

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/103285/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Liu, Zhenya and Wang, Shixuan 2017. Decoding Chinese stock market returns: three-state hidden semi-Markov model. *Pacific-Basin Finance Journal* 44 , pp. 127-149. 10.1016/j.pacfin.2017.06.007

Publishers page: <http://dx.doi.org/10.1016/j.pacfin.2017.06.007>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Accepted Manuscript

Decoding Chinese stock market returns: Three-state hidden semi-Markov model

Zhenya Liu, Shixuan Wang

PII: S0927-538X(16)30149-4
DOI: doi:[10.1016/j.pacfin.2017.06.007](https://doi.org/10.1016/j.pacfin.2017.06.007)
Reference: PACFIN 943

To appear in: *Pacific-Basin Finance Journal*

Received date: 16 September 2016
Revised date: 21 March 2017
Accepted date: 20 June 2017



Please cite this article as: Liu, Zhenya, Wang, Shixuan, Decoding Chinese stock market returns: Three-state hidden semi-Markov model, *Pacific-Basin Finance Journal* (2017), doi:[10.1016/j.pacfin.2017.06.007](https://doi.org/10.1016/j.pacfin.2017.06.007)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Decoding Chinese Stock Market Returns: Three-State Hidden Semi-Markov Model [☆]

Zhenya Liu^{a,b,*}, Shixuan Wang^b

^a*Renmin University of China, Beijing, 100872, P.R. China*

^b*University of Birmingham, Birmingham, B15 2TT, UK*

Abstract

In this paper, we employ a three-state hidden semi-Markov model (HSMM) to explain the time-varying distribution of the Chinese stock market returns since 2005. Our results indicate that the time-varying distribution depends on the hidden states, which are represented by three market conditions, namely the bear, sidewalk, and bull markets. We find that the inflation, the PMI, and the exchange rate are significantly related to the market conditions in China. A simple trading strategy based on expanding window decoding shows profitability with a Sharpe ratio of 1.14.

Keywords: Chinese Stock Market, Asset Return, Hidden Semi-Markov Model

JEL: C22, G12, G15

[☆]The research of Shixuan Wang was supported by the Economic and Social Research Council (UK) [grant number ES/J50001X/1] and a Royal Economics Society Junior Fellowship.

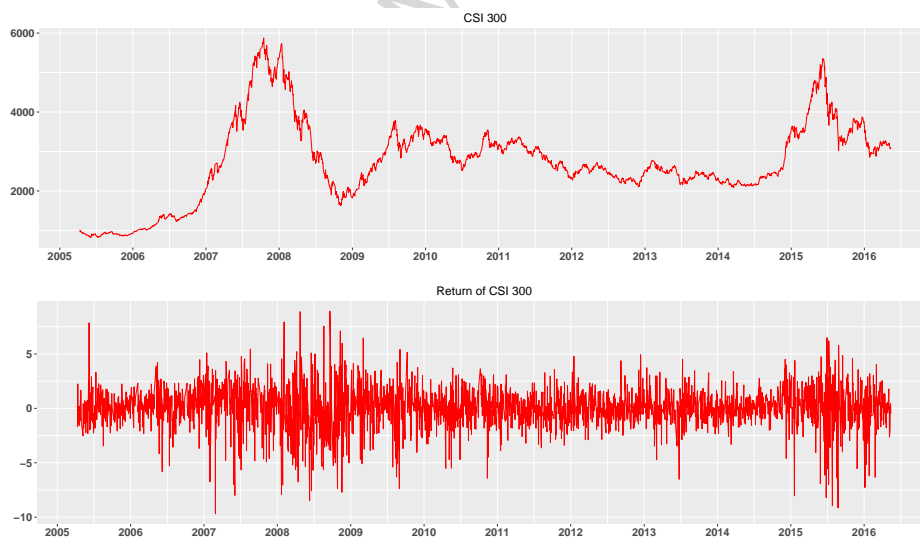
*Corresponding author

Email addresses: zhenya_liu@hotmail.com (Zhenya Liu), shixuan_wang@hotmail.com (Shixuan Wang)

1. Introduction

In this paper, the research question concerns what hidden states drive the time-varying distribution of the Chinese stock market returns. The literature on the Chinese stock market focuses on financial integration, speculative trading, government interventions, information asymmetry, and the relation with bank credit (e.g. Girardin & Liu, 2007; Mei et al., 2009; Los & Yu, 2008; Chan et al., 2008; Girardin & Liu, 2005). Less attention has been paid to the time-varying features of the Chinese stock market after 2005. We have observed that the Chinese stock market behaves quite differently across different periods since 2005. Between 2005 and 2009, the Chinese stock market index (CSI 300) increased approximately six times from 1003 (April 8th 2006) to 5877 (October 16th 2007), and then dropped to 1627.759 (April 11th 2008). Between 2010 and 2014, the CSI 300 had much less volatility and fluctuated between 2000 and 3500. From 2015 onwards, the market became highly volatile again (see Figure 1).

Figure 1: CSI 300 and its Returns



We employ a three-state hidden semi-Markov model (HSMM) to explain the time-varying distribution of the Chinese stock market returns. Based on the estimation by the expectation-maximization (EM) algorithm, the hidden states behind the return data are represented by the three market conditions, namely the bear, sidewalk, and bull markets. The underlying sequence of hidden states is globally decoded by the Viterbi algorithm. The evolution of the market conditions of the Chinese stock market over the last decade is then reviewed. Using Monte Carlo simulations, our three-state HSMM is compared with a stochastic volatility (SV) model

and a tGARCH(1,1) model with respect to three stylized facts, the fat tails, the “long-memory” and the Taylor effect. All three models can reproduce stylized facts of the “long-memory” and the Taylor effect, but tGARCH(1,1) fails to reduce the fat tails. Additionally, the information criteria indicate that our three-state HSMM provides a better performance than the two-state HSMM, the three-state hidden Markov model (HMM), and the two-state HMM.

We start our analysis by investigating the stylized facts of the CSI 300 returns. These stylized facts of asset returns in the developed markets are well documented in the literature (Granger & Ding, 1995; Pagan, 1996; Cont, 2001). They can be classified into two categories, namely distributional properties and temporal properties. Distributional properties relate to the non-Gaussianity of the distribution of asset returns, whilst temporal properties refer to the time dependence of asset returns and of the squared/absolute asset returns.

In the early studies exploring distributional properties, normal distributions with stationary parameters were often selected in order to model daily asset returns. However, Mandelbrot (1997) doubted the Gaussian hypothesis of asset returns and stated that stable Paretian distributions with characteristic exponents of less than 2 are better suited to fit the empirical distribution of assets (Mandelbrot's hypothesis). Fama (1965) undertook extensive testing on empirical data and found that extreme tail values are more frequent than the Gaussian hypothesis (a.k.a. leptokurtosis), which supports the Mandelbrot's hypothesis. In order to explain the notion of leptokurtosis, Fama tried two modified versions of the Gaussian model: a Gaussian mixture model and a non-stationary Gaussian model. However, his empirical evidence supports neither of them. Praetz (1972) and Blattberg & Gonedes (1974) employed t-distributions with small degrees of freedom in order to capture the fat-tail of the empirical distribution of asset returns. Granger & Ding (1995) suggested that the appropriate distribution is the double exponential distribution with zero mean and unit variance. Mittnik & Rachev (1993) inspected various stable distributions for asset returns and found that the Weibull distribution gave the best fit for the S&P 500 daily returns between 1982 and 1986.

In terms of temporal properties, the ARCH-family models are often used for volatility clustering. The original ARCH model was introduced by Engle (1982) in order to model non-constant variances. Bollerslev (1986) generalised the ARCH model by allowing past conditional variances to affect current conditional variances. Afterwards, variants of the GARCH were developed, including EGARCH, GJR, GARCH-M, and so forth. Bollerslev et al. (1992) comprehensively reviewed many types of GARCH models. As for the continuous-time set-up, stochastic volatility models were introduced by Taylor (1986) in an attempt to overcome the main drawback of the

Black-Scholes model characterised by a constant volatility. Stochastic volatility models facilitate analysis of a variety of option pricing problems. A review of the stochastic volatility models was conducted by Jäckel (2004).

The HMM is suitable to capture both distributional and temporal properties. The state process of the model evolves as a Markov chain, providing the channel of time dependency. Its distribution is a mixture of several distributions, enabling it to explain the fat tails. Rydén et al. (1998) adopted an HMM with component distributions as normal distributions (zero mean but different variance) in order to reproduce most of the stylized facts of the daily returns. However, the HMM fails to reproduce the slow decay in the autocorrelation function (ACF) of the squared returns. For the Chinese stock market, Girardin & Liu (2003) use a switch-in-the-mean-and-variance model (MSMH(3)-AR(5)) in order to examine the market conditions on the Shanghai A-share market from 1994 to 2002. They found three regimes: a speculative market, a bull market and a bear market.

There are two ways to improve the HMM. The first way is to change the component distribution into other types of distribution. Rogers & Zhang (2011) proposed a two-state HMM with non-Gaussian component distributions. They examined various component distributions. By using the Kolmogorov-Smirnov test, the symmetric hyperbolic distribution is found to be the most appropriate component distribution. With the inclusion of a regularisation term, they can reproduce the slow decay of the ACF in the absolute returns. Their model setting mainly focused on statistical properties and lacked meaning for the field of economics. The second way is to generalise the sojourn time distribution of the HMM. Bulla & Bulla (2006) modelled daily returns with the HSMM, which is a generalisation of the HMM by explicitly specifying the sojourn time distribution. They utilised both normal distributions and Student's t -distributions as the component distributions. The stylized facts of the daily returns were entirely reproduced by the HSMM. Their research focused on analysing the variances but ignored the means of the component distributions. We believe that the means of the component distributions are also worth investigating because they lead to different market conditions.

In this paper, we provide an interpretation of the hidden states, facilitating the economic explanations of the HSMM model. Additionally, we provide two simple applications of our three-state HSMM. In the first application, the three-state HSMM is used as a market condition classifier, and then an ordinal logit model is employed in order to analyse the links between the decoded market conditions and the macroeconomic variables. We find that the inflation, the PMI, and the exchange rate are significantly related to the market conditions in China. In

the second application, we design a simple trading strategy based on the expanding window decoding. The trading strategy shows profitability with a Sharpe ratio of 1.14.

The remainder of the paper is structured as follows. Section 2 describes our data and its descriptive statistics. Section 3 analyses the distributional and temporal properties of the Chinese stock market returns. Section 4 briefly introduces the HSMM, estimation method, decoding techniques and our model set-up. In Section 5, the estimation results and the decoding results are presented and their economic meanings are discussed, followed by the model evaluation and comparison in Section 6. Section 7 analyses the links between the decoded market conditions and the macroeconomic variables. Section 8 presents a simple trading strategy with a Sharpe ratio of 1.14. Section 9 summarises the paper.

2. Data

2.1. Data Information

The raw data is the closing price of the CSI 300¹, which is a free-floating weighted stock market index of 300 A-share stocks listed on both the Shanghai Stock Exchange and the Shenzhen Stock Exchange. The sample period is from April 8th 2005 (the launch date of the CSI 300) to May 13th 2016. The number of observations accounts for 2697 in total. Our data was downloaded from Wind. The daily return is defined as 100 times the first-order difference of the natural logarithm of the price series.

$$r_t = 100 \times (\log(P_t) - \log(P_{t-1})) \quad (1)$$

where P_t is the closing price of the CSI 300.

2.2. Descriptive Statistics

Table 1 presents the four moments of the daily returns of the CSI 300. The mean is roughly 0.042 and the standard deviation is 1.898. The third moment, skewness, shows that the daily return is negatively skewed. The fourth moment, kurtosis, is larger than the double of the normality benchmark. This implies that the daily returns of the CSI 300 have the leptokurtosis

¹We employ the CSI 300 because it consists of stocks in both the Shanghai Stock Exchange and the Shenzhen Stock Exchange, which is the most representative index market of the A Share. Other Chinese stock market indexes (Shanghai Composite Index, SSE 50, and CSI 500) lead to similar results. Additional results are available upon request.

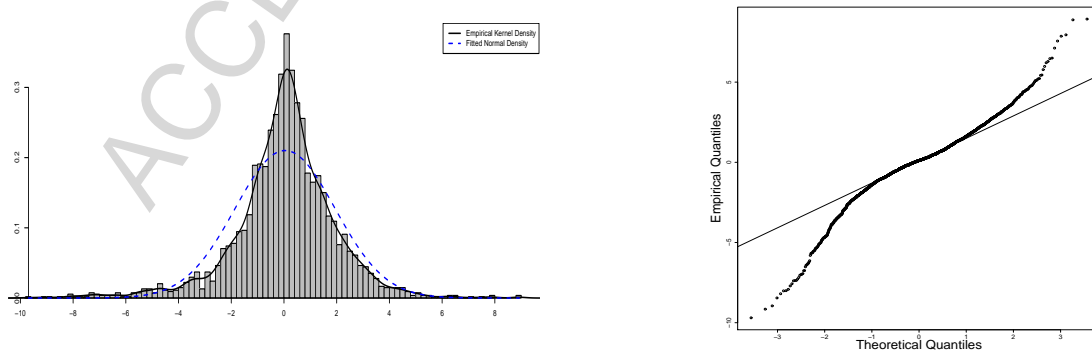
and the fat tails. The third and fourth moments indicate that the distribution of the daily returns deviates from the normal distribution. The non-Gaussianity can be confirmed by the Kolmogorov-Smirnov test with statistics of 0.953 and a P-value of approximately zero.

Table 1: Descriptive Statistics

Mean	0.042
Standard deviation	1.898
Skewness	-0.516
Kurtosis	6.089

In order to inspect non-Gaussianity, we fit a normal distribution to the empirical distribution and compare it to the empirical kernel density in Figure 2a. As it may be observed, the empirical kernel density has the leptokurtosis in the middle and the fat tails at the two sides. The empirical density is highly inconsistent with the fitted normal density. For the purpose of visualising the magnitude of the fat tails, Figure 2b shows the QQ plot of the empirical distribution to a theoretical normal distribution. While the empirical quantiles fit the normal quantiles in the middle part, they diverge at the two tails. The QQ plot confirms the heavy tail of the daily returns of the CSI 300.

Figure 2: Non-Gaussianity of the return distribution



(a) Empirical Kernel Density vs. Fitted Normal Density

(b) QQ Plot

3. Distributional and Temporal Properties

3.1. Distributional Properties

In order to study the distributional properties, we fit various parametric distributions to our empirical data, the daily returns of the CSI 300. Most parametric distributional types studied in the literature are considered here. Four evaluation tools (log likelihood, AIC, BIC, and the Kolmogorov-Smirnov test) are reported in Table 2 for all the fitted parametric distributions.

The normal distribution has the lowest log likelihood, and the highest AIC and BIC, which confirmed the non-Gaussianity shown in Section 2.2. A Student's t -distribution with a degree of freedom 2.145 gives a better fitting than a normal distribution as it can capture the fat-tail to some extent. However, the t -distribution is also rejected by the Kolmogorov-Smirnov Test. The double Weibull, which gave a good fit for S&P 500 (Mittnik & Rachev, 1993), is inferior to the t -distribution for the Chinese stock index returns. A double exponential distribution seems to be the best fitted distribution within the non-mixture distribution category. The Kolmogorov-Smirnov test cannot reject a double exponential distribution with a P-value of 79.83%. A symmetric hyperbolic distribution is rejected by the Kolmogorov-Smirnov test at the 5% significance level.

If mixture distributions are considered, the Gaussian mixture distribution with two components (Gaussian mixture (2)) is better than the double exponential distribution with a higher log likelihood, lower AIC and BIC, and a Kolmogorov-Smirnov test P-value of 86.20%. With an additional component, a Gaussian mixture distribution with three components (Gaussian mixture (3)) produces a higher log likelihood. It may be argued that the increase in likelihood comes from over-fitting by introducing more parameters. However, the AIC and BIC of Gaussian mixture (3) are lower than those of Gaussian mixture (2). Since the AIC and BIC penalise the additional number of parameters, this suggests that Gaussian mixture (3) is superior to Gaussian mixture (2) for Chinese stock index returns. Furthermore, a Kolmogorov-Smirnov test cannot reject Gaussian mixture (3) at the 5% level.

The study of the fitting of various parametric distributions suggests that Gaussian mixture (3) is a good candidate to capture the distributional properties of Chinese stock index returns, which provides an intuitive foundation for using the three-state HSMM in this paper.

Table 2: Various Parametric Distribution Fittings

Type of Distribution	Fitted Parameters	Log Likelihood	AIC	BIC	Kolmogorov-Smirnov Test (P-value)
Normal	mean	-5554.054	11112.110	11123.910	0.00%
	sd	1.897			
Student's t	df	-5396.709	10795.420	10801.320	0.00%
	2.145				
Double Weibull	alpha	-5446.249	10894.500	10900.400	0.00%
	-0.911				
Double Exponential	location	-5339.219	10682.440	10694.240	79.83%
	scale	1.332			
Symmetric Hyperbolic	location	-5366.661	10737.320	10749.120	4.70%
	scale	1.802			
Gaussian Mixture (2)	Component 1	-5350.699	10713.400	10748.800	86.20%
	mean	sd	weight		
	-0.232	2.861	33.86%		
	Component 2				
	mean	sd	weight		
	0.182	1.093	66.14%		
Gaussian Mixture (3)	Component 1	-5328.992	10675.980	10729.080	12.20%
	mean	sd	weight		
	-0.704	3.194	20.01%		
	Component 2				
	mean	sd	weight		
	0.029	0.610	23.90%		
	Component 3				
	mean	sd	weight		
	0.313	1.532	56.08%		

Note: We use a one-parameter double Weibull distribution $F(x) = \exp(-(-x)^{-\alpha})\mathbf{1}_{x \leq 0} + \mathbf{1}_{x > 0}$ the same as Cont (2001).

3.2. Temporal Properties

3.2.1. “Long-memory”

As can be seen in Figure 3a, the autocorrelation functions are insignificant ² for most lags with a small number of exceptions. Thus, daily returns are uncorrelated. Figure 3b and Figure 3c show that the autocorrelation functions of both squared returns and absolute returns are significant for all lags and decay slowly. This slowly decaying autocorrelation is referred to as the “long-memory” in the literature. Both squared returns and absolute returns are two types of volatility measure. The reason behind the “long-memory” could be volatility clustering, which results from the fact that the volatility of the past returns will affect the volatility of future returns for a considerably long period of time.

The temporal property of “long-memory” implies that there is some time dependence for the squared/absolute returns. This time dependence is very persistent for the volatility of returns. The GARCH-family models and the stochastic volatility models are usually used to capture volatility clustering. A Markov chain or semi-Markov chain is also capable of modelling volatility clustering in a discrete way. The advantage of a Markov chain or semi-Markov chain is that they can be associated with various distributions. Hence, the study of temporal properties gives us another incentive to use our three-state HSMM.

3.2.2. Taylor Effect

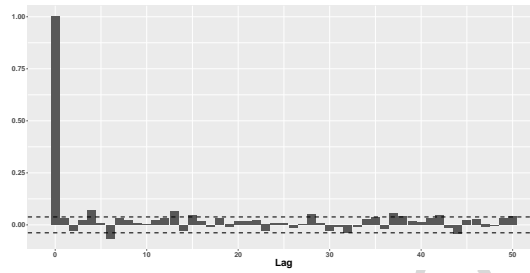
Taylor (1986) found that the autocorrelations of the power of absolute returns are the highest when the power coefficient is one. In a mathematical definition, this is represented as:

$$\text{corr}(|r_t|, |r_{t+k}|) > \text{corr}(|r_t|^\theta, |r_{t+k}|^\theta) \quad \text{for any } \theta \neq 1 \quad (2)$$

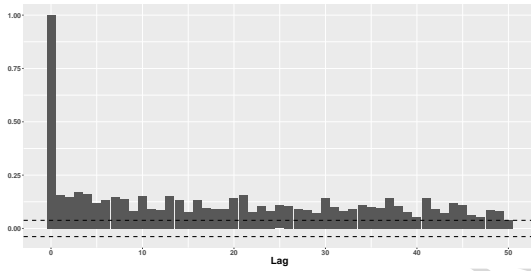
Figure 4 illustrates the Taylor effect for the daily returns of the CSI 300. One horizontal dimension is the lag number and the other is the power coefficient θ . The vertical dimension is the autocorrelation function values. The surface has the highest value in the middle where $\theta = 1$ for all lags. The surface is declining when θ deviates from 1 and reaches its lowest values at the sides of the space. $|r_t|^\theta$ is a volatility measure with different scales. The Taylor effect implies that the volatility measured by $|r_t|$ has the strongest time dependence.

²The 95% confidence band for the autocorrelation function is calculated by $\pm 1.96/\sqrt{N}$, where N is the sample size.

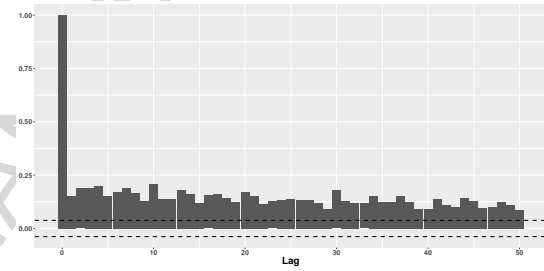
Figure 3: ACF of Original Returns, Squared Returns, and Absolute Returns



(a) ACF of Original Returns

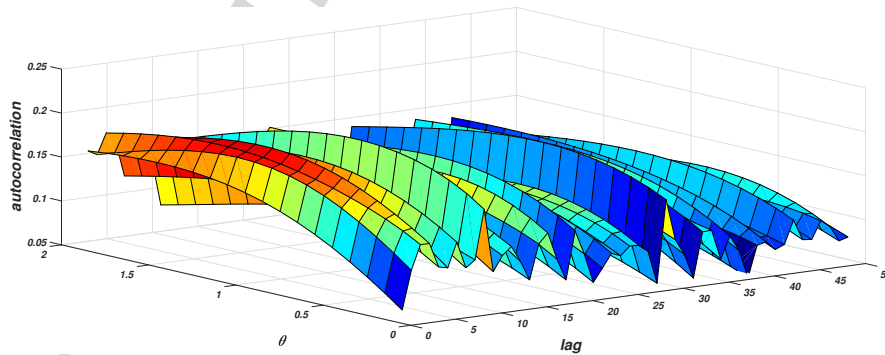


(b) ACF of Squared Returns



(c) ACF of Absolute Returns

Figure 4: Taylor Effect



In summary, the combination of a Gaussian mixture model and the “long-memory” time dependence can correspond to HMM/HSMM where the distributional dimension is a Gaussian mixture distribution and the time dependence is captured by a hidden Markov chain/hidden semi-Markov chain. The distributional properties and the temporal properties of our empirical data provide the intuitive support to adopt the HMM/HSMM for Chinese stock index returns. The HSMM is finally chosen because it is a generalisation of the HMM and Rydén et al. (1998) found that the HMM could not reproduce the stylized fact of the “long-memory”.

4. Methodology

4.1. Hidden Semi-Markov Model

One limitation of the HMM is that its sojourn time³ has to follow a geometric distribution (Yu, 2010; Bulla & Bulla, 2006). The HSMM generalises the HMM by allowing the sojourn time distribution to follow other distributions. In other words, the sojourn time d of a given state is explicitly specified for the HSMM. Similarly to the HMM, the HSMM also entails two processes, an unobservable state process $\mathbf{S}_1^T = \mathbf{s}_1^T$ and an observation process $\mathbf{X}_1^T = \mathbf{x}_1^T$, where \mathbf{s}_1^T is the notation for the realised states s_1, s_2, \dots, s_T and \mathbf{x}_1^T is the notation for observations x_1, x_2, \dots, x_T . The hidden state process \mathbf{S}_1^T is an unobservable semi-Markov chain with m states. The observation process \mathbf{X}_1^T is associated with the hidden state process through component distributions⁴. Equation 3 shows the component distribution for state i at time t .

$$\mathbb{P}_i(x_t) = \mathbb{P}(x_t | s_t = i) \quad \text{where } i \in \{1, 2, \dots, m\} \quad (3)$$

Equation 4 defines the state transition probability from state i to state j .

$$\gamma_{ij} = \mathbb{P}(s_{t+1} = j | s_t = i) \quad \text{where } i \neq j, \quad i, j \in \{1, 2, \dots, m\} \quad (4)$$

Unlike the HMM, the transition probability from one state to the same state in the HSMM is zero, i.e. $\gamma_{ij} = 0$. The sojourn time in the HSMM is controlled by the sojourn time distribution defined in Equation 5.

$$d_i(u) = \mathbb{P}(s_{t+u+1} \neq i, s_{t+1}^{t+u} = i | s_{t+1} = i, s_t \neq i) \quad (5)$$

where the variable u is the length of the sojourn time which can follow nonparametric or parametric distributions. The sojourn time distribution for each state i can follow different types of distribution or the same type of distribution but with different values of the parameters.

The transition probability matrix (TPM) has entries for the transition probabilities γ_{ij} at

³The sojourn time is also known as the dwell time, occupancy time, or duration time.

⁴The component distribution is also known as emission distribution, conditional distribution, or marginal distribution.

row i and column j . The diagonal elements in the TPM of the HSMM are zeros.

$$\mathbf{\Gamma} = \begin{pmatrix} 0 & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & 0 & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & 0 \end{pmatrix} \quad (6)$$

We estimate the model using the EM algorithm (see Appendix A). The most likely sequence of the states is globally decoded by the Viterbi algorithm (see Appendix B.1). Additionally, we use local decoding to compute the conditional probabilities for each state at time t given the observation (see Appendix B.2). Our implementation is based on the R package ‘hsmm’ (Bulla & Bulla, 2013).

4.2. Model Setting

The number of states in our HSMM is set to three. Section 3.1 shows that Gaussian mixture (3) is a better fit than Gaussian mixture (2) based on the likelihood and the information criteria. The normal distribution is chosen to be the component distribution for our HSMM. Other distributions could be considered, but the empirical results of the Chinese stock index returns show that the HSMM with normal components is sufficient to explain our data. Moreover, it is convenient to conduct various tests on a normal distribution, like the z-test in order to examine the significance of the mean. Additionally, the normal component distribution enables us to give a straightforward interpretation for the HSMM.

As for the sojourn time distribution, the logarithmic distribution is selected because it can produce stable estimation results while the EM algorithm may not converge under many other sojourn time distributions. The logarithmic distribution has only one parameter and can avoid overfitting by introducing more parameters. The negative Binomial distribution used by Bulla & Bulla (2006) is also a suitable candidate but it produces similar results as the logarithmic distribution. Other sojourn time distributions could be used, but the logarithmic distribution is sufficient for our data.

5. Empirical Results

5.1. Estimation Results

Through the EM algorithm, the parameters of the HSMM are estimated, including the parameters of the component distributions, transition probability matrix, and sojourn time distri-

bution. Table 3 presents the estimated parameters of the component distributions. Based on the estimated mean and standard deviation, it is able to compute one-sample z-statistics in order to test the significance of the mean. The formula to compute the z-statistics is as follows.

$$z_i = \frac{\bar{x}_i}{\sigma_i/\sqrt{n_i}} \quad \text{for } i \in \{1, 2, 3\} \quad (7)$$

where \bar{x}_i is the mean of state i , σ_i is the standard deviation of state i , and n_i is the sample size of state i . The one-sample z-test suggests that the mean of state 1 is significantly below 0 at the 1% significant level; the mean of state 3 is significantly above 0 at the 1% significant level; whilst the mean of state 2 is insignificant from 0.

Table 3: Component Distribution Parameters

	State 1	State 2	State 3
Mean	-0.510	-0.020	0.622
Std. Dev.	3.113	1.156	1.440
Sample Size	572	1430	695
z-statistics	3.918***	-0.654	11.387***

Note: *p<0.1; **p<0.05; ***p<0.01

The results indicate that the time-varying distribution of the returns depends on the hidden states, which can be interpreted as the market conditions. Specifically, state 1 corresponds to the bear market, state 2 corresponds to the sidewalk market, and state 3 corresponds to the bull market. We define the bear, sidewalk, and bull markets from the perspective of the distributional features.

Definition. A Bear Market

- The mean of the distribution of the daily returns conditional on a bull market is significantly less than 0.
- The frequency of the positive returns is expected to be larger than that of the negative returns.
- Because of the above statistical properties, the price in a bear market is generally decreasing.

Definition. A Sidewalk Market

- The mean of the distribution of the daily returns conditional on a sidewalk market should be insignificantly different from 0.

- It is expected to observe a roughly equal number of positive and negative returns.
- Because of the above statistical properties, the price in a sidewalk market stays in a band and shows a mean-reversion pattern.

Definition. A Bull Market

- The mean of the distribution of the daily returns conditional on a bull market should be significantly larger than 0.
- The frequency of the positive returns is expected to be larger than that of the negative returns.
- Because of the above statistical properties, the price in a bull market is generally increasing.

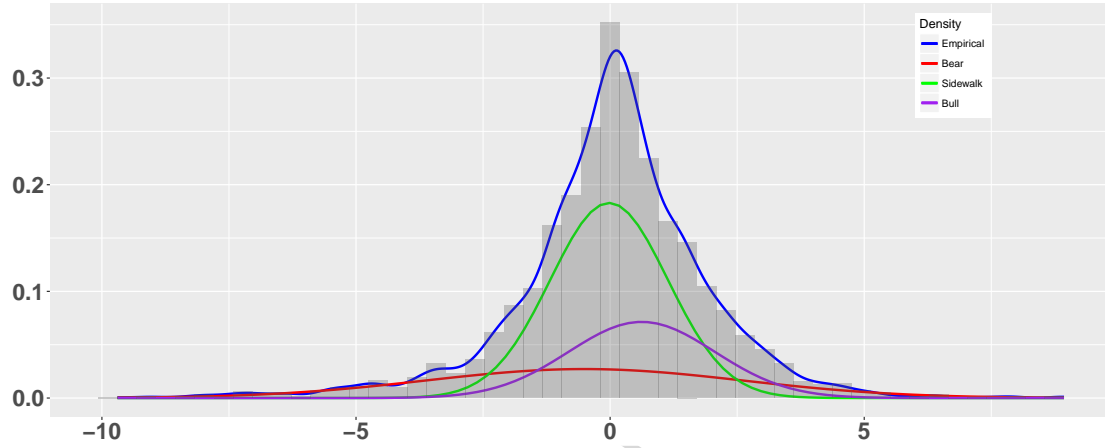
Table 4 presents the empirical frequency of the positive and negative returns for the fitted component distributions and confirms our interpretation of the three underlying states of the HSMM. As can be seen, the frequency of the positive return of state 3 is 67.19%, while the frequency of the negative return is 32.81%. There are negative returns in the bull market as well, but positive returns are more frequent. This statistical evidence empowers the price in the bull market to increase. Hence, state 3 can be regarded as a bull market according to its statistical features. Using the same logic, state 1 has a significant negative mean and corresponds to the bear market where the price shows a downward trend because the negative returns (52.97%) occur more often than the positive returns (47.03%). As for state 2, the frequency of the positive and negative returns is nearly the same at around 50%. State 2 corresponds to the sidewalk market where the price displays a mean-reversion pattern.

Table 4: Frequency of Positive and Negative Returns

	State 1	State 2	State 3
Positive Return Freq.	47.03%	50.49%	67.19%
Negative Return Freq.	52.97%	49.51%	32.81%

Based on the estimated parameters in the component distribution, Figure 5 displays the histogram of the daily returns of the CSI 300, the empirical density, and three fitted component distribution densities. By separating the empirical distribution into three component distributions, the HSMM is able to explain the leptokurtosis and fat tail effects. The over-peak in the middle part of the empirical distribution mainly results from the sidewalk market, whereas the bear market plays a vital role in the fat tails. The standard deviation of the bear market is 3.113, which is much higher than for the other two markets. Hence, the bear market is the most volatile market, followed by the bull market. Conversely, the sidewalk market is the most stable market.

Figure 5: Empirical and Component Density



The existing literature often ignores the analysis of the mean of component distribution. However, the component mean is important for price behaviour. Although the mean of state 3 (0.622) is very small, it is still significantly larger than zero. This small but significant positive mean ensures that positive returns occur more frequently than the negative returns, which is the key feature of the bull market. The same logic can be applied to state 1. The insignificant mean of state 2 ensures that its distribution is almost symmetrical around 0 and the frequency of positive returns and negative returns is nearly the same.

Table 5 presents the number of days, the number of times, and average sojourn time for different market conditions. Our results show that the bull market has a slightly longer sojourn time than the bear market. Additionally, the average sojourn time for the sidewalk market is the longest with 204.29 days, which is much longer than in the case the other two types of markets.

Table 5: Sojourn Information

	State 1 (Bear)	State 2 (Sidewalk)	State 3 (Bull)
Number of Days	572	1430	695
Number of Times	22	7	25
Average Sojourn	26.00	204.29	27.80

Table 6 gives the estimated transition probability matrix (TPM) of the HSMM for the CSI 300 returns. The sojourn time of the HSMM is controlled by the sojourn time distribution rather than by the diagonal entries in the TPM. Hence, the diagonal entries are all zeros for the HSMM. There are a

few interesting economic implications that can be drawn from the TPM.

Table 6: Transition Probability Matrix

From \ To	State 1 (Bear)	State 2 (Sidewalk)	State 3 (Bull)
State 1 (Bear)	0	0.02%	99.98%
State 2 (Sidewalk)	49.56%	0	50.44%
State 3 (Bull)	74.08%	25.92%	0

- After a bear market, it is highly likely (99.98%) that a bull market will follow. This situation often occurs at the end of a crisis when the market starts to recover. Once the market has reached the bottom, it is common that investors overreact (over-sell) and stocks are under-valued.
- A bear market and a bull market have equal possibility (around 50%) to occur after a sidewalk market. In other words, it is unclear whether a bull or bear market will follow after the price fluctuates within a certain range for a long period.
- At the end of a bull market, the market has a high probability (74.08%) to be bear and a low probability (25.92%) to be sidewalk. These circumstances usually ensue after a bubble burst, such as the financial crisis in 2008.

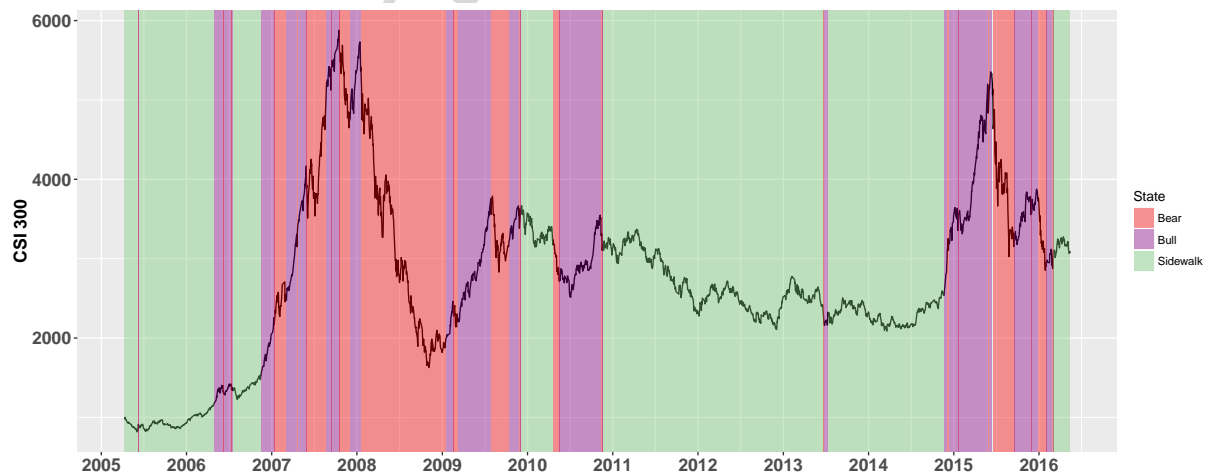
5.2. Decoding Results

The global decoding is conducted by the Viterbi algorithm. Figure 6 shows the global decoding states with reference to the CSI 300 original series, while Figure 7 is correlated with the daily returns of the CSI 300. The purple background represents the bull market, the red background denotes the bear market, and the green background stands for the sidewalk market. We review the evolution of the transition between the different market conditions for the Chinese stock market in our sample period.

- At the beginning of our sample period (April 8th 2005), the Chinese stock market was in a sidewalk market and lasted for about one year until April 27th 2006. After a short period of bull (April 28th 2006 to July 14th 2006), the market became sidewalk again and lasted for approximately 4 months (July 17th 2006 to November 13th 2006).
- At the end of 2006, the CSI 300 started to climb and reached its historically high peak at 5877.20 on October 10th 2007. Interestingly, it is common to believe that this period is a “pure” bull market, but our decoding results show that this period was not purely bull, but was in fact mixed with some periods of the bear market.

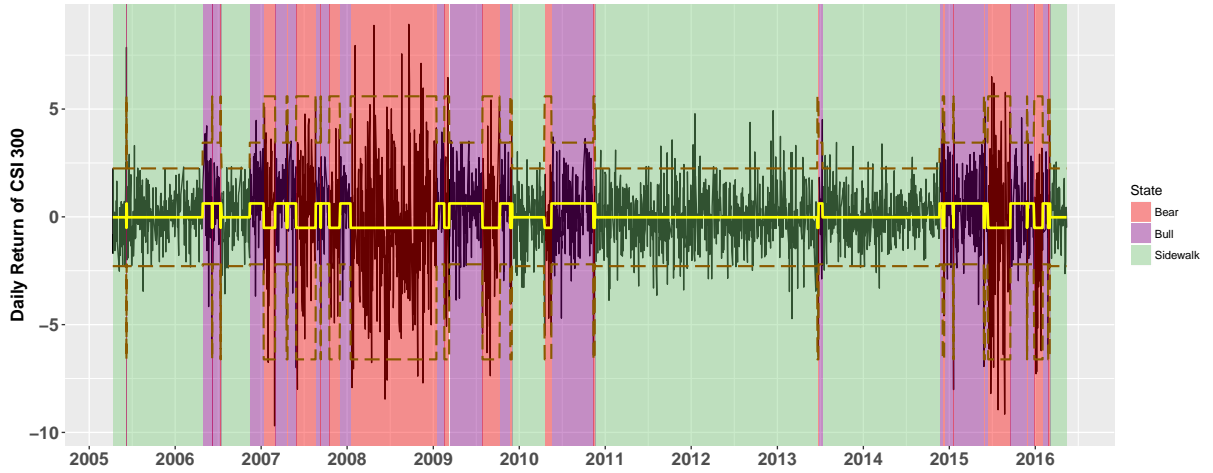
- After the financial crisis (March 12th 2007), the market went into a “pure” bear market and the CSI 300 dropped from its peak to the bottom of 1627.76 on November 4th 2008, which is the largest drop in the history of the Chinese stock market. After the market collapsed, it started to be bull and recover.
- Afterwards, the Chinese stock market experienced some periods of bear, sidewalk, and bull alternatively. It went into a remarkably long period of sidewalk from November 19th 2011 to November 20th 2014. During that period, the CSI 300 stayed in the range of 2000 to 3500 and displayed a mean-reversion pattern.
- From December 11th 2014, the Chinese stock market became bull and rocketed from 3183.01 to 5335.12 on June 12th 2015, which represented an astonishing increase of 67.61%. After this bull market, the Chinese stock market abruptly transited into a bear market. The CSI 300 shrank to 3025.70 on August 26th 2015, which was a dramatic 43.29% decrease.
- From September 18th 2015 to December 25th 2015, the Chinese stock market was a bull market over the course of three months. Afterwards, a short bear market and a short bull market occurred, followed by a sidewalk market.

Figure 6: CSI 300 and Market States (Global Decoding)



The local decoding (Figure 8) offers a more detailed probability of each state along with time in the sample period. The local decoding results confirm our understanding on the transition of the market conditions of the Chinese stock market. Before 2007, state 2 remained at a high level of probability. During 2007 and 2008, state 1 and state 3 alternatively reached high probabilities. After the financial crisis, state 1 remained at a high probability for about one year, while the other two remained low.

Figure 7: Daily Returns of CSI300 and Market States (Global Decoding)



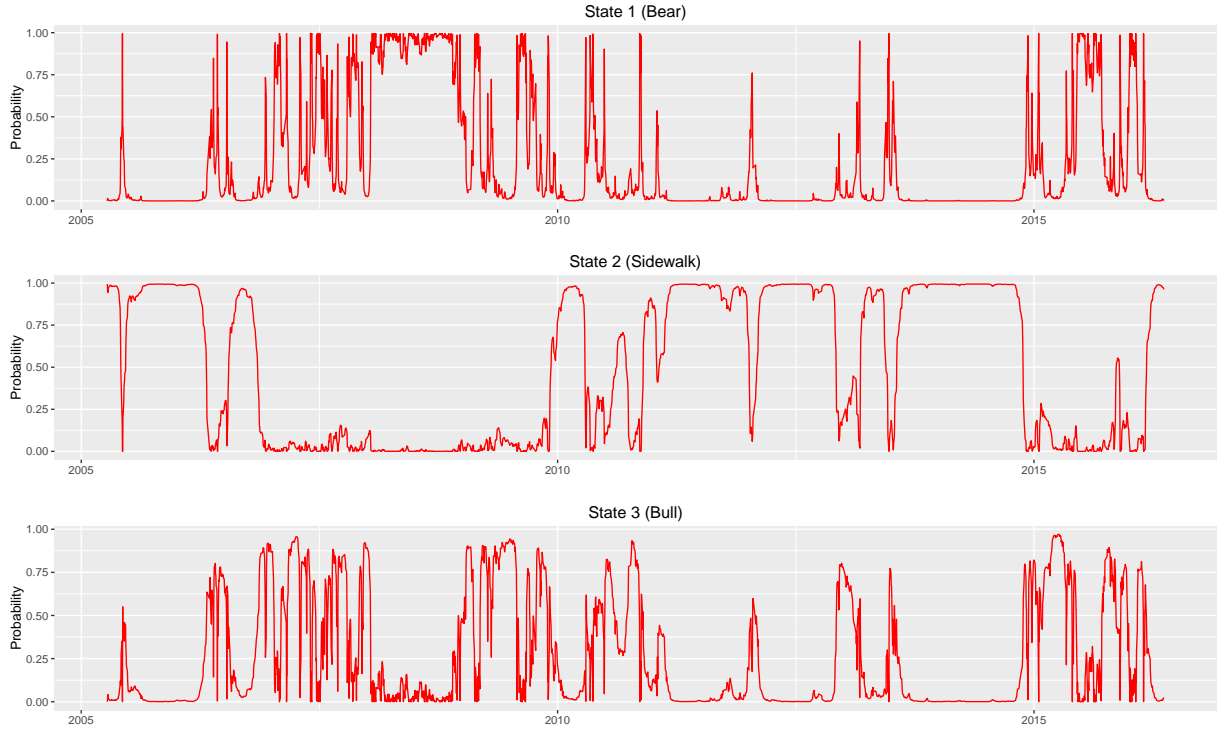
From 2011 to 2014, the probability of state 2 was almost 100% with a few exceptions. After 2015, the probability of state 3 reached a relatively high level again and was followed by a comparatively high level of state 1. State 2 has had the highest probability recently. The local decoding results are consistent with the global decoding results.

6. Model Evaluation and Comparison

In this section, we compare the empirical performance of our three-state HSMM with a stochastic volatility (SV) model and a t GARCH(1,1) model in terms of three stylized facts, namely the fat tails, the “long-memory”, and the Taylor effect.

The economic significance of the three stylized facts is as follows. Fat tails are related to the Value-at-Risk, which plays a vital important role in financial risk management. If the model cannot capture the correct left tail risk, the Value-at-Risk could be underestimated. In this circumstance, the investors may encounter an extra loss they do not expect. Figures 3a to 3c show that the return itself is not autocorrelated but that the squared return and the absolute return are autocorrelated and their ACF are slowly decaying. Note that the squared return and the absolute return are both volatility measures. Hence, the stylized fact of the “long-memory” is associated with volatility clustering, i.e. a large volatility tends to be followed by a large volatility and a small volatility tends to be followed by a small volatility. A good model should capture the persistence of the volatility in asset returns. The Taylor effect is a famous statistical observation. Taylor (1986) has initially found that the absolute return with power one has the highest autocorrelation. The following literatures treat the reproduction of the Taylor effect as

Figure 8: Local Decoding



an important benchmark (e.g. Rydén et al., 1998; Bulla & Bulla, 2006; Rogers & Zhang, 2011). If the model fails to reproduce the Taylor effect, then the data generation process in the model could not fully represent the empirical asset return.

Additionally, we also compare our model with a two-state HSMM, a three-state HMM, and a two-state HMM with respect to log likelihood, AIC, and BIC. This is for the model selection purpose and confirms our choice of a three-state HSMM.

6.1. Comparison with Other Volatility Models

Following Jacquier et al. (1994) and Kim et al. (1998), the stochastic volatility model is specified as

$$y_t = e^{h_t/2} \varepsilon_t \quad (8)$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma \eta_t \quad (9)$$

where y_t is the demeaned log return, h_t is the latent time-varying log volatility process, μ is the mean level of the log volatility, ϕ is the persistence coefficient for the volatility process, σ is the volatility of the log volatility, and ε_t and η_t are uncorrelated standard normal white noise shocks.

We use the MCMC method developed by Kastner & Frühwirth-Schnatter (2014) to estimate the stochastic volatility model. As for the mean level $\mu \in \mathbb{R}$, we choose the usual normal prior $\mu \sim \mathcal{N}(\log(\text{var}(y_t)), 1)$. The persistence parameter $\phi \in (-1, 1)$ is equipped with the Beta prior $(\phi + 1)/2 \sim \mathcal{B}(20, 1)$. In terms of the volatility of the log volatility $\sigma \in \mathbb{R}^+$, we choose $\sigma^2 \in 0.1 \times \chi_1^2$. For the MCMC setting, the thinning parameter is set to be 10, the burn-in parameter is 5000, and the number of draw is 55000. Table 7 presents the posterior draws of the parameters. The posterior mean is employed for the point estimation of the parameters.

The second benchmark model we consider is the tGARCH(1,1), which has the form

$$r_t = \mu + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim t(\nu) \quad (10)$$

$$\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (11)$$

where $\omega > 0$, $\alpha_1 > 0$, $\beta_1 > 0$, and $\alpha_1 + \beta_1 < 1$. r_t is the log return calculated in Equation 1. The conditional distribution is set to be the Student's t -distribution with the degree of freedom ν . We use the quasi-maximum likelihood method (Bollerslev & Wooldridge, 1992) to estimate the tGARCH(1,1) model, and the estimation results are reported in Table 8.

Table 7: Estimation for the SV Model

	Mean	Std. Error	5%	50%	95%
μ	0.933	0.198	0.622	0.932	1.255
ϕ	0.985	0.005	0.976	0.985	0.992
σ	0.141	0.017	0.115	0.140	0.171
$e^{\mu/2}$	1.602	0.160	1.365	1.594	1.873
σ^2	0.020	0.005	0.013	0.019	0.029

Table 8: Estimation for the tGARCH(1,1)

	Estimate	Std. Error	t-stat	P-value
μ	0.083	0.027	3.043	0.002
ω	0.026	0.010	2.719	0.007
α_1	0.056	0.009	6.368	0.000
β_1	0.939	0.009	102.187	0.000
ν	5.280	0.567	9.318	0.000

6.1.1. Fat Tail Reduction

Figure 9 shows the QQ plots of the log returns in the overall sample and three market conditions with respect to the theoretical normal distribution. The overall sample has tremendously fat tails since the QQ line deviates heavily from the diagonal line. After the separation of the whole sample into 3 states by our HSMM, the bear market and the bull market have close to normal distributions with slight tails, while the sidewalk market has modest tails. The QQ plots suggest that the distributions of the three market conditions are close to normal distributions. The reduction of fat tails can be confirmed by the kurtosis. The kurtosis of log returns is 6.089 in the overall sample, 3.002 in State 1 (Bear), 3.865 in State 2 (Sidewalk), and 3.283 in State 3 (Bull). The kurtosis of the three market conditions is close to 3. Hence, the assumption of the normal component distribution is suitable for our data. This confirms that the distributional property of the Chinese stock market returns could be a mixture of Gaussian distributions.

Figure 9: QQ Plots of the Log Returns

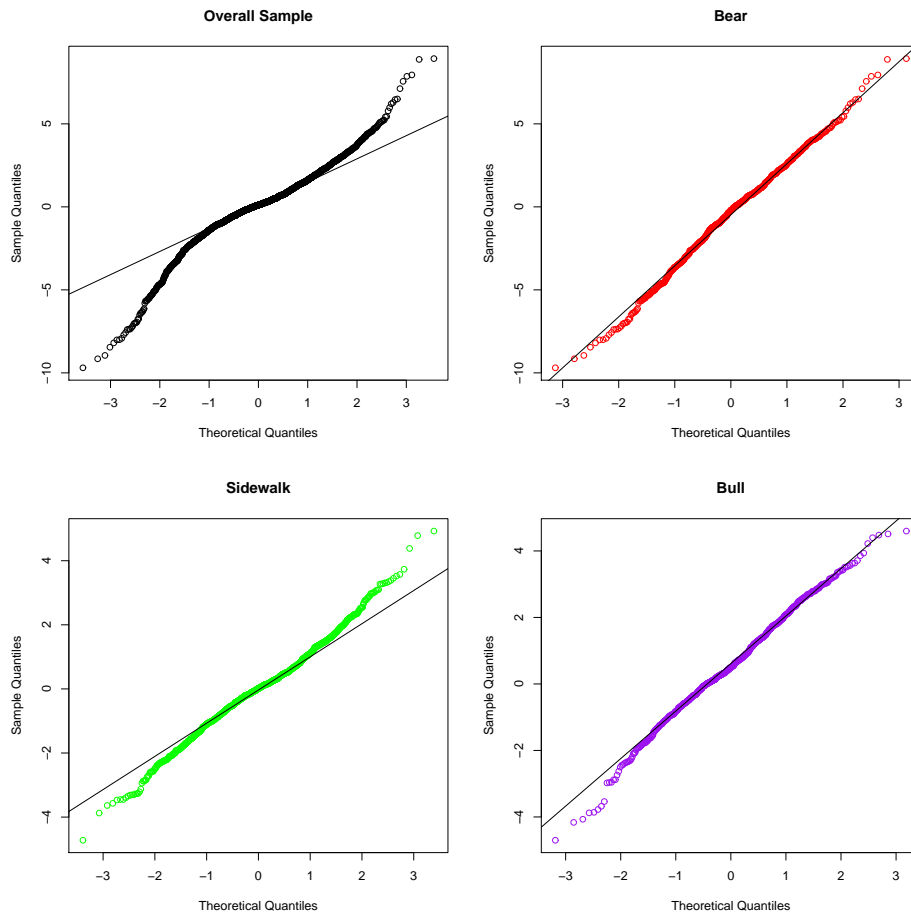


Figure 11 depicts the standardized residuals from three models and their QQ plot with respect to the theoretical normal distribution. The standardized residual of HSMM is defined as $(r_t - \bar{x}_i)/\sigma_i, i = 1, 2, 3$, the standardized residual of the SV model is ε_t in Equation 8, and the standardized residual of tGARCH is ε_t in Equation 10. Both the HSMM and the SV model can significantly reduce the fat tail, while the GARCH(1,1) still presents fat tails in its standardized residuals. The right tail of the standardized residuals in the SV model is slightly smaller than that in the HSMM. We further compare the kurtosis of the standardized residuals in the three models. The kurtosis of the standardized residuals is 3.560 in the HSMM, 3.285 in the SV model, and 4.678 in the tGARCH. In terms of fat tail reduction, the HSMM slightly underperforms the SV model, but they both outperform the tGARCH.

6.1.2. “Long-memory”

Rydén et al. (1998) could not reproduce the slow decay of the ACF of the squared or absolute returns by the HMM. It is interesting to examine the “long-memory” property of our model and two benchmark models. We simulate data from those three models based on the estimated parameters from our empirical CSI 300 log return data. The number of the Monte Carlo simulation accounts for 5000 repetitions. Figure 10 shows the empirical ACF and the model ACF for squared and absolute returns. The grey bars represent the empirical ACF while the red line is the model ACF.

Generally, all three models can reproduce the “long-memory” property because they all have slow decaying ACF. There are some differences between the ACF of the three models. Firstly, the model ACF of the HSMM is close to the empirical ACF before Lag 10, but it is slightly underestimated for large lags. Secondly, tGARCH has best fitting for the empirical ACF of the squared return, while the model ACF of the HSMM is underestimated and that of the SV model is overestimated. Thirdly, the SV model gives the best fitting for the empirical ACF of the absolute return, while the model ACF of the HSMM is still underestimated and that of the tGARCH is overestimated.

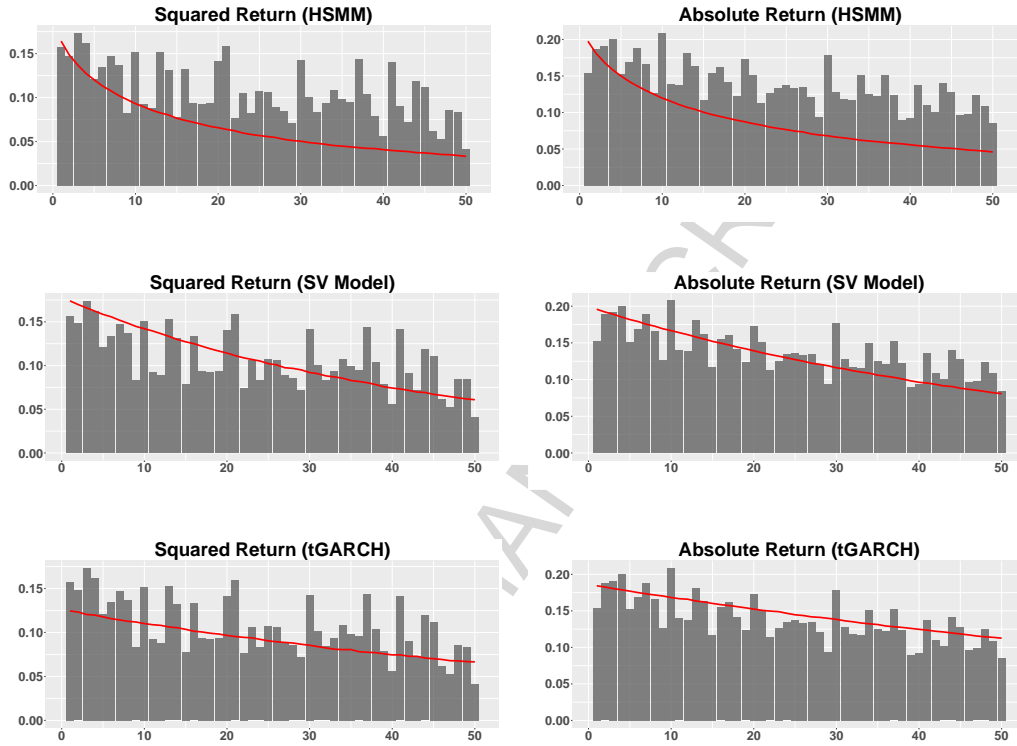
Our simulation results of the tGARCH are consistent with Ding et al. (1993). Their Monte Carlo simulation study also shows that ARCH type models can facilitate the “long-memory” property for both squared returns and absolute returns. Ding & Granger (1996) also derived the theoretical ACF for various GARCH(1,1) models and found them to be exponential decreasing.

6.1.3. Taylor Effect

The autocorrelation functions of different power values θ (i.e. $corr(|r_t|^\theta, |r_{t-k}|^\theta)$) for the three models are also simulated by the same Monte Carlo procedure with 5000 repetitions. Figure 12 displays the Taylor effect of all three models. The surface of the simulated Taylor effect is much smoother than that of the empirical Taylor effect in Figure 4.

Generally, all three models can reproduce the stylized fact of the Taylor effect, i.e. the autocorrelation function with power value $\theta = 1$ is the highest among the other power values. As can be seen in the

Figure 10: Empirical ACF and Model ACF



3-D plot, the surface decreases slowly when θ deviates from 1 and reaches the lowest values at two sides. Nevertheless, the shape of autocorrelation functions surface is different for the three models. The surface of the HSM has a more evident curvature before Lag 10 and becomes much flatter with larger lags. The surface of the SV model and the tGARCH has a consistent curvature in terms of the lag numbers. The tGARCH has a more blended surface than the SV model.

6.2. Comparison with the Hidden Markov Models

Our three-state HSM is compared with the two-state HSM used by Bulla & Bulla (2006), the three-state HMM adopted by Rydén et al. (1998), and the two-state HMM employed by Rydén et al. (1998) in terms of log likelihood, AIC, and BIC. Table 9 summarises the performance of all of the models for the return data of the CSI 300. Our three-state HSM has the highest log likelihood among the other models. This indicates that our model fits the empirical return of the CSI 300 better than the other models. However, it is not fair to evaluate model performance purely on the log likelihood because different models may have different numbers of parameters. Introducing more parameters usually increases the log likelihood but may result in overfitting.

Figure 11: Standardized Residuals and their QQ Plots

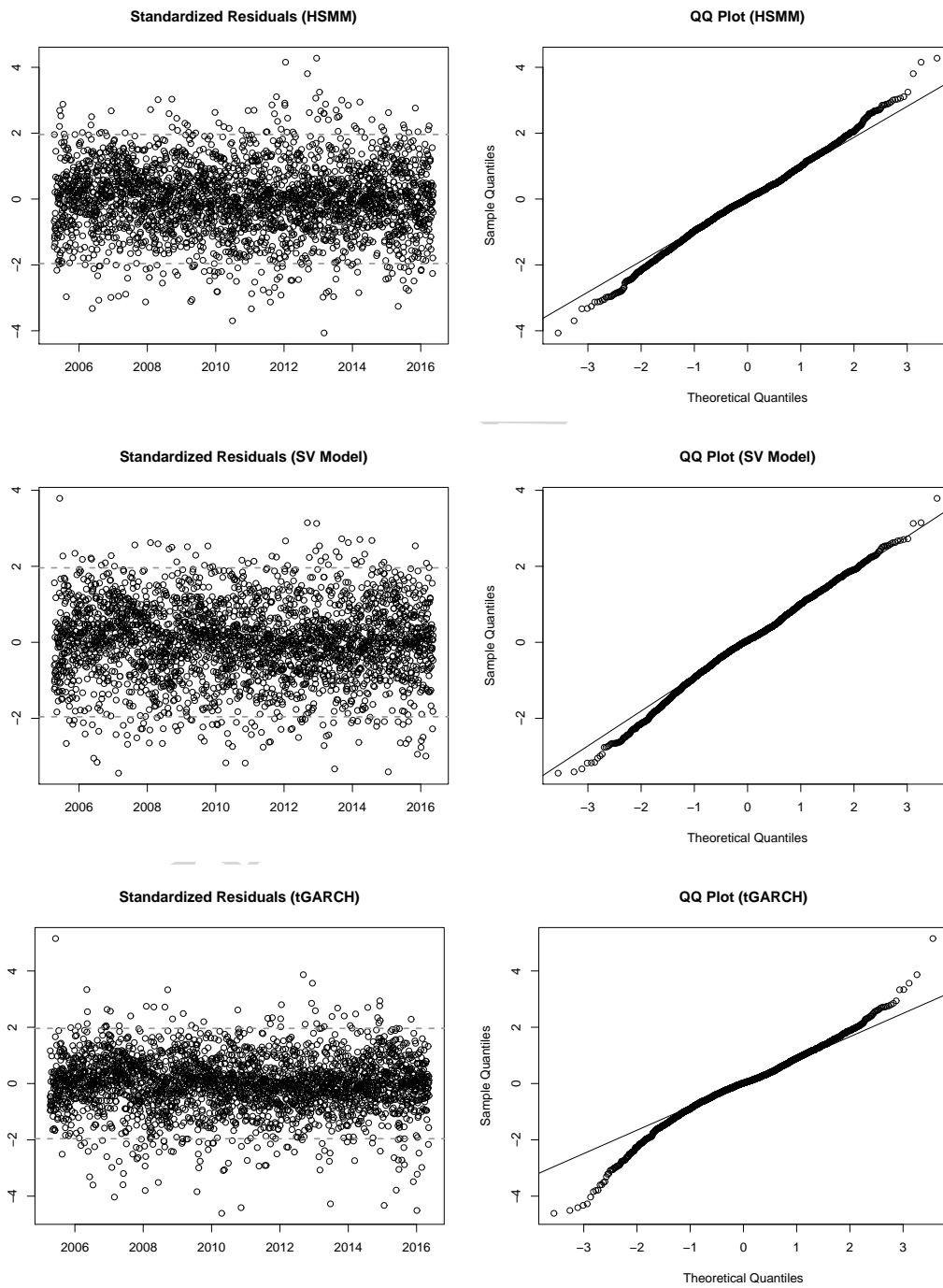
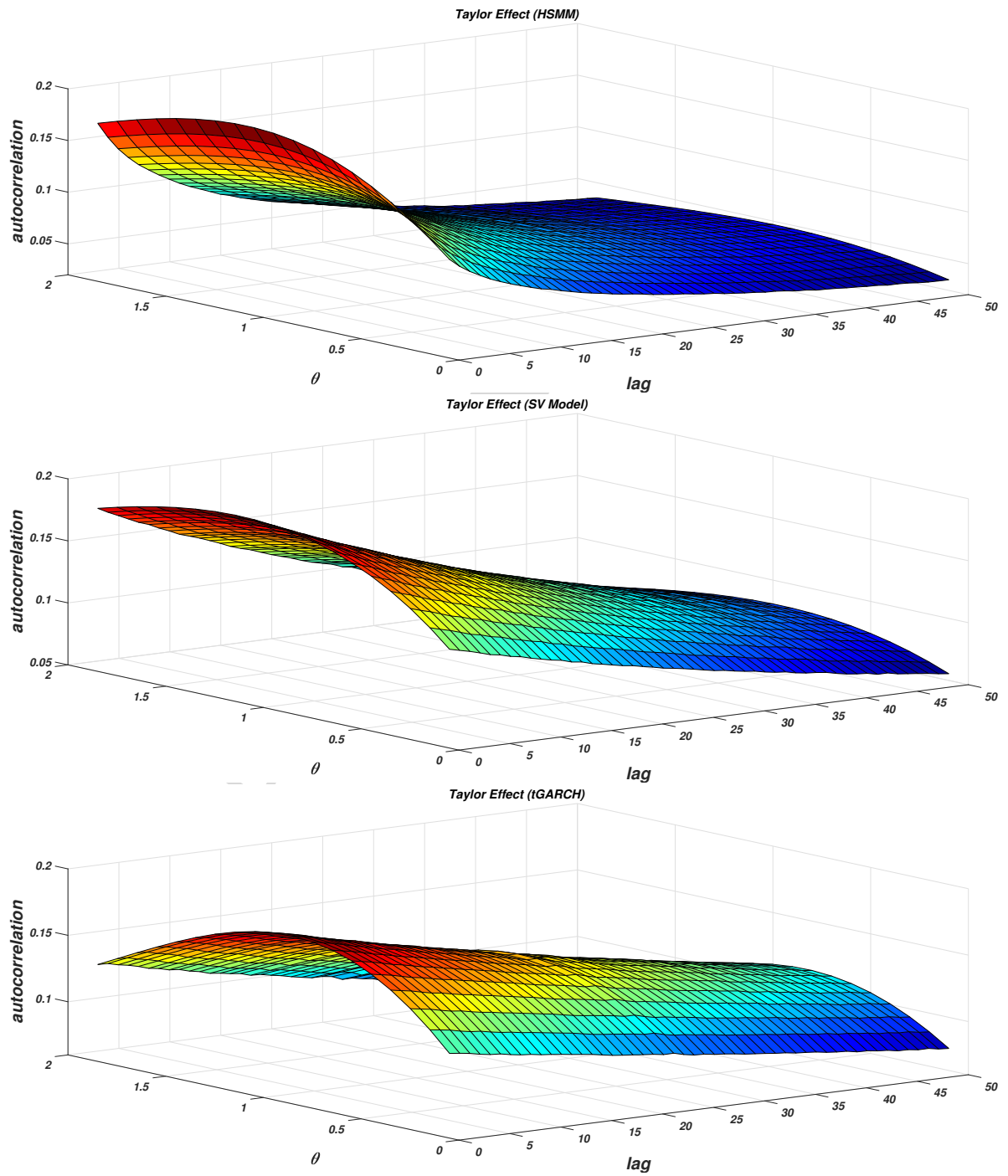


Figure 12: Taylor Effect



The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are fair model comparison tools. The AIC penalises the number of parameters and the BIC takes into consideration both the number of parameters as well as the sample size. A better model has a smaller AIC/BIC. The three-state HSMM has the smallest AIC and BIC. This implies that the HSMM does not over-fit the data and additional parameters are sensibly introduced to model the data.

Table 9: Model Comparison with Hidden Markov Models

	3-State HSMM	2-State HSMM	3-State HMM	2-State HMM
Iteration No.	130	18	234	45
Log Likelihood	-5178.911	-5213.159	-5186.953	-5229.820
AIC	10385.820	10440.320	10401.910	10473.640
BIC	10468.420	10481.620	10484.500	10514.940

7. Application 1: Links to Macroeconomic Variables

The first application of our three-state HSMM is to classify market conditions, and then investigate the links between macroeconomic variables and the market conditions. The three market conditions can be naturally ordered by the mean of the component distributions, i.e. bear in the low category, sidewalk in the middle category, and bull in the high category. It is straightforward to employ an ordinal logit model to examine the relationship between the decoded market conditions and the macroeconomic variables.

Since most economic variables have longer frequency than the daily data, we aggregate the daily market condition to the monthly market condition. The aggregation rule is to count the number of appearance of the three market conditions in a specific month, and then the one with the highest number of appearance represents the market condition for that month. Note that we remove the first month (April 2005) and the last month (May 2016) in our sample period because of the missing data.

We select eight important monthly macroeconomic variables shown in Table 10. The selected macroeconomic variables are from four different categories, including monetary policy, fiscal policy, economic growth, and the foreign exchange market. The selected eight variables have a low correlation with the others. In order to avoid multicollinearity, it is appropriate not to incorporate too many variables. In order to mitigate the impact of seasonality, we calculate the year-to-year % for most variables. There is no need to calculate the growth rate for the PMI since it is already expressed as a percentage. The monthly return (i.e. month-to-month %) of the exchange rate USD/RMB is a good measure of the valuation of the RMB. The sample period of all macroeconomic variables is from May 2005 to April 2016, in total 132 observations.

Table 10: Macroeconomic Variables Information

Code	Name	Transformation	Source
CPI	Consumer Price Index	Year-to-Year%	National Bureau of Statistics
LB	Loan Balance of Financial Institutions	Year-to-Year%	The People's Bank of China
GE	Government Expenditure	Year-to-Year%	National Bureau of Statistics
IFA	Investment of Fixed Assets	Year-to-Year%	National Bureau of Statistics
IVA	Industrial Value-added	Year-to-Year%	National Bureau of Statistics
RS	Total Retail Sales of Consumer Goods	Year-to-Year%	National Bureau of Statistics
PMI	Purchasing Managers' Index		National Bureau of Statistics
EXR	Exchange Rate (USD/RMB) - Monthly Avg.	Month-to-Month%	The People's Bank of China

Table 11 shows the estimation result of the ordinal logit model with the market condition as a dependent variable and the macroeconomic variables as independent variables. We also calculate the relative risk ratios, which allow an easier interpretation of the logit coefficients. We find that the year-to-year % of CPI (i.e. inflation), the PMI, and the monthly return of USD/RMB exchange rate are significantly related to the market condition of the Chinese stock market.

Table 11: Ordinal Logit Model

<i>Mkt. Cond.</i>	Estimation				Odds
	Value	Std. Error	z-stat	P-value	
CPI	-0.368**	0.162	-2.270	0.020	0.692**
LB	0.086	0.054	1.573	0.120	1.089
GE	0.004	0.012	0.355	0.720	1.004
IFA	-0.063	0.060	-1.065	0.290	0.939
IVA	-0.053	0.062	-0.856	0.390	0.948
RS	-0.122	0.113	-1.082	0.280	0.885
PMI	0.231**	0.098	2.361	0.020	1.260**
EXR	-1.015**	0.443	-2.290	0.020	0.363**
<i>Intercept</i>					
Bear Sidewalk	7.113	4.729	1.504	0.130	
Sidewalk Bull	10.159	4.773	2.128	0.030	

Note: *p<0.1; **p<0.05; ***p<0.01

Keeping all other variables constant, when the inflation increases by 1%, it is 0.692 times more likely to be in a higher category. In other words, a higher inflation indicates a more bearish market⁵. There are two reasons for the negative relationship between the inflation and the market condition. Firstly, investors expect that a high inflation would trigger a contractionary monetary policy imposed by the central bank, and then further limit the growth of the money supply. This is consistent with our ordinal logit regression results. There is weak evidence (P-value: 0.12) suggesting that the loan balance is positively related to a more bullish market. Secondly, a higher inflation increases the investors' expectation of the stock return and the discount rate, and shrinks the valuation of stocks.

If the PMI increases by 1%, the market is 1.26 times more likely to be a more bullish market. The significant positive relation between the PMI and the market condition is straightforward. The PMI is a leading indicator of the economic health of the manufacturing sector. A PMI of more than 50 (less than 50) indicates an expansion (contraction) of the manufacturing sector compared to the previous month. It is worth noting that Chinese investors pay great attention to the PMI, and make investment decisions based on the new release of the PMI.

If the RMB depreciates by 1% in terms of the USD, it is 0.363 times more likely to see a more bullish market. In short, the appreciation of the RMB induces a more bullish market. Generally speaking, when the domestic currency appreciates, it is intuitive to observe the increase of stock prices since the domestic assets dominating in the domestic currency go up. Additionally, a health economy usually comes up with a strong currency. During 2006 and 2007, it can be observed that the timing of the RMB appreciation is generally consistent with the bull market.

Interestingly, the growth rates of government expenditure, investment of fixed assets, industrial value-added, and retail sales exhibit an insignificant relationship with the market condition. Also, the positive relationship between the loan balance and the market condition is weak. This is because the official data of those economic indicators are released in the next month, not in the current month. Investors usually make decisions based on the economic prospect and use leading indicators, such as the PMI. It is possible to find significant relations between the market conditions and the lagged economic variables. For the purpose of simplicity, we leave future research studies to explore this possibility.

8. Application 2: Trading Strategy

In the second application, we design a simple trading strategy⁶ based on our three-state HSMM. In order to test the profitability of the trading strategy, we split the data into two parts, a training sample

⁵Compared with the bear market, the sidewalk market is a more bullish market. Compared with the bull market, the sidewalk market is a more bearish market.

⁶This is only a numerical demonstration of the trading strategy. Investors cannot directly trade the CSI 300 in China, but the index ETF can be its proxy.

(April 8th 2005 to December 31st 2013) and a testing sample (January 1st 2014 to May 13th 2016). The three-state HSMM is estimated by the data in the training sample.

In order not to use future information, we use the expanding window to recursively decode the most likely sequence of states. Specifically, we fix the start date of the window to April 8th 2005 and move the end date of the window to each date in the testing sample. The performance of the trading strategy is only evaluated for the testing sample.

For each expanding window, we conduct global decoding for the data and take the last decoded state in the window. The trading rule is as follows:

- If the last decoded state is the bear market → Short and hold in the next trading day
- If the last decoded state is the sidewalk market → No position in the next trading day
- If the last decoded state is the bull market → Long and hold in the next trading day

Figure 13 shows the cumulative return of the trading strategy in the upper panel, drawdown in the middle panel and trading signal in the lower panel. This trading strategy is profitable with an annualised return of 37.59% and a Sharpe ratio of 1.14⁷. The maximum drawdown occurred at -21.34% in January 2015. There are three remarkable periods. The trading strategy does not have any position (trading signal: 0) before December 2014; it takes a long position (trading signal: 1) in April 2014; and it takes a short position (trading signal: -1) from June 2015 to October 2015 and from January 2016 to March 2016. The majority of the profitability in the trading strategy comes from the short position.

For the robustness test, we conduct three other data split schemes and this trading strategy still shows high profitability. The robustness test results are shown in Appendix C.

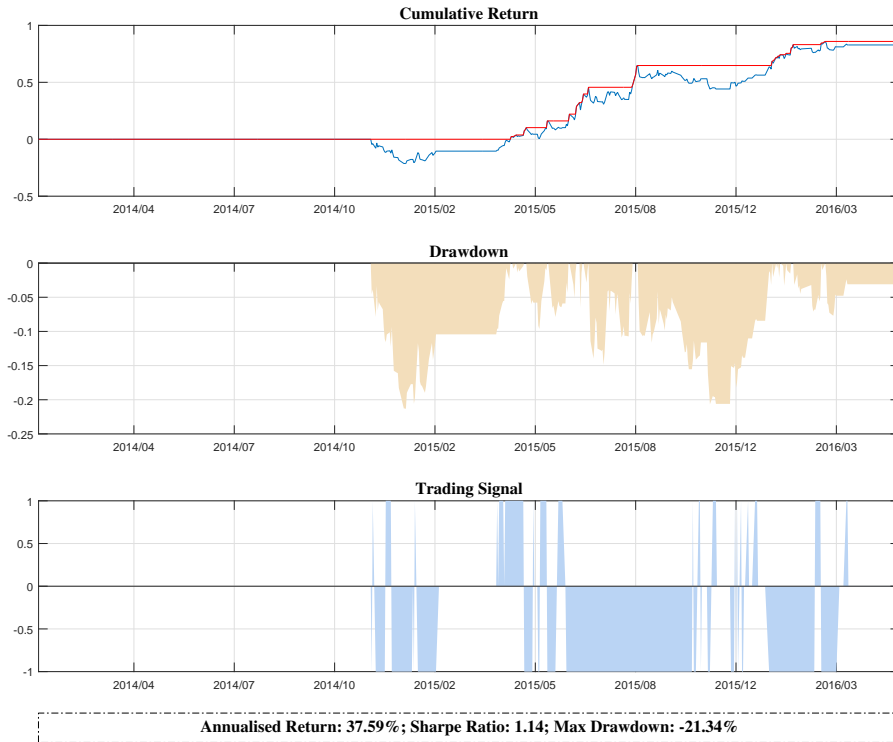
9. Conclusion

In this paper, we employ a three-state hidden semi-Markov model (HSMM) to explain the time-varying distribution of Chinese stock market returns. Our results indicate that the time-varying distribution of the Chinese stock market returns depends on the market conditions, namely the bear, sidewalk, and bull market. We reviewed the evolution of the market conditions in the Chinese stock market over the last decade. The most prominent periods are the bear market (January 16th 2008 to January 14th 2009), the long sidewalk market (November 19th 2011 to November 20th 2014), and a recent bull market (December 11th 2014 to May 27th 2015). Through Monte Carlo simulations, our three-state HSMM along with a SV model and a tGARCH(1,1) can reproduce the stylized facts of the “long-memory” and the Taylor effect, but tGARCH(1,1) fails to reduce the fat tails.

⁷The risk free rate in China is assumed to be a constant of 4.35% according to <http://www.global-rates.com/interest-rates/central-banks/central-bank-china/psc-interest-rate.aspx>.

Figure 13: Performance of the Simple Trading Strategy

(Trading Sample: Apr.2005 - Dec.2013; Testing Sample: Jan.2014 - May.2016)



In the first application, our three-state HSMM is used as a market condition classifier and the links between the decoded market conditions and the macroeconomic variables are identified by an ordinal logit model. We found that the inflation, the PMI, and the exchange rate were significantly related to the market condition of the Chinese stock market. In the second application, we design a simple trading strategy based on the expanding window decoding. The trading strategy shows profitability with a Sharpe ratio of 1.14 in the testing sample. Based on the results of three other data split schemes, we believe that the profitability of the trading strategy is robust.

One limitation of this paper is that our sample period is not very long. Future research may wish to explore the difference in terms of market conditions between China and other developed markets.

Appendix A. EM Algorithm

In this appendix, we provide technical details of the EM algorithm for the right-censored HSMM. We adopt the right-censored HSMM because the assumption of the classical HSMM that the last observation

always coincides with the exit from a state seems to be unrealistic for financial time series data (Bulla & Bulla, 2006). The right-censored setting releases the assumption that the last observation is always to be the end of a state. In other words, the last visited state will last for some time even after the last observation. There is no immediate jump to other states after the last observation. In the right-censored setting, the sojourn time in the last visited state is modelled by the survivor function, which shown in Equation A.1.

$$D_i(u) = \sum_{\nu > u} d_i(\nu) \quad (\text{A.1})$$

Guédon (2003) provided the complete-data likelihood for the right-censored HSMM. The complete-data likelihood function contains the observation \mathbf{X}_1^T and the state sequence \mathbf{S}_1^T , where $u - 1$ is the period that the last visited state will continue after the last observation. The last visited state will jump into other states at time $T + u$. The complete-data likelihood for the right-censored HSMM is shown in Equation A.2.

$$L_c(\mathbf{X}_1^T, \mathbf{s}_1^{T+u} | \boldsymbol{\theta}) = \mathbb{P}(\mathbf{X}_1^T = \mathbf{x}_1^T, \mathbf{S}_1^T = \mathbf{s}_1^T, \mathbf{S}_{T+1}^{T+u-1} = s_T, S_{T+u} \neq s_T | \boldsymbol{\theta}) \quad (\text{A.2})$$

The final likelihood function is obtained by summing all possible state sequences and all possible prolongation lengths of the last state, as indicated in Equation A.3.

$$L(\boldsymbol{\theta}) = \sum_{t=S_1, \dots, S_T} \sum_{u_{T+}} L_c(\mathbf{X}_1^T, \mathbf{s}_1^{T+u} | \boldsymbol{\theta}) \quad (\text{A.3})$$

where $\sum_{t=S_1, \dots, S_T}$ represents the summation of all possible state sequences, and $\sum_{u_{T+}}$ represents the summation over all possible additional sojourn time after time T .

It is difficult to compute the likelihood function in Equation A.2 because the underlying state sequence is unknown. It needs to consider all possibilities of the state sequence in order to compute the full likelihood, which is not realistic. The expectation-maximization (EM) algorithm provides a suitable procedure to deal with the missing data problem. The EM algorithm (Baum et al., 1970) is an iterative procedure to increase the likelihood until it reaches the convergence criteria. The EM algorithm iteratively conducts the E-step and M-step. Given an initial guess of the parameter vector $\boldsymbol{\theta}$, the E-step firstly computes the Q-function, which is the conditional expectation for the complete-data likelihood.

- **E-step**

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t-1)}) = \mathbb{E} \left\{ L_c(\mathbf{X}_1^T, \mathbf{s}_1^{T+u} | \boldsymbol{\theta}) | \mathbf{X}_1^T = \mathbf{x}_1^T, \boldsymbol{\theta}^{(t-1)} \right\} \quad (\text{A.4})$$

Based on the Q-function in Equation (12), the M-step aims to maximise the Q-function with respect to parameter $\boldsymbol{\theta}$.

- **M-step**

$$\boldsymbol{\theta}^{(t)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t-1)}) \quad (\text{A.5})$$

The parameter vector θ that maximises the Q-function in the M-step of the previous iteration will be used in the E-step of the next iteration. Along with every iteration, the likelihood is non-decreasing. The algorithm will stop once the convergence criterion is satisfied. Normally, the convergence criterion is the successive change of likelihood is less than a very small number. The EM algorithm is not guaranteed to reach the global maxima and it might be trapped in local maxima. Hence, it is necessary to try different initial values of parameter vector θ in order to check that the maximum reached is the global maximum rather than the local maximum.

In this paper, we focused on the economic interpretation of the HSMM rather than on the mathematical derivation. We direct the reader to the thesis of Bulla (2006) for the mathematical details of the EM algorithm of the HSMM. In his thesis, he decomposes the Q-function of the HSMM into four components, which correspond to the initial probabilities, transition probabilities, sojourn time, and component distributions. The E-step is implemented by the forward-backward algorithm. His decomposition framework facilitates the M-step in which the four components can be maximised individually. Bulla shows the close-form solutions for some common distributions for the sojourn time and component distributions, but the numerical solver could be applied if a closed solution does not exist.

Unlike the maximum likelihood method whereby the standard errors can be directly calculated by the Fisher information matrix (FIM), one drawback of the EM algorithm is that the FIM is not a by-product of the algorithm. It is highly unlikely to obtain the FIM by evaluating analytically the second-order derivatives of the marginal log-likelihood of the HSMM. Recent numerical methods are developed to get an approximation of the FIM (see Louis, 1982; Meng & Rubin, 1991; Jamshidian & Jennrich, 2000). However, all these methods have limitations (Meng, 2016). There is no generally accepted method to get diagnosis for the estimation of the HSMM by the EM algorithm. Hence, we do not provide the diagnosis of the estimated parameters in this paper.

Appendix B. Decoding Technique

It is interesting to decode the most likely states in the Markov chain. There are two decoding techniques for the HSMM, global decoding and local decoding. Global decoding aims to determine the most likely sequence of the states given the observations, while the local decoding computes the conditional probability of each state at times given the observations. Normally, they produce similar but not identical decoding results.

Appendix B.1. Global Decoding

The purpose of global decoding is to find the most likely sequence of states conditional on the observations. Mathematically speaking, global decoding intends to find a sequence of states with the

highest likelihood given the observations, which is shown in Equation B.1.

$$\hat{\mathbf{S}}_1^T = \arg \max_{\mathbf{S}_1^T} \mathbb{P}(\mathbf{S}_1^T | \mathbf{X}_1^T) \quad (\text{B.1})$$

The exhaustive attack method has the computational complexity level at $\mathcal{O}(m^T)$. This brute force method is not feasible for long sequence data. The Viterbi algorithm (Viterbi, 1967) was developed by utilising the Markov property of the HMM and HSMM. This is an efficient dynamic programming algorithm and its complexity level is $\mathcal{O}(m \times T)$. The Viterbi algorithm works in the following way.

$$\arg \max_{\mathbf{S}_1^T} \mathbb{P}(\mathbf{S}_1^T | \mathbf{X}_1^T) = \arg \max_{\mathbf{S}_1^T} \mathbb{P}(\mathbf{S}_1^T, \mathbf{X}_1^T) \quad (\text{B.2})$$

Define

$$\mu_t(S_t) = \max_{\mathbf{S}_1^{t-1}} \mathbb{P}(\mathbf{S}_1^t, \mathbf{X}_1^t) \quad (\text{B.3})$$

The recursion expression for $\max \mathbb{P}(\mathbf{S}_1^t, \mathbf{X}_1^t)$ can be derived.

$$\begin{aligned} \max_{\mathbf{S}_1^{t-1}} \mathbb{P}(\mathbf{S}_1^t, \mathbf{X}_1^t) &= \max_{\mathbf{S}_1^{t-1}} \mathbb{P}(X_t | S_t) \mathbb{P}(S_t | S_{t-1}) \mathbb{P}(\mathbf{S}_1^{t-1}, \mathbf{X}_1^{t-1}) \\ &= \max_{S_{t-1}} \left\{ \mathbb{P}(X_t | S_t) \mathbb{P}(S_t | S_{t-1}) \max_{\mathbf{S}_1^{t-2}} \mathbb{P}(\mathbf{S}_1^{t-1}, \mathbf{X}_1^{t-1}) \right\} \\ &= \max_{S_{t-1}} \mathbb{P}(X_t | S_t) \mathbb{P}(S_t | S_{t-1}) \mu_{t-1}(S_{t-1}) \end{aligned} \quad (\text{B.4})$$

The recursion equation for $\mu_t(S_t)$ is as follows

$$\mu_t(S_t) = \max_{S_{t-1}} \mathbb{P}(X_t | S_t) \mathbb{P}(S_t | S_{t-1}) \mu_{t-1}(S_{t-1}) \quad \text{for } t = 2, \dots, T \quad (\text{B.5})$$

Equation B.5 is able to compute the most likely trajectory for each state up to time t . At time T , the state with the highest $\mathbb{P}(\mathbf{S}_1^T, \mathbf{X}_1^T)$ is picked up and the corresponding trajectory is the solution for the Viterbi algorithm.

Appendix B.2. Local Decoding

The purpose of local decoding is to compute the conditional probabilities for each state at time t given the observation.

$$\hat{S}_t = \arg \max_{S_t} \mathbb{P}(S_t = i | \mathbf{X}_1^T = \mathbf{x}_1^T) \quad \text{for } i = 1, \dots, m \quad (\text{B.6})$$

The conditional probability in Equation B.6 can be decomposed into three terms.

$$\begin{aligned} \mathbb{P}(S_t = i | \mathbf{X}_1^T = \mathbf{x}_1^T) &= \mathbb{P}(S_{t+1} \neq i, S_t = i | \mathbf{X}_1^T = \mathbf{x}_1^T) \\ &\quad + \mathbb{P}(S_{t+1} = i | \mathbf{X}_1^T = \mathbf{x}_1^T) \\ &\quad - \mathbb{P}(S_{t+1} = i, S_t \neq i | \mathbf{X}_1^T = \mathbf{x}_1^T) \end{aligned} \quad (\text{B.7})$$

Define $\xi_t(i) = \mathbb{P}(S_t = i | \mathbf{X}_1^T = \mathbf{x}_1^T)$. It is able to obtain the recursion equation for the conditional probability.

$$\xi_t(i) = \mathbb{P}(S_{t+1} \neq i, S_t = i | \mathbf{X}_1^T = \mathbf{x}_1^T) + \xi_{t+1}(i) - \mathbb{P}(S_{t+1} = i, S_t \neq i | \mathbf{X}_1^T = \mathbf{x}_1^T) \quad (\text{B.8})$$

With Equation B.8, $\xi_t(i)$ can be computed based on the $\xi_{t+1}(i)$ and it is able to calculate the conditional probabilities at all time in a backward way (see details in Guédon (2003)).

Appendix C. Robustness Test of the Trading Strategy

We implemented three other data split schemes in order to test the profitability of our trading strategy. The first scheme is cutting the sample at the end of 2009 (i.e. trading sample: Apr.2005 - Dec.2009, testing sample: Jan.2010 - May.2016); the second scheme is at the end of 2012 (i.e. trading sample: Apr.2005 - Dec.2012, testing sample: Jan.2013 - May.2016); and the third scheme is at the end of 2014 (i.e. trading sample: Apr.2005 - Dec.2014, testing sample: Jan.2015 - May.2016). For each scheme, the HSM is estimated by using the training data, and then the trading strategy is tested for the period of the testing sample. The trading strategy still shows high profitability for all three schemes (see Figure C.14 to Figure C.16). The profitability of the trading strategy is believed to be robust.

Figure C.14: Performance of the Simple Trading Strategy

(Trading Sample: Apr.2005 - Dec.2009; Testing Sample: Jan.2010 - May.2016)

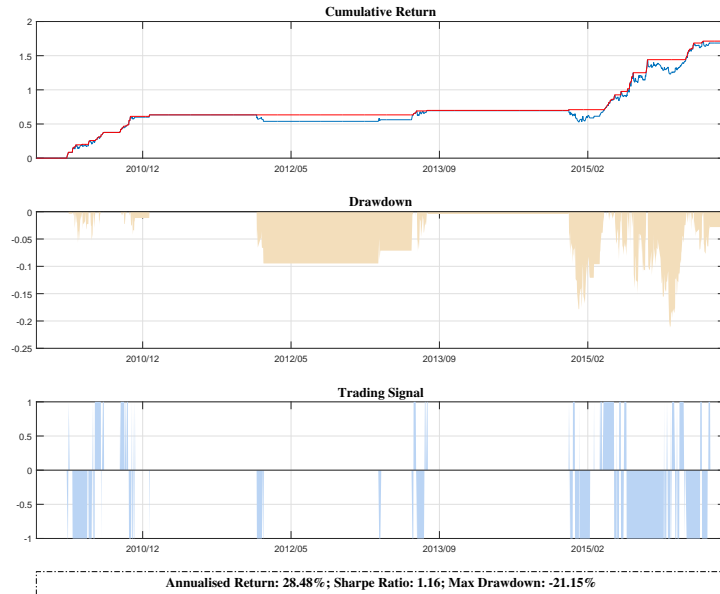


Figure C.15: Performance of the Simple Trading Strategy

(Trading Sample: Apr.2005 - Dec.2012; Testing Sample: Jan.2013 - May.2016)

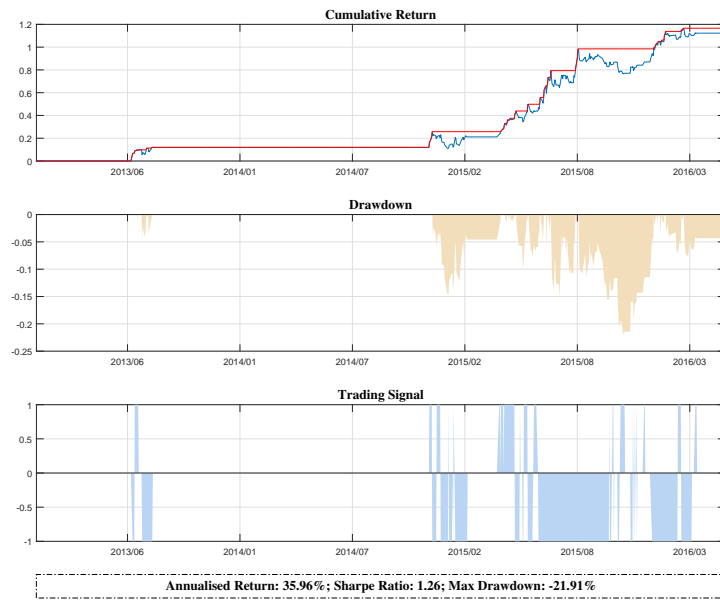
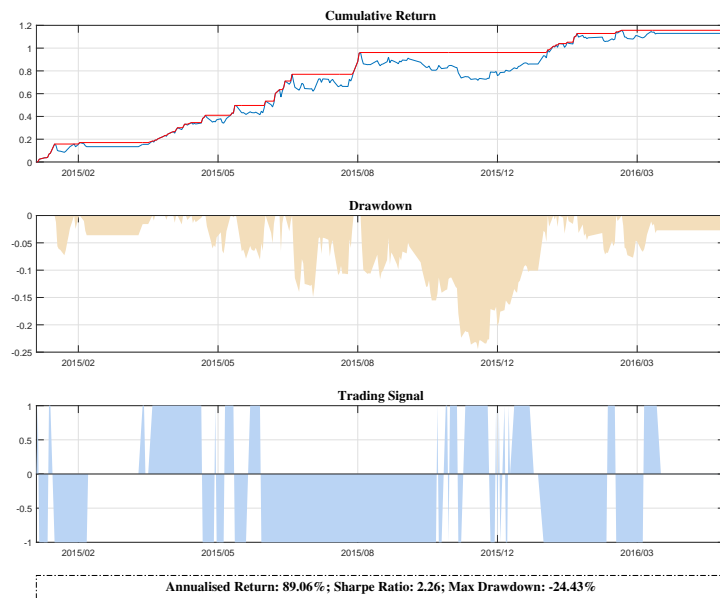


Figure C.16: Performance of the Simple Trading Strategy

(Trading Sample: Apr.2005 - Dec.2014; Testing Sample: Jan.2015 - May.2016)



Acknowledgement

The authors would like to thank the anonymous reviewer and the journal editor for their helpful remarks, which led to an improvement of the paper. Comments from Rasha Alsakka, Géraldine David, Haoshen Hu, Andrea Nocera, Claudio Morana, Eric Girardin, Anindya Banerjee, and Lajos Horváth are gratefully acknowledged. The responsibility of any error in the paper remains ours. Versions of the paper have been presented at the 14th INFINITI Conference on International Finance, Dublin, June 2016, the World Finance Conference, New York, July 2016, the Chinese Economic Association Annual Conference, Duisburg, September 2016, and seminars at the University of Birmingham (Departmental Seminar, Birmingham) and the Renmin University of China (Huang Da-Mundell Lecture Series, Beijing).

References

- 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*, 164–171.
- Blattberg, R. C., & Gonedes, N. J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *The journal of business*, *47*, 244–280.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*, 307–327.
- Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, *52*, 5–59.
- Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, *11*, 143–172.
- Bulla, J. (2006). *Application of hidden Markov models and hidden semi-Markov models to financial time series*. Ph.D. thesis.
- Bulla, J., & Bulla, I. (2006). Stylized facts of financial time series and hidden semi-markov models. *Computational Statistics & Data Analysis*, *51*, 2192–2209.
- Bulla, J., & Bulla, I. (2013). *hsmm: Hidden Semi Markov Models*. URL: <https://CRAN.R-project.org/package=hsmm> r package version 0.4.
- Chan, K., Menkveld, A. J., & Yang, Z. (2008). Information asymmetry and asset prices: Evidence from the china foreign share discount. *The Journal of Finance*, *63*, 159–196.

- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:2, 223–236.
- Ding, Z., & Granger, C. W. (1996). Modeling volatility persistence of speculative returns: a new approach. *Journal of econometrics*, 73, 185–215.
- Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1, 83–106.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, (pp. 987–1007).
- Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business*, 38, 34–105.
- Girardin, E., & Liu, Z. (2003). The chinese stock market: A casino with 'buffer zones'? *Journal of Chinese Economic and Business Studies*, 1, 57–70.
- Girardin, E., & Liu, Z. (2005). Bank credit and seasonal anomalies in china's stock markets. *China Economic Review*, 16, 465–483.
- Girardin, E., & Liu, Z. (2007). The financial integration of china: New evidence on temporally aggregated data for the a-share market. *China Economic Review*, 18, 354–371.
- Granger, C. W., & Ding, Z. (1995). Some properties of absolute return: An alternative measure of risk. *Annales d'Economie et de Statistique*, (pp. 67–91).
- Guédon, Y. (2003). Estimating hidden semi-markov chains from discrete sequences. *Journal of Computational and Graphical Statistics*, 12, 604–639.
- Jäckel, P. (2004). *Stochastic volatility models: past, present and future*. Chichester, UK: Wiley.
- Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20, 69–87.
- Jamshidian, M., & Jennrich, R. I. (2000). Standard errors for em estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 257–270.
- Kastner, G., & Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76, 408–423.
- Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, 65, 361–393.

- Los, C. A., & Yu, B. (2008). Persistence characteristics of the chinese stock markets. *International Review of Financial Analysis*, *17*, 64–82.
- Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, (pp. 226–233).
- Mandelbrot, B. B. (1997). *The variation of certain speculative prices*. Springer.
- Mei, J., Scheinkman, J. A., & Xiong, W. (2009). Speculative trading and stock prices: Evidence from chinese ab share premia. *Annals of Economics and Finance*, *10*, 225–255.
- Meng, L. (2016). Method for computation of the fisher information matrix in the expectation-maximization algorithm. *arXiv preprint arXiv:1608.01734*, .
- Meng, X.-L., & Rubin, D. B. (1991). Using em to obtain asymptotic variance-covariance matrices: The sem algorithm. *Journal of the American Statistical Association*, *86*, 899–909.
- Mittnik, S., & Rachev, S. T. (1993). Modeling asset returns with alternative stable distributions. *Econometric reviews*, *12*, 261–330.
- Pagan, A. (1996). The econometrics of financial markets. *Journal of empirical finance*, *3*, 15–102.
- Praetz, P. D. (1972). The distribution of share price changes. *The Journal of Business*, *45*, 49–55.
- Rogers, L. C. G., & Zhang, L. (2011). An asset return model capturing stylized facts. *Mathematics and Financial Economics*, *5*, 101–119.
- Rydén, T., Teräsvirta, T., Åsbrink, S. et al. (1998). Stylized facts of daily return series and the hidden markov model. *Journal of applied econometrics*, *13*, 217–244.
- Taylor, S. J. (1986). *Modelling financial time series*. world scientific.
- Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*, 260–269.
- Yu, S.-Z. (2010). Hidden semi-markov models. *Artificial Intelligence*, *174*, 215–243.

Highlights

- We explain the Chinese stock market returns by a three-state HSMM.
- The time-varying distribution depends on three market conditions (Bear/Sidewalk/Bull).
- The inflation, the PMI, and the exchange rate are related to the market condition.
- A simple trading strategy based on expanding window decoding shows profitability.

ACCEPTED MANUSCRIPT