Statistics and Its Interface Volume 12 (2019) 215–225

Estimation and testing nonhomogeneity of Hidden Markov model with application in financial time series

Mian Huang, Yue Huang*, and Kang He

Both homogeneous and nonhomogeneous Hidden Markov models (HMM) have been gaining increased attention in financial time series modeling. The homogeneous HMM assumes constant transition probabilities, while nonhomogeneous HMM assumes varying transition matrix depended on some covariates. While both assumptions may seem plausible in different applications, there is a lack of studies from a statistical inference aspect. In this paper, we study the nonhomogeneous hidden Markov model, and propose an estimation via a modified EM algorithm, the kernel regression and local likelihood techniques. The motivation for this new procedure is that it enables us to employ a generalized likelihood ratio test procedure to test whether the transition matrix actually depends on a specific covariate. We propose the CV method to select bandwidth and the BIC method to select number of states, and further propose conditional bootstrap method to assess the standard errors of the estimates. We conduct a simulation study to demonstrate our procedure, and show that the Wilk's type of phenomenon holds for the proposed model. Furthermore, we analyze S&P 500 Index return data. Our analysis reveals different patterns in bull and bear markets, and show that the time varying transitions are statistically significant.

AMS 2000 subject classifications: Primary 60J99, 62G08; secondary 62G10.
Keywords and phrases: Hidden Markov model, Nonhomogeneous transition matrix, Generalized likelihood ratio test, Kernel regression, EM algorithm.

1. INTRODUCTION

Hidden Markov model (HMM) has been successfully applied for several decades in pattern recognition and signal processing, especially in the field of automatic speech recognition [15]. In recent decades, HMM has been widely applied in the modelling of financial time series due to its superior capability of extracting the transition patterns under different market conditions. [12] introduced a hidden Markov model to the field of financial modelling and studied the changes in regime of U.S. real GNP through a Markov switching AR model. Following this seminal work, researchers extend traditional time series models in HMM framework to study the internal regime shifts. [10] analysed the real interest rate and found several patterns with shifts in both mean and variance of the series. [3] developed a Markov-ARCH model to examine the issue of volatility persistence in U.S. treasury market. [11] introduced a switching GARCH model to establish the conditional distribution of interest rates. Applications in modelling volatility and returns of stock markets can also be found in [23], [13] and [6].

As the phenomena of momentum, mean reverse and volatility clustering are obvious in the financial market, the transition matrix may vary over time or depend on an exogenous covariate when we employ an HMM approach. There are several studies focusing on the nonhomogeneous HMM in financial time series. [22] studied speculative attacks against EMS currencies by setting transition probabilities between "tranquil" and "speculative" to be a function of fundamentals and expectations. [20] verified that a nonhomogeneous HMM, where a multinomial logit specification is adopted, improves the predictive ability of standard HMM in modelling interest rates. [16] proposed to use a multinomial logit model to parameterize a nonhomogeneous HMM and applied it to estimate the empirical two-pillar Phillips curve for the euro area.

Although nonhomogeneous HMM may have some advantages over homogeneous HMM ([20]), to the best of our knowledge, there lacks of studies from statistical inference aspect. One may wonder whether a homogeneous HMM is appropriate for the data analysis, or whether the transition matrix does depend on a covariate, and therefore nonhomogeneous HMM can provide better estimation. In this paper, we will address this critical issue using the generalized likelihood ratio test developed in [9]. We first introduce the model definition of nonhomogeneous HMM, and then propose a modified EM algorithm using technique of kernel regression and local likelihood method. The transition probability functions and parameters of emission models are estimated simultaneously. Several works study nonparametric estimation for transition matrix in HMM, e.g., [19] developed a K-nearest-neighbor based nonparametric nonhomogeneous HMM, [25] introduced a flexible nonhomogeneous HMM for panel observed data based on B-spline smoothing method. Compared to the

*Corresponding author.

above approach, our method is able to define effective degree of freedom for kernel estimates, and conduct model selection based on information criteria. Using effective degree of freedom for kernel estimates, the BIC is proposed to select the number of states in nonhomogeneous HMM. We propose conditional bootstrap method for standard error estimation and confidence intervals. We further employ generalized likelihood ratio test statistics to test the nonhomogeneity hypothesis. Extensive simulations are conducted to demonstrate our methodologies and show that the Wilk's type of phenomenon holds for the proposed model. Finally, we apply our approach to the daily returns of S&P 500 index. Nonhomogeneous HMM provides smooth estimation of time-depended transition matrix in the analysis of stock return data, which reveals different patterns in bull and bear markets. The generalized likelihood ratio test indicates that the time varying transitions are statistically significant.

This paper proceeds as follows. In section 2, we develop estimation procedure for nonhomogeneous HMM and further study issues of model selection and inference. In section 3, we conduct simulation analysis to demonstrate our methodologies. Application to the daily returns of S&P 500 index is given in section4. Conclusion and discussion are presented in section 5.

2. ESTIMATION AND INFERENCE

2.1 Model definition and estimation

Given a stochastic sequence $\{(Y_t, U_t), t \in \mathbf{N}\}$, suppose the corresponding unobserved latent variable $\{S_t, t \in \mathbf{N}\}$ follows a nonhomogeneous finite state Markov chain with finite state space $\{1, 2, \ldots, S\}$. Throughout this paper, we assume that U_t is univariate and observable. Conditioning on $U_t = u$, the transition probability of $\{S_t, t \in \mathbf{N}\}$ is a $S \times S$ matrix $\Gamma(u)$, where the elements are $\gamma_{jk}(u) = P(S_{t+1} = k|S_t = j, U_t = u)$, for $j, k = 1, 2, \ldots, S$, with constraints $\sum_{k=1}^{S} \gamma_{jk}(u) = 1$ for $j = 1, 2, \ldots, S$. Let the initial distribution of $\{S_t, t \in \mathbf{N}\}$ be $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_S)$ with $\sum_{k=1}^{S} \delta_k = 1$. We assume that the initial probabilities and transition probabilities are all positive.

The observation couple is $\{U_t, Y_t\}$, where U_t is the covariate. Given $S_t = k$ and U_t, we assume that the observation Y_t follows a parametric density model $p_k(y|\theta_k), k = 1, 2, \ldots, S$. As a special case, when the parametric density is normal, $p_k(y|\theta_k) \equiv \phi(\mu_k, \sigma_k^2), k = 1, 2, \ldots, S$, and $\theta_k \equiv (\mu_k, \sigma_k^2)$, where ϕ is a normal density function. Let $\boldsymbol{\theta} = \{\theta_k, k = 1, 2, \ldots, S\}$, and $P(y|\boldsymbol{\theta})$ be a $S \times S$ diagonal matrix with diagonal elements $p_1(y|\theta_1), \ldots, p_S(y|\theta_S)$. We have described a nonhomogeneous HMM with unknown parameters $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$, and transition probability functions $\Gamma(\cdot) = \{\gamma_{jk}(\cdot), j, k = 1, 2, \ldots, S\}$.

A modified EM algorithm

In order to estimate the unknown parameters and unknown functions, we propose a modified EM algorithm, and introduce kernel regression method to deal with estimation of transition probability functions. Let $\{(y_t, u_t), t = 1, 2, \ldots, T\}$ be a finite realization of $\{(Y_t, U_t), t \in \mathbf{N}\}$.

Let $\mathbf{1}'$ be a $S \times 1$ column vector whose elements are all 1, the likelihood function for the observed data is

$$\ell(\boldsymbol{\delta}, \boldsymbol{\theta}, \Gamma(\cdot)) = \boldsymbol{\delta} P(y_1|\boldsymbol{\theta})\Gamma(u_1) \cdots \Gamma(u_{T-1}) P(y_T|\boldsymbol{\theta})\mathbf{1}'. \tag{1}$$

We introduce the latent variables $\boldsymbol{z_t} = (z_{t1}, \ldots, z_{tS})$ associated with S_t, where the element indicator is

$$z_{tk} = \begin{cases} 1 & \text{if } S_t = k, \\ 0 & \text{otherwise.} \end{cases}$$

Then following the derivation in the Appendix, the complete log-likelihood function for $\{(y_t, u_t, \boldsymbol{z}_t), t = 1, \ldots, T\}$ is given by

$$\begin{aligned} \mathcal{L} = & \sum_{t=2}^{T} \sum_{j=1}^{S} \sum_{k=1}^{S} z_{t-1,j} z_{tk} \log(\gamma_{jk}(u_{t-1})) \\ & + \sum_{k=1}^{S} z_{1k} \log(\delta_k) + \sum_{t=1}^{T} \sum_{k=1}^{S} z_{tk} \log(p_k(y_t|\boldsymbol{\theta})), \end{aligned} \tag{2}$$

where δ_k is the initial probability which equals to $P(S_1 = k)$, $\gamma_{jk}(u_{t-1})$ is the transition probability at time $t-1$ which equals to $P(S_t = k|S_{t-1} = j, U_{t-1} = u_{t-1})$ and $p_k(y_t|\boldsymbol{\theta})$ is the parametric density function followed by Y_t given $S_t = k$. For example, in case of normal distribution, when Y_t follows normal density given $S = k$, we have $p_k(y_t|\boldsymbol{\theta}) = p_k(y_t|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-(y_t-\mu_k)^2/2\sigma_k^2}$, where $\theta_k \equiv (\mu_k, \sigma_k^2)$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_S)$.

In the modified EM algorithm, we calculate the expectations of z_{tk} and $z_{t-1,j}z_{tk}$ in the E-step, denoted by r_{tk} and h_{tjk} respectively. Then estimate unknown parameters and functions by maximizing the complete log-likelihood function in the M-step. E-step and M-step are iterated until convergence.

E-step:

Given the entire observed sequence of $\{y_t, u_t\}$ and the estimates of $(\boldsymbol{\delta}, \boldsymbol{\theta}, \Gamma(\cdot))$ from the last iteration (or initial values for the first iteration), the expectation of the complete log-likelihood function is

$$\begin{aligned} E(\mathcal{L}) = & \sum_{t=2}^{T} \sum_{j=1}^{S} \sum_{k=1}^{S} h_{tjk} \log(\gamma_{jk}(u_{t-1})) \\ & + \sum_{k=1}^{S} r_{1k} \log(\delta_k) + \sum_{t=1}^{T} \sum_{k=1}^{S} r_{tk} \log(p_k(y_t|\boldsymbol{\theta})) \\ = & \ell_1(\boldsymbol{\delta}) + \ell_2(\Gamma(\cdot)) + \ell_3(\boldsymbol{\theta}), \end{aligned} \tag{3}$$

where r_{tk} is the conditional probability of being in state k at time t given the entire observed sequence, and h_{tjk} is the conditional probability of being in state j at time $t-1$ and in state k at time t given the entire observed sequence. In practice, r_{tk} and h_{tjk} can be efficiently calculated by the following forward-backward algorithm.

Define the $1 \times S$ vector of forward probabilities $\boldsymbol{\alpha}_t = (\alpha_{t1}, \ldots, \alpha_{tS})$ as

$$\boldsymbol{\alpha}_t = \boldsymbol{\delta} P(y_1|\boldsymbol{\theta})\Gamma(u_1)P(y_2|\boldsymbol{\theta})\cdots\Gamma(u_{t-1})P(y_t|\boldsymbol{\theta}),$$

and define the $1 \times S$ vector of backward probabilities $\boldsymbol{\beta}_t = (\beta_{t1}, \ldots, \beta_{tS})$ as

$$\boldsymbol{\beta}_t^{'} = \Gamma(u_t)P(y_{t+1}|\boldsymbol{\theta})\Gamma(u_{t+1})P(y_{t+2}|\boldsymbol{\theta})\cdots\Gamma(u_{T-1})P(y_T|\boldsymbol{\theta})\mathbf{1}^{'}.$$

Then, we have

$$\begin{aligned} r_{tk} &= \alpha_{tk}\beta_{tk}/\ell(\boldsymbol{\delta}, \boldsymbol{\theta}, \Gamma(\cdot)), \\ h_{tjk} &= \alpha_{t-1,j}\gamma_{jk}(u_{t-1})p_k(y_t|\theta_k)\beta_{tk}/\ell(\boldsymbol{\delta}, \boldsymbol{\theta}, \Gamma(\cdot)). \end{aligned} \tag{4}$$

M-step:

The maximization of $E(\mathcal{L})$ in (3) can be achieved by maximizing $\ell_1(\boldsymbol{\delta})$, $\ell_2(\Gamma(\cdot))$ and $\ell_3(\boldsymbol{\theta})$ respectively. Under the constraint $\sum_{k=1}^S \delta_k = 1$, maximizing the first term $\ell_1(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ gives

$$\hat{\delta}_j = \frac{r_{1j}}{\sum_{k=1}^S r_{1k}}. \tag{5}$$

The second term $\ell_2(\Gamma(\cdot))$ contains nonparametric functions $\gamma_{jk}(\cdot)$, and cannot be estimated directly. We propose estimation based on the ideas of kernel regression and local likelihood method, where the nonparametric functions are estimated in a pointwise manner. Suppose that we want to estimate $\gamma_{jk}(\cdot)$ at u, we first use local constant γ_{jk} to approximate $\gamma_{jk}(u)$. Then we construct local log-likelihood function as

$$L_2(\Gamma) = \sum_{t=2}^T \sum_{j=1}^S \sum_{k=1}^S h_{tjk} \log \gamma_{jk} K_h(u_{t-1} - u),$$

where $\Gamma = \{\gamma_{jk}\}, j, k = 1, 2, \ldots, S$, $K_h = h^{-1}K(\cdot/h)$ is a rescaled kernel of a kernel function $K(\cdot)$ with a bandwidth h. These local constants satisfy the constraint $\sum_{k=1}^S \gamma_{jk} = 1$.

Maximizing $L_2(\Gamma)$ with respect to Γ gives

$$\hat{\gamma}_{jk}(u) = \hat{\gamma}_{jk} = \frac{\sum_{t=2}^T h_{tjk} K_h(u_{t-1} - u)}{\sum_{t=2}^T r_{t-1,j} K_h(u_{t-1} - u)}. \tag{6}$$

Note that the rescaled kernel function will assign more weight to the observations close to u than the observations far away from u. The solution to the local likelihood function $\hat{\gamma}_{jk}$ is a local likelihood estimation for $\gamma_{jk}(\cdot)$ at u. We take u in a set of evenly distributed grid points, and further obtain the whole function by interpolation.

Finally, we maximize the third term $\ell_3(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, which is

$$\hat{\boldsymbol{\theta}} = \arg\max \ell_3(\boldsymbol{\theta}). \tag{7}$$

The maximization is depended on the density models $p_k(y_t|\boldsymbol{\theta})$ in $\ell_3(\boldsymbol{\theta})$, $k = 1, \cdots, S$. In some situations, there exists explicit solutions. For example, when Y_t follows a normal density, and $\theta_k = (\mu_k, \sigma_k^2)$, the explicit solutions for $k = 1, 2, \ldots, S$ are

$$\hat{\mu}_k = \frac{\sum_{t=1}^T r_{tk} \cdot y_t}{\sum_{t=1}^T r_{tk}},$$

and

$$\hat{\sigma}_k^2 = \frac{\sum_{t=1}^T r_{tk} \cdot (y_t - \hat{\mu}_k)^2}{\sum_{t=1}^T r_{tk}}.$$

In case an explicit solution does not exist, the maximization can be obtained by, for example, using Newton's method or gradient descent method, or by using other more advanced numerical methods.

2.2 Model selection and inference

Model selection

In practice, we need to consider two model selection issues, the selection of bandwidth and the selection of the number of hidden states. When the number of hidden states is known, cross validation (CV) method is used to determine the bandwidth. Due to the sequence dependence, we consider the CV approach proposed by Celeux and Durand (2008). In this CV procedure, we randomly delete part of the observations, and then estimate the parameters and compute the likelihood for the incomplete sequence. This process is repeated M times. Let $\mathcal{L}_m^{CV}$ be the likelihood of the incomplete sequence calculated in the mth time, and $\mathcal{L}^{Origin}$ be the likelihood calculated using the whole sequence. Then the bandwidth is selected such that $\sum_{m=1}^M |\mathcal{L}_m^{CV} - \mathcal{L}^{Origin}|$ is minimized. In the simulation, we take $M = 30$.

For a given number of states, the bandwidth is determined by the above CV method. Next we can use the BIC approach to determine the number of hidden states. The BIC has the form

$$-2\mathcal{L} + \log(T) \times DoF,$$

where $\mathcal{L}$ is the maximum log-likelihood, and DoF is the degree of freedom in the model.

To access the degree of freedom of smoothing functions, we consider the approach used by [9]. The degree of freedom for an one-dimensional smoothing function is

$$dfs = \tau_K h^{-1}|\Omega|\{K(0) - \frac{1}{2}\int K^2(t)dt\},$$

where Ω is the support of covariate U, and

$$\tau_K = \frac{K(0) - \frac{1}{2}\int K^2(t)dt}{\int\{K(t) - \frac{1}{2}K * K(t)\}^2 dt}.$$

Hence, the degree of freedom for the proposed model is $DoF = S(S-1) \times dfs + df(\boldsymbol{\theta}) + S - 1$, where $df(\boldsymbol{\theta})$ is the degree of freedom contributed by $\boldsymbol{\theta}$. Then the selected number of states is

$$S = \arg\min(-2\mathcal{L} + \log(T) \times DoF).$$

In practice, the implementation is to determine the optimal bandwidth by CV method under different S and then to determine S by the BIC. In Section 4, we show that this approach works well in the simulation setting.

Bootstrap standard error

To obtain standard errors for the unknown parameters and functions, we propose a parametric bootstrap method. Specifically, we fit the model and obtain estimates $\{\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}}, \hat{\Gamma}(u)\}$. We then take samples from the fitted model of size T, and refit the model again to obtain the bootstrap estimates. This allows us to calculate standard errors of the estimates, and further obtain their confidence intervals.

Testing hypothesis

As we study nonhomogeneous HMM, a question arises whether or not the transition matrix actually depend on the covariate. To test the homogeneous null hypothesis against nonhomogeneous alternatives, we use generalized likelihood ratio(GLR) statistics, which was introduced in [9]. The testing problem is

$$H_0 : \gamma_{jk}(u) = c_{jk} \qquad versus \qquad H_1 : \gamma_{jk}(u) \neq c_{jk}$$

for $j, k = 1, 2, \ldots, S$, where c_{jk} are unknown constants.

Let $\ell^*(H_0)$ and $\ell^*(H_1)$ be log-likelihood functions computed under null and alternative hypothesis respectively. Then a generalized likelihood ratio test statistic is defined as

$$\Pi = 2\{\ell^*(H_1) - \ell^*(H_0)\}.$$

Note that the null model is parametric and its alternative is semiparametric. Instead of investigating the asymptotic distribution of Π, we consider the conditional bootstrap method [4] to construct the null distribution under the following steps:

(a) Estimate $\hat{\boldsymbol{\delta}}_0, \hat{\boldsymbol{\theta}}_0, \hat{\Gamma}_0$ under homogeneous HMM.

(b) Generate bootstrap sample $\{(U_t, Y_t^*), t \in \mathbf{N}\}$ from the HMM model with estimated parameters $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\Gamma}_0$.

(c) For each bootstrap sample, we calculate the generalized likelihood test statistics, and then further obtain its approximate distribution.

We can verify the above conditional bootstrap method by checking whether the asymptotic null distribution is independent of the nuisance parameters Γ_0. The verification is performed via simulation study in section 4, and the result shows that Wilk's phenomenon holds in this setting. Therefore, we can obtain the approximate p-value of the test by calculating the percentile of the test statistics in the conditional bootstrap sample. A small p-value (e.g., less than 0.01) suggests a rejection of the null hypothesis, and hence we are confident that the nonhomogeneous HMM is more appropriate compared to the homogenous HMM.

3. SIMULATION STUDY

To illustrate the efficiency of estimates obtained by the modified EM algorithm, we perform numerical simulations in this section. Here, we specify the emission model $p_k(y|\theta_k)$ to be a normal distribution with mean μ_k and variance σ_k^2, which is denoted by $\mathcal{N}(\mu_k, \sigma_k^2)$. Thus, we have $\theta_k = \{\mu_k, \sigma_k^2\}, k = 1, 2, \ldots, S$.

To assess the performance of the estimated transition probability functions, we consider the square root of the average squared errors as

$$\mathrm{RASE}_{\gamma_{jk}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\{\hat{\gamma}_{jk}(u_t) - \gamma_{jk}(u_t)\}^2},$$

where $\{u_t, t = 1, \ldots, T\}$ are the grid points evenly in the range of the covariate U.

To assess the performance of estimated parameters, we use the root of mean squared error. For the mean μ, it is

$$\mathrm{RMSE}_{\mu} = \sqrt{\frac{1}{S}\sum_{k=1}^{S}\{\hat{\mu}_k - \mu_k\}^2}.$$

Similarly, we define $RMSE_{\sigma^2}$ for the variance σ^2.

In the simulation study, we consider two simulation settings.

Setting 1. A two-state nonhomogeneous HMM with $\boldsymbol{\delta} = (0.4, 0.6), \mu_1 = -1, \mu_2 = 1, \sigma_1^2 = 0.09, \sigma_2^2 = 0.09$,

$$\gamma_{11}(u) = 1/2\cos(3\pi u) + 1/2, \quad \gamma_{21}(u) = 1/2\sin(2\pi u) + 1/2.$$

Setting 2. A three-state nonhomogeneous HMM with $\boldsymbol{\delta} = (0.2, 0.3, 0.5), \mu_1 = -2, \mu_2 = 0, \mu_3 = 2, \sigma_1^2 = 0.09, \sigma_2^2 = 0.09, \sigma_3^2 = 0.09$,

$$\begin{aligned}
\gamma_{11}(u) &= 1/4\cos(3\pi u) + 0.3, \quad \gamma_{12}(u) = 1/4\sin(3\pi u) + 0.3,\\
\gamma_{21}(u) &= 1/4\cos(2\pi u) + 0.3, \quad \gamma_{22}(u) = 1/4\sin(2\pi u) + 0.3,\\
\gamma_{31}(u) &= 1/4\cos(\pi u) + 0.3, \quad \gamma_{32}(u) = 1/4\sin(\pi u) + 0.3.
\end{aligned}$$

We set the sample size $T = 200, 400, 800$, respectively. We generate sample sequence $\{Y_t\}$ with length of T, where covariate U is from a uniform distribution $U[0, 1]$. The Epanechnikov kernel is used for kernel regression. Initial values are obtained by the K-means algorithm. The grid points of kernel functions to be estimated are set to 100 points evenly distributed on $[0, 1]$.

We first examine the BIC approach to determine the number of states S. We run 100 simulations for each sample size respectively. For each simulated sample, nonhomogeneous HMMs are fitted with two, three, four states under the optimal bandwidths calculated from the CV method, and the corresponding BIC scores are calculated. Table 1 displays the frequencies of selected S by the BIC approach. The results show that the BIC approach works well to determine the number of states S in both simulation settings.

Table 1. Frequencies of selected S by the BIC

	Two-state model			Three-state model		
	S=2	S=3	S=4	S=2	S=3	S=4
T=200	96	2	2	2	95	3
T=400	98	2	0	0	97	3
T=800	99	1	0	0	98	2

Next we test the performance of estimation via RMSE and RASE. To save time, we use fixed optimal bandwidth instead of optimal bandwidth for each simulated data. We generate 20 datasets for a given sample size and determine a optimal bandwidth for each dataset using the CV method, then we choose the most frequently selected bandwidth $\hat{h}$ to be the fixed optimal bandwidth. Then we consider three different bandwidths: $\frac{2}{3}\hat{h}$, $\hat{h}$, $1.5\hat{h}$, which correspond to the under-smoothing, appropriate smoothing, and over-smoothing, respectively. For each sample size and each bandwidth, we conduct 500 simulations respectively. Table 2 displays the simulation results for the two-state setting estimated with a two-state model, which contains the mean and standard deviation (in parentheses) of RMSE_μ, RMSE_{σ^2}, $\text{RASE}_{\gamma_{11}}$ and $\text{RASE}_{\gamma_{21}}$. Table 3 displays the simulation results for the three-state setting estimated with a three-state model. The results show that the proposed procedure performs quite well across the wide range of bandwidths. We also compare our method to another nonparametric estimation method for transition matrix, the K-nearest-neighbor method(KNN). We conduct a simulation for the two-state nonhomogeneous HMM based on KNN method. The estimation procedure is similar to our method and the optimal parameter $\hat{k}$ is determined by the BIC. Table 4 displays the results under three different k: $\lfloor\frac{2}{3}\hat{k}\rfloor$, $\hat{k}$, $\lfloor 1.5\hat{k}\rfloor$. It is noted that the average squared errors of estimation for transition matrix based on KNN are larger than our method, which shows that our proposed procedure is better than the KNN method.

Then we test the accuracy of the standard error estimation via the parametric bootstrap method. We generate bootstrap samples from the fitted two-state nonhomogeneous HMM, and refit again to obtain the bootstrap estimates, where standard deviations can be calculated from. Table 5, 6, 7 and 8 summarize the performance of the stan-

Table 2. Mean and standard deviation (in parentheses) of RMSEs and RASEs for the two-state model

T	h	RMSE_μ	RMSE_{σ^2}	$\text{RASE}_{\gamma_{11}}$	$\text{RASE}_{\gamma_{21}}$
200	0.04	0.0270(0.0139)	0.0114(0.0058)	0.1023(0.0271)	0.1039(0.0261)
	0.06	0.0266(0.0142)	0.0116(0.0062)	0.1000(0.0271)	0.0933(0.0254)
	0.09	0.0270(0.0138)	0.0114(0.0057)	0.1393(0.0275)	0.1021(0.0234)
400	0.027	0.0188(0.0096)	0.0081(0.0044)	0.0834(0.0193)	0.0779(0.0195)
	0.04	0.0189(0.0099)	0.0080(0.0042)	0.0742(0.0178)	0.0718(0.0194)
	0.06	0.0182(0.0093)	0.0080(0.0042)	0.0812(0.0191)	0.0789(0.0183)
800	0.02	0.0134(0.0077)	0.0056(0.0029)	0.0675(0.0127)	0.0698(0.0141)
	0.03	0.0133(0.0073)	0.0057(0.0030)	0.0570(0.0135)	0.0594(0.0140)
	0.045	0.0132(0.0070)	0.0057(0.0030)	0.0562(0.0133)	0.0556(0.0138)

Table 3. Mean and standard deviation (in parentheses) of RMSEs and RASEs for the three-state model

T	h	RMSE_μ	RMSE_{σ^2}	$\text{RASE}_{\gamma_{11}}$	$\text{RASE}_{\gamma_{12}}$	$\text{RASE}_{\gamma_{21}}$	$\text{RASE}_{\gamma_{22}}$	$\text{RASE}_{\gamma_{31}}$	$\text{RASE}_{\gamma_{32}}$
200	0.06	0.0431	0.0171	0.1187	0.1295	0.1276	0.1228	0.1372	0.1133
		(0.0167)	(0.0071)	(0.0350)	(0.0348)	(0.0345)	(0.0341)	(0.0360)	(0.0328)
	0.09	0.0440	0.0184	0.1140	0.1176	0.1055	0.1050	0.1071	0.0895
		(0.0180)	(0.0077)	(0.0331)	(0.0323)	(0.0340)	(0.0327)	(0.0353)	(0.0320)
	0.135	0.0426	0.0181	0.1324	0.1246	0.1017	0.1079	0.0943	0.0756
		(0.0168)	(0.0081)	(0.0276)	(0.0278)	(0.0318)	(0.0374)	(0.0356)	(0.0340)
400	0.053	0.0290	0.0124	0.0871	0.0964	0.0919	0.0907	0.0985	0.0847
		(0.0125)	(0.0051)	(0.0230)	(0.0236)	(0.0256)	(0.0258)	(0.0242)	(0.0216)
	0.08	0.0294	0.0123	0.0805	0.0877	0.0807	0.0800	0.0851	0.0723
		(0.0131)	(0.0051)	(0.0212)	(0.0226)	(0.0245)	(0.0267)	(0.0246)	(0.0238)
	0.12	0.0286	0.0128	0.1077	0.1006	0.0799	0.0781	0.0729	0.0576
		(0.0125)	(0.0059)	(0.0210)	(0.0160)	(0.0212)	(0.0266)	(0.0279)	(0.0252)
800	0.04	0.0217	0.0092	0.0689	0.0737	0.0720	0.0736	0.0811	0.0701
		(0.0090)	(0.0040)	(0.0160)	(0.0168)	(0.0162)	(0.0165)	(0.0178)	(0.0162)
	0.06	0.0194	0.0085	0.0624	0.0690	0.0610	0.0594	0.0681	0.0568
		(0.0086)	(0.0035)	(0.0164)	(0.0154)	(0.0167)	(0.0187)	(0.0182)	(0.0164)
	0.09	0.0214	0.0090	0.0737	0.0773	0.0630	0.0594	0.0587	0.0478
		(0.0091)	(0.0038)	(0.0170)	(0.0156)	(0.0174)	(0.0183)	(0.0192)	(0.0168)

Table 4. Mean and standard deviation (in parentheses) of RMSEs and RASEs for the two-state model based on KNN

T	k	$RMSE_\mu$	$RMSE_{\sigma^2}$	$RASE_{\gamma_{11}}$	$RASE_{\gamma_{21}}$
200	19	0.0252(0.0141)	0.0112(0.0058)	0.1254(0.0253)	0.1192(0.0256)
	29	0.0272(0.0138)	0.0126(0.0064)	0.1155(0.0298)	0.1088(0.0233)
	43	0.0260(0.0148)	0.0124(0.0063)	0.1479(0.0293)	0.1140(0.0210)
400	35	0.0189(0.0094)	0.0079(0.0042)	0.0923(0.0188)	0.0895(0.0188)
	53	0.0190(0.0102)	0.0082(0.0043)	0.0843(0.0186)	0.0835(0.0165)
	79	0.0196(0.0106)	0.0082(0.0045)	0.1187(0.0209)	0.0942(0.0156)
800	62	0.0139(0.0072)	0.0057(0.0029)	0.0666(0.0125)	0.0676(0.0125)
	93	0.0134(0.0067)	0.0059(0.0031)	0.0627(0.0126)	0.0639(0.0118)
	139	0.0134(0.0068)	0.0059(0.0031)	0.0879(0.0163)	0.0765(0.0120)

Table 5. Standard errors of nonparametric functional estimates via the parametric bootstrap ($T = 200$ and $h = 0.06$)

		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\gamma_{11}(\cdot)$	SD	0.097	0.102	0.051	0.068	0.110	0.076	0.052	0.102	0.092
	SE	0.096	0.097	0.067	0.076	0.100	0.078	0.068	0.098	0.093
	Std	0.022	0.019	0.018	0.018	0.017	0.020	0.019	0.020	0.019
	Cov.P	0.940	0.956	0.870	0.926	0.918	0.960	0.878	0.940	0.954
$\gamma_{21}(\cdot)$	SD	0.090	0.045	0.044	0.090	0.117	0.095	0.046	0.047	0.096
	SE	0.085	0.051	0.050	0.084	0.102	0.086	0.053	0.053	0.087
	Std	0.023	0.018	0.018	0.017	0.018	0.022	0.020	0.020	0.025
	Cov.P	0.946	0.936	0.946	0.938	0.874	0.944	0.928	0.922	0.930

Table 6. Standard errors of nonparametric functional estimates via the parametric bootstrap ($T = 400$ and $h = 0.04$)

		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\gamma_{11}(\cdot)$	SD	0.073	0.091	0.040	0.061	0.094	0.055	0.035	0.093	0.076
	SE	0.074	0.084	0.046	0.059	0.087	0.060	0.045	0.083	0.077
	Std	0.017	0.016	0.015	0.016	0.016	0.016	0.014	0.014	0.015
	Cov.P	0.970	0.932	0.930	0.940	0.938	0.938	0.890	0.850	0.934
$\gamma_{21}(\cdot)$	SD	0.074	0.035	0.033	0.076	0.096	0.077	0.037	0.033	0.081
	SE	0.072	0.034	0.035	0.073	0.089	0.075	0.037	0.036	0.074
	Std	0.016	0.015	0.015	0.015	0.016	0.018	0.015	0.015	0.017
	Cov.P	0.960	0.945	0.965	0.950	0.935	0.960	0.950	0.955	0.925

Table 7. Standard errors of nonparametric functional estimates via the parametric bootstrap ($T = 800$ and $h = 0.03$)

		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\gamma_{11}(\cdot)$	SD	0.064	0.069	0.027	0.046	0.075	0.044	0.026	0.078	0.060
	SE	0.061	0.070	0.030	0.047	0.071	0.047	0.030	0.067	0.060
	Std	0.013	0.012	0.010	0.012	0.011	0.011	0.009	0.011	0.011
	Cov.P	0.950	0.945	0.925	0.950	0.940	0.940	0.935	0.850	0.955
$\gamma_{21}(\cdot)$	SD	0.061	0.028	0.026	0.063	0.078	0.057	0.025	0.024	0.060
	SE	0.059	0.026	0.025	0.059	0.073	0.058	0.026	0.028	0.059
	Std	0.012	0.010	0.010	0.011	0.012	0.011	0.011	0.010	0.012
	Cov.P	0.945	0.960	0.935	0.955	0.925	0.965	0.965	0.960	0.975

dard errors for the functional estimates of transition probabilities at nine points, the standard errors of the means and variances. The standard deviation of the 500 estimates, denoted by SD, can be viewed as the true standard error and served as benchmark. The sample average and the standard deviation of the 500 estimated standard errors using bootstrap, denoted by SE and Std, are then calculated. The coverage probabilities of the estimated standard errors with the confidence level of 0.95, denoted by Cov.P, are showed to verify the validity of our bootstrap method. As the results demonstrate, the proposed bootstrap procedure works reasonably well since the estimated standard error SE is close to the true standard error SD, the coverage probabilities are close to nominal 0.95.

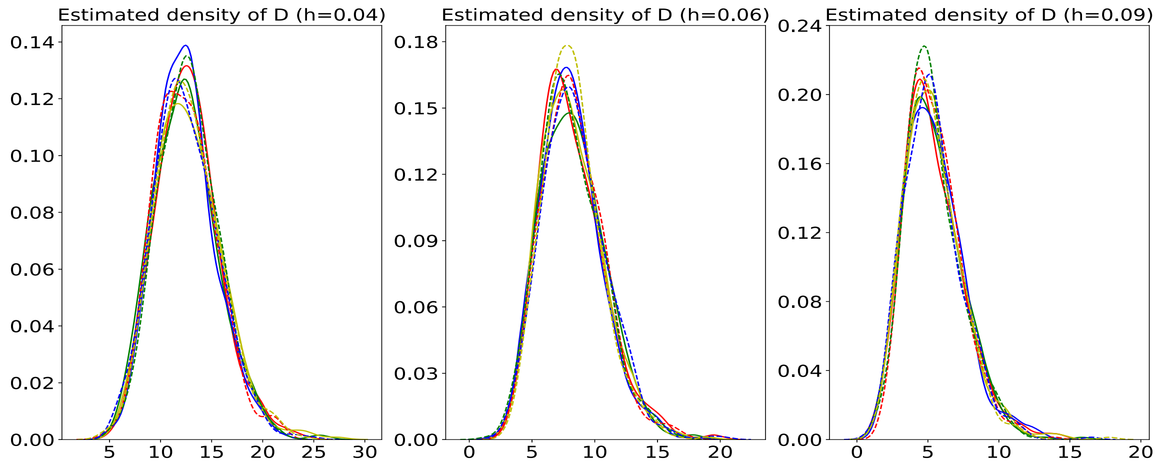

Figure 1. The estimated density of unconditional null distributions of Π (solid lines), and the estimated density of conditional null distributions of Π (dotted lines).

Table 8. Standard errors of estimates via the parametric bootstrap

		SD	SE(Std)	Cov.P
$T = 200, h = 0.06$	μ_1	0.028	0.029(0.005)	0.942
	μ_2	0.032	0.029(0.006)	0.924
	σ_1^2	0.012	0.012(0.003)	0.952
	σ_2^2	0.013	0.012(0.003)	0.956
$T = 400, h = 0.04$	μ_1	0.022	0.021(0.004)	0.945
	μ_2	0.022	0.020(0.004)	0.960
	σ_1^2	0.008	0.009(0.002)	0.895
	σ_2^2	0.010	0.009(0.002)	0.910
$T = 800, h = 0.03$	μ_1	0.014	0.014(0.002)	0.945
	μ_2	0.015	0.014(0.002)	0.935
	σ_1^2	0.006	0.006(0.001)	0.950
	σ_2^2	0.006	0.006(0.001)	0.960

In order to investigate whether the Wilk's type phenomenon holds for the proposed model, we conduct a simulation of testing in the two-state setting. Under the null hypothesis H_0, the transition probabilities γ_{jk} are constants. For 4 sets of different values, $(\gamma_{11}, \gamma_{21}) = \{(\frac{1}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}), (\frac{2}{3}, \frac{1}{3}), (\frac{2}{3}, \frac{2}{3})\}$, we first calculate the unconditional null distribution with $T = 200$ via 500 Monte Carlo simulations. As showed in Figure 1, the resulting 4 densities are quite close which means that the asymptotic distribution of the LRT statistic under the null hypothesis is not sensitive to the true value of Γ. Next, we select 4 typical samples generated from the 4 sets of values of $(\gamma_{11}, \gamma_{21})$ and calculate the conditional null distribution based on its 500 bootstrap samples respectively. As showed in Figure 1, the resulting 4 densities are quite close to the true null distribution which suggests that our conditional bootstrap method works reasonably well to approximate the true null distribution.

Furthermore, the power of the proposed test is also of interest. We evaluate the power functions in the two-state setting under a set of local alternatives with different λ:

$$
\begin{aligned}
H_0 :& \gamma_{11}(u) = c_{11}, \quad \gamma_{21}(u) = c_{21}, \\
H_1 :& \gamma_{11}(u) = 1/2\lambda\cos(3\pi u)/\sqrt{Th} + 1/2, \\
& \gamma_{21}(u) = 1/2\lambda\sin(2\pi u)/\sqrt{Th} + 1/2,
\end{aligned}
$$

and $\gamma_{12}(u) = 1-\gamma_{11}(u)$, $\gamma_{22}(u) = 1-\gamma_{21}(u)$, where $\lambda/\sqrt{Th} \in [0, 1]$. We calculate the power functions under three different significance levels: $0.01, 0.05, 0.1$, based on 500 simulations for sample size $T = 200, 400, 800$. The results in Figure 2 show that the powers increase rapidly as λ increases. When $\lambda = 0$, the alternative collapses into the null hypothesis, and the powers at $\lambda = 0$ are indeed the sizes of the test. As showed in Table 9, the sizes for the three significance levels are close to the nominal level. This shows that the proposed bootstrap procedure approximately provides the right levels of the test.

Table 9. The sizes of the test (The powers at $\lambda = 0$)

	0.01	0.05	0.1
$T = 200, h = 0.06$	0.018	0.063	0.122
$T = 400, h = 0.04$	0.024	0.074	0.126
$T = 800, h = 0.03$	0.020	0.064	0.118

4. APPLICATION TO S&P 500 INDEX DATA

In this section, we illustrate our model by analysing a S&P 500 index dataset. The data contains the daily returns of S&P 500 index from January 1st, 2014 to September 7, 2018. The price sequence during this period is showed in Figure 3. When HMM is employed for analysis, it is expected that the transition probability may be different in a bull and a bear market conditions. Hence, we analyse this

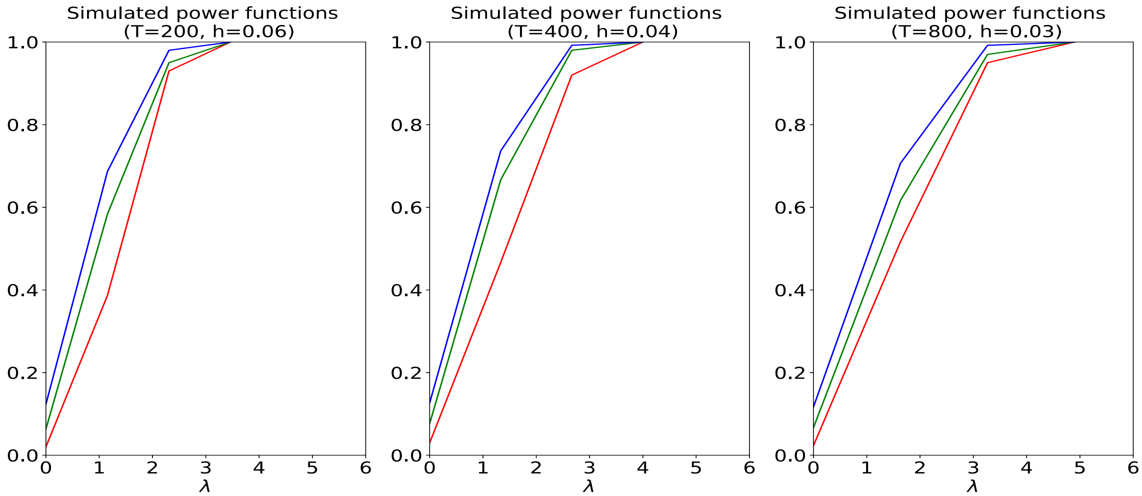

Figure 2. The power functions of the test against local alternatives.

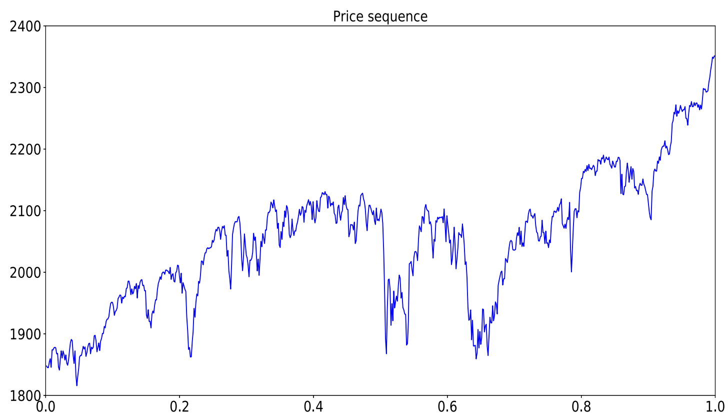

Figure 3. The price sequence.

dataset with the nonhomogeneous HMM. The main purpose is to study how the transition probabilities change over time. Therefore, the covariate is set to be the time. For convenience of analysis, we rescale time to the range from 0 to 1. The grid points of kernel functions to be estimated are set to 100 points evenly distributed on $[0, 1]$.

To determine the number of hidden states and the bandwidth, we first calculate the cross-validated likelihood under 8 different number of states S from 2 to 9, and different bandwidth h in the range of $[0.02, 0.1]$. The optimal bandwidths under different S are then determined. Then we calculate the BIC score under different S with corresponding optimal bandwidth and select the model of $S = 6$ and $h = 0.09$ with the smallest value of BIC score. Thus, we consider a six-state nonhomogeneous HMM and set the optimal bandwidth to be 0.09. We further compare the BIC scores of the selected model and a homogeneous HMM with constant transition probabilities. The BIC score for selected model and a homogeneous HMM is -5212.2 and -5087.1, respectively. This result shows that the return data is more likely to follow a nonhomogeneous HMM than a homogeneous HMM as the former model has a smaller BIC score.

Next, we test whether transition probabilities are functions of time by using the generalized likelihood ratio test. Based on 500 conditional bootstrap simulations, the resulting LRT statistics Π is 77.11 and the approximate p-value of the test is less than 0.01. This means that we can reject the null hypothesis and think that the transition probabilities change all the time, which also reflects that it is more appropriate to use a nonparametric HMM here compared to the homogenous HMM.

Based on the proposed estimation method, we obtain the estimated means of six states, and the corresponding standard errors in Table 10. The six states can be interpreted as strong bear, medium bear, weak bear, weak bull, medium bull, and strong bull patterns of the market. The time-depended estimates of transition probabilities are showed in Figure 4. Our method is able to provide smooth estimates of

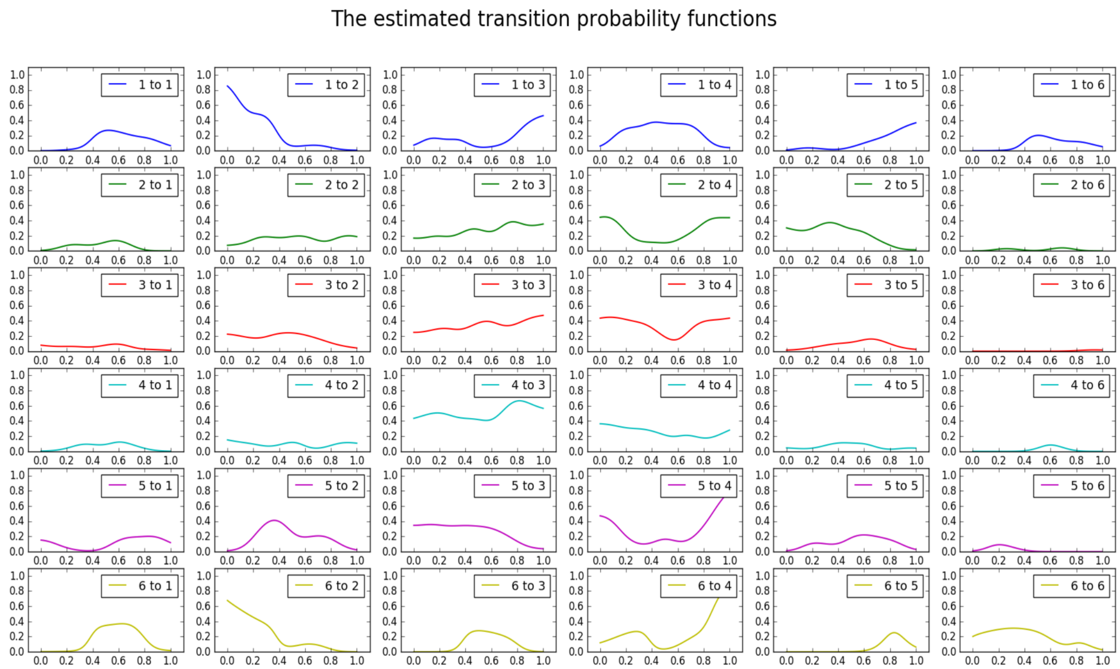

Figure 4. The estimated transition probability functions.

Table 10. Estimated means and standard errors

State	Mean	StD
State 1	-0.017	0.007
State 2	-0.007	0.003
State 3	-0.001	0.002
State 4	0.004	0.002
State 5	0.012	0.003
State 6	0.023	0.007

the transition probability functions. The results show significant different transition probability in bull and bear markets. From the price data showed in Figure 3, we can roughly identify a bull market during time interval [0.65,1] (Jan. 2016 - Feb. 2017), and a bear market during time interval [0.4, 0.65] (Apr. 2015 - Jan. 2016). We observe that in a bull market from Jan. 2016 to Feb. 2017, transition probabilities from state 3, 5, 6 to state 4 are increasing and relatively large. This suggests that bull market tends to rise slowly from medium/large positive movements and small negative movement. However, during the bear market from Apr. 2015 to Jan. 2016, the transition probabilities from state 5, 6 to state 4 are flat and relatively small. The transition probability from state 1 to state 4 is relatively large. This suggests that the market tends to rise slowly after big loss in bear market.

5. CONCLUSION

In this paper, we study nonparametric hidden Markov Model (HMM) with covariate dependent emission probabilities and transition probabilities. The advantage of nonparametric HMM is to allow the impact of covariate included in the whole model no matter what part is influenced. We propose a modified EM algorithm by combining the EM algorithm and the kernel regression method. This approach allows us to define effective number of parameters and further conduct model selection. We propose conditional bootstrap method for standard error, and demonstrate our methods via extensive simulations. A generalized likelihood ratio test procedure is introduced, and simulations show that the Wilk's type of phenomenon holds for the proposed model. The proposed model has also been applied to model the daily returns of S&P500 index. The testing procedure indicates that for the S&P500 dataset, the proposed nonparametric HMM is more appropriate than a parametric HMM statistically.

Asymptotic properties for the estimates are of interest; however, they are out of scope of this paper. To evaluate the accuracy of our proposed bootstrap method, we introduce and calculate coverage probability in simulation. The results show that the bootstrap method performs well in estimation of standard errors, as the coverage probability is close to nominal level.

APPENDIX

Derivation of Equation 2

If the hidden states $S_1, \ldots, S_T$ are observed, the complete log-likelihood function is

$$\mathcal{L} = \log(\delta_{S_1} p_{S_1}(y_t|\theta_{S_1})\gamma_{S_1,S_2}(u_1) \cdots \gamma_{S_{T-1},S_T}(u_{T-1}) \times p_{S_T}(y_T|\theta_{S_T}))$$

$$= \log\left(\delta_{S_1}\prod_{t=2}^{T}\gamma_{S_{t-1},S_t}(u_{t-1})\prod_{t=1}^{T}p_{S_t}(y_t|\theta_{S_t})\right)$$
$$= \log(\delta_{S_1}) + \sum_{t=2}^{T}\log(\gamma_{S_{t-1},S_t}(u_{t-1}))$$
$$+ \sum_{t=1}^{T}\log(p_{S_t}(y_t|\theta_{S_t})).$$

As the hidden states $S_1, \dots, S_T$ are missing, we introduce the latent variables $\boldsymbol{z}_t = (z_{t1}, \dots, z_{tS})$, where the element indicator is

$$z_{tk} = \begin{cases} 1 & \text{if } S_t = k, \\ 0 & \text{otherwise.} \end{cases}$$

Using the indicators, $\log(\delta_{S_1})$ can be rewritten as $\sum_{k=1}^{S} z_{1k}\log(\delta_k)$, and $\log(p_{S_t}(y_t|\theta_{S_t}))$ can be rewritten as $\sum_{k=1}^{S} z_{tk}\log(p_k(y_t|\theta_k))$. For the term of transition probability, indicator multiplication is formed as

$$z_{t-1,j}z_{tk} = \begin{cases} 1 & \text{if } S_{t-1} = j \text{ and } S_t = k, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $\log(\gamma_{S_{t-1},S_t}(u_{t-1}))$ can be rewritten as $\sum_{j=1}^{S}\sum_{k=1}^{S} z_{t-1,j}z_{tk}\log(\gamma_{jk}(u_{t-1}))$. We have completed the derivation of Equation 2.

ACKNOWLEDGEMENTS

The authors thank the editor, the associate editor and reviewers for their constructive comments that have led to a dramatic improvement of the earlier version of this article. Huang's research is supported by the Shanghai Young Talent Development Program, and National Natural Science Foundation of China (NNSFC), grant 11301324.

Received 11 January 2018

REFERENCES

[1] Aerts, M. and Claeskens, G. (1997). Local polynomial estimation in multiparameter likelihood models. *Journal of the American Statistical Association* **92** 1536–1545. MR1615263

[2] Ailliot, P., Bessac, J., Monbet, V. and Pene, F. (2015). Non-homogeneous hidden Markov-switching models for wind time series. *Journal of Statistical Planning and Inference* **160** 75–88. MR3315635

[3] Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics* **12** 309–316.

[4] Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* **95** 888–902. MR1804446

[5] Celeux, G. and Durand, J.-B. (2008). Selecting hidden Markov model state number with cross-validated likelihood. *Computational Statistics* **23** 541–564. MR2448181

[6] Choi, K. and Hammoudeh, S. (2010). Volatility behavior of oil, industrial commodity and stock markets in a regime-switching environment. *Energy Policy* **38** 4388–4399.

[7] Diebold, F. X., Lee, J.-H. and Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. *Business Cycles: Durations, Dynamics, and Forecasting* 144–165.

[8] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications* **66**. Chapman & Hall/CRC. MR1383587

[9] Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of statistics* 153–193. MR1833962

[10] Garcia, R., Perron, P. et al. (1991). *An analysis of the real interest rate under regime shifts*. Universite de Montreal, Departement de sciences economiques.

[11] Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42** 27–62.

[12] 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. MR0996941

[13] Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal* **5** 41–53. MR1988438

[14] Huang, M., Li, R. and Wang, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association* **108** 929–941. MR3174674

[15] Huang, X., Baker, J. and Reddy, R. (2014). A historical perspective of speech recognition. *Communications of the ACM* **57** 94–103.

[16] Kaufmann, S. (2015). K-state switching models with time-varying transition distributions–Does loan growth signal stronger effects of variables on inflation? *Journal of Econometrics* **187** 82–94. MR3347296

[17] Kim, C.-J., Piger, J. and Startz, R. (2008). Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics* **143** 263–273. MR2423067

[18] Maruotti, A. and Rocci, R. (2012). A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Statistics in medicine* **31** 871–886. MR2913866

[19] Mehrotra, R. and Sharma, A. (2005). A nonparametric non-homogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. *Journal of Geophysical Research: Atmospheres* **110**.

[20] Meligkotsidou, L. and Dellaportas, P. (2011). Forecasting with non-homogeneous hidden Markov models. *Statistics and Computing* **21** 439–449. MR2806620

[21] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications* **9** 141–142.

[22] Peria, M. S. M. (2002). A regime-switching approach to the study of speculative attacks: a focus on EMS crises. In *Advances in Markov-Switching Models* 159–194. Springer.

[23] Schaller, H. and Norden, S. V. (1997). Regime switching in stock market returns. *Applied Financial Economics* **7** 177–191. MR2478835

[24] Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association* **82** 559–567. MR0898359

[25] Titman, A. C. (2011). Flexible nonhomogeneous Markov models for panel observed data. *Biometrics* **67** 780–787. MR2829132

[26] Watson, G. S. (1964). Smooth regression analysis. *Sankhy: The Indian Journal of Statistics, Series A* **26** 359–372. MR0185765

Mian Huang
Shanghai University of Finance and Economics
No.777 Guoding Road
Yangpu District
Shanghai, 200433
China
E-mail address: huang.mian@mail.shufe.edu.cn

Yue Huang
Shanghai University of Finance and Economics
No.777 Guoding Road
Yangpu District
Shanghai, 200433
China
E-mail address: huangyue0711@gmail.com

Kang He
Shanghai University of Finance and Economics
No.777 Guoding Road
Yangpu District
Shanghai, 200433
China
E-mail address: hekang30@aliyun.com