

Electronic Thesis and Dissertation Repository

2-8-2018 10:00 AM

Some applications of higher-order hidden Markov models in the exotic commodity markets

Heng Xiong, *The University of Western Ontario*

Supervisor: Mamon, Rogemar., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

© Heng Xiong 2018

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



Part of the [Applied Statistics Commons](#), [Finance and Financial Management Commons](#), [Management Sciences and Quantitative Methods Commons](#), [Numerical Analysis and Scientific Computing Commons](#), and the [Statistical Models Commons](#)

Recommended Citation

Xiong, Heng, "Some applications of higher-order hidden Markov models in the exotic commodity markets" (2018). *Electronic Thesis and Dissertation Repository*. 5226.
<https://ir.lib.uwo.ca/etd/5226>

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.

Abstract

The liberalisation of regional and global commodity markets over the last several decades resulted in certain commodity price behaviours that require new modelling and estimation approaches. Such new approaches have important implications to the valuation and utilisation of commodity derivatives. Derivatives are becoming increasingly crucial for market participants in hedging their exposure to volatile price swings and in managing risks associated with derivative trading. The modelling of commodity-based variables is an integral part of risk management and optimal-investment strategies for commodity-linked portfolios. The characteristics of commodity price evolution cannot be captured sufficiently by one-state driven models even with the inclusion of multiple factors. This inspires the adoption of regime-switching methods to rectify the one-state multi-factor modelling inadequacies. In this research, we aim to employ higher-order hidden Markov models (HOHMMs) in order to take advantage of the latent information in the observed process recorded in the past. This hugely enhances and complements the regime-switching features of our approach in describing certain variables that virtually determine the value of some commodity derivatives such as contracts dependent on temperature, electricity spot price, and fish-price dynamics. Our push for the utility of the change-of-probability-measure technique facilitates the derivation of recursive filtering algorithms. This then establishes a self-tuning dynamic estimation procedure. Both the data-fitting and forecasting performances of various model settings are investigated.

This research work emerged from four related projects detailed as follows. (i) We start with an HMM to model the behaviour of daily average temperatures (DATs) geared towards the analysis of weather derivatives. (ii) The model in (i) is extended naturally by showcasing the capacity of an HOHMM-based approach to simultaneously describe the DATs' salient properties of mean reversion, seasonality, memory and stochasticity. (iii) An HOHMM-driven jump process augments the HOHMM-based de-seasonalised temperature process to capture price spikes, and the ensuing filtering algorithms under this modelling framework are constructed to provide optimal parameter estimates. (iv) Finally, a multi-dimensional HOHMM-modulated set up is built for futures price-curve dynamics pertinent to financial product valuation and risk management in the aquaculture sector. We examine the performance of this new modelling set up by considering goodness-of-fit and out-of-sample forecasting metrics with a detailed numerical demonstration using a multivariate

data set compiled by the Fish Pool ASA.

This research offers a collection of more flexible stochastic modelling approaches for pricing and risk analysis of certain commodity derivatives on weather, electricity and fish prices. The novelty of our techniques is the powerful capability to automate the parameter estimation. Consequently, we contribute to the development of financial tools that aid in selecting the appropriate and optimal model on the basis of some information criteria and within current technological advancements in which continuous flow of observed data are now readily accessible in real time.

Keywords: regime-switching model, commodity-derivatives valuation, change of reference probability measure, optimal parameter estimation, multivariate HOHMM filtering method

Co-Authorship Statement

I hereby declare that this thesis incorporates materials that are direct results of my main efforts during my doctoral study. All research outputs (jointly authored with Dr Rogemar Mamon) led to two published papers and two manuscripts under review in refereed journals, and these are detailed below.

The content of chapter 2 appeared as a full paper in the journal *Computational Management Science*; see reference [44] in chapter 4.

The results of chapter 3 were published in the *Journal of Computational Science*; see reference [37] in chapter 4.

Chapter 4 is based on a manuscript that is currently under review in the journal *Applied Energy*.

The source of chapter 5 is a manuscript under consideration for publication in the *Journal of Economic Dynamics and Control*.

An integrated-article format of this dissertation is adopted in accordance with the Western's thesis guidelines. As a lead author of all self-contained chapters, I am responsible for probing the feasibility of the proposed modelling frameworks, the implementation of filtering algorithms, and the completion of the manuscripts. Dr Mamon guided me conscientiously on the research plan, proposed general modelling development and formulation, interpretation involved in the empirical analysis of all aforementioned four projects, and in addressing competently the referees' concerns for successful journal publications.

I certify that this thesis is fully a product of my own work. It was conducted from September 2014 to present under the supervision of Dr Mamon at the University of Western Ontario.

London, Ontario

To my Dad in Heaven
"I miss you, my first & forever hero."

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Dr Rogemar Mamon for his innumerably valuable guidance on my graduate study. This doctoral research enjoys steady progress and is building momentum for potential impact due to his constructive comments and excellent advice. More importantly, his infinite patience, thorough knowledge and deep insights have carried me through some toughest times of my academic life since day one at Western.

I would also like to sincerely thank Dr Matt Davison for his invaluable help and support. He is always generous with his time to discuss some of my academic and professional issues. I am extremely appreciative of his well-thought out suggestions and tenacious encouragement throughout my graduate studies.

I gratefully acknowledge the useful feedback, precious time, and energy of my other thesis committee members: Dr Marcos Escobar-Anel, Dr Lars Stentoft, and Dr Francois Watier. Their helpful comments and suggestions certainly strengthen this dissertation at various levels. Furthermore, many thanks are due to the outstanding faculty and staff of the Department of Statistical and Actuarial Sciences (DSAS) for their support in various ways. I also sincerely thank the financial support provided by the DSAS and the Ontario Graduate Scholarship/Queen Elizabeth II Scholarship Program.

I take this chance to thank all my colleagues and friends in Canada. I am so blessed to be in your great company, affording moral support in the accomplishment of this lifetime milestone.

Last but not least, I would like to express my special thanks to my family, most especially my parents and grandma, for their selfless love, unconditional support, and unfailing belief on my ability to reach greater heights and possibilities. I will do my best to be a better man and make you all proud.

Contents

Abstract	i
Co-Authorship Statement	iii
Dedication	iv
Acknowledgements	v
List of Figures	x
List of Tables	xiii
List of Appendices	xv
1 Introduction	1
1.1 Research motivation and objectives	1
1.2 Literature review	2
1.2.1 Overview of HMMs	2
1.2.2 Extension to HOHMMs	5
1.2.3 Evaluation with forward-backward algorithm	6
1.2.4 Decoding with Viterbi algorithm	7
1.2.5 Training with expectation-maximisation algorithm	8
1.2.6 Change of measure method in HOHMMs	11
1.3 Structure of the thesis	12
1.3.1 Putting a price tag on temperature	13
1.3.2 A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics	13
1.3.3 A higher-order Markov chain-modulated model for electricity spot- price dynamics	14

1.3.4	Modelling and forecasting futures-prices curves in the Fish Pool market	14
	References	15
2	Putting a price tag on temperature	18
2.1	Introduction	18
2.2	Model description	20
2.2.1	Model for temperature derivatives	20
2.2.2	The HMM-modulated OU process	21
2.3	Recursive filtering	23
2.3.1	Change of reference probability measure	23
2.3.2	Calculation of recursive filters	24
2.4	Optimal parameter estimation	26
2.5	Numerical implementation	28
2.5.1	The deterministic component	28
2.5.2	The stochastic component	29
2.5.2.1	Initial values for the parameter estimates	31
2.5.2.2	Evolution of parameter estimates	31
2.5.3	Analysis of regime and model selection	32
2.5.3.1	Assessment of predicted temperature values	36
2.5.3.2	Likelihood-based model selection and error analysis	39
2.6	Application to the pricing of temperature-dependent contracts	41
2.6.1	Temperature-based derivatives	42
2.6.2	Pricing of temperature-based index futures and options	43
2.6.3	The price of risk and risk-neutral measure Q	44
2.6.4	Pricing of a temperature-based contract	48
2.6.5	Sensitivity analysis of temperature-based option price	50
2.7	Conclusion	57
	References	58
3	A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics	63
3.1	Introduction	63
3.2	Model description	67
3.2.1	Model for temperature derivatives	67
3.2.2	Ornstein-Uhlenbeck (OU) process in the temperature model	68

3.2.2.1	HMM-modulated OU process	68
3.2.2.2	HOHMM-modulated Ornstein-Uhlenbeck process	69
3.3	Recursive filtering	70
3.3.1	Reference probability measure	70
3.3.2	Calculation of recursive filters	72
3.4	Optimal parameter estimation	75
3.5	Numerical implementation	77
3.5.1	Analysis of the deterministic component	78
3.5.2	Analysis of the stochastic component	79
3.5.3	Validity of model evaluation	80
3.5.4	Implementation aspects of the HOHMM-OU filtering	81
3.5.4.1	Initial values for the parameter estimation	82
3.5.4.2	Evolution of parameter estimates and comparison under HMM and HOHMM settings	83
3.5.5	Model selection and other diagnostics	88
3.5.5.1	Assessment of predicted temperatures	88
3.5.5.2	Error analysis and model selection	91
3.6	Conclusion	95
	References	96

4	A higher-order Markov chain-modulated model for electricity spot-price dynamics	101
4.1	Introduction	101
4.2	Model formulation	105
4.2.1	Model for electricity spot price	105
4.2.2	The fusion of HOHMM, OU and compound Poisson models	106
4.3	Recursive filtering	109
4.3.1	Change of reference probability measure	109
4.3.2	Calculation of recursive filters	110
4.4	Optimal parameter estimation	113
4.5	Numerical application	117
4.5.1	Analysis of the deterministic component	117
4.5.2	Application of filtering algorithms for the HOHMM-OU process with jumps	119
4.5.2.1	Data processing with the filtering algorithm	119

4.5.2.2	Implementing the filtering procedure	121
4.5.3	Discussion of model performance	127
4.5.3.1	Forecasting and error analysis	127
4.5.3.2	Selection of suitable model setting	132
4.5.3.3	The valuation of expected future spot at delivery	135
4.6	Conclusion	138
	References	138
5	Modelling and forecasting futures-prices curves in the Fish Pool market	144
5.1	Introduction	144
5.2	Model setup	148
5.3	Filtering and parameter estimation	152
5.4	Numerical application	156
5.4.1	Fish pool exchange and data description	156
5.4.2	Implementation of filters and estimation	160
5.4.3	Model performance and selection	162
5.5	Conclusion	171
	References	172
6	Conclusion	180
6.1	Summary of research contributions	180
6.2	Further research directions	182
	References	183
Appendix A Derivations of the model's optimal parameter estimates in chapter 2		185
Appendix B Proofs of Propositions in chapter 2		189
Appendix C Derivations of the model's optimal parameter estimates in chapter 3		192
Appendix D Derivations of the model's optimal parameter estimates in chapter 4		196
Appendix E Derivation of the model's optimal parameter estimates in chapter 5		204
Appendix F Derivation of the d-step ahead forecasting formula in chapter 5		207
Curriculum Vitae		209

List of Figures

1.1	Reference measure optimal filter derivation for HOHMMs	12
2.1	Fitted seasonal component and actual observations	30
2.2	Deseasonalised stochastic component X_t	30
2.3	Evolution of parameter estimates for θ , α , and ξ under a 2-state HMM-based model	33
2.4	Evolution of transition probability under a 2-state HMM-based model	34
2.5	Evolution of parameter estimates for θ , α , and ξ under a 3-state HMM-based model	35
2.6	Evolution of transition probability under a 3-state HMM-based model	36
2.7	One-step ahead forecasts under a 2-state HMM-based model	37
2.8	Comparison of the expected HDD and actual HDD in a 2-state model	38
2.9	Comparison of one step ahead forecasts in 1-, 2-, and 3-state HMMs	38
2.10	Monthly call option price for varying strike HDD	51
2.11	HDD call option price for varying the market price of risk with initial state 1	51
2.12	HDD call option price for varying θ_1^*	54
2.13	HDD call option price for varying θ_2^*	55
2.14	HDD call option price for varying α_1	55
2.15	HDD call option price for varying α_2	56
2.16	HDD call option price for varying ξ_1	56
2.17	HDD call option price for varying ξ_2	57
3.1	Fitted seasonal component versus actual observations	79
3.2	Deseasonalised stochastic component X_t	80
3.3	Evolution of transition probability under a 2-state HOHMM-based model	83
3.4	Evolution of parameter estimates for κ , ϑ , and ϱ under a 2-state HOHMM-based model	85
3.5	Evolution of parameter estimates for κ , ϑ , and ϱ under a 3-state HOHMM-based model	86

3.6	Evolution of parameter estimates for κ , ϑ , and ρ under a 2-state HOHMM-based model with 95% confidence level	87
3.7	One-step ahead forecasts under a 3-state HOHMM-based model	89
3.8	Comparison of the expected HDD and actual HDD in a 3-state HOHMM-based model	90
3.9	Comparison of one-step-ahead forecasts in 1-, 2-, and 3-state HOHMM-based models	90
3.10	Evolution of AICs for the 1-, 2-, and 3-state HMM- and HOHMM-based models	94
4.1	Fitted seasonal component and actual observations	119
4.2	Zoom-in view of fitted seasonal component versus actual observed prices (Jan/01/2014 - Dec/31/2014)	120
4.3	Deseasonalised stochastic component of DESP	120
4.4	Evolution of parameter estimates for α , θ , and ξ under a 2-state HOHMM	123
4.5	Evolution of parameter estimates for μ_β , and σ_β	124
4.6	Evolution of parameter estimates for α , θ , and ξ under a 3-state HOHMM-based model	125
4.7	Evolution of parameter estimates for μ_β , and σ_β	126
4.8	One-step ahead forecasts for X_k and S_k under a 3-state HOHMM	130
4.9	Comparison of one-step ahead forecasts in 1-, 2-, 3-state HMM- and HOHMM-based models	131
4.10	Evolution of BIC values under the 1-, 2-, 3-state HMM- and HOHMM-based models	135
4.11	Expected future spot price on delivery (EFSP) under a 2-state HOHMM	137
5.1	Evolution of transition probabilities under a 2-state HOHMM for futures prices with expiry T^h	162
5.2	Evolution of transition probabilities under a 3-state HOHMM for prices of futures with expiry T^h	163
5.3	Evolution of parameter estimates for ξ , λ , ν , and σ under a 2-state HOHMM with expiry on 30 Dec 2016.	164
5.4	Evolution of parameter estimates for ξ , λ , ν , and σ under a 3-state HOHMM with expiry on 30 Dec 2016.	165
5.5	One-step ahead predictions for futures prices with expiries on 28 Oct 2016, 29 Nov 2016, and 29 Dec 2016	173

5.6	AIC, AICc and BIC for the 1-, 2-, 3-state HMM/HOHMM-based model for salmon futures prices	174
5.7	One-step ahead and out-of-sample predictions under the 3-state HOHMM for futures prices with expiries on 31 Jan 2017, 28 Feb 2017, and 31 Mar 2017	175

List of Tables

2.1	Descriptive statistics for daily average temperature(DAT)	29
2.2	Parameter estimates for the seasonality-component	29
2.3	Interval of standard errors for parameter estimates under 1-, 2-, 3-state HMM- and HOHMM-based models	34
2.4	Error analysis of HMM-based models	39
2.5	Bonferroni-adjusted p -values of a t -test in terms of RMSEs under HMM settings	40
2.6	Number of parameters under HMM settings	41
2.7	AIC-based selection analysis	41
2.8	Specifications of typical CME temperature-based contracts	42
2.9	Elements of a standardised temperature-based contract	43
2.10	Optimal parameter estimates for the 2-state HMM model with and without the market price of risk	50
2.11	HDD call option price for varying intensities p^c	54
3.1	Descriptive statistics for the DATs data set	78
3.2	Parameter estimates for the seasonality-component	78
3.3	Parameter estimates for the seasonality-component	81
3.4	Interval of standard errors for the parameter estimates under 1-, 2-, 3-state HMM- and HOHMM-based models	88
3.5	Results of error analysis on HMM-based models	92
3.6	Results of error analysis on HOHMM-based models	92
3.7	Bonferroni-corrected p -values for a paired t -test on RMSEs under HMM and HOHMM settings	92
3.8	Number of estimated parameters under the HMM and HOHMM settings	93
3.9	Results of error analysis for the HDD on HMM- and HOHMM-based models	94
4.1	Descriptive statistics for daily electricity spot price (DESP)	118
4.2	Parameter estimates for the seasonal component	118

4.3	Interval of standard errors for parameter estimates under 1-, 2-, 3-state HMM- and HOHMM-based models	128
4.4	Error analysis of HMM- and HOHMM-based models	132
4.5	Bonferroni-corrected p -values for the t -test performed on the RMSEs involving the DSP	132
4.6	Number of estimated parameters under HMM-OU with jumps and HOHMMs-OU with jumps	133
4.7	Comparison of selection criteria AIC	134
4.8	Optimal parameter estimates for the 2-state HOHMM	136
5.1	The data periods of future contracts (maturities of 1–6 months) covered by the moving window in the filtering procedure	158
5.2	Futures contracts with maturity up to 6 months and expiration on 29 Jan 2016	158
5.3	Descriptive statistics for log-futures prices (maturities of 1 – 6 months) . . .	159
5.4	Error analysis of HMM- and HOHMM-based models under 1-, 2-, and 3-state settings for salmon futures prices	167
5.5	Bonferroni-corrected p -values for the paired t -test performed on the RMSEs involving salmon futures prices	168
5.6	Number of estimated parameters under HMM and HOHMM settings for salmon futures prices	169
5.7	Out-of-sample error analysis of HMM- and HOHMM-based models under 1-, 2-, and 3- state settings for futures prices	171

List of Appendices

Appendix A: Derivations of the model's optimal parameter estimates in chapter 2 . .	184
Appendix B: Proofs of Propositions in chapter 2	188
Appendix C: Derivations of the model's optimal parameter estimates in chapter 3 . .	191
Appendix D: Derivations of the model's optimal parameter estimates in chapter 4 . .	195
Appendix E: Derivation of the model's optimal parameter estimates in chapter 5 . . .	203
Appendix F: Derivation of the expected value of d -step ahead forecasts in chapter 5 .	206

Chapter 1

Introduction

1.1 Research motivation and objectives

Amongst the most pressing concerns of today's contemporary society include climate change, worldwide political uncertainty, resiliency of power resource generation, food sustainability, and development of new or revised regulatory frameworks. Undoubtedly, these issues have impact on the financial, energy, agriculture, and aquaculture sectors. This research work delves on the development of mathematical models that support the creation of financial solutions, in the above-mentioned industries, to mitigate, hedge and transfer relevant risks closely tied to potential huge economic losses arising from such societal pressing issues.

The financial solutions refer to exotic commodity-based derivatives whose values are dependent on weather measurements, and prices in the electricity and fish markets. Taking suitable positions comprising the correct number of contracts in derivatives trading forms strategies in managing risk exposure to certain pertinent factors. Research studies in the aforementioned markets are based on one-state models. However, they might not accurately capture the stylised behaviours of price evolution, especially during periods of financial crises or when positive news affect market sentiments that drive prices suddenly. These situations require models with more capabilities to account for abrupt fluctuations of various statistics, and one simple but powerful way is the embedding of a regime-switching approach enriched by a memory-capturing mechanism.

In particular, to meet the goals of mathematical modelling developments in the context of considerations as described above, the following will be carried out: (i) Construct

an Ornstein-Uhlenbeck (OU) modelling framework modulated by hidden Markov models (HMM) and higher-order hidden Markov models (HOHMM) to adequately delineate important characteristics of market primitives. (ii) Add a compound-Poisson process into the HOHMM-OU setting to pick up the spike dynamics in the observed time series. (iii) Extend the HOHMM setting into a multi-dimensional framework tailored for multivariate data series. (iv) Develop self-calibrating algorithms and dynamic estimation procedures for the proposed models. (v) Derive recursive filters for quantities that are functions of both HMM and HOHMM, and use such filters to obtain models' optimal parameter estimates. (vi) Assess and compare performances under both HMM and HOHMM settings using error analyses and likelihood-based information criteria.

1.2 Literature review

In this section, we will first introduce the theory and assumptions of HMMs, and then extend the concepts to HOHMMs. Besides expanding the HMM analytical framework, we will review four fundamental results at an intuitive level for the HOHMM construction and filtering relevant to the objectives of our research. They consist of (i) evaluation (ii) decoding (iii) training, and (iv) change of measure method. The first three issues are notable problems and thoroughly addressed within HMMs; see [1], [19], and [3], whilst we will examine them under the HOHMM setting by following [2]. A summary of the change of probability measure technique will be given to reduce an HOHMM to an equivalent first-order representation for our filtering processes.

1.2.1 Overview of HMMs

Since initially introduced in the late 1960s, HMMs have been widely utilised as a ubiquitous instrument in many fields of engineering and physical sciences, such as speech recognition, data compression, and molecular physics. The comprehensive mathematical structure of HMMs constructs a solid theoretical foundation and provides a great performance in practice. The study of applying this tool in the financial field was pioneered by Hamilton [8]. There have been systematic methodologies and extensive applications of HMMs developed to capture variables in finance and economy since then.

The general mechanism of HMMs in finance can be concisely explained by the signal theory, well-known in engineering. Outputs produced from any processes in both theory and practice normally are considered as signals. In finance, we then can treat collected data as observed signals, such as prices and indices of underlying assets. However, these observations might contain noise or be distorted by various errors from their generating procedures. The estimation and measurements for real sources might be corrupted. It is then not possible to reveal the hidden state of the model only through the observed symbols. Fortunately, the application of HMMs is capable of filtering the noise out and unravelling the distortion. The hidden variables and information in our observations then can be efficiently estimated and further investigated.

Following Cappé et al. [3], an HMM comprises a bivariate discrete process $\{X_t, O_t\}$ with $t \in \mathbb{Z}_0^+$. Let $\{X_t\}$ be a Markov chain encapsulated in an observed process $\{O_t\}$, which is a corrupted version of $\{X_t\}$ due to some noise. Even though the latent process $\{X_t\}$ is hidden to the outside, it can be observable through $\{O_t\}$ under an HMM. As in Rabiner [29], a set of parameters $(N, n, \mathbf{\Pi}, \mathbf{\Lambda}, \mathbf{P})$ needs to be introduced in order to define an HMM completely.

- N , the number of hidden states in the model. Even though they are latent, physical significance to the states can be detected for many practical applications. We denote the state in the model as q_t at time t , and individual states as

$$S_t = \{s_1, s_2, \dots, s_N\}$$

- n , the number of distinct observation signals per hidden state at time t . For instance, “rainy day” and “non-rainy day” can be treated as two distinct observed symbols. We denote the signal as “ Y_t ” at time t . The individual signals are presented by

$$Y_t = \{y_1, y_2, \dots, y_n\}$$

- $\mathbf{\Pi} = \{\pi_{ji}\}$ stands for the state transition probability distribution that governs transitions among states, where

$$\pi_{ji} = P(q_{t+1} = s_i | q_t = s_j), \quad 1 \leq i, j \leq N,$$

and $\sum_{i=1}^N \pi_{ji} = 1, \pi_{ji} \geq 0$.

- $\mathbf{\Lambda} = \{\Lambda_j\}$ denotes the initial state distribution, where

$$\Lambda_j = P(q_1 = s_j), \quad 1 \leq j \leq N.$$

- $\mathbf{P} = \{p_i(m)\}$ denotes the probability distribution of observed symbols in hidden state i , where

$$p_i(m) = P(O_t = y_m \mid q_t = s_i), \quad 1 \leq i \leq N, 1 \leq m \leq n,$$

and $\sum_{m=1}^n p_i(m) = 1, \{p_i(m) \geq 0\}$.

Before implementing HMMs in generating observations, we need to explicitly state three important assumptions in the theory of HMMs [29] as follows:

- Assumption 1: Markov assumption

$$P(q_{t+1} = s_i \mid q_t = s_j, q_{t-1} = s_l, \dots, q_0 = s_h) = P(q_{t+1} = s_i \mid q_t = s_j) = \pi_{ji}.$$

We assume the Markov property holds. As shown in its definition, the current state s_t is independent of all the states prior to $t - 1$, given s_{t-1} . The observed symbols meets the Markov assumption as well with respect to their corresponding states.

- Assumption 2: Stationarity assumption

$$P(q_{t+1} = s_i \mid q_t = s_j) = P(q_{k+1} = s_i \mid q_k = s_j), \quad \text{for } 1 \leq i, j \leq N \text{ and } \forall k, t \in \mathbb{Z}_0^+.$$

The intuitive explanation for this assumption is that the state transition distribution over next state given current state does not change over time.

- Assumption 3: Output independence assumption

$$P(O \mid q_1, q_2, \dots, q_T, N, n, \mathbf{\Pi}, \mathbf{\Lambda}, \mathbf{P}) = \prod_{t=1}^T P(O_t \mid q_t, N, n, \mathbf{\Pi}, \mathbf{\Lambda}, \mathbf{P}).$$

Let $O = O_1, O_2, \dots, O_T$ be an observation sequence, in which T denotes the total number of observations. Every observation $O_t, 1 \leq t \leq T$, is one of the signals from Y_t . This assumption indicates that the current output observations are statistically independent of previous ones.

Once proper values are assigned to the set of parameters $(N, n, \mathbf{\Pi}, \mathbf{\Lambda}, \mathbf{P})$ under the aforementioned assumptions, the HMM then can be utilised to produce an observation sequence O , the general procedures of which include the following steps. We start with setting an initial state in terms of $\mathbf{\Lambda}$ and then choose $O_t = y_m$ according to $p_i(m)$. π_{ji} will be applied when transiting to another state $q_{t+1} = s_i$. By taking increment by 1 for t till T , we shall produce O through an appropriate HMM.

1.2.2 Extension to HOHMMs

HMM-based approaches and strategies are in widespread use with its structural simplicity and straightforward computation. In view of the definitions in Section 1.2.1, an HMM formulation is generally capable of resolving first-order transition among hidden states. However, it might be inadequate for some real-world applications. For instance, the limited validities in assumptions basically indicates that only the present information affects what you are about to express next in speech recognition, one of the most extensive fields of HMMs application. It seems improper since long-term dependence might be necessary to retain a robust speech structure. As presented in Juang and Rabiner [9], HMMs are not sufficient under higher levels of a recognition system, including syntactic or semantic processing.

Various literatures also find out the longer memory property in many financial series, such as the long term dependence in the volatility of S&P 500 returns [12] and in percentage changes on Treasury debt security yields [13]. For this sort of financial series, a growing body of evidence suggests that a regular HMM cannot effectively capture their stylised behaviours ([17] and [21]), thereby raising a need to expand the HMM literature. It results in some scholars examining applications of HOHMMs in the financial field. For instance, Siu et al. proposed a higher-order Markov-switching model with the drift and the volatility modulated by a discrete-time high-order Markov chain for measuring risk of a risky portfolio [28]. Xi et al. introduced an analysis of asset allocation strategies under both HMM and HOHMM settings, and concluded that the HOHMM-based approach outperforms the HMM-based strategy for certain levels of transaction costs [20].

As an extension of a normal Markov model, a higher-order Markov model builds longer memories into a Markov process by specifying that a current state depends on other previous states, not only on the previous one. Analogically, we can find out an HOHMM involves some features presented in Section 1.2.1 with two major differences. The state transition probability distribution becomes

$$\pi_{i_{t-k+1}, \dots, i_{t+1}} = P(q_{t+1} = s_{i_{t+1}} | q_t = s_{i_t}, \dots, q_{t-k+1} = s_{i_{t-k+1}}), \text{ for } 1 \leq k \leq t \leq T, \quad (1.1)$$

and the initial k states distribution is defined as

$$\pi_{i_1, \dots, i_k} = P(q_1 = s_{i_1}, \dots, q_k = s_{i_k}). \quad (1.2)$$

Equations (1.1) and (1.2) increase the number of parameters exponentially in terms of the order of the HOHMM setting. They lead to more complicated processes to address the three classic problems originally in HMMs, the results of which will be elaborated in the following sections.

1.2.3 Evaluation with forward-backward algorithm

Given an observation sequence $O = O_1, \dots, O_T$ and a parameter set $\mathcal{S} = (\mathbf{\Pi}, \mathbf{\Lambda}, \mathbf{P})$, we can solve the evaluation problem under the HOHMM setting by finding $P(O | \mathcal{S})$. Even though straightforward probabilistic arguments is one of options to address this issue, it is not feasible in practice considering the computation in order of N^T . Therefore, we adopt the forward-backward algorithm [29] to efficiently compute the probability of the observed outputs. We firstly define a forward variable as follows

$$\alpha_t(i_{t-k+1}, \dots, i_t) = P(O_1, \dots, O_t, q_{t-k+1} = s_{i_{t-k+1}}, \dots, q_t = s_{i_t} | \mathcal{S}), \text{ for } 1 \leq k \leq t \leq T.$$

It is a joint conditional probability that the subsequence of first t observed outcomes is O_1, \dots, O_t and the last m hidden states of an HOHMM at time t are $s_{i_{t-k+1}}, \dots, s_{i_t}$. The forward probability $\alpha_T(i_{T-k+1}, \dots, i_T)$ can then be calculated through the forward algorithm by the following recursion,

$$\begin{aligned} \alpha_{t+1}(i_{t-k+2}, \dots, i_{t+1}) &= \sum_{i_{t-k+1}=1}^N \left(\alpha_t(i_{t-k+1}, \dots, i_t) P(O_{t+1} | \mathcal{S}, q_{t+1} = s_{i_{t+1}}) \cdot \right. \\ &\quad \left. P(q_{t+1} = s_{i_{t+1}} | \mathcal{S}, q_{t-k+1} = s_{i_{t-k+1}}, \dots, q_t = s_{i_t}) \right) \\ &= \sum_{i_{t-k+1}=1}^N \alpha_t(i_{t-k+1}, \dots, i_t) \pi_{i_{t-k+1}, \dots, i_{t+1}} p_{i_{t+1}}(O_{t+1}), \end{aligned} \quad (1.3)$$

where $\alpha_k(i_1, \dots, i_k) = \Lambda_{i_1, \dots, i_k} \cdot \prod_{j=1}^k p_{i_j}(O_j)$, for $1 \leq j \leq k$. The required probability is then obtained by $P(O | \mathcal{S}) = \sum_{i_{T-k+1}, \dots, i_T=1}^N \alpha_T(i_{T-k+1}, \dots, i_T)$ as the sum of all forward variables.

Instead of computing variables forward in time, the backward algorithm proceeds in a similar way but backward in time. The backward variable is defined as

$$\beta_t(i_1, \dots, i_k) = P(O_{t+k}, \dots, O_T | \mathcal{S}, q_t = s_{i_t}, \dots, q_{t+k-1} = s_{i_{t+k-1}}), \text{ for } 1 \leq t \leq T - k.$$

As in the case of the forward algorithm, this joint conditional probability can be obtained through a recursive equation given by

$$\begin{aligned}
\beta_t(i_1, \dots, i_k) &= \sum_{i_{t+k}=1}^N \left(P(O_{t+k+1}, \dots, O_T \mid \mathcal{S}, q_{t+1} = s_{i_{t+1}}, \dots, q_{t+k} = s_{i_{t+k}}) \cdot \right. \\
&\quad \left. P(O_{t+k} \mid \mathcal{S}, q_{t+k} = s_{i_{t+k}}) \cdot P(q_{t+k} = s_{i_{t+k}} \mid \mathcal{S}, q_t = s_{i_t}, \dots, q_{t+k-1} = s_{i_{t+k-1}}) \right) \\
&= \sum_{j=1}^N \beta_{t+1}(i_2, \dots, i_k, j) \pi_{i_2, \dots, i_k, j} P_j(O_{t+k}), \tag{1.4}
\end{aligned}$$

where $\beta_{T-t}(i_1, \dots, i_k) = 1$, for $0 \leq t \leq k-1$, and $1 \leq i_1, \dots, i_k \leq N$. This quantity will be required for the development of the Baum-Welch algorithm [1] to solve the learning problem in Section 1.2.5.

1.2.4 Decoding with Viterbi algorithm

Given a particular HOHMM with the parameter set \mathcal{S} , we attempt to find out the most probable sequence of underlying hidden states, $Q = q_1, q_2, \dots, q_T$, from an observation sequence $O = O_1, O_2, \dots, O_T$. That is equivalent to calculate $\operatorname{argmax}_Q P(Q \mid \mathcal{S}, O)$. Since we have

$$\operatorname{argmax}_Q P(Q \mid \mathcal{S}, O) = \operatorname{argmax}_Q \frac{P(Q, O \mid \mathcal{S})}{P(O \mid \mathcal{S})},$$

and $P(O \mid \mathcal{S})$ does not depend on Q , if $\operatorname{argmax}_Q P(Q, O \mid \mathcal{S})$ is uncovered, we then resolve the decoding problem. As a technique based on dynamic programming method for verifying the existence of the best state sequence, the Viterbi algorithm [19] is an efficient approach to decode the observation sequence O . To facilitate the calculation, we define a variable as

$$\theta_t(i_{t-k+1}, \dots, i_t) = \max_{q_1, \dots, q_{t-k}} P(q_1 = s_{i_1}, \dots, q_t = s_{i_t}, O_1, \dots, O_t \mid \mathcal{S}),$$

where $1 \leq k \leq t \leq T$. $\theta_t(i_{t-k+1}, \dots, i_t)$ is the maximum probability of the partial observation sequence up to time t and the state sequence ending at s_{i_t} . The following recursive algorithm holds for $1 \leq k+1 \leq t \leq T$ and $1 \leq i_{t+1} \leq N$

$$\begin{aligned}
\theta_{t+1}(i_{t-k+2}, \dots, i_{t+1}) &= \max_{q_1, \dots, q_{t-k+1}} \left(P(q_{t+1} = s_{i_{t+1}}, O_{t+1} \mid \mathcal{S}, q_1 = s_{i_1}, \dots, q_t = s_{i_t}, O_1, \dots, O_t) \cdot \right. \\
&\quad \left. P(q_1 = s_{i_1}, \dots, q_t = s_{i_t}, O_1, \dots, O_t \mid \mathcal{S}) \right) \\
&= \max_{1 \leq i_{t-k+1} \leq N} \left(\theta_t(i_{t-k+1}, \dots, i_t) \cdot P(O_{t+1} \mid \mathcal{S}, q_{t+1} = s_{i_{t+1}}) \cdot \right. \\
&\quad \left. P(q_{t+1} = s_{i_{t+1}} \mid \mathcal{S}, q_{t-k+1} = s_{i_{t-k+1}}, \dots, q_t = s_{i_t}) \right) \\
&= p_{i_{t+1}}(O_{t+1}) \left(\max_{1 \leq i_{t-k+1} \leq N} \theta_t(i_{t-k+1}, \dots, i_t) \pi_{i_{t-k+1}, \dots, i_{t+1}} \right),
\end{aligned}$$

where for $1 \leq j \leq k$ and $1 \leq i_1, \dots, i_k \leq N$, the initialisation is given by

$$\begin{aligned} \theta_k(i_1, \dots, i_k) &= P(q_1 = s_{i_1} \dots q_k = s_{i_k}, O_1, \dots, O_k \mid \mathcal{S}) \\ &= P(q_1 = s_{i_1} \dots q_k = s_{i_k} \mid \mathcal{S}) \cdot \prod_{j=1}^k P(O_j \mid \mathcal{S}, q_j = s_{i_j}) \\ &= \Lambda_{i_1, \dots, i_k} \prod_{j=1}^k p_{i_j}(O_j) \end{aligned}$$

Eventually, we can find out $\operatorname{argmax}_{1 \leq q_{t-k+1}, \dots, q_t \leq N} \theta_T(i_{t-k+1}, \dots, i_t)$ through the Viterbi iteration. To recover entire most likely state sequence, we further define a back-tracked array that traces $\theta_t(i_{t-k+1}, \dots, i_t)$ for each t and i_{t-k+1}, \dots, i_t as

$$\phi_{t+1}(i_{t-k+2}, \dots, i_{t+1}) = \operatorname{argmax}_{1 \leq q_{t-k+1} \leq N} \theta_t(i_{t-k+1}, \dots, i_t) \cdot \pi_{i_{t-k+1}, \dots, i_{t+1}}.$$

Since we have $\phi_T(i_{T-k+1}, \dots, i_T) = \operatorname{argmax}_{1 \leq q_{t-k+1}, \dots, q_T \leq N} \theta_T(i_{t-k+1}, \dots, i_T)$, the remaining q_t then can be obtained recursively by setting $\phi_k(i_1, \dots, i_k) = 0$.

1.2.5 Training with expectation-maximisation algorithm

Both evaluation and decoding issues associated with HOHMMs in Sections 1.2.3 and 1.2.4 are solved under a measurable circumstance where the set of parameters \mathcal{S} is known in advance. However, we might encounter the situation in practice when parameters are not directly foreknown and have to be estimated. This is the training problem that adjusts model parameters by maximising the probability of the observed sequence under the given HOHMM. Regular maximum likelihood criteria might not generate an optimal solution especially with unknown state paths. $P(O \mid \mathcal{S})$, however, can be locally maximised under the HOHMM setting by the Baum-Welch algorithm [1], a special version of the expectation-maximum (EM) algorithm [4]. It provides an efficient method for researching parameter estimation in the further application of HOHMMs. We first present the general EM algorithm, and then describe results of the Baum-Welch algorithm under the HOHMM setting by following [2].

Generally, there are two main applications of the EM algorithm. One situation occurs when likelihood functions are not analytically solvable but can be tractable by assuming the occurrence of additional missing parameters. The other one happens where data is incomplete or missing owing to the possible limitation of observations. In our study of

HOHMMs, the state sequence Q is hidden and viewed as the missing data, whilst the process O is observed. The EM algorithm is presented in the discrete density case for the subsequent discussion of the Baum-Welch algorithm and to remain consistent with the previous discussion in evaluation and decoding. We assume the complete-data likelihood function $L(\mathcal{S} | Q, O) = P(Q, O | \mathcal{S})$, and have

$$P(Q, O | \mathcal{S}) = P(Q | O, \mathcal{S})P(O | \mathcal{S}). \quad (1.5)$$

Let $\widehat{\mathcal{S}}$ be the set of reestimated parameters that generates an improved model compared to that of \mathcal{S} . Considering equation (1.5) and the log likelihood function $l(\widehat{\mathcal{S}} | Q, O) = \log P(O | \widehat{\mathcal{S}}) = \sum_Q \log P(Q | \widehat{\mathcal{S}})$, we get

$$\log P(Q, O | \widehat{\mathcal{S}}) = \log P(Q | O, \widehat{\mathcal{S}}) + \log P(O | \widehat{\mathcal{S}}) \quad (1.6)$$

By taking a sum over Q and multiplying equation (1.6) by $P(Q | O, \mathcal{S})$, we can further transform it into

$$\begin{aligned} \log \frac{P(O | \widehat{\mathcal{S}})}{P(O | \mathcal{S})} &= \sum_Q P(Q | O, \mathcal{S}) \log P(O, Q | \widehat{\mathcal{S}}) - \sum_Q P(Q | O, \mathcal{S}) \log P(O, Q | \mathcal{S}) \\ &\quad + \sum_Q P(Q | O, \mathcal{S}) \left(\log \frac{P(Q | O, \mathcal{S})}{P(Q | O, \widehat{\mathcal{S}})} \right). \end{aligned} \quad (1.7)$$

It is straightforward to show that $\sum_Q P(Q | O, \mathcal{S}) \left(\log \frac{P(Q | O, \mathcal{S})}{P(Q | O, \widehat{\mathcal{S}})} \right) \geq 0$. Let $A(\mathcal{S}, \widehat{\mathcal{S}}) = \sum_Q P(Q | O, \mathcal{S}) \log P(O, Q | \widehat{\mathcal{S}})$. With the previous notation, we can rewrite equation (1.7) as

$$l(\widehat{\mathcal{S}} | O) - l(\mathcal{S} | O) \geq A(\mathcal{S}, \widehat{\mathcal{S}}) - A(\mathcal{S}, \mathcal{S}), \quad (1.8)$$

where the strict inequality holds unless $P(Q | O, \widehat{\mathcal{S}}) = P(Q | O, \mathcal{S})$ or $\widehat{\mathcal{S}} = \mathcal{S}$. The EM procedure is iteratively constructed from the sequence $\{\widehat{\mathcal{S}}^{(k)}\}_{k \geq 1}$ initiating with some value $\widehat{\mathcal{S}}^{(0)}$. Every iteration contains two steps:

1. Expectation step: calculate $A(\mathcal{S}, \widehat{\mathcal{S}}^{(k)})$.
2. Maximisation step: determine $\widehat{\mathcal{S}}^{(k+1)} = \operatorname{argmax}_{\mathcal{S}} A(\mathcal{S}, \widehat{\mathcal{S}}^{(k)})$.

When $\widehat{\mathcal{S}} \neq \mathcal{S}$ in equation (1.8), an improved estimate $\widehat{\mathcal{S}}$ is produced by the iteration until a stopping criterion is met to make $\widehat{\mathcal{S}} \approx \mathcal{S}$. Then a local maximum of the likelihood is

reached.

Two auxiliary variables need to be defined first in terms of forward and backward variables for properly presenting the Baum-Welch algorithm. We denote $\varepsilon_t(i_1, \dots, i_{k+1})$ as the probability of being in state $s_{i_1}, \dots, s_{i_{k+1}}$ at time $t, \dots, t+k$ correspondingly given the model and the observation sequence, where

$$\begin{aligned} \varepsilon_t(i_1, \dots, i_{k+1}) &= P(q_t = s_{i_1}, \dots, q_{t+k} = s_{i_{k+1}} \mid O, \mathcal{S}) \\ &= \frac{P(q_t = s_{i_1}, \dots, q_{t+k} = s_{i_{k+1}}, O_1, \dots, O_{t+k} \mid \mathcal{S})}{P(O \mid \mathcal{S})} \\ &= \frac{\alpha_{t+k-1}(i_1, \dots, i_k) \beta_{t+1}(i_2, \dots, i_{k+1}) \pi_{i_1, \dots, i_{k+1}} p_{i_{k+1}}(O_{t+k})}{P(O \mid \mathcal{S})}. \end{aligned} \quad (1.9)$$

The second variable $\xi_t(i_1, \dots, i_k)$ is the probability of being in state s_{i_1}, \dots, s_{i_k} at time $t, \dots, t+k-1$ correspondingly, given the observation sequence and the model. It can be expressed by $\varepsilon_t(i_1, \dots, i_{k+1})$ as

$$\xi_t(i_1, \dots, i_k) = \sum_{i_{k+1}=1}^T \varepsilon_t(i_1, \dots, i_{k+1}) \quad (1.10)$$

By summing $\varepsilon_t(i_1, \dots, i_{k+1})$ over t , we get the expectation number of transitions from the state sequence $s_{i_1}, \dots, s_{i_{k+1}}$. The expectation number of transitions from the state sequence s_{i_1}, \dots, s_{i_k} can be generated alike by summing up $\xi_t(i_1, \dots, i_k)$ over t . Re-estimation formulas then can be given by

$$\widehat{\varepsilon}_t(i_1) = \widehat{\xi}_t(i_1) \quad (1.11)$$

$$\widehat{\pi}_{i_1, \dots, i_{k+1}} = \frac{\sum_{t=1}^{T-k} \varepsilon_t(i_1, \dots, i_{k+1})}{\sum_{t=1}^{T-k} \sum_{i_{k+1}=1}^N \varepsilon_t(i_1, \dots, i_{k+1})} \quad (1.12)$$

$$\widehat{p}_j(y_k) = \frac{\sum_{t=1, s, t}^{T-k} O_{t=y_k} \xi_t(j)}{\sum_{k=1}^n \sum_{t=1, s, t}^{T-k} O_{t=y_k} \xi_t(j)} \quad (1.13)$$

The Baum-Welch training process thereby can be performed by assuming a starting set of parameters \mathcal{S}_0 . Forward and backward variables, α_t and β_t , are then computed according to equations (1.3) and (1.4) respectively, followed by calculating auxiliary variables ε_t and ξ_t using equations (1.9) and (1.10). The parameters are re-estimated iteratively by equations

(1.11) to (1.13) till estimates converge. As we mentioned before, the Baum-Welch algorithm is viewed as a special case of the EM algorithm with the same goal of maximising $P(O | \mathcal{S})$ by adjusting \mathcal{S} given an HOHMM. To deal with hidden states and incomplete data, the EM algorithm becomes significantly powerful in the parameter estimation and filtering processes of HOHMMs.

1.2.6 Change of measure method in HOHMMs

To address filtering problems of HOHMMs in our research, relevant processes and optimal filters can be generated generally through two ways: the semi-martingale method and the change of probability measure approach. The prior one is direct but extremely complicated, whilst the latter one is indirect but efficient in filtering applications. Zakai [22] pioneers the application of change of measure technique in stochastic filtering. Elliott et al. [6] derive optimal filters with this method for HMMs based on Girsanov's Theorem. Mamon et al. [14] develop closed-form solutions of the recursive filters for estimating optimally the parameters of a commodity price model via change of measure.

Considering the wide use of this approach mostly in HMMs, one additional but critical step ought to be implemented for our HOHMM setting. That is to transform a higher-order Markov chain into a first-order Markov chain rather than to derive the change of probability measure for HOHMMs directly. Kriouile et al. [10] develop an equivalent model specific to transiting a second-order discrete HMM to a regular HMM. An detailed example is described to convert a second-order two state Markov chain into a regular 2-state one with efficient recursive algorithms [15]. Based on these previous work for processing second-order HMMs, Du Preez extends order-reducing algorithms to a general HOHMM setting, which enables any HOHMM to be transformed into its corresponding first-order HMM [5]. The essential idea is similar to the common method that converts higher-order differential equations into a system of first-order differential equations. We follow their computing algorithms by introducing a mapping variable that transforms a higher-order Markov chain into a regular one. Then the change of probability measure method can be utilised to find our filtering algorithms.

Equivalent to the real world measure P , an ideal measure \tilde{P} is defined as a reference probability from a discrete-time version of Girsanov's Theorem [9]. Under \tilde{P} , observations are independent and identically random variables, and the Markov chain follows the same dy-

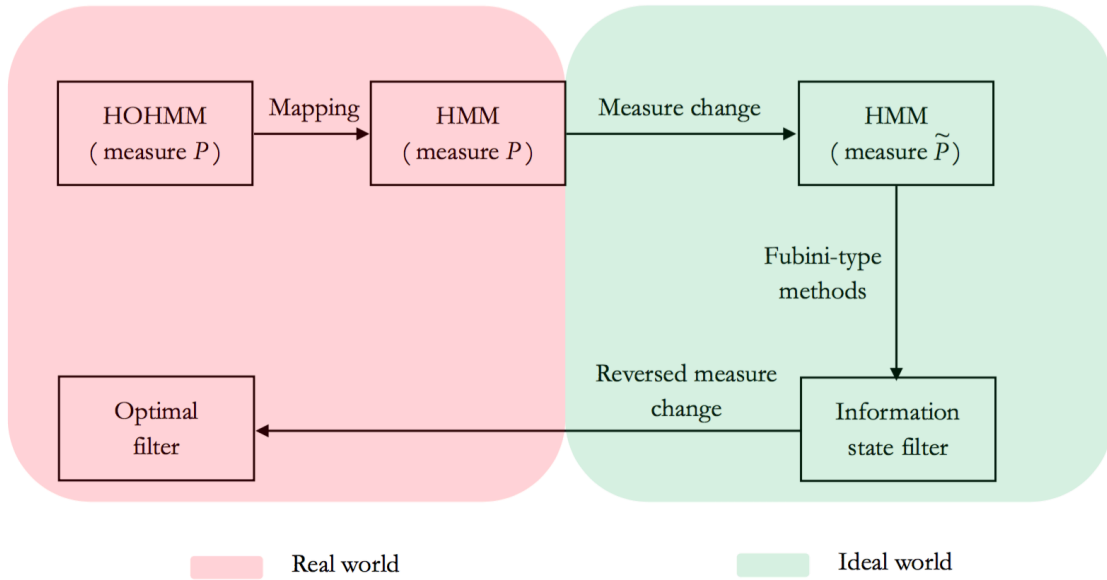


Figure 1.1: Reference measure optimal filter derivation for HOHMMs

namics as those under P . To find optimal filters, an easy-to-compute framework then can be carried out by Fubini's Theorem, which allows the interchange of expectations and summations [11]. The results and calculations under the ideal measure \tilde{P} can be traced back to the real world measure P by invoking a reverse measure change. The derivation process of reference probability optimal filters can be explicitly illustrated in Figure 1.1.

1.3 Structure of the thesis

This thesis consists of six chapters. An overview of HMMs and underpinnings of HOHMMs is given in this Chapter; see previous section. The main contents focus on the results of four related projects. We develop HMM and HOHMM settings for the dynamics of daily average temperatures (DATs) for weather derivatives in Chapters 2 and 3, respectively. The modelling, filtering and estimation problems for the respective electricity-spot and salmon futures prices are addressed in Chapters 4 and 5. Some concluding remarks are given in Chapter 6. The synopsis of the projects are briefly introduced below.

1.3.1 Putting a price tag on temperature

A model for the evolution of DATs is put forward to support the analysis of weather derivatives. The goal is to capture simultaneously the mean-reversion, seasonality and stochasticity properties of the DATs process. An OU process modulated by an HMM is proposed to model the both mean-reversion and stochasticity of a deseasonalised component. The seasonality part is modelled by a combination of linear and sinusoidal functions. OU-HMM filtering algorithms are established for the evolution of switching model parameters. Consequently, adaptive parameter estimates are obtained. Numerical implementation of the estimation technique using a data set compiled by the National Climatic Data Center (NCDC) was conducted. A sensitivity analysis of the option prices with respect to model parameters is included.

1.3.2 A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics

We develop a model for the evolution of DATs that could benefit the analysis of weather derivatives in finance and economics as well as the modelling of time series data in meteorology, hydrology and other branches of the sciences and engineering. Our focus is to capture the mean-reverting, seasonality, memory and stochastic properties of the temperature movement and other time series exhibiting such properties. To model both mean-reversion and stochasticity, a deseasonalised component is assumed to follow an OU process modulated by a higher-order hidden Markov chain, which takes into account short/long range dependence in the data. The seasonality part is modelled through a combination of linear and sinusoidal functions with appropriate coefficients and arguments. Furthermore, we put forward a parameter estimation approach that establishes recursive HOHMM filtering algorithms customised for the regime-switching evolution of model parameters. Quantities that are functions of HOHMC characterise these filters. Utilising the EM method in conjunction with the change of measure technique, optimal and self-updating parameter estimates are obtained. We illustrate the numerical implementation of our model and estimation technique using a 4-year Toronto DATs data set compiled by the NCDC. We perform pertinent model selection and validation diagnostics to assess the performance of our methodology. It is shown that a 2-state HOHMM-based model best captures the empirical characteristics of the temperature data under examination on the basis of various error-based and information-criterion metrics.

1.3.3 A higher-order Markov chain-modulated model for electricity spot-price dynamics

Over the last three decades, the electricity sector worldwide underwent massive deregulation. Power market participants have encountered a growing number of challenges due to competition and other pertinent factors. As electricity is a non-storable commodity, its price is extremely sensitive to changes in supply and demand. The evolution of electricity prices exhibits pronounced mean reversion and cyclical patterns, possesses extreme volatility and relatively frequently occurring spikes, and manifests presence of memory property. These observed features necessitate the development of models aimed to simultaneously capture such price characteristics for forecasting, risk management, and valuation of electricity-driven derivatives. This study tackles the modelling and estimation problems under a new paradigm that integrates the deterministic calendar seasons and stochastic factors governing electricity prices. The de-seasonalised component of our proposed model has both the jump and mean-reverting properties to account for spikes and periodic cycles alternating between lower price returns and compensating periods of higher price returns. The parameters of the de-seasonalised model components are also modulated by a higher-order hidden Markov chain (HOHMC) in discrete time. This provides a mechanism to extract latent information from historical data. The HOHMC's state is interpreted as the "state of the world" resulting from the interaction of various forces impacting the electricity market. Filters are developed to generate optimal estimates of HOHMC-relevant quantities using the observation process, and these provide online estimates of model parameters. Empirical demonstrations using the daily electricity spot prices, compiled by the Alberta Electric System Operator (AESO), show that our HOHMM approach has considerable merits in terms of price data fitting and forecasting metrics. Implications of our model to the pricing of an electricity forward contract are also examined.

1.3.4 Modelling and forecasting futures-prices curves in the Fish Pool market

This project aims to capture the evolution and salient features of fish-futures prices with a flexible and dynamic approach via a higher-order hidden Markov model (HOHMM). The parameters of a proposed futures-price model under a multivariate setting are governed by a discrete-time HOHMM to account for the random switching of market or economic states over time. Multi-dimensional filters derived, along with the application of the expectation-

maximisation (EM) algorithm, give rise to a complete on-line estimation of futures-price model parameters. Our numerical implementation analyses the Fish Pool's salmon futures prices. The goodness of fit and forecasting performance of our approach are assessed using appropriate statistical metrics. Our empirical findings illustrate that the HOHMM-based approach possesses better fitting capacity and excellent short-run predictive ability of futures prices vis-à-vis certain modelling benchmarks.

References

- [1] L. E. Baum, T. Petrie, G. Soules and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics* 41(1) (1970), 164–171.
- [2] W. Ching. *Markov Chains: Models, Algorithms and Applications* (2nd;2;2nd 2013; ed).
- [3] O. Cappé, E. Moulines and T. Ryden. *Inference in Hidden Markov Models* (2006).
- [4] A. P. Dempster, N. M. Laird and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1) (1977), 1–33.
- [5] J. A. Du Preez. Efficient training of high-order hidden markov models using first-order representations. *Computer Speech & Language*, 12(1)(1998), 23–39.
- [6] R. J. Elliott, L. Aggoun and J. B. Moore. *Hidden Markov Models: Estimation and Control* (Corr. 2nd. printing. ed.) (1997).
- [7] R. J. Elliott and H. Yang. Forward and backward equations for an adjoint process, *Festschrift for G. Kallianpur*, Springer Verlag, Berlin, Heidelberg, New York. (1992).
- [8] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57(2)(1989), 357–384.
- [9] B. H. Juang and L. R. Rabiner. *Spectral representations for speech recognition by neural networks-a tutorial* (1992).
- [10] A. Kriouile, J. Mari and J. Haon. Some improvements in speech recognition algorithms based on HMM (1990).

- [11] M. Loeve. Probability Theory, fourth edn, Springer Verlag, Berlin, Heidelberg, New York. (1978).
- [12] N. Lobato and N. E. Savin. Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics* 16(3) (1998), 261-268.
- [13] J. McCarthy. Tests of long-range dependence in interest rates using wavelets. *Quarterly Review of Economics and Finance* 44(1) (2004), 261–268.
- [14] R. Mamon, C. Erlwein, B. Gopaluni, Adaptive signal processing of asset price dynamics with predictability analysis, *Information Sciences*, 178 (1) (2008), 203–219.
- [15] I. L. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete-Valued Time Series* (1997).
- [16] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 77 (2) (1989), 257–286.
- [17] T. Rydén, T. Teräsvirta, and SÅsbrink. Stylized facts of daily return series and the hidden Markov model. *Journal of applied econometrics*. 5(1)(1998), 217–44.
- [18] T. Siu, W. Ching, E. Fung, M. Ng, X. Li, A high-order markov-switching model for risk measurement, *Computers and Mathematics with Applications*, 58 (1) (2009), 1–10.
- [19] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* 13(2)(1967), 260-269.
- [20] X. Xi, R. Mamon and M. Davison. A higher-order hidden markov chain-modulated model for asset allocation. *Journal of Mathematical Modelling and Algorithms in Operations Research* 13(1)(2014),59–85.
- [21] S. Yu, Z. Liu, M. S. Squillante, C. Xia and L. Zhang. A hidden semi-markov model for web workload self-similarity (2002).
- [22] M. Zakai. On the optimal filtering of diffusion processes. *Zeitschrift f r Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 11(3)(1969), 230–243.

Chapter 2

Putting a price tag on temperature

2.1 Introduction

We are all familiar with car and home insurance policies. The cost and overall value of these contracts are determined by the likelihood of occurrence of particular calamities, disasters and accidents, amongst other risk factors. But, how does one insure against the vagaries of weather? Specifically, what is the price of a contract whose value depends on how hot or cold it will be on a given day in the future?

We shall consider a modelling approach that provides support for temperature-linked contracts and weather-related derivatives. It is estimated that about 39.1% of the U.S. gross domestic product is weather-sensitive [13], and over 90% of weather derivatives are temperature-based [14]. The first weather-derivative transaction took place in 1997 executed by Aquila Energy and embedded in a power contract [8], and in 1999 the first exchange-traded temperature derivative was launched in the Chicago Mercantile Exchange (CME). The volume of weather derivatives, since then, has grown rapidly both in the exchange and over-the-counter (OTC) markets.

Weather-driven futures and options trading in the CME are written on the following indices: temperature heating degree days (HDD), cooling degree days (CDD) and cumulative average temperature (CAT). These indices are not tradeable or storable. Hence, the usual no-arbitrage valuation methodology is not necessarily valid in pricing such contracts [10]. Assuming the existence of an appropriate pricing measure, our aim is to develop a model that could accurately capture the salient features of the temperature data aiding the accurate pricing of weather derivatives.

In the literature, several studies were conducted to deal with the valuation of weather derivatives. The historical burn analysis (HBA) is adopted by most investors given its ease of replication [12]. This is because the HBA assumes that the distribution of the expected value of the temperature-based derivatives simply follow the historical data; there is no construction of any stochastic models and fitting for the dynamics of the underlying variables. It is argued, however, in [25] that HBA is bound to be biased and prone to large pricing errors. As a suitable alternative, time-series methods are proposed [25] and the marginal substitution value principle of mathematical economics could be employed for pricing [10]. Indifference pricing was also considered in [38], and a comparison of HBA, Black-Scholes-Merton approximation, and an equilibrium Monte Carlo simulation was examined in [32].

The Ornstein-Uhlenbeck (OU) process is tapped in reproducing the dynamic behaviour of the DATs and hence in generating the temperature-based indices. This is aligned to the concept of adopting a continuous-time stochastic process to capture the dynamics of temperature [11]. Incorporating seasonalities [1] enriches the model. In [3], DAT variations are modelled with an OU process with innovations following the generalised hyperbolic Levy process.

A one-state stochastic process cannot describe the behaviour of temperature with great precision and flexibility, especially when there are drastic changes in regimes attributed to occasional climatic changes. Thus, the regime-switching OU model is considered in [14], which seems to be the first paper embedding a regime-switching approach to capture the evolution of the time series temperature data. However, no dynamic estimation, not even a static one, was provided in [14] leaving a big gap in implementing a regime-switching approach involving observed data until the publication of [37] that employs a higher-order HMM. We note though that [37], extending the developments in [36], focuses only estimation and model fitting, and the pricing and risk measurement of pertinent contracts under such a model setting remain unaddressed. In this chapter, we apply hidden Markov model (HMM) filtering algorithms to provide optimal parameter model estimates for a successful implementation of the valuation and risk management of weather futures and option products.

Our exposition is organised in the following way. Section 2.2 presents a self-contained for-

mulation of the temperature modelling using a discrete-time HMM modulating the model parameters. In Section 2.3, we derive recursive filtering equations for various quantities of HMM via the change of reference probability measure method. Our self-updating parameter estimation scheme is laid out in Section 2.4. Numerical work demonstrating the applicability of our proposed model and estimation technique to a four-year Toronto temperature data set is detailed in 2.5; and in this section, mechanics of the best-model selection are also outlined through forecasting and penalised log-likelihood criteria. In Section 2.6, the valuation of a temperature option is discussed and a sensitivity analysis of the option price with respect to parameters of our proposed model is included. Section 2.7 gives some concluding remarks.

2.2 Model description

Notation: All vectors and matrices will be denoted by bold English/Greek letters in lowercase and bold capitalised English/Greek letters, respectively.

2.2.1 Model for temperature derivatives

As weather derivatives are written on HDD, CDD and CAT, we put forward a model to capture these measurements. HDD and CDD quantify the respective demands of heating and cooling of a particular location. Since temperature could be observed at any time of a given day, given the time interval $[\tau_1, \tau_2]$ with $\tau_1 < \tau_2$, we can regard it as an Itô process so that the respective continuous-time functional forms of HDD, CDD and CAT are

$$\begin{aligned} \text{HDD} &= \int_{\tau_1}^{\tau_2} \max(T_{\text{base}} - T_s, 0) ds, \\ \text{CDD} &= \int_{\tau_1}^{\tau_2} \max(T_s - T_{\text{base}}, 0) ds, \\ \text{and CAT} &= \int_{\tau_1}^{\tau_2} T_s ds, \end{aligned} \tag{2.1}$$

where T_{base} is the base temperature, usually given as 65°F, or 18°C in the market. In practice, the daily average temperatures (DATs) are measured on a daily basis so that the

temperature-based indices must be considered in discrete time defined by

$$\text{HDD} = \sum_{t=\tau_1}^{\tau_2} \max(T_{\text{base}} - T_t, 0), \quad \text{CDD} = \sum_{t=\tau_1}^{\tau_2} \max(T_t - T_{\text{base}}, 0), \quad \text{and} \quad \text{CAT} = \sum_{t=\tau_1}^{\tau_2} T_t, \quad (2.2)$$

where T_t is the DAT on day t computed as $\frac{T_{\text{max}} + T_{\text{min}}}{2}$. Clearly, CAT is simply the sum of DATs over the contract period. Weather contracts typically mature in a month or season. An HDD contract's period spans the period covering October in one year to April of the next year. A CDD contract's period covers the warm season, i.e., starting April until October of the same year. Overlapping months, October and April, in the CDD and HDD contracts are termed as transition or shoulder months. Following the rationale in [5], the needed indices as well as the dynamic representation of DATs will be obtained from T_t , which we are going to model.

Suppose (Ω, \mathcal{F}, P) is an underlying probability space for the process T_t given by,

$$T_t = X_t + S_t. \quad (2.3)$$

Equation (2.3) is composed of X_t that follows an OU process with HMM-driven parameters and S_t described by a deterministic function to capture the seasonal trends. Motivated by [5], the seasonal component is

$$S_t = at + b + \sum_{h=1}^3 \left[c_h \sin\left(d_h t \frac{2\pi}{365}\right) + e_h \cos\left(d_h t \frac{2\pi}{365}\right) \right], \quad (2.4)$$

where $d_1 = 1$, $d_2 = 2$, $d_3 = 4$ to account for the respective yearly, semi-annual and quarterly patterns.

2.2.2 The HMM-modulated OU process

The OU process X_t in (2.3) satisfies the stochastic differential equation (SDE)

$$dX_t = \alpha(\theta - X_t)dt + \xi dB_t, \quad (2.5)$$

where B_t is a standard Brownian motion under P . The solution, by Itô's lemma, of (2.5) is

$$X_t = X_s e^{-\alpha(t-s)} + (1 - e^{-\alpha(t-s)})\theta + \xi e^{-\alpha t} \int_s^t e^{\alpha u} dB_u \quad (2.6)$$

for $s \leq t$. The explicit representation of the temperature process T_t is then given by

$$T_t = S_t + X_s e^{-\alpha(t-s)} + (1 - e^{-\alpha(t-s)})\theta + \xi e^{-\alpha t} \int_s^t e^{\alpha u} dB_u. \quad (2.7)$$

The discretised versions of (2.6) and (2.7), by approximating the distributions of their respective stochastic-integral components, are given by

$$X_{k+1} = e^{-\alpha \Delta t_{k+1}} X_k + (1 - e^{-\alpha \Delta t_{k+1}})\theta + \xi \sqrt{\frac{1 - e^{-2\alpha \Delta t_{k+1}}}{2\alpha}} z_{k+1}, \quad (2.8)$$

where $\Delta t_{k+1} = t_{k+1} - t_k$ and $\{z_{k+1}\}$ is a sequence of independent and identically distributed (IID) standard normal random variables.

As indicated in Section 2.1, having constant parameters in the OU process might not be adequate in describing the dynamic switches of regimes resulting from the interactions of various market and economic factors. In the succeeding discussion, the parameters in (2.8) will be governed by a homogeneous Markov chain \mathbf{y}_k with finite states in discrete time to model the behaviour of X_t and hence, T_t .

To simplify the ensuing algebraic calculations, the state space is associated with the canonical basis of \mathbb{R}^N , which is $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$, $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$ with 1 in the i th position. Here, N is the total number of states and \top denotes the transpose of a matrix. The dynamics of \mathbf{y}_k is $\mathbf{y}_{k+1} = \mathbf{\Pi} \mathbf{y}_k + \boldsymbol{\zeta}_{k+1}$, where $\mathbf{\Pi} = (\pi_{ji}) \in \mathbb{R}^{N \times N}$, $\boldsymbol{\zeta}_{k+1}$ is a martingale increment, and $\pi_{ji} = P(\mathbf{y}_k = \mathbf{e}_j | \mathbf{y}_{k-1} = \mathbf{e}_i)$ with $\sum_{j=1}^N \pi_{ji} = 1$. Let $\mathbf{P} = (p_{ji}) \in \mathbb{R}^{N \times N}$ denote the intensity matrix for the continuous-time Markov chain process at time t , where, $p_{ji} = \lim_{\Delta t \rightarrow 0} \frac{\pi_{ji}(t, t+\Delta t)}{\Delta t}$, for $j \neq i$, (cf Cox and Miller [7]), with $p_{ji} \geq 0$ and $\sum_{i=1}^N p_{ji} = 0$. Given that our HMM filtering methodology is under the discrete-time framework, it is necessary to recall the connection of the rate matrix \mathbf{P} to the transition probability matrix $\mathbf{\Pi}$, which is its discrete-time counterpart providing inputs to the recursive filtering equations in Section 2.3 and eventually to the option pricing formulae in Section 2.6. From Grimmett and Stirzaker [27] (see as well Siu, et al [33]), $\mathbf{\Pi}$ is the exponential matrix of \mathbf{P} , i.e.,

$$\mathbf{\Pi} = \exp(\mathbf{P}) = \sum_{k=0}^{\infty} \frac{\mathbf{P}^k}{k!}. \quad (2.9)$$

A handy technique to convert an $N \times N$ matrix \mathbf{P} into an $N \times N$ $\mathbf{\Pi}$ is to take advantage of the diagonalisability of \mathbf{P} inherent in our modelling formulation. If \mathbf{P} has eigenvalues

b_1, b_2, \dots, b_N and corresponding eigenvectors $\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_N$, then $\mathbf{P} = \mathbf{J}\mathbf{B}\mathbf{J}^{-1}$, where \mathbf{B} is a diagonal matrix with b_1, b_2, \dots, b_N as its diagonal elements, and $\mathbf{J} = \{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_N\}$ is an invertible matrix. Furthermore,

$$\begin{aligned} \exp(\mathbf{P}) &= \exp(\mathbf{J}\mathbf{B}\mathbf{J}^{-1}) \\ &= \mathbf{I} + \mathbf{J}\mathbf{B}\mathbf{J}^{-1} + \mathbf{J}\frac{\mathbf{B}^2}{2!}\mathbf{J}^{-1} + \dots \\ &= \mathbf{J}\left(\mathbf{I} + \mathbf{B} + \frac{\mathbf{B}^2}{2!} + \dots\right)\mathbf{J}^{-1} \\ &= \mathbf{J}\exp(\mathbf{B})\mathbf{J}^{-1}. \end{aligned} \quad (2.10)$$

Both (2.9) and (2.10) can be evaluated numerically by almost all mathematical or statistical software packages nowadays.

We assume the process X_k in (2.8) have parameters that depend on \mathbf{y}_k , i.e.,

$$X_{k+1} = e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}} X_k + (1 - e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}})\theta(\mathbf{y}_k) + \xi(\mathbf{y}_k) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{y}_k)\Delta t_{k+1}}}{2\alpha(\mathbf{y}_k)}} z_{k+1}, \quad (2.11)$$

$$\text{and } T_{k+1} = S_{k+1} + e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}} X_k + (1 - e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}})\theta(\mathbf{y}_k) + \xi(\mathbf{y}_k) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{y}_k)\Delta t_{k+1}}}{2\alpha(\mathbf{y}_k)}} z_{k+1}. \quad (2.12)$$

With \mathbf{y}_k 's state space, $\alpha_k := \alpha(\mathbf{y}_k) = \langle \boldsymbol{\alpha}, \mathbf{y}_k \rangle$, $\theta_k := \theta(\mathbf{y}_k) = \langle \boldsymbol{\theta}, \mathbf{y}_k \rangle$, and $\xi_k := \xi(\mathbf{y}_k) = \langle \boldsymbol{\xi}, \mathbf{y}_k \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product in \mathbb{R}^N . Clearly, \mathbf{y}_k makes the models in (2.11) and (2.12) regime-switching.

2.3 Recursive filtering

2.3.1 Change of reference probability measure

We shall introduce an ideal probability measure \tilde{P} , where calculations of the recursive filters are manageable to carry out. This is because under \tilde{P} , the observations are independent, identically distributed (IID) random variables. The real measure P can be recovered via the change of measure technique. The state of the HMM will be estimated from the noisily observed data set. Under the reference measure \tilde{P} , the dynamics of \mathbf{y}_k remain unchanged but

are independent of the observed data series; see the underlying principle utilised in Elliott et al. [19].

A discrete-time version of the Girsanov's theorem enables us to back out P from \tilde{P} [19]. The real-world measure P is recovered by considering the Radon-Nikodym derivative

$$\left. \frac{dP}{d\tilde{P}} \right|_{\mathcal{F}_k} = \Psi_k = \prod_{l=1}^k \varphi_l, \quad k \geq 1,$$

where

$$\varphi_l = \frac{\phi \left\{ \xi(\mathbf{y}_{l-1})^{-1} \left[\left(\frac{1 - e^{-2\alpha(\mathbf{y}_{l-1})\Delta t_{l-1}}}{2\alpha(\mathbf{y}_{l-1})} \right)^{-\frac{1}{2}} \right] \beta(\mathbf{y}_{l-1}) \right\}}{\left(\xi(\mathbf{y}_{l-1}) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{y}_{l-1})\Delta t_{l-1}}}{2\alpha(\mathbf{y}_{l-1})}} \right) \phi(X_l)}. \quad (2.13)$$

In (2.13), ϕ is the probability density function of a standard normal random variable, $\varphi_0 = 1$, $\{\varphi_l, l \in \mathbb{Z}^+\}$ is an \mathcal{F}_l -martingale under P , and $\beta(\mathbf{y}_{l-1}) = X_l - e^{-\alpha(\mathbf{y}_{l-1})}X_{l-1} - (1 - e^{-\alpha(\mathbf{y}_{l-1})\Delta t_{l-1}})\theta(\mathbf{y}_{l-1})$. With the aid of (2.13), the derivation of the recursive filters for \mathbf{y}_k and related quantities under \tilde{P} is facilitated, and then the calculated results can be re-interpreted under P via a reverse change of measure.

2.3.2 Calculation of recursive filters

Suppose \mathcal{X}_k is the filtration generated by the observations of X_k , which 'conceal' the 'true' state of \mathbf{y}_k . We aim to obtain optimal estimates of quantities that are functions of \mathbf{y}_k under P through their adaptive filters calculated using the conditional expectations \mathcal{X}_k under \tilde{P} and using the Bayes' theorem.

So, for instance, to estimate a scalar quantity U under P , Bayes' theorem for conditional expectation gives

$$\mathbb{E}[U_k | \mathcal{X}_k] = \frac{\tilde{\mathbb{E}}[\Psi_k U_k | \mathcal{X}_k]}{\tilde{\mathbb{E}}[\Psi_k | \mathcal{X}_k]} \quad (2.14)$$

where $\mathbb{E}[\cdot]$ and $\tilde{\mathbb{E}}[\cdot]$ are the conditional expectations under P and \tilde{P} , respectively. As in [18], we consider an \mathcal{F}_l -adapted U_l , having the form

$$U_l = U_{l-1} + g_l + \langle \mathbf{h}_l, \mathbf{o}_l \rangle + r_l f(X_l) \quad (2.15)$$

where g , \mathbf{h}_l , and r_l are \mathcal{F} -predictable; and f is a scalar-valued function and $\mathbf{o}_l = \mathbf{y}_l + \mathbf{\Pi}\mathbf{y}_{l-1}$.

Write $\gamma_k(U_k) := \tilde{\mathbb{E}}[\Psi_k U_k | \mathcal{X}_k]$. Therefore, equation (2.14) can be expressed as $\frac{\gamma_k(U_k)}{\gamma_k(\mathbf{1})}$. Since $\gamma_k(U_k) = \gamma_k(U_k \langle \mathbf{1}, \mathbf{y}_k \rangle) = \langle \mathbf{1}, \gamma_k(U_k \mathbf{y}_k) \rangle$, the filter for any adapted process U has the representation

$$\mathbb{E}[U_k | \mathcal{X}_k] = \frac{\langle \mathbf{1}, \gamma_k(U_k \mathbf{y}_k) \rangle}{\langle \mathbf{1}, \gamma_k(\mathbf{y}_k) \rangle}.$$

Invoking Theorem (5.3) of [15] and tailoring it to our modelling framework and notation, the recursion for $\gamma_k(U_k \mathbf{y}_k)$ is

$$\begin{aligned} \gamma_k(U_k \mathbf{y}_k) &= \sum_{i=1}^N \Lambda^i(X_k) \left[\langle \mathbf{e}_i, \gamma_{k-1}(U_{k-1} \mathbf{y}_{k-1}) \rangle \mathbf{\Pi e}_i + \langle \mathbf{e}_i, \gamma_{k-1}(g_k \mathbf{y}_k - 1) \rangle \mathbf{\Pi e}_i \right. \\ &\quad \left. + (\text{diag}(\mathbf{\Pi e}_i) - \mathbf{\Pi e}_i \otimes \mathbf{\Pi e}_i) \gamma_{k-1}(h_k \langle \mathbf{e}_i, \mathbf{y}_{k-1} \rangle) + \gamma_{k-1}(r_k \langle \mathbf{e}_i, \mathbf{y}_{k-1} \rangle) f(X_k) \mathbf{\Pi e}_i \right], \end{aligned} \quad (2.16)$$

where \otimes stands for the tensor product of vectors, $\text{diag}(\cdot)$ is a diagonal matrix, and $\Lambda^i(X_k)$ is defined by

$$\Lambda^i(X_l) = \exp \left\{ - \frac{[e^{-\alpha_i \Delta(l-1)} X_{l-1} + (1 - e^{-\alpha_i \Delta(l-1)}) \theta_i] X_l - [e^{-\alpha_i \Delta(l-1)} X_{l-1} + (1 - e^{-\alpha_i \Delta(l-1)}) \theta_i]^2}{\left(\xi_i \sqrt{\frac{1 - e^{-2\alpha_i \Delta(l-1)}}{2\alpha_i}} \right)^2} - \frac{[e^{-\alpha_i \Delta(l-1)} X_{l-1} + (1 - e^{-\alpha_i \Delta(l-1)}) \theta_i]^2}{2 \left(\xi_i \sqrt{\frac{1 - e^{-2\alpha_i \Delta(l-1)}}{2\alpha_i}} \right)^2} \right\}. \quad (2.17)$$

For the estimation of model parameters in the next section, we require the recursive filters of the following scalar quantities.

- (i) The number of jumps from state \mathbf{e}_r to state \mathbf{e}_s up to time k :

$$\mathcal{J}_k^{sr} = \sum_{l=1}^k \langle \mathbf{y}_{l-1}, \mathbf{e}_r \rangle \langle \mathbf{y}_l, \mathbf{e}_s \rangle = \mathcal{J}_{k-1}^{sr} + \pi_{sr} \langle \mathbf{y}_{k-1}, \mathbf{e}_s \rangle + \langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{o}_k, \mathbf{e}_s \rangle. \quad (2.18)$$

- (ii) The number of occupations up to time k , i.e., the amount of time that \mathbf{y}_k spent in state \mathbf{e}_r :

$$\mathcal{O}_k^r = \sum_{l=1}^k \langle \mathbf{y}_{l-1}, \mathbf{e}_r \rangle = \mathcal{O}_{k-1}^r + \langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle. \quad (2.19)$$

- (iii) The auxiliary process dependent on \mathbf{y}_k for the function f up to time k in state \mathbf{e}_r :

$$\mathcal{T}_k^r(f) = \sum_{l=1}^k f(X_l) \langle \mathbf{y}_{l-1}, \mathbf{e}_r \rangle = \mathcal{T}_{k-1}^r(f) + f(X_k) \langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle. \quad (2.20)$$

The estimator for the state \mathbf{e}_i of \mathbf{y}_k is readily obtained by taking $U_k = U_0 = 1, g_k = h_k = r_k = 0$ in equation (2.15). It has the simplified form

$$\gamma_k(\mathbf{y}_k) = \sum_{i=1}^N \Lambda^i(X_k) \mathbf{\Pi e}_i \langle \mathbf{e}_i, \gamma_{k-1}(\mathbf{y}_{k-1}) \rangle. \quad (2.21)$$

By virtue of (2.16) with $U_0 = 0, U_k = \mathcal{J}_k^{sr}, g_k = \pi_{sr} \langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle, h_k = \langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle \mathbf{e}_s$ and $r_k = 0$, we get

$$\gamma_k(\mathcal{J}_k^{sr} \mathbf{y}_k) = \sum_{i=1}^N \Lambda^i(X_k) \langle \gamma_{k-1}(\mathcal{J}_{k-1}^{sr} \mathbf{y}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi e}_i + \Lambda^r(X_k) \gamma_{k-1}(\langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle) \pi_{sr} \mathbf{e}_s. \quad (2.22)$$

Setting $U_0 = 0, U_k = \mathcal{O}_k^r, g_k = \langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle$, and $h_k = r_k = 0$ in (2.16) and using (2.14), we obtain

$$\gamma_k(\mathcal{O}_k^r \mathbf{y}_k) = \sum_{i=1}^N \Lambda^i(X_k) \langle \gamma_{k-1}(\mathcal{O}_{k-1}^r \mathbf{y}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi e}_i + \Lambda^r(X_k) \gamma_{k-1}(\langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle) \mathbf{\Pi e}_r. \quad (2.23)$$

For the auxiliary process \mathcal{T}_k^r in (2.20), where f has the form $f(X) = X, f(X) = X^2$, or $f(X) = X_{l+1}X_l$, we utilise (2.16) with $U_k = \mathcal{T}_k^r(f), U_0 = g_k = h_k = 0$, and $r_k = \langle \mathbf{y}_{k-1}, \mathbf{e}_r \rangle$ to have

$$\gamma_k(\mathcal{T}_k^r(f) \mathbf{y}_k) = \sum_{i=1}^N \Lambda^i(X_k) \langle \gamma_{k-1}(\mathcal{T}_{k-1}^r(f) \mathbf{y}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi e}_i + \Lambda^r(X_k) \langle \gamma_{k-1}(\mathbf{y}_{k-1}), \mathbf{e}_r \rangle f(X_k) \mathbf{\Pi e}_r. \quad (2.24)$$

2.4 Optimal parameter estimation

We recall the Expectation-Maximisation (EM) algorithm [17] starting with a family of probability measures $\{P^\nu, \nu \in \Upsilon\}$ on (Ω, \mathcal{F}) , where Υ is the set of parameters. Let P^ν be absolutely continuous with reference to a fixed probability measure P^0 . Assume that the observations $\{x_1, x_2, \dots, x_k\}$ of X_k are available and $\mathcal{X}_k \subset \mathcal{F}_k$. The likelihood as a function of information in \mathcal{X}_k is

$$L(\nu) = \mathbb{E}^0 \left[\frac{dP^\nu}{dP^0} \middle| \mathcal{X}_k \right]. \quad (2.25)$$

The goal is to find the estimate of ν that maximises $L(\nu)$, $\widehat{\nu} \in \operatorname{argmax}_{\nu \in \Upsilon} L(\nu)$. Now, consider

$$Q(\nu; \widehat{\nu}_m) = \mathbb{E}_{\widehat{\nu}_m} \left[\log \frac{dP^\nu}{dP^{\widehat{\nu}_m}} \middle| \mathcal{X}_k \right]. \quad (2.26)$$

The EM algorithm is implemented as follows. First, let $m = 0$ and choose the initial value $\widehat{\nu}_0$ for ν . The first iteration is computed as $Q(\nu; \widehat{\nu}_0) = \mathbb{E}_{\widehat{\nu}_0} [\log \frac{dP^\nu}{dP^{\widehat{\nu}_0}} | \mathcal{X}_k]$, and $\widehat{\nu}_1$ is found to make $Q(\widehat{\nu}_1; \widehat{\nu}_0) \geq Q(\nu; \widehat{\nu}_0)$. Second, the E-step computes (2.26). Third, the M-step finds $\widehat{\nu}_{m+1}$ such that $Q(\widehat{\nu}_{m+1}; \widehat{\nu}_m) \geq Q(\nu; \widehat{\nu}_m)$; $\widehat{\nu}_{m+1} \in \operatorname{argmax}_{\nu \in \Upsilon} Q(\nu; \widehat{\nu}_m)$. The E-step and M-step are repeated until some stopping criterion is met, e.g., $|\widehat{\nu}_{m+1} - \widehat{\nu}_m| < \varepsilon$, where ε is sufficiently small.

Let $\delta(\mathbf{y}_k) = e^{-\alpha(\mathbf{y}_k)\Delta t_k}$ and $\eta(\mathbf{y}_k) = (1 - e^{-\alpha(\mathbf{y}_k)\Delta t_k})\theta(\mathbf{y}_k)$. From (2.8), X_{k+1} has a normal distribution with

$$\text{mean } \mu(\mathbf{y}_k) = \delta(\mathbf{y}_k)X_k + \eta(\mathbf{y}_k) \quad \text{and variance } \epsilon^2(\mathbf{y}_k) = \xi^2(\mathbf{y}_k) \frac{1 - e^{-2\alpha(\mathbf{y}_k)\Delta t_k}}{2\alpha(\mathbf{y}_k)}.$$

Hence, at state i ,

$$\alpha_i = -\frac{1}{\Delta t_k} \log \delta_i, \quad \theta_i = \frac{\eta_i}{1 - \delta_i} \quad \text{and} \quad \xi_i^2 = \frac{-\frac{2\epsilon_i^2}{\Delta t_k} \log \delta_i}{1 - \delta_i^2}. \quad (2.27)$$

The complete parameter estimation of (2.8) is attained by applying the EM algorithm and using the adaptive filters (2.18), (2.19) and (2.20). We then have the next result.

Proposition 2.4.1 *The EM estimates for model parameters are*

$$\widehat{\delta}_i = \frac{\widehat{\mathcal{T}}_k^i(X_{k-1}, X_k) - \eta_i \widehat{\mathcal{T}}_k^i(X_{k-1})}{\widehat{\mathcal{T}}_k^i(X_{k-1}^2)}, \quad (2.28)$$

$$\widehat{\eta}_i = \frac{\widehat{\mathcal{T}}_k^i(X_k) - \delta_i \widehat{\mathcal{T}}_k^i(X_{k-1})}{\widehat{\theta}_k^i}, \quad (2.29)$$

$$\widehat{\epsilon}_i^2 = \frac{\widehat{\mathcal{T}}_k^i(X_k^2) + \delta_i^2 \widehat{\mathcal{T}}_k^i(X_{k-1}^2) + \eta_i^2 \widehat{\mathcal{B}}_k^i + 2\eta_i^2 \delta_i \widehat{\mathcal{T}}_k^i(X_{k-1}) - 2\delta_i \widehat{\mathcal{T}}_k^i(X_{k-1}, X_k) - 2\eta_i \widehat{\mathcal{T}}_k^i(X_k)}{\widehat{\theta}_k^i}, \quad (2.30)$$

and

$$\widehat{\pi}_{ji} = \frac{\widehat{\mathcal{J}}_k^{ji}}{\widehat{\mathcal{O}}_k^i}. \quad (2.31)$$

Proof The proofs of (2.28)–(2.31) are outlined in Appendix A.

Remark 1: Having calculated δ_i , η_i and ϵ_i^2 in conjunction with (2.27), our proposed model in (2.8) is fully determined as $\{\widehat{\alpha}_i, \widehat{\xi}_i, \widehat{\theta}_i\}$ is explicitly available. The establishment of the recursive filters implies that parameter estimates are self-calibrating, i.e., they are updated automatically every time there is new information.

Remark 2: The filtering and parameter estimation procedure, as laid out above, is distinct from and more clear-cut than the one given in Erlwein and Mamon [19]. In this research work, the EM estimates are provided directly for each parameter appearing in the mean of the OU process; whereas in [19], the EM estimates rely first on the estimate of the entire drift component.

2.5 Numerical implementation

We implement the results from Section 2.4 on the four-year Toronto DATs (01 Jan 2011 – 31 Dec 2014 with 1461 data points) compiled by the NCDC. The seasonal component S_t is removed from T_t (see equation (2.3)) before applying the recursive filters to X_t . The data set is then grouped into batches of equal sizes to perform the filtering.

2.5.1 The deterministic component

The deterministic component S_t in equation (2.3) is fitted to the entire data set to obtain the estimates of its coefficients and arguments. Noting that S_t is a linear combination of sinusoidal functions, we use the built-in function of linear regression in the R software to accomplish the fitting process. This is complemented by a stepwise regression procedure to identify the dependent variable relevant to S_t . The R function ‘step’ is then applied to perform the selection of explanatory variables.

Mean	Std Deviation	Std Error	Min	Max	Skewness	Kurtosis
8.88	10.65	0.28	-19.25	31.5	-0.23	-0.92

Table 2.1: DATs' descriptive statistics

Parameter	Estimate	95% confidence interval
a	-0.0018	(-0.00235, -0.00127)
b	10.2111	(9.75806, 10.66405)
c_1	-5.2948	(-5.61554, -4.97416)
c_2	-0.6385	(-0.95550, -0.32341)
c_3	-0.4131	(-0.72794, -0.09821)
e_1	-12.7240	(-13.03831, -12.40979)

Table 2.2: Parameter estimates for the seasonality component S_t

Table 2.1 shows the DATs' descriptive statistics, which could guide in selecting initial values. Table 2.2 presents the parameter estimates of the deterministic S_t with an adjusted R-squared (coefficient of determination) of 0.835. This means 83.5% of the variation in the response T_t can be explained well by the model given the regressor variables in equation (2.4) except for e_2 and e_3 . Therefore, these variables were eliminated. The fitted seasonality component S_t is presented in Figure 2.1 together with the actual DATs. The characteristics of the plotted S_t are congruous with the temperature movement in both the summer and winter seasons. With the given values of T_t and S_t , X_t is calculated and then treated as the observation process for our filtering and parameter estimation implementation.

2.5.2 The stochastic component

The seasonal trend of the actual data is adequately captured even in the presence of X_t and some noise. The deseasonalised stochastic component $X_t = T_t - S_t$ from (2.3) is also depicted in Figure 2.2. We process X_t in 73 groups with 20 data points in each processing window, where $\Delta t = 1$ day. So, we are updating roughly every 3 weeks with the aid of the recursive filtering equations. Other filtering window sizes were also tested and we find that the size only slightly affects the results; similar outcomes are produced even with different window sizes.

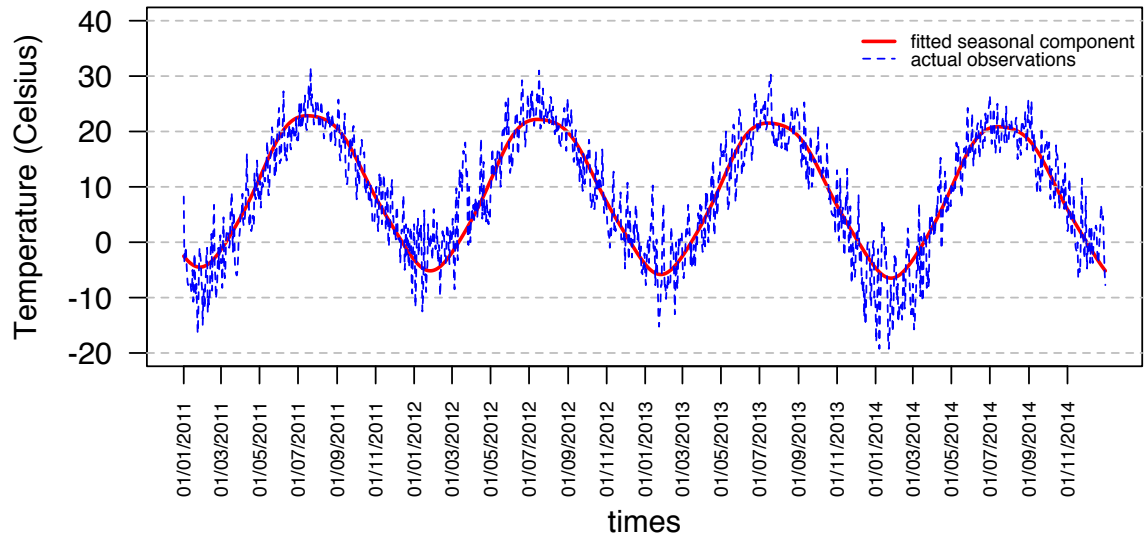
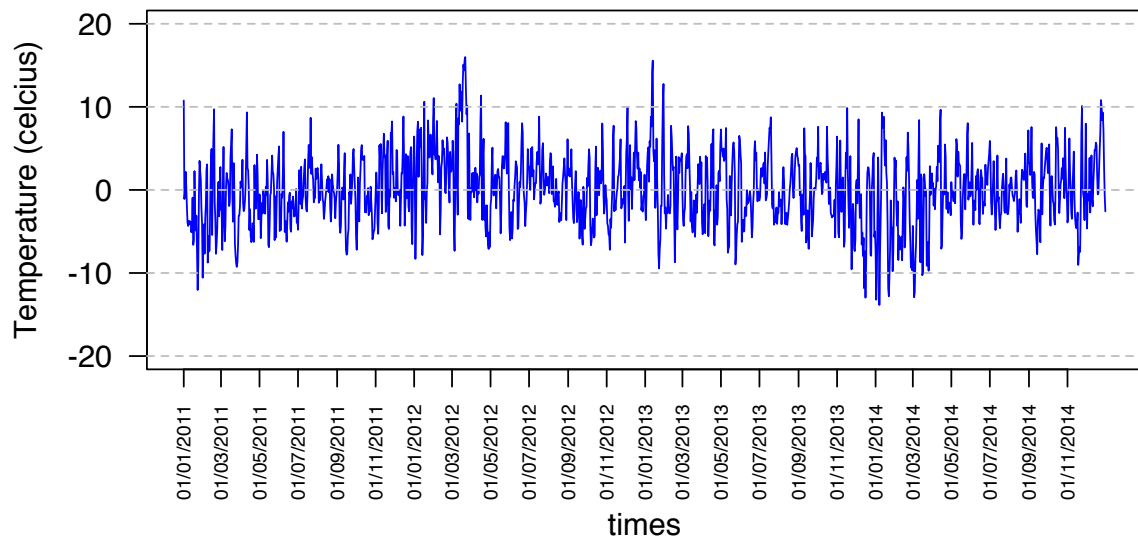


Figure 2.1: Fitted seasonal component and actual observations

Figure 2.2: Deseasonalised stochastic component X_t

2.5.2.1 Initial values for the parameter estimates

We assign $\frac{1}{N}$ for the initial transition probabilities. The other initial parameter values for the filtering equations are based on the parameter estimates for the X_t process under a single state. We consider the single-state version of (2.8), which is

$$X_{k+1} = \delta X_k + \eta + \epsilon z_{k+1}, \quad (2.32)$$

where

$$\delta = e^{-\alpha \Delta t_k}, \quad \eta = (1 - e^{-\alpha \Delta t_k}) \theta, \quad \epsilon^2 = \xi^2 \frac{1 - e^{-2\alpha \Delta t_k}}{2\alpha}.$$

Using equation (2.32), the likelihood function of X_k is

$$L(X_k; \delta, \eta, \epsilon) = \prod_{k=1}^m \frac{1}{\sqrt{2\pi\epsilon}} \exp\left(-\frac{(X_k - \eta - \delta X_{k-1})^2}{2\epsilon^2}\right), \quad (2.33)$$

where $1 \leq m \leq 1460$ in our case. Equation (2.33) entails the solution to

$$\operatorname{argmax}(\log L(X_k; \delta, \eta, \epsilon)) = \sum_{k=1}^m \left(\log \frac{1}{\sqrt{2\pi\epsilon}} - \frac{(X_k - \eta - \delta X_{k-1})^2}{2\epsilon^2} \right). \quad (2.34)$$

We use, with appropriate adjustments in our problem formulation, the R function ‘optim’ to solve equation (2.34). The results are $\widehat{\delta} = 0.6518$, $\widehat{\eta} = -0.001624$ and $\widehat{\epsilon} = 0.4271$. These are employed as benchmarks to select initial values for parameter estimates in the modelling framework with more than 1 regime. They serve as a good guide in launching the filtering recursions producing optimal values for various quantities leading to the EM parameter estimates of the model. We must also ensure that $\epsilon > 0$ to avoid bizarre outcomes when equation (2.34) is applied.

2.5.2.2 Evolution of parameter estimates

To process the data, we apply recursive filtering equations (2.22)-(2.24), and then use the results in 73 passes to feed into (2.28)-(2.31) to find the optimal parameter estimates. The implementation was made on two-state and three-state HMCs, and Figure 2.3 displays the movement, through each algorithm pass, of the optimal estimates for θ , α , and ξ under a 2-state set up. All three parameters exhibit convergence regardless of the choice of initial values provided they do not substantially deviate from the values suggested in our one-state ML estimation.

The choice of initial values could impact the speed of convergence, but eventually all parameter estimates approach certain unique values. The parameter estimates in the second state are characterised by lower mean-reverting speed, but higher mean-reverting level and volatility in comparison to those in state 1. Figure 2.4 depicts the transition probabilities with stable evolutions. A similar pattern in the evolution of parameter estimates for θ , α , and ξ under the 3-state model is shown in Figure 2.5. In this case, we see that the mean-reverting level and volatility in state 1 are the highest but with the lowest speed. The probability transition estimates π_{ji} are displayed in Figure 2.6. The behaviours of the parameter evolutions in a two-state setting are similar. Stability in the evolution of estimates for θ , α , ξ and Π is attained by the self-calibrating HMM algorithm after approximately 26 algorithm passes in both the 2-state and 3-state HMM settings. To quantify the variance of the various estimators, we calculate the Fisher information matrix $I(\mathbf{v}) = -\mathbb{E}_{\mathbf{v}}\left(\frac{\partial^2}{\partial \mathbf{v}^2} \log L(X; \mathbf{v})\right)$, which bounds the asymptotic variance of the MLEs and \mathbf{v} is a vector of parameters. The MLE is consistent and possesses an asymptotically normal sampling distribution [36]. Hence, we utilise the limiting distribution of the MLE for $\widehat{\mathbf{v}}$ to dynamically obtain the 95% confidence interval. For a generic (scalar) MLE $\widehat{v}_i \in \widehat{\mathbf{v}}$, this is given by $\widehat{v}_i \pm 1.96 \frac{1}{\sqrt{I(\widehat{\mathbf{v}})_{ii}}}$.

The entries of the Fisher information matrix are derived from the log-likelihood expressions given in the Appendix A and the results are summarised below.

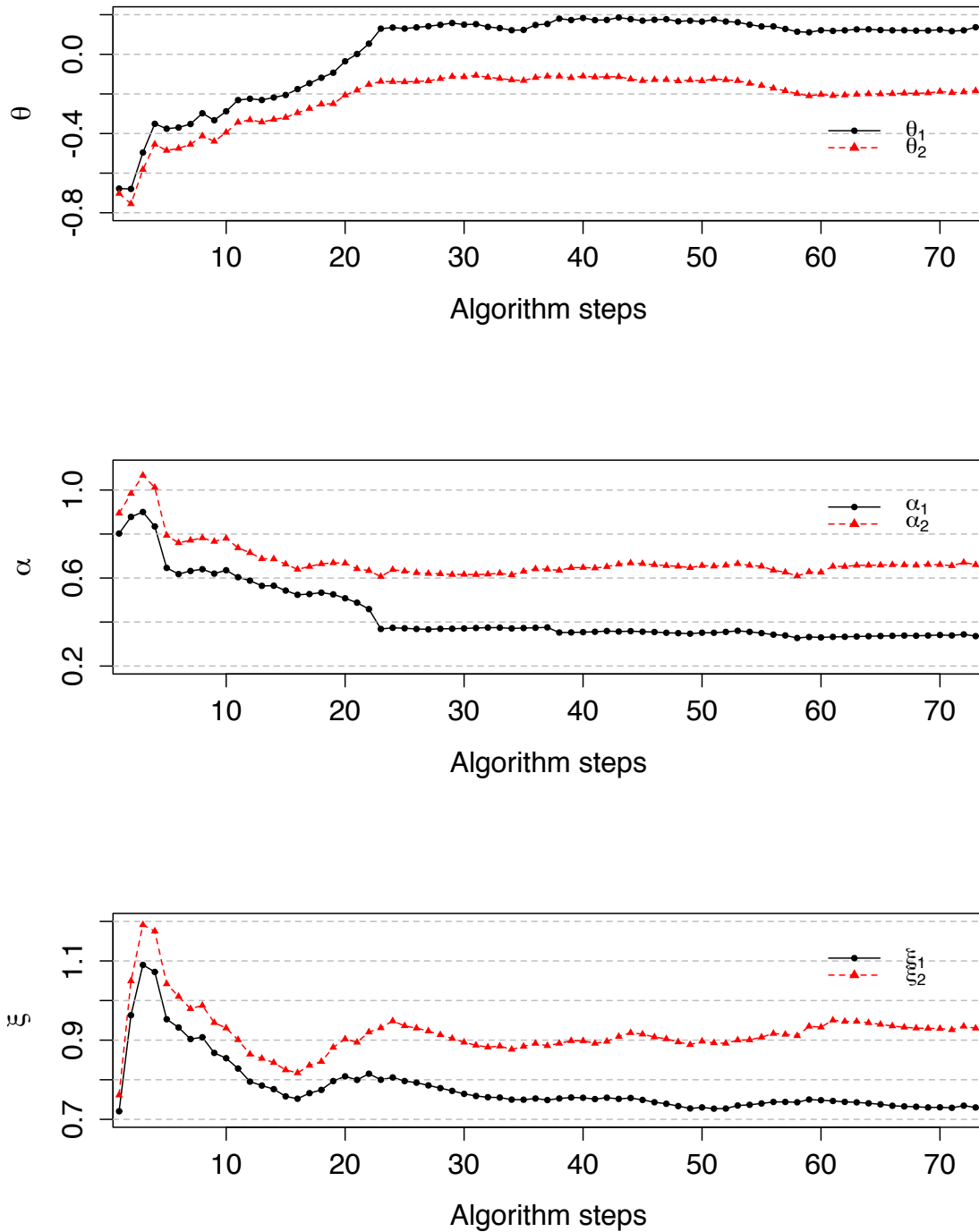
$$I(\pi_{ji}) = \frac{\widehat{\mathcal{F}}_k^{ji}}{\pi_{ji}^2}, \quad I(\delta_i) = \frac{\widehat{\mathcal{F}}_k^i(X_{k-1}^2)}{\epsilon_i^2}, \quad I(\eta_i) = \frac{\widehat{\mathcal{O}}_k^i}{\epsilon_i^2}, \quad \text{and}$$

$$I(\epsilon_i) = -\frac{\widehat{\mathcal{O}}_k^i}{\epsilon_i^2} + \frac{3}{\epsilon_i^4} \left(\widehat{\mathcal{F}}_k^i(X_k^2) + \widehat{\mathcal{O}}_k^i \vartheta_i^2 + \delta_i^2 \widehat{\mathcal{F}}_k^i(X_{k-1}^2) - 2 \widehat{\mathcal{F}}_k^i(X_k) \eta_i \right. \\ \left. - 2 \widehat{\mathcal{F}}_k^i(X_{k-1}, X_k) \delta_i + 2 \widehat{\mathcal{F}}_k^i(X_{k-1}) \eta_i \delta_i \right).$$

The standard errors (SEs) of all parameter estimates for the 1-, 2- and 3-state models are getting smaller as more algorithm passes are being run. The narrow ranges of the SEs shown in Table 2.3 indicate that precise estimates are achieved by the proposed estimation technique here (via HMM filtering and the EM algorithm).

2.5.3 Analysis of regime and model selection

In order to find out the best-fitting model that captures the dynamics of the process X_t , we perform a one-step ahead forecasting under the HMM setting, i.e., comparing the predictions of the stochastic component and DATs. This is supplemented with an AIC analysis

Figure 2.3: Evolution of parameter estimates for θ , α , and ξ under a 2-state model

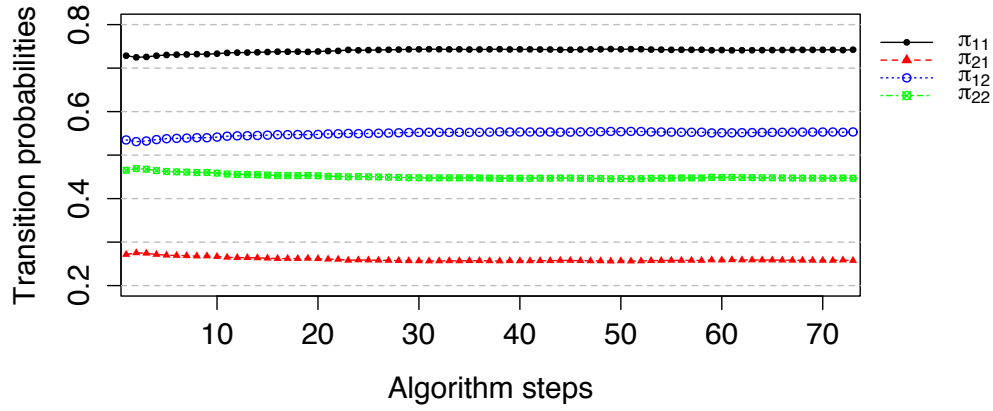
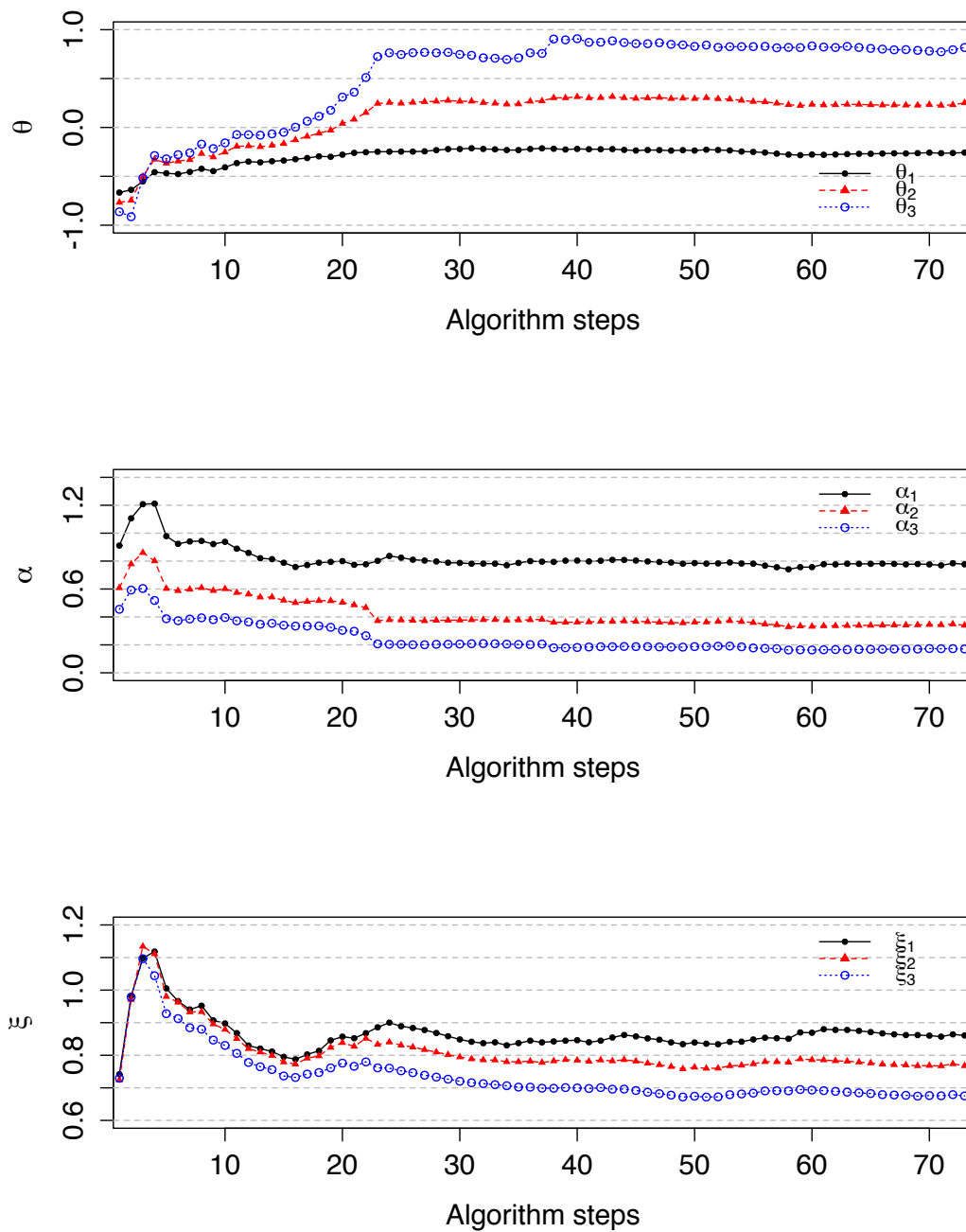


Figure 2.4: Evolution of transition probability under a 2-state model

Parameter Estimates	1-state model		2-state model		3-state model	
	Lower	Upper	Lower	Upper	Lower	Upper
$\widehat{\delta}_i$	3.62152×10^{-90}	0.84905	2.88315×10^{-93}	5.69605×10^{-2}	1.23351×10^{-89}	0.50619
$\widehat{\eta}_i$	3.13262×10^{-90}	0.81819	2.49571×10^{-93}	5.04276×10^{-2}	1.28888×10^{-89}	0.50535
$\widehat{\epsilon}_i$	2.21510×10^{-90}	0.54026	1.08302×10^{-93}	2.36474×10^{-2}	7.17870×10^{-90}	0.34230
$\widehat{\pi}_{ji}$	4.79291×10^{-90}	1.08185	1.24810×10^{-93}	3.60257×10^{-2}	3.52893×10^{-90}	0.34041

Table 2.3: Interval of standard errors for parameter estimates under 1-, 2- and 3-state models

Figure 2.5: Evolution of parameter estimates for θ , α , and ξ under a 3-state model

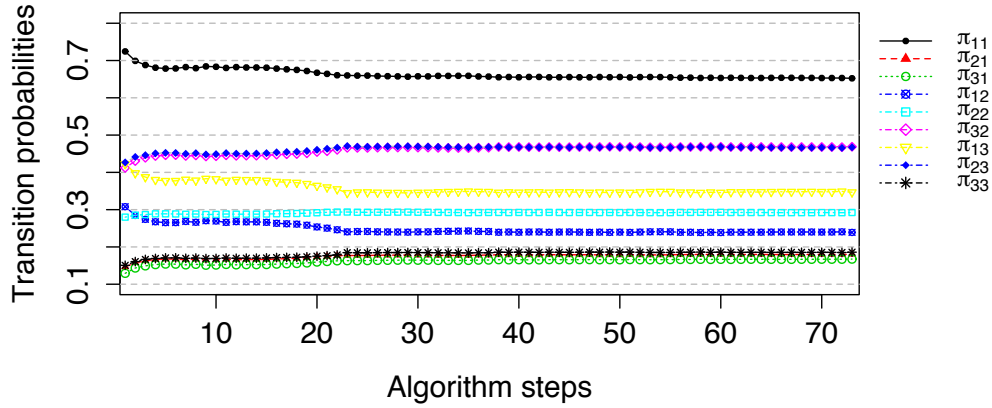


Figure 2.6: Evolution of transition probability under a 3-state model

as a guide in choosing the optimal number of states.

2.5.3.1 Assessment of predicted temperature values

We evaluate the one-step ahead predictions for DATs as an additional diagnostic for model fitting. The expected value of the X_{k+1} given the available information up to time k provides the forecast

$$\mathbb{E}(X_{k+1} | \mathcal{X}_k) = \mathbb{E}(\delta(\mathbf{y}_k) X_k + \eta(\mathbf{y}_k) + \epsilon(\mathbf{y}_k) z_{k+1} | \mathcal{X}_k) = \langle \delta, \widehat{\mathbf{y}}_k \rangle X_k + \langle \eta, \widehat{\mathbf{y}}_k \rangle. \quad (2.35)$$

Figure 2.7 illustrates the one-step ahead forecasts for the deseasonalised process X_k vis-à-vis DATs under a 2-state HMM set-up in accordance with (2.35).

A comparison of the observed seasonal HDD and the expected seasonal HDD for the entire period (2011-2014) under the 2-state model is presented in Figure 2.8. Clearly, the expected seasonal HDD obtained from the model follows closely the actual seasonal HDD. Furthermore, we examine an upward 6-month period, in a magnified view of forecasting, of the DATs under all 3 states settings as shown in Figure 2.9. Visually, all forecasts are quite close to the actual DATs and the temperatures' trends and dynamics are captured well by our proposed self-calibrating estimation approach.

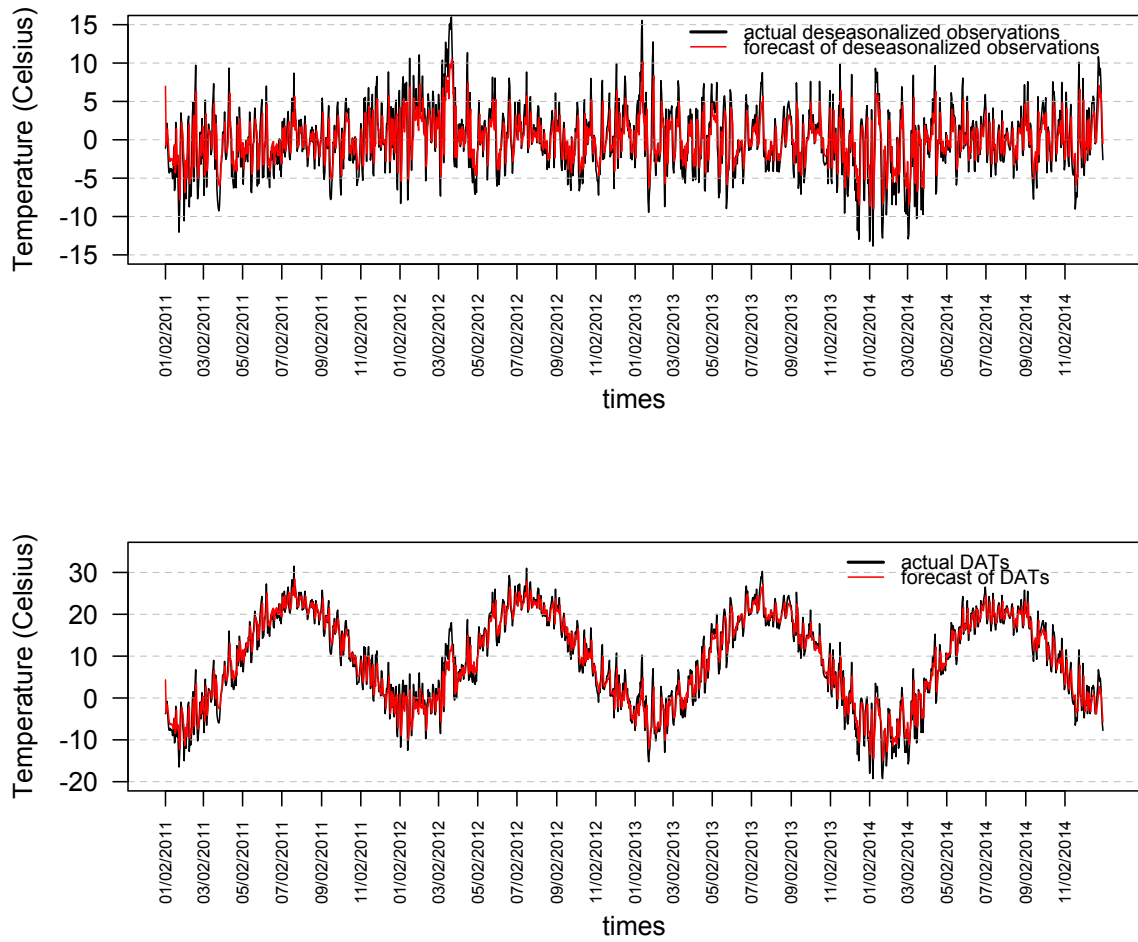


Figure 2.7: One-step ahead forecasts under a 2-state model

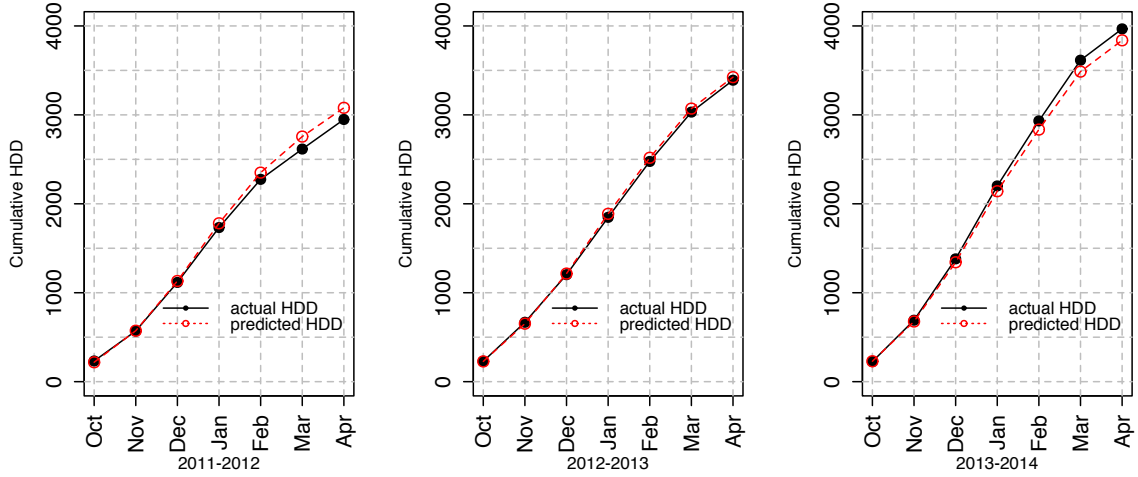


Figure 2.8: Comparison of the expected HDD and actual HDD in a 2-state model

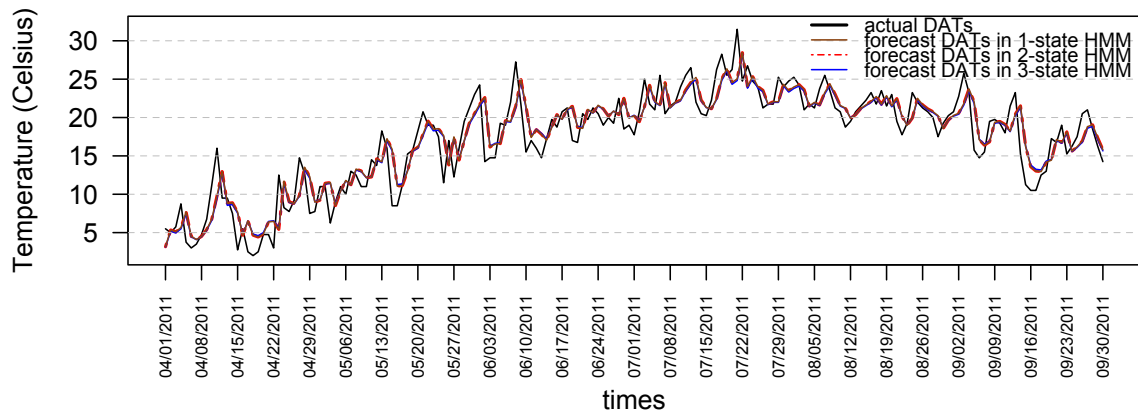


Figure 2.9: Comparison of one step ahead forecasts in 1-, 2-, and 3-state HMMs

HMM setting	Deseasonalised X_k			DATs T_k		
	1-state	2-state	3-state	1-state	2-state	3-state
MSE	0.42702	0.42700	0.42757	10.67538	10.67504	10.68929
RMSE	0.65346	0.65345	0.65389	3.26732	3.26727	3.26945
MAE	0.50467	0.50465	0.50584	2.52334	2.52324	2.52921
RAE	0.74435	0.74432	0.74608	0.27539	0.27538	0.27603

Table 2.4: Results of error analysis for X_k and T_k under different model settings

2.5.3.2 Likelihood-based model selection and error analysis

Visual inspection is not sufficient in deciding which model does the job well especially when the actual values are very close to the predicted values under several competing models. We, therefore, provide quantitative approaches in gauging the ‘best-performing’ model on the basis of information-criterion and minimised-error metrics. Following [19], we examine the mean-squared error (MSE), root-mean-squared error (RMSE), absolute-mean error (MAE), and relative-absolute error (RAE) in assessing the one-step ahead predictions under the 1-, 2- and 3-state HMMs. Suppose, $k = 1, 2, \dots$, X_k is the actual value of the process at time k , \widehat{X}_k is the predicted value at time k given available data up to time $k - 1$, \bar{X} is the mean of all X_k ’s, and $K = 1460$ is the total number of predicted observations. Then the error metrics are computed as

$$\text{MSE} = \frac{\sum_{k=1}^K (\widehat{X}_k - X_k)^2}{K}, \quad \text{RMSE} = \sqrt{\frac{\sum_{k=1}^K (\widehat{X}_k - X_k)^2}{K}},$$

$$\text{MAE} = \frac{\sum_{k=1}^K |\widehat{X}_k - X_k|}{K}, \quad \text{and} \quad \text{RAE} = \frac{\sum_{k=1}^K |\widehat{X}_k - X_k|}{\sum_{k=1}^K |\widehat{X}_k - \bar{X}|}.$$

Table 2.4 displays the various error values concerning the one-step ahead predictions for the model’s stochastic component X_k and observed DATs T_k . Whilst the differences in errors are generally small, the 2-state model outperforms the 1- and 3-state models under all 4 error measures. We also investigated the 4-state HMM but found no evidence of even further minimal improvements. Increasing the number of regimes may outweigh the benefit of model flexibility as there is a corresponding increase as well in the parameters to be estimated.

To evaluate the statistical significance for the mean difference of errors, a t -test involving three pairs of model settings is conducted on the RMSEs generated using the bootstrapping method. Firstly, a bootstrapped sample of the same size 1460 is generated by sampling with replacement from the original squared errors for each model setting. Secondly, the square root of the average squared errors for every bootstrapped sample is computed. After running the two-step bootstrapped procedure 10,000 times, the bootstrapped RMSEs are then produced for the paired t -test.

We report in Table 2.5 the Bonferroni-corrected p -values accounting for the pairwise settings. With a significance level of 5%, we see that the results for the 1- and 2-state settings are statistically different and the same conclusion holds for the results involving 1- and 3-state settings. The results for the 2- and 3-state settings though are not statistically different. This implies two things: (1) there is a benefit in adding regime-switching features to the model setting; and (2) a 2-state model is sufficient for the DATs series under examination. These findings support the observations in Section 2.5.3.1.

	1-state	1-state	2-state
t -test	vs	vs	vs
	2-state	3-state	3-state
p -value	0.01259	7.06352×10^{-5}	0.57193

Table 2.5: Bonferroni-adjusted p -values of a t -test applied on bootstrapped-generated RMSEs

To balance the model's goodness of fit and complexity, we also adopt the Akaike Information Criterion (AIC) in choosing the 'best-performing' model. The AIC is calculated as

$$\text{AIC} = -2 \log L(X; \mathbf{v}) + 2\omega, \quad (2.36)$$

where ω is the number of estimated parameters included in the model, and $\log L(X; \mathbf{v})$ denotes the model's loglikelihood evaluated using the data X and a set of parameters \mathbf{v} . The preferred model has the lowest AIC value. In terms of the observed X_k 's and parameter set \mathbf{v} , the loglikelihood is given by

$$\log L(X_k; \mathbf{v}) = \sum_{k=1}^B \left(\log \left(\frac{1}{\sqrt{2\pi}\epsilon(\mathbf{y}_{k-1})} \right) - \frac{X_k - \eta(\mathbf{y}_{k-1}) - \delta(\mathbf{y}_{k-1})}{2\epsilon^2(\mathbf{y}_{k-1})} \right), \quad (2.37)$$

where B is the number of observations in each algorithm pass.

From equation (2.36), a model that has high loglikelihood is favoured and a complex model (in the sense of having many parameters) is penalised. Looking at (2.11) and Π , we obtain the number of parameters to be estimated for each model setting and this is shown in Table 2.6. These inputs are necessary to complete the calculation of (2.36).

Setting	1-state	2-state	3-state	N -state
No. of parameters	3	8	15	$3N + (N - 1)N$

Table 2.6: Number of parameters to be estimated under HMM

With the aid of equation (2.37) and Table 2.6, we calculated the AIC values in Table 2.7 for the HMM settings as well as the setting with no regime-switching. Table 2.7 shows that both the 2- and 3-state HMM settings have lower AIC values than that of the 1-state model. It indicates that there is merit in embedding the regime-switching feature into the model. Also, the 2-state HMM model produces the lowest AIC value. So, this result together with the error analysis leads us to conclude that the 2-state HMM setting is the most suitable for our data set under examination.

Model	Number of parameters	AIC
1-state model	3	2899.7982
2-state HMM	8	2888.4031
3-state HMM	15	2897.7614

Table 2.7: AIC analysis

2.6 Application to the pricing of temperature-dependent contracts

The parameter estimation and error analysis performed in the previous sections indicate that a 2-state HMM framework presents a better performance in capturing the dynamics of DATs and HDD index. Therefore, we shall utilise the optimal estimates generated through the filtering recursions in pricing weather contracts. In this section, mathematical expressions for the price of HDD futures and options are derived with the consideration of the price of risk. Furthermore, a numerical example on pricing a European-call option via simulation is presented. A sensitivity analysis, under a 2-state HMM setting, is

Regions	Contract types	Contract periods	One contract size
North America (US & Canada)	Monthly HDD	Oct-Apr	\$20
	Monthly CDD	Apr-Oct	
	Seasonal HDD	Nov-Mar & Dec-Feb	
	Seasonal CDD	May-Sept & Jul-Aug	
Europe (Schengen Areas & UK)	Monthly CAT	Apr-Oct	€20 or £20
	Seasonal CAT	2-7 consecutive months from Apr-Oct	

Table 2.8: Specifications of typical CME temperature-based contracts

also implemented to examine the extent of each parameter's influence to the value of a temperature-based derivative.

2.6.1 Temperature-based derivatives

Although there are various underlying variables for weather derivatives such as temperature, air pressure, humidity, etc, the most commonly used underlying is a temperature-based index. Temperature-based futures and options are the main financial products trading in the CME covering many regions and countries around the globe. These temperature index-linked products quantify weather and make it as tradeable as some other regular underlying assets such as stocks and commodities. They are currently geared to average monthly and seasonal weather in several north American and European cities.

In the United States and Canada, the CME weather contracts are mainly written on the HDD or CDD index with a monthly or seasonal duration, whilst weather derivatives in Europe are normally based on the CAT index. Table 2.8 summarises the specifications of typical CME temperature-based futures and options. These products written on accumulated indices over a calendar duration are financially settled at the end of the index or contract period. The most common option style so far is a European-type option in both the OTC market and CME, which may only be exercised at its expiration date. A cap on the maximum payout is normally set for OTC weather contracts, whereas there is currently no such limit for those trading in the CME. In contrast to temperature-based option contracts, regular options traded in the CME typically specify index futures as the underlying variables. A standardised temperature-based contract normally contains five basic elements listed in Table 2.9.

Element	Description
Underlying variable	A temperature-based index or index futures
Contract period	Accumulated and consecutive calendar period
Contract (tick) size	A dollar amount attached to each HDD, CDD or CAT point
Land-based station	A specific station that reports observations including temperature for a particular area
Payoff function	A function that converts a tradeable temperature-based index into a cash flow in terms of the above four elements.

Table 2.9: Elements of a standardised temperature-based contract

2.6.2 Pricing of temperature-based index futures and options

In the literature, there are two mainstream approaches in modelling the underlying geared towards the pricing of temperature-driven contracts. The first approach is to employ a continuous-time process modelling the DATs. For instance, Bellini [3] proposed an extended OU model for DATs and to price HDD/CDD contracts assuming a Lévy-type noise process. Benth and Šaltytė-Benth [5] computed futures and option prices using a continuous-time autoregressive model for the temperature dynamics under a normality assumption. The second approach is to adopt a discrete-time process to model DATs considering that temperature indices are defined as discrete sums of DATs rather than integrals over a specific period. Moreno [30] demonstrated that it is not necessary to estimate DATs continuously and the goodness of fit using a discrete-time process is deemed satisfactory.

The first approach facilitates various computations for valuation via stochastic calculus and the latter one is mostly used in the actuarial field under a discrete-time framework. The pricing conducted in this chapter employs a continuous-time OU process but such a process is discretised in order to estimate the model parameters. Once parameter estimates are available, as in our case, any temperature-based derivative can be theoretically valued by taking the expectation of its discounted payoff at contract's maturity. We derive the pricing formulae for temperature-based futures and options, and include a specific pricing example via simulation.

The market of weather derivatives is relatively less liquid and this complication is exacerbated as well by the fact that the temperature itself cannot be either stored or traded. A zero risk premium could not be necessarily assumed because weather contract prices are

not independent of the regime shifts in our model and investors' risk aversions especially in such a somewhat illiquid market. Hence, the price of risk process λ_t of the underlying variable ought to be considered in order to determine a fair price for a temperature-based contract.

2.6.3 The price of risk and risk-neutral measure Q

In a complete market, there is a unique risk-neutral probability measure equivalent to the real-world probability measure, where the price of any contingent claim is the expected value of its payoff discounted at the risk-free rate. However, the weather-derivatives market is incomplete because the underlying variable (temperature index) is not storable or tradable. This hampers a straightforward adoption of the no-arbitrage arguments to price weather-derivative instruments. Notwithstanding this difficulty, a risk-neutral valuation can still be utilised in incomplete markets; see Xu et al. [38]. In an incomplete market, there are several equivalent-martingale or risk-neutral measures to choose from. More specifically, there is a risk-neutral measure Q equivalent to the objective measure P for pricing weather derivatives; see [31]. The rationale in selecting or constructing this Q is detailed in the subsequent discussion.

We first note that there is also an issue regarding the risk-neutral valuation for illiquid and incomplete markets such as the market for weather derivatives. This arises from the lack of observed market prices that supposedly could determine the risk-neutral parameters and the martingale measure Q . In any event, the price obtained under Q can be regarded as a benchmark similar to the idea of pricing other exotic and structured products not yet available in the market. This “benchmarking principle” was espoused in Gao et al. [25] in which the risk-neutral valuation was still employed to get no-arbitrage prices for a guaranteed annuity option; this was done even though market information is insufficient to reflect certain conditions concerning insurance-related contingent claims. Moreover, the risk-neutral valuation was applied to price a product dependent on an underlying (i.e., mortality rate) that is also non-tradeable and storable.

Akin to the pricing under Q is the concept of the market price of risk λ_t at time t , which links the P to the Q -dynamics of the underlying variable. The quantity λ_t could be assumed zero if only to attain simplicity and tractability. But, some recent findings indicate that λ_t associated with the temperature variable is significantly different from 0; see Xu et al. [38],

Benth et al. [4], and Elias et al. [14]. In this chapter, we shall investigate the impact of λ_t in the pricing of weather contracts. This in turn requires an explicit expression for T_t under Q . The formal justification for a risk-neutral measure Q utilised for weather-derivative pricing in our set up follows from Weron's work [31], which is anchored intimately on λ_t that is assumed a real-valued measurable and bounded process.

Through the Girsanov's theorem, there are probability measures Q and Q^λ equivalent to P , such that

$$B_t^{Q^\lambda} = B_t + \int_0^t \lambda_s ds = B_t + \lambda_t t, \quad (2.38)$$

where B^{Q^λ} is a standard Brownian motion under Q^λ and B_t is a P -Brownian motion driving our deseasonalised process X_t in equation (2.5). Benth and Šaltytė-Benth [5] stated the existence of a flexible class of risk-neutral probabilities (e.g., Q^λ) that can be obtained from the time-varying λ_t to fit the forward curves. As Q^λ is a risk-adjusted or risk-neutral probability measure (see page 151 of [31] and page 173 of [2]), this signifies that Q^λ coincide with the needed Q for the purpose of our valuation.

Referring to the construction arguments in [31], we obtain the condition $\theta^Q = \theta - \frac{\lambda_t \xi}{\alpha}$, where θ^Q is defined under Q . So extending [31] to our framework, λ_t links the parameters of our proposed model through

$$\theta^*(\mathbf{y}_t) = \theta(\mathbf{y}_t) - \frac{\lambda_t \xi(\mathbf{y}_t)}{\alpha(\mathbf{y}_t)}, \quad (2.39)$$

where $\alpha(\mathbf{y}_t)$, $\xi(\mathbf{y}_t)$, and $\theta(\mathbf{y}_t)$ are given in equation (2.11). Here, $\theta^*(\mathbf{y}_t)$ is the new parameter under Q that replaces $\theta(\mathbf{y}_t)$ in the OU process after taking into account λ_t , which encapsulates both the (i) price of risk due to regime switching and (ii) market price of risk (due to the shocks from B_t).

Remark 3: The introduction of Q essentially introduces a market price of risk parameter λ_t . As such, we refer to λ_t as a form of an 'insurance premium' for taking uncertainty risks as notably identified in (i) and (ii) above.

A rigorous construction of a regime-switching λ_t under a discrete-time OU process in the context of bond pricing is given in [20]. Such construction in our option-pricing framework is a separate issue and not within the scope of this research. For illustrative intent, we assume that λ_t is a non-zero constant for all regimes in the HMM setting, but we shall

investigate the effect of its variation in the succeeding sensitivity analysis. For weather-derivative valuation, we utilise

$$dX_t = \alpha(\mathbf{y}_t)(\theta^*(\mathbf{y}_t) - X_t)dt + \xi(\mathbf{y}_t)dB_t^Q \quad (2.40)$$

and

$$dT_t = dS_t + \alpha(\mathbf{y}_t)(\theta^*(\mathbf{y}_t) - X_t)dt + \xi(\mathbf{y}_t)dB_t^Q. \quad (2.41)$$

Consequently, we have the explicit form

$$T_t = S_t + X_s e^{-\alpha(\mathbf{y}_t)(t-s)} + (1 - e^{-\alpha(\mathbf{y}_t)(t-s)})\theta^*(\mathbf{y}_t) + \xi(\mathbf{y}_t)e^{-\alpha(\mathbf{y}_t)t} \int_s^t e^{\alpha(\mathbf{y}_u)u} dB_u^Q. \quad (2.42)$$

As asserted in Geman and Leonardi [26], the hypothesis for accumulated-degree days in pricing weather contracts on temperature indices being normally distributed is accepted by both researchers and practitioners. In Benth and Šaltytė-Benth [5], the DATs are modelled by an OU process having a seasonal volatility, and the HDD/CDD-index pricing was accomplished under the normality assumption. More recently, Alexandridis and Zapranis [2] verified the suitability of such a normality assumption. Under the same assumption together with the principles employed in [2] and [5], we derive pricing formulae for HDD futures and options under the OU-HMM settings.

It is assumed that the contract's period is $[\tau_1, \tau_2]$, the contract size is 1 monetary unit for each index point, and the risk-free interest rate r_f is constant with continuous compounding. The HDD futures price $F_H(t, \tau_1, \tau_2)$ is evaluated under two scenarios: (i) at time t before the contract period, i.e., $0 \leq t \leq \tau_1 < \tau_2$, and (ii) at a time t within the contract period, i.e., $0 \leq \tau_1 \leq t < \tau_2$.

From Benth and Šaltytė-Benth's reasoning [5], $e^{-r_f(\tau_2-t)}\mathbb{E}^Q\left[\int_{\tau_1}^{\tau_2} \max(T_{\text{base}} - T_v, 0) dv - F_H(t, \tau_1, \tau_2) \middle| \mathcal{F}_t\right] = 0$; and so the HDD futures price, as an \mathcal{F}_t -adapted stochastic process, must be defined as

$$F_H(t, \tau_1, \tau_2) = \mathbb{E}^Q \left[\int_{\tau_1}^{\tau_2} \max(T_{\text{base}} - T_v, 0) dv \middle| \mathcal{F}_t \right]. \quad (2.43)$$

For the European HDD call option, we consider two types: a call dependent on the HDD itself and a call dependent on the HDD futures price. The pricing results, whose proofs are given in Appendix B, are presented in Propositions 2.6.1–2.6.3 below under the model described in (2.41).

Proposition 2.6.1 *The HDD futures price for the contract period $0 \leq \tau_1 < \tau_2$ is given by*

$$F_H(t, \tau_1, \tau_2) = \mathbb{E}^Q[H | \mathcal{F}_t] = \begin{cases} \int_{\tau_1}^{\tau_2} A(t, v) D\left(\frac{M(t, v, X_t)}{A(t, v)}\right) dv & 0 \leq t \leq \tau_1 < \tau_2 \\ \int_{\tau_1}^t \max(T_{base} - T_u, 0) du & 0 \leq \tau_1 \leq t < \tau_2, \\ + \int_t^{\tau_2} A(t, v) D\left(\frac{M(t, v, X_t)}{A(t, v)}\right) dv \end{cases}$$

where

$$H = \int_{\tau_1}^{\tau_2} \max(T_{base} - T_v, 0) dv,$$

$$M(t, v, X_t) = T_{base} - S_v - X_t e^{\alpha(\mathbf{y}_t)(v-t)} + (1 - e^{\alpha(\mathbf{y}_t)(v-t)})\theta^*(\mathbf{y}_t),$$

$$\text{and } A^2(t, v) = \xi^2(\mathbf{y}_t) e^{2\alpha(\mathbf{y}_t)v} \int_t^v e^{2\alpha(\mathbf{y}_t)u} du, \quad D(x) = x\Phi(x) + \phi(x)$$

with Φ and ϕ are the cumulative and probability density functions of a standard normal distribution, respectively.

Proposition 2.6.2 *The price of a European call option written on an HDD futures price at time t is given by*

$$C_{F_H}(t, K, \tau_T) = e^{-r(\tau_T-t)} \int_K^{F_m} (F_H - K) g(F_H) dF_H,$$

where

$$F_H = \int_{\tau_1}^{\tau_2} A(t, v) D\left(\frac{M(t, v, X_t)}{A(t, v)}\right) dv, \quad F_m = \max(F_H),$$

τ_T is the expiration date, $t \leq \tau_T \leq \tau_1 \leq \tau_2$, K is the strike price, r is the risk-free interest rate, and $g(x)$ is the probability density function of a random variable X .

Proposition 2.6.3 *The price of a European call option written on HDD at time t is given by*

$$C_H(t, K, \tau_T) = e^{-r(\tau_T-t)} \int_K^{H_m} (H - K) g(H) dH,$$

where

$$H = \int_t^{\tau_T} \max(T_{\text{base}} - T_v, 0) dv, \quad H_m = \max(H),$$

τ_T is the expiration date, $t \leq \tau_T$, K is the strike price, r is the risk-free interest rate, and $g(x)$ is the probability density function of a random variable X .

2.6.4 Pricing of a temperature-based contract

When pricing HDD options in Alaton et al. [1], the distribution of HDD is Gaussian and the probability of $\max(T_{\text{base}} - T_v, 0) = 0$ is close to zero during the winter period. Adopting this assumption to Proposition 2.6.3 with M_c and A_c^2 for the mean and variance of H , respectively, setting H_m to infinity, and employing the method in deriving the result in Proposition 2.6.1, we have

$$\begin{aligned} C_H(t, K, \tau_T) &= e^{-r(\tau_T-t)} \int_K^{H_m} (H - K) g(H) dH \\ &= e^{-r(\tau_T-t)} \left[(M_c - K) \Phi\left(\frac{M_c - K}{A_c}\right) + \phi\left(\frac{M_c - K}{A_c}\right) \right] \\ &= e^{-r(\tau_T-t)} A_c D\left(\frac{M_c - K}{A_c}\right). \end{aligned} \quad (2.44)$$

The normal distribution for modelling HDD provides a good fit in several winter months for almost all locations [28]. However, it is almost impossible to obtain a good fit in modelling the HDD in other months with the use of the normal distribution most especially in December when cold temperatures give rise to heavy tails in the empirical distribution of the HDD. Therefore, this assumption is not quite realistic for Proposition 2.6.3 in pricing a one-month weather option contract. Although an analytic valuation formula can be generated due to the simplicity of a Gaussian distribution, such distribution is probably going to be adequate only in valuing a seasonal HDD option rather than a contract for one month.

To compute the prices of futures and options written on HDD, we note the path-dependent nature of the expected payoff, which can be calculated efficiently using the Monte-Carlo (MC) simulation method. We perform the valuation under a 2-state HMM model and include as well the price of the risk process. In particular, we evaluate

$$F_{H^d}^d(t, \tau_1, \tau_2) = V \mathbb{E}^Q [H^d | \mathcal{F}_t], \quad (2.45)$$

$$C_{H^d}^d(t, K, \tau_T) = e^{-r(\tau_T - t)} V \mathbb{E}^Q \left[\max(H^d - K, 0) \mid \mathcal{F}_t \right], \quad (2.46)$$

where V is the contract size, and H^d is defined as $\sum_{\tau_1}^{\tau_2} \max(T_{\text{base}} - T, 0)$ in (2.45) and $\sum_t^{\tau_T} \max(T_{\text{base}} - T, 0)$ in (2.46) consistent with the respective notations employed in Propositions 2.6.1 and 2.6.3.

We examine closely the valuation of HDD call options through a sensitivity analysis. Also, in order to make a valid comparison between the option prices generated under our model and the results in Elias et al. [14], we make certain conditions uniform between the contract in our study and that in [14]. Consider a European call with effectivity of 01 – 31 Dec 2012, and $t = 0$, $V = \$20$, and $K = \$580$ in equation (2.46). We set $r = 1\%$, which is the 1-year T-bill's yield as published by the Bank of Canada in 2012, to proxy the interest rate. Having optimal estimates for $\theta(\mathbf{y}_k)$, $\alpha(\mathbf{y}_k)$, and $\xi(\mathbf{y}_k)$ in Section 2.5, and granting λ_t is known or assigned, a value for $\theta^*(\mathbf{y}_t)$ in accordance with (2.39) is immediate. Moreover, T and H^d can be calculated based on equations (2.41) and (2.46), respectively. The optimal estimates for parameters to compute T and H^d are shown in Table 2.10.

Unlike a regular underlying variable, temperature varies geographically over swaths of a particular region, and this further adds to the difficulty of determining the appropriate level of λ_t . The market price of risk associated with the temperature variable is deemed significant in pricing weather futures and options. Risk premia, across a range of risk aversions, for several cities in America were established in terms of the contemporaneous correlation levels between aggregate dividends and temperature in [9]. Here, aggregate dividends refer to the total consumption of investors so that temperature-related contingent claims can be priced by considering the corresponding payoff and the utility function on consumption. In particular, risk premia for temperature-dependent call options in Chicago and New York were estimated to be 5.73% and 7.63% respectively.

Assuming that the risk premium solely depends on the market price of risk associated with the underlying variable, the average of the two risk premium values for the two cities (Chicago and New York) located close to Toronto, which is $\lambda_t = 6.68\%$, is utilised in our illustrative example.

In their pricing calculations, Elias et al. [14] assumed $r = 5\%$ in 2008, and used an expected

Parameter estimate	$\widehat{\delta}$	$\widehat{\eta}$	$\widehat{\epsilon}$	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{\xi}$	$\widehat{\theta}^*$
Regime 1	0.7146	0.0390	0.6229	0.1367	0.3361	0.7301	-0.0084
Regime 2	0.5166	-0.0889	0.6929	-0.1839	0.6604	0.9301	-0.2779

Table 2.10: Optimal parameter estimates for the 2-state HMM model with and without the market price of risk

HDD in their payoff function. In our case, after simulating the underlying variable H^d using the optimal estimates in Table 2.10, we perform 10,000 simulations to evaluate the expected payoff in (2.46). Then, by embedding λ_t into a 2-state HMM-based model, we get the call option price written on HDD as

$$C_{H^d}^d \left(0, 580, \frac{31}{365} \right) = e^{-0.01 \frac{31}{365}} 20 \mathbb{E} \mathcal{Q} \left[\max(H^d - 580, 0) \middle| \mathcal{F}_0 \right] = \$861.33. \quad (2.47)$$

If λ_t is not considered in equation (2.47), the price is just \$507.53. The SEs of the MC simulations with and without λ_t are 8.9369 and 6.3323 respectively, and the corresponding 95% confidence intervals are [843.81, 878.85] and [495.12, 519.94]. In this case, 10,000 simulations are sufficient to obtain a desired level of accuracy at the significance level of 5%.

Figure 2.10 exhibits the variation of the HDD option prices with and without the price of risk process, over a range of strike HDD. As the strike price K increases from 540 to 620, both option prices decline. They reach zero at $K = 623.10$ and $K = 605.49$ respectively. Indeed, the price of risk process does make a substantial difference in pricing an HDD call option for each strike HDD.

2.6.5 Sensitivity analysis of temperature-based option price

Sensitivity analysis entails the variation of one model parameter over a certain range, whilst holding the rest of the parameters fixed, in order to see the impact on the option price. Similar to the idea in Siu et.al [33], we shall do such an analysis under a 2-state HMM model, which turned out to be the best-fitting model given our data set. Using the pricing example highlighted in Section 2.6.4, we perform sensitivity analyses on the price of risk process, Markov chain's intensity matrix, and other model parameters.

Figure 2.11 shows the influence of λ_t on the HDD call option price by changing its value from 0 to 20%. The option price rises with an increasing λ_t , which is consistent with the

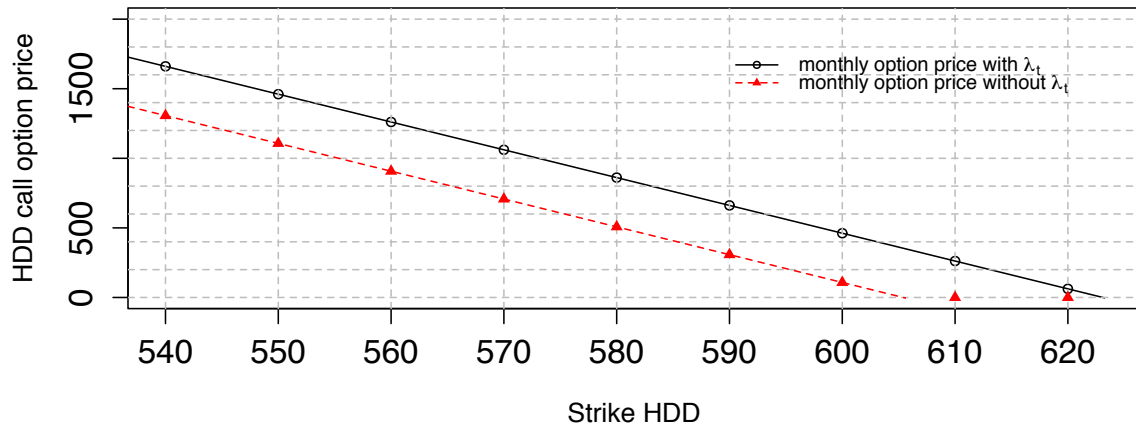


Figure 2.10: One-month call option price with initial state 1 versus HDD strike

results from Elias et. al [14] and it constitutes a significant portion of the temperature-based derivative's price.

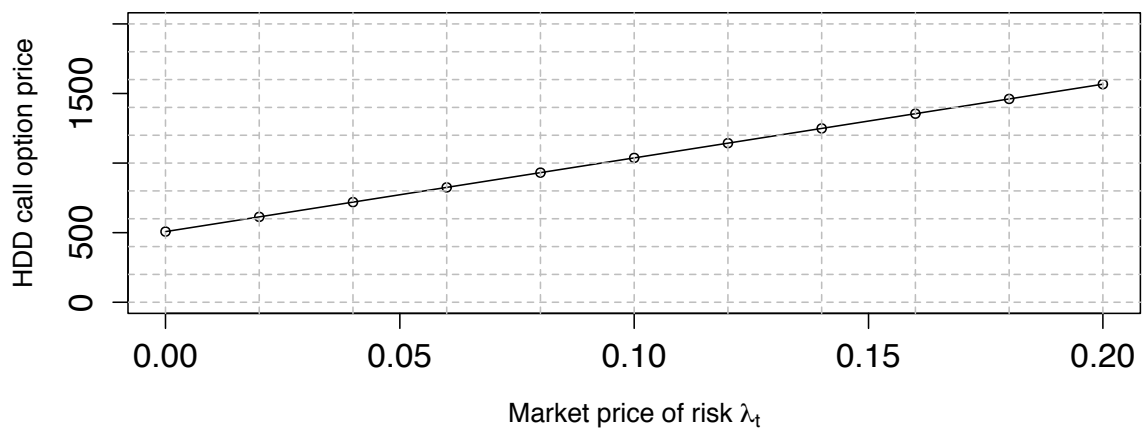


Figure 2.11: HDD call option price with initial state 1 versus market price of risk

To easily interpret our numerical results, we denote states 1 and 2 as “hot” and “cold” regimes, respectively, corresponding to the hot and cold fronts in the 2-state HMM-temperature

model setting. The intensity matrix \mathbf{P} is then given by

$$\mathbf{P} = \begin{pmatrix} -p & p \\ p & -p \end{pmatrix},$$

where p , the intensity parameter, is the rate of entering into/leaving from one state to another. The use of the matrix \mathbf{P} facilitates in probing the option-price sensitivity associated with the jump probability of the Markov chain as we only have one parameter to focus on, which is p in this case. For such a purpose, a convenient and practical representation of matrix $\mathbf{\Pi}$, defined in (2.9), is

$$\mathbf{\Pi} = \exp(\mathbf{P}) = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2} e^{-2p} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \quad (2.48)$$

In our numerical demonstration, the values of p are set to 0.1, 0.3, 0.5, 0.7, and 0.9 and our simulations consider starting states 1 and 2 at $t = 0$ separately. The tremendous impact of the intensity parameter is shown in Table 2.11; a comparison with the pricing results for the HDD call options with no-regime switching model is included. When the initial regime is state 2 (cold), the intensity negatively affects the HDD call option, i.e., a higher intensity implies a faster rate of switching into a “hot” state. This leads to a smaller HDD, and so a lower option price based on equation (2.46). When we start with a “cold” regime and during the month of December as considered in our example, HDD calculated from the two-state HMM model ought to be generally higher than the ones generated from simulation with initial state 1 (hot).

On the other hand, when the initial regime is state 2 (cold), lower p 's produce higher option prices. This is because when p is low, the probability of remaining in the initial state is high. Option prices produced by a no-regime switching model are very close to the ones generated from the 2-state HMM model with $p = 0.1$ and initial state 2 in Table 2.11. This is expected since a very small p indicates a very slow switching from the “cold” state to a “hot” state within the context of the December contract period.

We utilise an increment level of 0.2 for the sensitivity analyses involving θ^* , α , and ξ . The values of optimal estimates listed in Table 2.10 are regarded as starting benchmarks when adding or lowering increment levels. Figure 2.12 depicts the option prices computed through varying values of θ_1^* . It is apparent that a change in θ_1^* greatly influences the option

price. With initial state 1, as θ_1^* increases, the option price dramatically decreases and with initial state 2, we have the opposite conclusion for the option price. This is because when the mean-level θ^* is low (cold), HDD is high giving a high positive HDD payoff. Figure 2.13 depicts the impact of varying θ_2^* on the option price. Even though the prices decline along with the increase of θ_2^* similar to the impact trend of θ_1^* , the prices under θ_2^* are a bit more spread out upwards compared to those under θ_1^* . This is because θ_2^* represents the mean level under a cold state. Once we increase θ_2^* , even though option prices decline, any subsequent switches to the hot state is not sufficient to drastically perturb the price level under state 2.

Figures 2.14 and 2.15 display the respective plots of option prices by varying the values of α_1 and α_2 . Although a slight increase in option price can be identified with a corresponding increase in α_1 , no drastic fluctuations appear under both “cold” and “hot” initial regimes. The same can be observed in the option prices when α_2 is varied in Figure 2.15. However, the option prices can be seen to be a wee bit higher under initial state 2 for both α_1 and α_2 .

The impact of ξ_1 on the option price starting with the respective “hot” and “cold” states is illustrated in Figure 2.16. The effect of varying ξ_2 on the option price is also presented in Figure 2.17. The option price decreases with increasing ξ_1 irrespective of whether the initial state is 1 or 2, whilst the opposite is true for the option price when ξ_2 increases. Moreover, ξ_1 seems to have a slightly more effect on the variability of option prices than ξ_2 does given the same increment. An increasing volatility under state 1 (hot) leads to a rising temperature level, which in turn yields a decreasing option price. The results also suggest that the variation of ξ_2 under initial state 2 contributes more to the variation of the option prices when compared to those obtained from the same variation levels of ξ_2 under the initial state 1 as shown in Figures 2.16 and 2.17. Higher volatility ξ_2 under the initial “cold” state results to higher option prices along with the fact that the sensitivity analysis is conducted for the cold month period of December in our illustrative case.

Remark 4: The standard errors of option prices calculated in (2.47) of subsection 2.6.4, as well as of those displayed in Figure 2.11 and Table 2.11 of subsection 2.6.5 fall in the range [\$6.67, \$9.61].

Strike HDD		$K_c = 520$	$K_c = 530$	$K_c = 540$	$K_c = 550$	$K_c = 560$	$K_c = 570$	$K_c = 580$
1-state model		1993.40	1798.78	1607.25	1420.29	1239.71	1066.88	903.97
Initial state 1 (Hot state)	$p = 0.1$	1642.04	1442.20	1242.37	1042.54	842.71	642.88	443.05
	$p = 0.3$	1652.74	1452.91	1253.08	1053.25	853.42	653.59	453.76
	$p = 0.5$	1808.83	1609.00	1409.17	1209.34	1009.51	809.68	609.85
	$p = 0.7$	1918.75	1718.92	1519.09	1319.26	1119.43	919.60	719.77
	$p = 0.9$	1919.37	1719.54	1519.71	1319.88	1120.05	920.22	720.39
Initial state 2 (Cold state)	$p = 0.1$	2045.34	1845.51	1645.68	1445.85	1246.02	1046.19	846.36
	$p = 0.3$	2037.09	1837.26	1637.43	1437.60	1237.77	1037.94	838.11
	$p = 0.5$	1925.11	1725.28	1525.45	1325.62	1125.79	925.95	726.12
	$p = 0.7$	1908.94	1709.16	1509.28	1309.45	1109.62	909.78	709.95
	$p = 0.9$	1863.26	1663.43	1463.60	1263.77	1063.94	864.11	664.28

Table 2.11: HDD call option price for varying intensities p

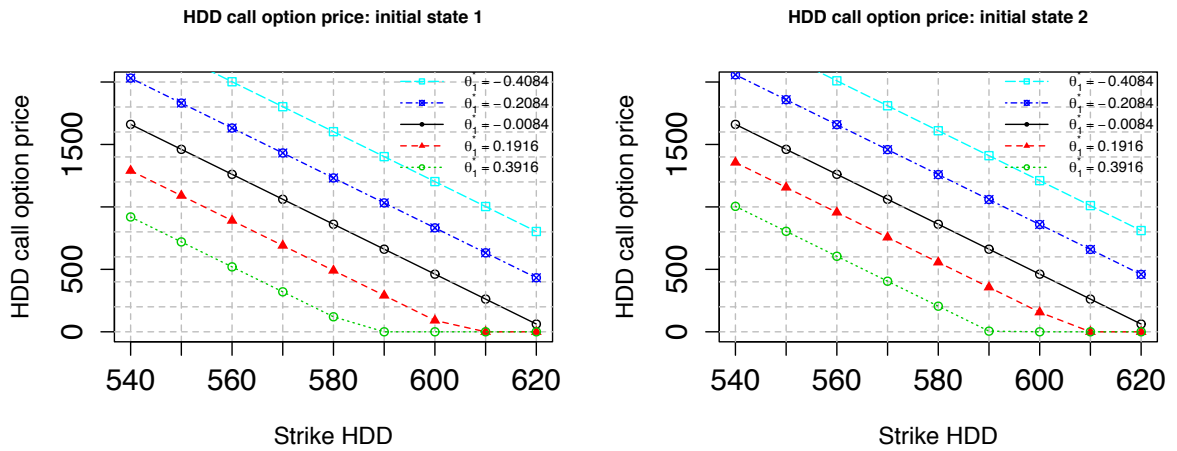


Figure 2.12: HDD call option price for varying θ_1^*

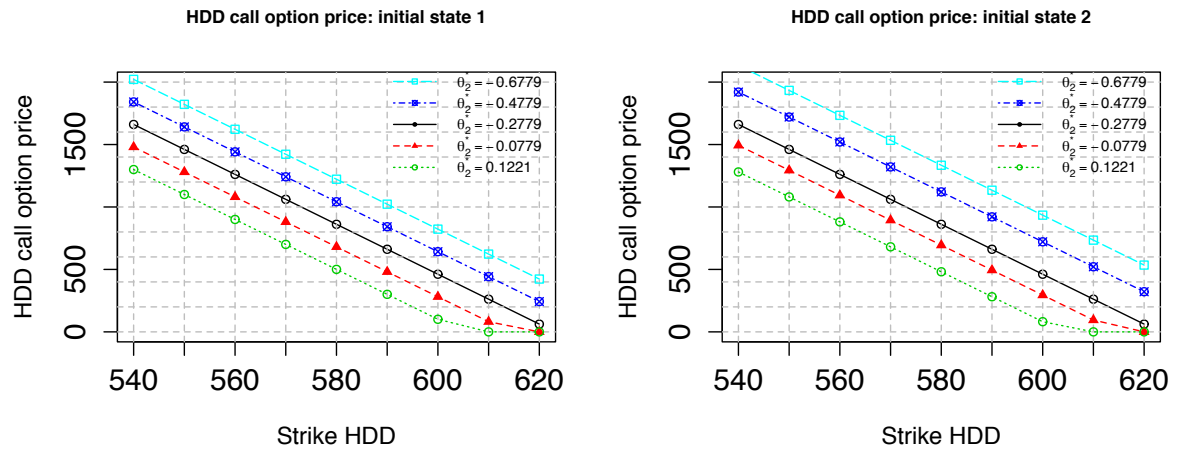


Figure 2.13: HDD call option price for varying θ_2^*

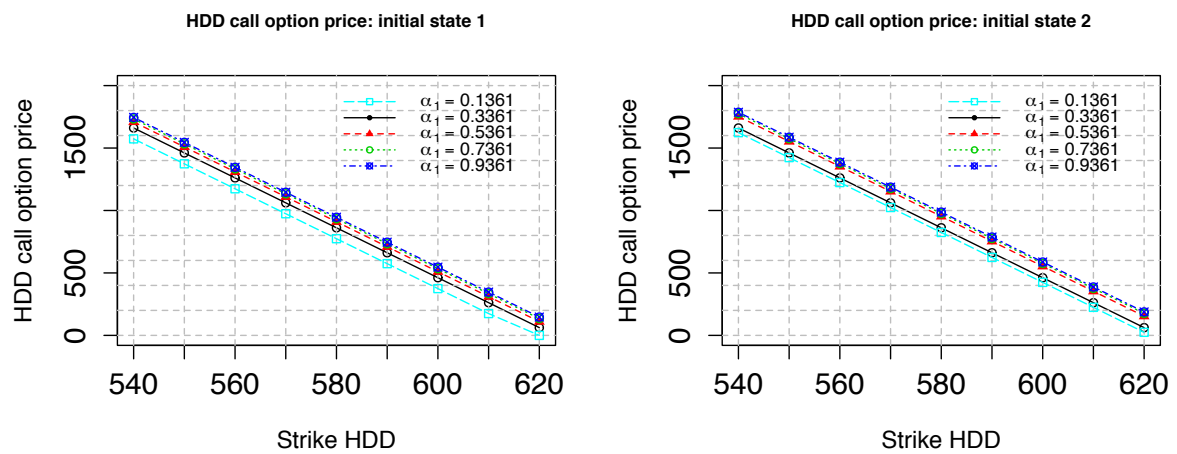


Figure 2.14: HDD call option price for varying α_1

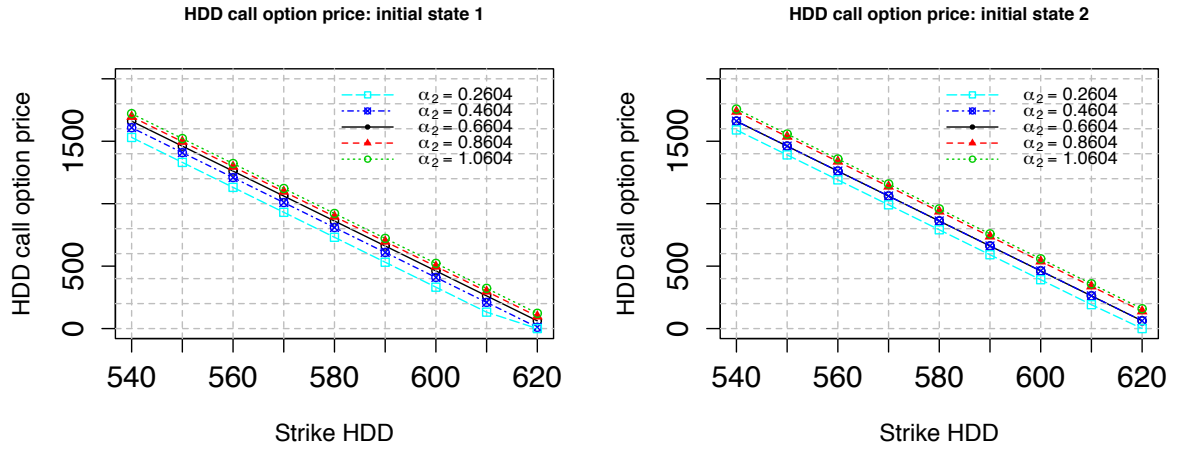


Figure 2.15: HDD call option price for varying α_2

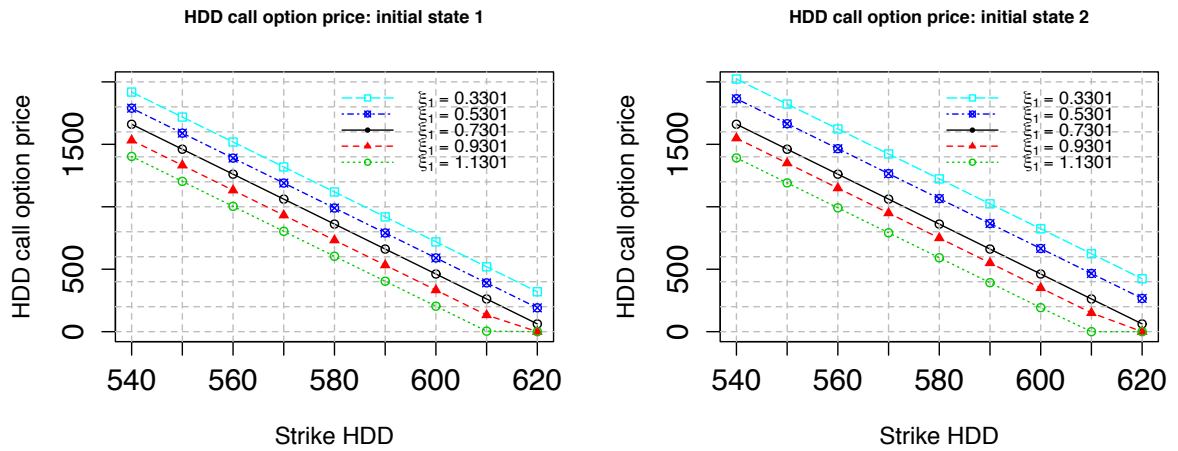
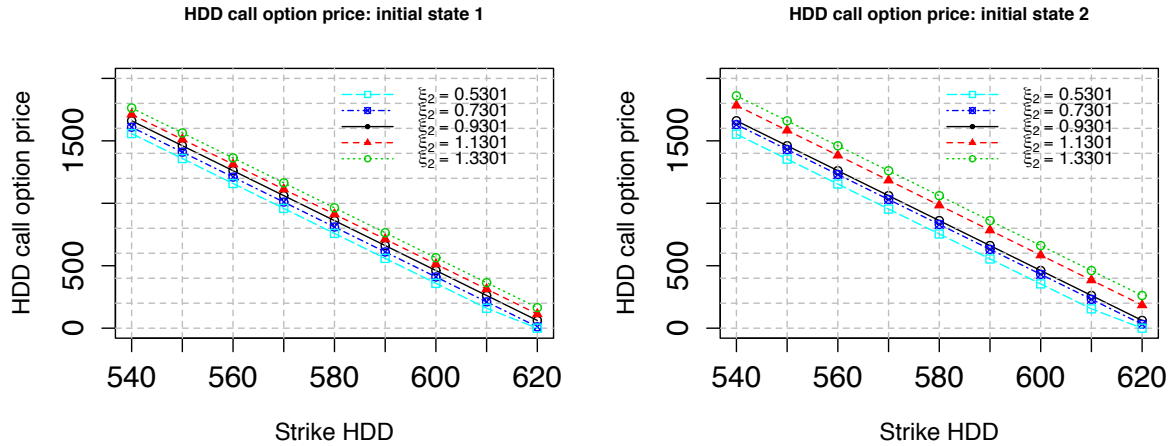


Figure 2.16: HDD call option price for varying ξ_1

Figure 2.17: HDD call option price for varying ξ_2

2.7 Conclusion

We implemented the parameter estimation of a temperature model, comprising of seasonal and stochastic components with mean reversion, designed to accurately capture the characteristics unique to the dynamics of the DATs. As the model parameters are modulated by HMM, the HMM-based filtering technique, via change of reference probability measure, along with the EM algorithm was employed to provide self-tuning parameters. We tested the model and estimation method on a 4-year Toronto's DATs. The one-step ahead forecasting errors and the AIC metrics were generated, and on their bases, the 2-state HMM is deemed to adequately model the 4-year Toronto DATs data set. Furthermore, we derived an expression for the price of HDD futures and option contracts by taking into account the price of risk process. Sensitivity analyses, under a 2-state HMM setting, were also performed by probing the behaviour of the option price as the intensity of the transition matrix and other model parameters were varied. We found that the mean-reverting level θ and volatility ξ have significant effects on pricing temperature-linked options.

Our work complements that in [14] and further elevates the use of regime-switching model for temperature modelling at a level that supports an interactive platform, given a data set, for the pricing and risk management of weather derivatives. A natural extension of this work is to investigate dynamic risk management under our setting and using filtered and optimal parameter estimates obtained from our estimation approach. The results could then be benchmarked with those from current estimation methodology for weather-linked

contracts. A stress-testing comparison using the framework put forward here and other modelling approaches could be pursued as well in the context of portfolio optimisation, in the spirit of [24] for instance; or risk measurement similar to the objectives of [25] but for weather derivatives instead; and modelling of weather futures prices constituting a multivariate time series resembling to the analysis in [12].

References

- [1] P. Alaton, B. Djehiche, D. Stillberger, On modelling and pricing weather derivatives, *Applied Mathematical Finance*, 9 (1) (2002) 1–20.
- [2] A. Alexandridis, A. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, Springer-Verlag, New York (2015).
- [3] F. Bellini, The weather derivatives market: modelling and pricing temperature, Ph.D. Thesis, (2005) University of Lugano, Switzerland.
- [4] F. Benth, W. Hardle and B. Cabrera, Pricing of asian temperature risk, in Anonymous Berlin, Heidelberg: Springer Berlin Heidelberg, (2011) 163–199.
- [5] F. Benth, J. Šaltytė-Benth, Weather derivatives and stochastic modelling of temperature, *International Journal of Stochastic Analysis*, (2011), 1–21. doi:10.1155/2011/576791
- [6] F. Benth, J. Šaltytė-Benth, Stochastic modelling of temperature variations with a view towards weather derivatives, *Applied Mathematical Finance*, 12 (1) (2005), 53–85.
- [7] D. R. Cox, H. D. Miller, *The Theory of Stochastic Processes*. London: Methuen, (1965).
- [8] S. Campbell, F. Diebold, Weather forecasting for weather derivatives, *Journal of the American Statistical Association*, 100 (469) (2005) 6–16.
- [9] M. Cao, J. Wei, Weather derivatives valuation and market price of weather risk, *Journal of Futures Markets*, 24 (11) (2004) 1065-1089.
- [10] W. Ching, T. Siu, L. Li, Pricing exotic options under a high-order Markovian regime switching model, *Journal of Applied Mathematics and Decision Sciences*, (2007) 1–15, doi:10.1155/2007/18014.

- [11] G. Considine, Introduction to weather derivatives, CME Group Web, <http://web.math.pmf.unizg.hr/amimica/pub/lapl.pdf>, Accessed March 2016
- [12] P. Date, R. Mamon, A. Tenyakov, Filtering and forecasting commodity futures prices under an HMM framework, *Energy Economics*, 40 (2013) 1001–1013.
- [13] M. Davis, Pricing weather derivatives by marginal value, *Quantitative Finance*, 1 (3) (2001) 305–308.
- [14] B. Dischel, At least: a model for weather risk, *Weather Risk Special Report, Energy and Power Risk Management*, March issue (1998) 3032.
- [15] G. Dorfleitner, M. Wimmer, The pricing of temperature futures at the Chicago Mercantile Exchange, *Journal of Banking and Finance*, 34(6) (2010) 1360–1370.
- [16] J. Dutton, Opportunities and priorities in a new era for weather and climate services, *Bulletin of the American Meteorological Society*, 83 (9) (2002) 1303-131.
- [17] R. Elias, M. Wahab, L. Fang, A comparison of regime-switching temperature modeling approaches for applications in weather derivatives, *European Journal of Operational Research*, 232 (3) (2014) 549–560.
- [18] R. Elliott, Exact adaptive filters for Markov chains observed in Gaussian noise, *Automatica*, 30 (1994) 1399–1408.
- [19] R. Elliott, L. Aggoun, J. Moore, J., *Hidden Markov Models: Estimation and Control*, Springer, New York (1995).
- [20] R. Elliott, T. Siu, A. Badescu, Bond valuation under a discrete-time regime switching term-structure model and its continuous-time extension, *Managerial Finance*, 37 (11) (2011) 1025-1047.
- [21] R. Elliott, V. Krishnamurthy, New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models, *IEEE Transactions on Automatic Control*, 44 (5) (1999) 938-051.
- [22] C. Erlwein, R. Mamon, An online estimation scheme for a HullWhite model with HMM-driven parameters, *Statistical Methods and Applications*, 18 (1) (2009) 87–107.

- [23] C. Erlwein, F. Benth, R. Mamon, HMM filtering and parameter estimation of an electricity spot price model, *Energy Economics*, 32 (5) (2010) 1034–1043.
- [24] C. Erlwein, R. Mamon, M. Davison, An examination of HMM-based investment strategies for asset allocation, *Applied Stochastic Models in Business and Industry*, 27 (3) (2011) 204–221.
- [25] H. Gao, R. Mamon, X. Liu, Risk measurement of a guaranteed annuity option under a stochastic modelling framework, *Mathematics and Computers in Simulation*, 132 (2017) 100–119.
- [26] H. Geman, M. Leonardi, Alternative approaches to weather derivatives pricing, *Managerial Finance*, 31(6)(2005) 46–72.
- [27] G. Grimmett, D. Stirzaker, *Probability and Random processes*, Oxford University Press, Oxford (2001).
- [28] S. Jewson, Introduction to weather derivative pricing, *Journal of Alternative Investments*, 7(2) (2004) 57-64.
- [29] S. Jewson, A. Brix, C. Ziehmann, *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*, Cambridge University Press, Cambridge (2005).
- [30] M. Moreno, Riding the temp, *Weather Derivatives*, FOW Special Supplement, December (2000).
- [31] J. Norris, *Markov chains*, Cambridge University Press, Cambridge & New York (1997).
- [32] T. Richards, M. Manfredo, D. Sanders, Pricing weather derivatives, *American Journal of Agricultural Economics*, 86(4) (2004) 1005–1017.
- [33] T. Siu, C. Erlwein, R. Mamon, The pricing of credit default swaps under a Markov-modulated Merton’s structural model, *North American Actuarial Journal*, 12 (1) (2008) 19–46.
- [34] A. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge and New York (1998).

- [35] R. Weron, Modeling and forecasting electricity loads and prices: A statistical approach. Hoboken, NJ; Chichester, England; John Wiley and Sons. (2006)
- [36] X. Xi, R. Mamon, Parameter estimation of an asset price model driven by a weak hidden Markov chain, *Economic Modelling*, 28 (1) (2011), 36–46.
- [37] H. Xiong, R. Mamon, A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics, *Journal of Computational Science*, 17 (2016) 47–61.
- [38] W. Xu, M. Odening, O. Musshoff, Indifference pricing of weather derivatives, *American Journal of Agricultural Economics*, 90(4) (2008) 979–993.

Chapter 3

A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics

3.1 Introduction

Over \$2.4 trillion in economic losses and nearly 2 million deaths globally have been reported as a result of weather-related hazards since 1971; see [34]. Catastrophes, such as floods, tornadoes and hurricanes, are well-publicised as a result of the heavy casualties and enormous property losses that they cause. But, whilst these weather-driven disasters are getting more attention, we note that even some slight aberrations of weather conditions, although non-catastrophic, could undeniably still be matters of consequence too. They may not produce immediately tremendous losses, yet their accompanying potential risks could materialise later in the future with more frequency as well as lead to more widespread impact on business and economy. Weather risk, for instance, is best depicted by crop yields that may drop substantially due to prolonged droughts or rainy season. Poor weather hugely affects planned construction schedules. Cold winters increase operational costs of energy consumers.

Dutton and John [13] concluded that weather and climate risks exist in nearly 33.3% of the private industry activities and 39.1% of the U.S. gross domestic product is weather-sensitive. In order to hedge the risks associated with such non-catastrophic weather conditions, the so-called weather derivatives were created. Their values depend on the underlying

variables governed by weather-related measurements, such as temperature, humidity, rainfall and snowfall. As documented in Considine [8], the first weather-derivative transaction took place in 1997, which was executed by Aquila Energy and embedded in a power contract. It is noted in Cao and Wei [9] that the Chicago Mercantile Exchange (CME) launched the first exchange-traded weather derivatives linked to temperature in 1999. Since then, the volume of weather derivatives grew rapidly both in the exchange and over-the-counter (OTC) markets. Notional values of weather derivative contracts processed in the exchange market are greater than those in the OTC market. Yet, interestingly, investors and other market participants are more actively involved in the OTC market according to the survey reported by Price Waterhouse Coopers (PwC); see [28].

In contrast to the traditional financial derivatives whose underlying assets are typically bonds and stocks, weather derivatives are dependent on non-negotiable underlying indices that have no price themselves. Such indices are introduced to quantify weather phenomena but casually employed to construct the basis of a weather derivative contract. As stated by Elias et al. [14], over 90% of weather derivatives are temperature-based. Weather futures and options traded in the CME are written on temperature heating degree day (HDD), cooling degree day (CDD) and cumulative average temperature (CAT) indices. The most distinctive feature of these temperature-based indices is that, unlike conventional financial underlying asset, they are not tradeable or storable. Therefore, the traditional no-arbitrage pricing approach, utilised in the Black-Scholes modelling framework, is not applicable to pricing weather derivatives [10]. Without delving into the challenges of derivative valuation concerning the choice of an appropriate pricing measure, which are outside the scope of this research, our aim is to develop a model that could accurately capture the salient features of the evolution of temperature data. Our proposed model then may be useful in accurately pricing weather derivatives.

In the literature, several studies were conducted to address the pricing of weather derivatives. It is asserted in Dorfleitner and Wimmer [12] that the historical burn analysis (HBA) is normally adopted by most investors given its straightforwardness and ease of replication. Nevertheless, Jewson and Brix [25] argued that HBA is bound to be biased and prone to large pricing errors, and so, time-series methods are proposed in [25] instead. Davis [10] dealt with weather derivative pricing by relying on the marginal substitution value principle of mathematical economics.

More recent studies considered the Ornstein-Uhlenbeck (OU) process in reproducing the dynamic behaviour of the temperature series. In comparison to simulating temperature indices in the HBA method, modelling the daily average temperature (DAT) directly generates more accurate outcomes since it contains more information under both regular and extreme conditions. The estimated DAT model is then utilised to develop corresponding temperature indices. These indices are used to determine the price of various weather derivatives together with their risk management. Dischel [11] first proposed the concept of adopting a continuous-time stochastic process to capture the dynamics of temperature. Alaton et al. [1] improved Dischel's model by incorporating seasonalities in the mean with a sinusoidal function. Benth and Saltyte-Benth [3] studied DAT variations with an OU process, where the noise follows a generalised hyperbolic Lévy process.

However, the application of a one-state stochastic process in almost all of the above-mentioned papers may not describe the behaviour of temperature with great accuracy, especially during drastic changes in regimes or emergence of entirely different states as in the occurrence of occasional climatic changes. This consideration inspires the extension of the typical OU mean-reverting process with the addition of the regime-switching feature. It seems that the research of Elias et al. [14] was by far the only paper that employs a regime-switching approach to model the stochastic behaviour of temperature. They applied lattices to construct corresponding models and concluded that the HDD and CDD generated by a two-state process were closer to the observed data than those obtained by a single stochastic process.

To take several steps of research progress farther away from the accomplishments in [14], the purpose of this work is to develop self-calibrating higher-order hidden Markov model (HOHMM) filtering algorithms that provide improved accuracy in capturing the behaviour and features of temperature, which will be beneficial for better valuation and risk management of weather futures and option products. An HOHMM of order k is an HMM that takes into account the HMM values up to lag k ; in the literature, HOHMM is also termed as a weak HMM as the memoryless assumption is weakened or relaxed to account for information revealed in the last k steps. The usual HMM has found many financial and economic applications, and pioneering developments, with the illustration of a two-state regime-switching model, are highlighted in Hamilton [22]. Elliott et al. [6] made significant landmarks in the estimation methodology of HMM via the change of measure technique for model identification after processing batches of data. Since then, researchers

put forward HMM-modulated models to address various problems in quantitative finance, insurance, economics, epidemiology, and other branches of the sciences and engineering. Erlwein and Mamon [18] is a precursor to our development of an HOHMM driving a mean-reverting process. We note that Mamon et al. [26] devised a 2-state HMM in modelling commodity prices and then analysed h -step ahead price predictions. Nonetheless, the memoryless property in the regular Markov assumption is a real drawback as models must also capture long-range dependency in the observed economic and financial data. To this end, Xi and Mamon [32], for instance, promoted the modelling of risky asset's log returns with the use of an HOHMM for a more flexible framework and possibility of better forecasting performance. This promising approach is primarily applied in speech recognition although its application in finance has started gradually to take momentum in recent years. For example, Ching et al. [7] priced exotic options assuming the variable's dynamics are under the HOHMM setting. With the aid of a higher-order Markov model, Siu et al. [28] investigated some aspects of risk management and examined the model's influence on risk measures through back testing.

The departure of our contributions from the current state of the art in Markov-switching temperature modelling is highlighted by the following accomplishments in the study: (i) Our proposed modelling approach simultaneously captures four important stylised facts of temperature evolution, which are the mean-reversion, seasonality, memory and stochasticity. (ii) Although adopted from the literature in Markovian regime-switching models for economic and financial variables, to the best of our knowledge, this research is the first to put forward a dynamic estimation procedure that readily gives parameter updates whenever a new set of data becomes available. (iii) The embedding of the HOHMM in the OU-modelling framework is also new in an attempt to capture the memory property in the temperature data series. (iv) Within the OU-modelling set up, we develop the filtering results relying on a transformation that converts a second-order HOHMM into a regular HMM [26]; hence, our approach enables the estimation of higher-order HOHMM to be implemented via an HOHMM-order reduction. (v) Finally, systematic and sufficient model implementation details are laid out including various diagnostics and validation based on techniques from statistical analysis and inference. Such details are beneficial for practitioners in building efficient computing platforms and interfaces with their business software as well as to other scientists in constructing a modelling framework that examines the movements and ability to control the dynamics of similar or related phenomena.

The remaining parts of this chapter are structured as follows. Section 3.2 presents the formulation of temperature modelling with a discrete-time higher-order hidden Markov chain governing the model parameters. In Section 3.3, the change of probability measure is applied to derive recursive filtering equations for quantities that are functions of HOHMM and necessary to carry out an online parameter estimation scheme in Section 3.4. The numerical implementation of our proposed model and estimation method is detailed in Section 3.5 involving a data set of daily temperatures collected at the Toronto Pearson Airport. The determination of the most appropriate model setting is given in Section 3.5 by comparing the respective forecasting performance and penalised-log likelihood of different competing set ups. Section 3.6 concludes.

3.2 Model description

3.2.1 Model for temperature derivatives

Majority of futures and option contracts on weather derivatives in the OTC and exchange-traded markets are written on HDD, CDD and CAT measurements. We, therefore, concentrate on the modelling of the above-mentioned indices. In particular, HDD and CDD were actualised to measure the demands of heating and cooling, respectively, of a certain location, and CAT is simply the sum of DATs over the contract period. Their corresponding values over the time interval $[\tau_1, \tau_2]$ with $0 \leq \tau_1 < \tau_2$ are given by

$$\begin{aligned} \text{HDD} &= \sum_{t=\tau_1}^{\tau_2} \max(T_{\text{base}} - T_t, 0), \\ \text{CDD} &= \sum_{t=\tau_1}^{\tau_2} \max(T_t - T_{\text{base}}, 0) \quad \text{and} \quad \text{CAT} = \sum_{t=\tau_1}^{\tau_2} T_t, \end{aligned}$$

where T_{base} is the base temperature, usually given as 65°F, or 18°C in the market, T_t is the daily average temperature on day t , computed as $\frac{T_{\text{max}} + T_{\text{min}}}{2}$. The contracts written on these indices have a weekly, monthly or seasonal duration. HDD contracts typically cover the cold months from October in one year to April of the following year, whilst the period of the CDD contracts is the warm season that runs from April to October of the same year. The overlapping months, October and April, are called the transition or shoulder months.

Some authors of previous papers (e.g., Dorfleitner and Wimmer [12]) attempted to price

temperature derivatives using an index-modelling approach by modelling the aforesaid indices directly. However, Benth et al.[2] argued that it is not appropriate to model the indices directly to derive option prices because the dynamics for the futures price could not even be obtained. Thus, for the valuation of temperature derivatives, we focus on modelling T_t that yields the needed indices as well as the dynamic representation of DATs.

Suppose (Ω, \mathcal{F}, P) is an underlying probability space that supports the modelling of the observed process T_t , the DAT on day t . Following Benth et al. [2],

$$T_t = X_t + S_t. \quad (3.1)$$

Equation (3.1) has two components: X_t that is assumed to follow an OU process with an HOHMM-governed parameters and S_t , which is a deterministic function devised to pin down the seasonal trends and DATs' mean reversion. As in Campbell and Diebold [5], the seasonal component is given by

$$S_t = at + b + \sum_{h=1}^3 \left[c_h \sin\left(d_h t \frac{2\pi}{365}\right) + e_h \cos\left(d_h t \frac{2\pi}{365}\right) \right], \quad (3.2)$$

where $d_1 = 1$, $d_2 = 2$, $d_3 = 4$ to cover the yearly, semi-annual and quarterly patterns, respectively.

3.2.2 Ornstein-Uhlenbeck (OU) process in the temperature model

We consider an OU model for the X_t component of the temperature model in (3.1). In the succeeding discussions, we denote all vectors by bold English/Greek letters in lowercase and all matrices are represented by bold capitalised English/Greek letters.

3.2.2.1 HMM-modulated OU process

Consider the stochastic differential equation (SDE) for the OU process X_t given by

$$dX_t = \alpha(\theta - X_t)dt + \xi dB_t, \quad (3.3)$$

where B_t is a standard Brownian motion on some probability space. By Itô's lemma and for $s \leq t$, the SDE in (3.3) has the solution

$$X_t = X_s e^{-\alpha(t-s)} + (1 - e^{-\alpha(t-s)})\theta + \xi e^{-\alpha t} \int_s^t e^{\alpha u} dB_u. \quad (3.4)$$

By utilising the Euler approximation, the discretised version of (3.4) is

$$\begin{aligned} X_{k+1} &= e^{-\alpha\Delta t_{k+1}} X_k + (1 - e^{-\alpha\Delta t_{k+1}}) \theta \\ &\quad + \xi \sqrt{\frac{1 - e^{-2\alpha\Delta t_{k+1}}}{2\alpha}} z_{k+1}, \end{aligned} \quad (3.5)$$

where $\Delta t_{k+1} = t_{k+1} - t_k$ and $\{z_{k+1}\}$ is a sequence of independent and identically distributed (IID) standard normal random variables.

We assume that we have a homogeneous Markov chain \mathbf{y}_k with finite states in discrete time. Its state space is associated with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$, where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top \in \mathbb{R}^N$ with 1 in the i th position, N stands for the total number of states, and \top denotes the transpose of a matrix. In a typical OU-HMM setting (e.g., Date et al. [9]), X_k with representation in (3.5) has parameters dependent on the Markov chain \mathbf{y}_k , i.e.,

$$\begin{aligned} X_{k+1} &= e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}} X_k + (1 - e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}}) \theta(\mathbf{y}_k) \\ &\quad + \xi(\mathbf{y}_k) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{y}_k)\Delta t_{k+1}}}{2\alpha(\mathbf{y}_k)}} z_{k+1}. \end{aligned} \quad (3.6)$$

In equation (3.6), the speed of mean-reversion α , the mean level of the process θ , and the volatility ξ are all governed by \mathbf{y}_k making the model regime switching. Given the representation of the Markov chain's state space, $\alpha_k := \alpha(\mathbf{y}_k) = \langle \boldsymbol{\alpha}, \mathbf{y}_k \rangle$, $\theta_k := \theta(\mathbf{y}_k) = \langle \boldsymbol{\theta}, \mathbf{y}_k \rangle$, and $\xi_k := \xi(\mathbf{y}_k) = \langle \boldsymbol{\xi}, \mathbf{y}_k \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product in \mathbb{R}^N .

3.2.2.2 HOHMM-modulated Ornstein-Uhlenbeck process

The second-order hidden Markov chain will be used as a prototype to develop the HOHMM setting; such prototype's simplicity will facilitate the discussion of concepts associated with a generalised Markov chain of lag k ($k = 1, 2, \dots$). Under this setting, the parameters α , θ and ξ are governed by a second-order Markov chain \mathbf{y}_k^w , which is defined on the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_k\}, P)$ with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$. Write $\mathcal{F}_k := \mathcal{F}_k^w \vee \mathcal{F}_k^B$ for the global filtration, where \mathcal{F}_k^w is the filtration generated by $\sigma\{\mathbf{y}_0^w, \dots, \mathbf{y}_k^w\}$ and \mathcal{F}_k^B is the filtration generated by B_t .

A discrete-time second-order hidden Markov chain \mathbf{y}_k^w at current time k depends on states that occurred at two prior lag times $k - 1$ and $k - 2$. The key principle in the filtering and estimation of HOHMM is the utility of a mapping that *transforms* an HOHMM into a

regular HMM. In our case, a mapping, say, ζ converts the second-order Markov chain into the usual Markov chain. The transformation ζ is defined as

$$\zeta(\mathbf{e}_b, \mathbf{e}_c) = \mathbf{e}_{bc}, \text{ for } 1 \leq b, c \leq N, \quad (3.7)$$

where \mathbf{e}_{bc} denotes a unit vector with 1 in its $((b-1)N+c)$ th position. Then, the new Markov chain satisfies the relation

$$\langle \zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w), \mathbf{e}_{bc} \rangle = \langle \mathbf{y}_{k+1}^w, \mathbf{e}_b \rangle \langle \mathbf{y}_k^w, \mathbf{e}_c \rangle \quad (3.8)$$

and has the semi-martingale representation

$$\zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) = \mathbf{\Pi} \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) + \delta_{k+1}^w, \quad (3.9)$$

where $\{\delta_{k+1}^w\}_{k \geq 1}$ is a martingale increment. So, $\mathbb{E}^P[\delta_{k+1}^w | \mathcal{F}_k^w] = \mathbf{0}$ and $\mathbf{\Pi}$ is the associated $N^2 \times N^2$ probability transition matrix.

3.3 Recursive filtering

3.3.1 Reference probability measure

Deriving recursive filters under the real-world probability P , where there is dependence amongst the observations, presents some difficulty. We construct an ideal probability measure Q that makes computations manageable because under such an ideal measure, the observed X_k 's are IID. The calculations are then reverted back to the measure P via an appropriate discretised version of the Girsanov density.

Suppose \mathbf{y}_k^w is a second-order hidden Markov chain that drives the model parameters of X_k . Write

$$H_{dcb} := P(\mathbf{y}_{k+1}^w = \mathbf{e}_d | \mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b),$$

where $k \geq 1$, and $d, c, b \in \{1, 2, \dots, N\}$. As in Siu et al. [28], \mathbf{H} is an associated $N \times N^2$ probability transition matrix given by

$$\mathbf{H} = \begin{pmatrix} h_{111} & h_{112} & \cdots & h_{11N} & \cdots & h_{1N1} & h_{1N2} & \cdots & h_{1NN} \\ h_{211} & h_{212} & \cdots & h_{21N} & \cdots & h_{2N1} & h_{2N2} & \cdots & h_{2NN} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ h_{N11} & \cdots & & h_{N1N} & \cdots & h_{NN1} & h_{NN2} & \cdots & h_{NNN} \end{pmatrix}.$$

Since we constructed ζ to transform a second-order Markov chain into a regular Markov chain, its corresponding $N^2 \times N^2$ matrix of transition probabilities $\mathbf{\Pi}$ can be defined in terms of \mathbf{H} as

$$\mathbf{\Pi} = \begin{pmatrix} h_{111} & \cdots & h_{11N} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & h_{121} & \cdots & h_{12N} & \cdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \vdots & \ddots & \vdots & \cdots & 0 & \cdots & \\ h_{N11} & \cdots & h_{N1N} & 0 & \cdots & 0 & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{N21} & \cdots & h_{N2N} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & 0 & \cdots & 0 & \cdots & h_{1N1} & \cdots & h_{1NN} \\ 0 & \cdots & 0 & \vdots & \ddots & \vdots & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & h_{NN1} & \cdots & h_{NNN} \end{pmatrix}$$

where $\pi_{ji} = h_{dcb} = P(\mathbf{y}_{k+1}^w = \mathbf{e}_d | \mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b)$ for $j = (d-1)N + c$, $i = (c-1)N + b$, and $\pi_{ji} = 0$ otherwise.

Under the HOHMM framework, the model for X_k in (3.6) is modified to

$$X_{k+1} = \kappa(\mathbf{y}_k^w) X_k + \vartheta(\mathbf{y}_k^w) + \varrho(\mathbf{y}_k^w) z_{k+1}, \quad (3.10)$$

where

$$\begin{aligned} \kappa(\mathbf{y}_k^w) &= e^{-\alpha(\mathbf{y}_k^w)\Delta t_{k+1}}, \quad \vartheta(\mathbf{y}_k^w) = (1 - e^{-\alpha(\mathbf{y}_k^w)\Delta t_{k+1}})\theta(\mathbf{y}_k^w), \\ \varrho(\mathbf{y}_k^w) &= \xi(\mathbf{y}_k^w) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{y}_k^w)\Delta t_{k+1}}}{2\alpha(\mathbf{y}_k^w)}}. \end{aligned}$$

As we do not observe the state of \mathbf{y}_k^w , it needs to be estimated and to this end, we work as noted above under an ideal reference probability measure Q whose existence is justified by the Kolmogorov's extension theorem. Again, under Q , $\{z_k\}_{k \geq 1}$ is a sequence of IID $N(0, 1)$ random variables and also independent of \mathbf{y}_k^w .

Let $\mathcal{X}_k = \sigma(X_0, X_1, \dots)$. We introduce the Girsanov's density to back out P from Q given

an \mathcal{X}_k -adapted process Ψ_k^w defined by

$$\Psi_k^w = \frac{dP}{dQ} \Big|_{\mathcal{X}_k^w} = \prod_{l=1}^k \varphi_l^w, \quad k \geq 1, \quad \Psi_0 = 1 \quad (3.11)$$

$$\text{and } \varphi_l^w = \frac{\phi \left\{ \varrho(\mathbf{y}_{l-1}^w)^{-1} [X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1}] \right\}}{\varrho(\mathbf{y}_{l-1}^w) \phi(X_{l-1})}.$$

3.3.2 Calculation of recursive filters

Our goal is to obtain optimal parameter estimates under the real-world measure P through self-updating filters that are functions of \mathbf{y}_k^w . The conditional expectation of any \mathcal{F}^w -adapted process given \mathcal{X}_k is then needed. As mentioned above, calculations performed under the Q measure could be done with ease. Moreover, the Bayes' theorem for conditional expectations provides a connection between values calculated under P and those calculated under Q .

We establish adaptive filtering processes, under P , for estimators of quantities that are functions of $\zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$ as per equations (3.8) and (3.9), but by doing the analysis and computations under measure Q . Write

$$\begin{aligned} s_k(cb) &:= P(\mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b \mid \mathcal{X}_k) \\ &= \mathbb{E}[\langle \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle \mid \mathcal{X}_k] \end{aligned} \quad (3.12)$$

and $\mathbf{s}_k = (s_k(11), \dots, s_k(cb), \dots, s_k(NN))^\top \in \mathbb{R}^{N \times N}$. From the Bayes' theorem for conditional expectation,

$$\mathbf{s}_k = \mathbb{E}[\zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{X}_k] = \frac{\mathbb{E}^Q[\Psi_k^w \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{X}_k]}{\mathbb{E}^Q[\Psi_k^w \mid \mathcal{X}_k]}. \quad (3.13)$$

Write $\mathbf{v}_k := \mathbb{E}^Q[\Psi_k^w \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{X}_k]$, and note that

$$\sum_{c,b=1}^N \langle \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle = \langle \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{1} \rangle = 1, \quad (3.14)$$

where $\mathbf{1}$ is an \mathbb{R}^{N^2} -vector with 1 in all of its entries. Then, we have

$$\begin{aligned} \sum_{c,b=1}^N \langle \mathbf{v}_k, \mathbf{e}_{cb} \rangle &= \sum_{c,b=1}^N \langle \mathbb{E}^Q[\Psi_k^w \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{X}_k], \mathbf{e}_{cb} \rangle \\ &= \mathbb{E}^Q \left[\Psi_k^w \sum_{c,b=1}^N \langle \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle \mid \mathcal{X}_k \right] \\ &= \mathbb{E}^Q[\Psi_k^w \mid \mathcal{X}_k]. \end{aligned} \quad (3.15)$$

Along with equations (3.14) and (3.15), the conditional expectation of $\zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$ under the real-world probability measure P could be written explicitly as

$$\mathbf{s}_k = \frac{\mathbf{v}_k}{\sum_{c,b=1}^N \langle \mathbf{v}_k, \mathbf{e}_{cb} \rangle} = \frac{\mathbf{v}_k}{\langle \mathbf{v}_k, \mathbf{1} \rangle}. \quad (3.16)$$

As noted in Xi and Mamon [32], a diagonal matrix is needed to put recursive processes in place under an HOHMM. Let \mathbf{D}_k be an $N^2 \times N^2$ diagonal matrix to estimate several functions of $\zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$, where

$$\mathbf{D}_k = \begin{pmatrix} d_k^1 & 0 & & \cdots & 0 \\ 0 & \ddots & 0 & & \vdots \\ \vdots & 0 & d_k^N & & \ddots \\ & & & \ddots & \\ & & & \ddots & d_k^1 & 0 & \vdots \\ \vdots & & & & 0 & \ddots & 0 \\ 0 & \cdots & & & \cdots & 0 & d_k^N \end{pmatrix}$$

with diagonal elements

$$d_k^i = \frac{\phi\left((X_k - \vartheta_i - \kappa_i X_{k-1}) \varrho_i^{-1}\right)}{\varrho \phi(X_k)}. \quad (3.17)$$

We define below various processes needed for the parameter estimation in the next section.

1. The number of jumps from state $(\mathbf{e}_r, \mathbf{e}_s)$ to state \mathbf{e}_t up to time k , $2 \leq l \leq k$ and $r, s, t = 1, \dots, N$, is denoted by

$$\mathbf{A}_k^{tsr} := \sum_{l=2}^k \langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle. \quad (3.18)$$

2. The number of occupations up to time k , which is the length of time that Markov chain \mathbf{y}^w spent in state \mathbf{e}_t , $2 \leq l \leq k$ and $t = 1, \dots, N$, is given by

$$\mathbf{B}_k^t := \sum_{l=2}^k \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle = \mathbf{B}_{k-1}^t + \langle \mathbf{y}_{k-1}^w, \mathbf{e}_t \rangle. \quad (3.19)$$

3. The number of occupations up to time k with the length of time that Markov chain \mathbf{y}^w spent in state $(\mathbf{e}_t, \mathbf{e}_s)$, $2 \leq l \leq k$ and $s, t = 1, \dots, N$, is calculated as

$$\mathbf{B}_k^{ts} := \sum_{l=2}^k \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \langle \mathbf{y}_{l-2}^w, \mathbf{e}_s \rangle. \quad (3.20)$$

4. The auxiliary process related to the Markov chain \mathbf{y}^w for the function f up to time k in state \mathbf{e}_t , $2 \leq l \leq k$, $t = 1, \dots, N$, is computed as

$$\begin{aligned} \mathbf{C}_k^t(f) &:= \sum_{l=2}^k f(X_l) \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle = \mathbf{C}_{k-1}^t(f) \\ &\quad + f(X_k) \langle \mathbf{y}_{k-1}^w, \mathbf{e}_t \rangle, \end{aligned} \quad (3.21)$$

where f has the form $f(X) = X$, $f(X) = X^2$, or $f(X) = X_{l-1}X_l$.

5. The conditional expectation of $\zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$ in Equation (3.16) can be written recursively in terms of the diagonal matrix \mathbf{D}_{k+1} as

$$\mathbf{s}_{k+1} = \mathbf{D}_{k+1} \mathbf{s}_k. \quad (3.22)$$

Let \mathbf{G}_k be any \mathcal{X}_k -measurable generic process defined in equations (3.18)-(3.21). From Bayes' theorem for conditional expectation, the 'best estimate' for \mathbf{G}_k is

$$\begin{aligned} \widehat{\mathbf{G}}_k &= \mathbb{E}^P [\mathbf{G}_k | \mathcal{X}_k] = \frac{\mathbb{E}^Q [\Psi_k^w \mathbf{G}_k | \mathcal{X}_k]}{\mathbb{E}^Q (\Psi_k^w | \mathcal{X}_k)} \\ &= \frac{\mathbb{E}^Q [\Psi_k^w \mathbf{G}_k | \mathcal{X}_k]}{\sum_{c,b=1}^N \langle \mathbf{v}_k, \mathbf{e}_{cb} \rangle} = \frac{\mathbb{E}^Q [\Psi_k^w \mathbf{G}_k | \mathcal{X}_k]}{\langle \mathbf{v}_k, \mathbf{1} \rangle}. \end{aligned} \quad (3.23)$$

In order to get dynamic updates for \mathbf{G}_k every time an observed value X_k arrives or a batch of X_k 's becomes available, we provide recursions via the vector quantity $\mathbf{G}_k \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)$ that will ultimately lead to an efficient updating of the scalar quantity $\widehat{\mathbf{G}}_k$.

Write

$$\gamma_k^w(\mathbf{G}_k) := \mathbb{E}^Q [\Psi_k^w \mathbf{G}_k | \mathcal{X}_k]. \quad (3.24)$$

By equation (3.14), the scalar $\gamma_k^w(\mathbf{G}_k)$ can be calculated using the vector quantity $\mathbf{G}_k \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)$ by noticing

$$\begin{aligned} \gamma_k^w(\mathbf{G}_k) &= \gamma_k^w(\mathbf{G}_k \langle \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{1} \rangle) \\ &= \gamma_k^w(\langle \mathbf{G}_k \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{1} \rangle) \\ &= \langle \gamma_k^w(\zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)), \mathbf{1} \rangle \\ &= \mathbb{E}^Q (\Psi_k^w \mathbf{G}_k \langle \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{1} \rangle | \mathcal{X}_k). \end{aligned} \quad (3.25)$$

Therefore, equation (3.23) can be further calculated as

$$\widehat{\mathbf{G}}_k = \frac{\mathbb{E}^Q [\Psi_k^w \mathbf{G}_k | \mathcal{X}_k]}{\langle \mathbf{v}_k, \mathbf{1} \rangle} = \frac{\gamma_k^w(\mathbf{G}_k)}{\langle \mathbf{v}_k, \mathbf{1} \rangle}, \quad (3.26)$$

which shows $\widehat{\mathbf{G}}_k$ can be obtained purely from calculations under Q and can be updated dynamically if we have recursions for $\gamma_k^w(\mathbf{G}_k)$. So, to implement (3.26) for the quantities \mathbf{A}_k^{tsr} , \mathbf{B}_k^t , \mathbf{B}_k^{ts} , and $\mathbf{C}_k^t(f)$, we utilise the semi-martingale representation (3.9) to derive their unnormalised recursive filtered estimates.

Adopting Xi and Mamon [32], let

$$\mathbf{K}_t, \quad 1 \leq t \leq N$$

be an $N^2 \times N^2$ matrix with \mathbf{e}_{it} on its $((i-1)N+t)$ th column and 0 elsewhere. Then we have the following result.

Proposition 1: The vector recursions involving the respective quantities in equations (3.18)–(3.21) are

$$\gamma_{k+1}^w(\mathbf{A}_{k+1}^{tsr} \zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{I} \mathbf{D}_{k+1} \gamma_k^w(\mathbf{A}_k^{tsr} \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + \langle \mathbf{s}_k, \mathbf{e}_{sr} \rangle d_{k+1}^t \langle \mathbf{I} \mathbf{e}_{sr}, \mathbf{e}_{ts} \rangle \mathbf{e}_{ts}, \quad (3.27)$$

$$\gamma_{k+1}^w(\mathbf{B}_{k+1}^t \zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{I} \mathbf{D}_{k+1} \gamma_k^w(\mathbf{B}_k^t \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + \mathbf{K}_t d_{k+1}^t \mathbf{I} \mathbf{s}_k, \quad (3.28)$$

$$\gamma_{k+1}^w(\mathbf{B}_{k+1}^{ts} \zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{I} \mathbf{D}_{k+1} \gamma_k^w(\mathbf{B}_k^{ts} \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + \langle \mathbf{s}_k, \mathbf{e}_{ts} \rangle d_{k+1}^t \mathbf{I} \mathbf{e}_{ts}, \quad (3.29)$$

$$\gamma_{k+1}^w(\mathbf{C}_{k+1}^t(f) \zeta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{I} \mathbf{D}_{k+1} \gamma_k^w(\mathbf{C}_k^t(f) \zeta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + f(X_{k+1}) \mathbf{K}_t d_{k+1}^t \mathbf{I} \mathbf{s}_k. \quad (3.30)$$

Proof The derivations are similar to those given in Mamon et al. [26].

3.4 Optimal parameter estimation

In this section, we present the optimal estimates for the parameters of our proposed temperature model using the maximum-likelihood approach. Since maximising the likelihood or

log-likelihood functions is not straightforward for such functions with complicated structures like ours, we employ the EM algorithm, which is an effective iterative method for parameter estimation especially for an exponential family of models [29].

We shall show that our EM results give optimal estimates in terms of the filters in (3.22) and (3.27)–(3.30). Consider a family of probability measures $\{P^{v^w}, v^w \in \Upsilon^w\}$ on (Ω, \mathcal{F}^w) . The algorithm is implemented by setting the initial P^{v_0} and then changing from P^{v_0} to P^{v^w} reflecting the variation of the data to the updated parameters.

We refer to $\Upsilon^w = \{\kappa_t, \vartheta_t, \varrho_t, h_{tsr}, 1 \leq t, s, r \leq N\}$ as the set of parameters under our HOHMM setting. The corresponding maximum likelihood (ML) estimator is

$$\widehat{v}^w \in \operatorname{argmax}_{v^w \in \Upsilon^w} L(v^w) \quad \text{and} \quad L(v^w) = \mathbb{E}^{v_0} \left[\frac{dP^{v^w}}{dP^{v_0}} \middle| \mathcal{X}_k \right].$$

The sequence of log likelihoods yields monotonically non-decreasing ML estimates, which locally converges [35].

A change of measure from P^{v_0} to P^{v^w} is performed to estimate the entries of matrix \mathbf{H} thereby giving $\mathbf{\Pi}$ updated entries through the filtering algorithms of the relevant quantities. Recall that \mathbf{y}^w is an HOHMC with transition matrix $\mathbf{H} = (h_{tsr})$ under P^{v^w} . Hence, a new probability measure $P^{\widehat{v}^w}$ must be constructed, under which \mathbf{y}^w is still an HOHMC but with transition matrix $\widehat{\mathbf{H}} = (\widehat{h}_{tsr})$, where $\widehat{h}_{tsr} = P^{\widehat{v}^w}(\mathbf{y}_{k+1}^w = \mathbf{e}_t | \mathbf{y}_k^w = \mathbf{e}_s, \mathbf{y}_{k-1}^w = \mathbf{e}_r)$ with $\widehat{h}_{tsr} \geq 0$.

From Elliott et al. [6], the appropriate density in conjunction with the EM algorithm for our successive estimation of the transition probabilities is

$$\begin{aligned} \frac{dP^{v^w}}{dP^{v_0}} \bigg|_{\mathcal{X}_k} &= \Lambda_k, \quad \Lambda_0 = 1 \quad \text{and} \\ \Lambda_k &= \prod_{l=2}^k \prod_{t,s,r=1}^N \left(\frac{\widehat{h}_{tsr}}{h_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle}. \end{aligned} \quad (3.31)$$

When $h_{tsr} = 0$, $\widehat{h}_{tsr} = 0$; in this case, we set $\frac{\widehat{h}_{tsr}}{h_{tsr}} = 1$. The resulting expression for \widehat{h}_{tsr} and those for the rest of the parameters, computed as well via the EM algorithm, are given in the succeeding summary.

Proposition 2: The EM estimates, at state t based on a data series up to time k ($k \geq 1$), for

the parameters of the model in (3.10) are as follows:

$$\begin{aligned}\widehat{\kappa}_t &= \frac{\widehat{\mathbf{C}}_k^t(X_{k-1}, X_k) - \vartheta^t \widehat{\mathbf{C}}_k^t(X_{k-1})}{\widehat{\mathbf{C}}_k^t(X_{k-1}^2)} \\ &= \frac{\gamma_k^w(\widehat{\mathbf{C}}_k^t(X_{k-1}, X_k)) - \vartheta^t \gamma_k^w(\widehat{\mathbf{C}}_k^t(X_{k-1}))}{\gamma_k^w(\widehat{\mathbf{C}}_k^t(X_{k-1}^2))},\end{aligned}\quad (3.32)$$

$$\begin{aligned}\widehat{\vartheta}_t &= \frac{\widehat{\mathbf{C}}_k^t(X_k) - \kappa^t \widehat{\mathbf{C}}_k^t(X_{k-1})}{\widehat{\mathbf{B}}_k^t} \\ &= \frac{\gamma_{k+1}^w(\widehat{\mathbf{C}}_k^t(X_k)) - \kappa^t \gamma_k^w(\widehat{\mathbf{C}}_k^t(X_{k-1}))}{\gamma_k^w(\widehat{\mathbf{B}}_k^t)},\end{aligned}\quad (3.33)$$

$$\begin{aligned}\widehat{\varrho}_t^2 &= \frac{\widehat{\mathbf{C}}_k^t(X_k^2) + \kappa_t^2 \widehat{\mathbf{C}}_k^t(X_{k-1}^2) + \vartheta_t^2 \widehat{\mathbf{B}}_k^t + 2\vartheta_t \kappa_t \widehat{\mathbf{C}}_k^t(X_{k-1})}{\widehat{\mathbf{B}}_k^t} \\ &\quad - \frac{2\kappa_t \widehat{\mathbf{C}}_k^t(X_{k-1}, X_k) - 2\vartheta_t \widehat{\mathbf{C}}_k^t(X_k)}{\widehat{\mathbf{B}}_k^t}\end{aligned}\quad (3.34)$$

$$\widehat{h}_{tsr} = \frac{\widehat{\mathbf{A}}_k^{tsr}}{\widehat{\mathbf{B}}_k^{sr}}, \quad \forall \text{ pairs } (t, s), t \neq s. \quad (3.35)$$

Proof The proofs of (3.32)-(3.35) can be found in the Appendix C.

Thus, when a temperature series is available at time k , we could get automatically new parameters κ_t , ϑ_t , ϱ_t , and h_{tsr} by running the filtering recursions of the Markov chain given in Proposition 1.

3.5 Numerical implementation

We implement our recursive HOHMM filtering algorithms on DAT series collected by the NCDC. The data were recorded from 01 Jan 2011 to 31 Dec 2014 comprising of 1461 data points. Since the DAT data have high presence of seasonality, we first remove the deterministic cyclical pattern S_t from T_t in equation (3.1), after which our filtering equations are implemented to obtain optimal parameter estimates of the OU component X_t . The data set is then divided into batches with an equal size to perform the implementation of the self-tuning process with recursive filtering equations under both HMM and HOHMM settings.

Mean	Sd Dev	SE	Min	Max	Skew	Kurtosis
8.88	10.65	0.28	-19.25	31.5	-0.23	-0.92

Table 3.1: Descriptive statistics for daily average temperature(DAT)

Parameter	Estimate	95% confidence interval
a	10.2111	(-0.0024, -0.0013)
b	-0.0018	(9.7581, 10.6641)
c_1	-5.2949	(-5.6155, -4.9742)
c_2	-0.6395	(-0.9555, -0.3234)
c_3	-0.4131	(-0.7279, -0.0982)
e_1	-12.7241	(-13.0383, -12.4098)

Table 3.2: Parameter estimates for the seasonality component S_t

3.5.1 Analysis of the deterministic component

We fit S_t described in (3.1) into the entire data set. Since S_t is a linear combination of sinusoidal functions and a linear mapping, the built-in function of the software R's linear regression is competent to execute this fitting process. In fact, a stepwise regression procedure was carried out to identify the best fitting model for S_t using the R function 'step' to perform the selection of explanatory variables.

Table 3.1 shows some descriptive statistics for the DAT, which guided us with certain references in choosing the initial values in the implementation process. The values of estimated parameters from best fitting our data to the specified S_t are exhibited in Table 3.2. The adjusted coefficient of determination R^2 shows that 83.5% of the variation in the response T_t can be well explained by the model with all the regressor variables in equation (3.2) except for e_2 and e_3 . Consequently, these predictors are eliminated during the process of variable selection on the basis of the scoring criteria built in the function 'step', including the adjusted R^2 and Akaike information criterion (AIC).

The fitted seasonality component S_t is plotted in Figure 3.1 together with the actual DAT, which displays high-seasonality property. The plot depicts four crests and four troughs occurring in summer and winter, respectively, of the 4-year period. The main characteristics of the two graphs neatly jibe with the temperature movements in the four seasons of each year. The remaining component X_t is treated as our observation process which we shall use

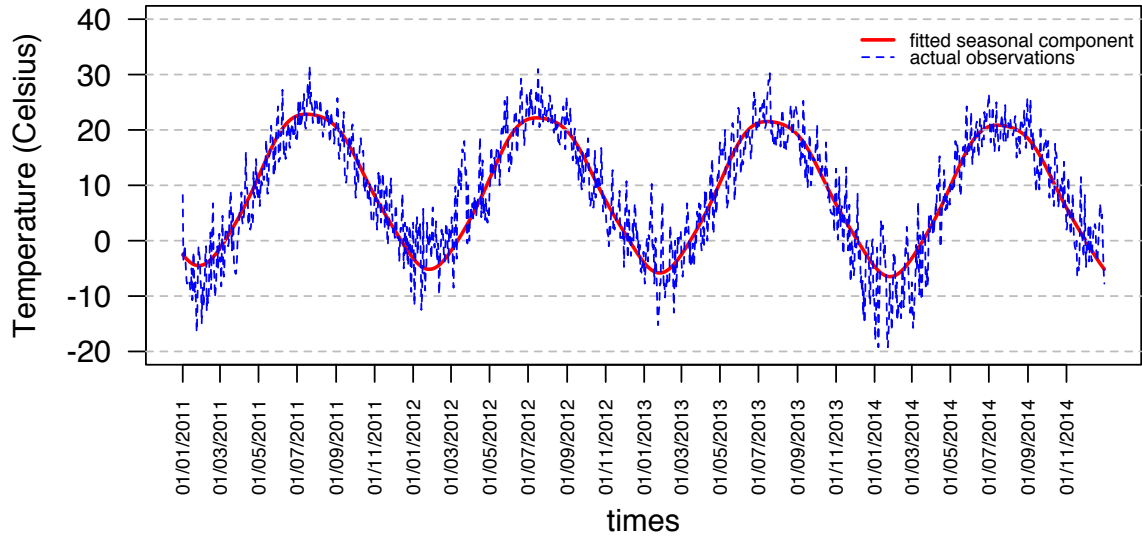


Figure 3.1: Fitted seasonal component versus actual DAT series

to test our proposed filtering and parameter estimation methods.

3.5.2 Analysis of the stochastic component

The cyclical component of the model captures very well the underlying seasonal trend of the actual data amidst the noisy values of X_t . The deseasonalised stochastic component $X_t = T_t - S_t$ is displayed in Figure 3.2.

We process the data set in 73 batches, and so there are 20 data points in each group with $\Delta t = 1$ day. This means that the parameters are updated roughly every 3 weeks. Other filtering-window sizes can certainly be adopted in the data processing. Our experiment shows that the choice of the window size has little effect on the numerical outcomes. A batch of 20 data points is fairly adequate to cover the arrival of new temperature recordings that might influence changes in temperature such as wind power, ocean currents, and other meteorological events.

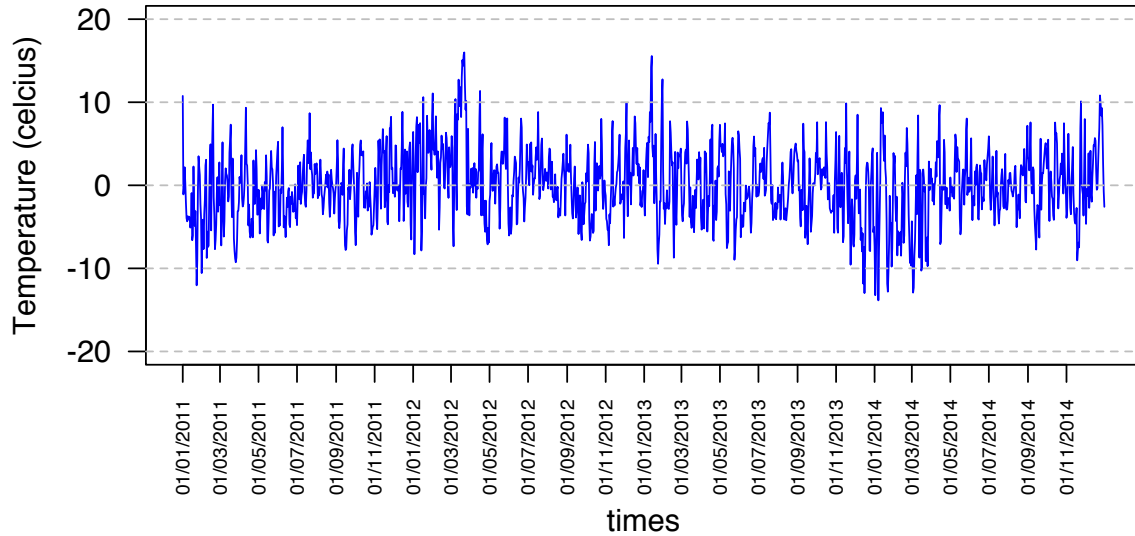


Figure 3.2: Deseasonalised stochastic component X_t

3.5.3 Validity of model evaluation

We are aware that a valid statistical analysis of a proposed model entails testing using a data set that was not used in the model estimation. Otherwise, if the model is not tested on a different data set from the same population, it is difficult to determine if any patterns found are simply due to chance. Thus, showing the formulated model's predictive power necessitates waiting for new data to come in. This ensures that there is no hand-tailoring of the predictive ability of the model to the data on hand as the upcoming weather data are not yet available.

Our online filtering method is consistent with the above statistical principle because we use the past and current data through X_t to obtain parameter updates that encapsulate the information up to present time t . Such parameter updates are used to process a new batch of accumulated information in order to obtain a succeeding new set of parameter updates. Also, utilising the same parameter updates, we obtain our predicted values for the calculation of error analysis metrics. Hence, the data set, used in prediction and calculation of log-likelihood measures for testing goodness of fit and assessing model complexity in section 3.5.5, is different from and does not overlap with that used in model estimation.

Parameter	[Min, Max]
a	[10.0900, 10.5300]
b	[-0.0027, -0.0017]
c_1	[-5.3220, -4.9600]
c_2	[-0.6500, -0.4489]
c_3	[-0.5062, -0.3337]
e_1	[-12.8700, -12.6900]

Table 3.3: Minimum and maximum parameter estimates for S_t 's coefficients covering twelve 4-year moving windows as described in Section 3.5.3

Our method may appear not in agreement though with the above principle if S_t is taken into account. This is because the S_t function must be calculated first using the entire data set before the X_t data series can be produced. We argue that there is no inconsistency here whenever the coefficients of the S_t component remain almost constant as time evolves. Such is exactly the case for the S_t 's coefficients estimated from our daily temperature readings covering the 4-year period under study (01 Jan 2011 – 31 Dec 2014) and coefficients estimated from the twelve prior 4-year period moving windows going backwards (i.e., Data Window I: 01 Jan 2007– 31 Dec 2010; Data Window II: 01 Dec 2006 – 30 Nov 2010; Data Window III: 01 Nov 2006 – 31 Oct 2010; etc). Table 3.3 displays the [minimum, maximum] values of the 12 estimates for each coefficient. Clearly, the variation is so small. Thus, we take the average of each set of 12 estimates as a proxy for a given coefficient value so that the use of values in Table 3.2 (as close approximations to the corresponding proxies) for S_t to determine X_t is justified in adherence to the above statistical principle.

3.5.4 Implementation aspects of the HOHMM-OU filtering

To detect the presence of memory, we estimate the order of differencing d in the autoregressive fractionally integrated moving average framework [16]. For $0 < d < 0.5$, there is presence of finite long memory, and short memory for $d = 0$. Estimating d could be done via the Geweke-Porter-Hudak estimator, approximate MLE, and the smoothed periodogram method with the R functions 'fdGPH', 'fdML', and 'fdSperio' respectively. We relied on 'fdSperio' proposed by Reisen [30] as it has no constraint on d and applies to non-stationary series, but the same cannot be said for the 'fdGPH' and 'fdML' functions. Our data set yields a rough estimate of $d = 1/4$, implying it possesses some form of memory. We acknowledge the limitation of our modelling, which is the imposition of a second-order

lag and our current HOHMM approach still does not include a mechanism to provide an optimal lag estimate based on the data series. However, for the purpose of illustrating our filtering implementation, we verified that our temperature data series exhibits memory, validating the appropriateness of the HOHMM.

Remark: Indeed, if for a given data set, a lag order of greater than 2 is formally necessary (in a statistical-inference sense), then the transformation in (3.7) can be repeatedly utilised until a 2nd-order HOHMM set up is obtained, and therefore, our current filtering and estimation results could be applied in a straightforward manner. Despite the capacity for sophistication of being able to include as many lags as needed, there is also the practical consideration, especially from the perspective of implementation in the industry, to balance between the benefit of having a flexible but complex model and the associated formidable computational cost. Of course, given the power of supercomputers, we envision that our optimal processing results in this chapter for large lags can be efficiently implemented someday and the dreaded curse of dimensionality can be appreciably alleviated. With the rapid continuing development in computing architectures, we hope to see that the complicated part of re-coding and extension of filtering algorithms as the lag becomes bigger could be facilitated with much ease.

3.5.4.1 Initial values for the parameter estimation

The filtering algorithms are implemented by first finding the initial estimates of parameters under the assumption that the X_t is a single state process. A value of $\frac{1}{N}$ is given to each non-zero element of the matrix $\mathbf{\Pi}$. We detail the procedures in setting initial values for κ , ϑ , and ϱ . Given that $\{z_{k+1}\}$ in (3.10) are IID standard normal, the likelihood function of X_{k+1} is

$$L(X_{k+1}; \kappa, \vartheta, \varrho) = \prod_{k=1}^m \frac{1}{\sqrt{2\pi\varrho}} \exp\left(-\frac{(X_{k+1} - \vartheta - \kappa X_k)^2}{2\varrho^2}\right) \quad (3.36)$$

where $1 \leq m \leq 1460$ in our case. Equivalently, our task is to seek for the maximisers of the sum of log-likelihood, i.e.,

$$\begin{aligned} & \operatorname{argmax}(\log L(X_{k+1}; \kappa, \vartheta, \varrho)) \\ &= \sum_{k=1}^m \left(\log \frac{1}{\sqrt{2\pi\varrho}} - \frac{(X_{k+1} - \vartheta - \kappa X_k)^2}{2\varrho^2} \right). \end{aligned} \quad (3.37)$$

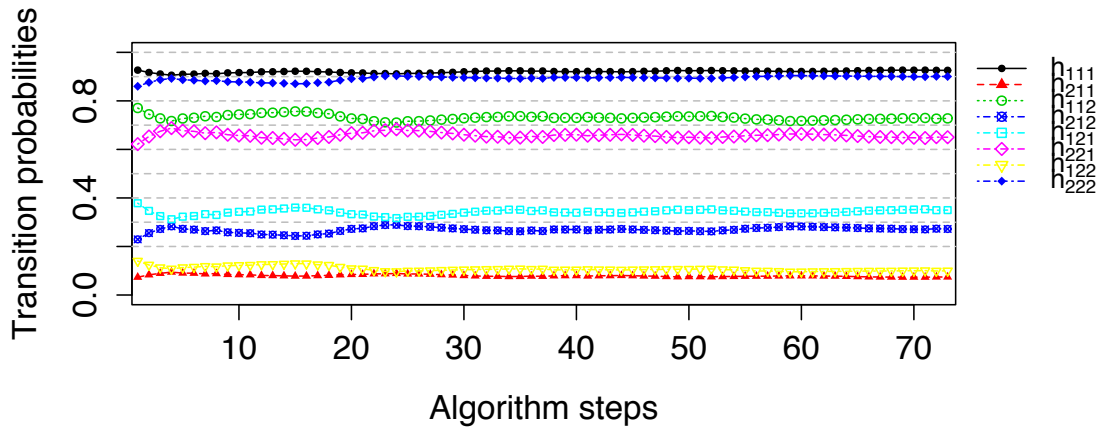


Figure 3.3: Evolution of transition probabilities under a 2-state HOHMM-based model

We employ the R function ‘optim’ to solve the optimisation problem in (3.37), and get $\kappa = 0.6518$, $\vartheta = -0.001624$ and $\varrho = 0.4271$ acting as benchmarks in selecting initial values for the parameter estimation of frameworks with more than 1 regime.

3.5.4.2 Evolution of parameter estimates and comparison under HMM and HOHMM settings

Propositions 1 and 2 aid in getting numerical results from running the self-tuning filtering algorithms (3.27)-(3.30) on batches of data. Filtered estimates are then fed correspondingly into parameter estimate representations in (3.32)-(3.35). The filters process new information, and in turn, optimally update the parameter estimates. One algorithm run constitutes an algorithm pass, and we have 73 passes in total. In every algorithm pass beyond the initial pass, the parameter estimates from the previous pass serve as initial parameter values for the succeeding pass. The filtering algorithm is implemented on the process X_t driven by two-state and three-state HOHMCs.

For a complete comparison of model settings, we also perform filtering implementations on two special cases: 2-state and 3-state HMM frameworks. Note that HMM is a special case of HOHMM when the lag order is 1. Although not plotted, similar dynamic characteristics for the evolution of parameter estimates under both the 2-state and 3-state HMM settings are obtained when compared to the results from the HOHMM settings. The dynamic movements of the estimates are depicted in Figures 3.3–3.5. These gradually converge to certain

unique values after approximately 25 passes in both the 2-state and 3-state HOHMM set ups.

Akin to the reliability of parameter estimates is the quantification of their variability. Therefore, we examine the variance of the various estimators via the Fisher information $I(v^w) = -\mathbb{E}_{v^w} \left[\frac{\partial^2}{\partial v_w^2} \log L(X; v^w) \right]$, which bounds the asymptotic variance of the ML estimates. The MLE is consistent and has an asymptotically normal sampling distribution [36]. So, we utilise the limiting distribution of the ML estimator \widehat{v}^w to obtain the 95% confidence interval for the estimated v^w in the form $\widehat{v}^w \pm 1.96 \frac{1}{\sqrt{I(\widehat{v}^w)}}$. The Fisher information involved in each estimator is derived straightforwardly from the log-likelihood functions in the EM algorithm calculations detailed in the Appendix C, and the final results of such derivations are listed below.

$$I(h_{tsr}) = \frac{\widehat{\mathbf{A}}_k^{tsr}}{h_{tsr}^2}, \quad (3.38)$$

$$I(\kappa_t) = \frac{\widehat{\mathbf{C}}_k^t(X_{k-1}^2)}{\varrho_t^2}, \quad (3.39)$$

$$I(\vartheta_t) = \frac{\widehat{\mathbf{B}}_k^t}{\varrho_t^2}, \quad (3.40)$$

$$\begin{aligned} I(\varrho_t) = & -\frac{\widehat{\mathbf{B}}_k^t}{\varrho_t^2} + \frac{3}{\varrho_t^4} (\widehat{\mathbf{C}}_k^t(X_k^2) \\ & + \widehat{\mathbf{B}}_k^t \vartheta_t^2 + \kappa_t^2 \widehat{\mathbf{C}}_k^t(X_{k-1}^2) - 2\widehat{\mathbf{C}}_k^t(X_k) \vartheta_t, \\ & - 2\widehat{\mathbf{C}}_k^t(X_{k-1}, X_k) \kappa_t + 2\widehat{\mathbf{C}}_k^t(X_{k-1}) \vartheta_t \kappa_t). \end{aligned} \quad (3.41)$$

The 95% confidence intervals for the parameters $\boldsymbol{\kappa}$, $\boldsymbol{\vartheta}$, and $\boldsymbol{\varrho}$ in the 2-state HOHMM-based model are shown in Figure 3.6.

The 95% confidence intervals generated for estimated parameters throughout the 73 algorithm passes are extremely narrow and this is attributed to the declining standard errors. Standard errors of all parameter estimates for the 1-, 2- and 3-state models under both the HMM and HOHMM settings were examined and they all become smaller as the the number of algorithm passes increases. The narrow ranges exhibited in Table 3.4 reflect that precise estimates are achieved by the EM algorithm with our filtering approach.

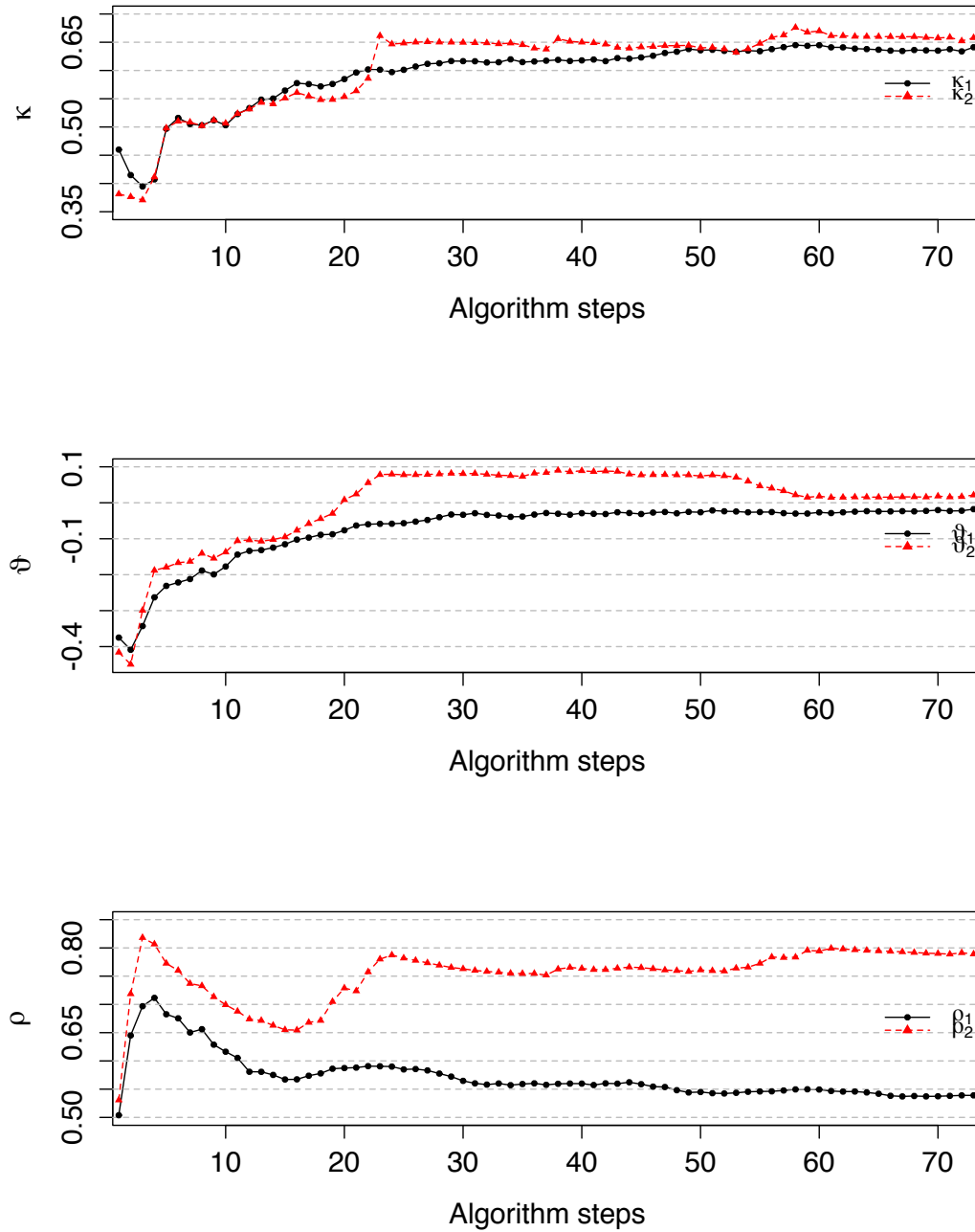


Figure 3.4: Evolution of parameter estimates for κ , θ , and ρ under a 2-state HOHMM-based model

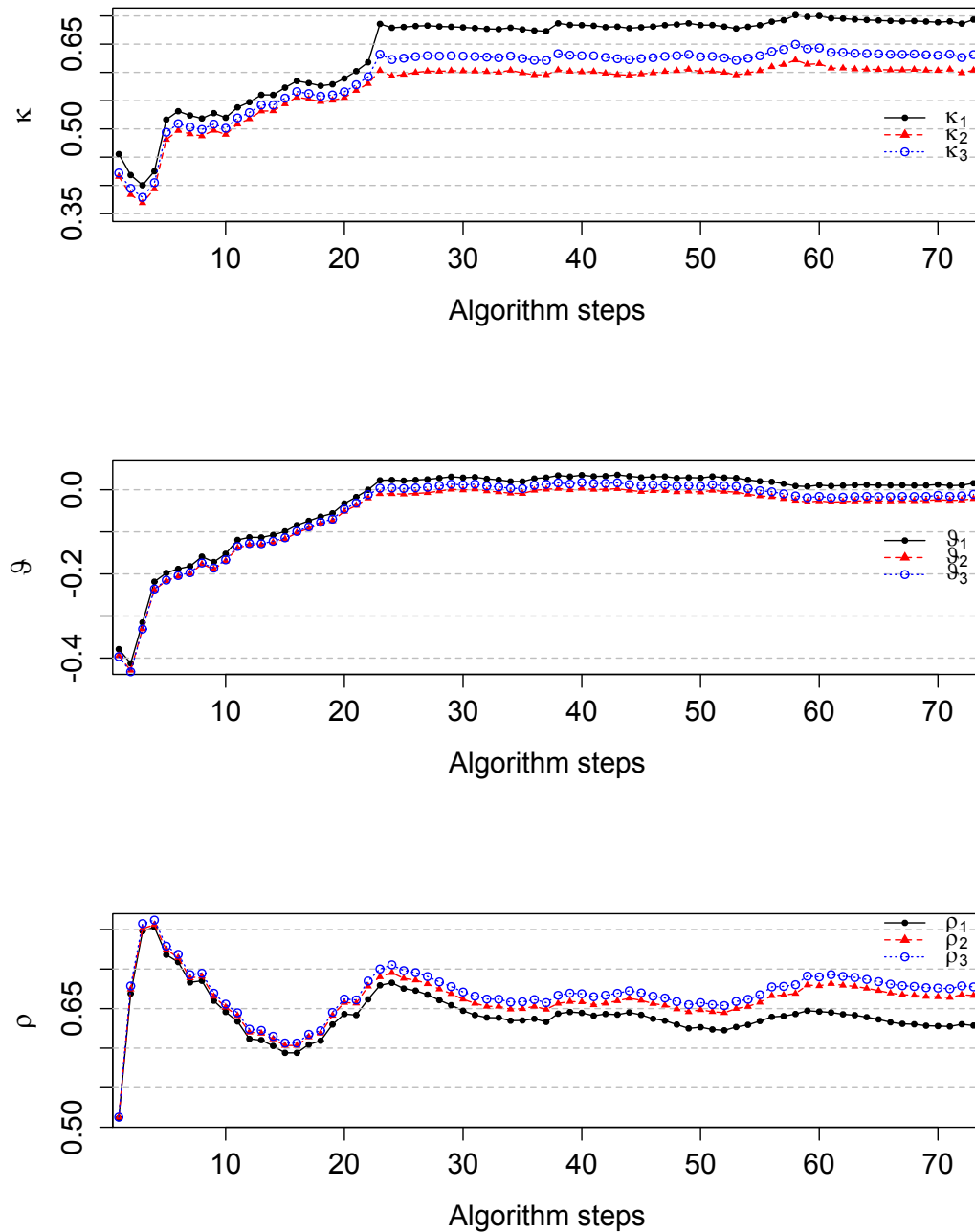


Figure 3.5: Evolution of parameter estimates for κ , θ , and ρ under a 3-state HOHMM-based model

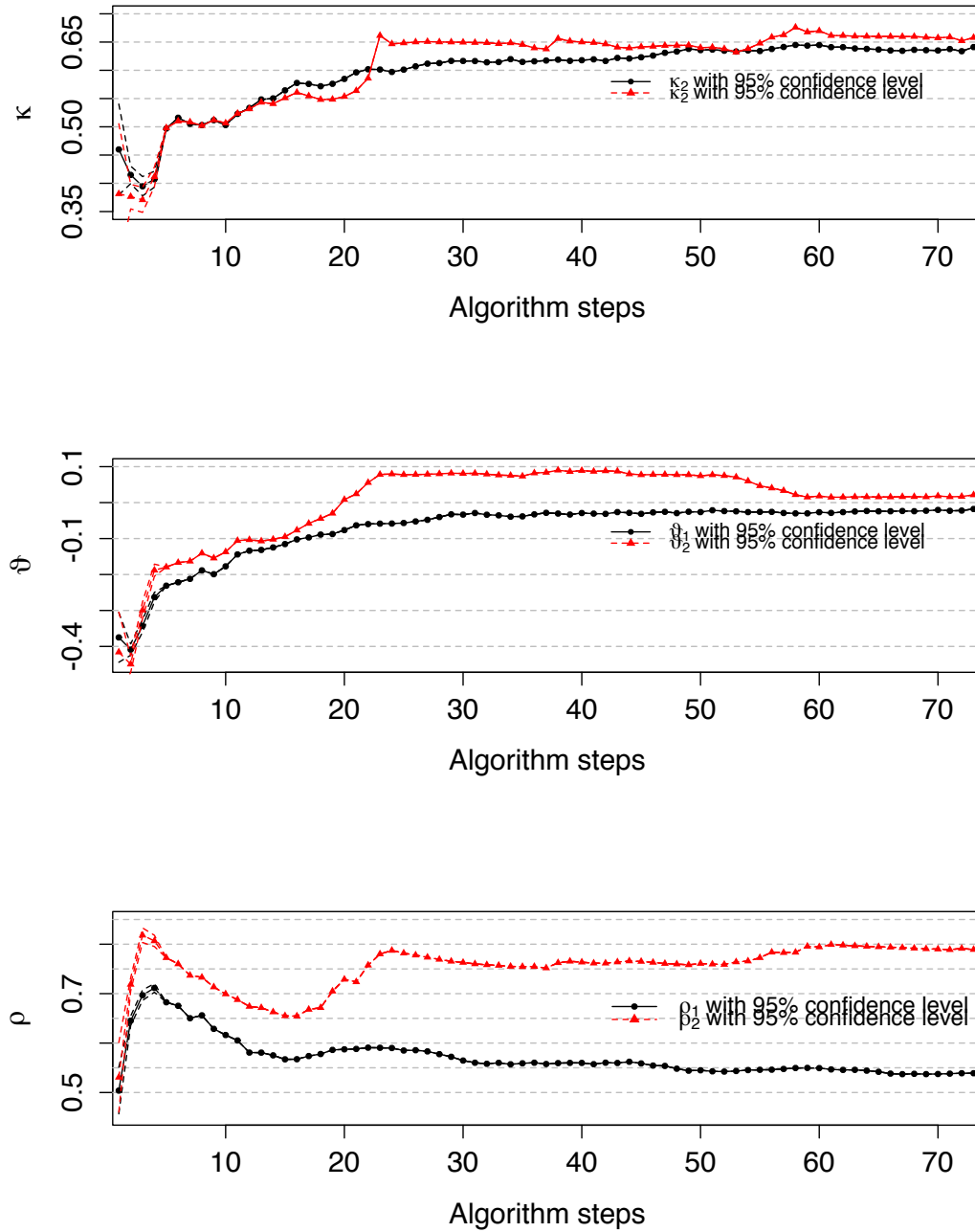


Figure 3.6: Evolution of parameter estimates for κ , θ , and ρ under a 2-state HOHMM-based model with 95% confidence level

HMM						
Parameter	1-state model		2-state model		3-state model	
Estimates	Bound of SE					
	Lower	Upper	Lower	Upper	Lower	Upper
$\widehat{\delta}_i$	$3.62152 * 10^{-90}$	0.84905	$2.88315 * 10^{-93}$	$5.69605 * 10^{-2}$	$1.23351 * 10^{-89}$	0.50619
$\widehat{\eta}_i$	$3.13262 * 10^{-90}$	0.81819	$2.49571 * 10^{-93}$	$5.04276 * 10^{-2}$	$1.28888 * 10^{-89}$	0.50535
$\widehat{\epsilon}_i$	$2.21510 * 10^{-90}$	0.54026	$1.08302 * 10^{-93}$	$2.36474 * 10^{-2}$	$7.17870 * 10^{-90}$	0.34230
\widehat{p}_{ji}	$4.79291 * 10^{-90}$	1.08185	$1.24810 * 10^{-93}$	$3.60257 * 10^{-2}$	$3.52893 * 10^{-90}$	0.34041

HOHMM						
Parameter	1-state model		2-state model		3-state model	
Estimates	Bound of SE					
	Lower	Upper	Lower	Upper	Lower	Upper
$\widehat{\kappa}_t$	$3.62152 * 10^{-90}$	0.84905	$1.21738 * 10^{-96}$	$6.33495 * 10^{-2}$	$1.53624 * 10^{-90}$	$7.84591 * 10^{-1}$
$\widehat{\vartheta}_t$	$3.13262 * 10^{-90}$	0.81819	$8.94466 * 10^{-97}$	$5.67838 * 10^{-2}$	$1.33044 * 10^{-90}$	$7.57304 * 10^{-1}$
$\widehat{\varrho}_t$	$2.21510 * 10^{-90}$	0.54026	$6.32678 * 10^{-97}$	$3.59050 * 10^{-2}$	$9.40836 * 10^{-91}$	$4.96797 * 10^{-1}$
\widehat{h}_{tsr}	$4.79291 * 10^{-90}$	1.08185	$4.74425 * 10^{-97}$	$7.03355 * 10^{-2}$	$6.79239 * 10^{-91}$	$8.85929 * 10^{-1}$

Table 3.4: Interval of standard errors for the parameter estimates under 1-, 2-, 3-state HMM- and HOHMM-based models

3.5.5 Model selection and other diagnostics

One-step ahead forecasts from both the HMM and HOHMM settings are compared. We also present a forecasting error analysis accompanied by an AIC analysis on the selection of optimal number of HMM and HOHMM states.

3.5.5.1 Assessment of predicted temperatures

To make predictions of the DAT values over a one-step ahead time interval, we evaluate the expected value of the observation process at time $k + 1$. Given X_{k+1} in (3.10), we have

$$\begin{aligned} \mathbb{E}[X_{k+1} | \mathcal{X}_k] &= \mathbb{E}[\kappa(\mathbf{y}_k^w) X_k + \vartheta(\mathbf{y}_k^w) + \varrho(\mathbf{y}_k^w) z_{k+1} | \mathcal{X}_k] \\ &= \langle \boldsymbol{\kappa}, \widehat{\mathbf{y}}_k^w \rangle X_k + \langle \boldsymbol{\vartheta}, \widehat{\mathbf{y}}_k^w \rangle. \end{aligned} \quad (3.42)$$

Figure 3.7 depicts the one-step ahead forecasts both for the deseasonalised data X_k and DATs under the 3-state HOHMM-based model using (3.42). The very short-term predictions for X_k and DATs are very close to the actual data series, and the same can also be said for X_k and DATs under the HMM-based models. We also present a comparison

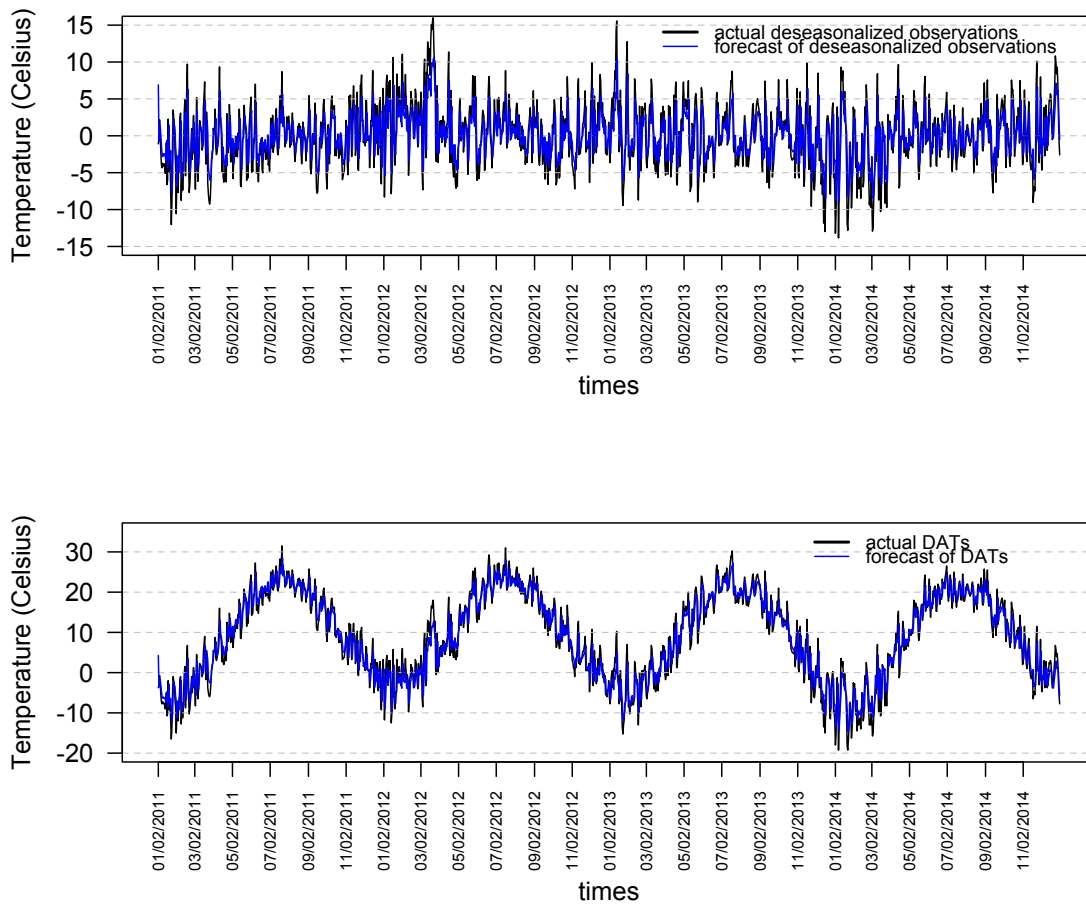


Figure 3.7: One-step ahead forecasts under a 3-state HOHMM-based model

between the observed seasonal HDD and the predicted seasonal HDD covering the entire period (2011-2014) under the 3-state HOHMM-based settings in Figure 3.8. Similar comparison was performed, though not shown, under the HMM-based settings. The seasonal HDD forecasts obtained from the proposed models follow closely the actual seasonal HDD. Furthermore, the magnified view of the predicted DATs under the HOHMM set ups is given in Figure 3.9 covering a 6-month period; DATs forecasts under the HMM setting are also generated but not shown here. All forecasts are relatively close to the actual DATs. The dynamics and trends of the temperature series are captured well by our filtering algorithms and self-calibrating parameter estimation method.

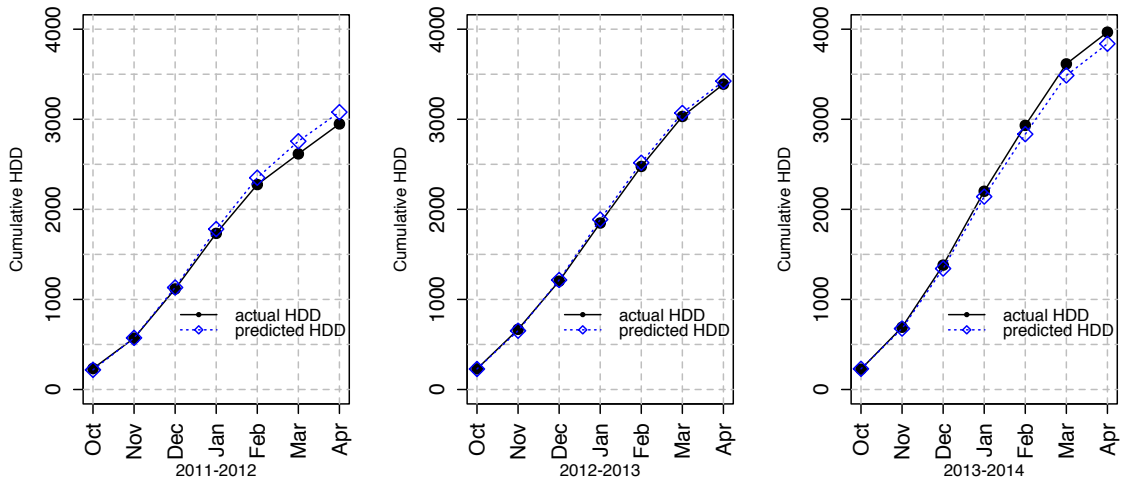


Figure 3.8: Comparison of the expected HDD and actual HDD in a 3-state HOHMM-based model

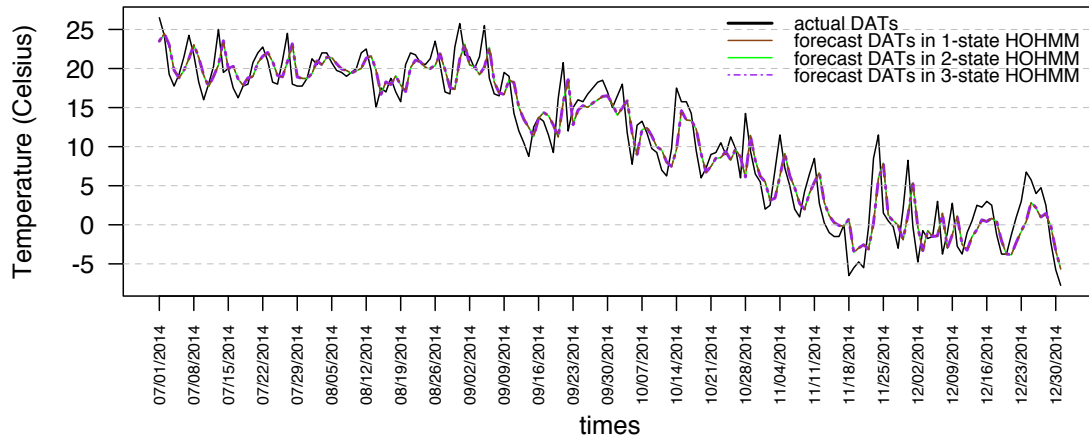


Figure 3.9: Comparison of one-step-ahead forecasts in 1-, 2-, and 3-state HOHMM-based models

3.5.5.2 Error analysis and model selection

We perform an error analysis to quantify the goodness of fit of various HMM and HOHMM settings using the criteria put forward in Erlwein et al. [19], which are the mean square error (MSE), root mean square error (RMSE), absolute mean error (MAE) and relative absolute error (RAE). Suppose X_k denotes the actual value at time k , \hat{X}_k stands for the one-step ahead prediction value at time k , \bar{Y} is the mean of X_k 's, and $n = 1460$ is the total number of predicted values.

$$\text{MSE} = \frac{\sum_{k=1}^n (\hat{X}_k - X_k)^2}{n}, \quad \text{RMSE} = \sqrt{\frac{\sum_{k=1}^n (\hat{X}_k - X_k)^2}{n}},$$

$$\text{MAE} = \frac{\sum_{k=1}^n |\hat{X}_k - X_k|}{n} \quad \text{and} \quad \text{RAE} = \frac{\sum_{k=1}^n |\hat{X}_k - X_k|}{\sum_{k=1}^n |X_k - \bar{X}|}.$$

Table 3.5 displays the error-analysis results involving X_k and DATs data under the HMM setting whilst Table 3.6 contains the error-analysis results under the HOHMM setting. Although the errors are generally small, the error metrics illustrate that the 2-state model outperforms the 1- and 3-state models under both HMM and HOHMM frameworks. In addition, the 1-, 2- and 3-state HOHMM-based models produce better forecasts than those generated by their corresponding HMM settings as far as error measures are concerned. The 4-state HMM-based and HOHMM-based settings were also examined, but no evidence of even minimal improvement is achieved by making the model more complex such as including 64 parameters in its probability transition matrix.

To determine if the error-mean differences are statistically significant in each pairwise setting, we perform a t -test using the bootstrapped method. We generate RMSEs for all possible paired HOHMM settings, and then calculate, using the R function 'p.adjust', the adjusted p -values with the Bonferroni's method to control the familywise error rate. For the 1-state HOHMM versus the 2-state HOHMM, the p -value is smaller than 0.01 so that we can conclude there is sufficient evidence of significant difference after adding one more regime into the one-state model. For the 2-state HOHMM vis-à-vis the 3-state HOHMM, the p -value is much greater than 0.05, meaning that we cannot reject the null hypothesis of no difference. The estimated p -values covering three pairs of model settings are shown in Table 3.7. On the basis of a 5% significance level, the 1- and 2-state, and 1-and 3-state models in both the HMM and HOHMM frameworks are statistically different, whilst the 2- and 3-state settings are not. This suggests that there is benefit to using a regime-switching

HMM setting	Deseasonalized X_t			DATs T_t		
	1-state	2-state	3-state	1-state	2-state	3-state
MSE	0.42702	0.42700	0.42887	10.67538	10.67504	10.72172
RMSE	0.65346	0.65345	0.65488	3.26732	3.26727	3.27440
MAE	0.50467	0.50465	0.50728	2.52333	2.52324	2.53640
RAE	0.74435	0.74432	0.74821	0.27539	0.27538	0.27681

Table 3.5: Results of error analysis on HMM-based models

HOHMM setting	Deseasonal X_t			DATs T_t		
	1-state	2-state	3-state	1-state	2-state	3-state
MSE	0.42702	0.42686	0.42698	10.6754	10.67142	10.67445
RMSE	0.65346	0.65334	0.65344	3.26732	3.26671	3.26718
MAE	0.50467	0.50448	0.50463	2.52334	2.52239	2.52315
RAE	0.74435	0.74407	0.74430	0.27539	0.27528	0.27537

Table 3.6: Results of error analysis on HMM-based models

model, and the 2- and 3-state HMM and HOHMM settings possess similar capability to capture the DAT dynamics. That is, a two-state model is sufficient for our data.

We complement our error analysis with a likelihood-based model selection analysis via the AIC. This estimates the Kullback-Leibler information under the ML paradigm, and given by

$$\text{AIC} = -2 \log L(X; \nu) + 2g,$$

where g is the number of estimated parameters in the model, and $\log L(X; \nu)$ denotes the log-likelihood function of the model given the data X and a set of parameters ν . A

Setting	HMM			HOHMM		
	1-state	1-state	2-state	1-state	1-state	2-state
t -test	vs	vs	vs	vs	vs	vs
	2-state	3-state	3-state	2-state	3-state	3-state
p -value	0.01259	7.06352×10^{-5}	0.57193	0.00678	0.01495	1.00000

Table 3.7: Bonferroni-corrected p -values for a paired t -test applied to RMSEs

model deemed the best and is chosen if it is able to balance between fitness (maximise the log-likelihood function) and complexity (minimise the penalty from too many parameters) yielding the lowest AIC value. As a function of the observation X_t , parameter sets v^* and v^w , the log-likelihood functions under the respective HMM and HOHMM settings are

$$\log L(X_k; v^*) = \sum_{k=1}^{\beta} \left[\log \left(\frac{1}{\sqrt{2\pi}\sigma(\mathbf{y}_k)} \right) - \frac{X_{k+1} - \mu(\mathbf{y}_k)}{2\sigma^2(\mathbf{y}_k)} \right], \quad (3.43)$$

$$\log L(X_k; v^w) = \sum_{k=1}^{\beta} \sum_{t=1}^N \langle \mathbf{y}_k^w, \mathbf{e}_t \rangle \left[\log \left(\frac{1}{\sqrt{2\pi}\varrho(\mathbf{y}_k)} \right) - \frac{X_{k+1} - \kappa(\mathbf{y}_k^w) X_k - \vartheta(\mathbf{y}_k^w)}{2\varrho^2(\mathbf{y}_k^w)} \right], \quad (3.44)$$

where β denotes the number of observations in each pass, and N is the number of regimes. From equation (3.10) and the matrix of transition probabilities, we can determine the total number of parameters in each of the HMM and HOHMM settings and these are presented in Table 3.8.

Setting	1-state	2-state	3-state	N -state
HMM	3	8	15	$3N + (N - 1)N$
HOHMM	3	10	27	$3N + (N - 1)N^2$

Table 3.8: Number of estimated parameters under the HMM and HOHMM settings

Employing equations (3.43) and (3.44), and Table 3.8, we compute the AIC values for each model as we run through the algorithm passes. Figure 3.10 illustrates the evolution of the computed AIC values for all candidate models. As the number of regimes grows in a model, there is a substantial increase in the number of parameters especially coming from the number of transition probabilities. Our results show that the 2-state HOHMM has the smallest AIC value and a more stable pattern in the entire data period. Therefore the 2-state model is the best-fitting model for the dynamics of our data set, which agrees with our error analysis.

Although it is typical in the statistical analysis of a proposed financial model to benchmark it against say, random walk, log-normal and ARCH-type models (see Tenyakov et al. [32] and Hardy [23], for example), these common benchmarks are incompatible with our

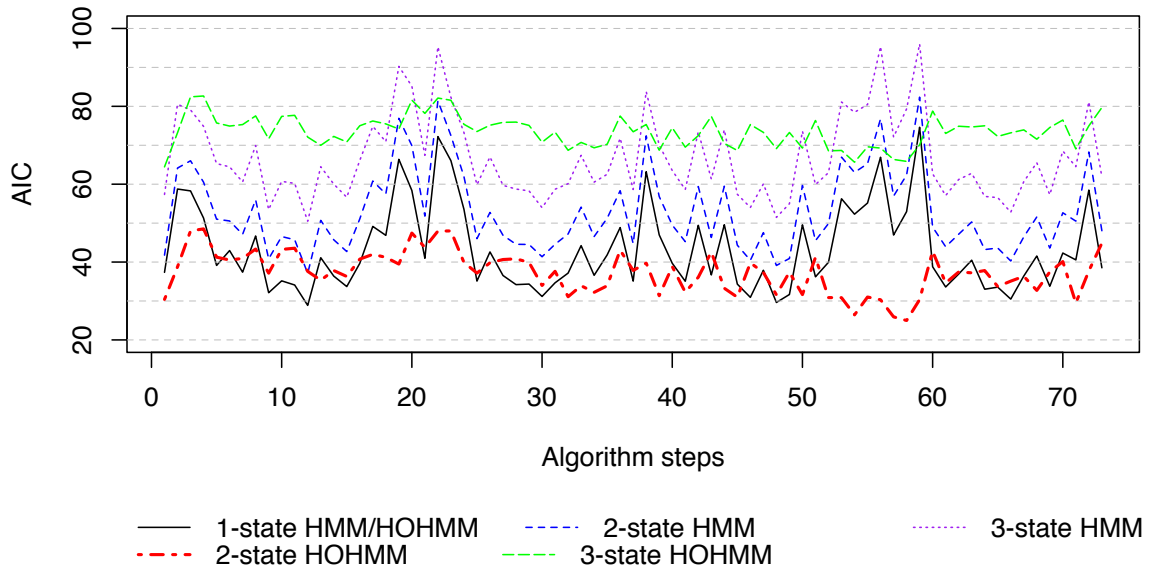


Figure 3.10: Evolution of AIC for 1-, 2-, and 3-state HMM- and HOHMM-based models

modelling situation. More specifically, these models for benchmarking do not have the capability to capture the important property of mean-reversion for temperature data and thus, they are inordinately disadvantaged in comparison to mean-reverting models. Hence, the only meaningful benchmark not involving any Markov chain in our case is the one-state OU process, where there is no switching of regimes. From Figure 3.10, the 1-state model (i.e., no Markov chain) is performing poorly relative to the 2- and 3-state models. This is also supported by the HDD error-analysis metrics shown in Table 3.9.

HDD	HMM setting			HOHMM setting		
	1-state	2-state	3-state	1-state	2-state	3-state
MSE	4725.2238	4724.6591	5754.0650	4725.2238	4705.0409	4724.9993
RMSE	68.7403	68.7362	75.8556	68.7403	68.5933	68.7386
MAE	51.4458	51.4237	57.3876	51.4458	51.3215	51.4243
RAE	0.0504	0.0504	0.0562	0.0504	0.0503	0.0504

Table 3.9: Results of error analysis for the HDD on HMM- and HOHMM-based models

3.6 Conclusion

The major contribution of this chapter is the development of a model flexible enough to describe a data set that exhibit seasonality, randomness, mean-reversion and memory. Such is the case for the dynamics of DATs series, which is the most utilised underlying variable in the creation of weather-dependent derivatives. In particular, we put forward a model following a deterministic seasonality component and the mean-reversion is governed by the OU process. The stochasticity is generated by the use of the Brownian motion in the diffusion component. Embedding an HOHMM facilitated the switching of parameters amongst various conceivable regimes and also captures short or long-range dependence.

Our estimation procedure for all the model parameters is successfully achieved through the extension of the HMM-OU filtering techniques applied to a proposed transformed HOHMM. With the change of probability measure and EM algorithm, self-tuning HOHMM recursive filters were obtained that supports online parameter estimation. Our results also encompass the special case of HOHMM with lag 1, which is the case for the usual HMM recursive filters. Our modelling methodology was tested on Toronto's DATs covering a 4-year period. Our post-modelling diagnostics reveal reliable one-step ahead forecasts under various settings. We must note though that the 2-state HOHMM provides the best framework for the data that we investigated. This is validated by the AIC analysis and the accompanying error analyses.

The current HMM-modulated models have been employed in many areas of finance and economics, such as the modelling of commodity futures prices, developing asset allocation strategies, valuing interest-rate products, and so on. HMM- and HOHMM-based methods in weather derivatives are virtually nonexistent until the publication of the 2-state regime-switching temperature model of Elias et al.[14]. This work further elevates that development temperature modelling for weather derivatives in three respects: (i) provision of dynamic estimation with efficient algorithms, (ii) generalisation of the usual HMM to take advantage of information beyond lag one, and (iii) empirical results that reinforce the choice of the optimal regime dimension. We believe that our proposed method offers an effective alternative to available approaches in efficient modelling and forecasting temperature dynamics with easily interpretable results.

A natural direction to pursue further is to price weather derivatives by finding a risk-neutral

measure and linking this pricing measure to the optimal estimates produced from our proposed filtering procedure. Whilst this is beyond the objectives of this study, we have laid down some groundwork to motivate that kind of research exploration. It would be also worth examining this HOHMM-based filtering algorithms and the parameter estimates in relation to the modelling of other weather measurements such as wind speed, level of rainfall, etc, either for the purpose of financial pricing or meteorological modelling and forecasting. Finally, an apparent weakness of this research is the modeller's prescription of HOHMM's lag order before the filtering algorithms could be implemented. It is hoped that this could be rectified in future research through the construction of statistical inference techniques that estimate the correct lag order as implied by the data.

References

- [1] P. Alaton, B. Djehiche, D. Stillberger, On modeling and pricing weather derivatives, *Applied Mathematical Finance*, 9 (1) (2002) 1–20.
- [2] F. Benth, J. Šaltytė-Benth, Weather derivatives and stochastic modelling of temperature, *International Journal of Stochastic Analysis*, 1–21 (2011). doi:10.1155/2011/576791
- [3] F. Benth, J. Šaltytė-Benth, Stochastic modelling of temperature variations with a view towards weather derivatives, *Applied Mathematical Finance*, 12 (1) (2005), 53–85.
- [4] J. Beran, *Statistics for Long-Memory Processes*, Chapman and Hall, New York (1994).
- [5] S. Campbell, F. Diebold, Weather forecasting for weather derivatives, *Journal of the American Statistical Association*, 100 (469) (2005) 6–16.
- [6] M. Cao, J. Wei, Weather derivatives: a new class of financial instruments, University of Toronto Web, <http://www.rotman.utoronto.ca/wei/research/JAI.pdf> (2003), Accessed March 2016.
- [7] W. Ching, T. Siu, L. Li, Pricing exotic options under a high-order Markovian regime switching model, *Journal of Applied Mathematics and Decision Sciences.*, (2007) 1–15, doi:10.1155/2007/18014.
- [8] G. Considine, Introduction to weather derivatives, CME Group Web, <http://web.math.pmf.unizg.hr/amimica/pub/lapl.pdf>, Accessed March 2016
- [9] P. Date, A. Tenyakov, R. Mamon, Filtering and forecasting commodity futures prices under an HMM framework, *Energy Economics*, 40 (2013) 1001-1013.
- [10] M. Davis, Pricing weather derivatives by marginal value, *Quantitative Finance*, 1 (3) (2001) 305–308.

- [11] B. Dischel, At least: a model for weather risk, *Weather Risk Special Report, Energy and Power Risk Management*, March issue (1998) 3032.
- [12] G. Dorfleitner, M. Wimmer, The pricing of temperature futures at the Chicago Mercantile Exchange, *Journal of Banking and Finance*, 34(6) (2010) 1360–1370.
- [13] J. Dutton, Opportunities and priorities in a new era for weather and climate services, *Bulletin of the American Meteorological Society*, 83 (9) (2002) 1303-131.
- [14] R. Elias, M. Wahab, L. Fang, L., A comparison of regime-switching temperature modeling approaches for applications in weather derivatives, *European Journal of Operational Research*, 232 (3) (2014) 549–560.
- [15] R. Elliott, Exact adaptive filters for Markov chains observed in Gaussian noise, *Automatica*, 30 (1994) 1399–1408.
- [16] R. Elliott, L. Aggoun, J. Moore, J., *Hidden Markov Models: Estimation and Control*, Springer, New York (1995).
- [17] R. Elliott, V. Krishnamurthy, New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models, *IEEE Transactions on Automatic Control*, 44 (5) (1999) 938-051.
- [18] C. Erlwein , R. Mamon, An online estimation scheme for a HullWhite model with HMM-driven parameters, *Statistical Methods and Applications*, 18 (1) (2009) 87–107.
- [19] C. Erlwein, F. Benth, R. Mamon, HMM filtering and parameter estimation of an electricity spot price model, *Energy Economics*, 32 (5) (2010) 1034–1043.
- [20] D. Gujarati, *Basic Econometrics* (4th ed), McGraw-Hill, New York (2003).
- [21] C. Granger, R. Joyeux, An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1 (1980) 49–64.
- [22] J. Hamilton, Rational expectations econometric analysis of changes in regime: an investigation of term structure of interest rates, *Journal of Economic Dynamics and Control*, 12 (1988) 385–423.
- [23] M. Hardy, A regime-switching model of long-term stock returns, *North American Actuarial Journal*, 5 (2) (2001) 41–53.

- [24] J. Hull, A. White, Pricing interest-rate-derivative securities, *Review of Financial Studies*, 3 (4) (1990) 573–592.
- [25] S. Jewson, A. Brix, C. Ziehmann, *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*, Cambridge University Press, Cambridge (2005).
- [26] R. Mamon, R. Elliott, *Hidden Markov Models in Finance: International Series in Operations Research and Management Science*, 104, Springer, New York, (2007).
- [27] R. Mamon, C. Erlwein, B. Gopaluni, Adaptive signal processing of asset price dynamics with predictability analysis, *Information Sciences*, 178 (1) (2008) 203–219.
- [28] Price Waterhouse Coopers, 2011 Weather risk derivative survey, Weather Risk Management Association,
<http://library.constantcontact.com/download/get/file/1101687496358-153/PwC+Survey+Final+Presentation+20110519PRESS.pdf>
(2011), Accessed March 2016
- [29] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 77 (2) (1989) 257–286.
- [30] V. Reisen, Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram, *Journal of Time Series Analysis*, 15 (1994) 335–350.
- [31] T. Siu, W. Ching, E. Fung, M. Ng, X. Li, A high-order markov-switching model for risk measurement, *Computers and Mathematics with Applications*, 58 (1) (2009) 1–10.
- [32] A. Tenyakov, R. Mamon, M. Davison, Modelling high-frequency FX rate dynamics: A zero-delay multi-dimensional HMM-based approach, *Knowledge-Based Systems*, 101 (2016) 142–155.
- [33] A. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge and New York (1998).
- [34] World Meteorological Organization (WMO) and the Centre for Research on the Epidemiology of Disasters (CRED) of the Catholic University of Louvain (UCL), Atlas

of Mortality and Economic Losses from Weather, Climate and Water Extremes 1970-2012, WMO Press, Belgium (2014).

- [35] C. Wu, On the convergence properties of the EM Algorithm, *Annals of Statistics*, 11 (1) (1983) 95–103.
- [36] X. Xi, R. Mamon, Parameter estimation of an asset price model driven by a weak hidden markov chain, *Economic Modelling* 28 (1) (2011) 36–46.

Chapter 4

A higher-order Markov chain-modulated model for electricity spot-price dynamics

4.1 Introduction

Electricity contributed greatly to technological revolution and hastened industrial progress. As one of the pivotal mankind's discoveries, it is indispensable in our modern society. It changed, in a significant way, almost every aspect of our daily lives. It would be hard to imagine a world without electronics, machines, equipments and inventions that are energised by electric power nowadays. Considering the advantages of economies of scale and steady industrial productivity, regulation of electricity supply was centralised; but, this inhibited any competition at all. In terms of economic efficiency, microeconomic theory maintains that introducing competition can motivate production innovation, increase supply diversity, and offer most benefits to consumers. Since privatisation and competition were first introduced to electric power systems in Chile [28], electricity markets have experienced restructuring and deregulations in many countries. Electricity is now a daily necessity; given its prime importance, it is a commodity actively trading in the financial markets. Consequently, we are witnessing increasing price uncertainties and risks in electricity-driven investment portfolios. So, models capable of adequately capturing electricity price dynamics for electricity- contract pricing and risk management are sought. Existing research studies attempt to formulate electricity pricing models that mimic model-pricing developments in the commodity market. However, due to the unique characteristics of

electricity spot prices, many financial models designed for regular commodities cannot be adapted necessarily to the electricity markets.

Compared to common and tangible assets in the capital markets, electricity is non-storable; when produced, it must be consumed almost instantly. This non-storability feature makes its spot price highly sensitive to demand and supply in real time. As a result, electricity has more dramatic price evolutions than those of other energy sources, such as crude oil and natural gas. Since electricity cannot be stored, electricity prices are largely dependent upon supply and demand, which exhibit pronounced seasonality patterns. Somewhat different from general seasonality, highly occurring multi-cyclical nature of prices is evident in electricity markets. Annual and quarterly patterns, for example, are attributed principally to the variations in temperature and duration of daylight, especially during winter and summer. Cyclical patterns also occur weekly, daily and even intra-daily caused by quantity demanded that is varying in time.

Similar to other energy commodities, it is also critical to incorporate the mean-reverting property in electricity prices as argued in Schwartz [31]. Owing to demand fluctuations, the price can deviate from the long-run mean. However, the production cost and supply adjustments are capable of hauling it back to the mean level over time, hence, the mean-reverting behaviour of electricity prices. Unbalanced supply-and-demand situation that happens expectedly or unexpectedly from time to time could bring about abrupt and large jumps on the prices. Capturing simultaneously the above-mentioned properties of the price dynamics is a great challenge in electricity spot-price modelling in the framework of any deregulated markets. Yet, this difficulty must be dealt with in order to address the concerns nestled in the trading and hedging of electricity-dependent derivatives.

Modelling electricity prices can be divided into two roughly categorical approaches. The first approach focuses on the modelling of the entire forward curve akin to the valuation of associated derivatives and short-term prediction of futures prices; see Islyayev and Date [19] and Fanelli et al. [12]. The second approach concentrates on model formulation of spot prices taking into account several and only the most relevant fundamental price drivers observed in electricity markets; see Ziel et al. [47]. The emphasis of our research is on the latter approach. Early research works mainly put forward stochastic models that only capture seasonality and mean reversion in prices. The seasonal systematic patterns are typically described by sinusoidal functions [23] and the Ornstein-Uhlenbeck (OU) process is

prevalently utilised to model the mean reversion as noted in Schwartz [31]. A two-factor model in an OU-inspired setting was developed by Schwartz and Smith [32] in an effort to explicitly portray seasonal patterns and short-term mean-reverting variations in prices at the equilibrium level. However, most of these early literature did not take into account the peculiarity of electricity prices and examine it thoroughly.

Motivated by the phenomenon of frequent but large price jumps observed in the electricity markets, scholars began to develop modelling capabilities that can handle spikes in electricity-price fluctuations. Deng [6] pioneered the use of jump-diffusion models with a mean-reverting process to replicate the distinct characteristics of electricity-spot prices; such models were fitted to data from the American markets. Benth et al. [1] added jumps to a general exponential multi-factor mean-reverting model and performed calibration employing data from the Nord-Pool market. In Seifert and Uhrig-Homburg [34], different models for the jump component in the electricity markets were explored and the effectiveness of different jump specifications compared; such models were applied to the European Energy Exchange (EEX) market. Many research proponents advanced the utility of Poisson-jump models, whilst others, in recent literature, argued that modelling the ‘jumpy’ attribute could be achieved better by introducing a regime-switching approach. Huisman and Mahieu [18] proposed a regime-shifting jump process with a recovery state and concluded that it performed better than a stochastic jump model whilst a two-regime-switching model with ‘abnormal’ and ‘normal’ states was shown superior to a Poisson-jump model in De Jong [7] in capturing electricity-price dynamics. Wu et al. [37] studied the Alberta electricity-spot market via a hidden Markov model with a multi-model identification approach. Alberta electricity pool prices are divided into five classes, and the last three classes with prices over \$100/MWh are deemed high-pool-price regions. It was found that price-forecasting performance could be improved substantially, especially in high-pool-price regions, by incorporating a Markovian regime-switching mechanism.

In the context of regime-switching approaches, hidden Markov models (HMMs) have been widely successful in many engineering, economic and financial applications; see Mamon and Elliott [20, 25]. HMMs are beneficial for modelling processes illustrating regime-switching dynamics via Markov chains with latent states. This approach was previously applied in the examination of electricity-price behaviour. A non-stationary model based on input-output HMM was developed to model time series of spot prices in the Spanish electricity market [15]. As well, Yu and Sheblé [46] utilised HMMs to efficiently model

price movements in the electricity market and provide good forecasts adjudged from the perspective of accuracy and dynamic information in the US market.

The assumption of first-order state transition dependency in HMM renders this model deficient in taking advantage of the information content of historical data. Thus, researchers introduced recently higher-order hidden Markov models (HOHMMs), also called weak hidden Markov models. An HOHMM is a doubly embedded stochastic process having an observation series and an underlying unobserved Markov chain. The probability distribution of the Markov chain's state transition at present depends not only on the most recent state but also on states at prior epochs into the past. The primary aim in practice is to obtain the best estimate of the Markov chain's current or future state, which represents the 'filtered' or 'predicted' state of the market or economic system. This estimation is performed by taking the conditional expectations of functions of the Markov chain and observed electricity spot prices that are regarded as offshoot of the interaction of many latent factors, such as market participants' actions, production and consumption, weather conditions, transmission network, etc. In our context, HOHMM setting is designed to capture the presence of memories in electricity market prices that will provide additional information in the estimation and forecasting of economic regimes and other model parameters.

An array of HOHMMs' applications can be found in speech recognition and finance, amongst other research areas. An HOHMM for piecewise linear processes was developed by Lee and Jean [22] to approximate the behaviour of a real process in speech recognition. Xiong and Mamon [43] demonstrated that an HOHMM captures more accurately the empirical characteristics of a data set on daily average temperature (DAT) when compared to the usual HMM. With the aid of HOHMMs, Xi and Mamon [38] devised a more flexible framework and showed better performance in forecasting the risky asset's log returns. Other HOHMM-related works in financial modelling include [39], [40], [41], and [42].

To the best of our knowledge, this research is the first to build an HOHMM-modulated electricity spot-price model. The regular HMM approach is extended in our approach and the HOHMM's prediction performance is examined. Our work can be viewed both as an update and extension of Erlwein et al.'s model construction [7] that was based on an exponential-OU process with a jump component under the HMM setting; albeit Erlwein et al.'s empirical application covers the Nord-Pool market whilst ours investigates the Alberta electricity market. The structure of the electricity industry in Alberta is unique in North

America and apparently different from that of Norway, as the former is developed from a set of uniquely distinguishable circumstances. In particular the Alberta Electric System Operator (AESO) relies on wind farms; and to avoid power shortages when wind drops off and does not blow consistently, coal and natural gas plants have to take up the slack.

Modelling the behaviour of electricity spot prices is of utmost importance because they serve as the underlying variable for the values of many traded electricity contracts and they are key indicators in the strategic planning and investment-decision making of various stake holders in the electricity market. Our main research contribution is the creation of an HOHMM-OU-Poisson framework that simultaneously delineates five salient properties of electricity spot prices. The implementation portion of this work highlights model validation and post-modelling diagnostics. Capitalising on Erlwein et al.'s use of HMM on electricity spot price modelling [7], this research further highlights HOHMC-modulated model parameters to address memory in time series data. Our recursive filters are also expressed more compactly in a matrix notation.

This chapter is organised as follows. In Section 4.2, we building an electricity spot price model whose parameters are governed by an HOHMM in discrete time. Through a change of reference probability technique, adaptive filters are presented in Section 4.3 for the states of the HOHMM and related quantities of the observation process. We outline the self-calibrating parameter estimation scheme in terms of the recursive filters via the EM algorithm in Section 4.4. Numerical implementation, which includes assessment of model's goodness of fit and prediction ability, is performed on a 4-year AESO data in Section 4.5. Implication to forward pricing of our proposed HOHMM-modulated electricity spot price model is also illustrated. Section 4.6 provides some concluding remarks.

4.2 Model formulation

4.2.1 Model for electricity spot price

As noted above, studies in the earlier literature employ time series models to describe the price process of electricity. An OU process with a predictable component is a germane example, and this was adopted by Lucia and Schwartz [23] to replicate the unfolding of the log-spot prices in the Nordic exchange. Yousef [45] applied mean-reverting diffusion

models to the Alberta electricity market; however, these models could not produce spikes in the dynamics of electricity spot prices. Invoking Benth et al. [1], we suppose the electricity spot price S_t at time t , defined on some probability space (Ω, \mathcal{F}, P) , is given by

$$S_t = D_t e^{X_t}. \quad (4.1)$$

Equation (4.1) comprises two components to handle seasonality, and stochasticity coming from normal perturbations and jump behaviours. The deterministic function D_t accounts for regularities in the price evolutions and periodic trends. The stochastic part X_t is assumed to be an OU process (to rationalise the tendency to return to a long-run mean) plus a compound Poisson process that deals with excessive volatilities and spikes.

Trigonometric functions were applied by various authors to capture processes' seasonality in different complex systems. In the power market for instance, Lucia and Schwartz [23] included sinusoidal and dummy variables for prices in the Nordic market. Linear trend and cosine functions were introduced to model spot prices' seasonality in the US electricity market by Geman and Roncoroni [14]. As in Xiong and Mamon [43, 44], daily temperature data series can be modelled by a combination of sinusoidal functions, with varying seasonal frequencies, and a linear trend. Following [43] and [1], D_t has the specification

$$D_t = at + b + \sum_{h=1}^3 \left[c_h \sin\left(d_h t \frac{2\pi}{365}\right) + e_h \cos\left(d_h t \frac{2\pi}{365}\right) + g_h \sin\left(d_h t \frac{2\pi}{7}\right) + j_h \cos\left(d_h t \frac{2\pi}{7}\right) \right], \quad (4.2)$$

where $d_1 = 1$, $d_2 = 2$, and $d_3 = 4$ to cover periodic trends including the annual, semi-annual, quarterly, monthly, weekly, and mid-week patterns. The constants a , b , c_h , e_h , g_h , and j_h are recovered from a data set of prices. Such a deterministic function $D(t)$ may be extended with the augmentation of more terms to elucidate other possible factors at play (e.g., effects of generation capacity and holidays). We argue, however, that since temperature is a major driver for electricity demand and contributes substantially to seasonal variations, equation (4.2) can sufficiently portray the seasonal characteristics of electricity spot prices.

4.2.2 The fusion of HOHMM, OU and compound Poisson models

In the ensuing exposition, we shall denote all vectors and matrices by bold small English/Greek letters and bold capitalised English/Greek letters, respectively. The deseasonalised component X_t in (4.2) follows an OU model and an additive compound-Poisson

component; these equip the capability to model mean reversion, price variations and spike dynamics. The process X_t in equation (4.1) satisfies the stochastic differential equation (SDE)

$$dX_t = \alpha(\theta - X_t)dt + \xi dW_t + dJ_t, \quad (4.3)$$

where α , θ , and ξ stand for the speed of mean reversion, mean-reverting level, and volatility, respectively. A Brownian motion $\{W_t\}$ models the random price fluctuations under a stable market condition, whilst $\{J_t\}$ is designed to pick up the price spikes. A compound Poisson process affords capacity to model jumps, as adduced in [11, 20] and similar to their settings, we let $dJ_t = \beta d\mathcal{P}_t$, where \mathcal{P}_t is a Poisson process with a constant intensity ζ and a normally distributed jump size $\beta \sim N(\mu_\beta, \sigma_\beta^2)$. For $s \leq t$, the continuous-time solution of equation (4.3), by Itô's lemma, is

$$\begin{aligned} X_t = & X_s e^{\alpha(t-s)} + (1 - e^{-\alpha(t-s)})\theta + \xi e^{-\alpha t} \int_s^t e^{\alpha u} dW_u \\ & + \int_s^t e^{-\alpha(t-u)} dJ_u. \end{aligned} \quad (4.4)$$

By approximating the distributions of the stochastic integrals in (4.4), the discrete-time version of the solution is

$$\begin{aligned} X_{k+1} = & X_k e^{-\alpha \Delta t_{k+1}} + (1 - e^{-\alpha \Delta t_{k+1}})\theta + \xi \sqrt{\frac{1 - e^{-2\alpha \Delta t_{k+1}}}{2\alpha}} z_{k+1} \\ & + \sum_{m=1}^{\mathcal{P}_{\Delta t_{k+1}}} e^{-\alpha(\Delta t_{k+1} - v_m)} \beta_m, \end{aligned} \quad (4.5)$$

where $\Delta t_{k+1} = t_{k+1} - t_k$ for $k \in \mathbb{Z}^* := \mathbb{Z}^+ \cup \{0\}$, v_m is the occurrence time of the m^{th} jump, and $\{z_{k+1}\}$ is a sequence of independent and identically distributed (IID) standard normal random variables.

To create a regime-switching model for electricity spot prices, we begin with a homogeneous Markov chain \mathbf{y}_k with a finite state space $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$, where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top \in \mathbb{R}^N$ with 1 in the i th position, \top is the matrix transpose operator, and N is the state-space dimension. Model parameters in equation (4.5) switch randomly, in accordance with the dictates of \mathbf{y}_k , amongst different electricity-market regimes as time progresses. With the canonical basis as \mathbf{y}_k 's state space, we have $\alpha_k := \alpha(\mathbf{y}_k) = \langle \boldsymbol{\alpha}, \mathbf{y}_k \rangle$, $\theta_k := \theta(\mathbf{y}_k) = \langle \boldsymbol{\theta}, \mathbf{y}_k \rangle$, $\xi_k := \xi(\mathbf{y}_k) = \langle \boldsymbol{\xi}, \mathbf{y}_k \rangle$, and $\beta_k := \beta(\mathbf{y}_k) = \langle \boldsymbol{\beta}, \mathbf{y}_k \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product in \mathbb{R}^N . The

stochastic process X_k in equation (4.5) then can be written as

$$\begin{aligned}
X_{k+1} = & X_k e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}} + \left(1 - e^{-\alpha(\mathbf{y}_k)\Delta t_{k+1}}\right) \theta(\mathbf{y}_k) \\
& + \xi(\mathbf{y}_k) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{y}_k)\Delta t_{k+1}}}{2\alpha(\mathbf{y}_k)}} z_{k+1} \\
& + \sum_{m=1}^{\mathcal{P}_{\Delta t_{k+1}}} e^{-\alpha(\mathbf{y}_k)(\Delta t_{k+1} - v_m)} \beta_m(\mathbf{y}_k).
\end{aligned} \tag{4.6}$$

To completely characterise the parameter estimation and filtering under the HOHMM setting, we concentrate on a second-order hidden Markov chain. Of course, in theory and principle, the estimation and filtering for the generalised Markov chain with order greater than 2 can be extended in a straightforward manner, notwithstanding the corresponding computational challenge. Suppose \mathbf{y}^w is a second-order hidden Markov chain regulating the evolution of α , θ , ξ , and β . We define $\mathcal{F}_k := \mathcal{F}_k^w \vee \mathcal{F}_k^z \vee \mathcal{F}_k^J$ as the global filtration, where \mathcal{F}_k^w , \mathcal{F}_k^z , and \mathcal{F}_k^J are filtrations generated by $\{\mathbf{y}_k^w\}$, $\{W_t\}$ and $\{J_t\}$, respectively. Under the probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_k\}, P)$, the discrete-time hidden Markov chain \mathbf{y}_k^w at the current step k depends on the information revealed at two prior steps $k-1$ and $k-2$. Following [28], write \mathbf{T} for the $\mathbf{R}^{N \times N^2}$ transition probability matrix, and specified as

$$\mathbf{T} := \begin{pmatrix} p_{111} & p_{112} & \cdots & p_{11N} & \cdots & p_{1N1} & p_{1N2} & \cdots & p_{1NN} \\ p_{211} & p_{212} & \cdots & p_{21N} & \cdots & p_{2N1} & p_{2N2} & \cdots & p_{2NN} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ p_{N11} & \cdots & & p_{N1N} & \cdots & p_{NN1} & p_{NN2} & \cdots & p_{NNN} \end{pmatrix},$$

where $p_{dcb} := P(\mathbf{y}_{k+1}^w = \mathbf{e}_d | \mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b)$ with $k \geq 1$ and $d, c, b \in \{1, 2, \dots, N\}$, which stands for the probability that the Markov chain will be in state d at time $k+1$, given that it is in state c at time k and in state b at time $k-1$.

A vital strategy in the estimation and filtering of HOHMMs is a mapping that converts an HOHMM into an HMM, and then the usual HMM filtering methods can then be conveniently applied. For our second-order HMC, let ϖ be a transformation defined by

$$\varpi(\mathbf{e}_b, \mathbf{e}_c) = \mathbf{e}_{bc}, \quad \text{for } 1 \leq b, c \leq N, \tag{4.7}$$

where \mathbf{e}_{bc} is a unit vector with 1 in its $((b-1)N + c)$ th position. We obtain a new Markov chain $\varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$, whose state space is the canonical basis of \mathbf{R}^{N^2} , and

$$\langle \varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w), \mathbf{e}_{bc} \rangle = \langle \mathbf{y}_{k+1}^w, \mathbf{e}_b \rangle \langle \mathbf{y}_k^w, \mathbf{e}_c \rangle. \tag{4.8}$$

The corresponding dynamics are

$$\varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) = \mathbf{A}\varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) + \mathbf{m}_{k+1}^w, \quad (4.9)$$

where $\{\mathbf{m}_{k+1}^w\}_{k \geq 1}$ is a sequence of martingale increments with $\mathbb{E}[\mathbf{m}_{k+1}^w | \mathcal{F}_k^w] = \mathbf{0}$ under the real-world measure P . The matrix $\mathbf{A} = (a_{ji})$ is the $\mathbf{R}^{N^2 \times N^2}$ probability transition matrix with entries $a_{ji} := p_{dcb} = P(\mathbf{y}_{k+1}^w = \mathbf{e}_d | \mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b)$ for $j = (d-1)N + c$, $i = (c-1)N + b$, and $a_{ji} = 0$ otherwise.

4.3 Recursive filtering

4.3.1 Change of reference probability measure

Deriving recursive filters, under the real-world probability P with the realistic supposition of dependent observations, could be computationally expensive and cumbersome. To this end, we shall perform the recursive filtering involving \mathbf{y}_k^w under an ideal reference probability measure \tilde{P} . Under \tilde{P} , the observed X_k 's are IID and \mathbf{y}_k^w has the same dynamics under both P and \tilde{P} . We assume further that the jump component of the stochastic process X_t is unaffected by the change of reference probability measure following Merton [29]. The estimated processes' dynamics, under measure \tilde{P} , can be recovered via a discrete-time version of the Girsanovs theorem [9] following a reverse measure change.

To aid the filtering computations in our HOHMM setting, we re-express X_{k+1} in equation (4.6) as

$$X_{k+1} = \kappa(\mathbf{y}_k^w) X_k + \vartheta(\mathbf{y}_k^w) + \varrho(\mathbf{y}_k^w) z_{k+1} + \tau(\mathbf{y}_k^w), \quad (4.10)$$

where

$$\kappa(\mathbf{y}_k^w) = e^{-\alpha(\mathbf{y}_k^w)\Delta t_{k+1}}, \quad (4.11)$$

$$\vartheta(\mathbf{y}_k^w) = (1 - e^{-\alpha(\mathbf{y}_k^w)\Delta t_{k+1}}) \theta(\mathbf{y}_k^w), \quad (4.12)$$

$$\varrho(\mathbf{y}_k^w) = \xi(\mathbf{y}_k^w) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{y}_k^w)\Delta t_{k+1}}}{2\alpha(\mathbf{y}_k^w)}}, \quad (4.13)$$

$$\tau(\mathbf{y}_k^w) = \sum_{h=1}^{\mathcal{P}_{\Delta t_{k+1}}} e^{-\alpha(\mathbf{y}_k^w)(\Delta t_{k+1}-v_h)} \beta_h(\mathbf{y}_k^w). \quad (4.14)$$

Akin to the change of measure is an ideal reference measure \tilde{P} , whose construction is justified by the Kolmogorov's extension theorem [3]. As noted earlier, the corresponding dynamics of J_k and \mathbf{y}_k^w are unaltered by the measure change. We may recover P from \tilde{P} given an \mathcal{F}_k^w -adapted process, for $k \geq 1$, through the Radon-Nikodym derivative

$$\Psi_k^w = \frac{dP}{d\tilde{P}} \Big|_{\mathcal{F}_k^w} = \prod_{l=1}^k \varphi_l^w \quad (4.15)$$

and

$$\begin{aligned} \varphi_l^w &= \frac{\phi \left\{ \left(\varrho^2(\mathbf{y}_{l-1}^w) \right)^{-1} \left[X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \tau(\mathbf{y}_{l-1}^w) \right] \right\}}{\left(\varrho^2(\mathbf{y}_{l-1}^w) \right) \phi(X_l)} \\ &= \exp \left(- \frac{\left(\vartheta(\mathbf{y}_{l-1}^w) + \kappa(\mathbf{y}_{l-1}^w) X_{l-1} + \tau(\mathbf{y}_{l-1}^w) \right) X_l}{\varrho^2(\mathbf{y}_{l-1}^w)} \right. \\ &\quad \left. - \frac{\left(\vartheta(\mathbf{y}_{l-1}^w) + \kappa(\mathbf{y}_{l-1}^w) X_{l-1} + \tau(\mathbf{y}_{l-1}^w) \right)^2}{2\varrho^2(\mathbf{y}_{l-1}^w)} \right), \end{aligned}$$

where ϕ is the probability density function of a standard normal random variable and $\Psi_0 = 1$, $\{\Psi_l, l \in \mathbb{Z}^+\}$ is an \mathcal{F}_l^w -adapted martingale under P .

4.3.2 Calculation of recursive filters

Under \tilde{P} where the observation process is IID, calculations and derivations needed for the filtering and parameter estimation are manageable to carry out. Suppose \mathcal{X}_k is the filtration generated by X_k , which will be used to estimate the HOHMC \mathbf{y}_k^w . The Bayes' theorem for conditional expectation is a handy tool in getting optimal estimates of various quantities under P , using the results from recursive filters under \tilde{P} .

In particular, let $\mathbf{g}_k = (g_k(1), g_k(2), \dots, g_k(cb), \dots, g_k(NN))^\top \in \mathbb{R}^{N^2}$, where $g_k(cb) := P(\mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b \mid \mathcal{X}_k) = \mathbb{E}[\langle \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle \mid \mathcal{X}_k]$. A filter for $\varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)$ under P is given by

$$\mathbf{g}_k := \mathbb{E}[\varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{X}_k] = \frac{\mathbb{E}^{\tilde{P}}[\Psi_k^w \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{X}_k]}{\mathbb{E}^{\tilde{P}}[\Psi_k^w \mid \mathcal{X}_k]}. \quad (4.16)$$

Write $\boldsymbol{\gamma}_k := \mathbb{E}^{\tilde{P}} [\Psi_k^w \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) | \mathcal{X}_k]$. Since $\sum_{c,b=1}^N \langle \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle = \langle \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{1} \rangle = 1$, where $\mathbf{1}$ is an \mathbb{R}^{N^2} vector with 1 in all of its entries, we have

$$\begin{aligned} \sum_{c,b=1}^N \langle \boldsymbol{\gamma}_k, \mathbf{e}_{cb} \rangle &= \sum_{c,b=1}^N \langle \mathbb{E}^{\tilde{P}} [\Psi_k^w \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) | \mathcal{X}_k], \mathbf{e}_{cb} \rangle \\ &= \mathbb{E}^{\tilde{P}} \left[\Psi_k^w \sum_{c,b=1}^N \langle \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle \mid \mathcal{X}_k \right] \\ &= \mathbb{E}^{\tilde{P}} [\Psi_k^w \mid \mathcal{X}_k]. \end{aligned} \quad (4.17)$$

Therefore, the filter of $\varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$ in equation (4.16) under P has an explicit representation given by

$$\mathbf{g}_k = \frac{\boldsymbol{\gamma}_k}{\sum_{c,b=1}^N \langle \boldsymbol{\gamma}_k, \mathbf{e}_{cb} \rangle} = \frac{\boldsymbol{\gamma}_k}{\langle \boldsymbol{\gamma}_k, \mathbf{1} \rangle}. \quad (4.18)$$

To derive recursive processes and estimate relevant quantities in terms of $\varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$, we need to construct two $N^2 \times N^2$ matrices by following Xi and Mamon [38]. These are \mathbf{K}_t with \mathbf{e}_{it} on its $((i-1)N+t)$ th column and 0 elsewhere for $1 \leq i, t \leq N$, and a diagonal matrix \mathbf{H}_k , where

$$\mathbf{H}_k = \begin{pmatrix} h_k^1 & 0 & & \cdots & 0 \\ 0 & \ddots & 0 & & \vdots \\ \vdots & 0 & h_k^N & \ddots & \\ & & & \ddots & \\ & & \ddots & h_k^1 & 0 & \vdots \\ \vdots & & & 0 & \ddots & 0 \\ 0 & \cdots & & \cdots & 0 & h_k^N \end{pmatrix}$$

with diagonal entries

$$h_k^i = \frac{\phi \left\{ (\varrho^2)^{-1} [X_k - \vartheta - \kappa X_{k-1} - \tau] \right\}}{(\varrho^2) \phi(X_k)}. \quad (4.19)$$

Following Xiong and Mamon [43], we define certain scalar processes of interest involving the HOHMC \mathbf{y}_k^w . These are as follows:

(i)

$$\mathcal{A}_k^{tsr} = \sum_{l=2}^k \langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle, \quad (4.20)$$

which refers to the number of jumps from state $(\mathbf{e}_r, \mathbf{e}_s)$ to state \mathbf{e}_t up to time k , where $2 \leq l \leq k$ and $r, s, t = 1, \dots, N$;

(ii)

$$\mathcal{B}_k^t = \sum_{l=2}^k \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle = \mathcal{B}_{k-1}^t + \langle \mathbf{y}_{k-1}^w, \mathbf{e}_t \rangle, \quad (4.21)$$

which gives to the occupation time up to time k or the length of time that \mathbf{y}_k^w spent in state \mathbf{e}_t , where $2 \leq l \leq k$ and $t = 1, \dots, N$;

(iii)

$$\mathcal{B}_k^{ts} = \sum_{l=2}^k \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \langle \mathbf{y}_{l-2}^w, \mathbf{e}_s \rangle, \quad (4.22)$$

which represents to the occupation time up to time k or the length of time that \mathbf{y}_k^w spent in state $(\mathbf{e}_t, \mathbf{e}_s)$, where $2 \leq l \leq k$ and $s, t = 1, \dots, N$;

(iv)

$$\begin{aligned} \mathcal{C}_k^t(f) &= \sum_{l=2}^k f(X_l) \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \\ &= \mathcal{C}_{k-1}^t(f) + f(X_k) \langle \mathbf{y}_{k-1}^w, \mathbf{e}_t \rangle, \end{aligned} \quad (4.23)$$

which is an auxiliary process related to \mathbf{y}_k^w for some function f up to time k in state \mathbf{e}_t , where $2 \leq l \leq k$, $t = 1, \dots, N$. Here, f takes the functional forms $f(X) = X$, $f(X) = (X)^2$ or $f(X) = X_{k-1}X_k$.

Consequently, the conditional expectation of $\varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$ in equation (4.18) can be written recursively as

$$\mathbf{g}_{k+1} = \mathbf{A}\mathbf{H}_{k+1}\mathbf{g}_k. \quad (4.24)$$

Suppose U_k is any \mathcal{X}_k -measurable process, denoting any of the quantities in equations (4.20)-(4.23). Write $D_k^w[U_k] := \mathbb{E}^{\tilde{P}}[\Psi_k^w U_k | \mathcal{X}_k]$ and $\widehat{U}_k := \mathbb{E}[U_k | \mathcal{X}_k]$. Here, \widehat{U}_k is regarded as the ‘best estimate’ of U_k . Similar to the steps in establishing equation (4.18), the conditional expectation of U_k given \mathcal{X}_k can be obtained using calculations that are entirely under \tilde{P} by observing that

$$\widehat{U}_k = \mathbb{E}[U_k | \mathcal{X}_k] = \frac{\mathbb{E}^{\tilde{P}}[\Psi_k^w U_k | \mathcal{X}_k]}{\mathbb{E}^{\tilde{P}}[\Psi_k^w | \mathcal{X}_k]} = \frac{D_k^w(U_k)}{\langle \boldsymbol{\gamma}_k, \mathbf{1} \rangle}. \quad (4.25)$$

Evaluating the numerator of (4.25) (cf Mamon et al. [26]) for each quantity defined in equations (4.20)–(4.23) and taking advantage of the semi-martingale representation in (4.9), recursive filtering equations are obtained that will provide self-calibrating estimates of the HOHMM parameters. This will be elaborated further in Section 4.4.

Proposition 1: The filtering recursions for the respective quantity in equations (4.20)–(4.23) are

$$D_{k+1}^w (\mathcal{A}_{k+1}^{tsr} \varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{A}\mathbf{H}_{k+1} D_k^w (\mathcal{A}_k^{tsr} \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + \langle \mathbf{g}_k, \mathbf{e}_{sr} \rangle h_{k+1}^t \langle \mathbf{A}\mathbf{e}_{sr}, \mathbf{e}_{ts} \rangle \mathbf{e}_{ts}, \quad (4.26)$$

$$D_{k+1}^w (\mathcal{B}_{k+1}^t \varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{A}\mathbf{H}_{k+1} D_k^w (\mathcal{B}_k^t \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + \mathbf{K}_t h_{k+1}^t \mathbf{A}\mathbf{g}_k, \quad (4.27)$$

$$D_{k+1}^w (\mathcal{B}_{k+1}^{ts} \varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{A}\mathbf{H}_{k+1} D_k^w (\mathcal{B}_k^{ts} \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + \langle \mathbf{g}_k, \mathbf{e}_{ts} \rangle h_{k+1}^t \mathbf{A}\mathbf{e}_{ts}, \quad (4.28)$$

and

$$D_{k+1}^w (\mathcal{C}_{k+1}^t(f) \varpi(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{A}\mathbf{H}_{k+1} D_k^w (\mathcal{C}_k^t(f) \varpi(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + f(X_{k+1}) \mathbf{K}_t h_{k+1}^t \mathbf{A}\mathbf{g}_k. \quad (4.29)$$

Proof The proofs of (4.26)–(4.29) are similar to those given in Mamon et al. [26].

4.4 Optimal parameter estimation

In this section, we derive the optimal estimates for the parameters of the stochastic process X_t through the maximum-likelihood approach. We employ the Expectation-Maximisation (EM) algorithm to deal with the somewhat involved nature of our recursive filtering relations, which aptly requires an efficient iterative approach. Refer to [7], [43], and [38], amongst others, for further details of the EM algorithm.

A key step in the EM-algorithm implementation is to characterise the probability density function (pdf) of X_t , which is needed in the maximisation of the appropriate expected

log-likelihood function given the information reflected in \mathcal{X}_t . Model parameters are governed by \mathbf{y}_t^w but suppose that the HOHMM and all model parameters remain unchanged over the infinitesimal interval $[s, t]$. Then, the regular HOHMM-modulated OU process without jumps has a normal distribution with mean $X_s e^{\alpha(t-s)} + (1 - e^{-\alpha(t-s)})\theta$ and variance $\frac{\xi^2(1 - e^{-2\alpha(t-s)})}{2\alpha}$.

The pdf of X_t can be derived completely if the distribution of the jump component J_t is also determined. As stated in Section 4.2.2, J_t relies on β and \mathcal{P} . We recall that $\beta \sim N(\mu_\beta, \sigma_\beta^2)$ and \mathcal{P} has a constant jump intensity ζ . Utilising the integral approximation $\int_s^t e^{-\alpha(t-u)} dJ_u \approx e^{-\alpha(t-u)}(J_t - J_s)$ in Erlwein et al. [7], the density of the jump component is

$$\begin{aligned} \Phi_{J_{t-s}}(x) &= \sum_{m=0}^{\infty} \frac{(\zeta(t-s))^m}{m!} e^{-\zeta(t-s)} \\ &\quad \times \phi(x; \mu_\beta e^{-\alpha(t-s)}, \sigma_\beta^2 e^{-2\alpha(t-s)}). \end{aligned} \quad (4.30)$$

By noting the stationarity of the compound Poisson process, equation (4.30) is also the density function of the increment $J_t - J_s$.

In equation (4.4), X_t is the sum of a regular OU process and a jump term. As in [17], and just focusing first on the non-switching setting, we can utilise the convolution of the OU and J_t 's densities to get the probability density of X_t given X_s , which turns out to be an expectation of a normal density. That is,

$$\begin{aligned} \Phi_{X_t|X_s}(x) &= \sum_{m=0}^{\infty} \frac{(\zeta(t-s))^m}{m!} e^{-\zeta(t-s)} \times \phi\left(x; X_s e^{\alpha(t-s)} \right. \\ &\quad \left. + (1 - e^{-\alpha(t-s)})\theta + \mu_\beta m e^{-\alpha(t-s)}, \right. \\ &\quad \left. \frac{\xi^2(1 - e^{-2\alpha(t-s)})}{2\alpha} + \sigma_\beta^2 m e^{-2\alpha(t-s)}\right) \\ &= \mathbb{E}^{\mathcal{P}_{\Delta t}} \left[\phi\left(x; X_s e^{\alpha(t-s)} + (1 - e^{-\alpha(t-s)})\theta \right. \right. \\ &\quad \left. \left. + \mu_\beta \mathcal{P}_{\Delta t} e^{-\alpha(t-s)}, \frac{\xi^2(1 - e^{-2\alpha(t-s)})}{2\alpha} \right. \right. \\ &\quad \left. \left. + \sigma_\beta^2 \mathcal{P}_{\Delta t} e^{-2\alpha(t-s)}\right) \right], \end{aligned} \quad (4.31)$$

where $\mathcal{P}_{\Delta t}$ denotes a Poisson counter.

Going back to the implementation of the EM algorithm under our HOHMM setting, let $\mathcal{P}_{\Delta t}$ be q and assume (for tractability) that \mathcal{P}_t is independent of other parameters in the model. From (4.10) and (4.31), the discrete-time process X_{k+1} , under the HOHMM framework, is set to follow a mixture of normal distributions with

$$\mu_X(\mathbf{y}_k^w) = \kappa(\mathbf{y}_k^w) X_k + \vartheta(\mathbf{y}_k^w) + \mu_{\beta(\mathbf{y}_k^w)} q \kappa(\mathbf{y}_k^w),$$

and

$$\sigma_X^2(\mathbf{y}_k^w) = \varrho^2(\mathbf{y}_k^w) + \sigma_{\beta(\mathbf{y}_k^w)}^2 q \kappa^2(\mathbf{y}_k^w).$$

We now compute the maximum likelihood estimates (MLEs) of model parameters via the EM algorithm. Define $\mathcal{E}^w = \{\kappa_t, \vartheta_t, \varrho_t, \mu_{\beta_t}, \sigma_{\beta_t}, p_{tsr}, 1 \leq t, s, r \leq N\}$ as the set of HOHMM-based parameters. Starting with the set \mathcal{E}_0^w of initial parameters, the recursive parameter updates yield an updated set $\widehat{\mathcal{E}}^w$ of MLEs, where $\widehat{\mathcal{E}}^w \in \operatorname{argmax}_{\mathcal{E}^w} L(\mathcal{E}^w)$ and $L(\mathcal{E}^w) = \mathbb{E}^{\mathcal{E}_0} \left[\frac{dP^{\mathcal{E}^w}}{dP^{\mathcal{E}_0}} \middle| \mathcal{X}_k \right]$. Following Xiong and Mamon's arguments [43], the estimation of the matrix \mathbf{T} of transition probabilities can also be facilitated by a change of measure from $P^{\mathcal{E}_0}$ to $P^{\mathcal{E}^w}$ and entries are updated automatically through the filtering processes. It should be noted that \mathbf{y}_k^w is still an HOHMC under both $P^{\mathcal{E}^w}$ and $P^{\widehat{\mathcal{E}}^w}$ but each has the corresponding transition matrices $\mathbf{T} = (p_{tsr})$ and $\widehat{\mathbf{T}} = (\widehat{p}_{tsr})$. To estimate the transition probabilities in succession, i.e., by substituting \mathbf{T} with $\widehat{\mathbf{T}}$, we utilise the likelihood function in combination with the EM algorithm and consider

$$\Lambda_k^{\mathbf{T}} = \frac{dP^{\mathcal{E}^w}}{dP^{\mathcal{E}_0}} \middle| \mathcal{X}_k = \prod_{l=2}^k \prod_{t,s,r=1}^N \left(\frac{\widehat{p}_{tsr}}{p_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle}, \quad (4.32)$$

where $\frac{\widehat{p}_{tsr}}{p_{tsr}} = 1$ for $p_{tsr} = 0$ and $\widehat{p}_{tsr} = 0$. The estimation of parameters given the observation process following equation (4.10) is accomplished by blending the EM algorithm and recursive filters in (4.26)-(4.29). The resulting outcomes for \widehat{p}_{tsr} and the rest parameters are given as follows.

Proposition 2: The EM estimates, at state t given a series of observations X_{k+1} for $k \geq 0$,

are

$$\begin{aligned}\widehat{\kappa}_t &= \frac{\widehat{\mathcal{C}}_k^t(X_k, X_{k+1}) - \vartheta_t \widehat{\mathcal{C}}_k^t(X_k) + \mu_{\beta_t} q \widehat{\mathcal{C}}_k^t(X_{k+1}) - \vartheta_t q \mu_{\beta_t} \widehat{\mathcal{B}}_k^t}{\widehat{\mathcal{C}}_k^t(X_k^2) + 2\mu_{\beta_t} q \widehat{\mathcal{C}}_k^t(X_k) + (\mu_{\beta_t} q)^2 \widehat{\mathcal{B}}_k^t} \\ &= \frac{D_k^w(\widehat{\mathcal{C}}_k^t(X_k, X_{k+1})) - \vartheta_t D_k^w(\widehat{\mathcal{C}}_k^t(X_k)) + \mu_{\beta_t} q D_k^w(\widehat{\mathcal{C}}_k^t(X_{k+1}))}{D_k^w(\widehat{\mathcal{C}}_k^t(X_k^2)) + 2\mu_{\beta_t} q D_k^w(\widehat{\mathcal{C}}_k^t(X_k)) + (\mu_{\beta_t} q)^2 D_k^w(\widehat{\mathcal{B}}_k^t)} \\ &\quad - \frac{\vartheta_t q \mu_{\beta_t} D_k^w(\widehat{\mathcal{B}}_k^t)}{D_k^w(\widehat{\mathcal{C}}_k^t(X_k^2)) + 2\mu_{\beta_t} q D_k^w(\widehat{\mathcal{C}}_k^t(X_k)) + (\mu_{\beta_t} q)^2 D_k^w(\widehat{\mathcal{B}}_k^t)},\end{aligned}\tag{4.33}$$

$$\begin{aligned}\widehat{\vartheta}_t &= \frac{\widehat{\mathcal{C}}_k^t(X_{k+1}) - \kappa_t \widehat{\mathcal{C}}_k^t(X_k) - \kappa_t \mu_{\beta_t} q \widehat{\mathcal{B}}_k^t(X_k)}{\widehat{\mathcal{B}}_k^t} \\ &= \frac{D_k^w(\widehat{\mathcal{C}}_k^t(X_{k+1})) - \kappa_t D_k^w(\widehat{\mathcal{C}}_k^t(X_k)) - \kappa_t \mu_{\beta_t} q D_k^w(\widehat{\mathcal{B}}_k^t)}{D_k^w(\widehat{\mathcal{B}}_k^t)},\end{aligned}\tag{4.34}$$

$$\begin{aligned}\widehat{\varrho}_t^2 &= \frac{\widehat{\mathcal{C}}_k^t(X_{k+1}^2) + \kappa_t^2 \widehat{\mathcal{C}}_k^t(X_k^2)}{\widehat{\mathcal{B}}_k^t} \\ &\quad + \frac{\widehat{\mathcal{B}}_k^t \left(\vartheta_t^2 + (\kappa_t \mu_{\beta_t} q)^2 + 2\vartheta_t \kappa_t \mu_{\beta_t} q - \sigma_{\beta_t}^2 \kappa_t^2 q \right)}{\widehat{\mathcal{B}}_k^t} \\ &\quad + \frac{\widehat{\mathcal{C}}_k^t(X_k) (2\mu_{\beta_t} q \kappa_t^2 + 2\vartheta_t \kappa_t)}{\widehat{\mathcal{B}}_k^t} \\ &\quad - \frac{(2\vartheta_t + 2\kappa_t \mu_{\beta_t} q) \widehat{\mathcal{C}}_k^t(X_{k+1}) + 2\kappa_t \widehat{\mathcal{C}}_k^t(X_{k+1} X_k)}{\widehat{\mathcal{B}}_k^t},\end{aligned}\tag{4.35}$$

$$\begin{aligned}\widehat{\mu}_{\beta_t} &= \frac{\widehat{\mathcal{C}}_k^t(X_{k+1}) - \vartheta_t \widehat{\mathcal{B}}_k^t(X_k) - \kappa_t \widehat{\mathcal{C}}_k^t(X_k)}{\kappa_t q \widehat{\mathcal{B}}_k^t(X_k)} \\ &= \frac{D_k^w(\widehat{\mathcal{C}}_k^t(X_{k+1})) - \vartheta_t D_k^w(\widehat{\mathcal{B}}_k^t(X_k)) - \kappa_t D_k^w(\widehat{\mathcal{C}}_k^t(X_k))}{\kappa_t q D_k^w(\widehat{\mathcal{B}}_k^t(X_k))}\end{aligned}\tag{4.36}$$

$$\begin{aligned}\widehat{\sigma}_{\beta_t}^2 &= \frac{\widehat{\mathcal{C}}_k^t(X_{k+1}^2) + \kappa_t^2 \widehat{\mathcal{C}}_k^t(X_k^2)}{\widehat{\mathcal{B}}_k^t \kappa_t^2 q} \\ &\quad + \frac{\widehat{\mathcal{B}}_k^t \left(\vartheta_t^2 + (\kappa_t \mu_{\beta_t} q)^2 + 2\vartheta_t \kappa_t \mu_{\beta_t} q - \varrho_t^2 \right)}{\widehat{\mathcal{B}}_k^t \kappa_t^2 q} \\ &\quad + \frac{\widehat{\mathcal{C}}_k^t(X_k) (2\mu_{\beta_t} q \kappa_t^2 + 2\vartheta_t \kappa_t)}{\widehat{\mathcal{B}}_k^t \kappa_t^2 q} \\ &\quad - \frac{(2\vartheta_t + 2\kappa_t \mu_{\beta_t} q) \widehat{\mathcal{C}}_k^t(X_{k+1}) + 2\kappa_t \widehat{\mathcal{C}}_k^t(X_{k+1} X_k)}{\widehat{\mathcal{B}}_k^t \kappa_t^2 q},\end{aligned}\tag{4.37}$$

$$\widehat{p}_{tsr} = \frac{\widehat{\mathcal{A}}_k^{tsr}}{\widehat{\mathcal{B}}_k^{sr}} = \frac{D_k^w(\mathcal{A}_k^{tsr})}{D_k^w(\mathcal{B}_k^{sr})}, \quad \forall \text{pairs } (t, s), t \neq s.\tag{4.38}$$

Proof See Appendix D.

For ease of numerical implementation, we assume $q = 1$ in Section 4.5. The filtering recursions in Proposition 1 yield new estimates for $\kappa_t, \vartheta_t, \varrho_t^2, \mu_{\beta_t}, \sigma_{\beta_t}^2$, and p_{tsr} , $1 \leq t, s, r \leq N$ whenever a new sequence of electricity spot prices becomes available up to time k .

The development of these recursive filters thereby implies that the parameter estimates under our HOHMM setting are self-calibrating. The results in Propositions 1 and 2 constitute further research progress relative to those given in Erlwein et al. [7], and Xiong and Mamon [43] in the following respects. Firstly, self-calibrating filtering algorithms for electricity spot prices in a regular HMM setting [7] are extended to a general HOHMM case. Secondly, we include a compound Poisson process to depict the spikes in the price dynamics; such inclusion was not a consideration item in the HOHMM setting of [43]. Thirdly, Erlwein et al. [7] went through the route of estimating first the entire drift component before being able to compute an estimate for the mean-reverting level κ . In our case, we directly dealt with the MLE of κ by providing an explicit solution as a function of filtering recursions. The required sequence of computations in recovering the model parameters is clarified.

4.5 Numerical application

To evaluate the performance of our proposed model in conjunction with its associated filtering algorithms in section 4.4, we implement them on the daily electricity spot prices (DESP) recorded by the Alberta Electric System Operator (AESO). The AESO data set contains 1461 observations covering a 4-year period from 01 Jan 2011 to 31 Dec 2014. In the context of this data set, let us first consider the seasonality and stochastic components as underscored in the DESP model in accordance with equation (4.1).

4.5.1 Analysis of the deterministic component

Alberta launched the first wholesale electricity market in Canada. It prescribes that if the wholesale electrical energy generated in the province is not consumed on site, it must flow through the Power Pool operated by the AESO [27]. The AESO, on behalf of Albertans, runs a fair and openly competitive electricity market and offers a reliable and economic

operation of the Alberta Interconnected Electric System (AIES) [5].

From Figure 4.1, we observe prices displaying strong mean reversion, high periodicities, and numerous spikes. Despite variations and jumps, DESP exhibit a long-run mean-reverting level and cyclical patterns. For instance, there are more extreme values and changes in temperature values during the winter and summer seasons, whilst off-peak periods exist in springs and autumns. The seasonality aspect mainly originates from market supply and demand, which depend on electricity generation capacity, human activities, and weather conditions. To execute the recursive filtering equations on the deseasonalised part of the model, we perform data fitting on the component D_t using the statistical software R's built-in regression functions; this then removes the discernible seasonal pattern following equation (4.1). A 'step' function is further applied in selecting the suitable number of explanatory variables in terms of the adjusted- R^2 and the Akaike information criterion (AIC). The descriptive statistics, presented in Table 4.1, guide the parameters' initialisation in the implementation procedure. Table 4.2 summarises the estimated parameters for D_t .

Mean	Std Deviation	Std Error	Median
67.53	89.77	2.35	32.34
Min	Max	Skewness	Kurtosis
5.61	669.89	2.98	9.74

Table 4.1: Descriptive statistics for daily electricity spot price (DESP)

Parameter	Estimate	95% confidence interval
a	-0.0181	(-0.0290, -0.0071)
b	80.7214	(71.5387, 89.9041)
c_1	-6.4621	(-12.9686, 0.0445)
c_2	8.9067	(2.4942, 15.3191)
g_2	-7.6311	(-14.0103, -1.2519)
e_1	-6.2641	(-12.6405, 0.1123)
j_1	-20.1426	(-26.5187, -13.7664)

Table 4.2: Parameter estimates for the seasonal component D_t

Figure 4.1 displays the plots of the fitted seasonality-component function and the actual DESP data. A zoom-in view of the data evolution from 01 Jan 2013 - 31 Dec 2014 is

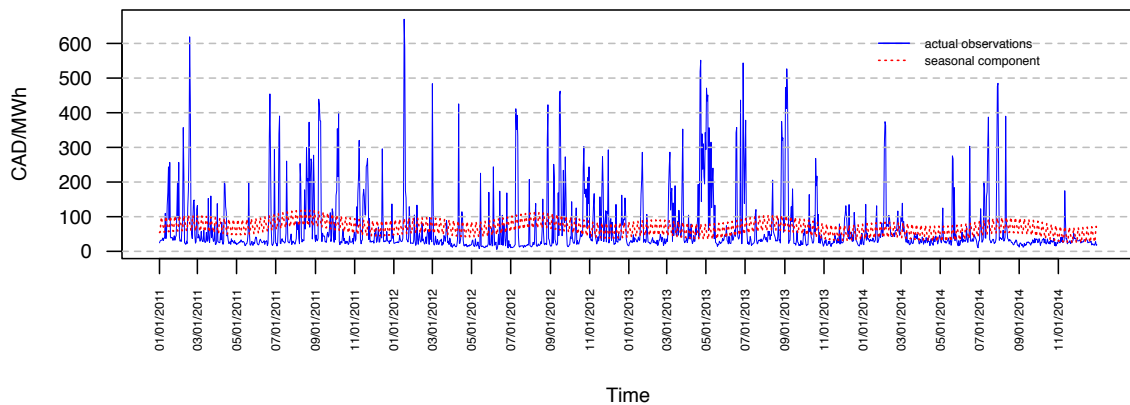


Figure 4.1: Fitted seasonal component and observed spot prices

presented in Figure 4.2. Even though the seasonal component D_t has a notably crucial role for DESP modelling, the spot-price behaviour is still hugely influenced by stochasticity; refer to Figures 4.1 and 4.2.

4.5.2 Application of filtering algorithms for the HOHMM-OU process with jumps

4.5.2.1 Data processing with the filtering algorithm

Electricity spot prices are gathered with daily frequency and so we assign $\Delta t = 1$. We process the observations in 73 batches with 20 data points in each batch. In this sense, the parameters are updated roughly every 3 weeks. Other filtering window sizes can be explored too; however, within the data set that we analyse, our experimentation produces similar outcomes, telling us that with different window sizes have negligible effect. We note that a 20-point data length for the HOHMM filtering procedure is fairly sufficient and not numerically onerous in processing the continual flow of new information including those resulting from abrupt changes in the DESP dynamics due to extreme weather, supply outages, and excess demand, amongst others. In general, practitioners in other fields applying online HOHMM-based filtering methods have the freedom to choose a data window size befitting their circumstances such as computational resources and frequency of data generation.

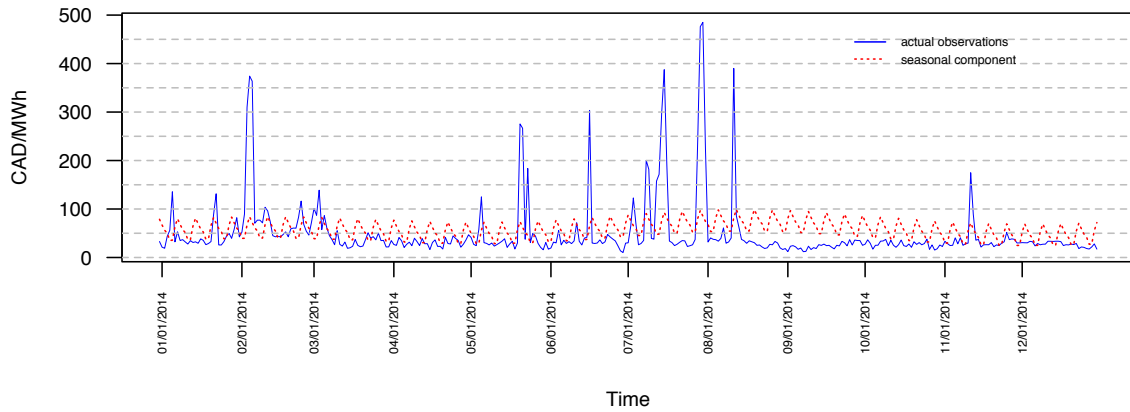


Figure 4.2: Zoom-in view of the fitted seasonal component versus observed prices (Jan/01/2014 - Dec/31/2014)

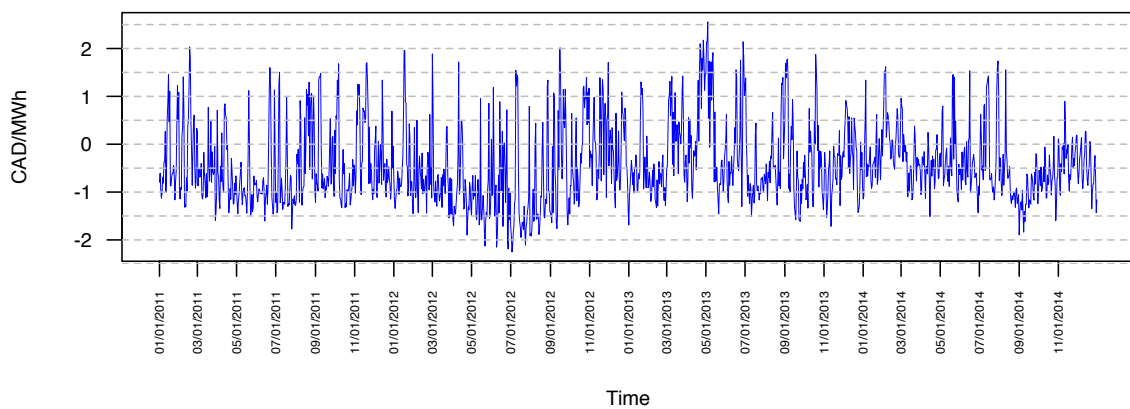


Figure 4.3: Deseasonalised stochastic component of DESP

HOHMM filtering calls for the verification of memory presence, and determination of initial values in connection with parameter estimation. To detect memory in our data, we evaluate the fractional-differencing parameter d in Granger and Joyeux's autoregressive fractionally integrated moving average (ARFIMA) methodology [16]. When $0 < d < 0.5$, there is a finite long memory in the data; and $d = 0$ signifies short memory. In computing d , one may employ the Geweke-Porter-Hudak estimator, approximated MLE, and smoothed periodogram approach; estimated values for d could be easily returned by utilising the R functions 'fdGPH', 'fdML', and 'fdSperio', respectively. In our case, we employ 'fdSperio', proposed by Reisen [30], since unlike the two former algorithms, the latter algorithm has no restriction and is applicable to a non-stationary process. Our de-seasonalised data set gives $\widehat{d} = 0.20$.

We wish to get the optimal estimates for the parameter set $\mathcal{E}^w = \{\kappa, \vartheta, \varrho, \mu_\beta, \sigma_\beta, \mathbf{p}\}$; for emphasis, these parameters are driven by a discrete-time finite-state HOHMC. This goal is accomplished by initialising estimates and updating subsequently the parameter set. Benchmarks for starting values are attained by treating X_t as a single-regime process. This implies that the transition probability matrix \mathbf{T} is identity. From (4.10), the likelihood function of X_{k+1} is

$$\begin{aligned} L(X_{k+1}; \kappa, \vartheta, \varrho, \mu_\beta, \sigma_\beta) \\ = \prod_{k=1}^m \frac{\exp\left(-\frac{(X_{k+1} - \vartheta - \kappa X_k - \mu_\beta q \kappa)^2}{2(\varrho^2 + \sigma_\beta^2 q \kappa^2)}\right)}{\sqrt{2\pi(\varrho^2 + \sigma_\beta^2 q \kappa^2)}}, \end{aligned} \quad (4.39)$$

where $1 \leq m \leq 1460$ in our implementation. This is equivalent to minimising the negative of the log likelihood, i.e.

$$\sum_{k=1}^m \left(\log \sqrt{2\pi(\varrho^2 + \sigma_\beta^2 q \kappa^2)} + \frac{(X_{k+1} - \vartheta - \kappa X_k - \mu_\beta q \kappa)^2}{2(\varrho^2 + \sigma_\beta^2 q \kappa^2)} \right). \quad (4.40)$$

The R function 'optim' is applied to equation (4.40) producing $\widehat{\kappa} = 0.6394$, $\widehat{\vartheta} = -0.1515$, $\widehat{\varrho} = 0.6443$, $\widehat{\mu}_\beta = -0.4945$, and $\widehat{\sigma}_\beta = 0.0103$.

4.5.2.2 Implementing the filtering procedure

Propositions 1 and 2, in conjunction with the initial parameter estimates, obtained in Section 4.5.2.1 are put into use for the estimation of the proposed model interfacing the

HOHMM, OU and jump attributes. Our dynamic parameter estimation proceeds by first evaluating the filtering recursions (4.26)-(4.29) and then employing their numerical outcomes to provide optimal estimates through equations (4.33)-(4.38); processing completion of one batch of data values constitutes one algorithm step or pass.

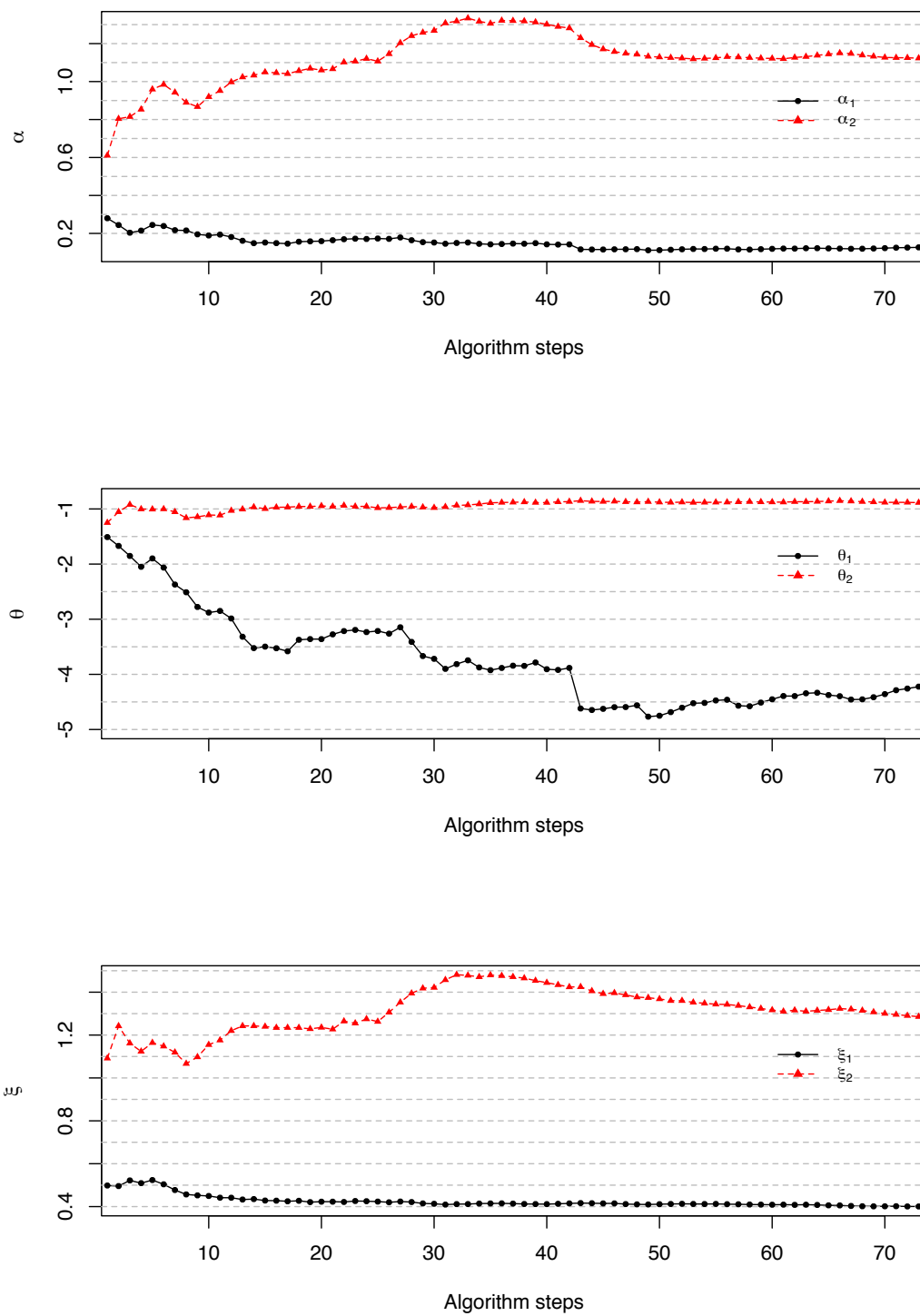
The final value in an algorithm step is used as the initial value for the filtering recursions to complete the succeeding algorithm step. Once the optimal estimates of the parameters in equation (4.10) are obtained through our self-calibrating process after certain number of algorithm steps, we can back out, via equations (4.11)-(4.13), the optimal parameter estimates of the proposed model with the original specifications in equation (4.5).

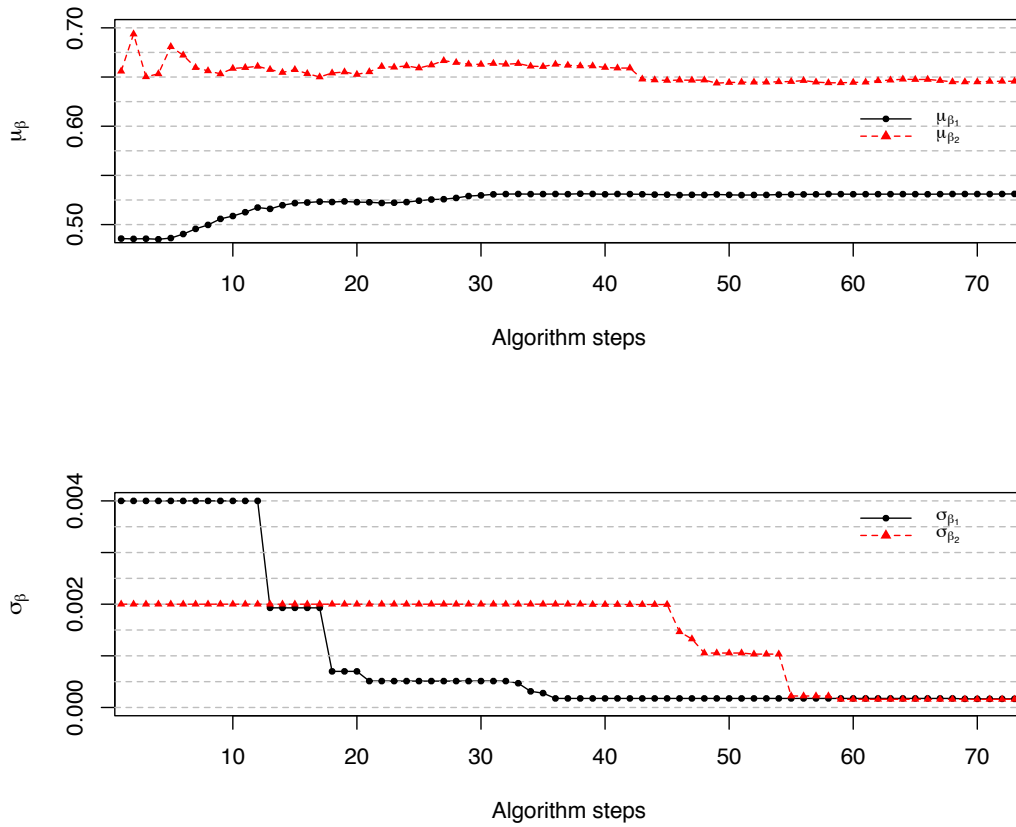
The filtering algorithms are implemented on the process X_t under the 1-,2-, and 3-state settings driven by both HOHMC and HMC. Estimated parameters evolve as shown in Figures 4.4-4.7. The parameter evolutions under the 2-state HOHMM (Figures 4.4-4.5) are different from those under the 3-state HOHMM (Figures 4.6-4.7). Nonetheless, there is gradual convergence to certain values after approximately 55 passes for each sequence of parameter estimates. Whilst the parameters' initial values might affect the convergence speed, as long as they do not deviate substantially from the benchmark values under the 1-state setting in section 4.5.2.1, the evolution of estimates for each parameter will eventually approach a specific value. Although not plotted here, parameter estimates under the HMMs exhibit similar realisations. It is worth mentioning that with HOHMM, the movements of the parameter estimates under the 1- and 2-state HOHMM settings exhibit analogous dynamic patterns and converge to similar optimal values; whilst the evolution of parameter estimates under 3-state HOHMM setting is distinct. It indicates that a 2-state HOHMM might be adequate to capture the dynamics of DESP and reflect the economic and market information. This is further supported by the analysis of model performance in subsection 4.5.3.

We quantify the variability of the parameter estimates by considering the variance of the estimators using the Fisher information $I(\mathcal{E}^w)$. This is given by

$$I(\mathcal{E}^w) = -\mathbb{E}^{\mathcal{E}^w} \left[\frac{\partial^2}{\partial \mathcal{E}_w^2} \log L(X; \mathcal{E}^w) \right].$$

This provides a bound on the asymptotic variance of the MLEs. The ML estimator is consistent and has an asymptotically normal sampling distribution; see [36]. We utilise the limiting distribution of the MLE, $\widehat{\mathcal{E}}^w$, to get the 95% confidence interval. For a generic

Figure 4.4: Evolution of parameter estimates for α , θ , and ξ under a 2-state HOHMM

Figure 4.5: Evolution of parameter estimates for μ_β , and σ_β under a 2-state HOHMM

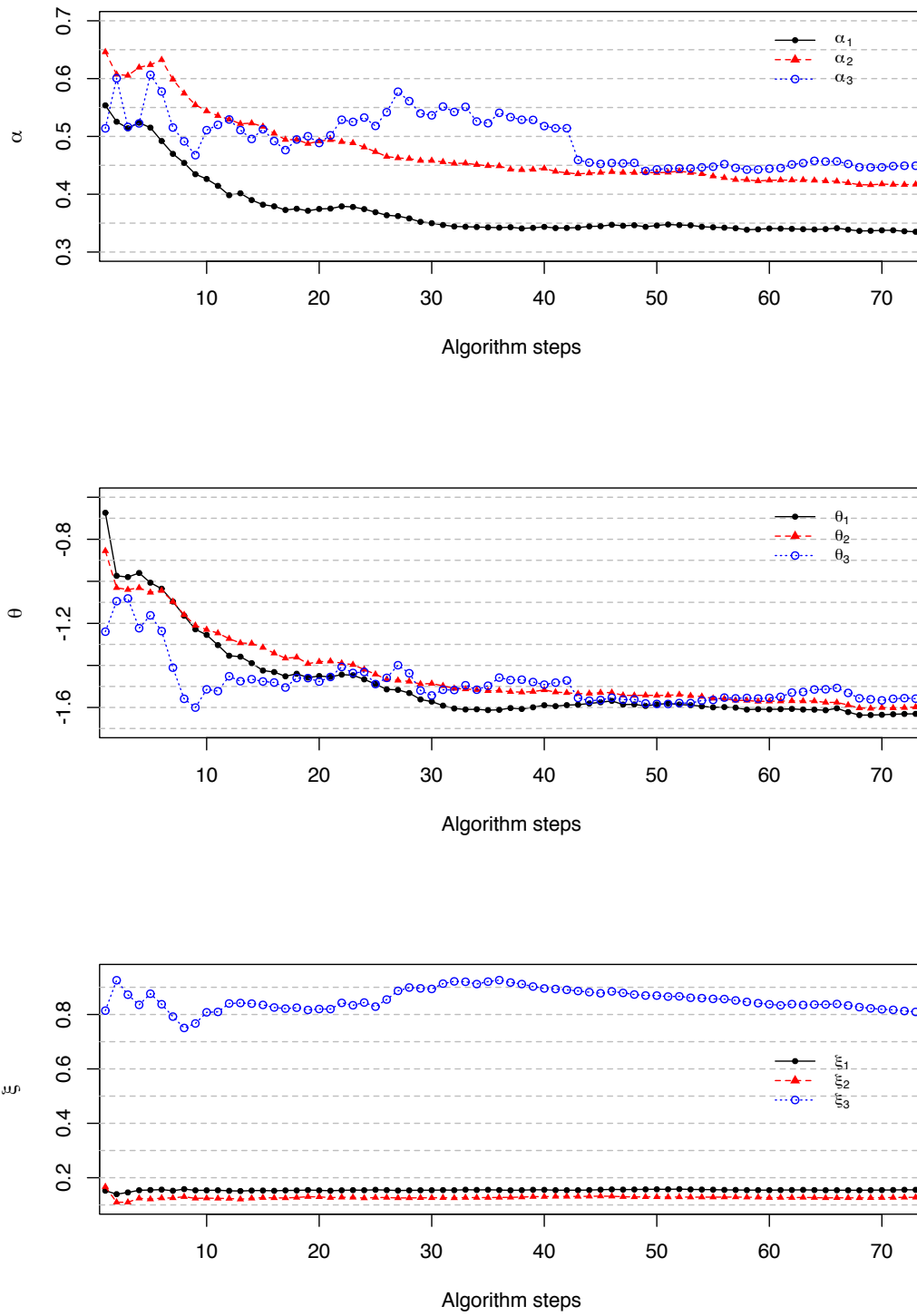
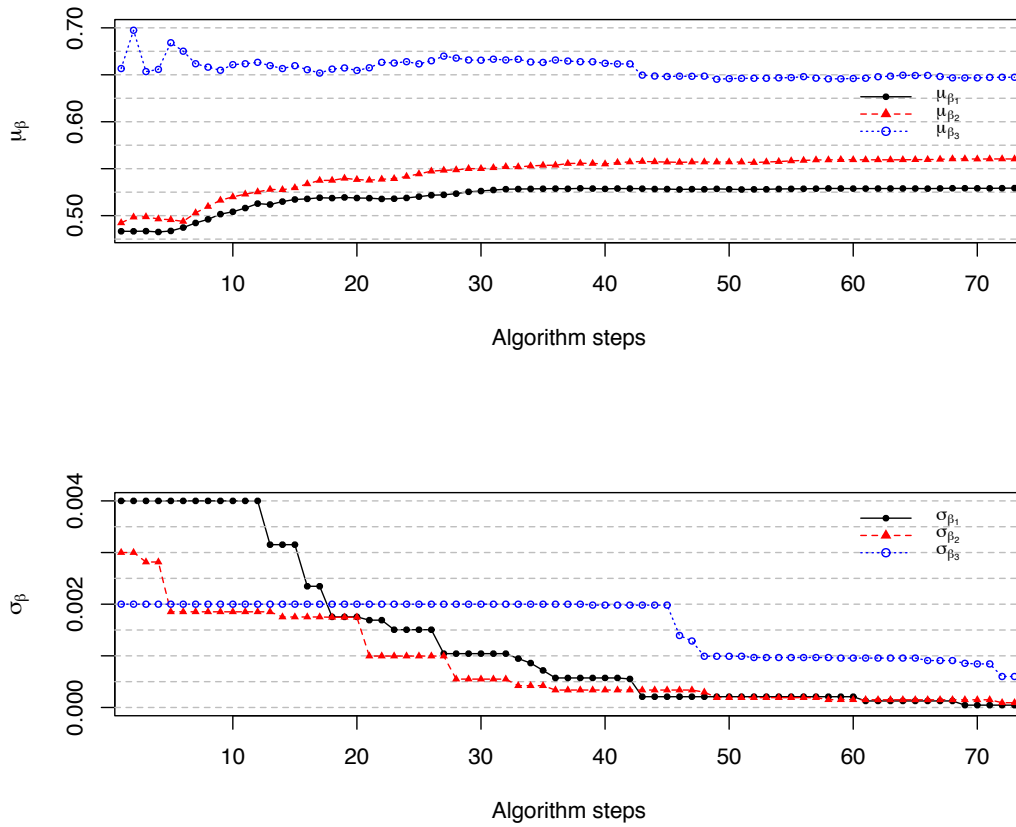


Figure 4.6: Evolution of parameter estimates for α , θ , and ξ under a 3-state HOHMM

Figure 4.7: Evolution of parameter estimates for μ_β , and σ_β under a 3-state HOHMM

(scalar) estimate ϵ^w , this is $\widehat{\epsilon}^w \pm 1.96 \frac{1}{\sqrt{I(\widehat{\epsilon}^w)}}$. Derivation of $I(\epsilon^w)$ makes use of the the log-likelihood functions, and this is delegated in the Appendix D. The results are summarise below.

$$\begin{aligned}
I(\widehat{p}_{tsr}) &= \frac{\widehat{\mathcal{A}}_k^{tsr}}{p_{tsr}^2}, & I(\widehat{\vartheta}_t) &= \frac{\widehat{\mathcal{B}}_k^t}{\varrho_t^2}, & I(\widehat{\mu}_{\beta_t}) &= \frac{\kappa_t q \widehat{\mathcal{B}}_k^t}{\varrho_t^2}, \\
I(\widehat{\kappa}_t) &= \frac{\widehat{\mathcal{C}}_k^t(X_k^2) + 2\mu_{\beta_t} q \widehat{\mathcal{C}}_k^t(X_k) + (\mu_{\beta_t} q)^2 \widehat{\mathcal{B}}_k^t}{\varrho_t^2}, \\
I(\widehat{\varrho}_t) &= -\frac{\widehat{\mathcal{B}}_k^t}{\varrho_t^2} + \frac{3}{\varrho_t^4} \left(\widehat{\mathcal{C}}_k^t(X_{k+1}^2) + \kappa_t^2 \widehat{\mathcal{C}}_k^t(X_k^2) \right. \\
&\quad + \widehat{\mathcal{B}}_k^t \left(\vartheta_t^2 + (\kappa_t \mu_{\beta_t} p)^2 + 2\vartheta_t \kappa_t \mu_{\beta_t} q - \sigma_{\beta_t}^2 \kappa_t^2 q \right) \\
&\quad + \widehat{\mathcal{C}}_k^t(X_k) (2\mu_{\beta_t} q \kappa_t^2 + 2\vartheta_t \kappa_t) \\
&\quad \left. - (2\vartheta_t + 2\kappa_t \mu_{\beta_t} p) \widehat{\mathcal{C}}_k^t(X_{k+1}) - 2\kappa_t \widehat{\mathcal{C}}_k^t(X_{k+1} X_k) \right), \\
I(\widehat{\sigma}_{\beta_t}) &= -\frac{1}{\varrho_t^2 \kappa_t^2 q} + \frac{3}{\varrho_t^4 \kappa_t^2 q} \left(\widehat{\mathcal{C}}_k^t(X_{k+1}^2) + \kappa_t^2 \widehat{\mathcal{C}}_k^t(X_k^2) \right. \\
&\quad + \widehat{\mathcal{B}}_k^t \left(\vartheta_t^2 + (\kappa_t \mu_{\beta_t} p)^2 + 2\vartheta_t \kappa_t \mu_{\beta_t} q - \sigma_{\beta_t}^2 \kappa_t^2 q \right) \\
&\quad + \widehat{\mathcal{C}}_k^t(X_k) (2\mu_{\beta_t} q \kappa_t^2 + 2\vartheta_t \kappa_t) \\
&\quad \left. - (2\vartheta_t + 2\kappa_t \mu_{\beta_t} p) \widehat{\mathcal{C}}_k^t(X_{k+1}) - 2\kappa_t \widehat{\mathcal{C}}_k^t(X_{k+1} X_k) \right).
\end{aligned}$$

Our recursive filtering is an adaptive approach and it has the mechanism for optimal estimates to be iteratively produced. As the estimates obtained manifest reasonably well convergence and stability properties the SEs become smaller as we go further down the algorithm steps. Moreover, Table 4.3 evinces that we have robust parameter estimates as implied by the substantially narrow ranges of all SEs throughout the algorithm passes.

4.5.3 Discussion of model performance

4.5.3.1 Forecasting and error analysis

We shall generate and assess the one-step ahead forecasts for X_t and DESP under the HOHMM-with-jump settings. To do these, we first compute the expected value of the

HMM						
Parameter Estimates	1-state model		2-state model		3-state model	
	Bound of SE					
	Lower	Upper	Lower	Upper	Lower	Upper
$\widehat{\kappa}_i$	1.0834×10^{-6}	0.3113	2.4645×10^{-8}	0.1574	4.45092×10^{-7}	0.3113
$\widehat{\vartheta}_i$	8.8910×10^{-7}	0.2568	2.0532×10^{-8}	0.2127	3.2596×10^{-7}	0.2568
$\widehat{\varrho}_i$	9.4071×10^{-8}	0.0254	3.1251×10^{-9}	0.1377	4.1097×10^{-8}	0.0254
$\widehat{\mu}_{\beta_i}$	1.0480×10^{-6}	0.3162	2.2175×10^{-9}	0.4088	4.0512×10^{-7}	0.1253
$\widehat{\sigma}_{\beta_i}$	6.7328×10^{-8}	0.0171	3.1251×10^{-9}	0.0346	2.6396×10^{-8}	0.0170
\widehat{p}_{ji}	6.3839×10^{-7}	0.2063	3.2075×10^{-8}	0.1785	6.3839×10^{-7}	1.6160

HOHMM						
Parameter Estimates	1-state model		2-state model		3-state model	
	Bound of SE					
	Lower	Upper	Lower	Upper	Lower	Upper
$\widehat{\kappa}_t$	1.0834×10^{-6}	0.3113	7.4602×10^{-7}	0.3827	2.2600×10^{-6}	0.3858
$\widehat{\vartheta}_t$	8.8910×10^{-7}	0.2568	6.3506×10^{-7}	0.2766	1.6974×10^{-6}	0.2676
$\widehat{\varrho}_t$	9.4071×10^{-8}	0.0254	1.0576×10^{-7}	0.4424	1.3078×10^{-7}	0.0221
$\widehat{\mu}_{\beta_t}$	1.0480×10^{-6}	0.3162	8.3156×10^{-8}	0.3357	2.0243×10^{-6}	0.3265
$\widehat{\sigma}_{\beta_t}$	6.7328×10^{-8}	0.0171	7.4834×10^{-8}	0.0243	9.1683×10^{-8}	0.0150
\widehat{q}_{tsr}	6.3839×10^{-7}	0.2063	3.7525×10^{-9}	1.2956	4.3980×10^{-9}	0.8859

Table 4.3: Interval of standard errors for parameter estimates under the 1-, 2-, 3-state HMMs and HOHMMs

stochastic component in (4.10). This gives

$$\begin{aligned}
\mathbb{E}[X_{k+1} | \mathcal{X}_k] &= \mathbb{E} \left[\kappa(\mathbf{y}_k^w) X_k + \vartheta(\mathbf{y}_k^w) + \varrho(\mathbf{y}_k^w) z_{k+1} \right. \\
&\quad \left. + \sum_{m=1}^{G\Delta t_k} e^{-\alpha(\mathbf{y}_k^w)(\Delta t_k - \tau_m)} \beta_m(\mathbf{y}_k^w) \middle| \mathcal{X}_k \right] \\
&= \langle \boldsymbol{\kappa}, \widehat{\mathbf{y}}_k^w \rangle X_k + \langle \boldsymbol{\vartheta}, \widehat{\mathbf{y}}_k^w \rangle \\
&\quad + q \langle \boldsymbol{\kappa}, \widehat{\mathbf{y}}_k^w \rangle \langle \boldsymbol{\mu}_\beta, \widehat{\mathbf{y}}_k^w \rangle.
\end{aligned} \tag{4.41}$$

Equation (4.41) brings about one-step ahead forecasts under the non-regime-switching model, and 2-state and 3-state settings driven by HOHMMs. Figure 4.8 presents the graphs showing the movements of the one-step ahead predictions for X_k and DESP under the 3-state HOHMM setting. A magnified view of the predicted DESP is depicted in Figure 4.9 covering a one-year period. The forecasts follow closely the actual dynamics of DESP during ‘normal’ periods. Models with regime-switching features outperform the 1-state model during periods with jump occurrences, and HOHMMs are more accurate than the HMMs in predicting spikes in DESP.

Visually, the DESP’s dynamics and spikes are captured very reasonably by our self-calibrating algorithm and filtering processes under the regime-switching setting. But, more formally, we wish to quantify the difference, in terms of accuracy and fitting performance, between the HOHMMs and HMMs. To assess the goodness of fit, as a by product of the predictability performance, an error analysis following the criteria in Erlwein et al. [7] is undertaken. We shall rely on the following error metrics: mean-squared error (MSE), root-mean-squared error (RMSE), absolute mean error (MAE), relative absolute error (RAE), and median absolute percent error (MdAPE). They are calculated as follows:

$$\begin{aligned}
\text{MSE} &= \frac{\sum_{k=1}^m (\widehat{X}_k - X_k)^2}{m}, & \text{RMSE} &= \sqrt{\frac{\sum_{k=1}^m (\widehat{X}_k - X_k)^2}{m}}, \\
\text{MAE} &= \frac{\sum_{k=1}^m |\widehat{X}_k - X_k|}{m}, & \text{RAE} &= \frac{\sum_{k=1}^m |\widehat{X}_k - X_k|}{\sum_{k=1}^m |\widehat{X}_k - \bar{X}|}, \\
\text{and MdAPE} &= \text{Md} \left| \frac{\widehat{X}_k - X_k}{X_k} \right|,
\end{aligned}$$

where X_k denotes the actual value of the observation at time k ; \widehat{X}_k stands for its corresponding prediction; \bar{X} is the mean of all X_k ’s; and $m = 1460$, which is the total number of

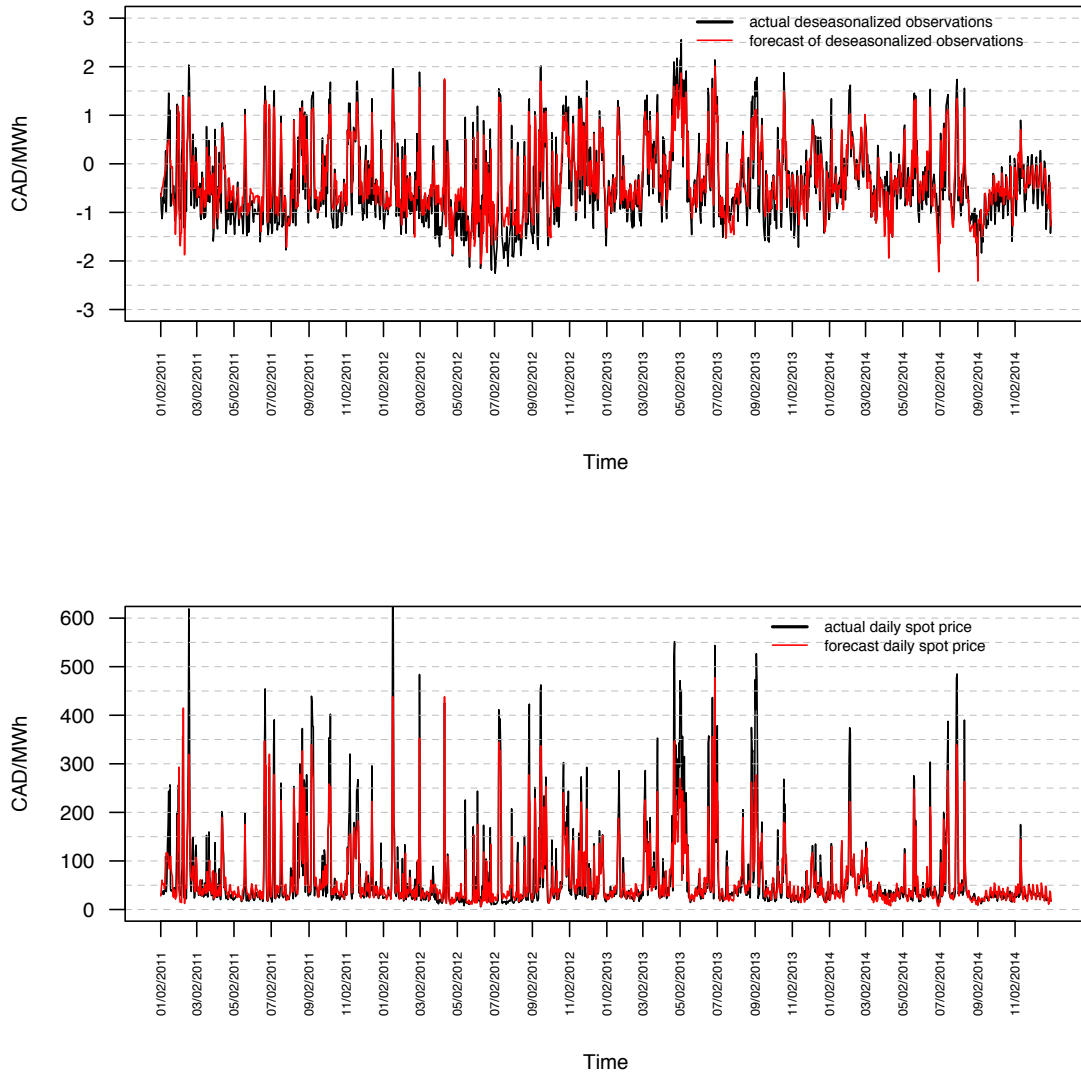


Figure 4.8: One-step ahead forecasts for X_k and S_k under a 3-state HOHMM

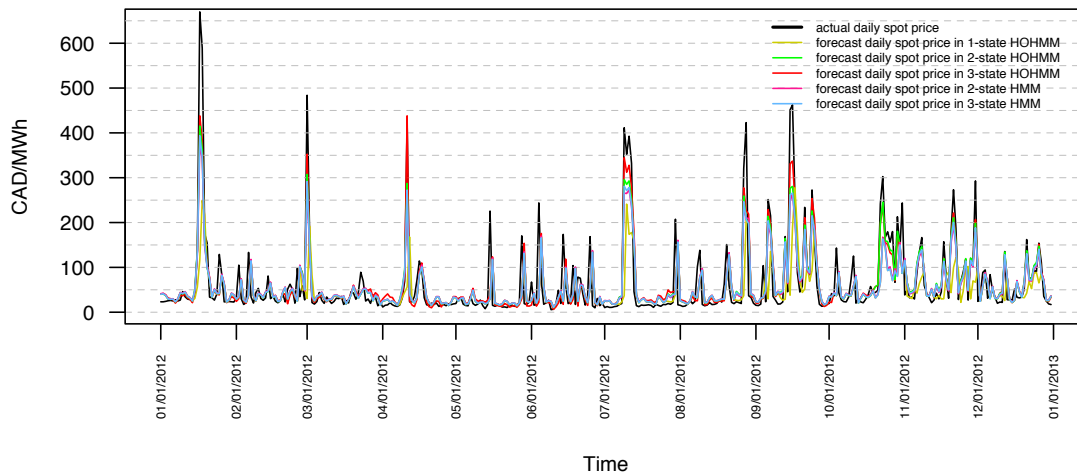


Figure 4.9: Comparison of one-step ahead forecasts in 1-, 2-, 3-state HMMs and HOHMMs

predicted values.

The results of our error analyses are presented in Table 4.4; they bespeak that a non-regime-switching model is less adequate than a regime-switching model in describing the dynamics of our data. Under the HMM setting, the 3-state framework outperforms the 1- and 2-state frameworks; whilst under the HOHMM setting, the 2-state model is the best in predicting the stochastic component and DESP. Furthermore, the HOHMM beats the regular HMM with respect to the four metrics; this finding agrees with the output revealed in Figure 4.9.

The 2-state HOHMM has the best over-all prediction amongst all state settings. The 4-state HMMs and HOHMMs were also tested but no significant improvement is achieved. Additionally, the computational cost of using HMMs and HOHMMs with state dimensions higher than 4 outweighs the benefits. Results of the t -test for the RMSEs' mean differences under various modelling set ups are displayed in Table 4.5. The adjusted p -values in our pairwise comparison were computed via the Bonferroni's method using the R function 'p.adjust'; this addresses the issue caused by familywise errors. The p -values for the comparison of 2-state versus 3-state HOHMMs are large; so, we cannot reject the null hypothesis of no RMSE-mean differences at a significance level of 5%. This agrees with Figures 4.8 and 4.9, where regimes 2 and 3 behave similarly. Thus, introducing a third state will generate minimal gain (if any). Error metrics for the 2-state and 3-state settings

HMM setting	Deseasonalised component X_t			Electricity daily spot price S_t		
	1-state	2-state	3-state	1-state	2-state	3-state
MSE	0.4152	0.2870	0.2828	5454.2	2840.2	2769.0
RMSE	0.6443	0.5357	0.5318	73.852	53.293	52.621
MdAPE	46.712	42.440	40.724	37.534	32.332	31.675
MAE	0.4828	0.4062	0.4016	36.738	27.883	27.259
RAE	0.7317	0.6156	0.6086	0.6445	0.4892	0.4782
HOHMM setting	Deseasonalised component X_t			Electricity daily spot price S_t		
	1-state	2-state	3-state	1-state	2-state	3-state
MSE	0.4152	0.2542	0.2768	5454.2	1906.9	1996.6
RMSE	0.6443	0.5042	0.5261	73.852	43.668	44.683
MdAPE	46.712	39.156	40.289	37.534	31.104	36.238
MAE	0.4828	0.3825	0.3938	36.738	24.253	24.373
RAE	0.7317	0.5797	0.5967	0.6445	0.4255	0.4276

Table 4.4: Error analysis of the HMM- and HOHMM-based models for DSP

Setting	HMM			HOHMM		
t -test	1-state	1-state	2-state	1-state	1-state	2-state
	vs	vs	vs	vs	vs	vs
	2-state	3-state	3-state	2-state	3-state	3-state
p -value	9.2771×10^{-12}	2.8685×10^{-15}	0.7771	1.0775×10^{-12}	1.3966×10^{-15}	0.8502

Table 4.5: Bonferroni-corrected p -values for the t -test performed on the RMSEs involving the DSP

in Table 4.4 show very close results. However, comparison tests demonstrate that the 1-state and 2-state settings' error metric values are statistically different, and the same can be said concerning those for the 1-state and 3-state settings. Additionally, Figure 4.9 and Table 4.4 lend support to such an outcome, which clearly suggests the merit of incorporating regime-switching features in the model for our data.

4.5.3.2 Selection of suitable model setting

We use an information selection criterion that emphasises the trade-off between bias and variance for our HMM and HOHMM settings. Similar to Erlwein et al. [7], and Xi and

Mamon [38]), we consider the Akaike Information Criterion (AIC), to provide a measure that balances the relative goodness of fit and complexity of various model settings. The AIC is given by

$$\text{AIC} = -2 \log L(X; \mathcal{E}^w) + 2l, \quad (4.42)$$

where l is the number of model parameters to be estimated, and $\log L(X; \mathcal{E}^w)$ is the log-likelihood function associated with the model. For the observation process X_{k+1} , the corresponding log-likelihood under the HOHMM setting is

$$\begin{aligned} \log L(X_{k+1}; \boldsymbol{\kappa}, \boldsymbol{\vartheta}, \boldsymbol{\varrho}, \boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta^2) &= \sum_{k=1}^B \sum_{t=1}^N \langle \mathbf{y}_k^w, \mathbf{e}_t \rangle \times \\ &\left(\log \left(2\pi \left(\varrho^2(\mathbf{y}_k^w) + \sigma_\beta^2 q \kappa^2(\mathbf{y}_k^w) \right)^{-\frac{1}{2}} \right) \right. \\ &\left. - \frac{\left(X_{k+1} - \vartheta(\mathbf{y}_k^w) - \kappa(\mathbf{y}_k^w) X_k - \boldsymbol{\mu}_\beta(\mathbf{y}_k^w) q \kappa(\mathbf{y}_k^w) \right)^2}{2 \left(\varrho^2(\mathbf{y}_k^w) + \sigma_\beta^2(\mathbf{y}_k^w) q \kappa^2(\mathbf{y}_k^w) \right)} \right), \end{aligned} \quad (4.43)$$

where B is the number of observed values and N is the number of states. The number of parameters to be estimated depends upon N , and this is summarised in Table 4.6. As expected, there is a significant increase in the number of parameters to be estimated as the number of states grows in a model.

Given the form of the AIC in equation (4.42), the selection principle hinges on minimising the AIC. From Table 4.7, the 1-state model has the highest AIC values. Clearly, the two-regime HOHMM-OU-jump setting outperforms all other settings given our data as model inputs.

Model	1-state	2-state	3-state	N -state
HMM	5	12	21	$N^2 + 4N$
HOHMM	5	14	33	$N^3 - N^2 + 5N$

Table 4.6: Number of estimated parameters under HMM-OU-jump and HOHMM-OU-jump settings

Model	Number of parameters	AIC
1-state model	5	3005.51
2-state HMM	12	2877.48
3-state HMM	21	2821.99
2-state HOHMM	14	2657.84
3-state HOHMM	33	2897.54

Table 4.7: Comparison of selection criteria AIC

The AIC metric lacks the asymptotic-consistency property as pointed out by Bozdogan [4]. This puts into question the AIC's optimality as the best-model criterion (cf [33]) when one is confronted by a collection of models with different dimensions and numbers of parameters. The Bayesian information criterion (BIC) is an alternative metric, which admits the data set's sample size on top of other inputs included in the AIC. As comprehensively discussed in Kuha [21], if the two criteria agree on the best-model choice, robustness in model selection is achieved. For this reason, we also adopt the BIC in evaluating the set of candidate models to choose from.

The AICs in Table 4.7 come from a static estimation. We complement this with the generation of BIC values generated by our dynamic parameter estimation method. The BIC is computed as

$$\text{BIC} = -l \log B + 2 \log L(X; \mathcal{E}). \quad (4.44)$$

By setting B as the number of observations in each algorithm pass when employing equation (4.43), a series of BIC values are obtained as the data set is processed in its entirety. Given the model choice-metric form in equation (4.44), the underlying principle of selection is to maximise the BIC function. Figure 4.10 depicts the evolution of the calculated BIC values for the 1-, 2-, and 3-state models.

For the 1-state model, we get higher BIC values for almost all of the periods spanned by the algorithm passes in the whole data set. We note, nonetheless, that for periods where jumps occur, the BIC values markedly drop under the 1-state model. The drastic decline of the BIC during DESP's spike occurrences insinuates that this criterion is unable to sustain robustness with one-state modelling. In contrast, the BIC values produced by the regime-switching models have more stable patterns. Furthermore, the 2-state HOHMM even produces BIC values higher than those from the 1-state HOHMM/HMM for algorithm steps

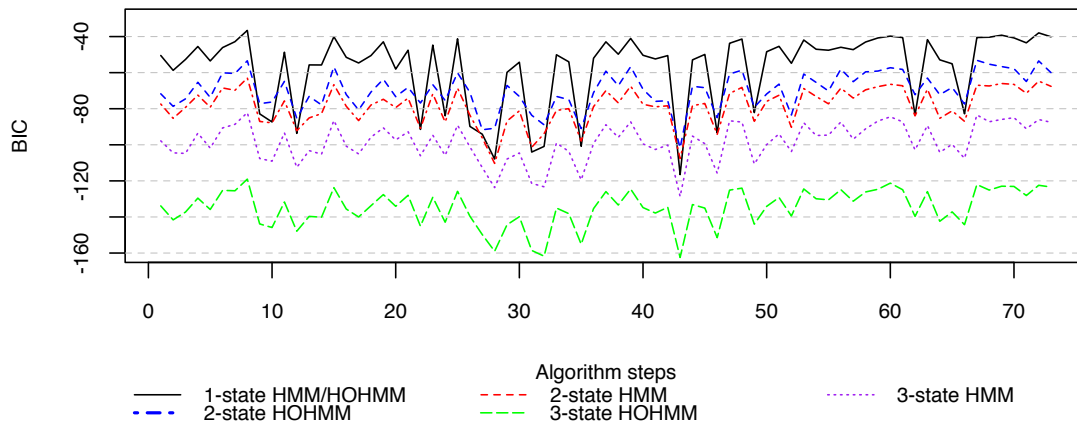


Figure 4.10: Evolution of BIC values under the 1-, 2-, 3-state HMM- and HOHMM-based models

that encompass periods of jump events. This tells us that the 1-state model is not able to capture the market well in situations where jumps could be present. On the other hand, the 2-state HOHMM can describe satisfactorily the dynamics of data with jump characteristics. This is also supported by the results from our error and AIC-based analyses. Undoubtedly, it is worthy to embed regime-switching capabilities, and in the case of our data set, the 2-state HOHMM setting is regarded as the most appropriate choice.

4.5.3.3 The valuation of expected future spot at delivery

As a consequence of electricity-market deregulation, power producers and wholesale buyers turn to derivatives and recent financial-product innovations in an effort to manage their risk exposure. Futures and forward written on electricity prices are commonly used for hedging, which entails the pricing of these contracts. Forward/futures price modelling can be categorised into two strands of study in the current literature. The first strand is to directly construct forward-price curves from market data (e.g., Fleten and Lemming [13]). The second strand is to derive forward prices as the expected future spot prices at delivery (EFSP) (e.g., Benth et al. [2]).

In this research, we first identify the best-fitting model for our DESP data. Then, optimal estimates using our filtering-based calibration are fed into the chosen model for EFSP.

Regime	$\widehat{\kappa}$	$\widehat{\vartheta}$	$\widehat{\varrho}$	$\widehat{\mu}_\beta$	$\widehat{\sigma}_\beta$	$\widehat{\alpha}$	$\widehat{\theta}$	$\widehat{\xi}$
State 1	0.7169	-0.4722	0.1281	0.5312	1.6664×10^{-4}	0.3328	-1.6679	0.1499
State 2	0.6388	-0.5643	0.6473	0.6457	1.5808×10^{-4}	0.4482	-1.5622	0.7965

Table 4.8: Optimal parameter estimates for the 2-state HOHMM

Let $G(t, T)$ denote the expected value of the spot price for delivery at future time T , and assume that the current time is t . With EFSP as the underlying variable, the theoretical forward price $F(t, T)$ of an electricity contract is computed as

$$F(t, T) = \lambda(t, T) + G(t, T), \quad (4.45)$$

where $\lambda(t, T)$ is the price of the risk process. In our OU process with jumps, $\lambda(t, T)$ has two contributing parts: (i) the price of risk due to switching of regimes and (ii) the premium due to the market price of risk. Under the real-world probability measure P , the EFSP as a function of HOHMMs is calculated as

$$\begin{aligned}
G(t, T) &= \mathbb{E}[S_T | \mathcal{X}_t] \\
&= D(T) \exp \left(X_t e^{-\alpha(\mathbf{y}_t^w)(T-t)} + \left(1 - e^{-\alpha(\mathbf{y}_t^w)(T-t)}\right) \theta(\mathbf{y}_t^w) \right. \\
&\quad \left. + \frac{\xi^2(\mathbf{y}_t^w) \left(1 - e^{-2\alpha(\mathbf{y}_t^w)(T-t)}\right)}{2\alpha(\mathbf{y}_t^w)} \right) \\
&\quad \times \mathbb{E} \left[\exp \left(\int_s^T e^{-\alpha(\mathbf{y}_t^w)(T-u)} dJ_u \mid \mathcal{X}_t \right) \right] \\
&= D(T) \exp \left(X_t e^{-\alpha(\mathbf{y}_t^w)(T-t)} + \left(1 - e^{-\alpha(\mathbf{y}_t^w)(T-t)}\right) \theta(\mathbf{y}_t^w) \right. \\
&\quad \left. + \frac{\xi^2(\mathbf{y}_t^w) \left(1 - e^{-2\alpha(\mathbf{y}_t^w)(T-t)}\right)}{2\alpha(\mathbf{y}_t^w)} \right) \\
&\quad \times \exp \left(\int_s^T \exp \left(\mu_\beta(\mathbf{y}_t^w) e^{-\alpha(\mathbf{y}_t^w)(T-u)} + \sigma_\beta^2(\mathbf{y}_t^w) e^{-2\alpha(\mathbf{y}_t^w)(T-t)} \right) \varsigma du \right. \\
&\quad \left. - \varsigma(T-t) \right). \quad (4.46)
\end{aligned}$$

Derivation details of (4.46) are similar to those given in Benth et al. [2]. The optimal parameter estimates in the evaluation of the EFSP in equation (4.46) are shown in Table 4.8.

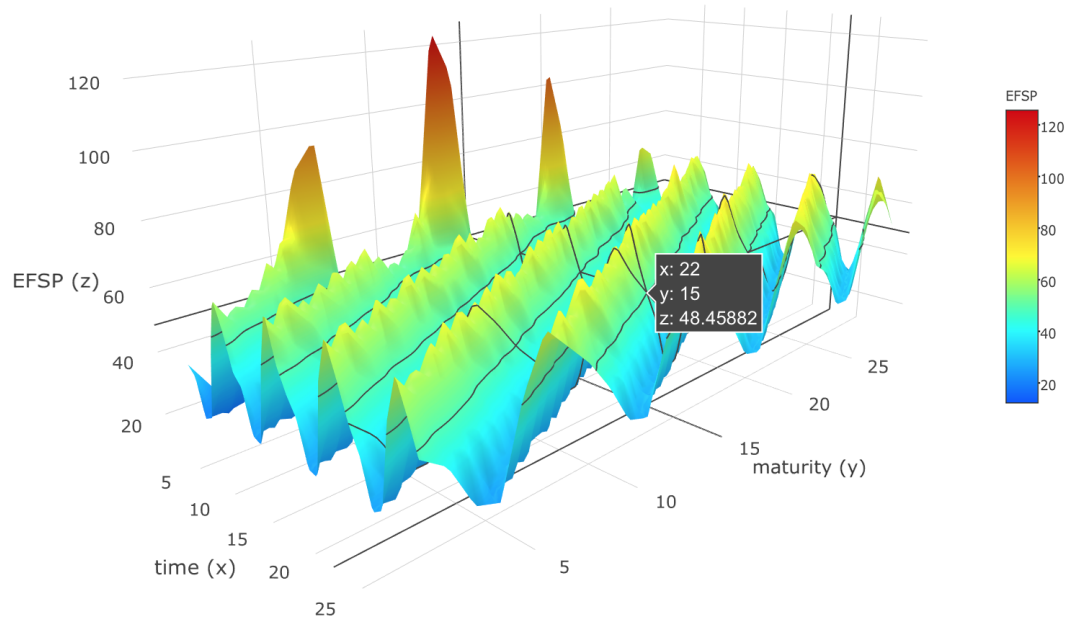


Figure 4.11: Expected future spot price on delivery (EFSP) under a 2-state HOHMM

It is reasonable to assume a constant Markov chain when pricing EFSP over a short-term delivery period. For a longer maturity of $T > 30$ days, EFSP has to be numerically computed with a dynamic Markov chain involved, as recommended in Erlwein et al. [7].

The reasonableness of the “constant Markov-chain” assumption for contracts with short delivery horizon remains disputable. In our implementation, we simulate dynamic states through an HOHMC to obtain estimated values of the EFSP for a general delivery period. Figure 4.11 shows the simulated EFSP with time t mapping the last 30 days of our dataset for maturities $T = 1, \dots, 30$ days. Salient features of the data are replicated by our proposed model such as the short-term fluctuations, seasonal patterns, and spikes of electricity prices. Our pricing application is of practical significance, as forward contract valuation is immediate once $\lambda(t, T)$'s estimate is available. Determining the appropriate $\lambda(t, T)$ is beyond the scope of this article. But, we certainly recognise its critical impact on the pricing of electricity contracts. This warrants a separate investigation that could utilise our current results and exposition serving the necessary groundwork.

4.6 Conclusion

A new framework, possessing regime-switching and data memory-capturing mechanisms, is proposed for the modelling and forecasting of electricity spot prices. The spot price is modelled as an exponential of an OU process, and augmented by a compound Poisson term to deal with the mean-reverting, stochastic and spiked components. The exponential part of the model is scaled by a suitable deterministic combination of trigonometric and linear functions to account for the cycles and other type of attributes of certain price determinants. The parameters in the exponential component are modulated by a higher-order hidden Markov chain in discrete time, which drives the random switching between different economic regimes.

The work of Erlwein et al. [7] appears to be the only online HMM filtering algorithms for the analysis of electricity spot prices. An improvement of their approach was accomplished in our model development. Empirical implementation based on our extended HOHMM filters confirmed that the predominant features of seasonality, mean reversion and extremely large spikes in electricity prices are all captured quite well. The various insights from this study reinforce the findings of Erlwein et al. [7]. We further showed that HOHMMs offer a much better fit than those achieved by the usual HMMs. The 2-state HOHMM outperforms other model settings for our data in accordance with an information-criterion evaluation.

This study showcased two key research contributions: (i) new filters that support self-updating parameter estimation of the HOHMM-OU-jump model thereby enriching the latest research of Xiong and Mamon [43] with the inclusion of spikes; and (ii) extension of the Erlwein et al.'s model [7] by incorporating HOHMMs for electricity price modelling geared towards derivative valuation and risk measurement. A direct and natural direction of this work is further empirical test of the modelling framework and estimation being put forward in the analysis of various contracts (more sophisticated than forwards) in the electricity market for investment and hedging. The HOHMM with lag order 2, which we fixed, is the focus of this research; certainly, the implementation of the HOHMM filtering with a higher lag order, and statistical inference for the estimation of the optimal lag are all worthwhile future research endeavours.

References

- [1] F. Benth, L. Ekeland, R. Hauge, B. Nielsen, A note on arbitrage-free pricing of forward contracts in energy markets, *Applied Mathematical Finance*, 10(4) (2003) 325–336.
- [2] F. Benth, A. Cartea, R. Kiesel, Pricing forward contracts in power markets by the certainty equivalence principle: Explaining the sign of the market risk premium, *Journal of Banking and Finance*, 32(10) (2008) 2006–2021.
- [3] R. Bwidehattacharya, E. Waymire, Kolmogorov’s extension theorem and Brownian motion, In: *A Basic Course in Probability Theory*, (eds.: S. Axler, K. Ribet), Springer, New York (2007) 129–140.
- [4] H. Bozdogan, Model selection and Akaike’s information criterion: The general theory and its analytical extensions, *Psychometrika*, 52(3) (1987) 345–370.
- [5] CCNMatthews Newswire, AESO 10-year plan identifies potential \$3.5 billion transmission investment, Toronto, Marketwired L.P. (2007) 1.
- [6] S. Deng, Pricing electricity derivatives under alternative stochastic spot price models, *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, (2000) 10. doi:10.1109/HICSS.2000.926755.
- [7] C. De Jong, The nature of power spikes: A regime-switching approach. *Studies in Nonlinear Dynamics and Econometrics* 10(3) (2006). DOI: <https://doi.org/10.2202/1558-3708.1361>.
- [8] C. Erlwein, F. Benth, R. Mamon, HMM filtering and parameter estimation of an electricity spot price model. *Energy Economics*, 32(5) (2010) 1034–1043.

- [9] R. Elliott, H. Yang, Forward and backward equations for an adjoint process, In: *Stochastic Processes*, (eds.: S. Cambanis, J. Ghosh, R. Karandikar, P. Sen), Springer, New York, NY, (1992) 61–69
- [10] R. Elliott, L. Aggoun, J. Moore, *Hidden Markov Models: Estimation and Control*, Springer, New York, 1995.
- [11] A. Escribano, J. Pena and P. Villaplana, Modelling electricity prices: International evidence. *Oxford Bulletin of Economics and Statistics*, 73(5) (2011) 622–650.
- [12] V. Fanelli, L. Maddalena, S. Musti, Modelling electricity futures prices using seasonal path-dependent volatility, *Applied Energy*, 173 (2016), 92–102.
- [13] S. Fleten, J. Lemming, Constructing forward price curves in electricity markets, *Energy Economics*, 25(5) (2003) 409–424.
- [14] H. Geman, A. Roncoroni, Understanding the fine structure of electricity prices, *Journal of Business*, 79(3) (2006) 1225–1261.
- [15] A. Gonzalez, A. Roque, J. Garcia-Gonzalez, Modeling and forecasting electricity prices with input-output hidden Markov models. *IEEE Transactions on Power Systems*, 20(1) (2005) 13–24.
- [16] C. Granger, R. Joyeux, An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1 (1980) 49–64.
- [17] F. Hanson, J. Westman, Stochastic analysis of jump-diffusions for financial log-return processes, In: *Stochastic Theory and Control. Lecture Notes in Control and Information Sciences*, (eds.: B. Pasik-Duncan), Springer, Berlin, Heidelberg, 280(2002) 169–83.
- [18] R. Huisman, R. Mahieu, Regime jumps in electricity prices. *Energy Economics*, 25(5) (2003) 425–434.
- [19] S. Islyayev, P. Date, Electricity futures price models: Calibration and forecasting. *European Journal of Operational Research*, 247(1) (2015) 144.
- [20] M. Jabłonska, H. Nampala and T. Kauranne, The multiple-mean-reversion jump-diffusion model for Nordic electricity spot prices. *Journal of Energy Markets*, 4(2) (2011) 3–25.

- [21] J. Kuha, AIC and BIC: Comparisons of Assumptions and Performance, *Sociological Methods and Research*, 33(2)(2004),188–229.
- [22] L. Lee, F. Jean, High-order hidden Markov model for piecewise linear processes and applications to speech recognition, *Journal of the Acoustical Society of America*, 140(2) (2016) EL204–EL210.
- [23] J. Lucia, E. Schwartz, Electricity prices and power derivatives: Evidence from the Nordic power exchange. *Review of Derivatives Research*, 5(1) (2002) 5–50.
- [24] R. Mamon, R. Elliott, *Hidden Markov Models in Finance*, International Series in Operations Research and Management Science, 104, Springer, New York, 2007.
- [25] R. Mamon, R. Elliott, *Hidden Markov Models in Finance: Further Developments and Applications*, International Series in Operations Research and Management Science, 209, Springer, New York, 2014.
- [26] R. Mamon, C. Erlwein, B. Gopaluni, Adaptive signal processing of asset price dynamics with predictability analysis, *Information Sciences*, 178 (1) (2008) 203–219.
- [27] Market Surveillance Administrator, Alberta wholesale electricity market, 2010 <http://albertamsa.ca/uploads/pdf/Reports/Reports/Alberta%20Wholesale%20Electricity%20Market%20Report%20092910.pdf> (accessed in June 2017).
- [28] W. Megginson, J. Netter, From state to market: A survey of empirical studies on privatization. *Journal of Economic Literature*, 39(2) (2001) 321–389.
- [29] R. Merton, Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1) (1976) 125–144.
- [30] V. Reisen, Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram, *Journal of Time Series Analysis*, 15 (1994) 335–350.
- [31] E. Schwartz, The stochastic behavior of commodity prices: Implications for valuation and hedging, *Journal of Finance*, 52(3)(1997) 923–973.
- [32] E. Schwartz, J. Smith, Short-term variations and long-term dynamics in commodity prices, *Management Science*, 46(7) (2000) 893–911.

- [33] G. Schwarz, Estimating the dimension of a model, *Annals of Statistics*, 6(1978) 461–464.
- [34] J. Seifert, M. Uhrig-Homburg, Modelling jumps in electricity prices: Theory and empirical evidence. *Review of Derivatives Research*, 10(1) (2007) 59–85.
- [35] T. Siu, W. Ching, E. Fung, M. Ng, X. Li, A high-order markov-switching model for risk measurement, *Computers and Mathematics with Applications*, 58 (1) (2009) 1–10.
- [36] A. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge and New York, 1998.
- [37] O. Wu, T. Liu, B. Huang, F. Forbes, Predicting electricity pool prices using hidden Markov models, *Ifac-Papersonline*, 28(8) (2015), 343–348.
- [38] X. Xi, R. Mamon, Parameter estimation of an asset price model driven by a weak hidden markov chain, *Economic Modelling*, 28(1) (2011) 36–46.
- [39] X. Xi, R. Mamon, Yield curve modelling using a multivariate higher-order HMM, In: *State-Space Models and Applications in Economics and Finance* (eds.: Y. Zeng, and Y. Wu), *Springers Series in Statistics and Econometrics for Finance 1*, (2013) 185–203.
- [40] X. Xi, R. Mamon, Capturing the regime-switching and memory properties of interest rates, *Computational Economics*, 44(3) (2014) 307–337.
- [41] X. Xi, R. Mamon, Parameter estimation in a WHMM setting with independent and volatility components, In: *Hidden Markov Models in Finance: Volume II (Further Developments and Applications)* (eds.: R. Mamon, R. Elliott), Springer (2014) 227–240.
- [42] X. Xi, R. Mamon, M. Davison, A higher-order hidden Markov chain-modulated model for asset allocation, *Journal of Mathematical Modelling and Algorithms in Operations Research*, 13(1) (2014) 59–85.
- [43] H. Xiong, R. Mamon, A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics, *Journal of Computational Science*, 17 (2016) 47–61.
- [44] H. Xiong, R. Mamon, Putting a price tag on temperature. *Computational Management Science*, in press (2017). DOI: 10.1007/s10287-017-0291-8

- [45] L. Yousef, Derivation and Empirical Testing of Alternative Pricing Models in Alberta's Electricity Market, ProQuest Dissertations Publishing, 2002.
- [46] W. Yu, G. Sheblé, Modeling electricity markets with hidden markov model, *Electric Power Systems Research*, 76(6)(2006) 445–451.
- [47] F. Ziel, R. Steinert, S. Husmann, Efficient modeling and forecasting of electricity spot prices, *Energy Economics*, 47(2015), 98–111.

Chapter 5

Modelling and forecasting futures-prices curves in the Fish Pool market

5.1 Introduction

Commodity futures markets provide a mechanism for investors to hedge a price risk or possibly profit from price changes at the futures contract's maturity. These markets, in essence, offer social and economic benefits through the efficient allocation of resources and making some form of insurance accessible to businesses. The analysis of commodity futures typically covers the energy and agriculture sectors with the intertwined objectives of modelling price formation and risk management. The literature on futures contracts dealing with the needs of the aquaculture and fisheries industries remain rather scanty, owing to the challenges in establishing efficient markets in these two sectors.

We note that frozen shrimp was the very first commodity from the seafood industry to be traded in the Chicago Mercantile Exchange back in the 1960s. Unfortunately, it was a thin market and only lasted for 3 years; see [19]. Three decades later, the Minneapolis Grain Exchange introduced the first exchange-traded shrimp futures contract in 1994 to help market participants hedge their risk exposure to shrimp prices. Alas, the market was also terminated in 2000 due to low trading volumes and lack of interest, as per the account of Quagraine and Engle [24].

The aquaculture-based futures markets in North America were short-lived and apparently failed unluckily. In contrast, the Fish Pool ASA, located in Norway, has been successful

since its launch in 2005. Furthermore, it has a strong connection with the rapidly expanding farmed salmon industry over the course of several decades. Primarily owned by Oslo Børs ASA, the Fish Pool has become a global exchange serving a venue to mitigate price risk and an instrument to reckon in sustaining the fish and seafood markets. It operates as a regulated marketplace for trading financial derivatives with the salmon price as the underlying variable. Fish Pool is a very active exchange in the trading of salmon derivatives. As reported in [23], the volume of salmon traded in the futures exchange has increased dramatically from approximately 3,600 tonnes in 2006 to 90,445 tonnes in 2016 with a record cash value of 5.8 billion Norwegian Krone (NOK).

The Fish Pool plays a critical role in the pricing and management of fresh farmed salmon. The creation and trading of salmon forward and futures contracts bring financial and economic improvements in terms of increasing the producers' profitability, enhancing market efficiency, and facilitating risk hedging. It is thereby vital to model the dynamics of the process that underlies the value of the futures contract. In this chapter, we put forward a regime-switching model that is capable of accurately describing the joint dynamics of salmon futures prices. Our modelling is geared towards the valuation other pertinent financial derivatives on salmon, hedging against volatile salmon price swings, and optimising harvesting and investment strategies of seafood resources.

A number of models for salmon prices are deterministic simply to keep the financial-modelling framework analytically tractable; see for example, Cacho [3], and Guttormsen [15]. In our case, we deal with, using stochastic processes, the uncertainty of market prices. Although Solibakke [27] built a stochastic volatility model for the Fish Pool market, their framework only considered the time-varying dynamics of front months contracts only, rather than the entire term structure of the the futures prices. Ewald [11] derived explicit formulae for prices of fish futures and call options under the assumption that the stock level of salmon follows a stochastic logistic growth. The assumption in [11] is somewhat problematic because such governing dynamics in the pricing framework refer to wild catches of fresh salmon from an open-access fishery source and not catches via fish farming. According to the UN's Food and Agriculture Organisation (FAO) [18], the farmed-fish industry has surpassed the wild-fish sector for a few decades now in terms of production volume. Moreover, as a widely traded commodity, farmed salmon, and not wild ones, form the foundation of the Fish Pool's market structure.

Certain studies in the literature feature the construction of salmon futures pricing framework starting with the spot-price evolution; then this is followed by estimating the spot-price model using a Kalman-filtering method (e.g., Schwartz [26]). Ankamah-Yeboah et al. [23] presented the price-formation development in the salmon/aquaculture futures market via a risk-premium model, whereby spot prices lead to price discovery of futures contracts. Asche et al [1] studied the spot-forward relationship by treating futures price with maturities up to six months as an unbiased estimator of spot prices. The popular two-factor commodity model for spot prices, with a stochastic convenience yield explicitly included, is widely applied to examine salmon futures prices and aquaculture-farming considerations; see details in Ewald et al. [13] and [12]. Alternatively, research investigations would also focus directly on futures prices when performing the analysis of the Fish Pool salmon market.

Of particular interest about the Fish Pool is that the salmon's spot prices and futures prices are observed weekly and daily, respectively. Thus, spot prices missing on certain days could be recovered from futures prices collected at a daily frequency. However, if the daily futures-price modelling begins with a spot price description, generating spot prices with daily frequency (from data with weekly frequency) is necessary, and this presents an added difficulty as noted in [1]. Another Fish Pool's striking characteristic is that on every trading day, closing futures prices, corresponding to various maturities that varies from months to years, are available. But then, as we move forward in time, the remaining time to maturity of the contract also decreases; and futures prices corresponding to the remaining time to maturity are available as well. This complication, in the context of multivariate modelling, creates a heavy computational challenge when capturing the term structure of futures prices at the Fish Pool whilst ensuring that every futures price in the entire data set is processed, without overlap (i.e., taken as input only once), in the estimation of model parameters. Such a complication identified above is certainly a gap in the current literature, which we shall address in Section 5.4 of this chapter. The usual approach so far to circumvent this issue in model estimation is simply to average actual maturities; see Ankamah-Yeboah et al. [23], Ewald et al. [13] and Ewald et al. [12].

Our research is distinguished from the aforementioned works by the following contributions: (i) Instead of using the classical two-factor model, we construct a Markovian regime-switching model to capture the memory and stochasticity of salmon futures prices. (ii) This chapter is the first to pin down the modelling of the term structure of salmon futures prices

without assuming approximate maturities of the contracts in the estimation procedure. (iii) We directly investigate the dynamic behaviour of salmon futures prices, thereby avoiding the use of weekly spot prices. (iv) As an alternative to the Kalman filtering method, we put forward a self-calibrating estimation method based on a regime-switching model modulated by a higher-order Markov chain, which is capable of modelling empirical distributions of virtually all shapes.

Markov regime-switching models capture the dynamics of price evolution more accurately and provide better forecasting performance as per the findings in Erlwein et al. [10], Date et al. [5], and Xiong and Mamon [38]. In the context of a Markovian regime-switching mechanism, hidden Markov models (HMMs) have been widely utilised encompassing many applications in engineering, and the natural and social sciences; for examples of applications concentrating in the fields of economics, insurance and finance, see Mamon and Elliott [20, 21]. HMMs drive the model's regime-switching attribute via Markov chains with hidden states; although the assumption of first-order state transition dependency in HMMs does not take advantage of information from historical past that may be beneficially useful. This gives rise to the concept of higher-order hidden Markov model (HOHMM), which is a doubly stochastic process in that there is a stochastic model for the observation process and for which the parameters are modulated by a higher-order hidden Markov chain (HOHMC). An ultimate aim is to obtain the best estimate of the latent Markov chain's current or future state. In Xi and Mamon [32], it was shown that the HOHMM setting yields better forecasting performance in modelling the risky assets log returns. Xiong and Mamon [37] found that HOHMM-based models could capture better the empirical characteristics of Toronto's daily average temperature compared to the capability of HMM-based models.

Our investigation reveals that Date et al. [5] is by far the only paper that utilised the HMM-filtering algorithms to model and forecast commodity futures prices. In [5], model calibration uses heating-oil futures price data; results showed that Markov regime-switching models outperformed a one-regime model. This work can be deemed as an extension and updated version of [5], but with more advancements in various aspects of futures-price modelling described as follows: (i) Model parameters in this chapter are modulated by an HOHMM, which has a memory-capturing "configuration" that may enable the extraction of possibly useful information from past data. (ii) HOHMM filtering algorithms, under a multivariate setting, are devised for estimation. (iii) The multi-step ahead forecasts are analysed to evaluate the prediction performance of all proposed models under both the HMM and

HOHMM settings. (iv) Finally, and most importantly, we design a crafty moving-window scheme for our data-filtering algorithm passes; this new data-processing scheme covers the whole life cycle of all futures contracts under study, and ensures that no data point is missed out or entered more than once into the filtering equations.

We start with the modelling of futures prices under a one-state set up, assuming that the spot price process is lognormally distributed. A finite-state HOHMM is then embedded into the framework, regulating the stochastic switching amongst regimes. Our best estimate (in the sense of an expected value given a history of past and current information) of a regime reflects a state or level generated by the interactions of competing market forces that affect the futures prices. Filters for the HOHMMs are established by first transforming the HOHMMs into HMMs, and then usual techniques are applied to obtain optimal parameter estimates of the HOHMM-based models.

The remainder of this chapter is structured as follows. Section 5.2 presents the formulation of the multi-dimensional model, the parameters of which are driven by a discrete-time HOHMC. In Section 5.3, adaptive filters for the states of the HOHMC and relevant quantities are derived through a change of reference probability technique. We then draw up a self-calibrating parameter-estimation scheme; this takes into account the recursive filters via the Expectation-Maximisation (EM) algorithm. Numerical application of our proposed models is conducted on a data set of daily salmon future prices collected from the Fish Pool in Section 5.4. Both proposed model's goodness of fit and prediction performance are assessed. Some concluding remarks are given in Section 5.5.

5.2 Model setup

Following Ross [25] and Schwartz [26], we begin with the assumption that the commodity price evolves under a risk-neutral probability measure Q . Let P_t denote the spot price of salmon at time t , and suppose P_t behaves in accordance with the stochastic differential equation (SDE)

$$dP_t = \alpha(\theta - \log P_t)P_t dt + \xi P_t dW_t^Q, \quad (5.1)$$

where α , θ and ξ are positive constants, and W_t^Q is a standard Brownian motion under Q .

Write $X_t := \log P_t$. By Itô's lemma, the log-spot price X_t is a mean-reverting stochastic process with stochastic dynamics

$$dX_t = \alpha(\mu - X_t)dt + \xi dW_t^Q, \quad (5.2)$$

where α measures the speed of mean reversion moving to the long-run mean level of log price, $\mu = \theta - \frac{\xi^2}{2\alpha}$, and ξ denotes the volatility of the process. Employing the Brownian-motion and Itô-isometry properties, the conditional distribution of X under Q at time T , where $T > t$, is normal with respective mean and variance given by

$$\mathbb{E}^Q [X_T | \mathcal{F}_t] = X_t e^{-\alpha(T-t)} + (1 - e^{-\alpha(T-t)})\mu, \quad (5.3)$$

$$\text{and } \text{Var}^Q [X_T | \mathcal{F}_t] = \frac{\xi^2 [1 - e^{-2\alpha(T-t)}]}{2\alpha}, \quad (5.4)$$

where \mathcal{F}_t is an W_t^Q -generated filtration.

From Manoliu and Tompaidis [17], the price F_t at time t of a futures contract with maturity T is the expected price of the underlying commodity at time T under Q . That is,

$$F_t = \mathbb{E}^Q [P_T | \mathcal{F}_t] = \mathbb{E}^Q [e^{X_T} | \mathcal{F}_t].$$

Using the results from equations (5.3) and (5.4), the log-expected value of the spot price is

$$\begin{aligned} G_t &= \log F_t \\ &= \mathbb{E}^Q [X_T | \mathcal{F}_t] + \frac{1}{2} \text{Var}^Q [X_T | \mathcal{F}_t] \\ &= X_t e^{-\alpha(T-t)} + (1 - e^{-\alpha(T-t)})\mu + \frac{\xi^2 (1 - e^{-2\alpha(T-t)})}{4\alpha}. \end{aligned} \quad (5.5)$$

As noted in Schwartz [26] and Weron [31], it is reasonable to assume that the dynamics of salmon spot prices follow a mean-reverting stochastic process under an objective measure, and to introduce the price of risk λ_t . We shall be working under the objective (or real-world probability) measure P when implementing a filtering-based estimation using observed market data in Sections 5.3 and 5.4. To facilitate the estimation, a change of measure will be employed; this is independent from the concept of risk-neutral measure involved in valuation. The construction of λ_t , connecting Q and P , in pricing temperature-based

derivatives is elaborated in Xiong and Mamon [38]. Elliott et al. [8] proposed a modified version of the Esscher transform that takes into account λ_t when implementing a discrete-time regime-switching Gaussian model and its continuous-time version. Research findings in [8] and [38] form the basis in our development of a regime-switching model for salmon futures prices under the arbitrage-free assumption. In particular, the log-spot price X_t under P evolves as

$$dX_t = \alpha(\mu^P - X_t)dt + \xi dW_t, \quad (5.6)$$

where W_t is a Wiener process under P , and $\mu^P = \mu + \frac{\lambda_t \xi}{\alpha}$. By Itô's lemma, along with equations (5.5), the dynamics of (5.6) under P are

$$dG_t = \left(\lambda_t \xi e^{-\alpha(T-t)} - \frac{1}{2} \xi^2 e^{-2\alpha(T-t)} \right) dt + \xi e^{-\alpha(T-t)} dW_t. \quad (5.7)$$

Salmon futures prices are assumed available at each time t^g , $g \in \mathbb{Z}^+$ with maturities T^1, T^2, \dots, T^g . Let $F_{t^h}^h$ be the price of a futures contract at time t^h with maturity T^h for $h = 1, \dots, g$. Using the Euler method and invoking the results in Date et al. [5], the g -dimensional discretisation of equation (5.7) is

$$\begin{aligned} G_{k+1}^h = & G_k^h + \frac{1}{\alpha^h} \left(1 - e^{-\alpha^h \Delta t_{k+1}^h} \right) \xi^h e^{-\alpha^h (T_h - t_{k+1}^h)} \left(\lambda_k^h - \frac{1}{4} \xi^h e^{-\alpha^h (T_h - t_{k+1}^h)} \left(1 + e^{-\alpha^h \Delta t_{k+1}^h} \right) \right) \\ & + \xi e^{-\alpha^h (T_h - t_{k+1}^h)} \sqrt{\frac{1 - e^{-2\alpha^h \Delta t_{k+1}^h}}{2\alpha^h}} z_{k+1}^h, \end{aligned} \quad (5.8)$$

where $\Delta t_{k+1}^h = t_{k+1}^h - t_k^h$ for $k \in \mathbb{Z}_0^+$, and $\{z_{k+1}^h\}$ are sequences of independent and identically distributed (IID) $N(0, 1)$ random variables.

Remark 1: *In the succeeding discussion, all vectors are denoted by bold lowercase letters and all matrices are represented by bold capitalised letters.*

To incorporate the impact of changes in market conditions and economic regimes on futures prices, we embed an HOHMM following the formulation in Xiong and Mamon [37]; moreover, the modelling framework is extended to a multi-dimensional set up. Let \mathbf{y}_k be a discrete-time homogeneous Markov chain with a finite-state space in \mathbb{R}^N under a probability space (Ω, \mathcal{F}, P) . We associate the state space with the canonical basis of \mathbb{R}^N , which is

$\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$, where $\mathbf{e}_i := (0, \dots, 0, 1, 0, \dots, 0)^\top \in \mathbb{R}^N$ with 1 in the i th position; also, \top stands for the transpose of a vector and N is the state-space dimension. To equip the model with a regime-switching capability, the g -dimensional process in equation (5.8) will have parameters that are governed by \mathbf{y}_k so that

$$\begin{aligned} G_{k+1}^h = & G_k^h + \frac{1}{\alpha^h(\mathbf{y}_k)} \left(1 - e^{-\alpha^h(\mathbf{y}_k)\Delta t_{k+1}^h} \right) \xi^h(\mathbf{y}_k) e^{-\alpha^h(\mathbf{y}_k)(T_h - t_{k+1}^h)} \left(\lambda^h(\mathbf{y}_k) \right. \\ & \left. - \frac{1}{4} \xi^h(\mathbf{y}_k) e^{-\alpha^h(\mathbf{y}_k)(T_h - t_{k+1}^h)} \left(1 + e^{-\alpha^h(\mathbf{y}_k)\Delta t_{k+1}^h} \right) \right) \\ & + \xi^h(\mathbf{y}_k) e^{-\alpha^h(\mathbf{y}_k)(T_h - t_{k+1}^h)} \sqrt{\frac{1 - e^{-2\alpha^h(\mathbf{y}_k)\Delta t_{k+1}^h}}{2\alpha^h(\mathbf{y}_k)}} z_{k+1}^h. \end{aligned} \quad (5.9)$$

Taking advantage of the algebraic simplification due to \mathbf{y}_k 's state-space representation, we have $\alpha^h(\mathbf{y}_k) = \langle \boldsymbol{\alpha}^h, \mathbf{y}_k \rangle$, $\lambda^h(\mathbf{y}_k) = \langle \boldsymbol{\lambda}^h, \mathbf{y}_k \rangle$, and $\xi^h(\mathbf{y}_k) = \langle \boldsymbol{\xi}^h, \mathbf{y}_k \rangle$ in equation (5.9), where $\langle \cdot, \cdot \rangle$ is the inner product in \mathbb{R}^N . The IID random variables $\{z_{k+1}^h\}$ are independent of the higher-order Markov chain \mathbf{y}_k .

To facilitate the conceptual discussion of the filtering processes, we focus on a second-order Markov chain \mathbf{y}_k^w to construct and implement the multi-dimensional HOHMM filtering method. Concentrating on a second-order \mathbf{y}_k^w as a prototype on this modelling breaks down the complexity in discerning the sequence and structure of algorithms in implementing a generalised HOHMC \mathbf{y}_k with a lag order k , $k \in \mathbb{N}^+$. A second-order \mathbf{y}_k^w is a Markov chain at the current step k that depends on information available at prior two steps $k-1$ and $k-2$. As in Siu et al. [28], we define an $\mathbb{R}^{N \times N^2}$ transition probability matrix \mathbf{P} , given as

$$\mathbf{P} = \begin{pmatrix} p_{111} & p_{112} & \cdots & p_{11N} & \cdots & p_{1N1} & p_{1N2} & \cdots & p_{1NN} \\ p_{211} & p_{212} & \cdots & p_{21N} & \cdots & p_{2N1} & p_{2N2} & \cdots & p_{2NN} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ p_{N11} & \cdots & & p_{N1N} & \cdots & p_{NN1} & p_{NN2} & \cdots & p_{NNN} \end{pmatrix},$$

where $p_{dcb} := P(\mathbf{y}_{k+1}^w = \mathbf{e}_d | \mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b)$ with $k \geq 1$ and $d, c, b \in \{1, 2, \dots, N\}$. The quantity p_{dcb} is interpreted as the probability that the Markov chain will be in state d given that currently it is in state c and was in state b immediately prior to being in the present state. The observation process in equation (5.9) can be expressed as

$$G_{k+1}^h = G_k^h + \nu^h(\mathbf{y}_k^w) + \sigma^h(\mathbf{y}_k^w) z_{k+1}^h, \quad (5.10)$$

where

$$\begin{aligned} \mathbf{v}^h(\mathbf{y}_k^w) &= \frac{1}{\alpha^h(\mathbf{y}_k^w)} \left(1 - e^{-\alpha^h(\mathbf{y}_k^w)\Delta t_{k+1}^h} \right) \xi^h(\mathbf{y}_k^w) e^{-\alpha^h(\mathbf{y}_k^w)(T_h - t_{k+1}^h)} \\ &\quad \times \left(\lambda^h(\mathbf{y}_k^w) - \frac{1}{4} \xi^h(\mathbf{y}_k^w) e^{-\alpha^h(\mathbf{y}_k^w)(T_h - t_{k+1}^h)} \left(1 + e^{-\alpha^h(\mathbf{y}_k^w)\Delta t_{k+1}^h} \right) \right) = \langle \mathbf{v}^h(k), \mathbf{y}_k^w \rangle, \end{aligned} \quad (5.11)$$

where $\mathbf{v}^h(k) = (v_1^h(k), v_2^h(k), \dots, v_N^h(k))^\top \in \mathbb{R}^N$, and

$$\sigma^h(\mathbf{y}_k^w) = \xi^h(\mathbf{y}_k^w) e^{-\alpha^h(\mathbf{y}_k^w)(T_h - t_{k+1}^h)} \sqrt{\frac{1 - e^{-2\alpha^h(\mathbf{y}_k^w)\Delta t_{k+1}^h}}{2\alpha^h(\mathbf{y}_k^w)}} = \langle \sigma^h(k), \mathbf{y}_k^w \rangle, \quad (5.12)$$

where $\sigma^h(k) = (\sigma_1^h(k), \sigma_2^h(k), \dots, \sigma_N^h(k))^\top \in \mathbb{R}^N$. (5.13)

The stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_k\}, P)$ serves as the modelling background and supports the stochastic processes considered in our framework. The global filtration \mathcal{F}_k is defined as $\mathcal{F}_k := \mathcal{F}_k^w \vee \mathcal{F}_k^z$, where \mathcal{F}_k^w and \mathcal{F}_k^z are filtrations generated by \mathbf{y}_k^w and W_t (a P -Wiener process), respectively.

It is important to note that \mathbf{y}_k^w is latent rather than directly observable from the salmon futures market. In particular, the state of the HOHMM is hidden in the noisy observation process G_{k+1}^h , which is evolving under P . Following the ideas common in papers [32]-[37], a transformation that converts \mathbf{y}_k^w into a regular Markov chain is employed, after which the usual HMM-filtering estimations apply. Consider the mapping $\eta(\mathbf{e}_b, \mathbf{e}_c) := \mathbf{e}_{bc}$, for $1 \leq b, c \leq N$, where \mathbf{e}_{bc} is an \mathbb{R}^{N^2} unit vector with 1 in its $((b-1)N + c)$ th position. The semi-martingale representation of the new Markov chain $\eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$ is then given by

$$\eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) = \mathbf{B}\eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) + \boldsymbol{\epsilon}_{k+1}^w, \quad (5.14)$$

where $\{\boldsymbol{\epsilon}_{k+1}^w\}_{k \geq 1}$ is a martingale increment, and \mathbf{B} is a $\mathbb{R}^{N^2 \times N^2}$ probability transition matrix with entries $\mathbf{b}_{ji} := p_{dcb}$ for $j = (d-1)N + c$, $i = (c-1)N + b$, and $\mathbf{b}_{ji} = 0$ otherwise.

5.3 Filtering and parameter estimation

We rely on Elliott et al. [6] in adopting an approach that introduces an equivalent probability measure \tilde{P} under which the observation process is a sequence of IID random variables.

The ensuing calculations are facilitated by the IID assumption. So, filters are computed under \tilde{P} , and then they are related back to P with the justification via the Girsanov's theorem.

Remark 2: *The reference measure \tilde{P} for filtering is not the same as the equivalent risk-neutral measure Q used in pricing. The purpose of having a \tilde{P} is to circumvent the direct calculations under P , which necessitates hard semi-martingale computations.*

Under our modelling framework, the suitable Radon-Nikodým derivative of P with respect to \tilde{P} is constructed as

$$\begin{aligned}\Psi_k &:= \frac{dP}{d\tilde{P}} \Big|_{\mathcal{F}_k} = \prod_{h=1}^g \prod_{l=1}^k \varphi_l^h, \quad k \geq 1, \quad \Psi_0 = 1, \\ \varphi_l^h &= \frac{\phi\left(\left(\sigma^h(\mathbf{y}_{l-1}^w)\right)^{-1} \left(G_l^h - G_{l-1}^h - \nu^h(\mathbf{y}_{l-1}^w)\right)\right)}{\sigma^h(\mathbf{y}_{l-1}^w) \phi\left(G_l^h\right)} \\ &= \frac{1}{\sigma^h(\mathbf{y}_{l-1}^w)} \exp\left(\frac{1}{2}\left(\left(G_l^h\right)^2 - \left(\sigma^h(\mathbf{y}_{l-1}^w)\right)^{-2} \times \left(G_l^h - G_{l-1}^h - \nu^h(\mathbf{y}_{l-1}^w)\right)^2\right)\right),\end{aligned}$$

where ϕ is the probability density function of an $N(0, 1)$ random variable, and \mathcal{F}_k is the filtration generated by the observation process G_k . Every parameter driving the multivariate observation process in equation (5.10) is modulated by the same \mathbf{y}_k^w . Even though the correlation of salmon futures prices is not explicitly modelled, it is implicitly encapsulated in the underlying higher-order Markov chain regulating to G_k 's multivariate dynamics with possibly dependent variates.

The general principle of obtaining optimal estimates of pertinent quantities under P for a multi-dimensional HOHMM setting is first to construct filters, which are conditional expectations under the reference measure \tilde{P} involving functions of $\eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)$. Then, the filters under \tilde{P} will be used to recover the model's parameter estimates under P .

Write $\mathbf{q}_k = (q_k(11), \dots, q_k(cb), \dots, q_k(NN))^T \in \mathbb{R}^{N \times N}$, where

$$q_k(cb) := P(\mathbf{y}_k^w = \mathbf{e}_c, \mathbf{y}_{k-1}^w = \mathbf{e}_b \mid \mathcal{F}_k) = \mathbb{E}[\langle \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle \mid \mathcal{F}_k]. \quad (5.15)$$

By the Bayes' theorem for conditional expectations, a filter of $\eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)$ under P can be computed as

$$\mathbf{q}_k = \mathbb{E}[\eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{F}_k] = \frac{\mathbb{E}^{\tilde{P}}[\Psi_k \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \mid \mathcal{F}_k]}{\mathbb{E}^{\tilde{P}}[\Psi_k \mid \mathcal{F}_k]}. \quad (5.16)$$

Let $\mathbf{s}_k := \mathbb{E}^{\tilde{P}} [\Psi_k \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) | \mathcal{F}_k]$. Since $\sum_{c,b=1}^N \langle \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle = \langle \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{1} \rangle = 1$, where $\mathbf{1} := (1, \dots, 1)^\top \in \mathbb{R}^{N^2}$, we have

$$\begin{aligned} \sum_{c,b=1}^N \langle \mathbf{s}_k, \mathbf{e}_{cb} \rangle &= \sum_{c,b=1}^N \langle \mathbb{E}^{\tilde{P}} [\Psi_k \eta(\mathbf{y}_k, \mathbf{y}_{k-1}) | \mathcal{F}_k], \mathbf{e}_{cb} \rangle \\ &= \mathbb{E}^{\tilde{P}} \left[\Psi_k \sum_{c,b=1}^N \langle \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w), \mathbf{e}_{cb} \rangle \mid \mathcal{F}_k \right] \\ &= \mathbb{E}^{\tilde{P}} [\Psi_k \mid \mathcal{F}_k]. \end{aligned} \quad (5.17)$$

Plugging in the result from equation (5.17) into equation (5.16), together with the construction of \mathbf{s}_k , yields

$$\mathbf{q}_k = \frac{\mathbf{s}_k}{\sum_{c,b=1}^N \langle \mathbf{s}_k, \mathbf{e}_{cb} \rangle} = \frac{\mathbf{s}_k}{\langle \mathbf{s}_k, \mathbf{1} \rangle}. \quad (5.18)$$

As with Xi and Mamon [32], we also define the following relevant quantities:

$$\mathcal{A}_k^{tsr} = \sum_{l=2}^k \langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle, \quad (5.19)$$

$$\mathcal{B}_k^t = \sum_{l=2}^k \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle = \mathcal{B}_{k-1}^t + \langle \mathbf{y}_{k-1}^w, \mathbf{e}_t \rangle, \quad (5.20)$$

$$\mathcal{B}_k^{ts} = \sum_{l=2}^k \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \langle \mathbf{y}_{l-2}^w, \mathbf{e}_s \rangle, \quad (5.21)$$

$$\mathcal{C}_k^t(f(G_k^h)) = \sum_{l=2}^k f(G_l^h) \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle = \mathcal{C}_{k-1}^t(f(G_{k-1}^h)) + f(G_k^h) \langle \mathbf{y}_{k-1}^w, \mathbf{e}_t \rangle, \quad (5.22)$$

where $r, s, t = 1, \dots, N$, $2 \leq l \leq k$, $1 \leq h \leq g$, and f has the form $f(G^h) = G^h$, or $f(G^h) = (G^h)^2$. Equations (5.19), (5.20) and (5.21) refer, respectively, to the Markov chain's number of jumps from state $(\mathbf{e}_r, \mathbf{e}_s)$ to \mathbf{e}_t , amount of time spent in state \mathbf{e}_t , and occupation time in state $(\mathbf{e}_t, \mathbf{e}_s)$ up to time k . The auxiliary process $\mathcal{C}_k^t(f)$ in equation (5.22) is the level sum for the states \mathbf{e}_t .

We define two $N^2 \times N^2$ matrices \mathbf{C} and \mathbf{D} (similar to Xiong and Mamon [37]) for deriving the filters of $\eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$. Recursive filtering relations for $\zeta_{k+1}^w \left(\mathcal{A}_{k+1}^{tsr} \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)$, $\zeta_{k+1}^w \left(\mathcal{B}_{k+1}^t \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)$, $\zeta_{k+1}^w \left(\mathcal{B}_{k+1}^{ts} \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)$, and $\zeta_{k+1}^w \left(\mathcal{C}_{k+1}^t(f) \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)$ are obtained with the aid of equations (5.14) and (5.18).

Proposition 5.3.1 *Let \mathbf{C}_{k+1} be an diagonal matrix*

$$\mathbf{C}_{k+1} = \begin{pmatrix} c_{k+1}^1 & 0 & & \cdots & 0 \\ 0 & \ddots & 0 & & \vdots \\ \vdots & 0 & c_{k+1}^N & \ddots & \\ & & & \ddots & \\ & & & & c_{k+1}^1 & 0 & \vdots \\ \vdots & & & & 0 & \ddots & 0 \\ 0 & \cdots & & & \cdots & 0 & c_{k+1}^N \end{pmatrix},$$

where

$$c_{k+1}^i = \prod_{h=1}^g \frac{1}{\sigma^h(\mathbf{y}_k^w)} \exp \left(\frac{1}{2} \left((G_{k+1}^h)^2 - (\sigma_i^h(\mathbf{y}_k^w))^{-2} \right. \right. \\ \left. \left. \times (G_{k+1}^h - G_k^h - v_i^h(\mathbf{y}_k^w))^2 \right) \right),$$

for $k \geq 0$ and $1 \leq b, c \leq N$. Then

$$\mathbf{s}_{k+1} = \mathbf{B}\mathbf{C}_{k+1}\mathbf{s}_k \quad (5.23)$$

$$\zeta_{k+1}^w (\mathcal{A}_{k+1}^{ISR} \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{B}\mathbf{C}_{k+1} \zeta_k^w (\mathcal{A}_k^{ISR} \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) \\ + \langle \mathbf{s}_k, \mathbf{e}_{sr} \rangle c_{k+1}^t \langle \mathbf{B}\mathbf{e}_{sr}, \mathbf{e}_{ts} \rangle \mathbf{e}_{ts} \quad (5.24)$$

$$\zeta_{k+1}^w (\mathcal{B}_{k+1}^t \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{B}\mathbf{C}_{k+1} \zeta_k^w (\mathcal{B}_{k+1}^t \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) \\ + \mathbf{D}_t c_{k+1}^t \mathbf{B}\mathbf{s}_k \quad (5.25)$$

$$\zeta_{k+1}^w (\mathcal{B}_{k+1}^{ts} \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{B}\mathbf{C}_{k+1} \zeta_k^w (\mathcal{B}_{k+1}^{ts} \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) \\ + \langle \mathbf{s}_k, \mathbf{e}_{ts} \rangle c_{k+1}^t \mathbf{B}\mathbf{e}_{ts} \quad (5.26)$$

$$\zeta_{k+1}^w (\mathcal{C}_{k+1}^t (f^h) \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)) = \mathbf{B}\mathbf{C}_{k+1} \zeta_k^w (\mathcal{C}_{k+1}^t (f^h) \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w)) + f(G_{k+1}^h) \mathbf{D}_t c_{k+1}^t \mathbf{B}\mathbf{s}_k, \quad (5.27)$$

where \mathbf{D}_t , $1 \leq t \leq N$ is an $N^2 \times N^2$ matrix with \mathbf{e}_{it} on its $((i-1)N+t)$ th column and 0 elsewhere.

Proof The derivations are analogous to the proofs provided in Xi and Mamon [32], or Xiong and Mamon [37].

Remark 3: Expressions for our recursive filters (5.23)-(5.27) in matrix notation are more compact and 'neater' than those presented in Erlwein and Mamon [10], engendering a

more efficient evaluation using software with intensive matrix-manipulation capability. Additionally, these results are extensions of the filtering equations derived by Date et al. [5] for commodity futures prices under a regular HMM framework.

To estimate the optimal parameters, we adopt the Expectation-Maximisation (EM) algorithm for the multi-dimensional HOHMM framework. The estimates of $v^h(\mathbf{y}_k^w)$, $\sigma^h(\mathbf{y}_k^w)$, and the transition probability \mathbf{P} are functions of the recursive filters in Proposition 5.3.1.

Proposition 5.3.2 For a g -dimensional observation process G_{k+1}^h in equation (5.10), $k \geq 0$, $1 \leq h \leq g$, the EM estimates of its parameters are given by

$$\widehat{p}_{tsr} = \frac{\widehat{\mathcal{A}}_{k+1}^{tsr}}{\widehat{\mathcal{B}}_{k+1}^{sr}} = \frac{\zeta_{k+1}^w \left(\mathcal{A}_{k+1}^{tsr} \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)}{\zeta_{k+1}^w \left(\mathcal{B}_{k+1}^{sr} \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)}, \quad \forall \text{pairs } (t, s), t \neq s, \quad (5.28)$$

$$\widehat{v}_t^h = \frac{\widehat{\mathcal{C}}_{k+1}^t}{\widehat{\mathcal{B}}_{k+1}^t} = \frac{\zeta_{k+1}^w \left(\mathcal{C}_{k+1}^t (f^h) \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)}{\zeta_{k+1}^w \left(\mathcal{B}_{k+1}^t \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)}, \quad (5.29)$$

$$\widehat{\sigma}_t^h = \left(\frac{\zeta_{k+1}^w \left(\mathcal{C}_{k+1}^t \left((f^h) \right)^2 \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)}{\zeta_{k+1}^w \left(\mathcal{B}_{k+1}^t \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)} - \frac{2 f_t^h \zeta_{k+1}^w \left(\mathcal{C}_{k+1}^t (f^h) \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)}{\zeta_{k+1}^w \left(\mathcal{B}_{k+1}^t \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)} + \frac{\left(f_t^h \right)^2 \zeta_{k+1}^w \left(\mathcal{B}_{k+1}^t \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)}{\zeta_{k+1}^w \left(\mathcal{B}_{k+1}^t \eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \right)} \right)^{\frac{1}{2}}. \quad (5.30)$$

Proof See Appendix E.

Once the quantities in Proposition 5.3.1 are all determined, model parameter estimates are immediate from Proposition 5.3.2.

5.4 Numerical application

5.4.1 Fish pool exchange and data description

Our numerical implementation utilises data from the Fish Pool ASA exchange market. The Fish Pool contracts, mainly composed of futures and options, are settled financially rather than settled by physical delivery of salmon. All contracts traded at the Fish Pool ASA are settled on a monthly basis with the Fish Pool Index (FPI) as a reference price reflecting the actual spot price of fresh Atlantic salmon. FPI is a weighted average (cf [14]) of three

indices (weights given in parenthesis): Nasdaq Salmon Index (85%), Fish Pool European Buyers Index (10%), and Statistics Norway customs statistics (5%). To avoid issues concerning deliverable grades, the FPI makes use of the weighted average of the most traded weight categories: 3-4 kg, 4-5 kg, 4-6 kg with contributions of 30%, 40%, and 30%, respectively, to the averaging procedure.

Spot prices are accessible at the Fish Pool on a weekly basis for immediate delivery. On the other hand, forward prices are available on a daily basis, reflecting the expectations of market participants for future trading. The interest rate is assumed deterministic rather than stochastic because forward prices in the Fish Pool are equivalent to futures prices for numerical application. To the best of our knowledge, research studies in the literature on pricing salmon futures would transform daily futures prices into weekly or monthly frequency. This is evident from the futures price modelling formulation of Ankamah-Yeboah et al. [23], Asche et al. [1], and Ewald et al. [13], which contains weekly spot prices.

Remark 4: *Our approach differs significantly from the current methodology on salmon futures price modelling. Instead of averaging futures prices to generate a proxy for the weekly spot prices as the common practice in various papers, we directly use daily futures price observations for modelling and forecasting the dynamics of futures prices in the near or medium-term horizons.*

We customise a filtering method tailored to our proposed multivariate model; this is then implemented on data concerning futures contracts with maturities up to 6 months. These are short-term contracts that are more frequently traded than others in the salmon market; see Ankamah-Yeboah et al. [23] and Ewald et al. [13]. We consider a data set of daily log-return series of futures prices collected by the Fish Pool; this data set comprises 1515 data points. The 12 maturity dates are denoted by T^h , where $h = 1, \dots, 12$, and they are set to be the last business day in the months of January 2016, February 2016, ..., December 2016. When the filtering procedure is carried out on the multivariate data, a moving window is going over time t^h until the entire trajectory of futures price curves is exhaustively processed. Futures, with maturities from 1 to 6 months, are traded on any business date between the starting t^h and ending t^h as shown in Table 5.1.

h -dimension	Starting t^h	Ending t^h	Maturity T^h
1	03Aug2015	28Jan2016	29Jan2016
2	01Sept2015	26Feb2016	29Feb2016
3	01Oct2015	30Mar2016	31Mar2016
4	02Nov2015	28Apr2016	29Apr2016
5	01Dec2015	30May2016	31May2016
6	01Jan2016	29Jun2016	30Jun2016
7	01Feb2016	28Jul2016	29Jul2016
8	01Mar2016	30Aug2016	31Aug2016
9	01Apr2016	29Sept2016	30Sept2016
10	02May2016	28Oct2016	31Oct2016
11	01Jun2016	29Nov2016	30Nov2016
12	01Jul2016	29Dec2016	30Dec2016

Table 5.1: Illustrating the data periods of future contracts (maturities of 1–6 months) covered by the moving window in the filtering procedure

To further elucidate the relevance of the information in Table 5.1, consider futures contracts with expiry data T^1 (29 Jan 2016) and refer to Table 5.2.

Futures contract	Time to maturity	Trading date
6 – month	$T^1 - t_1^1,$	$t_1^1 : 03Aug2015$
	$T^1 - t_2^1,$	$t_2^1 : 04Aug2015$
	\vdots	\vdots
5 – month	$T^1 - t_{21}^1,$	$t_{21}^1 : 31Aug2015$
	$T^1 - t_{22}^1,$	$t_{22}^1 : 01Sept2015$
	\vdots	\vdots
1 – month	$T^1 - t_{43}^1,$	$t_{43}^1 : 30Sept2015$
	\vdots	\vdots
	$T^1 - t_{108}^1,$	$t_{108}^1 : 04Jan2016$
1 – month	\vdots	\vdots
	$T^1 - t_{126}^1,$	$t_{126}^1 : 28Jan2016$

Table 5.2: Futures contracts with maturity up to 6 months and expiration on 29 Jan 2016

If a market participant enters into a futures contract on any business day between t_1^1 (03 Aug 2015) and t_{21}^1 (31 Aug 2015), then he owns has a 6-month futures contract with maturity date T^1 ; In this case, the contract could have between 151 to 180 days until expiration. A similar reasoning can be made when owning 5-month, \dots , 1-month contracts until the expiration T^1 is reached. Our algorithm, with moving window going through time points t^h , is applied to futures contracts with maturity dates T^2, T^3, \dots, T^{12} . This proposed data-reorganisation scheme ensures that for times to maturity $T^h - t_k^h$, $1 \leq k \leq 126$ (k expressed as number of trading days), we have 12 data points corresponding to each t_k . More importantly, this scheme will create a multivariate time series without missing or causing a double entry of any data point.

Remark 5: *A novelty of this work arises from realistically treating the remaining time to maturity $T^h - t_k^h$ to be varying. This gives the benefit of using all available raw information without any additional data transformation in the process of modelling and parameter estimation.*

Table 5.3 presents the descriptive statistics of our data set. Low volatility as well as skewness and kurtosis of low magnitude are observed for the log-futures price data.

Maturity T^h	Mean	Sd Dev	Min	Max	Skew	Kurtosis
29Jan2016	3.8814	0.0868	3.7796	4.0993	0.8528	-0.2463
29Feb2016	3.8987	0.0844	3.7773	4.0535	0.1215	-1.2253
31Mar2016	3.9511	0.1012	3.8022	4.1431	0.2419	-1.0862
29Apr2016	3.9572	0.0978	3.8308	4.1026	0.2493	-1.5814
31May2016	3.9883	0.1090	3.8351	4.2047	0.1825	-1.2799
30Jun2016	3.9951	0.1227	3.8199	4.2413	0.4316	-0.9328
29Jul2016	4.0411	0.1498	3.8133	4.3470	0.5050	-0.7739
31Aug2016	4.0108	0.0727	3.8712	4.1109	-0.4151	-1.3396
30Sept2016	3.9949	0.0535	3.8918	4.0673	-0.7087	-0.5837
31Oct2016	4.0309	0.0503	3.9551	4.1636	1.2693	0.8055
30Nov2016	4.1498	0.0584	4.0518	4.287	0.6630	-0.3971
30Dec2016	4.2321	0.0705	4.1431	4.3307	0.0789	-1.7328

Table 5.3: Descriptive statistics for log-futures prices (maturities of 1 – 6 months)

5.4.2 Implementation of filters and estimation

As pointed out in Xiong and Mamon [37], the autoregressive fractionally integrated moving average (ARFIMA) method could be utilised to verify presence of memory (in the sense of current data value's dependence on previous data points). For our data set, fractional-differencing parameter u is assessed against these benchmarks: $u = 0$ implies the time series data exhibits short memory, and $0 < u < 0.5$ implies there is finite long memory in the time series data. We use the R function 'fdSperio', which is suitable for non-stationary processes, in examining the memory property of our data set. We obtained $\widehat{u} = 0.0633$ indicating short memory. This rationalises our use of the 2nd-order HOHMM set-up, which is sufficient to tackle short memory.

Coming up with the appropriate initial parameter values in the implementation of our self-calibrating estimation could make use of the least-square method suggested in Erlwein and Mamon [10] or the likelihood maximisation mentioned in Date et al. [5]. In our case, we adopt the latter method. Considering G_t as a one-state process, the maximiser of the associated log-likelihood function, given a series of observations, is given by

$$\operatorname{argmax}(\log L(G_{k+1}; \nu, \sigma)) = \prod_{k=1}^m \left(\log \frac{1}{\sqrt{2\pi\sigma}} - \frac{(G_{k+1} - G_k - \nu)^2}{2\sigma^2} \right), \quad (5.31)$$

where $m = 120$ in this numerical application. For a g -dimensional setting, the log-likelihood to be maximised must simultaneously take into account all futures prices spanning maturities T^h , $1 \leq h \leq g$, where $g = 12$. The aim is to find maximisers ν^h and σ^h for equation (5.10). Given that $\{z_{k+1}^h\}$ is a sequence of IID $N(0, 1)$, we must solve the optimisation problem

$$\operatorname{argmax}(\log L(G_{k+1}; \nu, \sigma)) = \prod_{h=1}^g \prod_{k=1}^m \left(\log \frac{1}{\sqrt{2\pi\sigma^h}} - \frac{(G_{k+1}^h - G_k^h - \nu^h)^2}{2(\sigma^h)^2} \right). \quad (5.32)$$

Invoking Date et al. [5], it is reasonable to assume that the speed of mean reversion is fixed and independent of the Markov chain when modelling the log-futures prices of a commodity. In fact, we validated that the estimates of α in our model remains relatively stable using prices of the short-term futures. A fixed α is justified and provides some simplification in the computation. We emphasise that both λ and ξ are still governed by \mathbf{y}^w , and their estimated values are required to update $\nu(\mathbf{y}^w)$ and $\sigma(\mathbf{y}^w)$ in equation (5.10). We address

problem (5.32) using both R functions ‘nlm’ and ‘optim’, generating initial parameter estimates $\widehat{\alpha} = 0.9864$, $\widehat{\lambda} = 4.2257$, and $\widehat{\xi} = 0.1855$; these estimates are used as benchmarks in selecting starting values for the multi-regime HOHMMs.

Each variate in our multivariate HOHMM has the same number of regimes and every variate’s switching is driven by the same HOHMC’s transition matrix \mathbf{P} . The time increment $\Delta t = 1$ trading day in the processing of our data, which is divided into 24 batches. Each batch, for one algorithm, has 5 vectors (i.e., $5 t_k^h$ for all T^h , $h = 1, 2, \dots, 12$), and it gives a window size covering an entire week of trading. One week of multivariate data is deemed sufficient to accumulate new information that could impact the futures market, such as substantial changes in supply and demand, climactic conditions (e.g., sea-surface temperature, water currents, etc), economic and political events, amongst other market factors. Thus, our algorithm formulation updates quantities that are functions of HOHMM every trading week through the online filters (5.23)-(5.27).

The filtered estimates are then fed into equations (5.28)-(5.30) to get new parameter estimates. The estimates from the most recent algorithm pass serve as initial parameter values for the next algorithm pass via the recursive filtering equations. We experimented on various combinations of different initial parameter values and other filtering windows. It is found that, for majority of the times during the filtering process, convergence of parameter estimates could be achieved without any iteration failures. When this is true, window sizes and starting values only impact the convergence speed, but they do not substantially produce different results.

Figures 5.1 and 5.2 depict the movement of the entries for the transition probability matrix \mathbf{P} under the 2-state and 3-state settings, respectively. The jumps in the evolution of probabilities reflect possible market-state changes and price fluctuations. Under both 2-state and 3-state settings, noticeable changes occur from the 11th to 15th algorithm pass; and after approximately 19 algorithm passes, stability of transition probability evolution is attained. Evidence of regime switching is also detected in the evolution of other parameter estimates for the process G_t in equation (5.10).

When estimates of ν and σ are generated and with the the value of α computed from solving the aforementioned optimisation in (5.32), the evolution of the estimates for λ and ξ can be obtained through equations (5.11) and (5.12). Figures 5.3 and 5.4 show the evolution of

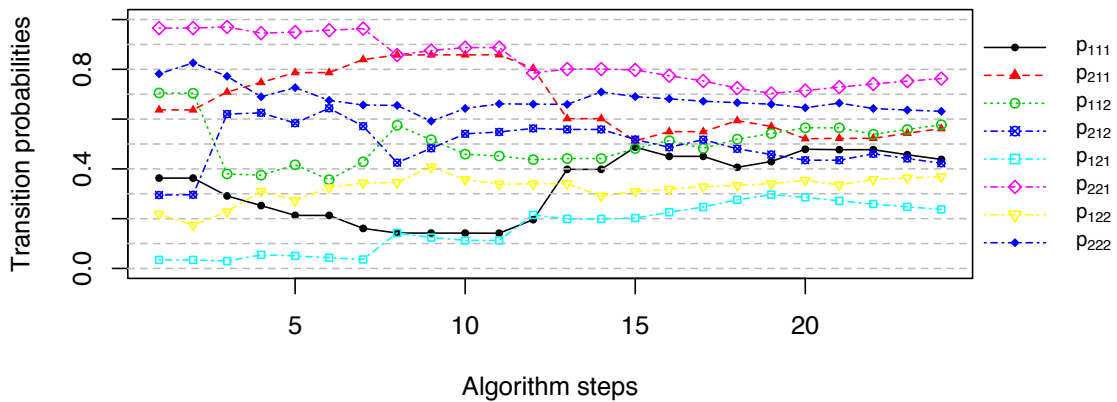


Figure 5.1: Evolution of transition probabilities under a 2-state HOHMM for futures prices with expiry T^h .

parameters, under both 2- and 3-state HOHMM set ups, using prices from contracts with expiration on the last maturity date T^{12} . The small magnitude of volatilities ξ and σ is consistent with the descriptive statistics of the data portrayed in Table 5.3. Figures 5.1 and 5.2 illustrate that the behaviour of parameter estimates is relatively stable after a period of dramatic switches. Also, the general downward pattern of parameter dynamics is a shared characteristic of Figures 5.3 and 5.4.

5.4.3 Model performance and selection

For a comprehensive evaluation of model performance, we also implement our filtering algorithm on 2- and 3-state HMM frameworks, which are two special cases of an HOHMM with lag order 1. It is a common post-diagnostic check to compare the forecasting performance of a proposed model with other benchmarked models, such as the random walk and ARCH-type models; see Hardy [16]. However, these typical benchmarks are incompatible with the HMM and HOHMM frameworks. This is because we directly model multivariate futures prices and incorporate memory property of the market into the setting, whereas random walk and ARCH models deal with univariate spot prices. We also rely on Mamon et al. [22] whereby it was found that ARCH and GARCH models are unable to beat the regular HMMs' performance with respect to capturing short- and medium-term predictability of a data series. To find the 'best-performing' model that captures the main features of the

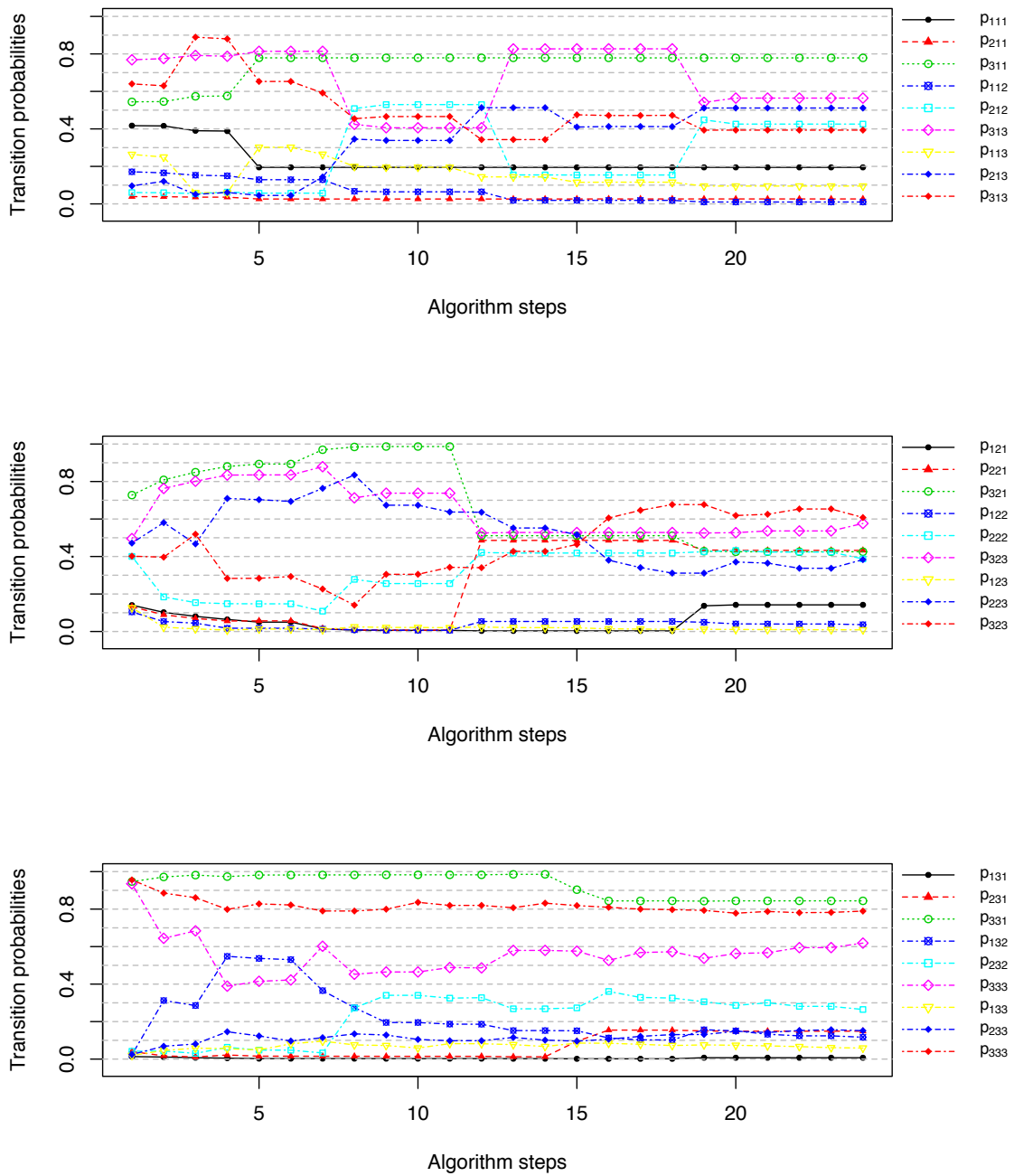


Figure 5.2: Evolution of transition probabilities under a 3-state HOHMM for prices of futures with expiry T^h .

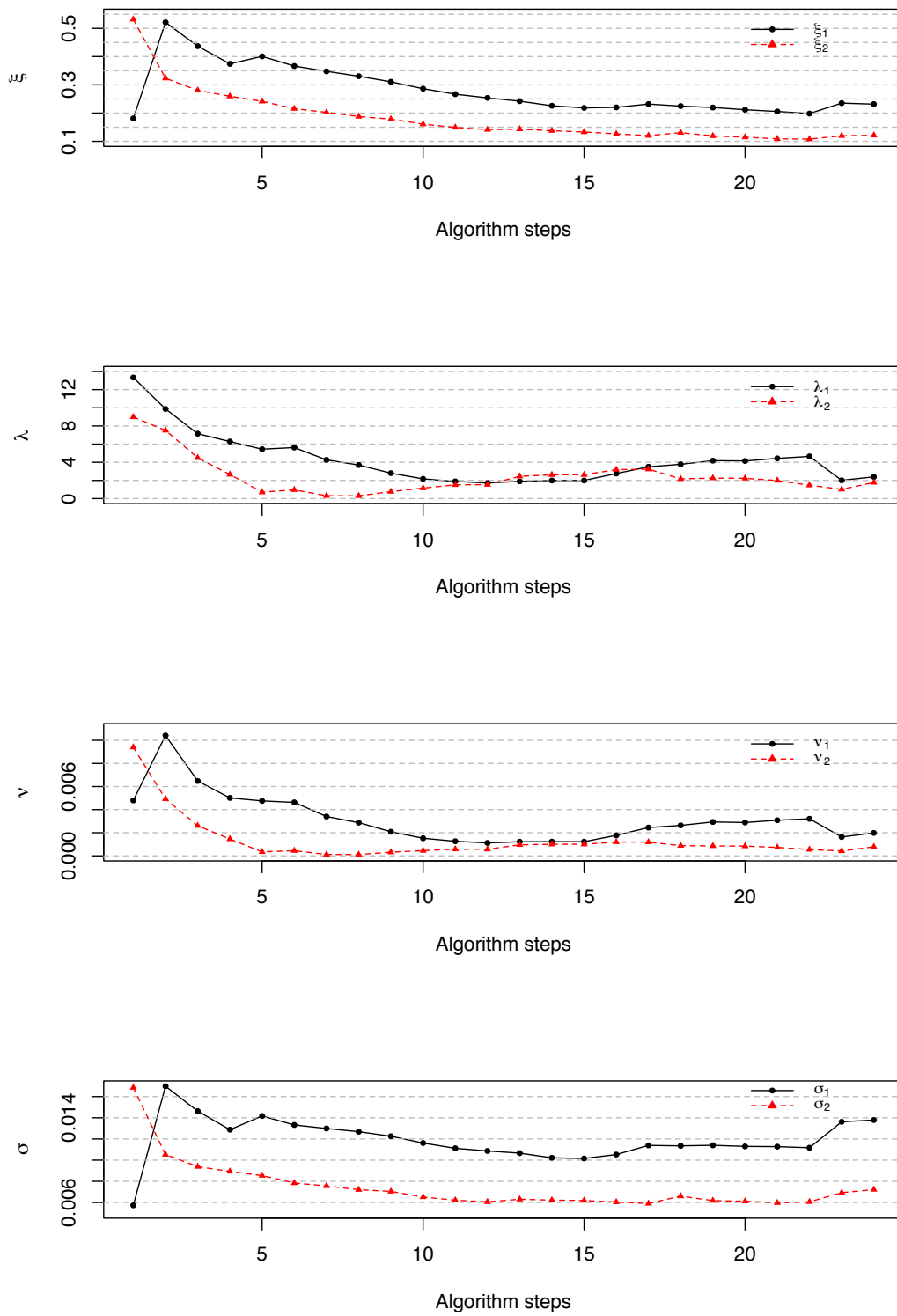


Figure 5.3: Evolution of parameter estimates ξ , λ , ν , and σ under a 2-state HOHMM for prices of futures with expiry 30 Dec 2016.

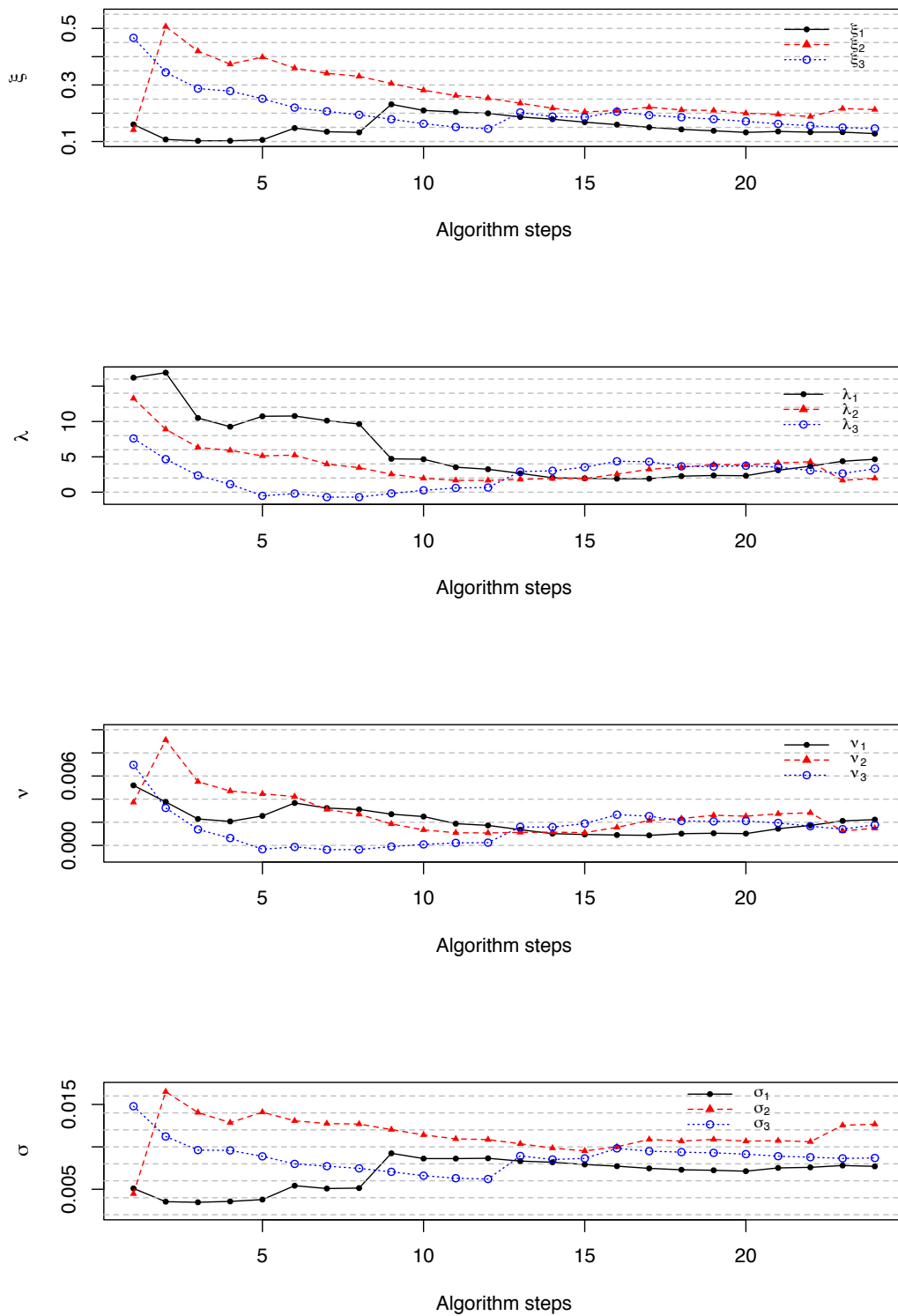


Figure 5.4: Evolution of parameter estimates ξ , λ , ν , and σ under a 3-state HOHMM for prices of futures with expiry 30 Dec 2016.

salmon futures prices, we evaluate the prediction of G_t under the 1-, 2-, and 3-state HMMs and HOHMMs. The one-step ahead forecasts in Date et al. [5] are extended; additionally, the d -step ahead predictions for all proposed models are assessed.

Proposition 5.4.1 *Given information up to time k , the ‘best’ estimate of the d -step ahead forecast of the g -dimensional observation process $G_{k+1}^h = \log F_{k+1}^h$ is*

$$\mathbb{E} \left[F_{k+d}^h \mid \mathcal{F}_k \right] = F_k^h \prod_{l=1}^d \sum_{j,i=1}^N \langle \mathbf{B}^{l-1} \mathbf{q}_l, \mathbf{e}_{ji} \rangle \exp \left(v_j^h + \frac{(\sigma_j^h)^2}{2} \right). \quad (5.33)$$

Proof See Appendix F.

Figure 5.5 depicts the one-step ahead prediction, taking $d = 1$ in equation (5.33), for salmon futures prices under the 3-state HOHMM setting with maturity dates of 28 Oct 2016, 29 Nov 2016 and 29 Dec 2016. We can observe that the one-step ahead forecasts are quite close to the actual market data. Although not shown here, similar patterns can be found for the one-step ahead forecasts under the HMM setting as well. The forecasts, with very short term, from our proposed models follow closely the actual prices at the Fish Pool. The trends and dynamics of salmon futures prices, from visual inspection, are captured well by our filtering algorithms and estimation procedure.

A formal way of quantifying forecasts’ quality is through an error analysis, i.e, examining the goodness of fit for all the proposed HMMs and HOHMMs. Following the criteria in Erlwein et al. [7] and Date et al. [5, 4], we evaluate the root-mean-squared error (RMSE), absolute-mean error (AME), relative-absolute error (RAE), and mean-absolute-percent error (MAPE) of the proposed and competing multi-dimensional models. We extend further the assessment of the error metrics to the case of d -step ahead forecasts.

The results of error analyses for $d = 1, 2, \dots, 5$, i.e., spanning the entire next week’s trading days, are presented in Table 5.4. For all d , the 3-state model outperforms other HOHMM-state settings. Except for the 1-step ahead forecasting case, where the 2-state HMM yields better fit than those obtained from the 1-and 3-state HMMs, the HOHMM with 3 states is adjudged better than any HMM set up in terms of goodness of fit. Compared to the HMM settings, the HOHMM settings produce smaller errors in one-step ahead predictions, whilst no pronounced difference is detected as the forecasting horizons become longer. The 4-state HMM and HOHMM settings, which were examined as well, effect only negligible improvements. Therefore, there is no practical motivation to pursue an HOHMM with a

<i>d</i> -step ahead	HMM setting	RMSE	MAPE	AME	RAE
1	1-state	0.5104	0.5195	0.2962	0.0498
	2-state	0.5036	0.4926	0.2859	0.0481
	3-state	0.5055	0.5051	0.2876	0.0484
2	1-state	0.8062	0.9130	0.5181	0.0878
	2-state	0.7852	0.8718	0.4941	0.0840
	3-state	0.7842	0.8654	0.4909	0.0834
3	1-state	1.0685	1.2607	0.7104	0.1213
	2-state	1.0388	1.2053	0.6791	0.1165
	3-state	1.0387	1.2021	0.6776	0.1162
4	1-state	1.3048	1.5741	0.8804	0.1514
	2-state	1.2655	1.5013	0.8397	0.1453
	3-state	1.2650	1.4903	0.8338	0.1443
5	1-state	1.5298	1.8622	1.0334	0.1788
	2-state	1.4782	1.7672	0.9812	0.1713
	3-state	1.4779	1.7537	0.9738	0.1701
<i>d</i> -step ahead	HOHMM setting	RMSE	MAPE	AME	RAE
1	1-state	0.5104	0.5195	0.2962	0.0498
	2-state	0.5029	0.4964	0.2824	0.0476
	3-state	0.5021	0.4926	0.2804	0.0472
2	1-state	0.8062	0.9130	0.5181	0.0878
	2-state	0.7851	0.8131	0.4614	0.0787
	3-state	0.7847	0.8114	0.4605	0.0785
3	1-state	1.0685	1.2607	0.7104	0.1213
	2-state	1.0389	1.1290	0.6372	0.1100
	3-state	1.0386	1.1276	0.6364	0.1099
4	1-state	1.3048	1.5741	0.8804	0.1514
	2-state	1.2653	1.4192	0.7959	0.1392
	3-state	1.2651	1.4188	0.7956	0.1391
5	1-state	1.5298	1.8622	1.0334	0.1788
	2-state	1.4778	1.7024	0.9480	0.1678
	3-state	1.4777	1.7024	0.9480	0.1678

Table 5.4: Error analysis of HMM- and HOHMM-based models under 1-, 2-, and 3-state settings for salmon futures prices

Setting	HMM			HOHMM		
	1-state	1-state	2-state	1-state	1-state	2-state
<i>t</i> -test	vs	vs	vs	vs	vs	vs
	2-state	3-state	3-state	2-state	3-state	3-state
<i>p</i> -value	4.3026×10^{-7}	4.3225×10^{-4}	0.8564	2.5657×10^{-5}	1.6071×10^{-8}	1.0000

Table 5.5: Bonferroni-corrected *p*-values for the paired *t*-test performed on the RMSEs involving salmon futures prices

number of states greater than 3. In fact, it is also prohibitive to consider a large number of states since the size of the transition probability entries becomes unwieldy even with just one regime addition; see Figures 5.1 and 5.2. Notwithstanding the small fitting errors seen in general for the 1-, 2- and 3-state HMM and HOHMM settings, HMM and HOHMM with a regime-switching feature clearly outperform the special case, 1-state model, across all forecasting metrics. This indicates the indisputable benefit of incorporating regime-switching and memory-capturing capabilities into models. Overall, the 3-state HOHMM is the best modelling framework in accurately describing the dynamics of our salmon futures-price data set.

The statistical significance of the error-mean differences in each pairwise setting of our proposed models is evaluated by a *t*-test based on a 95% confidence level. The adjusted *p*-values are computed with the Bonferroni's approach to control the family wise error rate. Table 5.5 displays the estimated Bonferroni-corrected *p*-values for the three pairs of model settings.

For the 2-state HMM versus 3-state HMM and the 2-state HMM versus 3-state HOHMM models, the *p*-values are larger than 0.05. Therefore, the null hypothesis of no significant difference cannot be rejected. This tells us that the 2-state and 3-state HMM settings (also 2-state and 3-state HOHMM settings) have similar capability in capturing the dynamics of the observation process although Table 5.4 shows the 3-state is slightly better than the 2-state setting. For the 1-state versus 2-state and 1-state versus 3-state setting under HMM (also, 1-state versus 2-state and 1-state versus 3-state setting under HOHMM) the *p*-values are appreciably smaller than 0.01. So, the RMSE differences between the HMM-based switching and no-switching models (also, between the HOHMM-based switching and no-switching models) are statistically significant.

Model	1-state	2-state	3-state	N -state
HMM	2	6	12	$N^2 + 2N$
HOHMM	2	8	24	$N^3 - N^2 + 2N$

Table 5.6: Number of estimated parameters under HMM and HOHMM settings for salmon futures prices

The results of our error analyses will be reinforced by information-criterion assessments in identifying the ‘best-performing’ model. Information-criteria assessment techniques aim to provide a balance between model’s goodness of fit and complexity (i.e., increased number of parameters). To put it simply, the information-criterion metrics provide a comprehensive trade-off between bias and variance of our HMM and HOHMM settings. Here, we make use of the Akaike information criterion (AIC), the small-sample-size corrected version of AIC (AICc), and the Bayesian information criterion (BIC).

The AIC emphasises the penalty for increasing the number of parameters. The AICc is the AIC with a correction for a relatively small size of model parameters by increasing the penalty for model complexity. The BIC strengthens the AIC and AICc by selecting the ‘best’ model from a set of candidate models with a different penalty for the addition of parameters. The AIC, AICc and BIC are computed as

$$\text{AIC} = -2 \log L(\Theta) + 2l, \quad (5.34)$$

$$\text{AICc} = -2 \log L(\Theta) + 2l + \frac{2l(l+1)}{m-l-1}, \quad (5.35)$$

$$\text{BIC} = -2 \log L(\Theta) + 2l \log m, \quad (5.36)$$

where l is the number of model parameters to be estimated as summarised in Table 5.6. In (5.36), m is the number of data points, and $\log L(\Theta)$ is the log-likelihood function associated with the model given by

$$\log L(\Theta) = \sum_{h=1}^g \sum_{k=1}^B \sum_{t=1}^N \langle \mathbf{y}_k^w, \mathbf{e}_t \rangle \left(\log \left(\frac{1}{\sqrt{2\pi}\sigma^h(\mathbf{y}_k^w)} \right) - \frac{(G_{k+1} - G_k - v(\mathbf{y}_k^w))^2}{2(\sigma^h(\mathbf{y}_k^w))^2} \right) \quad (5.37)$$

with B being the number of observations in each algorithm pass. The model that gives the smallest information-criterion assessment value obtained using equations (5.34)– (5.36) is preferred. We compute the values of AIC, AICc and BIC for the 1-, 2- and 3-state HMMs and HOHMMs after each algorithm step. The evolution of the calculated values of these information-criterion metrics is plotted in Figure 5.6. We observe that the the 1-state

model generates the largest values and the most volatile patterns throughout the entire data period. Even though the 3-state HOHMM produces higher values than those from other regime-switching models at the earlier part of the algorithm pass, it generally has stable patterns and smaller AIC/AICc/BIC values thereafter. These findings fortify the results of the above-mentioned error analyses, selecting the 3-state HOHMM as the best model for the data set studied in this chapter.

Probing the models' forecasting performance entails as well running the model on data set covering the out-of-sample period, i.e., the data period not used in model estimation. The empirical evidence from the in-sample performance is commonly less reliable than the those generated from analysing forecasts in the out-of-sample period, which better reflects the information available in 'real time'; see White [30], and Stock and Watson [29]. As the d -step ahead prediction is forward-looking and does not overlap with the period encompassing the data used for the filtering and estimation of parameters, our approach is distinctly consistent with the out-of-sample forecasting.

This section on numerical demonstration culminates with an appraisal of all model settings on their one-step ahead prediction ability using price data of futures with maturities after 30 Dec 2016. Specifically, this out-of-sample forecasting exercise focuses on futures with maturities up to 6 months that expire on the last business day of Jan 2017, Feb 2017, ..., Jun 2017. Figure 5.7 exhibits the comparison between the actual and predicted futures prices under the 3-state HOHMM. It shows that the movement of the one-step ahead predictions resemble the actual data. Such a result is also in agreement with what Figure 5.5 reveals. The results of the out-of-sample error analysis under the 1-, 2- and 3-state HMMs and HOHMMs are also reported in Table 5.7. The 2- and 3-state HMMs and HOHMMs incontestably outperform the 1-state model with respect to all forecasting metrics whilst the 3-state HOHMM has the smallest forecast errors. Summing up altogether the pertinent evidence from error analyses, information-criteria evaluations, and out-of-sample predictions, it is ascertained that there is merit in putting forward the regime-switching settings for the modelling and forecasting of Fish Pool's futures prices. The 3-state HOHMM turns out to be the best-fitting model for the chosen sample data within the intent of our empirical analysis.

	HMM setting			HOHMM setting		
	1-state	2-state	3-state	1-state	2-state	3-state
RMSE	0.4517	0.4369	0.4434	0.4517	0.4370	0.4367
MAPE	0.4745	0.4250	0.4256	0.4745	0.4250	0.4239
AME	0.3231	0.2889	0.2890	0.3231	0.2889	0.2881
RAE	0.0742	0.0665	0.0666	0.0742	0.0665	0.0663

Table 5.7: Out-of-sample error analysis of HMM- and HOHMM-based models under 1-, 2-, and 3- state settings for futures prices

5.5 Conclusion

In this work, we put forward a multi-dimensional HOHMM-based approach in examining the evolution of salmon futures prices. It is assumed that the parameters of the arbitrage-free log-futures prices model are driven by a second-order hidden Markov chain in discrete time. The usual approach to solve for the MLEs is to utilise the classic Kalman filtering under a univariate setting, which is common in the price discovery research of the aquaculture literature. But in light of the correlated multivariate nature of the data for salmon futures prices, we developed suitable and congruous self-calibrating filtering algorithms, given in matrix representations, for optimal parameter estimation and short-term price prediction.

To illustrate the implementability and prediction performance of our approach, we provided a numerical demonstration utilising fresh-farmed salmon futures prices publicly available at the Fish Pool. The HOHMM settings were tested on an extensive data set of daily futures contracts that are actively traded in the market and mature within 6 months. The innovation of our implementation, in conjunction with our self-tuning estimation framework, emanates from the design of data processing scheme that exhausts every possible raw price information despite the complexity brought by the varying maturities of futures contracts trading in the Fish Pool market. Parameter estimates were generated by filtering recursions under the paradigm of shifting market regimes. We conducted post model diagnostics for our proposed HOHMM setting, and also benchmarked it with the regular HMM setting. A comparative analysis of model-validation metrics was performed concentrating on error analysis, charting dynamic information-criteria evolution, and d -day ahead

forecasts for $d = 1, \dots, 5$. It is shown that the inclusion of a regime-switching and memory-description features into the model brings about advantages in attaining better goodness of fit and price-forecasts accuracy of salmon futures prices. The 3-state HOHMM has the utmost sufficiency to describe the essential attributes of the data set gathered for the empirical analysis in this chapter.

This work contributes to the widening of the collection of available quantitative techniques, enabling further the use of financial technologies in the management of price risk and volatilities in the fisheries sector and aquaculture industry. Our results could be extended in a number of directions, such as applying the proposed framework to develop dynamic hedging strategies and explore optimisation of portfolios with direct exposure to prices of fish and seafood.

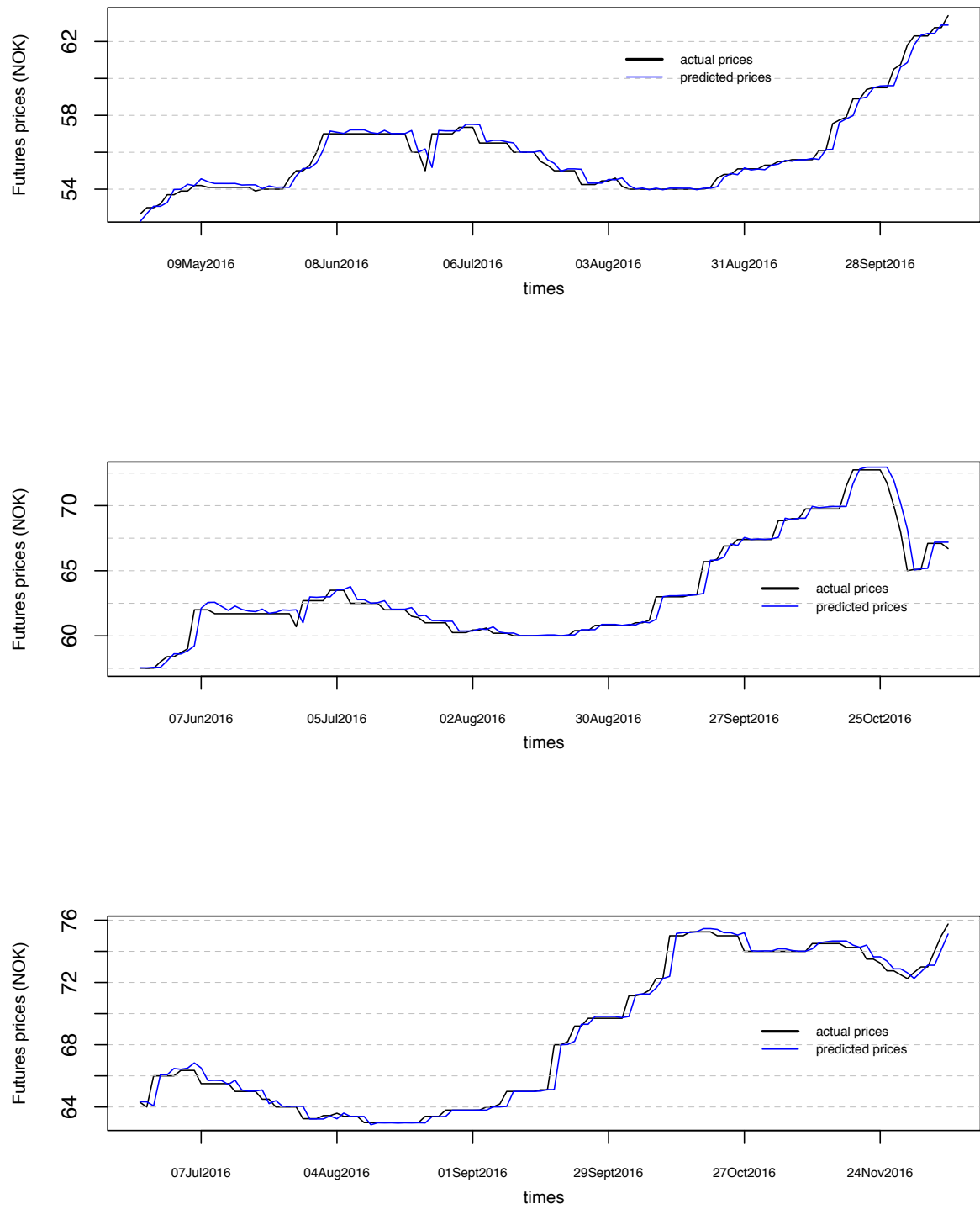


Figure 5.5: One-step ahead predictions for futures prices with expiries on 28 Oct 2016, 29 Nov 2016, and 29 Dec 2016.

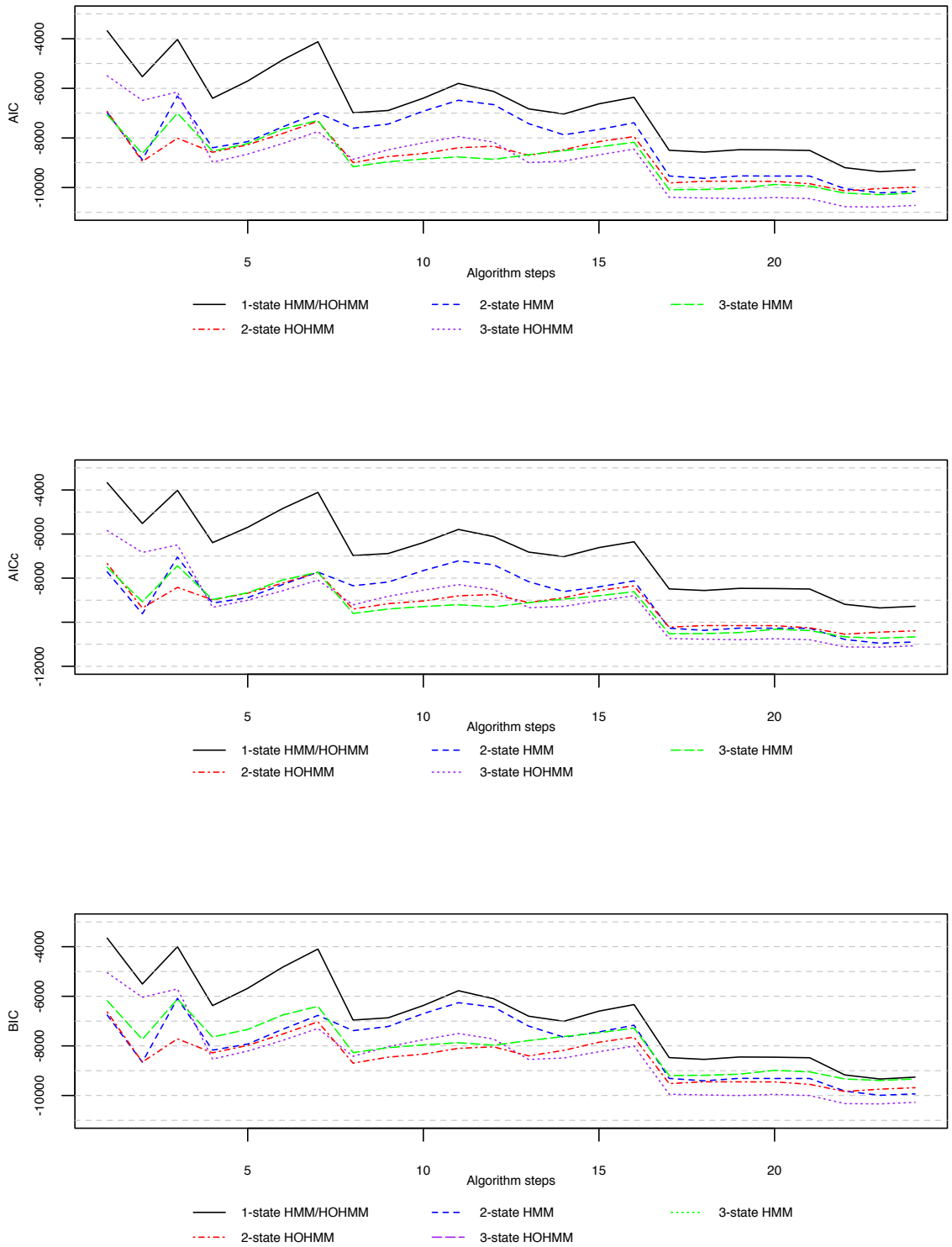


Figure 5.6: AIC, AICc and BIC for the 1-, 2-, 3-state HMM/HOHMM-based model for salmon futures prices.

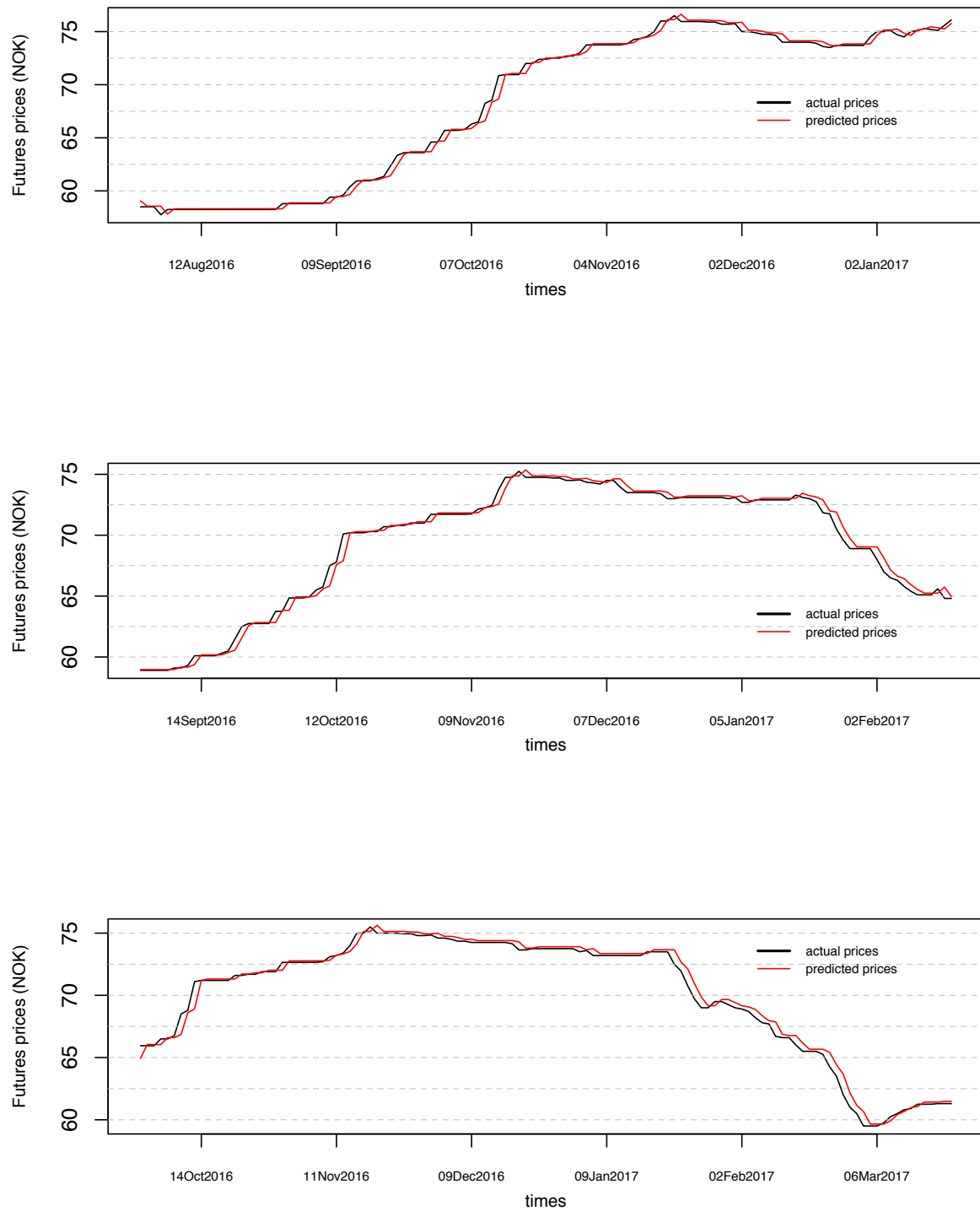


Figure 5.7: One-step ahead and out-of-sample predictions under the 3-state HOHMM for futures prices with expiries on 31 Jan 2017, 28 Feb 2017, and 31 Mar 2017.

References

- [1] F. Asche, B. Misund, A. Oglend, The spot-forward relationship in the Atlantic salmon market, *Aquaculture Economics and Management*, 20(2)(2016), 222–234.
- [2] I. Ankamah-Yeboah, M. Nielsen, R. Nielsen, Price formation of the salmon aquaculture futures market, *Aquaculture Economics and Management*, (2017) 1–24. DOI: 10.1080/13657305.2016.1189014
- [3] O. Cacho, Systems modelling and bioeconomic modelling in aquaculture, *Aquaculture Economics and Management*, 1(1-2)(1997), 45–64.
- [4] P. Date, L. Jalen, R. Mamon, A partially linearised sigma point filter for latent state estimation in nonlinear time series models, *Journal of Computational and Applied Mathematics*, 233(2010) 2675–2682.
- [5] P. Date, R. Mamon, A. Tenyakov, Filtering and forecasting commodity futures prices under an HMM framework, *Energy Economics*, 40(2013) 1001–1013.
- [6] R. Elliott, L. Aggoun, J. Moore, J., *Hidden Markov Models: Estimation and Control*, Springer, New York (1995).
- [7] C. Erlwein, F. Benth, R. Mamon, HMM filtering and parameter estimation of an electricity spot price model, *Energy Economics*, 32 (5) (2010) 1034–1043.
- [8] R. J. Elliott, T. Kuen Siu, A. Badescu, Bond valuation under a discrete-time regime-switching term-structure model and its continuous-time extension, *Managerial Finance*, 37(11)(2011) 1025–1047.
- [9] C. Erlwein, R. Mamon, M. Davison, An examination of HMM-based investment strategies for asset allocation, *Applied Stochastic Models in Business and Industry*, 27(3)(2011), 204–221.

- [10] C. Erlwein, R. Mamon, An online estimation scheme for a Hull-White model with HMM-driven parameters, *Statistical Methods and Applications*, 18(1)(2009) 87–107.
- [11] C. Ewald, Derivatives on nonstorable renewable resources: Fish futures and options, not so fishy after all, *Natural Resource Modeling*, 26(2) (2013), 215–236.
- [12] C. Ewald, R. Ouyang, T. K. Siu, On the market-consistent valuation of fish farms: Using the real option approach and salmon futures, *American Journal of Agricultural Economics*, 99(1)(2017), 207–224.
- [13] C. Ewald, R. Ouyang, T. K. Siu, The market for salmon futures: An empirical analysis of the Fish Pool using the Schwartz multi-factor model. *Quantitative Finance*, 16(12)(2016), 1823–1842.
- [14] Fish Pool Index. <http://fishpool.eu/price-information/spot-prices/fish-pool-index/> (accessed Oct 2017).
- [15] A. Guttormsen, Faustman in the sea: Optimal rotation in aquaculture. *Marine Resource Economics*, 23(4)(2008), 401–410.
- [16] M. Hardy, A regime-switching model of long-term stock returns, *North American Actuarial Journal*, 5 (2) (2001), 41–53.
- [17] M. Manoliu, S. Tompaidis, Energy futures prices: Term structure models with Kalman filter estimation, *Applied Mathematical Finance*, 9(1)(2002) 21–43.
- [18] Market competition between farmed and wild fish: A literature survey, *Food and Agriculture of the United Nations*. <http://www.fao.org/3/a-i5700e.pdf> (accessed Oct 2017).
- [19] J. Martínez–Garmendia, J. Anderson, Hedging performance of shrimp futures contracts with multiple deliverable grades, *Journal of Futures Markets*, 19(8)(1999), 957–990.
- [20] R. Mamon, R. Elliott, *Hidden Markov Models in Finance*, International Series in Operations Research and Management Science, 104, Springer, New York, 2007.
- [21] R. Mamon, R. Elliott, *Hidden Markov Models in Finance: Further Developments and Applications*, International Series in Operations Research and Management Science, 104, Springer, New York, 2014.

- [22] R. Mamon, C. Erlwein, R. Gopaluni, Adaptive signal processing of asset price dynamics with predictability analysis. *Information Sciences*, 178(2008), 203–219.
- [23] Press release, Fish Pool turned over 90000 tons of salmon last year, 2017, <http://ilaks.no/fish-pool-snudde-over-90-000-tonn-laks-ifjor/> (accessed Oct 2017).
- [24] K. Quagraine, C. Engle, A latent class model for analysing preferences for catfish, *Aquaculture Economics and Management*, 10(1)(2006), 1–14.
- [25] S. Ross, S. Hedging long run commitments: Exercises in incomplete market pricing, *Economic Notes: Economic Review of Banca Monte dei Paschi di Siena*. 26(2)(1997) 385–420.
- [26] E. Schwartz, E. The Stochastic behavior of commodity prices: Implications for valuation and hedging, *Journal of Finance*, 52(3)(1997) 923–973.
- [27] P. Solibakke, Scientific stochastic volatility models for the salmon forward market: Forecasting (un-)conditional moments. *Aquaculture Economics and Management*, 16(3)(2012), 222–249.
- [28] T. Siu, W. Ching, E. Fung, M. Ng, X. Li, A high-order Markov-switching model for risk measurement, *Computers and Mathematics with Applications*, 58(1) (2009) 1–10.
- [29] J. Stock, M. Watson, *Introduction to Econometrics* (3rd ed.). Boston: Addison-Wesley. (2011)
- [30] H. White, A reality check for data snooping. *Econometrica*, 68(5)(2000), 1097–1126.
- [31] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. Hoboken, NJ; Chichester, England: John Wiley and Sons. (2006).
- [32] X. Xi, R. Mamon, Parameter estimation of an asset price model driven by a weak hidden Markov chain, *Economic Modelling* 28(1) (2011) 36–46.
- [33] X. Xi, R. Mamon, Yield curve modelling using a multivariate higher-order HMM. In Zeng, Y. and Wu, S. (eds), *State-Space Models and Applications in Economics and Finance*. New York, Springer (2013) 185–203.

- [34] X. Xi, R. Mamon, M. Davison, A higher-order hidden Markov chain-modulated model for asset allocation, *Journal of Mathematical Modelling and Algorithms in Operations Research* 13(1) (2014) 59–85.
- [35] X. Xi, R. Mamon, Parameter estimation in a WHMM setting with independent and volatility components. In Mamon, R. and Elliott, R (eds), In: *Hidden Markov Models in Finance: Volume II (Further Developments and Applications)*. New York, Springer (2014) 227–240.
- [36] X. Xi, R. Mamon, Capturing the regime-switching and memory properties of interest rates, *Computational Economics*, 44(3) (2014), 307–337.
- [37] H. Xiong, R. Mamon, A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics, *Journal of Computational Science*, 17 (2016), 47–61.
- [38] H. Xiong, R. Mamon, Putting a price tag on temperature, *Computational Management Science*, (2017), <https://doi.org/10.1007/s10287-017-0291-8>

Chapter 6

Conclusion

6.1 Summary of research contributions

In this thesis, we developed more extensions of the the regime-switching models governed by HOHMCs. The corresponding filtering algorithms to support the new models' dynamic parameter estimation were fully constructed. The usual Markov assumption in regime-switching approaches was relaxed in our HOHMM setting by having dependency of data beyond the first-order lag times. In turn, such a relaxation of the Markov assumption equips a capacity to exploit information from time series data recorded in the past. Various empirical implementations were put forward, and they were specifically designed to capture some stylised behaviours of commodity prices and indices for support of derivative valuation and risk management. Modelling set ups were especially tailored for the dynamics of variables that primarily affect values of weather, electricity, and aquaculture-based contracts.

The higher-order hidden Markov process modulates the parameters in discrete time to capture the random shifts amongst different economic regimes resulting from the interaction of various factors. We derived, via some 'idealised' reference probability measure, the recursive filters for the state of the Markov chain and auxiliary quantities of the observation process. EM parameter estimates are expressed in terms of the adaptive filters, which produces self-calibrating modelling methodology. The performance of the four proposed models was assessed and benchmarked against present standard models in the literature and current practice, comparing various statistical metrics that quantify the model's goodness of fit (i.e., minimised forecasting errors) and the balance between model's maximised likelihood and inherent complexity.

This research work offers further developments and valuable insights in both the theoretical and practical aspects of regime-switching models covering recent financial innovations in the commodity markets. As a recapitulation, the accomplishments of this research began with the modelling in Chapter 2 of the DATs under a discrete-time HMM; an online parameter estimation scheme and evaluation of a temperature-based option with parameter sensitivity analysis were also presented. The formulation of a temperature model with a discrete-time HOHMC governing the model parameters was introduced in Chapter 3. Recursive filtering equations for various quantities as a function of HOHMM were aided smoothly by the probability-measure-change technique, and we underscored a numerical implementation on a data set of 4-year Toronto's DATs. We created a modelling set-up combining OU and jump processes, both of which are modulated by a HOHMC, in the depiction of electricity spot-price movement in Chapter 4. The model was applied to a deseasonalised series of AESO-collected data. Chapter 5 advanced the use of multi-dimensional HOHMM in the modelling and very short-term prediction of futures prices in the aquaculture industry. An empirical illustration was carried out on salmon futures prices in the Fish Pool market for the purpose of model validation.

Over all, the significant research contributions in this thesis highlight the: (i) generalisation of the regular HMM-OU framework to a modelling set up that possesses the means to extract information from the data observed beyond a unit time lag, hence incorporating the flexibility to capture data's memory property; (ii) development of dynamic model calibration procedures, through HOHMM-filtering, that addresses the need for financial technologies for automation in the context of today's artificial-intelligence-driven world; (iii) valuation and risk measurement of temperature-linked derivatives under a regime-switching approach; (iv) enrichment of the HOHMM-OU model with a compound Poisson process to accurately chart electricity spot price dynamics; and (v) design of multi-dimensional HOHMM filters and predictors in the analysis of salmon futures prices, whereby the dynamic estimation picks up and processes all available raw data points once only and no observed values are discarded or transformed into proxy values that yield more approximation errors and uncertainty of model results.

6.2 Further research directions

We recognise that, despite the extensions achieved in our proposed models, shortcomings in our approaches remain that will open avenues for research investigations. We laid down some of the ground work to push further the theoretical and practical considerations involved in HOHMM filtering techniques to bolster their potential benefit as new kinds of data series unfold in the areas of finance, engineering, and social sciences. Several offshoots and ramifications of this research work are itemised below.

- As indicated in this thesis, we just utilised an HOHMC with a lag order 2, as a paradigm that simplifies the elaboration of details in the HOHMM-setting development. This helps in the rough quantification of the computational complexity when reducing the lag order from, say k to $k - 1$, for $k = 2, 3, \dots$. Ideally, however, it is desirable to come up with statistical inference techniques that will give the HOHMM's optimal lag order \widehat{k} suitable for a given data set.
- The construction of our HOHMM filtering algorithms is aligned with the discrete-time nature of the observed time series. As an alternative, we have yet to see the development of the HOHMMs and their filters in continuous time. For example, recent regime-switching models with continuous-time HMM filters were formulated in [1]. Although this type of filters requires discretisation and integral approximations, there are instances that filter derivations are more straightforward because they are standard continuous-time stochastic calculus computations. Thus, there is merit to consider continuous-time filtering for HOHMMs, which is still an inchoate area in stochastic modelling.
- Our HOHMM-based modelling set ups have a Gaussian-noise term. Similar to HMM-based set ups as espoused in [2], filtering recursions under non-Gaussian noise terms could be considered for an even greater flexibility of capturing a variety of shapes of statistical distributions in the context of big data and new kinds of data given recent financial-market innovations and regulatory developments. Implementation of new filters under non-normal correlated multivariate models is also anticipated to entail more advanced computing platforms to attain efficiency of self-tuning parameter estimation.
- A natural direction in modelling the DATs under the HOHMM framework is the pricing and dynamic hedging of other weather-linked products. For instance, the

modelling of weather futures prices can be pursued by multivariate HOHMM-based filtering algorithms. The filtered-based EM parameter estimates require a connection with a suitable risk-neutral measure via the construction a quantity that encapsulates the risk premium (i.e., price and switching risks). Optimisation of portfolios containing products linked to weather indices is another closely related problem that could benefit from our HOHMM approach.

- With regard to the modelling of electricity-spot, further applications to the pricing of more complicated electricity derivatives other than just forward contracts could be explored. Applications that consider both weather and electricity derivatives in tandem will be of a particular interest to practitioners. For instance, a dynamic risk management strategy involving electricity and temperature-based futures could be supported by the HOHMM filtering algorithms in finding optimal hedge ratios.
- Climate change in recent times significantly disturbs salmon population dynamics and leads to harvest uncertainty. For instance, sea surface temperature (SST), with new peculiarities, is a key factor that affects the fish futures market; see L. Little et al. [3]. A regime-switching SST model that includes spatial and temporal variables could be proposed, and the HOHMM-filtering recursions for parameter estimation could be derived. Furthermore, co-integration between SST and fish futures prices could be examined to offer risk managers with possible toolkits in dealing with the problem of estimating the optimal fish-farm harvesting time, and in engineering appropriate hedging strategies.

References

- [1] S. Grimm, An interest-rate model with regime-switching mean-reversion level, PhD thesis, (2016), <https://www.genealogy.math.ndsu.nodak.edu/id.php?id=211815>.
- [2] L. Jalen, R. Mamon, Parameter estimation in a WHMM setting with independent and volatility components. In Mamon, R. and Elliott, R (eds), In: Hidden Markov Models in Finance: Volume II (Further Developments and Applications). New York, Springer (2014) 241–261.
- [3] L. Little, A. Hobday, J. Parslow, C. Davies, R. Grafton, Funding climate adaptation strategies with climate derivatives, *Climate Risk Management*, 8(C)(2015), 9–15.

Appendix A

Derivations of the model's optimal parameter estimates in chapter 2

A.1 Optimal estimate for δ

Define a new measure $P^{\widehat{\nu}}$ via

$$\left. \frac{dP^{\widehat{\nu}}}{dP^{\nu^*}} \right|_{\mathcal{X}_k} = \Psi_k^\delta = \prod_{l=1}^k \varphi_l^\delta.$$

where

$$\varphi_l^\delta = \frac{\exp\left(-\frac{1}{2\epsilon^2(\mathbf{y}_{l-1})} (X_l - \widehat{\delta}(\mathbf{y}_{l-1}) X_{l-1} - \eta(\mathbf{y}_{l-1}))^2\right)}{\exp\left(-\frac{1}{2\epsilon^2(\mathbf{y}_{l-1})} (X_l - \delta(\mathbf{y}_{l-1}) X_{l-1} - \eta(\mathbf{y}_{l-1}))^2\right)}.$$

Therefore, the log likelihood for Ψ_k^δ is

$$\begin{aligned} \log \Psi_k^\delta &= \sum_{l=1}^k \left(-\frac{\widehat{\delta}^2(\mathbf{y}_{l-1}) X_{l-1}^2 - 2X_l \widehat{\delta}(\mathbf{y}_{l-1}) X_{l-1} + 2\eta(\mathbf{y}_{l-1}) \widehat{\delta}(\mathbf{y}_{l-1}) X_{l-1}}{2\epsilon^2(\mathbf{y}_{l-1})} + \mathbf{R}(\delta(\mathbf{y}_{l-1})) \right) \\ &= \sum_{l=1}^k \left(\sum_{i=1}^n \langle \mathbf{y}_{l-1}, \mathbf{e}_i \rangle \left(-\frac{\widehat{\delta}_i^2 X_{l-1}^2 - 2X_l \widehat{\delta}_i X_{l-1} + 2\eta_i \widehat{\delta}_i X_{l-1}}{2\epsilon_i^2} + \mathbf{R}(\delta_i) \right) \right). \end{aligned}$$

Since $\mathbf{R}(\delta_i)$ does not contain $\widehat{\delta}_i$, such remainder has no affect on the result of the derivation.

From equations (2.19) and (2.20), and considering $\widehat{U}_l = \mathbb{E}(U_l | \mathcal{X}_k)$, we have

$$\begin{aligned} L(\widehat{\delta}_i) &= \sum_{i=1}^N \mathbb{E} \left[-\frac{1}{2\epsilon_i^2} \left(\widehat{\delta}_i^2 \mathcal{T}_k^i(X_{k-1}^2) - 2\widehat{\delta}_i \mathcal{T}_k^i(X_{k-1}, X_k) + 2\eta_i \widehat{\delta}_i \mathcal{T}_k^i(X_{k-1}) \right) \middle| \mathcal{X}_k \right] + \mathbf{R}(\delta_i) \\ &= \sum_{i=1}^N -\frac{1}{2\epsilon_i^2} \left(\widehat{\delta}_i^2 \mathcal{T}_k^i(X_{k-1}^2) - 2\widehat{\delta}_i \mathcal{T}_k^i(X_{k-1}, X_k) + 2\eta_i \widehat{\delta}_i \mathcal{T}_k^i(X_{k-1}) \right) + \mathbf{R}(\delta_i). \end{aligned}$$

We differentiate $L(\widehat{\delta}_i)$ with respect to $\widehat{\delta}_i$ and set the result to zero giving

$$\widehat{\delta}_i = \frac{\widehat{\mathcal{F}}_k^i(X_{k-1}, X_k) - \eta_i \widehat{\mathcal{F}}_k^i(X_{k-1})}{\widehat{\mathcal{F}}_k^i(X_{k-1}^2)}.$$

A.2 Optimal estimate for η

Define a new measure $P^{\widehat{\nu}}$ through $\left. \frac{dP^{\widehat{\nu}}}{dP^{\nu^*}} \right|_{\mathcal{X}_k} = \Psi_k^\eta = \prod_{l=1}^k \varphi_l^\eta$,

where $\varphi_l^\eta = \frac{\exp\left(-\frac{1}{2\epsilon^2(\mathbf{y}_{l-1})} (X_l - \delta(\mathbf{y}_{l-1}) X_{l-1} - \widehat{\eta}(\mathbf{y}_{l-1}))^2\right)}{\exp\left(-\frac{1}{2\epsilon^2(\mathbf{y}_{l-1})} (X_l - \delta(\mathbf{y}_{l-1}) X_{l-1} - \eta(\mathbf{y}_{l-1}))^2\right)}$ leading to the log likelihood

$$\log \Psi_k^\eta = \sum_{l=1}^k \left(-\frac{\widehat{\eta}^2(\mathbf{y}_{l-1}) - 2X_l \widehat{\eta}(\mathbf{y}_{l-1}) + 2\widehat{\eta}(\mathbf{y}_{l-1}) \delta(\mathbf{y}_{l-1}) X_{l-1}}{2\epsilon^2(\mathbf{y}_{l-1})} + \mathbf{R}(\eta(\mathbf{y}_{l-1})) \right).$$

Invoking equations (2.19) and (2.20) and then taking expectation of the log likelihood involving X_k , we obtain $L(\widehat{\eta}) = \mathbb{E}[\log \Psi_k^\eta | \mathcal{X}_k]$ with

$$\mathbb{E}[\log \Psi_k^\eta | \mathcal{X}_k] = \sum_{l=1}^k \mathbb{E} \left[\left(\sum_{i=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}, \mathbf{e}_i \rangle}{2\epsilon_i^2} (\widehat{\eta}_i^2 - 2X_l \widehat{\eta}_i + 2\widehat{\eta}_i \delta_i X_{l-1}) \right) + \mathbf{R}(\eta_i) \right) \middle| \mathcal{X}_k \right].$$

Differentiation of $L(\widehat{\eta})$ and setting the result to 0, we get $\widehat{\eta}_i = \frac{\widehat{\mathcal{F}}_k^i(X_k) - \delta_i \widehat{\mathcal{F}}_k^i(X_{k-1})}{\widehat{\mathcal{O}}_k^i}$.

A.3 Optimal estimate for ϵ^2

Construct a new measure $P^{\widehat{\nu}}$ by setting

$$\left. \frac{dP^{\widehat{\nu}}}{dP^{\nu^*}} \right|_{\mathcal{X}_k} = \Psi_k^{\epsilon^2} = \prod_{l=1}^k \varphi_l^{\epsilon^2},$$

where $\varphi_l^{\epsilon^2} = \frac{\epsilon(\mathbf{y}_{l-1}) \exp\left(-\frac{1}{2\widehat{\epsilon^2}(\mathbf{y}_{l-1})} (X_l - \delta(\mathbf{y}_{l-1}) X_{l-1} - \eta(\mathbf{y}_{l-1}))^2\right)}{\widehat{\epsilon}(\mathbf{y}_{l-1}) \exp\left(-\frac{1}{2\widehat{\epsilon^2}(\mathbf{y}_{l-1})} (X_l - \delta(\mathbf{y}_{l-1}) X_{l-1} - \eta(\mathbf{y}_{l-1}))^2\right)}$. The log likelihood of $\Psi_k^{\epsilon^2}$ is calculated as

$$\log \Psi_k^{\epsilon^2} = \sum_{l=1}^k \left(\log \left(\frac{1}{\widehat{\epsilon}(\mathbf{y}_{l-1})} \right) + \left(-\frac{1}{2\widehat{\epsilon^2}(\mathbf{y}_{l-1})} \right) (X_l - \delta(\mathbf{y}_{l-1}) X_{l-1} - \eta(\mathbf{y}_{l-1}))^2 + \mathbf{R}(\epsilon(\mathbf{y}_{l-1})) \right),$$

where $\mathbf{R}(\epsilon^2)$ does not have $\widehat{\epsilon}$. From equations (2.19)–(2.20), we have the expectation of the log likelihood as a function of X_k , denoted by $L(\widehat{\epsilon^2})$, which is given by

$$\mathbb{E} \left[\log \Psi_k^{\epsilon^2} \mid \mathcal{X}_k \right] = \sum_{l=1}^k \mathbb{E} \left[\sum_{i=1}^N \langle \mathbf{y}_{l-1}, \mathbf{e}_i \rangle \left(\log \left(\frac{1}{\widehat{\epsilon}(\mathbf{y}_{l-1})} \right) + \left(-\frac{1}{2\widehat{\epsilon^2}(\mathbf{y}_{l-1})} \right) (X_l - \delta(\mathbf{y}_{l-1}) X_{l-1} - \eta(\mathbf{y}_{l-1}))^2 \right) \right] + \mathbf{R}(\epsilon_i).$$

Differentiating $L(\widehat{\epsilon^2})$ with respect to $\widehat{\epsilon^2}$ and equating the result to 0 yield

$$\widehat{\epsilon}_i^2 = \frac{\widehat{\mathcal{T}}_k^i(X_k^2) + \delta_i^2 \widehat{\mathcal{T}}_k^i(X_k^2) + \eta_i^2 \widehat{\mathcal{B}}_k^i + 2\eta_i^2 \delta_i \widehat{\mathcal{T}}_k^i(X_{k-1}) - 2\delta_i \widehat{\mathcal{T}}_k^i(X_{k-1}, X_k) - 2\eta_i \widehat{\mathcal{T}}_k^i(X_k)}{\widehat{\theta}_k^t}.$$

A.4 Optimal estimate for π_{ji}

Define the Radon-Nikodym derivative of $P^{\widehat{v}^*}$ with respect to P^{v^*} as $\left. \frac{dP^{\widehat{v}^*}}{dP^{v^*}} \right|_{\mathcal{X}_k} = \Psi_k^\pi = \prod_{l=1}^k \varphi_l^\pi$,

where $\varphi_l^\pi = \prod_{j,i=1}^n \left(\frac{\widehat{\pi}_{ji}}{\pi_{ji}} \right)^{\langle \mathbf{y}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{y}_l, \mathbf{e}_j \rangle}$. Taking expectation of the log likelihood in conjunction with equation (2.18), we have

$$\begin{aligned} \mathbb{E} [\log \varphi_l^\pi \mid \mathcal{X}_k] &= \mathbb{E} \left[\sum_{l=1}^k \sum_{j,i=1}^n \log \left(\frac{\widehat{\pi}_{ji}}{\pi_{ji}} \right)^{\langle \mathbf{y}_{l-1}, \mathbf{e}_j \rangle \langle \mathbf{y}_l, \mathbf{e}_i \rangle} \mid \mathcal{X}_k \right] \\ &= \mathbb{E} \left[\left(\sum_{l=1}^k \sum_{j,i=1}^n (\log \widehat{\pi}_{ji} - \log \pi_{ji}) \langle \mathbf{y}_{l-1}, \mathbf{e}_j \rangle \langle \mathbf{y}_l, \mathbf{e}_i \rangle \right) \mid \mathcal{X}_k \right] \\ &= \mathbb{E} \left[\sum_{j,i=1}^n \log \widehat{\pi}_{ji} \mathcal{J}_k^{tsr} \right] + \mathbf{R}(\pi_{ji}), \end{aligned}$$

where $R(\pi_{ji})$ is free of $\widehat{\pi}_{ji}$. Noting $\sum_{i=1}^n \widehat{\pi}_{ji} = 1$ and introducing a Lagrange multiplier ς , we maximise

$$L(\widehat{\pi}_{ji}, \varsigma) = \sum_{j,i=1}^n \log \widehat{\pi}_{ji} \mathcal{J}_k^{ji} + \varsigma \left(\sum_{j=1}^n \widehat{\pi}_{ji} - 1 \right) + R(\pi_{ji}).$$

Differentiation of $L(\widehat{\pi}_{ji}, \varsigma)$ with respect to $\widehat{\pi}_{ji}$ and ς and then setting each partial derivative to 0 would result to

$$\frac{1}{\widehat{\pi}_{ji}} \widehat{\mathcal{J}}_k^{ji} + \varsigma = 0.$$

Since $\sum_{j=1}^n \widehat{\pi}_{ji} = 1$ and $\sum_{j=1}^n \mathcal{J}_k^{ji} = \mathcal{O}_k^i$, we have $\sum_{j=1}^n \widehat{\pi}_{ji} = \frac{\sum_{j=1}^n \widehat{\mathcal{J}}_k^{ji}}{-\varsigma} = \frac{\mathcal{O}_k^i}{-\varsigma} = 1$, which can

be re-expressed as $\sum_{j=1}^n \widehat{\pi}_{ji} = \frac{\sum_{j=1}^n \widehat{\mathcal{J}}_k^{ji}}{\widehat{\mathcal{O}}_k^i}$. Therefore, the optimal estimate for p is given by

$$\widehat{\pi}_{ji} = \frac{\widehat{\mathcal{J}}_k^{ji}}{\widehat{\mathcal{O}}_k^i}.$$

Appendix B

Proofs of Propositions in chapter 2

B.1 Proof of Proposition 2.6.1

Proof The HDD futures price is calculated as the expected value of HDD over the contract period given the current value at time t under the risk-neutral measure Q . Based on the the Fubini-Tonelli theorem, expectation and integration can be interchanged. We firstly perform the proof for the case of $0 \leq t \leq \tau_1 < \tau_2$ as follows

$$\begin{aligned} F_H(t, \tau_1, \tau_2) &= \mathbb{E}^Q \left[H \middle| \mathcal{F}_t \right] \\ &= \mathbb{E}^Q \left[\int_{\tau_1}^{\tau_2} \max(T_{\text{base}} - T_v, 0) \, dv \middle| \mathcal{F}_t \right] \\ &= \int_{\tau_1}^{\tau_2} \mathbb{E}^Q \left[\max(T_{\text{base}} - T_v, 0) \middle| \mathcal{F}_t \right] \, dv. \end{aligned}$$

To solve for the futures price, it is necessary to compute the value of $\mathbb{E}^Q \left[\max(T_{\text{base}} - T, 0) \middle| \mathcal{F}_t \right]$. We suppose $\alpha(\mathbf{y}_t)$ is deterministic so that $\int_s^t e^{\alpha u} dB_u$ in equation (2.7) is a Wiener process (cf Benth and Šaltytė-Benth [2]). The integral $\int_s^t e^{\alpha(\mathbf{y}_u)u} dB_u^Q$ in equation (2.42), when evaluated using the optimal parameter estimates under the HMM settings, is then a Wiener process that incorporates deterministic regime-switching with filtration extended to time t . Under the normality assumption, $T_{\text{base}} - T_v$ is normally distributed with $\mathbb{E}^Q \left(\max(T_{\text{base}} - T, 0) \middle| \mathcal{F}_t \right) = M(t, v, X_t)$, and $\text{Var} \left(\max(T_{\text{base}} - T, 0) \middle| \mathcal{F}_t \right) = A^2(t, v)$. Write $M := M(t, v, X_t)$ and $A^2 :=$

$A^2(t, v)$. Let $T^* = \frac{T_{\text{base}} - T - M}{A^2}$, which has a standard normal distribution as well. Then

$$\begin{aligned}
\mathbb{E}^Q \left[\max(T_{\text{base}} - T, 0) \middle| \mathcal{F}_t \right] &= \int_{-\infty}^{T_{\text{base}}} - (T_{\text{base}} - T) \frac{1}{\sqrt{2\pi}} e^{-\frac{(T - (T_{\text{base}} - M))^2}{2A^2}} dT \\
&= \int_{-\frac{M}{A}}^{\infty} (AT^* + M) \frac{1}{\sqrt{2\pi}A} e^{-\frac{(T_{\text{base}} - AT^* - M - (T_{\text{base}} - M))^2}{2A^2}} A dT^* \\
&= \int_{-\frac{M}{A}}^{\infty} M \frac{1}{\sqrt{2\pi}} e^{-\frac{(T^*)^2}{2}} dT^* + \int_{-\frac{M}{A}}^{\infty} AT^* \frac{1}{\sqrt{2\pi}} e^{-\frac{(T^*)^2}{2}} dT^* \\
&= M \left(1 - \int_{-\infty}^{-\frac{M}{A}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(T^*)^2}{2}} dT^* \right) + A \left(-\frac{1}{\sqrt{2\pi}} e^{-\frac{(T^*)^2}{2}} \bigg|_{-\frac{M}{A}}^{\infty} \right) \\
&= M \left(1 - \Phi \left(-\frac{M}{A} \right) \right) + A \frac{1}{\sqrt{2\pi}} e^{-\frac{(\frac{M}{A})^2}{2}} \\
&= M\Phi \left(\frac{M}{A} \right) + A\phi \left(\frac{M}{A} \right).
\end{aligned}$$

Therefore, in terms of $M(t, v, X_t)$, $A^2(t, v)$, and $D(x) = x\Phi(x) + \phi(x)$, the HDD futures price for the period $0 \leq t \leq \tau_1 < \tau_2$ is given by

$$\begin{aligned}
F_{H_f}(t, \tau_1, \tau_2) &= \int_{\tau_1}^{\tau_2} M(t, v, X_t) \Phi \left(\frac{M(t, v, X_t)}{A(t, v)} \right) + A(t, v) \phi \left(\frac{M(t, v, X_t)}{A(t, v)} \right) dv \\
&= \int_{\tau_1}^{\tau_2} A(t, v) \left(\frac{M(t, v, X_t)}{A(t, v)} \Phi \left(\frac{M(t, v, X_t)}{A(t, v)} \right) + \phi \left(\frac{M(t, v, X_t)}{A(t, v)} \right) \right) dv \\
&= \int_{\tau_1}^{\tau_2} A(t, v) D \left(\frac{M(t, v, X_t)}{A(t, v)} \right) dv.
\end{aligned}$$

For the scenario $0 \leq \tau_1 \leq t < \tau_2$, based on the result from the first case and the fact that if Z is \mathcal{F}_t -measurable, $\mathbb{E}^Q(Z | \mathcal{F}_t) = Z$, we have

$$\begin{aligned}
F_H(t, \tau_1, \tau_2) &= \mathbb{E}^Q \left[\int_{\tau_1}^{\tau_2} \max(T_{\text{base}} - T_v, 0) dv \middle| \mathcal{F}_t \right] \\
&= \mathbb{E}^Q \left[\int_{\tau_1}^t \max(T_{\text{base}} - T_v, 0) dv + \int_t^{\tau_2} \max(T_{\text{base}} - T_v, 0) dv \middle| \mathcal{F}_t \right] \\
&= \int_{\tau_1}^t \max(T_{\text{base}} - T_u, 0) du + \mathbb{E}^Q \left[\int_t^{\tau_2} \max(T_{\text{base}} - T_v, 0) dv \middle| \mathcal{F}_t \right] \\
&= \int_{\tau_1}^t \max(T_{\text{base}} - T_u, 0) du + \int_t^{\tau_2} A(t, v) D \left(\frac{M(t, v, X_t)}{A(t, v)} \right) dv.
\end{aligned}$$

B.2 Proof of Proposition 2.6.2

Proof Since F_H is independent of \mathcal{F}_t , its conditional expectation given \mathcal{F}_t is equal to $\mathbb{E}^Q[F_H]$. Thus,

$$\begin{aligned}
 C_{F_H}(t, K, \tau_T) &= e^{-r(T-t)} \mathbb{E}^Q \left[\max \left(\int_{\tau_1}^{\tau_2} A(t, v) D \left(\frac{M(t, v, X_t)}{A(t, v)} \right) dv - K, 0 \right) \middle| \mathcal{F}_t \right] \\
 &= e^{-r(T-t)} \mathbb{E}^Q \left[\max(F_H - K, 0) \middle| \mathcal{F}_t \right] \\
 &= e^{-r(T-t)} \mathbb{E}^Q (\max(F_H - K, 0)) \\
 &= e^{-r(T-t)} \int_K^{F_m} (F_H - K) g(F_H) dF_H.
 \end{aligned}$$

B.3 Proof of Proposition 2.6.3

Proof It can be straightforwardly derived from the HDD formula and employing similar arguments in the proof of Proposition 2.6.2.

Appendix C

Derivations of the model's optimal parameter estimates in chapter 3

C.1 Optimal estimate for κ

Write $\kappa := (\kappa_1, \kappa_2, \dots, \kappa_N)^\top \in \mathbb{R}^N$. To update the estimates $\widehat{\kappa} = (\widehat{\kappa}_1, \widehat{\kappa}_2, \dots, \widehat{\kappa}_N)^\top \in \mathbb{R}^N$, define a new measure $P^{\widehat{\nu}^w}$ in accordance with equation (3.11) and via $\left. \frac{dP^{\widehat{\nu}^w}}{dP^{\nu^w}} \right|_{\mathcal{X}_k} = \Psi_k^{w\kappa} =$

$$\prod_{l=1}^k \varphi_l^{w\kappa}, \text{ where } \varphi_l^{w\kappa} = \frac{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} (X_l - \vartheta(\mathbf{y}_{l-1}^w) - \widehat{\kappa}(\mathbf{y}_{l-1}^w) X_{l-1})^2\right)}{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} (X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1})^2\right)}. \text{ This means that}$$

$$\log \Psi_k^{w\kappa} = \sum_{l=1}^k \left[-\frac{\widehat{\kappa}^2(\mathbf{y}_{l-1}^w) X_{l-1}^2 - 2X_l \widehat{\kappa}(\mathbf{y}_{l-1}^w) X_{l-1} + 2\vartheta(\mathbf{y}_{l-1}^w) \widehat{\kappa}(\mathbf{y}_{l-1}^w) X_{l-1}}{2\varrho^2(\mathbf{y}_{l-1}^w)} + \mathbf{R}(\kappa(\mathbf{y}_{l-1}^w)) \right],$$

where $\mathbf{R}(\kappa_t)$ is independent of $\widehat{\kappa}_t$. We consider the expectation of the log likelihood, i.e., $L(\widehat{\kappa}) = \mathbb{E}(\log \Psi_k^{w\kappa} \mid \mathcal{X}_k)$. By virtue of (3.21),

$$\begin{aligned} \mathbb{E}[\log \Psi_k^{w\kappa} \mid \mathcal{X}_k] &= \sum_{l=1}^k \mathbb{E} \left[\left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\varrho_t^2} (\widehat{\kappa}_t^2 X_{l-1}^2 \right. \right. \right. \\ &\quad \left. \left. \left. - 2X_l \widehat{\kappa}_t X_{l-1} + 2\vartheta_t \widehat{\kappa}_t X_{l-1} \right) + \mathbf{R}(\kappa_t) \right) \middle| \mathcal{X}_k \right]. \end{aligned}$$

Differentiation of $L(\widehat{\kappa})$ with respect to $\widehat{\kappa}$ and setting the resulting expression to 0, we get the optimal estimate of $\widehat{\kappa}$, given the observations X_{k+1} , as $\widehat{\kappa}_t = \frac{\widehat{\mathbf{C}}_k^t(X_{k-1}, X_k) - \vartheta_t \widehat{\mathbf{C}}_k^t(X_{k-1})}{\widehat{\mathbf{C}}_k^t(X_{k-1}^2)}$.

C.2 Optimal estimate for ϑ

Associated with the change from $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_N)^\top \in \mathbb{R}^N$ to $\widehat{\boldsymbol{\vartheta}} = (\widehat{\vartheta}_1, \widehat{\vartheta}_2, \dots, \widehat{\vartheta}_N)^\top \in \mathbb{R}^N$ is a new measure $P^{\widehat{\boldsymbol{\vartheta}}}$, based on equation (3.11), via $\frac{dP^{\widehat{\boldsymbol{\vartheta}}}}{dP^{\boldsymbol{\vartheta}}}\Big|_{\mathcal{X}_k} = \Psi_k^{w\widehat{\boldsymbol{\vartheta}}} = \prod_{l=1}^k \varphi_l^{w\widehat{\boldsymbol{\vartheta}}}$, where

$$\varphi_l^{w\widehat{\boldsymbol{\vartheta}}} = \frac{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \widehat{\vartheta}(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1}\right)^2\right)}{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1}\right)^2\right)}. \text{ Thus,}$$

$$\log \Psi_k^{w\widehat{\boldsymbol{\vartheta}}} = \sum_{l=1}^k \left(-\frac{\widehat{\vartheta}^2(\mathbf{y}_{l-1}^w) - 2X_l \widehat{\vartheta}(\mathbf{y}_{l-1}^w) + 2\widehat{\vartheta}(\mathbf{y}_{l-1}^w) \kappa(\mathbf{y}_{l-1}^w) X_{l-1}}{2\varrho^2(\mathbf{y}_{l-1}^w)} + \mathbf{R}(\vartheta(\mathbf{y}_{l-1}^w)) \right),$$

where $\mathbf{R}(\vartheta)$ is a remainder not containing $\widehat{\vartheta}_t$. Applying equations (3.19) and (3.21), the expectation of the log likelihood depending on X_k is $L(\widehat{\boldsymbol{\vartheta}}) = \mathbb{E}(\log \Psi_k^{w\widehat{\boldsymbol{\vartheta}}} | \mathcal{X}_k)$, where

$$\mathbb{E}[\log \Psi_k^{w\widehat{\boldsymbol{\vartheta}}} | \mathcal{X}_k] = \sum_{l=1}^k \mathbb{E} \left[\left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\varrho_t^2} (\widehat{\vartheta}_t^2 - 2X_t \widehat{\vartheta}_t + 2\widehat{\vartheta}_t \kappa_t X_{l-1}) \right) + \mathbf{R}(\vartheta_t) \right) \middle| \mathcal{X}_k \right].$$

We differentiate $L(\widehat{\boldsymbol{\vartheta}})$ with respect to $\widehat{\vartheta}$ and equate the result to 0. The optimal estimate $\widehat{\vartheta}$ may be derived as $\widehat{\vartheta}_t = \frac{\widehat{\mathbf{C}}_k^t(X_k) - \kappa_t \widehat{\mathbf{C}}_k^t(X_{k-1})}{\widehat{\mathbf{B}}_k^t}$.

C.3 Optimal estimate for ϱ

Consider a transformation from $\boldsymbol{\varrho} = (\varrho_1, \varrho_2, \dots, \varrho_N)^\top \in \mathbb{R}^N$ to $\widehat{\boldsymbol{\varrho}} = (\widehat{\varrho}_1, \widehat{\varrho}_2, \dots, \widehat{\varrho}_N)^\top \in \mathbb{R}^N$ using the Radon-Nikodym derivative in equation (3.11), i.e.,

$$\frac{dP^{\widehat{\boldsymbol{\varrho}}}}{dP^{\boldsymbol{\varrho}}}\Big|_{\mathcal{X}_k} = \Psi_k^{w\widehat{\boldsymbol{\varrho}}} = \prod_{l=1}^k \varphi_l^{w\widehat{\boldsymbol{\varrho}}},$$

where $\varphi_l^{w\widehat{\boldsymbol{\varrho}}} = \frac{\varrho(\mathbf{y}_{l-1}^w) \exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1}\right)^2\right)}{\widehat{\varrho}(\mathbf{y}_{l-1}^w) \exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1}\right)^2\right)}$. So,

$$\log \Psi_k^{w\widehat{\boldsymbol{\varrho}}} = \sum_{l=1}^k \left[\log \left(\frac{1}{\widehat{\varrho}(\mathbf{y}_{l-1}^w)} \right) + \left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \right) (X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1})^2 + \mathbf{R}(\varrho(\mathbf{y}_{l-1}^w)) \right].$$

where $R(\varrho_t)$ does not include $\widehat{\varrho}_t$. By equations (3.19) and (3.21),

$$\begin{aligned} \mathbb{E} \left[\log \Psi_k^{w\varrho} \mid \mathcal{X}_k \right] &= \sum_{l=1}^k \mathbb{E} \left[\sum_{t=1}^N \langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \left(\log \left(\frac{1}{\widehat{\varrho}(\mathbf{y}_{l-1}^w)} \right) \right. \right. \\ &\quad \left. \left. + \left(-\frac{1}{2\widehat{\varrho}^2(\mathbf{y}_{l-1}^w)} \right) (X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1})^2 \right) \right] + R(\varrho_t). \end{aligned}$$

We differentiate $L(\widehat{\varrho})$ with respect to $\widehat{\varrho}$ ignoring the remainder and equate the derivative to 0. The optimal estimate for parameter $\widehat{\varrho}$ is given by

$$\begin{aligned} \widehat{\varrho}_t^2 &= \frac{\widehat{\mathbf{C}}_k^t(X_k^2) + \widehat{\mathbf{B}}_k^t \vartheta_t^2 + \kappa_t^2 \widehat{\mathbf{C}}_k^t(X_{k-1}^2) + 2\widehat{\mathbf{C}}_k^t(X_{k-1}) \vartheta_t \kappa_t}{\widehat{\mathbf{B}}_k^t} \\ &\quad - 2 \frac{\widehat{\mathbf{C}}_k^t(X_k) \vartheta_t + \widehat{\mathbf{C}}_k^t(X_{k-1}, X_k) \kappa_t}{\widehat{\mathbf{B}}_k^t}. \end{aligned}$$

C.4 Optimal estimate for h

Consider a new measure $P^{\widehat{v}^w}$ defined in equation (3.31) via $\left. \frac{dP^{\widehat{v}^w}}{dP^{v^w}} \right|_{\mathcal{X}_k} = \mathbf{A}_k^h$,

where $\mathbf{A}_k^h = \prod_{l=2}^k \prod_{t,s,r=1}^N \left(\frac{\widehat{h}_{tsr}}{h_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle}$. By equation (3.18), the expectation of the log-likelihood on X_k is

$$\mathbb{E} \left[\log \mathbf{A}_k^h \mid \mathcal{X}_k \right] = \mathbb{E} \left(\sum_{l=2}^k \sum_{t,s,r=1}^N \log \left(\frac{\widehat{h}_{tsr}}{h_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle} \mid \mathcal{X}_k \right),$$

where $R(h_{tsr})$ is a remainder independent of \widehat{h}_{tsr} . With the constraint $\sum_{t=1}^N \widehat{h}_{tsr} = 1$, we introduce the Lagrange multiplier ρ and obtain the function to maximise as

$$L(\widehat{h}_{tsr}, \rho) = \sum_{t,s,r=1}^N \log \widehat{h}_{tsr} \mathbf{A}_k^{tsr} + \rho \left(\sum_{t=1}^N \widehat{h}_{tsr} - 1 \right) + R(h_{tsr}).$$

Differentiating $L(\widehat{h}_{tsr}, \rho)$ with respect to h_{tsr} and ρ and setting the derivatives to 0, we have

$$\frac{1}{\widehat{h}_{tsr}} \widehat{\mathbf{A}}_k^{tsr} + \rho = 0. \text{ Since } \sum_{t=1}^N \widehat{h}_{tsr} = 1 \text{ and } \sum_{t=1}^N \mathbf{A}_k^{tsr} = \mathbf{B}_k^{sr}, \text{ we get } \sum_{t=1}^N \widehat{h}_{tsr} = \frac{\sum_{t=1}^N \widehat{\mathbf{A}}_k^{tsr}}{-\rho} = \frac{\mathbf{B}_k^{sr}}{-\rho} =$$

1, which can be rewritten as $\sum_{t=1}^N \widehat{h}_{tsr} = \frac{\sum_{t=1}^N \widehat{\mathbf{A}}_k^{tsr}}{\widehat{\mathbf{B}}_k^{sr}}$. This means that the optimal estimate \widehat{h}_{tsr} is given by $\widehat{h}_{tsr} = \frac{\widehat{\mathbf{A}}_k^{tsr}}{\widehat{\mathbf{B}}_k^{sr}}$.

Appendix D

Derivations of the model's optimal parameter estimates in chapter 4

D.1 Optimal estimate for κ

Define the Radon-Nikodym derivative of $P^{\widehat{\mathcal{E}}^w}$ with respect to $P^{\mathcal{E}^w}$ following equation (4.15)

as $\frac{dP^{\widehat{\mathcal{E}}^w}}{dP^{\mathcal{E}^w}} \Big|_{\mathcal{F}_k} = \Psi_k^{wk} = \prod_{l=1}^k \varphi_l^{wk}$, where

$$\varphi_l^{wk} = \frac{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \widehat{\kappa}(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w)\right)^2\right)}{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w)\right)^2\right)}.$$

Thus,

$$\begin{aligned} \log \Psi_k^{wk} &= \sum_{l=1}^k \left(-\frac{\widehat{\kappa}^2(\mathbf{y}_{l-1}^w) (X_{l-1}^2 + 2\mu_{\beta(\mathbf{y}_{l-1}^w)} q X_{l-1} + (\mu_{\beta(\mathbf{y}_{l-1}^w)} q)^2)}{2\varrho^2(\mathbf{y}_{l-1}^w)} \right. \\ &\quad \left. - \frac{2\widehat{\kappa}(\mathbf{y}_{l-1}^w) (X_{l-1} + \mu_{\beta(\mathbf{y}_{l-1}^w)} q) (\vartheta(\mathbf{y}_{l-1}^w) - X_l)}{2\varrho^2(\mathbf{y}_{l-1}^w)} + \mathbf{R}(\kappa(\mathbf{y}_{l-1}^w)) \right) \\ &= \sum_{l=1}^k \left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\varrho_t^2} \left(\widehat{\kappa}^2 (X_{l-1}^2 + 2\mu_{\beta} q X_{l-1} + (\mu_{\beta} q)^2) \right. \right. \right. \\ &\quad \left. \left. \left. + 2\widehat{\kappa} (X_{l-1} + \mu_{\beta} q) (\vartheta - X_l) \right) \right) + \mathbf{R}(\kappa_t) \right), \end{aligned}$$

where $R(\kappa_t)$ is a remainder free of $\widehat{\kappa}_t$. With equations (4.21) and (4.23), we consider the expectation of the log likelihood, i.e., $L(\widehat{\kappa}) = \mathbb{E} [\log \Psi_k^{w\kappa} | \mathcal{X}_k]$, where

$$\mathbb{E} [\log \Psi_k^{w\kappa} | \mathcal{X}_k] = \sum_{l=1}^k \mathbb{E} \left[\left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\varrho_t^2} (\widehat{\kappa}^2 (X_{l-1}^2 + 2\mu_{\beta} q X_{l-1} + (\mu_{\beta} q)) \right. \right. \right. \\ \left. \left. \left. + 2\widehat{\kappa} (X_{l-1} + \mu_{\beta} q) (\vartheta - X_l) \right) + R(\kappa_t) \right) \middle| \mathcal{X}_k \right].$$

By differentiating $L(\widehat{\kappa})$ with respect to $\widehat{\kappa}$ and setting the result to 0, we get the optimal estimate $\widehat{\kappa}$, which is

$$\widehat{\kappa}_t = \frac{\widehat{\mathcal{C}}_k^t(X_{k-1}, X_k) - \vartheta_t \widehat{\mathcal{C}}_k^t(X_{k-1}) + \mu_{\beta_t} q \widehat{\mathcal{C}}_k^t(X_k) - \vartheta_t q \mu_{\beta_t} \widehat{\mathcal{B}}_k^t}{\widehat{\mathcal{C}}_k^t(X_{k-1}^2) + 2\mu_{\beta_t} q \widehat{\mathcal{C}}_k^t(X_{k-1}) + (\mu_{\beta_t} q)^2 \widehat{\mathcal{B}}_k^t}.$$

D.2 Optimal estimate for ϑ

Define a new measure $P^{\widehat{\mathcal{E}}^w}$ based on equation (4.15). Consider the construction $\frac{dP^{\widehat{\mathcal{E}}^w}}{dP^{\mathcal{E}^w}} \Big|_{\mathcal{X}_k} =$

$$\Psi_k^{w\vartheta} = \prod_{l=1}^k \varphi_l^{w\vartheta}, \text{ where}$$

$$\varphi_l^{w\vartheta} = \frac{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \widehat{\vartheta}(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w)\right)^2\right)}{\exp\left(-\frac{1}{2\varrho^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w)\right)^2\right)}.$$

So,

$$\log \Psi_k^{w\vartheta} = \sum_{l=1}^k \left(-\frac{\widehat{\vartheta}^2(\mathbf{y}_{l-1}^w) - 2X_l \widehat{\vartheta}(\mathbf{y}_{l-1}^w) + 2\kappa(\mathbf{y}_{l-1}^w) \widehat{\vartheta}(\mathbf{y}_{l-1}^w) (X_{l-1} + \mu_{\beta(\mathbf{y}_{l-1}^w)} q)}{2\varrho^2(\mathbf{y}_{l-1}^w)} \right. \\ \left. + R(\vartheta(\mathbf{y}_{l-1}^w)) \right) \\ = \sum_{l=1}^k \left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\varrho_t^2} (\widehat{\vartheta}^2 - 2X_l \widehat{\vartheta} + 2\kappa \widehat{\vartheta} (X_{l-1} + \mu_{\beta} q)) \right) \right. \\ \left. + R(\vartheta_t) \right),$$

where $R(\vartheta_t)$ is a remainder that does not contain $\widehat{\vartheta}_t$.

On the basis of equations (4.21) and (4.23), the expectation of the log-likelihood depending on X_k is $L(\widehat{\vartheta}) = \mathbb{E} \left[\log \Psi_k^{w\vartheta} \mid \mathcal{X}_k \right]$, where

$$\mathbb{E} \left[\log \Psi_k^{w\vartheta} \mid \mathcal{X}_k \right] = \sum_{l=1}^k \mathbb{E} \left[\left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\varrho_t^2} (\widehat{\vartheta}^2 - 2X_l \widehat{\vartheta} + 2\kappa \widehat{\vartheta} (X_{l-1} + \mu_\beta q)) \right) + R(\vartheta_t) \right) \middle| \mathcal{X}_k \right].$$

Solving $\frac{\partial L(\widehat{\vartheta})}{\partial \widehat{\vartheta}}$ and equating the result to 0, we get the optimal estimate

$$\widehat{\vartheta}_t = \frac{\widehat{\mathcal{C}}_k^t(X_k) - \kappa_t \widehat{\mathcal{C}}_k^t(X_{k-1}) - \kappa_t \mu_{\beta_t} q \widehat{\mathcal{B}}_k^t(X_{k-1})}{\widehat{\mathcal{B}}_k^t}.$$

D.3 Optimal estimate for ϱ

Define a new measure $P^{\widehat{\mathcal{E}}^w}$, using equation (4.15), by setting $\frac{dP^{\widehat{\mathcal{E}}^w}}{dP^{\mathcal{E}^w}} \Big|_{\mathcal{X}_k} = \Psi_k^{w\varrho} = \prod_{l=1}^k \varphi_l^{w\varrho}$,

where

$$\varphi_l^{w\varrho} = \frac{\varrho(\mathbf{y}_{l-1}^w) \exp \left(-\frac{(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w))^2}{2(\widehat{\varrho}^2(\mathbf{y}_{l-1}^w) + \sigma_{\beta(\mathbf{y}_{l-1}^w)}^2 q \kappa^2(\mathbf{y}_{l-1}^w))} \right)}{\widehat{\varrho}(\mathbf{y}_{l-1}^w) \exp \left(-\frac{(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w))^2}{2(\varrho^2(\mathbf{y}_{l-1}^w) + \sigma_{\beta(\mathbf{y}_{l-1}^w)}^2 q \kappa^2(\mathbf{y}_{l-1}^w))} \right)}.$$

From which, we have

$$\begin{aligned} \log \Psi_k^{w\varrho} &= \sum_{l=1}^k \left(\log \left(\frac{1}{\widehat{\varrho}(\mathbf{y}_{l-1}^w)} \right) + \mathbf{R}(\varrho(\mathbf{y}_{l-1}^w)) \right. \\ &\quad \left. - \frac{\left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w) \right)^2}{2 \left(\widehat{\varrho}^2(\mathbf{y}_{l-1}^w) + \sigma_{\beta(\mathbf{y}_{l-1}^w)}^2 q \kappa^2(\mathbf{y}_{l-1}^w) \right)} \right) \\ &= \sum_{l=1}^k \left(\sum_{t=1}^N \left(\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \left(\log \left(\frac{1}{\widehat{\varrho}(\mathbf{y}_{l-1}^w)} \right) \right) \right. \right. \\ &\quad \left. \left. - \frac{\left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w) \right)^2}{2 \left(\widehat{\varrho}^2(\mathbf{y}_{l-1}^w) + \sigma_{\beta(\mathbf{y}_{l-1}^w)}^2 q \kappa^2(\mathbf{y}_{l-1}^w) \right)} \right) + \mathbf{R}(\varrho_t) \right), \end{aligned}$$

where $\mathbf{R}(\varrho_t)$ is independent of $\widehat{\varrho}_t$. As per (4.21) and (4.23), we have

$$\begin{aligned} \mathbb{E} \left[\log \Psi_k^{w\varrho} \mid \mathcal{X}_k \right] &= \sum_{l=1}^k \mathbb{E} \left[\left(\sum_{t=1}^N \left(\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \left(\log \left(\frac{1}{\widehat{\varrho}(\mathbf{y}_{l-1}^w)} \right) \right) \right. \right. \right. \\ &\quad \left. \left. - \frac{1}{2 \left(\widehat{\varrho}^2(\mathbf{y}_{l-1}^w) + \sigma_{\beta(\mathbf{y}_{l-1}^w)}^2 q \kappa^2(\mathbf{y}_{l-1}^w) \right)} \right) \right. \\ &\quad \left. \times \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} \right. \right. \\ &\quad \left. \left. - \mu_{\beta(\mathbf{y}_{l-1}^w)} q \kappa(\mathbf{y}_{l-1}^w) \right)^2 \right) \Bigg| \mathcal{X}_k \right] \\ &\quad + \mathbf{R}(\varrho_t). \end{aligned}$$

Equating to 0 the mathematical derivative (with respect to $\widehat{\varrho}$) of $L(\widehat{\varrho}) = \mathbb{E} \left[\log \Psi_k^{w\varrho} \mid \mathcal{X}_k \right]$, our optimal estimate of $\widehat{\varrho}^2$ is

$$\begin{aligned} \widehat{\varrho}_t^2 &= \frac{\widehat{\mathcal{C}}_k^t(X_k^2) + \kappa_t^2 \widehat{\mathcal{C}}_k^t(X_{k-1}^2) + \widehat{\mathcal{B}}_k^t \left(\vartheta_t^2 + (\kappa_t \mu_{\beta_t} p)^2 + 2\vartheta_t \kappa_t \mu_{\beta_t} q - \sigma_{\beta_t}^2 \kappa_t^2 q \right)}{\widehat{\mathcal{B}}_k^t} \\ &\quad + \frac{\widehat{\mathcal{C}}_k^t(X_{k-1}) (2\mu_{\beta_t} q \kappa_t^2 + 2\vartheta_t \kappa_t)}{\widehat{\mathcal{B}}_k^t} \\ &\quad - \frac{(2\vartheta_t + 2\kappa_t \mu_{\beta_t} p) \widehat{\mathcal{C}}_k^t(X_k) + 2\kappa_t \widehat{\mathcal{C}}_k^t(X_{k-1} X_k)}{\widehat{\mathcal{B}}_k^t}. \end{aligned}$$

D.4 Optimal estimate for μ_β

Starting with equation (4.15), define a new measure $P^{\widehat{\mathcal{E}}^v}$ through $\frac{dP^{\widehat{\mathcal{E}}^v}}{dP^{\mathcal{E}^v}} \Big|_{\mathcal{X}_k} = \Psi_k^{w\mu_\beta} = \prod_{l=1}^k \varphi_l^{w\mu_\beta}$, where

$$\varphi_l^{w\mu_\beta} = \frac{\exp\left(-\frac{1}{2\sigma^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \widehat{\mu}_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w)\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2(\mathbf{y}_{l-1}^w)} \left(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w) X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w)\right)^2\right)}.$$

Thus,

$$\begin{aligned} \log \Psi_k^{w\mu_\beta} &= \sum_{l=1}^k \left(-\frac{\left(\widehat{\mu}_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w) - 2X_l + 2\left(\kappa(\mathbf{y}_{l-1}^w) + \vartheta(\mathbf{y}_{l-1}^w)\right)\right)}{2\sigma^2(\mathbf{y}_{l-1}^w)} \right. \\ &\quad \left. \times \left(\widehat{\mu}_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w)\right) + \mathbf{R}(\mu_{\beta(\mathbf{y}_{l-1}^w)}) \right) \\ &= \sum_{l=1}^k \left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\sigma_t^2} \widehat{\mu}_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w) \left(\widehat{\mu}_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w) - 2X_l \right. \right. \right. \\ &\quad \left. \left. \left. + 2\left(\kappa(\mathbf{y}_{l-1}^w) + \vartheta(\mathbf{y}_{l-1}^w)\right)\right) \right) \right) + \mathbf{R}(\mu_{\beta_l}), \end{aligned}$$

where $\mathbf{R}(\mu_{\beta_l})$ does not contain $\widehat{\mu}_{\beta_l}$. Again, by (4.21) and (4.23),

$$\begin{aligned} \mathbb{E} \left[\log \Psi_k^{w\mu_\beta} \mid \mathcal{X}_k \right] &= \sum_{l=1}^k \mathbb{E} \left[\left(\sum_{t=1}^N \left(-\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\sigma_t^2} \widehat{\mu}_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w) \right. \right. \right. \\ &\quad \left. \left. \left. \left(\widehat{\mu}_{\beta(\mathbf{y}_{l-1}^w)} q\kappa(\mathbf{y}_{l-1}^w) - 2X_l + 2\left(\kappa(\mathbf{y}_{l-1}^w) + \vartheta(\mathbf{y}_{l-1}^w)\right)\right) \right) \right) \mid \mathcal{X}_k \right] + \mathbf{R}(\mu_{\beta_l}). \end{aligned}$$

Following the same optimisation procedure as above,

$$\widehat{\mu}_{\beta_l} = \frac{\widehat{\mathcal{C}}_k^l(X_k) - \vartheta_l \widehat{\mathcal{B}}_k^l(X_{k-1}) - \kappa_l \widehat{\mathcal{C}}_k^l(X_{k-1})}{\kappa_l q \widehat{\mathcal{B}}_k^l(X_{k-1})}.$$

D.5 Optimal estimate for σ_β

From equation (4.15), define a new measure $P^{\widehat{\mathcal{E}}^w}$ by setting $\frac{dP^{\widehat{\mathcal{E}}^w}}{dP^{\mathcal{E}^w}} \Big|_{\mathcal{X}_k} = \Psi_k^{w\sigma_\beta} = \prod_{l=1}^k \varphi_l^{w\sigma_\beta}$, where

$$\varphi_l^{w\sigma_\beta} = \frac{\sigma_{\beta(\mathbf{y}_{l-1}^w)} \exp\left(-\frac{(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w)X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)}q\kappa(\mathbf{y}_{l-1}^w))^2}{2(\varrho^2(\mathbf{y}_{l-1}^w) + \widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)}^2)q\kappa^2(\mathbf{y}_{l-1}^w)}\right)}{\widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)} \exp\left(-\frac{(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w)X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)}q\kappa(\mathbf{y}_{l-1}^w))^2}{2(\varrho^2(\mathbf{y}_{l-1}^w) + \sigma_{\beta(\mathbf{y}_{l-1}^w)}^2)q\kappa^2(\mathbf{y}_{l-1}^w)}\right)}.$$

This yields the log likelihood

$$\begin{aligned} \log \Psi_k^{w\sigma_\beta} &= \sum_{l=1}^k \left(\log \left(\frac{1}{\widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)}} \right) \right. \\ &\quad \left. - \frac{(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w)X_{l-1} - \mu_{\beta(\mathbf{y}_{l-1}^w)}q\kappa(\mathbf{y}_{l-1}^w))^2}{2(\varrho^2(\mathbf{y}_{l-1}^w) + \widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)}^2)q\kappa^2(\mathbf{y}_{l-1}^w)} \right. \\ &\quad \left. + \mathbf{R}(\sigma_{\beta(\mathbf{y}_{l-1}^w)}) \right) \\ &= \sum_{l=1}^k \sum_{t=1}^N \left(\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \left(\log \left(\frac{1}{\widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)}} \right) \right) \right. \\ &\quad \left. - \frac{(X_l - \vartheta(\mathbf{y}_{l-1}^w) - \kappa(\mathbf{y}_{l-1}^w)X_l - \mu_{\beta(\mathbf{y}_{l-1}^w)}q\kappa(\mathbf{y}_{l-1}^w))^2}{2(\varrho^2(\mathbf{y}_{l-1}^w) + \widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)}^2)q\kappa^2(\mathbf{y}_{l-1}^w)} \right) \\ &\quad + \mathbf{R}(\sigma_{\beta_t}), \end{aligned}$$

where $\mathbf{R}(\varrho_t)$ does not depend on $\widehat{\varrho}_t$. Invoking equations (4.21) and (4.23),

$$\begin{aligned}
\mathbb{E} \left[\log \Psi_k^{w\sigma_\beta} \mid \mathcal{X}_k \right] &= \sum_{l=1}^k \mathbb{E} \left(\left(\sum_{t=1}^N \left(\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \left(\log \left(\frac{1}{\widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)}} \right) \right. \right. \right. \right. \\
&\quad \left. \left. \left. - \frac{1}{2 \left(\varrho^2 \left(\mathbf{y}_{l-1}^w \right) + \widehat{\sigma}_{\beta(\mathbf{y}_{l-1}^w)}^2 q \kappa^2 \left(\mathbf{y}_{l-1}^w \right) \right) \right) \right) \right) \Bigg| \mathcal{X}_k \right) + \mathbf{R} \left(\sigma_{\beta_t} \right) \\
&= \sum_{t=1}^N \left(\log \left(\frac{1}{\widehat{\sigma}_\beta} \right) - \left(\frac{1}{2 \left(\varrho^2 + \widehat{\sigma}_\beta^2 q \kappa^2 \right)} \right) \left(X_l - \vartheta - \kappa X_{l-1} - \mu_\beta q \kappa \right)^2 \right) \\
&\quad + \mathbf{R} \left(\sigma_{\beta_t} \right).
\end{aligned}$$

The optimal estimate for $\widehat{\sigma}_\beta^2$ is

$$\begin{aligned}
\widehat{\sigma}_{\beta_t}^2 &= \frac{\widehat{\mathcal{E}}_k^t \left(X_k^2 \right) + \kappa_t^2 \widehat{\mathcal{E}}_k^t \left(X_{k-1}^2 \right) + \widehat{\mathcal{B}}_k^t \left(\vartheta_t^2 + \left(\kappa_t \mu_{\beta_t} q \right)^2 \right)}{\widehat{\mathcal{B}}_k^t \kappa_t^2 q} \\
&\quad + \frac{\widehat{\mathcal{B}}_k^t \left(2\vartheta_t \kappa_t \mu_{\beta_t} q - \varrho_t^2 \right) + \widehat{\mathcal{E}}_k^t \left(X_{k-1} \right) \left(2\mu_{\beta_t} q \kappa_t^2 + 2\vartheta_t \kappa_t \right)}{\widehat{\mathcal{B}}_k^t \kappa_t^2 q} \\
&\quad - \frac{\left(2\vartheta_t + 2\kappa_t \mu_{\beta_t} q \right) \widehat{\mathcal{E}}_k^t \left(X_k \right) + 2\kappa_t \widehat{\mathcal{E}}_k^t \left(X_{k-1} X_k \right)}{\widehat{\mathcal{B}}_k^t \kappa_t^2 q}.
\end{aligned}$$

D.6 Optimal estimate for p_{tsr}

To define a new measure $P^{\widehat{\mathcal{E}}^w}$ described in equation (4.32), consider the Radon-Nikodym derivative $\left. \frac{dP^{\widehat{\mathcal{E}}^w}}{dP^{\mathcal{E}^w}} \right|_{\mathcal{X}_k} = \Lambda_k^p$, where

$$\Lambda_k^p = \prod_{l=2}^k \prod_{t,s,r=1}^N \left(\frac{\widehat{p}_{tsr}}{p_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle}.$$

With the aid of equation (4.20),

$$\begin{aligned}
\mathbb{E} \left[\log \Lambda_k^p \mid \mathcal{X}_k \right] &= \mathbb{E} \left[\sum_{l=2}^k \sum_{t,s,r=1}^N \log \left(\frac{\widehat{p}_{tsr}}{p_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle} \mid \mathcal{X}_k \right] \\
&= \mathbb{E} \left[\left(\sum_{l=2}^k \sum_{t,s,r=1}^N (\log \widehat{p}_{tsr} - \log p_{tsr}) \langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle \right) \mid \mathcal{X}_k \right] \\
&= \mathbb{E} \left[\sum_{t,s,r=1}^N \log \widehat{p}_{tsr} \mathcal{A}_k^{tsr} \mid \mathcal{X}_k \right] + \mathbf{R}(p_{tsr}),
\end{aligned}$$

where $\mathbf{R}(p_{tsr})$ is a \widehat{p}_{tsr} -free expression. With the constraint $\sum_{t=1}^N \widehat{p}_{tsr} = 1$, we introduce the Lagrange multiplier ρ and consider the function

$$L(\widehat{p}_{tsr}, \rho) = \sum_{t,s,r=1}^N \log \widehat{p}_{tsr} \mathcal{A}_k^{tsr} + \rho \left(\sum_{t=1}^N \widehat{p}_{tsr} - 1 \right) + \mathbf{R}(p_{tsr}).$$

Differentiating $L(\widehat{p}_{tsr}, \rho)$ with respect to p_{tsr} and ρ , and then setting the resulting mathematical derivatives to 0, we have

$$\frac{1}{\widehat{p}_{tsr}} \mathcal{A}_k^{tsr} + \rho = 0.$$

Given that $\sum_{t=1}^N \widehat{p}_{tsr} = 1$ and $\sum_{t=1}^N \mathcal{A}_k^{tsr} = \mathcal{B}_k^{sr}$, we have $\sum_{t=1}^N \widehat{p}_{tsr} = \frac{\sum_{t=1}^N \mathcal{A}_k^{tsr}}{-\rho} = \frac{\mathcal{B}_k^{sr}}{-\rho} = 1$, and consequently, $\sum_{t=1}^N \widehat{p}_{tsr} = \frac{\sum_{t=1}^N \mathcal{A}_k^{tsr}}{\mathcal{B}_k^{sr}}$. Henceforth,

$$\widehat{p}_{tsr} = \frac{\mathcal{A}_k^{tsr}}{\mathcal{B}_k^{sr}}.$$

Appendix E

Derivation of the model's optimal parameter estimates in chapter 5

E.1 Optimal estimate for ν^h

Following the EM estimation given in Xi and Mamon [32], we define the Radon-Nikodym derivative of $P^{\widehat{\Theta}^w}$ with respect to P^{Θ^w} as $\frac{dP^{\widehat{\Theta}^w}}{dP^{\Theta^w}} \Big|_{\mathcal{F}_k} = \Psi_k^{w\nu^h} = \prod_{l=1}^k \varphi_l^{w\nu^h}$, where

$$\varphi_l^{w\nu^h} = \frac{\exp\left(-\frac{1}{2(\sigma^h)^2(\mathbf{y}_{l-1}^w)} \left(G_l - \widehat{\nu}^h(\mathbf{y}_{l-1}^w) - G_{l-1}\right)^2\right)}{\exp\left(-\frac{1}{2(\sigma^h)^2(\mathbf{y}_{l-1}^w)} \left(G_l - \nu^h(\mathbf{y}_{l-1}^w) - G_{l-1}\right)^2\right)}.$$

Thus,

$$\begin{aligned} \log \Psi_k^{w\nu^h} &= \sum_{l=1}^k \frac{\widehat{\nu}^h(\mathbf{y}_{l-1}^w) (2G_l - 2G_{l-1} - \widehat{\nu}^h(\mathbf{y}_{l-1}^w))}{2(\sigma^h(\mathbf{y}_{l-1}^w))^2} + \mathbf{R}(\nu^h(\mathbf{y}_{l-1}^w)) \\ &= \sum_{l=1}^k \sum_{t=1}^N \left(\frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2(\sigma_t^h)^2} \widehat{\nu}^h(\mathbf{y}_{l-1}^w) (2G_l - 2G_{l-1} - \widehat{\nu}^h(\mathbf{y}_{l-1}^w)) + \mathbf{R}(\nu_t^h) \right), \end{aligned}$$

where $\mathbf{R}(\nu_t^h)$ is a $\widehat{\nu}_t^h$ -free expression. With equations (5.20) and (5.22), we consider the expectation of the log likelihood, i.e., $L(\widehat{\nu}^h) = \mathbb{E}[\log \Psi_k^{w\nu^h} \mid \mathcal{F}_k]$, where

$$\mathbb{E}[\log \Psi_k^{w\nu^h} \mid \mathcal{F}_k] = \sum_{l=1}^k \mathbb{E} \left[\sum_{t=1}^N \frac{\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle}{2\sigma_t^2} \widehat{\nu}^h(\mathbf{y}_{l-1}^w) (2G_l - 2G_{l-1} - \widehat{\nu}^h(\mathbf{y}_{l-1}^w)) \Big| \mathcal{F}_k \right] + \mathbf{R}(\nu_t^h).$$

By differentiating $L(\widehat{\boldsymbol{v}}^h)$ with respect to $\widehat{\boldsymbol{v}}^h$ and setting the result to 0, we get

$$\widehat{\boldsymbol{v}}_t^h = \frac{\widehat{\mathcal{C}}_k^t}{\widehat{\mathcal{B}}_k^t} = \frac{\zeta_k^w \left(\mathcal{C}_k^t(f^h) \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \right)}{\zeta_k^w \left(\mathcal{B}_k^t \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \right)}.$$

E.2 Optimal estimate for σ^h

Define a new measure $P^{\widehat{\Theta}^w}$ via $\frac{dP^{\widehat{\Theta}^w}}{dP^{\Theta^w}} \Big|_{\mathcal{F}_k} = \Psi_k^{w\sigma^h} = \prod_{l=1}^k \varphi_l^{w\sigma^h}$, where

$$\varphi_l^{w\sigma^h} = \frac{\sigma^h(\mathbf{y}_{l-1}^w) \exp\left(-\frac{(G_l - \nu^h(\mathbf{y}_{l-1}^w) - G_{l-1})^2}{2(\widehat{\sigma}^h(\mathbf{y}_{l-1}^w))^2}\right)}{\widehat{\sigma}^h(\mathbf{y}_{l-1}^w) \exp\left(-\frac{(G_l - \nu^h(\mathbf{y}_{l-1}^w) - G_{l-1})^2}{2(\sigma^h(\mathbf{y}_{l-1}^w))^2}\right)}.$$

Consequently,

$$\begin{aligned} \log \Psi_k^{w\sigma^h} &= \sum_{l=1}^k \left(\log \frac{1}{\widehat{\sigma}^h(\mathbf{y}_{l-1}^w)} - \frac{(G_l - \nu^h(\mathbf{y}_{l-1}^w) - G_{l-1})^2}{2(\widehat{\sigma}^h(\mathbf{y}_{l-1}^w))^2} + R(\sigma^h(\mathbf{y}_{l-1}^w)) \right) \\ &= \sum_{l=1}^k \sum_{t=1}^N \left(\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \left(\log \frac{1}{\widehat{\sigma}^h(\mathbf{y}_{l-1}^w)} - \frac{(G_l - \nu^h(\mathbf{y}_{l-1}^w) - G_{l-1})^2}{2(\widehat{\sigma}^h(\mathbf{y}_{l-1}^w))^2} \right) \right) + R(\sigma_t^h), \end{aligned}$$

where $R(\sigma_t^h)$ is independent of $\widehat{\sigma}_t^h$. From equations (5.20) and (5.22), we have

$$\mathbb{E} \left[\log \Psi_k^{w\sigma^h} \mid \mathcal{F}_k \right] = \sum_{l=1}^k \mathbb{E} \left[\sum_{t=1}^N \left(\langle \mathbf{y}_{l-1}^w, \mathbf{e}_t \rangle \left(\log \frac{1}{\widehat{\sigma}^h(\mathbf{y}_{l-1}^w)} - \frac{(G_l - \nu^h(\mathbf{y}_{l-1}^w) - G_{l-1})^2}{2(\widehat{\sigma}^h(\mathbf{y}_{l-1}^w))^2} \right) \right) \Big| \mathcal{F}_k \right] + R(\sigma_t^h).$$

When the above expression is differentiated with respect to $\widehat{\sigma}^h$ and the resulting derivative is set to 0, the optimal estimate is

$$\widehat{\sigma}_t^h = \left(\frac{\zeta_k^w \left(\mathcal{C}_k^t(f^h)^2 \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \right) - 2f_t^h \zeta_k^w \left(\mathcal{C}_k^t(f^h) \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \right)}{\zeta_k^w \left(\mathcal{B}_k^t \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \right)} + \frac{(f_t^h)^2 \zeta_k^w \left(\mathcal{B}_k^t \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \right)}{\zeta_k^w \left(\mathcal{B}_k^t \eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) \right)} \right)^{\frac{1}{2}}.$$

E.3 Optimal estimate for p_{tsr}

To define a new measure $P^{\widehat{\Theta}^v}$, consider the Radon-Nikodym derivative $\left. \frac{dP^{\widehat{\Theta}^v}}{dP^{\Theta^w}} \right|_{\mathcal{F}_k} = \Lambda_k^p$, where

$$\Lambda_k^p = \prod_{l=2}^k \prod_{t,s,r=1}^N \left(\frac{\widehat{p}_{tsr}}{p_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle}.$$

By equation (5.19),

$$\begin{aligned} \mathbb{E} \left[\log \Lambda_k^p \mid \mathcal{F}_k \right] &= \mathbb{E} \left[\sum_{l=2}^k \sum_{t,s,r=1}^N \log \left(\frac{\widehat{p}_{tsr}}{p_{tsr}} \right)^{\langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle} \mid \mathcal{F}_k \right] \\ &= \mathbb{E} \left[\left(\sum_{l=2}^k \sum_{t,s,r=1}^N (\log \widehat{p}_{tsr} - \log p_{tsr}) \langle \mathbf{y}_{l-2}^w, \mathbf{e}_r \rangle \langle \mathbf{y}_{l-1}^w, \mathbf{e}_s \rangle \langle \mathbf{y}_l^w, \mathbf{e}_t \rangle \right) \mid \mathcal{F}_k \right] = \mathbb{E} \left[\sum_{t,s,r=1}^N \log \widehat{p}_{tsr} \mathcal{A}_k^{tsr} \mid \mathcal{F}_k \right] + \mathbf{R}(p_{tsr}), \end{aligned}$$

where $\mathbf{R}(p_{tsr})$ is a remainder free of \widehat{p}_{tsr} . The constraint $\sum_{t=1}^N \widehat{p}_{tsr} = 1$ is dealt with by introducing a Lagrange multiplier ϱ and the function

$$L(\widehat{p}_{tsr}, \varrho) = \sum_{t,s,r=1}^N \log \widehat{p}_{tsr} \mathcal{A}_k^{tsr} + \varrho \left(\sum_{t=1}^N \widehat{p}_{tsr} - 1 \right) + \mathbf{R}(p_{tsr}).$$

Differentiating $L(\widehat{p}_{tsr}, \varrho)$ with respect to p_{tsr} and ϱ , and then setting the resulting mathematical derivatives to 0, we obtain

$$\frac{1}{\widehat{p}_{tsr}} \mathcal{A}_k^{tsr} + \varrho = 0.$$

Noting that $\sum_{t=1}^N \widehat{p}_{tsr} = 1$ and $\sum_{t=1}^N \mathcal{A}_k^{tsr} = \mathcal{B}_k^{sr}$, we then have $\sum_{t=1}^N \widehat{p}_{tsr} = \frac{\sum_{t=1}^N \mathcal{A}_k^{tsr}}{-\varrho} = \frac{\mathcal{B}_k^{sr}}{-\varrho} = 1$. Hence,

$$\widehat{p}_{tsr} = \frac{\mathcal{A}_k^{tsr}}{\mathcal{B}_k^{sr}}.$$

Appendix F

Derivation of the d -step ahead forecasting formula in chapter 5

From equation (5.14), the conditional expectation of the new Markov chain $\eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w)$, given information up to time k , is

$$\mathbb{E}[\eta(\mathbf{y}_{k+1}^w, \mathbf{y}_k^w) \mid \mathcal{F}_k] = \mathbf{B}\eta(\mathbf{y}_k^w, \mathbf{y}_{k-1}^w) = \mathbf{B}\mathbf{q}_k. \quad (\text{F.1})$$

Since all non-zero entries in \mathbf{B} are elements of \mathbf{P} , equation (F.1) gives $\mathbb{E}[\mathbf{y}_{k+1}^w \mid \mathcal{F}_k] = \mathbf{P}\mathbf{q}_k$. The construction of \mathbf{P} along with equation (5.15) leads to

$$\mathbb{E}[\eta(\mathbf{y}_{k+d}^w, \mathbf{y}_{k+d-1}^w) \mid \mathcal{F}_k] = \mathbf{B}^d \mathbf{q}_k, \quad (\text{F.2})$$

and from equation (F.2),

$$\mathbb{E}[\mathbf{y}_{k+d}^w \mid \mathcal{F}_k] = \mathbf{P}\mathbf{q}_{k+d-1} = \mathbf{P}\mathbf{B}^{d-1} \mathbf{q}_k. \quad (\text{F.3})$$

Employing $G_{k+d}^h - G_k^h = \log \frac{F_{k+d}^h}{F_k^h}$, and equations (5.10) and (F.3),

$$\mathbb{E} \left[\log \frac{F_{k+d}^h}{F_k^h} \mid \mathcal{F}_k \right] = \langle \mathbf{v}^h, \mathbb{E}[\mathbf{y}_{k+d}^w \mid \mathcal{F}_k] \rangle = \langle \mathbf{v}^h, \mathbf{P}\mathbf{B}^{d-1} \mathbf{q}_k \rangle. \quad (\text{F.4})$$

As per equation (5.10), $\log \frac{F_{k+d}^h}{F_k^h}$ is characterised by a mixture of normal distributions having the density function $\sum_{j,i=1}^N \langle \mathbf{q}_{k+d-1}, \mathbf{e}_{ji} \rangle \phi(G; \nu_j^h, \sigma_j^h)$. The 1-step ahead forecasts for F_{k+1}^h is then

$$\mathbb{E} [F_{k+1}^h \mid \mathcal{F}_k] = F_k^h \sum_{j,i=1}^N \langle \mathbf{q}_k, \mathbf{e}_{ji} \rangle \exp \left(\nu_j^h + \frac{(\sigma_j^h)^2}{2} \right). \quad (\text{F.5})$$

For $d = 2$, assume that $F_{k+1}^h = \mathbb{E}[F_{k+1}^h | \mathcal{F}_k]$. Thus,

$$\mathbb{E}[F_{k+2}^h | \mathcal{F}_k] = F_k^h \prod_{l=1}^2 \sum_{j,i=1}^N \langle \mathbf{B}^{l-1} \mathbf{q}_k, \mathbf{e}_{ji} \rangle \exp \left(v_j^h + \frac{(\sigma_j^h)^2}{2} \right). \quad (\text{F.6})$$

The d -step ahead forecasts for F_{k+d}^h can be obtained straightforwardly through

$$\mathbb{E}[F_{k+d}^h | \mathcal{F}_k] = F_k^h \prod_{l=1}^d \sum_{j,i=1}^N \langle \mathbf{B}^{l-1} \mathbf{q}_k, \mathbf{e}_{ji} \rangle \exp \left(v_j^h + \frac{(\sigma_j^h)^2}{2} \right) \quad (\text{F.7})$$

by the principle of mathematical induction.

Curriculum Vitae

Name: Heng Xiong

Post-Secondary Education and Degrees: Ph.D. in Statistics (Financial Modelling), 2014 - 2018

University of Western Ontario, London, ON

Degrees:

M.Sc. in Statistics (Financial Modelling), 2011 - 2014

University of Western Ontario, London, ON

M.Sc. in Engineering Business Management, 2005 - 2006

University of Warwick, Coventry, UK

B.Sc. in Electronic Science and Technology, 2001 - 2005

Huazhong University of Science and Technology, P.R. China

Honours and Awards: Western Graduate Research Scholarship, 2011 - 2018

Ontario Graduate Scholarship, 2017 - 2018

Graduate Student Teaching Award, 2012 - 2013

Related Work Experience: Research and Teaching Assistant, 2011 - 2018

University of Western Ontario

Publications:

[1] H. Xiong, R. Mamon, A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics, *Journal of Computational Science*, 17 (2016) 47–61.

[2] H. Xiong, R. Mamon, Putting a price tag on temperature, *Computational Management Science*, in press (2017). DOI: 10.1007/s10287-017-0291-8.

[3] H. Xiong, R. Mamon, A higher-order Markov chain-modulated model for electricity spot-price dynamics, submitted to *Applied Energy*, 2017.

[4] H. Xiong, R. Mamon, Modelling and forecasting futures-prices curves in the Fish Pool market, submitted to *Journal of Economic Dynamics and Control*, 2017.