Latent state recognition by an enhanced hidden Markov model

Yuan Yao[*]   Yi Cao[†]   Jia Zhai[‡]   Junxiu Liu[§]   Mengyuan Xiang[¶]   Lu Wang[‖]

March, 2020

Abstract

In this paper, we start from relaxing assumptions of traditional hidden Markov model then develop a novel framework for decoding the latent states, from which the dynamics of multi-variable financial data is generated. To construct the framework, we model the observed variables as a p-order vector autoregressive process, allow the latent state to evolve through a semi-Markov chain, and shrink the auto-regression and covariance matrices via a penalized maximization likelihood method. Using the 50-dimensional simulated data, the 12-dimensional 5-minute order book data of the Chinese CSI 300 index component stocks, the 49-dimensional daily data of U.S. industry portfolio, and 1-dimensional hourly data of four primary foreign exchange rates, our empirical analyses show that the proposed model outperforms the alternative model in accurately recognizing anomalous events and achieves better sharp ratio in a pseudo trading strategy via the latent states. The superior performance is across the data frequency of minute, hour and daily, the dimension of one, 12, and 50, the data type of stock, foreign exchange rate, and industry portfolio.

JEL code: C63, F47

Keywords: LASSO; Vector-autoregressive Model; Hidden Markov Model.

---

[*]Institute of Management Science and Engineering, Business School, Henan University, Jinming District, Kaifeng, Henan 475004, China. Email: prof.yuanyao@gmail.com.

[†]Corresponding author: Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS, UK. Email: jason.caoyi@gmail.com.

[‡]Xian Jiaotong-Liverpool University, 111 Ren'ai Road Suzhou Industrial Park Suzhou Jiangsu Province P. R. China 215123. Email: jia.zhai1982@gmail.com.

[§]School of Computing, Engineering and Intelligent Systems, Ulster University, Magee campus BT48 7JL, UK. Email: j.liu1@ulster.ac.uk.

[¶]Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS, UK. Email: mengyuan.xiang@ed.ac.uk.

[‖]Director of Asian Pacific, American Higher Education Alliance; CEO, Linkup Education Technology Co. Ltd., 40 Bell Chime, Irvine, CA, 92618. Email: lwang@linkupedu.com.

1 Introduction

The boom and crash of financial markets have intense inference on the economical and social aspects of our daily life. Such changes are usually associated with events such as great inflation, dot-com bubble, financial crisis, and even the market manipulation Hamilton (2010); Tsang and Chen (2018). However, those events are usually not directly observable from the market variables. One may attempt to identify those events by certain changes of the statistical properties of selected market variables. This is usually termed as 'regime change detection', or 'latent state detection' Piger (2009). Regime change is significant to investors as well as the regulators and policy makers. Incorporating regime change forecasting and latent state identification is helpful for investors avoid losses and policy makers in releasing required policy changes Kritzman et al. (2012). Being able to recognising the latent states also help regulators in monitoring the market for maintaining the stability.

1.1 Econometrics background

Most of the studies on regime change detection or latent state recognition are based on the time series econometric models. This study follows this trend and extends existing work to construct a more generic framework of multivariable time series. It is well known that the return of the stock price is close to but not fully following the white noise process. A large number of studies prove that two strong characteristics are associated with the price return series: the correlation and clustering. Akgiray (1989) shows the evidence of a significant level of statistical dependence among the stock returns. Andersen et al. (2001) find that the realized volatility and correlation show strong temporal dependences and can be well modelled by long-memory processes. Brown and Warner (1985) also recognize the auto-correlation in daily returns and further indicate the change of the covariance due to the market event. Consequently, modelling the stock price return by an autoregressive process has been confirmed by theoretical and empirical studies such as the work of Engle (1982); Nelson (1991) and the vector autoregressive model for the multi-variable financial return in recent studies as Cubadda et al. (2017); Kalli and Griffin (2018); Billio et al. (2019).

However, the autoregressive models fail to capture the clustering of the price return, which is usually called "volatility clustering" as the study of Granger and Machina (2006). This failure is mainly due to that the intrinsic state generating the observed variables often switches among two or more mechanisms, which are essentially determined by the market conditions. For example, the

trading behaviours during an anomalous market condition are expected to follow a different autoregressive process from that of a peaceful market. The portfolio performance in a stable economy shall follow a separated autoregressive process from the one in a recession period. Despite the fact the intrinsic state being not observable on the market, it can be estimated by statistical modelling, i.e., the Markov-switching model as the work of Cai (1994); Rydén et al. (1998); Kim et al. (1999); Maheu and McCurdy (2000); Diebold and Inoue (2001); Kim et al. (2008). An extended Markov-switching model with data clusters is proposed and applied in identifying energy price states Dias and Ramos (2014). Zhou and Mamon (2012) examine and show that the performance of interest rate model with Markov regime switching is better than the models without the regime characteristics. Lin et al. (2017) further show the consideration of Markovian regime switching enhances the performances of industry time series prediction. Markov-switching model holds two basic assumptions: 1) the existence of latent states that generate the observed variables. The latent state evolves following a Markovian process: the current state at time t is only dependent on the most recent state at time t-1; 2) The observed variables are conditional independent random variables generated according to different latent states.

The assumptions are unreasonably strong in real world, hence many studies extend the Markov-switching model by relaxing either of the two restrictive assumptions. A Gaussian hidden Markov model (G-HMM) is proposed by Bilmes et al. (1998) with the mixture Gaussian distribution that describes the observed sequence so that the volatility clustering is incorporated in the model. An autoregressive hidden Markov model (AR-HMM) equips the traditional HMM with stochastic dependence between observations in the study of Ephraim et al. (1989). The AR-HMM is widely applied in single dimension signal processing areas, i.e., speech recognition Hu and Wang (2004); Gannot et al. (1998) and clinical data Stanculescu et al. (2013); Dang et al. (2017). A vector autoregressive hidden Markov model (VAR-HMM) is proposed in Hamilton (1989, 1990) where the observed multi-variables are described as a vector autoregressive process with selected orders. The VAR-HMM is proved to be valuable in identifying the regime switching, which, if ignored, provides a significant impact on the cost of portfolio allocation when a risk-free asset is available to hold as the studies of Ang and Bekaert (2002); Luo et al. (2015).

The three extensions of HMM, G-HMM, AR-HMM, and VAR-HMM contribute to an identical strand: explicitly modelling the correlations in observed sequence such that the assumption 2) can be released. Another strand is to drop the Markovian constraint of assumption 1) on the latent state to demonstrate a sensible and practical idea of modelling the data in real world. In many

applications, the latent state more likely sticks in one or two conditions than switches to some others. To comply with this, Bulla and Bulla (2006) propose a hidden semi-Markov model (HsMM) to consider the temporal high order dependence through explicitly modelling the probability of staying at a latent state for n time units instead of a presumed geometric distribution of state transition in HMM. The HsMM model has been successfully applied in more than thirty scientific areas and has been proved outperforming in change-point recognition, regime-switching identification and anomalous state detection as the review study in Yu (2010). There is an additional strand of the HMM extension for achieving a more stable estimates of the covariance matrices as the work of Fiecas et al. (2017). The HMM suffers from the instability of estimated covariance matrices, which is even not assured to be invertible when the data to be modelled is high-dimensional with respect to the length of the data available (see also the work of Johnson et al. (2002)). A shrinkage estimation of HMM (abbreviated as sh-HMM) via a penalized maximization likelihood method for providing an invertible, positive-definite, and smaller covariance matrices is initialized in Ledoit and Wolf (2004), further discussed in Sancetta (2008), and completely proposed in Fiecas et al. (2017). The sh-HMM has been proved to be particularly useful in analyzing multi-variable financial data, i.e., adaptive portfolio selection as the work of Nystrup et al. (2018).

1.2 Proposed contribution

This study builds upon and embraces all three strands of HMM extensions. By 1) considering the observed variables following a vector autoregressive process of selected orders; 2) allowing the latent state to evolve through a semi-Markov chain; and 3) finally shrinking the autoregressive and covariance matrices via a penalized maximization likelihood method, we provide a general framework for modelling multi-variable financial time series and recognizing the underlying latent state behind the observed variables. Our main contribution is the formulation of a regularized framework with the abandonment of two primary assumptions in traditional HMM. As far as we aware, this is the first work that adopts primary extensions of the HMM together as one single model to overcome the unnecessary constraint and unrealistic assumption, and generate a stable regularized estimation of the auto-correlation and covariance matrices of the practical financial data.

We also contribute to the literature by offering comprehensive empirical evidence with four types of data. Following the Fiecas et al. (2017), we firstly evaluate the model by simulated time series with dimension of 50, sample size of 1000 and two latent states. Our proposed model achieves higher

precision and recall than the alternative model. We further evaluate the model by three practical dataset: the high frequency order book of selected stocks, the U.S. industry portfolio, and selected popular foreign exchange rates. For the order book data, the well-known events of "circuit-breaker" on 4 and 7 Jan 2016 are tested for the latent state recognition evaluation. The empirical studies show that our proposed model accurately identifies the chaos but latent states of the markets at around 13:12 and 13:33 on 4 Jan 2016, and 9:42 and 9:58 on 7 Jan 2016. The accuracy can be observed at minute-level. However, the benchmark model recognizes relatively messy latent states. For the U.S. industry portfolio data from 1 Jul 1926 to 28 Jun 2019, the empirical results show that the proposed model accurately identifies all big events from 1926 to 2019 including the great recession in 1930s, the great inflation in 1970s, 2000's dot-com bubble, and financial crisis in 2007-2008. To have a closer look, we re-run the models focusing on the data from 3 Jan 2007 to 28 Jun 2019. Our proposed model can have a more detailed recognition of the financial crisis from 2008 to 2009, the double-dip recession at 2011, the shocks at 2016 (Chinese stock market crash, OPEC cut, and Brexit), and the uncertainty at 2019. We select four popular foreign exchange rates from Jan 2009 to Aug 2015 with hourly interval as the last empirical study data. Due to high sensitivity of the FX rate to the economic and social events, we carry out a trading strategy based on the latent state following the work of Nystrup et al. (2020) for a clear comparison. Our proposed model achieves the highest sharp-ratio among all benchmark models. Therefore, the accurate recognition of our proposed model is regardless of data dimension, frequency, event type or duration.

This paper is organized as follows. In Section 2, we give the details of the structure of the proposed model, then discuss the regularization method, and finally show the revised EM method for the parameter estimations. In Section 3, we discuss the model performance evaluated by simulated data. In Section 4, we provide a thorough empirical study on 5-second frequency micro-structure order book data of the stocks listed in Shanghai Stock Exchange from 2005 to 2016, and the U.S. 49 industry portfolio return data from 1926 to 2019. The Section 5 gives a brief conclusion and discussion.

2 The proposed model: Lasso-VHsMM

In this study, we propose a vector autoregressive (VAR) hidden semi-Markov model with Lasso regularization to model the dynamics of multi-variable financial data. We call the proposed model as Lasso-VHsMM. The design of Lasso-VHsMM follows the vector autoregressive hidden Markov

model framework in Hamilton (1989); Francq and Zakoıan (2001); Monbet and Ailliot (2017), and extends the Markov process to be a semi-Markov chain by allowing the duration for each state to be a random variable, and further applies the Lasso regularization on both the hidden state covariance matrix following Fiecas et al. (2017) and the vector autoregressive coefficients following Tibshirani (1996).

2.1 Model structure

To develop the Lasso-VHsMM, it is convenient to follow the standard process in empirical finance. First, we are interested in observing multiple financial time series simultaneously as a vector. We represent the multivariable vector at time t as $\mathbf{y}_t \in \mathbb{R}^d$, where d is the vector dimension, as well as the number of financial data we observe, and t is the time t$=1,...,T$.

Second, we assume the variables $\mathbf{y}_t$ are determined by fewer unobserved latent variables. The latent variables are denoted by $S_t \in \{1,...,M\}$, where M is the number of latent variables and is less than d. We usually call the latent variables as latent states or hidden states in the hidden Markov model. The assumption of observed variables being influenced by fewer hidden states is usually made in many financial applications, for example the order book modelling Jiang et al. (2019), trading behaviour detection Cao et al. (2014), and financial regime estimation Chopin and Pelgrin (2004). We allow the hidden states to be a semi-Markov process by explicitly defining the state duration as a random variable following the work in Yu (2010); Van der Hoek and Elliott (2019). According to the semi-Markov process, a state i may stay for n times with a probability $r_i(\text{n})$=P(stay n times at latent state i), where $n \in \mathcal{D}$, $\mathcal{D} = \{1,...,D\}$, and D is a pre-defined maximum duration for a state. Therefore we have the duration density of the latent state as $\boldsymbol{r} = [r_1,...,r_M]$. An important improvement of the hidden semi-Markov model (hsMM) to traditional hidden Markov model (hMM) is that the latter allows only one observation per state while the former allows each state emitting a sequence of observation vectors Yu (2010); Van der Hoek and Elliott (2019). The length of the sequence at state i is determined by the length of time remaining in state i, which is the duration or sojourn time $r_i(\text{n})$. The $r_i(\text{n})$ is defined as a random variable and assumes an integer value. After defining the hidden states S_t, we can define the transition probability. After defining the hidden state S_t, we represent the transition probability from state i at time t to j at time $t+1$ as q_{ij}, for i,j=1,...,M. Considering the state duration, the transition probability from state i with duration n to state j can be represented by $q_{(i,n)(j)}$. Consequently, the prior state i starts at $t-n+1$ and ends at t, with the duration n, and then transits to

state j at the time $t + 1$ according to the transition probability $q_{(i,n)(j)}$. It can be defined as $q_{(i,n)(j)} = P[S_{t+1} = j | S_{t-n+1:t} = i] = q_{ij} r_i(\text{n})$, $i \neq j$. Thus, the self-transition probability, $q_{(i,n)(i)}$ is zero. The state transition probability is subject to $\sum_{j \in S, j \neq i} \sum_{n \in \mathcal{D}} q_{(i,n)(j)} = 1$. Consequently, the state is dependent on both the previous state and its duration. We represent the $\boldsymbol{Q} = \{ q_{(i,n)(j)}, i, j = 1, ..., M \}$ as the hidden state transition matrix. When the state is at i, n observations are emitted with the probability of $b_{i,n}(\mathbf{y}_{t-n+1:t}) = P[\mathbf{y}_{t-n+1:t} | S_{t-n+1:t} = i]$. We assume that the latent states are associated with different prior probabilities, denoted as $\boldsymbol{\delta} = [\delta_{1,n}, ..., \delta_{1,n}]$, where the $\delta_{i,n} = P[S_{t-n+1:t} = i]$. The $\boldsymbol{\delta}$ represents the probability of the initial latent state with its duration before the first observation $\mathbf{y}_1$ is obtained. In most applications, the distribution of the state duration $r_i(n)$ can be modelled by exponential familiy distributions, such as the Gaussian Ariki and Jack (1989), Poisson Russell and Moore (1985), and Gamma distribution Levinson (1986). Following the study of Mitchell and Jamieson (1993), the probability mass function for duration of state i can be expressed as $r_i(n) = \frac{1}{B(\theta_i)} \xi(n) \exp(-\sum_{p=1}^{P} \theta_{i,p} S_p(n))$, where P is the number of natural parameters, $\theta_{i,p}$ is the p-th natural parameter for state i and $\boldsymbol{\Theta_i} = [\theta_{i,1}, ..., \theta_{i,p}]$, $S_p(n)$ and $\xi(n)$ are sufficient statistic, and $B(\theta_i)$ is a normalizing term and can be obtained by $B(\theta_i) = \sum_{n=1}^{D} \xi(n) \exp(-\sum_{p=1}^{P} \theta_{i,p} S_p(n))$. In this study, we follow the method in Freguson (1980) to find the new duration parameters for state i by maximizing $\sum_{n=1}^{D} \hat{r}_i(n) \log r_i(n)$ subject to the constraint $\sum_{n=1}^{D} r_i(n) = 1$.

Third, we assume a p order autoregressive dependence among the observation vector $\mathbf{y}_t$ as

$$\boldsymbol{y_t} = \mu_i + \sum_{k=1}^{p} \boldsymbol{A}_{ki} \boldsymbol{y}_{t-k} + \epsilon_{ti} \tag{1}$$

where ϵ_{ti} follows Gaussian process $\mathcal{N}(0, \boldsymbol{\Sigma}_i)$; $i = 1, ..., M$, and $t = 1, ..., n$; $\mu_i \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_i \in \mathbb{R}^{d \times d}$ are conditional mean and covariance matrix of $\boldsymbol{y}_t$ given the previously observed data $\boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-p}$ and the current state S_i; the $\boldsymbol{A}_{ki}$ is the k-order matrix of autocorrelation on the condition of $S_t = i$. Introducing the Vector auto-regression (VAR) model to HsMM is to capture the linear interdependences among the observed multiple time series. As the VAR model has been successfully applied in numerous financial modelling examples, i.e., market impact estimation Hautsch and Huang (2012); Jiang et al. (2019), the vector auto-regression has been proved as a generic feature in multi-variable time-series modelling. Adding VAR to HsMM makes the samples drawn from the model more reflecting the natural features of multi-variable financial data. We abbreviate the VAR plus HsMM as the VHsMM model.

Consequently, the data generation mechanism of VHsMM model can be represented as following

steps. Step 1, the model generates an initial state $S_1 = i, i \in 1, ..., M$ according to the initial state distribution $\boldsymbol{\delta}_i$; Step 2, a duration time n is generated with the probability $r_i(n)$ drawn from state duration density $\boldsymbol{r}$; Step 3, a sequence of observations $\boldsymbol{y_1}, ..., \boldsymbol{y_n} \in \mathbb{R}^d$ are selected according to the p order autoregressive model in equation 1; Step 4, the next hidden state, $S_{n+1} = j$, is selected corresponding to the probability q_{ij} from the state transition matrix $\boldsymbol{Q}$. As the state i lasts for n time units, we also represent the initial state S_1 as $S_{1:n}$. Then the data generation process iterates until the last observation.

After the construction of the basic structure of VHsMM model, we represent all parameters as a tuple $\boldsymbol{\theta} = \{\boldsymbol{\delta}, \boldsymbol{r}, \boldsymbol{Q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{A}\}$, where $\boldsymbol{\delta}$ is the prior probability of latent state with $M - 1$ free parameters; $\boldsymbol{r}$ is the duration density of the latent state with $M(D - 1)$ free parameters; $\boldsymbol{Q}$ is the latent state transition matrix with $M(M - 2)$ free parameters; $\boldsymbol{\mu}$ is the conditional mean with Md free parameters; $\boldsymbol{\Sigma}$ is the covariance matrix with $\frac{Md(d+1)}{2}$ free parameters; and $\boldsymbol{A}$ is the k order auto-regression matrix with Mpd^2 free parameters.

2.2 Model regularization

We intend to apply our proposed model as a general framework in modelling different financial data, for example the order book data, which may have second or minute-level frequency, the U.S. industry portfolio data, which are usually daily or monthly data. To enhance the feasibility of our model to cope with financial data with large frequency range, we use the LASSO regularization to penalize the observed variables and the latent states. Our motivation lies in the following two aspects.

Model Applicability. If we model the N-level order book as the work in Hautsch and Huang (2012), we use the price and depth of both bid and ask sides at each level with buy or sell indicator to combine a vector of $4N + 2$ dimensions. If we choose $N = 5$ and use five-second order book data [1], the number of parameters ($4N + 2 = 22$) is close to the one of observations ($8 \times 60/5 = 96$) in a trading day, which makes the covariance matrix $\boldsymbol{\Sigma}_i$ not invertible and the VHsMM not be well estimated. If we are interested in 10-level order book, the number of parameters increases to 42. Thus we have to either narrow down the variables or extend the window to cover 4 to 5 trading days for a reliable estimation of the model, as the work of Hautsch and Huang (2012). Conversely, if using high-frequency data with five-second interval (as the work of Jiang et al. (2019)), we may face

---

[1]To alleviate the high-frequency noise on the market, order book data is usually downsampled to five-minute interval to reflect a robust and true market dynamics Gençay et al. (2001); Aldridge (2013)

a problem of a sparse matrix of auto-correlation or covariance, which we consider in the following section. Similarly, U.S. industry portfolio data contains average return in more than 40 industry sections. If using monthly data, we need at least the data over ten years ($12 \times 10 = 120$) for a reliable model estimation. It is acceptable to intuit or group the variables and model them separately as the work in Hautsch and Huang (2012). However, if equipped with a steady statistical method rather than a heuristic search to select the effective and high-quality variables and shrink the negligible ones, the model's reliability and applicability can be reasonably enhanced across many financial scenarios.

Matrix Sparsity. One popular idea is to assume that only a limited number of effective and meaningful variables at any time that are worth investigating Hastie et al. (2005); John Lu (2010); Chinco et al. (2019). For example, if modelling the cross-sectional stock return as Chinco et al. (2019) or the order book changes as Hautsch and Huang (2012), the auto-regression matrix $\boldsymbol{A}$ is sparse as many observations of the stock return and bid/ask price/depth change are close to zero. Hence there is a requirement to remove the unnecessary correlations in $\boldsymbol{A}$ and only keep the most relevant details in the model. To achieve this, people either intuit the candidate variables or rely on the Akaike or Bayesian information criterion (AIC or BIC) of Akaike (1974); Schwarz et al. (1978). In the former case, intuition may lead to arbitrarily selected unexpected candidates. In the latter case, those criteria are useful in selecting the best subsets. However, when the number of variable grows, selecting by AIC or BIC becomes a non-convex NP hard problem Natarajan (1995), and is therefore "highly impractical" as indicated by Candès et al. (2009). For example, selecting from 22 variables requires a search from subsets of 2^{22} combinations.

After defining all components and parameters, we represent the overall structure of the proposed Lasso-VHsMM model in the Figure 2.1.

2.2.1 Implementation

To regularize the estimation for covariance matrix $\boldsymbol{\Sigma}$, we follow the work proposed by Yuan and Huang (2009); Sancetta (2008); Ledoit and Wolf (2004). The regularized covariance matrix, represented as $\boldsymbol{\Sigma}^r$ is generated by a combination of a weighted sample covariance matrix, $\hat{\boldsymbol{\Sigma}}$, estimated by the original maximum likelihood and the target matrix αI_p,

$$\boldsymbol{\Sigma}^r = (1 - W)\hat{\boldsymbol{\Sigma}} + W\alpha I_p \tag{2}$$

Figure 1: Illustration of the proposed Lasso-VHsMM. This figure shows how the working mechanism of Lasso-VHsMM model. The first hidden state S_1 and duration probability r_1 are selected according to the transition probability q. S_1 lasts for $n=3$ time units at the probability r_1 and emits three observations $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$. The hidden state transits to S_2 according to the transition probability and stays for $n=4$ time units with five emitted observations $(\mathbf{y}_4, \mathbf{y}_5, \mathbf{y}_6, \mathbf{y}_7, \mathbf{y}_8)$ according to the emission probability. The hidden states then transit to $S_3, ..., S_n$ until the final observation $\mathbf{y}_n$.
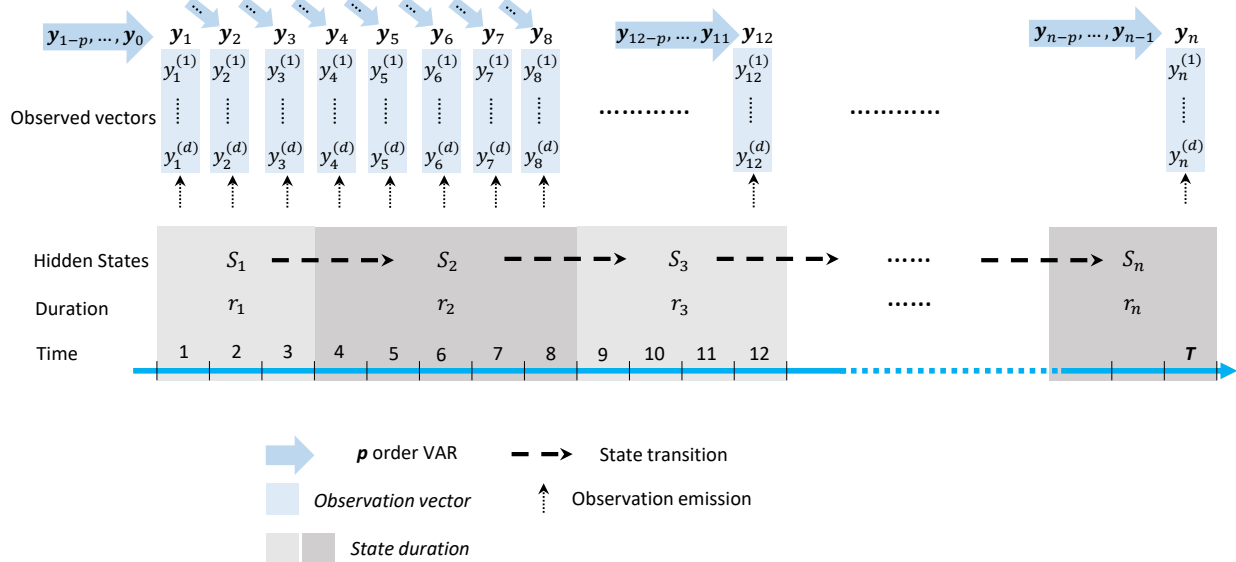
with regard to the constraint of trace of $\mathbf{\Sigma}^r$ equal to the trace of $\hat{\mathbf{\Sigma}}$, where $W = \frac{\lambda}{1+\lambda} \in [0,1]$. This shrinkage estimation is proposed by Ledoit and Wolf (2004) to obtain an optimized parameter by $\operatorname{argmin}_{0 \leq W \leq 1} \mathbb{E}_{\theta_0} ||\Sigma^r - \Sigma_0||^2$. The regularization penalty parameter λ is then selected adaptable to the underlying distribution, particularly $\lambda_0 = \frac{W_0}{1-W_0}$. Following the work of Ledoit and Wolf (2004), the covariance matrix shrinkage is formulated as a penalized maximum likelihood estimation with regularization penalty λ selected iteratively based on the observations to achieve a minimized optimization error for the covariance matrix $\mathbf{\Sigma}$ in terms of the Frobenius norm. Note that, when we have $W = 0$, the case of no shrinkage, we have the $\mathbf{\Sigma}^r = \hat{\mathbf{\Sigma}}$. The scaling factor α is selected to satisfy the constraint of the equal trace: $\operatorname{tr}(\mathbf{\Sigma}^r) = \operatorname{tr}(\hat{\mathbf{\Sigma}})$. Overall, the regularization guarantees the positive definiteness of the matrix $\mathbf{\Sigma}^r$, maintains the size of $\mathbf{\Sigma}^r$ the same as the sample covariance matrix $\hat{\mathbf{\Sigma}}$, and enhances the regularity of the matrix $\mathbf{\Sigma}^r$. For example, when we increase the value of λ, the condition number of the regularized covariance matrix is usually close to one, which shows less dispersion between the largest and the smallest eigenvalues of the matrix.

The penalty for the auto-regression coefficients $\mathbf{A}$ follows the traditional Least Absolute Shrinkage and Selection Operator (LASSO) Tibshirani (1996) to identify the most relevant coefficients. The LASSO coefficients can be estimated by minimizing the objective function with a roughness

9

penalty λ_a placed on the sum of the absolute value of the auto-regression coefficients as the "l_1 penalty". The penalty tuning parameter λ is used to control the strength of the regularization. A small value of λ usually leads to a more strict selection.

$$a^r = \arg\min_a ||\boldsymbol{y}_{p+1:T} - \mu + \sum_p^{k=1} a_k^T \boldsymbol{y}_{p+1-k:T-k}||_2^2 + \lambda_a ||a||_1 \tag{3}$$

where the $|| \cdot ||_1$ is the Manhattan norm, the $|| \cdot ||_2$ is the Euclidean norm; and the penalty $\lambda_a \geq 0$ controls the strength of the regularization. The constraint of the penalty combines the model estimation and the variable selection as a single process. The LASSO regularization select coefficients by zeroing negligible coefficients and shrinking other relevant ones.

2.2.2 Penalty value selection

In this study, we select the regularization penalty λ by minimizing one step ahead mean square forecast error (MSFE) following the work of Bańbura et al. (2010); Nicholson et al. (2017); Baek et al. (2017). We divide all data observations into three time periods: the initialization period [1:T_1], the validation period ($T_1 : T_2$], and the testing period ($T_2 : T_3$], as illustrated in Figure 2. At the first step, we estimate our model by using the data in the initialization period, from 1 up to time $t=T_1$ and forecast the $\hat{\boldsymbol{y}}_{T_1+1}^{\lambda_i}$ for $i=1,...,N$, which denotes all possible penalty values. At the second step, we add one more observation at the time $t=T_1+1$ to estimate the model and forecast the $\hat{\boldsymbol{y}}_{T_1+2}^{\lambda_i}$ for $i=1,...,N$. We then repeat this process until the time $t=T_2-1$. The penalty value can be selected by minimizing the MSFE with respective to the λ:

$$MSFE(\lambda_i) = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2-1} ||\hat{\boldsymbol{y}}_{t+1}^{\lambda_i} - \boldsymbol{y}_{t+1}||_F^2 \tag{4}$$

where the $|| \cdot ||_F$ is the Frobenius norm defined as $||\boldsymbol{A}||_F = \sqrt{tr(\boldsymbol{A}^T \boldsymbol{A})}$. We follow the work in Friedman et al. (2010); Nicholson et al. (2017) to select the penalty parameter λ and λ_a from a grid of potential values, which start from the smallest value for which every coefficient is forced to zero, then increase in log-linear steps. We follow the "rule of thumb" in Nicholson et al. (2017) that a grid depth of $25\lambda_{min}$ and 10 grid points to search an appropriate performance.

Figure 2: Illustration of the dataset partition for penalty parameter validation. This figure shows the mechanism for penalty parameter λ estimation.
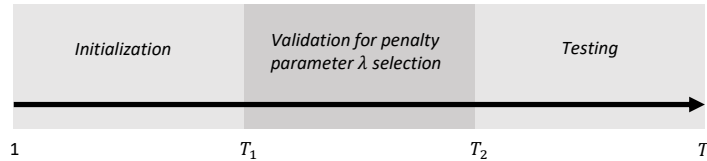
2.3 Model parameter estimation

To estimate the parameters in the proposed Lasso-VHsMM model, we follow the general methods of penalized likelihood estimation by the work of Green (1990); De Pierro (1995) under Expectation-Maximization (EM) framework in traditional hidden Markov model in Baum et al. (1970) and Dempster et al. (1977).

For the Expectation step (E-step), we follow the work of Yu (2010) to estimate the standard forward-backward variables. We define the multivariable autoregressive density as $f_{j,n}(\boldsymbol{y}_{t+1:t+n}) = P(\boldsymbol{y}_{t+1:t+n}|S_{t+1:t+n} = j)$, under the condition of the hidden state j continuing for n time units duration. Based on this, we can define the forward variable as $\alpha_t(j,n) = P(S_{t-n+1:t} = j, \boldsymbol{y}_{1:t}|\boldsymbol{\theta})$, where the j=1,...,M, t=1,...,T, and n=1,...,$\min(D,t)$ (D is the maximum duration for a state defined in Section 2.1). The forward variable $\alpha_t(j,n)$ denotes the probability density generated by the model parameters $\boldsymbol{\theta}$ under the condition of the hidden state staying at j for n time units until t with the observations from time 1 to t. With the initialization of the forward variable $\alpha_0(j,n) = \delta_j$, j=1,...,M, we can further define the recursion formula as

$$\alpha_t(j,n) = \sum_{i=1}^{M} \sum_{n'=1}^{\min(D,t)} \alpha_{t-n}(i,n')q_{i,j}r_i(n)f_{j,n}(\boldsymbol{y}_{t-n+1:t}) \tag{5}$$

where t=1,...,T; q is the latent state transition probability. The backward variable can be defined similarly as well by $\beta_t(j,n) = P(\boldsymbol{y}_{t+1:T}|S_{t-n+1:t}, \boldsymbol{\theta})$, where j=1,...,M; t=1,...,T; and n={1, ..., $\min(D,t)$}. The backward variable $\beta_t(j,n)$ denotes the probability of seeing the observation $\boldsymbol{y}_{t+1:T}$ when the latent state S lasts for n time units from $t-n+1$ to t with the model parameters $\boldsymbol{\theta}$. Following the traditional setting in the work of De Pierro (1995) and Yu (2010), we start with the initialization of $\beta_T(j,n)$=1, we can therefore calculate the recursion formula as

$$\beta_t(j,n) = \sum_{i=1}^{M} \sum_{n'=1}^{\min(D,T-t)} q_{ji}r_j(n)f_{i,n'}(\boldsymbol{y}_{t+1:t+n'}\beta_{t+n'}(i,n')) \tag{6}$$

11

To calculate the expected parameters at the E-step, we define three supplementary variables, $\xi_t(i,j)$ as the probability of the state transition from i at t to j at $t+1$ given the model parameters with the observation from $t = 1$ to T; $\eta_t(j,n)$ as the probability of the state staying at j for n given the model parameters with the observation from $t = 1$ to T; and $\gamma_t(j)$ as the state staying at j at time t given the model parameters with the observation from $t = 1$ to T. Therefore, we can derive

$$
\begin{aligned}
\xi_t(i,j) &= P(S_t = i, S_{t+1} = j, \boldsymbol{y}_{1:T}|\boldsymbol{\theta}) \tag{7}\\
&= \sum_{n'=1}^{\min(D,t)} \sum_{n=1}^{\min(D,T-t)} \alpha_t(i,n') q_{ij} f_{j,n}(\boldsymbol{y}_{t+1:t+n}) \beta_{t+n}(j,n)\\
\eta_t(j,n) &= P(S_{t-n+1:t} = j, \boldsymbol{y}_{1:T}|\boldsymbol{\theta}) = \alpha_t(j,n)\beta_t(j,n) \tag{8}\\
\gamma_t(j) &= P(S_t = j, \boldsymbol{y}_{1:T}|\boldsymbol{\theta}) = \sum_{n=1}^{\min(D,t,T-t)} \eta_t(j,n) \tag{9}
\end{aligned}
$$

Based on those defined variables, we can then calculate the expected model parameters after each iteration as

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) &= E_{\boldsymbol{\theta}^{(l)}} \{\log [P_{\boldsymbol{\theta}}(\boldsymbol{y}_1,...,\boldsymbol{y}_T,S_1,\ldots,S_T)] | \boldsymbol{y}_1,...,\boldsymbol{y}_T\}\\
&= E_{\boldsymbol{\theta}^{(l)}} \{\log [P_{\boldsymbol{\theta}}(S_1,\ldots,S_T)] | \boldsymbol{y}_1,...,\boldsymbol{y}_T\} + E_{\boldsymbol{\theta}^{(l)}} \{\log [P_{\boldsymbol{\theta}}(\boldsymbol{y}_1,...,\boldsymbol{y}_T|S_1,\ldots,S_T)] | \boldsymbol{y}_1,...,\boldsymbol{y}_T\}\\
&= \left[\sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j\neq i} \frac{\xi_t(i,j)}{\gamma_t(j)\log q_{ij}}\right] + \left[\sum_{i=1}^{M} \gamma_0(i)\log\delta_i\right] + \left[\sum_{t=1}^{T}\sum_{j=1}^{M}\sum_{n=1}^{D} \frac{\eta_t(j,n)}{\gamma_t(j)\log r_j(n)}\right]\\
&\quad + \left[\sum_{t=1}^{T}\sum_{j=1}^{M} \gamma_t(j)\log P(\boldsymbol{y}_t|\boldsymbol{y}_{t-1:\max(1,t-p)},\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j,\boldsymbol{A}_j)\right], \tag{10}
\end{aligned}
$$

where the $\boldsymbol{\theta}^{(l)}$ represents the model parameters at the l-th iteration, the $P(\boldsymbol{y}_t|\boldsymbol{y}_{t-1:\max(1,t-p)},\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j,\boldsymbol{A}_j)$ denotes the probability density of the p-th order autoregressive process, which is dependent on the latent states.

After the derivation of the model parameters at the E-step, we can easily calculate the maximized value of parameters by separating each one in $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)})$. Therefore we can have,

$$
\begin{aligned}
\delta_j &= \frac{\gamma_0(j)}{\sum_j \gamma_0(j)} \tag{11}\\
q_{ij} &= \frac{\sum_t \xi_t(i,j)}{\sum_{j\neq i}\sum_t \xi_t(i,j)} \tag{12}\\
r_j(n) &= \frac{\sum_t \eta_t(j,n)}{\sum_n \sum_t \eta_t(j,n)} \tag{13}
\end{aligned}
$$

12

We can update $\boldsymbol{\mu}_j$ as the traditional unpenalized conditional mean, and matrix of autoregressive process $\boldsymbol{A}_j$ as well by the least square regression in LASSO regularized VAR model, in which the weight $\gamma_t(j)$ has been combined with the probability density of the observation $P(\boldsymbol{y}_t|\boldsymbol{y}_{t-1:t-p})$. We follow the pathwise coordinate descent optimization method as the work of Friedman et al. (2007) for updating the values of $\boldsymbol{\mu}_j$, and $\boldsymbol{A}_j$. The Σ_j can be updated as the discussion of equation 2, in which as a combination of weighted variance of the error of VAR model with a constrained identity matrix $W\alpha I_p$. Following the proof of the asymptotic properties of the EM method in the work of Bickel et al. (1998); Barbu and Limnios (2009); Trevezas and Limnios (2011), the estimated parameters of the model $\hat{\boldsymbol{\theta}}_T$ is strongly consistent when the time T achieves to infinity. Particularly, the Trevezas and Limnios (2011) shows the consistency and asymptotic normality for the estimated parameters for hidden semi-Markov model with finite latent states. It indicates a positive probability of state transition as $P(S_{t+\tau} = j|S_t = i) > 0$, and a finite conditional density of the state duration $r_i(n)$, when states $i, j \in \{1, ..., M\}$, and τ is a positive integer.

3   Simulations

To evaluate the performance of the proposed Lasso-VHsMM model, we firstly use the simulated data to measure the accuracy of the parameters that are retrieved from the model. The simulated data is a multivariable time series of dimension $d=50$, generated from $M=2$ latent states with the sample size $T=1000$. The parameter for simulation is summarized in the Table 1.

Table 1: Simulation parameters. This table contains the parameter values for simulating the multivariable time series data to evaluate the performance of Lasso-VHsMM

Parameters	Values
Data dimension	$d=50$
Sample size	$T=500$
Number of states	$M=2$
Mean of states	$\boldsymbol{\mu}=\boldsymbol{0}_{50\times1}$
Max state duration	$D=30$

We construct the transition probability matrix as

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{14}$$

which indicates a self-transition is not allowed. It is due to the explicitly defined state duration density $\boldsymbol{r}_i(n)$, that assigns a certain likelihood of staying at a state i for n time units. Accordingly,

we simulate the time series considering the dense and sparse matrices of the covariance $\mathbf{\Sigma}$ and auto-regression $\mathbf{A}$ following the work of Fiecas et al. (2017). For the sparse matrix, we define the sparse covariance matrices for state 1 and 2 as

$$\mathbf{\Sigma}_{1,ij}^{Sparse} = \begin{cases} e^{-|i-j|}, & if|i-j| < 2 \\ 0, & otherwise \end{cases} \tag{15}$$

$$\mathbf{\Sigma}_{2,ij}^{Sparse} = \begin{cases} e^{-2|i-j|}, & if|i-j| < 2 \\ 0, & otherwise \end{cases} \tag{16}$$

$$\mathbf{A}_1^{Sparse} = \begin{bmatrix} 0.1 & 0.05 & 0 & \dots & \dots & 0 \\ 0.05 & 0.1 & 0.05 & 0 & \dots & 0 \\ 0 & 0.05 & 0.1 & 0.05 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0.05 & 0.1 & 0.05 \\ 0 & \dots & \dots & 0 & 0.05 & 0.1 \end{bmatrix}, \tag{17}$$

$$\mathbf{A}_2^{Sparse} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & 0 \end{bmatrix}. \tag{18}$$

For the dense matrices, we have

$$\mathbf{\Sigma}_{1,ij}^{Dense} = e^{-|i-j|}, \tag{19}$$

$$\mathbf{\Sigma}_{2,ij}^{Dense} = e^{-2|i-j|}, \tag{20}$$

$$\mathbf{A}_{1,ij}^{Dense} = \frac{1}{10}e^{-|i-j|}, \tag{21}$$

$$\mathbf{A}_{2,ij}^{Dense} = \frac{1}{10}e^{-2|i-j|}. \tag{22}$$

We estimate two models under the proposed framework: one is Lasso-VHsMM model (with regularization) and another is VHsMM model (without regularization) using the simulated data. We choose the first order autoregressive dependence $p=1$. For the Lasso-VHsMM model, we select the regularization penalty λ and λ_a for the covariance matrix Σ and auto-regression matrix A respectively by the method discussed in Section 2.2.2: minimizing one step ahead mean square forecast error (MSFE). To select the appropriate values for λ_a, we choose a range of $\lambda_a \in [0.1, 100]$

and 20 grid points with equal log-scale to search an appropriate penalty. Similarly, we choose a range of $\lambda \in [0.0001, 1]$ with 100 grid points for an appropriate search.

Following the work of Bańbura et al. (2010); Nicholson et al. (2017); Baek et al. (2017), we divide the simulated data points to two time periods by $T_1=800$. Therefore the data in the initialization period $[1:T_1=800]$ is to estimate the models, while the remained data in the period $(T_1=800:1000]$ is for model validation. The simulation process is as: 1) simulation of a time series by the parameter in Table 1 and equation 14-22; 2) estimate and validate the Lasso-VHsMM and VHsMM model by the simulated data; 3) calculate the Frobenius Norm (or Euclidean norm) of the differences between the estimated parameters and the real parameters for the simulated the time series. We repeat this process for 1000 times and calculate average values of the Frobenius Norm.

Table 2: The error of estimated parameters under Frobenius Norm. This table contains the Frobenius Norm (the averaged value with standard deviation in the parentheses) of the difference between the estimated and real parameters by 1000 simulated data.

	Sparse		**Dense**	
Parameter	VHsMM	Lasso-VHsMM	VHsMM	Lasso-VHsMM
$\|\hat{\boldsymbol{A}}_1 - \boldsymbol{A}_1\|_F$	3.1242 (0.2447)	0.3089 (0.0004)	3.5463 (0.2326)	0.3217 (0.0009)
$\|\hat{\boldsymbol{A}}_2 - \boldsymbol{A}_2\|_F$	3.3935 (0.2156)	0.1088 (0.0006)	3.3719 (0.2778)	0.1277 (0.0006)
$\|\hat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1\|_F$	3.0332 (0.1446)	1.0027 (0.0028)	3.1183 (0.1015)	2.0196 (0.0837)
$\|\hat{\boldsymbol{\Sigma}}_2 - \boldsymbol{\Sigma}_2\|_F$	3.0851 (0.1199)	1.1083 (0.0041)	3.0098 (0.1180)	1.9472 (0.0200)
$\|\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1\|_F$	0.4190 (0.0544)	0.3228 (0.0022)	0.4276 (0.0506)	0.3133 (0.0020)
$\|\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2\|_F$	0.4057 (0.0505)	0.3027 (0.0016)	0.4024 (0.0513)	0.3068 (0.0012)
$\|\hat{\boldsymbol{r}}_1 - \boldsymbol{r}_1\|_F$	0.1547 (0.0303)	0.1284 (0.0047)	0.2233 (0.0324)	0.1789 (0.0027)
$\|\hat{\boldsymbol{r}}_2 - \boldsymbol{r}_2\|_F$	0.1615 (0.0286)	0.1298 (0.0018)	0.1732 (0.0240)	0.1129 (0.0011)

Table 3: An example of latent state identification confusion matrix. The Lasso-VHsMM and VHsMM model are estimated by the simulated time series (observations) for identifying the latent states that generate the observations.

				Real		
	Lasso-VHsMM			State 1	State 2	**Precision**
		Estimated	State 1	177	6	**96.7213%**
			State 2	10	125	
Sparse			**Recall**	**94.6524%**		
	VHsMM			State 1	State 2	**Precision**
		Estimated	State 1	176	10	**94.6237%**
			State 2	11	121	
			Recall	**94.1176%**		
				Real		
	Lasso-VHsMM			State 1	State 2	**Precision**
		Estimated	State 1	171	14	**92.4324%**
			State 2	16	117	
Dense			**Recall**	**91.4439%**		
	VHsMM			State 1	State 2	**Precision**
		Estimated	State 1	168	16	**91.3043%**
			State 2	19	115	
			Recall	**89.8396%**		

Table 4: Average precision, recall, and accuracy of latent state identification. This table contains the state identification by the estimated Lasso-VHsMM and VHsMM model. The identifications are on 1000 simulated time series. The precision and recall for each time series has been calculated as the Table 3 and the average precision, recall, and state classification accuracy across all 1000 time series are illustrated.

		Precision	Recall	Accuracy
Sparse	**Lasso-VHsMM**	**97.27%**	**95.19%**	**94.97%**
	VHsMM	95.63%	94.09%	93.08%
Dense	**Lasso-VHsMM**	**93.99%**	**91.49%**	**90.88%**
	VHsMM	91.26%	88.36%	89.62%

The results are illustrated in tables 2 and 3. In Table 2 we can clearly observe that the errors of covariance and autoregression matrices generated by Lasso-VHsMM model are highly reduced compared to the ones by VHsMM model. This is especially obvious in estimations of sparse matrices (the columns under "**Sparse**"), which follows our expectation of the "matrix sparsity" discussed in Section 2.2. For the dense matrices, the Lasso-VHsMM model also outperforms the VHsMM model in estimating Σ and A with relatively less significance. For the estimation of the conditional mean μ of the VAR component, and the state duration density r, Lasso-VHsMM model achieves slightly better average results than the VHsMM model with much lower standard deviations. From the last four rows in Table 2, we can observe that the standard deviations of the estimated parameters by the Lasso-VHsMM model are consistently one-tenth of the ones by VHsMM model. This shows a stable better performance of the Lasso-VHsMM model in tracking the time series. In Table 3, we show an example of the confusion matrices of the identification of the latent states by two models. For both the dense and sparse matrices, the Lasso-VHsMM model achieves better state identification performance in precision and recall. We repeat this identification on 1000 simulated time series and generate the average precision, recall, and the accuracy of state classification as shown in Table 4. It's clear that the Lasso-VHsMM model consistently outperforms the VHsMM model in recognising the latent states in multivariable time series. Note that the "Accuracy" column in Table 4 show an averaged accuracy of correctly identifying both the state 1 and 2.

4   Financial Data Analysis

We apply the Lasso-VHsMM model to detect anomalous trading behaviour and identify market regime switching. The financial datasets we use in the empirical analysis include three types: 1) 5-second order book data of component stocks of Chinese CSI 300 index that are listed in Shanghai Stock Exchange with the time period from August 2005 to August 2016; 2) daily U.S. industry portfolio data from 1 Jul 1926 to 28 Jun 2019; 3) 1-hour foreign exchange rates from Jan 2009

to Aug 2015. Those three applications cover a broad spectrum of financial scenarios with data frequencies of 5-second, one-hour and daily respectively. We choose the three applications due to their similar underlying task: to decode the latent incentives behind the observed phenomenon.

4.1 Benchmark Models

To evaluate the result of the proposed Lasso-VHsMM model, we compare its financial data analysis result with two well-established models: Shrink Hidden Markov Model (shrink-HMM) following the study of Fiecas et al. (2017), and Markov Chain Monte Carlo (MCMC) for continuous-time asset pricing models following the study of Johannes and Polson (2010); Verhofen (2005). Following the configurations in Section 5 of Fiecas et al. (2017), we calculate the Akaike information criterion (AIC) to determine the states and yield the lowest AIC with two states. We estimate the shrinkage weights for state 1 and 2 and use the estimated model to capture the stability and high-risk states of U.S. industry. Following the regime switching configuration in Section 5.3 of Johannes and Polson (2010) and Section 4 of Verhofen (2005), we consider the time series follows a stochastic differential equation and the drift and diffusion are driven by a continuous, discrete state Markov Chain. We impose the vector of mean return and the variance-covariance matrix. The Markov Chain is generated by the Gibbs sampler with 110,000 iterations.

4.2 Pseudo trading strategy

As the previous study points out Fiecas et al. (2017), the latent state identification is merely an unsupervised coarse classification and "will not allow us to detect the precise" changes in the data. To obtain a clearer comparison among different models, we follow the idea of the trading strategy in Nystrup et al. (2020). As the data in the empirical studies is financial data, we implement a long-short pseudo trading strategy according to the identified latent states of the financial markets. In the stable state (the normal state), we long the underlying financial asset; while in the volatile state (the anomalous state), we short the underlying financial asset. We assume the transaction cost is 1% of a trade. The final total value of the asset position is calculated after the final date. The return of the trading strategy is then computed based on the initial asset, transaction cost and the final value of the asset. In the pseudo trading strategy, we run each model for 1000 times and calculate the standard deviation of the 1000 return and then find the sharp ratio for each model. We assume we can long and short the U.S. industry portfolio index and the FX in the trading strategy although those assets might not be purchased at the same time with no additional cost.

17

However, the assumption of the purchasability is to compare the results of the Lasso-VHsMM, shrink-HMM, and MCMC models rather than a study of the profitability of a trading strategy.

4.3 Trading Behaviour

The concept of "disruptive trading behaviour" has been defined by the US Commodity Futures Trading Commission (CFTC) in 2013 CFTC (May 20 2013) as the buy or sell orders with intention of cancellation prior to the execution with "reckless disregard" for the market integration and regulation. The disruptive trading activities have been discussed in Cao et al. (2014, 2015); Zhai et al. (2017); Wang et al. (2019) as the market manipulation activities, undertaken through carefully designed purchasing and selling order sequences as a means of inducing market price movements to follow their expectations, thus resulting in an anomalous market fluctuation. A common way to detect the anomalous trading behaviours is modelling the market price as the work of Cao et al. (2014); Zhai et al. (2017, 2018) with the known manipulative cases. Unfortunately, on one hand, real disruptive trading examples are rarely publicly disclosed due to the prohibition from regulatory rules; on the other hand, the market price is the expected convergence of all aggregated market activities Choi et al. (2019) and inspecting the market price only can not reflect the true incentive of the trading behaviour. For example, several jumps of price may not lead to an anomalous trading action, but a sequence of such jumps with certain patterns may be.

Consequently, to identify whether the intention behind certain trading behaviours is anomalous, we analyze both the prices and depths on both bid and ask sides of the three-level order book following the work of Hautsch and Huang (2012); Jiang et al. (2019). We obtain the order book data with five-second interval from China Security Market Trade & Quote of the China Stock Market & Accounting Research (CSMAR) database. We select all component stocks of the Chinese CSI 300 index that are listed in Shanghai Stock Exchange with the exclusion of the finance sector and the companies listed less than three years. There are 148 stocks from August 2005 to August 2016 selected as the final dataset. We randomly select 30 stocks and summarize the descriptive statistics in the Table Appendix-1.

The market price p_t and depth v_t in bid and ask sides of a 3-level order book are combined as a multi-variable time series. Therefore we have a $d = 4 \times 3 = 12$ dimension order book vector (OBV) defined as

$$OBV_t = \left[p_t^{a,1}, p_t^{b,1}, v_t^{a,1}, v_t^{b,1}, p_t^{a,2}, p_t^{b,2}, v_t^{a,2}, v_t^{b,2}, p_t^{a,3}, p_t^{b,3}, v_t^{a,3}, v_t^{b,3}, \right] \qquad (23)$$

where t denotes the time; $p_t^{a/b,k}$ represents the market ask (a) or bid (b) price at the k-th level order book; and $v_t^{a/b,k}$ represents the market ask or bid depth (indicated by total outstanding volume) at the k-th level order book.

We then define the observation of the Lasso-VHsMM model $\boldsymbol{y}_t$ as the log return of each dimension of OBV_t.

$$\boldsymbol{y}_t = \left[\log \frac{p_t^{a,1}}{p_{t-1}^{a,1}}, \log \frac{p_t^{b,1}}{p_{t-1}^{b,1}}, \log \frac{v_t^{a,1}}{v_{t-1}^{a,1}}, \log \frac{v_t^{b,1}}{v_{t-1}^{b,1}}, \ldots \ldots, \log \frac{v_t^{a,3}}{v_{t-1}^{a,3}}, \log \frac{v_t^{b,3}}{v_{t-1}^{b,3}} \right] \tag{24}$$

We show four examples of the correlation among the log returns in equation 24 by heatmap in Figure Appendix-1. The heatmaps show that in the lag 0 heatmap (first row in Figure Appendix-1), half of the dimensions have moderately strong, positive correlations and the other half, in the contrast, have weak correlations. The lag 1 heapmap, however, shows weak correlations across most dimensions. The sparse correlation supports the motivation of matrix sparsity in Section 2.2 on the regularized auto-regression matrix of the Lasso-VHsMM model.

Evaluating the Lasso-VHsMM model on real data is not easy because the latent states are usually not observable. As stated in Fiecas et al. (2017), the model does "not allow us to detect the precise date that specific companies are affected". Therefore, we test our proposed model by the order book data on a specific day in Shanghai Stock Exchange: the circuit break days of 4 and 7 Jan 2016 Wei (2017). The stock market circuit breaker has been employed in Shanghai Stock Exchange for four days 4-7 Jan 2016. On 4 and 7 Jan 2019, the breaker has melted due to the turbulence on the market. As the turbulence event has been publicly discovered, we clearly know that the trading behaviours at around 13:33 on 4 Jan 2016, and 9:40 and 9:58 on 7 Jan 2016 are extremely anomalous. We use the order book data of 4 and 7 Jan 2019 as the testing data to evaluate our proposed model. The order book data from Jun to Dec 2015 is used to estimate the model.

Similarly as the simulation experiments in Section 3, we estimate the Lasso-VHsMM and VHsMM model with first order autoregressive dependence $p=1$, and then test them by the same data respectively. The regularization penalties are selected by the method discussed in Section 2.2.2: minimizing one step ahead mean square forecast error (MSFE). The penalty λ_a is selected by a 20 points grid from a range of $\lambda_a \in [0.1, 100]$; while the penalty λ is selected by a 100 points grid from a range of $[0.0001, 1]$.

We show the 12 order book variables in equation 24 and the decoded latent states of four

Figure 3: The examples of order book data and latent states decoded by Lasso-VHsMM and VHsMM models. Four stocks are selected as examples with ID 600005, 600008, 600009, 600010. The figure shows the 12 order book variables in equation 24 with 5-second interval. Panel A shows the data from 13:00 to 13: 40 on Jan 4, 2016. The colour lines are the 12 dimensions of the order book variables in equation 23. The grey area are the identified latent states by the Lasso-VHsMM model (left column) and VHsMM model (right column). The X and Y-axis represent the physical time t, and the log return of order book price and volume $\boldsymbol{y}_t$, defined in equation 24, respectively.
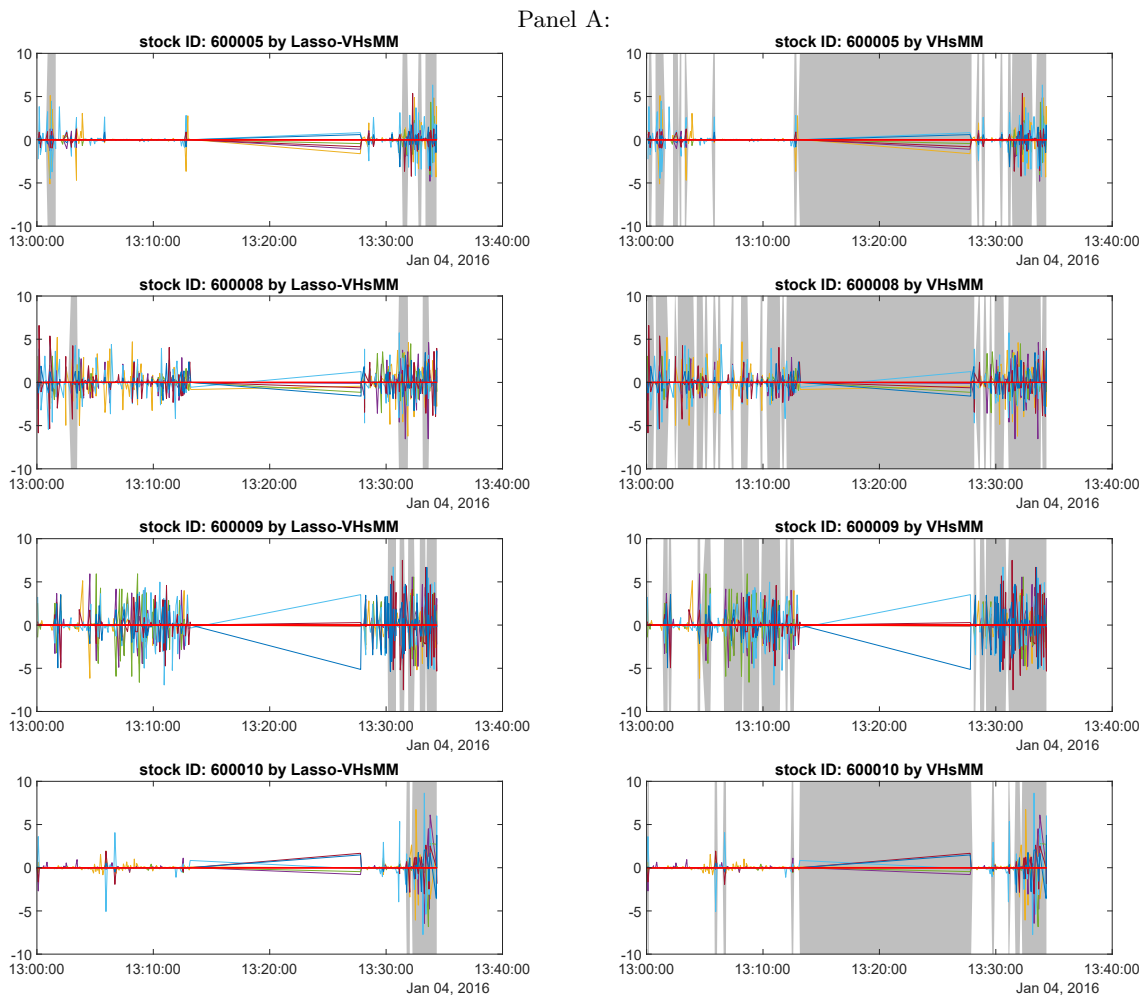
Panel A:

example stocks in Figure 3. The Panel A of Figure 3 shows the data from 13:00 to 13:40 on Jan 4, 2016, on which the circuit breaker has been melt for the first time. The Panel B of Figure 3 shows the data from 9:20 to 10:00 on Jan 7, 2015, when the circuit breaker has been melt for the second time. The figures on the left column contains the latent states decoded by the Lasso-VHsMM model and the ones on the right column is by the VHsMM model.

The first circuit breaker melt at 13:33 on Jan 4, 2016 due to the anomalous trading behaviours several minutes before the time. It is clear that on the left column at Panel A, the Lasso-VHsMM model accurately recognizes the anomalous latent states (grey area) lasting from approximate 13:30 to 13:33, which is immediately leading to the melt of the circuit breaker. On the right column of

20

Panel B: The examples of order book data and latent states decoded by Lasso-VHsMM and VHsMM models. Four stocks are selected as examples with ID 600005, 600008, 600009, 600010. The figure shows the 12 order book variables in equation 24 with 5-second interval. Panel B shows the data from 9:20 to 10:00 on Jan 7, 2016. The colour lines are the 12 dimensions of the order book variables in equation 23. The grey area are the identified latent states by the Lasso-VHsMM model (left column) and VHsMM model (right column). The X and Y-axis represent the physical time t, and the log return of order book price and volume $\boldsymbol{y}_t$, defined in equation 24, respectively.
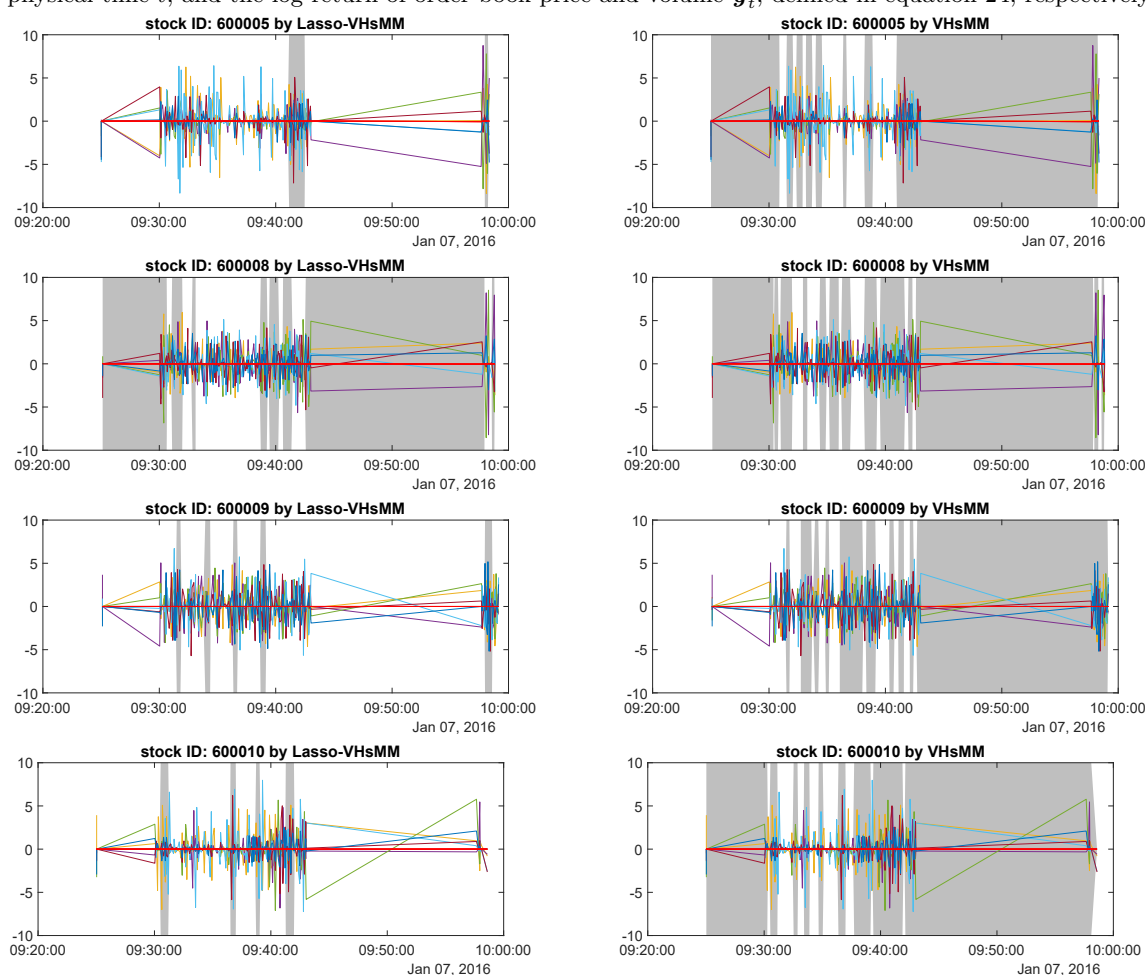
Panel A, the VHsMM model, however, identifies much more anomalous latent states that generate volatile behaviours. A turbulent market on Jan 4, 2016 makes the VHsMM model identifying the anomalous latent states more aggressively than the Lasso-VHsMM model, which results in a more chaotic latent state recognition.

Panel B shows the data from 9:20 to 10:00 on Jan 7, 2016, on which a more turbulent market can be observed as the record in Wikipedia (2019). On the right column, the VHsMM model identifies almost all market behaviours as anomalous state; while on the left column, the Lasso-VHsMM model detects the anomalous states at around 9:40, 9:58 and a few other time points. We see that, though the results in two columns have some similar chaotic latent states as the stock 600008, the Lasso-VHsMM model yields a more stable recognition of underlying anomalous latent states than that of the VHsMM model. More examples in Figure Appendix-2 show the similar results.

We point out that the reconstruction of the latent states from the observed variables is a "coarse classification" as the statement in work of Fiecas et al. (2017). This classification unable to provide us an exactly accurate detection with dates and times that specific events occur and affect the market. However, this recognition by the Lasso-VHsMM model enables us mostly accurately to capture the transition in the financial market between the states of normal and anomalous.

4.4 US Portfolio

Following the work of Fiecas et al. (2017); Nystrup et al. (2018), we apply the Lasso-VHsMM and VHsMM model to the U.S. industry portfolio data, which is publicly available at Kenneth R. French (2019). The data consists of daily return of 49 different industry sections taken from NYSE, NASDAQ, and AMEX. The data covers a considerable long time period from 1 Jul 1926 to 28 Jun 2019 and has 49,030 records with $d = 49$ dimensions.

We apply the Lasso-VHsMM, VHsMM, shrink-HMM, and MCMC models on the data to reconstruct the latent states behind the observed daily returns. We optimize the Lasso-VHsMM model by the method in Section 2.2.2: minimizing one step ahead mean square forecast error (MSFE) with the same penalty configurations for λ and λ_a as Section 4.3: 20 points grid from a range of $\lambda_a \in [0.1, 100]$ for penalty λ_a and 100 points grid from a range of $[0.0001, 1]$ for the penalty λ.

We show the recognized anomalous latent states in Figure 6. In the Panel A, we show the recognized anomalous states by the Lasso-VHsMM model (left column) on the data across 93 years from 1926 to 2019. We see that the anomalous state occurred during the big events including the great recession in 1930s Wikipedia (2019), the great inflation in 1970s, dot-com bubble in 2000,

Figure 4: The examples of U.S. industry portfolio data and latent states decoded by Lasso-VHsMM, VHsMM, Shrink-HMM, and MCMC models. The figure shows the daily return of 49 different industry sections. The colour lines are the 49-dimensional U.S. industry portfolio data. The grey area are the identified latent states by models. Panel A shows the data from 1 Jul 1926 to 28 Jun 2019 by the Lasso-VHsMM and VHsMM models; Panel B shows the data from 3 Jan 2007 to 28 Jun 2019 by the Lasso-VHsMM model, VHsMM model, Shrink-HMM, and MCMC models. The X and Y-axis represent the year, and the log return of the U.S. industry index, respectively.

Panel A:

Panel B:

23
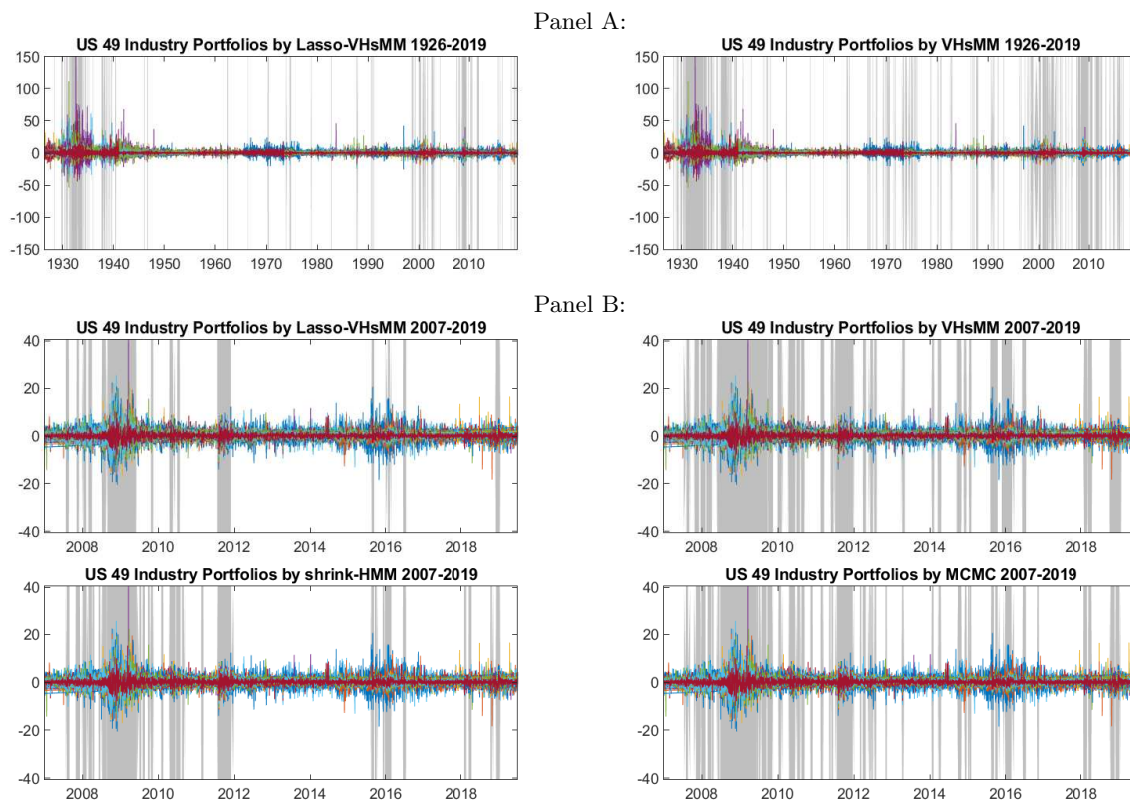
as well as the global financial crisis of 2008. To have a closer look, we run the Lasso-VHsMM, shrink-HMM, and MCMC models again on the data from 3 Jan 2007 to 28 Jun 2019 and show the results in Panel B. From the results of Lasso-VHsMM model on the left column, we have a more detailed recognition of many important events, which include the financial crisis from 2008 to 2009, the double-dip recession at 2011, the shocks at 2016 (Chinese stock market crash, OPEC cut, and Brexit), and the uncertainty at 2019. We compare the states on the left and right columns in Panel A and B. We see that, although the latent states have some similarities, the Lasso-VHsMM model achieves more interpretable and stable recognition of the underlying states. We argue that, such a recognition is not an accurate classification but an unsupervised learning. The Lasso-VHsMM model does not allow us precisely forecast the events through the latent events, but provides a powerful method to capture the transition in the U.S. industry between the stable position and the anomalous status, which is highly risky to the market as the work of Fiecas et al. (2017). In Panel B, we also show the results of VHsMM, shrink-HMM, and MCMC models. We can clearly observe that the shrink-HMM achieves quite similar latent state identification as Lasso-VHsMM except some more tiny events around late 2009 and early 2010. However, the MCMC model identifies most of the tiny rather than the primary states.

In addition to the state identification, Table 5 shows the results of the long-short pseudo trading strategy using different models. The purpose of the strategy is to compare the different latent state identification rather than a study of the profitability. The results clearly show that the Lasso-VHsMM achieves the best sharp-ratio and is favorable to all other models in identifying the stable and volatile financial states.

Table 5: This table shows the results of the long-short pseudo trading strategy based on the U.S. industry portfolio index with Lasso-VHsMM, VHsMM, shrink-HMM, and MCMC models.

	Return	Risk	Sharpe Ratio	Turnover
Lasso-VHsMM	0.1775	0.0985	1.8025	1.2499
VHsMM	0.1235	0.1448	0.8530	5.5133
shrink-HMM	0.1519	0.1096	1.3864	6.2483
MCMC	0.0972	0.2055	0.4732	6.4059

4.5 Foreign exchange rate

In addition to the empirical study of the multivariate data, we consider an application to four primary foreign exchange rates: EUR/USD, GBP/USD, GBP/EUR and GBP/JPY from 2009 to 2015. The choice of the FX rates is mainly due to that the FX rate is highly liquid, volatile and

sensitive to most economic events. The choice of FX rate also supplements the empirical study by one-dimensional financial data in addition to the 12-dimensional data of stocks and 49-dimensional data of U.S. portfolio. The Figure 5 shows the identified states of the rate of EUR/USD and GBP/USD using the Lasso-VHsMM, shrink-HMM, and MCMC models. The results show that the Lasso-VHsMM model effectively identifies the primary latent states and is not over-sensitive to insignificant changes. The shrink-HMM model identifies the most primary states and some tiny changes as well. However, the MCMC model is very sensitive to the one-dimensional FX data and identifies most volatile events. For a further comparison, we once again carry out the pseudo trading strategy on the FX rates. The results are reported in the Table 6. The Lasso-VHsMM model achieves the best sharp-ratio by identifying the primary latent states, while the shrink-HMM achieves the second best performance with more identified latent states. However, the MCMC model identifies most volatile states and therefore worsens the performance with over-sensitive trading activities. The latent state identification results for the rates of GBP/EUR and GBP/JPY are illustrated in Figure Appendix-3 in the appendix for saving the space. The similar results can be observed

Table 6: This table shows the results of the long-short pseudo trading strategy based on the FX rates of EUR/USD and GBP/USD with Lasso-VHsMM, shrink-HMM, and MCMC models.

EUR/USD	Return	Risk	Sharpe Ratio	Turnover
Lasso-VHsMM	0.1585	0.1318	1.2025	1.2499
shrink-HMM	0.1525	0.1369	1.1139	4.2483
MCMC	0.0991	0.1752	0.5653	6.4059
GBP/USD	Return	Risk	Sharpe Ratio	Turnover
Lasso-VHsMM	0.1869	0.1054	1.7729	1.2499
shrink-HMM	0.1451	0.1087	1.3348	4.2483
MCMC	0.1009	0.2178	0.4632	6.4059

5 Conclusion

The traditional hidden Markov model has previously been substantially extended to relax either the Markovian assumption of the latent states or the conditional independent randomness assumption of the observed variables. However, a general framework seeking to release all of those assumptions has not yet to be developed. Thus, this study represents a novel approach in addressing this issue. We extend the traditional hidden Markov model framework and develop a vector autoregressive (VAR) hidden semi-Markov model with Lasso regularization (Lasso-VHsMM). Our

Figure 5: The examples of FX data and latent states decoded by Lasso-VHsMM, shrink-HMM, and MCMC models. The figure shows the daily return of two FX rate: EUR/USD and GBP/USD. The left column shows the data of EUR/USD from 1 Jan 2009 to 31 Aug 2015; The right column shows the data of GBP/USD from 1 Jan 2009 to 31 Aug 2015. The grey area are the identified latent states by the models. The X and Y-axis represent the physical time, and the log return of the hourly FX rate, respectively.
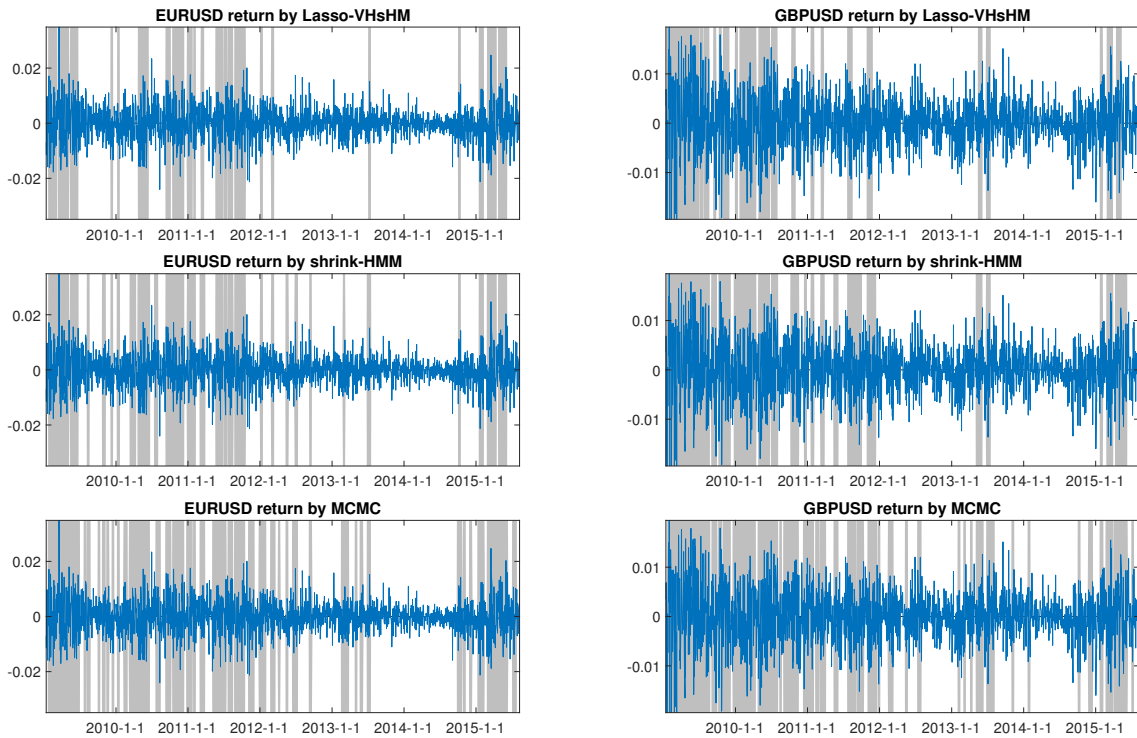
theoretical contribution includes three strands of extensions. In the proposed Lasso-VHsMM model, we consider the observed variables following a p-order vector autoregressive process, allow the latent states evolving via a semi-Markov chain, and shrink the auto-regression and covariance matrices by a penalized maximization likelihood method. This design particularly suits the practical financial data, which is usually multi-dimensional, auto-correlative, tends to be generated by certain one or two latent states, and more importantly, contains sparse (or close to sparse) autoregressive and covariance matrices. The empirical study of the model include four parts: the 50-dimensional simulated data; the 5-second interval, 12-dimensional order book data of stocks; the 49-dimensional daily data of U.S. industry portfolio; and the 1-hour interval, 1-dimensional data of four popular foreign exchange rates. We show empirically that the Lasso-VHsMM model consistently outperforms the alternative models in recognizing the latent states of the anomalous events. The latent states of the simulated data are identified by the Lasso-VHsMM model by 94.65% and 91.44% on sparse and dense autoregression matrix respectively. The primary events on the financial market microstructure, FX markets, and the U.S. industry portfolio can be identified accurately by Lasso-VHsMM model. The long-short trading strategy based on the latent states identified by the Lasso-VHsMM shows significantly better sharp-ratio than the one by shrink-HMM and MCMC models. These empirical findings substantiate the novelty of our model as a step towards formulating a general framework for modelling multi-variable financial data.

References

Akaike, H., 1974. A new look at the statistical model identification. In: Selected Papers of Hirotugu Akaike. Springer, pp. 215–222.

Akgiray, V., 1989. Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. Journal of business, 55–80.

Aldridge, I., 2013. High-frequency trading: a practical guide to algorithmic strategies and trading systems. Vol. 604. John Wiley & Sons.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of financial economics 61 (1), 43–76.

Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. The Review of Financial Studies 15 (4), 1137–1187.

Ariki, Y., Jack, M., 1989. Enhanced time duration constraints in hidden markov modelling for phoneme recognition. Electronics Letters 25 (13), 824–825.

Baek, C., Davis, R. A., Pipiras, V., 2017. Sparse seasonal and periodic vector autoregressive modeling. Computational Statistics & Data Analysis 106, 103–126.

Bańbura, M., Giannone, D., Reichlin, L., 2010. Large bayesian vector auto regressions. Journal of applied Econometrics 25 (1), 71–92.

Barbu, V. S., Limnios, N., 2009. Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis. Vol. 191. Springer Science & Business Media.

Baum, L. E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. The annals of mathematical statistics 41 (1), 164–171.

Bickel, P. J., Ritov, Y., Ryden, T., et al., 1998. Asymptotic normality of the maximum-likelihood estimator for general hidden markov models. The Annals of Statistics 26 (4), 1614–1635.

Billio, M., Casarin, R., Rossini, L., 2019. Bayesian nonparametric sparse var models. Journal of econometrics 212 (1), 97–115.

Bilmes, J. A., et al., 1998. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. International Computer Science Institute 4 (510), 126.

Brown, S. J., Warner, J. B., 1985. Using daily stock returns: The case of event studies. Journal of financial economics 14 (1), 3–31.

Bulla, J., Bulla, I., 2006. Stylized facts of financial time series and hidden semi-markov models. Computational Statistics & Data Analysis 51 (4), 2192–2209.

Cai, J., 1994. A markov model of switching-regime arch. Journal of Business & Economic Statistics 12 (3), 309–316.

Candès, E. J., Plan, Y., et al., 2009. Near-ideal model selection by l1 minimization. The Annals of Statistics 37 (5A), 2145–2177.

Cao, Y., Li, Y., Coleman, S., Belatreche, A., McGinnity, T. M., 2014. Adaptive hidden markov model with anomaly states for price manipulation detection. IEEE transactions on neural networks and learning systems 26 (2), 318–330.

Cao, Y., Li, Y., Coleman, S., Belatreche, A., McGinnity, T. M., 2015. Detecting wash trade in financial market using digraphs and dynamic programming. IEEE transactions on neural networks and learning systems 27 (11), 2351–2363.

CFTC, May 20 2013. Antidisruptive practices authority no. 3038-ad96. Release, Commodity Exchange Act.

Chinco, A., Clark-Joseph, A. D., Ye, M., 2019. Sparse signals in the cross-section of returns. The Journal of Finance 74 (1), 449–492.

Choi, J. H., Larsen, K., Seppi, D. J., 2019. Information and trading targets in a dynamic market equilibrium. Journal of Financial Economics 132 (3), 22–49.

Chopin, N., Pelgrin, F., 2004. Bayesian inference and state number determination for hidden markov models: an application to the information content of the yield curve about inflation. Journal of Econometrics 123 (2), 327–344.

Cubadda, G., Guardabascio, B., Hecq, A., 2017. A vector heterogeneous autoregressive index model for realized volatility measures. International Journal of Forecasting 33 (2), 337–344.

Dang, S., Chaudhury, S., Lall, B., Roy, P. K., 2017. Learning effective connectivity from fmri using autoregressive hidden markov model with missing data. Journal of neuroscience methods 278, 87–100.

De Pierro, A. R., 1995. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. IEEE transactions on medical imaging 14 (1), 132–137.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39 (1), 1–22.

Dias, J. G., Ramos, S. B., 2014. Dynamic clustering of energy markets: An extended hidden markov approach. Expert Systems with Applications 41 (17), 7722–7729.

Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of econometrics 105 (1), 131–159.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society, 987–1007.

Ephraim, Y., Malah, D., Juang, B.-H., 1989. On the application of hidden markov models for enhancing noisy speech. IEEE Transactions on Acoustics, Speech, and Signal Processing 37 (12), 1846–1856.

Fiecas, M., Franke, J., von Sachs, R., Tadjuidje Kamgaing, J., 2017. Shrinkage estimation for multivariate hidden markov models. Journal of the American Statistical Association 112 (517), 424–435.

Francq, C., Zakoıan, J.-M., 2001. Stationarity of multivariate markov–switching arma models. Journal of Econometrics 102 (2), 339–364.

Freguson, J., 1980. Variable duration models for speech. In: Proc. Symposium on the Application of Hidden Markov Models to Text and Speech, 1980.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., et al., 2007. Pathwise coordinate optimization. The annals of applied statistics 1 (2), 302–332.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software 33 (1), 1.

Gannot, S., Burshtein, D., Weinstein, E., 1998. Iterative and sequential kalman filter-based speech enhancement algorithms. IEEE Transactions on speech and audio processing 6 (4), 373–385.

Gençay, R., Dacorogna, M., Muller, U. A., Pictet, O., Olsen, R., 2001. An introduction to high-frequency finance. Elsevier.

Granger, C. W., Machina, M. J., 2006. Structural attribution of observed volatility clustering. Journal of Econometrics 135 (1-2), 15–29.

Green, P. J., 1990. On use of the em algorithm for penalized likelihood estimation. Journal of the Royal Statistical Society: Series B (Methodological) 52 (3), 443–452.

Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica: Journal of the Econometric Society, 357–384.

Hamilton, J. D., 1990. Analysis of time series subject to changes in regime. Journal of econometrics 45 (1-2), 39–70.

Hamilton, J. D., 2010. Regime switching models. In: Macroeconometrics and time series analysis. Springer, pp. 202–209.

Hastie, T., Tibshirani, R., Friedman, J., Franklin, J., 2005. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer 27 (2), 83–85.

Hautsch, N., Huang, R., 2012. The market impact of a limit order. Journal of Economic Dynamics and Control 36 (4), 501–522.

Hu, G., Wang, D., 2004. Monaural speech segregation based on pitch tracking and amplitude modulation. IEEE Transactions on neural networks 15 (5), 1135–1150.

Jiang, Y., Cao, Y., Liu, X., Zhai, J., 2019. Volatility modeling and prediction: the role of price impact. Quantitative Finance, 1–17.

Johannes, M., Polson, N., 2010. Mcmc methods for continuous-time financial econometrics. In: Handbook of Financial Econometrics: Applications. Elsevier, pp. 1–72.

John Lu, Z., 2010. The elements of statistical learning: data mining, inference, and prediction. Journal of the Royal Statistical Society: Series A (Statistics in Society) 173 (3), 693–694.

Johnson, R. A., Wichern, D. W., et al., 2002. Applied multivariate statistical analysis. Vol. 5. Prentice hall Upper Saddle River, NJ.

Kalli, M., Griffin, J. E., 2018. Bayesian nonparametric vector autoregressive models. Journal of econometrics 203 (2), 267–282.

Kenneth R. French, 2019. Industry portfolios. [Online; accessed 25-August-2019].

    URL `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

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.

Kim, C.-J., Piger, J., Startz, R., 2008. Estimation of markov regime-switching regression models with endogenous switching. Journal of Econometrics 143 (2), 263–273.

Kritzman, M., Page, S., Turkington, D., 2012. Regime shifts: Implications for dynamic strategies (corrected). Financial Analysts Journal 68 (3), 22–39.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis 88 (2), 365–411.

Levinson, S. E., 1986. Continuously variable duration hidden markov models for automatic speech recognition. Computer Speech & Language 1 (1), 29–45.

Lin, L., Wang, F., Xie, X., Zhong, S., 2017. Random forests-based extreme learning machine ensemble for multi-regime time series prediction. Expert Systems with Applications 83, 164–176.

Luo, C., Seco, L., Wu, L.-L. B., 2015. Portfolio optimization in hedge funds by ogarch and markov switching model. Omega 57, 34–39.

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

Mitchell, C. D., Jamieson, L. H., 1993. Modeling duration in a hidden markov model with the exponential family. In: 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 2. IEEE, pp. 331–334.

Monbet, V., Ailliot, P., 2017. Sparse vector markov switching autoregressive models. application to multivariate time series of temperature. Computational Statistics & Data Analysis 108, 40–51.

Natarajan, B. K., 1995. Sparse approximate solutions to linear systems. SIAM journal on computing 24 (2), 227–234.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica: Journal of the Econometric Society, 347–370.

Nicholson, W. B., Matteson, D. S., Bien, J., 2017. Varx-l: Structured regularization for large vector autoregressions with exogenous variables. International Journal of Forecasting 33 (3), 627–651.

Nystrup, P., Boyd, S., Lindström, E., Madsen, H., 2018. Multi-period portfolio selection with drawdown control. Annals of Operations Research, 1–27.

Nystrup, P., Lindström, E., Madsen, H., 2020. Learning hidden markov models with persistent states by penalizing jumps. Expert Systems with Applications 150, 113307.

Piger, J., 2009. Econometrics: Models of regime changes.

Russell, M., Moore, R., 1985. Explicit modelling of state occupancy in hidden markov models for automatic speech recognition. In: ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 10. IEEE, pp. 5–8.

Rydén, T., Teräsvirta, T., Åsbrink, S., 1998. Stylized facts of daily return series and the hidden markov model. Journal of applied econometrics 13 (3), 217–244.

Sancetta, A., 2008. Sample covariance shrinkage for high dimensional dependent data. Journal of Multivariate Analysis 99 (5), 949–967.

Schwarz, G., et al., 1978. Estimating the dimension of a model. The annals of statistics 6 (2), 461–464.

Stanculescu, I., Williams, C. K., Freer, Y., 2013. Autoregressive hidden markov models for the early detection of neonatal sepsis. IEEE journal of biomedical and health informatics 18 (5), 1560–1570.

Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288.

Trevezas, S., Limnios, N., 2011. Exact mle and asymptotic properties for nonparametric semi-markov models. Journal of Nonparametric Statistics 23 (3), 719–739.

Tsang, E., Chen, J., 2018. Regime change detection using directional change indicators in the foreign exchange market to chart brexit. IEEE Transactions on Emerging Topics in Computational Intelligence 2 (3), 185–193.

Van der Hoek, J., Elliott, R. J., 2019. Introduction to Hidden Semi-Markov Models. Vol. 445. Cambridge University Press.

Verhofen, M., 2005. Markov chain monte carlo methods in financial econometrics. Financial Markets and Portfolio Management 19 (4), 397–405.

Wang, Q., Xu, W., Huang, X., Yang, K., 2019. Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. Neurocomputing 347, 46–58.

Wei, S., 2017. Public attention helps stocks survive during circuit-breaker period in china. In: 3rd International Symposium on Social Science ISSS 2017. Atlantis Press.

Wikipedia, 2019. 201516 Chinese stock market turbulence. `http://en.wikipedia.org/w/index.php?title=2015\%E2\%80\%9316\%20Chinese\%20stock\%20market\%20turbulence&oldid=908160852`, [Online; accessed 24-August-2019].

Wikipedia, 2019. Great recession. [Online; accessed 25-August-2019].
URL `https://en.wikipedia.org/wiki/Great_Recession`

Yu, S.-Z., 2010. Hidden semi-markov models. Artificial intelligence 174 (2), 215–243.

Yuan, M., Huang, J. Z., 2009. Regularized parameter estimation of high dimensional t distribution. Journal of Statistical Planning and Inference 139 (7), 2284–2292.

Zhai, J., Cao, Y., Ding, X., 2018. Data analytic approach for manipulation detection in stock market. Review of Quantitative Finance and Accounting 50 (3), 897–932.

Zhai, J., Cao, Y., Yao, Y., Ding, X., Li, Y., 2017. Computational intelligent hybrid model for detecting disruptive trading activity. Decision Support Systems 93, 26–41.

Zhou, N., Mamon, R., 2012. An accessible implementation of interest rate models with markov-switching. Expert Systems with Applications 39 (5), 4679–4689.

6 Appendix

Table Appendix-1 Summary of a randomly selected sample of stocks This table summarizes the stock code, starting date, the number of observations, the market value (MV), the book-to-market ratio (B/M), turnover, and the industry of 60 randomly selected sample out of 148 stocks. The market value, B/M ratio and turnover are annual average from 2005 to 2015. Unit of market value is 10 billion RMB. Industries include Agriculture, Forestry, Stockbreeding and Fishing (A); Mining (B); Manufacturing (C); Utilities including Electricity, Heating, Gas and Water (D); Construction (E); Whole sale and Retail (F); Transportation, Storage and Post (G); Information Technology (I); Real Estate (K); Renting and Business Services (L); Culture, Sport and Entertainment (R) and Other Industries (S). For stocks with listing code 600019, 600021, 600737, 600795, 600886, 601727 and 601899, the end dates are 24/06/2016, 23/08/2016, 26/08/2016, 30/08/2016, 31/12/2015, 30/08/2016, and 30/08/2016, respectively. For the rest the end date is 31/08/2016.

Code	Start	No. Obs.	MV	B/M	Turnover	Industry	Code	Start	No. Obs.	MV	B/M	Turnover	Industry
600009	20050802	6,096,670	3.58	0.45	283.26	G	600704	20050802	5,146,162	1.00	0.52	587.21	F
600018	20050823	6,217,505	11.44	0.40	159.38	G	600737	20050802	5,496,981	1.33	0.28	604.65	C
600019	20050819	6,814,311	12.63	1.01	189.52	C	600739	20050802	5,888,146	2.52	0.38	487.77	F
600021	20050802	5,789,256	1.30	0.59	302.61	D	600795	20050802	6,464,099	4.68	0.58	333.29	D
600028	20050802	7,003,746	79.53	0.68	192.68	B	600827	20050812	4,986,806	1.40	0.50	419.97	F
600031	20050802	6,572,498	6.08	0.27	334.38	C	600871	20050802	3,990,858	3.73	0.29	717.59	B
600037	20050802	6,144,264	1.57	0.42	398.61	I	600886	20050810	4,961,217	2.62	0.51	309.34	D
600048	20060801	6,254,424	7.22	0.43	358.59	K	600895	20050802	6,052,374	1.77	0.35	362.96	S
600050	20050802	7,095,527	12.19	0.63	317.67	I	600900	20050816	5,727,339	13.93	0.47	162.79	D
600060	20050802	6,202,587	1.17	0.58	502.96	C	601006	20060802	6,118,540	13.99	0.49	225.04	G
600104	20050802	6,648,977	15.45	0.61	175.37	C	601018	20100929	3,039,754	4.97	0.60	237.15	G
600118	20050802	5,948,568	1.79	0.14	573.16	C	601111	20060821	5,858,692	9.97	0.50	320.24	G
600150	20050802	5,811,173	4.87	0.37	355.97	C	601117	20100108	3,956,957	3.62	0.55	463.91	E
600153	20050802	5,528,810	1.86	0.47	356.93	F	601118	20110110	3,342,477	2.85	0.31	914.63	A
600221	20050802	6,335,269	2.86	0.60	422.25	G	601186	20080311	4,981,514	10.61	0.74	319.67	E
600332	20050802	5,457,670	1.70	0.30	480.44	C	601333	20061225	5,745,430	3.13	0.83	341.86	G
600372	20050802	4,187,902	2.01	0.21	467.09	C	601390	20071204	5,265,225	12.88	0.73	281.27	E
600373	20050802	4,466,210	0.92	0.38	769.24	R	601600	20070508	5,552,114	12.04	0.51	283.03	C
600376	20050802	5,449,450	1.54	0.60	561.86	K	601607	20100310	3,771,642	2.58	0.50	427.32	F
600383	20050802	6,216,054	3.16	0.56	408.03	K	601608	20120709	1,890,581	1.70	0.50	732.06	C
600406	20050802	4,900,218	2.38	0.13	298.76	I	601618	20090922	3,955,802	6.88	0.71	282.5	E
600415	20050802	5,069,522	2.57	0.26	285.57	L	601668	20090730	4,415,087	13.59	0.79	267.55	E
600446	20050802	4,271,413	0.68	0.21	614.79	I	601669	20111019	2,588,456	5.94	0.71	738.88	E
600489	20050802	6,445,691	3.23	0.23	633.41	B	601727	20081208	3,981,268	7.82	0.36	325.29	C
600519	20050802	5,808,568	17.78	0.15	111.26	C	601766	20080819	4,582,147	10.22	0.35	107.81	B
600570	20050810	3,458,298	1.32	0.11	434.42	I	601857	20071106	5,315,030	217.61	0.54	399.32	L
600585	20050802	5,763,471	8.82	0.45	241.54	C	601888	20091016	3,471,201	3.26	0.22	476.93	B
600606	20050802	4,784,083	2.30	0.52	686.33	K	601899	20080430	5,039,212	7.87	0.31	333.03	G
600688	20050802	5,462,120	5.44	0.34	554.19	C	601919	20070627	4,911,125	10.25	0.47	1,283.00	D
600690	20050802	6,018,562	3.54	0.34	282.22	C	601989	20091217	3,707,862	10.05	0.42	376.28	C

Figure Appendix-1 The four examples of the correlation heatmap of the log return (lag 0 and 1) of 12 dimensions of OBS in equation 24. Four stocks are selected as examples of the strength of the correlation. The first row shows the lag 0 log-return correlation of stock ID 600005, 600008, 600009, 600010 respectively. The second row shows the lag 1 log-return of those stocks correspondingly.
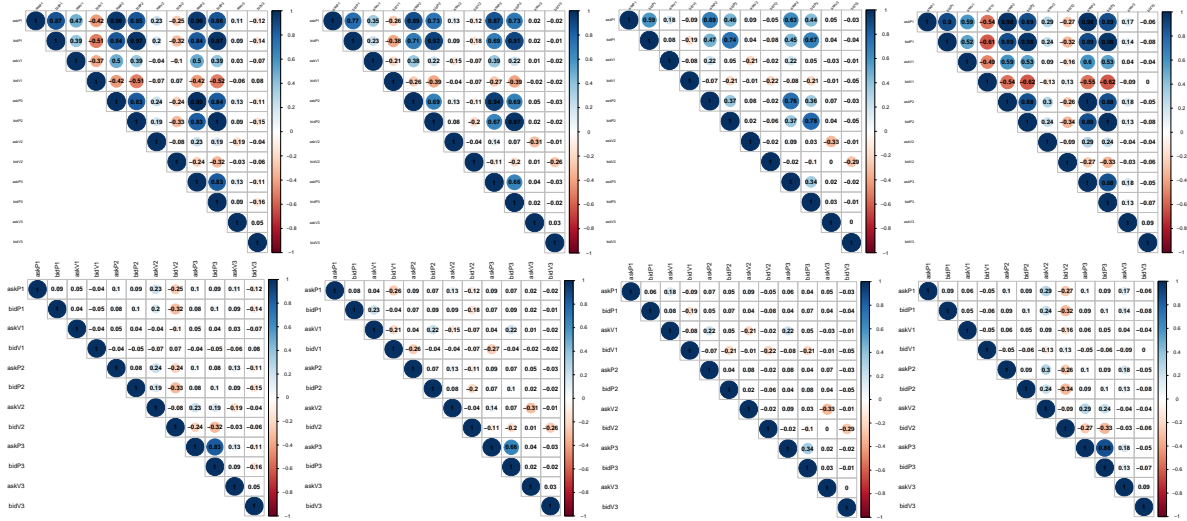

Table Appendix-2 This table shows the results of the long-short pseudo trading strategy based on the FX rates of GBP/EUR and GBP/JPY with Lasso-VHsMM, shrink-HMM, and MCMC models.

GBP/EUR	Return	Risk	Sharpe Ratio	Turnover
Lasso-VHsMM	0.2045	0.1414	1.44622	1.2499
shrink-HMM	0.1602	0.1362	1.17641	4.2483
MCMC	0.0968	0.1911	0.50671	6.4059
GBP/JPY	Return	Risk	Sharpe Ratio	Turnover
Lasso-VHsMM	0.1618	0.1088	1.48736	1.2499
shrink-HMM	0.1603	0.1279	1.25303	4.2483
MCMC	0.1021	0.1861	0.54858	6.4059

Figure Appendix-2 Panel A The examples of order book data and latent states decoded by Lasso-VHsMM and VHsMM models. Five stocks are ID 600011, 600018, 600019, 600021, 600022. Panel A shows the data from 13:00 to 13: 40 on Jan 4, 2016; The grey area are the identified latent states by the Lasso-VHsMM model (left column) and VHsMM model (right column). The X and Y-axis represent the physical time t, and the log return of order book price and volume $\boldsymbol{y}_t$, defined in equation 24, respectively.
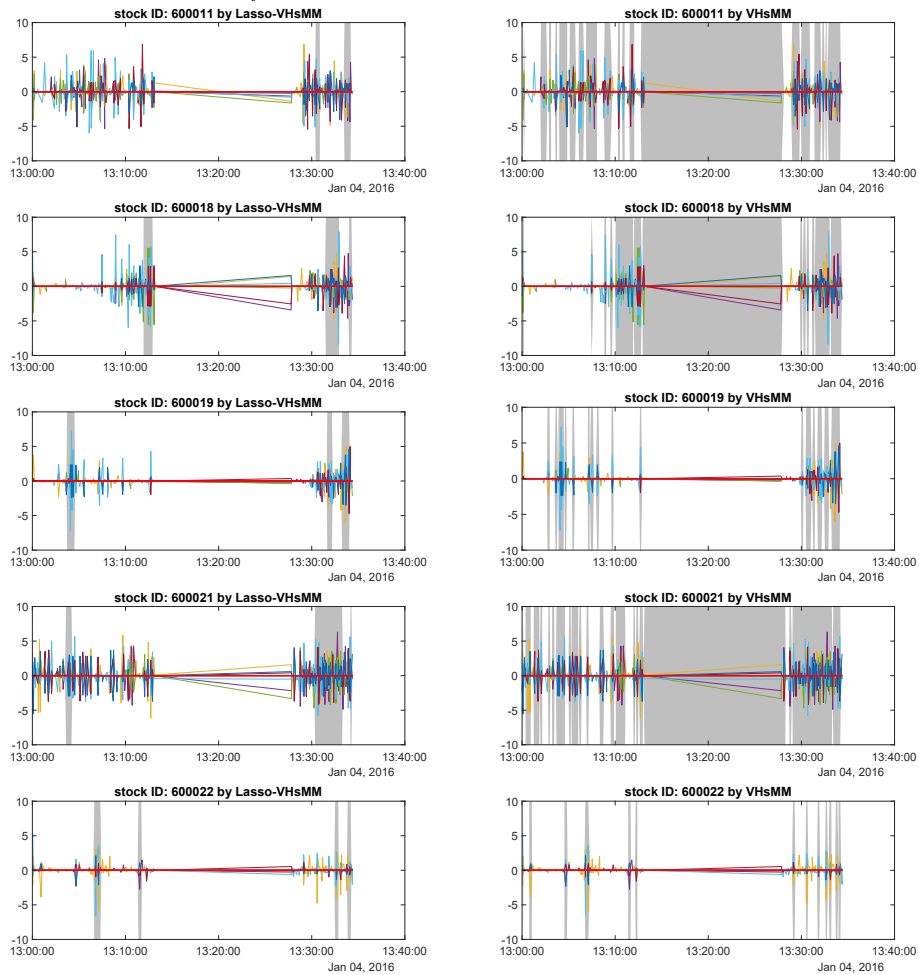
Panel B shows the data from 9:20 to 10:00 on Jan 7, 2016. The grey area are the identified latent states by the Lasso-VHsMM model (left column) and VHsMM model (right column). The X and Y-axis represent the physical time t, and the log return of order book price and volume $\boldsymbol{y}_t$, defined in equation 24, respectively.
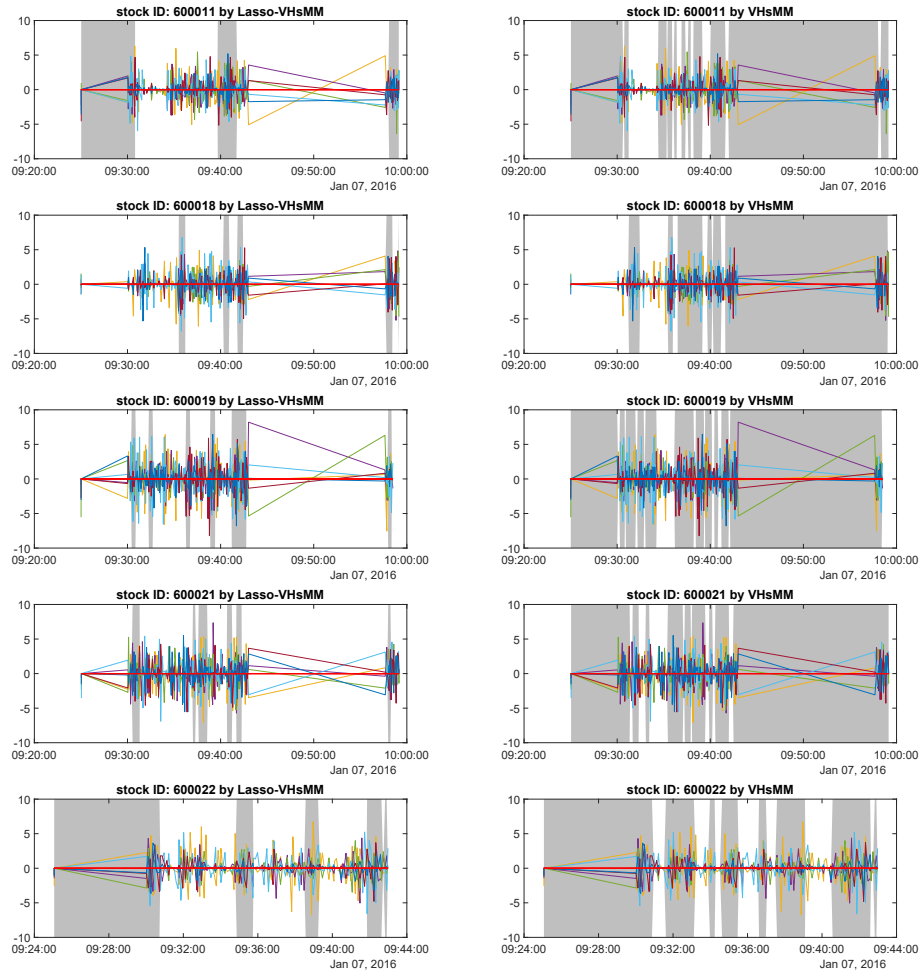
Panel B:

Figure Appendix-3 The examples of FX data and latent states decoded by Lasso-VHsMM, shrink-HMM, and MCMC models. The figure shows the daily return of two FX rate: GBP/EUR and GBP/JPY. The left column shows the data of GBP/EUR from 1 Jan 2009 to 31 Aug 2015; The right column shows the data of GBP/JPY from 1 Jan 2009 to 31 Aug 2015. The grey area are the identified latent states by the models. The X and Y-axis represent the physical time, and the log return of hourly FX rate, respectively.