MODELING STOCK INDEX USING FINITE STATE MARKOV CHAIN

Hammad Hassan Mirza¹, Mian Sajid Nazir² and Ghulam Ali³

Abstract

The Existing stock price models are mainly based on time series methodologies which are hard to estimate and involve lots of assumptions. This study, in contrast, assumes that the stock prices follow stochastic process that possesses Markov dependency with finite state transition probabilities and proposes an alternate methodology for stock price modeling. For this purpose, daily stock index data from Pakistan Stock Exchange (PSE) is collected from 2010-2015 and categorized in to 10 state spaces. Based on the results of state transition model, the study highlights the most probable state of return and also its transition into another state. Further, the study used Monte Carlo method of stock index simulations both Markov chain and original stock index. The analysis shows that it is possible to model and forecast stock index by capturing the underlying Markov process. The results of the study are helpful for investors in selecting the right time of making investment and for academician to think about more sophisticated methods of state identification.

Keywords: Markov Chain, Stock Index, Finite State Space, Weak-form-Efficiency, Monte Carlo Simulation.

JEL Classification: G100

Introduction

Portfolio optimization problem remains one of the most challenging topics for financial researchers of this century. The Markowitz classical approach of mean-variance analysis was mainly based on maximization of returns while keeping the variance at a constant level. Fama (1965) empirically confirmed the accuracy of Bachelier (1914) random walk behavior of stock prices where stock price reacts according to the arrival of new information. Analysis of financial time series and investigating stock price behaviors has also been a subject of study in finance.

¹ PhD/Assistant Professor, Noon Business School, University of Sargodha. Email: hammad.hassan@uos.edu.pk

² PhD/Assistant Professor, Department of Management Sciences, COMSATS University, Islamabad, Lahore Campus. Email: snazir@cuilahore.edu.pk

³ PhD/Associate Professor, Noon Business School, University of Sargodha. Email: ghulam.ali@uos.edu.pk

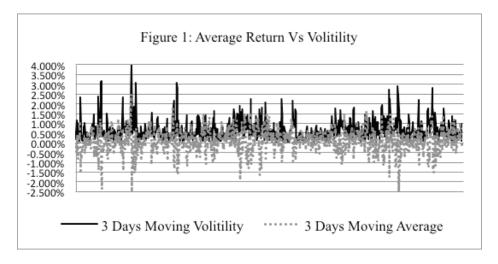
Several researchers (Fama 1965; Nelson 1991; Mandelbrot, 1997) argued that stock return distribution is leptokurtic and the classical as well as most of the conventional analysis techniques ignore the state transition property of asset returns. Therefore, the markets which are characterized by regime switching property of asset return, more appropriate technique is to analyze their state transition property instead of conventional methods of time series analysis (Zhou & Yin 2003; Yin & Zhou, 2004; Guidolin & Timmermann, 2007; Bae et al., 2014). The present study suggest Markov switching model to be more appropriate method of analysis is such scenario.

Markov chain is an important concept in modeling conceptual processes which has evolved over a period of time. It assumes that future values depend only on current observation while the knowledge of history is irrelevant. Markov switching models are based on the idea of transition of one state in to another state which is governed by Markov process. Wikipedia defines Discrete Time Markov Chain (DTMC) in the following words;

> "DTMC is a random process that undergoes transitions from one state to another on a state space. It must possess a property that is usually characterized as memorylessness: the probability distribution of the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of memorylessness is called the Markov property".

The idea of construction and analysis of Markov chain is related with the concept of market efficiency and the Markov process seems to be consistent with weak form of market efficiency. According to the market efficiency theories, the stock prices react according to the information and in case of weak form of market efficiency, the past price information is fully incorporated in the current price. This implies that it is near to impossible that investor gets abnormal profit over a long period of time using the trading rules based on historical price information. The Markov chain assumes that prices keep changing within specified but unobserved price range (the present study calls it price regime or state) which forms psychological price barriers and in order to break the barriers the market needs information. The present study assumes that potential information pushes the prices from one regime to another regime. Technically, the prices follow a pattern and there could be several ways of analyzing the price pattern and use it for future predictions. The previous studies have applied time series methodologies including Exponentially Weighted Moving Average (EWMA), Auto Regressive Integrated Moving Average (ARIMA), Generalized Auto Regressive Conditional Hetroscedastic (GARCH) and technical chart analysis techniques but only a handful of researchers have used regime switching methodology for analysis. The first step is to identify the regimes, second is to analyze and record the frequency of transition of one state into another, third is to calculate the transitional probabilities and finally use the probabilities for further price forecasting. The detail of this process is discussed in the methodology section.

As stated earlier, few researchers have used the regime switching Markov chain methodology in analyzing the stock prices, the distinction of the idea presented here is the introduction of several finite state spaces instead of conventional method of 3 major state spaces. PSE is among the highly volatile market of the region which is more volatile during low negative return regimes and less volatile during positive return periods. This is evident in figure 1 which compares the 3 days moving average returns of PSE with 3 days volatility. The figure depicts that during the high negative return periods the volatility graph is at its peak while returns are more stable during the period of positive returns. As conventional approaches like EWMA, ARIMA and GARCH are not much effective as the parameters of these models are constants and introducing Ito's process ⁴ in theses model are extremely complex. Therefore, the present study suggests the use of regime switching models to be more appropriate in case of PSE.



Literature Review

The existing literature provides a good deal of debate on the idea of regime shifts and Markov switching models. Since the time of Goldfeld and Quandt (1973) who appears to be the pioneers in addressing the existence of regimes and introduced the regime switching regression for estimation. Goldfeld and Quandt (1973 b) latter addressed the issues in structural shifts by switching regression. Hamilton (1989) improved the model of Goldfeld and Quandt (1973) by allowing regime shifts in dependent variable and introduced Markov Switching Autoregressive (MSAR) model.

⁴ Ito process is based on Itô calculus, named after Kiyoshi Itô, which extends the methods of calculus to stochastic processes such as Brownian motion (see Wiener process). It has important applications in mathematical finance and stochastic differential equations. The details of this complex idea is beyond the scope of this paper.

Research

The application of regime switching models in case of stock market returns was first introduced by Turner et al. (1989) who capture the regime shift behavior in mean and variance of stock market returns using MSAR. Latter, several researchers (for instance, Chu et al., 1996; Schaller & Norden, 1997 & Nishiyama, 1998 etc) studied and highlighted the regime switching property of stock returns. Both Cheu et al. (1994) and Schaller and Norden (1997) found regime shifts in stock returns. They investigated the relationship between stock returns and their volatility using MSAR and found a nonlinear relationship in returns and volatility. Nishiyima (1998) investigated the existence of different regimes in aggregate stock returns and their mean variance properties in five industrialized countries. He focused more on volatility shifts rather than mean shifts whiles identifying the switching behavior and found consistent volatility based regime shifts in all countries. Similarly Maheu and McCudy (2000) used regime switching model in US stock market. Wang and Theobald (2007) applied MS regression in East Asian countries and found that stock returns in these countries are characterized by two and three regimes. Ismail and Zaidi (2008) examined the regime switching model in more detail in Malaysia. They used non-linear MSAR framework to capture regime shift behavior in Bursa Malaysia. Laha (2006) investigated regime switching behavior in India by using hidden Markov model under Bayesian framework. Kumar (2006) analyzed the weekly data using Markov switching vector error correction model (MS-VECM) and found the existence of two different regimes identified on the basis of stock prices and trading volume. Researchers have also attempted to model a nonlinear structure in time series data. For instance, Turner et al. (1989), Schaller and Norden (1997), Hamilto and Lin (1996) and Gordon and St-Amour (2000) modeled the nonlinear dynamics of stock market volatility. The evidence from developing economies regarding the application of Markov chains is still very limited. Although MC is an emerging technique of modeling stock returns however, to the best of authors' knowledge previous researchers have ignored this method of modeling stock returns especially in emerging markets.

Methodology

By Markov process the present study means a stochastic process {X (t); t ε T} having Markov property for a finite set of points (t_0,t_1,...,t_9). Let X be the finite state space Markov chain with states(1,2,...,10). Given a particular time event t the chain X is in state i and p_{ij} ⁵ denotes the conditional probability that X will be in state j at time t+1 given it was observed in state i at time t. In a similar way p_ij^((n)) represents the probability that X would transit from state i to state j after n transitions, given p_ij^((n))>0. Further, if A is a transition probability matrix of Markov Chain with finite state space then the elements of p_ij^((n)) have ergodic properties. The ergodic properties 6 of Markov chain include irreducibility, aperiodicity and time homogeneity. In short Markov chain is a

⁵ It implicitly assumed here that Markov chain in time homogeneous as the quantity p_ijis independent of time t.

⁶ If (x0, x1....xn) is an irreducible, time homogeneous, discrete space Markov chain, with stationary distribution π , then 1/n $\sum (i=1)^n [f(x, i)] \Rightarrow E[f(x)]$ for any bounded function f in $\in \mathbb{R}$.

process where for every n and t1<t2<t3.....tn, we have:

$$P(x(tn) \le xn | x(t) t \le t n-1) = P(x(tn) \le xn | x(t n-1))$$
(1)

The Markov property implies that the probability distribution of future prices does not follow any particular path which is followed by the price in the past therefore; investor cannot predict the future prices just by observing the past prices. To construct the Markov chain the present study taken the daily KSE100 index of PSE from 2010 to 2015. The log returns are then classified into 10 states based on the range of returns with the difference of 10%. The daily movement of returns for five years is closely observed to identify by the pattern of movement of index from one state of return to another. Before calculating the transition probability matrix, several tests are conducted to ensure the presence of Markov property. Initially, to test the dependence of a state on another, chi square test of independence is conducted. To check the stationarity of states, unit root test is also conducted and finally to verify the Markov property, AR(1) and AR(2) models are estimated ⁷. Based on the transition of one state in to another a frequencies are calculated which are latter used to calculate transition probabilities. The transition state frequency is converted into transition probabilities as below:

If A is the transition probability matrix of an irreducible, aperiodic finite state Markov chain then $\lim_{m \to \infty} A^t = \pi = \begin{bmatrix} n_1 & n_2 & \cdots & p_m \end{bmatrix}$

$$\lim_{t \to \infty} A^{t} = \pi = \begin{bmatrix} p_1 & p_2 & \dots & p_m \end{bmatrix}$$

Where
$$0 < p_i < 1$$
 and $\sum_{i=1}^{m} p_i = 1$ (3)

The Markov chain with above property is said to be ergodic and possesses a limiting distribution π (Baht, 1972). Based on the above mentioned methodology the statistical analysis is conducted and results are presented in the next section. In second phase, Monte Carlo (MC) simulation method is used to simulate random future data. MC model is given as below:

$$I_t = I_{t-1} * e^r \tag{4}$$

Where It is the index at time t, It-1 is the previous value of index and r is the rate of return. Where r consists of drift factors defined by $(\mu - \sigma^2/2)$ at time t and a random variance σW t. Hence,

$$I_t = I_{t-1} * e \left[\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right] \dots (5)$$

⁷ In order to have a Markov property, only AR(1) should be significant and not AR(2)

The above MC equation is used to generate ten series each from Markov chain and actual stock index and the results therefrom are discussed in the next section. *Statistical Analysis*

Initially ten states are designed using a constant range of 0.1% starting from 3% to -4% based on the daily index returns. Table 1 presents the descriptive analysis of ten states. The last three columns show the mean return of each state and volatility measured by standard deviation and coefficient of variation. The descriptive analysis of states shows that 4th and 6th states are highly volatile however it is not clear that which initial state is most probable to be transited in to state 4 or 6. Before analyzing the chain sequence in the mentioned below states it is important to estimates that whether or not the chain possesses the Markov property.

State (S)	State Range (% Returns)	Mean	SD	CV
S ₁	0.030 and above	0.0311	0.0010	0.03215
S_2	Below 0.030 to 0.021	0.0232	0.0012	0.05172
S ₃	Below 0.021 to 0.010	0.0128	0.0026	0.20312
S_4	Below 0.010 to 0.001	0.0043	0.0024	0.5581
S_5	0	0.0000	0.0000	0.0000
S_6	Below 0 to -0.009	-0.0042	0.0027	0.6428
\mathbf{S}_7	Below -0.009 to -0.020	-0.0145	0.0026	0.1793
S_8	Below -0.020 to -0.030	-0.0251	0.0026	0.1035
S_9	Below -0.030 to -0.040	-0.0328	0.0017	0.0518
\mathbf{S}_{10}	Below -0.040	-0.0425	0.0024	0.0564

Table 1.Descriptive Analysis of States

At first step the chi-square (χ^2) test of independence of states given the current state is conducted, the estimated value of the test is significant at the level less than 1% (*i.e.* [$\chi = 497.44$] > $\chi^2_{.001}$) which shows that the identified states are dependent on its first lag, which is a necessary condition for a Markov chain. However, in order to fulfill the Markov chain requirement the identified states must be correlated with its first lag which means that in order to predict the next state of returns the only information required is the knowledge of the current state of return.

Table 2:Test of Stationarity and Auto Regression

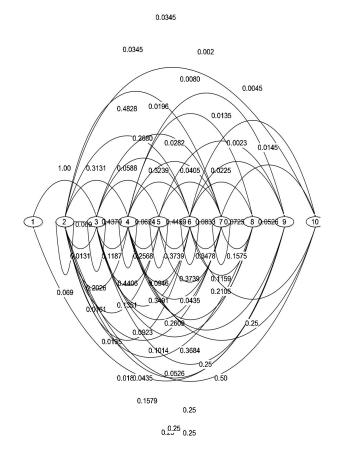
Dependent variable is current state and LAG1 represents the state at time t-1 and LGA2 represents the state at time t-2. To check the stationarity of states Unit root test is conducted which remain significant at level. Unit root t-stat is -31.97871 and p-value is 0.000. Parentheses contain (standard Error) and [t-statistic]. **** shows the level of significance at the level 1% or less.

Variable	Model 1	Model 2
	0.1175***	
LAG 1	(0.0276)	
	[4.2547]	
		0.0105
LAG2		(0.0278)
		[0.3803]
	6.0271***	6.7581***
Constant	(0.1924)	(0.1939)
	[31.30]	[34.850]
R Squared	0.0138	0.001
Adjusted R Squared	0.01305	0.000
Drubin Watson	1.9973	1.7652
F Stat	18.1024	0.1447
Prob F	0.000	0.7037

This assumption is tested using two methods. Firstly the test of autocorrelation using first, second and third lag is applied and secondly the test of auto regression using the 1st, 2nd and 3rd order is applied. The former shows a significant autocorrelation between the current and first lagged value (i.e. AC1=0.12; sig 0.000) and very weak correlation with the second and third lagged value (i.e. AC2=0.01; sig 0.000 and AC3=0.01; sig 0.000) while the latter confirmed that only AR (1) is significant. The results of auto regression are presented in table 2. The results of table 2 confirms the presence of Markov property in the states identified in this study as only the value of first lag is significant. The next step is to construct a state transition matrix. Following is the transition probability matrix 'A' which is measured on the basis of frequency of transition of one state into another. Each component of A is p_ij, where p is the probability of transition of state i into state j, with i,j=1,2,3....10.

$\mathbb{A} = \begin{bmatrix} 0.00\% & 6.90\% & 6.90\% & 31.03\% & 0.00\% & 48.28\% & 3.45\% & 0.00\% & 0.00\% & 3.00\% & 1.31\% & 20.26\% & 43.79\% & 5.88\% & 26.80\% & 1.96\% & 0.00\% & 0.00\% & 0.00\% & 0.00\% & 0.00\% & 0.00\% & 1.61\% & 11.87\% & 44.06\% & 6.24\% & 32.39\% & 2.82\% & 0.80\% & 0.20\% & 0.00\% & 1.35\% & 13.51\% & 25.68\% & 9.46\% & 44.59\% & 4.05\% & 0.00\% & 1.35\% & 0.00\% & 1.80\% & 9.23\% & 34.91\% & 5.41\% & 37.39\% & 8.33\% & 2.25\% & 0.23\% & 0.23\% & 0.00\% & 15.79\% & 5.26\% & 36.84\% & 0.00\% & 21.05\% & 15.79\% & 0.00\% & 5.26\% & 0.00\% & 0.$
--

Based on the transition probability matrix following is the diagrammatic depiction of regime switching Markov chain (Figure 2). The arrows at the top show the probability of transition from high return state towards lower return state while arrows at lower side shows the probability of transition from low return state towards the high return state. The self-directed arrows at the lower side show the probability that the current state shall persist.





For example, M is the initial state space which shows the return on index is in state 1 with p = 1.00 at a particular point in time, and then it will shift from state 1 to state 3 with P=1.00 after first transition (given as M*A, where A is transition matrix given above). After 6th transition the probability of states are presented in (5) below. Following is the output of transition vectors using MatLab.

 $\begin{array}{l} >> M = [1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ >> M * A = [0.0000 & 0.0001 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ >> M * A^2 = [0.0000 & 0.0131 & 0.2026 & 0.4379 & 0.0588 & 0.2680 & 0.0196 & 0.0000 & 0.0000 & 0.0000 \\ >> M * A^3 = [0.0000 & 0.0171 & 0.1286 & 0.3995 & 0.0601 & 0.3357 & 0.0437 & 0.0110 & 0.0023 & 0.0019 \\ >> M * A^4 = [0.0006 & 0.0208 & 0.1194 & 0.3873 & 0.0582 & 0.3425 & 0.0516 & 0.0139 & 0.0030 & 0.0027 \\ >> M * A^5 = [0.0007 & 0.0221 & 0.1184 & 0.3846 & 0.0574 & 0.3431 & 0.0531 & 0.0146 & 0.0031 & 0.0030 \\ >> M * A^6 = [0.0008 & 0.0223 & 0.1183 & 0.3842 & 0.0572 & 0.3431 & 0.0533 & 0.0147 & 0.0031 & 0.0031 \\ \end{array}$

Similarly, if the initial state is S10 with P=1.00 then after 5th transition, the probability distribution of state are given in (6) below:

$$\begin{split} >> N &= \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \\ >> N * A &= \begin{bmatrix} 0.0000 & 0.2500 & 0.0000 & 0.5000 & 0.0000 & 0.2500 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix} \\ >> N * A^2 &= \begin{bmatrix} 0.0000 & 0.0298 & 0.0997 & 0.3852 & 0.0447 & 0.3761 & 0.0435 & 0.097 & 0.0016 & 0.097 \end{bmatrix} \\ >> N * A^3 &= \begin{bmatrix} 0.0004 & 0.0232 & 0.1141 & 0.3856 & 0.0563 & 0.3460 & 0.0536 & 0.0147 & 0.0027 & 0.0034 \end{bmatrix} \\ >> N * A^4 &= \begin{bmatrix} 0.0007 & 0.0225 & 0.1173 & 0.3841 & 0.0571 & 0.3437 & 0.0536 & 0.0148 & 0.0031 & 0.0031 \end{bmatrix} \\ >> N * A^5 &= \begin{bmatrix} 0.0008 & 0.0224 & 0.1180 & 0.3840 & 0.0572 & 0.3432 & 0.0534 & 0.0147 & 0.0031 & 0.0031 \end{bmatrix} \end{split}$$

The probability distribution of states given above is equal to the stationary transition vector T. If probability of return stays in 1st state is 1.00 at time t then there is 38.41% chance that return would be in state 4 and 34.31% chance that return would be in state 6 and so on so forth. Finally, based on the transitional probability matrix expected return of state j is calculated given state i and using these return forecasts future value of KSE index are calculated. Figure 3 shows the comparison of actual index and forecasted index. It is evident from figure 3 that based on Markov chain model some sort of intuition can be taken regarding the stock indices.

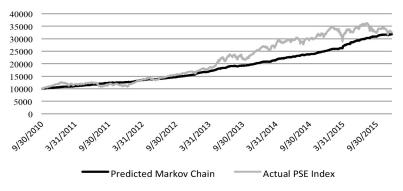
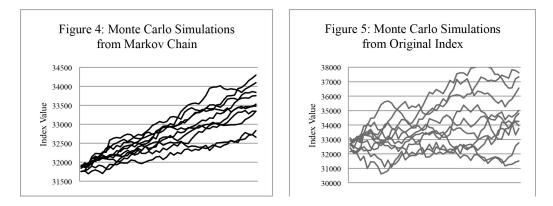


Figure 3: PSE Index (Actual vs Predicted)

Further based on Monte Carlo approach random simulations are run for each of the series i.e. predicted Markov chain and actual KSE index, the results are depicted in the figure 4 and 5 respectively. The smoothness of Markov chain simulation is due to the fact that additional variance has been absorbed in single state and index prediction can be made with lesser volatility.



Conclusion and Direction for Future Research

The present study attempts to investigate the presence of Markov property in PSE using the data of KSE 100 index. Based on the daily index returns ten return states are identified. To identify the Markov property test of dependence of current state on the first lagged state is estimated using AR (1) and AR (2) models. The stock transition from one state to another state is calculated and based on the relative frequency of transitions a state transition matrix is identified. This transition matrix is further employed to calculate the next expected return given a particular state at present. The returns are converted in to predicted index which showed that a Markov chain is suitable for modeling stock indices. The methodology used in this paper is preliminary in nature because the authors did not find any evidence of the use of Markov chain in developing economies especially in Pakistan. Therefore, the authors find it justified to start modeling the index using discrete time finite state MC. Due to its power of capturing behavioral factors in price modeling, Markov chains are emerging as an alternate way of analyzing the time series data. In future, researchers can analyze and model a portfolio of individual stocks instead of stock index. There is still need to improve the method of identification of states in Markov chains. Instead of discrete time finite state models more complex continuous time models can be studies. Last but not least, future researchers can use Markov Chain Monte Carlo (MCMC) method under Bayesian framework, in stock price modeling and forecasting.

References

- Bachelier, L., & Bachelier, L. J. B. A. (1914). Le jeu, la chance et le hasard. E. Flammarion.
- Bae, G. I., Kim, W. C., & Mulvey, J. M. (2014). Dynamic asset allocation for varied financial markets under regime switching framework. *European Journal of Operational Research*, 234(2), 450-458.
- Bhat, U. N., & Miller, G. K. (1972). Elements of applied stochastic processes. J. Wiley.
- Chu, C. S. J., Santoni, G. J., & Liu, T. (1996). Stock market volatility and regime shifts in returns. *Information Sciences*, 94(1-4), 179-190.
- Fama, E. F. (1965). The behavior of stock-market prices. The journal of Business, 38(1), 34-105.
- Goldfeld, S. M., & Quandt, R. E. (1973). A Markov model for switching regressions. Journal of econometrics, 1(1), 3-15.
- Goldfeld, S., & Quandt, R. (1973). The estimation of structural shifts by switching regressions. In *Annals of Economic and Social Measurement, Volume 2, number 4* (pp. 475-485). NBER.
- Gordon, S., & St-Amour, P. (2000). A preference regime model of bull and bear markets. American Economic Review, 90(4), 1019-1033.
- Guidolin, M., & Timmermann, A. (2007). Asset allocation under multivariate regime switching. Journal of Economic Dynamics and Control, 31(11), 3503-3544.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica: *Journal of the Econometric Society*, 357-384.
- Hamilton, J. D., & Lin, G. (1996). Stock market volatility and the business cycle. Journal of Applied Econometrics, 11(5), 573-593.
- Ismail, M. T., & Isa, Z. (2008). Identifying regime shifts in Malaysian stock market returns. *International Research Journal of Finance and Economics*, 15, 44-57.
- Kumar, A. (2008). A Markov-Switching Vector Error Correction Model of the Indian Stock Prices and Trading Volume.
- Maheu, J. M., & McCurdy, T. H. (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics*, 18(1), 100-112.
- Mandelbrot, B. B. (1997). The variation of certain speculative prices. In *Fractals and scaling in finance* (pp. 371-418). Springer, New York, NY.
- Markowitz, H. M. (1959). Portfolio Selection: Efficient Diversification of Investments. Cowles Foundation for Research in Economics, Yale University
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica: *Journal of the Econometric Society*, 347-370.
- Nishiyama, K. (1998). Some evidence on regime shifts in international stock markets. *Managerial Finance*, 24(4), 30-55.
- Schaller, H., & Norden, S. V. (1997). Regime switching in stock market returns. Applied Financial Economics, 7(2), 177-191.
- Turner, C. M., Startz, R., & Nelson, C. R. (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25(1), 3-22.

- Wang, P., & Theobald, M. (2008). Regime-switching volatility of six East Asian emerging markets. Research in International Business and Finance, 22(3), 267-283.
- Yin, G., & Zhou, X. Y. (2004). Markowitz's mean-variance portfolio selection with regime switching: From discrete-time models to their continuous-time limits. *IEEE Transactions on automatic control*, 49(3), 349-360.
- Zhou, X. Y., & Yin, G. (2003). Markowitz's mean-variance portfolio selection with regime switching: A continuous-time model. *SIAM Journal on Control and Optimization*, 42(4), 1466-1482.