{\partial}{\partial Q} \left( \frac{1}{2} \text{tr}(\Psi Q^{-1}) \right) = -Q^{-1} \Psi Q^{-1} + \frac{1}{2} \text{diag}(Q^{-1} \Psi Q^{-1})$$

(d)

$$\begin{aligned}
\frac{\partial}{\partial Q} \left( \frac{1}{2} \sum_{t=1}^T (R_t - M_{st})' \Delta_{st}^{-1} \Gamma^{-1} Q^{-1} \Gamma^{-1} \Delta_{st}^{-1} (R_t - M_{st}) \right) \\
= \frac{1}{2} \sum_{t=1}^T \frac{\partial}{\partial Q} \left( R_t^* \Gamma^{-1} Q^{-1} \Gamma^{-1} R_t^* \right),
\end{aligned}$$

where  $R_t^* = \Delta_{st}^{-1}(R_t - M_{st})$ . Recall that  $\Gamma = \text{diag}\left(\frac{1}{\sqrt{Q_{11}}}, \dots, \frac{1}{\sqrt{Q_{NN}}}\right)$  is a function of  $Q$ . Proceed,

$$\begin{aligned}
& \frac{1}{2} \sum_{t=1}^T \frac{\partial}{\partial Q} \left( R_t^* \Gamma^{-1} Q^{-1} \Gamma^{-1} R_t^* \right) \\
&= \frac{1}{2} \sum_{t=1}^T \left( \frac{\partial R_t^* \Gamma^{-1}}{\partial Q} Q^{-1} \Gamma^{-1} R_t^* \right) + \frac{1}{2} \sum_{t=1}^T \left( R_t^* \Gamma^{-1} Q^{-1} \frac{\partial \Gamma^{-1} R_t^*}{\partial Q} \right) \\
&\quad + \frac{1}{2} \sum_{t=1}^T \left( R_t^* \Gamma^{-1} \frac{\partial Q^{-1}}{\partial Q} \Gamma^{-1} R_t^* \right) \\
&= \frac{1}{2} \sum_{t=1}^T \text{diag} \left( \frac{1}{2} Q_{ii}^{-1/2} R_{ti}^* Q_i^{-1} \Gamma^{-1} R_t^* \right) + \frac{1}{2} \sum_{t=1}^T \text{diag} \left( R_t^* \Gamma^{-1} Q_i^{-1} \frac{1}{2} Q_{ii}^{-1/2} R_{ti}^* \right) \\
&\quad + \frac{1}{2} \sum_{t=1}^T \left( R_t^* \Gamma^{-1} \frac{\partial Q^{-1}}{\partial Q} \Gamma^{-1} R_t^* \right) \\
&= \frac{1}{2} \sum_{t=1}^T \text{diag} \left( Q_{ii}^{-1/2} R_{ti}^* Q_i^{-1} \Gamma^{-1} R_t^* \right) + \frac{1}{2} \sum_{t=1}^T \left( R_t^* \Gamma^{-1} \frac{\partial Q^{-1}}{\partial Q} \Gamma^{-1} R_t^* \right) \\
&= \frac{1}{2} \sum_{t=1}^T \text{diag} \left( Q_{ii}^{-1/2} R_{ti}^* Q_i^{-1} \Gamma^{-1} R_t^* \right) + \\
&\quad \frac{1}{2} \sum_{t=1}^T \left( -2Q^{-1} \Gamma^{-1} R_t^* R_t^* \Gamma^{-1} Q^{-1} + \text{diag} \left( Q^{-1} \Gamma^{-1} R_t^* R_t^* \Gamma^{-1} Q^{-1} \right) \right) \\
&= \frac{1}{2} \text{diag} \left( Q_{ii}^{-1/2} (Q^{-1} \Gamma^{-1})_i (R^* R^*)_i \right) \\
&\quad - Q^{-1} \Gamma^{-1} (R^* R^*) \Gamma^{-1} Q^{-1} + \frac{1}{2} \text{diag} \left( Q^{-1} \Gamma^{-1} (R^* R^*) \Gamma^{-1} Q^{-1} \right),
\end{aligned}$$

where  $R^* = (R_1^*, R_2^*, \dots, R_T^*)'$ .

The Geodesic Lagrangian Monte Carlo is carried out as follows.

(a)

$$\text{vech}(V_0) \sim N(0, G^{-1}(Q)),$$

where  $Q$  is the current value  $Q_0$ . Matrix  $G(Q) = D'(Q^{-1} \otimes Q^{-1})D$  and  $G^{-1}(Q) = D^+(Q \otimes Q)D^+$ . Matrix  $D$  satisfies  $\text{vec}(V) = D\text{vech}(V)$  and  $D^+ = (D'D)^{-1}D'$ .

(b) Set  $V = V_0$  and  $Q = Q_0$ . Send  $V$  and  $Q$  to the flow through the following steps. For  $l = 1, \dots, L$ , where  $L$  is the leapfrog length,

i.

$$vech(V) = vech(V) + \frac{\epsilon}{2} G^{-1}(Q) vech \left( \frac{\partial}{\partial Q} \left[ \log p(Q | \cdot) + \frac{N+1}{2} \log |Q| \right] \right)$$

ii.

$$Q = Q^{1/2} \exp(\epsilon Q^{-1/2} V Q^{-1/2}) Q^{1/2}$$

and

$$V = V Q^{-1/2} \exp(\epsilon Q^{-1/2} V Q^{-1/2}) Q^{1/2}$$

iii.

$$vech(V) = vech(V) + \frac{\epsilon}{2} G^{-1}(Q) vech \left( \frac{\partial}{\partial Q} \left[ \log p(Q | \cdot) + \frac{N+1}{2} \log |Q| \right] \right)$$

And  $\epsilon$  is the leapfrog stepsize.

(c) Accept  $Q$  with probability  $\min(1, \exp(E(Q_0, V_0) - E(Q, V)))$ .

4.  $S | M, \Delta, Q, P$

The state variables  $S = \{s_1, \dots, s_T\}$  are randomly drawn by using the forward filtering backward sampling method in Chib (1996).

5.  $P | S$

Parametrise the prior of  $P_j$ , the  $j$ th column of the transition matrix  $P$ , as  $Dir(\alpha_{j1}, \alpha_{j2})$  for  $j = 1, 2$ . Its conditional posterior is given by

$$p(P_j | \cdot) \sim Dir(\alpha_{j1} + n_{j1}, \alpha_{j2} + n_{j2}),$$

where vector  $(n_{j1}, n_{j2})$  records the numbers of switches from state  $j$  to state 1 and 2, respectively.

6.  $m_k | M_k, v_k^2$  for  $k = 1, 2$ .

Parametrise the prior  $p(m_k) \sim N(m_0, v_0^2)$ . Its conditional posterior is

$$\begin{aligned} p(m_k | \cdot) &\propto p(m_k) p(M_k | m_k, v_k^2) \\ &\propto \exp \left\{ -\frac{(m_k - m_0)^2}{2v_0^2} \right\} \exp \left\{ -\frac{\sum_{i=1}^N (\mu_{ik} - m_k)^2}{2v_k^2} \right\}. \end{aligned}$$

Therefore,

$$m_k | \cdot \sim N(\bar{m}, \bar{v}^2),$$

where  $\bar{v}^2 = (v_0^{-2} + Nv_k^{-2})^{-1}$  and  $\bar{m} = \bar{v}^2 \left( v_0^{-2}m_0 + v_k^{-2} \sum_{i=1}^N \mu_{ik} \right)$

7.  $v_k^2 \mid M_k, m_k$  for  $k = 1, 2$ .

Parametrise the prior  $p(v_k^2) \sim IG(v_0/2, s_0/2)$ . Its conditional posterior is

$$p(v_k^2 \mid \cdot) \propto p(v_k^2)p(M_k \mid m_k, v_k^2)$$

Standard derivation shows

$$v_k^2 \mid \cdot \sim IG(\bar{v}/2, \bar{s}/2),$$

where  $\bar{v} = v_0 + N$  and  $\bar{s} = s_0 + \sum_{i=1}^N (\mu_{ik} - m_k)^2$ .

8.  $\zeta_k \mid \Delta_k, b_k^2$  for  $k = 1, 2$ .

Parametrise the prior  $\zeta_k \sim N(m_0, v_0^2)$ . Its conditional posterior is

$$\begin{aligned} p(\zeta_k \mid \cdot) &\propto p(\zeta_k)p(\Delta_k \mid \zeta_k, b_k^2) \\ &\propto \exp \left\{ -\frac{(\zeta_k - m_0)^2}{2v_0^2} \right\} \exp \left\{ -\frac{\sum_{i=1}^N (\log(\sigma_{ik}) - \zeta_k)^2}{2b_k^2} \right\} \end{aligned}$$

Standard deviation shows

$$\zeta_k \mid \cdot \sim N(\bar{m}, \bar{v}^2),$$

where  $\bar{v}^2 = (v_0^{-2} + Nb_k^{-2})^{-1}$  and  $\bar{m} = \bar{v}^2 \left( v_0^{-2}m_0 + b_k^{-2} \sum_{i=1}^N \log(\sigma_{ik}) \right)$ .

9.  $b_k^2 \mid \Delta_k, \zeta_k$  for  $k = 1, 2$ .

Parametrise the prior  $p(b_k^2) \sim IG(v_0/2, s_0/2)$ . Its conditional posterior is

$$p(b_k^2 \mid \cdot) \propto p(b_k^2)p(\Delta_k \mid \zeta_k, b_k^2)$$

Standard deviation shows

$$b_k^2 \mid \cdot \sim IG(\bar{v}/2, \bar{s}/2),$$

where  $\bar{v} = v_0 + N$  and  $\bar{s} = s_0 + \sum_{i=1}^N (\log(\sigma_{ik}) - \zeta_k)^2$ .

## A.7 Inference from MMS-IC

The algorithm is very similar to that of MMS-CC in Appendix A.6. Some differences are:

1.  $Q_k$  for  $k = 1, 2$ .

It is similar to drawing  $Q$  in Appendix A.6. The difference is that drawing  $Q_k$  only uses part of the data for  $s_t = k$ .

2. An additional array  $W = (w_1, \dots, w_T)$  as the state variables for the Markov process of correlations. Same method is applied as in Chib (1996).

## A.8 Inference from MMS-DC

The algorithm is very similar to that of MMS-CC in Appendix A.6. Some differences are:

1. Drawing  $M_k$  and  $\Delta_k$  are similar as in Appendix A.6. A small difference is that drawing  $M_1$  and  $\Delta_1$  uses the data for  $s_t = 1$  or 2, and  $M_2$  and  $\Delta_2$  for  $s_t = 3$  or 4.
2.  $Q_k$  for  $k = 1, 2$ . It is similar to drawing  $Q$  in Appendix A.6. The difference is that drawing  $Q_1$  only uses part of the data for  $s_t = 1$  or 3, while  $Q_2$  for  $s_t = 2$  or 4.
3. State vector  $S$  is drawn using the same method as in Chib (1996), but with 4 states instead of 2.
4. Each row of the transition matrix has 4 elements and drawn from a Dirichlet distribution. It is similar as in Appendix A.6.