Pacific-Basin Finance Journal 44 (2017) 127–149

Contents lists available at ScienceDirect

Pacific-Basin Finance Journal

journal homepage: www.elsevier.com/locate/pacfin

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

Zhenya Liu[a,b,*], Shixuan Wang[c]

[a] Renmin University of China, Beijing 100872, PR China
[b] University of Birmingham, Birmingham B15 2TT, UK
[c] Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, UK.

ARTICLE INFO

JEL classification:
C22
G12
G15

Keywords:
Chinese stock market
Asset return
Hidden semi-Markov model

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.

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 and Liu, 2007; Mei et al., 2009; Los and Yu, 2008; Chan et al., 2008; Girardin and 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 Fig. 1).

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, which are 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.

[☆] 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.
E-mail addresses: zhenya_liu@hotmail.com (Z. Liu), shixuan_wang@hotmail.com (S. Wang).

http://dx.doi.org/10.1016/j.pacfin.2017.06.007
Received 16 September 2016; Received in revised form 21 March 2017; Accepted 20 June 2017
Available online 23 June 2017
0927-538X/

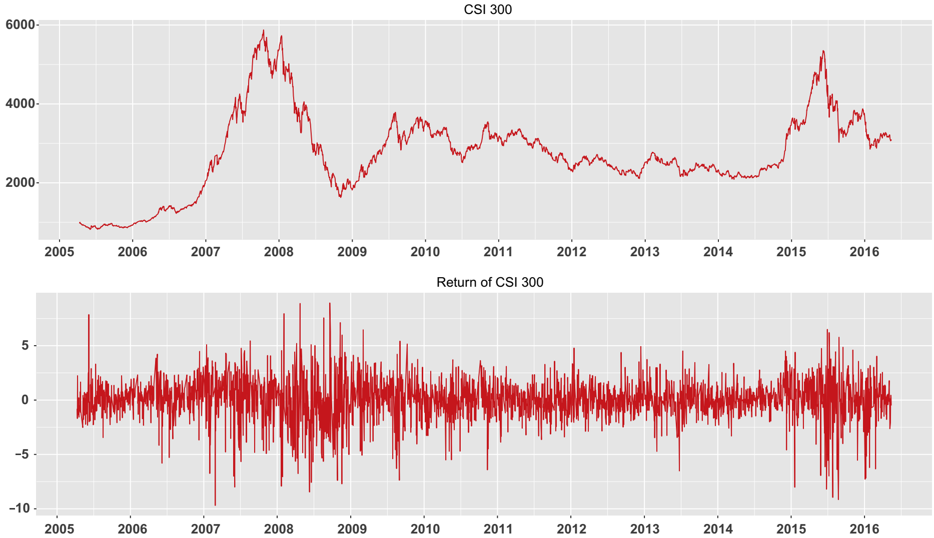

Fig. 1. CSI 300 and its returns.

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 and 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 and 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 and Ding (1995) suggested that the appropriate distribution is the double exponential distribution with zero mean and unit variance. Mittnik and 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 and 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 and 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 and 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