

Western  Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

7-31-2012 12:00 AM

Further applications of higher-order Markov chains and developments in regime-switching models

Xiaojing Xi

The University of Western Ontario

Supervisor

Rogemar Mamon

The University of Western Ontario

Graduate Program in Applied Mathematics

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Xiaojing Xi 2012

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Other Applied Mathematics Commons](#)

Recommended Citation

Xi, Xiaojing, "Further applications of higher-order Markov chains and developments in regime-switching models" (2012). *Electronic Thesis and Dissertation Repository*. 678.

<https://ir.lib.uwo.ca/etd/678>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

**FURTHER APPLICATIONS OF HIGHER-ORDER MARKOV
CHAINS AND DEVELOPMENTS IN REGIME-SWITCHING
MODELS**

(Spine title: Higher-order Markov chains and regime-switching models)

(Thesis format: Integrated Article)

by

Xiaojing Xi

Graduate Program in Applied Mathematics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies

The University of Western Ontario

London, Ontario, Canada

© Xiaojing Xi 2012

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Examiners:

Supervisor:

Dr. Adam Metzler

Dr. Rogemar Mamon

Dr. Mark Reesor

Co-Supervisor:

Dr. Hao Yu

Dr. Marianito Rodrigo

Dr. Anatoliy Swishchuk

The thesis by

Xiaojing Xi

entitled:

**Further applications of higher-order Markov chains and developments in regime-switching
models**

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date

Chair of the Thesis Examination Board

Abstract

We consider higher-order hidden Markov models (HMM), also called weak HMM (WHMM), to capture the regime-switching and memory properties of financial time series. A technique of transforming a WHMM into a regular HMM is employed, which in turn enables the development of recursive filters. With the use of the change of reference probability measure methodology and EM algorithm, a dynamic estimation of model parameters is obtained. Several applications and extensions were investigated. WHMM is adopted in describing the evolution of asset prices and its performance is examined through a forecasting analysis. This is extended to the case when the drift and volatility components of the logreturns are modulated by two independent WHMMs that do not necessarily have the same number of states. Numerical experiment is conducted based on simulated data to demonstrate the ability of our estimation approach in recovering the “true” model parameters. The analogue of recursive filtering and parameter estimation to handle multivariate data is also established. Some aspects of statistical inference arising from model implementation such as the assessment of model adequacy and goodness of fit are examined and addressed. The usefulness of the WHMM framework is tested on an asset allocation problem whereby investors determine the optimal investment strategy for the next time step through the results of the algorithm procedure. As an application in the modeling of yield curves, it is shown that the WHMM, with its memory-capturing mechanism, outperforms the usual HMM. A mean-reverting interest rate model is further developed whereby its parameters are modulated by a WHMM along with the formulation of a self-tuning parameter estimation. Finally, we propose an inverse Stieltjes moment approach to solve the inverse problem of calibration inherent in an HMM-based regime-switching set-up.

Keywords: Higher-order Markov chain, filtering, change of reference probability method, asset price modeling, forecasting, asset allocation, multivariate data, term structure of interest rates, inverse problem in finance

Co-Authorship Statement

I hereby declare that this thesis incorporates materials that are direct results of several joint research collaborations. My research outputs with my supervisor, Dr. Marianito Rodrigo (co-supervisor) and Dr. Matt Davison have already led to published papers and manuscripts under review in various peer-reviewed journals and refereed book chapters. These are detailed below.

The content of chapter 2 appeared as a published article (co-authored with my supervisor, Dr. Rogemar Mamon) in the journal *Economic Modelling*; see reference [25] of chapter 3.

Chapter 3 is currently being converted into a manuscript (co-authored with Dr. Mamon) and will be submitted for publication in the journal *Systems and Control Letters*.

Materials of chapter 4 are based on a manuscript (co-authored with Drs. Mamon and Davison) that is presently under revision for the *Journal of Mathematical Modelling and Algorithms*. This chapter already includes modifications to the original submission in an effort to address comments and suggestions of two referees.

Chapter 5 is based on a manuscript (co-authored with Dr. Mamon) that was submitted (by invitation) as a book chapter to a peer-reviewed monograph “*State-Space Models and Applications in Economics and Finance*” edited by S. Wu and Y. Zeng and will be featured as part of the new Springer Series *Statistics and Econometrics for Finance*.

Chapter 6 originated from a paper that is currently under review in the journal *Statistics and Computing*.

The results of chapter 7 are contained in a paper that was already accepted in a refereed monograph “*Advances in Statistics, Probability and Actuarial Science*” edited by S. Cohen,

D. Madan, T. Siu and H. Yang, and will be published by World Scientific.

As the first author of all the papers emanating from this research, I was in-charge of model implementation, filtering algorithm developments, data collection and analysis, literature review and completing the first draft of the manuscripts. With the exception of my supervisor and co-supervisors guidance on modeling framework formulations, I certify that this dissertation is fully a product of my own work.

Acknowledgments

I would like to sincerely thank all of the people who helped in the various stages of this research. First and foremost, I would like to gratefully acknowledge my supervisor Dr. Rogemar Mamon for his guidance and indispensable support in completing this research. I greatly admire him for his accessibility and patience, his professionalism and scientific insight. His experience, suggestions and encouragement were priceless. I am also appreciative of the help and support of my co-supervisor Dr. Marianito Rodrigo.

Sincere thanks are due to Dr. Matt Davison, Dr. Adam Metzler and Dr. Mark Reesor for leading the Financial math group and providing a productive and fun research environment. I wish to extend my warmest thanks to my course instructors Dr. Rob Corless and Dr. Greg Reid. Their kind support and guidance have been of great value in this study. The departmental staff has also been tremendous; my sincere thanks to Audrey Kager and Karen Foullong for answering all my questions and resolving issues of concern. I am grateful to my current and former colleagues in the PhD program for making the dark, depressing hours seem brighter.

Finally my deepest gratitude goes to my family members, specially my parents for their constant supports and encouragements throughout my life. I would like particularly to extend my sincere gratitude to my loving husband for giving me strength during hard times. Without all your loves, encouragements, and understanding it would never be possible for me to continue my studies.

Table of Content

Certificate of Examination	ii
Abstract	iii
Co-Authorship Statement	iv
Acknowledgments	vi
List of Figures	xii
List of Tables	xiii
1 Introduction	1
1.1 Background and motivation	1
1.2 Research objectives	3
1.3 Literature	4
1.4 Structure of the thesis	5
References	9
2 Parameter estimation of an asset price model driven by a WHMM	12
2.1 Introduction	12
2.2 Description of a weak hidden Markov model	15
2.3 Change of reference probability measure	18
2.4 Calculation of recursive filters	20
2.5 Parameter estimation	23

2.6	Numerical results	25
2.7	Conclusion	35
	References	37
3	Parameter estimation in a WHMM setting with independent drift and volatility components	40
3.1	Introduction	40
3.2	Model background	43
3.3	Filters and parameter estimation	46
3.4	A simulation study	52
3.5	Conclusion	56
	References	57
4	A weak hidden Markov chain-modulated model for asset allocation	60
4.1	Introduction	60
4.2	Filtering and parameter estimation	66
4.3	Forecasting indices	73
4.4	A switching investment strategy	79
4.5	A mixed investment strategy	83
4.6	Evaluating the portfolio performance	86
4.7	Conclusion	97
	References	99
5	Yield curve modeling using a multivariate higher-order HMM	103
5.1	Introduction	103
5.2	Filtering and parameter estimation	108
5.3	Implementation	114
5.4	Forecasting and error analysis	120
5.5	Conclusion	125

References	127
6 An interest rate model incorporating memory and regime-switching	131
6.1 Introduction	131
6.2 Model construction	136
6.3 Filters and parameter estimation	140
6.4 Implementation	145
6.5 Forecasting and error analyses	153
6.6 Conclusion	160
References	162
7 Parameter estimation of a regime-switching model	167
7.1 Introduction	167
7.2 Regime-switching model setup	170
7.3 Derivation of a system of Dupire-type PDEs	173
7.4 Inverse Stieltjes moment problem	175
7.5 Numerical implementation and results	178
7.6 Implementation to “practical data”	186
7.7 Conclusion	192
References	195
8 Concluding remarks	198
8.1 Summary and commentaries	198
8.2 Further research directions	200
References	203
A Proof of equation (2.16)	204
B Proof of Proposition 2.4.1	205
B.1 Proof of equation (2.22)	205

B.2	Proof of equation (2.23)	206
B.3	Proof of equation (2.24)	207
C	Proof of Proposition 2.5.1	208
C.1	Proof of equation (2.27)	208
C.2	Proof of equation (2.28)	209
C.3	Proof of equation (2.29)	210
D	Proof of Proposition 6.3.2	212
D.1	Proof of equation (6.25)	212
D.2	Proof of equation (6.26)	213
D.3	Proof of equation (6.27)	214
E	Proof of equation (6.36)	216
	Curriculum Vitae	218

List of Figures

2.1	Evolution of estimates for \mathbf{f} , σ and \mathbf{A} -matrix under the 2-state setting	28
2.2	Evolution of estimates for the transition probabilities under the 3-state setting	29
2.3	Evolution of estimates for \mathbf{f} and σ under the 3-state setting	30
3.1	Evolution of parameter estimates under different model settings	55
4.1	Evolution of parameter estimates under the 2-state setting	76
4.2	Actual data and one-step ahead forecasts for Russell 3000 growth and value indices (left), and zoom-in view for the period Jul 02 – Dec 05 (right)	77
4.3	Actual returns and one-step ahead forecasts for returns of Russell 3000 growth (left) and value (right) indices	78
4.4	AIC, AICc and BIC values for the 1-, 2- and 3-state models	80
4.5	Numbers of the intervals switching strategy has the highest and the second highest terminal values for varying transaction cost	82
4.6	Optimal weights for Russell 3000 value and growth indices in the WHMM-based mixed strategy with $\nu = 0.08$	85
4.7	Evolution of optimal weights for Russell 3000 growth index in the WHMM-based mixed strategy with varying ν 's	85
4.8	Switching, mixed, pure growth and pure value strategies comparison between 1995 and 2010	87
5.1	Evolution of estimates for transition probabilities through algorithm steps under the 2-state setting	117

5.2	Evolution of estimates for f through algorithm steps under the 2-state setting . . .	118
5.3	Evolution of estimates for σ through algorithm steps under the 2-state setting . . .	119
5.4	AIC for the 1-, 2-, 3- and 4-state models	122
5.5	One-step ahead predicted values (%) versus actual Treasury yields (%) under a 2-state WHMM setting	124
6.1	Evolution of estimates for transition probabilities under the 2-state setting . . .	148
6.2	Evolution of parameter estimates under the 2-state setting	149
6.3	Evolution of parameter estimates under the 3-state setting	151
6.4	Plot of actual and one-step ahead forecasts in a 2-state WHMM	154
6.5	Plots of AIC values for 1-,2-,3- and 4-state WHMMs	159
7.1	Actual and estimated call prices: $\lambda_1 = 0.25, \lambda_2 = 2$ and $n = 2$	188
7.2	Actual and estimated call prices: $\lambda = 2$ and $n = 7$	192
7.3	Actual and estimated call prices: $\lambda_1 = 0.25, \lambda_2 = 1$ and $n = 7$	194

List of Tables

2.1	Segregation of the period of actual data into two states	26
2.2	Segregation of the period of actual data into three states	26
2.3	Range of SEs for each parameter under the 1-, 2- and 3-state settings	31
2.4	Error analysis of WHMM and HMM models under the 2-state setting	33
2.5	Error analysis of WHMM and HMM models under the 3-state setting	34
3.1	Comparison of 1-step ahead forecast errors	56
3.2	Comparison of 5-step ahead forecast errors	56
4.1	Summary statistics of Russell 3000 growth returns	74
4.2	Summary statistics of Russell 3000 value return	74
4.3	Summary statistics of Russell 3000 return	74
4.4	Error measures for one-step ahead forecasts under 1-, 2- and 3-state WHMM set-ups	78
4.5	Performance comparison for WHMM- and HMM-based switching strategies with varying transaction costs.	82
4.6	Performance comparison between WHMM- and HMM-based mixed strategies with varying transaction costs.	86
4.7	Sharpe ratio for five investment strategies using 15 intervals. Numbers inside the parentheses are standard errors.	88
4.8	Jensen's alpha for four investment strategies using 15 intervals. Numbers inside the parentheses are standard errors.	90

4.9	AR for four investment strategies using 15 intervals. Numbers inside the parentheses are standard errors.	91
4.10	p -values for the Jarque-Bera test of normality on data given in Tables 4.7 - 4.9	93
4.11	p -values for a one-tailed significance test on the performance results shown in Tables 4.8 - 4.9	93
4.12	p -values for a Wilcoxon rank sum test on the performance results shown in Tables 4.7 - 4.9	93
4.13	Performance evaluation for 10000 bootstrapped datasets with 5bps transaction cost	95
4.14	Performance evaluation for 10000 bootstrapped datasets with 30bps transaction cost	96
5.1	Descriptive summary statistics and data segregation into two states	115
5.2	Segregation of data into three states	115
5.3	Parameter estimates at the end of final algorithm step for $N = 3$	117
5.4	RMSE for one-step ahead predictions versus actual values	122
5.5	Error analysis of WHMM and HMM models under the 1-state setting	125
5.6	Error analysis of WHMM and HMM models under the 2-state setting	125
5.7	Error analysis of WHMM and HMM models under the 3-state setting	125
5.8	Error analysis of WHMM and HMM models under the 4-state setting	126
6.1	Possible segregation of data into 2 states	146
6.2	Possible segregation of data into 3 states	146
6.3	Estimates of H under different estimators	147
6.4	Range of SEs for each parameter under the 1-, 2-, 3- and 4-state settings	152
6.5	Error analysis for 2-state setting	157
6.6	Error analysis for 3-state setting	157
6.7	Error analysis for 4-state setting	158

6.8	p -values for a one-tailed significance test on the comparison of WHMM states based on forecasting errors shown in Tables 6.5-6.7	158
6.9	p -values for a one-tailed significance test on the comparison of HMM and WHMM based on forecasting errors shown in Tables 6.5-6.7	158
7.1	Example 1: Estimated parameters for different λ and n with $\sigma_1 = 0.1$ and $\sigma_2 = 0.3$	185
7.2	Example 2: Estimated parameters for different λ and n with $\sigma_1 = 0.1$ and $\sigma_2 = 0.3$	187
7.3	Example 1: Estimated parameters for different λ and n	191
7.4	Example 2: Estimated parameters for different λ and n	193

Chapter 1

Introduction

1.1 Background and motivation

A higher-order hidden Markov model (HMM) or the so-called weak HMM (WHMM) is an extension of the usual HMM in which the hidden process (i.e., not observed) is a higher-order Markov chain. WHMMs are also known as hidden semi-Markov models or variable-duration HMM in other areas of engineering and the physical sciences.

A WHMM is a Markov chain model that is dependent on prior states. Hence, the higher the order, the greater is the dependency and so more information about the past is captured by this type of model. Solberg [20] comments on the idea of higher-order Markov chain and states “The real significance of higher-order Markov chain is to establish that the Markov assumption is not really as restrictive as it first appears”. Barbu and Limnios [1] remark “The main drawback of HMMs comes from the Markov property, which requires that the sojourn time in a state be geometrically distributed. This makes the HMMs too restrictive from a practical point of view. Thus, [with WHMM] we have a model that combines the flexibility of a semi-Markov process with the modeling capacity of HMMs”. This cognizance from experts in the area of Markov chain modeling highlights the advantages obtained when longer past state sequence is taken into account into HMMs.

The memoryless property of the original HMM can be formulated as follows. If the past and current information of a process are known, the statistical behavior of future evolution of the process is determined by its present state, and therefore, the states of the past and the future are conditionally independent. Nonetheless, there are many situations in real life where HMMs memoryless property seems unwarranted and indefensible. For instance, the presence of memory in asset prices, interest rates and other time series of financial variables is well documented. The HMM can capture more information from the past by weakening its Markovian hypothesis and extending the dependency to any number of prior epochs, thus giving rise to WHMMs. Such models are certainly appropriate for financial time series where memories are evident.

In the majority of WHMM applications, parameter estimation is at the core of its implementation. The estimation procedure for WHMM is much more complicated than the observable weak Markov chain (WMC). The underlying WMC in WHMM is neither observed nor can be measured directly. Instead, we are given the evolution in time of the observations distorted in noise. From an engineering perspective, one may view the observed process as a received signal and the hidden WMC as an emitted signal. In WHMM, the number of involved model parameters increases exponentially as the number of states and length of order increase. Needless to say, the large number of parameters complicates the estimation procedure and increases the computational burden. Thus, designing efficient computing algorithms which can facilitate the estimation is desired. On the other hand, accurate forecasting of financial variables is an important consideration in the application of WHMM to many types of decision-making endeavors. Theoretically speaking, since WHMM can capture more historical information of the unobservable market state, it should outperform the usual HMM on data fitting provided there is presence of memory in the data-generating process. It is therefore worthwhile to investigate if advantages and benefits of employing WHMM exist in the context of practical financial applications.

1.2 Research objectives

To tackle some of the major aforementioned problems above, it is the fundamental theme of this work to widen the literature on the applications of WHMMs capable of capturing the memory property in financial time series. This research comprises of both theoretical and numerical investigations. The main objectives and scope of this thesis are as follows:

- Develop a methodology to estimate parameters of WHMM: The higher-order Markov chain is transformed into a regular Markov chain. Signal filtering techniques are then utilized to filter out the hidden signal. Recursive filters are derived for the state of the Markov chain and other auxiliary processes related to the Markov chain. With the EM algorithm, we provide recursive estimates for the parameters of several financial models.
- Demonstrate the accessibility and applicability of the proposed models and estimation techniques: Recursive filters and estimation methods are implemented on simulated data and market data to recover model parameters. The developed algorithms are run on batches of data to reduce computational expenses.
- Evaluate the performance of WHMM in fitting and forecasting and address related statistical inference issues in model implementation: The short-term forecasts under WHMM setting are compared to those from regular HMM using different performance metrics. As well, statistical tests are applied to determine significance of results under various financial modeling considerations.
- Illustrate the benefits and accurate modeling of WHMM to financial applications: We consider the modeling of asset prices involving both univariate and multivariate data and the term structure of interest rates. Application of WHMM to asset allocation is also examined.
- Development of an estimation technique for calibration of regime-switching models: This thesis aims to contribute to the further development of regime-switching models

by constructing a method that will compute model parameters given market price data. Such a problem is a central concern in option pricing and hedging.

1.3 Literature

This section surveys available literature on higher-order HMMs sketching a backdrop against which this study has taken place. Note that in the context of a particular application, more specific perspective on the relationship of our research to existing literature on regime-switching models, original HMMs and WHMMs shall be provided in the beginning of each succeeding chapter.

The theory and algorithms pertaining to WHMM were first enriched by the application of WHMM in speech recognition. WHMM-based approach was first proposed by Ferguson [8]. In his pioneering work, such approach was called explicit-duration HMM. In contrast to the implicit duration of HMMs, the state of duration is dependent on the current state of the underlying higher-order Markov process. Russel and Moore [16] investigated WHMM in using a Poisson distribution to model duration. Levison [13] further explored the model with continuous duration by employing a gamma distribution in the modeling of speech segment durations.

Guédon and Coccozza-Thivent [9] proposed the adoption of the EM algorithm to WHMMs in estimating the duration parameters. In their work, WHMMs were presented with state occupancy modeled by a gamma distribution as put forward in [13] and the observation process is modeled by a mixed Gaussian distribution. Kriouile, et al. [10] derived an extended Baum-Welch re-estimation algorithm for second-order discrete HMMs. Ferguson [8] pointed out that the state and the duration time in one state of a WHMM can be embedded into a complex state of HMM. Ferguson's idea was exploited in Krishnamurthy, et al. [11] by reformulating a higher-order scalar state into a first-order 2-vector HMM. With such a reformulation, signals and parameters can be estimated by the Baum-Welch algorithm. Since then, similar approaches

started to appear in the literature including Ramesh and Wilpon [15] and Sin and Kim [19] using Viterbi algorithm.

As previously mentioned, computational complexity is a common problem in the applications of WHMM mainly due to the large number of parameters. This drew researchers' attention to construct efficient estimation algorithms for WHMMs. Du Preez [5] developed a Fast Incremental Training algorithm which can reduce a WHMM with any order to a regular HMM. Yu and Kobayashi [22] proposed a forward-backward algorithm in which the notion of a state together with its remaining sojourn time is used to define the forward-backward variables. Bulla, et al. [3] developed a software package for the statistical software **R**, which allows for the simulation and maximum likelihood estimation of WHMMs. Overfitting is another issue that may be dealt with cross-validate technique, see Elliott, et al. [6].

Since the 1990s, the WHMM was applied to various fields including electrocardiography by Thoraval, et al. [21], hand writing recognition by Kundu, et al. [12], genes recognition in DNA by Burge and Karlin [4], among others. In recent years starting in 2000, the applications of WHMM have been increasing with the advent of technological advancements. They were widely applied in areas of growing human interests such as wireless internet traffic (Yu, et al. [23]), protein structure prediction (Schmidler, et al. [18]), rain event time series (Sansom and Thomson [17]), MRI sequence analysis (Faisan, et al. [7]), financial time series (Bulla and Bulla [2]) and classification of musics (Liu, et al. [14]).

1.4 Structure of the thesis

This thesis is composed of eight chapters including this Introduction. The rest of the material are compilations of the related research outputs on WHMMs and regime-switching models. The contents of the subsequent chapters are detailed below.

In chapter 2, we introduce the concept of WHMM in an attempt to capture more accurately the evolution of a risky asset. The logreturns of assets are modulated by a WMC with finite state space. In particular, the optimal states estimates of the second-order Markov chain and parameters estimates of the model are given in terms of the discrete-time filters for the state of the Markov chain, the number of jumps, occupation time and auxiliary processes. We provide a detailed implementation of the model a financial time series dataset along with the analysis of the h -step ahead forecasts. The results of our error analysis suggest that within the dataset studied and considering longer predictive horizons, WHMM gives a better forecasting performance than the traditional HMM.

An extension of the WHMM for logreturns of assets in which the drift and volatility are governed by two independent WMCs is given in chapter 3. A detailed example is provided to demonstrate the transformation of an extended WHMM into a regular WHMM. Filtering methods and EM algorithm are implemented on simulated data to recover the “true” parameters. Error analyses of the h -step ahead predictions are provided to assess model performance for different combination of states.

In chapter 4, we present an analysis of asset allocation strategies when the asset returns are driven by a discrete-time WHMM. The “switching” and “mixed” strategies are studied. We use a multivariate filtering approach in conjunction with the EM algorithm to obtain estimates of model parameters. This, in turn, aids investors in determining the optimal strategy for the next time step. Numerical implementation is carried out on the Russell 3000 value and growth indices data. The respective performances of portfolios under particular trading strategies are benchmarked against three classical investment measures.

A multivariate higher-order Markov model for the structure of interest rates is developed in chapter 5. The multivariate filtering technique and EM algorithm are adopted to obtain optimal estimates of model parameters. We assess the goodness of fit of the one-step-ahead forecasts

and apply the Akaike information criterion (AIC) in finding the optimal number of economic regimes. The filtering algorithms were implemented on a dataset consisting of approximately 3 years of daily US-Treasury yields. Our empirical results show that based on the AIC and root-mean-square error metric, a two-state WHMM is deemed as the most appropriate in describing the term structure dynamics within the dataset and period of study. Moreover, an analysis of the h -day ahead predictions generated from WHMM is compared with those generated from the regular HMM. By including a memory-capturing mechanism, the WHMM outperforms the HMM in terms of low forecasting errors.

In chapter 6, an Ornstein-Uhlenbeck interest rate model whose mean-reverting level, speed of mean reversion and volatility are all modulated by a WMC. We derive the filters of the WMC and other auxiliary processes through a change of reference probability measures. Optimal estimates of model parameters are computed by employing the EM algorithm. We examine the h -step ahead forecasts under our proposed set-up and compare them to those under the usual Markovian regime-switching framework. Our numerical results generated from the implementation of WMC-based filters on a ten-year dataset of weekly short-term maturity Canadian yield rates give better goodness-of-fit performance than that from the HMM, and indicate that a two-state WMC is adequate to model the data.

In chapter 7, we address the problem of model calibration under a regime-switching model setting. A method is proposed to recover the time-dependent parameters of the Black-Scholes option pricing model when the underlying stock price dynamics are modeled by a finite-state continuous-time Markov chain. The coupled system of Dupire-type partial differential equations is derived and formulated as an inverse Stieltjes moment problem. A numerical illustration is included to show how to apply our proposed method on financial data. The accuracy of the calculation is examined and sensitivity analyses are undertaken to study the behavior of the estimated results when model parameters are varied.

A summary of the findings of the thesis as well as possible future works motivated by this research is presented in chapter 8.

References

- [1] V. Barbu and N. Limnios. *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications*. Springer, New York, 2008.
- [2] J. Bulla and I. Bulla. Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics and Data Analysis*, 51:2192–2209, 2006.
- [3] J. Bulla, I. Bulla, and O. Nenadi. hsmm-An **R** package for analyzing hidden semi-Markov models. *Computational Statistics and Data Analysis*, 54:611–619, 2010.
- [4] C. Burge and S. Karlin. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, 268:78–94, 1997.
- [5] J. A. du Preez. *Efficient high-order hidden Markov modeling*. PhD thesis, University of Stellenbosch, 1997.
- [6] R. J. Elliott, W. C. Hunter, and B. M. Jamieson. Financial signal processing: a self calibrating model. *International Journal of Theoretical and Applied Finance*, 4:567–584, 2003.
- [7] S. Faisan, L. Thoraval, J. P. Armspach, and F. Heitz. Hidden semi-Markov event sequence models: application to brain functional MRI sequence analysis. In *International Conference on Image Processing*, volume 1, pages 880–883, Rochester, NY, 2002.

- [8] J. D. Ferguson. Variable duration models for speech. In *Symp. Application of Hidden Markov Models to Text and Speech*, pages 143–179, Princeton, NJ, 1980. Institute for Defense Analyses.
- [9] Y. Guédon and C. CoCozza-Thivent. Explicit state occupancy modeling by hidden semi-Markov models: application of Derin’s scheme. *Computer Speech and Language*, 4:167–192, 1990.
- [10] A. Kriouile, J. F. Mari, and J. P. Haton. Some improvements in speech recognition based on HMM. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 545–548, Albuquerque, NM, 1990.
- [11] V. Krishnamurthy, J. B. Moore, and S. H. Chung. Hidden fractal model signal processing. *Signal Processing*, 24:177–192, 1991.
- [12] A. Kundu, Y. He, and M. Y. Chen. Efficient utilization of variable duration information in HMM based HWR systems. In *International Conference on Image Processing*, volume 3, pages 304–307, Santa Barbara, CA, 1997.
- [13] S. E. Levinson. Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech and Language*, 1:29–45, 1986.
- [14] X. B. Liu, D. S. Yang, and X. O. Chen. New approach to classification of Chinese folk music based on extension of HMM. In *International Conference on Audio, Language, and Image Processing*, pages 1172–1179, Shanghai, China, 2008.
- [15] P. Ramesh and J. G. Wilpon. Modeling state durations in hidden Markov models for automatic speech recognition. In *IEEE international conference on Acoustics, speech and signal processing*, pages 381–384, Washington, DC, 1992.

- [16] M. J. Russel and R. K. Moor. Explicit modeling of state occupancy in hidden Markov models for automatic speech recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5–8, Tampa, FL, 1985.
- [17] J. Sansom and P. Thomson. Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability*, 38:142–157, 2001.
- [18] S. C. Schmidler, J. S. Liu, and D. L. Brutlag. Bayesian segmentation of protein secondary structure. *Journal of Computational Biology*, 7:233–248, 2000.
- [19] B. Sin and J. H. Kim. Nonstationary hidden Markov model. *Signal Processing*, 46:31–46, 1995.
- [20] J. Solberg. *Modeling Random Processes for Engineers and Managers*. John Wiley & Sons, New Jersey, 2009.
- [21] L. Thoraval, G. Carrault, and F. Mora. Continuously variable duration hidden Markov models for ECG segmentation. *Engineering in Medicine and Biology Society*, 4:529–530, 1992.
- [22] S. Z. Yu and H. Kobayashi. An efficient forwardbackward algorithm for an explicit duration hidden Markov model. *IEEE Signal Processing Letters*, 10:11–14, 2003.
- [23] S. Z. Yu, B. L. Mark, and H. Kobayashi. Mobility tracking and traffic characterization for efficient wireless internet access. In *Multiaccess, Mobility and Teletraffic for Wireless Communications*, pages 279–290, Norwell, MA, 2000. Kluwer Academic Publishers.

Chapter 2

Parameter estimation of an asset price model driven by a WHMM

2.1 Introduction

In this chapter, we introduce the concept of higher-order HMM or WHMM and how it extends the regime-switching framework. The key ideas are presented including the notation and rationale for building financial models enriched by WHMM.

In financial modeling, it is well known that the parameters of a model for the evolution of financial data tend to change over time. Various Markov-switching models have been proposed to describe the behavior of business cycles or volatility regimes. The idea of regime-switching models can be traced back to the early works of Quandt [17] and Quandt and Goldfeld [11]. In an influential paper, Hamilton [13] puts forward Markov-switching methods in the modeling of non-stationary time series. Turner, et al. [21] argue that in a model, either the mean or variance, or both may exhibit differences between two regimes. Chu, et al. [3] apply a Markov-switching model to market returns and examine the variation in volatility for different return regimes. The results of their analysis show that the stock returns are best characterized by a model containing six regimes. Bollen, et al. [1] introduce a regime-switching model with independent shifts in

mean and variance and examine its ability to capture the dynamics of foreign exchange rate.

The mathematical challenge akin to regime-switching models largely boils down to obtaining the optimal estimation of the required number of parameters and the parameters themselves, which are governed by a discrete-time Markov chain. A previous study by Elliott, et al. [4], provides not only recursive estimates of the Markov chain but also continual, recursively self-updating estimates for all parameters of the model. HMM filtering methods are quite popular in statistics and engineering and have been widely applied to financial problems. Elliott and van der Hoek [7] adopt an HMM filtering-based method in the examination of an asset allocation problem. More recently, Erlwein, et al. [9] develop and analyze investment strategies relying on HMM approaches. In Elliott, et al. [5], HMM filtering techniques are applied to an interest rate model and an explicit expression for the price of zero-coupon bonds is provided. Furthermore, Erlwein and Mamon [8] derive and implement the filters of a Hull-White interest rate model in which the interest rate's volatility, mean-reverting level and speed of mean-reversion are governed by a Markov chain in discrete time. The investigation of Elliott, et al. [6], based on the gauge transformation, provides a robust form of filtering equations which offers substantial improvement over classical filtering by avoiding numerical approximations to stochastic integrals, for a continuous-time HMM.

In recent years, there has been a considerable attention in financial time series that are observed to possess memories and usually modeled by stochastic evolution equations. While the traditional HMM already brings a certain degree of modeling sophistication since it is able to capture the switching of parameters between regimes, it is felt that the usual Markov assumption is inadequate. It is for this reason that a WHMM is appropriate when memories are present. For instance, the second-order Markov chain has the effect of having the next state dependent on the two prior states. Of course, the higher the order of the chain, the more extended the dependency and consequently, more information from the past is incorporated into

the Markov chain model. As mentioned in Solberg [20], the real significance of higher-order Markov chains is the demonstration that the Markov assumption is not really as restrictive as it first appears. One is not limited to a dependence on just one previous time epoch. In principle, the dependency can be extended to any number of prior epochs. Obviously, the drawback is that there is a practical price to pay for the enlargement of the number of states and the estimation of parameters becomes more involved.

Recursive filtering equations for a discrete-time WHMM with finite state space and *discrete-range* observations are derived in Luo and Tsoi [14]. These filters are used to re-estimate the parameters of the model. An application of WHMM in risk measurement of a risky portfolio can be found in the paper of Siu, et al. [19], who also examine the higher-order effect of the underlying Markov chain via backtesting. In Siu, et al. [18], a method to recover spot rates and credit ratings is developed using a double higher-order HMM. For valuation of derivatives, Ching, et al. [2] investigated the problem of pricing exotic options under a WHMM setting.

In this chapter, we introduce a WHMM-modulated model for the logreturns of a risky asset. More specifically, we assume that the the mean and variance of the logreturns from the risky asset are governed by a discrete-time, finite-state WHMM. We first derive the filters for the discrete-time, *continuous-range* observations and obtain the optimal estimates for the parameters. Second, we test the applicability and effectiveness of the WHMM in capturing the empirical features of stock index data, S&P500, spanning the period 1997-2010. Third, in terms of evaluating the model's predictability performance, we compare our WHMM with the regular HMM under different numbers of states to ascertain the benefits derived from the proposed WHMM-based asset price model.

This chapter is structured as follows. Section 2.2 gives the WHMM formulation. We introduce the change of reference probability technique in Section 2.3, which forms the underpinnings of the filtering and estimation employed in this chapter. Sections 2.4 and 2.5 set out the details of

the steps for the filtering method and the optimal recursive parameter estimation. An empirical investigation involving a data set along with the forecasting analysis is presented in section 2.6. Finally, section 2.7 concludes.

2.2 Description of a weak hidden Markov model

Owing to its simplicity and along with the fact that any diffusion can be approximated by a Markov chains, the theory of Markov chain has found abundant applications in the modeling of complex and dynamical systems including the financial market. An accessible and brief survey of Markov chains and an account of its ubiquity in several branches of science are given in Haigh [12]. In addition to the objectives specified in section 2.1, it is the intent of this paper to illustrate the usefulness of higher-order hidden Markov chains in economic modeling highly intertwined to the interest of financial analysts and engineers.

In engineering, for example, the charge, $Q(t)$, at time t at a fixed point in an electrical circuit is of interest. However, due to error in the measurement of $Q(t)$, it cannot really be measured but rather just a noisy version of it. The aim is to filter the noise out of our observations. Similarly, in financial economics, we wish to answer the question if financial data such as asset prices and stock indices contain information about latent variables? If so, how might their behavior in general and in particular their dynamics be estimated? We shall present a methodology to address this problem.

In the succeeding discussion, all vectors will be denoted by bold letters in lowercase and all matrices will be denoted by English or Greek letters in uppercase. We assume all stochastic processes are defined on a complete probability space (Ω, \mathcal{F}, P) , where P is a real-world probability. Let $\mathbf{x} = \{\mathbf{x}_k\}_{k \geq 0}$ be a discrete-time Markov chain with N states. We associate the state space of \mathbf{x}_k with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\} \subset \mathbb{R}^N$, where the \mathbf{e}_i 's are unit vectors in \mathbb{R}^N with unity in the i th element and zero elsewhere. We use $\langle \mathbf{x}_k, \mathbf{e}_i \rangle = 1_{\{\mathbf{x}_k = \mathbf{e}_i\}}$ to emphasize

the identification of \mathbf{x}_k with the canonical basis, where $\langle \mathbf{b}, \mathbf{c} \rangle$ represents the Euclidean scalar product in \mathbb{R}^N of the vectors \mathbf{b} and \mathbf{c} . The state process \mathbf{x} may represent the state of an economy. If $N = 3$ for example, $\langle \mathbf{x}_k, \mathbf{e}_1 \rangle$, $\langle \mathbf{x}_k, \mathbf{e}_2 \rangle$ and $\langle \mathbf{x}_k, \mathbf{e}_3 \rangle$ represent the “best”, “second-best” and “worst” economic state, respectively. We suppose \mathbf{x}_0 is given, or its distribution known.

We say that process \mathbf{x} is a weak Markov chain of order $n \geq 1$, if its value at the time $k + 1$ depends on its value in the previous n time steps. That is,

$$\begin{aligned} P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \mathbf{x}_1 = x_1, \dots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k) \\ = P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_{k-n+1} = x_{k-n+1}, \dots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k). \end{aligned} \quad (2.1)$$

When $n = 1$, the usual or regular Markov chain is recovered.

Remark: *To simplify the discussion and present a complete characterization of the parameter estimation, we only concentrate on a weak Markov chain of order 2.*

Under the second-order Markov chain, we have

$$\begin{aligned} P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \mathbf{x}_1 = x_1, \dots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k) \\ = P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k). \end{aligned} \quad (2.2)$$

Write

$$a_{lmv} := P(\mathbf{x}_{k+1} = \mathbf{e}_l | \mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v), \quad (2.3)$$

where $l, m, v \in \{1, 2, \dots, N\}$. Then we have the associated $N \times N^2$ transition matrix

$$\mathbf{A} = \begin{pmatrix} a_{111} & a_{112} & \dots & a_{11N} & \dots & a_{1N1} & \dots & a_{1NN} \\ a_{211} & a_{212} & \dots & a_{21N} & \dots & a_{2N1} & \dots & a_{2NN} \\ & & \dots & & & & \dots & \\ a_{N11} & a_{N12} & \dots & a_{N1N} & \dots & a_{NN1} & \dots & a_{NNN} \end{pmatrix}.$$

Let y_k , $k \geq 1$, denote the observation process which is the sequence of logreturns of asset prices. It has to be noted that we do not observe \mathbf{x} from the financial market directly. Instead, there exists a function h such that

$$y_{k+1} = h(\mathbf{x}_k, z_{k+1}) = f(\mathbf{x}_k) + \sigma(\mathbf{x}_k)z_{k+1}, \quad k \geq 1. \quad (2.4)$$

The $\{z_k\}_{k \geq 1}$ in equation (2.4) is a sequence of independent identically distributed (IID) standard normal random variables independent of \mathbf{x} . We assume there are some vectors $\mathbf{f} = (f_1, f_2, \dots, f_N)^\top$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)^\top$ such that $f(\mathbf{x}_k) = \langle \mathbf{f}, \mathbf{x}_k \rangle$ and $\sigma(\mathbf{x}_k) = \langle \boldsymbol{\sigma}, \mathbf{x}_k \rangle$ represent the mean and volatility of y_k at time k , respectively. This assumption comes naturally from the canonical state space implying that a non-linear function of the chain can be represented as linear function of the chain via the scalar product. Here, \top denotes the transpose of a matrix. We shall further assume $\sigma_i > 0$ for every $1 \leq i \leq N$. Let $\{\mathcal{F}_k\}_{k \geq 0}$ denote the complete filtration generated by \mathbf{x} , $\{\mathcal{B}_k\}_{k \geq 0}$ denote the complete filtration generated by y and $\{\mathcal{H}_k\}_{k \geq 0}$ denote the complete filtration generated by \mathbf{x} and y . The model in equation (2.4) under the usual HMM was also the starting point of an empirical study in [15] devoted to the analysis of inflation rate movement.

The main idea of constructing filtering equations for the WHMM is to embed the second-order Markov chain into a first-order Markov chain, and then apply the already known methods for regular HMMs. To do this, we define a mapping ξ by

$$\xi(\mathbf{e}_r, \mathbf{e}_s) = \mathbf{e}_{rs}, \quad \text{for } 1 \leq r, s \leq N,$$

where \mathbf{e}_{rs} is a unit vector in \mathbb{R}^{N^2} with 1 in its $((r-1)N + s)$ th position. The mapping ξ groups two time steps of \mathbf{x} to form a new first-order Markov chain. Note that

$$\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle = \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle$$

represents the identification of \mathbf{x} at the current and previous time steps with the canonical basis. Let $\mathbf{\Pi}$ be an $N^2 \times N^2$ matrix, which represents the transition probability matrix of the new Markov chain $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$. It can be reconstructed from the matrix \mathbf{A} and is given by

$$\mathbf{\Pi} = \begin{pmatrix} a_{111} & \dots & a_{11N} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{121} & \dots & a_{12N} & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & a_{1N1} & \dots & a_{1NN} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{N11} & \dots & a_{N1N} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{N21} & \dots & a_{N2N} & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & a_{NN1} & \dots & a_{NNN} \end{pmatrix},$$

where

$$\pi_{ij} = \begin{cases} a_{lmv} & \text{if } i = (l-1)N + m, \quad j = (m-1)N + v \\ 0 & \text{otherwise.} \end{cases}$$

Now at time k , each nonzero element in $\mathbf{\Pi}$ represents the probability

$$\pi_{ij} = a_{lmv} = P(\mathbf{x}_k = \mathbf{e}_l | \mathbf{x}_{k-1} = \mathbf{e}_m, \mathbf{x}_{k-2} = \mathbf{e}_v)$$

and each zero represents an impossible transition. Following Siu, et al. [19], under the probability measure P , the weak Markov chain \mathbf{x} has the semi-martingale representation

$$\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi} \xi(\mathbf{x}_{k-1}, \mathbf{x}_{k-2}) + \mathbf{v}_k, \quad (2.5)$$

where $\{\mathbf{v}_k\}_{k \geq 1}$ is a sequence of \mathbb{R}^{N^2} -martingale increments with $E[\mathbf{v}_k | \mathcal{F}_k] = 0$.

2.3 Change of reference probability measure

In this section, we aim to estimate $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ given the observed data y_k under the real world probability P . In reality, y_k are not independent of each other. By the Kolmogorov's Extension

Theorem, there exists a reference probability measure \bar{P} , under which the observed data are independent, and thus, the calculations are easier to perform. We first present the relation between the real world probability measure P and the reference probability measure \bar{P} and then estimate $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under \bar{P} .

Under the ideal measure \bar{P} ,

- (i) $\{y_k\}_{k \geq 1}$ is a sequence of $N(0, 1)$ IID random variables, which are independent of \mathbf{x}_k , and
- (ii) $\{\mathbf{x}_k\}_{k \geq 0}$ is a weak Markov chain such that (2.5) holds and $\bar{E}[\mathbf{v}_k | \mathcal{F}_k] = \mathbf{0}$.

Write $\phi(z)$ for the probability density function of a standard normal random variable Z . To construct P from \bar{P} , we define the processes λ_l and Λ_k by

$$\lambda_l := \frac{\phi(\sigma(\mathbf{x}_{l-1})^{-1}(y_l - f(\mathbf{x}_{l-1})))}{\sigma(\mathbf{x}_{l-1})\phi(y_l)}, \quad (2.6)$$

and

$$\Lambda_k := \prod_{l=1}^k \lambda_l, \quad k \geq 1, \quad \Lambda_0 = 1. \quad (2.7)$$

To back out the probability measure P , we consider the Radon-Nikodým derivative Λ_k and set

$$\left. \frac{dP}{d\bar{P}} \right|_{\mathcal{H}_k} = \Lambda_k. \quad (2.8)$$

It could be shown that under P , the sequence $\{z_k\}$ is a sequence of IID standard normal random variables, where

$$z_k = \sigma(\mathbf{x}_{k-1})^{-1}(y_k - f(\mathbf{x}_{k-1})), \quad k \geq 1. \quad (2.9)$$

That is, the probability laws of \mathbf{x} under P and \bar{P} are the same; see Elliott, et al. [4] further. While we need the estimates of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under P , we shall perform all calculations under the reference probability measure \bar{P} . We could then employ the Bayes' theorem for conditional expectation, which relates conditional expectations under two different measures.

Let \mathbf{p}_k denote the conditional distribution of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ given \mathcal{Y}_k under P , so that $\mathbf{p}_k = (p_k^{11}, \dots, p_k^{ij}, \dots, p_k^{NN})^\top$, $1 \leq i, j \leq N$, is a vector in \mathbb{R}^{N^2} and

$$\begin{aligned} p_k^{ij} &= P(\mathbf{x}_k = \mathbf{e}_i, \mathbf{x}_{k-1} = \mathbf{e}_j | \mathcal{Y}_k) \\ &= E[\langle \mathbf{x}_k, \mathbf{e}_i \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_j \rangle | \mathcal{Y}_k] \\ &= E[\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij} \rangle | \mathcal{Y}_k]. \end{aligned}$$

Write

$$\mathbf{q}_k := \bar{E}[\Lambda_k \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_k]. \quad (2.10)$$

Since $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ takes values on the canonical basis of indicator functions, we have

$$\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1} \rangle = \sum_{i,j=1}^N \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij} \rangle = 1, \quad (2.11)$$

where $\mathbf{1}$ is an \mathbb{R}^{N^2} vector of 1's. Therefore,

$$\begin{aligned} \langle \mathbf{q}_k, \mathbf{1} \rangle &= \bar{E}[\Lambda_k \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1} \rangle | \mathcal{Y}_k] \\ &= \bar{E}[\Lambda_k | \mathcal{Y}_k] \end{aligned} \quad (2.12)$$

Invoking the Bayes' theorem for conditional expectation and equation (2.12), we get the explicit form for the conditional distribution

$$\mathbf{p}_k = \frac{\mathbf{q}_k}{\langle \mathbf{q}_k, \mathbf{1} \rangle}. \quad (2.13)$$

2.4 Calculation of recursive filters

The method we utilize to estimate the unknown model parameters is based on the estimation of the state process; this is $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ in our case. Looking at equation (2.13), we need a recursive

(iv) $T_k^r(g)$, the level sum for the state \mathbf{e}_r ,

$$T_k^r(g) = \sum_{l=1}^k g(y_l) \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle. \quad (2.20)$$

Here g is a function that takes the form $g(y) = y$ or $g(y) = y^2$.

For any \mathcal{H} -adapted process \mathbf{X}_k , the filter of \mathbf{X}_k is defined as $\hat{\mathbf{X}}_k := E[\mathbf{X}_k | \mathcal{Y}_k]$. We write $\gamma(\mathbf{X})_k := \bar{E}[\Lambda_k \mathbf{X}_k | \mathcal{Y}_k]$. Again, from Bayes' theorem for conditional expectation and equation (2.13), we have

$$\hat{\mathbf{X}}_k^{rst} = \frac{\bar{E}[\Lambda_k \mathbf{X}_k^{rst} | \mathcal{Y}_k]}{\bar{E}[\Lambda_k | \mathcal{Y}_k]} = \frac{\bar{E}[\Lambda_k \mathbf{X}_k^{rst} | \mathcal{Y}_k]}{\langle \mathbf{q}_k, \mathbf{1} \rangle}. \quad (2.21)$$

It would be difficult to estimate the quantities J_k^{rst} , O_k^{rs} , O_k^r and $T_k^r(g)$ directly. However, we could take advantage of the semi-martingale representation in equation (2.5) to obtain recursive filter relations for the vector quantities $J_k^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^{rs} \xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^r \xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ and $T_k^r(g) \xi(\mathbf{x}_k, \mathbf{x}_{k-1})$. The recursive filters of these vector processes are given in the following proposition.

Proposition 2.4.1 *Let \mathbf{V}_r , $1 \leq r \leq N$, be an $N^2 \times N^2$ matrix such that the $((i-1)N+r)$ th column of \mathbf{V}_r is \mathbf{e}_{ir} for $i = 1, 2, \dots, N$ and zero elsewhere. If \mathbf{B} is the diagonal matrix defined in equation (2.14) then*

$$\gamma\left(J^{rst} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)\right)_{k+1} = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma\left(J^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)_k + b_{k+1}^r \langle \mathbf{\Pi} \mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \mathbf{q}_k, \mathbf{e}_{st} \rangle \mathbf{e}_{rs}, \quad (2.22)$$

$$\gamma\left(O^{rs} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)\right)_{k+1} = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma\left(O^{rs} \xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)_k + b_{k+1}^r \langle \mathbf{q}_k, \mathbf{e}_{rs} \rangle \mathbf{\Pi} \mathbf{e}_{rs}, \quad (2.23)$$

$$\gamma\left(O^r \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)\right)_{k+1} = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma\left(O^r \xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)_k + b_{k+1}^r \mathbf{V}_r \mathbf{\Pi} \mathbf{q}_k, \quad (2.24)$$

$$\gamma\left(T^r(g) \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)\right)_{k+1} = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma\left(T^r(g) \xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)_k + b_{k+1}^r g(y_{k+1}) \mathbf{V}_r \mathbf{\Pi} \mathbf{q}_k. \quad (2.25)$$

Proof See Appendix B.

The recursive filters given in Proposition 2.4.1 provide updates to the estimates of the vector processes every time new information arrives. Each recursion involves the state process

$\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$. Similar to equation (2.12), we can relate the vector recursive processes to the scalar quantities of interest. For instance, the scalar quantity $\gamma(J^{rst})_k$, can be calculated by noting that

$$\begin{aligned} \langle \gamma(J^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k, \mathbf{1} \rangle &= \bar{E}[\Lambda_k J_k^{rst} \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1} \rangle | \mathcal{Y}_k] \\ &= \bar{E}[\Lambda_k J_k^{rst} | \mathcal{Y}_k] \\ &= \gamma(J^{rst})_k. \end{aligned}$$

The values for the other scalar quantities can be computed similarly.

2.5 Parameter estimation

In this section, we describe the estimation of the asset price model parameters, \mathbf{f} and σ in equation (2.4). Unlike the usual HMM, we estimate the transition matrix \mathbf{A} instead of $\mathbf{\Pi}$. We make use of the Expectation-Maximization (EM) algorithm. This algorithm offers an alternative method to maximize the conditional pseudo log-likelihood. The parameter updates are expressed in terms of the recursion in Proposition 2.4.1.

We first recall the EM algorithm. Let $\{P_\theta, \theta \in \Theta\}$ be a family of probability measures on a measurable space (Ω, \mathcal{F}) which is absolutely continuous with respect to a fixed probability measure P_0 ; Θ is some parameter space. Let $\mathcal{Y} \subset \mathcal{F}$. The likelihood function for estimating the parameter θ based on the information encapsulated in \mathcal{Y} is given by

$$L(\theta) = E_0 \left[\frac{dP_\theta}{dP_0} \middle| \mathcal{Y} \right]$$

while the maximum likelihood estimator of θ is defined by

$$\hat{\theta} \in \arg \max_{\theta \in \Theta} L(\theta).$$

The most likely value of the parameter θ is the one that maximizes this conditional expectation.

The MLE is, however, difficult to compute. By employing the EM algorithm, we can obtain the “maximizer” iteratively. The procedure is described below:

Step 1. Set $m = 0$ and choose $\hat{\theta}_0$.

Step 2. (E-step) Set $\theta^* = \hat{\theta}_m$ and compute

$$Q(\theta, \theta^*) = E_{\theta^*} \left[\log \frac{dP_\theta}{dP_{\theta^*}} \middle| \mathcal{Y} \right].$$

Step 3. (M-step) Find

$$\hat{\theta}_{m+1} \in \arg \max_{\theta \in \Theta} Q(\theta, \theta^*).$$

Step 4. Replace m by $m + 1$ and repeat the procedure beginning with step 2 until a stopping criterion is satisfied.

It is shown in Wu [22] that the sequence of estimates $\{\hat{\theta}_m\}$ gives nondecreasing values of the likelihood function and it converges to a local maximum of the expected log-likelihood. Since the EM algorithm does not identify the global maximum of the likelihood function, we may have to test several initial values at a wide range to illustrate that they all converge to the same value; this converged value is still not necessarily a global maximum but it gives us some assurance that we have a maximum value over a wide range of the parameter space. Suppose our model is determined by a set of estimated parameters $\hat{\theta} = \{\hat{a}_{rst}, \hat{f}_r, \hat{\sigma}_r, 1 \leq r, s, t \leq N\}$ which maximizes the corresponding conditional log-likelihood function.

Consider the case of estimating the transition matrix \mathbf{A} . The EM algorithm involves a change of measure from P_θ to $P_{\hat{\theta}}$. Under P_θ , \mathbf{x} is a weak Markov chain with transition matrix \mathbf{A} . Under measure $P_{\hat{\theta}}$, \mathbf{x} remains a weak Markov chain with transition matrix $\hat{\mathbf{A}} = (\hat{a}_{rst})$, which means $P_{\hat{\theta}}(\mathbf{x}_{k+1} = \mathbf{e}_r | \mathbf{x}_k = \mathbf{e}_s, \mathbf{x}_{k-1} = \mathbf{e}_t) = \hat{a}_{rst}$. Therefore, $\hat{a}_{rst} \geq 0$ and $\sum_{r=1}^N \hat{a}_{rst} = 1$. To replace the

parameter \mathbf{A} by $\hat{\mathbf{A}}$ in the weak Markov chain \mathbf{x} , we define the Radon-Nikodým derivative of $P_{\hat{\theta}}$ respect to P_{θ} :

$$\left. \frac{dP_{\hat{\theta}}}{dP_{\theta}} \right|_{\mathcal{G}_k} = \Gamma_k, \quad (2.26)$$

where

$$\Gamma_k = \prod_{l=2}^k \prod_{r,s,t=1}^N \left(\frac{\hat{a}_{rst}}{a_{rst}} \right)^{\langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle}.$$

In case $a_{rst} = 0$, take $\hat{a}_{rst} = 0$ and $\hat{a}_{rst}/a_{rst} = 1$. Refer to [14] for the justification why \mathbf{x} has transition matrix $\hat{\mathbf{A}}$ under $P_{\hat{\theta}}$. The optimal estimates for the model parameters, $\hat{\mathbf{f}}$, $\hat{\sigma}$ and $\hat{\mathbf{A}}$ are given by the following proposition.

Proposition 2.5.1 *If the set of parameters $\{\hat{a}_{rst}, \hat{f}_r, \hat{\sigma}_r\}$ determines the model then the EM estimates for these parameters are given by*

$$\hat{a}_{rst} = \frac{\hat{J}_k^{rst}}{\hat{O}_k^{st}} = \frac{\gamma(J^{rst})_k}{\gamma(O^{st})_k}, \quad \forall \text{ pairs } (r, s), r \neq s, \quad (2.27)$$

$$\hat{f}_r = \frac{\hat{T}_k^r}{\hat{O}_k^r} = \frac{\gamma(T^r(y))_k}{\gamma(O^r)_k}, \quad (2.28)$$

$$\hat{\sigma}_r^2 = \frac{\hat{T}_k^r(y^2) - 2\hat{f}_r \hat{T}_k^r(y) + \hat{f}_r^2 \hat{O}_k^r}{\hat{O}_k^r}, \quad \text{and } \hat{\sigma}_r = \sqrt{\hat{\sigma}_r^2}. \quad (2.29)$$

Proof See Appendix C.

Remark: *The estimator given in (2.27) is defined for the elements, a_{rst} , where $r \neq s$. However, one can compute the estimated values for a_{sst} by noting that $\forall s, t, \sum_{r=1}^N a_{rst} = 1$.*

2.6 Numerical results

We implement the recursive filters derived in the previous section on the daily logreturns series of S&P500. The data were recorded from September 1997 to March 2010; the dataset then consists of 3156 data points. Preliminary diagnostics would reveal that the evolution of the log return undergoes several distinct regimes characterized by states with high and low means as

well as high and low volatilities. To capture the behavior of regime switching, we assume that the log return's mean \mathbf{f} and volatility σ are governed by an WHMM \mathbf{x} . Given the daily asset price process S_k , we have

$$y_{k+1} = \log \frac{S_{k+1}}{S_k} = \langle \mathbf{f}, \mathbf{x}_k \rangle + \langle \sigma, \mathbf{x}_k \rangle z_{k+1}.$$

We segregate the observation data into different groups according to the level of mean and volatility. Of course, this is a qualitative way of selecting states. A more formal method of doing this is through a sequential analysis, see Wu [23], that deals with change-point problems. Tables 2.1 and 2.2 display descriptive statistics for *possible* segregations of actual data into either two or three states. From these segregations we can see that the log return y_k has a lower volatility when the mean is positive. When the market is bearish, i.e., associated with y_k having negative mean, equity investment is more risky and thus, we expect a higher volatility.

1st state	2nd state
Sept 1997-Jan 2000	Feb 2000-Sept 2003
Mean: 7.33×10^{-4}	Mean: -3.78×10^{-4}
Variance: 1.25×10^{-4}	Variance: 2×10^{-4}
Oct 2003-Dec 2006	Jan 2007-Mar 2010
Mean: 4.04×10^{-4}	Mean: -2.44×10^{-4}
Variance: 4.48×10^{-5}	Variance: 3.01×10^{-4}

Table 2.1: Segregation of the period of actual data into two states

1st state	2nd state	3rd state
Sept 1997-Jan 2000	Oct 2003-Dec 2006	Feb 2000-Sept 2003
Mean: 7.33×10^{-4}	Mean: 4.04×10^{-4}	Mean: -3.78×10^{-4}
Variance: 1.25×10^{-4}	Variance: 4.48×10^{-5}	Variance: 2×10^{-4}
		Jan 2007-Mar 2010
		Mean: -2.44×10^{-4}
		Variance: 3.01×10^{-4}

Table 2.2: Segregation of the period of actual data into three states

The implementation starts with the assignment of initial values for f_r and σ_r , $r = 1, \dots, N$. All non-zero entries in the transition matrix $\mathbf{\Pi}$ were assigned an initial value of $1/N$. In using the recursive filter equations, we process (i.e., apply recursive filters) the data in batches of 20 observation points. This means the parameters are roughly updated monthly. Each algorithm run, which processes a batch of 20 data points, constitutes what we call one complete algorithm step or an algorithm pass. At the end of each step, new estimates for \mathbf{f} , $\boldsymbol{\sigma}$ and \mathbf{A} are obtained. From the matrix \mathbf{A} , we get $\mathbf{\Pi}$. These new estimates are in turn used as initial estimates for the successive parameter estimation using the recursive filter equations. The frequency of parameter updating usually depends on the nature of observation data and the dictates of the financial market. We also experimented processing the data with different batch lengths in our multi-pass procedure, and it appears that monthly updating is sufficient for this particular case judging from the small forecasting error criterion.

Figure 2.1 shows three plots depicting the evolution of estimates for \mathbf{f} , $\boldsymbol{\sigma}$ and the transition matrix \mathbf{A} under the two-state weak Markov chain setting. For the three-state weak Markov chain setting, the transition matrix has $3^3 = 27$ elements and their evolutions are shown in three separate plots in Figure 2.2. Figure 2.3 displays the plots for the dynamics of the mean and volatility under the 3-state WHMM set-up. In both 2-state and 3-state WHMM modeling frameworks, the movements of the mean and volatility exhibit similar patterns. It is worth noting that through this multi-pass recursive algorithm, parameters appear to stabilize after approximately six passes. Our experiment indicates that this stability is achieved regardless of the choice of the initial parameter values. We note nonetheless that the speed of convergence is sensitive to the initial choice of parameter values. We note that on step 140, there is a change in the trend of the estimated probabilities and kinks in the estimated f and σ . This coincides with the market crisis that occurred in mid 2008. Thus, the filters are able to adapt to market conditions that prevailed in that period. We derive the explicit formula of the Fisher information to measure the variability of parameter estimates. The Fisher information $I(\theta)$ is

Figure 2.1: Evolution of estimates for \mathbf{f} , σ and \mathbf{A} -matrix under the 2-state setting

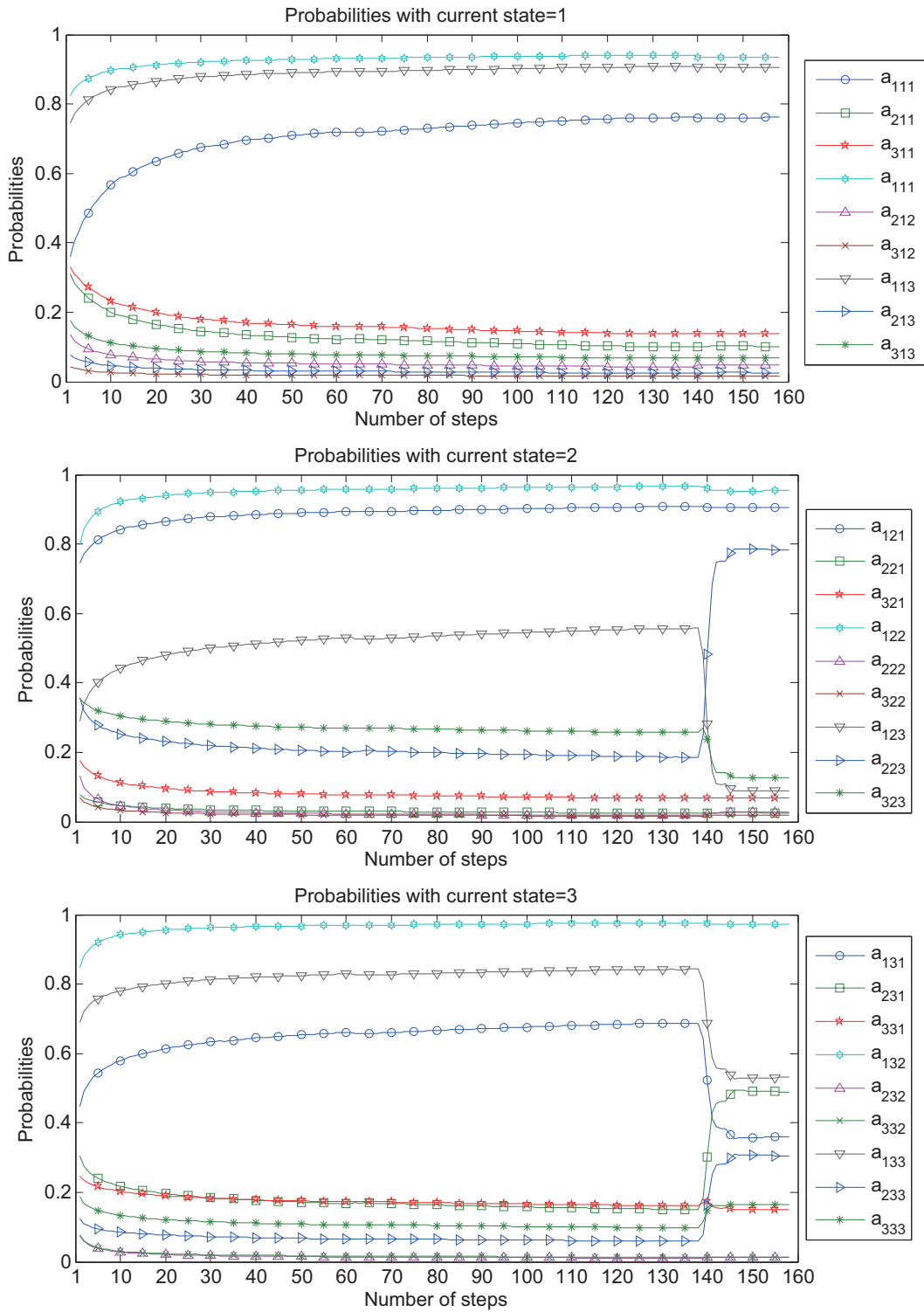


Figure 2.2: Evolution of estimates for the transition probabilities under the 3-state setting

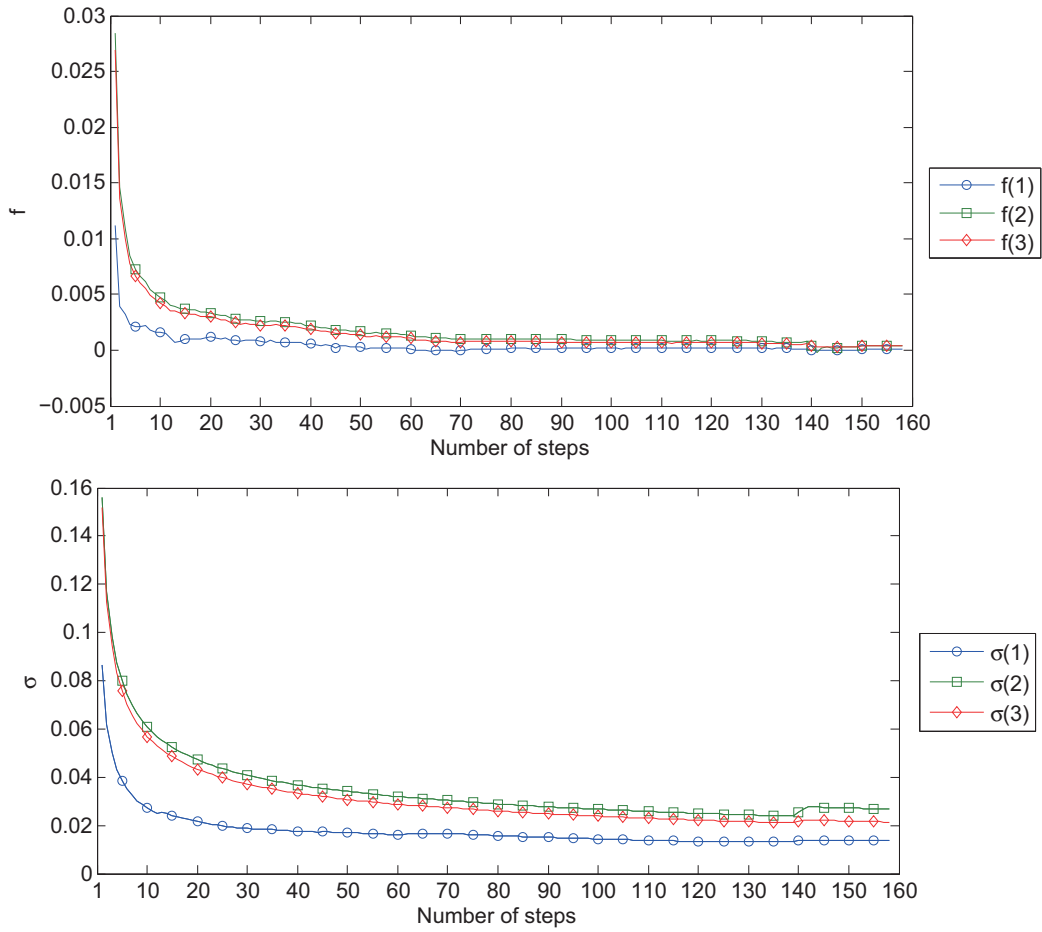


Figure 2.3: Evolution of estimates for \mathbf{f} and $\boldsymbol{\sigma}$ under the 3-state setting

Parameter estimate	Range of standard errors		
	1-state	2-state	3-state
\hat{a}_{rst}	$[3.22 \times 10^{-14}, 3.19 \times 10^{-9}]$	$[1.18 \times 10^{-12}, 1.26 \times 10^{-7}]$	$[8.21 \times 10^{-12}, 1.25 \times 10^{-5}]$
\hat{f}_r	$[5.18 \times 10^{-18}, 5.74 \times 10^{-13}]$	$[2.99 \times 10^{-16}, 1.32 \times 10^{-10}]$	$[2.26 \times 10^{-14}, 7.57 \times 10^{-9}]$
$\hat{\sigma}_r$	$[2.59 \times 10^{-18}, 2.87 \times 10^{-13}]$	$[1.50 \times 10^{-16}, 6.75 \times 10^{-11}]$	$[1.13 \times 10^{-14}, 4.04 \times 10^{-9}]$

Table 2.3: Range of SEs for each parameter under the 1-, 2- and 3-state settings

defined as the negative expectation of the second derivative of the log-density for a parameter θ . The inverse of the Fisher information is used to calculate the variance associated with the maximum-likelihood estimates. The sampling distribution of a maximum likelihood estimator is asymptotically normal and its variance can be calculated from $I^{-1}(\theta)$; see Garthwaite, et al. [10], for example. Following equations (C.1), (C.7) and (C.8) for $1 \leq r, s, t \leq N$, the closed-form expressions for the Fisher information of each parameter are given by

$$I(a_{rst}) = \frac{\hat{f}_k^{rst}}{a_{rst}^2}, \quad I(f_r) = \frac{\hat{O}_k^r}{\sigma_r^{-2}} \quad \text{and} \quad I(\sigma_r) = -\frac{\hat{O}_k^r}{\sigma_r^2} + \frac{3(\hat{T}_k^r(y_k^2) - 2\hat{T}_k^r(y_k)f_r + f_r^2)}{\sigma_r^{-4}}.$$

We provide the range of tabulated SEs over the entire algorithm steps for each parameter in Table 2.3 under the WHMM with $N = 1, 2, 3$.

Since the model via recursive formulas is self-updating and quickly produces reasonable parameter estimates, we can use it to forecast asset prices over an h -day ahead horizon. The semi-martingale representation of a weak Markov chain \mathbf{x} in equation (2.5) suggests that

$$E[\xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_k] = \mathbf{\Pi} \mathbf{p}_k, \quad \text{where } \mathbf{p}_k = E[\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_k].$$

This tells us that

$$E[\xi(\mathbf{x}_{k+h}, \mathbf{x}_{k+h-1}) | \mathcal{Y}_k] = \mathbf{\Pi}^h \mathbf{p}_k, \quad \text{for } h = 1, 2, \dots \quad (2.30)$$

Recall that \mathbf{A} is defined by $a_{mv} = P(\mathbf{x}_{k+1} = \mathbf{e}_l | \mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v)$, so that

$$E[\mathbf{x}_{k+1} | \mathcal{Y}_k] = \mathbf{A} \mathbf{p}_k. \quad (2.31)$$

Equations (2.31) and (2.30) then imply that

$$E[\mathbf{x}_{k+h}|\mathcal{Y}_k] = \mathbf{A}\mathbf{p}_{k+h-1} = \mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k. \quad (2.32)$$

Using equation (2.32), the best estimate of the logarithmic increment y_{k+h} given available information at time k is

$$E[y_{k+h}|\mathcal{Y}_k] = \langle \mathbf{f}, \mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k \rangle.$$

On the other hand, the conditional variance of y_{k+h} is given by

$$\text{Var}[y_{k+h}|\mathcal{Y}_k] = \mathbf{f}^\top \text{diag}(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)\mathbf{f} + \boldsymbol{\sigma}^\top \text{diag}(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)\boldsymbol{\sigma} - \langle \mathbf{f}, \mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k \rangle^2,$$

where $\text{diag}(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)$ is a diagonal matrix whose diagonal entries are the components of the vector $\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k$.

Conditional on the information structure \mathcal{Y}_k , the observation process y_{k+h} has a mixed normal distribution with explicit representation

$$\sum_{i,j=1}^N \langle \mathbf{p}_{k+h-1}, \mathbf{e}_{ij} \rangle \phi(y; f_i, \sigma_i).$$

Therefore the best estimate of the asset price at time $k+h$ based on available information at time k is given by

$$E[S_{k+h}|\mathcal{Y}_k] = S_k \sum_{i,j=1}^N \langle \mathbf{\Pi}^{h-1}\mathbf{p}_k, \mathbf{e}_{ij} \rangle \exp\left(f_i + \frac{\sigma_i^2}{2}\right). \quad (2.33)$$

Equation (2.33) is used to compute the h -day ahead forecasts of the S&P500 index values. In a related study, Mamon, et al. [16] compared the predictability performance under the Diebold-Kilian metric of two- and three-state HMM with the predictability performance implied by their chosen benchmarked models, namely, autoregressive conditional heteroscedasticity (ARCH(1)) and generalized ARCH (GARCH(1,1)) models. Their results suggest that

h -day ahead	RMSE		AME		RAE		APE	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
1	14.6652	14.6471	10.4380	10.4249	0.0663	0.0663	0.0092	0.0092
2	20.1931	20.1873	14.8358	14.8329	0.0943	0.0943	0.0130	0.0130
3	23.9568	23.9544	17.8275	17.8331	0.1133	0.1133	0.0157	0.0157
4	27.2519	27.2521	20.3194	20.3276	0.1291	0.1292	0.0179	0.0179
5	30.1839	30.1862	22.4197	22.4284	0.1425	0.1425	0.0197	0.0197
6	32.7025	32.7057	24.4092	24.4189	0.1551	0.1552	0.0215	0.0215
7	34.9630	34.9673	26.0477	26.0606	0.1655	0.1656	0.0229	0.0229
8	36.8720	36.8770	27.4169	27.4313	0.1742	0.1743	0.0241	0.0242
9	38.7440	38.7496	28.6977	28.7143	0.1824	0.1825	0.0253	0.0253
10	40.5210	40.5268	30.0238	30.0416	0.1908	0.1909	0.0265	0.0266
11	42.3122	42.3185	31.4179	31.4378	0.1997	0.1998	0.0278	0.0278
12	43.9374	43.9445	32.6621	32.6803	0.2076	0.2077	0.0289	0.0289
13	45.8010	45.8090	34.1693	34.1893	0.2171	0.2173	0.0303	0.0303
14	47.6995	47.7079	35.7466	35.7676	0.2272	0.2273	0.0316	0.0317
15	49.4687	49.4772	37.0384	37.0573	0.2354	0.2355	0.0328	0.0328
16	51.0220	51.0303	38.2533	38.2738	0.2431	0.2432	0.0339	0.0339
17	52.6908	52.6991	39.5625	39.5788	0.2514	0.2515	0.0350	0.0350
18	54.3287	54.3370	40.7496	40.7646	0.2590	0.2591	0.0361	0.0361
19	55.7159	55.7239	41.8855	41.8977	0.2662	0.2663	0.0371	0.0371
20	57.1569	57.1648	43.0776	43.0913	0.2738	0.2738	0.0381	0.0381
21	58.5421	58.5499	44.0828	44.0959	0.2801	0.2802	0.0390	0.0390
22	59.7281	59.7360	45.0667	45.0793	0.2864	0.2865	0.0399	0.0399
23	61.0026	61.0106	46.0674	46.0786	0.2928	0.2928	0.0409	0.0409
24	62.2606	62.2689	47.1080	47.1190	0.2994	0.2994	0.0418	0.0418
25	63.4862	63.4945	48.0434	48.0530	0.3053	0.3054	0.0426	0.0426

Table 2.4: Error analysis of WHMM and HMM models under the 2-state setting

compared to the benchmarked models, HMM models produce higher measures of short- and medium-run predictability.

In this empirical implementation, we compare the forecasting performance of WHMM with that of the regular HMM. The goodness of fit of the h -day ahead forecasts ($h = 1, \dots, 25$) to the actual data is evaluated using the root mean square error (RMSE), absolute mean error (AME), relative absolute error (RAE) and absolute percentage error (APE). These forecasts errors under the two- and three-state model settings are given in Table 2.4 and 2.5, respectively.

h -day ahead	RMSE		AME		RAE		APE	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
1	14.8066	14.6767	10.5542	10.4603	0.0671	0.0665	0.0093	0.0092
2	20.3007	20.1974	14.9244	14.8386	0.0948	0.0943	0.0131	0.0130
3	24.0228	23.9588	17.8700	17.8281	0.1136	0.1133	0.0157	0.0157
4	27.2972	27.2530	20.3319	20.3181	0.1292	0.1291	0.0179	0.0179
5	30.2163	30.1841	22.4184	22.4177	0.1425	0.1425	0.0197	0.0197
6	32.7255	32.7022	24.3957	24.4071	0.1550	0.1551	0.0215	0.0215
7	34.9782	34.9624	26.0275	26.0448	0.1654	0.1655	0.0229	0.0229
8	36.8819	36.8711	27.4015	27.4131	0.1741	0.1742	0.0241	0.0241
9	38.7495	38.7428	28.6763	28.6940	0.1822	0.1824	0.0253	0.0253
10	40.5234	40.5197	29.9903	30.0185	0.1906	0.1908	0.0265	0.0265
11	42.3108	45.0483	31.3680	31.8622	0.1993	0.1833	0.0278	0.0284
12	43.9309	47.0441	32.6089	33.3315	0.2072	0.1917	0.0289	0.0297
13	45.7893	49.2302	34.1124	34.8385	0.2168	0.2004	0.0302	0.0311
14	47.6846	51.3417	35.6899	36.5875	0.2268	0.2104	0.0316	0.0326
15	49.4517	53.3269	36.9841	37.9214	0.2350	0.2181	0.0327	0.0337
16	51.0056	54.8892	38.2056	39.2742	0.2428	0.2259	0.0338	0.0350
17	52.6736	56.8324	39.5141	40.7638	0.2511	0.2345	0.0350	0.0363
18	54.3113	58.6921	40.7013	42.0401	0.2587	0.2418	0.0360	0.0374
19	55.6988	60.1808	41.8339	43.1875	0.2659	0.2484	0.0370	0.0384
20	57.1400	61.8393	43.0299	44.4809	0.2735	0.2558	0.0381	0.0396
21	58.5253	63.5352	44.0464	45.7187	0.2799	0.2630	0.0390	0.0407
22	59.7102	64.9906	45.0458	46.9500	0.2862	0.2700	0.0399	0.0418
23	60.9835	66.7393	46.0340	48.2065	0.2925	0.2773	0.0408	0.0430
24	62.2403	68.4247	47.0681	49.4898	0.2991	0.2846	0.0418	0.0441
25	63.4650	69.9037	48.0087	50.6459	0.3051	0.2913	0.0426	0.0451

Table 2.5: Error analysis of WHMM and HMM models under the 3-state setting

Compared to the two-state WHMM, the regular two-state HMM yields lower forecast errors in the short forecasting horizon. However, the difference is too small to yield any practical significance. In all forecasting metrics (RMSE, MAE, RAE and APE), both WHMM and HMM forecasts errors are almost identical in the short- and medium- forecasting horizons. The WHMM clearly outperforms the HMM over a long forecasting horizon. For the three-state models, HMM gives slightly better fit to the actual data than the WHMM for the short forecasting horizon only. The benefit of employing a 3-state WHMM can be seen in the medium- and long-forecasting horizons in which the WHMM forecasting errors are much lower than the HMM forecasting errors.

The special case of one-state modeling framework for both the WHMM of order 2 and HMM, which correspond to the second-order and first-order autoregressive models, respectively, was also investigated. We found that the forecasting errors are generally higher than those exhibited by the two-state and three-state WHMM and HMM. This suggests that there is merit to exploring regime-switching models incorporating memories to accurately capture the dynamics of the data series.

2.7 Conclusion

In this chapter, we proposed a weak Markov chain-modulated model for asset prices. By transforming a WHMM into a regular HMM, we developed general recursive filters for discrete-time, continuous-range observations and utilized the EM algorithm in conjunction with the change of probability measure to re-estimate the parameters in our model. We analyzed the h -day ahead predictions and results from WHMM were compared to those from the regular HMM. Our empirical findings show that by including memories in the model, the two- and three-state WHMM outperforms the HMM in terms of low forecasting errors in the long forecasting horizon.

In this study, we only consider a weak Markov chain of order 2. However, in many financial time series applications such as GARCH and ARCH models, lag 1 or lag 2 already appears to be sufficient. Of course, it would be interesting to examine third-order or even n th-order HMM models with $n \geq 4$. The challenge here is to deal with the complexity of the estimation and the notation would certainly become unwieldy. As the number of states increases, the number of parameters also increases exponentially. A natural direction of this research is to develop a statistical methodology that will determine the optimal number of states and lags of the weak Markov chain.

References

- [1] N. Bollen, S. Gray, and R. Whaley. Regime switching in foreign exchange rates: evidence from currency option prices. *Journal of Econometrics*, 94:239–276, 2000.
- [2] W. K. Ching, T. K. Siu, and L. M. Li. Pricing exotic options under a higher-order Markovian regime switching model. *Journal of Applied Mathematics and Decision Sciences*, pages 1–15, 2007. Article ID 18014.
- [3] C. J. Chu, G. J. Santoni, and T. Liu. Stock market volatility and regime shift in returns. *Information Sciences*, 94:179–190, 1996.
- [4] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer, New York, 1995.
- [5] R. J. Elliott, W. C. Hunter, and B. M. Jamieson. Financial signal processing: a self calibrating model. *International Journal of Theoretical and Applied Finance*, 4:567–584, 2003.
- [6] R. J. Elliott, W. P. Malcolm, and A. H. Tsoi. Robust parameter estimation for asset price models with Markov modulated volatilities. *Journal of Economics Dynamics and Control*, 27:1391–1409, 2003.
- [7] R. J. Elliott and J. van der Hoek. An application of hidden Markov models to asset allocation problems. *Finance and Stochastics*, 3:229–238, 2003.

- [8] C. Erlwein and R. S. Mamon. An online estimation scheme for a Hull-White model with HMM-driven parameters. *Statistical Methods and Applications*, 18:87–107, 2009.
- [9] C. Erlwein, R. S. Mamon, and M. Davison. An examination of HMM-based investment strategies for asset allocation. *Applied Stochastic Models in Business and Industry*, 27:204–221, 2009.
- [10] P. Garthwaite, I. Jolliffe, and B. Jones. *Statistical Inference*. Oxford Science Publication, Oxford, 2nd edition, 2002.
- [11] S. M. Goldfeld and R. E. Quandt. A Markov model for switching regression. *Journal of Econometrics*, 1:3–16, 1973.
- [12] J. Haigh. Introduction to Markov chains-the finite case. *Significance* 7, 7:88–89, 2010.
- [13] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and business cycle. *Econometrica*, 57:357–384, 1989.
- [14] S. Luo and A. H. Tsoi. Filtering of hidden weak Markov chain-discrete range observation. In R. S. Mamon and R. J. Elliott, editors, *Hidden Markov Models in Finance*, pages 106–119. Springer, New York, 2007.
- [15] R. S. Mamon and Z. Duan. A self-tuning model for inflation rate dynamics. *Communications in Nonlinear Science and Numerical Simulation*, 15:2521–2528, 2010.
- [16] R. S. Mamon, C. Erlwein, and R. B. Gopaluni. Adaptive signal processing of asset price dynamics with predictability analysis. *Information Sciences*, 178:203–219, 2008.
- [17] R. E. Quandt. The estimation of parameters of linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 55:873–880, 1958.
- [18] T. K. Siu, W. K. Ching, and E. Fung. Extracting information from spot interest rates and credit ratings using double higher-order hidden Markov models. *Computational Economics*, 26:251–284, 2005.

- [19] T. K. Siu, W. K. Ching, E. Fung, M. Ng, and X. Li. A high-order Markov-switching model for risk measurement. *Computers and Mathematics with Applications*, 58:1–10, 2009.
- [20] J. Solberg. *Modelling Random Processes for Engineers and Managers*. John Wiley & Sons, New Jersey, 2009.
- [21] C. M. Turner, R. Startz, and C. R. Nelson. A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25:3–22, 1989.
- [22] C. Wu. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11:95–103, 1983.
- [23] Yanhong Wu. *Inference for Change Point and Post Change Means After a CUSUM Test*. Springer, New York, 2005.

Chapter 3

Parameter estimation in a WHMM setting with independent drift and volatility components

3.1 Introduction

In this chapter, we consider a quite general extension of WHMM-modulated asset price model by allowing the drift and volatility to be driven by two independent WMCs not necessarily having the same number of states. That is, for instance, the drift may have 3 states while the volatility has only two states and the asset returns components are driven by different Markov chains.

In recent years, Markovian regime switching models have received considerable interests in economics, finance and actuarial science. In a pioneering work, Hamilton [10] proposed a class of discrete-time Markov switching autoregressive time series models. In Hamilton's model, the parameters are modulated by a discrete-time, finite state Markov chain so that the parameters have different values in a particular time period according to the state of the chain in that period. This idea gives a natural and simple way to model the cyclical behavior or the impact of

changes in the financial market on financial series dynamic. Many empirical studies reveal that Markovian regime-switching model can provide a better description of economic and financial series than that from a single regime model. Hardy [11] proposed a Markov regime-switching lognormal model for stock returns, and implemented the model on S&P500 and the Toronto Stock Exchange 300 indices; it was found that the regime-switching model performed better than the GARCH model. Comparison of the fit to the data between the regime switching model and GARCH model is presented. Lange and Rohbek [13] gave a survey on regime switching in econometric time series modeling and the differences between observation switching and Markov switching are discussed. Ang and Timmermann [1] discussed the impact of regime switching on equilibrium asset prices and suggested that regimes exist in a variety of financial series such as fixed income, equity and currency markets. Veronesi [23] considered a three-state model in which the asset prices are highest in the “good” regime and lowest in the “bad” regime. Calvet and Fisher [3] suggested that regimes can account for time-varying state dependent and asymmetric reaction of equilibrium stock prices to news. A well-known class of such models is the HMM, in which a hidden Markov chain is adopted to describe the random transition of the hidden state of an economy. HMM can provide a reasonably realistic description of some important empirical features such as heavy-tails of the distribution of the returns and time-varying conditional volatility. Due to their empirical successes, HMMs have been widely adopted in modeling financial series dynamics. Rydén, et al. [20] considered an HMM for modeling daily return series, and investigated the capability of HMM to capture the series’ stylized facts. By applying the model to S&P500 daily returns, the results suggest that the HMM can describe most of the stylized facts except for the slowly decaying autocorrelation function of the absolute return. Early studies of HMM to financial time series include Tyssedal and Tjotheim [22], Pagan and Schwert [19] and Sola and Leroux [14]. The monograph by Elliott, et al. [5] provides a comprehensive discussion of parameter estimation under the HMM framework using filtering techniques. Since then, many researchers follow and apply this technique to finance and economics. Elliott, et al. [6] applied robust filtering equations

for a continuous-time HMM to estimate the volatility of a risky asset. Erlwein, et al. [17] derived and implemented the filters for logreturn of commodity prices, and compared the HMM to ARCH and GARCH models with respect to the prices' predictability. The HMM filtering method is applied to many other financial problems, for example, asset allocation [7, 9], interest rate modeling [8, 27], option pricing [15], and so on.

Although popular, the simple homogeneous Markov switching model is memoryless, which seems inadequate for real-world data. It is well-known in the economic and finance literature that the states of an economy and financial series present long-range dependence, which can be reified through a simple plot of an empirical autocorrelation function. Motivated by this empirical phenomenon, we consider a WHMM, as a more flexible alternative to HMM. The basic idea of an n th order HMM is that the behavior of the underlying Markov chain at the present time depends on its behavior in the past n time steps. WHMM is popular in speech and text recognition, but is rather new in probing the finer structures of the financial market. Bulla and Bulla [2] examined the fit of WHMMs to 18 daily sector return series and suggested that the stylized fact of slowly decaying autocorrelation can be described better by WHMMs. Yu, et al. [26] explored the use of WHMMs to capture the long-range dependence property. They derived the recursive formula for the autocovariance function over different time scales and the estimator of the Hurst parameter. Their empirical results demonstrate that the model can capture long-range dependence if one state is heavy-tailed distributed. By weakening the Markov assumption, WHMMs provide a simple and flexible way to describe the duration dependence through their dependence on backward recurrence time. This has led some authors to study the ideas of WHMMs in the fields of financial derivatives, for example, risk management [21], option pricing [4], interest rate modeling [12], and asset returns [25].

Mamon and Jalen [18] proposed a method based on employing tensors to transform two independent chains into one Markov chain so that the regular filtering technique can be applied.

This method is then implemented on two stock indices for illustration. In this paper, we investigate a WHMM in the situation that drift and volatility of the given data are driven by two independent WMCs. In particular, we suppose that the rate of return of a risky asset is governed by a WHMM with two underlying WMCs. The transformation method based on tensors is adopted and the filtering technique of WHMMs with one chain is then applied. Numerical study based on simulated observation data is given to demonstrate the effectiveness of this method. We also provide error analyses for different combination of states through the h -step ahead prediction performance.

This chapter is organized as follows. In section 3.2, we present the modeling framework of WHMM. The dynamics of a risky asset price under the WHMM extension is described. By utilizing a measure-change method, we derive recursive filters for the state of WMC and other processes of interests. Parameter estimation based on EM algorithm is established. In section 3.3, we implement the filters under the proposed extended set-up on simulated data. The method is examined using different algorithm starting values. We generate one- and five-step ahead forecasts for different models and compare the forecasting performance via four error metrics. The chapter ends with a conclusion section.

3.2 Model background

Now, we present a WHMM for modeling asset prices, where the drift and volatility have independent probability behavior. In the sequel, all vectors will be denoted by bold English or Greek letters in lowercase and all matrices will be denoted by bold letters in uppercase. Fix a complete probability space (Ω, \mathcal{F}, P) , where P is a real world probability measure. Define a discrete-time weak Markov chain $\{\mathbf{x}_k\}$, $k \geq 0$ on (Ω, \mathcal{F}, P) with a finite space $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$. The states of the chain represent different states of economy. Without loss of generality, the points in \mathcal{S} can be identified using the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\} \subset \mathbb{R}^N$, where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$ and \top denotes the transpose of a vector. The expression

$\langle \mathbf{x}_k, \mathbf{e}_i \rangle$ represents the event that the economy is in state i at time t and $\langle \cdot, \cdot \rangle$ stands for the inner product in \mathbb{R}^N .

In the succeeding discussion, we concentrate on a WMC of order 2 to simplify the discussion and present a complete characterization of the parameter estimation. The probability of the next time step for a second-order WMC depends on the information on current and previous time steps. Let $\mathbf{A} \in \mathbb{R}^{N \times N^2}$ denote the transition probability matrix of WMC \mathbf{x}_k . Each entry $a_{lmv} := P(\mathbf{x}_{k+1} = \mathbf{e}_l | \mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v)$, $l, m, v \in 1, \dots, N$ is the transition probability that the chain \mathbf{x} enters state l given that the current and previous states were states m and v , respectively. The salient idea in the filtering method for WHMM is that, a second-order Markov chain is transformed into a first-order Markov chain through a mapping ξ , and then we may apply the regular filtering method. The mapping ξ is defined by

$$\xi(\mathbf{e}_r, \mathbf{e}_s) = \mathbf{e}_{rs}, \text{ for } 1 \leq r, s \leq N,$$

where \mathbf{e}_{rs} is an \mathbb{R}^{N^2} -unit vector with unity in its $((r-1)N + s)$ th position. The identification of the new first-order Markov chain with the canonical basis is given by

$$\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle = \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle.$$

We further assume that the new Markov chain has a new transition probability matrix, $\mathbf{\Pi} \in \mathbb{R}^{N^2 \times N^2}$, given by

$$\pi_{ij} = \begin{cases} a_{lmv} & \text{if } i = (l-1)N + m, \quad j = (m-1)N + v \\ 0 & \text{otherwise.} \end{cases}$$

Each non-zero element π_{ij} represents the probability

$$\pi_{ij} = a_{lmv} = P(\mathbf{x}_k = \mathbf{e}_l | \mathbf{x}_{k-1} = \mathbf{e}_m, \mathbf{x}_{k-2} = \mathbf{e}_v),$$

and each zero represents an impossible transition. It may be shown that the new Markov chain $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ has the semi-martingale representation

$$\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi}\xi(\mathbf{x}_{k-1}, \mathbf{x}_{k-2}) + \mathbf{v}_k, \quad (3.1)$$

where $\{\mathbf{v}_k\}_{k \geq 1}$ is a sequence of \mathbb{R}^{N^2} martingale increments.

Let S_k , $k \geq 1$, denote a series of asset prices and y_k denote the logarithmic increments. In the previous study Xi and Mamon [25], we discussed the case where the drift and volatility of y_k are governed by the same hidden Markov chain. In particular, y_k is assumed to have the dynamics

$$y_{k+1} = f(\mathbf{x}_k) + \sigma(\mathbf{x}_k)z_{k+1} = \langle \mathbf{f}, \mathbf{x}_k \rangle + \langle \boldsymbol{\sigma}, \mathbf{x}_k \rangle z_{k+1}. \quad (3.2)$$

The sequence $\{z_k\}$ is a sequence of $N(0, 1)$ IID random variables, which are independent of the \mathbf{x} -process. In this study, we consider the case when the drift and volatility have independent states and probabilistic behavior. That is, we assume y_k has the dynamics

$$y_{k+1} = \langle \mathbf{f}, \mathbf{x}_k^1 \rangle + \langle \boldsymbol{\sigma}, \mathbf{x}_k^2 \rangle z_{k+1}, \quad (3.3)$$

where \mathbf{x}^i is an N_i -state WMC on state space \mathcal{S}_i with transition matrix $\mathbf{A}_i \in \mathbb{R}^{N_i \times N_i}$. Suppose the drift and volatility have the form $\mathbf{f} = (f_1, f_2, \dots, f_{N_1}) \in \mathbb{R}^{N_1}$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_{N_2}) \in \mathbb{R}^{N_2}$ respectively. In order to apply the regular WHMM filtering technique, we aim to re-formulate the hidden WMCs so that the dynamic of y_k is in the same form as in equation (3.2). Let \otimes denote tensor product. Following the idea in [18], we transform the two chains, \mathbf{x}_k^1 and \mathbf{x}_k^2 , into a new WMC \mathbf{x}_k using Kronecker product or tensor product, i.e., $\mathbf{x}_k = \mathbf{x}_k^1 \otimes \mathbf{x}_k^2$. Then \mathbf{x}_k is an $N_1 N_2$ -state WMC with transition matrix $\mathbf{A} = \mathbf{A}_1 \otimes \mathbf{A}_2$. Write $\mathbf{1}_N$ for the vector $(1, 1, \dots, 1) \in \mathbb{R}^N$. The

reformulated drift and volatility are given by

$$\boldsymbol{\alpha} = \mathbf{f} \otimes \mathbf{1}_{N_2},$$

$$\boldsymbol{\eta} = \mathbf{1}_{N_1} \otimes \boldsymbol{\sigma}$$

Therefore the dynamics of y_k in equation (3.3) can be rewritten as

$$y_{k+1} = \langle \boldsymbol{\alpha}, \mathbf{x}_k \rangle + \langle \boldsymbol{\eta}, \mathbf{x}_k \rangle z_{k+1}. \quad (3.4)$$

We demonstrate the workings of this transformation through a numerical example in section 3.4.

3.3 Filters and parameter estimation

Under the real world measure P , we cannot observe the hidden state of the economy \mathbf{x}_k directly. Instead, we are given market observations \mathbf{y}_k , which contain information about \mathbf{x}_k . Since the unknown drift and volatility are highly dependent on the WMC, the estimation of parameters reduced to “filtering” the hidden WMC out of the observations. However, the derivation of filters under P is complicated. Exploiting the Kolmogorov’s Extension theorem, we note that there exists a reference probability measure \bar{P} under which

- \mathbf{y}_k ’s are $N(0, 1)$ IID random variables and
- \mathbf{x} is a finite state WMC satisfying (3.1) and $\bar{E}[\mathbf{v}_k | \mathcal{Y}_k] = 0$.

Under the measure \bar{P} , \mathbf{y} does not depend on \mathbf{x} , and therefore it is more convenient to evaluate the filtered estimates. The calculation starts with the reference probability measure and then we perform a measure change to construct the real-world measure P . Consider a \mathcal{Y}_k -adapted

process Λ_k , $k \geq 0$ defined by

$$\lambda_l = \frac{\phi\left(\sigma(\mathbf{x}_{l-1})^{-1}(y_l - f(\mathbf{x}_{l-1}))\right)}{\sigma(\mathbf{x}_{l-1})\phi(y_l)}, \quad (3.5)$$

$$\Lambda_k = \prod_{l=1}^k \lambda_l, \quad k \geq 1, \quad \Lambda_0 = 1, \quad (3.6)$$

where $\phi(z)$ is the probability density function of a standard normal random variable Z . Define the Radon-Nikodým derivative of P with respect to \bar{P} by

$$\frac{dP}{d\bar{P}}\Big|_{\mathcal{Y}_k} := \Lambda_k. \quad (3.7)$$

Suppose X_k is a \mathcal{Y}_k -adapted process. Write $\hat{X}_k := E[X_k|\mathcal{Y}_k]$ and $\gamma(X_k) := \bar{E}[\Lambda_k X_k|\mathcal{Y}_k]$. Then by Bayes' theorem, the unnormalized filter of X_k is

$$\hat{X}_k = \frac{\bar{E}[\Lambda_k X_k|\mathcal{Y}_k]}{\bar{E}[\Lambda_k|\mathcal{Y}_k]} = \frac{\gamma(X_k)}{\gamma(1)}. \quad (3.8)$$

Let us derive the conditional expectation of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ given \mathcal{Y}_k under P . Write

$$p_k^{ij} := P(\mathbf{x}_k = \mathbf{e}_i, \mathbf{x}_{k-1} = \mathbf{e}_j|\mathcal{Y}_k) = E[\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij} \rangle|\mathcal{Y}_k] \quad (3.9)$$

with $\mathbf{p}_k = (p_k^{11}, \dots, p_k^{ij}, \dots, p_k^{NN}) \in \mathbb{R}^{N^2}$. Bayes' theorem for conditional expectation (for example, see p.22 of Elliott, et al. [5]) implies

$$\mathbf{p}_k = E[\xi(\mathbf{x}_k, \mathbf{x}_{k-1})|\mathcal{Y}_k] = \frac{\gamma(\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))}{\gamma(1)}. \quad (3.10)$$

Note that

$$\sum_{i,j=1}^N \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij} \rangle = \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1}_{N^2} \rangle = 1. \quad (3.11)$$

Let $\mathbf{q}_k = \gamma(\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))$ so that

$$\langle \mathbf{q}_k, \mathbf{1}_{N^2} \rangle = \bar{E}[\Lambda_k \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1}_{N^2} \rangle | \mathcal{Y}_k] = \gamma(1). \quad (3.12)$$

From equations (3.10) and (3.12), we get the conditional distribution of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under P as

$$\mathbf{p}_k = \frac{\mathbf{q}_k}{\langle \mathbf{q}_k, \mathbf{1}_{N^2} \rangle}. \quad (3.13)$$

In order to estimate the state process $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, we shall investigate the recursion for the process \mathbf{q}_k . Define the diagonal matrix $\mathbf{B}_k \in \mathbb{R}^{N^2 \times N^2}$ by

$$\mathbf{B}_k = \text{diag}(b_k^1, \dots, b_k^N, \dots, b_k^1, \dots, b_k^N) \quad (3.14)$$

where $\text{diag}(\mathbf{v})$ is a diagonal matrix whose diagonal entries are the components of the vector \mathbf{v} and

$$b_k^i = \frac{\phi((y_k - f_i)/\sigma_i)}{\sigma_i \phi(y_k)}. \quad (3.15)$$

To estimate the parameters of the model, we first present a set of quantities that are useful for the derivation of estimates. Three of these processes are related to the state process and one is related to both the state and observation processes. These quantities are defined by, for $r, s, t = 1, \dots, N$,

- J_k^{rst} , the number of jumps from $(\mathbf{e}_s, \mathbf{e}_t)$ to state \mathbf{e}_r up to time k ,

$$J_k^{rst} = \sum_{l=1}^k \langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle \quad (3.16)$$

- O_k^{rs} , the occupation time of \mathbf{x} spent in state $(\mathbf{e}_r, \mathbf{e}_s)$ up to time k ,

$$O_k^{rs} = \sum_{l=1}^k \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_s \rangle \quad (3.17)$$

- O_k^r , the occupation time spent by \mathbf{x} in state \mathbf{e}_r up to time k ,

$$O_k^r = \sum_{l=1}^k \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \quad (3.18)$$

- $T_k^r(g)$, the level sum for the state \mathbf{e}_r ,

$$T_k^r(g) = \sum_{l=1}^k g(y_l) \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle. \quad (3.19)$$

Here, g is a function with the form $g(y) = y$ or $g(y) = y^2$, for $1 \leq l \leq k$.

The following proposition presents the recursive formulas for the vectors $\gamma(J_k^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))$, $\gamma(O_k^{rs} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))$, $\gamma(O_k^r \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))$ and $\gamma(T_k^r(g) \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))$, which are the unnormalized filtered estimates of J_k^{rst} , O_k^{rs} , O_k^r and $T_k^r(g)$ respectively. The recursive relation of these vector processes and \mathbf{q}_k under a multi-dimensional observation set-up are given in the following proposition.

Proposition 3.3.1 *Let \mathbf{V}_r , $1 \leq r \leq N$ be an $N^2 \times N^2$ matrix such that the $((i-1)N+r)$ th column of \mathbf{V}_r is \mathbf{e}_{ir} for $i = 1 \dots N$ and zero elsewhere. If \mathbf{B} is the diagonal matrix defined in equation (3.14) then*

$$\mathbf{q}_{k+1} = \mathbf{B}_{k+1} \mathbf{\Pi} \mathbf{q}_k \quad (3.20)$$

and

$$\gamma(J_{k+1}^{rst} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)) = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma(J_k^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1})) + b_{k+1}^r \langle \mathbf{\Pi} \mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \mathbf{q}_k, \mathbf{e}_{st} \rangle \mathbf{e}_{rs}, \quad (3.21)$$

$$\gamma(O_{k+1}^{rs} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)) = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma(O_k^{rs} \xi(\mathbf{x}_k, \mathbf{x}_{k-1})) + b_{k+1}^r \langle \mathbf{q}_k, \mathbf{e}_{rs} \rangle \mathbf{\Pi} \mathbf{e}_{rs}, \quad (3.22)$$

$$\gamma(O_{k+1}^r \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)) = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma(O_k^r \xi(\mathbf{x}_k, \mathbf{x}_{k-1})) + b_{k+1}^r \mathbf{V}_r \mathbf{\Pi} \mathbf{q}_k, \quad (3.23)$$

$$\gamma(T_{k+1}^r(g) \xi(\mathbf{x}_{k+1}, \mathbf{x}_k)) = \mathbf{B}_{k+1} \mathbf{\Pi} \gamma(T_k^r(g) \xi(\mathbf{x}_k, \mathbf{x}_{k-1})) + g(y_{k+1}^g) b_{k+1}^r \mathbf{V}_r \mathbf{\Pi} \mathbf{q}_k. \quad (3.24)$$

Proof The proof is given in Appendices A and B, where $N = N^1 \times N^2$.

Similar to equation (3.11), the unnormalized filter estimates of $\gamma(J_k^{rst})$, $\gamma(O_k^{rs})$, $\gamma(O_k^r)$ and $\gamma(T_k^r(g))$ can be determined by taking the inner products with $\mathbf{1}_{N^2}$. For example,

$$\gamma(J_k^{rst}) = \langle \gamma(J_k^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1})), \mathbf{1}_{N^2} \rangle.$$

The normalized estimates are obtained by dividing $\gamma(J_k^{rst})$ by $\gamma(1)$, that is, $\hat{J}_k^{rst} = \gamma(J_k^{rst})/\gamma(1)$.

We now briefly illustrate the EM algorithm for estimating the optimal parameters using the filters in Proposition 3.3.1. The parameters in our model are given by the set

$$\theta = \{a_{rst}, f_r, \sigma_r, 1 \leq r, s, t \leq N\}.$$

The EM method provides an estimation technique based on two stages: the expectation and maximization. The algorithm proceeds by initially selecting any set of parameters, denoted by θ_0 , for the model. The change to a new set of parameters is described by a change of probability measure from P_0 to P_θ . This means that the likelihood function for estimating the parameter θ based on the given information \mathcal{Y} is

$$L(\theta) = E_0 \left[\frac{dP_\theta}{dP_0} \middle| \mathcal{Y} \right].$$

The logarithm of the Radon-Nykodým derivative of the new measure with respect to the old measure is then calculated. A set of parameters that maximize the conditional log-likelihood is then determined. It is shown in [24] that the sequence of the estimated log-likelihood is monotonically increasing and the sequence of estimates converges to a local maximum of the expectation of the estimated likelihood function. This method provides a self-tuning approximation of the maximum likelihood estimate. As an example, let us consider the case of estimating the transition matrix. Note that the non-zero entries of $\mathbf{\Pi}$ are the same as the entries of \mathbf{A} . We estimate the matrix \mathbf{A} then construct $\mathbf{\Pi}$ for the calculation of filters. We first perform

a change of measure from P_θ to $P_{\hat{\theta}}$ for this method. Under P_θ , \mathbf{x} is a WMC with transition matrix $\mathbf{A} = (a_{rst})$. In [16], it is proved that under $P_{\hat{\theta}}$, \mathbf{x} is still a WMC and the transition matrix is $\hat{\mathbf{A}} = (\hat{a}_{rst})$. To replace \mathbf{A} by $\hat{\mathbf{A}}$, the likelihood function is

$$\frac{dP_\theta}{dP_0} \Big|_{\mathcal{B}_k} = \Gamma_k^A,$$

$$\Gamma_k^A = \prod_{l=2}^k \prod_{r,s,t=1}^N \left(\frac{\hat{a}_{rst}}{a_{rst}} \right)^{\langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle}.$$

In case $a_{rst} = 0$, take $\hat{a}_{rst} = 0$ and $\hat{a}_{rst}/a_{rst} = 1$. The estimates are expressed in terms of the recursions provided in equations (3.21)-(3.24) and given in the following proposition.

Proposition 3.3.2 *Suppose the observation is d -dimensional and the set of parameters $\{\hat{a}_{rst}, \hat{\mathbf{f}}_r, \hat{\sigma}_r\}$ determines the dynamics of $y_k, k \geq 1$. Then the EM estimates for these parameters are given by*

$$\hat{a}_{rst} = \frac{\hat{J}_k^{rst}}{\hat{O}_k^{st}} = \frac{\gamma(J_k^{rst})}{\gamma(O_k^{st})}, \quad \forall \text{ pairs } (r, s), r \neq s, \quad (3.25)$$

$$\hat{f}_r = \frac{\hat{T}_k^r}{\hat{O}_k^r} = \frac{\gamma(T_k^r(y))}{\gamma(O_k^r)}, \quad (3.26)$$

$$\hat{\sigma}_r = \sqrt{\frac{\hat{T}_k^r(y^2) - 2\hat{f}_r \hat{T}_k^r(y) + \hat{f}_r^2 \hat{O}_k^r}{\hat{O}_k^r}}. \quad (3.27)$$

Proof See Appendix C.

Given the observation up to time k , new parameters $\hat{a}_{rst}(k), \hat{f}_r(k), \hat{\sigma}_r(k), 1 \leq r, s, t \leq N$ are provided by equations (3.25)-(3.27). The recursive filters for the unobserved Markov chain and related processes in Proposition 3.3.1 can easily get updated every time new information arrives. Thus, the parameter estimation is self-tuning.

3.4 A simulation study

In this section, we present a simulation study to illustrate the filtering technique when the drift and volatility of a process are governed by independent hidden WMCs.

An example of model setup

Suppose we are given a set of data generated from a process with a 3-state drift and 2-state volatility: $\mathbf{f} = (f_1, f_2, f_3)$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$. The re-formulated drift and volatility are

$$\boldsymbol{\alpha} = \mathbf{f} \otimes \mathbf{1}_2 = (f_1, f_1, f_2, f_2, f_3, f_3)$$

$$\boldsymbol{\eta} = \mathbf{1}_3 \otimes \boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_1, \sigma_2, \sigma_1, \sigma_2).$$

Note that the new WMC \mathbf{x} has 6 states. Instead of estimating all values in $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$, we only estimate \mathbf{f} and $\boldsymbol{\sigma}$, then reformulate $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ for the recursive filters. In this way, the algorithm estimates fewer parameters than the actual 6-state model, but it is rich enough to capture all information.

The steps of the algorithm are as follows:

1. Initialize \mathbf{f} , $\boldsymbol{\sigma}$, \mathbf{A}_1 and \mathbf{A}_2 .
2. Construct $\boldsymbol{\alpha}$, $\boldsymbol{\eta}$, \mathbf{A} and $\boldsymbol{\Pi}$.
3. Calculate the filters in Proposition 3.3.1 using $\boldsymbol{\alpha}$, $\boldsymbol{\eta}$, \mathbf{A} and $\boldsymbol{\Pi}$.
4. After a batch of y_k values, compute new estimates of \mathbf{f} , $\boldsymbol{\sigma}$ and \mathbf{A} using the recursive filters in Proposition 3.3.2.
5. Construct the new $\boldsymbol{\alpha}$, $\boldsymbol{\eta}$ and $\boldsymbol{\Pi}$, and use these estimates as the initial values for the next batch of data points. Repeat from step. 3

The algorithm allows us to generate new estimates when the new information arrives. Therefore the model is considered as self-tuning.

We illustrate the proposed scheme using simulated data. For the simulated observation data, two sets of 1000 point WMCs were generated with the following parameter values:

$$\mathbf{A}_1 = \begin{pmatrix} 0.8 & 0.8 & 0.8 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.8 & 0.8 & 0.8 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.8 & 0.8 & 0.8 \end{pmatrix},$$

$$\mathbf{A}_2 = \begin{pmatrix} 0.7 & 0.7 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.7 & 0.7 \end{pmatrix}.$$

The initial states of both WMCs are state 1. The true values of drift and volatility are $\mathbf{f} = (0.04, 0, -0.02)$ and $\boldsymbol{\sigma} = (0.02, 0.05)$, respectively. The dynamics in equation (3.3) yield 1000 simulated observation points which can be considered as daily returns of an asset price. Our on-line algorithm runs in batches consisting of 10 data points and produces a set of updated parameter estimates. We call one batch of data as one complete algorithm pass or step. In our simulation study, 100 algorithm steps were run and the parameters were updated every two weeks.

An important aspect to consider when implementing the EM algorithm is to determine the number of states. Erlwein and Mamon [8] determined the optimal number of regimes using the Akaike information criteria and found that a two-state HMM outperforms other multi-state HMMs in capturing the dynamics of the Canadian short rates proxied by the 30-day T-bill yields. Indeed, the number of states can be any reasonable value indicated by the data. In our case, we run the algorithm using different choices for the number of states to obtain the parameter estimates. In order to compare the performance of the models, we compare the 1-step ahead forecasts together with different goodness-of-fit measures (e.g. RMSE).

Figure 3.4 displays the evolution of \mathbf{f} and $\boldsymbol{\sigma}$ estimates under various WHMM settings. Initial

values of \mathbf{f} and $\boldsymbol{\sigma}$ are indicated below the corresponding plots. These values are chosen based on the true parameter values. All initial entries of the transition matrix, \mathbf{A} , are set to be $1/N$. The parameters become stable after approximately 5 steps in plots 3.1(a)-3.1(c). In plot 3.1(d), convergence is achieved after 25 steps. Our experiment shows that the convergence can be achieved with other valid starting values. However, the choice of initial parameter values and the model setting affect the speed of convergence.

The semi-martingale representation of a WMC in equation (3.1) and the definition of \mathbf{A} lead to

$$E[\mathbf{x}_{k+h}|\mathcal{Y}_k] = \mathbf{A}\boldsymbol{\Pi}^{h-1}\mathbf{p}_k, \quad h \geq 1.$$

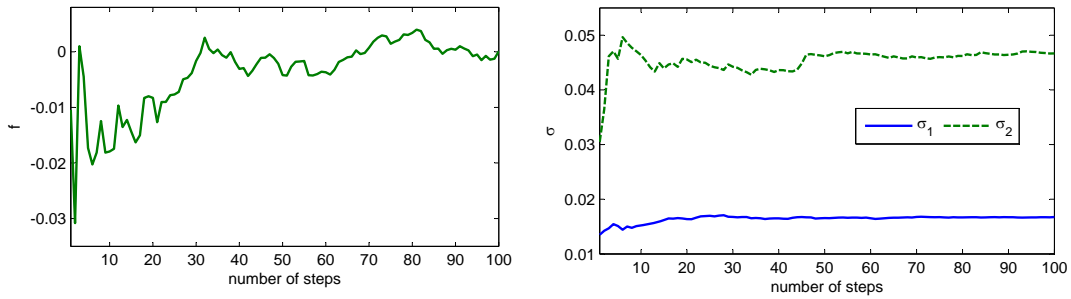
The h -step ahead forecasts of the logarithmic increment y_k is

$$E[y_{k+h}|\mathcal{Y}_k] = \langle \mathbf{f}, \mathbf{A}\boldsymbol{\Pi}^{h-1}\mathbf{p}_k \rangle,$$

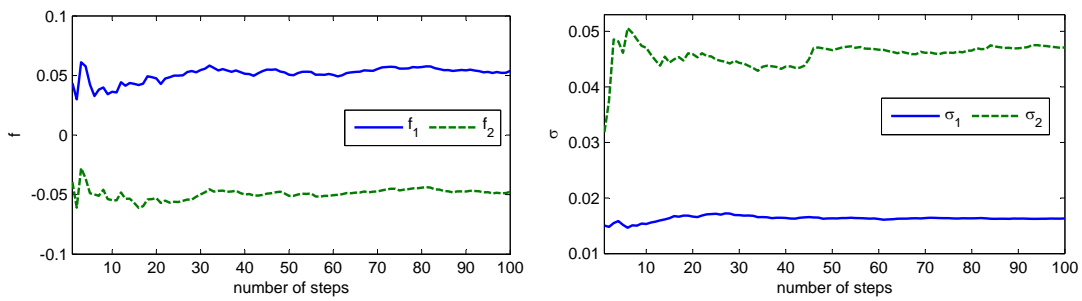
and the conditional variance of y_{k+h} is

$$\text{Var}[y_{k+h}|\mathcal{Y}_k] = \mathbf{f}^\top \text{diag}(\mathbf{A}\boldsymbol{\Pi}^{h-1}\mathbf{p}_k)\mathbf{f} + \boldsymbol{\sigma}^\top \text{diag}(\mathbf{A}\boldsymbol{\Pi}^{h-1}\mathbf{p}_k)\boldsymbol{\sigma} - \langle \mathbf{f}, \mathbf{A}\boldsymbol{\Pi}^{h-1}\mathbf{p}_k \rangle^2,$$

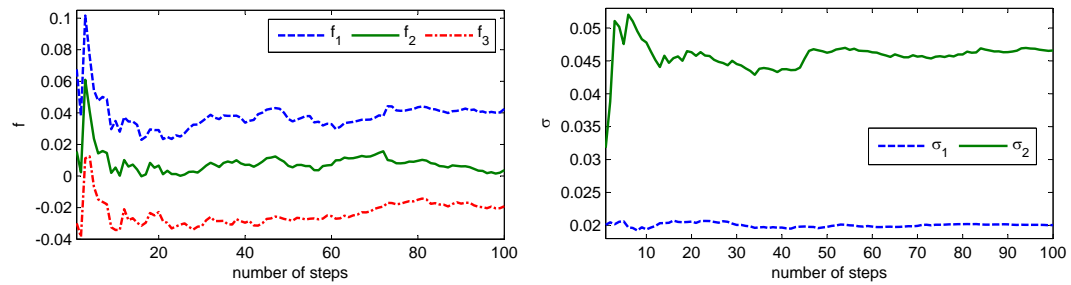
We compare the forecasting performance of WHMM with different initial settings. To assess the goodness-of-fit of the h -step ahead forecasts, we evaluate the RMSE, AME, RAE and APE for $h = 1$ and $h = 5$. These errors are reported in Tables 3.1 and 3.2. In both cases of $h = 1$ and $h = 5$, the model with 3-state drift and 2-state volatility gives the best fit in terms of lowest forecasting errors. Note that the 1-state drift and volatility model produces the highest errors in all metrics. It is because the single regime model is not able to capture the dynamics of the data series. Furthermore, the model with 2-state drift and 3-state volatility yields higher errors than all 2-state volatility models. Since the data is simulated using 2-state volatility, a model with more states will cause overestimation of parameters and produce large errors.



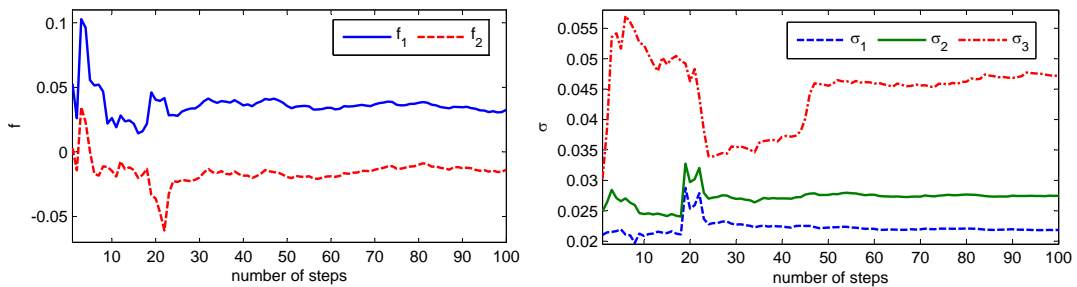
(a) 1-state drift and 2-state volatility with initial values: $f = 0$ and $\sigma = (0.01, 0.04)$



(b) 2-state drift and 2-state volatility with initial values: $\mathbf{f} = (0.03, -0.03)$ and $\sigma = (0.01, 0.03)$



(c) 3-state drift and 2-state volatility with initial values: $\mathbf{f} = (0.03, 0, -0.03)$ and $\sigma = (0.01, 0.03)$



(d) 2-state drift and 3-state volatility with initial values: $\mathbf{f} = (0.03, 0)$ and $\sigma = (0.01, 0.02, 0.03)$

Figure 3.1: Evolution of parameter estimates under different model settings

\mathbf{f}	σ	RMSE	AME	RAE	APE
1-state	1-state	1.3911	1.0233	0.1010	0.0440
1-state	2-state	1.0780	0.7001	0.0691	0.0282
2-state	3-state	1.0842	0.7010	0.0690	0.0819
2-state	1-state	1.0787	0.7006	0.0692	0.0820
2-state	2-state	1.0824	0.7088	0.0700	0.0285
3-state	2-state	1.0742	0.6854	0.0687	0.0279

Table 3.1: Comparison of 1-step ahead forecast errors

\mathbf{f}	σ	RMSE	AME	RAE	APE
1-state	1-state	2.6274	1.8770	0.1854	0.0775
1-state	2-state	2.4653	1.7449	0.1723	0.0705
2-state	3-state	2.4717	1.7477	0.1726	0.0706
2-state	1-state	2.4653	1.7450	0.1723	0.0705
2-state	2-state	2.4689	1.7470	0.1725	0.0706
3-state	2-state	2.4648	1.7442	0.1722	0.0704

Table 3.2: Comparison of 5-step ahead forecast errors

3.5 Conclusion

We extended a WHMM to model the evolution of a risky asset by considering the drift and volatility of the logarithmic increments governed by two independent hidden Markov chains. A tensor-based technique was employed to transform the two independent WMCs into a new WMC. Filtered estimates of the drift, volatility and the state of the new WMC were derived based on EM algorithm. Numerical examples on simulated data were given. Other than the true number of states, we also considered other combinations of state for drift and volatility and analyzed forecasting errors. Our empirical results suggest that using the number of states and the initial values indicated by the observation data will make the model perform better. Indeed, the starting values affects the algorithm performance. This is why we determine the correct number of states for each parameter when implementing the estimation filters on market observations via the RMSEs. It is worth an exploring the effectiveness and efficiency of the proposed filtering and parameter estimation technique on other models in finance and economics.

References

- [1] A. Ang and A. Timmermann. Regime changes and financial markets. *Annual Review of Financial Economics*, 4, 2012. forthcoming.
- [2] J. Bulla and I. Bulla. Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics and Data Analysis*, 51:2192–2209, 2006.
- [3] L. E. Calvet and A. J. Fisher. Multifrequency news and stock returns. *Journal of Financial Economics*, 86:178–212, 2007.
- [4] W. K. Ching, T. K. Siu, and L. M. Li. Pricing exotic options under a higher-order Markovian regime switching model. *Journal of Applied Mathematics and Decision Sciences*, pages 1–15, 2007. Article ID 18014.
- [5] R. Elliott, J. Moore, and L. Aggoun. *Hidden Markov Models: Estimation and Control*. Springer-Verlag, New York, 1995.
- [6] R. J. Elliott, W. P. Malcolm, and A. H. Tsoi. Robust parameter estimation for asset price models with Markov modulated volatilities. *Journal of Economics Dynamics and Control*, 27:1391–1409, 2003.
- [7] R. J. Elliott, T. K. Siu, and A. Badescu. On mean-variance portfolio selection under a hidden Markovian regime-switching model. *Economic Modelling*, 27:678–686, 2010.
- [8] C. Erlwein and R. Mamon. An online estimation scheme for a Hull-White model with HMM-driven parameters. *Statistical Methods and Applications*, 18(1):87–107, 2009.

- [9] C. Erlwein, R. Mamon, and M. Davison. An examination of HMM-based investment strategies for asset allocation. *Applied Stochastic Models in Business and Industry*, 27:204–221, 2011.
- [10] J. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–84, 1989.
- [11] M. R. Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5:41–53, 2001.
- [12] J. Hunt and P. Devolder. Semi-Markov regime switching interest rate models and minimal entropy measure. *Physica A: Statistical Mechanics and its Applications*, 390:3767–3781, 2011.
- [13] T. Lange and A. Rohbek. An introduction to regime switching time series model. In *Handbook of Financial Time Series*, pages 871–887, Berlin, Heidelberg, 2009. Springer.
- [14] B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and Their Applications*, 40:127–143, 1992.
- [15] C. C. Liew and T. K. Siu. A hidden Markov-switching model for option valuation. *Insurance: Mathematics and Economics*, 47:374–384, 2010.
- [16] S. Luo and A. H. Tsoi. Filtering of hidden weak Markov chain-discrete range observation. In R.S. Mamon and R.J. Elliott, editors, *Hidden Markov Models in Finance*, pages 106–119. Springer, New York, 2007.
- [17] R. S. Mamon, C. Erlwein, and R. B. Gopaluni. Adaptive signal processing of asset price dynamics with predictability analysis. *Information Sciences*, 178:203–219, 2008.
- [18] R. S. Mamon and L. Jalen. Parameter estimation in a regime-switching model when the drift and volatility are independent. In *Proceedings of 5th International Conference*

on Dynamic Systems and Applications, pages 291–298, Atlanta, Georgia, USA, 2008. Dynamic Publishers, Inc.

- [19] A. R. Pagan and G. W. Schwert. Alternative models for conditional stock volatility. *Journal of Econometrics*, 45:267–290, 1990.
- [20] T. Rydén, T. Terasvirta, and S. Asbrink. Stylized facts of daily return series and the hidden Markov model. *Journal Applied Econometric*, 13:217–244, 1998.
- [21] T. Siu, W. Ching, E. Fung, M. Ng, and X. Li. A high-order Markov-switching model for risk measurement. *Computers and Mathematics with Applications*, 58:1–10, 2009.
- [22] J. S. Tyssedal and D. Tjostheim. An autoregressive model with suddenly changing parameters and an application to stock market prices. *Applied Statistics*, 37:353–369, 1988.
- [23] P. Veronesi. How does information affect stock returns? *Journal of Finance*, 55:807–837, 2009.
- [24] C. Wu. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11:95–103, 1983.
- [25] X. Xi and R. S. Mamon. Parameter estimation of an asset price model driven by a weak hidden Markov chain. *Economic Modelling*, 28:36–46, 2011.
- [26] S. Yu, Z. Liu, M. S. Squillante, C. H. Xia, and L. Zhang. A hidden semi-Markov model for web workload self-similarity. In *Proceedings of 21st IEEE International Performance, Computing, and Communications Conference*, pages 65–72, Phoenix, AZ, 2002.
- [27] N. Zhou and R. Mamon. An accessible implementation of interest rate models with Markov-switching. *Expert Systems with Applications*, 39(5):4679–4689, 2012.

Chapter 4

A weak hidden Markov chain-modulated model for asset allocation

4.1 Introduction

This chapter extends the WHMM featured in chapter 2 to its multivariate version. We suppose that we have a multivariate financial time series and each series is driven by the same WMC although the white noise-driving process is different for every data series. Filtering algorithms are developed and we apply our results in the evaluation of several competing trading strategies within the context of an asset allocation problem.

It is well documented in the asset allocation literature that the inclusion of market regime-switching dynamics has considerable impact on the optimal portfolio strategy of individual investors. In financial portfolio management, the portfolio risk cannot be entirely eliminated although it can be controlled with an optimal asset allocation strategy combining different types and amount of investment. Each investment has its own unique risk and return characteristics. A well-designed investment aims at the maximization of expected portfolio return while controlling the level of risk. A fundamental example of the single-period mean-variance asset allocation problem is given by the Nobel prize-winning work of Markowitz [24], in which the

variance is employed as a measure of risk and the efficient allocation of wealth among different investment classes is provided. Since practical asset allocation problems involve intertemporal decisions, Samuelson [30] and Merton [26] considered asset allocation problem in a multi-period model and in a continuous-time model, respectively. In particular, Merton utilized stochastic optimal control theory to derive a closed-form solution for an asset allocation strategy under certain assumptions. A key assumption of these early works in the literature is that the dynamics of asset returns are linear processes with constant coefficients. However, the state of the economy and the financial market varies randomly over time. Investors are concerned with regime-switching uncertainty affecting the portfolio return. Such uncertainty affects the future payoffs and therefore could alter the optimal asset allocation. Ang and Bekaert [3] introduced a regime-switching model with time-varying correlations and volatilities for asset allocation. They reported evidence of shifting regimes in the US, UK and German equity markets. In a subsequent study, Ang and Bekaert [4] expanded the lists of markets and assets for investigation of optimal asset allocation under a regime-switching framework. Their out-of-sample test shows that the regime-switching strategy dominates a non-regime dependent strategy. Bauer, et al. [5] observed the tendency of changing correlations and volatility among assets, and considered a regime-switching technique for portfolio optimization. An out-of-sample backtesting was applied on a six-asset portfolio consisting of equities, bonds, commodities and real estate. The results demonstrated a significant information gain by using a regime-switching strategy. Guidolin and Timmermann [18] considered asset allocation decisions under a regime-switching model for asset returns with four separate regimes. It was found that the optimal allocations vary considerably across these states and change over time as investors revise their estimate of the state probabilities.

Various works that support model assumptions in which parameters change over time in accordance with the evolution of an unobserved Markov chain have been proposed. Elliott and van der Hoek [15] put forward a model for the rates of asset returns driven by a Markov chain

in discrete time. Their work features filtering and prediction techniques in the model identification and outlines how their method could be applied to the asset allocation problem using mean-variance type utility criterion. Graflund and Nilsson [17] investigated dynamic portfolio selection within a Markovian switching framework. Their results highlight the economic importance of regimes and suggests that ignoring the regime will require significant compensation. In the study of Ammann and Verhofen [2], Markov Chain Monte Carlo methods were applied to estimate a multivariate regime-switching model. Two clearly separable regimes characterized by different mean returns, volatilities and correlations were found. The results of their out-of-sample backtest suggests that the buy-and-hold strategy based on regime-switching model can be profitable. Bulla, et al. [7] focused on daily stock market return series at five major regional markets over the last four decades. They presented an out-of sample performance analysis with transaction costs taken into account and concluded that the strategy is improved by considering a Markovian switching model. Erlwein, et al. [16] developed and compared investment strategies in allocating funds to either growth or value stocks, whose price dynamics are driven by a hidden Markov model (HMM). Their investigation shows that the HMM-based strategies are more stable and outperform the pure growth strategy in terms of higher Sharpe ratios and lower variance of the performance. Elliott, et al. [14] considered a mean-variance portfolio selection problem where the appreciation rate of the risky asset is modulated by a continuous-time Markov chain. They employed the gauge transformation technique to obtain robust filters and developed the filter-based EM algorithm in calculating the estimates of the unknown parameters. An explicit solution to the mean-variance portfolio problem is derived using the filtering results.

While the original HMM is able to capture the switching of market or economy states, and hence leads to its widespread use, there is growing evidence that many financial series possess memories. There has been considerable interest in the study of long-range dependence present in the state of economy. In long-range dependent processes, the coupling between values at

different times is stronger than that of short-range dependent processes, and hence it exhibits a slow decay of autocorrelation. The analysis of long memory was pioneered by Hurst [20]. Long-range dependence is often presented by models that have been devised to be self-similar, and they can be described by heavy-tailed distributions. The presence of long memory in asset prices and returns has important implications in modern financial economics. A number of studies have investigated the presence of long-range dependence in financial series. Lobato and Savin [21], and Ray and Tsay [28] found evidence of long-range dependence in the volatility of returns on S&P500. McCarthy, et al. [25] found evidence of long memory in percentage change in yields on Treasury debt securities. The dependence properties of asset returns have motivated many researchers to develop stochastic models for this phenomenon. A study by Dajcman [11] examined a time varying long memory parameter for eight European stock market returns by using an auto regressive fractionally integrated moving average model. Couillard and Davison [10] noted that caution must be taken in statistical tests of long range Hurst process like dependence. Maheu [23] suggested that GARCH models can in some circumstances account for the long-memory property found in financial market volatility. There is a vast literature in modeling long-range dependence using single-state stochastic models. Yu, et al. [34] developed a recursive formula for estimating the Hurst parameter for a second-order HMM. Their numerical experiment on Web server workload showed that the second-order HMM can capture long-range dependence properties when the distribution of at least one state is heavy tailed. As an extension of the HMM, higher-order HMM has attracted more attention because of its capacity to a capture memory effect. It is suggested by Rydén, et al. [29] that a HMM cannot described the stylized fact of the very slowly decay in autocorrelation of returns. Bulla and Bulla [6] explored the goodness of fit of two WHMMs to daily return series from 18 pan-European sector indices. Their analysis shows a significantly improved fit of the autocorrelation and hence illustrates that the long-range dependence can be described better by WHMM. In this chapter we consider a WHMM, in particular a second-order HMM, for an asset allocation problem. As mentioned in Solberg [32], the real significance of WHMM

is to establish that the Markov chain assumption is not really as restrictive as it first appears. One is not limited to a dependence on just one prior time epoch but can make the dependency extend to any finite number of prior epochs, thereby capturing more information from the past. This, in turn, widens the literature on models that aim to reflect long-range dependence in financial models. In the WHMM model the transition matrix between one state and the next state is itself dependent on the information in the prior states. An n th-order Markov chain is dependent on the prior n states. The higher the order, the more extended the dependency, and therefore more information from the past can be reflected. Xi and Mamon [33] proposed a WHMM for discrete-time continuous-range observations and provided a detailed implementation of the model to a financial dataset. Hess [19] considered conditional CAPM strategies based on regime forecasts from an autoregressive Markov regime-switching behavior with lag two. The improvement of the portfolio performance by using the proposed strategy is examined through in-sample and out-of-sample analyses. An application of higher-order Markovian switching models for risk measurement is presented by Siu, et al. [31]. Another application of WHMM, on option pricing can be found in Ching, et al. [9].

In this chapter, we investigate optimal investment strategies for asset allocation under a weak Markov-switching framework. In particular, we assume the log returns of risky assets are modulated by a second-order multivariate Markov chain, whose current behavior depends on its behavior at the previous two time steps. The states of the weak Markov chain are interpreted as states of the economy. Compared to our previous research in [33], we extend the single-variate WHMM to a multivariate case by modifying the Radon-Nikodým derivative. This extension allows us to investigate the application of WHMM involving multiple financial series, such as those occurring in asset allocation. We use the same asset allocation strategies from [16]. Compared to their research, we relax the Markov assumption by increasing the first-order HMM to second-order HMM. The filtering technique for WHMM is implemented on updated market data, which includes the period of the subprime crisis. The numerical results show how a

WHMM captures information during the crisis period and affects the strategy. The WHMM has the advantage that it can capture the long-range dependence of the states of the market, and therefore it is more appropriate when memories are evident in financial series. From the investors' view, tactical investment decisions require the the evaluation of the expected future payoff on risky assets. More economic insights can be gained if relevant historical information can be incorporated into the unobservable market state; this will be beneficial to investors from both the economic and statistical perspectives. Although a higher-order Markov chain, more specifically a Markov chain of order higher than two, leads to more information incorporated in the HMM, the number of model parameters involved increases exponentially. Ching, et al. [8] apply a higher-order multivariate HMM to a sequence of multivariate categorical data and show that an n th-order, s -variate, N -state Markov chain model requires ns^2N^2 parameters. To facilitate the dynamic estimation of this huge number of parameters, we use a transformation that converts a WHMM into a regular HMM thereby enabling the estimation algorithm to perform smoothly. The transformation, which is essentially a mapping of states, is employed to eventually recover the required number of parameters. Our asset allocation strategies rely on the estimates of parameters and forecasted returns through the mathematical techniques of WHMM's.

This chapter is organized as follows. In section 4.2, we present the multidimensional WHMM filtering and estimation techniques. WHMM filtering procedure is applied to the Russell 3000 value and growth indices data, whose logreturns are assumed to follow a normal distribution with regime-switching dependent on two previous time epochs. The EM algorithm is then applied to obtain the online recursive estimates of the model parameters. In section 4.3, we utilize the optimal estimates to forecast the two indices and conclude that a two-state WHMM is sufficient to capture the characteristics of our data based on four error metrics. We investigate an investment strategy switching between the Russell 3000 Value and Growth indices in section 4.4. The switching decision determined by the one-step ahead forecasts return of each index.

In section 4.5, a mixed investment strategy is considered. The optimal weights of investment between the two indices are obtained by solving a mean-variance problem under the regime-switching setting. The estimation of the optimal weights incorporates the parameter estimates as well as the states of a weak Markov chain. Portfolio performance is investigated in section 4.6, where we use three classical measures for benchmarking. Furthermore, a bootstrap analysis is used to compare the stability of portfolios with various level of transaction costs. Section 4.7 concludes the chapter.

4.2 Filtering and parameter estimation

Let (Ω, \mathcal{F}, P) be a complete probability space under which \mathbf{x}_k is a Markov chain with finite-state space in discrete time ($k = 0, 1, 2, \dots$). To simplify the discussion and present a complete characterization of the parameter estimation, we only consider a weak Markov chain of order 2. That is,

$$\begin{aligned} P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \dots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k) \\ = P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k). \end{aligned}$$

Without loss of generality, the N -state weak Markov chain takes value from the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\} \subset \mathbb{R}^N$, where \mathbf{e}_i is the vector with unity in the i th element and zero elsewhere. We interpret $\langle \mathbf{x}_k, \mathbf{e}_i \rangle$ as the event that the economy is in the state i at time k . Here $\langle \cdot, \cdot \rangle$ denotes the inner product in \mathbb{R}^N . The element a_{lmv} , $l, m, v \in 1, \dots, N$, of the transition probability matrix \mathbf{A} refers to the probability that the process enters state l given that the current state is the m th state and the previous state was in v .

Instead of studying the weak Markov chain directly, we introduce a mapping ξ , to embed the second-order Markov chain into the first-order Markov chain, and then apply the regular filtering method. This idea is an analogy to the embedding of higher-order ODEs into systems

of first-order ODEs and solving the system by regular methods, see Abell and Braslton [1]. The mapping ξ is defined by

$$\xi(\mathbf{e}_r, \mathbf{e}_s) = \mathbf{e}_{rs}, \text{ for } 1 \leq r, s \leq N,$$

where \mathbf{e}_{rs} is an \mathbb{R}^{N^2} -unit vector with unity in its $((r-1)N + s)$ th position. Note that

$$\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle = \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle$$

represents the identification of the new first-order Markov chain with the canonical basis. We also define the new $N^2 \times N^2$ transition probability matrix $\mathbf{\Pi}$ of the new Markov chain by

$$\pi_{ij} = \begin{cases} a_{lmv} & \text{if } i = (l-1)N + m, \quad j = (m-1)N + v \\ 0 & \text{otherwise.} \end{cases}$$

Note that at time k , each non-zero element π_{ij} represents the probability

$$\pi_{ij} = a_{lmv} = P(\mathbf{x}_k = \mathbf{e}_l | \mathbf{x}_{k-1} = \mathbf{e}_m, \mathbf{x}_{k-2} = \mathbf{e}_v),$$

and each zero represents an impossible transition. It is known [31] that the new Markov chain $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ has a semi-martingale representation

$$\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi} \xi(\mathbf{x}_{k-1}, \mathbf{x}_{k-2}) + \mathbf{v}_k, \quad (4.1)$$

where $\{\mathbf{v}_k\}_{k \geq 1}$ is a sequence of \mathbb{R}^{N^2} martingale increments.

In a comprehensive monograph on HMM, MacDonald and Zucchini [22] devoted a section of introducing WHMM and give a detailed example of transforming a second-order two-state Markov chain into a regular two-state Markov chain. An efficient recursive algorithm for

computing the likelihood from consecutive observations under a second-order HMM is given. However, no application of the second-order WHMM MLE is given in this book. Du Preez, et al. [27] developed a computing algorithm to reduce any higher-order HMM to a corresponding first-order HMM. The algorithm is applied to language recognition. In contrast to their research objectives, we focus on financial time series applications. In particular, we obtain the reduced first-order HMM and estimate parameters of logreturns of asset prices.

Let $\mathbf{y}_k = (y_k^1, y_k^2, \dots, y_k^d)$ be a d -dimensional process. Each component y_k^g , $1 \leq g \leq d$, is the sequence of log returns of an asset price with the dynamics

$$y_{k+1}^g = f^g(\mathbf{x}_k) + \sigma^g(\mathbf{x}_k)z_{k+1}^g.$$

Here, each $\{z_k^g\}$ is a sequence of $N(0, 1)$ IID random variables and independent of \mathbf{x} . The function f^g and σ^g are determined by the vectors $\mathbf{f}^g = (f_1^g, f_2^g, \dots, f_N^g)^\top$ and $\boldsymbol{\sigma}^g = (\sigma_1^g, \sigma_2^g, \dots, \sigma_N^g)^\top$ in \mathbb{R}^N , and $f^g(\mathbf{x}_k) = \langle \mathbf{f}^g, \mathbf{x}_k \rangle$ and $\sigma^g(\mathbf{x}_k) = \langle \boldsymbol{\sigma}^g, \mathbf{x}_k \rangle$ represent the mean and variance of y_k^g , respectively and \top denotes the transpose of a matrix. Note that all components of the vector observation process have the same underlying weak Markov chain. In this chapter, we do not model the correlation among assets explicitly. However, the two asset prices are governed by the same hidden WMC, and thus they are correlated implicitly. Actual filter with correct correlation structure will presumably be better. So, one may view that this study is a lower bound for the validity of a larger study.

It must be noted that we do not observe the underlying weak Markov chain from the financial market directly. Under the real world measure P , the state \mathbf{x}_k is contained in the noisy observations $\mathbf{y}_k, k \geq 1$. We aim to “filter” the noise out of the observations. By the Kolmogorov Extension Theorem, there exists a reference probability measure \bar{P} , under which the observation \mathbf{y}_k are multivariate $N(0, 1)$ IID random variables and therefore \bar{P} is deemed to be an easier measure to work with. The filters are derived under the reference measure. We perform a mea-

sure change to construct the real-world measure P from the ideal-world measure \bar{P} by invoking a discrete-time version of Girsanov's theorem. Let $\phi(z)$ denote the probability density function of a standard normal random variable z . For each component g , define

$$\lambda_l^g = \frac{\phi(\sigma^g(\mathbf{x}_{l-1})^{-1}(y_l^g - f^g(\mathbf{x}_{l-1})))}{\sigma^g(\mathbf{x}_{l-1})\phi(y_l^g)},$$

and the Radon-Nikodým derivative of P with respect to \bar{P} , $\frac{dP}{d\bar{P}}|_{\mathcal{Y}_k} = \Lambda_k$, is given by

$$\Lambda_k = \prod_{g=1}^d \prod_{l=1}^k \lambda_l^g, \quad k \geq 1, \quad \Lambda_0 = 1.$$

To obtain the estimates of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under the real world measure, we first perform all calculations under the reference probability measure \bar{P} . Then, Bayes' theorem is employed to relate the conditional expectation under two different measures. Note that we can also consider another reference probability measure \tilde{P} under which the y_k^g are $N(0, \sigma^2)$, $\sigma \neq 1$. In that case, we define

$$\tilde{\lambda}_l^g = \frac{\phi(\sigma^g(\mathbf{x}_{l-1})^{-1}(y_l^g - f^g(\mathbf{x}_{l-1})))}{\phi(\sigma^g(\mathbf{x}_{l-1})^{-1}y_l^g)}.$$

Based on our numerical experiment, since $\tilde{\lambda}$ is much larger than λ , the speed of convergence with $\tilde{\lambda}$ is 10 steps slower than using λ .

Let us derive the conditional expectation of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ given \mathcal{Y}_k under P . Write

$$p_k^{ij} = P(\mathbf{x}_k = \mathbf{e}_i, \mathbf{x}_{k-1} = \mathbf{e}_j | \mathcal{Y}_k) = E[\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij} \rangle | \mathcal{Y}_k], \quad (4.2)$$

with $\mathbf{p}_k = (p_k^{11}, \dots, p_k^{ij}, \dots, p_k^{NN}) \in \mathbb{R}^{N^2}$. Using Bayes' theorem, we have

$$\mathbf{p}_k = E[\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_k] = \frac{\bar{E}[\Lambda_k \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_k]}{\bar{E}[\Lambda_k | \mathcal{Y}_k]}. \quad (4.3)$$

1. J^{rst} , the number of jumps from $(\mathbf{e}_s, \mathbf{e}_t)$ to state \mathbf{e}_r up to time k .
2. O_k^{rs} , the occupation time of the weak Markov chain spent in state $(\mathbf{e}_r, \mathbf{e}_s)$ up to time k ,
3. O_k^r , the occupation time spent by the weak Markov chain in state \mathbf{e}_r up to time k ,
4. $T_k^r(h)$, the level sum for the state \mathbf{e}_r , where h is a function with the form $h(y) = y$ or $h(y) = y^2$.

To obtain on-line estimates for the quantities of the above four related process, we shall take advantage of the semi-martingale representation in (4.1) and result in (4.7). We can then obtain recursive equations for the vector quantities $J_k^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^r\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ and $T_k^r(h)\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$. The recursive relation of these vector processes and \mathbf{q}_k under a multi-dimensional observation set-up are given in the following proposition.

Proposition 4.2.1 *Let $\mathbf{V}_r, 1 \leq r \leq N$ be an $N^2 \times N^2$ matrix such that the $((i-1)N+r)$ th column of \mathbf{V}_r is \mathbf{e}_{ir} for $i = 1 \dots N$ and zero elsewhere. If \mathbf{B} is the diagonal matrix defined in equation (4.5), then*

$$\mathbf{q}_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k \quad (4.8)$$

and

$$\begin{aligned} \gamma(J^{rst}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(J^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\ &\quad + b_{k+1}^r \langle \mathbf{\Pi}\mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \mathbf{q}_k, \mathbf{e}_{st} \rangle \mathbf{e}_{rs}, \end{aligned} \quad (4.9)$$

$$\begin{aligned} \gamma(O^{rs}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\ &\quad + b_{k+1}^r \langle \mathbf{q}_k, \mathbf{e}_{rs} \rangle \mathbf{\Pi}\mathbf{e}_{rs}, \end{aligned} \quad (4.10)$$

$$\begin{aligned} \gamma(O^r\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^r\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\ &\quad + b_{k+1}^r \mathbf{V}_r \mathbf{\Pi}\mathbf{q}_k, \end{aligned} \quad (4.11)$$

$$\begin{aligned} \gamma(T^r(h)\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(T^r(h)\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\ &\quad + h(y_{k+1}^g) b_{k+1}^r \mathbf{V}_r \mathbf{\Pi}\mathbf{q}_k. \end{aligned} \quad (4.12)$$

Proof See Appendices A and B for an analogous proof of each estimate under the single observation setting. ■

Similar to equation (4.4), by summing the components, equations (4.9) to (4.12) give expressions for $\gamma(J^{rst})_k$, $\gamma(O^{rs})_k$, $\gamma(O^r)_k$ and $\gamma(T^r(g))_k$.

Now we make use of the EM algorithm to estimate the optimal parameters. The calculation is similar to the technique as in single observation set-up. The estimates are expressed in terms of the recursions in equations (4.9)-(4.12), which are provided in the following proposition.

Proposition 4.2.2 *Suppose the observation is d -dimensional and the set of parameters $\{\hat{a}_{rst}, \hat{f}_r^g, \hat{\sigma}_r^g\}$ determines the dynamics of y_k^g , $k \geq 1$, $1 \leq g \leq d$, then the EM estimates for these parameters are given by*

$$\hat{a}_{rst} = \frac{\hat{J}_k^{rst}}{\hat{O}_k^{st}} = \frac{\gamma(J^{rst})_k}{\gamma(O^{st})_k}, \quad \forall \text{ pairs } (r, s), \quad r \neq s, \quad (4.13)$$

$$\hat{f}_r^g = \frac{\hat{T}_k^r}{\hat{O}_k^r} = \frac{\gamma(T^r(y^g))_k}{\gamma(O^r)_k}, \quad (4.14)$$

$$\hat{\sigma}_r^g = \sqrt{\frac{\hat{T}^r((y^g)^2)_k - 2\hat{f}_r^g \hat{T}^r(y^g)_k + (\hat{f}_r^g)^2 \hat{O}_k^r}{\hat{O}_k^r}}. \quad (4.15)$$

Proof See Appendix C for an analogous proof of each estimate under the single observation setting. ■

Given the observation up to time k , new parameters $\hat{a}_{rst}(k)$, $\hat{f}_r^g(k)$, $\hat{\sigma}_r^g(k)$, $1 \leq r, s, t \leq N$ are given by equations (4.13)-(4.15). The recursive filters for the unobserved Markov chain and the related process in Proposition 4.2.1 can be re-evaluated using the new estimates. Consequently, it allows the algorithm to update the parameters automatically.

4.3 Forecasting indices

Suppose an investor wants to choose a portfolio with two investments to diversify his/her risk. In order to have such diversification the two assets should act differently during different periods in the economic cycle. For example, growth and value stocks tend to perform well at different times of the economic cycle, so switching between the classes at appropriate times may add value. We apply the iterative procedure derived in the previous section to two weekly datasets of stock indices: Russell 3000 growth and Russell 3000 value indices. The data were recorded from June 1995 to December 2010; thus there are 783 data points in each dataset. Both indices are constructed based on the Russell 3000 index, in which the underlying companies are all incorporated in the U.S and representing approximately 98% of the investable U.S. equity market. Companies within the Russell 3000 that exhibit higher price-to-book and forecasted earnings are used to form the Russell 3000 growth index. This subindex therefore measures the performance of the broad growth segment of the US equity market. The Russell 3000 value index includes Russell 3000 companies with lower price-to-book value and lower forecasted growth values. Therefore the Russell 3000 value index measures the performance of the value stocks in the US equity market.

The regime-switching models are developed to capture particular behavior of the evolution of an asset price. We segregate the observation data into four intervals to investigate the index values and returns. Tables 4.1-4.3 provide descriptive statistics of the Russell 3000 Index together with the growth- and value-subindices for the entire period as well as the subperiods. The descriptive statistics demonstrate the possible segregation of the actual data into different states according to the levels of mean and volatility. We find the subperiods characterized by different levels of mean and volatility. For example, we can see that the log return y_k has a higher volatility when the mean is negative, and vice versa. If the data has only one state, the model will collapse to one regime. As a result, the estimated parameters of each state will be close to each other.

	Entire data	06/95-07/98	07/98-09/03	09/03-08/08	09/05/08-12/31/10
Max	0.1659	0.0669	0.1659	0.0429	0.1090
Min	-0.1806	-0.0462	-0.1683	-0.0522	-0.1806
Median	0.0026	0.0052	0.0007	0.0012	0.0034
Mean	0.0010	0.0048	-0.0008	0.0009	0.0005
Std	0.0299	0.0196	0.0379	0.0184	0.0399
Skewness	-0.4478	-0.0281	-0.0721	-0.3730	-0.8401
Kurtosis	5.0859	0.1581	2.6943	0.2936	3.7011

Table 4.1: Summary statistics of Russell 3000 growth returns

	Entire data	06/95-07/98	07/98-09/03	09/03-08/08	09/05/08-12/31/10
Max	0.1381	0.0554	0.0719	0.0615	0.1381
Min	-0.2167	-0.0551	-0.1162	-0.0609	-0.2167
Median	0.0025	0.0052	-0.0007	0.0028	0.0031
Mean	0.0011	0.0042	0.0001	0.0011	-0.0005
Std	0.0266	0.0168	0.0258	0.0187	0.0465
Skewness	-0.8031	-0.2649	-0.2585	-0.4757	-0.7317
Kurtosis	8.1906	0.5851	1.6751	0.7746	4.0000

Table 4.2: Summary statistics of Russell 3000 value return

We consider the two indices as a two-dimensional observation process. The dynamics of the log returns are given by

$$y_{k+1}^{RV} = \log \frac{RValue(k+1)}{RValue(k)} = f^{RV}(\mathbf{x}_k) + \sigma^{RV}(\mathbf{x}_k)w_{k+1}^{RV}$$

$$y_{k+1}^{RG} = \log \frac{RGrowth(k+1)}{RGrowth(k)} = f^{RG}(\mathbf{x}_k) + \sigma^{RG}(\mathbf{x}_k)w_{k+1}^{RG}$$

	Entire data	06/95-07/98	07/98-09/03	09/03-08/08	09/05/08-12/31/10
Max	0.1204	0.0611	0.0995	0.1204	0.1204
Min	-0.1986	-0.0507	-0.1267	-0.1986	-0.1986
Median	0.0024	0.0048	-0.0008	0.0038	0.0042
Mean	0.0012	0.0045	-0.0002	-0.0015	0.0000
Std	0.0272	0.0177	0.0302	0.0465	0.0426
Skewness	-0.6741	-0.1051	-0.2204	-0.6765	-0.8040
Kurtosis	5.8949	0.2693	2.0180	2.9595	3.9150

Table 4.3: Summary statistics of Russell 3000 return

where

$$\mathbf{f}^{RV} = (f_1^{RV}, \dots, f_N^{RV}) \in \mathbb{R}^N, \mathbf{f}^{RG} = (f_1^{RG}, \dots, f_N^{RG}) \in \mathbb{R}^N,$$

$$\boldsymbol{\sigma}^{RV} = (\sigma_1^{RV}, \dots, \sigma_N^{RV}) \in \mathbb{R}^N, \boldsymbol{\sigma}^{RG} = (\sigma_1^{RG}, \dots, \sigma_N^{RG}) \in \mathbb{R}^N,$$

are governed by the same WHMM \mathbf{x} . Here, w_k^{RV} and w_k^{RG} are $N(0, 1)$ IID random variables independent of each other. The data are processed in batches of 10 observation points. At the end of each pass through the data, \mathbf{f} , $\boldsymbol{\sigma}$, \mathbf{A} and $\boldsymbol{\Pi}$ are updated with new estimates using the formulas given in the previous section. These new estimates are in turn used as initial parameters for the next pass. This means the parameters are updated roughly every two and a half months. We process the data in batches in order to lower computational expenses. Furthermore, the use of batches is consistent with the idea of suboptimal schemes; see page 15 of Elliott, et al. [13]. Investors can choose any length of a batch to update their information according to their needs. In our numerical experiment, we find that updating parameters every two and half months is sufficient to capture market information. While using batches with less numbers of data points improves forecasting errors slightly, it does not lead to a better portfolio performance. Figure 4.1 displays the plot of the evolution of \mathbf{f}^{RV} , \mathbf{f}^{RG} , $\boldsymbol{\sigma}^{RV}$, $\boldsymbol{\sigma}^{RG}$ and the transition matrix \mathbf{A} under the two-state WHMM setting.

The optimal investment strategy is developed based on the forecasts of index returns. To assess the predictive performance of the model, we calculate the one-step ahead forecasts for both indices through the following equations:

$$E[\text{RValue}_{k+1} | \mathcal{Y}_k] = \text{RValue}_k \sum_{i,j=1}^N \langle \mathbf{p}_k, \mathbf{e}_{ij} \rangle \exp(f_i^{RV} + (\sigma_i^{RV})^2/2) \quad (4.16)$$

$$E[\text{RGrowth}_{k+1} | \mathcal{Y}_k] = \text{RGrowth}_k \sum_{i,j=1}^N \langle \mathbf{p}_k, \mathbf{e}_{ij} \rangle \exp(f_i^{RG} + (\sigma_i^{RG})^2/2). \quad (4.17)$$

The left panel of Figure 4.2 depicts a comparison between the actual Russell 3000 growth and

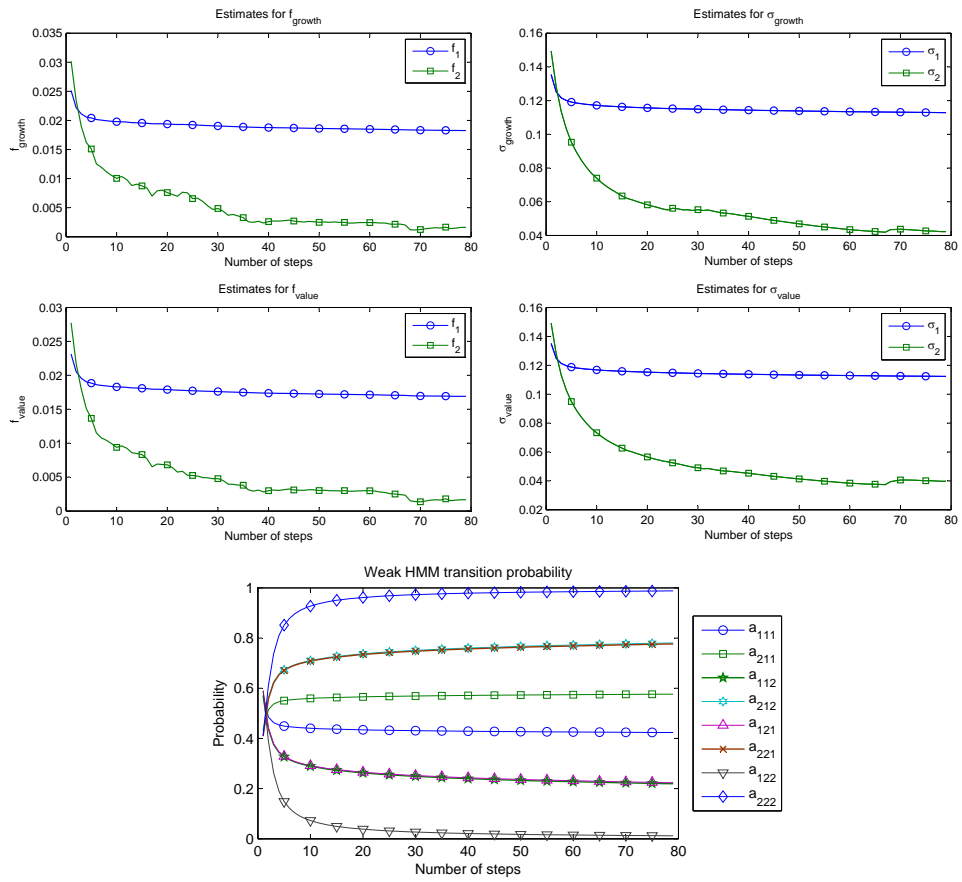


Figure 4.1: Evolution of parameter estimates under the 2-state setting

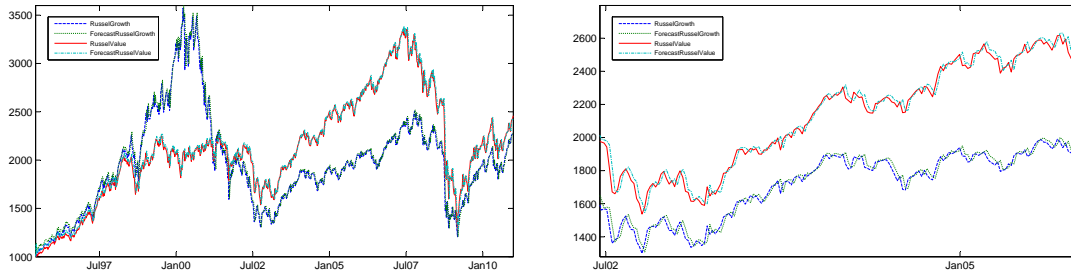


Figure 4.2: Actual data and one-step ahead forecasts for Russell 3000 growth and value indices (left), and zoom-in view for the period Jul 02 – Dec 05 (right)

value indices data and the one-step ahead forecasts. We can observe from the zoom-in view (right panel) highlighting a period of over 3 years how close the forecasts are to the actual data. We see that we have a reasonable prediction performance from both plots based on a strong positive relation between actual indices and one-step ahead forecasts. Any diffusion-type model will suggest that tomorrow's level is close to today's; the WHMM we present is no exception. Note that the result of prediction in Figure 4.2 is similar to that of an AR(1) process applied to the indices. This is because the 1-state WHMM is an AR(1)-process. As regular AR(p) models are not able to describe the temporal properties, TP1 and TP2, in Rydén, et al. [29], we assume WHMM can perform better in forecasting. In this study, the only forecasting performance we care about is to test how much better we can manage a portfolio using WHMMs; This is not a purely classical econometrics exercise. Scatter plots for actual returns versus one-step ahead forecast returns are shown in Figure 4.3. The plots form a circle pattern centering at zero. This is consistent with the fact that even if the states could be forecasted with greater accuracy, we will not get the correct sign of the return much more than half the time because the volatility term is much higher than the drift term. In light of Figures 4.3, how can we discern the usefulness of the WHMM? We need to examine the performance of an entire strategy.

We also run the filtering algorithm with different number of states. To assess the goodness of fit of the one-step ahead forecasts, we use four criteria: RMSE, AME, RAE and APE, for

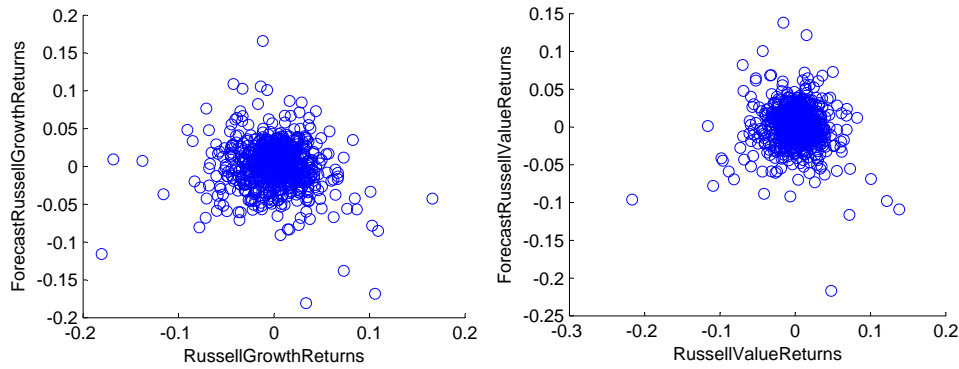


Figure 4.3: Actual returns and one-step ahead forecasts for returns of Russell 3000 growth (left) and value (right) indices

	1-state WHMM	2-state WHMM	3-state WHMM
RAE value	0.0966	0.0976	0.1044
APE value	0.0188	0.0192	0.0216
AME value	38.6423	39.0510	41.7675
RMSE value	54.4720	54.6768	56.7974
RAE growth	0.1159	0.1170	0.1241
APE growth	0.0215	0.0218	0.0241
AME growth	43.0295	43.4276	46.0558
RMSE growth	64.4588	64.7098	66.5664

Table 4.4: Error measures for one-step ahead forecasts under 1-, 2- and 3-state WHMM set-ups

$N = 1$, $N = 2$ and $N = 3$. The results of this error analysis are given in Table 4.4. The results show that the two-state tends to outperform the three-state model in all forecasting metrics. Although the one-state model has a slight improvement, the APE has smaller value under the two-state model. We further adopt three information criteria, namely, the Akaike information criterion (AIC), the small-sample-size corrected version of AIC (AICc) and Bayesian information criterion (BIC) to measure the relative goodness of fit of a statistical model. Information criteria offer a relative measure of lost information described by the trade-off between bias and

variance in the model construction. The formulas of these information criteria are:

$$\begin{aligned} \text{AIC} &= 2s - 2\mathcal{L}(\theta), \\ \text{AICc} &= \text{AIC} + \frac{2s(s+1)}{n-s-1}, \\ \text{BIC} &= s\log(n) - 2\mathcal{L}(\theta), \end{aligned}$$

where s is the number of parameters, n is the number of data points and $\mathcal{L}(\theta)$ denotes the log-likelihood function of the model. The preferred model is the one that gives the smallest criteria value. For the vector observation process \mathbf{y}_k in each pass, the log-likelihood of the parameter set θ is given by

$$\mathcal{L}(\theta) = \sum_{l=1}^{\# \text{ in batch}} \sum_{i=1}^d \sum_{r=1}^N \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \left(-\frac{1}{2} \log(2\pi\sigma^i(\mathbf{x}_{l-1})^2) - \frac{(y_l^i - f^i(\mathbf{x}_{l-1}))^2}{2\sigma^i(\mathbf{x}_{l-1})^2} \right). \quad (4.18)$$

The calculated values for the 1-, 2- and 3-state models after each algorithm step are presented in Figure 4.4. Observe that the 1-state model produces the smallest AIC values except around step 68 which is corresponding to the period of subprime crisis. This implies that the 1-state model is not able to describe the market during the crisis period. Therefore, the error analysis confirms the capability of the two-state WHMM in capturing the characteristics of the dataset.

4.4 A switching investment strategy

There are various asset allocation strategies that one can devise. Here we focus on a dynamic asset allocation which assumes active changes to an investment based on short-term market forecasts for returns. The switching strategy utilizes the forecasted risk-adjusted returns of the indices as signals to switch investments between the Russell 3000 growth and Russell 3000 value index. The forecasted risk-adjusted return is calculated by dividing the forecasted returns by the realized volatility of each index covering the previous 20 data points.

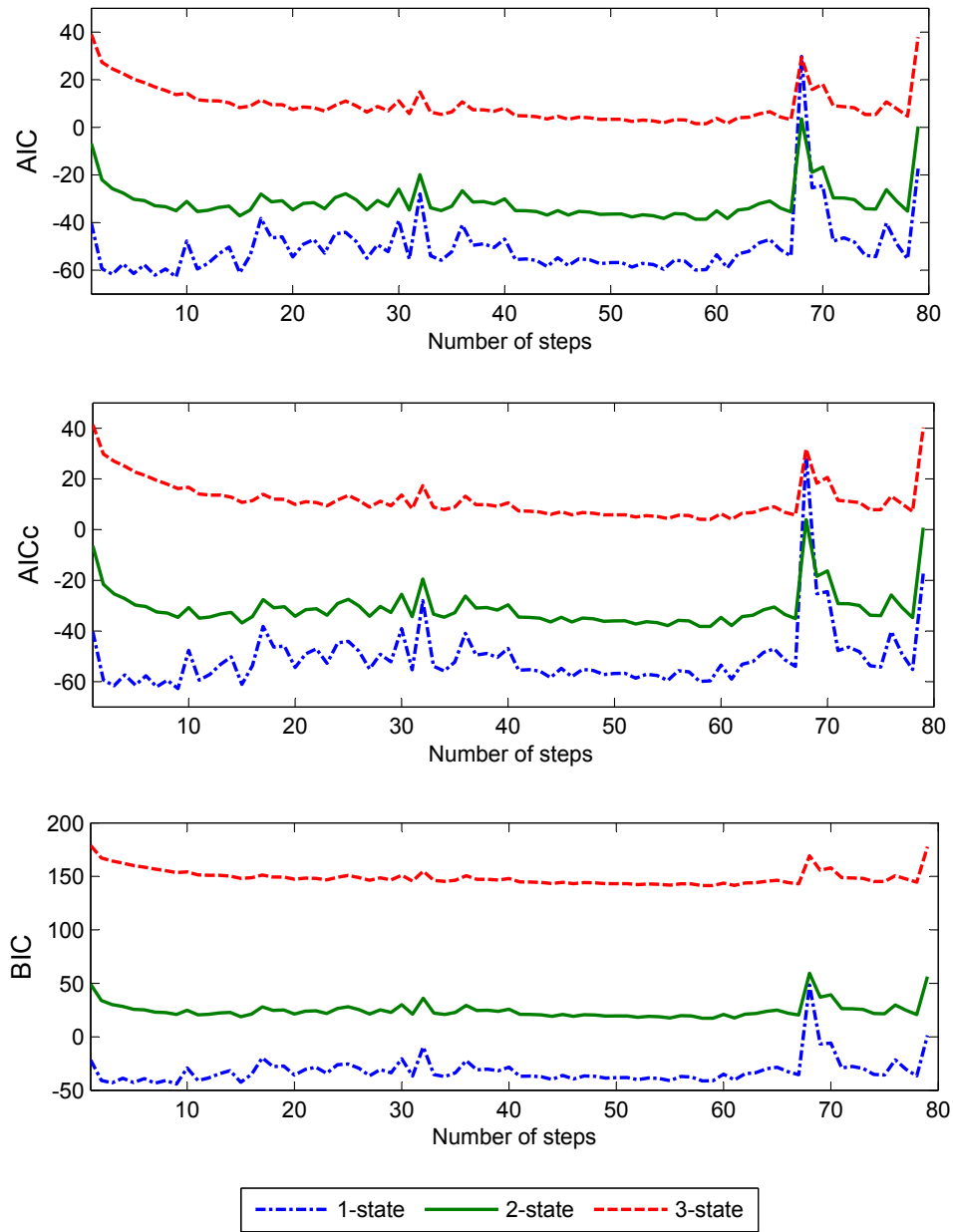


Figure 4.4: AIC, AICc and BIC values for the 1-, 2- and 3-state models

We implement the switching strategy on a 15-year dataset recorded from June 1995 to December 2010. The observation data is divided into 15 intervals and each interval covers roughly one year. We suppose a hypothetical starting investment of \$100 and then apply the forecasting method on the period considered. At the beginning of each interval, the signals from the 1-step ahead forecasted risk-adjusted return on both indices are compared. The full amount is invested in the index with the higher forecasted risk-adjusted return. We also assume that the transaction cost is a fixed percentage of total investment. When asset allocation changes, this transaction cost is subtracted from total investment.

The overall performance of the switching strategy is compared with that of the pure investment strategy on the basis of the log-return of the terminal wealth. Let X^{RG} and X^{RV} denote the differences of log-returns from the switching and pure strategies, which are defined by

$$X_i^{RG} = \log \frac{SW_i}{100} - \log \frac{RG_i}{100} \quad (4.19)$$

$$X_i^{RV} = \log \frac{SW_i}{100} - \log \frac{RV_i}{100}, \quad (4.20)$$

for $i = 1, 2, \dots, 15$. Here, SW_i denotes the terminal wealth of the portfolio with switching strategy at the end of the i th interval. RG_i and RV_i denote the terminal wealth of the investment, holding 100% of Russell 3000 growth index and Russell 3000 value index, respectively, at the end of the i th interval. The portfolio performance under varying transaction costs from 5 basis points (1bp=0.01%) to 70 bps is presented in Table 4.5. In addition, we present the performance of both switching and pure indices strategies using the usual HMM forecasts. Our study shows that the WHMM-based switching strategy has higher values in $\text{Mean}(X^{RV})$ and $\text{Mean}(X^{RG})$ than those from HMM-based switching strategy yielding negative values. WHMM-based strategy shows higher $\text{std}(X^{RV})$ and lower $\text{std}(X^{RG})$ than those based on HMM strategy. As we can observe, WHMM-based $\text{Mean}(X^{RV})$ and $\text{Mean}(X^{RG})$ are positive and slightly decrease as transaction cost increases. This means that on average the log return from the switching strategy is

Transaction Cost	5 bps		20 bps		50 bps		70 bps	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
Mean (X^{RG}) %	3.9106	-0.6895	3.5163	-0.9281	2.7261	-1.4062	2.1979	-1.7258
Std (X^{RG}) %	2.4102	6.7138	2.4701	6.5882	2.7086	6.3598	2.9403	6.2259
Mean (X^{RV}) %	11.2393	-9.5292	10.8450	-9.7677	10.0547	-10.2458	9.5266	-10.5654
Std (X^{RV}) %	18.4701	15.8552	18.2859	16.0045	17.9291	16.3117	17.7003	16.5226

Table 4.5: Performance comparison for WHMM- and HMM-based switching strategies with varying transaction costs.

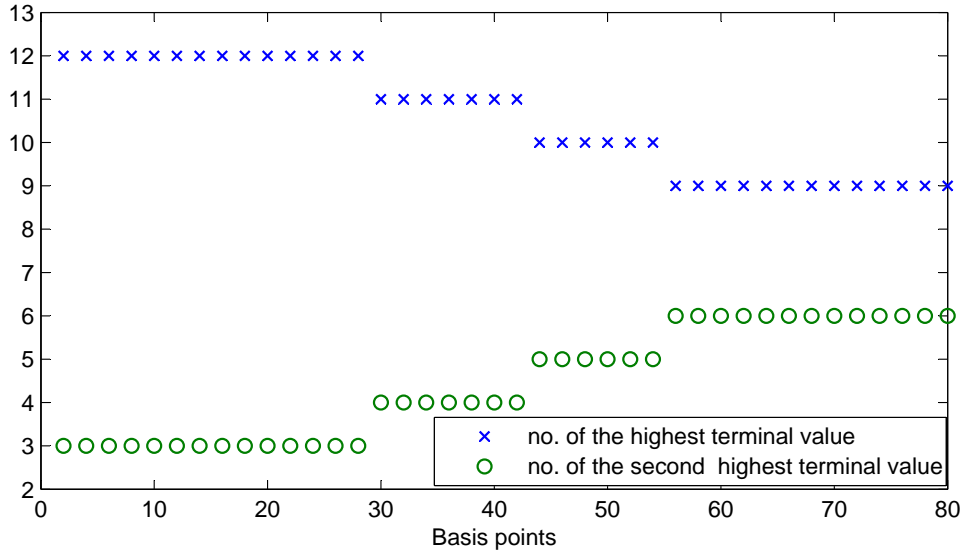


Figure 4.5: Numbers of the intervals switching strategy has the highest and the second highest terminal values for varying transaction cost

higher than that from the pure index investments. Compared with the pure growth and value strategies, the WHMM switching strategy has either the highest or the second highest terminal value in 15 intervals. Figure 4.5 displays number of the intervals in which switching strategy has the highest and the second highest terminal values for transaction cost varying from 1bp to 80bps. We observe, however, high values of $\text{std}(X^{RV})$ and $\text{std}(X^{RG})$, which indicate a high risk of employing the switching strategy. We next introduce a mixed strategy to address the diversification of risks.

4.5 A mixed investment strategy

Selecting an investment strategy is similar to the asset allocation decision problem in that one tries to maximize strategy return while controlling portfolio risk. The risk is evaluated in terms of the variance of the portfolio's return. The mean-variance problem entails maximizing the expected portfolio's return and minimizing the variance of the portfolio's return. In this section, we investigate a mixed asset allocation strategy on two assets whose dynamics are modulated by WHMM. With this strategy investors determine the optimal weight of each asset to allocate based on the estimated parameters and the state of the weak Markov chain. The development here follows the applications of results in Erlwein, et al. [16] and Elliott and van der Hoek [15]. Suppose an investor is guided by an optimization equation MV, which is a linear combination of the expected portfolio's return and variance of the portfolio's return. Let $\mathbf{w} = (w_g, w_v)$ denote the weight of Russell 3000 growth and value indices, respectively. To solve this mean-variance problem, we wish to estimate the optimal \mathbf{w} which maximizes the function

$$MV(\mathbf{w}) = vE[w_g y_{k+1}^{RG} + w_v y_{k+1}^{RV} | \mathcal{Y}_k] - \text{Var}[w_g y_{k+1}^{RG} + w_v y_{k+1}^{RV} | \mathcal{Y}_k],$$

where y^{RG} and y^{RV} are the logreturns of Russell 3000 growth and value indices, and v is a non-negative risk aversion factor. The optimal weights are given in the following proposition.

Proposition 4.5.1 *Let $v > 0$ be the risk aversion factor. Suppose that neither short selling nor borrowing is allowed. Strictly speaking, since we have to rebalance the portfolio every year using the filters, \mathbf{w} is time-dependent. The optimal weight w_g is given by*

$$w_g = \begin{cases} \frac{v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k \rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k \rangle) + 2\langle \sigma^{RV}, \hat{\mathbf{x}}_k \rangle^2}{2(\langle \sigma^{RG}, \hat{\mathbf{x}}_k \rangle^2 + \langle \sigma^{RV}, \hat{\mathbf{x}}_k \rangle^2)} & \text{when } -2\langle \sigma^{RV}, \hat{\mathbf{x}}_k \rangle^2 < v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k \rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k \rangle) \\ & < 2\langle \sigma^{RG}, \hat{\mathbf{x}}_k \rangle^2 \\ 1 & \text{when } v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k \rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k \rangle) > 2\langle \sigma^{RG}, \hat{\mathbf{x}}_k \rangle^2 \\ 0 & \text{when } v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k \rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k \rangle) < -2\langle \sigma^{RV}, \hat{\mathbf{x}}_k \rangle^2 \end{cases},$$

and the optimal weight w_v is given by $w_v = 1 - w_g$.

Proof See [16] for proof. ■

It has to be emphasized that, the optimal weight depends on the state of the embedded MC $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ by noting that

$$\langle \hat{\mathbf{x}}_k, \mathbf{e}_r \rangle = \sum_{i=1}^N E[\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ri} \rangle | \mathcal{Y}_k], \text{ for } r = 1 \dots N,$$

where $\mathbf{e}_{ri} = \xi(\mathbf{e}_r, \mathbf{e}_i)$.

Note that the weights belong to the interval $[0, 1]$ since neither short selling nor borrowing is allowed. Similar to the previous section, we divide the observation data into 15 intervals. For each interval, the optimal weights are calculated for each time k utilizing the optimal parameters and the estimated states of the weak Markov chain. Investors can allocate investment using different parameter updating frequencies depending on their goal. To achieve consistency in comparison with the switching and pure index strategies, the weights employed for each index is updated at the beginning of each interval. Transaction costs will also be considered. To gauge the strategy performance, we shall focus on the terminal value of the portfolio.

Figure 4.6 exhibits a plot of optimal weights for Russell 3000 growth and value indices. The risk aversion factor ν is a scaling constant which is chosen by the investor. Here, we allow this factor to vary from $\nu = 0$ (totally avoiding risk) to $\nu = 5$ (seeking some risk). The evolution of optimal weights for Russell 3000 growth index with different values of ν is shown in Figure 4.7. When ν is small, the investor is relatively conservative. The switching of market's regime has less impact on his/her asset allocation as can be viewed from the stable variation of weights for the Russell 3000 growth index. The investor with higher ν appears to aggressively react to market regime switching. In the left panel of Figure 4.2, we see that there are roughly two different states: Russell 3000 growth index has higher risk than the value index before September 2001 and it has lower risk after that. Consequently, the weight to allocate in growth index is

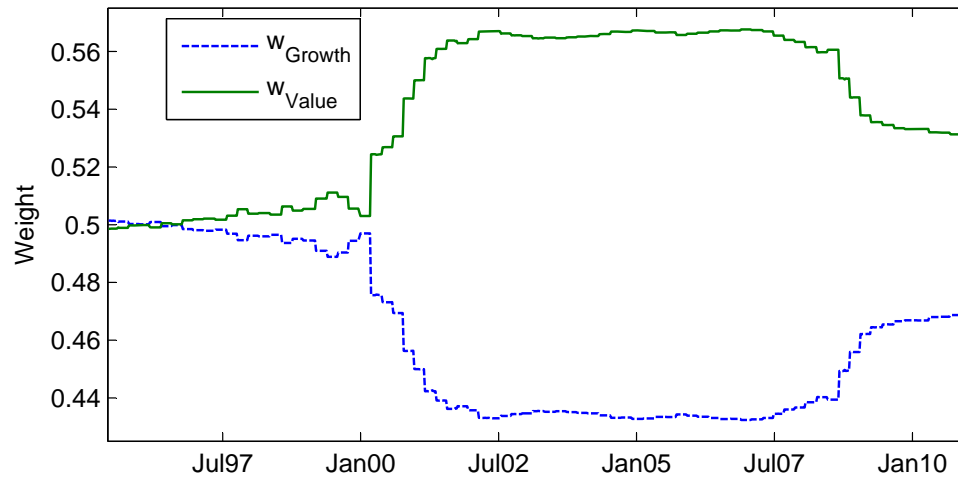


Figure 4.6: Optimal weights for Russell 3000 value and growth indices in the WHMM-based mixed strategy with $\nu = 0.08$

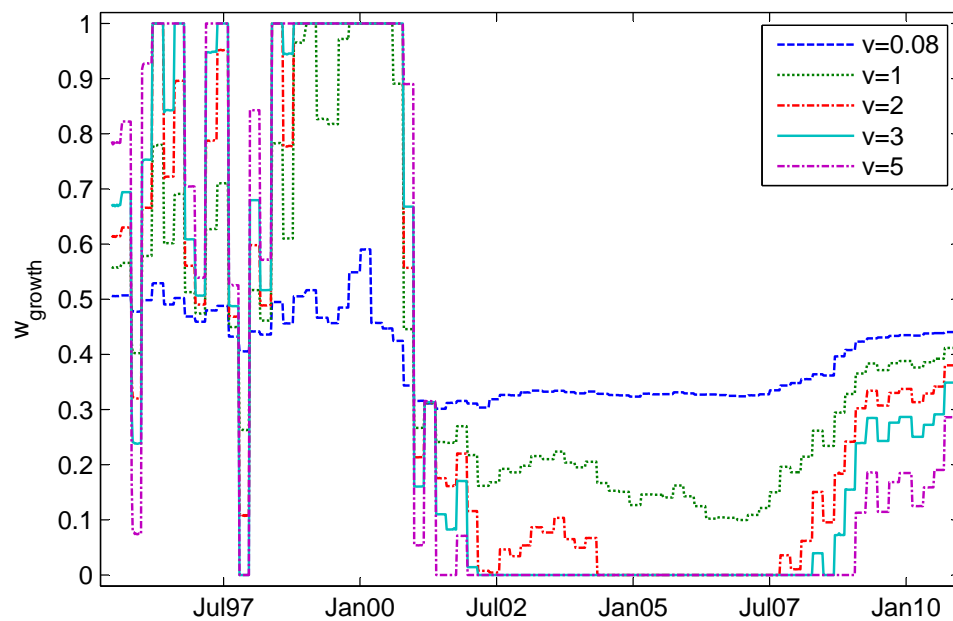


Figure 4.7: Evolution of optimal weights for Russell 3000 growth index in the WHMM-based mixed strategy with varying ν 's

Transaction Cost	5 bps		20 bps		50 bps		70 bps	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
Mean (X^{RG}) %	4.5606	2.9287	3.5750	1.8686	1.5993	-0.2564	0.2789	-1.6767
Std (X^{RG}) %	10.664	11.8866	10.3087	11.5399	9.7089	10.9647	9.4039	10.6797
Mean (X^{RV}) %	-2.7680	-5.9108	-3.7536	-6.9710	-5.7292	-9.0960	-7.0497	-10.5163
Std (X^{RV}) %	8.5098	7.6077	8.9849	8.1914	10.0169	9.4298	10.7537	10.2950

Table 4.6: Performance comparison between WHMM- and HMM-based mixed strategies with varying transaction costs.

higher than 0.5 before this time and it drops below 0.5 when the index has less uncertainty.

Table 4.6 shows the overall performance of the mixed strategy, which is compared with the pure growth and pure value strategies with $\nu = 0.08$, obtained through the analogue formulas of (4.19) and (4.20). The standard deviations of the differences of returns, $\text{std}(X^{RG})$ and $\text{std}(X^{RV})$, are lower than that of using switching strategy as we expected. $\text{Mean}(X^{RG})$ and $\text{Mean}(X^{RV})$ decrease as transaction cost increases. Compared to the mixed strategy based on the forecasts under the usual HMM framework, the WHMM-based mixed strategy produces higher values in both mean and standard deviation. The WHMM setting certainly carries more opportunities to explore the trade off between expected return and risk, which means higher risk may lead to higher return.

Figure 4.8 presents the evolution of investment under the switching, mixed and pure index strategies. Each subplot covers three years of data. We can see that based on the value of the investments, the mixed strategy does not always outperform other strategies. It is not straightforward to establish from the plots which strategy is the best. We shall then evaluate the portfolio performance through some classical measures in investment.

4.6 Evaluating the portfolio performance

In this section, we carry out performance comparisons among portfolio allocation strategies developed in the previous sections using historical data and simulated data. We evaluate the

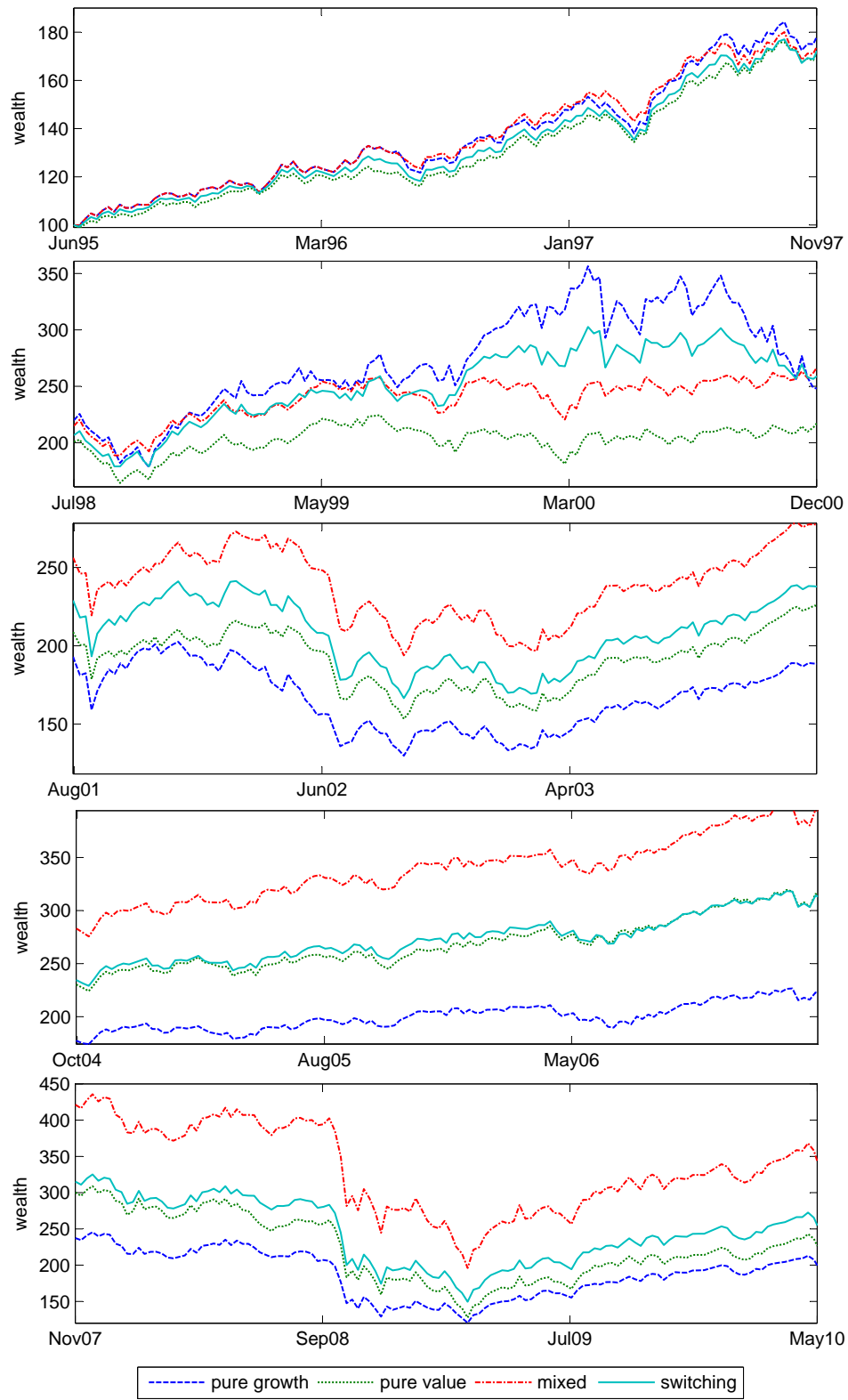


Figure 4.8: Switching, mixed, pure growth and pure value strategies comparison between 1995 and 2010

Period	Switching strategy	Mixed strategy	Pure Russell 3000 value	Pure Russell 3000 growth	Pure Russell 3000 index
1	0.2164	0.2016	0.1645	0.2164	0.2020
2	0.2254	0.2178	0.2254	0.2020	0.2177
3	-0.0083	-0.0298	-0.0523	-0.0083	-0.0292
4	0.0999	0.1485	0.0999	0.1699	0.1491
5	0.0141	0.0095	0.0141	0.0058	0.0184
6	-0.0751	-0.1347	-0.0751	-0.1652	-0.1360
7	-0.1488	-0.1930	-0.1488	-0.2394	-0.1984
8	0.1252	0.1304	0.1252	0.1331	0.1311
9	0.0206	-0.0393	0.0206	-0.1104	-0.0475
10	0.0701	0.0630	0.0555	0.0701	0.0633
11	0.0976	0.0701	0.0976	0.0346	0.0653
12	-0.0839	-0.1471	-0.1904	-0.0839	-0.1405
13	-0.1293	-0.1294	-0.1293	-0.1264	-0.1292
14	0.0259	0.0243	0.0259	0.0219	0.0241
15	0.6620	0.5511	0.4190	0.6620	0.5581
Mean	0.0741 (4.96×10^{-4})	0.0495 (4.74×10^{-4})	0.0435 (3.92×10^{-4})	0.0521 (5.42×10^{-4})	0.0499 (4.7×10^{-4})
Std	0.1976 (5.99×10^{-4})	0.1890 (4.46×10^{-4})	0.1577 (3.10×10^{-4})	0.2168 (5.82×10^{-4})	0.1906 (4.50×10^{-4})
Mean/Std	0.3751 (2.24×10^{-3})	0.2622 (2.55×10^{-3})	0.2755 (2.66×10^{-3})	0.2405 (2.53×10^{-3})	0.2618 (2.50×10^{-3})

Table 4.7: Sharpe ratio for five investment strategies using 15 intervals. Numbers inside the parentheses are standard errors.

portfolio performance through a benchmark. In this case, the Russell 3000 index is a natural benchmark since both the Russell 3000 growth and value indices are its subindices. The comparison of four portfolios with the benchmark is made using three classical measures on returns; these are the Sharpe ratio, Jensen's alpha and the appraisal ratio.

The first measure is the Sharpe ratio, denoted by SR, and

$$SR = \frac{E[R_{\text{portfolio}} - R_{\text{riskfree}}]}{\sqrt{\text{Var}(R_{\text{portfolio}} - R_{\text{riskfree}})}}$$

where R_{riskfree} is the risk-free interest rate (downloaded from the Federal Reserve board website). This SR is used to characterize how well the return of an asset compensates the investors

for the risk taken. The higher the Sharpe ratio the higher is the return with the same level of risk. In Table 4.7, we tabulate the Sharpe ratio of five investment strategies using the dataset divided into 15 intervals. Note that the switching strategy has the same Sharpe ratio with that of one of either pure value or pure growth strategy. At the beginning of each interval, the switching strategy allocates to one of the subindices with a number of shares depending on the value of the switching investment and the chosen index at previous time step. Hence, the switching portfolio and the chosen subindex have the same return in one interval. Such similarity is also true in other measures since all the calculations are based on the returns. The differences among the strategies are small. Out of the 15 intervals, the switching strategy shows a better performance than the benchmark in 11 intervals and the mixed strategy outperforms the benchmark in 6 intervals. Both the WHMM switching and WHMM mixed strategies have higher risk-adjusted mean than the benchmark. In particular, the switching strategy shows the highest risk-adjusted mean in all of the strategies.

We calculate Jensen's alpha, which is often used to measure the abnormal return of a portfolio over the expected return. This is denoted by α_J and it is the constant in the regression model,

$$\alpha_J = R_{\text{portfolio}} - [R_{\text{riskfree}} - \beta_{\text{portfolio}}(R_{\text{benchmark}} - R_{\text{riskfree}})].$$

A positive alpha indicates the portfolio has a higher marginal return. Table 4.8 shows the Jensen's alpha for four allocation strategies. Although the differences among the values of α are very small, there are 11 and 5 positive α 's out of 15 for the switching and mixed strategies, respectively. It indicates the marginal returns in these periods are higher than that of the benchmark.

Finally, we consider Treynor and Black's appraisal ratio (AR), also known as the information

Period	Switching Strategy ($\times 10^{-4}$)	Mixed Strategy ($\times 10^{-4}$)	Pure Russell 3000 Value ($\times 10^{-4}$)	Pure Russell 3000 Growth ($\times 10^{-4}$)
1	1.2306	-0.0417	1.230	-1.3637 6
2	3.6729	0.0300	-3.7262	3.6729
3	4.0153	-0.1148	4.0153	-4.3166
4	-5.6901	-0.0249	5.0587	-5.6901
5	6.1203	-2.5431	-10.2510	6.1203
6	17.7209	0.8786	-21.5255	17.7209
7	12.7347	1.5509	-14.0599	12.7347
8	-0.1632	-0.0380	0.1686	-0.1632
9	11.3998	1.4556	-12.270	11.3998 7
10	1.8036	-0.1018	1.8036	-1.5748
11	6.3168	0.9435	-6.3031	6.3168
12	13.8661	-1.7868	13.8661	-14.7742
13	-2.1714	-0.2851	1.9350	-2.1714
14	-0.4850	-0.0225	0.5308	-0.4850
15	18.0944	-0.9831	18.0944	-18.4459
Mean	5.8977 (0.0183)	-0.0722 (0.0028)	-1.4289 (0.0261)	0.5987 (0.0245)
Std	7.3348 (0.0099)	1.0969 (0.0022)	10.3850 (0.0182)	9.6750 (0.0169)
Mean/Std	8040.7842 (26.1277)	-658.2883 (28.2625)	-1375.9094 (28.1958)	618.8129 (28.4250)

Table 4.8: Jensen's alpha for four investment strategies using 15 intervals. Numbers inside the parentheses are standard errors.

Period	Switching Strategy	Mixed Strategy	Pure Russell 3000 Value	Pure Russell 3000 Growth
1	0.1539	-0.1761	-0.1571	0.1539
2	-0.0239	-0.0252	-0.0239	0.0233
3	0.0863	-0.1354	-0.0921	0.0863
4	-0.1729	-0.0946	-0.1729	0.1781
5	-0.0130	-0.2037	-0.0130	-0.0227
6	0.1810	0.1199	0.1810	-0.1746
7	0.2562	0.1838	0.2562	-0.2658
8	-0.0528	-0.0547	-0.0528	0.0529
9	0.3399	0.3356	0.3399	-0.3478
10	0.0251	-0.0042	-0.0206	0.0251
11	0.1213	0.1122	0.1213	-0.1222
12	0.3355	-0.3366	-0.3338	0.3355
13	-0.0718	-0.0686	-0.0718	0.0703
14	0.0260	0.0307	0.0260	-0.0250
15	0.3435	-0.3505	-0.3388	0.3435
Mean	0.1023 (4.03×10^{-4})	-0.0445 (4.66×10^{-4})	-0.0235 (4.88×10^{-4})	0.0207 (4.90×10^{-4})
Std	0.1633 (2.14×10^{-4})	0.1869 (3.19×10^{-4})	0.1926 (3.17×10^{-4})	0.1953 (3.27×10^{-4})
Mean/Std	0.6263 (27.1×10^{-4})	-0.2381 (29.41×10^{-4})	-0.1220 (29.01×10^{-4})	0.1060 (28.63×10^{-4})

Table 4.9: AR for four investment strategies using 15 intervals. Numbers inside the parentheses are standard errors.

ratio. It is defined as the ratio between relative return and the relative risk and is given by

$$AR = \frac{E[R_{\text{portfolio}} - R_{\text{benchmark}}]}{\sqrt{\text{Var}(R_{\text{portfolio}} - R_{\text{benchmark}})}}.$$

The formula is very similar to the Sharpe ratio. Whereas the Sharpe ratio measures return relative to a riskless asset, the AR looks at returns relative to a risky benchmark. The higher the AR, the higher is the active return of the portfolio given the same risk level. Table 4.9 reports the AR of four investment strategies. The switching strategy outperforms the mixed strategy in 11 intervals. In particular, we have the highest mean and lowest standard deviation under this measure. We observe higher mean for switching strategy in each performance measure. A *t*-test is carried out to assess whether the means of portfolio under various performance measures are statistically different. In order to run a *t*-test, each of the two data means sets being compared should follow a normal distribution. Table 4.10 presents the *p*-values for the Jarque-Bera normality test for the portfolio measures. The *p*-value for Jensen's alpha and Appraisal ratio of all portfolios are high which suggests there is not sufficient evidence to indicate that these data sets are coming from a non-normal distribution. Moreover, at 0.05 significance level, we can reject the null hypothesis that under the Sharpe ratio, the data on switching, mixed, pure growth and pure index strategies are from normal distribution. We test the difference between each pair of portfolios for the two measures. Table 4.11 shows the *p*-value for a one-tailed paired *t*-tests of significance assuming unequal variances. Comparing the switching and mixed strategies, the *p*-values in the first column are very small. This tells us that the difference in means under these performance measures of these two strategies is highly significant. The same can be said for the comparison of switching and pure growth strategies. Comparing the mixed and the pure value strategies, the *p*-values are large so that we cannot reject the null hypothesis, i.e., we cannot reject that the two means are equal. Similar conclusion can be made when we compare the mixed and pure growth strategies. Moreover, the switching strategy has the best performance for the period considered. In addition to the *t*-test, we use the Wilcoxon

	Switching strategy	Mixed strategy	Pure Russell Value	Pure Russell Growth	Pure Russell Index
Sharpe Ratio	0.0026	0.0350	0.3136	0.0111	0.0322
Jensen's Alpha	0.4092	0.3608	0.5000	0.5000	-
Appraisal Ratio	0.4929	0.5000	0.5000	0.5000	-

Table 4.10: p -values for the Jarque-Bera test of normality on data given in Tables 4.7 - 4.9

	Switching vs Mixed	Switching vs Pure growth	Switching vs Pure value	Mixed vs Pure growth	Mixed vs Pure value
Jensen's alpha	0.0035	0.0174	0.0515	0.3113	0.3967
Appraisal ratio	0.0149	0.1127	0.0321	0.1791	0.3821

Table 4.11: p -values for a one-tailed significance test on the performance results shown in Tables 4.8 - 4.9

rank sum test for the significance of the differences, and the p -values are reported in Table 4.12. Wilcoxon test does not rely on the normality assumption and so it complements our use of the t -test. The results suggest that the Jensen's alpha and Appraisal ratio for switching strategy are significantly different from that of mixed strategy. It is consistent with the results from t -test. Next, we give a simulation analysis in conjunction with the three portfolio measures.

We are interested in the statistical inference of the above portfolio measures for each portfolio strategy. The bootstrap is a way of finding the sampling distribution from one sample path. Introduced by Efron and Tibshirani [12], it is a technique allowing estimation of the sample distribution of almost any statistic. This method can be implemented when the sample could be assumed to be drawn from an independent and identically distributed population. The bootstrap method constructs a number of resamples of the observation datasets with equal size by random sampling with replacement from the original dataset. The datasets used for the bootstrapping

	Switching vs Mixed	Switching vs Pure growth	Switching vs Pure value	Mixed vs Pure growth	Mixed vs Pure value
Sharpe Ratio	0.6935	0.6934	0.8681	0.9835	0.9669
Jensen's alpha	0.0161	0.0648	0.1054	0.6187	0.6783
Appraisal ratio	0.0344	0.2899	0.0742	0.2998	0.6783

Table 4.12: p -values for a Wilcoxon rank sum test on the performance results shown in Tables 4.7 - 4.9

are the Russell 3000 index, the growth subindex and the value subindex. Each of the original datasets contains 784 data points. The procedure of the simulation is as follows:

1. We divide the datasets into 16 intervals and each interval contains 49 weeks.
2. A resample is created by repeatedly sampling with replacement from these 16 intervals. This means that we randomly pick one interval for the resample path and put the interval back for drawing again. As a result, any interval can be drawn more than once, or not at all. The resample has the same size as the original data.
3. The three measures are calculated for the new resample path.

The construction is repeated 10,000 times and the statistics of the three classical measures are obtained. Table 4.13 shows the statistics of the portfolios, Sharpe ratio, Jensen's alpha and AR with a transaction cost amounting to 5bps. Table 4.14 presents the same analysis with transaction cost of 30bps. When 5 bps transaction costs are introduced, the WHMM switching and mixed strategies generate higher mean Sharpe ratio and AR from the 10,000 bootstrap sample paths than those from other strategies. Only the pure growth strategy leads to a negative mean in the Jensen's alpha measure. The standard deviation of switching strategy from the bootstrapped samples is higher than that in the mixed strategy in all cases. The estimated 95% confidence intervals for the mixed strategy are smaller than those in the switching strategy. Apparently, the WHMM mixed strategy is more stable than the WHMM switching strategy in terms of the standard deviation and 95% confidence interval under the 5bps transaction cost. A comparison of the mean and variance of the portfolio returns is also presented. The switching strategy outperforms other strategies with the highest variance nonetheless. On the other hand, the pure Russell 3000 index strategy has the lowest variance, but the 95% confidence interval is bigger than those in both WHMM strategies. It indicates that both WHMM strategies are more stable than the benchmark in terms of the confidence interval. When transaction costs are set to 30 bps, the switching strategy produces the highest mean in all cases. Since the mixed strategy is the most costly strategy, it has a lower mean than pure value strategy under all measures.

Sharpe ratio	Mean ($\times 10^{-2}$)	Std ($\times 10^{-2}$)	95% Con.Int. ($\times 10^{-2}$)
Switching	3.1617	0.1711	[3.1498 3.1737]
Mixed	2.8455	0.1321	[2.8363 2.8548]
Pure Value	2.8137	0.1430	[2.8038 2.8238]
Pure Growth	1.7502	0.0829	[1.7445 1.7561]
Pure Index	2.2978	0.1056	[2.2904 2.3052]
Jensen's α	Mean ($\times 10^{-3}$)	Std ($\times 10^{-3}$)	95% Con.Int. ($\times 10^{-3}$)
Switching	0.4037	0.0607	[0.3995 0.4080]
Mixed	0.2259	0.0483	[0.2225 0.2293]
Pure Value	0.2673	0.0370	[0.2647 0.2699]
Pure Growth	-0.0430	0.0506	[-0.0465 -0.0395]
AR	Mean ($\times 10^{-3}$)	Std ($\times 10^{-3}$)	95% Con.Int. ($\times 10^{-3}$)
Switching	1.4270	0.4287	[1.3970 1.4570]
Mixed	-2.3559	0.3149	[-2.3779 -2.3338]
Pure Value	-7.0522	0.5095	[-7.0879 -7.0165]
Pure Growth	-6.2734	0.3956	[-6.3010 -6.2457]
Mean & Std return	Mean ($\times 10^{-3}$)	Std ($\times 10^{-3}$)	95% Con.Int. ($\times 10^{-3}$)
Switching	0.9250	0.0429	[0.9220 0.9280]
Mixed	0.8682	0.0244	[0.8665 0.8699]
Pure Value	0.7770	0.0065	[0.7765 0.7774]
Pure Growth	0.7829	0.0074	[0.7824 0.7834]
Pure Index	0.8967	0.0051	[0.8964 0.8971]

Table 4.13: Performance evaluation for 10000 bootstrapped datasets with 5bps transaction cost

However, it still has positive means in both Jensen's alpha and AR measures. Both WHMM-based strategies outperform the benchmark with 30 bps transaction cost in terms of higher Sharpe ratio, positive Jensen's alpha and positive AR. The switching strategy is the most risky judging the standard deviation of the measures. The mixed strategy shows a lower variance than the pure sub-indices strategies in Sharpe ratio and Jensen's alpha, and a lower variance than the pure growth strategy in the AR. Furthermore, the smaller 95% confidence interval of mixed strategy indicates that it is more stable than the switching strategy in the case of 30bps transaction cost. Therefore, we could conclude that the WHMM switching strategy gives a higher mean return, however the WHMM mixed strategy is less risky and more stable.

Sharpe Ratio	Mean ($\times 10^{-2}$)	Std ($\times 10^{-2}$)	95% Con.Int. ($\times 10^{-2}$)	
Switching	3.7352	0.3169	[3.7130	3.7574]
Mixed	2.4876	0.1220	[2.4791	2.4962]
Pure Value	2.8781	0.1380	[2.8684	2.8877]
Pure Growth	1.5679	0.1173	[1.5597	1.5761]
Pure Index	2.1326	0.1176	[2.1244	2.1408]
Jensen's α	Mean ($\times 10^{-3}$)	Std ($\times 10^{-3}$)	95% Con.Int. ($\times 10^{-3}$)	
Switching	5.0273	0.7980	[4.9714	5.0832]
Mixed	1.6992	0.5069	[1.6637	1.7347]
Pure Value	2.6856	0.4149	[2.6565	2.7146]
Pure Growth	-0.2362	0.5838	[-0.2770	-0.1953]
AR	Mean ($\times 10^{-3}$)	Std ($\times 10^{-3}$)	95% Con.Int. ($\times 10^{-3}$)	
Switching	7.5903	0.8554	[7.5304	7.6501]
Mixed	1.3412	0.3479	[1.3169	1.3656]
Pure Value	3.4708	0.2721	[3.4517	3.4898]
Pure Growth	-5.8516	0.5834	[-5.8925	-5.8108]
Mean & Std return	Mean ($\times 10^{-3}$)	Std ($\times 10^{-3}$)	95% Con.Int. ($\times 10^{-3}$)	
Switching	1.0447	0.0673	[1.0400	1.0494]
Mixed	0.7382	0.0193	[0.7369	0.7396]
Pure Value	0.7669	0.0058	[0.7665	0.7674]
Pure Growth	0.6304	0.0108	[0.6297	0.6312]
Pure Index	0.7236	0.0070	[0.7231	0.7241]

Table 4.14: Performance evaluation for 10000 bootstrapped datasets with 30bps transaction cost

4.7 Conclusion

Asset allocation strategies for growth and value stocks under a weak hidden Markovian regime-switching setting are examined. We suppose that the mean and volatilities of the price indices returns are modulated by a discrete-time multivariate WHMM process. Recursive optimal estimates by filtering multidimensional observations are given for the state and various processes related to the underlying second-order Markov chain. The parameters of the model, including the transition probabilities, the drift and the variance parameters in the multidimensional observations, can be re-estimated and the forecasts can be obtained using the estimates. We investigated two investment strategies: a switching strategy and a mixed strategy, using the weekly Russell 3000 growth and value indices data from 1995 to 2010. The switching strategy made use of the one-step ahead forecasted return for both indices and invested into the index with higher risk-adjusted forecasted return for each time interval. The mixed strategy lead to a mean-variance optimization problem, in which the optimal weights for each index were calculated using the estimated drifts and variance.

We compared both WHMM strategies with the HMM-based approach. For certain levels of transaction costs, the WHMM-based strategies outperform the HMM-based strategies in terms of the higher differences of log-return between the tested strategy and the pure strategies. The WHMM switching strategy never gives the worst performance in the time interval considered. The evolution of the optimal weights represents the investors' reaction to regime-switching in the market. And thus the mixed strategy has a lower variance of the return. Performance comparisons of the four portfolio strategies with the benchmark using Sharpe ratio, Jensen's alpha and AR were presented. When compared to the benchmark, which is the pure Russell 3000 index strategy, both WHMM strategies have higher risk-adjusted return. The switching strategy has higher marginal and relative returns than the benchmark. Furthermore, the bootstrap analysis with different transaction costs demonstrates that with 5 bps transaction cost, both WHMM-based strategies have higher return and are more stable than the benchmark in terms

of higher values in the performance measures and smaller confidence intervals. In the case of 30 bps transaction costs, the WHMM strategies still have higher returns, but the switching strategy is less stable and the mixed strategy is more stable than the benchmark.

References

- [1] M. L. Abell and J. P. Braslton. *Introductory Differential Equations: With Boundary Value Equations*. Academic press, California, 2009.
- [2] M. Ammann and M. Verhofen. The effect of market regimes on style allocation. *Financial Markets Portfolio Management*, 20:309–337, 2006.
- [3] A. Ang and G. Bekaert. International asset allocation with regime shifts. *Review of Financial Studies*, 15:1137–1187, 2002.
- [4] A. Ang and G. Bekaert. Timing and diversification: A state-dependent asset allocation approach. *Financial Analysts Journal*, 60:1137–1187, 2004.
- [5] R. Bauer, R. Haerden, and R. Molenaar. Timing and diversification: A state-dependent asset allocation approach. *Journal of Investing*, 13:72–80, 2004.
- [6] J. Bulla and I. Bulla. Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics and Data Analysis*, 51:2192–2209, 2006.
- [7] J. Bulla, S. Mergner, I. Bulla, A. Sesboüé, and C. Chesneau. Markov switching asset allocation: Do profitable strategies exist? *Journal of Asset Management*, 12:310–321, 2011.
- [8] W. Ching, M. K. Ng, and E. S. Fung. Higher-order multivariate Markov chains and their applications. *Linear Algebra and its Applications*, 428:492–507, 2008.

- [9] W. K. Ching, T. K. Siu, and L. M. Li. Pricing exotic options under a higher-order hidden Markov model. *Journal of Applied Mathematics and Decision Sciences*, pages 1–15, 2007. Article ID 18014.
- [10] M. Couillard and M. Davison. A comment on measuring the Hurst exponent in financial time series. *Physica A: Statistical Mechanics and its Applications*, 348:404–418, 2005.
- [11] S. Dajcman. Time-varying long-range dependence in stock market returns and financial market disruptions - a case of eight European countries. *Applied Economics Letters*, 19:953–957, 2012.
- [12] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press, 1994.
- [13] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer, New York, 1995.
- [14] R. J. Elliott, T. K. Siu, and A. Badescu. On mean-variance portfolio selection under a hidden Markovian regime-switching model. *Economic Modelling*, 27:678–686, 2010.
- [15] R. J. Elliott and J. van der Hoek. An application of hidden markov models to asset allocation problems. *Finance and Stochastics*, 1:229–238, 1997.
- [16] C. Erlwein, R. Mamon, and M. Davison. An examination of HMM-based investment strategies for asset allocation. *Applied Stochastic Models in Business and Industry*, 27:204–221, 2011.
- [17] A. Graflund and B. Nilsson. Dynamic portfolio selection: The relevance of switching regimes and investment horizon. *European Financial Management*, 9:179–00, 2003.
- [18] M. Guidolin and A. Timmermann. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31:3503–3544, 2007.

- [19] M. K. Hess. Timing and diversification: A state-dependent asset allocation approach. *European Journal of Finance*, 12:189–204, 2006.
- [20] H. E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–770, 1951.
- [21] I. N. Lobato and N. E. Savin. Real and spurious long memory in stock market data. *Journal of Business and Economic Statistics*, 16:261–266, 1988.
- [22] I. L. MacDonald and W. Zucchini. *Hidden Markov and other models for discrete-valued time series*. Chapman & Hall, London, 1997.
- [23] J. Maheu. Can GARCH models capture long-range dependence? *Studies in Nonlinear Dynamics and Econometrics*, 9:1, 2005.
- [24] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.
- [25] J. McCarthy, R. DiSario, H. Saraoglu, and H. C. Li. Tests of long-range dependence in interest rates using wavelets. *Quarterly Review of Economics and Finance*, 44:180–189, 2004.
- [26] R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3:373–413, 1971.
- [27] J. A. Du Preez, E. Barnard, and D. M. Weber. Efficient high-order hidden Markov modeling. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2911–2914, 1998.
- [28] B. Ray and R. Tsay. Long range dependence in daily stock volatilities. *Journal of Business and Economic Statistics*, 18:254–262, 2000.
- [29] T. Rydén, T. Teräsvirta, and S. Åsbrink. Stylized facts of daily return series and the hidden Markov model. *Journal Applied Econometrics*, 13:217–244, 1998.

- [30] P. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51:239–46, 1969.
- [31] T. K. Siu, W. K. Ching, E. Fung, M. Ng, and X. Li. A high-order Markov-switching model for risk measurement. *Computers and Mathematics with Applications*, 58:1–10, 2009.
- [32] J. Solberg. *Modelling Random Processes for Engineers and Managers*. Wiley & Sons, Inc., New Jersey.
- [33] X. Xi and R. Mamon. Parameter estimation of an asset price model driven by a weak hidden Markov chain. *Economic Modelling*, 28:36–46, 2011.
- [34] S. Yu, Z. Liu, M. S. Squillante, C. H. Xia, and L. Zhang. A hidden semi-Markov model for web workload self-similarity. In *21st IEEE International Performance, Computing, and Communications Conference*, pages 65–72, 2002.

Chapter 5

Yield curve modeling using a multivariate higher-order HMM

5.1 Introduction

This chapter considers the application of a higher-order Markov chain to the modeling of the term structure of interest rates. This is a further application of the theoretical developments in chapter 4. However, instead of considering a bivariate series with emphasis on investment strategies, the application here is centered on a multivariate series of yields and an error analysis is extended to a multidimensional WHMM setting.

Interest rate modeling is a central consideration in financial markets given its fundamental importance in pricing, risk management and investment. Classical models for the short-term interest rate, such as those proposed by Merton [23], Vasicek [30], Cox, et al. [5], and Hull and White [16], assume deterministic parameters. However, we all know that the economy and market are subjected to dynamic, and in some cases, significant changes. Such changes have substantial impact on the evolution of interest rates. Research works in recent years focus on the development of appropriate quantitative models suited for time-varying model parameters to accurately capture the behavior of various financial variables and economic indicators. The

introduction of regime-switching models provided some ways of incorporating the impact of market and economic changes on interest rates. Hamilton's works (cf. [13]) set forth the impetus for the construction of regime-switching-based methods in the modeling of non-stationary time series. Under such methods, values of the model parameters at a particular moment depends on the state of an underlying Markov chain at that moment. A study by Smith [27] found evidence that volatility depends on the level of the short rate and supports Markov-switching model over a stochastic volatility model. Landen [18] developed an HMM framework for the short-term interest rates, in which the mean and variance are governed by an underlying Markov process. In practice, the underlying state of the market and volatilities are unobservable and so parameter estimation for these Markov-switching models presents some challenges both from the practical and mathematical standpoints.

In a comprehensive work, Elliott, et al. [9] provided recursive self-updating estimates for the Markov chain as well as the model drift and diffusion parameters modulated by the same Markov chain. Elliott, et al. [10] proposed a multivariate HMM for the short rate process and HMM filtering techniques are employed in their implementation. Erlwein and Mamon [11] considered a Hull-White interest rate model in which the interest rate's volatility, mean-reverting level and speed of mean-reversion are all governed by a Markov chain in discrete time. The HMM filters are derived and implemented on a financial dataset. Their analysis of prediction errors together with the aid of the Akaike information criterion shows that a two-regime model is sufficient to describe the interest rate dynamics in their study. More recent developments on regime-switching literature focus on extending various commonly known models. Hunt and Devolder [17] studied an extension of the Ho and Lee model under a semi-Markov regime-switching framework, and an application of their proposed extension to the pricing of European bond options was given. Zhou and Mamon [32] investigated the Vasicek, CIR and Black-Karasinski models with the parameters of the short rate being modulated by a finite-state Markov chain in discrete time. A quasi-maximum likelihood method is utilized

to estimate model parameters and implementation of their algorithms was carried out on the Canadian yield rates.

Some recent studies examine the integration of regime-switching models with other modeling approaches to obtain a richer methodology. A four-state model to capture rate dynamics in the US spot and forward rate markets was proposed by Guidolin and Timmermann [12]; their out-of-sample forecasting exercise show evidence that, at short horizons, combining regime-switching forecasts with simpler univariate time-series forecasts can help reduce the root mean squared forecast errors. Meligkotsidou and Dellaportas [22] adopted a Bayesian forecasting methodology of discrete-time finite state-space HMM with non-constant transition matrix in modeling monthly data on rates of return series; the results of their Markov chain Monte Carlo algorithms indicate that nonhomogeneous HMMs improve the predictive ability of the model when compared against a standard homogeneous HMM.

Other papers on regime-switching models advance new approaches in detecting further evidence of regime shifts in the market. Startz and Tsang [29] constructed an unobserved component model in which the short-term interest rate is composed of a stochastic trend and a stationary cycle; results from their model-based measures suggest that allowing for regime switching in shock variances improves model performance. Audrino and Mederos [1] proposed a smooth transition tree model that combines regression trees and GARCH models to describe regime switches in the short-term interest rate series; their empirical results provide evidence of the power of the model in forecasting the conditional mean and variance. Utilizing an adapted unit-root test, Holmes, et al. [15] found evidence that Australian and New Zealand interest rates can switch between regimes characterized by differences in mean, variance and persistence.

For a review of models of term structure of interest rates under regime-switching setting, including earlier regime-switching models of short-term interest rate in discrete time, and recent

Markov-switching models in continuous-time, refer to Nieh, et al. [24]. Whilst the original HMM can reasonably model the impact of structural changes in the financial time series, there is a need to also develop quantitative models that are able to capture time series memory. Processes with long-memory characteristics have stronger coupling between values at different times than that of short-memory processes, and they can be described by heavy-tailed distributions. Mandelbrot [20] demonstrated applications of stochastic processes with long memory in economics and finance. Cajueiro and Tabak [2, 3] showed evidence of long-range dependence in the LIBOR and US interest rates. McCarthy, et al. [21] probed the presence of long memory in corporate bond yield spreads and found strong evidence that such presence exists. Numerous studies have developed stochastic models to capture the long-range dependence property in financial time series. Maheu [19] concluded that GARCH models can capture the long-memory property in volatility of financial prices under some circumstances. Dajcman [6] proposed an autoregressive fractionally integrated moving average (AFIMA) model for eight European stock market returns. Duan and Jacob [8] suggested that inclusion of long-range dependence in their model improves significantly model fitting performance on real interest rate data. It seems, however, that the existing literature on long-memory property of time series mainly concentrates on single-state stochastic models.

This chapter contributes to the widening of literature on the development of models that are able to capture not only regime-switching but also short- or long-term dependence in the HMM that modulates regime switches. We put forward a WHMM to model the movement of the term structure of interest rates. As Solberg [28] pointed out, the real significance of WHMM is to rectify the weakness of the usual HMM. In certain instances, HMM's memoryless property seems unwarranted for many stochastic processes observed in real-life applications. WHMM generalizes HMM and therefore, the memoryless property implied by the Markov assumption is not really as restrictive as it first appears. By using WHMM, the probability of current state does not depend on just one prior time epoch but on any finite number of prior epochs, and

so more information from the past is taken into account. The higher-order Markov chain and its applications in finance have been investigated by a number of authors, and these include, among others, Xi and Mamon [31] for returns of risky assets; Siu, et al. [26] for risk measurement; Ching, et al. [4] for exotic option pricing, and Siu, et al. [25] for spot rates and credit ratings.

In this chapter, we consider a multivariate WHMM for the evolution of the term structure of interest rates. Extending the formulation in [10], the short term rate can be rewritten as a function of a discrete-time WMC. We assume that the drift and diffusion terms of the yield values are governed by a second-order multivariate Markov chain. The states of the WMC are associated with the states of the market, whose current behavior depends on the behavior at the previous two time steps. We utilize a transformation that converts a WHMM into a regular HMM allowing us to apply the usual HMM estimation algorithm.

The remainder of this chapter is organized as follows. Section 5.2 describes the formulation of the multivariate modeling framework. The derivation of the filters for the states of WMC and other related processes through a measure change is presented. The recursive estimations are obtained by applying the EM algorithm. The implementation of this proposed model is given in section 5.3. The dataset involved in our numerical study consists of daily US Treasury yields. In section 5.4, we provide a discussion on how to select the optimal number of states. Using some criteria, we conclude that a two-state WHMM is sufficient to capture the market dynamics of our data. We also present an analysis of h -day ahead forecasts under the 1-, 2-, 3- and 4-state settings. Forecasting errors generated under the WHMM are compared with those generated under the regular HMM. With WHMM, being a device to capture both memory and regime-switching properties in the data series, we obtain better results than those produced by the HMM in terms of lower forecasting errors. We conclude with some remarks in section 5.5.

5.2 Filtering and parameter estimation

We assume all processes in our modeling set-up are supported by complete probability space (Ω, \mathcal{F}, P) . Let $\{\mathbf{x}_t\}$, $t \geq 0$ be a continuous-time weak Markov chain with finite space $\mathbf{S} = \{s_1, s_2, \dots, s_N\}$. Without loss of generality, we identify the points in \mathbf{S} with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\} \subset \mathbb{R}^N$, where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$ and \top denotes the transpose of a vector. The representation $\langle \mathbf{x}_t, \mathbf{e}_i \rangle$ refers to the event that the economy is in state i at time t . Here, $\langle \cdot, \cdot \rangle$ stands for the inner product in \mathbb{R}^N . We suppose the short rate process \mathbf{r}_t is a function of the unobservable Markov chain \mathbf{x}_t , such that $\mathbf{r}_t = r(\mathbf{x}_t) = \langle \mathbf{r}, \mathbf{x}_t \rangle$ for some vector $\mathbf{r} \in \mathbb{R}^N$. At time t , a zero-coupon bond expiring at time $t + \tau_g$, $g = 1, \dots, d$, has the price

$$F^g(\mathbf{x}_t, t) = E \left[\exp \left(- \int_t^{t+\tau_g} r(\mathbf{x}_s) ds \right) \middle| \mathcal{F}_t \right].$$

It was shown in Siu, et al. [26] that the yield values in discrete-time can be expressed as $\langle \mathbf{f}^g, \mathbf{x}_k \rangle = -\frac{1}{\tau_g} \log F^g(\mathbf{x}_k, t_k)$, and $\mathbf{x}_k = \mathbf{x}_{t_k}$, as a discrete-time version of the state process \mathbf{x}_k . Let $\mathbf{y}_k = (y_k^1, y_k^2, \dots, y_k^d)$ denote the d -dimensional yield process. Each component y_k^g , $1 \leq g \leq d$, is part of the sequence of yield values and has dynamics

$$y_{k+1}^g = f^g(\mathbf{x}_k) + \sigma^g(\mathbf{x}_k) z_{k+1}^g.$$

Each sequence $\{z_k^g\}$ is a sequence of $N(0, 1)$ IID random variables, which are independent of the \mathbf{x} -process. More specifically, the functions f^g and σ^g are given by the vectors $\mathbf{f}^g = (f_1^g, f_2^g, \dots, f_N^g)^\top$ and $\boldsymbol{\sigma}^g = (\sigma_1^g, \sigma_2^g, \dots, \sigma_N^g)^\top$ in \mathbb{R}^N , $f^g(\mathbf{x}_k) = \langle \mathbf{f}^g, \mathbf{x}_k \rangle$ and $\sigma^g(\mathbf{x}_k) = \langle \boldsymbol{\sigma}^g, \mathbf{x}_k \rangle$ represent the mean and variance of y_k^g , respectively. Note that we do not model the correlation among yields explicitly. However, all components of the vector observation process are modulated by the same underlying WMC, and thus they are correlated implicitly. Actual filter with correct correlation structure will presumably be better. So, one may view that this study is a lower bound for the validity of a larger study.

Our attention will solely be on a WMC of order 2 to simplify the discussion and present a complete characterization of the parameter estimation. The probability of the next time step for the WMC given the previous information is

$$\begin{aligned} P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \dots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k) \\ = P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k). \end{aligned}$$

Each entry of the transition probability matrix $\mathbf{A} := (a_{lmv}) \in \mathbb{R}^{N \times N^2}$, where $l, m, v \in 1, \dots, N$, refers to the probability that the process enters state l given that the current and previous states were in states m and v , respectively. The salient idea in the filtering of WHMM is that, a second-order Markov chain is transformed into a first-order Markov chain through a mapping ξ , and then we may apply the regular filtering method. The mapping ξ is defined by

$$\xi(\mathbf{e}_r, \mathbf{e}_s) = \mathbf{e}_{rs}, \text{ for } 1 \leq r, s \leq N,$$

where \mathbf{e}_{rs} is an \mathbb{R}^{N^2} -unit vector with unity in its $((r-1)N + s)$ th position. Note that

$$\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle = \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle$$

is the identification of the new first-order Markov chain with the canonical basis of \mathbb{R}^{N^2} . The new $N^2 \times N^2$ transition probability matrix $\mathbf{\Pi}$ of the new Markov chain is defined by

$$\pi_{ij} = \begin{cases} a_{lmv} & \text{if } i = (l-1)N + m, \quad j = (m-1)N + v \\ 0 & \text{otherwise.} \end{cases}$$

Here, each non-zero element π_{ij} represents the probability

$$\pi_{ij} = a_{lmv} = P(\mathbf{x}_k = \mathbf{e}_l | \mathbf{x}_{k-1} = \mathbf{e}_m, \mathbf{x}_{k-2} = \mathbf{e}_v),$$

and each zero represents an impossible transition. It may be shown that the new Markov chain $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ has the semi-martingale representation

$$\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi}\xi(\mathbf{x}_{k-1}, \mathbf{x}_{k-2}) + \mathbf{v}_k, \quad (5.1)$$

where $\{\mathbf{v}_k\}_{k \geq 1}$ is a sequence of \mathbb{R}^{N^2} martingale increments.

Under the real world measure P (the market measure in the context of this discussion), the underlying WMC is not observed directly. Instead, the state \mathbf{x}_k is contained in the noisy market observations $\mathbf{y}_k, k \geq 1$. We aim to “filter” the noise out of the observed market values. However, the derivation of filters under P is complicated. By Kolmogorov’s extension theorem, there exists a reference probability measure \bar{P} under which the \mathbf{y}_k ’s are $N(0, 1)$ IID random variables and therefore \bar{P} is an easier measure to work with. Now, we perform a measure change to construct the real-world measure P from the ideal-world measure \bar{P} by invoking a discrete-time version of Girsanov’s theorem. Let $\phi(z)$ denote the probability density function of a standard normal random variable z . For each component g , write

$$\lambda_l^g := \frac{\phi(\sigma^g(\mathbf{x}_{l-1})^{-1}(y_l^g - f^g(\mathbf{x}_{l-1})))}{\sigma^g(\mathbf{x}_{l-1})\phi(y_l^g)}.$$

The Radon-Nikodým derivative of P with respect to $\bar{P}, \frac{dP}{d\bar{P}}|_{\mathcal{F}_k} := \Lambda_k$, is defined by

$$\Lambda_k = \prod_{g=1}^d \prod_{l=1}^k \lambda_l^g, \quad k \geq 1, \quad \Lambda_0 = 1.$$

To obtain the estimates of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under the real world measure, we first perform all calculations under the reference probability measure \bar{P} . Calculations under the two measures are linked by the Bayes’ theorem for conditional expectation.

where

$$b_k^i = \prod_{g=1}^d \frac{\phi((y_k^g - f_i^g)/\sigma_i^g)}{\sigma_i^g \phi(y_k^g)}.$$

Notation: For any \mathcal{Y}_k -adapted process X_k , write $\hat{X}_k := E[X_k|\mathcal{Y}_k]$ and $\gamma(X)_k := \bar{E}[\Lambda_k X_k|\mathcal{Y}_k]$.

Invoking Bayes' theorem again, we get

$$\hat{X}_k = \frac{\gamma(X)_k}{\bar{E}[\Lambda_k|\mathcal{Y}_k]}. \quad (5.7)$$

To estimate the parameters of the model, recursive filters will be derived for several quantities of interest. For $r, s, t = 1, \dots, N$, let J_k^{rst} denote the number of jumps from $(\mathbf{e}_s, \mathbf{e}_t)$ to state \mathbf{e}_r up to time k , that is,

$$J_k^{rst} = \sum_{l=1}^k \langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle;$$

O_k^{rs} represents the occupation time of the WMC spent in state $(\mathbf{e}_r, \mathbf{e}_s)$ up to time k , that is,

$$O_k^{rs} = \sum_{l=1}^k \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_s \rangle;$$

O_k^r denotes the occupation time spent by the weak Markov chain in state \mathbf{e}_r up to time k , that is,

$$O_k^r = \sum_{l=1}^k \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle;$$

$T_k^r(g)$ is the level sum for the state \mathbf{e}_r , that is,

$$T_k^r(h) = \sum_{l=1}^k h(y_l) \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle.$$

Here, h is a function with the form $h(y) = y$ or $h(y) = y^2$.

The quantities considered in the above four related processes are needed in the estimation of model parameters as illustrated in Proposition 5.2.1 below. We shall take advantage of the semi-

martingale representation in (5.1) and best estimate of an adapted process X in (5.7) to obtain recursive formulas for the vector quantities $\gamma\left(J_k^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)$, $\gamma\left(O_k^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)$, $\gamma\left(O_k^r\xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)$ and $\gamma\left(T_k^r(h)\xi(\mathbf{x}_k, \mathbf{x}_{k-1})\right)$. The recursive relation of these vector processes and \mathbf{q}_k under a multi-dimensional observation set-up are given in the following proposition.

Proposition 5.2.1 *Let \mathbf{V}_r , $1 \leq r \leq N$ be an $N^2 \times N^2$ matrix such that the $((i-1)N+r)$ th column of \mathbf{V}_r is \mathbf{e}_{ir} for $i = 1 \dots N$ and zero elsewhere. If \mathbf{B} is the diagonal matrix defined in equation (5.6) then*

$$\mathbf{q}_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k \quad (5.8)$$

and

$$\gamma(J^{rst}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(J^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \langle \mathbf{\Pi}\mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \mathbf{q}_k, \mathbf{e}_{st} \rangle \mathbf{e}_{rs}, \quad (5.9)$$

$$\gamma(O^{rs}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \langle \mathbf{q}_k, \mathbf{e}_{rs} \rangle \mathbf{\Pi}\mathbf{e}_{rs}, \quad (5.10)$$

$$\gamma(O^r\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^r\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \mathbf{V}_r \mathbf{\Pi}\mathbf{q}_k, \quad (5.11)$$

$$\gamma(T^r(h)\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(T^r(h)\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + h(y_{k+1}^g)b_{k+1}^r \mathbf{V}_r \mathbf{\Pi}\mathbf{q}_k. \quad (5.12)$$

Proof See Appendices A and B for an analogous proof for each of the filters under the single-observation setting. ■

Similar to equation (5.5), we determine the normalized filter estimates of $\gamma(J^{rst})_k$, $\gamma(O^{rs})_k$, $\gamma(O^r)_k$ and $\gamma(T^r(h))_k$ by summing up the components of the vector expressions given in equations (5.9) to (5.12).

We adopt the EM algorithm to estimate the optimal parameters. The calculation is similar to the technique for the single-observation set-up. The estimates are expressed in terms of the recursions provided in equations (5.9)-(5.12) and given in the following proposition.

Proposition 5.2.2 *Suppose the observation is d -dimensional and the set of parameters $\{\hat{\mathbf{a}}_{rst}, \hat{f}_r^g, \hat{\sigma}_r^g\}$*

determines the dynamics of y_k^g , $k \geq 1$, $1 \leq g \leq d$. Then the EM estimates for these parameters are given by

$$\hat{a}_{rst} = \frac{\hat{J}_k^{rst}}{\hat{O}_k^{st}} = \frac{\gamma(J^{rst})_k}{\gamma(O^{st})_k}, \quad \forall \text{ pairs } (r, s), \quad r \neq s, \quad (5.13)$$

$$\hat{f}_r^g = \frac{\hat{T}_k^r}{\hat{O}_k^r} = \frac{\gamma(T^r(y^g))_k}{\gamma(O^r)_k}, \quad (5.14)$$

$$\hat{\sigma}_r^g = \sqrt{\frac{\hat{T}^r((y^g)^2)_k - 2\hat{f}_r^g \hat{T}^r(y^g)_k + (\hat{f}_r^g)^2 \hat{O}_k^r}{\hat{O}_k^r}}. \quad (5.15)$$

Proof See Appendix C for an analogous proof of each estimate under the single observation setting. ■

Given the observation up to time k , new parameters $\hat{a}_{rst}(k)$, $\hat{f}_r^g(k)$, $\hat{\sigma}_r^g(k)$, $1 \leq r, s, t \leq N$ are provided by equations (5.13)-(5.15). The recursive filters for the unobserved Markov chain and related processes in Proposition 5.2.1 can easily get updated every time new information arrives. Thus, we obtain a dynamic parameter estimation.

5.3 Implementation

We implement the recursive filters derived in the previous section on yields of 3-month and 6-month US T-bills, 1-year and 5-year US T-notes, and 20-year and 30-year US bonds. The dataset of yields, compiled by the Bank of Canada, contains 718 daily vector observations from 22 December 2008 to 31 October 31 2011. The evolution of yields underwent several regimes as evidenced by the changes in parameter values and the summary descriptive statistics (see Table 5.1) signifying that the data are coming from a distribution with heavy tails relative to the normal distribution. In particular, we see that the values of excess kurtosis for the yield curves are higher than those from a normal distribution. Regime-switching models are designed to capture this type of data behavior. Tables 5.1 and 5.2 display possible segregations of the actual data into either two or three states based on mean and volatility levels. This preliminary analysis of the actual data reveals that yield volatilities exhibit some degree of correlation to

the mean and maturity. From Table [?] and assuming we have two distinct states corresponding to Dec/08-July/10 and August/10-Nov/11, we see that when maturity is short, yield volatilities are higher with relatively high means; when maturity is long, yield volatilities are higher with lower means.

Maturity	Overall			Dec/08-July/10		August/10-Nov/11	
	Mean	STD	Ex. Kurtosis	Mean	STD	Mean	STD
3-month	0.13	0.06	0.57	0.14	0.07	0.12	0.04
6-month	0.23	0.09	1.97	0.26	0.10	0.20	0.05
1-year	0.43	0.16	0.22	0.45	0.12	0.41	0.20
5-year	3.39	0.57	0.70	3.33	0.42	3.45	0.71
20-year	4.07	0.46	7.88	4.14	0.38	3.98	0.53
30-year	4.12	0.50	6.65	4.15	0.50	4.09	0.48

Table 5.1: Descriptive summary statistics and data segregation into two states

Maturity	Dec/08-Feb/09		March/09-April/11		May/11-Nov/11	
	Mean	STD	Mean	STD	Mean	STD
3-month	0.17	0.07	0.13	0.04	0.09	0.04
6-month	0.34	0.08	0.20	0.04	0.18	0.06
1-year	0.54	0.11	0.38	0.14	0.44	0.19
5-year	3.11	0.47	3.41	0.57	3.71	0.53
20-year	3.97	0.43	4.14	0.44	4.01	0.50
30-year	3.86	0.55	4.30	0.38	4.01	0.50

Table 5.2: Segregation of data into three states

The dataset on yield values is a six-dimensional observation process, whose dynamics are given by

$$y_{k+1}^g = f^g(\mathbf{x}_k) + \sigma^g(\mathbf{x}_k)z_{k+1}^g, \quad g = 1, \dots, 6,$$

where $\mathbf{f}^g = (f_1^g, \dots, f_N^g) \in \mathbb{R}^N$ and $\sigma^g = (\sigma_1^g, \dots, \sigma_N^g) \in \mathbb{R}^N$ are governed by the WHMM \mathbf{x} . The implementation procedure starts with choosing the initial values for \mathbf{f}^g and σ^g , $g = 1, \dots, 6$. All non-zero entries in the transition matrix $\mathbf{\Pi}$ are set to $1/N$. The data are processed in 71 batches, and a batch consists of 10 yield vectors. Each algorithm run through a batch of data is termed as one complete pass or algorithm step. At the end of each step, new estimates for \mathbf{f} , σ and \mathbf{A} are computed. From the estimates of \mathbf{A} , we construct $\mathbf{\Pi}$. These new estimates are

in turn used as initial parameter values in the succeeding batch data processing that employs the recursive filter equations. This self-tuning algorithm allows a fortnightly update of the parameters.

Figure 5.1 illustrates the evolution of the transition probabilities under the two-state setting. The plot in the top panel shows the probabilities of staying in the same regime as the previous step. The plot in the bottom panel shows the probabilities of switching to a different state from the previous step. Except for some jumps in the probability values around the 50th algorithm pass, the bond market is quite stable as demonstrated by the relatively smooth evolution of probabilities. The large changes correspond to yield fluctuations over a brief period of time, e.g., 6-month T-bill rate increased from 0.29 on 31 Dec 2010 to 0.61 on 03 Jan 2011, and 20-year T-bond rate increased from 4.13 on 31 Dec 31 2010 to 4.55 on 05 Jan 2011. These significant changes during a short time span constitute evidence of states switching captured by the WHMM. Additionally, these market changes are reflected in the dynamics of parameter estimates.

Figures 5.2 and 5.3 show plots depicting the movement through time of the optimal parameter estimates for each yield vector under the 2-state WHMM set-up. Furthermore, the values of \mathbf{f} and σ are positively correlated with the yield maturity, i.e., the longer the maturity, the higher the mean and volatility levels. The evolution of parameters for 1-, 5- and 20-year yields support our preliminary analysis that the 1-year and 5-year yields have states characterized by high (low) means and high (low) volatilities. The 20-year yield series, however, has states characterized by low (high) means and high (low) volatilities. Such consistent behavioral mean-volatility relationship patterns are not necessarily present for yields of instruments that have either very short or very long maturities. It is worth mentioning that parameters appear to stabilize after approximately seven steps through this online algorithm. The same patterns are produced regardless of the choice of the initial values. The choice of the initial parameter values though

can affect the speed of convergence. For the 3-state setting, we report the final estimates of \mathbf{A} , \mathbf{f} and σ after the final algorithm step in Table 5.3.

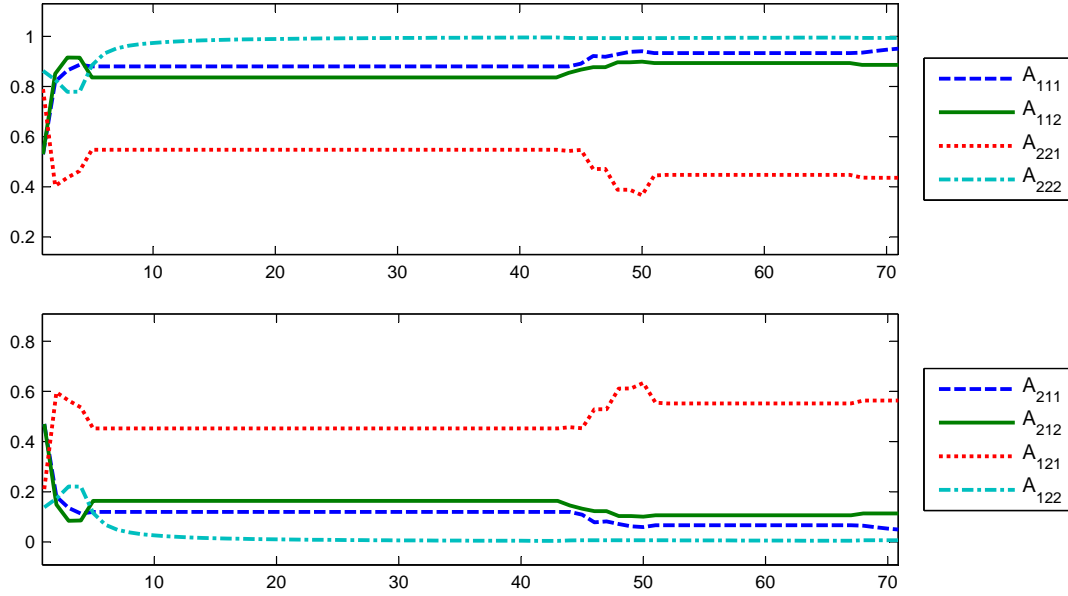


Figure 5.1: Evolution of estimates for transition probabilities through algorithm steps under the 2-state setting

Final estimation:

A matrix:	$\begin{pmatrix} 0.818 & 0.848 & 0.836 & 0.029 & 0.000 & 0.005 & 0.041 & 0.007 & 0.010 \\ 0.091 & 0.076 & 0.082 & 0.942 & 1.000 & 0.989 & 0.041 & 0.008 & 0.010 \\ 0.091 & 0.076 & 0.082 & 0.029 & 0.000 & 0.006 & 0.918 & 0.985 & 0.980 \end{pmatrix}$		
f matrix:	$\begin{pmatrix} 0.06 & 0.13 & 0.07 \\ 0.53 & 0.23 & 0.50 \\ 0.87 & 0.43 & 0.91 \\ 3.04 & 3.39 & 3.12 \\ 3.79 & 4.08 & 3.85 \\ 3.56 & 4.14 & 3.62 \end{pmatrix}$	$\sigma \text{ matrix:}$	$\begin{pmatrix} 0.16 & 0.06 & 0.07 \\ 0.08 & 0.08 & 0.05 \\ 0.02 & 0.16 & 0.05 \\ 0.35 & 0.57 & 0.20 \\ 0.48 & 0.45 & 0.21 \\ 0.44 & 0.49 & 0.21 \end{pmatrix}$

Table 5.3: Parameter estimates at the end of final algorithm step for $N = 3$

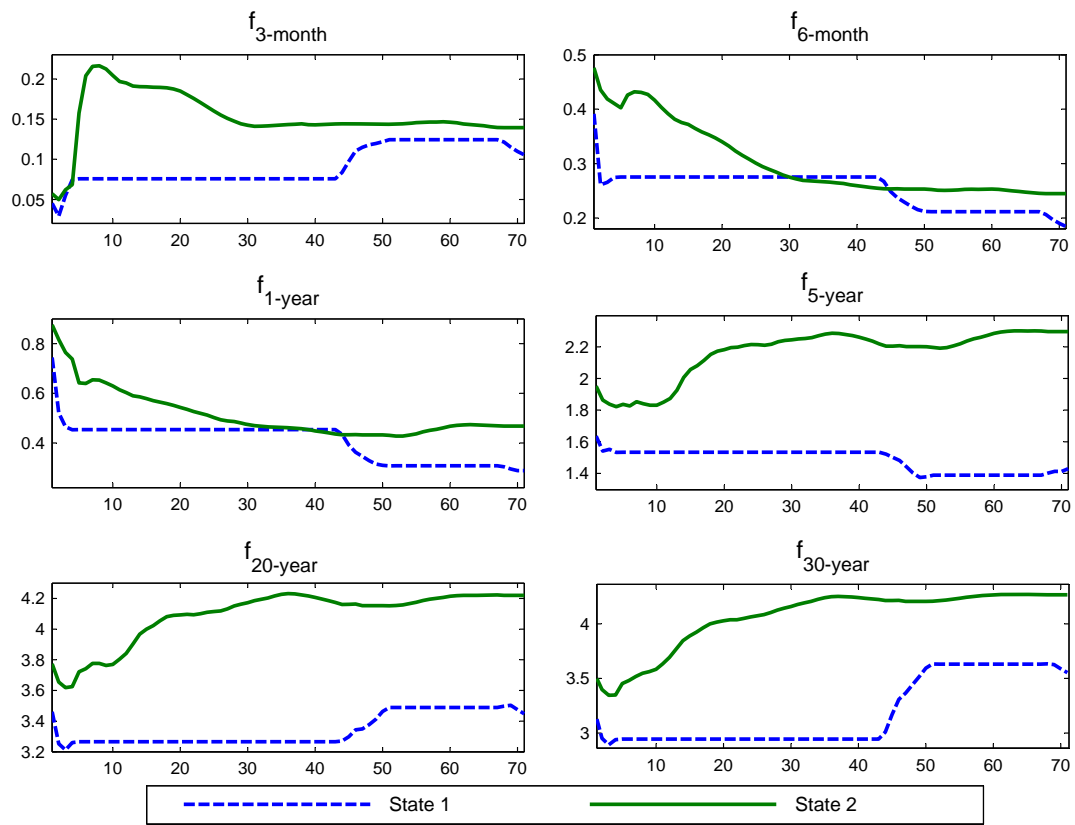


Figure 5.2: Evolution of estimates for f through algorithm steps under the 2-state setting

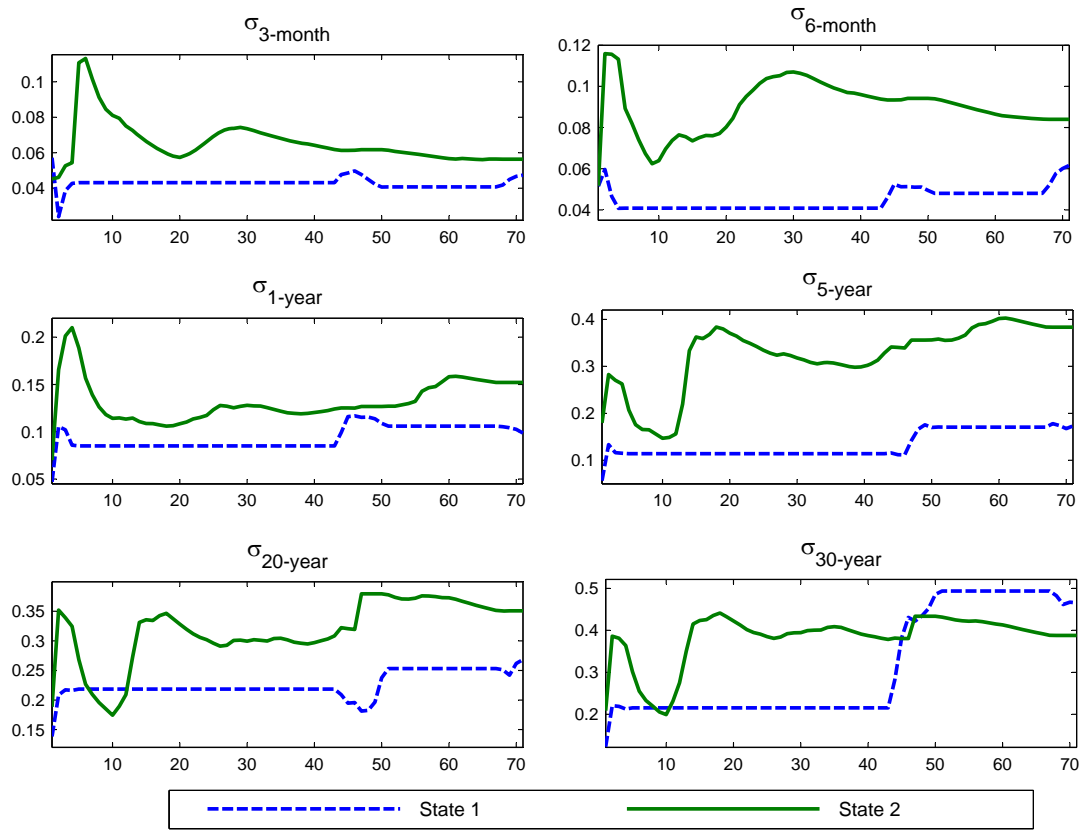


Figure 5.3: Evolution of estimates for σ through algorithm steps under the 2-state setting

5.4 Forecasting and error analysis

In this section, we shall use the model parameter estimates to forecast yield values covering an h -day ahead horizon. The semi-martingale representation of \mathbf{x} in (5.1) leads to

$$E[\xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_k] = \mathbf{\Pi} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi} \mathbf{p}_k. \quad (5.16)$$

Furthermore, we have

$$E[\xi(\mathbf{x}_{k+h}, \mathbf{x}_{k+h-1}) | \mathcal{Y}_k] = \mathbf{\Pi}^h \mathbf{p}_k, \text{ for } h = 1, 2, \dots \quad (5.17)$$

Recall that $\mathbf{\Pi}$ is constructed from \mathbf{A} , which is defined by

$$a_{lmv} = P(\mathbf{x}_{k+1} = \mathbf{e}_l | \mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v),$$

so that equation (5.16) gives

$$E[\mathbf{x}_{k+1} | \mathcal{Y}_k] = \mathbf{A} \mathbf{p}_k. \quad (5.18)$$

Hence, from equations (5.17) and (5.18),

$$E[\mathbf{x}_{k+h} | \mathcal{Y}_k] = \mathbf{A} \mathbf{p}_{k+h-1} = \mathbf{A} \mathbf{\Pi}^{h-1} \mathbf{p}_k. \quad (5.19)$$

Using equation (5.19), the best estimate of the h -step ahead predicted yields y_{k+h}^i given available information up to time k is

$$\hat{y}_{k+h}^i = E[y_{k+h}^i | \mathcal{Y}_k] = \langle \mathbf{f}^i, \mathbf{A} \mathbf{\Pi}^{h-1} \mathbf{p}_k \rangle, \text{ for } 1 \leq i \leq d. \quad (5.20)$$

The conditional variance for the predicted yields are calculated using

$$\text{Var}[y_{k+h}^i | \mathcal{Y}_k] = (\mathbf{f}^i)^\top \text{diag}(\mathbf{A} \mathbf{\Pi}^{h-1} \mathbf{p}_k) \mathbf{f}^i + (\boldsymbol{\sigma}^i)^\top \text{diag}(\mathbf{A} \mathbf{\Pi}^{h-1} \mathbf{p}_k) \boldsymbol{\sigma}^i - \langle \mathbf{f}^i, \mathbf{A} \mathbf{\Pi}^{h-1} \mathbf{p}_k \rangle^2, \quad (5.21)$$

where $\text{diag}(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)$ is a diagonal matrix whose diagonal entries are the components of the vector $(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)$.

The determination of the optimal number of states given a particular dataset is an important statistical inference problem. Hardy [14], and Erlwein and Mamon [11] applied the AIC to determine the optimal number of regimes in HMM-based models. The AIC is a measure of the relative goodness of fit of a statistical model. It offers a relative measure of lost information described by the trade-off between bias and variance in the model construction. The AIC is calculated as

$$\text{AIC} = 2s - 2\mathcal{L}(\theta),$$

where s is the number of parameters and $\mathcal{L}(\theta)$ denotes the likelihood function of the model. The preferred model is the one that gives the minimum AIC value. For the vector observation process \mathbf{y}_k in each pass, the log-likelihood of the parameter set θ is given by

$$\mathcal{L}(\theta) = \sum_{l=1}^{\# \text{ in batch}} \sum_{i=1}^d \sum_{r=1}^N \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \left(-\frac{1}{2} \log \left(2\pi\sigma^i(\mathbf{x}_{l-1})^2 \right) - \frac{\left(y_l^i - f^i(\mathbf{x}_{l-1}) \right)^2}{2\sigma^i(\mathbf{x}_{l-1})^2} \right). \quad (5.22)$$

The calculated AIC values for the 1-, 2-, 3- and 4-state models after each algorithm step are presented in Figure 5.4. Both 1- and 2-state models are reasonable in capturing the dynamics of our data gauging from this criterion with the 1-state model producing the smallest AIC values. The results indicate that both 1- and 2-state models perform better than the 3- and 4-state models. The model we proposed requires the estimation of $(N^2 - 1)N + 2mN$ parameters, where m is the number of securities. As N becomes gradually higher, the number of needed estimations rises quickly leading to higher AIC's. Nonetheless, the AIC cannot measure how well a model fits the actual time series data. In order to assess the goodness of fit of the one-step ahead forecasts, we evaluate the RMSE for the 1-, 2-, 3- and 4-state WHMM-based term structure models. The results of this error analysis are given in Table 5.4.

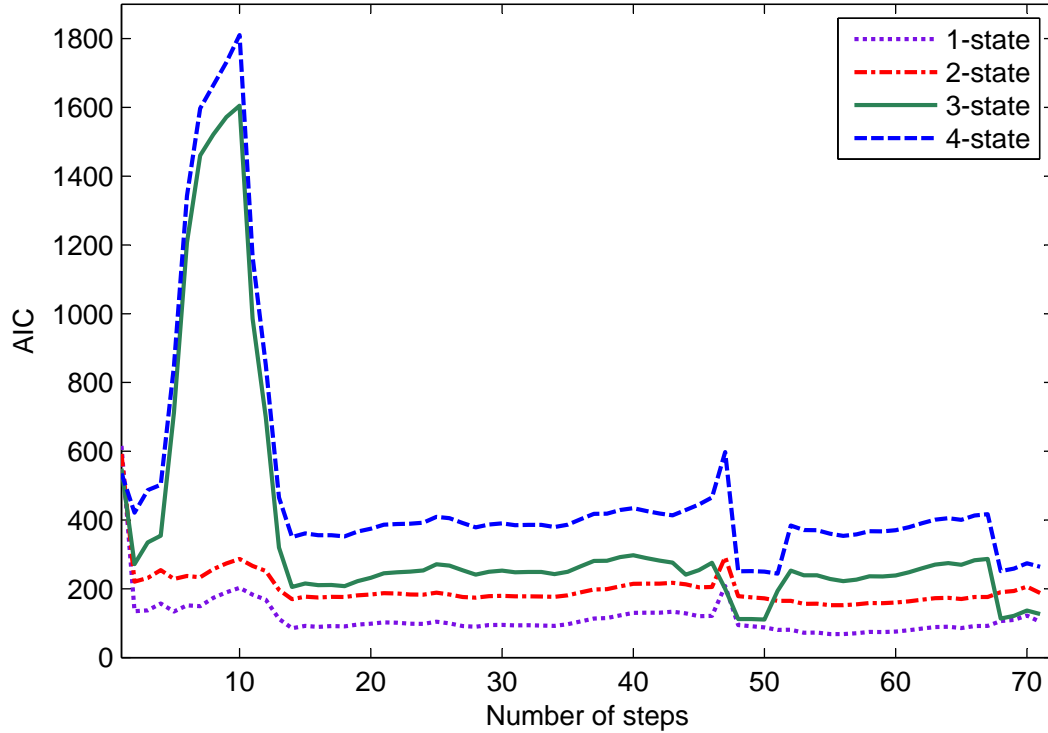


Figure 5.4: AIC for the 1-, 2-, 3- and 4-state models

Clearly, the 2-state model outperforms the model with no switching in terms of lower forecasting errors. The large improvement in the error implies that the models with regime switching can generate better price forecasts. The comparison of error measures also shows that the 4-state model is able to forecast the short-maturity yields better than the 2-state model. However, the improvement is not significant. Since a larger number of state increases the complexity of parameter estimation, a 2-state model is sufficient to model the yield values.

State setting	3-month	6-month	1-year	5-year	20-year	30-year
1	0.0558	0.0864	0.1631	0.4923	0.4363	0.4732
2	0.0539	0.0821	0.1458	0.3418	0.2994	0.3590
3	0.0558	0.0854	0.1619	0.4921	0.4357	0.4722
4	0.0524	0.0806	0.1426	0.3656	0.3194	0.3710

Table 5.4: RMSE for one-step ahead predictions versus actual values

Figure 5.5 exhibits the actual yields and 1-step ahead forecasted yields for the 3- and 6-month T-bills, 1- and 5-year T-notes, and 20- and 30-year T-bonds. The 99% confidence intervals for the predicted yields is also displayed and were calculated using $E[y_{k+1}^i | \mathcal{Y}_k] \pm$

2.575 $\sqrt{\text{Var}[y_{k+1}^i | \mathcal{Y}_k]}$. The resulting forecasts follow the actual data quite well. Empirical results confirm that the WHMM can capture most of the market dynamics.

In [31], the forecasting performance of the one-dimensional WHMM is compared with that of the regular HMM using the dataset on S&P500 prices. The results suggest that the WHMM outperforms the HMM over a long forecasting horizon. In this empirical implementation, we also evaluate the goodness of fit of the h -day ahead forecasts using the RMSE and APE as benchmarks. The multi-dimensional WHMM-based term structure model is compared with the regular multi-dimensional HMM model using these two criteria. The RMSE for an h -day ahead prediction of y^i , $i = 1, \dots, d$ is given by

$$\text{RMSE}(i, h) = \sqrt{\frac{1}{M} \sum_{k=1}^M (y_{k+h}^i - \hat{y}_{k+h}^i)^2},$$

where M is the number of forecast points. Similarly, the APE for an h -day ahead prediction of y^i is defined by

$$\text{APE}(i, h) = \frac{1}{M} \sum_{k=1}^M \left| \frac{y_{k+h}^i - \hat{y}_{k+h}^i}{y_{k+h}^i} \right|.$$

Following the idea in Date, et al. [7] in measuring the prediction performance of multivariate models, we calculate the average RMSE (AvRMSE) and average APE (AvAPE) over six yields. AvRMSE(h) is the average of RMSE(i, h) over yield values with different maturities, i.e.,

$$\text{AvRMSE}_h = \frac{1}{d} \sum_{i=1}^d \text{RMSE}(i, h).$$

On the other hand, AvAPE(h) denotes the average of APE(i, h) over yield values with different maturities, i.e.,

$$\text{AvAPE}_h = \frac{1}{d} \sum_{i=1}^d \text{APE}(i, h).$$

These error analyses are displayed in Tables 5.5-5.8. In Table 5.5, the one-state WHMM coincides with the one-state HMM. Under the WHMM framework, memory is a property of the

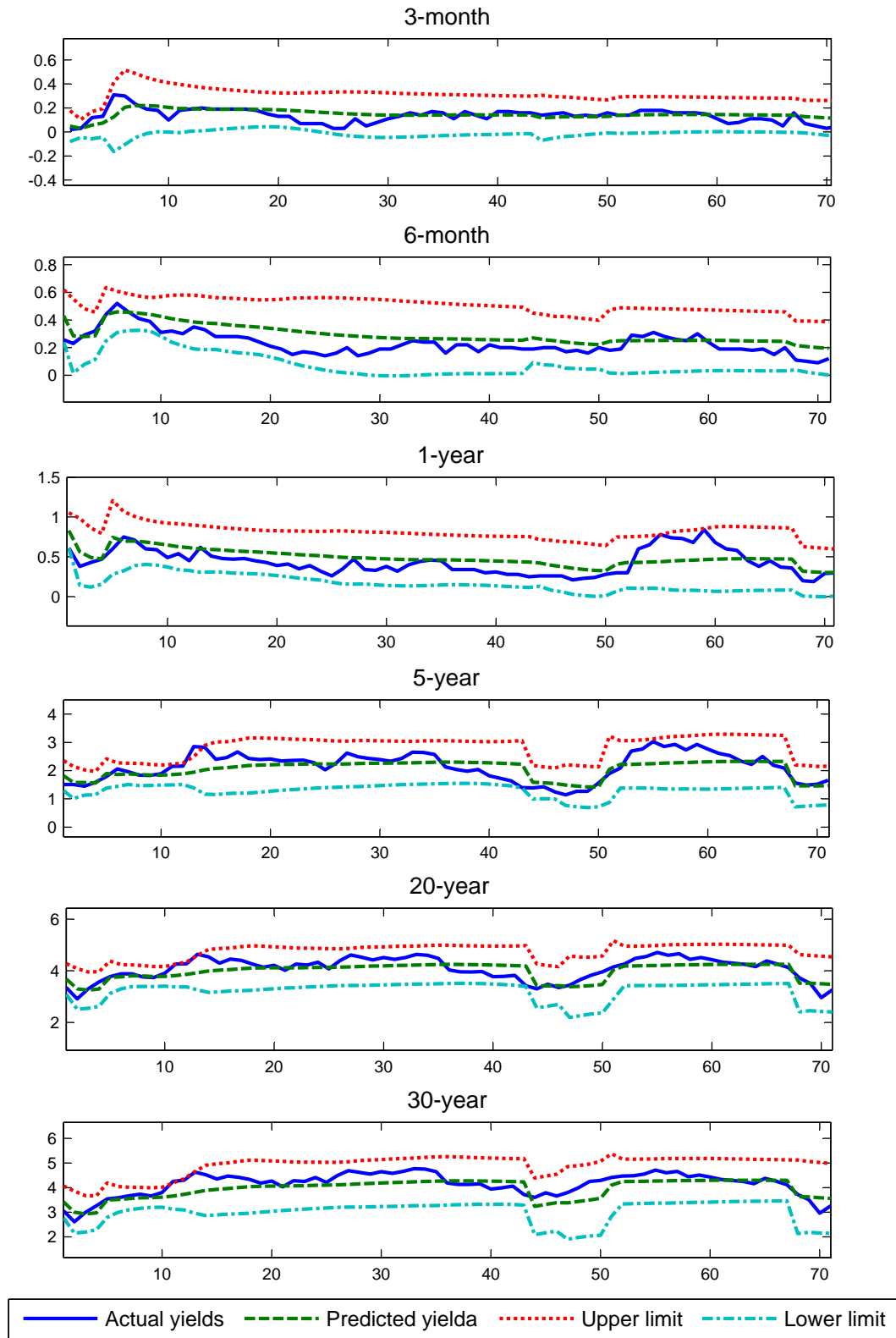


Figure 5.5: One-step ahead predicted values (%) versus actual Treasury yields (%) under a 2-state WHMM setting

	<i>h</i> -day ahead						
	1	2	3	4	5	6	7
AvRMSE _{<i>h</i>} of WHMM/HMM	0.2845	0.2862	0.2825	0.2883	0.2842	0.2828	0.2943
AvAPE _{<i>h</i>} of WHMM/HMM	0.2843	0.2687	0.2690	0.2722	0.2817	0.3073	0.3018

Table 5.5: Error analysis of WHMM and HMM models under the 1-state setting

	<i>h</i> -day ahead						
	1	2	3	4	5	6	7
AvRMSE _{<i>h</i>} of WHMM	0.2137	0.2175	0.2173	0.2263	0.2241	0.2812	0.2351
AvRMSE _{<i>h</i>} of HMM	0.2738	0.2773	0.2742	0.2815	0.2794	0.3054	0.2897
AvAPE _{<i>h</i>} of WHMM	0.2465	0.2352	0.2393	0.2436	0.2546	0.2812	0.2769
AvAPE _{<i>h</i>} of HMM	0.2789	0.2643	0.2656	0.2694	0.2794	0.3054	0.2998

Table 5.6: Error analysis of WHMM and HMM models under the 2-state setting

underlying market state process. A one-state Markov chain stays in only one state throughout the progression of time. That is, there is no memory of visiting other states in the previous steps. This is why WHMM collapses to the HMM set-up under the one-state setting. The 2-state WHMM gives a better fit than the HMM in terms of lower forecasting errors with respect to both metrics. The differences between errors from the WHMM and HMM models within the 3-state setting, shown in Table 5.7, are too small to make any practical significance. The 4-state WHMM seems to outperform the regular HMM in the long-horizon forecasting under the RMSE but not for the APE metric.

5.5 Conclusion

In this chapter, we put forward a multivariate WHMM-driven term structure model where the means and volatilities of vector observations are governed by a second-order Markov chain in

	<i>h</i> -day ahead						
	1	2	3	4	5	6	7
AvRMSE _{<i>h</i>} of WHMM	0.2839	0.2857	0.2819	0.2877	0.2838	0.2725	0.2851
AvRMSE _{<i>h</i>} of HMM	0.2855	0.2841	0.2844	0.2864	0.2866	0.2328	0.2451
AvAPE _{<i>h</i>} of WHMM	0.2828	0.2673	0.2676	0.2707	0.2804	0.2782	0.2763
AvAPE _{<i>h</i>} of HMM	0.2855	0.2867	0.2670	0.2773	0.2875	0.2849	0.2805

Table 5.7: Error analysis of WHMM and HMM models under the 3-state setting

	<i>h</i> -day ahead						
	1	2	3	4	5	6	7
AvRMSE _{<i>h</i>} of WHMM	0.2219	0.2252	0.2231	0.2320	0.2318	0.2307	0.2399
AvRMSE _{<i>h</i>} of HMM	0.2098	0.2057	0.2182	0.2291	0.2368	0.2437	0.2514
AvAPE _{<i>h</i>} of WHMM	0.2485	0.2364	0.2388	0.2416	0.2544	0.2805	0.2775
AvAPE _{<i>h</i>} of HMM	0.2086	0.2113	0.2129	0.2172	0.2218	0.2346	0.2406

Table 5.8: Error analysis of WHMM and HMM models under the 4-state setting

discrete time. The proposed model is tested on time series data of yields covering 3- and 6-month US T-bills, 1- and 5-year US T-notes, and 20- and 30-year US T-bonds. A multivariate filtering technique along with the EM algorithm was employed in the optimal estimation of parameters. The algorithms were run on batches of data and parameters are updated when new information arrives thereby making the model self-tuning. The empirical results of the implementation of filters and parameter estimation demonstrate the adequacy of the proposed model in capturing market dynamics and regime changes in the data. We applied the AIC to determine the optimal number of regimes and assessed the goodness of fit of the one-step ahead forecasts generated by the 1-, 2-, 3- and 4-state models. We found that within the dataset examined, a two-state model is deemed sufficient to capture the term structure dynamics. An analysis of the *h*-day ahead predictions was also presented and results from WHMM were compared with those from the regular HMM. The numerical results in this chapter manifest the merits of WHMM as it outperforms the HMM in terms of low forecasting errors. We attribute such improved performance to building a model that takes into account both regime switching and memories in the data series.

References

- [1] F. Audrino and M. C. Mederos. Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics*, 26:999–1022, 2011.
- [2] D. O. Cajueiro and B. M. Tabak. Long-range dependence and multifractality in the term structure of LIBOR interest rates. *Physica A: Statistical Mechanics and its Applications*, 373:603–614, 2007.
- [3] D. O. Cajueiro and B. M. Tabak. Time-varying long-range dependence in US interest rates. *Chaos, Solitons and Fractals*, 34:360–367, 2007.
- [4] W. K. Ching, T. K. Siu, and L. M. Li. Pricing exotic options under a higher-order hidden Markov model. *Journal of Applied Mathematics and Decision Sciences*, pages 1–15, 2007. Article ID 18014.
- [5] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.
- [6] S. Dajcman. Time-varying long-range dependence in stock market returns and financial market disruptions - a case of eight European countries. *Applied Economics Letters*, 19:953–957, 2012.
- [7] P. Date, L. Jalen, and R. Mamon. A partially linearized sigma point filter for latent state estimation in nonlinear time series models. *Journal of Computational and Applied Mathematics*, 233:2675–2682, 2010.

- [8] J. C. Duan and K. Jacobs. A simple long-memory equilibrium interest rate model. *Economics Letters*, 53:317–321, 1996.
- [9] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer, New York, 1995.
- [10] R. J. Elliott, W. Hunter, and B. Jamieson. Financial signal processing: A self calibrating model. *International Journal of Theoretical and Applied Finance*, 4:567–584, 2001.
- [11] C. Erlwein and R. S. Mamon. An online estimation scheme for a Hull-White model with HMM-driven parameters. *Statistical Methods and Applications*, 18:87–107, 2009.
- [12] M. Guidolin and A. Timmermann. Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, 150:297–311, 2009.
- [13] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and business cycle. *Econometrica*, 57:357–384, 1989.
- [14] M. Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5:41–53, 2001.
- [15] M. J. Holmes, R. Dutu, and X. Cui. Real interest rates, inflation and the open economy: A regime-switching perspective on Australia and New Zealand. *International Review of Economics and Finance*, 18:351–360, 2009.
- [16] J. Hull and A. White. Pricing interest-rate derivative securities. *Review of Financial Studies*, 3:573–592, 1990.
- [17] J. Hunt and P. Devolder. Semi-Markov regime switching interest rate models and minimal entropy measure. *Physica A*, 390:3767–3781, 2011.
- [18] C. Landén. Bond pricing in a hidden Markov model of the short rate. *Finance Stochastics*, 4:371–389, 2000.

- [19] J. Maheu. Can GARCH models capture long-range dependence? *Studies in Nonlinear Dynamics and Econometrics*, 9:Article 1, 2005.
- [20] B. B. Mandelbrot. When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, 53:225–236, 1971.
- [21] J. McCarthy, C. Pantalone, and H. C. Li. Investigating long memory in yield spreads. *Journal of Fixed Income*, 19:73–81, 2009.
- [22] L. Meligkotsidou and P. Dellaportas. Forecasting with non-homogeneous hidden Markov models. *Statistics and Computing*, 21:439–449, 2011.
- [23] R. C. Merton. Theory of rational option pricing. *Bell Journal of Economics*, 4:141–183, 1973.
- [24] C. C. Nieh, S. Wu, and Y. Zeng. Regime shifts and the term structure of interest rates. In C. F. Lee and J. Lee, editors, *Handbook of Quantitative Finance and Risk Management*, pages 1121–1134. Springer, 2010.
- [25] T. K. Siu, W. K. Ching, and E. Fung. Extracting information from spot interest rates and credit ratings using double higher-order hidden Markov models. *Computational Economics*, 26:251–284, 2005.
- [26] T. K. Siu, W. K. Ching, E. Fung, M. Ng, and X. Li. A higher-order Markov-switching model for risk measurement. *Computers and Mathematics with Applications*, 58:1–10, 2009.
- [27] D. Smith. Markov-switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economic Statistics*, 20:183–197, 2002.
- [28] J. Solberg. *Modelling Random Processes for Engineers and Managers*. Wiley & Sons, Inc., New Jersey, 2009.

- [29] R. Startz and K. P. Tsang. An unobserved components model of the yield curve. *Journal of Money, Credit and Banking*, 42:1613–1640, 2010.
- [30] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.
- [31] X. Xi and R. Mamon. Parameter estimation of an asset price model driven by a weak hidden Markov chain. *Economic Modelling*, 28:36–46, 2011.
- [32] N. Zhou and R. S. Mamon. An accessible implementation of interest rate models with Markov-switching. *Expert Systems with Applications*, 39:4679–4689, 2012.

Chapter 6

An interest rate model incorporating memory and regime-switching

6.1 Introduction

This chapter is closely related to chapter 5 as the main theme is still the modeling of the term structure of interest rates. However, instead of modeling directly the evolution of yields, we focus on the modeling of the short rate driven by a diffusion process whose parameters are modulated by a WHMM. As the short rate process is really not observed, it is being proxied by the returns series from a fixed income instrument with a very short maturity (30-day T-bill in this case) in our numerical implementation. Other proxies for the short rate are the LIBOR rates; see Filipović [16].

Various models for the evolution of the short rate, forward rate and the LIBOR rate have been put forward in the last three decades. The modeling of interest rates is a paramount consideration in finance as the theoretical construction of the yield curve hinges on it. Various interest rates are also prime economic indicators monitored and controlled by regulatory authorities such as the Feds and other central banks. The short rate is the interest rate at which a loan can be charged for an infinitesimally short period. Classical models describing the short rate

dynamics include those developed by Vasicek [38], Cox, et al. [8], and Hull and White [22]. These models are able to capture the mean-reverting property of the interest rate process. Such property suggests that interest rates' high and low values are temporary and the value will tend to move to an average level over time. The Vasicek and CIR models assume a constant mean-reverting level. The Hull-White model is an extension of the Vasicek model in which the parameters are deterministic functions of time. Methodologies have been developed in order to capture the dynamic behavior of financial and economic variables, e.g., Hamilton [19]. A regime-switching-based approach posits that an economic environment varies and shifts between different regimes as time progresses. Such approach has the flexibility and capability of handling changes in economic states by allowing model parameters to stochastically vary. More specifically, Markov-switching models have parameters that change over time in accordance with the dynamics of an unobserved Markov chain.

Hamilton [19] pioneered the research in this field with a particular focus to economic modeling. A study by Smith [34] found evidence that the volatility depends on the level of the short rate and supports a Markov-switching model over a stochastic volatility model. Landen [25] developed an HMM for short-term interest rates, in which the rate's mean and variance are governed by a Markov process. In practice, the data fitting performance of these Markov-switching models is a central concern. As the Markov chain is not directly observed, hence it is hidden, there is a need to devise optimal, efficient and self-updating estimation techniques. As can be expected, this kind of estimation for the model parameters and Markov chain's unobservable state presents some challenges both from the practical and mathematical standpoints. In a comprehensive work, Elliott, et al. [13] provided recursive estimates of the Markov chain together with the drift and volatility parameters in a Markov-based modeling framework. Elliott, et al. [11] proposed a multivariate HMM for the short rate and filtering methods were applied to obtain optimal estimates of the parameters. Many of the developments in the regime-switching literature focus on extending various commonly known models. Elliott and Mamon [12] pro-

posed a Vasicek model where the mean-reverting level depends on a continuous-time Markov chain. Erlwein and Mamon [14] considered a Hull-White interest rate model in which the interest rate's volatility, mean-reverting level and speed of mean-reversion are all governed by a Markov chain in discrete time. The HMM filters are derived and implemented on a financial dataset and their analysis of the prediction errors along with the use of AIC shows that a two-regime model is sufficient to describe the interest rate dynamics. Zhou and Mamon [41] investigated the Vasicek, CIR and Black-Karasinski short-rate models whose parameters are modulated by a finite-state and discrete-time Markov chain. A quasi-maximum likelihood method is utilized to estimate model parameters and implementation of their algorithms was carried out on the Canadian yield rates.

Some recent studies integrate regime-switching models with other modeling approaches to obtain an enhanced methodology. Guidolin and Timmermann [18] proposed a four-state model for the US spot and forward rates; their forecasting experiment shows evidence that, at short horizons, combining regime-switching forecasts with simpler univariate time-series forecasts is able to reduce forecasting errors. A Bayesian forecasting methodology involving a discrete-time HMM with non-constant transition matrix in modeling monthly data on rates of return series was used by Meligkotsidou and Dellaportas [30]; the nonhomogeneous HMMs showed improved predictive ability over a standard homogeneous HMM. Other papers on regime-switching models feature new approaches in detecting further evidence of regime shifts in the market. Startz and Tsang [36] constructed a model in which the short-term interest rate consists of two components: a stochastic trend and a stationary cycle; their results suggest that allowing for regime switching in shock variances improves model performance. In the work by Audrino and Mederos [2], a smooth transition tree model that combines regression trees and GARCH models is employed to describe regime switches in the short-term interest rate series; their empirical results provide evidence of the power of the model in forecasting the conditional mean and variance. Utilizing an adapted unit-root test, Holmes, et al. [21] found

evidence that Australian and New Zealand interest rates switch between regimes in mean and variance. For an exhaustive review of term structure models under a regime-switching setting, see Nieh, et al. [31].

While the regular HMM is quite popular and succeeded in capturing the dynamics of many financial and economic processes, its major drawback is the memoryless property of the underlying Markov process. In this chapter, we propose a WHMM to address such deficiency of the usual HMM. As argued by Solberg [35], the real significance of WHMM is to establish that the assumption of memoryless property is not really as restrictive as it first appears. By using a higher-order Markov chain, the probability of the current state is not only dependent on just one prior time epoch but on any finite number of prior epochs. This model offers an alternative to long-range dependence models designed to capture data memory present in financial time series. Mandelbrot [27] demonstrated the benefits of incorporating memory in economics and finance applications. Cajueiro and Tabak [4, 5] showed evidence of long-range dependence in LIBOR and US interest rates. To quantify the presence of memory, Matteo [28] gave a detailed description of the generalized Hurst exponent approach and carried out an empirical analysis across different markets. McCarthy, et al. [29] examined corporate bond yield spread data and found strong evidence that data memory exists. Processes with long memory characteristics have stronger coupling between values at different time than that of short-memory processes. As illustrated in the empirical work of Bouchaud and Potters [3], long-memory processes are not normally distributed but have fatter tails and higher peaks around the mean.

Other studies have developed models to deal directly with the long-range dependence property in financial time series. Maheu [26] concluded that GARCH models can capture the long-memory property in financial volatility under some circumstances. Dajcman [9] proposed an autoregressive fractionally integrated moving average model for the returns of eight European stock market returns. Duan and Jacob's long-range dependence model [10] yields a signif-

icantly improved fit when applied to real interest rate data. However, most of the existing models in time series that take into account the data memory property have stationary parameters, which seem inadequate in real-world applications. This serves as another motivation to introduce the idea of higher-order HMMs. The underlying Markov process for these models offers a simple way yet rich enough to describe the evolution of market variables with dynamic parameters and capture as well the memory property through the dependence on the backward time recurrence. To the best of our knowledge, the embedding of WHMM into available interest rate modeling approaches in the context of dynamic parameter estimation is so far non-existent. It is our intent to show the usefulness and merits of a WHMM-based interest rate model. Attempting to accomplish a similar goal, Hunt and Devolder [23] constructed an extension of the Ho and Lee model under a semi-Markov regime-switching framework aimed to capture the data's long-memory property. An application of their proposed extension to the pricing of European bond options was given. A few applications of higher-order Markov chain in finance were considered in recent years and include modeling applications for returns of risky assets (Xi and Mamon [40]), risk management (Siu, et al. [33]), exotic option pricing (Ching, et al. [7]), and spot rates and credit ratings (Siu et al. [32]). Except for [40], none of these deal with the problem of efficient and systematic parameter estimation.

In this chapter, we investigate the development of a mean-reverting interest rate model under a weak Markov-switching framework. Following the approach in [14], the dynamics of the short-term rate can be re-written as a function of a discrete-time WMC. In particular, we assume the level and speed of mean reversion, and the volatility are governed by a second-order WMC, whose current behavior depends on the behavior exhibited at the previous two time steps. The WHMM could capture the presence of dependence in the states of the market, and therefore it is more appropriate when financial series exhibit memories. A transformation, which is fundamentally a mapping of states, is employed to convert a WHMM into a regular HMM. The estimation algorithm updates parameter estimates as soon as new data points become available.

Thus, what we come up with is a “self-tuning” estimation procedure.

This chapter is structured as follows. The next section describes the formulation of the modeling framework. A discussion on how to incorporate a WMC in the single-factor Hull-White model is presented. In section 6.3, we derive the filters for the states of the underlying WMC and other auxiliary processes through a change of reference probability technique. The recursive estimates of the speed and level of mean reversion, volatility and transition probability matrix are computed by utilizing the EM algorithm. The filtering technique and parameter estimation are then implemented in section 6.4 on the Canadian T-bill rates. In section 6.5, the h -step ahead forecasts under WHMM are generated and compared to those obtained under the usual HMM. We discuss the determination of the most appropriate number of states and how to provide standard errors on our model parameter estimates. Finally, section 6.6 provides some concluding remarks.

6.2 Model construction

For short rate models, the instantaneous spot rate r_t is the state variable. The stochastic differential equation (SDE) describing the dynamics of r_t in the Hull-White model [22] has the form

$$dr_t = \mu_t(\beta_t - r_t)dt + \zeta_t dW_t. \quad (6.1)$$

The parameters μ_t , β_t and ζ_t are deterministic functions of t and W_t is a standard Brownian motion. Originally, this kind of process was studied in the physics literature, and is known as a particular case of the Ornstein-Uhlenbeck process. The solution of this SDE is

$$r_t = r_0 e^{-\mu t} + (1 - e^{-\mu t})\beta + \zeta \int_0^t e^{-\mu(t-u)} dW_u.$$

This implies that, for any $s < t$,

$$r_t = r_s e^{-\mu(t-s)} + (1 - e^{-\mu(t-s)})\beta + \zeta \int_s^t e^{-\mu(t-u)} dW_u. \quad (6.2)$$

Note that, $E[r_t] \rightarrow \beta$ as $t \rightarrow \infty$. This property is referred to as the mean reversion of the short rate. The respective mean-reverting level and speed of mean reversion are β_t and μ_t .

In the subsequent discussion, all vectors and matrices are written in boldface English or Greek letters; vectors are in lowercase while matrices are in uppercase. Assume all processes in our modeling set-up are supported by a complete probability space (Ω, \mathcal{F}, P) . Let $\{\mathbf{x}_t\}$, $t \geq 0$, be a continuous-time WMC with finite space $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$. Without loss of generality, the points in \mathcal{S} can be identified with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\} \subset \mathbb{R}^N$, where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$ and \top denotes the transpose of a vector. The expression $\langle \mathbf{x}_t, \mathbf{e}_i \rangle$ represents the event that the economy is in state i at time t . Here, $\langle \cdot, \cdot \rangle$ stands for the inner product in \mathbb{R}^N .

Now, assume the parameters μ_t, β_t and ζ_t are governed by the WMC \mathbf{x} and therefore, the model parameters are switching among different economic regimes through time. The SDE in (6.1) for r_t can be re-written as

$$dr_t = \mu(\mathbf{x}_t)(\beta(\mathbf{x}_t) - r_t)dt + \zeta(\mathbf{x}_t)dW_t,$$

with $\mu(\mathbf{x}_t) = \langle \boldsymbol{\mu}, \mathbf{x}_t \rangle$, $\beta(\mathbf{x}_t) = \langle \boldsymbol{\beta}, \mathbf{x}_t \rangle$ and $\zeta(\mathbf{x}_t) = \langle \boldsymbol{\zeta}, \mathbf{x}_t \rangle$. For small $t - s$ such that \mathbf{x} is constant over $[s, t]$, the solution in (6.2) is

$$r_t = r_s e^{-\mu(\mathbf{x}_t)(t-s)} + (1 - e^{-\mu(\mathbf{x}_t)(t-s)})\beta(\mathbf{x}_t) + \int_s^t e^{-\mu(\mathbf{x}_t)(t-u)} \zeta(\mathbf{x}_u) dW_u. \quad (6.3)$$

Write

$$\alpha(\mathbf{x}_k) := e^{-\mu(\mathbf{x}_k)\Delta t_{k+1}}, \quad (6.4)$$

$$\eta(\mathbf{x}_k) := \beta(\mathbf{x}_k)(1 - e^{-\mu(\mathbf{x}_k)\Delta t_{k+1}}), \quad (6.5)$$

$$\sigma(\mathbf{x}_k) := \zeta(\mathbf{x}_k) \sqrt{(2\mu(\mathbf{x}_k))^{-1}(1 - e^{-2\mu(\mathbf{x}_k)\Delta t_{k+1}})} \quad (6.6)$$

$$\text{and} \quad \Delta t_{k+1} := t_{k+1} - t_k.$$

Equation (6.3) can be used to obtain a discrete-time representation of the interest rate process. While the short rate is a key interest rate and essential to no-arbitrage valuation, it cannot be observed directly. Short rates are proxied instead by the yields from short-term maturity fixed-income instruments as they are very liquid. The argument and evidence supporting the validity of such proxies is given in Filipović [16] and Chapman, et al. [6]. With the yield rate as proxy for the short rate and employing the newly defined parameters in equations (6.4)-(6.6), the observed yield value has dynamics

$$y_{k+1} = \alpha(\mathbf{x}_k)y_k + \eta(\mathbf{x}_k) + \sigma(\mathbf{x}_k)z_{k+1}. \quad (6.7)$$

The sequence $\{z_k\}$ is a sequence of $N(0, 1)$ IID random variables, which are independent from the \mathbf{x} -process.

Again, we shall focus on a WMC of order 2 to simplify the discussion and present a complete characterization of the parameter estimation. The probability involved in the next time step for a second-order WMC depends on the information on the current and previous time steps. The transition probability matrix $\mathbf{A} \in \mathbb{R}^{N \times N^2}$ is defined by

$$a_{lmv} := P(\mathbf{x}_{k+1} = \mathbf{e}_l | \mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v), \quad l, m, v \in 1, \dots, N. \quad (6.8)$$

Each entry of \mathbf{A} refers to the probability that the process enters state l given that the current

and previous states are m and v , respectively. The pertinent idea in the filtering method for WHMM is that a second-order Markov chain is transformed into a first-order Markov chain through a mapping ξ , and after which we may apply the regular filtering method. The mapping ξ is defined by

$$\xi(\mathbf{e}_r, \mathbf{e}_s) = \mathbf{e}_{rs}, \text{ for } 1 \leq r, s \leq N,$$

where \mathbf{e}_{rs} is an \mathbb{R}^{N^2} -unit vector with unity in its $((r-1)N + s)$ th position. The identification of the new first-order Markov chain with the canonical basis is given by

$$\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle = \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle.$$

The new transition probability matrix, $\mathbf{\Pi} \in \mathbb{R}^{N^2 \times N^2}$, of the new Markov chain is then

$$\pi_{ij} = \begin{cases} a_{lmv} & \text{if } i = (l-1)N + m, \quad j = (m-1)N + v \\ 0 & \text{otherwise.} \end{cases}$$

Here, each non-zero element π_{ij} refers to the probability

$$\pi_{ij} = a_{lmv} = P(\mathbf{x}_k = \mathbf{e}_l | \mathbf{x}_{k-1} = \mathbf{e}_m, \mathbf{x}_{k-2} = \mathbf{e}_v),$$

and each zero represents an impossible transition. The new Markov chain $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ has the semi-martingale representation

$$\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi} \xi(\mathbf{x}_{k-1}, \mathbf{x}_{k-2}) + \mathbf{v}_k, \tag{6.9}$$

where $\{\mathbf{v}_k\}_{k \geq 1}$ is a sequence of \mathbb{R}^{N^2} -martingale increments.

6.3 Filters and parameter estimation

Under the real world measure P , the true state of the underlying WMC \mathbf{x}_k is neither observed nor measured directly. Instead, it is contained in the noisy market observations y_k with “real world” dynamics given by equation (6.7). Our objective is to “filter” the noise out of the observation process in the best possible way. However, the derivation of filters under P is not straightforward. So, we take advantage of the Kolmogorov’s extension theorem that justifies the existence of a reference probability measure \bar{P} under which the y_k ’s are $N(0, 1)$ IID random variables. Mathematically, \bar{P} is an easier measure to work with. Thus, we perform a change of probability measure to construct the real-world measure P from the ideal-world measure \bar{P} by invoking the discrete-time version of the Girsanov’s theorem. Under P , the sequence of z_k ’s is a sequence of IID standard normal random variables, where $z_k = \sigma(\mathbf{x}_k)^{-1}(y_{k+1} - \alpha(\mathbf{x}_k)y_k - \eta(\mathbf{x}_k))$. Write

$$\lambda_l := \frac{\phi(\sigma(\mathbf{x}_l)^{-1})(y_{l+1} - \alpha(\mathbf{x}_l)y_l - \eta(\mathbf{x}_l))}{\sigma(\mathbf{x}_l)\phi(y_l)}, \quad (6.10)$$

where $\phi(z)$ denotes the probability density function of a standard normal random variable Z . The Radon-Nikodým derivative of P with respect to \bar{P} , $\frac{dP}{d\bar{P}}|_{\mathcal{F}_k} := \Lambda_k$, is defined by

$$\Lambda_k = \prod_{l=1}^k \lambda_l, \quad k \geq 1, \quad \Lambda_0 = 1, \quad (6.11)$$

and $\{\mathcal{F}_k\}$ is the filtration generated by the observation process y_k .

It has to be noted that Erlwein and Mamon [14] considered a reference measure \tilde{P} , under which the observation y_k is a sequence of $N(0, \sigma_k^2)$ IID random variables. We experimented processing datasets using filters derived under the \tilde{P} setting and then back out the real world measure P . We found that the parameter estimation algorithms under \tilde{P} have faster convergence than those under \bar{P} . This rationalizes the slightly modified construction of the Radon-Nikodým derivative as proposed in our formulation above.

To estimate the state of the new first-order Markov chain, $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under the real world measure, we first perform all calculations under the reference probability measure \bar{P} . Calculations under the two measures are then linked via Bayes' theorem for conditional expectation.

Let us derive the conditional expectation of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ given \mathcal{Y}_k under P . Write

$$p_k^{ij} := P(\mathbf{x}_k = \mathbf{e}_i, \mathbf{x}_{k-1} = \mathbf{e}_j | \mathcal{Y}_k) = E[\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij} \rangle | \mathcal{Y}_k] \quad (6.12)$$

with $\mathbf{p}_k = (p_k^{11}, \dots, p_k^{ij}, \dots, p_k^{NN}) \in \mathbb{R}^{N^2}$. Bayes' theorem tells us that

$$\mathbf{p}_k = E[\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_k] = \frac{\bar{E}[\Lambda_k \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_k]}{\bar{E}[\Lambda_k | \mathcal{Y}_k]}. \quad (6.13)$$

Writing $\mathbf{1}$ for the vector $(1, \dots, 1)^\top \in \mathbb{R}^{N^2}$, we see that

$$\sum_{i,j=1}^N \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij} \rangle = \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1} \rangle = 1.$$

Let $\mathbf{q}_k = \bar{E}[\Lambda_k \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_k]$ so that

$$\langle \mathbf{q}_k, \mathbf{1} \rangle = \bar{E}[\Lambda_k \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1} \rangle | \mathcal{Y}_k] = \bar{E}[\Lambda_k | \mathcal{Y}_k]. \quad (6.14)$$

From equations (6.13) and (6.14), we get the conditional distribution of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under P as

$$\mathbf{p}_k = \frac{\mathbf{q}_k}{\langle \mathbf{q}_k, \mathbf{1} \rangle}. \quad (6.15)$$

In order to estimate dynamically the state process $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, we derive a recursion for the

The function g in the level sum takes the form $g(y) = y$, $g(y) = y^2$ or $g(y) = y_{l-1}y_l$, for $2 \leq l \leq k$.

The four quantities above are needed in the estimation of model parameters as manifested in Proposition 6.3.1. By the semi-martingale representation in (6.9) and the best estimate of an adapted process X in (6.17), we can obtain recursive formulas for the vector quantities $J_k^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^r\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ and $T_k^r(g)\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$. The recursive relations of these vector processes and \mathbf{q}_k are given in the following proposition.

Proposition 6.3.1 *Let \mathbf{V}_r , $1 \leq r \leq N$ be an $N^2 \times N^2$ matrix such that the $((i-1)N+r)$ th column of \mathbf{V}_r is \mathbf{e}_{ir} for $i = 1 \dots N$ and zero elsewhere. If \mathbf{B} is the diagonal matrix defined in equation (6.16) then*

$$\mathbf{q}_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k \quad (6.18)$$

and

$$\gamma(J^{rst}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(J^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \langle \mathbf{\Pi}\mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \mathbf{q}_k, \mathbf{e}_{st} \rangle \mathbf{e}_{rs}, \quad (6.19)$$

$$\gamma(O^{rs}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \langle \mathbf{q}_k, \mathbf{e}_{rs} \rangle \mathbf{\Pi}\mathbf{e}_{rs}, \quad (6.20)$$

$$\gamma(O^r\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^r\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \mathbf{V}_r \mathbf{\Pi}\mathbf{q}_k, \quad (6.21)$$

$$\gamma(T^r(g)\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(T^r(g)\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + g(y_{k+1})b_{k+1}^r \mathbf{V}_r \mathbf{\Pi}\mathbf{q}_k. \quad (6.22)$$

Proof The proof is given in Appendix B.

Similar to the representation in equation (6.15), it is straightforward to determine the normalized filter estimates of $\gamma(J^{rst})_k$, $\gamma(O^{rs})_k$, $\gamma(O^r)_k$ and $\gamma(T^r(g))_k$ by summing the components of the vector expressions given in equations (6.19) to (6.22).

We adopt the EM algorithm and the filters in Proposition 6.3.1 to estimate the optimal model

parameters given by the set

$$\theta = \{a_{rst}, \alpha_r, \eta_r, \sigma_r; 1 \leq r, s, t \leq N\}.$$

The algorithm proceeds by selecting a set of initial parameters, θ_0 , for the model. The change to the updated parameter is described by a change of probability measure from P_{θ_0} to P_θ . That is, the likelihood function for estimating a parameter θ based on the information reflected in the yield values is

$$L(\theta) = E_{\theta_0} \left[\frac{dP_\theta}{dP_{\theta_0}} \middle| \mathcal{Y} \right].$$

The logarithm of the Radon-Nykodým derivative of the new measure with respect to the old measure is then calculated. A set of parameters $\hat{\theta}$ that maximizes the conditional log-likelihood is determined. It is shown in [39] that the sequence of the log-likelihoods is monotonically increasing and the associated sequence of estimates converges to a local maximum of the expectation of the likelihood function. Consider the estimation of the transition matrix. The non-zero entries of $\mathbf{\Pi}$ are the same as the entries of \mathbf{A} . We estimate the matrix \mathbf{A} and then construct $\mathbf{\Pi}$ for the computation of filters. We first perform a change of measure from P_θ to $P_{\hat{\theta}}$. Under P_θ , \mathbf{x} is a WMC with transition matrix $\mathbf{A} = (a_{rst})$. Under $P_{\hat{\theta}}$, \mathbf{x} is still a WMC and the transition matrix is $\hat{\mathbf{A}} = (\hat{a}_{rst})$. To replace \mathbf{A} by $\hat{\mathbf{A}}$, we use the likelihood function

$$\begin{aligned} \frac{dP_\theta}{dP_{\theta_0}} \middle| \mathcal{Y}_k &= \Gamma_k^A, \\ \Gamma_k^A &= \prod_{l=2}^k \prod_{r,s,t=1}^N \left(\frac{\hat{a}_{rst}}{a_{rst}} \right)^{\langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle}. \end{aligned} \quad (6.23)$$

In case $a_{rst} = 0$, take $\hat{a}_{rst} = 0$ and $\hat{a}_{rst}/a_{rst} = 1$. The model parameters are provided in the following proposition.

Proposition 6.3.2 *The EM estimates \hat{a}_{rst} , $\hat{\alpha}_r$, $\hat{\eta}_r$ and $\hat{\sigma}_r$, given the sequence of observations*

$y_k, k \geq 1$, are given by

$$\hat{a}_{rst} = \frac{\hat{J}_k^{rst}}{\hat{O}_k^{st}} = \frac{\gamma(J^{rst})_k}{\gamma(O^{st})_k}, \quad \forall \text{ pairs } (r, s), \quad r \neq s, \quad (6.24)$$

$$\hat{\alpha}_r = \frac{\hat{T}_k^r(y_{k-1}, y_k) - \eta^r \hat{T}_k^r(y_{k-1})}{\hat{T}_k^r(y_{k-1}^2)} = \frac{\gamma(T_k^r(y_{k-1}, y_k))_k - \eta^r \gamma(T_k^r(y_{k-1}))_k}{\gamma(T_k^r(y_{k-1}^2))_k}, \quad (6.25)$$

$$\hat{\eta}_r = \frac{\hat{T}_k^r(y_k) - \alpha^r \hat{T}_k^r(y_{k-1})}{\hat{O}_k^r} = \frac{\gamma(T^r(y_k))_k - \alpha^r \gamma(T^r(y_{k-1}))_k}{\gamma(O^r)_k}, \quad (6.26)$$

$$\hat{\sigma}_r^2 = \frac{\hat{T}_k^r(y_k^2) + \alpha_r^2 \hat{T}_k^r(y_{k-1}) + \eta_r^2 \hat{O}_k^r - 2\alpha_r \hat{T}_k^r(y_k y_{k-1}) - 2\eta_r \hat{T}_k^r(y_k) + 2\eta_r \alpha_r \hat{T}_k^r(y_{k-1})}{\hat{O}_k^r}. \quad (6.27)$$

Proof See Appendix C for proof of (6.24) and Appendix D for proofs of equations (6.25)-(6.27).

We see that having observations up to time k , new parameters $\hat{a}_{rst}(k), \hat{\alpha}_r(k), \hat{\eta}_r(k), \hat{\sigma}_r(k), 1 \leq r, s, t \leq N$ are then obtained using equations (6.24)-(6.27). In turn, the recursive filters for the unobserved Markov chain and related processes in Proposition 6.3.1 produce parameter updates each time new information arrives. This gives rise to a dynamic parameter estimation.

6.4 Implementation

The recursive filters specified in Proposition 6.3.1 are tested on yield rates of 30-day Canadian Treasury bills compiled by the Bank of Canada. The dataset consists of weekly T-bill yields recorded every Friday between 11 March 2002 to 09 March 2012. There are 505 data points. Tables 6.1 and 6.2 display the summary descriptive statistics of the data and possible segregation into either two states or three states, respectively. The evolution of yields undergoes several regimes characterized by different parameter values. This is supported by the summary statistics in the possible grouping periods obtained by using a least-square method assuming a one-state setting to estimate the parameters α, η and σ in each designated interval. Then, we recover the model parameters μ, β and ζ from equations (6.4)-(6.6) with $\Delta t = 1/52$. The least-square parameter estimation was carried out using the MATLAB function 'lsqcurvefit'.

	Overall	11Mar02–28Mar08	04Apr08–09Mar12
Sample Mean	0.021515	0.030109	0.008233
Sample Std	0.012979	0.007680	0.006837
α_{fit}	0.999087	0.997781	0.988183
η_{fit}	5.26×10^{-5}	4.26×10^{-5}	7.79×10^{-5}
σ_{fit}	0.001587	0.001323	0.004161
μ_{fit}	0.047500	0.115500	0.618100
β_{fit}	0.057619	0.019216	0.006593
ζ_{fit}	0.005723	0.004774	0.015093

Table 6.1: Possible segregation of data into 2 states

	Overall	11Mar02–30Sep05	07Oct05–30Jan09	06Feb09–09Mar12
Sample Mean	0.021515	0.024567	0.033035	0.005409
Sample Std	0.012979	0.003173	0.010039	0.003306
α_{fit}	0.999087	0.991748	0.999262	0.9990297
η_{fit}	5.26×10^{-5}	2.1×10^{-4}	3.3×10^{-5}	2.39×10^{-9}
σ_{fit}	0.001587	0.002184	0.001298	0.003629
μ_{fit}	0.047500	0.430900	0.038400	0.507000
β_{fit}	0.057619	0.026032	0.044706	2.47×10^{-7}
ζ_{fit}	0.005723	0.007906	0.004680	0.013148

Table 6.2: Possible segregation of data into 3 states

The preliminary analysis demonstrates possible segregation of the actual data into different states in accordance with the values of mean-reverting level, rate of mean reversion and variance. In particular, we see that the yield values y_k has a low mean-reverting speed, high mean-reverting level and low variance when the sample mean of yields is high. When the sample mean of yields is low, the estimated mean-reverting level is low, and both mean-reverting speed and the estimated variance are high.

Before we apply our filtering equations to the T-bill yield dataset, we ensure that the dataset does indeed exhibit the memory property. One way to find out presence of memory or long-range dependence is through the evaluation of the Hurst exponent. Let H be the Hurst exponent of an observed process y_k , and $\rho(q)$ denote the sample autocovariance function with time lag q , i.e., $\rho(q) = cov(y_k, y_{k+q})$. Then y_k has long memory if there exists a constant $c > 0$ such

	R/S analyses	Wavelet analyses	2nd-order Derivative	Variance vs Level
H	0.6376	0.7132	0.7414	0.7733

Table 6.3: Estimates of H under different estimators

that $\lim_{h \rightarrow \infty} \rho(q)/(cq^{2H-2}) = 1$ for $H \in (0.5, 1)$. There are several methods to estimate H in the financial and economic literature. The seminal work by Hurst [24] on re-scaled range statistical analysis R/S presented an estimator for H . Abry, et al. [1] used wavelet analysis to estimate H . Taqqu, et al. [37] discussed and compared nine different estimators in their research. In order to test the presence of long-range dependence in our data, we estimate H using four different estimators. The first two are the R/S and wavelet estimators. The third approach is based on the second-order discretized derivative and the last estimator is based on the slope of the loglog plot of the data's level versus variance. The MATLAB function 'wfbmesti' is used to find the estimates. The results are shown in Table 6.3. All four estimators give values over 0.5. Hence, this indicates that the data possesses some lag dependence.

The implementation procedure starts by selecting initial values for the parameters. All non-zero entries in the transition matrix $\mathbf{\Pi}$ are set to $1/N$. The initial values of other parameters are $\alpha_r = 0.99$, $\eta_r = 0.02$, $\sigma_r = 0.1$, $r = 1, \dots, N$. The dataset is processed in batches of 12 data points. An algorithm run, which processes a batch of data points, constitutes one complete algorithm step. At the end of each algorithm step, new parameter estimates are obtained and they are utilized iteratively as new initial values for the next batch. Since there are 12 weekly data points in a batch, the parameters are updated every 3 months. Different sizes of data-processing window (other than 12) were also tried, but they produce similar result.

Figures 6.1 and 6.2 display the evolution of estimates of \mathbf{A} , $\boldsymbol{\mu}$, $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ under the 2-state WHMM setting; the values for estimates are in percentage. The plot in the top panel of Figure 6.1 shows the probabilities of staying in the same regime at the next time step. The plot in the bottom panel shows the probabilities of switching in the next step to a state different from the state in the previous step. The noticeable changes in the transition probabilities from

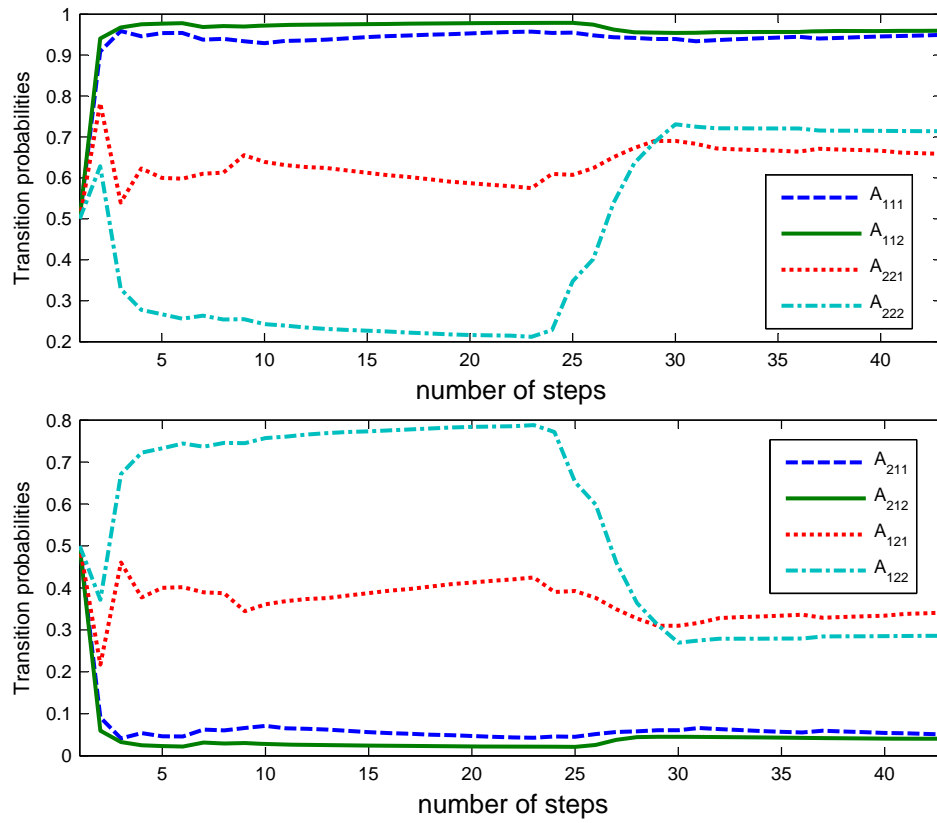


Figure 6.1: Evolution of estimates for transition probabilities under the 2-state setting

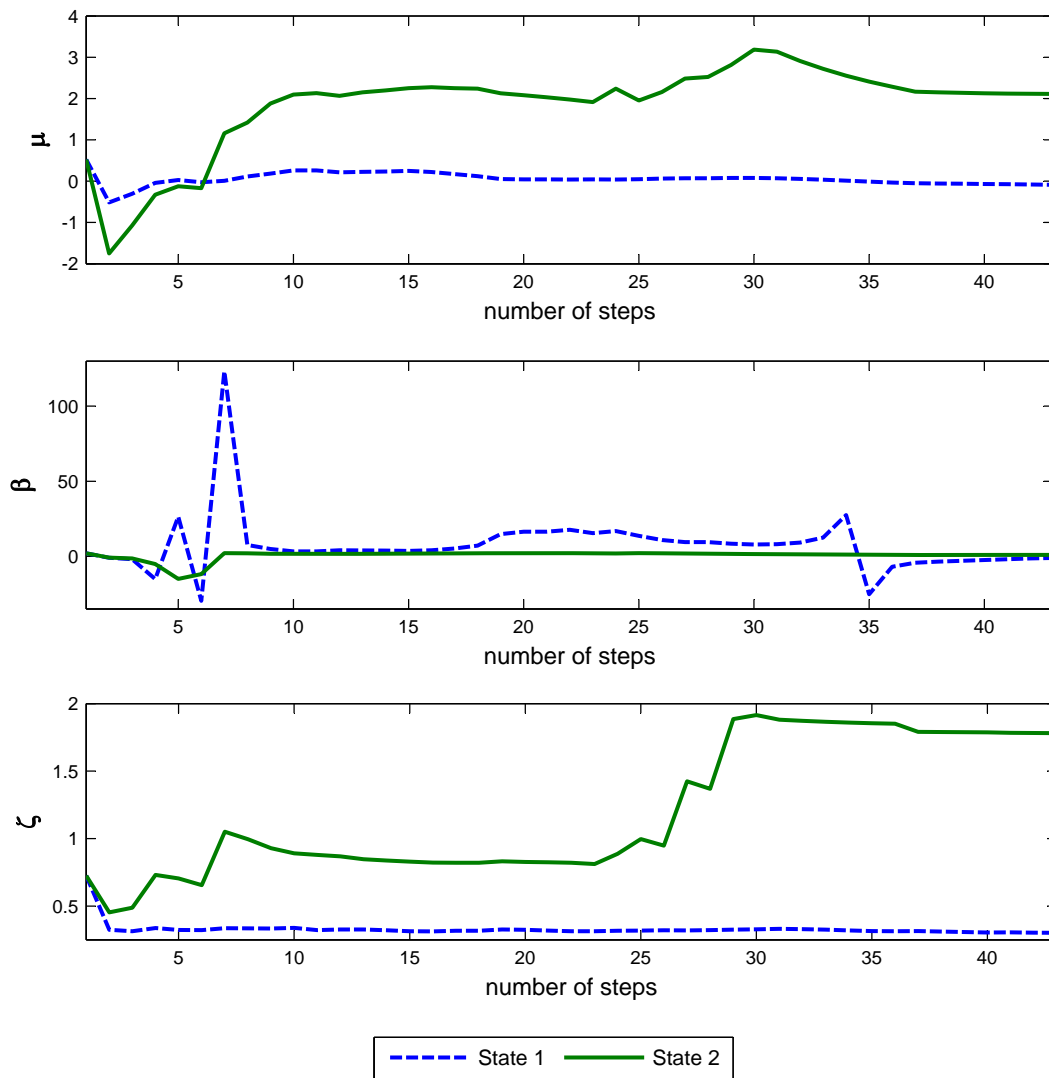


Figure 6.2: Evolution of parameter estimates under the 2-state setting

the 25th to the 30th algorithm step correspond to the drastic decline of the T-bill rate from 2.5% in June 2008 to 0.1% in September 2009, a period of economic meltdown brought about by the US subprime mortgage crisis. These significant market changes signal the occurrence of a regime switch in the market. The filtering algorithms are able to pick these up and they are reflected in the dynamics of the parameter estimates. We observe in Figure 6.2 that the movement of the optimal parameters through time still supports our preliminary analyses concerning the regime characteristics: state 1 has low mean-reverting speed, high-mean reverting level and low variance (until the 35th step); and the reverse is true for state 2. Figure 6.3 depicts the movement of μ , β and ζ through time under the three-state WHMM set up. We observe similar state characteristics for μ and ζ based on the dynamics of the parameters however, β does not follow the same evolution pattern as that in the 2-state setting. Note that under the three-state setting, the evolution of the parameters in states 1 and 2 is similar and the parameter values are close while the movement of parameters in state 3 is distinct. This suggests that a two-state WHMM is sufficient to capture the underlying market information. This is further backed up by a statistical-inference-based reasoning in the next section. It is worth mentioning that the parameters become stable after approximately 6 steps. Our experiment shows that this convergence can be achieved as long as the initial choice of parameter values are in a reasonable range consistent with equations (6.4)-(6.6). Apparently, the algorithms will not work for out-of-range or invalid initial values such as $\alpha_r < 0$ or $\sigma_r < 0$. Furthermore, just like any other algorithm the choice of initial values may affect the speed of convergence.

In order to measure the variability of parameter estimates, we derive the explicit formula of the Fisher information for each parameter. The Fisher information $I(\theta)$ is defined as the negative expectation of the second derivative of the log-density for a parameter θ . The inverse of the Fisher information is used to calculate the variance associated with the maximum-likelihood estimates. The sampling distribution of a maximum likelihood estimator is asymptotically normal and its variance can be calculated from $I^{-1}(\theta)$; see Garthwaite, et al. [17], for example.

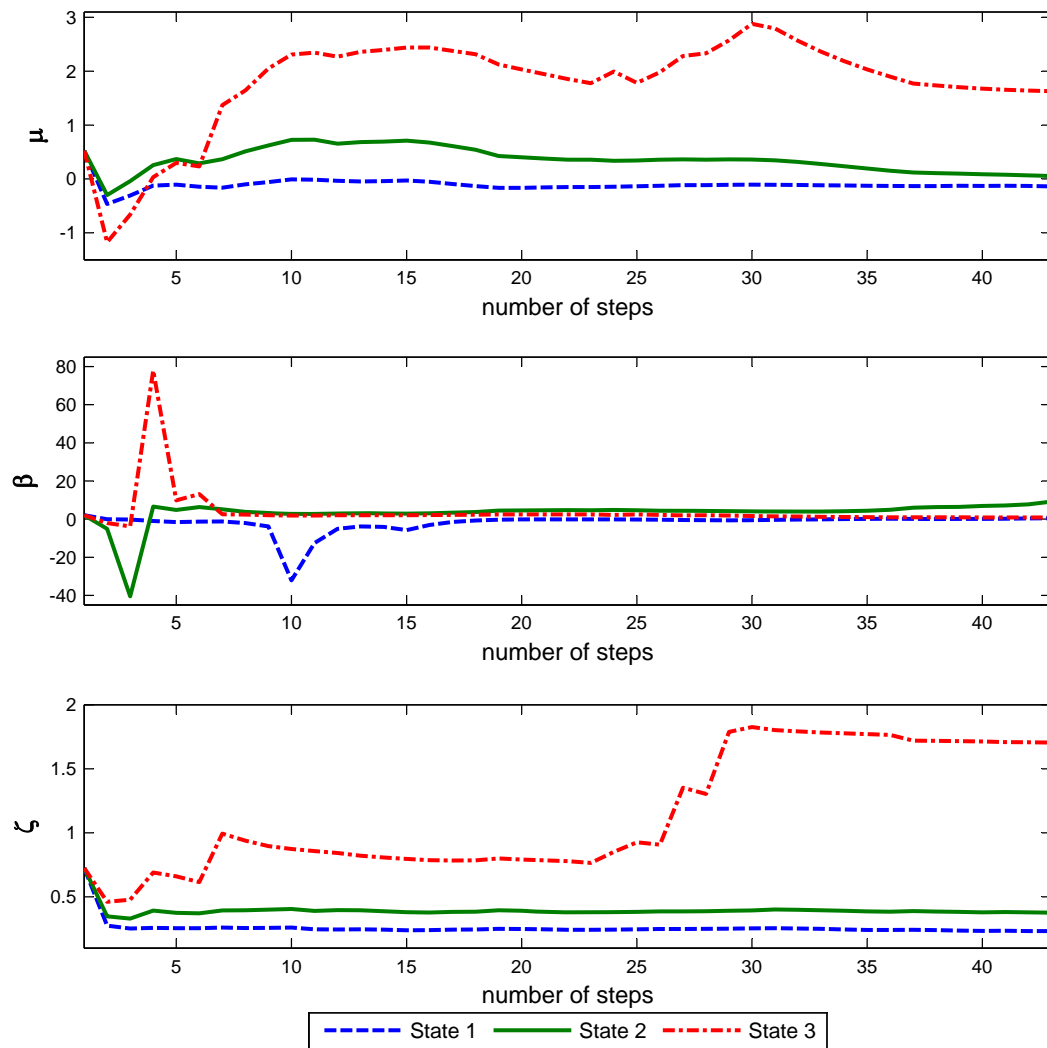


Figure 6.3: Evolution of parameter estimates under the 3-state setting

Parameter estimate	Range of standard errors	
	1-state	2-state
\hat{a}_{rst}	$[1.86 \times 10^{-62}, 1.91 \times 10^{-7}]$	$[3.54 \times 10^{-64}, 3.77 \times 10^{-7}]$
$\hat{\alpha}_r$	$[5.89 \times 10^{-66}, 3.03 \times 10^{-10}]$	$[1.71 \times 10^{-66}, 1.03 \times 10^{-9}]$
$\hat{\eta}_r$	$[5.00 \times 10^{-65}, 2.73 \times 10^{-9}]$	$[1.47 \times 10^{-65}, 8.00 \times 10^{-9}]$
$\hat{\sigma}_r$	$[6.00 \times 10^{-65}, 3.27 \times 10^{-9}]$	$[1.76 \times 10^{-65}, 9.61 \times 10^{-9}]$
	3-state	4-state
\hat{a}_{rst}	$[1.10 \times 10^{-63}, 2.00 \times 10^{-4}]$	$[8.32 \times 10^{-63}, 0.2896]$
$\hat{\alpha}_r$	$[1.52 \times 10^{-66}, 3.16 \times 10^{-7}]$	$[2.51 \times 10^{-66}, 4.27 \times 10^{-4}]$
$\hat{\eta}_r$	$[1.3 \times 10^{-65}, 2.55 \times 10^{-6}]$	$[2.14 \times 10^{65}, 0.0036]$
$\hat{\sigma}_r$	$[1.56 \times 10^{-65}, 3.06 \times 10^{-6}]$	$[2.57 \times 10^{-65}, 0.0043]$

Table 6.4: Range of SEs for each parameter under the 1-, 2-, 3- and 4-state settings

Following equations (C.1), (D.1), (D.2) and (D.3) for $1 \leq r, s, t \leq N$, the closed-form expressions for the Fisher information of each parameter are given by

$$I(a_{rst}) = a_{rst}^{-2} \hat{J}_k^{rst} \quad (6.28)$$

$$I(\alpha_r) = \sigma_r^{-2} \hat{T}_k^r(y_{k-1}^2) \quad (6.29)$$

$$I(\eta_r) = \sigma_r^{-2} \hat{O}_k^r \quad (6.30)$$

$$I(\sigma_r) = -\sigma_r^{-2} \hat{O}_k^r + 3\sigma_r^{-4} \left[\hat{T}_k^r(y_k^2) + \alpha_r^2 \hat{T}_k^r(y_{k-1}^2) + \eta_r^2 \hat{O}_k^r - 2\alpha_r \hat{T}_k^r(y_{k-1} y_k) - 2\eta_r \hat{T}_k^r(y_k) + 2\eta_r \alpha_r \hat{T}_k^r(y_{k-1}) \right]. \quad (6.31)$$

In Erlwein and Mamon [14], the inverse of Fisher information was utilized to compute the confidence interval for each parameter estimates. In our case, the $I^{-1}(\theta)$ estimates, which give the standard errors (SEs), turn out to be extremely small making the confidence intervals (CIs) very narrow. So, plotting the parameter estimates with the CIs is impractical. However, we provide the range of tabulated SEs over the entire algorithm steps for each parameter in Table 6.4 under the WHMM with $N = 1, \dots, 4$.

6.5 Forecasting and error analyses

In this section, we make use of the recursive parameter estimation technique for observations y_k to forecast yield values covering an h -week ahead horizon. The semi-martingale representation of \mathbf{x} in (6.9) leads to

$$E[\xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_k] = \mathbf{\Pi} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi} \mathbf{p}_k. \quad (6.32)$$

Recall that $\mathbf{\Pi}$ is constructed from \mathbf{A} defined by

$$a_{lmv} = P(\mathbf{x}_{k+1} = \mathbf{e}_l | \mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v),$$

so that equation (6.32) gives

$$E[\mathbf{x}_{k+h} | \mathcal{Y}_k] = \mathbf{A} \mathbf{\Pi}^{h-1} \mathbf{p}_k, \quad h \geq 1. \quad (6.33)$$

The one-step ahead forecasted yields of the T-bill rates are calculated by

$$\begin{aligned} E[y_{k+1} | \mathcal{Y}_k] &= E[\alpha(\mathbf{x}_k) y_k + \eta(\mathbf{x}_k) + \sigma(\mathbf{x}_k) z_{k+1} | \mathcal{Y}_k] \\ &= \langle \boldsymbol{\alpha}, \hat{\mathbf{x}}_k \rangle y_k + \langle \boldsymbol{\eta}, \hat{\mathbf{x}}_k \rangle. \end{aligned} \quad (6.34)$$

Similarly, the two-step forecasted yields are given by

$$\begin{aligned} E[y_{k+2} | \mathcal{Y}_k] &= E[\alpha(\mathbf{x}_{k+1}) y_{k+1} + \eta(\mathbf{x}_{k+1}) + \sigma(\mathbf{x}_{k+1}) z_{k+2} | \mathcal{Y}_k] \\ &= \langle \boldsymbol{\alpha}, \mathbf{A} \mathbf{p}_k \rangle \langle \boldsymbol{\alpha}, \hat{\mathbf{x}}_k \rangle y_k + \langle \boldsymbol{\eta}, \mathbf{A} \mathbf{p}_k \rangle. \end{aligned} \quad (6.35)$$

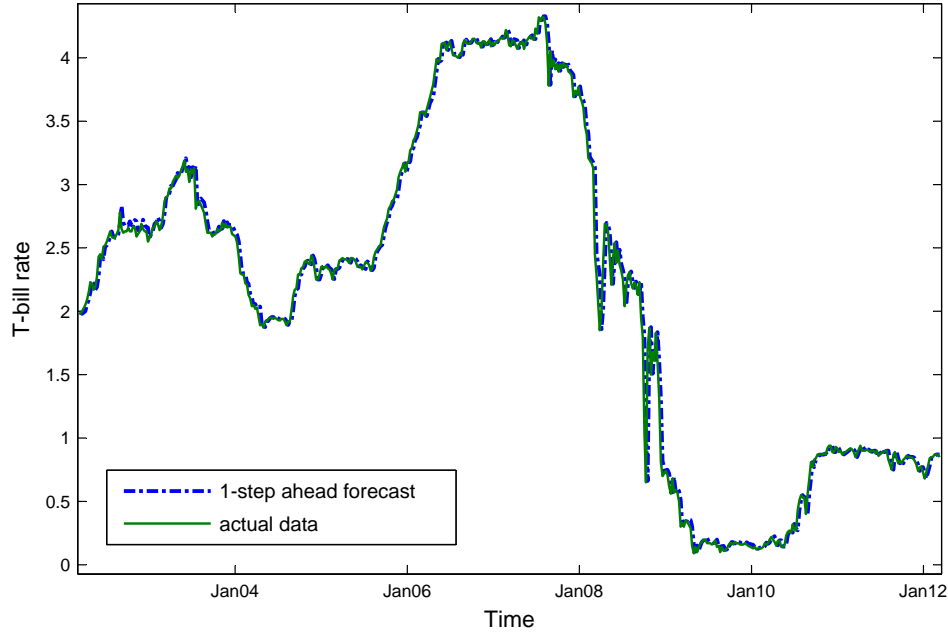


Figure 6.4: Plot of actual and one-step ahead forecasts in a 2-state WHMM

Following equations (6.33), (6.34) and (6.35), for any $h \geq 3$, the h -step ahead prediction of y_k is given by,

$$\begin{aligned}
 E[y_{k+h}|\mathcal{D}_k] &= \prod_{i=1}^{h-1} \langle \alpha, \mathbf{A}\Pi^{i-1}\mathbf{p}_k \rangle (\langle \alpha, \hat{\mathbf{x}}_k \rangle y_k + \langle \eta, \hat{\mathbf{x}}_k \rangle) \\
 &\quad + \sum_{i=1}^{h-2} \prod_{j=i}^{h-2} \langle \alpha, \mathbf{A}\Pi^j\mathbf{p}_k \rangle \langle \eta, \mathbf{A}\Pi^{i-1}\mathbf{p}_k \rangle + \langle \eta, \mathbf{A}\Pi^{h-2}\mathbf{p}_k \rangle.
 \end{aligned} \tag{6.36}$$

Proof See Appendix E for the proof of equation (6.36).

Figure 6.4 displays the one-step ahead forecasts for T-bill rates under a two-state WHMM setting. We observe that the resulting forecasts follow the actual data very closely. Empirical results confirm that our WHMM-based self-tuning algorithms capture the dynamics of the T-bill yields very well. The filtering and parameter estimation algorithms are also implemented with different number of states.

In Xi and Mamon [40], the forecasting performance of the WHMM is compared with that of the regular HMM using the S&P500 index dataset. They concluded that the WHMM outperforms

the HMM over a long forecasting horizon. We found that by including regime switching in our model, we obtain closer forecasts than those provided by the 1-state model on the basis of smaller forecasting errors. In recognition of the presence of long-range dependence revealed by the data, we aim to compare the predictability performance of the WHMM-based Hull-White model with that of the regular HMM-based model. In order to assess the goodness of fit of the h -step ahead forecasts, we evaluate the RMSE, AME, RAE and APE for the 1-, 2-, 3- and 4-state models. The RMSE, AME, RAE and APE for an h -step ahead prediction of y_k are calculated by

$$\text{RMSE}(h) = \sqrt{\frac{1}{M} \sum_{k=1}^M (y_{k+h} - \hat{y}_{k+h})^2}, \quad (6.37)$$

$$\text{AME}(h) = \frac{1}{M} \sum_{k=1}^M |y_{k+h} - \hat{y}_{k+h}|, \quad (6.38)$$

$$\text{RAE}(h) = \frac{\sum_{k=1}^M |y_{k+h} - \hat{y}_{k+h}|}{\sum_{k=1}^M |y_{k+h} - \bar{y}|}, \quad (6.39)$$

$$\text{APE}(h) = \frac{1}{M} \sum_{k=1}^M \left| \frac{y_{k+h} - \hat{y}_{k+h}}{y_{k+h}} \right|, \quad (6.40)$$

where M is the total number of forecasting points and \bar{y} is the mean of y_k over the forecasting period considered in the analysis. The WHMM and the regular HMM are compared using the four criteria in equations (6.37)–(6.40). The results of the error analyses are reported in Tables 6.5–6.7. The WHMM outperforms the regular HMM in terms of lower forecasting errors with respect to all four metrics. Note that the 4-state WHMM appears to have a slightly better fit than the 2-state WHMM over a short-time horizon. We perform a t -test of statistical significance for the mean difference of errors and the p -values are reported in Table 6.8. The p -values for comparing the 2-state and 3-state WHMM are large. Therefore we cannot reject the null hypothesis that the differences in means of the errors are equal. This is consistent with Figure 6.3 where states 1 and 2 almost coincide. Thus, a third regime to capture the data is not going to make a difference. We note that when comparing the 2- and 4-state WHMM, the

p -values are very small for the RMSE, RAE and APE metrics. It suggests that there is merit in considering a 4-state WHMM. Nonetheless, the benefit from such an increased number of states may be outweighed by the cost associated with the available resources such as computing time and memory, and may not be worth the further efforts. In other words, the error difference may be statistically significant but they may not necessarily be practically significant. In the third column of Table 6.8, we observe small p -values under the AME, RAE and APE measures. This implies that a 4-state WHMM has a better forecasting performance than a 3-state WHMM, which is in agreement with the fact that the 4-state WHMM is statistically different from the 2-state WHMM and both the 3-state and 2-state WHMMs have almost equal modeling capability when assessed under these fitting metrics.

A t -test is also carried out to evaluate whether the means of each set of forecasting errors under the WHMM are statistically different from those under the regular HMM. Table 6.9 exhibits the p -values for the one-tailed paired t -tests of significance for each pair of forecasting errors assuming unequal variance. The p -values for all metrics under the 2-, 3- and 4-state settings are very small. This tells us that the mean differences of the forecasting errors between the WHMM and HMM are all highly significantly. By including a mechanism to capture memory, the WHMM has a better fitting performance than the regular HMM for $N = 2, 3, \text{ and } 4$.

In order to find the most appropriate number of states inferred from the dataset given a WHMM or HMM set-up, we follow the approach used in Hardy [20] and Erlwein and Mamon [14]. We compare the AIC values for the 1-, 2- 3- and 4-state models. The AIC, motivated by the Kullback-Leiber information, is a function of the number of model parameters and the log-likelihood function of the model. The model selection criterion is calculated as

$$AIC = 2s - 2\mathcal{L}(\theta),$$

h	RMSE		AME		RAE		APE	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
1	0.1035	0.1047	0.0545	0.0556	0.0501	0.0511	0.0540	0.0552
2	0.1036	0.1048	0.0546	0.0556	0.0501	0.0510	0.0541	0.0552
3	0.1037	0.1048	0.0547	0.0556	0.0501	0.0509	0.0542	0.0553
4	0.1038	0.1048	0.0547	0.0555	0.0500	0.0508	0.0543	0.0554
5	0.1041	0.1051	0.0551	0.0558	0.0503	0.0509	0.0548	0.0558
6	0.1047	0.1056	0.0557	0.0563	0.0507	0.0513	0.0556	0.0565
7	0.1049	0.1056	0.0560	0.0564	0.0509	0.0513	0.0560	0.0569
8	0.1051	0.1058	0.0562	0.0565	0.0510	0.0513	0.0562	0.0570
9	0.1051	0.1058	0.0561	0.0564	0.0508	0.0511	0.0562	0.0569
10	0.1050	0.1056	0.0560	0.0562	0.0506	0.0508	0.0561	0.0568
11	0.1051	0.1055	0.0560	0.0560	0.0504	0.0505	0.0562	0.0568
12	0.1048	0.1049	0.0556	0.0555	0.0500	0.0499	0.0560	0.0566

Table 6.5: Error analysis for 2-state setting

h	RMSE		AME		RAE		APE	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
1	0.1036	0.1056	0.0546	0.0568	0.0502	0.0523	0.0544	0.0589
2	0.1037	0.1056	0.0547	0.0568	0.0502	0.0521	0.0545	0.0589
3	0.1038	0.1056	0.0548	0.0567	0.0502	0.0520	0.0547	0.0589
4	0.1039	0.1055	0.0548	0.0566	0.0501	0.0517	0.0547	0.0590
5	0.1042	0.1057	0.0552	0.0568	0.0504	0.0519	0.0553	0.0594
6	0.1048	0.1061	0.0558	0.0572	0.0508	0.0521	0.0560	0.0601
7	0.1050	0.1061	0.0561	0.0573	0.0510	0.0521	0.0564	0.0605
8	0.1052	0.1063	0.0563	0.0574	0.0511	0.0521	0.0567	0.0607
9	0.1052	0.1062	0.0562	0.0572	0.0508	0.0518	0.0566	0.0606
10	0.1051	0.1060	0.0561	0.0570	0.0507	0.0515	0.0565	0.0604
11	0.1052	0.1057	0.0561	0.0567	0.0506	0.0511	0.0566	0.0604
12	0.1049	0.1050	0.0557	0.0561	0.0501	0.0504	0.0564	0.0601

Table 6.6: Error analysis for 3-state setting

h ahead	RMSE		AME		RAE		APE	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
1	0.1034	0.1047	0.0540	0.0562	0.0497	0.0517	0.0524	0.0575
2	0.1035	0.1048	0.0541	0.0562	0.0497	0.0516	0.0525	0.0575
3	0.1036	0.1048	0.0542	0.0562	0.0497	0.0515	0.0526	0.0576
4	0.1037	0.1048	0.0542	0.0561	0.0496	0.0513	0.0527	0.0577
5	0.1041	0.1051	0.0547	0.0565	0.0499	0.0515	0.0532	0.0582
6	0.1047	0.1056	0.0553	0.0569	0.0504	0.0519	0.0540	0.0589
7	0.1049	0.1058	0.0556	0.0571	0.0505	0.0519	0.0544	0.0593
8	0.1051	0.1060	0.0558	0.0573	0.0506	0.0520	0.0546	0.0595
9	0.1050	0.1060	0.0556	0.0572	0.0503	0.0518	0.0545	0.0595
10	0.1049	0.1058	0.0556	0.0570	0.0502	0.0515	0.0544	0.0594
11	0.1051	0.1057	0.0555	0.0568	0.0501	0.0512	0.0545	0.0594
12	0.1048	0.1051	0.0552	0.0563	0.0496	0.0506	0.0543	0.0591

Table 6.7: Error analysis for 4-state setting

	2-state WHMM vs 3-state WHMM	2-state WHMM vs 4-state WHMM	3-state WHMM vs 4-state WHMM
RMSE	0.3662	8.67×10^{-4}	0.3266
AME	0.3469	0.0656	0.0312
RAE	0.2536	0.0077	0.0016
APE	0.1362	1.30×10^{-4}	8.21×10^{-6}

Table 6.8: p -values for a one-tailed significance test on the comparison of WHMM states based on forecasting errors shown in Tables 6.5-6.7

	RMSE WHMM vs HMM	AME WHMM vs HMM	RAE WHMM vs HMM	APE WHMM vs HMM
2-state	1.12×10^{-3}	1.48×10^{-2}	2.37×10^{-3}	9.03×10^{-3}
3-state	1.06×10^{-5}	4.65×10^{-6}	8.92×10^{-7}	4.44×10^{-11}
4-state	3.94×10^{-4}	4.80×10^{-7}	5.64×10^{-10}	1.69×10^{-12}

Table 6.9: p -values for a one-tailed significance test on the comparison of HMM and WHMM based on forecasting errors shown in Tables 6.5-6.7

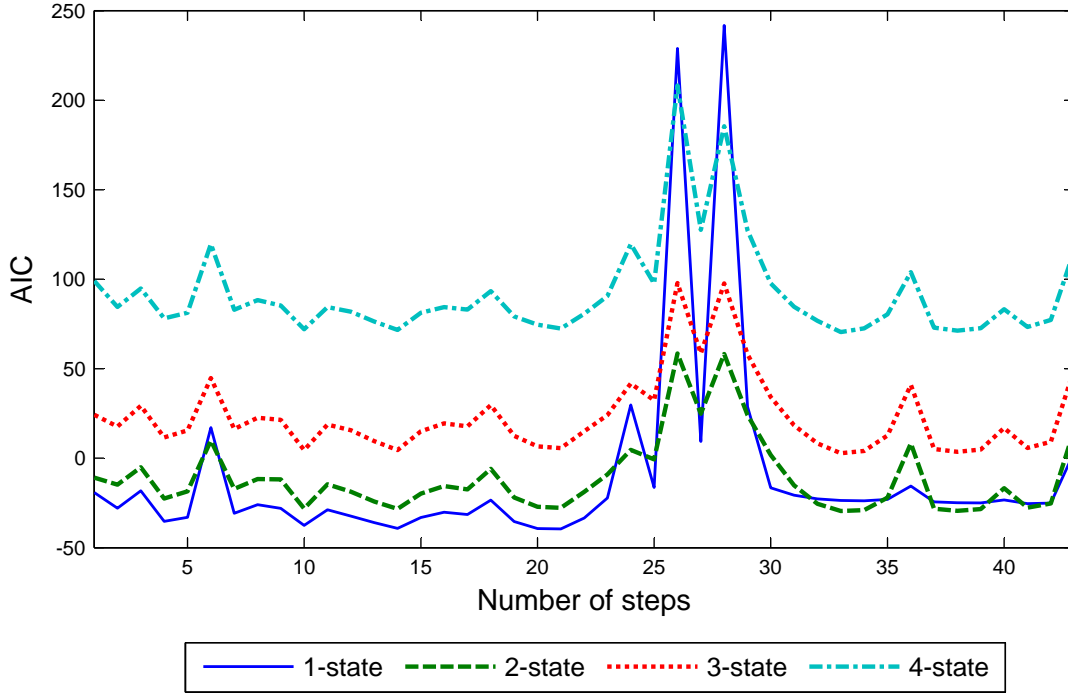


Figure 6.5: Plots of AIC values for 1-, 2-, 3- and 4-state WHMMs

where s is the number of parameters and $\mathcal{L}(\boldsymbol{\theta})$ denotes the log-likelihood function of the model given a vector of parameters $\boldsymbol{\theta}$. The preferred model has the minimum AIC value. For the observation process y_k in each pass, the log-likelihood of the vector of parameters $\boldsymbol{\theta}$ is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{l=1}^{\# \text{ in batch}} \sum_{r=1}^N \langle \mathbf{x}_l, \mathbf{e}_r \rangle \left(-\frac{1}{2} \log(2\pi\sigma^2(\mathbf{x}_l)) - \frac{(y_{l+1} - \alpha(\mathbf{x}_l)y_l - \eta(\mathbf{x}_l))^2}{2\sigma^2(\mathbf{x}_l)} \right).$$

The AIC for each model is evaluated using the parameter estimates given at the end of each algorithm step. This means that we obtain an AIC value after an algorithm run processing one batch of data points. The evolution of AIC values for the 1-, 2-, 3- and 4-state models after each algorithm step is presented in Figure 6.5. The AIC offers a relative measure of lost information described by the tradeoff between bias and variance in the model construction. In our model setting, there are $(N^2 - 1)N + 3N$ parameters needed to be estimated. A larger number of states increases the complexity of the parameter estimation and leads to a rapid increase in the AIC values. Although the 1-state model produces the smallest AIC values for majority of the steps,

it has the highest AIC values from the 25th to the 30th step. It is evident that there is merit in using a regime-switching model when market instability is expected to occur as the AIC values appear to be robust even in times of turbulent market conditions. While Table 6.8 shows that a 4-state WHMM has a fitting performance better than the 2- or 3-state WHMMs, the 2-state model is still reasonable in capturing the dynamics of our data and performs better than the 3- and 4-state models with respect to the AIC. Hence, the 2-state WHMM is adjudged as the most suitable for our dataset.

6.6 Conclusion

A regime-switching Hull-White model is developed in which the level and speed of mean reversion together with the volatility are governed by a second-order Markov chain in discrete time. The inclusion of the previous two time-steps in the Markov chain features the model's capacity to capture presence of memory in the data. By transforming a WHMM into a regular HMM, we were able to present general recursive filters. With the aid of the EM algorithm and change of reference probability measures, the model parameters are dynamically estimated. The proposed model is tested on a financial time series data of 30-day Canadian T-bill rates compiled during a 10-year period. The WHMM h -step ahead predictions are calculated and compared to those from the regular HMM under the 2-, 3- and 4-state settings. Our empirical results demonstrate that by utilizing a higher-order HMM, a better fit is obtained on the basis of forecasting errors way lower than those produced by the usual HMM. The choice of the most appropriate number of states was validated by the AIC analysis in conjunction with the goodness-of-fit exercise for the one-step ahead forecasts. We found that in the context of the dataset examined, the 2-state WHMM is sufficient to capture the interest rate dynamics. This result is consistent with the previous findings in Erlwein and Mamon [14] but we got a much improved fit given the smaller forecasting errors. The WHMM-based parameter estimation is also robust given the substantially low standard errors all throughout the entire algorithm steps. The WHMM-driven Hull-White filtering and parameter estimation technique that we

developed can be applied to a wide variety of processes exhibiting mean-reversion and presence of memory in finance, economics, engineering and other areas of the mathematical sciences. It may also be extended to a multivariate-data setting in a straightforward manner similar in idea to the extension for the non-mean-reverting regular HMM case carried out in Erlwein, et al. [15] although, of course, the computations will undoubtedly become more involved as the dimension of the data and lag dependence become bigger.

References

- [1] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Transaction on Information Theory*, 44(1):2–15, 1998.
- [2] F. Audrino and M. Medeiros. Modeling and forecasting shortterm interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics*, 26(6):999–1022, 2011.
- [3] J. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2004.
- [4] D. Cajueiro and B. Tabak. Long-range dependence and multifractality in the term structure of LIBOR interest rates. *Physica A: Statistical Mechanics and its Applications*, 373:603–614, 2007.
- [5] D. Cajueiro and B. Tabak. Time-varying long-range dependence in US interest rates. *Chaos, Solitons and Fractals*, 34:360–367, 2007.
- [6] D. Chapman, J. Long, and N. Pearson. Using proxies for the short rate: When are three months like an instant? *Review of Financial Studies*, 12:763–806, 1999.
- [7] W. Ching, T. Siu, and L. Li. Pricing exotic options under a higher-order hidden Markov model. *Journal of Applied Mathematics and Decision Sciences*, pages 1–15, 2007. Article ID 18014.

- [8] J. Cox, J. Ingersoll, and S. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.
- [9] S. Dajcman. Time-varying long-range dependence in stock market returns and financial market disruptions - a case of eight European countries. *Applied Economics Letters*, 19:953–957, 2012.
- [10] J. Duan and K. Jacobs. A simple long-memory equilibrium interest rate model. *Economics Letters*, 53(3):317–321, 1996.
- [11] R. Elliott, W. Hunter, and B. Jamieson. Financial signal processing: A self calibrating model. *International Journal of Theoretical and Applied Finance*, 4:567–584, 2001.
- [12] R. Elliott and R. Mamom. An interest rate model with a Markovian mean reverting level. *Quantitative Finance*, 2:454–458, 2002.
- [13] R. Elliott, J. Moore, and L. Aggoun. *Hidden Markov models: Estimation and Control*. Springer-Verlag, New York, 1995.
- [14] C. Erlwein and R. Mamon. An online estimation scheme for a Hull-White model with HMM-driven parameters. *Statistical Methods and Applications*, 18(1):87–107, 2009.
- [15] C. Erlwein, R. Mamon, and M. Davison. An examination of HMM-based investment strategies for asset allocation. *Applied Stochastic Models in Business and Industry*, 27:204–221, 2011.
- [16] D. Filipović. *Term Structure Models: A Graduate Course*. Springer, Heidelberg-London-New York, 2005.
- [17] P. Garthwaite, I. Jolliffe, and B. Jones. *Statistical Inference*. Oxford Science Publication, Oxford, 2nd edition, 2002.
- [18] M. Guidolin and A. Timmermann. Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, 150(2):297–311, 2009.

- [19] J. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–84, 1989.
- [20] M. Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53, 2001.
- [21] M. Holmes, R. Dutu, and X. Cui. Real interest rates, inflation and the open economy: A regime-switching perspective on Australia and New Zealand. *International Review of Economics and Finance*, 18(2):351–360, 2009.
- [22] J. Hull and A. White. Pricing interest-rate-derivative securities. *Review of Financial Studies*, 3(4):573–92, 1990.
- [23] J. Hunt and P. Devolder. Semi-Markov regime switching interest rate models and minimal entropy measure. *Physica A: Statistical Mechanics and its Applications*, 390:3767–3781, 2011.
- [24] H. Hurst. Long-term storage capacity of reservoirs. *American Society of Civil Engineers*, 116:770–808, 1951.
- [25] C. Landén. Bond pricing in a hidden Markov model of the short rate. *Finance and Stochastics*, 4(4):371–389, 2000.
- [26] J. Maheu. Can GARCH models capture long-range dependence? *Studies in Nonlinear Dynamics and Econometrics*, 9(4):1, 2005.
- [27] B. Mandelbrot. When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, 53(3):225–36, 1971.
- [28] T. Matteo. Multi-scaling in finance. *Quantitative Finance*, 7(1):21–36, 2007.
- [29] J. McCarthy, C. Pantalone, and H. Li. Investigating long memory in yield spreads. *Journal of Fixed Income*, 19(1):73–81, 2009.

- [30] L. Meligkotsidou and P. Dellaportas. Forecasting with non-homogeneous hidden Markov models. *Statistics and Computing*, 21:439–449, 2011.
- [31] C. Nieh, S. Wu, and Y. Zen. Regime shifts and the term structure of interest rates. In C. Lee and J. Lee, editors, *Handbook of Quantitative Finance and Risk Management*, pages 1121–1134. Springer, 2010.
- [32] T. Siu, W. Ching, E. Fung, and M. Ng. Extracting information from spot interest rates and credit ratings using double higher-order hidden Markov models. *Computational Economics*, 26:69–102, 2005.
- [33] T. Siu, W. Ching, E. Fung, M. Ng, and X. Li. A high-order Markov-switching model for risk measurement. *Computers and Mathematics with Applications*, 58:1–10, 2009.
- [34] D. Smith. Markov-switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economic Statistics*, 20(2):183–97, 2002.
- [35] J. Solberg. *Modelling Random Processes for Engineers and Managers*. Wiley & Sons, New Jersey, 2009.
- [36] R. Startz and K. Tsang. An unobserved components model of the yield curve. *Journal of Money, Credit and Banking*, 42(8):1613–1640, 2010.
- [37] M. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: An empirical study. *Fractals*, 3:785–798, 1995.
- [38] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.
- [39] C. Wu. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11:95–103, 1983.
- [40] X. Xi and R. Mamon. Parameter estimation of an asset price model driven by a weak hidden Markov chain. *Economic Modelling*, 28(1-2):36–46, 2011.

- [41] N. Zhou and R. Mamon. An accessible implementation of interest rate models with Markov-switching. *Expert Systems with Applications*, 39(5):4679–4689, 2012.

Chapter 7

Calibration of a regime-switching model using an inverse Stieltjes moment approach

7.1 Introduction

A central problem in finance that is somewhat different from the focus of chapters 2 to 6, which makes use of historical data, is the calibration of pricing models. Calibration means that we wish to obtain estimates of parameters given current market option prices. The recovery of model parameters given current observed market derivative prices is termed as an inverse problem in finance.

Volatility, for instance, is an important but unobservable parameter, whose estimate is necessary when pricing derivatives and enables us to understand price dynamics. Traders calculate implied volatilities from market data for option valuation as well as use them as a guide to monitor the market's sentiments. In the present work, we focus on recovering the parameters of a regime-switching model from European call option prices.

A number of approaches have been proposed to deal with this type of problem. In a pioneering paper, Dupire [12] verified empirically that different strikes and maturities lead to different implied volatilities for options on a given asset. Boyle and Thangaraj [6], as well as Andersen and Brotherton-Ratcliffe [1], obtained local implied volatilities by numerically implementing Dupire's equation. Rodrigo and Mamon [24] gave a new expression for the volatility by deriving a semi-explicit solution of Dupire's equation. They also provided a different formula in [23], which makes use of the so-called inverse Stieltjes moment approach. Bouchouev and Isakov [4] reduced the identification of the volatility to an inverse parabolic problem with the final observation. Deng, et al. [10] employed an optimal control framework with a new terminal condition to solve this kind of inverse problem.

Recently, considerable attention has been given to the use of regime-switching models, or HMMs, in finance. In an HMM, the model parameters switch among unobservable states of the economy and are governed by a Markov process. A regime-switching volatility is a simple way to incorporate stochastic volatilities. It has the ability to capture long-term and fundamental changes in the economic mechanism that generates the data. Significant empirical evidence from the literature lends support for the appropriateness of regime-switching models. For instance, Chu, et al. [8] advocated the use of these models to describe returns and volatility dynamics in the stock market. Turner, et al. [26] argued that either the mean or variance, or both, may exhibit differences between two regimes. The investigation of Engel and Hamilton [19], Bekaert and Hodrick [3], and Engel and Hakkio [18] documented regime switching in major foreign exchange rates. Dahlquist and Gray [9] and Ang and Bekaert [2] showed that various foreign, short-term interest rates are well described by regime-switching models. Some applications of regime-switching models modulated by a hidden Markov chain can be found in the work of Elliott and Mamon [16], as well as in Elliott and Kopp [15].

Regime-switching models have achieved growing importance in various financial problems as

they can capture a richer set of empirical and theoretical characteristics of a market. They have also enriched the developments in option pricing theory. For example, Elliott, et al. [14] developed a method to price options based on a regime-switching random Esscher transform. In turn, this method was used by Ching, et al. [7] to price exotic options under a hidden Markov model with long-range dependence in the states of an economy, which is known as a higher-order HMM. Mamon and Rodrigo [22] presented closed-form solutions for European option values when the dynamics of both the short rate and the volatility of the underlying price process are modulated by a continuous-time Markov chain. Boyle and Draviam [5] derived the system of partial differential equations (PDEs) of Black-Scholes type that governs the dynamics of European options in a regime-switching framework and price exotic options by solving the coupled PDEs numerically. Duan, et al. [11] developed a family of option pricing models which are based on the GARCH process and the variance-updating schemes also depend on a second factor orthogonal to asset innovations. Other works that feature regime-switching models in other applications include Siu, et al. [25] for credit default swaps, Elliott and van der Hoek [17] for asset allocations, and Elliott and Mamon [16] for short-term interest rates.

The above studies in option pricing under a regime-switching framework serve as motivation for investigating the inverse problem of recovering the volatilities when they are governed by HMMs. There is a relatively limited amount of literature on estimating regime-switching parameters using market data. In this chapter, we extend the inverse Stieltjes moment approach in [23] by assuming that the volatility of the underlying asset is governed by a continuous-time Markov chain. In this model, the unobservable parameters are the volatilities in each state and the intensity probabilities of the hidden Markov chain. We start with the well-known system of Black-Scholes-type PDEs and derive the coupled system of Dupire-type PDEs that governs the dynamics of European option prices.

The rest of the chapter is organized as follows. In section 7.2, we recall the regime-switching

model setup. In section 7.3, we derive the system of Dupire-type PDEs describing the dynamics of European option prices under this setup. We formulate the inverse Stieltjes moment problem in section 7.4, and also discuss how our proposed method could determine the model parameters. In section 7.5, we exhibit an implementation to a set of “theoretical data” which were generated by solving the coupled, Dupire-type PDEs. In section 7.6, numerical results for “practical data”, which is obtained from market data are presented. We conclude with a brief summary in section 7.7.

7.2 Regime-switching model setup

We wish to value a European option within the standard Black-Scholes market with two basic securities consisting of a riskless asset (a bond whose value is B_t at time $t \geq 0$) and a risky asset (a stock whose price is S_t at time t). Moreover, we assume that the economic state of the world is modeled by a finite-state Markov chain \mathbf{x}_t that evolves in continuous time. This implies that the bank rate process r_t and the stock’s volatility σ_t and rate of return μ_t are governed by Markov chain dynamics.

Without loss of generality, we may take the state space of \mathbf{x}_t to be the finite set $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ of canonical vectors in \mathbb{R}^N . Assume that \mathbf{x}_t is homogeneous in time and has intensity matrix $\mathbf{A} = (a_{ij})$, i.e.,

$$a_{ji} \geq 0 \quad \text{for } j \neq i, \quad \sum_{i=1}^N a_{ij} = 0 \quad \text{for each } j = 1, \dots, N.$$

If $\mathbf{p}_t = E[\mathbf{x}_t] = (p_t^1, \dots, p_t^N)^*$ where $*$ is the transpose operator, then \mathbf{p}_t satisfies

$$\frac{d\mathbf{p}_t}{dt} = \mathbf{A}\mathbf{p}_t.$$

It can be shown [13] that \mathbf{x}_t has a semi-martingale representation

$$\mathbf{x}_t = \mathbf{x}_0 + \int_0^t \mathbf{A}\mathbf{x}_u du + \mathbf{M}_t,$$

where \mathbf{M}_t is a martingale.

Suppose that $r_t = \langle \mathbf{r}, \mathbf{x}_t \rangle$ for some given vector $\mathbf{r} = (r_1, \dots, r_N)^*$ in \mathbb{R}^N with $r_1, \dots, r_N > 0$. Here, $\langle \cdot, \cdot \rangle$ denotes the usual inner product in \mathbb{R}^N . Then \$1 invested at time zero becomes

$$B_t = e^{\int_0^t r_u du} \quad (7.1)$$

at time t . In addition, suppose that the rate of return μ_t and the volatility σ_t depend on the state \mathbf{x}_t , i.e., there exist vectors $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)^*$ and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)^*$ in \mathbb{R}^N (with $\mu_i, \sigma_i > 0$ for all $i = 1, \dots, N$) such that $\mu_t = \langle \boldsymbol{\mu}, \mathbf{x}_t \rangle$ and $\sigma_t = \langle \boldsymbol{\sigma}, \mathbf{x}_t \rangle$. Then the dynamics of the stock are described by the stochastic differential equation

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t,$$

where W_t is a Brownian motion on a filtered probability space denoted by $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$ and $(\mathcal{F}_t)_{t \geq 0}$ is taken to be the natural filtration. W_t is independent of \mathbf{x}_t . It can be shown that S_t is expressible as

$$S_t = S_0 e^{\int_0^t (\mu_u - \sigma_u^2/2) du + \int_0^t \sigma_u dW_u}. \quad (7.2)$$

If the bond and stock dynamics are given by equations (7.1) and (7.2), respectively, and if at time $t \in [0, T]$ we have $S_t = S$ and $\mathbf{x}_t = \mathbf{x}$, then the price of a European call option with expiry T and strike price K is

$$c(t, S, T, K, \mathbf{x}) = E^Q \left[e^{-\int_t^T r_u du} (S_T - K)^+ \mid S_t = S, \mathbf{x}_t = \mathbf{x} \right] \quad (7.3)$$

where $(z)^+ = \max(z, 0)$ and E^Q denotes the expectation evaluated under a risk-neutral measure Q . We remark that regime switching leads to an incomplete market, which can be completed by the introduction of Arrow-Debreu securities [20] related to the cost of switching. Thus, in equation (7.3) we are assuming that we are already working under a risk-neutral mea-

sure Q . Just like in the classical Black-Scholes case, we assume that $\mu = r$ in the stock price dynamics under Q ; hence the rate of return will not appear in equation (7.3). We do not rule out the dependence of the market price of risk on the state \mathbf{x}_t at time t . But, irrespective of whether or not we assume a special or functional form for the market price of risk that depends on \mathbf{x}_t , or some other more general dependence which we do not know, the information from the market should be implicitly reflected in the parameters that we want to estimate. That is, we do not know what the exact dependence is but what is important to us are the parameter estimates that should encapsulate this information.

Define $c_i(t, S, T, K) = c(t, S, T, K, \mathbf{e}_i)$ for each $i = 1, \dots, N$. We note that $r_i = \langle \mathbf{r}, \mathbf{e}_i \rangle$ and $\sigma_i = \langle \boldsymbol{\sigma}, \mathbf{e}_i \rangle$. It can be shown in [22] that c_1, \dots, c_N satisfy a system of coupled PDEs of Black-Scholes type in the variables t and S , namely

$$\frac{\partial c_i}{\partial t} + \frac{1}{2} \sigma_i^2 S^2 \frac{\partial^2 c_i}{\partial S^2} + r_i S \frac{\partial c_i}{\partial S} - r_i c_i + \sum_{j=1}^N a_{ji} c_j = 0 \quad (i = 1, \dots, N), \quad (7.4)$$

together with the terminal conditions

$$c_i(T, S, T, K) = (S - K)^+ \quad (i = 1, \dots, N). \quad (7.5)$$

Let $\mathbf{c} = (c_1, \dots, c_N)^*$, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_N)$, and $\mathbf{R} = \text{diag}(r_1, \dots, r_N)$. Then equations (7.4) and (7.5) can be recast in matrix form as

$$\frac{\partial \mathbf{c}}{\partial t} + \frac{1}{2} S^2 \boldsymbol{\Sigma}^2 \frac{\partial^2 \mathbf{c}}{\partial S^2} + S \mathbf{R} \frac{\partial \mathbf{c}}{\partial S} - \mathbf{R} \mathbf{c} + \mathbf{A}^* \mathbf{c} = \mathbf{0}, \quad (7.6)$$

$$\mathbf{c}(T, S, T, K) = (S - K)^+ \mathbf{1}, \quad (7.7)$$

respectively, where $\mathbf{0}$ is the N -dimensional zero vector and $\mathbf{1}$ is the N -dimensional vector all of whose components are equal to one.

Our aim here is to solve the inverse problem of recovering the parameters of the underlying model from market data. The inverse problem was first considered by Dupire [12], who showed that if the prices of a European call option were known for all strike prices and maturity dates, then the volatility surface can be recovered from market data. In our case, instead of a local volatility function, we wish to recover the volatility matrix Σ , the transition intensity matrix \mathbf{A} , and the rate matrix \mathbf{R} .

It is important to note that actual market option prices are quoted for varying strikes and times to maturity. Thus, since we want to utilize a PDE-based approach to solve the inverse problem, we must first derive a system of PDEs similar to equation (7.6) but with the independent variables being the time to maturity and the strike price. In other words, we wish to obtain the analogue of Dupire's equation for the system of PDEs given in equation (7.6), which is the goal of the next section.

7.3 Derivation of a system of Dupire-type PDEs

First, we show that c_1, \dots, c_N are homogeneous functions of degree one with respect to S and K , i.e.,

$$c_i(t, \lambda S, T, \lambda K) = \lambda c_i(t, S, T, K) \quad (i = 1, \dots, N) \quad (7.8)$$

for all $\lambda > 0$. To prove equation (7.8), we will use a uniqueness argument by showing that $c_i(t, \lambda S, T, \lambda K)$ and $\lambda c_i(t, S, T, K)$ for all $i = 1, \dots, N$ satisfy the following final-value problem for $v(t, x, u, y)$:

$$\frac{\partial v_i}{\partial t} + \frac{1}{2} \sigma_i^2 x^2 \frac{\partial^2 v_i}{\partial x^2} + r_i x \frac{\partial v_i}{\partial x} - r_i v_i + \sum_{j=1}^N a_{ji} v_j = 0 \quad (i = 1, \dots, N), \quad (7.9)$$

$$v_i(T, x, u, y) = \lambda(x - y)^+ \quad (i = 1, \dots, N). \quad (7.10)$$

Let c_1, \dots, c_N be a solution of (7.4) and (7.5). Take $v_i(t, x, u, y) = \lambda c_i(t, S, T, K)$ where $x = S$,

$u = T$, and $y = K$. Then it is easy to see that v_1, \dots, v_N satisfy equations (7.9) and (7.10). Now take $v_i(t, x, u, y) = c_i(t, S, T, K)$ where $x = S/\lambda$, $u = T$, and $y = K/\lambda$. Again, it is straightforward to verify that v_1, \dots, v_N satisfy equations (7.9) and (7.10). Thus, the homogeneity condition (7.8) follows from the uniqueness of the solution of the final-value problem.

Invoking Euler's theorem on homogeneous functions, we obtain

$$S \frac{\partial c_i}{\partial S} + K \frac{\partial c_i}{\partial K} = c_i \quad (i = 1, \dots, N).$$

Differentiating the above equation with respect to S and K gives

$$S \frac{\partial^2 c_i}{\partial S^2} = -K \frac{\partial^2 c_i}{\partial S \partial K}, \quad K \frac{\partial^2 c_i}{\partial K^2} = -S \frac{\partial^2 c_i}{\partial K \partial S} \quad (i = 1, \dots, N),$$

respectively. It follows that

$$S^2 \frac{\partial^2 c_i}{\partial S^2} = K^2 \frac{\partial^2 c_i}{\partial K^2} \quad (i = 1, \dots, N)$$

and equation (7.4) becomes

$$\frac{\partial c_i}{\partial t} + \frac{1}{2} \sigma_i^2 K^2 \frac{\partial^2 c_i}{\partial K^2} - r_i K \frac{\partial c_i}{\partial K} + \sum_{j=1}^N a_{ji} c_j = 0 \quad (i = 1, \dots, N), \quad (7.11)$$

with the same terminal condition (7.5). In matrix form we therefore have

$$\frac{\partial \mathbf{c}}{\partial t} + \frac{1}{2} K^2 \Sigma^2 \frac{\partial^2 \mathbf{c}}{\partial K^2} - K \mathbf{R} \frac{\partial \mathbf{c}}{\partial K} + \mathbf{A}^* \mathbf{c} = \mathbf{0}, \quad (7.12)$$

$$\mathbf{c}(T, S, T, K) = (S - K)^+ \mathbf{1}. \quad (7.13)$$

Finally, letting $\mathbf{c}(t, S, T, K) = \bar{\mathbf{c}}(u, S, K)$ where $u = T - t$ in equations (7.12), (7.13) yields

$$\frac{\partial \bar{\mathbf{c}}}{\partial u} = \frac{1}{2} K^2 \Sigma^2 \frac{\partial^2 \bar{\mathbf{c}}}{\partial K^2} - K \mathbf{R} \frac{\partial \bar{\mathbf{c}}}{\partial K} + \mathbf{A}^* \bar{\mathbf{c}}, \quad (7.14)$$

$$\bar{\mathbf{c}}(0, S, K) = (S - K)^+ \mathbf{1}, \quad (7.15)$$

respectively. Note that equations (7.14) and (7.15) is now an initial-value problem. When $N = 1$, we have

$$\frac{\partial \bar{c}_1}{\partial u} = \frac{1}{2} \sigma_1^2 K^2 \frac{\partial^2 \bar{c}_1}{\partial K^2} - r_1 K \frac{\partial \bar{c}_1}{\partial K}, \quad (7.16)$$

which is Dupire's equation with a constant volatility [12]. Hence, equation (7.14) is the analogue of Dupire's equation for the regime-switching case.

7.4 Inverse Stieltjes moment problem

Having derived the system of PDEs in the appropriate independent variables u and K , we now proceed to solve the inverse problem of parameter estimation via the inverse Stieltjes moment method first proposed in Rodrigo and Mamon [23].

To simplify the ensuing notation, we shall write $\mathbf{c}(u, K)$ instead of $\bar{\mathbf{c}}(u, S, K)$. We will also assume for simplicity that $\mathbf{R} = r\mathbf{I}$ for some given $r > 0$, where \mathbf{I} is the identity matrix; the method can be easily extended to the more general case. As in Rodrigo and Mamon [23], let us define the n th moment of the call price by

$$m_n^{(i)}(u) = \int_0^\infty K^n c_i(u, K) dK \quad (i = 1 \dots N),$$

where n is a nonnegative integer. Multiplying both sides of equation (7.14) by K^n and integrating over $(0, \infty)$, we formally obtain

$$\begin{aligned} \int_0^\infty K^n \frac{\partial \mathbf{c}}{\partial u} dK &= \frac{1}{2} \int_0^\infty K^{n+2} \Sigma^2 \frac{\partial^2 \mathbf{c}}{\partial K^2} dK - \int_0^\infty r K^{n+1} \frac{\partial \mathbf{c}}{\partial K} dK \\ &+ \int_0^\infty K^n \mathbf{A}^* \mathbf{c} dK. \end{aligned} \quad (7.17)$$

Assuming that the call price decays to zero sufficiently fast as $K \rightarrow \infty$, we deduce that

$$\begin{aligned}\int_0^\infty K^n \frac{\partial \mathbf{c}}{\partial u} dK &= \frac{d\mathbf{m}_n}{du}, \\ \int_0^\infty K^{n+1} \frac{\partial \mathbf{c}}{\partial K} dK &= -(n+1)\mathbf{m}_n, \\ \int_0^\infty K^{n+2} \frac{\partial^2 \mathbf{c}}{\partial K^2} dK &= (n+1)(n+2)\mathbf{m}_n.\end{aligned}$$

Thus, equation (7.17) simplifies to

$$\frac{d\mathbf{m}_n}{du} - r(n+1)\mathbf{m}_n = \left[\frac{1}{2}(n+1)(n+2)\mathbf{\Sigma}^2 + \mathbf{A}^* \right] \mathbf{m}_n.$$

Considering any N consecutive moments gives

$$\begin{aligned}\frac{d\mathbf{m}_n}{du} - r(n+1)\mathbf{m}_n &= \left[\frac{(n+1)(n+2)}{2}\mathbf{\Sigma}^2 + \mathbf{A}^* \right] \mathbf{m}_n, \\ \frac{d\mathbf{m}_{n+1}}{du} - r(n+2)\mathbf{m}_{n+1} &= \left[\frac{(n+2)(n+3)}{2}\mathbf{\Sigma}^2 + \mathbf{A}^* \right] \mathbf{m}_{n+1}, \\ &\vdots \\ \frac{d\mathbf{m}_{n+N-1}}{du} - r(n+N)\mathbf{m}_{n+N-1} &= \left[\frac{(n+N)(n+N+1)}{2}\mathbf{\Sigma}^2 + \mathbf{A}^* \right] \mathbf{m}_{n+N-1}.\end{aligned}\tag{7.18}$$

Since the option prices are assumed to be observed, the moments are also known and we wish to estimate \mathbf{A} and $\mathbf{\Sigma}$. Given that \mathbf{A} is an intensity matrix, each of its N diagonal entries, say a_{ii} , is expressible as a sum of the entries in the i th column, so there are essentially $N^2 - N$ unknown entries of \mathbf{A} . Together with the N unknown diagonal entries of $\mathbf{\Sigma}$, we therefore have a total of N^2 parameters to estimate. Note that (7.18) is a linear system of N^2 equations in N^2 unknowns.

To explain the basic idea, let us take $N = 2$. Then (7.18) gives

$$\begin{pmatrix} \frac{(n+1)(n+2)}{2}m_n^{(1)} & 0 & m_n^{(2)} - m_n^{(1)} & 0 \\ 0 & \frac{(n+1)(n+2)}{2}m_n^{(2)} & 0 & m_n^{(1)} - m_n^{(2)} \\ \frac{(n+2)(n+3)}{2}m_{n+1}^{(1)} & 0 & m_{n+1}^{(2)} - m_{n+1}^{(1)} & 0 \\ 0 & \frac{(n+2)(n+3)}{2}m_{n+1}^{(2)} & 0 & m_{n+1}^{(1)} - m_{n+1}^{(2)} \end{pmatrix} \times \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ a_{21} \\ a_{12} \end{pmatrix} = \begin{pmatrix} \frac{dm_n^{(1)}}{du} - r(n+1)m_n^{(1)} \\ \frac{dm_n^{(2)}}{du} - r(n+1)m_n^{(2)} \\ \frac{dm_{n+1}^{(1)}}{du} - r(n+2)m_{n+1}^{(1)} \\ \frac{dm_{n+1}^{(2)}}{du} - r(n+2)m_{n+1}^{(2)} \end{pmatrix}. \quad (7.19)$$

Note that from (7.15) we see that

$$\mathbf{m}_n(0) = \int_0^\infty K^n (S - K)^+ \mathbf{1} dK = \frac{S^{n+2}}{(n+1)(n+2)} \mathbf{1}.$$

Let

$$M_n^{(i)}(u) = \int_0^u m_n^{(i)}(s) ds, \quad (i = 1, 2).$$

To incorporate the initial conditions, we integrate (7.19) over $[0, u]$ with respect to a dummy variable s to get

$$\begin{pmatrix} \frac{(n+1)(n+2)}{2}M_n^{(1)} & 0 & M_n^{(2)} - M_n^{(1)} & 0 \\ 0 & \frac{(n+1)(n+2)}{2}M_n^{(2)} & 0 & M_n^{(1)} - M_n^{(2)} \\ \frac{(n+2)(n+3)}{2}M_{n+1}^{(1)} & 0 & M_{n+1}^{(2)} - M_{n+1}^{(1)} & 0 \\ 0 & \frac{(n+2)(n+3)}{2}M_{n+1}^{(2)} & 0 & M_{n+1}^{(1)} - M_{n+1}^{(2)} \end{pmatrix} \times \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ a_{21} \\ a_{12} \end{pmatrix} = \begin{pmatrix} m_n^{(1)} - \frac{S^{n+2}}{(n+1)(n+2)} - r(n+1)M_n^{(1)} \\ m_n^{(2)} - \frac{S^{n+2}}{(n+1)(n+2)} - r(n+1)M_n^{(2)} \\ m_{n+1}^{(1)} - \frac{S^{n+3}}{(n+2)(n+3)} - r(n+2)M_{n+1}^{(1)} \\ m_{n+1}^{(2)} - \frac{S^{n+3}}{(n+2)(n+3)} - r(n+2)M_{n+1}^{(2)} \end{pmatrix}. \quad (7.20)$$

In summary, given $c_1(u, K)$ and $c_2(u, K)$ where $0 \leq u \leq T$ and $K \geq 0$, we compute $m_n^{(1)}$, $m_n^{(2)}$, $m_{n+1}^{(1)}$, and $m_{n+1}^{(2)}$ (for a fixed nonnegative integer n), as well as their integrals over $[0, u]$ for some u in $[0, T]$. We then solve the linear system (7.20) for the unknown parameters σ_1 , σ_2 , a_{21} , and a_{12} . Note that $a_{11} = -a_{21}$ and $a_{22} = -a_{12}$ by the definition of \mathbf{A} . In addition, the choice of u in $[0, T]$ should not matter since in this framework \mathbf{A} and $\mathbf{\Sigma}$ are constant matrices.

7.5 Numerical implementation and results

To test the accuracy of the inverse Stieltjes moment method, we need to have “observed” option prices. The “observed” option prices can be taken to be the solution generated by the initial-value problem (7.14), (7.15) (after specifying some matrices \mathbf{A} and $\mathbf{\Sigma}$). Then we try to recover \mathbf{A} and $\mathbf{\Sigma}$ using the moment method.

As a trial run, suppose that $\mathbf{A} = \mathbf{0}$, i.e., there is no switching among regimes. Then equation (7.14) reduces to a system of uncoupled Dupire equations. Hence, each component of \mathbf{c} solves Dupire’s equation, i.e., if $\mathbf{c} = (c_1, \dots, c_N)^*$, then

$$\begin{aligned} c_i(u, K) &= S \Phi(d_1^{(i)}(u, K)) - K e^{-ru} \Phi(d_2^{(i)}(u, K)) \quad (i = 1, \dots, N), \\ d_1^{(i)}(u, K) &= \frac{\log(S/K) + (r + \sigma_i^2/2)u}{\sigma_i \sqrt{u}}, \\ d_2^{(i)}(u, K) &= \frac{\log(S/K) + (r - \sigma_i^2/2)u}{\sigma_i \sqrt{u}}, \end{aligned} \quad (7.21)$$

where Φ denotes the cumulative distribution function of a standard normal variable. Now we assume that $r = 0.02$, $S = 20$, $u = 0$, $T = 1$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.3$. We take 50 values for strike price ranging from 0 to 60 and 100 values for the time to maturity ranging from 0 to 1. Then the moments and their integrals are calculated and moment method in (7.20) is applied. The estimated parameters are: $a_{21} = 6.67 \times 10^{(-4)}$, $a_{12} = 3.15 \times 10^{(-4)}$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2957$.

However, for the more realistic case $\mathbf{A} \neq \mathbf{0}$, there is no known explicit solution of (7.14)-(7.15) in general. So we will have to solve this problem numerically to generate the “observed” option prices. This implies that we have to truncate the interval $(0, \infty)$ to some finite interval $(0, K_{\max})$ where $K_{\max} > 0$, and then impose reasonable boundary conditions for \mathbf{c} at $K = 0$ and $K = K_{\max}$. If the right endpoint K_{\max} is sufficiently large, and recalling that each component of \mathbf{c} tends to zero as $K \rightarrow \infty$, then we can assign a positive but small value to each component. However, the boundary condition at the left endpoint $K = 0$ is not clear. When $N = 1$, the Black-Scholes formula evaluated at $K = 0$ gives S for the call price. For $N > 1$, it is not certain whether each component of \mathbf{c} will also have the value S . To get around this problem, we will solve (7.14)-(7.15) numerically for $K \in [0, K_{\max}]$ and $u \in [0, T]$ by formulating an explicit method with implicit boundary conditions.

Discretize the variables by

$$u \simeq u^i, \quad K \simeq K^j, \quad \mathbf{c} \simeq \mathbf{c}^{i,j} = (c_1^{i,j}, c_2^{i,j})^*,$$

where

$$u^i = i\Delta u, \quad \Delta u = \frac{T}{I} \quad (i = 0, \dots, I)$$

and

$$K^j = j\Delta K, \quad \Delta K = \frac{K_{\max}}{J} \quad (j = 0, \dots, J).$$

Using an explicit scheme, we discretize equation (7.14) to get

$$\begin{aligned} \frac{c_1^{i+1,j} - c_1^{i,j}}{\Delta u} &= \frac{1}{2}\sigma_1^2(K^j)^2 \frac{c_1^{i,j-1} - 2c_1^{i,j} + c_1^{i,j+1}}{(\Delta K)^2} - rK^j \frac{c_1^{i,j+1} - c_1^{i,j}}{\Delta K} \\ &\quad + a_{11}c_1^{i,j} + a_{21}c_2^{i,j}, \\ \frac{c_2^{i+1,j} - c_2^{i,j}}{\Delta u} &= \frac{1}{2}\sigma_2^2(K^j)^2 \frac{c_2^{i,j-1} - 2c_2^{i,j} + c_2^{i,j+1}}{(\Delta K)^2} - rK^j \frac{c_2^{i,j+1} - c_2^{i,j}}{\Delta K} \\ &\quad + a_{12}c_1^{i,j} + a_{22}c_2^{i,j}. \end{aligned}$$

This is equivalent to

$$\begin{aligned} \mathbf{c}^{i+1,j} = & \mathbf{c}^{i,j} + \frac{1}{2} \frac{\Delta u}{(\Delta K)^2} (K^j)^2 \Sigma^2 (\mathbf{c}^{i,j-1} - 2\mathbf{c}^{i,j} + \mathbf{c}^{i,j+1}) \\ & - \frac{r\Delta u}{\Delta K} K^j (\mathbf{c}^{i,j+1} - \mathbf{c}^{i,j}) + \Delta u \mathbf{A}^* \mathbf{c}^{i,j}, \end{aligned}$$

for all $i = 1, \dots, I-1$ and $j = 1, \dots, J-1$. This solves for the option prices at time $i+1$ in the open interval $(0, K_{\max})$ using the option prices calculated at time i over the closed interval $[0, K_{\max}]$. The initial condition is determined by

$$\mathbf{c}^{0,j} = (S - K^j)^+ \mathbf{1} \quad (j = 0, \dots, J),$$

which includes the values at both endpoints.

To determine the boundary values at $K = 0$, we use second-order Taylor expansions at $(u, 0)$ in the continuous variables, i.e.,

$$\begin{aligned} \mathbf{c}(u, \Delta K) & \simeq \mathbf{c}(u, 0) + (\Delta K) \frac{\partial \mathbf{c}}{\partial K}(u, 0) + \frac{1}{2} (\Delta K)^2 \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, 0), \\ \mathbf{c}(u, 2\Delta K) & \simeq \mathbf{c}(u, 0) + (2\Delta K) \frac{\partial \mathbf{c}}{\partial K}(u, 0) + \frac{1}{2} (2\Delta K)^2 \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, 0), \\ \mathbf{c}(u, 3\Delta K) & \simeq \mathbf{c}(u, 0) + (3\Delta K) \frac{\partial \mathbf{c}}{\partial K}(u, 0) + \frac{1}{2} (3\Delta K)^2 \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, 0). \end{aligned}$$

Define

$$\alpha_L(u) = \frac{\partial \mathbf{c}}{\partial K}(u, 0), \quad \beta_L(u) = \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, 0).$$

Then in discretized variables we get

$$\begin{aligned} \mathbf{c}^{i,1} & = \mathbf{c}^{i,0} + \Delta K \alpha_L(u^i) + \frac{1}{2} (\Delta K)^2 \beta_L(u^i), \\ \mathbf{c}^{i,2} & = \mathbf{c}^{i,0} + 2\Delta K \alpha_L(u^i) + 2(\Delta K)^2 \beta_L(u^i), \\ \mathbf{c}^{i,3} & = \mathbf{c}^{i,0} + 3\Delta K \alpha_L(u^i) + \frac{9}{2} (\Delta K)^2 \beta_L(u^i), \end{aligned}$$

valid for all $i = 1, \dots, I$. In matrix form, this is the same as

$$\begin{pmatrix} 1 & \Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & 2\Delta K & 2(\Delta K)^2 \\ 1 & 3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_1^{i,0} \\ \alpha_L^{(1)}(u^i) \\ \beta_L^{(1)}(u^i) \end{pmatrix} = \begin{pmatrix} c_1^{i,1} \\ c_1^{i,2} \\ c_1^{i,3} \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & \Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & 2\Delta K & 2(\Delta K)^2 \\ 1 & 3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_2^{i,0} \\ \alpha_L^{(2)}(u^i) \\ \beta_L^{(2)}(u^i) \end{pmatrix} = \begin{pmatrix} c_2^{i,1} \\ c_2^{i,2} \\ c_2^{i,3} \end{pmatrix}.$$

Note that the vectors $\alpha_L(u^i)$ and $\beta_L(u^i)$ are not known, which is why we need to solve these two linear systems to obtain the really desired quantity $\mathbf{c}^{i,0}$. Moreover, $\mathbf{c}^{i,1}$, $\mathbf{c}^{i,2}$, and $\mathbf{c}^{i,3}$ on the right-hand sides are known if we first solve the PDEs in the interior $(0, K_{\max})$.

Similarly, at the right boundary point $K = K_{\max}$, we expand

$$\begin{aligned} \mathbf{c}(u, K_{\max} - \Delta K) &\simeq \mathbf{c}(u, K_{\max}) + (-\Delta K) \frac{\partial \mathbf{c}}{\partial K}(u, K_{\max}) \\ &\quad + \frac{1}{2}(-\Delta K)^2 \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, K_{\max}), \\ \mathbf{c}(u, K_{\max} - 2\Delta K) &\simeq \mathbf{c}(u, K_{\max}) + (-2\Delta K) \frac{\partial \mathbf{c}}{\partial K}(u, K_{\max}) \\ &\quad + \frac{1}{2}(-2\Delta K)^2 \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, K_{\max}), \\ \mathbf{c}(u, K_{\max} - 3\Delta K) &\simeq \mathbf{c}(u, K_{\max}) + (-3\Delta K) \frac{\partial \mathbf{c}}{\partial K}(u, K_{\max}) \\ &\quad + \frac{1}{2}(-3\Delta K)^2 \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, K_{\max}). \end{aligned}$$

Defining

$$\alpha_R(u) = \frac{\partial \mathbf{c}}{\partial K}(u, K_{\max}), \quad \beta_R(u) = \frac{\partial^2 \mathbf{c}}{\partial K^2}(u, K_{\max}),$$

we obtain

$$\begin{aligned}\mathbf{c}^{i,J-1} &= \mathbf{c}^{i,J} - \Delta K \alpha_R(u^i) + \frac{1}{2}(\Delta K)^2 \beta_R(u^i), \\ \mathbf{c}^{i,J-2} &= \mathbf{c}^{i,J} - 2\Delta K \alpha_R(u^i) + 2(\Delta K)^2 \beta_R(u^i), \\ \mathbf{c}^{i,J-3} &= \mathbf{c}^{i,J} - 3\Delta K \alpha_R(u^i) + \frac{9}{2}(\Delta K)^2 \beta_R(u^i),\end{aligned}$$

or, in matrix form,

$$\begin{pmatrix} 1 & -\Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & -2\Delta K & 2(\Delta K)^2 \\ 1 & -3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_1^{i,J} \\ \alpha_R^{(1)}(u^i) \\ \beta_R^{(1)}(u^i) \end{pmatrix} = \begin{pmatrix} c_1^{i,J-1} \\ c_1^{i,J-2} \\ c_1^{i,J-3} \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & -\Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & -2\Delta K & 2(\Delta K)^2 \\ 1 & -3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_2^{i,J} \\ \alpha_R^{(2)}(u^i) \\ \beta_R^{(2)}(u^i) \end{pmatrix} = \begin{pmatrix} c_2^{i,J-1} \\ c_2^{i,J-2} \\ c_2^{i,J-3} \end{pmatrix}.$$

As before, the vectors $\alpha_R(u^i)$ and $\beta_R(u^i)$ are not known, so we solve these two linear systems to obtain the really desired quantity $\mathbf{c}^{i,J}$. Moreover, $\mathbf{c}^{i,J-1}$, $\mathbf{c}^{i,J-2}$, and $\mathbf{c}^{i,J-3}$ on the right-hand sides are known if we first solve the PDEs in the interior $(0, K_{\max})$.

Summarizing, the explicit algorithm incorporating implicit boundary conditions that we propose can be formulated as follows:

1. Set

$$\mathbf{c}^{0,j} = (S - K^j)^+ \mathbf{1} \quad (j = 0, \dots, J).$$

2. For all $i = 0, \dots, I - 1$ do

(a) For all $j = 1, \dots, J - 1$ do

$$\begin{aligned} \mathbf{c}^{i+1,j} = & \mathbf{c}^{i,j} + \frac{1}{2} \frac{\Delta u}{(\Delta K)^2} (K^j)^2 \Sigma^2 (\mathbf{c}^{i,j-1} - 2\mathbf{c}^{i,j} + \mathbf{c}^{i,j+1}) \\ & - \frac{r\Delta u}{\Delta K} K^j (\mathbf{c}^{i,j+1} - \mathbf{c}^{i,j}) + \Delta u \mathbf{A}^* \mathbf{c}^{i,j}. \end{aligned}$$

(b) Solve

$$\begin{pmatrix} 1 & \Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & 2\Delta K & 2(\Delta K)^2 \\ 1 & 3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_1^{i+1,0} \\ \alpha_L^{(1)}(u^{i+1}) \\ \beta_L^{(1)}(u^{i+1}) \end{pmatrix} = \begin{pmatrix} c_1^{i+1,1} \\ c_1^{i+1,2} \\ c_1^{i+1,3} \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & \Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & 2\Delta K & 2(\Delta K)^2 \\ 1 & 3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_2^{i+1,0} \\ \alpha_L^{(2)}(u^{i+1}) \\ \beta_L^{(2)}(u^{i+1}) \end{pmatrix} = \begin{pmatrix} c_2^{i+1,1} \\ c_2^{i+1,2} \\ c_2^{i+1,3} \end{pmatrix}$$

for $\mathbf{c}^{i+1,0} = (c_1^{i+1,0}, c_2^{i+1,0})^*$.

(c) Solve

$$\begin{pmatrix} 1 & -\Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & -2\Delta K & 2(\Delta K)^2 \\ 1 & -3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_1^{i+1,J} \\ \alpha_R^{(1)}(u^{i+1}) \\ \beta_R^{(1)}(u^{i+1}) \end{pmatrix} = \begin{pmatrix} c_1^{i+1,J-1} \\ c_1^{i+1,J-2} \\ c_1^{i+1,J-3} \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & -\Delta K & \frac{1}{2}(\Delta K)^2 \\ 1 & -2\Delta K & 2(\Delta K)^2 \\ 1 & -3\Delta K & \frac{9}{2}(\Delta K)^2 \end{pmatrix} \begin{pmatrix} c_2^{i+1,J} \\ \alpha_R^{(2)}(u^{i+1}) \\ \beta_R^{(2)}(u^{i+1}) \end{pmatrix} = \begin{pmatrix} c_2^{i+1,J-1} \\ c_2^{i+1,J-2} \\ c_2^{i+1,J-3} \end{pmatrix}$$

for $\mathbf{c}^{i+1,J} = (c_1^{i+1,J}, c_2^{i+1,J})^*$.

A stability criterion for this scheme is

$$\Delta u \leq (\Delta K)^2 \min\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}\right).$$

Although explicit schemes are generally slower than implicit schemes, the programming is

straightforward for the former compared to the latter since the linear system to be solved for the implicit scheme is not anymore tridiagonal.

In the following simulations, we take the parameter values to be $r = 0.02$, $S = 20$, $u = 0$, $T = 1$, $K_{\max} = 60$, $\sigma_1 = 0.1$, and $\sigma_2 = 0.3$. We use 1200 nodes to discretize time axes and 120 nodes to discretized strike axes, i.e., $\Delta u = \frac{1}{1200}$ and $\Delta K = 0.5$, in solving the PDEs (7.14), (7.15) numerically. The numerical solution contains a larger sized dataset which usually does not exist in practice; in reality there is only a set of small data points corresponding to time and strike nodes. In our example, we pick 13-time and 21-strike nodes from the solutions and these prices are used as market data in the inverse Stieltjes moment approach. Next, in order to calculate the truncated moments and their integrals accurately, we interpolate the call prices from time to maturity- and strike-direction. The size of the dataset increases to 500 by 500 points after interpolation. Here, we use a Matlab built-in function called Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) for the interpolation procedure. Note that in general that the market data is not equally spaced through time to maturity and strike prices. Therefore, the number of nodes interpolated between two prices depends on the differences in time to maturity and strike price of the two prices. Finally we solve the algebraic system (7.20) for the “unknown” parameters.

First, let us suppose that the intensity matrix \mathbf{A} is of the form

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix}$$

where $\lambda > 0$. In Table 7.1, we present the estimated parameters for different values of λ and n . Additionally, we evaluate the errors for the estimated option prices using the RMSE, which helps us to analyze the sensitivity of λ and n . From (7.20), the estimated parameters should be independent of n . The error, which occurs by solving the Dupire PDEs, certainly affects

λ	Estimated parameter	n		
		2	3	4
0.25	(λ_1, λ_2)	(0.2567,0.2240)	(0.2523,0.1995)	(0.2498,0.1477)
	(σ_1, σ_2)	(0.0977,0.3002)	(0.0987,0.2989)	(0.0994,0.2964)
	(rmse ₁ ,rmse ₂)	(0.0021,0.0058)	(0.0013,0.0069)	(7.70×10 ⁻⁴ ,0.0094)
0.5	(λ_1, λ_2)	(0.5073,0.4686)	(0.5016,0.4420)	(0.4982,0.3844)
	(σ_1, σ_2)	(0.0978,0.3001)	(0.0989,0.2989)	(0.0997,0.2964)
	(rmse ₁ ,rmse ₂)	(0.0013,0.0051)	(6.49×10 ⁻⁴ ,0.0060)	(1.71×10 ⁻⁴ ,0.0083)
1	(λ_1, λ_2)	(1.0088,0.9570)	(0.9995,0.9205)	(0.9939,0.8571)
	(σ_1, σ_2)	(0.0981,0.2999)	(0.0994,0.2988)	(0.1002,0.2964)
	(rmse ₁ ,rmse ₂)	(8.4×10 ⁻⁴ ,0.0042)	(7.22×10 ⁻⁴ ,0.0049)	(8.87×10 ⁻⁴ ,0.0065)
2	(λ_1, λ_2)	(2.0094,1.9369)	(1.9900,1.9001)	(1.9782,0.19101)
	(σ_1, σ_2)	(0.0987,0.2999)	(0.1004,0.2990)	(0.1015,0.2968)
	(rmse ₁ ,rmse ₂)	(0.0015,0.0034)	(0.0015,0.0019)	(0.0020,0.0047)
5	(λ_1, λ_2)	(4.9470,5.0349)	(4.9522,4.8768)	(4.8405,4.8238)
	(σ_1, σ_2)	(0.1021,0.3017)	(0.1047,0.3007)	(0.1061,0.2992)
	(rmse ₁ ,rmse ₂)	(0.0043,0.0051)	(0.0041,0.0046)	(0.0024,0.0028)
8	(λ_1, λ_2)	(7.7259,8.6577)	(7.5802,8.3862)	(7.5168,8.1685)
	(σ_1, σ_2)	(0.1069,0.3062)	(0.1102,0.3041)	(0.1116,0.3024)
	(rmse ₁ ,rmse ₂)	(0.0162,0.0179)	(0.0143,0.0160)	(0.0120,0.0139)

Table 7.1: Example 1: Estimated parameters for different λ and n with $\sigma_1 = 0.1$ and $\sigma_2 = 0.3$

the calculation. To rectify this to some extent, we utilize larger degrees of the moment, i.e., $n \geq 2$, which would put more weight on option prices in the calculation of moments and consequently reduce the impact of the errors in the Dupire PDEs. In this example, the column corresponding to $n = 2$ shows very good agreement between the estimated values and the actual values assigned when $\lambda \leq 2$. However, if the value of n is too large, higher calculation error occurs when the moments are calculated, and it also affects the estimated results. As can be observed, in the case of $\lambda \leq 2$, most of the RMSEs are larger when higher n is used. Furthermore, the accuracy of the estimation is also affected by the value of the intensity rate. As the intensity rate increases, the market tends to switch more frequently between the states and thus, it is more difficult to capture the information from the market data reducing the accuracy of the estimation. Again, this problem can be corrected by using larger degrees of the moment. Therefore, when $\lambda \geq 2$, RMSE is smaller when higher n is used.

In our second numerical experiment, we assume different intensity rates for state 1 and state 2, so the intensity matrix is of the form

$$\mathbf{A} = \begin{pmatrix} -\lambda_1 & \lambda_2 \\ \lambda_1 & -\lambda_2 \end{pmatrix}.$$

Other values of parameters remain unchanged. The estimated parameters for different values of (λ_1, λ_2) and n are presented in Table 7.2. The difference between the intensity parameters, λ_1 and λ_2 is apparently noticeable on the estimated parameters. When the difference is low, the estimated results closely agree to the actual values for all degrees of moments. In cases where the differences between the intensity parameters are quite substantial, for example $\lambda_1 = 0.25$, $\lambda_2 = 5$, the estimated results are inaccurate for low degree of moments. However, by using a higher degree of moment, e.g., $n = 4$, we still obtain the results closer to actual parameters.

After we estimate the unknown parameters, we calculate the call option prices in each states by solving (7.6) and (7.7). Figure 7.1 shows the estimated option prices and the actual values using $\lambda_1 = 0.25$, $\lambda_2 = 2$ and $n = 2$. Note that the computed values agree very well with the actual data.

7.6 Implementation to “practical data”

In practice, unlike the theoretical data sets used in section 7.4, there is only one set of option prices. In this section, we demonstrate how to split one set of option prices into two sets which would correspond to the two regimes and apply the inverse Stieltjes moment approach. In order that we have the “true” values of the parameters to benchmark with, we simulate option prices with different times to maturity and strike prices via Monte Carlo simulation method.

The parameters used for this particular implementation are $r = 0.02$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.3$. This allows us to compare the numerical estimates and “true” values for the volatility. Let

(λ_1, λ_2)	Estimated parameter	n		
		2	3	4
(0.25, 0.5)	(λ_1, λ_2)	(0.2578,0.4708)	(0.2529,0.4474)	(0.2500,0.3956)
	(σ_1, σ_2)	(0.0976,0.3001)	(0.0987,0.2990)	(0.0993,0.2965)
	(rmse ₁ ,rmse ₂)	(0.0022,0.0051)	(0.0014,0.0060)	(8.21×10 ⁻⁴ ,0.0084)
(0.25, 1)	(λ_1, λ_2)	(0.2604,0.9602)	(0.2541,0.9414)	(0.2506,0.8910)
	(σ_1, σ_2)	(0.0974,0.2997)	(0.0985,0.2989)	(0.0993,0.2968)
	(rmse ₁ ,rmse ₂)	(0.0024,0.0039)	(0.0015,0.0046)	(8.94×10 ⁻⁴ ,0.0064)
(0.25, 2)	(λ_1, λ_2)	(0.2672,1.9039)	(0.2676,1.9134)	(0.2521,1.8780)
	(σ_1, σ_2)	(0.0970,0.2979)	(0.0983,0.2982)	(0.0091,0.2970)
	(rmse ₁ ,rmse ₂)	(0.0027,0.0026)	(0.0017,0.0028)	(0.0010,0.0037)
(0.25, 5)	(λ_1, λ_2)	(0.3044,1.8150)	(0.2772,3.5618)	(0.2612,4.3568)
	(σ_1, σ_2)	(0.0954,0.2253)	(0.0974,0.2693)	(0.0987,0.2870)
	(rmse ₁ ,rmse ₂)	(0.0047,0.0145)	(0.0024,0.0053)	(0.0014,0.0021)
(1, 0.5)	(λ_1, λ_2)	(1.0059,0.4630)	(0.9982,0.4296)	(0.9934,0.3593)
	(σ_1, σ_2)	(0.0983,0.3001)	(0.0996,0.2987)	(0.1004,0.2960)
	(rmse ₁ ,rmse ₂)	(0.0010,0.0053)	(8.89×10 ⁻⁴ ,0.0063)	(0.0011,0.0085)
(1, 2)	(λ_1, λ_2)	(1.0159,1.9301)	(1.0028,1.9159)	(0.9951,1.8564)
	(σ_1, σ_2)	(0.0977,0.2993)	(0.0991,0.2989)	(0.1000,0.2971)
	(rmse ₁ ,rmse ₂)	(7.96×10 ⁻⁴ ,0.0029)	(4.447×10 ⁻⁴ ,0.0033)	(5.71 × 10 ⁻⁴ ,0.0042)
(1, 5)	(λ_1, λ_2)	(1.0518,4.3958)	(1.0194,4.6501)	(1.0007,4.8092)
	(σ_1, σ_2)	(0.0962,0.2887)	(0.0983,0.2984)	(0.0996,0.2973)
	(rmse ₁ ,rmse ₂)	(0.0015,0.0017)	(7.26×10 ⁻⁴ ,0.0013)	(1.84×10 ⁻⁴ ,0.0011)
(1, 8)	(λ_1, λ_2)	(1.1127,2.0694)	(1.0491,5.1333)	(1.0115,6.8723)
	(σ_1, σ_2)	(0.0946,0.2077)	(0.0974,0.2598)	(0.0992,0.2851)
	(rmse ₁ ,rmse ₂)	(0.0044,0.0082)	(0.0014,0.0026)	(5.94×10 ⁻⁴ ,8.18×10 ⁻⁴)
(2, 5)	(λ_1, λ_2)	(2.0416,4.8157)	(2.0013,4.8841)	(1.9788,4.9018)
	(σ_1, σ_2)	(0.0974,0.2980)	(0.0996,0.2991)	(0.1009,0.2994)
	(rmse ₁ ,rmse ₂)	(5.11×10 ⁻⁴ ,0.0014)	(6.81 × 10 ⁻⁴ ,0.0015)	(9.45 × 10 ⁻⁴ ,0.0017)
(2, 8)	(λ_1, λ_2)	(2.0934,7.4342)	(2.0207,7.7349)	(1.9793,8.0197)
	(σ_1, σ_2)	(0.0961,0.2937)	(0.0990,0.2974)	(0.1006,0.3008)
	(rmse ₁ ,rmse ₂)	(6.35×10 ⁻⁴ ,4.58×10 ⁻⁴)	(1.62×10 ⁻⁴ ,5.15×10 ⁻⁴)	(3.91×10 ⁻⁴ ,6.29×10 ⁻⁴)

Table 7.2: Example 2: Estimated parameters for different λ and n with $\sigma_1 = 0.1$ and $\sigma_2 = 0.3$.

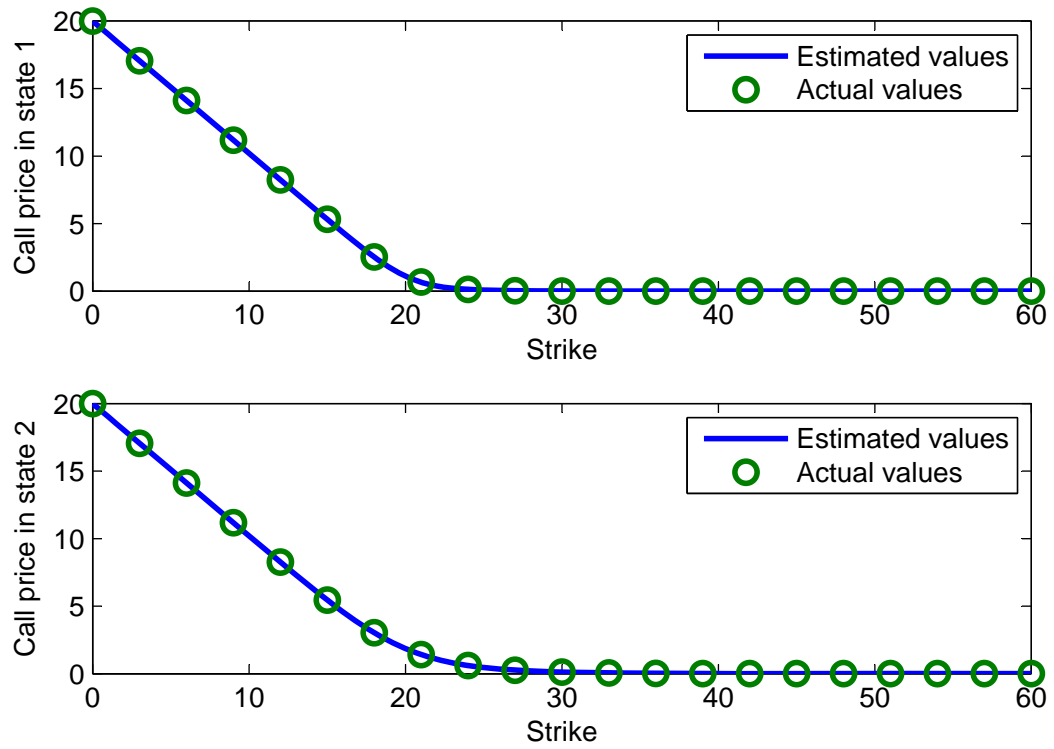


Figure 7.1: Actual and estimated call prices: $\lambda_1 = 0.25, \lambda_2 = 2$ and $n = 2$.

$\bar{C}(K, \tau)$ denote the call option prices obtained from Monte Carlo simulation with 5000 sample paths. The grid for call prices usually vary with the market data. In our example, we use the same grid as described in section 4, which is uniform with 9 nodes on the time axis and 21 nodes on the strike price axis. Let $\bar{C}_1(K, \tau)$ and $\bar{C}_2(K, \tau)$ denote the call prices corresponding to state 1 and 2 which are obtained from \bar{C} .

In splitting $\bar{C}(K, \tau)$ into $\bar{C}_1(K, \tau)$ and $\bar{C}_2(K, \tau)$, we first need to know information of the underlying stock and the behavior of the hidden Markov chain. We adopt the filtering technique from Mamon, et al. [21], and apply to a set of historical prices of the underlying stock. The recursive filters lead to optimal estimates for the volatilities $\hat{\sigma}_1$ and $\hat{\sigma}_2$, the intensity matrix \hat{Q} and the respective probabilities of each state \hat{x}_1 and \hat{x}_2 at time t . Here, the historical data is simulated using $S_{t_0} = 20$, $t_0 = 0$, $t_1 = 1$ and $\Delta t = \frac{1}{500}$. Note that $S_t = S_{t_1}$ is the spot price of the underlying stock.

Next, the estimates $\hat{\sigma}_1$ and $\hat{\sigma}_2$ and \hat{Q} are used to solve the Dupire PDE system in (7.16) to get the “theoretical” call prices $C_1(K, \tau)$ and $C_2(K, \tau)$. Values of other parameters for solving the PDE are $K_{min} = 10^{-3}$, $K_{max} = 60$, $\Delta K = \frac{3}{25}$ and $\Delta t = \frac{1}{500}$. The difference between the two sets of prices, $\Delta C(K, \tau) = C_2(K, \tau) - C_1(K, \tau)$, is then calculated. In order to make the comparison valid between $\Delta C(K, \tau)$ and $\bar{C}(K, \tau)$, the values of ΔC are selected in such a way that they correspond to the grid of \bar{C} . This assumes that the difference between $\bar{C}_1(K, \tau)$ and $\bar{C}_2(K, \tau)$ is equal to $\Delta C(K, \tau)$. It implies that

$$\bar{C}_2(K, \tau) - \bar{C}_1(K, \tau) = \Delta C(K, \tau). \quad (7.22)$$

Furthermore, we suppose that the market price \bar{C} is an expectation of the two prices from each state and thus

$$\bar{C}_1(K, \tau)\hat{x}_1 + \bar{C}_2(K, \tau)\hat{x}_2 = \bar{C}(K, \tau). \quad (7.23)$$

Solving equations (7.22) and (7.23) gives us the call prices corresponding to state 1 and 2. Once the two sets of option prices are obtained, the inverse Stieltjes moment approach is applied to recover the unknown parameters as we did in section 7.4. Similarly, we perform sensitivity analysis when the intensity rates are the same for both states and when they are different.

Based on the derivation of the inverse Stieltjes moment approach, the estimation results will depend on the agreement of data to the Dupire PDEs. Since $\bar{C}_1(K, \tau)$ and $\bar{C}_2(K, \tau)$ are obtained based on the estimated differences and probabilities, they do not satisfy the Dupire PDEs as well as the theoretical data. The error, which occurs by substituting $\bar{C}_1(K, \tau)$ and $\bar{C}_2(K, \tau)$ into the Dupire PDEs, certainly affects the calculation and hence the estimated results could be relatively unstable at times. To rectify this to some extent, we utilize larger degrees of the moment, i.e., $n \geq 5$, which would put more weight on option prices in the calculation of moments and consequently reduce the impact of the errors in the Dupire PDEs. Table 7.3 depicts the estimated parameters for different intensity rates λ and n . To estimate the option prices, we use the estimated σ_1 , σ_2 and \mathbf{Q} to solve (7.16) and substitute the solutions into equation (7.23). Additionally, we evaluate the errors for the estimated option prices using the RMSE, which are also reported in Table 7.3. In the case of $\lambda = 0$, the standard Black-Scholes model without regime switching is recovered, and it has the largest error. When λ is small, the estimated option prices have good agreement with the market prices. As the intensity rate increases, the market becomes more unstable and hence the RMSEs increase as well. This result is similar to the impact of λ on the numerical implementation to the theoretical data. In Figure 7.2, a comparison of the market and estimated option prices with $\lambda = 2$ and $n = 7$ is shown.

The estimated parameters and RMSE with different intensity rates for each state are shown in Table 7.4. Similar to the result shown in Table 7.2, the estimation error is affected by the difference between the two intensity rates. We can observe higher RMSEs in the cases of larger differences between λ_1 and λ_2 . Again the degree of moment, n , somehow effects the estima-

λ	Estimated value	n		
		5	6	7
0	λ	0.2613	0.3358	0.4111
	σ_1	0.1959	0.1062	0.1033
	σ_2	0.2133	0.1939	0.1763
	RMSE	1.4529	1.4454	1.4290
0.25	λ	0.2657	0.2899	0.2768
	σ_1	0.2114	0.1940	0.1735
	σ_2	0.3018	0.3160	0.3141
	RMSE	0.3086	0.4147	0.3713
0.5	λ	0.5846	0.5873	0.6153
	σ_1	0.1708	0.1697	0.1639
	σ_2	0.3130	0.3528	0.3904
	RMSE	0.5554	0.5421	0.5482
1	λ	0.8411	0.7625	0.7878
	σ_1	0.1530	0.1981	0.1997
	σ_2	0.3560	0.3903	0.3386
	RMSE	0.0428	0.0773	0.3753
2	λ	2.5568	2.2888	2.2948
	σ_1	0.1297	0.1617	0.1206
	σ_2	0.3167	0.3591	0.2746
	RMSE	0.1406	0.1904	0.1651
5	λ	5.1631	4.9145	4.5840
	σ_1	0.1761	0.1641	0.1368
	σ_2	0.3302	0.3253	0.2924
	RMSE	0.5679	0.5411	0.5144

Table 7.3: Example 1: Estimated parameters for different λ and n .

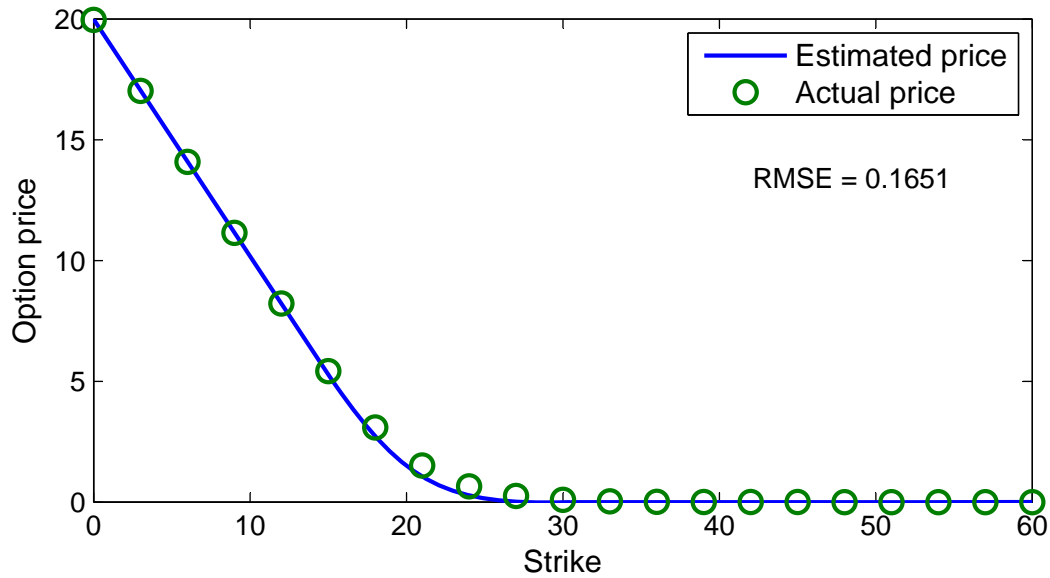


Figure 7.2: Actual and estimated call prices: $\lambda = 2$ and $n = 7$.

tion. However, it appears that a particular pattern for the RMSEs as a result of this impact does not really emerge. We present a comparison of the actual and estimated option prices in Figure 7.3 when $\lambda_1 = 0.25$, $\lambda_2 = 1$ and $n = 7$.

7.7 Conclusion

We developed a methodology based on the inverse Stieltjes moment technique to recover the parameters of a regime-switching model from option prices. In particular, the volatility of the asset price, which is the underlying variable of the option, switches over time and modulated by a continuous-time, finite-state Markov chain. The coupled system of Dupire-type PDEs was derived from the well-known coupled system of Black-Scholes PDEs. The inverse Stieltjes moment approach was adopted to formulate the PDEs forming a linear system of equations for the volatilities and the intensity parameters. We demonstrated how to apply this method to “theoretical data”, which were obtained by solving the Dupire PDEs, and “practical data” obtained from market data. Numerical results were presented to illustrate the accuracy of our

(λ_1, λ_2)	Estimated parameter	n		
		5	6	7
(0.25, 0.5)	(λ_1, λ_2)	(0.1891,0.4951)	(0.2631,0.5950)	(0.2202,0.6399)
	(σ_1, σ_2)	(0.1373,0.2791)	(0.1893,0.2605)	(0.1171,0.2705)
	RMSE	0.3398	0.5176	0.6128
(0.25, 1)	(λ_1, λ_2)	(0.2244,1.1918)	(0.2300,0.8251)	(0.1900,0.9501)
	(σ_1, σ_2)	(0.1670,0.2793)	(0.1555,0.2955)	(0.1366,0.3360)
	RMSE	0.3693	0.2912	0.2374
(0.25, 2)	(λ_1, λ_2)	(0.1933,1.2001)	(0.2126,1.4492)	(0.2268,1.3190)
	(σ_1, σ_2)	(0.1729,0.2233)	(0.1701,0.2383)	(0.1763,0.2053)
	RMSE	0.3054	0.3002	0.3093
(0.25, 5)	(λ_1, λ_2)	(0.1725,1.8142)	(0.1768,2.1447)	(0.1859,2.1059)
	(σ_1, σ_2)	(0.1772,0.3454)	(0.1788,0.2581)	(0.1741,0.2760)
	RMSE	0.8698	0.9121	0.9342
(1, 0.5)	(λ_1, λ_2)	(1.0440,0.6132)	(0.7531,0.5906)	(0.6788,0.3740)
	(σ_1, σ_2)	(0.1232,0.2845)	(0.1841,0.3932)	(0.1451,0.3800)
	RMSE	0.1183	0.7557	1.0029
(1, 2)	(λ_1, λ_2)	(0.7180,1.5622)	(1.0843,2.6256)	(1.4354,1.8084)
	(σ_1, σ_2)	(0.1833,0.2877)	(0.1621,0.3567)	(0.1425,0.3327)
	RMSE	0.5907	0.5139	0.4401
(1, 5)	(λ_1, λ_2)	(0.9290,2.5717)	(1.4728,3.1720)	(0.9066,6.3060)
	(σ_1, σ_2)	(0.1062,0.2192)	(0.1173,0.2543)	(0.1219,0.2412)
	RMSE	0.7413	0.5533	0.7658
(2, 1)	(λ_1, λ_2)	(1.9702,0.8243)	(1.8701,1.5095)	(1.8611,0.9318)
	(σ_1, σ_2)	(0.1748,0.2691)	(0.1549,0.4179)	(0.1352,0.3570)
	RMSE	0.3975	0.4297	0.5018
(2, 5)	(λ_1, λ_2)	(1.6245,3.3135)	(1.2278,4.3946)	(1.8332,4.3734)
	(σ_1, σ_2)	(0.1363,0.2522)	(0.1123,0.3441)	(0.1262,0.3569)
	RMSE	0.3884	0.2724	0.1648

Table 7.4: Example 2: Estimated parameters for different λ and n .

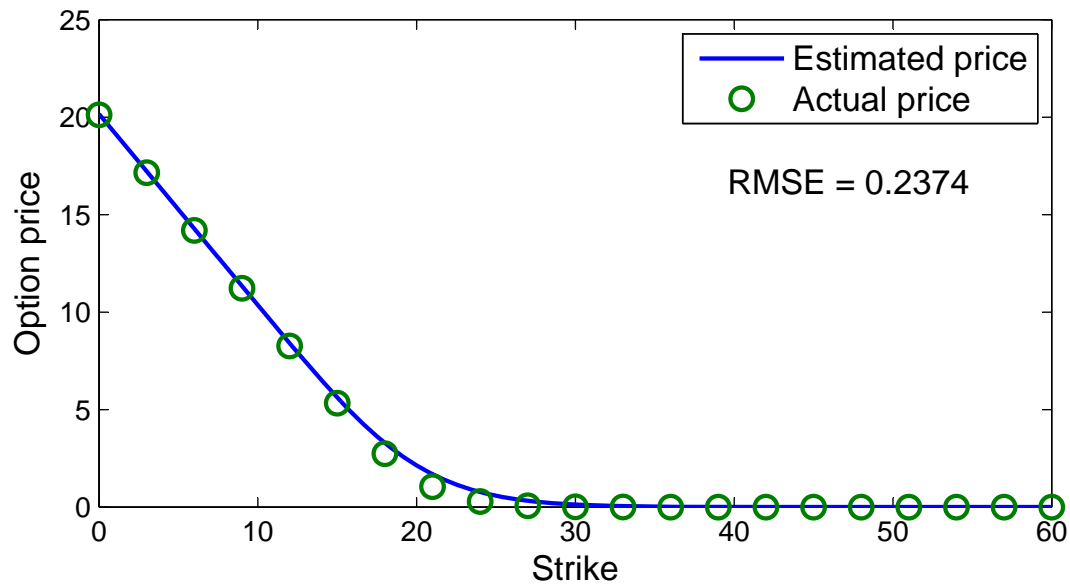


Figure 7.3: Actual and estimated call prices: $\lambda_1 = 0.25$, $\lambda_2 = 1$ and $n = 7$.

method. We also performed various analyses for both cases when the intensity parameters of the intensity matrix \mathbf{A} and the degree of the moment n is varied. Our findings based on the numerical experiments on the two types of data sets indicate the following: (i) for a single intensity rate λ , the higher the intensity rate the lower accuracy of the method, and (ii) for two different intensity parameters λ_1 and λ_2 , the greater the difference between these intensity rates the less accurate the estimation.

References

- [1] L. B. G. Andersen and R. Brotherton-Ratcliffe. The equity option volatility smile: An implicit finite-difference approach. *Journal of Computational Finance*, 1(2):5–37, 1998.
- [2] A. Ang and G. Bekaert. Regime switches in interest rates. *Journal of Business and Economic Statistics*, 20(2):163–82, 2002.
- [3] G. Bekaert and R. J. Hodrick. On biases in the measurement of foreign exchange risk premiums. *Journal of International Money and Finance*, 12(2):115–138, 1993.
- [4] I. Bouchouev and V. Isakov. The inverse problem of option pricing. *Inverse Problems*, 13(5):L11, 1997.
- [5] P. Boyle and T. Draviam. Pricing exotic options under regime switching. *Insurance: Mathematics and Economics*, 40(2):267–282, 2007.
- [6] P. P. Boyle and D. Thangaraj. Volatility estimation from observed option prices. *Decisions in Economics and Finance*, 23(1):31–52, 2000.
- [7] W. K. Ching, T. K. Siu, and L. Li. Pricing exotic options under a higher-order hidden Markov model. *Journal of Applied Mathematics and Decision Sciences*, pages 1–15, 2007. Article ID 18014.
- [8] C. J. Chu, G. J. Santoni, and T. Liu. Stock market volatility and regime shifts in returns. *Information Sciences*, 94:179–190, 1996.

- [9] M. Dahlquist and S. F. Gray. Regime-switching and interest rates in the European monetary system. *Journal of International Economics*, 50(2):399–419, 2000.
- [10] Z. Deng, J. Yu, and L. Yang. An inverse problem of determining the implied volatility in option pricing. *Journal of Mathematical Analysis and Applications*, 340(1):16–31, 2008.
- [11] J. Duan, I. Popova, and P. Ritchken. Option pricing under regime switching. *Quantitative Finance*, 2:116–132, 2002.
- [12] B. Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.
- [13] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer, Berlin, 1994.
- [14] R. J. Elliott, L. Chan, and T. K. Siu. Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1(4):423–432, 2005.
- [15] R. J. Elliott and P. E. Kopp. *Mathematics of Financial Markets*. Springer, Berlin Heidelberg New York, 1999.
- [16] R. J. Elliott and R. S. Mamon. An interest rate model with a Markovian mean reverting level. *Quantitative Finance*, 2:454–458, 2002.
- [17] R. J. Elliott and J. van der Hoek. An application of hidden Markov models to asset allocation problems. *Finance and Stochastics*, 1:229–238, 1997.
- [18] C. Engel and C. S. Hakkio. The distribution of exchange rates in the EMS. *International Journal of Finance and Economics*, 1(1):55–67, 1996.
- [19] C. Engel and J. D. Hamilton. Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review*, 80(4):689–713, 1990.
- [20] X. Guo. Information and option pricing. *Quantitative Finance*, 1:38–44, 2001.

- [21] R. S. Mamon, C. Erlwein, and R. B. Gopaluni. Adaptive signal processing of asset price dynamics with predictability analysis. *Information Sciences*, 187:203–219, 2008.
- [22] R. S. Mamon and M. R. Rodrigo. Explicit solutions to European options in a regime-switching economy. *Operations Research Letters*, 33(6):581–586, 2005.
- [23] M. R. Rodrigo and R. S. Mamon. Recovery of time-dependent parameters of a Black-Scholes-type equation: An inverse Stieltjes moment approach. *Journal of Applied Mathematics*, 2007, 2007. Article ID 62098.
- [24] M. R. Rodrigo and R. S. Mamon. A new representation of the local volatility surface. *International Journal of Theoretical and Applied Finance*, 11(7):691–703, 2008.
- [25] T. K. Siu, C. Erlwein, and R. S. Mamon. The pricing of credit default swaps under a Markov-modulated Merton’s structural model. *North American Actuarial Journal*, 12(1):229–238, 2008.
- [26] C. M. Turner, R. Startz, and C. R. Nelson. A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25(1):3–22, 1989.

Chapter 8

Concluding remarks

8.1 Summary and commentaries

In this thesis, we consider further theoretical developments of regime-switching models modulated by higher-order HMMs together with their efficient and dynamic parameter estimation. Special emphasis was given to various applications in finance in the context of risk management, asset allocation and the inverse problem involved in pricing derivatives. Throughout the course of this study, we exploited the power and capabilities of WHMMs and found opportunities demonstrating that they outperform the usual HMM under various criteria such as goodness of fit, model adequacy and investment performance metrics. The benefits from WHMMs stem from the fact that they are able to capture more information from the past by weakening the Markov assumption and extending the dependency to any number of prior epochs. This feature of WHMMs is appropriate for financial time series which usually exhibit a memory property. Filtering methods via the change of reference probability measures along with the EM algorithm were employed in the estimation of parameters using historical data. To complement our contributions to parameter estimation using past information for regime-switching models, we address the inverse problem under a regime-switching model for equity options. Below is a brief summary of what have been accomplished in the chapters containing the research dealt with in this thesis.

In chapter 2, we provided the background of and motivation for WHMM. A transformation mapping is introduced that permits the embedding of WHMM into a first-order HMM. Thus, general recursive filters for the discrete-time, continuous-range observations based on traditional HMM filtering method could be applied. By including a mechanism that captures memory in the states of the model, the HMM is outperformed by the WHMM with respect to low forecasting errors over the long forecasting horizons.

The case when drift and volatility components of asset returns are driven by two different WHMM is discussed in chapter 3. The two independent WMCs are converted into a new WMC through a tensor-product-based technique. Numerical examples on the simulated data demonstrate the accuracy of our estimation algorithms as “true” parameters were recovered.

In chapter 4, the filtering of a multivariate WHMM is developed and its application in asset allocation is considered. Parameter estimates were derived in the context of vector observations. Two WHMM-based investment strategies (switching and mixed) were proposed and compared to their counterparts under the HMM setting. For certain levels of transaction costs, the WHMM-based strategies produce better results than those from the HMM-based strategies on the basis of higher differences of logreturn values and benchmarks on pure strategies.

Another application of multivariate HMM was given in chapter 5 dealing with the term structure of interest rates. The proposed model and filtering algorithms were tested on time series data of yields on 3- and 6-month US T bills, 1- and 5-year US T-notes, and 20- and 30-year US T-bonds. Empirical results of our implementation of the filtering and parameter estimation methods illustrate the reasonability and sufficiency of our modeling approach in capturing market dynamics and regime changes in the data. Such results also provide support for the merits of employing WHMM over HMM when the data have memory property.

Continuing the theme of interest rate modeling, a regime-switching Hull-White model is presented in chapter 6. In contrast to regime-switching interest rate models in the literature, we considered a model where its mean level, speed of mean reversion and volatility are all driven by a WMC. Recursive filters and estimates for the parameters were derived. The estimation algorithms were implemented on a data series of 30-day Canadian T-bill yields (as proxy for the short rate process) compiled during a 10-year period. By utilizing a higher-order HMM, a better fit was obtained than those from the usual HMM.

In chapter 7, a method based on the inverse Stieltjes moment technique is put forward to recover parameters of a Markov-switching model from option prices. The coupled system of Dupire-type PDEs was established using the well-known coupled Black-Scholes PDEs. Then, the inverse moment approach was applied to formulate the PDEs forming a linear system of equations for the volatility and intensity parameters. We demonstrated the accuracy of our proposed method by applying it to “theoretical” and “practical data”.

8.2 Further research directions

Possible extensions and powerful enhancements can further be investigated as offshoots of these research works. We detail them below.

- As noted in this thesis, we concentrated on a WHMM of order 2 to temper the computational complexity involved in the implementation of filtering parameter estimation. An alternative methodology to the transform suggested here that can handle the estimation of WHMM with a lag higher than 2 would be desirable.
- The signal dynamics of WHMM can be made more general and nonlinear. Research on the WHMM context but similar to that of an HMM with nonlinear dynamics discussed in section 4.4 of Elliott, et al. [2] would be a natural extension. The development of unnormalized filters and recursive filters could be the aim of the investigation.

- A similar extension can be afforded to WHMM based on Hughes, et al. [4] proposing a nonhomogeneous HMM in which the transition probabilities depend on the history of states and the covariate of the underlying Markov chain.
- So far, we only considered WHMMs with discrete-time observations. The models can be studied under the continuous-time framework. In this case, we have an intensity matrix and its estimation along with the estimation of other parameters can be expected to be more complex given that the estimation of joint occupation times of the WMC and the observation process will be needed. See Elliott, et al. [2] for the usual HMM case.
- A class of WHMM-driven models for non-Gaussian observations could be examined. For example, a jump diffusion model or constant elasticity of variance (CEV) model have parameters driven by a WHMM.
- Some previous papers [1, 5, 3] suggested GARCH-type models with parameters driven by HMM. It would be worth examining the benefits of replacing HMM by WHMM.
- As noted in this thesis, the speed of convergence of parameter estimates is greatly affected by the initial values chosen. Thus, an efficient and systematic method or solid heuristics in setting appropriate initial values would be very helpful.
- With regard to the asset allocation problem, it would be interesting to explore applications that involve other types of portfolios such as those that cover currencies and other kinds of commodities and see if WHMM still offers some advantages and up to what extent.
- The inclusion of correlation between the driving noise (Brownian motion) of each asset return in the multivariate WHMM setting would lead to a more accurate estimation and effective risk management. The online estimation of correlation using the WHMM filtering approach will be a worthwhile investigation.

- The calibration method considered in chapter 7 is developed under a continuous-time HMM setting and the PDE was discretized in the implementation stage. Its extension to the WHMM setting and measuring the extent of the merits of this new setting over the added complexity would be an appealing research endeavor.

References

- [1] M. Carrasco and X. Chen. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39, 2002.
- [2] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer, New York, 1995.
- [3] M. Haas, S. Mittnik, and M. S. Paolella. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2:493–530, 2004.
- [4] J. P. Hughes, P. Guttorp, and S. P. Charles. A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48:15–30, 1999.
- [5] F. Klaassen. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27(2):363–394, 2002.

Appendix A

Proof of equation (2.16)

From the definition of q_k in equation (2.10), we have

$$\begin{aligned}
 q_{k+1} &= \bar{E}[\Lambda_{k+1} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_{k+1}] \\
 &= \bar{E}[\Lambda_k \lambda_{k+1} \mathbf{\Pi} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_{k+1}] \\
 &= \sum_{l,m,v=1}^N b_{k+1}^m \bar{E}[\Lambda_k \langle \mathbf{\Pi} \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{mv} \rangle | \mathcal{Y}_k] \mathbf{e}_{lm} \\
 &= \sum_{l,m,v=1}^N b_{k+1}^m \langle \mathbf{\Pi} \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \mathbf{q}_k, \mathbf{e}_{mv} \rangle \mathbf{e}_{lm} \\
 &= \mathbf{B}_{k+1} \mathbf{\Pi} \mathbf{q}_k.
 \end{aligned}$$

Appendix B

Proof of Proposition 2.4.1

B.1 Proof of equation (2.22)

Using the expressions in (2.6)-(2.7) and the definition in (2.22), we obtain

$$\begin{aligned}
& \gamma(J^{rst} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} \\
&= \bar{E}[\Lambda_{k+1} J_{k+1}^{rst} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_{k+1}] \\
&= \bar{E}[\Lambda_k \lambda_{k+1} (J_k^{rst} + \langle \mathbf{x}_{k+1}, \mathbf{e}_r \rangle \langle \mathbf{x}_k, \mathbf{e}_s \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_t \rangle) \Pi \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_{k+1}] \\
&= \sum_{l,m,v=1}^N b_{k+1}^m \bar{E}[\Lambda_k J_k^{rst} \langle \Pi \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{mv} \rangle | \mathcal{Y}_k] \mathbf{e}_{lm} \\
&\quad + b_{k+1}^r \bar{E}[\Lambda_k \langle \Pi \mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle | \mathcal{Y}_k] \mathbf{e}_{rs} \\
&= \sum_{l,m,v=1}^N b_{k+1}^m \langle \Pi \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \gamma(J_k^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k, \mathbf{e}_{mv} \rangle \mathbf{e}_{lm} + b_{k+1}^r \langle \Pi \mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \mathbf{q}_k, \mathbf{e}_{st} \rangle \mathbf{e}_{rs} \\
&= \mathbf{B}_{k+1} \Pi \gamma(J^{rst} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \langle \Pi \mathbf{e}_{st}, \mathbf{e}_{rs} \rangle \langle \mathbf{q}_k, \mathbf{e}_{st} \rangle \mathbf{e}_{rs}
\end{aligned}$$

B.2 Proof of equation (2.23)

Using the expressions in (2.6)-(2.7) and the definition in (2.23), we get

$$\begin{aligned}
& \gamma(O^{rs}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} \\
&= \bar{E}[\Lambda_{k+1} O_k^{rs} \xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_{k+1}] \\
&= \bar{E}[\Lambda_k \lambda_{k+1} (O_k^{rs} + \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle) \xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_{k+1}] \\
&= \sum_{l,m,v=1}^N b_{k+1}^m \bar{E}[\Lambda_k O_k^{rs} \langle \mathbf{\Pi} \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{mv} \rangle | \mathcal{Y}_k] \mathbf{e}_{lm} \\
&\quad + \sum_{l=1}^N b_{k+1}^r \bar{E}[\Lambda_k \langle \mathbf{\Pi} \mathbf{e}_{rs}, \mathbf{e}_{lr} \rangle \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle | \mathcal{Y}_k] \mathbf{e}_{lr} \\
&= \sum_{l,m,v=1}^N b_{k+1}^m \langle \mathbf{\Pi} \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \gamma(O_k^{rs} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k, \mathbf{e}_{mv} \rangle \mathbf{e}_{lm} + b_{k+1}^r \sum_{l=1}^N \langle \mathbf{\Pi} \mathbf{e}_{rs}, \mathbf{e}_{lr} \rangle \langle \mathbf{q}_k, \mathbf{e}_{rs} \rangle \mathbf{e}_{lr} \\
&= \mathbf{B}_{k+1} \mathbf{\Pi} \gamma(O^{rs} \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \langle \mathbf{q}_k, \mathbf{e}_{rs} \rangle \mathbf{\Pi} \mathbf{e}_{rs}
\end{aligned}$$

B.3 Proof of equation (2.24)

From the expressions in (2.6)-(2.7) and the definition in (2.24), we have

$$\begin{aligned}
& \gamma(O^r \xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} \\
&= \bar{E}[\Lambda_{k+1} O_{k+1}^r \xi(\mathbf{x}_{k+1}, \mathbf{x}_k) | \mathcal{Y}_{k+1}] \\
&= \bar{E}[\Lambda_k \lambda_{k+1} (O_k^r + \langle \mathbf{x}_k, \mathbf{e}_r \rangle) \Pi \xi(\mathbf{x}_k, \mathbf{x}_{k-1}) | \mathcal{Y}_{k+1}] \\
&= \sum_{l,m,v=1}^N b_{k+1}^m \bar{E}[\Lambda_k O_k^r \langle \Pi \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{mv} \rangle | \mathcal{Y}_k] \mathbf{e}_{lm} \\
&\quad + \sum_{l,m=1}^N b_{k+1}^r \bar{E}[\Lambda_k \langle \Pi \mathbf{e}_{rm}, \mathbf{e}_{lr} \rangle \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rm} \rangle | \mathcal{Y}_k] \mathbf{e}_{lr} \\
&= \sum_{l,m,v=1}^N b_{k+1}^m \langle \Pi \mathbf{e}_{mv}, \mathbf{e}_{lm} \rangle \langle \gamma(O_k^r \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k, \mathbf{e}_{mv} \rangle \mathbf{e}_{lm} + \sum_{l,m=1}^N b_{k+1}^r \langle \Pi \mathbf{e}_{rm}, \mathbf{e}_{lr} \rangle \langle \mathbf{q}_k, \mathbf{e}_{rm} \rangle \mathbf{e}_{lr} \\
&= \mathbf{B}_{k+1} \Pi \gamma(O^r \xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k + b_{k+1}^r \mathbf{V}_r \Pi \mathbf{q}_k
\end{aligned}$$

The proof of recursive formulas (2.25) follow similar proof of equation (2.24) by using the definition of λ_l and evaluating the resulting conditional expectation under \bar{P} .

Appendix C

Proof of Proposition 2.5.1

C.1 Proof of equation (2.27)

To derive an optimal estimate for a_{rst} , we consider a new measure $P_{\hat{\theta}}$, which is defined in equation (2.26). This means that

$$\begin{aligned} \log\left(\frac{dP_{\hat{\theta}}}{dP_{\theta}}\middle|\mathcal{Y}_k\right) &= \sum_{l=2}^k \sum_{r,s,t=1}^N [\log(\hat{a}_{rst}) - \log(a_{rst})] \langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle \\ &= \sum_{r,s,t=1}^N \log \hat{a}_{rst} J^{rst} + \text{Remainder}, \end{aligned} \quad (\text{C.1})$$

where the Remainder is independent of \hat{a}_{rst} . Write $E[\cdot] = E_{\hat{\theta}}[\cdot]$. Taking expectation of equation (C.1) we have

$$E\left[\log\left(\frac{dP_{\hat{\theta}}}{dP_{\theta}}\right)\middle|\mathcal{Y}_k\right] = \sum_{r,s,t=1}^N \log \hat{a}_{rst} \hat{J}^{rst} + \text{Remainder}. \quad (\text{C.2})$$

The optimal estimate of a_{rst} is the value that maximizes the log-likelihood in equation (C.1) subject to the constraint $\sum_{r=1}^N \hat{a}_{rst} = 1$. Now, we introduce the Lagrange multiplier β with the function:

$$L(\hat{a}_{rst}, \beta) = \sum_{r,s,t=1}^N \log \hat{a}_{rst} \hat{J}^{rst} + \beta \left(\sum_{r=1}^N \hat{a}_{rst} - 1 \right) + \text{Remainder}. \quad (\text{C.3})$$

Differentiating equation (C.3) with respect to \hat{a}_{rst} and β , and equating the derivatives to 0, we

get

$$\frac{1}{\hat{a}_{rst}} \hat{j}^{rst} + \beta = 0 \quad (\text{C.4})$$

and

$$\sum_{r=1}^N \hat{a}_{rst} = 1. \quad (\text{C.5})$$

Rewriting (C.4) yields

$$\hat{a}_{rst} = -\frac{\hat{j}^{rst}}{\beta}. \quad (\text{C.6})$$

Consequently, from equations (C.5) and (C.6) we have

$$\sum_{r=1}^N \hat{a}_{rst} = -\sum_{r=1}^N \frac{\hat{j}^{rst}}{\beta} = -\frac{\hat{O}^{st}}{\beta} = 1.$$

Hence, the Lagrange multiplier has the value $\beta = -\hat{O}^{st}$. From equation (C.6), the optimal estimates for \hat{a}_{rst} is

$$\hat{a}_{rst} = \frac{\hat{j}^{rst}}{\hat{O}^{st}},$$

which is the desired equation (2.27) that we wanted to show.

C.2 Proof of equation (2.28)

Given the parameters $\mathbf{f} = (f_1, f_2, \dots, f_N)^\top \in \mathbb{R}^N$ and we wish to perform a change to $\hat{\mathbf{f}} = (\hat{f}_1, \hat{f}_2, \dots, \hat{f}_N)^\top \in \mathbb{R}^N$. Consider a new measure $P_{\hat{\theta}}$ defined by

$$\frac{dP_{\hat{\theta}}}{dP_{\theta}} \Big|_{\mathcal{B}_k} = \Gamma_k^f = \prod_{l=1}^k \lambda_l^f,$$

where

$$\lambda_l^f = \exp\left(\frac{1}{2\sigma(\mathbf{x}_{l-1})^2} (f(\mathbf{x}_{l-1})^2 - \hat{f}(\mathbf{x}_{l-1})^2 - 2y_l f(\mathbf{x}_{l-1}) + 2y_l \hat{f}(\mathbf{x}_{l-1}))\right).$$

Therefore, we have

$$\begin{aligned}
E\left[\log \frac{dP_{\hat{\theta}}}{dP_{\theta}} \Big| \mathcal{Y}_k\right] &= E\left[\sum_{l=1}^k \log \lambda_l^f \Big| \mathcal{Y}_k\right] \\
&= E\left[\sum_{l=1}^k \frac{2y_l \hat{f}(\mathbf{x}_{l-1}) - \hat{f}(\mathbf{x}_{l-1})^2}{2\sigma(\mathbf{x}_{l-1})^2} + \text{Remainder} \Big| \mathcal{Y}_k\right] \\
&= \sum_{l=1}^k E\left[\sum_{r=1}^N \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \frac{2y_l \hat{f}_r - \hat{f}_r^2}{2\sigma_r^2} \Big| \mathcal{Y}_k\right] + \text{Remainder} \\
&= \sum_{l=1}^k E\left[\frac{2T_k^r(y) \hat{f}_r - O_k^r \hat{f}_r^2}{2\sigma_r^2} \Big| \mathcal{Y}_k\right] + \text{Remainder} \\
&= \sum_{l=1}^k \frac{2\hat{T}_k^r(y) \hat{f}_r - \hat{O}_k^r \hat{f}_r^2}{2\sigma_r^2} + \text{Remainder}, \tag{C.7}
\end{aligned}$$

where the Remainder does not involve \hat{f} . We differentiate the above expression and set its derivative to 0. This gives the optimal choice for \hat{f}_i given the observation data y so that we have

$$\hat{f}_r = \frac{\hat{T}_k^r(y)}{\hat{O}_k^r}.$$

C.3 Proof of equation (2.29)

To perform a change from $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)^\top \in \mathbb{R}^N$ to $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_N)^\top \in \mathbb{R}^N$, we define the Radon-Nikodým derivative as

$$\frac{dP_{\hat{\theta}}}{dP_{\theta}} \Big|_{\mathcal{Y}_k} = \Gamma_k^\sigma = \prod_{l=1}^k \lambda_l^\sigma,$$

where

$$\lambda_l^\sigma = \frac{\sigma(\mathbf{x}_{l-1})}{\hat{\sigma}(\mathbf{x}_{l-1})} \exp\left(\frac{1}{2\sigma(\mathbf{x}_{l-1})^2} (y_l - f(\mathbf{x}_{l-1}))^2 - \frac{1}{2\hat{\sigma}(\mathbf{x}_{l-1})^2} (y_l - f(\mathbf{x}_{l-1}))^2\right).$$

Hence,

$$E\left[\log \left(\frac{dP_{\hat{\theta}}}{dP_{\theta}}\right) \Big| \mathcal{Y}_k\right] = E\left[\sum_{l=1}^k \left(-\log \hat{\sigma}(\mathbf{x}_{l-1}) - \frac{(y_l - f(\mathbf{x}_{l-1}))^2}{2\hat{\sigma}(\mathbf{x}_{l-1})^2}\right) + \text{Remainder} \Big| \mathcal{Y}_k\right]$$

$$\begin{aligned}
&= E \left[- \sum_{l=1}^k \sum_{r=1}^N \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \left(\log \hat{\sigma}_r + \frac{(y_l - f_r)^2}{2\hat{\sigma}_r^2} \right) \middle| \mathcal{D}_k \right] + \text{Remainder} \\
&= \sum_{r=1}^N \left(- \log \hat{\sigma}_r \hat{\mathcal{O}}_k^r - \frac{1}{2\hat{\sigma}_r^2} (\hat{T}_k^r(y^2) - 2\hat{T}_k^r(y)f_r + f_r^2) \right) + \text{Remainder}, \quad (\text{C.8})
\end{aligned}$$

where the Remainder is independent of $\hat{\sigma}$. We differentiate the above expression in $\hat{\sigma}_r$ and equate the result to zero. Solving the equation we get the optimal choice of $\hat{\sigma}^2$:

$$\hat{\sigma}_r^2 = \frac{1}{\hat{\mathcal{O}}_k^r} (\hat{T}_k^r(y^2) - 2\hat{T}_k^r(y)\hat{f}_r + \hat{f}_r^2 \hat{\mathcal{O}}_k^r),$$

which is the result in equation (2.29).

Appendix D

Proof of Proposition 6.3.2

D.1 Proof of equation (6.25)

Given the parameter $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^\top \in \mathbb{R}^N$, we wish to update the estimates to $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N)^\top \in \mathbb{R}^N$. Consider a new measure $P_{\hat{\theta}}$ defined by

$$\frac{dP_{\hat{\theta}}}{dP_{\theta}} \Big|_{\mathcal{Y}_k} = \Lambda_k^{\hat{\alpha}} = \prod_{l=1}^k \lambda_l(\hat{\alpha}_{l-1}, y_l),$$

where

$$\lambda_l(\hat{\alpha}_{l-1}, y_l) = \exp \left\{ -\frac{(y_l - \hat{\alpha}(\mathbf{x}_{l-1})y_{l-1} - \eta(\mathbf{x}_{l-1}))^2 - (y_l - \alpha(\mathbf{x}_{l-1})y_{l-1} - \eta(\mathbf{x}_{l-1}))^2}{2\sigma(\mathbf{x}_{l-1})^2} \right\}.$$

Write $E[\cdot] = E_{\hat{\theta}}[\cdot]$. Therefore, we have

$$\begin{aligned} E \left[\log \frac{dP_{\hat{\theta}}}{dP_{\theta}} \Big|_{\mathcal{Y}_k} \right] &= E \left[\sum_{l=1}^k \log \lambda_l(\alpha_{l-1}, y_l) \Big|_{\mathcal{Y}_k} \right] \\ &= E \left[\sum_{l=1}^k -\frac{(\hat{\alpha}(\mathbf{x}_{l-1})^2 y_{l-1}^2 - 2\hat{\alpha}(\mathbf{x}_{l-1})y_{l-1}y_l + \hat{\alpha}(\mathbf{x}_{l-1})\eta(\mathbf{x}_{l-1})y_{l-1})}{2\sigma(\mathbf{x}_{l-1})^2} + \mathbf{R} \Big|_{\mathcal{Y}_k} \right] \\ &= \sum_{l=1}^k E \left[\sum_{r=1}^N -\frac{\langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle}{2\sigma_r^2} (\hat{\alpha}_r^2 y_{l-1}^2 - 2\hat{\alpha}_r y_{l-1}y_l + \hat{\alpha}_r \eta_r y_{l-1}) \Big|_{\mathcal{Y}_k} \right] + \mathbf{R} \end{aligned}$$

$$\begin{aligned}
&= \sum_{r=1}^N E \left[-\frac{1}{2\sigma_r^2} \left(\hat{\alpha}_r^2 T_k^r(y_{k-1}^2) - 2\hat{\alpha}_r T_k^r(y_{k-1}y_k) + \hat{\alpha}_r \eta_r T_k^r(y_{k-1}) \right) \middle| \mathcal{Y}_k \right] + \mathbf{R} \\
&= \sum_{r=1}^N -\frac{1}{2\sigma_r^2} \left(\hat{\alpha}_r^2 \hat{T}_k^r(y_{k-1}^2) - 2\hat{\alpha}_r \hat{T}_k^r(y_{k-1}y_k) + \hat{\alpha}_r \eta_r \hat{T}_k^r(y_{k-1}) \right) + \mathbf{R}, \tag{D.1}
\end{aligned}$$

where \mathbf{R} does not involve $\hat{\alpha}$. We differentiate the above expression and set its derivative to 0.

This gives the optimal choice for $\hat{\alpha}_i$ given the observation data y_k . We get

$$\hat{\alpha}_r = \frac{\hat{T}_k^r(y_{k-1}, y_k) - \eta_r \hat{T}_k^r(y_{k-1})}{\hat{T}_k^r(y_{k-1}^2)}.$$

D.2 Proof of equation (6.26)

Given the parameter $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_N)^\top \in \mathbb{R}^N$, we wish to obtain the update $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \hat{\eta}_2, \dots, \hat{\eta}_N)^\top \in \mathbb{R}^N$. Consider a new measure $P_{\hat{\boldsymbol{\theta}}}$ defined by

$$\frac{dP_{\hat{\boldsymbol{\theta}}}}{dP_{\boldsymbol{\theta}}} \Big|_{\mathcal{Y}_k} = \Lambda_k^\alpha = \prod_{l=1}^k \lambda_l(\hat{\boldsymbol{\eta}}_{l-1}, y_l),$$

where

$$\lambda_l(\hat{\boldsymbol{\eta}}_{l-1}, y_l) = \exp \left\{ -\frac{(y_l - \boldsymbol{\eta}(\mathbf{x}_{l-1}))y_{l-1} - \hat{\boldsymbol{\eta}}(\mathbf{x}_{l-1}))^2 - (y_l - \boldsymbol{\alpha}(\mathbf{x}_{l-1}))y_{l-1} - \boldsymbol{\eta}(\mathbf{x}_{l-1}))^2}{2\sigma(\mathbf{x}_{l-1})^2} \right\}.$$

Therefore, we have

$$\begin{aligned}
E \left[\log \frac{dP_{\hat{\boldsymbol{\theta}}}}{dP_{\boldsymbol{\theta}}} \middle| \mathcal{Y}_k \right] &= E \left[\sum_{l=1}^k \log \lambda_l(\boldsymbol{\alpha}_{l-1}, y_l) \middle| \mathcal{Y}_k \right] \\
&= E \left[\sum_{l=1}^k -\frac{(\hat{\boldsymbol{\eta}}(\mathbf{x}_{l-1}))^2 - 2\hat{\boldsymbol{\eta}}(\mathbf{x}_{l-1})y_l + 2\boldsymbol{\alpha}(\mathbf{x}_{l-1})\hat{\boldsymbol{\eta}}(\mathbf{x}_{l-1})y_{l-1}}{2\sigma(\mathbf{x}_{l-1})^2} + \mathbf{R} \middle| \mathcal{Y}_k \right] \\
&= \sum_{l=1}^k E \left[\sum_{r=1}^N -\frac{\langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle}{2\sigma_r^2} (\hat{\eta}_r^2 - 2\hat{\eta}_r y_l + 2\alpha_r \hat{\eta}_r y_{l-1}) \middle| \mathcal{Y}_k \right] + \mathbf{R}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{r=1}^N E \left[-\frac{1}{2\sigma_r^2} \left(\hat{\eta}_r^2 O_k^r - 2\hat{\eta}_r T_k^r(y_k) + 2\alpha_r \hat{\eta}_r T_k^r(y_{k-1}) \right) \middle| \mathcal{Y}_k \right] + \mathbf{R} \\
&= \sum_{r=1}^N -\frac{1}{2\sigma_r^2} \left(\hat{\eta}_r^2 \hat{O}_k^r - 2\hat{\eta}_r \hat{T}_k^r(y_k) + 2\alpha_r \hat{\eta}_r \hat{T}_k^r(y_{k-1}) \right) + \mathbf{R}, \tag{D.2}
\end{aligned}$$

where \mathbf{R} does not involve $\hat{\eta}$. We differentiate the above expression and set its derivative to 0.

This gives the optimal choice for $\hat{\eta}_i$ given the observation data y_k . We obtain

$$\hat{\eta}_r = \frac{\hat{T}_k^r(y_k) - \alpha^r \hat{T}_k^r(y_{k-1})}{\hat{O}_k^r}.$$

D.3 Proof of equation (6.27)

To perform a change from $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)^\top \in \mathbb{R}^N$ to $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_N)^\top \in \mathbb{R}^N$, we define the Radon-Nikodým derivative

$$\left. \frac{dP_{\hat{\theta}}}{dP_{\theta}} \right|_{\mathcal{Y}_k} = \Lambda_k^\sigma = \prod_{l=1}^k \lambda_l(\hat{\sigma}_{l-1}, y_l),$$

where

$$\lambda_l(\hat{\sigma}_{l-1}, y_l) = \frac{\sigma(\mathbf{x}_{l-1})}{\hat{\sigma}(\mathbf{x}_{l-1})} \exp \left\{ \frac{(y_l - \alpha(\mathbf{x}_{l-1})y_{l-1} - \eta(\mathbf{x}_{l-1}))^2}{2\sigma(\mathbf{x}_{l-1})^2} - \frac{(y_l - \alpha(\mathbf{x}_{l-1})y_{l-1} - \eta(\mathbf{x}_{l-1}))^2}{2\hat{\sigma}(\mathbf{x}_{l-1})^2} \right\}.$$

Hence,

$$\begin{aligned}
E \left[\log \left(\frac{dP_{\hat{\theta}}}{dP_{\theta}} \right) \middle| \mathcal{Y}_k \right] &= E \left[\sum_{l=1}^k -\log \hat{\sigma}(\mathbf{x}_{l-1}) - \frac{[y_l - \alpha(\mathbf{x}_{l-1})y_{l-1} - \eta(\mathbf{x}_{l-1})]^2}{2\hat{\sigma}(\mathbf{x}_{l-1})^2} \middle| \mathcal{Y}_k \right] + \mathbf{R} \\
&= E \left[-\sum_{l=1}^k \sum_{r=1}^N \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \left(\log \hat{\sigma}_r + \frac{1}{2\hat{\sigma}_r^2} (y_l^2 + \alpha_r^2 y_{l-1}^2 + \eta_r^2 O_k^r \right. \right. \\
&\quad \left. \left. - 2\alpha_r y_{l-1} y_l - 2\eta_r y_l + 2\eta_r \alpha_r y_{l-1}) \right) \middle| \mathcal{Y}_k \right] + \mathbf{R}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{r=1}^N \left(-\log \hat{\sigma}_r \hat{\mathcal{O}}_k^r - \frac{1}{2\hat{\sigma}_r^2} (\hat{T}_k^r(y_k^2) + \alpha_r^2 \hat{T}_k^r(y_{k-1}^2) + \eta_r^2 \hat{\mathcal{O}}_k^r \right. \\
&\quad \left. - 2\alpha_r \hat{T}_k^r(y_{k-1}y_k) - 2\eta_r \hat{T}_k^r(y_k) + 2\eta_r \alpha_r \hat{T}_k^r(y_{k-1})) \right) + \mathbf{R}, \tag{D.3}
\end{aligned}$$

where \mathbf{R} is independent of $\hat{\sigma}$. We differentiate the above expression in $\hat{\sigma}_r$ and equate the result to zero. Solving the equation we get the optimal choice of $\hat{\sigma}^2$, which is

$$\hat{\sigma}_r^2 = \frac{\hat{T}_k^r(y_k^2) + \alpha_r^2 \hat{T}_k^r(y_{k-1}^2) + \eta_r^2 \hat{\mathcal{O}}_k^r - 2\alpha_r \hat{T}_k^r(y_k y_{k-1}) - 2\eta_r \hat{T}_k^r(y_k) + 2\eta_r \alpha_r \hat{T}_k^r(y_{k-1})}{\hat{\mathcal{O}}_k^r}$$

and this agrees with equation (6.27).

Appendix E

Proof of equation (6.36)

We use mathematical induction to prove (6.36) for $h \geq 3$. Following equations (6.34) and (6.35), we have, when $h = 3$,

$$\begin{aligned}
 E[y_{k+3}|\mathcal{Y}_k] &= \langle \boldsymbol{\alpha}, \mathbf{p}_{k+2} \rangle E[y_{k+2}|\mathcal{Y}_k] + \langle \boldsymbol{\eta}, \mathbf{p}_{k+2} \rangle \\
 &= \langle \boldsymbol{\alpha}, \mathbf{p}_{k+2} \rangle (\langle \boldsymbol{\alpha}, \mathbf{p}_{k+1} \rangle E[y_{k+1}|\mathcal{Y}_k] + \langle \boldsymbol{\eta}, \mathbf{p}_{k+1} \rangle) + \langle \boldsymbol{\eta}, \mathbf{p}_{k+2} \rangle \\
 &= \prod_{i=1}^2 \langle \boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\Pi}^{i-1} \mathbf{p}_k \rangle (\langle \boldsymbol{\alpha}, \hat{\mathbf{x}}_k \rangle y_k + \langle \boldsymbol{\eta}, \hat{\mathbf{x}}_k \rangle) + \langle \boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\Pi} \mathbf{p}_k \rangle \langle \boldsymbol{\eta}, \mathbf{A} \mathbf{p}_k \rangle + \langle \boldsymbol{\eta}, \mathbf{A}\boldsymbol{\Pi} \mathbf{p}_k \rangle.
 \end{aligned}$$

Therefore, the statement is true for $h = 3$. Assume the statement is true for $h = m$, i.e.,

$$\begin{aligned}
 E[y_{k+m}|\mathcal{Y}_k] &= \prod_{i=1}^{m-1} \langle \boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\Pi}^{i-1} \mathbf{p}_k \rangle (\langle \boldsymbol{\alpha}, \hat{\mathbf{x}}_k \rangle y_k + \langle \boldsymbol{\eta}, \hat{\mathbf{x}}_k \rangle) \\
 &\quad + \sum_{i=1}^{m-2} \prod_{j=i}^{m-2} \langle \boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\Pi}^j \mathbf{p}_k \rangle \langle \boldsymbol{\eta}, \mathbf{A}\boldsymbol{\Pi}^{i-1} \mathbf{p}_k \rangle + \langle \boldsymbol{\eta}, \mathbf{A}\boldsymbol{\Pi}^{m-2} \mathbf{p}_k \rangle.
 \end{aligned}$$

We demonstrate that equation (6.36) is true for $h = m + 1$.

$$\begin{aligned}
 E[y_{k+m+1}|\mathcal{Y}_k] &= \langle \boldsymbol{\alpha}, \mathbf{p}_{k+m} \rangle E[y_{k+m}|\mathcal{Y}_k] + \langle \boldsymbol{\eta}, \mathbf{p}_{k+m} \rangle \\
 &= \langle \boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\Pi}^{m-1} \mathbf{p}_k \rangle \prod_{i=1}^{m-1} \langle \boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\Pi}^{i-1} \mathbf{p}_k \rangle (\langle \boldsymbol{\alpha}, \hat{\mathbf{x}}_k \rangle y_k + \langle \boldsymbol{\eta}, \hat{\mathbf{x}}_k \rangle)
 \end{aligned}$$

$$\begin{aligned}
& + \langle \alpha, \mathbf{A}\Pi^{m-1}\mathbf{p}_k \rangle \sum_{i=1}^{m-2} \prod_{j=i}^{m-2} \langle \alpha, \mathbf{A}\Pi^j\mathbf{p}_k \rangle \langle \eta, \mathbf{A}\Pi^{i-1}\mathbf{p}_k \rangle \\
& + \langle \alpha, \mathbf{A}\Pi^{m-1}\mathbf{p}_k \rangle \langle \eta, \mathbf{A}\Pi^{m-2}\mathbf{p}_k \rangle + \langle \eta, \mathbf{A}\Pi^{m-1}\mathbf{p}_k \rangle \\
= & \prod_{i=1}^m \langle \alpha, \mathbf{A}\Pi^{i-1}\mathbf{p}_k \rangle (\langle \alpha, \hat{\mathbf{x}}_k \rangle y_k + \langle \eta, \hat{\mathbf{x}}_k \rangle) + \sum_{i=1}^{m-2} \prod_{j=i}^{m-1} \langle \alpha, \mathbf{A}\Pi^j\mathbf{p}_k \rangle \langle \eta, \mathbf{A}\Pi^{i-1}\mathbf{p}_k \rangle \\
& + \langle \alpha, \mathbf{A}\Pi^{m-1}\mathbf{p}_k \rangle \langle \eta, \mathbf{A}\Pi^{m-2}\mathbf{p}_k \rangle + \langle \eta, \mathbf{A}\Pi^{m-1}\mathbf{p}_k \rangle \\
= & \prod_{i=1}^m \langle \alpha, \mathbf{A}\Pi^{i-1}\mathbf{p}_k \rangle (\langle \alpha, \hat{\mathbf{x}}_k \rangle y_k + \langle \eta, \hat{\mathbf{x}}_k \rangle) \\
& + \sum_{i=1}^{m-1} \prod_{j=i}^{m-1} \langle \alpha, \mathbf{A}\Pi^j\mathbf{p}_k \rangle \langle \eta, \mathbf{A}\Pi^{i-1}\mathbf{p}_k \rangle + \langle \eta, \mathbf{A}\Pi^{m-1}\mathbf{p}_k \rangle.
\end{aligned}$$

Therefore, by the principle of mathematical induction, the statement in equation (6.36) is true for $h \geq 3$.

Curriculum Vitae

Name: Xiaojing Xi

Post-Secondary Education and Degrees: Wilfrid Laurier University
Waterloo, On, CA
2003 - 2006 B.A.

Wilfrid Laurier University
Waterloo, On, CA
2006 - 2008 M.Sc.

University of Western Ontario
London, On, CA
2008 - 2012 Ph.D

Honours and Awards: Western Graduate Research Scholarship
2008-2012

Related Work Experience: Teaching Assistant
Wilfrid Laurier University
2006 - 2008

Teaching Assistant
University of Western Ontario
2008 - 2012

Publications:

- X. Xi, M. Rodrigo and R.S. Mamon, Parameter estimation of a regime-switching model using an inverse Stieltjes moment approach, In: *Advances in Statistics, Probability and Actuarial Science – Festschrift volume in honour of Robert Elliotts 70th Birthday*, S. Cohen, D. Madan, T. Siu and H. Yang (editors), World Scientific Publishing. pp. TBA. Accepted August 2011.

- X. Xi and R.S. Mamon, Parameter estimation of an asset price model driven by a weak hidden Markov chain, *Economic Modelling*, 28:36-46, 2011.

Submitted/ Under Revision Papers:

- X. Xi, R.S. Mamon and M. Davison, Weak hidden Markov chain-modulated model for asset allocation, under revision in *Journal of Mathematical Modelling and Algorithms*, 2012.
- X. Xi and R.S. Mamon, An interest rate model incorporating memory and regime-switching, Submitted to *Statistics and Computing*, 2012.
- X. Xi and R.S. Mamon, Yield curve modeling using a multivariate higher-order HMM, Submitted to *State-Space Models and Applications in Economics and Finance*, 2012.