8-5-2014 12:00 AM

Estimation of Hidden Markov Models and Their Applications in Finance

Anton Tenyakov, *The University of Western Ontario*

Supervisor: Professor Rogemar Mamon, *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

ESTIMATION OF HIDDEN MARKOV MODELS AND THEIR APPLICATIONS IN FINANCE

(Thesis format: Integrated-Article)

by

Anton Tenyakov

Graduate Program in Statistics and Actuarial Science

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

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

Movements of financial variables exhibit extreme fluctuations during periods of economic crisis and times of market uncertainty. They are also affected by institutional policies and intervention of regulatory authorities. These structural changes driving prices and other economic indicators can be captured reasonably by models featuring regime-switching capabilities. Hidden Markov models (HMM) modulating the model parameters to incorporate such regime-switching dynamics have been put forward in recent years, but many of them could still be further improved. In this research, we aim to address some of the inadequacies of previous regime-switching models in terms of their capacity to provide better forecasts and efficiency in estimating parameters. New models are developed, and their corresponding filtering results are obtained and tested on financial data sets.

The contributions of this research work include the following: (i) Recursive filtering algorithms are constructed for a regime-switching financial model consistent with no-arbitrage pricing. An application to the filtering and forecasting of futures prices under a multivariate set-up is presented. (ii) The modelling of risk due to market and funding liquidity is considered by capturing the joint dynamics of three time series (Treasury-Eurodollar spread, VIX and S&P 500 spread-derived metric), which mirror liquidity levels in the financial markets. HMM filters under a multi-regime mean-reverting model are established. (iii) Kalman filtering techniques and the change of reference probability-based filtering methods are integrated to obtain hybrid algorithms. A pairs trading investment strategy is supported by the combined power of both HMM and Kalman filters. It is shown that an investor is able to benefit from the proposed interplay of the two filtering methods. (iv) A zero-delay HMM is devised for the evolution of multivariate foreign exchange rate data under a high-frequency trading environment.

Recursive filters for quantities that are functions of a Markov chain are derived, which in turn provide optimal parameter estimates. (v) An algorithm is designed for the efficient calculation of the joint probability function for the occupation time in a Markov-modulated model for asset returns under a general number of economic regimes. The algorithm is constructed with accessible implementation and practical considerations in mind.

Keywords: Markov chain, change of measure, multivariate HMM filtering, oil future prices, Ornstein-Uhlenbeck process, liquidity, TED spread, VIX, financial distress, pairs trading, high-frequency data, regime-switching algorithms

Co-Authorship Statement

I hereby declare that this thesis incorporates materials that are direct results of my main efforts.

The content of chapter 2 was used as a basis of a full paper (co-authored with Dr. Rogemar Mamon and Dr. Paresh Date), which was published in the journal *Energy Economics*.

Chapter 3 is a modified version of the paper (co-authored with Dr. Rogemar Mamon and Dr. Matt Davison) submitted for publication in the *International Journal of Forecasting*.

The research results from chapter 4 were taken from a manuscript (co-authored with my supervisor) that is about to be submitted to the journal *Quantitative Finance*.

The source of chapter 5 is an article that is currently being finalised to undergo peer review in the journal *Annals of Operations Research*.

Chapter 6 was converted to a short paper for submission to *Systems and Control Letters*.

Please note that this thesis employed an integrated-article format following Western's thesis guidelines. This means that each chapter can be read independently as it does not rely on other chapters. Every chapter is deemed to be self-contained and can stand on its own.

With the exception of guidance on modelling framework formulations and occasional suggestions on numerical experiments from my supervisor, as well as inputs from my supervisor's collaborators, Drs. Date and Davison, I

certify that this document is a product of my own work. This research was conducted from May 2011–present under the supervision of Dr. Rogemar Mamon at the University of Western Ontario.

London, Ontario

This thesis is dedicated to my parents and sister.
Thank you for your support, love and encouragement.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Rogemar Mamon, for his continued support, undiminished motivation and enthusiasm, immense knowledge, and most especially, his long-lasting patience. His guidance has sustained the steady progress of my doctoral research.

I take this opportunity to thank Dr. Matt Davison for his invaluable help on the two research projects that became part of this thesis, and for his insightful ideas and suggestions on some aspects of my professional endeavours.

I am also grateful for the advice I received from Dr. Alexey Kuznetsov during the time I studied for my Master's degree at York University. His leading me to make the right choice and his occasional follow-through steering me to my PhD completion are much appreciated.

I acknowledge the financial support provided by the Department of Statistical and Actuarial Sciences, Western University; the Ontario Graduate Scholarship; and the Queen Elizabeth II Scholarship Program.

Finally, I convey my thanks to all my friends and colleagues who provided great company for all these years I spent in London. Thank you for making academic life bearable and for the great memories that will never be forgotten.

Contents

List of Tables

List of Figures

List of Appendices

1

HMM and quantitative finance

1.1 Introduction

A hidden Markov model (HMM) refers to a statistical model designed to capture the hidden states of a system and their evolution, which are governed by a Markov process. An HMM may be formulated in a simple state-space form, and much of the earlier works in this area focused on solving the problem of nonlinear optimization with the utility of forward-backward algorithms. Owing to recent advances, our implementation of the HMM filtering in this thesis, via the change of probability measures, will be carried out using the quicker forward-only-driven algorithms, filters. Before mathematical finance became an established area, pioneering developments and applications of HMM have occurred in engineering, speech recognition, image processing, and the fields of biological and physical sciences.

The motivation of using HMM to address problems in financial modelling can be clearly understood on the basis of signal estimation, which is popular in engineering. For instance, an engineer may be interested in determining at a given time the charge at a fixed point in an electric circuit. However, due to errors in measurement or some other underlying unknwon factors, the charge cannot really be measured but rather just a noisy version of it. Our goal is to "filter" the noise out from the series of observed values in the best possible way. In finance, we take prices and series of financial data and economic indicators as given. We wish to determine if such observed data do contain information about latent or hidden variables, and if so, how would we estimate their dynamics. See further Mamon and Elliott [45] or Mamon et al. [46].

Needless to say, the efficient estimation of financial variable's dynamics and accurate estimation of parameters have tremendous impact to the valuation of derivatives, risk management, and asset allocation, among other financial modelling endeavours. The object of this research is to augment further developments that could add value to engineering and other allied fields, and applications of HMMs in finance providing easy access for implementation in the industry. Specifically, we explore further extensions of HMM-based models together with filtering methods via measure change when estimating parameters. The extensions and developments are application-specific and we describe them below.

Remarks

1. *It has to be noted that the filtering algorithms for HMM adopted in this research make use of change of probability measures. It is important to emphasise that, although similar in principle, this has nothing to do with changing measure from real-world probability to risk-neutral probability employed in derivative pricing.*

2. *The change of measure in this research is to demonstrate that we can facilitate the calculation of filters under some mathematically "idealised world" (i.e., the new measure) under which the observations are independent and identically distributed random variables.*

3. *Fubini's theorem then applies and allows the interchanges of expectations and summations. Calculations are then related back to the real-world measure via an inverse change of measure through the construction of an appropriate Radon-Nikodým derivative, which is a discrete-time version of the Girsanov theorem.*

1.2 Research problems examined in this thesis

1.2.1 Filtering and forecasting commodity futures prices under an HMM framework

In this work, we propose a model for the evolution of arbitrage-free futures prices under a regime-switching framework. The estimation of model parameters is carried out using the hidden Markov filtering algorithms. Comprehensive numerical experiments on real financial market data are provided to illustrate the effectiveness of our algorithm. In particular, the model is calibrated with data from heat oil futures and its forecasting performance as well as statistical validity is investigated. The proposed model is parsimonious, self-calibrating and can be useful in predicting futures prices.

1.2.2 Filtering of an HMM-driven multivariate Ornstein-Uhlenbeck model with application to forecasting market liquidity

Following Boudt et al. [5], it is interesting to note that the T-bill ED (TED) spread is directly correlated with market stability. TED is calculated as the difference between the interest rates linked on interbank loans and on short-term US T-bills. Currently, its computation makes use of three-month LIBOR and the three-month T-bill yield rates. An increasing TED spread usually portends a meltdown of a stock market as it is taken as a sign of liquidity withdrawal. As described in Bloomberg, the TED spread can gauge perceived credit risk in the general economy since T-bills are risk-free instruments and the credit risk of lending to commercial banks is encapsulated in LIBOR. The rising of TED spread indicates that lenders view default counterparty risk to be rising as well. Thus, lenders would require a higher rate of interest or settle for lower returns on safer instruments such as T-bills. When the default risk of banks is decreasing, TED spread is falling; see Krugman [40].

We investigate the modelling of risk due to market and funding liquidity by capturing the joint dynamics of three time series: the Treasury-Eurodollar spread, the VIX and a metric derived from the S&P 500 spread. We propose a two-regime mean-reverting model in explaining the behavior of three time series, which mirror liquidity levels for financial markets. An expectation-maximisation algorithm in conjunction with multivariate filters is employed to construct optimal parameter estimates of the proposed

model. The selection of the modelling set-up is justified by balancing the best-fit criterion and model complexity. The model performance is demonstrated on historical market data by producing accurate prediction of market illiquidity states.

1.2.3 Hybrid filters

Kalman-type filters played an important role in mathematical finance; see Date and Ponomareva [10]. However, a major drawback of these filters is that the parameters of the model such as drift, volatility, mean-reversion parameter, transition matrix, etc. have to be either already known or estimated separately. Once the model parameters are estimated or imposed, they usually remain static for a long time, and what is updated is simply the estimate of the state of the underlying variable in question. This is not a serious concern in electrical engineering and physics as the systems in these areas are quite stable and therefore, the parameters are bound not to change very drastically. However, it is a different story in finance where dynamic switching is necessary to deal with extreme market forces and effects of financial crises, business cycles and unanticipated events brought by human activities and sentiments.

To enhance the capabilities of Kalman filter, it is possible to combine other filters with it in estimating parameters recursively. Nonetheless, this is far from ideal since previous formulations (cf. Logothetis and Krisnamurthy [43]) demonstrated that applications of the filters have to be done one after another, which is tantamount to performing a consecutive application of two separate modelling for the same data. Such approach has two major disadvantages: (i) it is quite slow and does not have the speed near that of the two approaches combined, and (ii) its aggregated error is high if one has to work with a huge quantity of data.

We expand the idea on automated pair trading approach proposed by Elliott et al. [18] by synthesising the Kalman and multi-regime dynamic filters. This provides a powerful tool in capturing the heavy-tailed distribution mean-reverting processes. An expectation-maximisation algorithm jointly with Kalman filtering produces the most efficient parameter estimation for the succeeding trading procedure. We address the practitioners' primary concern of filtering implementation. The performance of the algorithms is evaluated on the historical spread between Coca-Cola Company and PepsiCo Inc. indices.

1.2.4 Zero-delay HMM and high-frequency data

High-frequency trading (HFT) started to emerge in the 1990s in response to advances in computer technology and their adoption by the exchanges. From the original rudimentary order processing to the current state-of-the-art all-inclusive trading systems, HFT has evolved into a billion-dollar industry. As the foreign exchange (FX) market is liquid, thousands of transaction ticks are generated per business day. Data vendors like Reuters transmit more than 275,000 prices per day on FX spot rates, Dacorogna et al. [9].

As FX derivatives (e.g., futures and swaps) are one of the most highly frequent traded assets in the market, there is a strong demand for good predictions of the future movement of the FX rates as well as the estimation of FX volatilities. Indeed, FX rates are very volatile and not easy to model and predict. Many successful HFT strategies run on FX, equities, futures, and derivatives, see Aldridge [2]. It is documented in Cheung [8] for example, that from time to time FX rates exhibit spikes or regime changes.

We develop a zero-delay HMM to capture the evolution of multivariate FX rate data under a frequent trading environment. Recursive filters for the Markov chain and pertinent quantities are derived, and subsequently employed to obtain estimates for model parameters. The rationale of zero-delay HMM hinges on the idea that with fast trading, available information must be incorporated immediately in the evolution equations of the financial variables being modelled. Our proposed model is compared with the usual one-step delay HMM, GARCH and random walk models using likelihood-based criteria and error-type metrics. Parameter estimation both under the static and dynamic settings are carried out as well as in the models used as benchmarks in a comparative analysis. Implementation details are provided. We include a numerical illustration of the methodology applied to the currency data on UK sterling pounds and US dollars both against the Japanese yen. Our empirical results demonstrate greater fitting capacity and forecasting power of the zero-delay HMM over the comparators included in our analysis.

Static filtering algorithms have been available to academic and industry users for at

least a decade. Market risk managers have been applying these algorithms to estimate quantiles of portfolio distribution under Markov-switching model with typically 2 regimes. We develop an extension of the approach proposed by Hardy [34] and show how to numerically construct a density function for total asset's return distribution in a more complex environment, i.e. when an underlying HMM has more than two states. Our technique is straightforward and can easily be applied to various risk measure estimations.

1.3 Overview of hidden Markov models (HMMs)

In this section, we introduce the assumptions and concepts relevant to the theory of HMM. We hint at the versatility of the class of Markov models, which is a powerful machinery in modelling a variety of real-world phenomena. The discussion of simple HMM frameworks are given and extended to provide insights on how the estimation of parameters is carried out.

1.3.1 What is an HMM?

Consider a Markov chain $\{X_k\}$, where k is a non-negative integer. Suppose $\{X_k\}$ embedded in signals corrupted by some noise. Indeed, $\{X_k\}$ is hidden due to noise and not observable in practice. The Markov chain is often assumed to take values on a finite set, but this can be relaxed in general, allowing for an arbitrary state space. What is observed in the market or real world is a process $\{Y_k, k \geq 0\}$, which is a function of $\{X_k\}$. The time series $\{Y_k\}$ is a distorted version of $\{X_k\}$ due to some noise assumed to have a distribution, say, Gaussian or Poisson. The process $\{Y_k\}$ is a series of signals containing the "true" state or regime of an economy.

Following the formulation in Cappe et al. [7], an HMM is a bivariate discrete process $\{X_k, Y_k\}$, where $\{X_k\}$ is a Markov chain, and conditional on $\{X_k\}$, $\{Y_k\}$ is a sequence of independent random variables. The conditional distribution of Y_k only depends on X_k. The respective state spaces of $\{X_k\}$ and $\{Y_k\}$ are denoted by $\mathbf{X}$ and $\mathbf{Y}$.

From Figure 1.1, the distribution of the variable X_{k+1} given the whole information $(X_0, X_1, ..., X_{k-1}, X_k) \equiv \{\mathcal{X}\}$ until time k depends only on the value of X_k. Similarly,

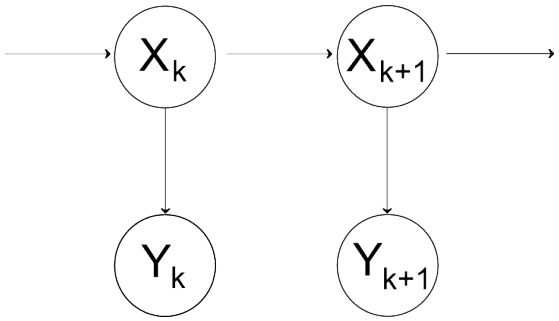


Figure 1.1: A depiction of the structure of a simple HMM, where $\{X_k\}$ is a hidden Markov chain and $\{Y_k\}$ is an observation process

the distribution of Y_{k+1} conditional on $Y_0, Y_1, ..., Y_{k-1}, Y_k$ and $X_0, X_1, ..., X_{k-1}, X_k$ depends only on X_k. Any HMM can be defined through a functional representation known as a general state-space model:

$$X_{k+1} = a(X_k, U_k) \tag{1.1}$$

$$Y_k = b(X_k, V_k), \tag{1.2}$$

where $\{U_k\}$ and $\{V_k\}$ are independent sequences of random variables that are independent of the initial distribution of X_0; a and b are measurable functions. It is considered that $\{U_k\}$ and $\{V_k\}$ belong to the same family of distributions such as Gaussian, Gamma, Poisson, etc; see Cappe et al. [7]; Elliott et al. [15]; Mamon et al. [45]; and Hamilton and Raj [27]. Such distributions are typically assumed for two reasons: computational convenience and ease of interpretation of the model. It is important to realize that the Markov property assumption is restrictive and it is not warranted in certain applications. This leads to the consideration of the so-called higher-order Markov chains or weak Markov chains described in Siu et al. [51, 52]; Xi and Mamon [55, 56]. In this case, the process $\{X_k\}$ is a weak Markov chain of order $n \geq 1$, if its value at the present time k depends on its value in the previous n time steps. More formally, this means that we have

$$X_{k+1} = a(X_k, ...X_{k-n+1}, U_k) \tag{1.3}$$

$$Y_k = b(X_k, ...X_{k-n+1}, V_k), \tag{1.4}$$

where $\{U_k\}$ and $\{V_k\}$ are random noises.

A natural extension of the HHM is to consider self-dependence of the emission process

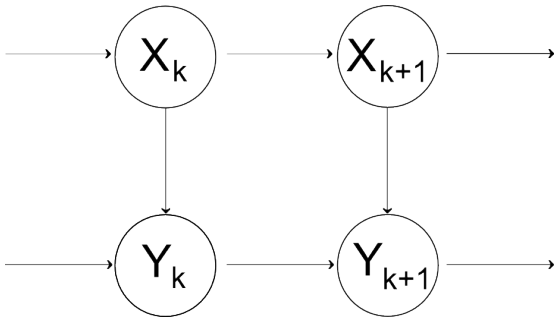


Figure 1.2: Discrete HMM with a dependence among emissions. A depiction of a structure of an emission-dependent HMM, where $\{X_k\}$ is a hidden Markov chain and $\{Y_k\}$ is an observation process

(cf. Cappe et al. [7]), another name for the observation process $\{Y_k\}$, which is mostly used in mathematical biology, genetics and computer science. This new model is illustrated in Figure 1.2. Under the state-space form representation, the model can be written as

$$X_{k+1} = a(X_k, U_k) \tag{1.5}$$

$$Y_{k+1} = b(X_{k+1}, Y_k, V_{k+1}). \tag{1.6}$$

Hamilton ([25] and [26]) coined the term "Markov-switching model" for such a model in equations (1.8) and (1.9). This manner of modeling dependences was used in econometric analysis to handle nonstationary time series. Markov-switching models have many similarities with basic HMMs. The computational machinery used for both models is virtually the same. However, they do differ in the statistical analysis.

Originally, the main areas of applications of HMMs are speech recognition and machine learning. In Fine et al. [23], hierarchical HMM is developed in an effort to improve algorithms that use stochastic context-free gammas. The idea is to put additional sources of dependency in the model. It was proposed that states emit sequences rather than single symbols. Therefore, every state is composed of substates which by themselves are composed of substates as well. To illustrate this, we consider a graphical representation of a simple one substate type model with multiple levels of dependencies. In a state-space form, the model featured in Figure 1.3 can be written as

$$C_{k+1} = a(C_k, U_k). \tag{1.7}$$

$$X_{k+1} = b(C_{k+1}, X_k, V_{k+1}) \tag{1.8}$$

$$Y_{k+1} = c(X_{k+1}, C_{k+1}, W_{k+1}). \tag{1.9}$$

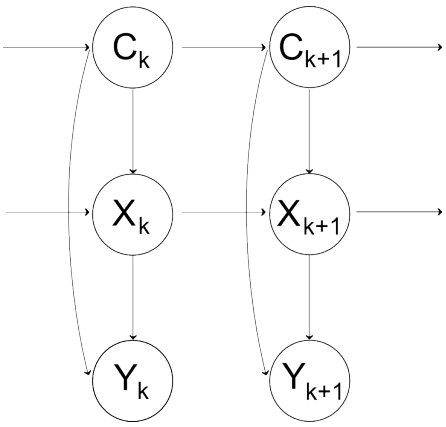


Figure 1.3: An example of a structure of a hierarchal HMM, where $\{C_k\}$ is a hidden Markov chain, and $\{Y_k\}$ and $\{X_k\}$ are correlated processes driven by$\{C_k\}$

Working with this type of model, however, is computationally intensive and, in some cases, it is not possibile to find a tractable solution. The only subclass of hierarchical HMMs that is widely used is Gaussian. In particular, a conditionally Gaussian linear state-space model (CGLSSM) has the state-space form

$$W_{k+1} = A(C_{k+1})W_k + R(C_{k+1})U_k \tag{1.10}$$

$$Y_{k+1} = B(C_k)W_k + S(C_k)V_k. \tag{1.11}$$

In equations (1.10) and (1.11), $\{C_k\}_{k\geq 0}$ is a Markov chain, $W_0 \sim N(\mu_\nu, \Sigma_\nu)$ and A, B, R and S are known matrix-valued functions of appropriate dimensions satisfying additional regularity conditions. Such model is well-known due to its significance in implementing the Kalman filter (cf. Kalman [36]) and use in computer engineering. Kalman filter is outlined briefly in the subsequent chapter. Kalman filtering equations together with Monte-Carlo methods are needed to work with CGLSSM.

To date, papers on HMMs abound with applications to speech recognition, genetics, biology, finance and economics. In the literature, HMMs have different classifications according to the type of noise, type of applications, etc. However, almost all of them can be obtained from the models described in equations (1.10)-(1.11). Hence, as alluded in Cappe et al. [7], if we learn the main principles of working with the above model

formulation, in theory, we can apply our machinery to almost every possible situation and get some solutions.

1.4 Synopsis of main results in HMM filtering

We summarise the main achievements in the area of HMM filtering. Results described in this chapter form the starting point of our contributions to the field. The first significant progress in HMM filtering was made in the mid 60s. From then on, major results have been established, which include the (i) forward-backward method (Baum et al. [4]), (ii) Baum-Welch filter [4], (iii) Viterbi algorithm [53], (iv) Expectation-Maximisation (EM) algorithm (Dempster et al. [11]), (v) Markov-switching model (Hamilton [25, 26]), and (vi) measure change-based filters (Zakai [57]; Elliott et al. [15]; Elliott [13]; Mamon and Elliott [45]; Elliott [14]; Siu et al. [51]; Erlwein[20]).

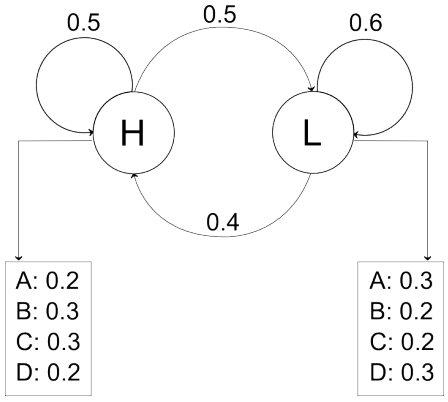


Figure 1.4: An example of a simple HMM with a discrete alphabet of emissions

The main results described in this chapter are discussed on an intuitive level. For the rigorous details, formal proofs and technical conditions, see the original papers. We adopt notation of Ewens and Grant [19]. Without loss of generality, we study one of the simplest HHMs and then try to explain Viterbi, forward-backward and Baum-Welch algorithms based on a model example taken from bioinformatics. Such type of HMM is used as a simple model for gene decoding. The model under consideration, as depicted in Fig 1.4, consists of two states $\{S_1 = H, S_2 = L\}$ (high and low). It has a transition matrix $\mathbf{A} = \{a_{ij}\}$, $i, j = 1, 2$ with an alphabet of emissions $\{A, B, C, D\}$ and

corresponding emission probabilities $b_i(a) = P(a|S_i)$ so that for example, $b_H(A) = 0.2$, $b_H(B) = 0.3$, and so on. We denote the set of emmision probabilities as $\mathbf{B}$. It is also considered that the initial distribution π is known.

1.4.1 Viterbi algorithm

Given some observed sequence $O = O_1, O_2, O_3, \ldots, O_T$ of outputs, we wish to compute efficiently the state sequence $Q = q_1, q_2, q_3, ...q_T$ that has the highest conditional probability given O. That is, calculate

$$\arg\max_Q \ P(Q|O). \tag{1.12}$$

For arbitrary t and i, define

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} \ P(\{q_1, q_2, ..., q_{t-1}, q_t = S_i\} \ \cap \ \{O_1, O_2, O_3, ...O_t\})$$

So, $\delta_t(i)$ is the maximum probability of all possible ways to end up in state S_i at time t having observed the sequence $O_1, O_2, O_3, ...O_t$. Then

$$\max_Q \ P(Q \cap O) = \max_i \ \delta_T(i). \tag{1.13}$$

Our aim is to find a sequence Q for which the maximum conditional probability in equation (1.12) is achieved. Since

$$\max_Q \ P(Q|O) = \max_Q \ \frac{P(Q \cap O)}{P(O)}$$

and the denominator on the right-hand side does not depend on Q,

$$\arg\max_Q \ P(Q|O) = \arg\max_Q \ \frac{P(Q \cap O)}{P(O)} = \arg\max_Q \ P(Q \cap O).$$

Hence, we have the algorithm as follows:

- Initialisation step: $\delta_1(i) = \pi_i b_i(O_1)$ for $1 \leq i \leq N$.
- Induction step: $\delta_t(j) = \max_{1 \leq i \leq N} \ \delta_{t-1}(i) a_{ij} b_j(O_t)$, $2 \leq t \leq T$, $\ 1 \leq j \leq N$

To recover the q_is, define

$$\psi_T := \arg \max_{1 \leq i \leq N} \ \delta_T(i)$$

and put $q_T = S_{\psi_T}$. Then q_T is the final state in the required state sequence. The remaining q_t are found recursively by using

$$\psi_t = \arg \max_{1 \leq i \leq N} \ \delta_t(i) a_{i\psi_{t+1}} \ \text{ and then putting } \ q_t = S_{\psi_t}.$$

Going back to our example, suppose we observed a sequence CCBABDCAA. From the Viterbi algorithm, we conclude that the most likely sequence of states emitted is HHHLLLLLL.

1.4.2 Forward-backward algorithm

Let $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ be the full set of parameters. Given λ, we address the question of how to calculate efficiently $P(O|\lambda)$, which is the probability of some given sequence of observed outputs. We consider an efficient method of calculating this by defining

$$\alpha(t, i) = P(O_1, O_2, O_3, ..., O_t, q_t = S_i). \tag{1.14}$$

This is a joint probability that the sequence of observations seen up to and including time t is $O_1, O_2, O_3, ..., O_t$, and that the state of the HMM at time t is S_i. The $\alpha(T, i)$ are called forward probabilities. The **forward** algorithm can be described in the following way:

- Initialisation: $\alpha(1, i) = \pi_i b_i(O_1)$.
- Induction: $\alpha(t + 1, i) = \sum_{j=1}^{N} \alpha(t, j) a_{ji} b_i(O_{t+1})$.
- Termination: $P(0) = \sum_{i=1}^{N} \alpha(T, i)$.

This algorithm requires computations in the order of TN^2, and thus, is feasible in practice even for problems involving large dimensions.

The second part of the forward-backward algorithm is the **backward** algorithm. In the above, we calculated successively

$$\alpha(1, \cdot), \alpha(2, \cdot), \ldots, \alpha(T, \cdot)$$

forward in time. In the backward algorithm, we calculate another quantity but backward in time. This will not be used to solve $P(0) = \sum_{i=1}^{N} \alpha(T, i)$ but it will be needed instead in the Baum-Welch algorithm in subsection 1.4.3.1. The goal of the backward algorithm is to calculate the probability $\beta(t, i)$ defined by

$$\beta(t, i) = P(O_{t+1}, O_{t+2}, ..., O_T | q_t = S_i), \tag{1.15}$$

for $1 \leq t \leq T-1$. For convenience and without loss of generality, we set $\beta(T, j)$ to 1, for all j. We then compute equation (1.15) working backwards from $t = T-1$. The induction step of the procedure entails the equation

$$\beta(t-1, i) = \sum_{j=1}^{N} a_{ij} b_j(O_t) \beta(t, j).$$

It may be shown that

$$P(q_t = k, q_{t+1} = l | O, \lambda) = \frac{\alpha(t, k) a_{kl} b_l(O_{t+1}) \beta(t+1, l)}{P(O|\lambda)}. \tag{1.16}$$

Equation (1.16) provides the impetus for the development of the Baum-Welch algorithm. Rabiner [49] generalised the idea of Baum et al. [4] and obtained normalised recursive equations for the forward-backward algorithm covering the case $\mathbf{X} = \{1, \ldots, r\}$.

1.4.3 Expectation-maximization algorithms

Both Dempster et al. [11] and Baum et al. [4] explained the main principles of finding the parameters of models in instances of incomplete data. These principles and algorithms became the cornerstone of further applications of HMMs as they can be viewed as a subcategory of models for missing data. We describe the results for the Baum-Welch and expectation-maximisation (EM) algorithms.

1.4.3.1 Baum-Welch algorithm

Baum-Welch (BW) algorithm is designed to maximise $P(O_1, O_2, O_3, \ldots, O_n)$. To describe a general BW algorithm, we expand the notation stated at the beginning of this chapter to:
$O^1 : O_1^1, O_2^1, O_3^1, ..., O_{m_1}^1$
$O^2 : O_1^2, O_2^2, O_3^2, ..., O_{m_2}^2$
$O^3 : O_1^3, O_2^3, O_3^3, ..., O_{m_3}^3$
.....
$O^n : O_1^n, O_2^n, O_3^n, ..., O_{m_n}^n$.

This means that we are provided with n training sequences of possibly different lengths $m_1, m_2, \cdots, m_n$. Suppose $a_{kl}^{(0)}$ and $b_k^{(0)}$ are initial parameters such that $P^0(O_r) \geq 0$ for

all training sequences. This restriction can be relaxed to ensure faster convergence of the algorithm. Here, $P^{(s)}(O_r)$ is the probability of observing the sequence O^r after s estimation steps. To proceed with the calculation, we consider initial values for the parameters given by

$$\begin{aligned}\hat{\pi} &= \text{expected proportion of times in state} \\ &\quad S_i \text{ at the first time point given } \{O^d\},\end{aligned} \tag{1.17}$$

$$\hat{a_{jk}} = \frac{E(N_{jk}|\{O^d\})}{E(N_j|\{O^d\})}, \tag{1.18}$$

and

$$\hat{b_i}(a) = \frac{E(N_i(a)|\{O^d\})}{E(N_i|\{O^d\})}, \tag{1.19}$$

where N_{jk} is the number of times $q_t^d = S_j$ and $q_{t+1}^d = S_k$ for some d and t; N_i is the number of times $q_t^d = S_i$ for some d and t; and $N_i(a)$ equals to the number of times $q_t^d = S_i$ if it emits symbol a, for some d and t.

Equations (1.17), (1.18) and (1.19) constitute the induction step of the recursive algorithm in the estimation of $\hat{\pi}, \hat{a_{jk}}$ and $\hat{b_i}(a)$. Convergence is achieved since $P(\{O^d\}|\hat{\lambda^{(n+1)}}) \geq P(\{O^d\}|\hat{\lambda}^{(n)})$, where $\hat{\lambda^{(n)}}$ is the set of parameters used in the n^{th} step.

To complete the calculations of equations (1.17), (1.18), and (1.19), write

$$\zeta_t^d(i,j) = P(q_t = k, q_{t+1} = l|O, \lambda) := \frac{\alpha(t,k)a_{kl}b_l(O_{t+1})\beta(t+1,l)}{P(O|\lambda)} \tag{1.20}$$

Equation (4.15) can be computed from (1.16). Let $I_t^d(i)$ be indicator variables defined by

$$I_t^d(i) = 1 \text{ if } q_t^d = S_i \text{ and } I_t^d(i) = 0 \text{ otherwise.}$$

Then $\sum_d \sum_t I_t^d(i)$ represents the number of times S_i is visited. Hence, the expected number of times S_i is visited, given $\{O^d\}$, is

$$\sum_d \sum_t E(I_t^d(i)|O^d) = \sum_d \sum_t \sum_{j=1}^N \zeta_t^d(i,j). \tag{1.21}$$

Thus, the expected number of times S_i is visited, given $\{O^d\}$, is

$$\sum_d \sum_t \sum_{j=1}^N \zeta_t^d(i,j).$$

Similarly, the expected number of transitions from S_i to S_j, given $\{O^d\}$, is

$$\sum_d \sum_t \zeta_t^d(i,j)$$

To calculate the numerator of equation (1.19), we define a new indicator random variable by

$$I_t^d(i,a) = 1 \text{ if } q_t^d = S_i \text{ and } O_t^d = a \ ; \ I_t^d(i) = 0 \text{ otherwise.}$$

Consequently,

$$E(N_i(a)|\{O^d\}) = \sum_d \sum_t \sum_{O_t^d = a} \sum_{j=1}^{N} \zeta_t^d(i,j).$$

1.4.3.2 The EM algorithm

Dempster et al. [11] laid down the foundations of a very general class of incomplete-data models, which includes HHMs and Markov-switching models. As previously mentioned, the Baum-Welch algorithm can be derived as a particular case of the EM algorithm. We follow Cappe et al. [7] closely in outlining the EM algorithm.

Assume that we have a σ-finite measure μ on (X,χ) and a family $\{f(\cdot;\theta)\}_{\theta\in\Theta}$ of non-negative μ-integrable functions on X. This family is indexed by a parameter $\theta \in \Theta$, where Θ is a subset of R^{d_θ}, for some integer d_θ. The task under consideration is the maximization, with respect to the parameter θ, of

$$L(\theta) := \int f(x;\theta)\mu(dx).$$

The function $f(\cdot;\theta)$ can be thought of as an unnormalised probability density with respect to μ.

In the study of HMMs, we shall consider the joint probability function f of two random variables X and Y where the latter is observed while the former is not. Then X is referred to as the missing data, f is a complete-data likelihood, and L is the density of Y alone, which is the likelihood available in estimating θ. In general, we assume $L(\theta)$ is positive. Thus, maximizing $L(\theta)$ is equivalent to maximising the log-likelihood

$$l(\theta) := \log\ L(\theta).$$

We also associate to each function $f(\cdot;\ \theta)$ the probability density function $p(\cdot;\ \theta)$, with respect to a measure μ, defined by

$$p(x;\theta) = \frac{f(x,\theta)}{L(\theta)}.$$

This tells us that $p(x,\theta)$ is a conditional density of X given Y.

A central underpinning of Dempster et al.'s methodology [11] is the EM algorithm, which starts with the family $\{Q(\cdot;\theta^{'})\}_{\theta\in\Theta}$ of real-valued functions on Θ indexed by $\theta^{'}$, and defined by

$$Q(\theta,\theta^{'}) = \int \log\ f(x;\theta)p(x;\theta^{'})\mu(dx). \tag{1.22}$$

The related regularity conditions are stated in subsection 1.4.3.2. When we encounter $0.\log 0$ in equation (1.22), we assign it a value of 0. The quantity $Q(\theta,\theta^{'})$ may be interpreted as the expectation of $\log\ f(X;\theta)$ when X is distributed according to the probability density function $p(\cdot;\ \theta)$. With the previous notation, it is possible to rewrite equation (1.22) as

$$Q(\theta,\theta^{'}) = l(\theta) - H(\theta,\theta^{'}), \tag{1.23}$$

where

$$H(\theta,\theta^{'}) = -\int \log\left(p(x;\theta)\right)p(x;\theta^{'})\mu(dx). \tag{1.24}$$

From equation (1.24), the difference

$$H(\theta,\theta^{'}) - H(\theta^{'},\theta^{'}) = -\int \log\ \frac{p(x;\theta)}{p(x;\theta^{'})}p(x;\theta^{'})\lambda(dx) \tag{1.25}$$

is recognised as a Kullback-Leibler divergence (distance) between probability density functions $p(x;\ \theta)$ and $p(x;\ \theta')$.

We complete our introduction to the EM algorithm by stating the regularity assumptions of the proposed framework.

- The parameter set Θ is an open subset of R^{d_θ}, for some integer d_θ.
- For any $\theta\in\Theta$, $L(\theta)$ is positive and finite.
- For any $(\theta,\theta^{'})\in\Theta\times\Theta$, $\int |\nabla_\theta\ \log\ p(x;\theta)|p(x;\theta^{'})\mu(dx)$ is finite, where ∇_θ is the gradient with respect to θ.

The convergence of the EM algorithm can be justified as follows. Under the assumptions of subsection 1.4.3.2, for any $(\theta, \theta^{'}) \in \Theta \times \Theta$,

$$l(\theta) - l(\theta^{'}) \geq Q(\theta; \theta^{'}) - Q(\theta^{'}; \theta^{'}), \tag{1.26}$$

where the inequality is strict unless $p(\cdot;\ \theta)$ and $p(\cdot;\ \theta^{'})$ are equal μ - a.e. In addition, we assume that (i) the mapping $\theta \rightarrow L(\theta)$ is continuously differentiable on Θ and (ii) for any $\theta^{'} \in \Theta$, $\theta \rightarrow H(\theta; \theta^{'})$ is continuously differentiable on Θ. Then for any $\theta^{'} \in \Theta$, $\theta \rightarrow Q(\theta; \theta^{'})$ is continuously differentiable on Θ and

$$\nabla_{\theta} l(\theta^{'}) = \nabla_{\theta} Q(\theta; \theta^{'})|_{\theta=\theta'}.$$

The EM algorithm is an iterative construction of the sequence $\{\theta^i\}_{i\geq 1}$ of parameter estimates given an initial guess θ^0. Each iteration is divided into two parts:

- Expectation step: Determine $Q(\theta; \theta^i)$.
- Maximisation step: Chose θ^{i+1} to be the value of $\theta \in \Theta$ that maximises $Q(\theta; \theta^i)$.

The EM algorithm is a very powerful tool when working with incomplete data and as shown in the sequel it supports the filter recursions.

1.4.4 Change of measure approach

This approach incorporates the techniques discussed above with one additional step of doing a change of probability measure. The basic idea of the method is to facilitate the calculations in an easier framework, where under a new probability measure $\widetilde{P}$, all observations are independent and perhaps, identically random variables. Therefore, all calculations take place under the $\widetilde{P}$ measure, where Fubini's theorem permits the interchange of expectations and summations (Loeve [42]). The construction of the new measure $\widetilde{P}$ follows from the Girsanov theorem. The procedure can be understood more clearly by considering Figures 1.5 and 1.6.

To illustrate the reference probability measure approach in a very simple situation we consider an example from Elliott [15]. Suppose that in a coin-tossing experiment, the probability of heads is p and the probability of tails is q. A probability space that

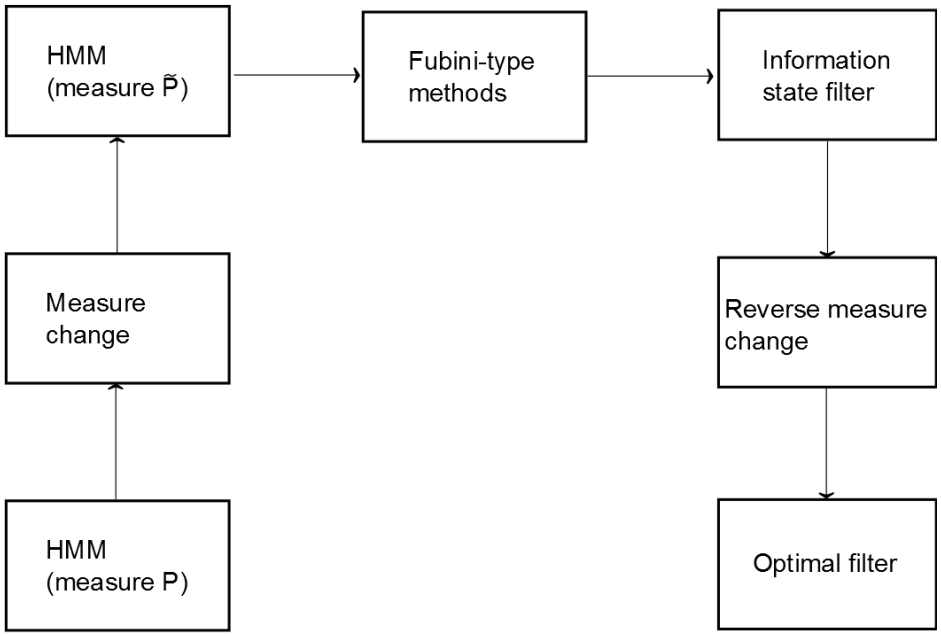


Figure 1.5: Reference probability optimal filter derivation

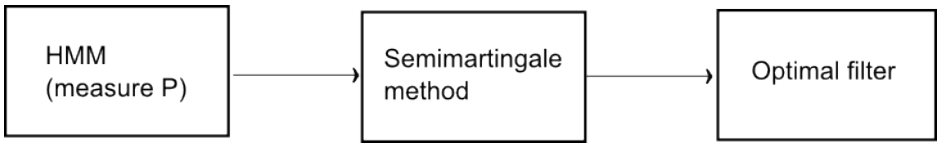


Figure 1.6: Direct optimal filter derivation

describes the outcomes of one throw of such a coin is $\Omega = \{H, T\}$ with probability measure P where $P(H) = p,\ P(T) = 1 - p$. We suppose p is neither 0 nor 1. Suppose further that we wish to adjust our statistics in our experiment to that of a fair coin. We can achieve this mathematically by introducing a new probability measure $\widetilde{P}$ such that $\widetilde{P}(H) = \widetilde{P}(T) = 0.5$. This implies that the event $\{T\}$ has been weighted by a factor $\dfrac{\widetilde{P}(T)}{P(T)} = \dfrac{1}{2\widetilde{p}}$. The function $\dfrac{\widetilde{P}(\cdot)}{P(\cdot)}$ is the Radon- Nikodym derivative of the fair (uniform) $\widetilde{P}$ measure with respect to P.

We note that the function $\dfrac{\widetilde{P}(\cdot)}{P(\cdot)}$ can be used to define $\widetilde{P}$ because

$$\widetilde{P}(\cdot) = \frac{\widetilde{P}(\cdot)}{P(\cdot)} P(\cdot).$$

Calculations performed under $\widetilde{P}$ can always be related back to the real-world measure P by invoking an inverse change of measure.

1.5 Kalman filter

Gaussian state-space models (GSSM) are an important subclass of Markov-switching models. They have been examined quite extensively as tools in handling time series with state-dependent coefficients. The estimation involved is relatively straightforward and applications are ubiquitous in various areas of engineering and the sciences including aerodynamics, physics, speech recognition, economics, genetics and finance, among others.

Linear Gaussian state-space models (LGSSM) form a subclass of their own. This can be largely explained by a huge variety of applications of data filtering in physics. Many laws of physics can be expressed as linear ordinary differential equation (ODE) or system of ODEs. Kalman ([36] and [37]) made filtering algorithms accessible that demonstrated filtering applications in the accurate estimation of the exact coordinates of moving objects (e.g., plane, missile, etc.). In the last two decades, there has been considerable attention in applying Kalman filters in the modeling and estimation of the dynamics of various quantities and variables in finance (cf. Date and Ponomareva [10],and Wells [54]).

To describe the Kalman filter, consider the state-space model

$$X_{k+1} = A_k X_k + R_k U_k \tag{1.27}$$

$$Y_k = B_k X_k + S_k V_k, \tag{1.28}$$

where $\{U_k\}_{k\geq 0}$ and $\{V_k\}_{k\geq 0}$ are uncorrelated sequences of white noise, i.e., they have zero mean and identity covariance matrices. The initial distribution of X, denoted by X_0, is assumed uncorrelated with $\{U_k\}_{k\geq 0}$ and $\{V_k\}_{k\geq 0}$, and $E(X_0) = 0$ and $\mathrm{Cov}(X_0) = \Sigma$.

Define the following:
$\hat{X}_{k|n}$: = one-step prediction of X_k given $Y_0, ..., Y_n$
$\sigma_{k|n} := \mathrm{Cov}(X_k - \hat{X}_{k|n})$
$\epsilon_k := Y_k - \hat{Y}_{k|k-1}$
S_k^t and B_k^t are the corresponding transposes of S_k and B_k.

Kalman filtering algorithm:
For $k = 0, \ldots, n$, the steps involved are:

- Initialisation step: If $k = 0$, set $\hat{X}_{k|k-1} = 0$ and $\Sigma_{k|k-1} = \Sigma$; otherwise set
  $\hat{X}_{k|k-1} = A_{k-1}\hat{X}_{k-1|k-1}$,
  $\Sigma_{k|k-1} = A_{k-1}\Sigma_{k-1|k-1}A_{k-1}^t + R_{k-1}R_{k-1}^t$.

- Iteration step: Calculate the
  Innovation $\epsilon_k = Y_k - B_k\hat{X}_{k|k-1}$,
  Innovation covariance $\Gamma_k = B_k\Sigma_{k|k-1}B_k^t + S_kS_k^t$,
  Kalman gain $K_k = \Sigma_{k|k-1}B_k^t\Gamma_k^{-1}$,
  Filtering state estimation $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k\epsilon_k$, and
  Filtering error covariance $\Sigma_{k|k} = \Sigma_{k|k-1} - K_kB_k\Sigma_{k|k-1}$.

The fundamental principles and theoretical underpinnings of Kalman filtering discussed above will serve as the starting point of the last two projects on the comparison of HMM filtering with other techniques, and combining HMM filtering with other filtering methods.

1.6 HMMs in finance, actuarial science and economics

This chapter underscores the various applications of HMMs to financial, actuarial and economic modeling. We focus on how researchers bridge the gap between mathematical developments and frameworks of applications in practice.

1.6.1 Motivation of using HMMs

Phenomena of stochastic nature abound in the sciences and social sciences and finance and economics are of no exception. Black and Scholes [6] proposed in their seminal work to use stochastic processes in modeling stock prices. A geometric Brownian motion (GBM) was employed implying that the log-returns are assumed normally distributed. This leads to an analytic expression for the price of European call and put options.

The assumption of normality is made for model tractability. However, such assumption means that asset price changes are not serially correlated, which is not entirely correct in reality. Merton [48] extended Black and Scholes' model to cover the case of time-dependent parameters affording greater flexibility. The presence of volatility smiles or smirks, however, confirms that the constant volatility in the Black-Scholes-Merton's GBM framework is not satisfied.

With market upheavals and recent financial crisis (e.g., 2008-2009 period), the pure diffusion approach to derivative pricing is deemed inadequate especially in modeling the dynamics of variables underlying long-term investments. Among the reasons are:

- The long-run log returns follow a distribution with heavier tails compared to those of the normal distribution.
- The distribution of log returns is completely different in various stages of the economy. For example, during recession data exhibit huge spikes with negative drift. During a stable or booming period the reverse is true, i.e., volatility is relatively low and positive drift is positive.
- Independence of increments of log returns is violated. Indeed, a strong Markov property is not realistic and models with different kind of dependencies have to be considered.

Diffusion models may be useful only in capturing the behavior of variables in the short run but alternative models must be sought in the analysis of data used for long-term financial instruments.

Several approaches have been proposed to rectify the normality assumption.

- Fat-tailed distributions could be incorporated in model construction. Also, empirical density of log-returns has a much higher concentration around the mean compared to the Gaussian distribution. Lack of tractability and more involved estimation of parameters are the drawbacks of this approach.

- Adding a jump component into the model could be employed to give provision for market crashes and other structural changes. Such models are termed jump-diffusion models and surveyed in Kou [39].

- A broader category of models, the Levy-type models, is an appropriate alternative. These models consider stochastic processes that have infinitely divisible distributions. They are quite rich in structure as Brownian motion and jump-diffusion models are special cases. Nonetheless, there are some pros and cons to work with Levy processes. It is quite easy to find a process that describes the data perfectly. But, more often than not, it is very hard, and in some cases, almost impossible, to calibrate the model making parameter estimation a herculean challenge. The techniques of working with Levy processes are completely different compared to the current approach of working with diffusion models (Kyprianou [41], and Asmussen [3]). New methods in this area provide general results. But, since there is an associated increase of estimation complexity and diminishing tractability with general results, they are not widely embraced by practitioners.

- A simple but powerful solution is through the use of mixture distributions with weights assigned to certain distributions (Giacomini et al. [24]). A mixture of distributions can create a distribution that captures stylized features of data observed during economic downturns or market recovery (e.g., 1997-1998 and 2009-2010 periods). The distribution gives various shapes reflecting different levels of skewness and excess kurtosis.

- The approaches that utilise jump-diffusion models and mixture distributions could be synthesized together. The idea is to use a Markov chain that governs the dynamics of parameters. The state of the Markov chain models the state

of the market or economy, and combination of states produced a mixture of distributions. Its sophistication can be increased further by adding a jump process whose intensity can also depend on a Markov chain. See Erlwein et al. [22].

1.6.2 Examples of HMM applications

Hamilton popularized Markov-switching models in a series of papers (cf. Hamilton [30], [31], [32], [33]) and described in details [27] how to work with autoregressive time series with parameters driven by a hidden Markov chain. His theoretical developments were based on a two-state Markov chain model. The estimation of model coefficients using the EM algorithm was established and the implementation was demonstrated using US data via maximum likelihood. ARCH models with regime-dependent coefficients were also presented in Hamilton and Susmel [29]. In [30], a class of state-space models that extends the regular HMM was introduced. All known HMMs including the Kalman filter Kalman [36] and time-series models within the same state-space framework were classified. Hamilton and Raj [28] came up with a collection of results and applications in econometrics and finance of time series with regime-dependent coefficients. Hamilton's contributions in HMMs have found continued applications in econometrics, finance and insurance.

Hardy [34] further developed and applied Hamilton's methodology [27] to the S&P 500 and TSE 300 indexes. The new models called regime-switching autoregressive (RSAR(1)) and regime-switching log normal (RSLN-3) models were compared to the existing ones, such as the independent lognormal model (ILN), AR(1), ARCH(1) and GARCH(1,1). It was established that the two-regime model has the best fit using a number of different criteria, such as the AIC (see Akaike [1]) and BIC (see Schwartz [50]) and the likelihood ratio test (see Klugman et al. [38]). Pricing formulas for long-term vanila options were derived. Calculations of popular risk measures such as the value-at-risk (VaR) and and conditional tail expectation (CTE) under multi-regime framework were performed. Sufficient information were given to enable the generalizations to models with more than two regimes. Interestingly, while a three-regime model was considered, it was indicated that such number of regimes is more than necessary for the data used in the empirical study.

Elliott et, al. [15] made substantial contributions to the HHM theory and applications

of Markov chains in finance. His approach is quite different from that of Hamilton in which a change of reference probability measures is at the core of the estimation and filtering procedure. Such methodology was pioneered in electrical engineering by Elliott and his colleagues, and its versatility and power were later promoted in financial applications. In Elliott et al. [16], a regular Ornstein-Uhlenbeck process with a mean-reverting level is modulated by a Markov chain. Based on HMM technique described in [15], dynamic filters for the model parameters were developed in continuous-time settings; simulations showed good results of modeling capability. Erlwein and Mamon [21] pursued this line of inquiry and designed filters in discrete-time setting in the estimation of HMM-driven parameters. The HMM measure-change approach was also applied in the pricing of commodities (Elliott et al. [17]; Erlwein et al. [22]; and Mamon et al. [47]). Another important application of HMMs is in risk measurement such as the works of Hardy in VaR and CTE computations. In Siu et al. [52], a change of measure approach developed by Elliott was used to find VaR on simulated data generated by a weak Markov chain. In Xi and Mamon [55], filtering and forecasting were carried out under a weak HMM setting. Applications to asset allocation are highlighted in Erlwein et al. [20]. Whilst the strategy in Guidolin and Timmermann [12] does not feature a reweighing of the portfolio, it describes the elegant approach of filtering in conjunction with the change of measure. We note that it is one of the few papers suggesting 4 as the appropriate number of regimes for the data in its numerical investigation; in majority of the case studies in the literature, 2 or 3 regimes are deemed sufficient.

1.7 Structure of the thesis

This thesis contains the following contributions. Chapter 2 presents the filtering and forecasting of futures prices under a multivariate HMM set-up. In chapter 3, we consider the filtering of an HMM-driven multivariate Ornstein-Uhlenbeck model with a special focus on forecasting market liquidity. Chapter 4 deals with the development of a pairs trading strategy under a multi-regime environment. An application of zero-delay HMMs to high-frequency foreign exchange trading is described in chapter 5. Chapter 6 generalises static calibration algorithm procedures for a regime-switching model with any number of regimes and describes the procedure of obtaining the distribution function of the total return in multi-regime environment. A summary and possible extensions of the research results in this thesis are given in chapter 7.

1.8 References

[1] Akaike, H., 1974. A new look at statistical model identification, IEEE Transactions on Automatic Control 19, 716 -723. 23

[2] Aldridge, I., 2010. High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems, John Wiley and Sons, Hoboken, New Jersey. 5

[3] Asmussen, S., 2000. Ruin Probabilities, Word Scientific, Singapore. 22

[4] Baum, L., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Annals of Mathematical Statistics, 41, 164–17. 10, 13

[5] Boudt, K., Paulus, E., Rosenthal, D., 2010. Funding liquidity, market liquidity and TED spread: a two-regime model, SSRN Working Paper Series, Social Science Electronic Publishing, Rochester, New York. 3

[6] Box, G., Draper, N., 1987. Empirical Model-Building and Response Surfaces, Wiley, New York. 21

[7] Cappe, O., Moulines, E., Ryden, T., 2007. Inference in Hidden Markov Models, Springer-Verlag Inc., Secaucus, New Jersey. 6, 7, 8, 9, 15

[8] Cheung, Y., 2005. Exchange rates and Markov switching dynamics, Journal of Business & Economic Statistics 23(3), 314–320. 5

[9] Dacorogna, M., Gencay, R., Muller, U., Olsen, R., 2001. An Introduction to High-Frequency Finance, Academic Press, San Diego, California. 5

[10] Date, P., Ponomareva, K., 2011. Linear and non-linear filtering in mathematical finance: a review, IMA Journal of Management Mathematics 22, 195–211. 4, 19

[11] Dempster, A., Laird, N., Rubin, D., 1977. Maximum Likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, 39, 1–38. 10, 13, 15, 16

[12] Guidolin, M., Timmermann, A., 2003. Value at risk and expected shortfall under regime switching, EFA 2004 Maastricht Meetings, 2983. 24

[13] Elliott, R., 1993. New finite-dimensional filters and smoothers for noisily observed Markov chains, IEEE Transactions on Information Theory 39, 265–271. 10

[14] Elliott, R., Kopp, E., 2004. Mathematics of Financial Markets, Springer, New York. 10

[15] Elliott, R., Aggoun, L., Moore, J., 1995. Hidden Markov Models. Estimation and control, Springer, New York. 7, 10, 17, 23, 24

[16] Elliot, R., Fisher, P., Platen, E., 1999. Hidden Markov filtering for a mean reverting interest rate model, Decision and Control, Proceedings of the 38th IEEE Conference 3, 2782–2787. 24

[17] Elliott, R., Sick, G., Stein, M., 2000. Pricing electricity calls, Working Paper, University of Alberta. 24

[18] Elliott, R., van der Hoek, J., Malcolm, W., 2005. Pairs Trading, Quantitative Finance 5(3), 271–276. 4

[19] Ewens, J.,Grant, G., 2001. Statistical Methods In Bioinformatics, Springer, New York. 10

[20] Erlwein, C., Mamon, R., Davison, M., 2011. An examination of HMM-based investment strategies for asset allocation, Applied Stochastic Models in Business and Industry 27, 204-221. 10, 24

[21] Erlwein, C., Mamon, R., 2007. An online estimation scheme for a Hull-White model with HMM-driven parameters, Statistical Methods and Applications, 18(1) 87–107. 24

[22] Erlwein, C., Benth, F., Mamon, R., 2010. HMM filtering and parameter estimation of an electricity spot price model, Energy Economics, 32(5) 1034–1043. 23, 24

[23] Fine, S., Singer, W., Tishby, N., 1998. The hierarchical hidden Markov model: analysis and applications, Machine Learning, 32, 41–62. 8

[24] Giacomini, R., Gottschling, A., Haefke, C., White, H., 2008. Mixtures of t-distributions for finance and forecasting, Journal of Econometrics, Elsevier, 144(1), 175–192. 22

[25] Hamilton, J., 1988. Rational expectations econometric analysis of changes in regime: an investigation of term structure of interest rates, Journal of Economic Dynamics and Control, 12, 385–423. 8, 10

[26] Hamilton, J., 1989. A new approach to the econometric analysis of nonstationary time series and the business cycle, Econometrica, 57, 357–384. 8, 10

[27] Hamilton, J., 1994. Time Series Analysis, Princeton University Press, Princeton, New Jersey. 7, 23

[28] Hamilton, J., Raj, B., 2002. Advances in Markov-Switching Models: Applications in Business Cycle Research and Finance, Physica-Verlag, New York. 23

[29] Hamilton, J., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime, Journal of Econometrics, 64, 307–333. 23

[30] Hamilton, J., 1994. State-space models, Handbook of Econometrics, 4, 4–50. 23

[31] Hamilton, J., 1993. Estimation, inference, and forecasting of time series subject to changes in regime, Handbook of Statistics, 11. 23

[32] Hamilton, J., 1990. Analysis of time series subject to changes in regime, Journal of Econometrics, Elsevier, 45(1-2), 39–70. 23

[33] Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 57(2), 357–84. 23

[34] Hardy, M., 2001. A regime switching model of long term stock returns, North American Actuarial Journal Society of Actuaries, 2(5), 11–26. 6, 23

[35] Hyndman, C., Elliott, R., 2007. Parameter estimation in commodity markets: a filtering approach, Journal of Economic Dynamics and Control, 31, 2350–2373.

[36] Kalman, R., 1960. A new approach to linear filtering and prediction problem, Transactions of the ASME–Journal of Basic Engineering 82, 35–45. 9, 19, 23

[37] Kalman, R., Bucy, R., 1961. New results in linear filtering and prediction theory, Transactions of the ASME–Journal of Basic Engineering 83, 95–108. 19

[38] Klugman, S., Panjer, H., Willmot, G., 2003. Loss Models: From Data To Decisions, Wiley, New York. 23

[39] Kou, S., 2008. Jump-diffusion models for asset pricing in financial engineering, Handbooks in OR and MS, 15, 73–116. 22

[40] Krugman P., 2008. The Conscience of a liberal - Mission not accomplished, not yet anyway, New York Times, March 12. 3

[41] Kyprianou, A., 2006. Introductory Lectures on Fluctuations of Levy Processes with Applications, Springer-Verlag, Berlin. 22

[42] Loeve, M., 1978. Probability Theory, Springer-Verlag, New York. 17

[43] Logothetis, A., Krisnamurthy, V., 1996. An adaptive hidden Markov model/Kalman filter algorithm for narrowband interference suppression with applications in multiple access communications, Statistical Signal and Array Processing, 8th IEEE Signal Processing Workshop, 490–493. 4

[44] Luo, S., Tsoi, A., 2007. Filtering of hidden weak Markov chain: discrete range observation, R.S. Mamon, R.J. Elliott (Eds.), Hidden Markov Models in Finance, Springer, New York , 106–119.

[45] Mamon R., Elliott, R., 2007. Hidden Markov Models in Finance, Springer, New York. 1, 7, 10

[46] Mamon R., Elliott, R., 2014. Hidden Markov Models in Finance: Further Developments and Applications (Volume II), Springer, New York. 1

[47] Mamon, R., Erlwein, C., Gopaluni, R., 2007. Adaptive signal processing of asset price dynamics with predictability analysis, Information Sciences, 178(1), 203–219, 2007. 24

[48] Merton., R., 1971. Theory of rational option pricing, Bell Journal of Economics and Management Science 4 (1), 141–183. 21

[49] Rabiner, L., 1989. A tutorial on hidden Markov models with selected applications in speech recognition, IEE Proceedings , 77 , 257–285. 13

[50] Schwartz, G., 1978. Estimating the dimensions of a model, Annals of Statistics, 6, 461–464. 23

[51] Siu, T., Ching, W., Fung, E., 2005. Extracting information from spot interest rates and credit ratings using double higher-order hidden Markov models, Computational Economics, 26 , 251–284. 7, 10

[52] Siu, T., Ching, W., Fung, E., Ng, M., Li, X., 2001. A high-order Markov-switching model for risk measurement, Computers and Mathematics with Applications, 58(1), 1–10. 7, 24

[53] Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, IEEE Transactions on Information Theory, 13 (2), 260–269. 10

[54] Wells, C., 1996. The Kalman Filter in Finance, Kluwer Academic Publishers, Dordrecht. 19

[55] Xi, X., Mamon, R., 2011. Parameter estimation of an asset price model driven by a weak hidden Markov chain, Economic Modelling, 28, 36–46. 7, 24

[56] Xi, X., Mamon, R., 2014. Parameter estimation in a WHMM setting with independent and volatility components, in: Hidden Markov Models in Finance: Volume II (Further Developments and Applications) (eds.: Mamon, R. and Elliott, R), Springer, 227–240. 7

[57] Zakai, M., 1969. On the optimal filtering of diffusion processes, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 11 (3), 230–243. 10

2

Filtering and forecasting commodity futures prices under an HMM framework

2.1 Introduction

In recent years, there have been various deregulations occurring in the electricity, oil and natural gas markets. Apparently, prices of these commodities reflect financial risks that are borne out by the market participants (sellers and buyers). Such risks are important considerations when proposing a model for the price evolution of these commodities especially in designing energy derivative contracts.

The modelling of commodity futures prices and their underlying variables was studied by various authors in the light of various financial modelling considerations and objectives. Cortazar et al. [8] proposed a multicommodity model for futures prices that allows the use of long-maturity futures prices available for one commodity to estimate futures prices of another commodity; Kalman filtering was used in the model estimation. Nakajima and Ohashi [22] put forward a commodity pricing model that incorporates the effect of linear relations among commodity spot prices, and provided a condition under which such linear relations represent cointegration; using crude oil and heating oil market data, Kalman filtering was also utilised to estimate the model parameters. In Antonio et al. [1], the inclusion of jump component is carried out to explain the behaviour of oil prices; this, however, creates difficulties in the estimation of

state variables, and so particle filters were applied instead of Kalman filters. Mirantes et al. [21] formulated a generalised multi-factor model (n non-seasonal factors and m seasonal factors) for the stochastic behaviour of commodity prices, which nests the deterministic seasonal models; the seasonal factors are trigonometric components driven by random processes. A one-factor regime-switching model was developed in Chen and Forsyth [7] but the objective was to capture the risk-adjusted natural gas spot price dynamics; regression was used in model calibration using both market data on futures and options on futures. Back et al. [2] conducted an extensive analysis covering samples of soybean, corn, heating oil and natural gas options, and provided evidence that seasonality in volatility is an important aspect to consider when valuing futures contracts; an appropriate seasonality adjustment significantly reduces pricing errors in these markets and yields more improvement in valuation accuracy than increasing the number of stochastic factors.

The contributions of this research differ from the previous works mentioned above. We aim to address (i) the development of a model for the evolution of arbitrage-free futures prices suitable for valuation of commodity derivatives and (ii) provision of a regime-switching framework with HMM-based dynamic estimation for the modelling of multivariate commodity price time series along with the investigation of its various implementation issues. Specifically, we proposed an approach for the estimation of latent state variables in a model employed in futures pricing. The model is adapted from Manoliu and Tompaidis [20] under a framework that is consistent with no-arbitrage pricing. The methodology in [20] leads to a state-space formulation of the futures price model suited for Kalman filtering and maximum likelihood method. More specifically, the state variable for the spot and futures model is an Ornstein-Uhlenbeck process designed to capture mean-reversion and observed term structure of volatilities and correlation. Under this model, the futures prices are lognormally distributed. We start from a lognormal spot price process and derive a multivariate equation for futures prices. Instead of using a constant parameter (possibly multi-factor) mean-reverting process, we allow the model parameters to be modulated by a finite-state hidden Markov chain in discrete time. The parameters could then switch dynamically amongst economic regimes representing the interactions of various factors including mean-reversion and cyclical patterns (seasonality) in commodity prices.

The usual method of finding the maximum likelihood parameter estimates (MLEs)

in conjunction with Kalman filtering is to numerically maximise the likelihood function. In Elliott and Hyndman [11], a filter-based implementation of the expectation maximisation (EM) algorithm that can be used to find the MLEs is presented. Such approach makes use of the change of measure technique to evaluate filters under an ideal measure and relate the calculations back to the real-world through the Baye's theorem. In recent years, linear and non-linear filtering have found a large number of applications in finance. A recent survey of developments in this area along with various implementation details in the context of financial modelling is featured in Date and Ponomareva [9].

Considering the multivariate nature of datasets for correlated futures prices, we utilise the filtering and estimation for vector observations put forward in Erlwein et al. [15]. The novelty of this work stems from the utilisation of all possible price information from the futures market to obtain model parameters. Our estimation procedures are designed to suitably calculate the $h-$step ahead forecasts of various related financial variables. We formulate a model that is compatible with the Erlwein, et al.'s [15] framework.

This chapter is structured as follows. Section 2.2 presents the formulation of the model for the evolution of arbitrage-free futures prices. In section 2.3, the filtering algorithms for parameter estimation are outlined. We provide in section 2.4 a numerical implementation by applying the algorithms to a data set of futures prices. We investigate the forecasting performance of our approach in predicting log returns and future prices. Finally, some concluding remarks are given in section 2.5.

2.2 Arbitrage-free evolution of futures prices

In this section, we provide a brief outline of the development of an arbitrage-free model of futures price dynamics. Modelling the arbitrage-free price evolution is essential to appropriately price securities in the commodity markets such as spread options. Specification of an arbitrage-free model is necessary to be consistent with the risk-neutral approach in pricing. The development is based on reference of Manoliu and Tompaidis [20] and the omitted proofs follow using a univariate version of Itô's lemma in a straightforward fashion.

We assume that the log-spot price ζ_t follows a single-factor mean-reverting process under the risk neutral measure (or $\mathbb{Q}$ measure, in conventional notation), i.e.,

$$d\zeta_t = (\alpha - \kappa\zeta_t)dt + \theta dW_t. \tag{2.1}$$

Here, α, κ and $\theta > 0$ are assumed constants and W_t is a $\mathbb{Q}$-Wiener process. The spot price is considered to be a latent state, i.e., unobservable. We assume further that at each time prices of m futures are available with maturities $T_1, T_2, \cdots, T_m$. The price of the futures contract with maturity T_i , at time $t < \min_i(T_i)$, is denoted by $F^i(t)$ and can be written as

$$F^i(t) = \mathbb{E}_{\mathbb{Q}}(e^{\zeta^i}|\mathcal{F}_t) \;=\; \exp\left(\mathbb{E}_{\mathbb{Q}}(\zeta^i|\mathcal{F}_t) + \frac{1}{2}\operatorname{Var}_{\mathbb{Q}}(\zeta^i|\mathcal{F}_t)\right),$$

where we denote ζ_{T_i} by ζ^i for notational brevity and use the fact that e^{ζ^i} is log-normal. Here, $\{\mathcal{F}_t\}$ is the filtration generated by W_t. This leads to a closed-form expression for $F^i(t)$ given by

$$F^i(t) = \exp\left(e^{-\kappa(T_i-t)}\zeta_t + \frac{\alpha}{\kappa}(1 - e^{-\kappa(T_i-t)}) + \frac{\theta^2}{4\kappa}\left(1 - e^{-2\kappa(T_i-t)}\right)\right). \tag{2.2}$$

Our modelling formulation is consistent with the log-spot price modelling assumptions in Manoliu and Tompaidis's paper [20]. Under the assumption of no-arbitrage valuation, the benefits from holding the physical asset, called convenience yield, are reflected in the futures price (cf. Hull [18]). Convenience yield in a commodity is analogous to a dividend in an asset that provides a known income; and dividends would naturally lead to a corresponding adjustment in the price of the underlying asset of a futures contract. Whilst we did not explicitly model the dynamics of the convenience yield as it is not the intent of this paper to quantify it, we see that it is implicitly taken into account through the log-spot price ζ in the closed-form expression for the futures price in equation (2.2).

Furthermore, at a fixed time t, $F^i(t)$ can be an increasing or a decreasing function of maturity T_i, depending on the choice of parameters, which can easily be seen from equation (2.2). Futures prices decreasing (respectively, increasing) with maturity reflects backwardation (respectively, contango). Typically, being able to model both these situations adequately is a reason for modelling net convenience yield explicitly (possibly as a stochastic process). We can achieve a switch between the two situations in our framework through updating model parameters via self calibration as well as through

regime switching, as will be made clear in the subsequent sections.

Next, we assume that the log-spot price follows a mean-reverting process with a different drift function and the same volatility in the objective measure (or $\mathbb{P}$ measure):

$$d\zeta_t = (\widetilde{\alpha} - \kappa\zeta_t)dt + \theta d\widetilde{W}_t, \tag{2.3}$$

where $\widetilde{W}_t$ is a $\mathbb{P}$-Wiener process. We assume the market to be arbitrage-free, which implies that there exists a *price of risk* process λ_t such that $\tilde{\alpha} - \alpha = \lambda_t\theta$ holds. For the time being, we assume $\lambda_t =: \lambda$ to be a constant. We apply Itô's lemma to $F^i(t)$, using equations (2.2) and (2.3), which leads to the following arbitrage-free dynamics for log-futures price:

$$d\left(\log F^i(t)\right) = e^{-\kappa(T_i-t)}\left(\lambda\theta - \frac{\theta^2}{2}e^{-\kappa(T_i-t)}\right)dt + \theta e^{-\kappa(T_i-t)}d\tilde{W}_t. \tag{2.4}$$

In the subsequent discussion, we shall assume that the parameters λ and θ are dependent on the current regime, and regime switching is allowed over time. However, the discussion on regime switching is postponed to section 3 and we will assume the parameters to be constant for the purpose of this section.

For calibration and forecasting purposes of a multivariate time series of futures prices, we use a moment-matching procedure to implement equation (2.4) in discrete time. We suppose that observation times $t_1 \leq t_2 \leq \ldots \leq t_N$ are equally spaced and $t_{k+1} - t_k =: \Delta$. To write the dynamics of a vector of futures prices in a compact form, let $\text{vec}\{a^i\}$ denote a vector with a^i at its ith element. Then, at time k, the arbitrage-free evolution of the futures price vector of log-returns is given by

$$\text{vec}\left\{y_k^i\right\} \quad = \quad \text{vec}\left\{f_k^i\right\} + \text{vec}\{q_k^i\}z_k, \tag{2.5}$$

where

$$f_k^i := \frac{\lambda\theta}{\kappa}e^{-\kappa(T_i-t_k)}\left(1 - e^{-\kappa\Delta}\right) - \frac{\theta^2}{4\kappa}e^{-2\kappa(T_i-t_k)}\left(1 - e^{-2\kappa\Delta}\right), \tag{2.6}$$

$$y_k^i := \log\frac{F^i(t_k)}{F^i(t_{k-1})}, \tag{2.7}$$

$$q_k^i = \theta e^{-\kappa(T_i-t_k)}\sqrt{\frac{1 - e^{-2\kappa\Delta}}{2\kappa}} \tag{2.8}$$

and $\{z_k\}$ is a sequence of independent Gaussian random variables with zero mean and unit variance. The above discrete-time implementation preserves the exact distribution

of $\log F^i(t_{k+1})$, conditional on $\mathcal{F}_k$ (i.e., information up to time t_k). Hence, it is preferred over more conventional Euler discretisation.

Although our formulation assumptions are similar to those in Manoliu and Tompaidis in [20], our approach differs significantly. The work in [20] relies on modelling the latent spot price evolution explicitly and then using Kalman filtering methodology to extract this latent price. Single as well as multi-factor models are used in [20] and a non-parametric seasonality adjustment is suggested for forecasting. In our case, the spot price does not feature in the futures equation and is modelled only implicitly. We incorporate regime switching to allow for factors such as seasonality and use a self-calibrating filter which is adapted from Erlwein, et al. [15] for our particular model structure. It will be demonstrated that a single-factor, two-regime model gives satisfactory results for a chosen data set.

Our aim is to obtain estimates for the parameters λ, κ and θ along with the transition probabilities if any of these parameters are governed by a discrete-time finite-state Markov chain. We solve this problem in two steps:

1. Initial parameter estimation: In this step, we assume that the transition probability matrix is identity over the observed time series and identify the initial parameter estimates for λ, κ and θ by maximising the likelihood of the observed time series. Let these initial estimates be denoted by λ_0, κ_0 and subsection 2.3.1.

2. Update of parameters and transition probabilities using a self-calibrating filter: To make the implementation tractable, we assign an appropriate fixed value for κ and hence, κ is assumed to be independent of the Markov chain. The estimates of λ and θ are updated using a self-calibrating filter. The implementation steps for this filter are derived in subsection 2.3.2.

2.3 Filtering and model parameter estimation

2.3.1 Initial estimates of parameters

To find the initial estimates of parameters, we assume that the system is operating in a single regime, i.e., the transition probability matrix for the Markov chain is identity. Suppose that data on futures prices is available for times $k = 1, 2, \ldots, n$, from which a vector-valued time series of log returns, $\text{vec}(y_1^i), \text{vec}(y_2^i), \ldots \text{vec}(y_n^i)$ can be constructed

for $i = 1, 2, \ldots, m$.

In equation (2.5) the components f_k^i and q_k^i are parametrised by θ, κ and λ. Since we are assuming that z_k are IID standard normal random variables, the likelihood function of y_k^i is given by

$$L\left(y_k^i; \lambda, \kappa, \theta\right) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi} q_k^i} \exp\left(-\frac{(r_k^i)^2}{2(q_k^i)^2}\right),$$

where

$$r_k^i := y_k^i - f_k^i. \tag{2.9}$$

For each futures maturity T_i, we can therefore find estimates of the parameters λ, κ and θ by maximising the likelihood of observations, *i.e.*, by solving the non-convex optimisation problem

$$\min_{\lambda,\kappa,\theta>0} \sum_{k=1}^{n} \left(\log q_k^i + \frac{(r_k^i)^2}{2(q_k^i)^2}\right). \tag{2.10}$$

To find a common set of parameters λ, κ and θ which maximise the likelihood of observations for all futures with all available maturities $(T_1, T_2, \cdots, T_m)$ *simultaneously*, one may follow multi-objective optimisation approach and seek a set of parameters such that likelihood for any T_i cannot be improved upon without decreasing the likelihood of observations for a different T_i. As we are going to update the parameters using a self-calibrating filter later, we reject this numerically involved approach. Instead, a simpler approach is sought to obtain a set of parameters which minimise the sum of negative log likelihood of observations for all futures, *i.e.* we solve the following optimisation problem:

$$\min_{\lambda,\kappa,\theta>0} \sum_{i=1}^{m} \sum_{k=1}^{n} \left(\log q_k^i + \frac{(r_k^i)^2}{2(q_k^i)^2}\right). \tag{2.11}$$

To obtain initial values, let $\{\lambda_0,\ \kappa_0, \theta_0\}$ be any set of locally minimising arguments. In the succeeding dynamic estimation, *i.e.*, updating, we shall keep κ fixed and assume that the remaining two parameters λ and θ depend on a finite-state Markov chain. The estimation and update of the transition probabilities of this Markov chain and values of other parameters corresponding to different states are discussed in the next section. The arbitrage-free futures price dynamics structure is preserved under the Markov-switching set-up because λ, being the price of risk, is dependent on f_k^i and q_k^i as shown

in equations (2.6) and (2.8). In Elliott et al. [10], for instance, a regime-switching random Esscher transform is constructed, which is dependent on the time-varying drift and volatility levels, to determine an equivalent martingale pricing measure.

2.3.2 Derivation of self-calibrating filter

We reformulate the problem in the standard form used in the literature on regime-switching models; see, for example, Buffington and Elliott [3]; Elliott et al. ([10] and [13]); Erlwein et al. ([14] and [15]), amongst others. Recall that the ith component of the vector in equation (2.5) can be written as

$$y_{k+1}^i = f_k^i + q_k^i z_{k+1}. \tag{2.12}$$

Let $\mathbf{x}_k$ be a finite-state homogeneous Markov chain in discrete time, i.e., $k = 0, 1, 2, \ldots$. The semi-martingale representation of $\mathbf{x}_k$ is given by

$$\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \boldsymbol{\epsilon}_{k+1}, \tag{2.13}$$

where $\mathbf{\Pi}$ is the transition probability matrix and $\boldsymbol{\epsilon}_{k+1}$ is martingale increment. Our observation process is $m-$dimensional (one price observation for each maturity) and the component i follows the dynamics given in (2.12). As mentioned earlier, z_k are IID random variables which are also independent from the Markov chain $\mathbf{x}_k$ driving the regime-switching dynamics of the mean and volatility parameters for each observation component.

To simplify considerably the algebra involved in the filtering equations, we associate the state space of $\mathbf{x}_k$ with the canonical basis of $I\!R^N$, which is the set of unit vectors $\mathbf{e}_r$, $r = 1, 2, \ldots, N$ and $\mathbf{e}_r$ is a vector having 1 in its rth entry and 0 elsewhere. So in equation (2.12), $f^i(\mathbf{x}_k) = \langle \mathbf{f}_k^i, \mathbf{x}_k \rangle$ and $q^i(\mathbf{x}_k) = \langle \mathbf{q}_k^i, \mathbf{x}_k \rangle$, where $\mathbf{f}_k^i = (f_1^i, f_2^i, \ldots, f_N^i)^\top \in I\!R^N$ and $\mathbf{q}_k^i = (q_1^i, q_2^i, \ldots, q_N^i)^\top \in I\!R^N$. The notation $\langle \cdot, \cdot \rangle$ is the usual scalar product and $\top$ denotes the transpose of a vector.

To obtain the optimal estimates of $\mathbf{x}_k$ using the observation process, we employ the change of reference probability technique. In this technique, we perform the calculations under $\widetilde{P}$ measure whereby the observations y_k's are $N(0, 1)$ IID sequence of random variables, and y_k is independent from $\mathbf{x}_k$. All components of the $m-$dimensional observation process have the same underlying Markov chain.

The change of reference probability in our framework utilises a discrete-time version of the Girsanov's theorem. The real-world measure P, under which we observe our measurements, can be recovered from $\widetilde{P}$ through the construction of the Radon-Nikodym derivative

$$\Lambda_k := \frac{dP}{d\widetilde{P}}\Big|_{\mathcal{F}_k} = \prod_{i=1}^{m}\prod_{l=1}^{k}\lambda_l^i,\ k \geq 1,$$

where

$$\Lambda_0 = 1 \ \text{ and } \ \lambda_l^i = \frac{\phi[q^i(\mathbf{x}_{l-1})^{-1}(y_l^i - f^i(\mathbf{x}_{l-1}))]}{q^i(\mathbf{x}_{l-1})\phi(y_l^i)}$$

with ϕ being the $N(0,1)$ density. All parameters are dependent on the same Markov chain and react to the same underlying price information. In some sense, the components are correlated through the Markov chain. Whilst the noise terms for the individual components are uncorrelated, the correlation structure of the futures prices are encapsulated in each noise. This simplification is made to make the model tractable.

We present the filter equations under a multivariate setting. Our goal is to provide adaptive filters for the estimates of the states and other auxiliary processes related to the Markov chain. Let $\mathcal{F}_k^y$ be the filtration generated by the log-returns process y_k. To find the conditional distribution of $\mathbf{x}_k$ given $\mathcal{F}_k$ under P, we write

$$\widehat{p}_k^i := P(\mathbf{x}_k = \mathbf{e}_r|\mathcal{F}_k^y) = E[\langle \mathbf{x}_k, \mathbf{e}_r\rangle\,|\mathcal{F}_k^y],$$

where $\widehat{\mathbf{p}}_k = (\widehat{p}_k^1, \widehat{p}_k^2, \ldots, \widehat{p}_k^N)^\top \in I\!R^N$. Now,

$$\widehat{\mathbf{p}}_k = E[\mathbf{x}_k|\mathcal{F}_k^y] = \frac{\widetilde{E}[\Lambda_k\mathbf{x}_k|\mathcal{F}_k^y]}{\widetilde{E}[\Lambda_k|\mathcal{F}_k^y]}$$

by the Bayes' theorem for conditional expectation. Let $\mathbf{c}_k = \widetilde{E}[\Lambda_k\mathbf{x}_k|\mathcal{F}_k^y]$ and note that $\sum_{r=1}^{N}\langle \mathbf{x}_k, \mathbf{e}_r\rangle = 1$. Thus,

$$\sum_{r=1}^{N}\langle \mathbf{c}_k, \mathbf{e}_r\rangle = \sum_{r=1}^{N}\Big\langle \widetilde{E}[\Lambda_k\mathbf{x}_k|\mathcal{F}_k^y], \mathbf{e}_r\Big\rangle = \widetilde{E}\left[\Lambda_k\sum_{r=1}^{N}\langle \mathbf{x}_k, \mathbf{e}_r\rangle\,\middle|\,\mathcal{F}_k^y\right] = \widetilde{E}[\Lambda_k|\mathcal{F}_k^y]. \quad (2.14)$$

The construction of $\mathbf{c}_k$ along with equation (5.6) yields

$$\widehat{\mathbf{p}}_k = \frac{\mathbf{c}_k}{\sum_{r=1}^{N} \langle \mathbf{c}_k, \mathbf{e}_r \rangle}.$$

Write $\widehat{G}_k := E[G|\mathcal{F}_k^y]$ for any $\mathcal{F}_k^y-$adapted process G_k. We denote the conditional expectation under $\widetilde{P}$ of $\Lambda_k G_k$ by $\gamma_k(G_k) := \widetilde{E}[\Lambda_k G_k|\mathcal{F}_k^y]$. The adaptive filters will enable the model parameters to adjust to current market conditions. We give the recursive filters for: (i) $(J^{sr}\mathbf{x})_k$, the process related to the Markov chain's jumps up to time k; (ii) $(O^r\mathbf{x})_k$, the process related to Markov chain's occupation time; and (iii) $(T^r(g)\mathbf{x})_k$, an auxiliary process related to $\mathbf{x}$ and for some function g.

The results for the recursive filters, which are modifications of those given in Erlwein et al. [15] when there is only one uniform source of noise for each component of the observation vector, are presented in the Appendix.

The Expectation-Maximisation (EM) algorithm is applied to calculate the optimal estimates of the model parameters. Such calculations result to expressions that involve the use of adaptive filters related to the Markov chain process provided in Proposition 1. Given the recursive filters in equations (2.16), (5.11), (5.12) and (5.13), the model parameters are updated every time new information arrives. Proposition 2 in the Appendix, whose proof is described in Erlwein et al. [15], gives the optimal parameter estimates computed using the EM algorithm in terms of the filters.

In subsection 2.4.1, we discuss the updating of λ (as f is affine in λ) and then use the nonlinear relationship between the parameters to arrive at an updated value of θ. It has to be noted that we fix κ, update λ, after which update θ given an estimate of f^i, i.e., the updates do satisfy constraints (at least approximately). The constraint, in fact, is our route to updating θ as there is no direct means to update it.

2.4 Numerical implementation

We illustrate our method by applying it to the dataset of daily log-returns series of heat oil future contracts compiled by Data Stream. The data were recorded from 19 June 2009 to 17 August 2011 with ten maturity dates. These maturity dates, denoted

	T=29/07/11	T=30/06/11	T= 31/05/11	T=29/04/11	T=31/03/11
Minimum	-0.0369	-0.0372	-0.03739	-0.03721	-0.0369
Maximum	0.0474	0.0478	0.0481	0.04812	0.0480
Median	0.0001	0.0001	0.0001	0.0000	0.0000
Mode	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	0.0003	0.0003	0.0003	0.0003	0.0003
Std Dev	0.0146	0.0147	0.0149	0.0150	0.0150
Skewness	-0.0360	-0.0342	-0.0312	-0.0252	-0.0219
Kurtosis	0.0054	-0.0023	-0.0116	-0.0190	-0.0329
	T=28/02/11	T=31/01/11	T=31/12/10	T=30/11/10	T=29/10/10
Minimum	-0.0366	-0.0369	-0.0372	-0.0379	-0.0384
Maximum	0.0477	0.0479	0.0481	0.0486	0.0490
Median	0.0000	0.0000	0.0000	0.0000	0.0000
Mode	0.000	0.000	0.0000	0.0000	0.0000
Mean	0.0003	0.0003	0.0002	0.0002	0.0003
Std Dev	0.0151	0.0152	0.0154	0.0157	0.0160
Skewness	-0.0238	-0.0254	-0.0242	-0.0240	-0.0240
Kurtosis	-0.0479	-0.0579	-0.0745	-0.0899	-0.1089

Table 2.1: Descriptive statistics for the log-returns of futures price for the entire dataset

by T_i, $i = 1, \ldots, 10$, are the last trading days of the month from 29 October 2010 to 29 July 2011. Table 2.1 show the descriptive statistics of the entire futures prices data given their maturities. In Tables 2.2 and 2.3, the descriptive statistics for the log-returns of futures price for the respective periods of 29/06/2009–14/07/2009 and 15/07/2009–24/07/2009 are shown. The suitability of the regime-switching model is clearly demonstrated by Tables 2.2 and 2.3, where the sample moments are statistically different for the two non-overlapping periods.

Indeed, the log-returns of our multivariate data undergo regime changes in mean and volatility levels. To model such regime-switching behaviour, we assume that for every maturity date, log-return's mean f and volatility q, the daily price process $F^i(t_k)$ is of the form

$$y^i_{k+1} = \ln \frac{F^i(t_k)}{F^i(t_{k-1})} = f(x_k) + q(x_k) z_{k+1}$$

corresponding to each maturity date T_i.

	T=29/07/11	T=30/06/11	T= 31/05/11	T=29/04/11	T=31/03/11
Minimum	-0.0268	-0.0273	-0.0278	-0.0283	-0.02870
Maximum	0.0070	0.0071	0.0072	0.0072	0.0069
Median	-0.0086	-0.0087	-0.0088	-0.0088	-0.0092
Mean	-0.0104	-0.0106	-0.0107	-0.0108	-0.0110
Std Dev	0.0106	0.0108	0.0109	0.0110	0.0111
Skewness	-0.0353	-0.0481	-0.0518	-0.0694	-0.0960
Kurtosis	-0.8898	-0.8852	-0.8639	-0.8379	-0.8404
	T=28/02/11	T=31/01/11	T=31/12/10	T=30/11/10	T=29/10/10
Minimum	-0.0292	-0.0297	-0.0304	-0.0314	-0.0324
Maximum	0.0066	0.0064	0.0062	0.0057	0.0055
Median	-0.0096	-0.0101	-0.0104	-0.0108	-0.0109
Mean	-0.0112	-0.0114	-0.0116	-0.0119	-0.0122
Std Dev	0.0111	0.0113	0.0114	0.0116	0.0119
Skewness	-0.1342	-0.1574	-0.1832	-0.2347	-0.2802
Kurtosis	-0.8255	-0.8300	-0.8142	-0.7892	-0.7674

Table 2.2: Descriptive statistics for the log-returns of futures price for the period 29/06/2009–14/07/2009

	T=29/07/11	T=30/06/11	T= 31/05/11	T=29/04/11	T=31/03/11
Minimum	0.0009	0.0009	0.0009	0.0009	0.0009
Maximum	0.0265	0.0268	0.0270	0.0271	0.0271
Median	0.0114	0.0115	0.0116	0.0117	0.0117
Mean	0.0128	0.0130	0.0132	0.0132	0.0132
Std Dev	0.0091	0.0091	0.0093	0.0094	0.0093
Skewness	0.0293	0.0352	0.0396	0.0390	0.0389
Kurtosis	-1.3725	-1.3875	-1.4187	-1.4349	-1.4351
	T=28/02/11	T=31/01/11	T=31/12/10	T=30/11/10	T=29/10/10
Minimum	0.0009	0.0009	0.0009	0.0009	0.0009
Maximum	0.0270	0.0270	0.0272	0.0275	0.0276
Median	0.0116	0.0116	0.0117	0.0118	0.0120
Mean	0.0132	0.0132	0.0133	0.0135	0.0136
Std Dev	0.0093	0.0093	0.0094	0.0095	0.0096
Skewness	0.0385	0.0387	0.0291	0.0289	0.0226
Kurtosis	-1.4360	-1.4355	-1.4264	-1.4420	-1.4953

Table 2.3: Descriptive statistics for the log-returns of futures price for the period 15/07/2009–24/07/2009

2.4.1 Computing initial parameter estimates

As mentioned in subsection 4.3.1, we first assume that the system operates under a one-regime setting. This allows us to find starting values λ_0, κ_0 and θ_0. To simplify the implementation, we assume that κ is constant. Thus, using $\kappa(t) = \kappa_0$ for all time $t > 0$, the evolution of the process λ_t is derived under the multi-regime modelling set-up. This would in turn update θ_t given an estimate of f^i as shown in equation (2.6).

Whilst we made the assumption that κ is constant to achieve simplicity in the implementation, such assumption is actually justified empirically. That is, κ appears not to depend on a Markov chain and remains constant through time for any quantity of regimes. We found that if the size of the data used to estimate the initial parameters is varied, both λ_0 and θ_0 change but κ appears stable. As it is computationally intensive to calculate starting values of λ, κ, θ for various combinations of window sizes for our dataset, we randomly select a few sub-dataset samples from the original data and then use these samples to estimate initial values. The processing of data via a moving lag window is discussed in subsection 2.4.2. Our numerical results show that there is not much perturbation in the estimated κ values. This indeed supports the assumption that κ can be taken as constant; see Table 2.4.

Following the calculation of the initial parameters of the model outlined in subsection 2.3.1, we first solve the minimisation problem in (2.11) using a standard function fminsearch in MATLAB. The built-in algorithm in fminsearch is quite fast, but it provides arguments which minimise the function in (2.11) only locally. Since this function exhibits fluctuating behaviour, we have to deal with many local extreme points. This implies that one has to search for a global minimum at least within a reasonably big subset of $\mathbb{R}^3$. We observed that the function fminsearch always finds an extreme point which is closest to the initial value supplied. A subset of $\mathbb{R}^3$ for the space of the initial values is therefore considered and a minimum point in that space is determined. The results of our estimations for five sub-dataset samples are displayed in Table 2.4 and again, it is clear that the assumption of a constant κ is reasonable. In our HMM filtering implementation, we set $\kappa = 0.00057$, being the average of the calculated initial values for κ_0. It has to be noted that the parameter θ does not vary a lot either. This fact provides additional support for the validity of the model and accuracy of the filtering algorithms. As one will see in subsection 2.4.2, the values of q_k^i do not change

λ_0	κ_0	θ_0
0.0618	5.832×10^{-4}	0.0140
0.0166	5.458×10^{-4}	0.0161
0.0528	5.411×10^{-4}	0.0153
0.0248	5.594×10^{-4}	0.0149
0.0343	5.693×10^{-4}	0.0165

Table 2.4: Estimation of initial values of κ, λ and θ for five sub-dataset samples

significantly after only several iterations. This is consistent with equation (2.8) since an estimate of q_k^i gives an update of θ.

2.4.2 Implementation of self-calibrating filter

The HMM filtering algorithms were implemented with a moving lag window of various time steps to obtain estimated values for f^i and q^i. That is, we experimented to process data points in batches of 1-5 data points per batch for each algorithm step. In particular, we apply the recursive filtering equations in Proposition 1 in processing a batch of data points. Consequently, this gives estimates of the filters for various quantities related to the Markov chain that are used to provide EM estimates for model parameters in accordance with Proposition 2. The two-step process of calculating the filters and computing EM estimates constitutes the completion of one algorithm step. The two-step process is then repeated for the next moving lag window of data points. The final filtered values in the previous algorithm step are employed as initial values for the filtering equations in the succeeding algorithm step.

Each moving lag window for our filtering is assessed on the basis of goodness-of-fit metric, which is the root mean square error (RMSE). The model corresponding to a given number of regimes and moving lag window with the lowest RMSE is deemed as the most appropriate for the dataset. The results of our RMSE computations are summarised in Tables 2.5 and 2.6 for the one-regime and two-regime settings, respectively. The starting values of q are uniform for all regimes as these do not affect the convergence of the filtering algorithms.

Given any number of regimes, a pattern emerges from Tables 2.5-2.6 in that the best

Window size	Num. of regimes	St. value of f	St. value of q	RMSE
1	1	0.0000	0.02	0.312131
2	1	0.0000	0.02	0.250943
3	1	0.0000	0.02	0.342138
4	1	0.0000	0.02	0.341751
5	1	0.0000	0.02	0.350081
1	1	-0.0001	0.04	0.156124
2	1	-0.0001	0.04	0.143903
3	1	-0.0001	0.04	0.120332
4	1	-0.0001	0.04	0.110002
5	1	-0.0001	0.04	0.123421

Table 2.5: RMSE results given number of regimes, size of filtering window and starting values model parameters under a one-state setting

Window size	Num. of regimes	St. value of f_i	St. value of σ_i	RMSE
1	2	[-0.01 +0.01]	0.02	0.155199
2	2	[-0.01 +0.01]	0.02	0.105657
3	2	[-0.01 +0.01]	0.02	0.115599
4	2	[-0.01 +0.01]	0.02	0.091993
5	2	[-0.01 +0.01]	0.02	0.127700
4	2	[-0.01 +0.01]	0.03	0.090471
4	2	[-0.02 +0.02]	0.04	0.084289
4	2	[-0.05 +0.05]	0.08	0.121808
4	2	[-0.03 +0.03]	0.04	0.101739
4	3	[-0.02 0 +0.02]	0.02	0.089001
4	3	[-0.03 0 +0.03]	0.04	0.132655

Table 2.6: RMSE results given number of regimes, size of filtering window and starting values model parameters under a two-state and three-state settings

No. of regimes N	Likelihood value L	No. of parameters d	BIC
1	9278.33	2	9270.15
2	9691.38	6	9666.86
3	9700.23	8	9663.45

Table 2.7: Likelihood-based model selection analysis

RMSE value is for a window of size four. For any starting value of f^i and q^i, we observe erratic trends of the transition probability matrix. In some instances, it has constant zeros and ones especially for models with low number of states. But starting values for parameters f and q can be chosen randomly. The only restriction for the initial values is to obtain convergence in the first iteration of the filtering algorithm. These starting values for f^i and q^i cannot be either too small or too large. Working with simulated data, we also found that that algorithm converges faster to the "true" values provided the number of regimes is chosen correctly. In our case, the numerical implementation of filters under the two-regime model yields the most stable parameter estimates.

2.4.3 Discussion of numerical results

The initial parameters were estimated using random subsets of the whole data focusing on the first 6 months. The HMM filtering algorithms were implemented to the remaining datasets with a moving window of four time steps to obtain estimated values of f^i and q^i for one-, two- and three-regime models. The evolution of transition probabilities is depicted in Figure 2.1. The plots of f and q in Figure 2.2 behave as expected. Considering that the data corresponds to the period when the economy was slowly recovering from the subprime financial crisis, we observed as anticipated, slowly decaying values of mean levels and constant behaviour after that period. A similar pattern for volatilities is obtained and there is a relatively big uncertainty at earlier periods but they leveled off not too long after a certain point. From the graph of f (Figure 2.2a) under one-regime framework, it is noticeable that the "true" values of f^is are always underestimated.

To evaluate the statistical significance of the RMSE values, we perform an F test. The estimated value of the F statistic for the comparison of RMSEs between one- and two-regime models is higher than the quantile value of the F distribution based on a 95%

confidence level where the p-value is 3.91×10^{-7}. For the comparison of RMSEs between two- and three-regime models, the p-value is 3.57×10^{-11}. Hence, there is merit in using a regime-switching framework. Considering that a two-state setting produces the best RMSE in Table 2.6, we conclude that the two-state Markov switching model is the most appropriate for our data.

This is further backed up by a likelihood-based selection criterion. The popular criterion for model selection is the Akaike information criterion (AIC). However, it is argued in Schwartz [23] that AIC may underestimate the optimal number of parameters, and thus, the Bayesian information criterion (BIC) is proposed as a robust alternative. The BIC metric is given

$$\text{BIC} = \ln L - \frac{1}{2} d \ln b,$$

where L is the likelihood function, d is the number of parameters in a model and b is the number of datapoints, respectively. The main idea is to choose the model with the highest BIC value. From Table 3.30, the BIC analysis indicates that the three-regime model is marginally better than the two-regime model. However, we maintain that the decision to choose a three-regime model does not outweigh the burden of model and computational complexity.

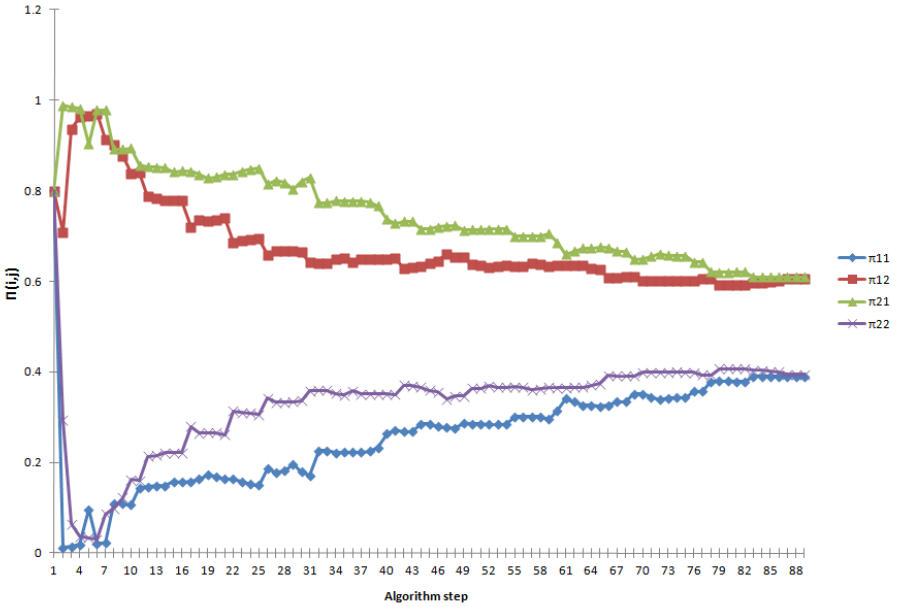
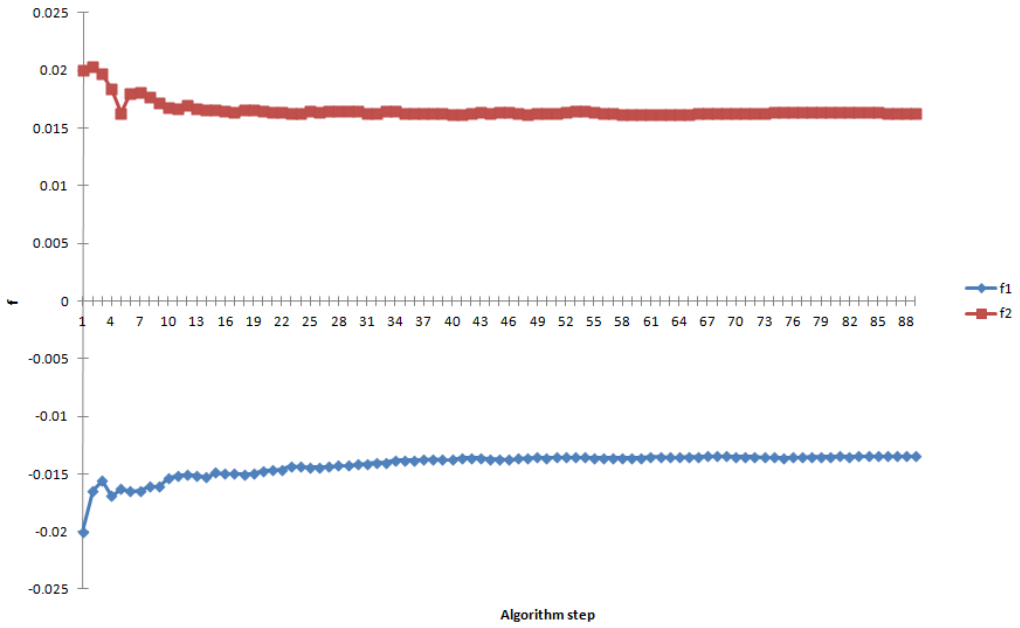
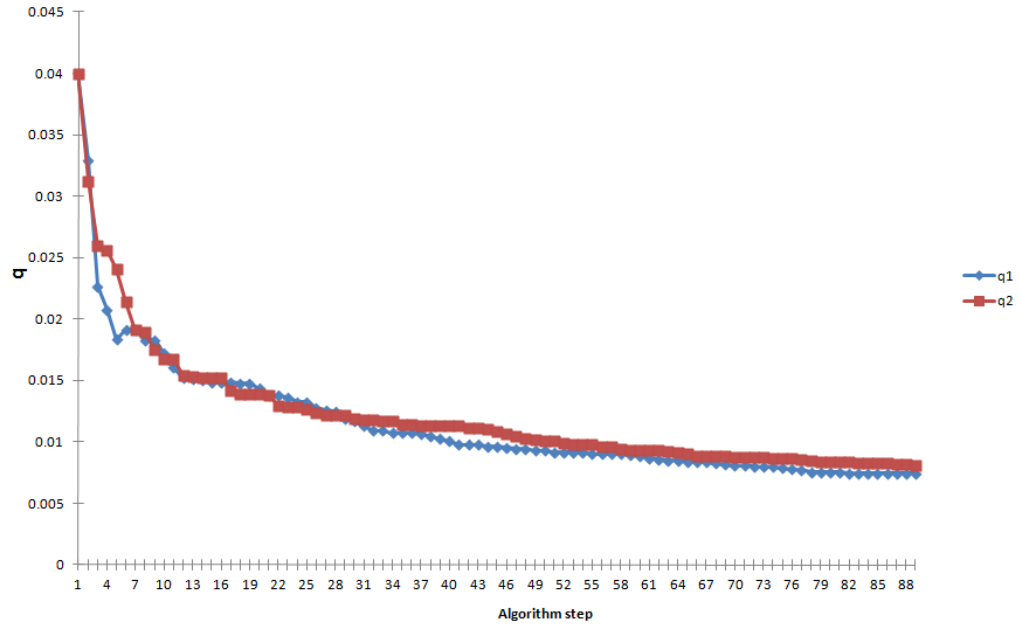


Figure 2.1: Evolution of transition probabilities



(a) Dynamics of f^i levels



(b) Dynamics of q^i levels

Figure 2.2: Parameter estimates using data prices of futures contracts with expiry 29/07/2011

The proposed model in this research is estimated by the maximum likelihood method. Hence, it appears that a likelihood ratio test (LRT) would be more appropriate in the comparison of embedded models, i.e., a lower dimensional model (say 2 regimes) is a restricted version of the higher dimensional model (say 3 regimes). However, it is noted in Hardy [17] that the likelihood ratio test is not a valid test for the number of regimes in a regime-switching model. In particular, there are theoretical problems concerning the consistency of the estimator (i.e., asymptotic level of the test) since the regularity conditions are not satisfied under the null hypothesis; thus, the chi-square theory underpinning the LRT does not apply, see Gassiat and Keribin [16]. This is further substantiated in Chen et al. [4] and [5], where it is explicitly stated that the required LRT's regularity conditions are not fulfilled under mixture problem. In fact, it is also asserted in Chen and Kalbfleish [6] that the asymptotic properties of likelihood ratio statistics for testing the number of subpopulations are complicated and difficult to establish. So, whilst there are modified LRTs tailored to regime-switching models, their implementation is quite involved and we opted to keep the model selection assessment simple by adhering to the BIC-based evaluation.

Finally, as can be seen from equation (2.6), the evolution of the market price of risk λ through time and maturity can be inferred and constructed once the estimates of other parameters are fully determined. Our numerical implementation with the use of the HMM filtering techniques and estimation of initial parameter values produce the λ dynamics given in Figure 2.4.

2.4.4 Prediction performance

The futures prices are treated as a 10-dimensional observation process. From equation (2.12), we have the dynamics of the vector process of price returns $y_k^i = \ln \frac{F^i(t_k)}{F^i(t_{k-1})}$ $i = 1, 2, \ldots, 10$. Therefore, the one-step ahead forecasts for $F^i_{t_{k+1}}$ is obtained through the forecast equation

$$E[F^i_{t_{k+1}}|\mathcal{F}_{t^i_k}] = F^i_{t_k} \sum_{j=1}^{N} \langle \hat{\mathbf{x}}_k, e_i \rangle \exp\left(f_j + \frac{q_j^2}{2}\right), \tag{2.15}$$

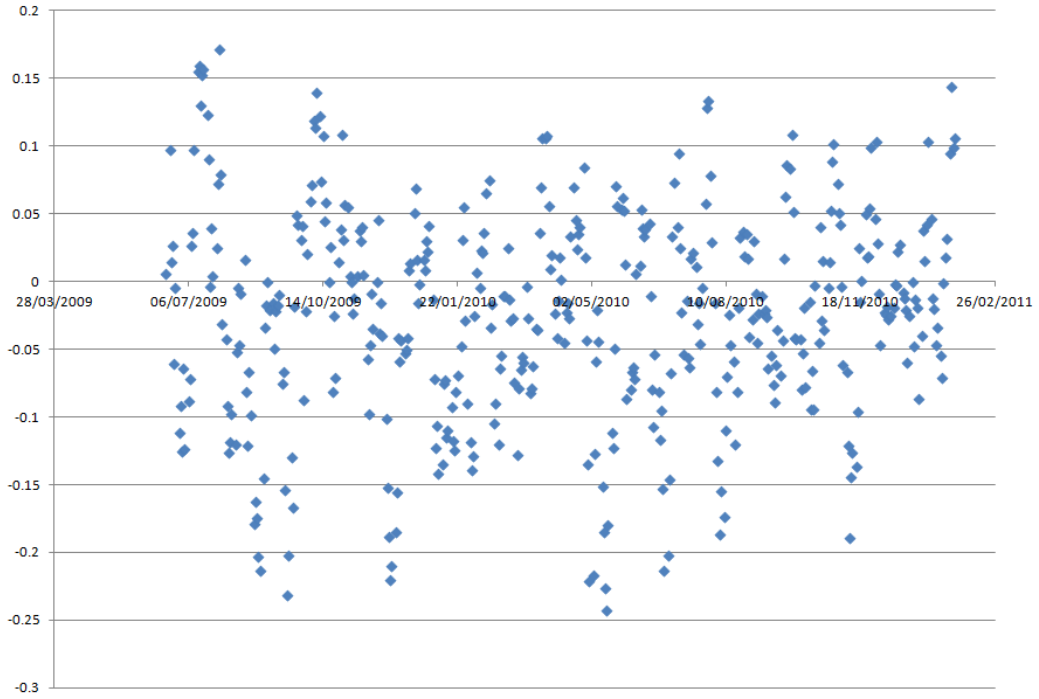
where $\hat{\mathbf{x}}_k$ is the estimate for the unconditional distribution of the Markov chain. We utilise the estimates of $f^i_{t_k}$ and $q^i_{t_k}$ to get one-step ahead forecasts for $F^i_{t_{k+1}}$. The results are shown in Figure 5a. To complement the RMSE metrics in Tables 2.5 and 2.6, the criteria in Hyndman and Koehler [19] in assessing the goodness of fit of the one-step

Model setting	MAPE	MdAPE	MdRAE
One-state model	0.02769	0.02090	1.18122
Two-state model	0.02395	0.01964	0.88347
Three-state model	0.02590	0.01921	0.94448

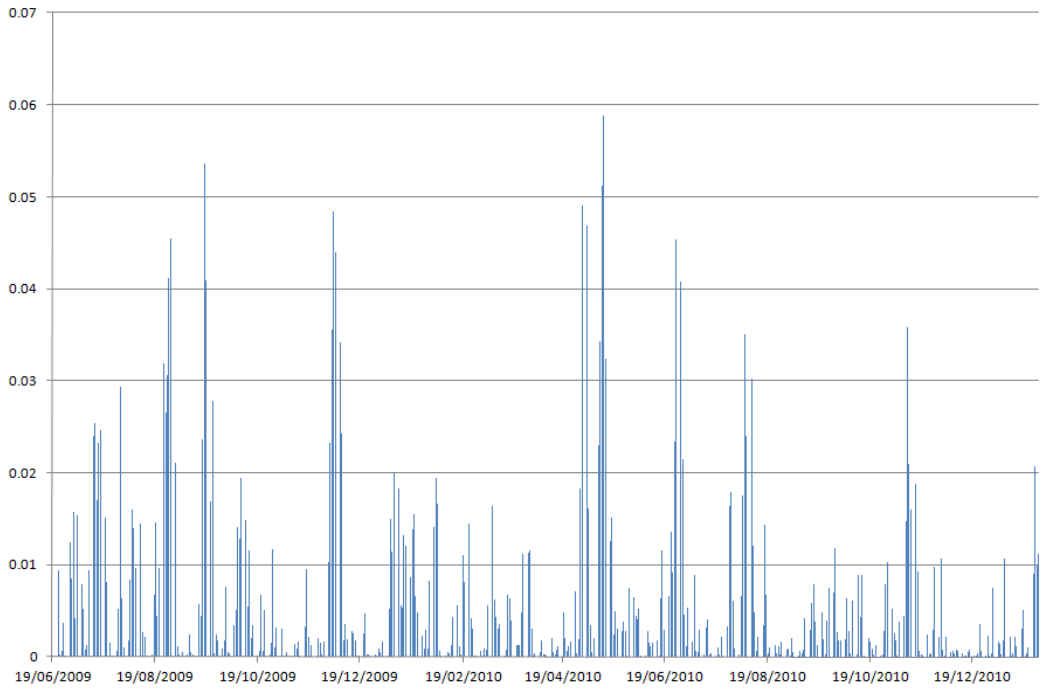
Table 2.8: Further error analysis

ahead forecasts are adopted. The mean absolute percentage error (MAPE), median absolute percentage error (MdAPE) and median relative absolute error (MdAPE), for the 1-, 2- and 3-state HMM-based models are evaluated. The models are compared using these three criteria. The results of this error analysis are presented in Table 2.8. The two-state model outperforms both the one-state and three-state models under the MAPE and MdRAE. The three-state model outperforms the two-state model under the MdAPE albeit the improvement is minimal.

Figure 2.5a displays the plots of the actual and one-step ahead predictions. A Q-Q plot depicted in Figure 2.5b strongly supports the initial assumption of using Brownian motion as a source of uncertainty in the model. From the plot of residuals against time shown in Figure 2.6a as well as Figure 2.6b, it is apparent that the assumption of constant variance is very reasonable. Whilst we would not generally expect this behaviour, one can see that from our dataset the variance does not change much for all regimes; see Figure 2.2b.



(a) Residuals



(b) Squared residuals

Figure 2.6: Residual analysis supporting the one-step ahead forecasting

2.5 Conclusion

In this work, we performed an integration of various modelling ideas to model the evolution of arbitrage-free futures prices. In particular, the initialised one-state model gives poor prediction, whilst the use of a multi-state regime switching model will necessitate finding the rate of mean reversion κ using a mechanism other than direct self calibration. A calibration of a multiple regime model with a good prediction performance is possible only through a combination of the two calibration methodologies as elaborated in subsections 2.3.1 and 2.3.2. We provided an approach and algorithms capable of providing parameter estimates for an arbitrage-free commodity futures price model. Given a data set of futures prices, the recovery of parameter estimates is carried out under the assumption that parameters shift dynamically according to the state of the economy modulated by a hidden Markov chain.

To illustrate the numerical feasibility of our approach, we focused on two- and three-regime switching modelling frameworks. We detailed the solution of determining appropriate initial parameter values by considering an approximation to a non-convex optimisation problem. Assumptions on the model parameters are verified empirically. Self-calibrating HMM filtering algorithms are then able to produce dynamic parameter estimates reflecting the switching of economic regimes. The performance of the model is deemed adequate based on the analysis of one-step ahead forecasts and the accompanying post-model diagnostics. We benchmarked our results with those from the one-state setting and both the goodness-of-fit and information criterion metrics validate the merits of using a model with regime-switching feature. Although the focus of our application is on daily data, we experimented as well on weekly data (say, Wednesday's prices). The parameter estimates as expected would change but the trends of the volatilities and drifts as depicted by their graphs remain the same. The plots of the transition probabilities have similar decaying behaviour as in Figure 2.1. Nevertheless, their evolution through algorithm steps is different. We observed that the near terminal probability values would still approach the 0.6 and 0.4 limits.

The relationship of this research with the existing recent works on modelling commodity futures price under some model characteristics is summarised in Table 2.9. Note that our method may also be employed for modelling the futures price evolution

Paper	RS***?	Mean-reversion?	SP**** observed?	FP****** observed?
[11]	No	No*	Yes	Yes
[14]	Yes	Yes	Yes	No
[20]	No	Yes	No**	Yes
This chapter	Yes	Yes	No**	Yes

*–*uses geometric Brownian motion for spot price*

**–*can include spot price as a zero-maturity futures price in both cases*

***–*regime switching*

****–*spot price*

*****–*forward price*

Table 2.9: Comparison of this research with recent existing works

of other commodities such metals, agricultural products and raw materials. The proposed model is shown to be very parsimonious, with a two-regime one-factor model, which provides adequate performance in one step ahead out-of-sample forecasting on 10 measurement variables, having only 6 free parameters and no non-parametric seasonality adjustments. The algorithms proposed in this research project provides a very useful alternative to the existing methods of futures price modelling and forecasting. Accurate forecasting commodity futures prices has important implications in various financial modelling applications such as the pricing of commodity spread options and calculation of quantile risk measures of commodity futures portfolios.

2.6 References

[1] Antonio, F., Aiube, L., Keshar, T., Baidya, N., Americo, E., Tito, H., 2008. Analysis of commodity prices with the particle filter, Energy Economics 30, 597–605. 30

[2] Back, J., Prokopzuk, M., Rudolf, M., 2013. Seasonality and the valuation of commodity options, Journal of Banking and Finance 37, 273–290. 31

[3] Buffington, J., Elliott, R., 2002. American options with regime switching, International Journal of Theoretical and Applied Finance 5, 497–514. 37

[4] Chen, H., Chen, J., Kalbfleisch, J., 2001. A modified likelihood ratio test for homogeneity in finite mixture models, Journal of the Royal Statistical Society Series B (Statistical Methodology) 63, 19–29. 49

[5] Chen, H., Chen, J., Kalbfleisch, J., 2004. Testing for a finite mixture models with two components, Journal of the Royal Statistical Society Series B (Statistical Methodology) 66, 95–115. 49

[6] Chen, J., Kalbfleisch, J., 2005. Modified likelihood ratio test in finite mixture models with structural parameter, Journal of Statistical Planning and Inference 129, 93–107. 49

[7] Chen, Z., Forsyth, P., 2010. Implications of a regime-switching model on natural gas storage valuation and optimal operation, Quantitative Finance 10, 159–176. 31

[8] Cortazar, G., Milla, C., Severino, F., 2008. A multicommodity model of futures prices: Using futures prices of one commodity to estimate the stochastic process of another, Journal of Futures Market 28, 537–560. 30

[9] Date, P., Ponomareva, K., 2011. Linear and non-linear filtering in mathematical finance: A review, IMA Journal of Management Mathematics 22, 195–211. 32

[10] Elliott, R., Chan, L., Siu, T., 2005. Option pricing and Esscher transform under regime switching, Annals of Finance 4, 423–432. 37

[11] Elliott, R., Hyndman, C., 2007. Parameter estimation in commodity markets: A filtering approach, Journal of Economic Dynamics and Control 31, 2350–2373. 32, 53

[12] Elliott, R., Chan, L., Siu, T., 2005. Option pricing and Esscher transform under regime switching, Annals of Finance 1, 423–432. 37

[13] Elliott, R., Siu, T., Chan, L., 2008. A PDE approach for risk measures for derivatives with regime switching, Annals of Finance 4, 55–74. 37

[14] Erlwein, C., Benth, F., Mamon, R., 2010. HMM filtering and parameter estimation of electricity spot price model, Energy Economics 32, 1034–1043. 37, 53

[15] Erlwein, C., Mamon, R., Davison, M., 2011. An examination of HMM-based investment strategies for asset allocation, Applied Stochastic Models in Business and Industry 27, 204–221. 32, 35, 37, 39

[16] Gassiat, E. and Keribin, C., 2000. The likelihood ratio test for the number of components in a mixture with Markov regime, ESAIM: Probability and Statistics 4, 25–52. 49

[17] Hardy, M., 2003. Investment Guarantees: Modelling and Risk Management for eEquity-Linked Life Insurance, John Wiley & Sons, Inc, New Jersey. 49

[18] Hull, J., 2011. Options, Futures, and Other Derivatives, Prentice Hall, Boston. 33

[19] Hyndman, R., Koehler, A., 2006. Another look at measures of forecast accuracy, International Journal of Forecasting 22, 679–688. 49

[20] Manoliu, M., Tompaidis, S., 2002. Energy futures: term structure models with Kalman filter estimation, Applied Mathematical Finance 9, 21–43. 31, 32, 33, 35, 53

[21] Mirantes, A., Poblacion, J., Serna, G., 2012. Stochastic seasonal behaviour of natural gas prices, European Financial Management 18, 410–443. 31

[22] Nakajima, K., Ohashi, K., 2012. A cointegrated commodity pricing model, Journal of Futures markets 32, 995–1033. 30

[23] Schwartz, G., 1978. Estimating the dimension of a model, Annals of Statistics 6, 461–464.

2.7 Appendix

Recursive filters and EM updates

Proposition 1: *Define the diagonal matrix* $\mathbf{B}$ *whose* B_{ij} *entry is given by*

$$B_{ij} = \begin{cases} \prod_{i=1}^{m} \frac{\phi\left(\frac{y_{k+1}^h - f_i^h}{q_i^h}\right)}{q_i^h \phi(y_{k+1}^h)} & for\ i = j \\ 0 & otherwise \end{cases}.$$

Then

$$\mathbf{c}_{k+1} = \mathbf{\Pi B c}_k, \tag{2.16}$$

$$\gamma\left(J^{(sr)}\mathbf{x}\right)_l = \mathbf{\Pi B}(y_l)\gamma\left(J^{(sr)}\mathbf{x}\right)_{l-1} + \langle \mathbf{c}_{l-1}, \mathbf{e}_r\rangle \left(\frac{\phi\left(\frac{y_l^i - f_r^i}{\sigma_r^i}\right)}{q_r^i \phi(y_l^i)}\right)^m \pi_{sr}\mathbf{e}_s, \tag{2.17}$$

$$\gamma\left(O^{(r)}\mathbf{x}\right)_l = \mathbf{\Pi B}(y_l)\gamma\left(O^{(r)}\mathbf{x}\right)_{l-1} + \langle \mathbf{c}_{l-1}, \mathbf{e}_r\rangle \left(\frac{\phi\left(\frac{y_l^i - f_r^i}{q_r^i}\right)}{\sigma_r^i \phi(y_l^i)}\right)^m \mathbf{\Pi e}_r, \tag{2.18}$$

and

$$\gamma\left(T^{(r)}\left(g(y)\right)\mathbf{x}\right)_l = \mathbf{\Pi B}(y_l)\gamma\left(T^{(r)}(g)\mathbf{x}\right)_{l-1} + \langle \mathbf{c}_{l-1}, \mathbf{e}_r\rangle \left(\frac{\phi\left(\frac{y_l^i - f_r^i}{\sigma_r^i}\right)}{q_r^i \phi(y_l^i)}\right)^m g(y_l^i)\mathbf{\Pi e}_r, \tag{2.19}$$

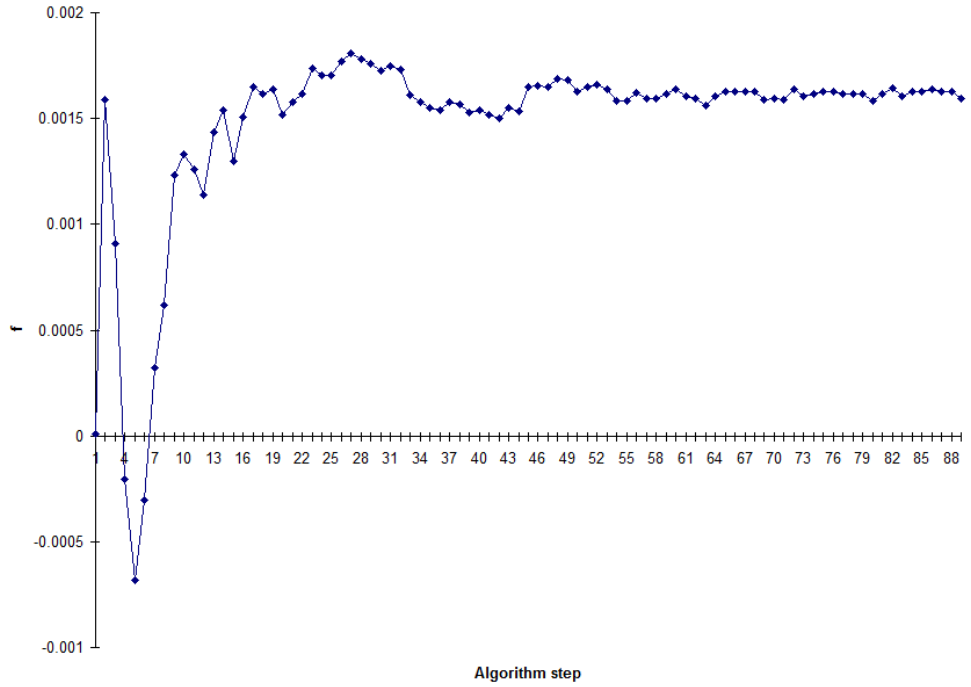
where $g(y_l^i) = y_l^i$ *or* $\left(y_l^i\right)^2$.

Proposition 2: *Consider a multivariate dataset* $y_1^i,\ y_2^i, \ldots, y_k^i,\ 1 \le i \le m$ *observed up to time* k. *If the set of parameters* $\left\{\widehat{\pi}_{sr}, \widehat{f}_r^i, \widehat{q}_r^i\right\}$ *characterises the model then the EM estimates are given by*

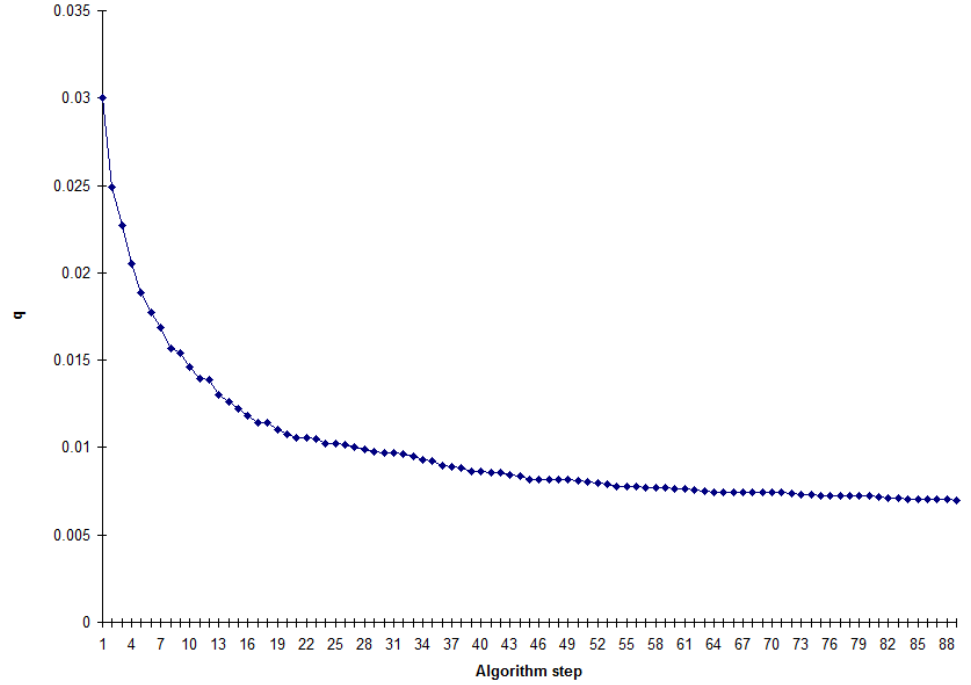
$$\widehat{\pi}_{sr} = \frac{\gamma\left(J^{(sr)}\right)_k}{\gamma\left(O^{(r)}\right)_k} \tag{2.20}$$

$$\widehat{f}_r^i = \frac{\gamma\left(T^{(r)}\left(y^i\right)\right)_k}{\gamma\left(O^{(r)}\right)_k} \tag{2.21}$$

$$\widehat{q}_r^i = \sqrt{\frac{\gamma\left(T^{(r)}\left(y^i\right)^2\right)_k - 2f_r^i\gamma\left(T^{(r)}(y^i)\right)_k + (f_r^i)^2\gamma\left(O^{(r)}\right)_k}{\gamma\left(O^{(r)}\right)_k}}. \tag{2.22}$$

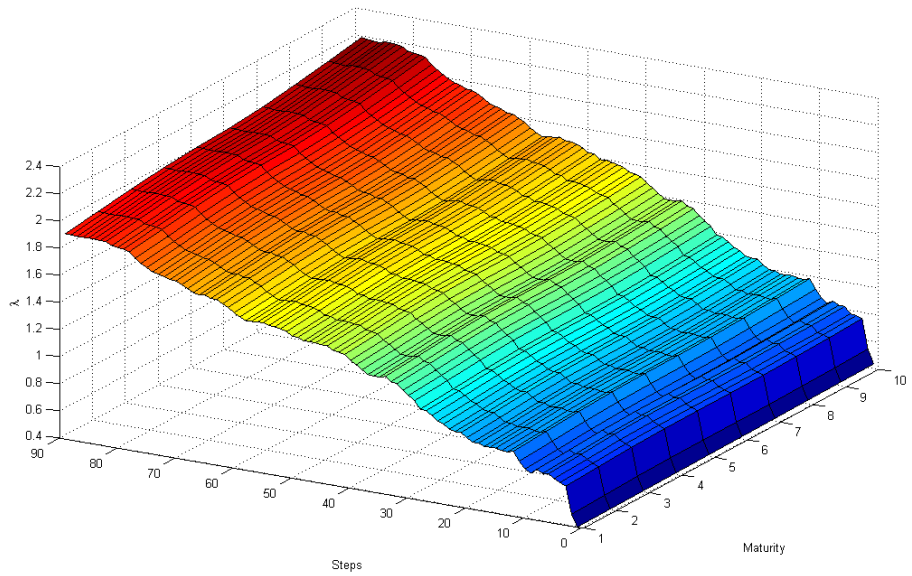


(a) Dynamics of f

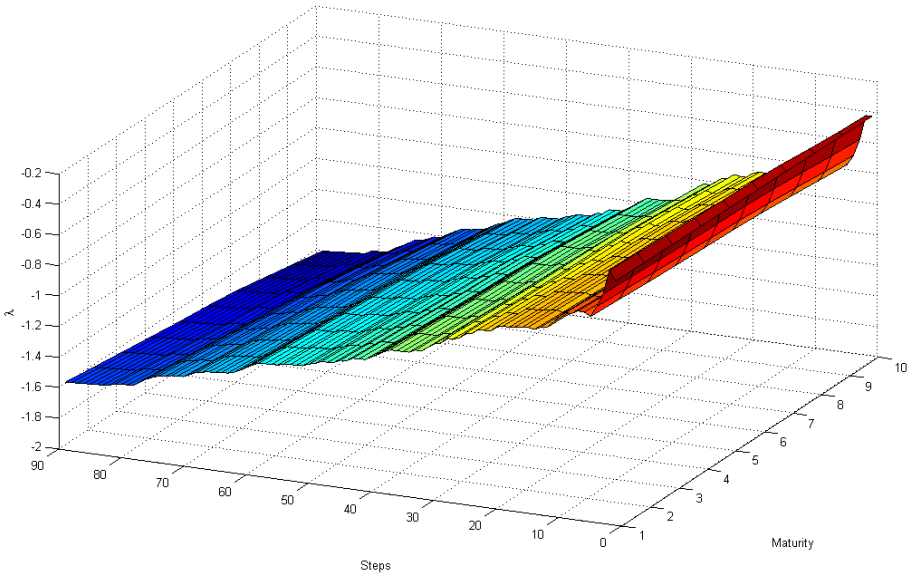


(b) Dynamics of q

Figure 2.3: Parameter estimates using data prices of futures contracts with expiry 29/07/2011 under a one-regime Markov chain

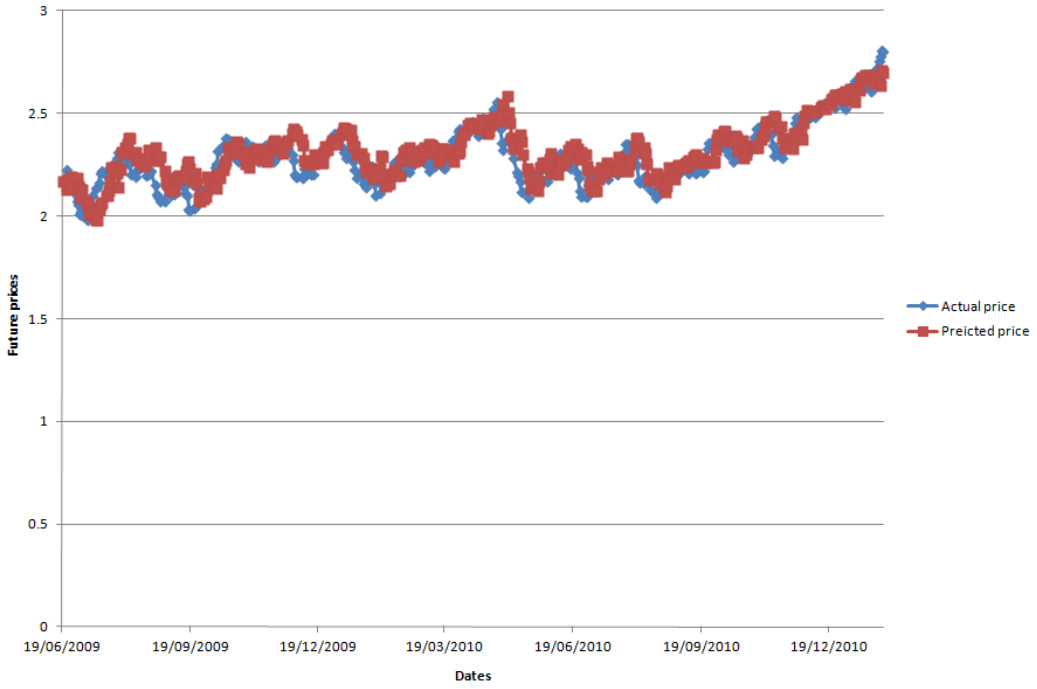


(a)

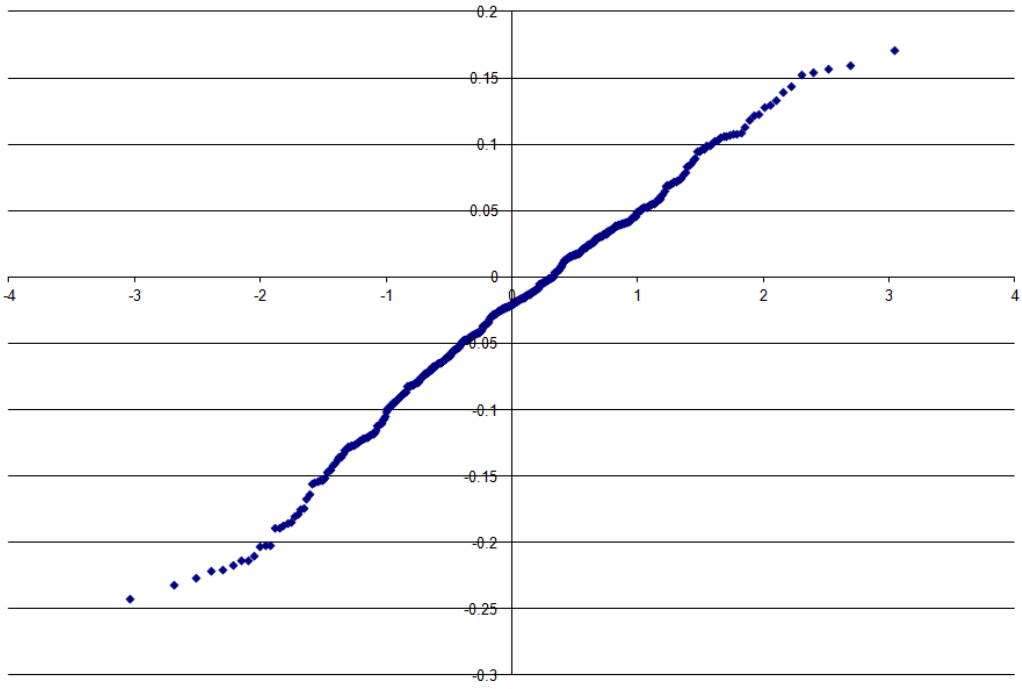


(b)

Figure 2.4: Dynamics of λ_t process under a 2-state setting corresponding to regime 1 in (a) and regime 2 in (b)



(a) One-step predictions



(b) Q-Q plot

Figure 2.5: One-step ahead forecasts and normal analysis of residuals

3

Filtering of an HMM-driven multivariate Ornstein-Uhlenbeck model with application to forecasting market liquidity

3.1 Introduction

The sale of an asset always has, to a greater or lesser degree, an effect on the market. A business's capacity to own sufficient liquid assets for the purpose of meeting its financial obligations is termed its liquidity. If an asset is able to be sold without producing drastic movements in its price and so with minimum loss of value is said to be liquid. Cash and cashable instruments are examples of liquid assets that can be used to meet immediate financial needs. Although currencies are liquid assets, even major currencies can at times suffer from severe illiquidity when they must be exchanged in the foreign exchange market. The US dollar and US dollar-linked assets, for instance, could experience market illiquidity if countries holding trillions of dollars of US bonds start dumping US dollar bonds.

The importance of dealing with liquidity problems is motivated by recent developments based on Basel III, which the US Federal Reserve uses as a liquidity requirement guideline for financial institutions. Basel III directives also require the diversification of counterparty risk and stress testing that could identify unusual market liquidity condi-

tions. The goal of regulation is to prevent investments that are particularly susceptible to sudden liquidity shifts.

In 2007, the world was deemed to have experienced the worst financial crisis since the 1930s. This crisis originated in the United States and spread across the global financial markets within less than a year. Some big financial organisations and banks declared bankruptcy. The downfall of Lehman Brothers was the most calamitous high-profile default of this crisis; see Gorton [19]. Whilst many financial market events in 2007-2008 were considered to be a direct consequence of improper credit risk management, it is also believed that the main trigger of economic turmoil was the inability to predict liquidity in the markets. In 2008, AIG had a huge portfolio of CDS and CDO that was originally rated AAA but backed by subprime loans. As a result of financial instability, the AIG products were downgraded and the company had to post additional collateral for its positions. These events are believed to be the main trigger of the liquidity crisis that began in September 2008, essentially bringing AIG to a level of bankruptcy but eventually bailed out by the US government. It is widely well-accepted that, ironically, efforts to mitigate credit/counterpaty risk could create additional illiquidity, which on its own causes instability in the financial industry. In this chapter, we propose a method of quantifying and forecasting illiquidity in the financial market. As financial turbulence cannot be avoided, warning systems that aid the prediction of economic crunches are necessary to prepare market participants to deal with future instability.

As pointed out in Goyenko [20], research on liquidity risk could be traced to two sources, namely, market liquidity and funding liquidity. Although both market and funding liquidity risks reinforced each other mutually, their mechanisms have different drivers (cf. Brunnermeier and Pedersen [5]). Chordia, et al. [8] argued that market liquidity, which refers to the easiness of trading, is asset-specific and influenced by variables that are market-wide and firm-specific. On the other hand, as asserted in [5], funding liquidity is agent-specific; this could depend overall on the borrowing constraints of dealers, hedge funds, investment banks and the availability of arbitrage capital. We wish to clarify that in this work, we develop an approach that examine several underlying mechanisms that drive both types of illiquidity.

It is set forth in [20] and the references therein that the T-bill–Eurodollar (TED)

and volatility index (VIX) are commonly used indicators of funding liquidity. Both indicators are further eleborated below, but it is worth noting that they are known not to be significantly impacted by stock (market) illiquidity. For this reason, a market liquidity metric based on bid and ask prices of stocks, as detailed below, is employed as another indicator.

It is documented in Boudt et al. [4] that the TED spread could also be considered a major indicator of market stability. The TED spread is calculated as the difference between the interest rate linked to interbank loans and the yield on short-term US T-bills. Currently, its computation makes use of the three-month London Interbank Offer Rate (LIBOR) and the three-month T-bill yield rates. An increasing TED spread usually portends a stock market meltdown as it is taken as a sign of liquidity withdrawal. The TED spread, as described in Boudt et al. [4], can gauge perceived liquidity risk in the general economy since T-bills are risk-free instruments and the funding liquidity risk of lending to commercial banks is encapsulated by LIBOR. A rising TED spread indicates that lenders view the default counterparty risk to be rising as well. Thus, lenders either require a higher rate of interest or settle for lower returns on safer instruments such as T-bills. Conversely, when the default risk of banks decreases, the TED spread falls; see Krugman [22].

We aim to use hidden Markov models (HMMs) driving a mean-reverting process in the analysis of the joint movements of important economic indicators to forecast liquidity and illiquidity states of the financial market. In this chapter, we utilise observed TED spread data as market signals and filter out the state of the economy and subsequently liquidity levels. Filtering results could be useful in assessing near-future market stability. The proposed idea is very similar to that of Abiad [1] wherein a regime-switching approach is used as an early warning device in identifying and characterising periods of currency crises. It must be recognised, however, that a noticeable TED spread movement cannot be taken as a pure indication of extreme illiquidity/market downturns caused by severe illiquidity. Whilst fluctuations in the TED spread may happen due to some significant underlying factors, these fluctuations are sometimes caused by pure noise alone. In the late 1990s, with the world battling the dot-com bubble and other financial upheavals, more instability and uncertainty in the behaviour of the TED spread was observed.

The second indicator for liquidity levels that we consider is the VIX. This is a trademarked ticker symbol for the Chicago Board Options Exchange (CBOE)'s market volatility index and measures the implied volatility of S&P 500 index options (www.cboe.com/micro/VIX). Using historical data, VIX appears to capture some periods of illiquidity that were not picked up by the TED spread. The third indicator we consider is a metric based on the evolution of the S&P 500. At the end of October 2012, market illiquidity was felt to be brought about by cautious trading as speculators and traders' anxiously anticipated the result of the US presidential election. This illiquidity was captured by an S&P 500-based bid-ask spread metric but not by the TED spread. This fact can be explained by the absence of direct causation effect between the proposed metric and the TED spread. For this reason, for a reasonably adequate study modelling liquidity can be accomplished by investigating the TED spread dynamics along with other indicators such as the VIX and an S&P 500-driven measure.

There have been many attempts to model and explain illiquidity such as those put forward in van der End and Tabbae [27]; Mancini et al. [23]; and Vayanos et al.[28], amongst others. Whilst these proposed modelling approaches include Monte Carlo simulations to demonstrate their implementability, they are nonetheless built on simple assumptions for tractability and do not offer the capacity for dynamic calibration using market data. This leaves a huge gap between model implementation which unifies theoretical approaches and real data. In this work, we attempt to address this gap by explaining how to fit the model with the data. With the aid of filtered estimates, we provide a description of the data dynamics with emphasis on the effect of illiquidity shocks.

In forecasting illiquidity, we use a discrete-time Markov chain assumed to modulate the parameters of a mean-reverting process so that several economic regimes can be embedded into the model. As mentioned in Brunnermeier [5], the economy has a liquidity "self-stabilising effect", which is either a "loss" or a "margin" type; see Brunnermeier and Pedersen [6] for additional discussion. It is, therefore, reasonable to look at the Ornstein-Uhlenbeck (OU) process as a simple model for the TED spread and thus liquidity level in general, as self stabilising and the mean-reverting properties of a process are closely linked. More specifically, a market downturn with a falling spiral effect could elicit a fire sale amongst borrowers, which in turn decreases prices and further worsen funding conditions. Such downturn effects could be associated with episodes of

deflation and periods of poor economic growth. This is based on the assumption that instances of a fall in the general price level (e.g., CPI, GNP deflator) roughly coincide with market pull backs in lending due to tighter borrowing constraints. As such, the evolution of indicators for these observed economic events can be captured by an OU process that exhibits temporary low and high value levels.

Goyenko [20] showed a strong correlation between TED and VIX as major indicators in illiquidity estimation; in the same paper, a measure for the evaluation of stock illiquidity was also presented using market bid and ask prices. Whilst it could be avouched that a high degree of correlation between indicators is suggestive of their dependency on a common underlying factor, the correct and ease of identification of such an underlying factor, if it exists at all, remains elusive. Our work is based on similar assumptions in [20], but instead of finding correlation between the major indicators of illiquidity, we incorporate as much information as possible into our model by simultaneously using three market variables integrated by a set of multidimensional dynamic filters. For stock illiquidity, the S&P 500-based spread is used. The main consideration of this paper is the prediction of illiquidity level based on previous information contained in a joint time series of indicators. The dynamic filtering algorithm's structure enables the finding of expected state probability of the illiquidity level at the next time steps.

Our main contribution is the development of an HMM-driven model tailored for capturing and predicting liquidity risk levels. The model's fit and forecasting power are examined using historical data series. Detailed empirical implementation procedures are provided along with the discussion of various aspects concerning model validation and other post-diagnostic modelling considerations. Our numerical results demonstrate that the proposed model has satisfactory capacity in identifying periods of liquidity crises. This modelling tool shows promise in aiding the prediction of economic crunches occurring over a short-time horizon.

The chapter is organised in the following way. Section 3.2 gives an overview of the modelling set up including the HMM formulation and introduction of the change of measure concept for the filtering technique. A description of the mathematical filtering equations is also presented. We specify in section 3.3 the data used for the numerical estimation and prediction experiments. The process of recursive parameter calcula-

tion together with the discussion of the econometric interpretation of the dynamics of estimates are delineated in section 3.4. Finally, section 3.5 concludes.

3.2 Modelling setup

An Ornstein-Uhlenbeck (OU) process r_t is any process that satisfies the stochastic differential equation (SDE)

$$dr_t = \theta(\mu - r_t)dt + \sigma dW_t, \tag{3.1}$$

where W_t is a standard Brownian motion defined on some probability space $(\Omega, \mathcal{F}, P)$, and θ, μ and σ are positive constants. The parameter μ is the mean level to which the process tends to move to over time, whilst θ is the speed of mean reversion, and σ is the volatility. Assuming that θ, μ and σ are constants, the solution of (4.1) by Itō's lemma is given by

$$r_t = r_0 e^{-\theta t} + (1 - e^{-\theta t})\mu + \sigma e^{-\theta t} \int_0^t e^{\theta s} dW_s, \tag{3.2}$$

where r_0 is the initial value at time $t = 0$.

In the sequel, it is assumed that θ, μ and σ will be time-dependent; hence, we respectively denote them by θ_t, μ_t and σ_t. To capture the switching of economic regimes, we also assume that the values of parameters θ_t, μ_t and σ_t are modulated by a discrete-time Markov chain with a finite-state space. We regard the state of the underlying Markov chain as the regime of an economy, or more specifically a liquidity regime dependent on major factors causing economic turbulence. In particular, the scenario when μ_t and σ_t are in the "worst" regime (i.e., very high μ_t and σ_t) corresponds to very unstable periods of the global financial crisis; in this instance, μ_t reaches a high value with σ_t considerably spiking up creating a completely unstable behaviour for θ_t.

A distinct contribution of this work is the detailed implementation of parameter estimation under a multivariate OU setting which extends the one-dimensional framework of Erlwein and Mamon [16]. We consider d OU processes; each process is denoted by $r_t^{(g)}$ with component $g \in \{1, \ , \ldots, d\}$. All vectors and matrices are written in bold lowercase and uppercase letters, respectively. Following the idea developed by Elliott

et al. [13] let us assume that $(\Omega, \mathcal{F}, P)$ is a probability space under which $\boldsymbol{x}_k$ is a homogeneous Markov chain with a finite-state space in discrete time. Thus, $\boldsymbol{x}_k$ evolves according to the equation

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Pi}\boldsymbol{x}_k + \boldsymbol{v}_{k+1}, \tag{3.3}$$

where $\boldsymbol{\Pi}$ is a transition matrix and $\boldsymbol{v}_{k+1}$ is a martingale increment, i.e., $E[\boldsymbol{v}_{k+1}|\mathcal{F}_k] = 0$, where $\mathcal{C}_k = \mathcal{F}_k \vee \mathcal{R}_k$. Here, $\mathcal{F}_k = \sigma\{\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_k\}$ is the filtration generated by $\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_k$ and $\mathcal{R}_k$ is the filtration generated by the $\{\boldsymbol{r}_k\}$ process.

With the closed-form solution in (3.2), each component of the d-dimensional observation process can be written as

$$\begin{aligned} r_{k+1}^{(g)} &= r_k^{(g)} e^{-\theta^{(g)}(\boldsymbol{x}_k)\Delta t} + (1 - e^{-\theta^{(g)}(\boldsymbol{x}_k)\Delta t})\mu^{(g)}(\boldsymbol{x}_k) \\ &\quad + \sigma^{(g)}(\boldsymbol{x}_k) e^{-\theta^{(g)}(\boldsymbol{x}_k)\Delta t} \int_{t_k}^{t_{k+1}} e^{\theta^{(g)}(\boldsymbol{x}_k)s} dW_s, \end{aligned} \tag{3.4}$$

where $\boldsymbol{\mu}^{(g)} = (\mu_1^{(g)}, \mu_2^{(g)}, ..., \mu_N^{(g)})^\top, \boldsymbol{\sigma}^{(g)} = (\sigma_1^{(g)}, \sigma_2^{(g)}, ..., \sigma_N^{(g)})^\top, \boldsymbol{\theta}^{(g)} = (\theta_1^{(g)}, \theta_2^{(g)}, ..., \theta_N^{(g)})^\top \in \mathbb{R}^N$ and $\Delta t = t_{k+1} - t_k$. For ease of calculation, the state space of $\boldsymbol{x}_k$ is associated with the canonical basis of $\mathbb{R}^N$, which is the set of unit vectors $\boldsymbol{e}_h$ in which $\boldsymbol{e}_h$ is a vector having 1 in its h^{th} entry and 0 elsewhere, $h = 1, 2, \ldots, N$. So in equation (3.4), $\mu^{(g)}(\boldsymbol{x}_k) = \langle \boldsymbol{\mu}_k^{(g)}, \boldsymbol{x}_k \rangle$, $\theta^{(g)}(\boldsymbol{x}_k) = \langle \boldsymbol{\theta}_k^{(g)}, \boldsymbol{x}_k \rangle$ and $\sigma^{(g)}(\boldsymbol{x}_k) = \langle \boldsymbol{\sigma}_k^{(g)}, \boldsymbol{x}_k \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual scalar product and $\top$ denotes the transpose of a vector.

If $\boldsymbol{x}_k$ is constant on a small time interval Δt then using the property of a Gaussian distribution and the Itô isometry, the variance of $r_{k+1}^{(g)}$ in (3.4) is

$$\int_{t_k}^{t_{k+1}} e^{2\theta^{(g)}(\boldsymbol{x}_k)s} ds = \frac{1 - e^{-2\theta^{(g)}(\boldsymbol{x}_k)\Delta t}}{2\theta^{(g)}(\boldsymbol{x}_k)}. \tag{3.5}$$

Equation (3.4) has the representation

$$r_{k+1}^{(g)} = \nu^{(g)}(\boldsymbol{x}_k) r_k^{(g)} + \zeta^{(g)}(\boldsymbol{x}_k) + \xi^{(g)}(\boldsymbol{x}_k)\omega_{k+1}^{(g)}, \quad 1 \le g \le d, \tag{3.6}$$

where $\omega_k^{(1)}, \omega_k^{(2)}, ..., \omega_k^{(d)}$ are independent standard Gaussian random variables and

$$\nu^{(g)}(\boldsymbol{x}_k) = e^{-\theta^{(g)}(\boldsymbol{x}_k)\Delta t}, \tag{3.7}$$

$$\zeta^{(g)}(\boldsymbol{x}_k) = (1 - e^{-\theta^{(g)}(\boldsymbol{x}_k)\Delta t})\mu^{(g)}(\boldsymbol{x}_k), \tag{3.8}$$

$$\xi^{(g)}(\boldsymbol{x}_k) = \sigma^{(g)}(\boldsymbol{x}_k)\sqrt{\frac{1 - e^{-2\theta^{(g)}(\boldsymbol{x}_k)\Delta t}}{2\theta^{(g)}(\boldsymbol{x}_k)}}. \tag{3.9}$$

The succeeding calculations are inspired by the approach described in Elliott et al. [13], where filters are derived under some equivalent probability measure $\bar{P}$. Under this ideal measure, the observations are independent and identically distributed random variables making the calculations of conditional expectations easy. The filters, which are conditional expectations, are then related back to the real-world by the use of Bayes' theorem for conditional expectation. The ideal measure $\bar{P}$ is equivalent to the real-world measure P via the Radon-Nikodym derivative constructed as

$$\Lambda_K = \frac{dP}{d\bar{P}}\bigg|_{\mathcal{C}_K} = \prod_{g=1}^{d}\prod_{k=1}^{K}\lambda_k^{(g)},\ K \geq 1,\ \Lambda_0 \equiv 1, \tag{3.10}$$

where

$$2\ln(\lambda_k^{(g)}) = -\frac{r_k^{(g)}\left(r_{k-1}^{(g)}\nu^{(g)}(\boldsymbol{x}_{k-1}) + \zeta^{(g)}(\boldsymbol{x}_{k-1})\right) - \left(r_{k-1}^{(g)}\nu^{(g)}(\boldsymbol{x}_{k-1}) + \zeta^{(g)}(\boldsymbol{x}_{k-1})\right)^2}{\xi^{(g)}(\boldsymbol{x}_{k-1})^2} \tag{3.11}$$

Write the conditional probability of $\boldsymbol{x}_k$ given $\mathcal{R}_k$ under P as

$$\widehat{p}_k^i := P(\boldsymbol{x}_k = \boldsymbol{e}_h|\mathcal{R}_k) = E[\langle \boldsymbol{x}_k, \boldsymbol{e}_h\rangle|\mathcal{R}_k],$$

where $\widehat{\mathbf{p}}_k = (\widehat{p}_k^1, \widehat{p}_k^2, \ldots, \widehat{p}_k^N)^\top \in I\!R^N$. Now,

$$\widehat{\mathbf{p}}_k = E[\boldsymbol{x}_k|\mathcal{R}_k] = \frac{\bar{E}[\Lambda_k\boldsymbol{x}_k|\mathcal{R}_k]}{\bar{E}[\Lambda_k|\mathcal{R}_k]}$$

by Bayes' theorem for conditional expectation. Let $\boldsymbol{c}_k = \bar{E}[\Lambda_k\boldsymbol{x}_k|\mathcal{R}_k]$ and note that $\sum_{i=1}^{N}\langle \boldsymbol{x}_k, \boldsymbol{e}_h\rangle = 1$. Thus,

$$\sum_{i=1}^{N}\langle \boldsymbol{c}_k, \boldsymbol{e}_i\rangle = \sum_{i=1}^{N}\langle \bar{E}[\Lambda_k\boldsymbol{x}_k|\mathcal{R}_k], \boldsymbol{e}_i\rangle = \bar{E}\left[\Lambda_k\sum_{i=1}^{N}\langle \boldsymbol{x}_k, \boldsymbol{e}_i\rangle\bigg|\mathcal{R}_k\right] = \bar{E}[\Lambda_k|\mathcal{R}_k]. \tag{3.12}$$

Consequently, equation (5.6) implies that

$$\widehat{\mathbf{p}}_k = \frac{\boldsymbol{c}_k}{\sum_{i=1}^{N}\langle \boldsymbol{c}_k, \boldsymbol{e}_i\rangle}.$$

Similar to Erlwein et al. [18] or Erlwein and Mamon [16], we define the following quantities:

$$\mathcal{J}^{js}_{k+1}\boldsymbol{x} = \sum_{n=1}^{k+1}\langle \boldsymbol{x}_{n-1}, \boldsymbol{e}_j\rangle\langle \boldsymbol{x}_n, \boldsymbol{e}_s\rangle \tag{3.13}$$

$$\mathcal{O}^{j}_{k+1}\boldsymbol{x} = \sum_{n=1}^{k+1}\langle \boldsymbol{x}_{n}, \boldsymbol{e}_j\rangle \tag{3.14}$$

$$\mathcal{T}^{j}_{k+1}(f)\boldsymbol{x} = \sum_{n=1}^{k+1}\langle \boldsymbol{x}_{n-1}, \boldsymbol{e}_j\rangle f(r_n)\ ,\ 1 \le j \le N. \tag{3.15}$$

Equations (5.7) and (5.8) are the respective number of jumps from $\boldsymbol{e}_s$ to $\boldsymbol{e}_j$ and the amount of time that $\boldsymbol{x}$ occupies the state $\boldsymbol{e}_j$ up to $k+1$. The quantity $\mathcal{T}^{j}_{k+1}(f)$ is an auxiliary process that depends on the function f of the observation process; in our case, f takes the form $f(r) = r$, $f(r) = r^2$ or $f(r) = r_{k+1}r_k$.

Other than generalising the framework in Erlwein and Mamon [16], our contribution includes expressing recursive filtering equations compactly though matrix notation. This allows efficient computation and decreases parameter estimation time using vector-optimised mathematical packages (e.g., MATLAB by The Mathworks). Define the diagonal matrix $\boldsymbol{D}(r_k)$ with elements $d_{i,j}$ by

$$(d_{ij}(r_k)) = \begin{cases} \prod_{g=1}^{d} \exp\left(-\frac{r_k^{(g)}\left(r_{k-1}^{(g)}\nu_i^{(g)}+\zeta_i^{(g)}\right)-\left(r_{k-1}^{(g)}\nu_i^{(g)}+\zeta_i^{(g)}\right)^2}{2\xi_i^{(g)}}\right) & \text{for } i=j \\ 0 \quad \text{otherwise.} \end{cases} \tag{3.16}$$

For any process G_k, we denote the conditional expectation, under $\bar{P}$, of $\Lambda_k G_k$ by $\gamma(G)_k := \bar{E}[\Lambda_k G_k|\mathcal{R}_k]$. We provide recursive filters for $\boldsymbol{c}_k$, $\gamma(\mathcal{J}^{j,i}\boldsymbol{x})_k$, $\gamma(\mathcal{O}^{i}\boldsymbol{x})_k$ and $\gamma(\mathcal{T}^{i}(f)^{(g)}\boldsymbol{x})_k$.

Theorem 1: *Let $\boldsymbol{D}$ be the matrix defined in (4.22). Then*

$$\boldsymbol{c}_k = \boldsymbol{\Pi}\boldsymbol{D}\boldsymbol{c}_{k-1} \tag{3.17}$$

$$\gamma(J^{j,i}\boldsymbol{x})_k = \boldsymbol{\Pi}\boldsymbol{D}(r_k)\gamma(J^{j,i}\boldsymbol{x})_{k-1} + \langle \boldsymbol{c}_{k-1}, \boldsymbol{e}_i\rangle\langle \boldsymbol{D}(r_k)\boldsymbol{e}_i, \boldsymbol{e}_i\rangle\pi_{ji}\boldsymbol{e}_j \tag{3.18}$$

$$\gamma(O^{i}\boldsymbol{x})_k = \boldsymbol{\Pi}\boldsymbol{D}(r_k)\gamma(O^{i}\boldsymbol{x})_{k-1} + \langle \boldsymbol{c}_{k-1}, \boldsymbol{e}_i\rangle\langle \boldsymbol{D}(r_k)\boldsymbol{e}_i, \boldsymbol{e}_i\rangle\boldsymbol{\Pi}\boldsymbol{e}_i \tag{3.19}$$

$$\gamma(T^{i}(f)^{(g)}\boldsymbol{x})_k = \boldsymbol{\Pi}\boldsymbol{D}(r_k)\gamma(T^{i}(f)^{(g)}\boldsymbol{x})_{k-1} + \langle \boldsymbol{c}_{k-1}, \boldsymbol{e}_i\rangle\langle \boldsymbol{D}(r_k)\boldsymbol{e}_i, \boldsymbol{e}_i\rangle f(r_k^{(g)})\boldsymbol{\Pi}\boldsymbol{e}_i. \tag{3.20}$$

Proof *The proof follows similar derivations of the filtering equations as those provided in Elliott et al.[13], Erlwein et al. [18] or Erlwein and Mamon [16].*

■

To obtain the model parameter estimates, we use the Expectation-Maximisation (EM) algorithm [11]. The EM estimation for the multi-regime setting is very similar to that in the one-dimensional case illustrated in Erlwein and Mamon [16], so the proof of the next theorem is omitted.

As indicated in the above discussion, the model parameters $\boldsymbol{\nu}^{(g)}, \boldsymbol{\zeta}^{(g)}$ and $\boldsymbol{\xi}^{(g)}$ have estimates that depend on the filters of quantities given in Theorem 1. These dynamic parameter estimates are given as follows.

Theorem 2: *If multivariate data set with row components* $r_1^{(g)}, r_2^{(g)}, \cdots, r_K^{(g)}$ $1 \leq g \leq d$ *is drawn from the model described in equation (4.3) then the EM parameter estimates are*

$$\widehat{\pi}_{ji} = \frac{\gamma\left(J^{j,i}\right)_k}{\gamma\left(O^i\right)_k} \tag{3.21}$$

$$\widehat{\nu}_i^{(g)} = \frac{\gamma\left(T^i\left(r_{k+1}^{(g)}, r_k^{(g)}\right)\right)_k - \zeta_i^{(g)}\gamma\left(T^i\left(r^{(g)}\right)\right)_k}{\gamma\left(T^i\left((r^{(g)})^2\right)\right)_k} \tag{3.22}$$

$$\widehat{\zeta}_i^{(g)} = \frac{\gamma\left(T^i\left(r^{(g)}\right)\right)_{k+1} - \widehat{\nu}_i^{(g)}\gamma\left(T^i\left(r^{(g)}\right)\right)_k}{\gamma\left(O^i\right)_k} \tag{3.23}$$

$$\begin{aligned}\widehat{\xi}_i^{(g)} \quad = \quad & \frac{\gamma\left(T^i\left((r^{(g)})^2\right)\right)_{k+1} + (\widehat{\nu}_i^{(g)})^2\gamma\left(T^i\left((r^{(g)})^2\right)\right)_k + (\widehat{\zeta}_i^{(g)})^2\gamma\left(O^i\right)_k}{\gamma\left(T^i\left((r^{(g)})^2\right)\right)_k} \\ & -2\frac{\widehat{\nu}_i^{(g)}\gamma\left(T^i\left(r_{k+1}^{(g)}, r_k^{(g)}\right)\right)_k + \widehat{\zeta}_i^{(g)}\gamma\left(\mathfrak{T}^i\left(r^{(g)}\right)\right)_{k+1} + \widehat{\nu}_i^{(g)}\widehat{\zeta}_i^{(g)}\gamma\left(T^i\left(r^{(g)}\right)\right)_k}{\gamma\left(O^i\right)_k}\end{aligned} \tag{3.24}$$

Proof The derivations of (5.17) - (3.24), which generalise the filters for the univariate OU case, are straightforward based on Erlwein and Mamon [16].

■

Remarks

1. *To implement the recursive equations in Theorem 1 in providing the dynamic updating of the estimates* (3.21)–(3.24) *under Theorem 2, note that* $\gamma(H^i)_k = \gamma(H^i\langle \mathbf{1}, \boldsymbol{x}_k\rangle) = \langle \mathbf{1}, \gamma(H^i \boldsymbol{x}_k)\rangle$, *for some function* H *which may denote* J, O *or* T.

2. *Equation (3.22) in Theorem 2 contains the parameter* $\zeta^{(g)}$, *which must be known prior to achieving a workable recursion. In practice, the sequence of equations (3.22) and (3.23) can be implemented in reverse order. That is,* $\widehat{\zeta}^{(g)}$ *can be estimated using the previous knowledge of* $\nu^{(g)}$. *This latter implementation was adopted in the empirical part of this project, resulting in significant stability in parameter estimates.*

3.3 Description of data for implementation

To model the levels of liquidity, we use three monthly time series data covering the period of 30 April 1998–30 April 2013; data points are recorded at the last trading day of each month. These data sets are: (i) TED spread obtained from Bloomberg, (ii) S&P 500 VIX compiled by the CBOE, and (iii) calculated average spread of S&P 500 based on the data collected by Bloomberg. The indicator in (iii), MktIll (a short form for "Market Illiquidity"), was adopted from Goyenko [20] and defined by

$$\text{MktIll} = 2\frac{Bid - Ask}{Bid + Ask} \tag{3.25}$$

where Bid and Ask are the respective bid and ask prices.

The choice of the end-of-month time series data sets in our analysis is mainly due to convenience as they are readily available from all data sources. To ensure that we are not missing possible anomalous patterns in the data, we also look at the time series values recorded at other days of the month. Figure 3.1 shows the dynamics of the TED spread for the data collected on the last day of the month (TED-30) and on the 11th of each month (TED-11); if the 11th is not a trading day we utilise the value of the previous trading day. Except for a few time points that correspond to the recession period of the late 90s and 2007-09 crisis, the behaviour of the two time series is almost identical. Although not shown here, the graph of VIX and Mktill data series display a very similar pattern. The main purpose of this research is not to accurately predict the dynamics of individual variables, but to capture the joint effects of these variables to forecast illiquidity. Whilst we use monthly discretisation, our method works for any

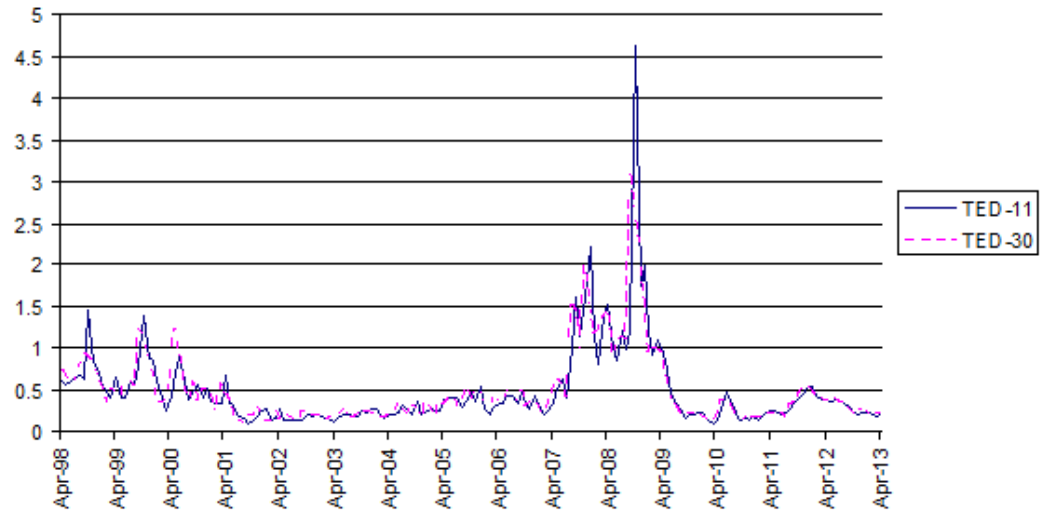


Figure 3.1: Plot of TED recorded on 11th or last trading day of the month

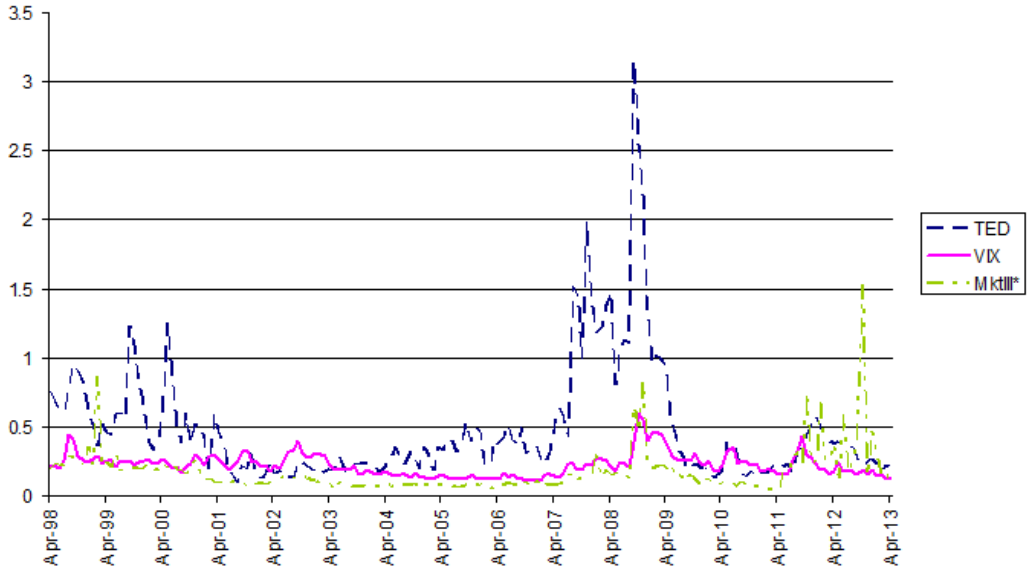


Figure 3.2: Plot of TED, VIX and MktIll $\times$ 100

discretisation frequency. The important consideration is to select a discretisation grid (monthly in our case) just fine enough to capture all the major economic breaks and in-

stabilities, and without creating distortions or introducing extra noise in the data. Data sets with weekly and quarterly frequency were also considered. The general behaviour of the data remains the same, and therefore monthly observations are appropriate for our intended application given the correct number of calculations involved in the window processing of data points.

The data set for our filtering applications is formed by constructing a matrix with a dimension of 181×3 over the period 30 April 1998 – 30 April 2013 with the TED, VIX and MktIll in the first, second and third columns, respectively. Figure 3.2 displays a visualisation of the movements of the TED spread, VIX and MktIll $\times 100$ variables. Note that we use MktIll $\times 100$ to scale the magnitude of MktIll and make it comparable to that of the TED and VIX. The instability of the TED spread in the late 1990s - early 2000s has been limited to the information technology bubble, political dispute surrounding the 2000 US presidential election, political and financial crises in post-Soviet Russia, and a recession in Japan; see Bianchi et al. [3] and Sipley [26], for example. The dot-com price bubble persisted through 1997-2000 climaxing in March 2000. It is worth noting that these three indicators pin down the occurrence of the financial market instability directly affecting liquidity. However, each indicator captures this instability at different moments and with different durations. The superiority of one measure over the others is therefore not clear. This is usually the case whenever the duration of the market crash is short and rapid economic recovery is expected by many.

On the other hand, the subprime mortgage crisis in 2008-2009 was captured by all three measures at once. The noticeably unusual spike in the TED spread during June 2007 seems to clearly herald the coming of extreme financial meltdown that happened in August-September 2008. From the plot of the trivariate series, we also observe the cyclical behaviour of the economy shifting from stable to unstable states. This provides support for using a multivariate version of the OU process in modelling the generating process of the underlying data.

3.4 Numerical application

3.4.1 Calculation of estimates and other implementation assumptions

Several approaches may be employed to find initial parameters for filtering algorithms. These include the methods in Erlwein and Mamon [16]; Erlwein et al. [17]; and Date

and Ponamareva [9], amongst others. Good starting parameter values are necessary to stabilise the filtering algorithm procedure. However, estimating initial parameters is not straightforward considering the nature of the data and other factors. Whilst none of the initial-value estimation algorithms must be disregarded, the choice of which to adopt mainly depends on (i) achieving stable performance and (ii) relative ease of implementation. In this project, we combine the above-mentioned approaches to generate reasonable initial estimates.

To choose the number of states in a regime-switching model, statistical inference-based methods such as the Akaike criterion information (AIC) [2], Bayes' information criterion (BIC) described in Schwarz [24] and Hardy [21], or the CHull metric mentioned in Ceulemans and Kiers [7] may be utilised. These criteria are independent of the nature of the data; they are general tools that can be applied to any data set with an ultimate goal of selecting the model that optimally balances goodness of fit and model complexity. To simplify the discussion, make the mathematics tractable and provide insightful interpretations, one may posit that the economy can only have two states - a "crisis" regime associated with abnormally high indicator values and a "regular" regime. A transitional state may be created and persists over some time due to the weighted combination of volatilities under the above two regimes.

Brokers who have long-term positions in different securities hedge their portfolios differently depending whether to expect a future crash or rally. In the trading world the value of the financial contract can only go up, go down or stay the same. In general, it is assumed that every stock always has a positive growth, i.e., it earns more money than what one can get from a risk-neutral investment. Therefore, even when the potential growth on a stock, mutual fund and other risky investment portfolios is *minimal* but the level of liquidity is high, the economy is still deemed to be in the good (or "high") state. Consequently, when the percentage change in the value of the index or any other major indicators of the financial state of the country (GDP, for example) is relatively close to the risk-neutral rate, we regard the economy to be in the "high" state. Furthermore, we rely on the results of Boudt [4] and Dionne [12] advancing two-state models in investigating liquidity. Any three-state model will be shown later as a special case of the two-regime framework, where the third state is between the "high" and "low" states.

Finding parameter estimates via the likelihood maximisation procedure is a tedious

endeavour but such a procedure provides the best results for dynamic modelling, if it can be accomplished. We shall use the first 40 points of the multidimensional data set to calculate the starting parameters for our filters. For simplicity, when obtaining initial filter values, it is assumed at the outset that the set of true parameters

$$\Xi = \{\pi_{ij}, \nu^{(g)}(\boldsymbol{x}_i), \zeta^{(g)}(\boldsymbol{x}_i), \xi^{(g)}(\boldsymbol{x}_i)\}$$

is homogeneous, i.e., the values of the set Ξ do not change when subsets of the data are chosen in a sense that true parameters of the model stay the same for data subsets of any type or size. Whilst this may be a strong assumption, the filters will eventually adapt and parameters will change accordingly as the number of algorithm passes is increased. The likelihood function, conditional on knowing which state the process $\boldsymbol{x}_i$ is in, is given by

$$L = \prod_{g=1}^{d} \prod_{i=1}^{K} \left\{ \frac{1}{\sqrt{2\pi((\xi^{(g)}(\boldsymbol{x}_i))^2}} \exp\left(- \frac{(r_{i+1}^{(g)} - \nu^{(g)}(\boldsymbol{x}_i) r_i^{(g)} - \zeta^{(g)}(\boldsymbol{x}_i))^2}{2(\xi^{(g)}(\boldsymbol{x}_i))^2} \right) \right\} \tag{3.26}$$

or

$$L = \prod_{g=1}^{d} \prod_{i=1}^{K} \phi_{g,i}, \tag{3.27}$$

where

$$\phi_{g,i} = \phi \left(\frac{r_{i+1}^{(g)} - \nu^{(g)}(\boldsymbol{x}_i) r_i^{(g)} - \zeta^{(g)}(\boldsymbol{x}_i)}{\xi^{(g)}(\boldsymbol{x}_i)} \right), \tag{3.28}$$

and ϕ stands for the density of the standard Gaussian distribution. See analogous concepts in Erlwein and Mamon [16] or Hardy [21]

As the sequence of the states $\boldsymbol{x}_i$ is hidden, a recursive algorithm similar to the one proposed in Hardy [21] is used. The idea of the algorithm is to calculate the most probable set Ξ by building the likelihood function using recursions and to apply standard computer routines for maximisation of the function over a desired set of parameters. Although Hardy's method was designed for the geometric Brownian motion model, we extend it in a straightforward manner to handle our multidimensional data set assumed to follow the OU process. Additionally, the definition of the log-likelihood function is extended by using the sum of the log-likelihood functions for the TED, VIX and MktIll data sets. Adopting the notation of this article, the density function of the process at time t in Hardy's algorithm, given the whole set of parameters including the state of the Markov chain at time t, is changed to $\phi_{g,i}(r_{i+1}, r_i, \Xi)$.

	ν_1	ν_2	ζ_1	ζ_2	ξ_1	ξ_2	π_{12}	π_{21}
TED	0.5178	1.4795	0.1478	-0.6130	0.6119	1.3191	0.6010	0.0076
VIX	0.5709	1.5484	0.4852	-0.1123	0.0212	0.4137	0.6010	0.0076
MktIll	0.5428	1.3090	0.0079	0.0016	0.0006	0.0001	0.6010	0.0076

Table 3.1: Initial parameter estimates for the filtering algorithms under the two-state setting

The results of the initial parameter estimation under the two-state model are provided in Table 3.1. The values of parameters ν_1 and ν_2, encapsulating the speed of mean reversion, can be considered to be almost the same for all three variables. This empirical result that these parameters have uniformly close estimated values is no coincident and gives additional strong support to the hypothesis about the dependency of TED, VIX and MktIll on the same underlying factor.

The HMM filtering algorithms can only give a local maximum, and at times could be extremely unstable to implement. Such limitation can be rectified by choosing initial estimates that fit the data very well and working in double precision arithmetic. It is also possible to employ a symbolic package such as Wolfram Researches' Mathematica, but in that case the speed of the computation drops dramatically. The static log-likelihood maximisation approach appears to yield initial parameters that afford appreciable stability for our OU-based filters.

The starting values for the one-regime model are obtained by simply maximising the likelihood function (4.35), taking into account that $\boldsymbol{x} = 1$, i.e., the system always operates under one state. The results of this optimisation are exhibited in Table 3.2. As expected, the initial parameter values for the single state model lie between the corresponding estimates for the two-regime model. The only parameter which does not follow this observation is ξ; but even then such ξ values corresponding to the three indicators produce a stable convergence for the filtering procedure outlined in the next subsection.

	ν	ζ	ξ
TED	1.2201	-0.1475	2.0761
VIX	0.5911	0.1295	0.0223
MktIll	0.8580	0.0023	0.0001

Table 3.2: Initial parameter estimates for the filtering algorithms under the one-state setting

3.4.2 Filtering procedure

In the estimation of the parameters of the underlying OU process, we employ the method described in section 6.2 using the starting parameters in Tables 3.1 and 3.2. The data set described in section 4.4 contains columns consisting of $g = 1$ (TED), 2 (VIX), and 3 (MktIll). There are 141 time points considered in our filtering application. The first 40 time points for the three vectors of data are used for the initialisation discussed in subsection 3.4.1. The predictive power of the model is tested on the last 60 monthly observations from the middle of the financial crisis (30 May 2008) up to the end of the time series data (30 April 2013). All results are analysed and evaluated using a combination of both intuitive and rigorous statistical approaches for decision making.

The dynamics of the estimates θ, μ and σ are computed by first producing the estimates of ν, ζ and ξ. Then using equations (4.14), (4.15) and (4.16), we back out the values of the desired model parameters. The OU filters described in section 6.2 were implemented with a moving window spanning vectors of data which extends the idea of the procedure in Erlwein and Mamon [16]. More specifically, vectors of data are processed through the recursive equations of the filter to obtain the best estimates (in the sense of conditional expectation) after several time points. Once the parameters are estimated from a batch of vector of data points, they become starting values for the next recursion, and so on. The size of the processing window is determined by likelihood-maximum or other statistical criterion. Owing to the complex nature of the data and the filtering equations, we employ the smallest window possible (3 points per window in our case) that gives stability to the algorithms. Whilst this choice results in a relatively high volatility, the outputs contain ample information about parameter fluctuations.

Figures 3.3–3.6 provide output for the parameter estimates of the OU process corresponding to TED spread. An implication that can be drawn from the behaviour of

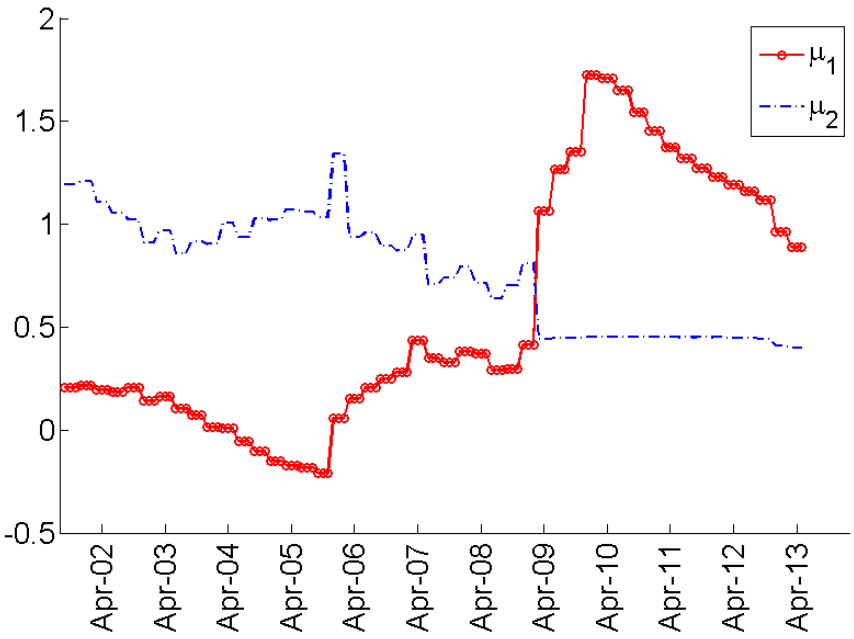


Figure 3.3: Evolution of the mean-level estimates for the TED spread data

transition probabilities in Figure 3.6 is that, over our study window, major illiquidity events do not happen very often, but when they do, they do not last very long and are not severe until the financial market collapse in 2007-2008; see Figures 3.3-3.6 or 3.10. After the crisis in 2008, the structure of the economy changed completely. This fact is supported by Figure 3.3, wherein the mean-reverting levels are switched. This phenomenon occasionally arise in similar filtering applications (e.g., Xi and Mamon [29] or Xi and Mamon [30]). In our case, this anomaly can be explained by the presence of higher volatility levels during times of greater uncertainty. This is substantiated by Figure 3.5 as the level of σ reaches the highest level in 2008.

The other odd behaviour shown by the filtering results is the negativity of θ in Figure 3.4. Even though the formulation of the OU process does not allow the parameter θ to be negative, the multi-regime construction of OU process, proposed by Elliott and

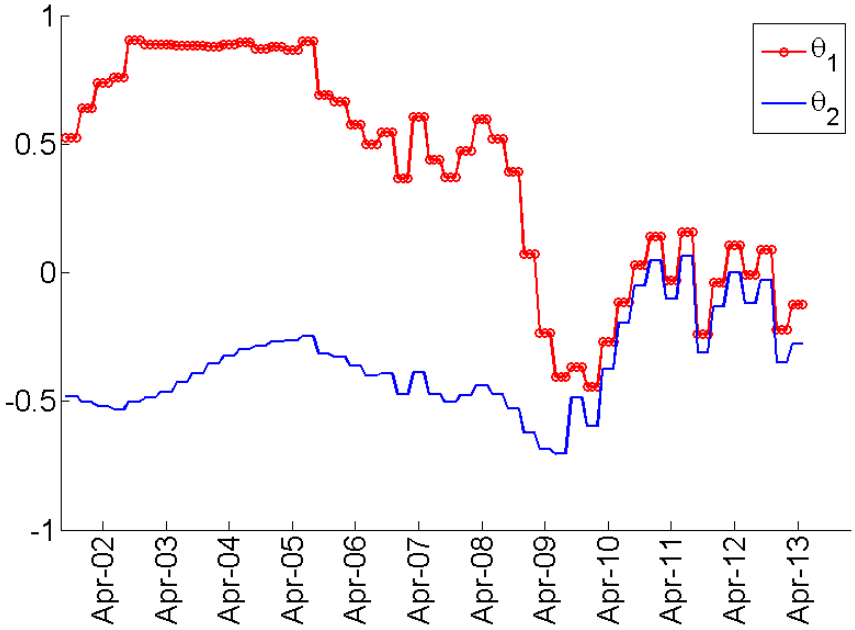


Figure 3.4: Evolution of the speed of mean reversion for the TED spread data

Wilson [15], does not restrict θ to be always positive. From an empirical perspective, getting $\theta \leq 0$ is justified by the fact that the OU process becomes unstable as can be seen during the 2007–2008 period when there was a sudden unfolding of several related financial and economic events leading to the crisis, and exacerbated by too much uncertainty and unpredictability of the economy. Getting a negative value for the speed of mean-reversion has the respective interpretations that the process is actually repelled from the mean level. However, during stable periods, the speed of reversion remains positive and this parameter is interpreted in the usual sense.

The results of the dynamic parameter estimation based on filtering under the one-regime framework are illustrated in Figures 3.7 –3.9. The dynamics of the parameters look similar to those of the two-regime model. This fact can be interpreted as an excellent fit of the 2-state HMM-modulated OU model to the data set. We use the AIC and BIC tailored to several previous works on filtering, in particular, Date et al. [10] and Xi and Mamon [30] to show that the proposed two-regime model provides a better

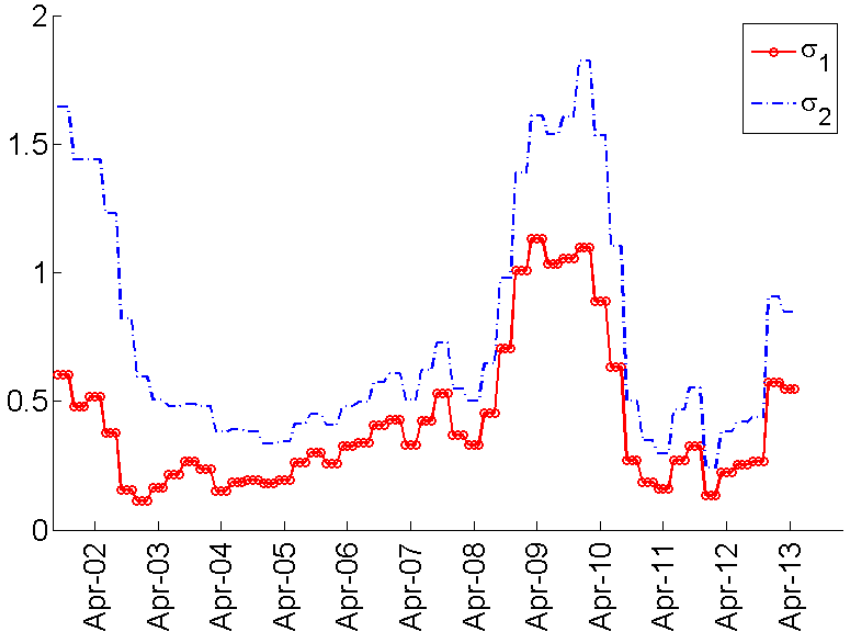


Figure 3.5: Evolution of the volatility levels for the TED spread data

explanation of the data compared to the one-regime setting. The AIC and BIC metrics are computed as

$$AIC = \ln L - p \tag{3.29}$$

and

$$BIC = \ln L - \frac{1}{2} p \ln \kappa, \tag{3.30}$$

where L is the log-likelihood function for the entire multivariate data set, κ is the number of observations and p is the number of parameters in a model. With the calculated value for the log-likelihood function of the last 141 vector of monthly observations, both the AIC and BIC signifies that the two-regime model significant outperforms the one-state model; see Table 5.3.

The general trend of the behaviour of the data during the crisis in 2007-2008 is captured by the model. Nonetheless, due to extreme volatility movements, getting a perfect fit during this period is a challenge. Given the initial parameter estimates, we report that it takes 5-6 algorithm steps for the OU filters to adjust and maintain some stability.

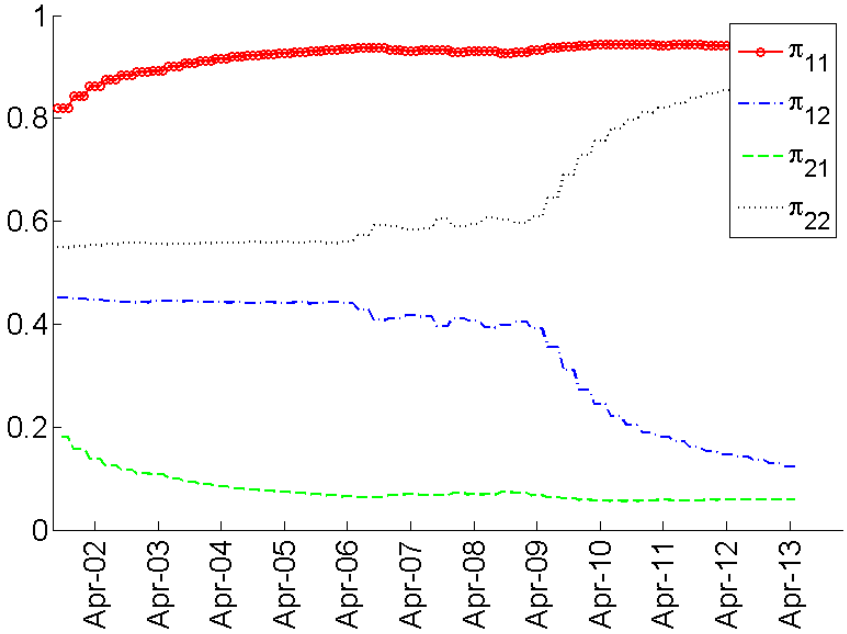


Figure 3.6: Evolution of the filtered transition probabilities obtained from the multivariate data

3.4.3 Filtering and forecasting illiquidity

There are many approaches in modeling illiquidity based on the TED spread or other major economic factors. However, these approaches either look at a certain threshold of the TED spread as benchmark for illiquidity (see, Boudt et al. [4] and Krugman [22]) or correlate the TED spread with another major economic variable (see Goyenko [20]).

In this work, we introduce a new approach which is naturally suited for dynamic filtering algorithms. It relies on the dynamics of $\widehat{\mathbf{p}}_k = E[\boldsymbol{x}_k|\mathcal{C}_k]$. As previously specified,

Regimes	Log-likelihood	BIC	AIC	Number of parameters
I	491.0379	446.4991	482.0379	9
II	569.6856	470.7104	549.6856	20

Table 3.3: Comparison of selection criteria for single- and 2-state regime models

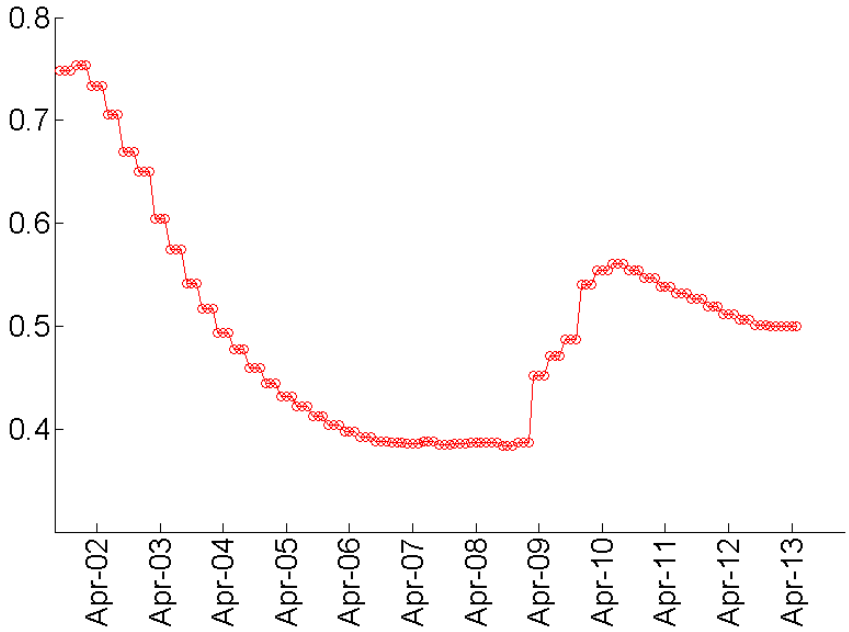


Figure 3.7: Evolution of the mean-reverting level under the one-state setting using the TED spread data

the two-regime model is instructive in that each regime corresponds to illiquid and liquid states of the market. We put forward that if $\widehat{\mathbf{p}}_k(1) = \langle \widehat{\mathbf{p}}, \boldsymbol{e}_1 \rangle \gg 0.5$, i.e., the probability of being in regime 1 is very high, the market is extremely liquid and it is therefore easy to buy and sell every contract. But, if $\widehat{\mathbf{p}}_k(1) \ll 0.5$, which is equivalent to $\widehat{\mathbf{p}}_k(2) = \langle \widehat{\mathbf{p}}, \boldsymbol{e}_2 \rangle \gg 0.5$, then the market is very illiquid, which typically corresponds to recession or period of economic crisis brought about by some major financial events.

We implement the above approach with the objective of determining regimes of illiquidity through the HMM filtering of data. The output of this implementation is displayed in Figure 3.10; the upper panel shows the joint plot for the evolving estimates of $\mu_i, \theta_i, \pi_{ij}$ and $\hat{p}_k(1)$ whilst the lower panel shows the TED evolution plotted on the same time scale. We omit the graphs of the two other indicators (VIX and MktIll) to avoid overcrowding the lower panel of this illustration. They are displayed in Figure 3.2 and similar demonstrations can be produced using various combination of results in Figures 3.2–3.6. Two salient points are as follows. First, every point estimate of

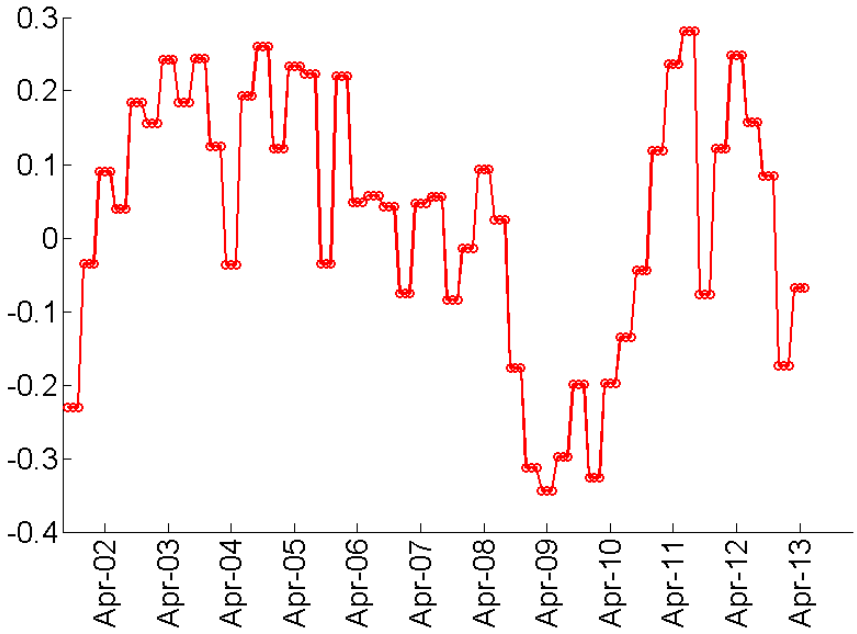


Figure 3.8: Evolution of the speed of mean reversion under the one-state setting using the TED spread data

each model parameter was obtained via a data processing procedure based on a deemed optimal filtering window of size three. Clearly, this inherent dependence on the data processing window size is a source of variability for modelling results. Second, from the side-by-side comparison in Figure 3.10, we delineate four major trigger events of the 2007–2008 liquidity crisis captured by the model with the period of their occurrences charted against the time axis. These events are the hedge fund crash, total loss underestimation by the federal regulators in the US, major collapse of the economy in September 2008 and recovery stage. One important finding of our modelling implementation is the pre-crisis stage that is captured only by the illiquidity state process, but not by the dynamics of any of the model parameters. Thus, it is essential to monitor all parameters, indicators and any other metrics simultaneously.

It is asserted in Brunnermeier [5] that the hedge fund crash was one of the major triggers of the 2007–2008 liquidity collapse. The outcome obtained in our model is consistent with Brunnermeier's argument; in fact, our results provide further support

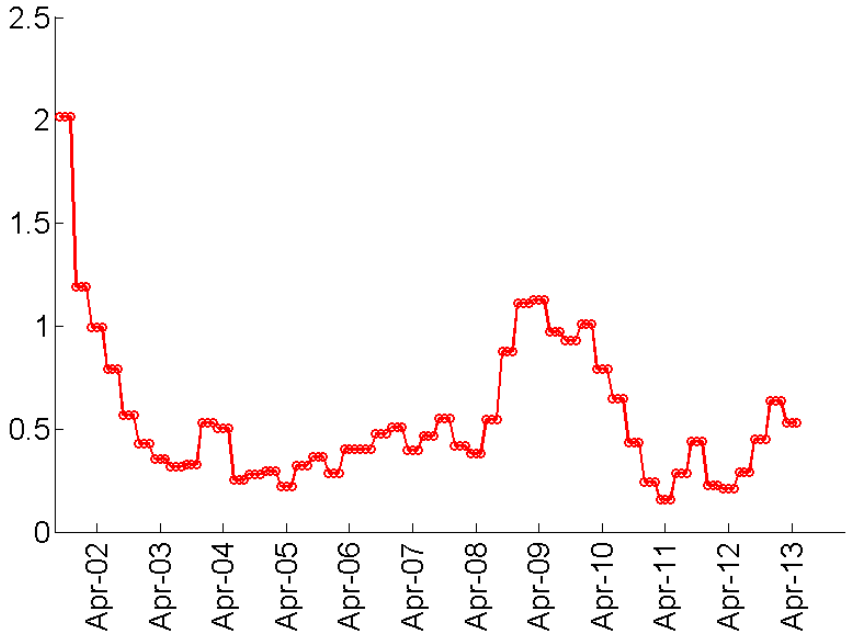


Figure 3.9: Evolution of the volatility under the one-state setting using the TED spread data

to the findings that the hedge fund crash event triggered the chain of events leading to the deepening of the financial crisis up to January–February 2009. The severity of the crisis is seemingly marked by the underestimation of the total loss by federal regulators. After this event, return to the "normal" liquidity level did not happen without the painful experience of several major defaults in the financial industry and new drastic measures taken by the regulators. The proposed model and modelling approach capture these dynamics exceptionally well. In Winter 2009, after many bankruptcies and government bailouts, trading levels started to pick up. Subsequently, financial uncertainty decreased, and trading volume rose from already low level comparable to that of the pre-crisis stage. Around this time, the considerable restructuring of the general economy caused the behaviour of the variables we are examining to change altogether with levels, movements and model parameter estimates completely different from those of the past.

The following technique is used to predict the state of liquidity. First, $\widehat{\mathbf{p}}_k$ is computed

for some k, and second, the expectation of $\widehat{\mathbf{p}}_{k+1}$ given $\mathcal{C}_k$ is calculated as

$$E[\widehat{\mathbf{p}}_{k+1}|\mathcal{C}_k] = \Pi_k \widehat{\mathbf{p}}_k, \tag{3.31}$$

where Π_k is defined by equation (4.2).

Figure 3.11 depicts the dynamics of the prediction values $E[\widehat{\mathbf{p}}_k(1)|\mathcal{C}_{k-1}]$ and shows the estimates of $\widehat{\mathbf{p}}_k(1)$ obtained by applying the filtering algorithms on the last 60 points of the data set. The drastic change in the movement of estimated probabilities between the fourth and fifth time points corresponds to a significant drop in the values of all the three variables (TED, VIX, MktIll) in April-May 2009; such twist was perfectly captured by the dynamics of $\widehat{\mathbf{p}}_k(1)$.

To distinguish the liquid from the illiquid state, we propose a criterion that hinges on $\widehat{\mathbf{p}}_k(i) \gg 0.5$ for $i = 1,\ 2$. If $\widehat{\mathbf{p}}_k(1) > 0.6$, it is assumed that there is enough evidence to conclude that the level of liquidity in the market is high, and traders can take positions with little or no probability of acquiring additional risks due to market or funding liquidity. So, the higher the $\widehat{\mathbf{p}}_k(1)$, the higher the liquidity level. Whenever $\widehat{\mathbf{p}}_k(1) < 0.4$, the financial markets are assumed illiquid, and therefore, additional capital must be infused to deal with the financial distress.

Based on empirical evidence, typifying exactly the liquidity state when $0.4 < \widehat{\mathbf{p}}_k(i) < 0.6$, $i = 1,\ 2$ is not an easy endeavour. This situation is characterised by a very high level of uncertainty regarding market directions over a short time. On the one hand, the "state of uncertainty" signals the occurrence of future hard times. On the other hand, it can also be viewed as a sign that economic stability is forthcoming after an economic downturn. The case in point here is the period of early 2009, when regulators used all possible schemes to stabilise the market sentiments and provided instant artificial liquidity to help markets function the way they were intended to be. We note that our proposed model gives somewhat overoptimistic estimates for $E[\widehat{\mathbf{p}}_k(1)]$ during the last period of the market crash in 2008, and this requires some adjustment. However, the estimates for the $\widehat{\mathbf{p}}_k(1)$ still remain at the 0.4 level, coinciding with what was previously argued concerning "state of uncertainty" and artificial liquidity. Liquidity state prediction for the last 48 data points is accurate in the sense that the predictions jibe very well with the classification of the liquidity state estimates.

The "state of uncertainty" can be viewed either as a third regime in a two-state model, which is interpreted as the lowest/worst bound for the "high" regime and upper/best bound for the "low" regime. This can be explained from an econometric point of view. Recession and upturn times in the economy are generally followed by short periods of market anxiety. During these unstable periods, liquidity can rise and fall quite frequently because speculators do not have stable expectations for the long-term horizons and short-term government interventions can provide only temporally relief. It is rather difficult to capture that "state" as it has in a way the characteristics of either regime. Of course, the stability of a possible separate three-regime model must be investigated as well, and could certainly be an alternative model. Nevertheless, preliminarily results in our case reveal that recursive algorithms do not provide even an approximate convergence for finding the starting parameters for the three-state model. Thus, we rule out the dynamic three-regime setting as inappropriate for this data set.

3.5 Concluding remarks

In this work, we developed an HMM-based modelling approach in assessing levels of market and funding liquidity risks. The structure of the proposed model incorporates major econometric assumptions concerning factors of economic recovery. We provided a detailed methodology on how to extract information from major economic indicators, and linking these to the short-term prediction of market illiquidity or liquidity. The methodology employed made use of newly developed multivariate HMM recursive filtering algorithms expressed in matrix representations. Effects of mean-reversion and liquidity state dependency were also explored.

The model's implementability and forecasting performance were investigated using market data. Results were analysed against statistical metrics and interpreted by examining underlying historical financial events. We found that the one-regime model significantly underperforms compared to a two-regime model. Undoubtedly, a simple OU process cannot capture all the very complex features of the liquidity risk in the financial market. A technique for liquidity-state estimation naturally consistent with dynamic HMM filtering algorithms was put forward and its validity was evaluated using past data.

An improvement that could be done with our modelling approach is the further examination of the two-state model. Its predictability of liquidity becomes uncertain if

the conditional probability of the Markov chain falls in the range $[0.4, 0.6]$. Despite our empirical and economic reasoning to support our assumption and conclusions under this scenario, additional analysis of this particular aspect is a promising research direction. Our preliminary results suggest that the three-state model cannot be fitted given the data we examined. There is however a possibility that the three-regime model may work by adding some other economic variables portraying clear multi-regime behaviour.

Our suggested modelling construction and empirical work used monthly data. Building on our results, further analysis of data with different frequency could be carried out to open avenues for modelling methods and insights about liquidity risk over a long or very short-time periods. These entail establishing new drivers, factors and determinants of liquidity to be included in the filtering experiments. The HMM-driven OU process may have to be tweaked to accommodate these new inputs leading to new filters.

We put forward and empirically tested a new way of estimating and predicting liquidity levels in the financial market. This approach provides a quantitative methodology that supports economic interpretation of our liquidity proxy variables. The liquid/illiquid regimes pinpointed by our regime-switching modelling approach accurately correspond to those identified by practitioners. This research therefore addressed the missing link between economical and mathematical modelling of liquidity. The methodology it contains could be useful for traders, economists, regulators and policy makers.

The current recommended modelling and estimation set up can be effectively exploited under sophisticated trading-scheme environments. For example, underlying variables involved in trading, valuation or reserve calculation for financial derivative contracts, are known to follow the OU process. Our filtering equations can be employed to provide dynamic parameter estimates both for pricing and risk management. Regulators may also consider this model to study the impact of different constraints on the economy.

3.6 References

[1] Abiad, A., 2007. Early warning systems for currency crises: A regime-switching approach, in Mamon, R. and Elliott, R. (eds) Hidden Markov Models in Finance, Springer, New York, 155–184. 62

[2] Akaike, H., 1974. A new look at the statistical model identification, IEEE Transactions on Automatic Control 19(6), 716–723. 73

[3] Bianchi, R., Drew, M., and Wijeratne, T. 2009. Systemic risk, the TED spread and hedge fund returns, International Journal of Business and Economics 1(1), 59–78. 72

[4] Boudt, K., Paulus, E., Rosenthal, R., 2010. Funding liquidity, market liquidity and TED spread: A two-regime model, working paper, SSRN. http://dx.doi.org/10.2139/ssrn.1668635. 62, 73, 80

[5] Brunnermeier, M., 2009. Deciphering the liquidity and credit crunch 2007-2008, Journal of Economic Perspectives 23(1), 77–100. 61, 63, 82

[6] Brunnermeier, M., Pedersen, L., 2008. Market liquidity and funding liquidity, Review of Financial Studies 22(6), 2201–2238. 63

[7] Ceulemans, E., Kiers, H., 2006. Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method, British Journal of Mathematical and Statistical Psychology 59, 133–150. 73

[8] Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity, Journal of Finance 56(2), 501–530. 61

[9] Date, P., Ponomareva, K., 2011. Linear and nonlinear filtering in mathematical finance: A review, IMA Journal of Management Mathematics 22, 195–211. 73

[10] Date, P., Mamon, R., Tenyakov, A., 2013. Filtering and forecasting commodity futures prices under an HMM framework, Energy Economics 40, 1001–1013. 78

[11] Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM Algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39(1), 1–38. 69

[12] Dionne, G., Chun, O. M., 2013. Default and liquidity regimes in the bond market during the 2002-2012 period, Canadian Journal of Economics 46(4), 1160–1195. 73

[13] Elliott, R., Aggoun, L., Moore, J., 1995. Hidden Markov Models: Estimation and Control, Springer, New York. 66, 67, 69

[14] Elliott, R., Hunter, W., Jamieson, B., 2001. Financial signal processing: A self-calibrating model, International Journal of Theoretical and Applied Finance 4, 567–584.

[15] Elliott, R., Wilson, C., 2007. The term structure of interest rates in a hidden Markov setting, In Mamon, R. and Elliott, R. (eds) Hidden Markov Models in Finance, Springer, New York. 78

[16] Erlwein, C., Mamon, R., 2009. An online estimation scheme for Hull-White model with HMM-driven parameters, Statistical Methods and Applications 18(1), 87–107 65, 68, 69, 72, 74, 76

[17] Erlwein, C., Benth, F., Mamon, R., 2010. HMM filtering and parameter estimation of an electricity spot price model, Energy Economics 32(5), 1034–1043 72

[18] Erlwein, C., Mamon, R., Davison, M., 2011. An examination of HMM-based investment strategies for asset allocation, Applied Stochastic Models in Business and Industry 27, 204–221. 68, 69

[19] G. Gorton., 2012. Misunderstanding Financial Crises: Why We Don't See Them Coming, Oxford University Press, New York. 61

[20] Goyenko, R., 2013. Treasury liquidity and funding liquidity: Evidence from mutual fund returns, available at SSRN: http://dx.doi.org/10.2139/ssrn.2023187. 61, 64, 70, 80

[21] Hardy, M., 2002. A regime-switching model of long-term stock returns, North American Actuarial Journal 6(1), 171–173. 73, 74

[22] Krugman, P., 2008. The Coincidence of a liberal - Mission not accomplished, not yet anyway, New York Times, March 12. 62, 80

[23] Mancini L., Ranaldo A., Wrampelmeyer J., 2012. The foreign exchange market: Not as liquid as you may think, http://www.voxeu.org/article/foreign-exchange-market-not-liquid-you-may-think . 63

[24] Schwarz, G., 1978. Estimating the dimension of a model, Annals of Statistics 6(2), 461–464. 73

[25] Shreve, S., 2004. Stochastic Calculus for Finance II: Continuous Time Models, Springer, New York.

[26] Sipley R., 2009. Market Indicators: The Best-Kept Secret to More Effective Trading and Investing, Bloomberg Press, New York. 72

[27] van der End, J., W., Tabbae M., 2012. When liquidity risk becomes a systemic issue: Empirical evidence of bank behaviour, Journal of Financial Stability 8, 107–120. 63

[28] Vayanos D., Wand J., 2013. Market liquidity–Theory and empirical evidence, in Constantinides, G., Stulz, R., Harris, M., eds., Financial Markets and Asset Pricing, Elsevier's Handbook of the Economics of Finance, 1289–1361. 63

[29] Xi, X., Mamon, R., 2011. Parameter estimation of an asset price model driven by a weak hidden Markov chain, Economic Modelling 28, 36–46. 77

[30] X. Xi, Mamon, R., 2013. Yield curve modelling using a multivariate higher-order HMM, in Zeng, Y. and Wu, S., State-Space Models and Applications in Economics and Finance, Springers Series in Statistics and Econometrics for Finance 1, 185–202. 77, 78

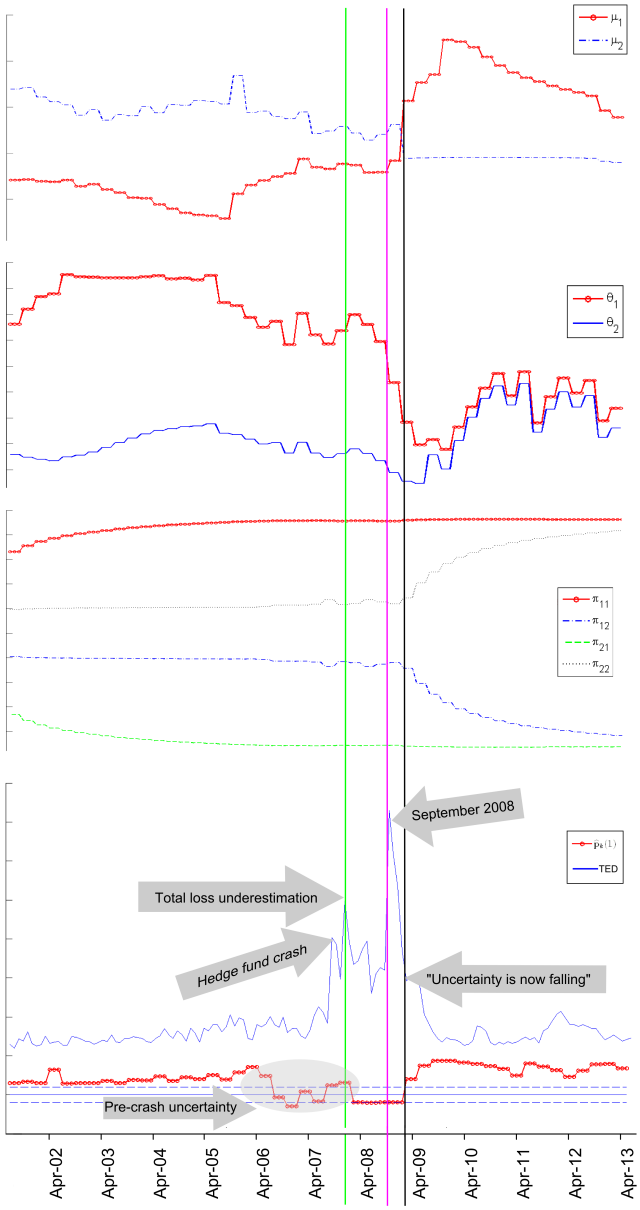


Figure 3.10: Side-by-side comparison between behaviour of model parameter estimates and movement of the TED spread along with the identification of major financial market events through time

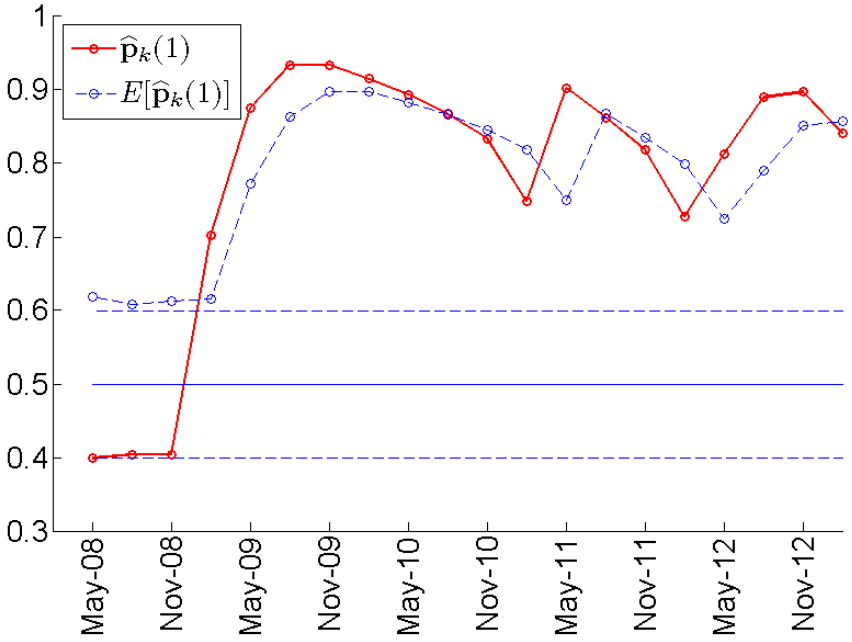


Figure 3.11: Evolution of the estimated liquidity-state probabilities and one-step ahead forecasts of liquidity-state probabilities

4

Pairs trading: An integrated Kalman-HMM approach

4.1 Introduction

Pairs trading is an investment strategy used to exploit financial markets that are out of equilibrium. It consists of a long position in one security and a short position in another security in a predetermined ratio (Elliott et al. [8]). This creates a hedge against the sector and the overall market where the stocks belong. If the market or sector crashes, a trader experiences a gain on the short position and a loss on the long position leaving the profit close to zero in spite of the large move. Traders bet on the direction of the stocks relative to each other. This type of strategy is effectively employed by hedge funds as it is possible to monitor for deviations in prices, automatically changing the positions to use market inefficiencies earning some profit.

The paper by Elliott et al. [8] describes two automated algorithms for setting a pairs trading strategy. The strategy requires parameter estimation support. The first algorithm is based on the smoother approach (Shumway and Stoffer [15]), and the other algorithm is based on dynamic filtering along with the EM-Algorithm (Elliott and Krishnamurthy [7]). It is shown that both algorithms work rather well on simulated data with a computational advantage for the latter. Unfortunately, due to the structure of the Kalman filter, the performance of the dynamic filtering algorithm is limited to white-noise type of models. The other criticism as Steele [16] pointed out is that the model in Elliott et al. [8] is theoretical and does not confirm market's stylised facts. For

example, the normality of returns on an equity spread (say, returns on a portfolio that is long Dell and short HP in equal dollar amounts), can be rejected yet such normality assumption is implicit in the posited model.

In our approach we will extend the idea of automatic pairs trading algorithm proposed by Elliott and Krishnamurthy [7] under a model with non-normal noise. We use a model with a hidden Markov chain modulating parameters to capture the non-normality of returns of the spread portfolio. We note that Levy-type processes could provide excellent statistical fit but they are difficult to interpret from financial perspective. Instead, we modify and integrate methods that are based on dynamic filters of Elliott et al. [6] , Elliott and Krishnamurthy [7], and matrix extension proposed by Mamon and Erlwein [11].

There are many examples demonstrating markedly better fit to financial data when using a hidden Markov model(HMM) compared to a simple autoregressive process. We extend the Gaussian mean reverting model by allowing parameters to be governed by a hidden Markov chain. The advantages of modelling non-normality from mixed-normal models are: (i) modelling framework and methodology are composed of simple methods supporting the normal-noise case and (ii) putting economic interpretation for the sudden changes in the behaviour of the process is easy. From the practitioners point of view, the most desirable characteristic of the HMM modulated model is tractability for applications in the financial industry.

Whilst pairs trading is deemed relative safe for market or sector arbitrageurs, no one is safe from the price swings within the sector. Such could ruin short-term profit opportunity and, in some cases, bankrupt a financially stable hedger, see for example Goldstein [5]. History teaches us that the best way to deal with this type of risk is either to unwind a short position whilst taking minimal losses or better track a constantly changing reverting level of the portfolio; the worst case scenario is described in Khandani and Lo [13]. The former approach is taken by aggressive high frequency algorithmic traders as their positions are usually neutral at the end of each trading day. However most market participants use pairs trading as a hedge against the sector and therefore they are mostly risk averse. Making model parameters flexible to adapt to market changes will decrease the size of the traders positions, and therefore potential risks will decrease as well. We illustrate that at any given time the hedgers position comprises of a short and

a long leg within the spread, with some drawback that the potential profit will decrease.

The structure of this chapter is as follows. Section 4.2 presents the modelling framework for the evolution of the underlying pairs or spread portfolio. In section 4.3, we outline the construction of the filtering algorithms for the parameter estimation as well as the trading strategy. We investigate the effectiveness of our approach in section 4.4 by setting trades based on simulated and historical data with the spread exhibiting a strong non-normal behaviour. Section 4.5 concludes.

4.2 Modelling setup

We set up the framework by first outlining the differences between observation, state and hidden-state processes. We then explain and justify the modification of the trading process.

4.2.1 Observation, state and hidden state processes

An Ornstein-Uhlenbeck (OU) process r_t is a process that satisfies the stochastic differential equation (SDE)

$$dr_t = \theta(\mu - r_t)dt + \sigma dW_t, \tag{4.1}$$

where W_t is a standard Brownian motion defined on some probability space $(\Omega, \Re, P)$, and θ, μ and σ are constants independent of W_t. The parameter μ is the mean level to which the process tends to move to over time, whilst θ is the speed of mean reversion, and σ is the volatility. In the following it is assumed that θ, μ and σ are not constant, it is considered that the parameters are dependent on some underlying Markov process $\{\boldsymbol{x}_k\}$. So, $\theta = \theta(\boldsymbol{x}_k), \mu = \mu(\boldsymbol{x}_k)$ and $\sigma = \sigma(\boldsymbol{x}_k)$, where

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Pi}\boldsymbol{x}_k + \boldsymbol{v}_{k+1}, \tag{4.2}$$

with $\boldsymbol{\Pi}$ and $\boldsymbol{v}_{k+1}$ being the transition matrix and a martingale increment respectively. That is, $E[\boldsymbol{v}_{k+1}|\mathcal{F}_k] = 0$, where $\mathcal{H}_k = \mathcal{F}_k \vee \mathcal{C}_k$. Here, $\mathcal{F}_k = \sigma\{\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_k\}$ is the filtration generated by $\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_k$ and $\mathcal{C}_k$ is the filtration generated by the $\{\boldsymbol{r}_k\}$ process. The states of $\boldsymbol{x}_k$ are mapped with canonical basis of $I\!R^N$, which is the set of unit vectors $\mathbf{e}_h,\ h = 1, 2, \ldots, N$ with $\mathbf{e}_h = (0,\ \ldots,\ 1,\ \ldots,\ 0)^\top$. that is, the hth component of $\mathbf{e}_h$ is 1, and 0 elsewhere.

Using the discretisation described in Erlwein [9] or Tenyakov at al. [17], it may be shown that if the parameters of the OU process modulated by the hidden Markov model (HMM) is constant over the small interval Δt, then equivalent to equation (4.1) is then result

$$r_{k+1} = \nu(\boldsymbol{x}_k)r_k + \zeta(\boldsymbol{x}_k) + \xi(\boldsymbol{x}_k)\omega_{k+1}, \tag{4.3}$$

where

$$\nu(\boldsymbol{x}_k) = e^{-\theta(\boldsymbol{x}_k)\Delta t}, \tag{4.4}$$

$$\zeta(\boldsymbol{x}_k) = (1 - e^{-\theta(\boldsymbol{x}_k)\Delta t})\mu(\boldsymbol{x}_k), \tag{4.5}$$

$$\xi(\boldsymbol{x}_k) = \sigma(\boldsymbol{x}_k)\sqrt{\frac{1 - e^{-2\theta(\boldsymbol{x}_k)\Delta t}}{2\theta(\boldsymbol{x}_k)}}. \tag{4.6}$$

In equations (4.3)-(4.16), $\mu(\boldsymbol{x}_k) = \langle \boldsymbol{\mu}_k, \boldsymbol{x}_k \rangle$, $\theta(\boldsymbol{x}_k) = \langle \boldsymbol{\theta}_k, \boldsymbol{x}_k \rangle$ and $\sigma(\boldsymbol{x}_k) = \langle \boldsymbol{\sigma}_k, \boldsymbol{x}_k \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual scalar product and $\top$ denotes the transpose of a vector. Such representation of the model parameters is the offshoot of choosing the Markov chain's state space as the canonical basis of $\mathbb{R}^N$.

Following Elliott et al. [8], the observation process y_k follows the state process r_k observed in some Gaussian noise with volatility parameter α. So,

$$y_k = r_k + \alpha Z_k, \tag{4.7}$$

where Z_k is a standard Gaussian random variable independent of $\mathcal{C}_k$.

To summarise, in discretised form the equations for the dynamic behaviour of the data are

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Pi}\boldsymbol{x}_k + \boldsymbol{v}_{k+1}, \text{ the hidden state process;}$$

$$r_{k+1} = \nu(\boldsymbol{x}_k)r_k + \zeta(\boldsymbol{x}_k) + \xi(\boldsymbol{x}_k)\omega_{k+1}, \text{ the state process;}$$

$$y_k = r_k + \alpha Z_k, \text{ the observation process.}$$

4.2.2 The trading strategy

We define $\mathcal{Y}_k = \sigma\{y_0, y_1, ..., y_k\}$, the whole information available from the market up to time k. The ultimate goal is to compute the quantity

$$\hat{r}_{k|k-1} = E[r_k|\mathcal{Y}_{k-1}]. \tag{4.8}$$

Write

$$\hat{r}_{k|k-1}(i) := E[r_k|\mathcal{Y}_{k-1}, R_k = i], \tag{4.9}$$

where $R_k = i$ represents the hidden Markov process being in the i^{th} state at time k.

The quantity $\hat{r_k}(i)$ is interpreted as the most possible value of r_k given the observed information up to time k and considering that the process is in state i at time k.

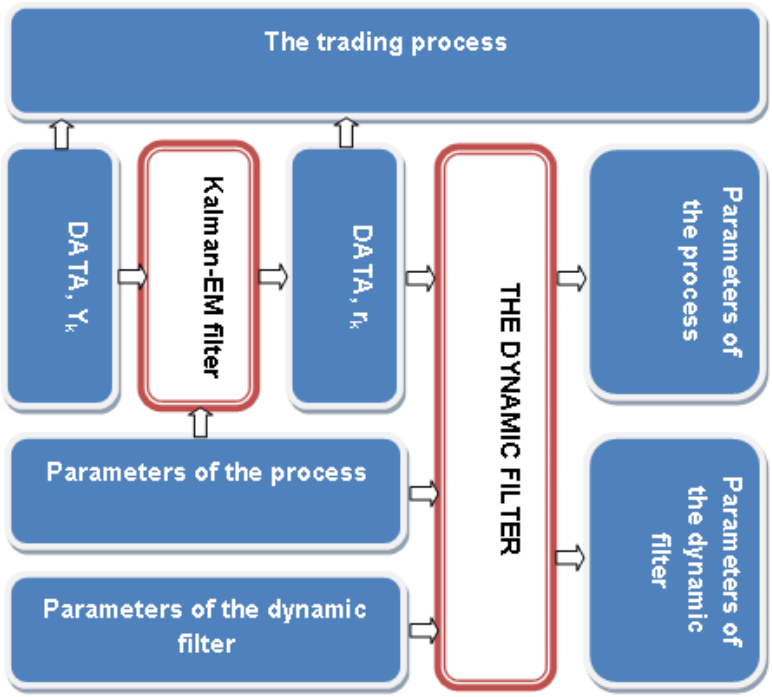


Figure 4.1: Trading strategy

We define the unconditional probability of being in state i as θ_i. The trading process can be described as follows: (i) firstly, we find the values i and $i+1$ such that $\hat{r}_{k|k-1}(i) < y_k < \hat{r}_{k|k-1}(i+1)$; (ii) secondly, the probabilities are aggregated subject to $\Theta_1 = \sum_{j=1}^{i} \theta_j$ and $\Theta_2 = \sum_{j=i+1}^{N} \theta_j$; and (iii) lastly, the positions are changed in a

manner similar to that in Elliott et al. [8] with the key difference that the portfolio consists of two parts proportional to Θ_1 and Θ_2. The pairs trading strategy is formed using two stocks with the predetermined ratio 1:1 as the stocks are taken from the same sector with a long position on one asset and a short in the other. The positions are set in the same way that the spread portfolio is either short or long depending on the expected level of the difference in the stock prices. If the level is lower than the spread then it is expected that the correction will occur so that the trader takes a long position on the spread portfolio, and vice versa.

The schematics of the trading process is displayed in Figure 4.1. If Ntnl represents the total notional amount for the investment in the strategy, $\text{Ntnl}_l = \text{Ntnl}\,\Theta_l$ for $l = 1, 2$ respectively. It could be thought of that the two trades are carried out simultaneously, i.e., $y_k < \hat{r}_{k|k-1}(i+1)$ for Ntnl_2 and $\hat{r}_{k|k-1}(i) < y_k$ for Ntnl_1. If $\hat{r_k}(i) < y_k$ the spread is assumed to be too large, so a long position in the spread portfolio is taken. Similarly, when $y_k < \hat{r}_{k|k-1}(i+1)$ a short trade is entered. This trading strategy leads to a substantial decrease in the overall risk exposure.

4.3 Filtering approach: extended Kalman and dynamic filters

In this section, we provide Kalman filtering results within the multi-regime set up described in equations from subsection 4.2.1. As well, we briefly recall the dynamic filtering results for the OU process.

4.3.1 HMM extended Kalman filter

We give the derivation of the conditional mean and variance of the state process r_k. The results are represented by recursive relations.

$$\begin{aligned}
\hat{r}_{k|k-1} &= \sum_{i=1}^{N} \hat{r}_{k|k-1}(i)\theta_i \qquad (4.10)\\
&= \sum_{i=1}^{N} E[r_k|\mathcal{Y}_{k-1}, R_k = i]\theta_i \\
&= \sum_{i=1}^{N} E[\nu(\boldsymbol{e}_i)r_{k-1} + \zeta(\boldsymbol{e}_i) + \xi(\boldsymbol{e}_i)\omega_{k+1}|\mathcal{Y}_{k-1}, R_k = i]\theta_i \\
&= \sum_{i=1}^{N} E[\nu(\boldsymbol{e}_i)r_{k-1} + \zeta(\boldsymbol{e}_i)|\mathcal{Y}_{k-1}, R_k = i]\theta_i \\
&= \sum_{i=1}^{N} \left(\nu(\boldsymbol{e}_i)E[r_{k-1}|\mathcal{Y}_{k-1}, R_k = i] + \zeta(\boldsymbol{e}_i)\right)\theta_i \\
&= \sum_{i=1}^{N} \theta_i\nu(\boldsymbol{e}_i)E[r_{k-1}|\mathcal{Y}_{k-1}, R_k = i] + \sum_{i=1}^{N} \zeta(\boldsymbol{e}_i)\theta_i \\
&\quad \text{using the independence of increments property} \\
&= E[r_{k-1}|\mathcal{Y}_{k-1}] \sum_{i=1}^{N} \theta_i\nu(\boldsymbol{e}_i) + \sum_{i=1}^{N} \zeta(\boldsymbol{e}_i)\theta_i \\
&= \hat{r}_{k-1}\hat{\nu} + \hat{\zeta}.
\end{aligned}$$

$$\begin{aligned}
\Sigma_{k+1|k} &= E[(r_{k+1} - \hat{r}_{k+1})^2 |\mathcal{Y}_k] \qquad (4.11)\\
&= E[(\nu(\boldsymbol{x}_k)r_k + \zeta(\boldsymbol{x}_k) + \xi(\boldsymbol{x}_k)\omega_{k+1} - \hat{r}_{k+1})^2 |\mathcal{Y}_k] \\
&= E[\Big(\nu(\boldsymbol{x}_k)r_k + \zeta(\boldsymbol{x}_k) + \xi(\boldsymbol{x}_k)\omega_{k+1} - \hat{r}_k\hat{\nu} - \hat{\zeta}\Big)^2 |\mathcal{Y}_k] \\
&= E[E[(\nu(\boldsymbol{x}_k)r_k + \xi(\boldsymbol{x}_k)\omega_{k+1} - \hat{r}_k\hat{\nu})^2 |\mathcal{Y}_k, R_{k+1}]] \\
&= E[E[(\nu(\boldsymbol{x}_k)r_k - \hat{r}_k\hat{\nu})^2 |\mathcal{Y}_k, R_{k+1}]] + E[\xi^2] \\
&= E[\nu^2]\Sigma_{k|k} + E[\xi^2] \\
&= \hat{\nu^2}\Sigma_{k|k} + \hat{\xi^2}.
\end{aligned}$$

The Kalman updating formula is defined as

$$\hat{r}_{k+1} = \hat{r}_{k+1|k} + M_{k+1}\left(y_{k+1} - \hat{r}_{k+1|k}\right), \qquad (4.12)$$

where M_{k+1} is determined below.

Whilst minimising the conditional variance, we get

$$\begin{aligned} E[(r_{k+1} - \hat{r}_{k+1})^2 \,|\mathcal{Y}_k] &= E[\left(r_{k+1} - \hat{r}_{k+1|k} - M_{k+1}[y_{k+1} - \hat{r}_{k+1|k}]\right)^2 |\mathcal{Y}_k] \\ &= E[\left(r_{k+1} - \hat{r}_{k+1|k} - M_{k+1}[r_{k+1} + dZ_{k+1} - \hat{r}_{k+1|k}]\right)^2 |\mathcal{Y}_k] \\ &= (1 - M_{k+1})^2 \Sigma_{k+1|k} + M^2\alpha^2. \end{aligned}$$

So, the optimal value for M is

$$M_{k+1} = \frac{\Sigma_{k+1|k}}{\Sigma_{k+1|k} + \alpha^2}. \tag{4.13}$$

Substituting the result from equation (4.13) into equation (4.13), we obtain

$$\Sigma_{k+1|k} = \Sigma_{k+1|k}\alpha^2 = \Sigma_{k+1|k} - M_{k+1}\Sigma_{k+1|k}. \tag{4.14}$$

Equations (4.10)-(4.14) look very similar to the static case when the parameters of the OU process are constants. If the set of parameters $\mathcal{S} = \{\nu_i, \zeta_i, \theta_i \text{ for } i = 1, \ldots, N\}$ is provided, the values of $\hat{x}_{k+1|k}$ or $\hat{x}_{k|k}$ can be found by recursively applying equations (4.10)-(4.14).

We use the sample standard deviation estimate to calculate $\hat{\alpha}$ given by

$$\hat{\alpha} = \sqrt{\frac{\sum_{j=1}^{n} (y_j - \hat{r}_j)^2}{n-1}}. \tag{4.15}$$

4.3.2 Parameter estimation

The following calculations are based on the methods outlined in Elliott et al. [6]. To simplify the notation, we write $\hat{r}_k := \hat{r}_{k|k}$. All filtering equations are established under the reference probability measure $\bar{P}$ and the results are inverted back to the real-world measure P. This is justified by the Bayes' theorem and the associated Radom-Nikodym derivative with this measure change is given by

$$\Lambda_K = \frac{dP}{d\bar{P}}\Big|_{\hat{\mathcal{H}}_K} = \prod_{k=1}^{K} \lambda_k, \;\; K \geq 1, \;\; \Lambda_0 \equiv 1, \tag{4.16}$$

where

$$2\ln(\lambda_k) = -\frac{\hat{r}_k\left(\hat{r}_{k-1}\nu(\boldsymbol{x}_{k-1}) + \zeta(\boldsymbol{x}_{k-1})\right) - \left(\hat{r}_{k-1}\nu(\boldsymbol{x}_{k-1}) + \zeta(\boldsymbol{x}_{k-1})\right)^2}{\xi(\boldsymbol{x}_{k-1})^2}, \tag{4.17}$$

and $\hat{\mathcal{C}}_K$ stands for the filtration of the process $\hat{r}_k$ up to time K.

Write the conditional probability of $\boldsymbol{x}_k$ given $\hat{\mathcal{C}}_k$ under P as

$$\beta_k^i := P(\boldsymbol{x}_k = \boldsymbol{e}_h|\hat{\mathcal{C}}_k) = E[\langle \boldsymbol{x}_k, \boldsymbol{e}_h\rangle|\hat{\mathcal{C}}_k],$$

where $\widehat{\theta}_k = (\widehat{\theta}_k^1, \widehat{\theta}_k^2, \ldots, \widehat{\theta}_k^N)^\top \in I\!R^N$. Now,

$$\widehat{\boldsymbol{\beta}}_k = E[\boldsymbol{x}_k|\hat{\mathcal{C}}_k] = \frac{\bar{E}[\Lambda_k \boldsymbol{x}_k|\hat{\mathcal{C}}_k]}{\bar{E}[\Lambda_k|\hat{\mathcal{C}}_k]}$$

by the Bayes' theorem for conditional expectation.

Set $\boldsymbol{c}_k = \bar{E}[\Lambda_k \boldsymbol{x}_k|\hat{\mathcal{C}}_k]$ and notice that $\sum_{i=1}^N \langle \boldsymbol{x}_k, \boldsymbol{e}_h\rangle = 1$. So,

$$\sum_{i=1}^N \langle \boldsymbol{c}_k, \boldsymbol{e}_i\rangle = \sum_{i=1}^N \langle \bar{E}[\Lambda_k \boldsymbol{x}_k|\hat{\mathcal{C}}_k], \boldsymbol{e}_i\rangle = \bar{E}\left[\Lambda_k \sum_{i=1}^N \langle \boldsymbol{x}_k, \boldsymbol{e}_i\rangle \middle| \hat{\mathcal{C}}_k\right] = \bar{E}[\Lambda_k|\hat{\mathcal{C}}_k]. \tag{4.18}$$

Therefore, equation (4.18) implies that

$$\widehat{\boldsymbol{\beta}}_k = \frac{\boldsymbol{c}_k}{\sum_{i=1}^N \langle \boldsymbol{c}_k, \boldsymbol{e}_i\rangle}.$$

Following Erlwein et al. [11] and Erlwein and Mamon [9], we define

$$\mathcal{J}_{k+1}^{js}\boldsymbol{x} = \sum_{n=1}^{k+1} \langle \boldsymbol{x}_{n-1}, \boldsymbol{e}_j\rangle\langle \boldsymbol{x}_n, \boldsymbol{e}_s\rangle \tag{4.19}$$

$$\mathcal{O}_{k+1}^{j}\boldsymbol{x} = \sum_{n=1}^{k+1} \langle \boldsymbol{x}_n, \boldsymbol{e}_j\rangle \tag{4.20}$$

$$\mathcal{T}_{k+1}^{j}(f)\boldsymbol{x} = \sum_{n=1}^{k+1} \langle \boldsymbol{x}_{n-1}, \boldsymbol{e}_j\rangle f(\hat{r}_n)\ ,\ 1 \le j \le N. \tag{4.21}$$

Equations (4.19) and (4.20) represent the number of jumps from $\boldsymbol{e}_s$ to $\boldsymbol{e}_j$, and the amount of time that $\boldsymbol{x}$ occupies the state $\boldsymbol{e}_j$ up to $k+1$, respectively. The quantity $\mathcal{T}^j_{k+1}(f)$ is an auxiliary process dependent on the function f; in our calculations $f(r) = r$ or $f(r) = r^2$ or $f(r) = \hat{r}_{k+1}\hat{r}_k$.

To find more compact and efficient representations of the filtering equations we define the matrix $\boldsymbol{D}(r_k)$ with elements $d_{i,j}$ by

$$(d_{ij}(r_k)) = \begin{cases} \exp\left(-\frac{\hat{r}_k(\hat{r}_{k-1}\nu_i+\zeta_i)-(\hat{r}_{k-1}\nu_i+\zeta_i)^2}{2\xi_i}\right) \text{ for } i=j \\ 0 \quad \text{otherwise.} \end{cases} \tag{4.22}$$

For any process G_k, we denote the conditional expectation, under $\bar{P}$, of $\Lambda_k G_k$ by $\gamma(G)_k := \bar{E}[\Lambda_k G_k|\hat{\mathcal{C}}_k]$. We provide recursive filters for $\boldsymbol{c}_k$, $\gamma(\mathcal{J}^{j,i}\boldsymbol{x})_k$, $\gamma(\mathcal{O}^i\boldsymbol{x})_k$ and $\gamma(\mathcal{T}^i(f)\boldsymbol{x})_k$.

Theorem 1: *Let $\boldsymbol{D}$ be the matrix defined in (4.22). Then*

$$\boldsymbol{c}_k = \boldsymbol{\Pi}\boldsymbol{D}\boldsymbol{c}_{k-1} \tag{4.23}$$

$$\gamma(J^{j,i}\boldsymbol{x})_k = \boldsymbol{\Pi}\boldsymbol{D}(r_k)\gamma(J^{j,i}\boldsymbol{x})_{k-1} + \langle \boldsymbol{c}_{k-1}, \boldsymbol{e}_i\rangle\langle \boldsymbol{D}(r_k)\boldsymbol{e}_i, \boldsymbol{e}_i\rangle \pi_{ji}\boldsymbol{e}_j \tag{4.24}$$

$$\gamma(O^i\boldsymbol{x})_k = \boldsymbol{\Pi}\boldsymbol{D}(r_k)\gamma(O^i\boldsymbol{x})_{k-1} + \langle \boldsymbol{c}_{k-1}, \boldsymbol{e}_i\rangle\langle \boldsymbol{D}(r_k)\boldsymbol{e}_i, \boldsymbol{e}_i\rangle \boldsymbol{\Pi}\boldsymbol{e}_i \tag{4.25}$$

$$\gamma(T^i(f)\boldsymbol{x})_k = \boldsymbol{\Pi}\boldsymbol{D}(r_k)\gamma(T^i(f)\boldsymbol{x})_{k-1} + \langle \boldsymbol{c}_{k-1}, \boldsymbol{e}_i\rangle\langle \boldsymbol{D}(r_k)\boldsymbol{e}_i, \boldsymbol{e}_i\rangle f(r_k)\boldsymbol{\Pi}\boldsymbol{e}_i. \tag{4.26}$$

Proof *The proof follows similar derivations of the filtering equations in Elliott [6], Erlwein et al. [11] or Erlwein and Mamon [9].*

■

Theorem 2: *If the data set with components $r_1, r_2, \ldots, r_K$ is drawn from the model described in equation (4.3) then the EM parameter estimates are*

$$\widehat{\pi}_{ji} = \frac{\gamma\left(J^{j,i}\right)_k}{\gamma\left(O^i\right)_k} \tag{4.27}$$

$$\widehat{\nu}_i = \frac{\gamma\left(T^i\left(r_{k+1}, r_k\right)\right)_k - \zeta_i\gamma\left(T^i\left(r\right)\right)_k}{\gamma\left(T^i\left((r)^2\right)\right)_k} \tag{4.28}$$

$$\widehat{\zeta}_i = \frac{\gamma\left(T^i\left(r\right)\right)_{k+1} - \widehat{\nu}_i\gamma\left(T^i\left(r\right)\right)_k}{\gamma\left(O^i\right)_k} \tag{4.29}$$

$$\begin{aligned}\widehat{\xi}_i \;=\; & \frac{\gamma\left(T^i\left(r^2\right)\right)_{k+1} + \widehat{\nu}_i^2\gamma\left(T^i\left(r^2\right)\right)_k + \widehat{\zeta}_i^2\gamma\left(O^i\right)_k}{\gamma\left(T^i\left(r^2\right)\right)_k} \\ & -2\frac{\widehat{\nu}_i\gamma\left(T^i\left(r_{k+1}, r_k\right)\right)_k + \widehat{\zeta}_i\gamma\left(\mathfrak{T}^i\left(r\right)\right)_{k+1} + \widehat{\nu}_i\widehat{\zeta}_i\gamma\left(T^i\left(r\right)\right)_k}{\gamma\left(O^i\right)_k}.\end{aligned} \tag{4.30}$$

Proof The derivations of (4.27) - (4.30) can bee seen in Erlwein and Mamon [9].

■

To get the model parameter estimates, we use the Expectation-Maximisation (EM) algorithm, Dempster et al. [4]. The estimates of ν, ζ and ξ can be obtained from Theorem 2, and their updates can be obtained by applying Theorem 1.

4.4 Numerical application

4.4.1 Preliminary results

In this section, we first validate the modelling approach on a simulated data set. We show how the newly developed filters respond to the underlying data with single-regime behaviour. The results will illustrate the possibility of using the new methodology not only during turbulent periods, when the data display distinct multi-regime spikes in log-returns, but also during relatively calm times.

The OU process has a normal distribution and the increments $\dfrac{r_{k+1} - r_k}{r_k}$, conditional on r_k are also normally distributed. From (4.3), the observation process y_k can be expressed as

$$y_k = \nu r_{k-1} + \eta + \xi\omega_k + \alpha Z_k, \tag{4.31}$$

which is equivalent to

$$y_k \overset{\text{dist}}{=} \nu r_{k-1} + \eta + \sqrt{\xi^2 + \alpha^2}\hat{\omega}_k. \tag{4.32}$$

Clearly, equation (4.32) has an OU functional form too, where $\hat{\omega}_k$ is distributed as

$N(0,1)$. The simulated histogram of $\frac{r_{k+1}-r_k}{r_k}$ is shown in Figure 4.2. Figure 4.3 provides an additional support of the normality assumption.

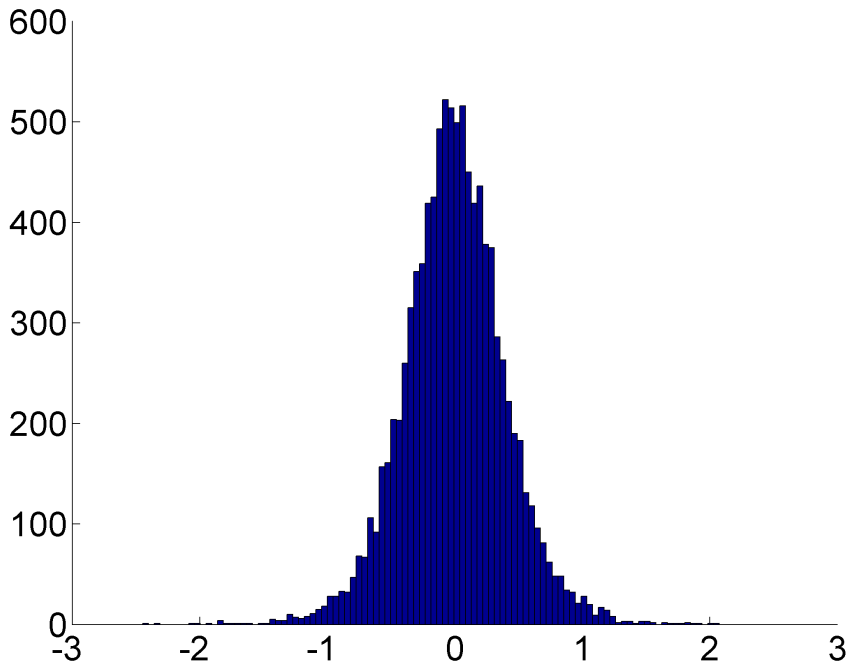


Figure 4.2: Histogram of $\frac{r_{k+1}-r_k}{r_k}$

Unfortunately, real data in practice often do not satisfy the normality hypothesis. Our task then is to come up with a model that is robust enough to incorporate different types of data distributions. Figures 4.4-4.5 depict a histogram and Q-Q plot, respectively, of data simulated from the model with Markov-modulated coefficients. This simulated data set violates the assumption of normality; it apparently produces very heavy tails. This fact attests that the Markov-driven model is able to capture non-Gaussian dynamics of real data. Consequently, the proposed approach has the potential to be employed extensively for long periods characterised by a mixture of calm and turbulent financial times.

For an additional simulation, we consider a simplified version of the discretised model in (4.3). However, the process will be generated without the regime-switching behaviour

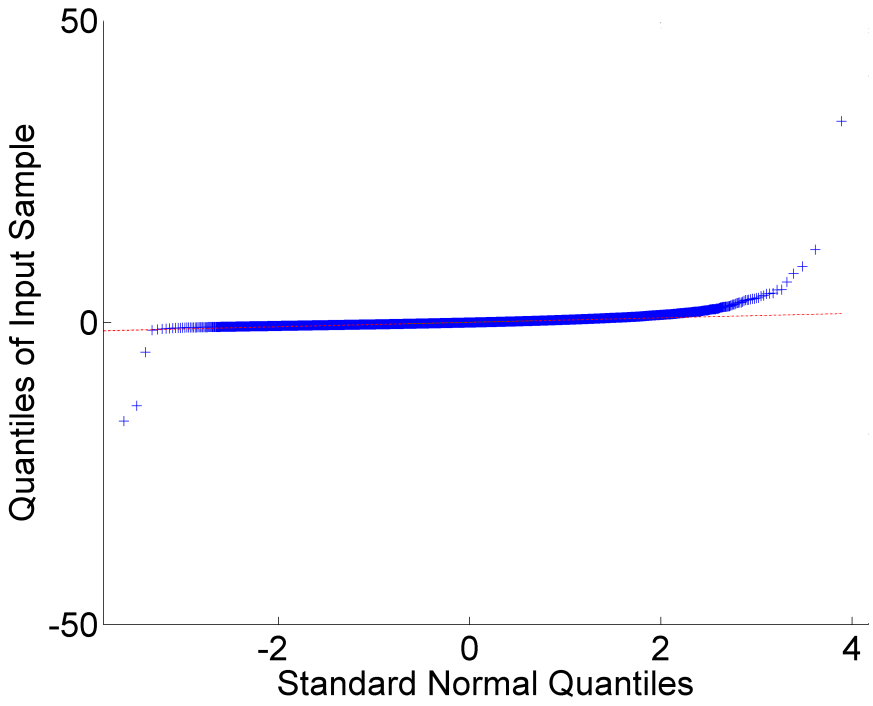


Figure 4.3: Q-Q plot of $\frac{r_{k+1}-r_k}{r_k}$

according to

$$\begin{aligned} r_{k+1} &= \nu r_k + \zeta + \xi\omega_{k+1}, \\ y_k &= r_k + \alpha Z_k, \end{aligned} \tag{4.33}$$

where $\nu = 0.6, \zeta = 0.15, \xi = 0.05$ and $\alpha = 0.4$.

From Figure 4.6, we conclude that the filters of Elliott and Krishnamurthy [7] produce quite noisy, but precise parameter values. The estimates are fairly stable and will not cause a significant loss if one decides to use them for financial market trading.

We will use the above results to benchmark the proposed method and algorithm by assessing the implementation speed, relative error in the estimated EM estimates and to comment on other features of the algorithm. Figures 4.7 - 4.10 present the dynamics of the parameters estimated using the new filters under the two-state model. Although the main purpose of the algorithm is to replicate the dynamics of the coefficients modulated by the HMM, we found that the one-regime (stationary) parameters are estimated very

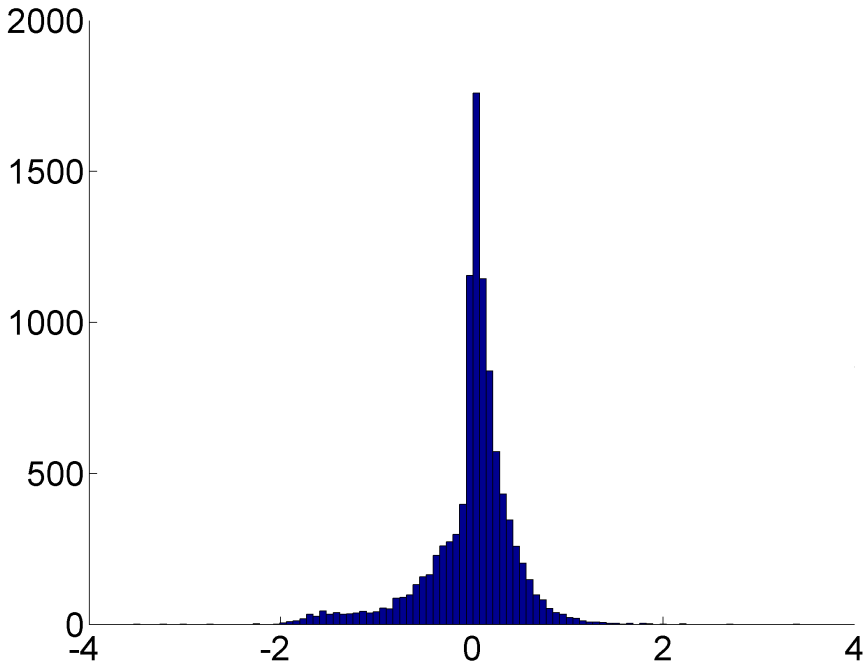


Figure 4.4: Histogram of $\frac{r_{k+1} - r_k}{r_k}$

accurately and it only takes about half time compared to that in the Kalman-EM algorithm of Elliott and Krisnamurthy [7].

Figures 4.7 - 4.9 show that one of the regimes produce "divergence" in all estimates. Yet the unconditional probability of being in the "divergent" regime is almost zero as illustrated in Figure 4.9. It is apparent that the filtered parameters have better precision than those given by the Kalman-EM filters; see Figure 4.6. Moreover, the increase in the computational time can be explained by the numerical implementation of the new method's structure. Comprehensive details about implementation of the multi-regime filter procedure can be found in Tenyakov and Mamon [17], Erlwein et al. [11] or Tenyakov and Mamon [18].

4.4.2 Analysis of the data

By construction, the Kalman-EM filter (cf. Elliott [8]) is not able capture all stylised characteristics of data following non-normal distribution. Thus, this presents difficulty

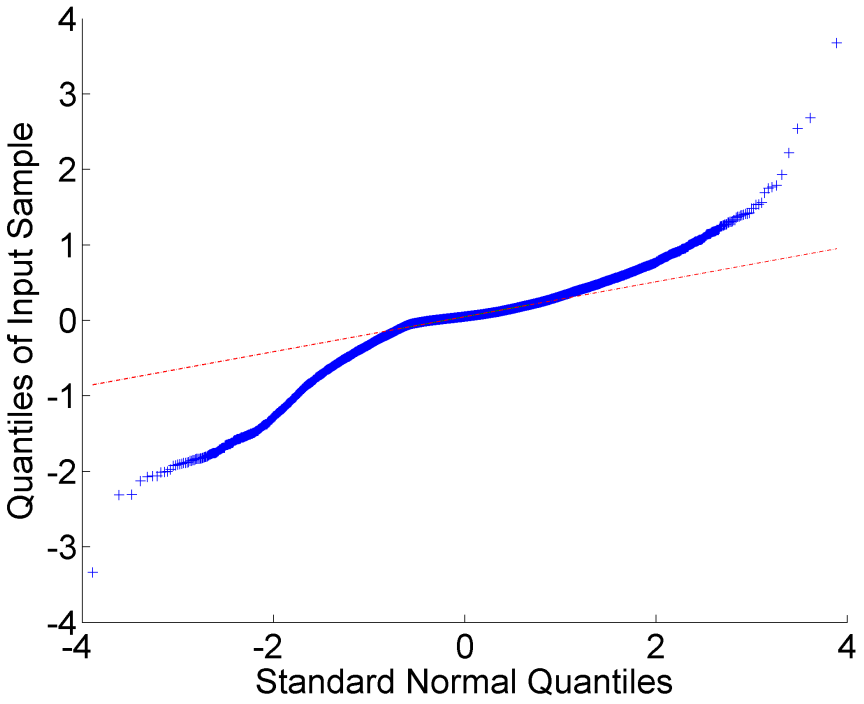


Figure 4.5: Q-Q plot of $\frac{r_{k+1} - r_k}{r_k}$

when implementing trading strategy based on Gaussian models. We pick the KO - PEP pair of stocks (i.e., Coca-Cola Co and PepsiCo Inc) to test our integrated filtering approach. The data of daily NYSE closing prices were chosen randomly from Bloomberg covering 24 October 2012 to 07 January 2014. The data set has 300 data points and contains subsets with price spreads exhibiting distinct non-Gaussian spikes and appear to have multi-regime behaviour; see Figure 4.11.

Daily closing data were used in the study as intra-day prices have additional liquidity and high-frequency-type noise embedded in them. It is also known that intra-day data are observed in uneven intervals making them not straightforward for immediate analysis. For our filtering procedure, we use the modified spread (SPR) between KO and PEP as

$$\text{SPR} = \text{KO} - \text{PEP} - 30. \tag{4.34}$$

The adjustment coefficient of 30 was chosen solely for convenience of representation.

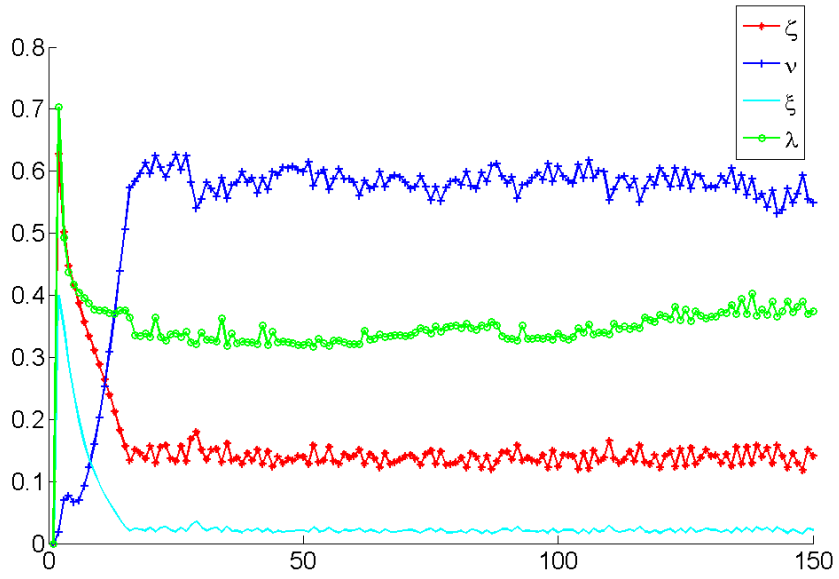


Figure 4.6: Single regime dynamic filtered parameter estimates using simulated data

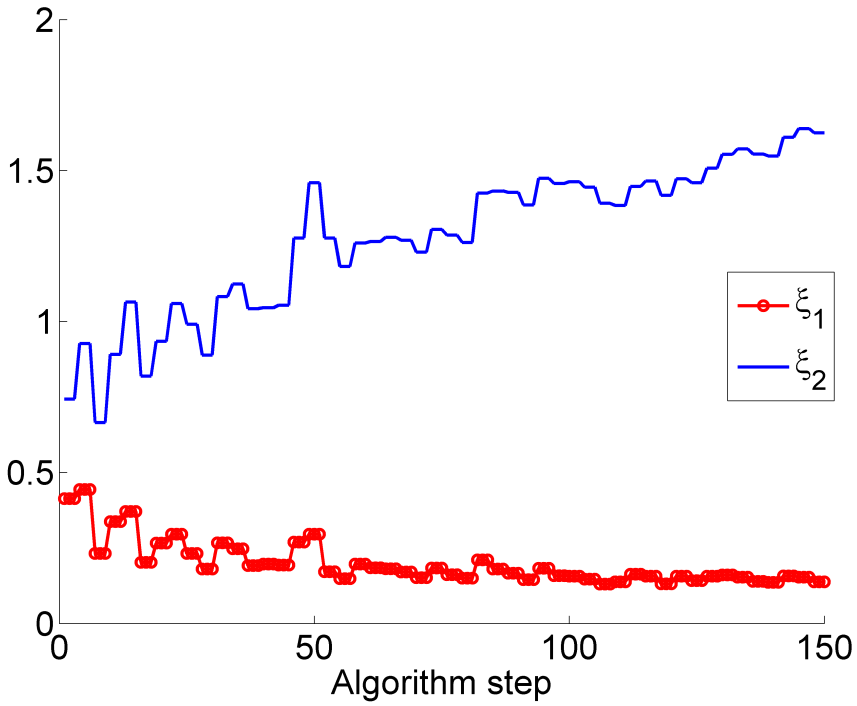


Figure 4.7: Evolution of the estimated ζ_1 and ζ_2 under 2-regime HMM

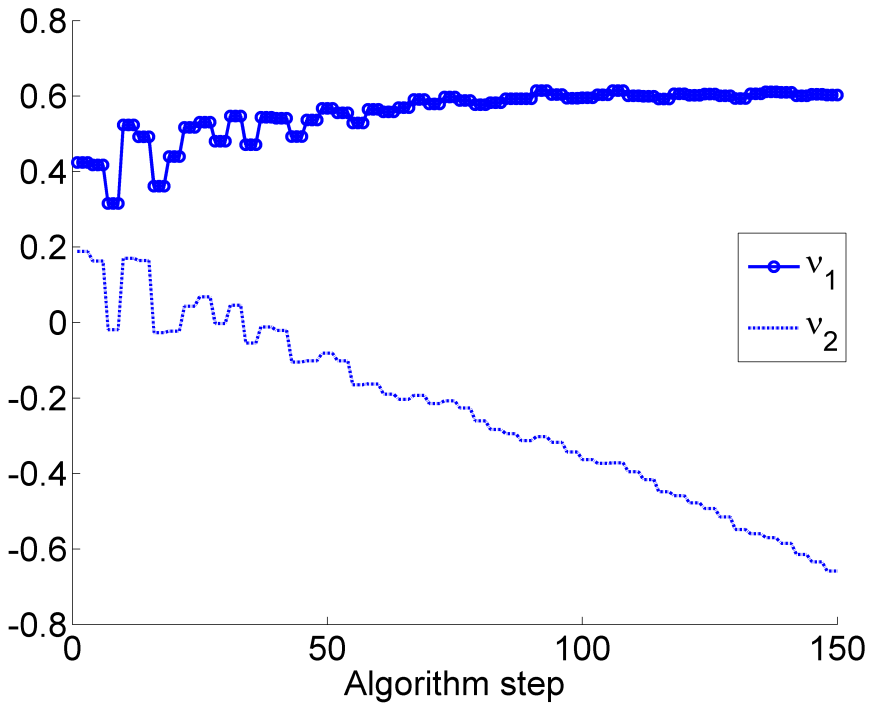


Figure 4.8: Evolution of the estimated ν_1 and ν_2 under a 2-regime HMM

For the real trading procedure, the spread KO − PEP is used. Note, that even with 30 as an adjustment, this does not result to negative values in the differences and the main characteristic of the data set is preserved. We attempted to find the rationale for the major spikes in SPR. No explanation from data sources such as Bloomberg and the Internet can be found. The occurrence of spikes appear data-specific, and looking at the same time series of spread from 2002, we see that the spikes happened again quite frequently.

4.4.3 Initialisation of the algorithm

To implement filtering algorithms, suitable parameters for initialisation are needed. An advantage of our proposed approach is the complete automation of the initialisation stage. The financial modeller may choose from various algorithms, some of which are described in Erlwein et al. [9, 10], or Date and Ponomareva [3], amongst others. For our purpose, a simple log-likelihood maximisation would suffice and we adopted the procedure in Date and Ponomareva [3], which is also described in Hardy [12]. We di-

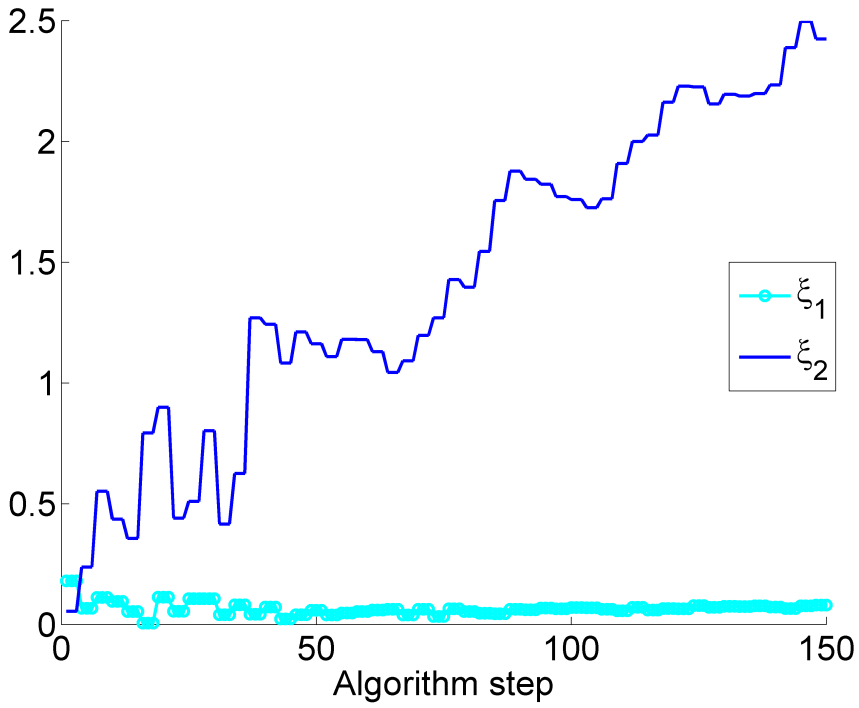


Figure 4.9: Evolution of the estimated ξ_1 and ξ_2 under a 2-regime HMM

vide our data into two parts, one part serves as a training subset, the remaining part is utilised for the trading procedure and validation.

To carry out the initialisation step, we assume that the data set follows the simple OU model, and so $\alpha = 0$ in (4.7), and the other parameters are not modulated by a Markov chain. The likelihood function is then given by

$$L = \prod_{i=1}^{N_{trn}} \left\{ \frac{1}{\sqrt{2\pi\xi^2}} \exp\left(-\frac{(r_{i+1} - \nu r_i - \zeta^2}{2\xi)^2} \right) \right\} \tag{4.35}$$

or

$$L = \prod_{i=1}^{N_{trn}} \phi_i, \tag{4.36}$$

where

$$\phi_i = \phi\left(\frac{r_{i+1} - \nu r_i - \zeta}{\xi} \right). \tag{4.37}$$

The MLEs resulting from equation (4.36) are chosen as starting values for the estimation and trading procedure detailed in the later sections. To choose the optimal number

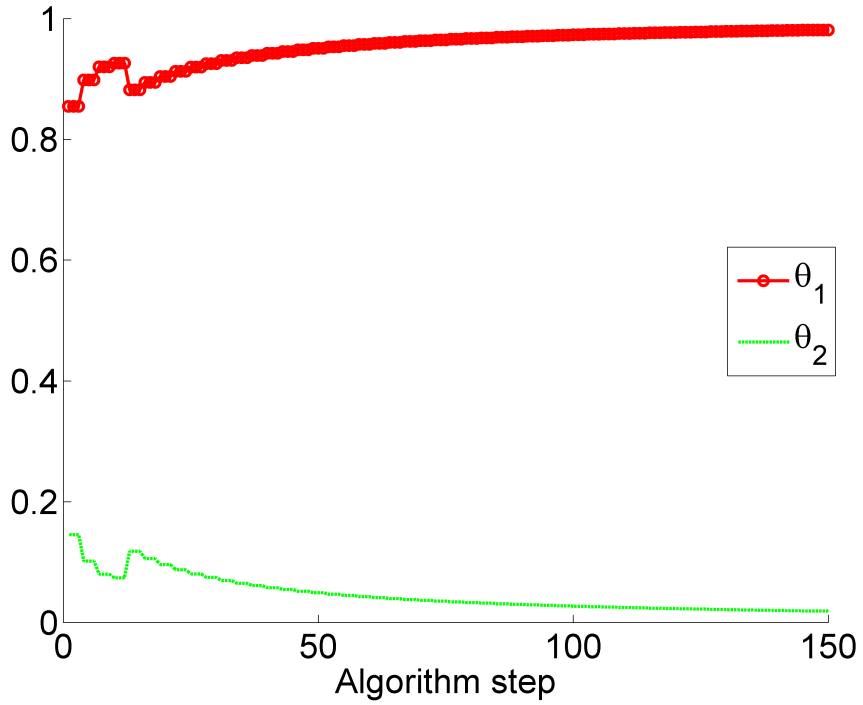


Figure 4.10: Evolution of the estimated θ_1 and θ_2 under a 2-regime HMM

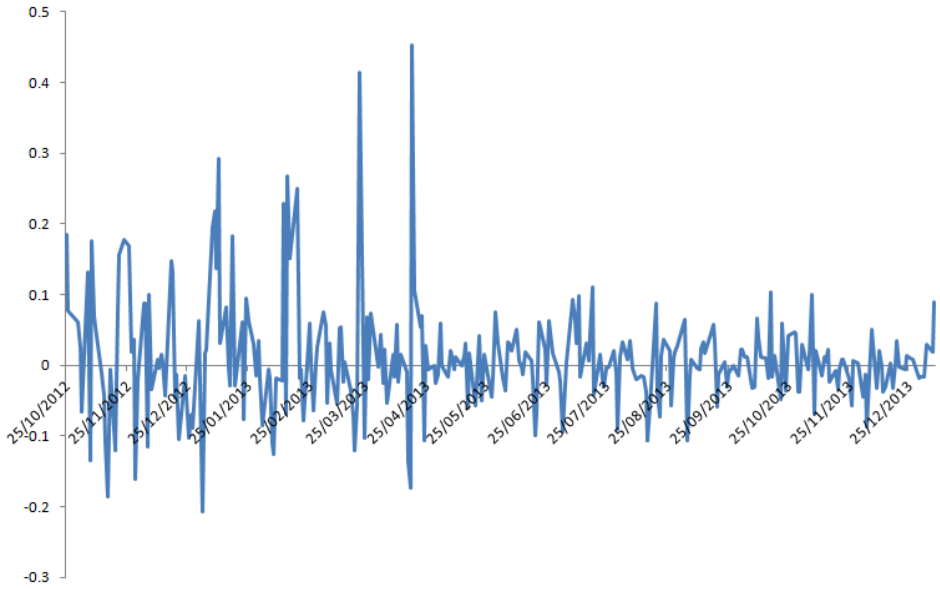


Figure 4.11: Spikes in the log of price spreads

	ν	ζ	ξ	π_1
PEP-KO(35)	0.9682	0.1225	0.2596	0.5

Table 4.1: Initial parameter estimates for the multi-regime filtering algorithm. The same values are used for all regimes (e.g., $\nu = \nu_1 = \nu_2$, etc).

of regimes, several model selection criteria could be employed such as the Akaike information criterion [1], C-hull criterion [2], and the Bayes information criterion [14]. Typically, the optimal number of regimes based on these criteria is two. The "curse of dimensionality" is still bearable at this level. In our new approach, the number of parameters will more than double if the number of regimes is increased to three. When the data set does not show extreme spikes (i.e., about 3 standard deviations or more from the mean), any increase in the number of regimes will expectedly result to accurate prediction of spread levels (this is called overfitting), however such setting is also penalised for its complexity.

We do not consider the development of a filtering algorithm under a one-regime model here as this was already considered in Elliott et al. [8]. Notwithstanding, this one-regime model is still used to benchmark the results of our suggested method. The initial estimates for the parameters of the filters are shown in Table 4.1. Following Elliott [6, 8] concerning the bounds of the parameter values (i.e., $0 < \nu < 1$ and ζ as well as $\xi > 0$), we conclude that the data set could be modelled by the OU process.

We use the same values to initialise the benchmark algorithm. Nonetheless, in Elliott [8] or Elliott and Krishnamurthy [7], there is no apparent indication on how to select the starting values. The convergence of estimates is expected as long as the initialisation provided the starting values are reasonably close to the actual model parameters. In this study, it turns out that Elliott's dynamic filter does not produce any convergence and hence, parameters are adjusted to produce stable convergence on the training data set. The "stabilised" parameters are shown in Table 4.2.

4.4.4 Numerical application

In this section we outline the trading method in conjunction with the filtering algorithms. This procedure is performed in several steps. The optimal parameters are

	ν	ζ	ξ
PEP-K O(35)	0.8731	1.5527	0.1364

Table 4.2: Initial parameter estimates used in applying the dynamic filtering algorithm of Elliott and Krishnamurthy [7]

found using the EM algorithm and then we find the estimate $\hat{r}_t$ and compare it to the observed y_t. The comparison will result will establish the trading position.

Before applying the filters to the data, the user may decide if smoothing the data, without introducing noise to the underlying process, is needed taking into consideration additional work and time for "cleaning" and smoothing. This is pertinent to high-frequency data but not necessary for our own purpose. Another important question is the length of data series to be included in one pass of the algorithm. Some insights on this issue are given in Tenyakov and Mamon [17, 18].

By taking a "practical" look at the data plotted on Figure 4.12, it seems that the process produces 2-3 major and approximately 10-15 minor jumps every 50 points. The data time series does not appear to have a constant mean-reverting level But, certain mean stabilisation seems possible in the horizon. If the processing window of $50/15 = 3.33$ or bigger, the minor changes of the data set's behaviour may be missed by the filter. This judgement on the window processing size is supported by the numerical work in Tenyakov and Mamon [18] or Erlwien et al. [11].

We apply our new algorithm (combining the Kalman and multi-regime filters) using starting parameters given in Table 4.1, which were estimated using the data subset from 16 January 2013 to 07 January 7 2014 with a moving filtering window of three points. The first 50 points of the data were "cleaned up" (i.e., processed using the Kalman estimation algorithm) and then the dynamic multi-regime filters are applied to the Kalman-filtered estimates to produce the parameters of our proposed algorithm.

The next step in our algorithm is different from that in the pairs trading paper of Elliott et al. [8]. We emphasise that the following shift happens not by one point, but by the size of the moving window which was established to be three. Whilst the the long shift might affect two possible position changes between the first and third value.

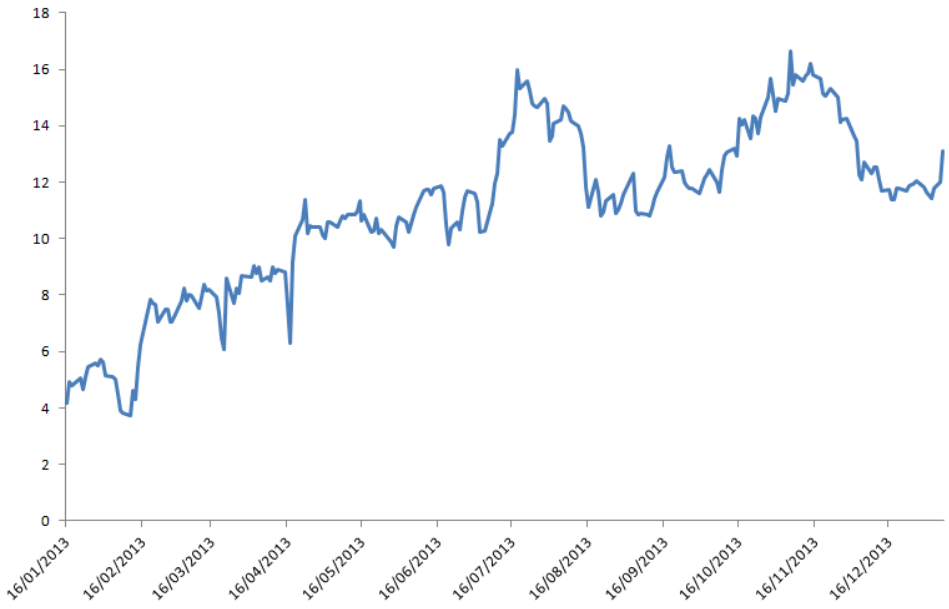


Figure 4.12: The dynamics of the spread SPR in the data subset used for parameter estimation

The step is repeated until all the data points are completely processed.

The plot of the estimated values of the spread in each state under the two-regime model is depicted in Figure 4.13. We also plot positions where the estimates from both regimes, $\hat{r}_t(1)$ and $\hat{r}_t(2)$, are either greater or smaller than the corresponding value for r_t. Points marked with $*$ indicates $\hat{r}_t(i) > r_t$ for all i, and with $+$ indicates the reverse inequality relation. It is visible that the predicted values of $\hat{r}_t$ follow the dynamic of the process with a very precisely. Since we chose one-regime model to begin with, it takes some some time for the algorithm to show relative stability in the spread process signifying further support for the two-regime behaviour of the data. Questions about fitness and model error must be settled using some statistical methods and criteria. Since the model is developed for financial trading purposes, profit will be the main criterion for choosing the best model in the context of this chapter.

We consider 4 trading strategies: (i) aggressive trade in which all earned extra capital being is reinvested in the stock; (ii) normal trade in which only one unit of stock the stock is kept and all earned capital is invested in the money market; (iii) safe trade is a

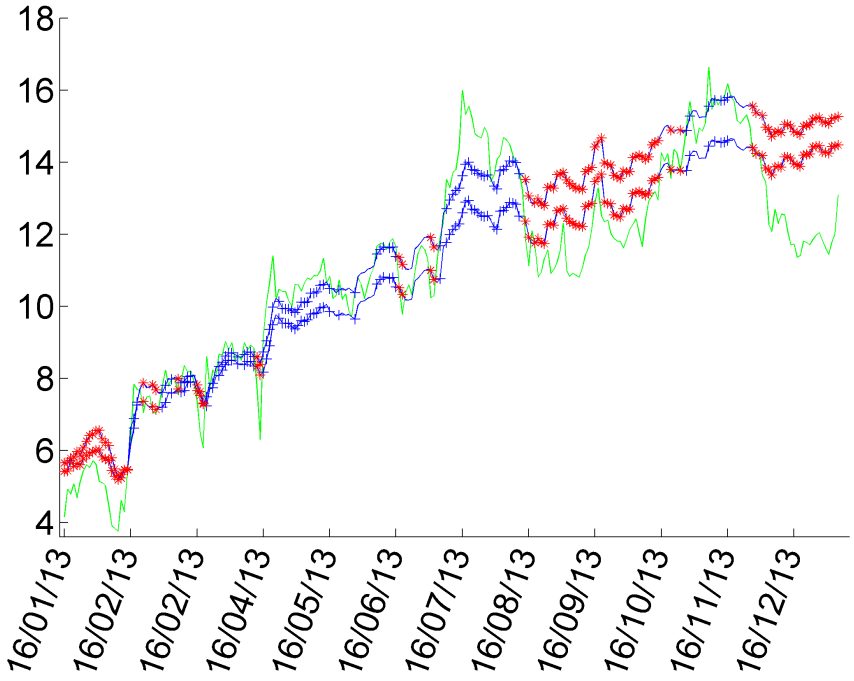


Figure 4.13: Data processed via the dynamic filtering algorithm

	(i) aggressive	(ii)normal	(iii)safe	(iv)imaginary
Profit,$	16.6361	16.0067	15.8564	18.4892

Table 4.3: Pairs trading profits using the dynamic approach with interest rate of 0.01%/per day and initial capital of zero

variation of the normal trade and employed when there is uncertainty in the positions, i.e. $\hat{r}_t(1) < r_t < \hat{r}_t(2)$, in which case all the capital is invested in the money market; and finally, (iv) imaginary trade in which the normal trade is performed on every data point of the "processed" data subset as described above; of course, this type of data set is not available in real time and, therefore, the result of this trade is used only for comparison.

Our results show that the aggressive strategy does not necessarily produce a significant increase in profits. In investing all earned capital in the spread portfolio, the main part of the profit is collected whilst the investor takes a long position on the portfolio.

Considering the long waiting times and the not significant increase in the value of the spread portfolio, the overall profit is consequently not significant as well.

One reason for the decrease of portfolio value is due to losses taken during the "incorrect" trades. Even though the decline of the profit is minimised by keeping a double position as explained in subsection 4.2.2, the increase of the portfolio value offsets the safe position. The little difference in the trading results between the safe and normal position can be explained by the behaviour of θ_t. If θ_t is always close to 0.5, this strategy corresponds to taking long and short positions with almost equal proportions of the portfolio. This means that the best strategy is doing nothing whilst investing all previous profits in the money market account.

As expected, the imaginary strategy outperforms all other trading strategies although the earned margin profit is not as high as one would predict. We rationalised this on the basis of the size of the moving window, which sets the maximum profit strategy. If the frequency of changing positions increases, it is hard to say anything about the amount that will be earned or lost as a consequence of the frequent trade. Any extra profit or loss is treated as a result of pure noise and it cannot be controlled within our set-up.

Using the values in Table 4.2, we estimate the parameters of our proposed modelling framework. The dynamics of the implied parameters ν, ζ, α and ξ calculated using the single-regime dynamic approach are shown in Figures 4.14-4.17. The graph of the "best" estimates of the process defined as $E[r_{k+1}] = \hat{\nu}E[r_k|Y_k]+\hat{\zeta}$ is given in Figure 4.18.

We tried using different sets of parameters as starting values, but some divergence in the parameters' behaviour does not disappear. This is to be expected anyhow because the data set dictates the dynamics of the parameters. That is, if convergence has to occur, almost any set of values can be assumed as starting values for the algorithms.

4.5 Conclusions and directions for further research

This work improved the performance of the pairs trading strategy proposed by Elliott et al. [8]. The adaptive power of two popular filtering approaches, namely, the Kalman

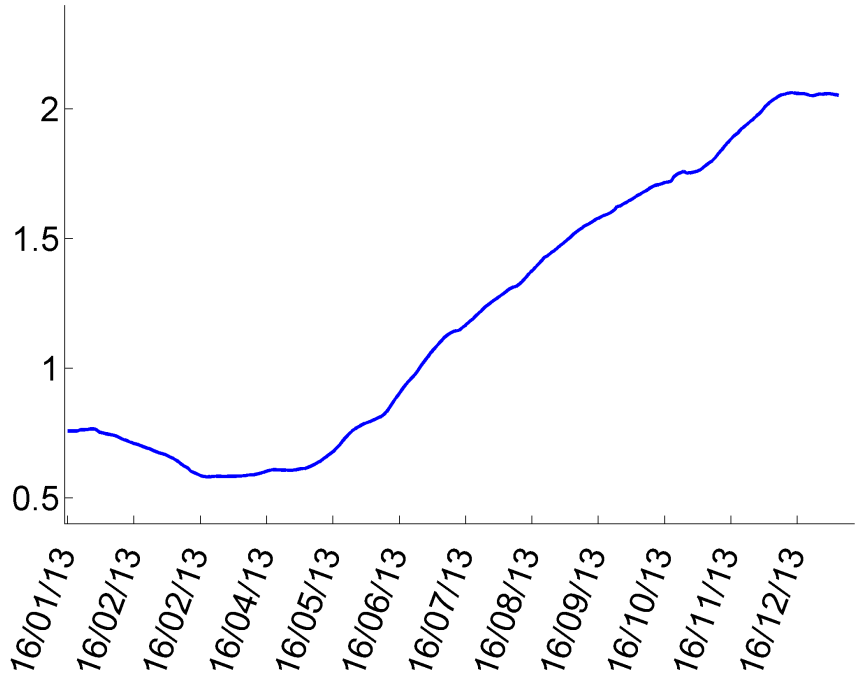


Figure 4.14: Evolution of the implied ν

and HMM filters were synthesised. Based on empirical evidence, the new hybrid algorithm outperforms each of the individual filtering technique. The proposed algorithm is more robust and well-suited to the pairs trading method, and requires the same computational resources to those of the conventional Kalman filters on Gaussian-type data sets.

We explored how significant the gain of having multi-dimensional filters in terms of accuracy without diminishing execution time and increasing complexity of the optimisation methods. Our applications proved successful in trading simulation. Under the assumptions of no transaction costs and bid-ask spreads, the back-testing trade performance showed that a hypothetical trader could manage to earn a non-zero profit with positive probability. The trading technique has the potential to yield positive gain especially when the transaction fees are low. Although our approach based on the combination of two filtering methods were tested on historical prices of Coca-Cola Co and PepsiCo Inc, practitioners could adopt this to other data sets and exploit arbitrage opportunities due to temporary market inefficiencies or other factors.

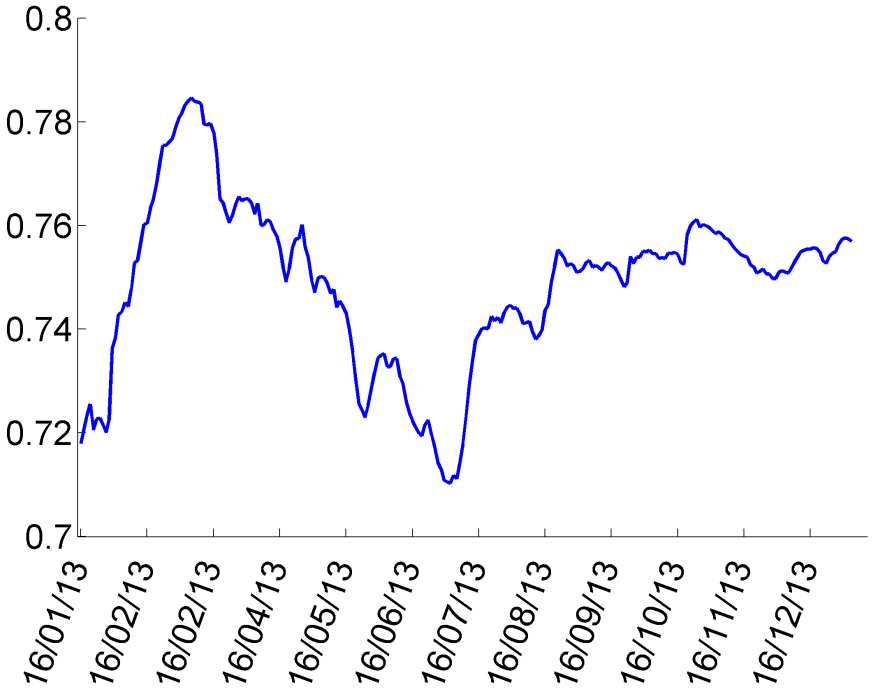


Figure 4.15: Evolution of the implied ζ

The inclusion of transaction costs for pairs trading under the multi-regime framework is open for further investigation. The cost structure needs to be defined and embedded into the procedure. Also, the use of high-frequency data may be considered as this is the case in current practice for hedging funds companies and other institutional investors. Further examination is required on how to produce filtered parameter estimates within their specified constraints without, or at least with minimal, additional numerical optimisation procedures.

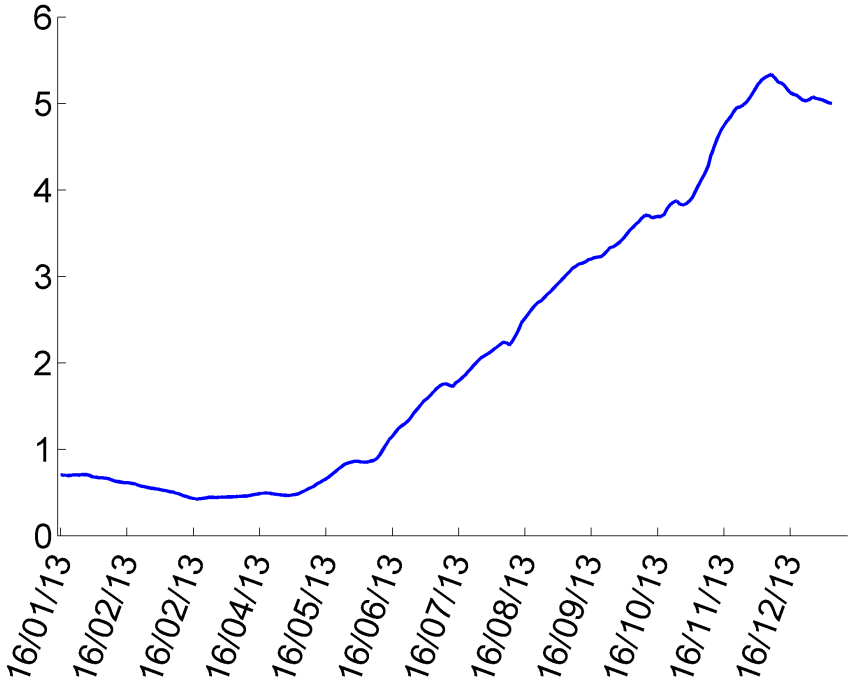


Figure 4.16: Evolution of the implied ξ

4.6 References

[1] Akaike, H., 1974. A new look at the statistical model identification, IEEE Transactions on Automatic Control 19(6), 716–723. 111

[2] Ceulemans, E., Kiers, H., 2006. Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method, British Journal of Mathematical and Statistical Psychology 59, 133–150. 111

[3] Date, P., Ponomareva, K., 2011. Linear and nonlinear filtering in mathematical finance: A review, IMA Journal of Management Mathematics 22, 195–211. 108

[4] Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM Algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39(1), 1-38. 102

[5] Goldstein, S. 2009. German billionaire reportedly commits suicide. Market Watch,

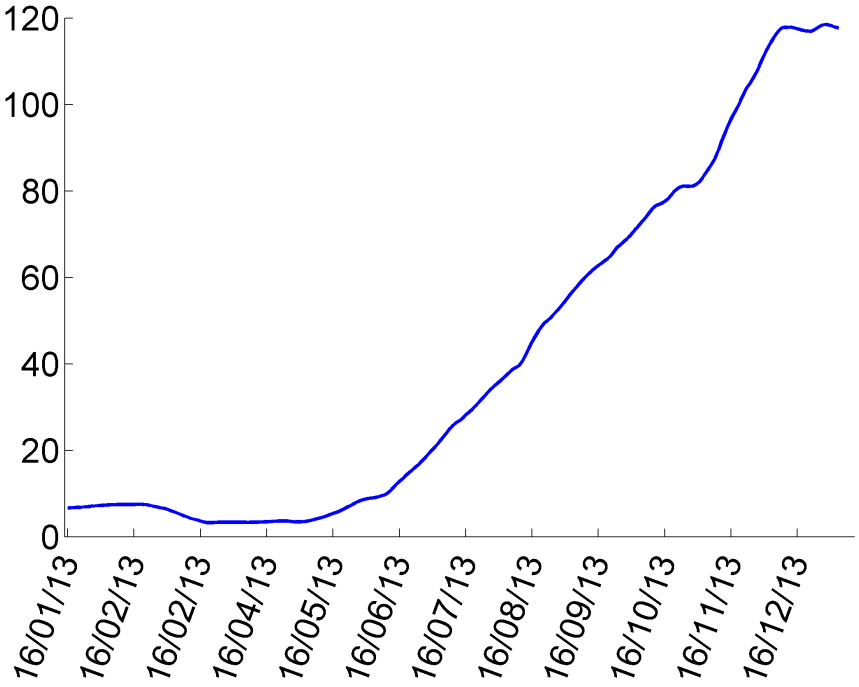


Figure 4.17: Evolution of the implied α

The Wall Street Journal. http://www.marketwatch.com/story/german-billionaire-said-to-commit-suicide-after-vw-losses . 93

[6] Elliott, R., Aggoun, L., Moore, J., 1995. Hidden Markov Models: Estimation and Control, Springer, New York. 93, 99, 101, 111

[7] Elliott, R., Krishnamurthy., 1999. New finite-dimensional filters for parameter estimation of discrete linear Gaussian models, IEEE Transactions of Automatic Control 44(5), 938–951. xii, 92, 93, 104, 105, 111, 112

[8] Elliott, R., van der Hoek, J., Malcolm, W., 2005. Pairs Trading, Quantitative Finance 5(3), 271–276. 92, 95, 97, 105, 111, 112, 115

[9] Erlwein, C., Mamon, R., 2009. An online estimation scheme for Hull-White model with HMM-driven parameters, Statistical Methods and Applications 18(1), 87–107. 95, 100, 101, 102, 108

[10] Erlwein, C., Benth, F., Mamon, R., 2010. HMM filtering and parameter estimation of an electricity spot price model, Energy Economics 32(5), 1034–1043. 108

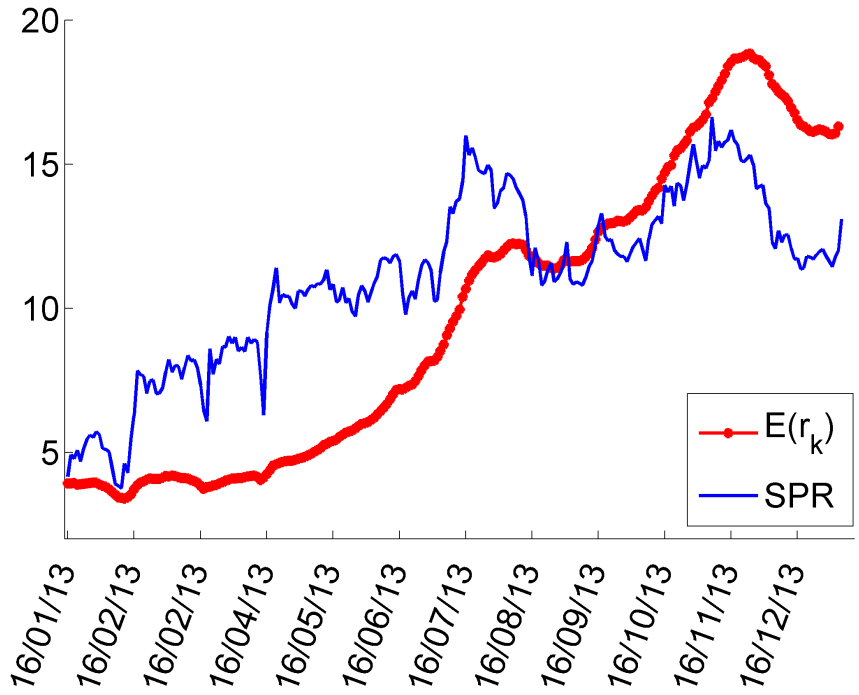


Figure 4.18: Comparison between $\widehat{r}_k$ and SPR

[11] Erlwein, C., Mamon, R., Davison, M., 2011. An examination of HMM-based investment strategies for asset allocation, Applied Stochastic Models in Business and Industry 27, 204–221. 93, 100, 101, 105, 112

[12] Hardy, M., 2002. A regime-switching model of long-term stock returns, North American Actuarial Journal 6(1), 171–173. 108

[13] Khandani, A., Lo, A., 2011. What happened to the quants in August 2007? Evidence from factors and transactions data, Journal of Financial Markets 14(1), 1–46. 93

[14] Schwarz, G., 1978. Estimating the dimension of a model, Annals of Statistics 6(2), 461–464. 111

[15] Shumway, R., Stoffer, D., 1982. An approach to time series smoothing and forecasting using the EM algorithm, Journal of Time Series Analysis 3(4), 253–264. 92

[16] Steele, M., 2012. Financial Time Series, statistics course, http://www-stat.wharton.upenn.edu/steele/Courses/434/434Context/PairsTrading/PairsTrading.html . 92

[17] Tenyakov, A., Mamon, R., Davison, M. 2014. Filtering of an HMM-driven multivariate Ornstein-Uhlenbeck model with application to forecasting market liquidity, working paper, University of Western Ontario, London, Canada. 95, 105, 112

[18] Tenyakov, A., Mamon, R., 2013. Modelling high-frequency exchange rates: A zero-delay multidimensional HMM-based approach, working paper, University of Western Ontario, London, Canada. 105, 112

[19] X. Xi, Mamon, R., 2013. Yield curve modelling using a multivariate higher-order HMM, in Zeng, Y. and Wu, S., State-Space Models and Applications in Economics and Finance, Springers Series in Statistics and Econometrics for Finance 1, 185–202.

5

Modelling high-frequency FX rate dynamics: A zero-delay multi-dimensional HMM-based approach

5.1 Introduction

Trading that entails days and weeks to complete decades ago are now being carried out in a fraction of seconds. There is apparently a shift in the market in recent years from the traditional long-term buy-and-hold strategy to short-term trading that involves fast execution of quote orders by computer programmes creating a cascade of buying and selling securities. Such high-frequency trading (HFT) employs modern technological tools and algorithms to achieve rapid trading. Trading firms, such as hedge funds, rake in profits from moving in and out of positions in microseconds to trade stocks, bonds and futures taking advantage of minute price differences detected by computers. Critics say high-speed trading exaggerate wild price swings, but advocates assert it helps provide price discovery and liquidity to the marketplace.

Shah [36] reported that computerised HFT has been making inroads in the market for currency derivatives. As indicated by the Boston-based consulting firm Aite Group, HFT already accounts for up to 30% of activity in the global FX market, mostly in heavily traded currencies like the dollar. In the same report, it was also noted that

Credit Suisse, for example, maintains an advanced execution services system for its lines of FX products globally, and such automated system allows its traders to manage option risks. Indeed, whilst the use of HFT is widespread in equities and commodities market, banks and hedge funds have been continuously making a push into niche areas including FX derivatives and the more thinly emerging market currencies. Given these developments and the fact that the value of currency derivatives depend on the movements of FX rates, there is therefore a strong demand for fast and reliable modelling of FX rates' future evolution as well as the accurate estimation of their volatilities.

A dependable FX rate model is essential in the pricing of FX derivatives as well as in the risk management and optimisation of portfolios containing assets and products that have FX rate exposures. In addition to the ability of the model to adequately replicate the stylised features of the FX process, there is also a need to be able to dynamically implement the model with relative ease and accessibility. As pointed out in Meese and Rogoff [31], a simple random walk (RW) describes the FX rate dynamics better than most of the suggested modelling approaches. This has the implication that majority of previously proposed FX rate models are unsatisfactory in terms of their out-of-sample forecasting and statistical fitting performance. In a paper that examines several FX rate models showing their advantages and disadvantages and illustrating the non-existence of a strong correlation between the behaviour of major economic factors and the FX rate dynamics, Nailliu and King [32] also found that only a few models can do better than the RW model in a comparative survey covering the last 3 decades of data.

In this chapter, we propose an alternative model along with its implementation methodology in explaining better the behaviour of FX rates. We show that under some performance metrics, this proposed model outperforms the RW model. Our methodology and modelling formulation reflect the present frequent trading practices of major players in the FX market as well as the mechanism to capture the appropriate state of the whole economy as time goes by. Using various statistical tests and empirical data, we show certain advantages of a zero-delay multivariate HMM in terms of its fitting and prediction capacity. As stated in Nailliu and King [32], it is possible to predict FX spot rates more accurately than those produced by the RW model if information about the structure of the market trading is aptly incorporated in the model. Inspired by this realisation, we introduce a model, given its simplicity and implementation attainability

with respect to the current computing technologies, that may be easily adopted by the practitioners. This augments the currently used time series based models such as the ARCH/GARCH and stochastic volatility (SV) models. Within the context of FX rate modelling, a vast literature has sprung from Engel's ARCH model [15] as noted in Maheu and McCurdy [30] and a discussion of the SV approach is given in Taylor [37] .

Certain information that drastically affects market prices of frequently traded assets has an instant influence in the dynamics of other financial variables such as FX rates. As elaborated in Haldane [22], traders engaging in HFT worldwide via computer programmes could automatically change their positions and bring some over-all 40,000 spot-price changes in less than a second. Thus, the assumption of current price dependence on previous asset information recorded on a time lag of one day or wider frequency, which is the usual framework for many discrete-time HMM-based approaches, is deemed insufficient. With liquid trading that affords volumes of readily available information nowadays, the utility of the zero-delay model is well justified.

In a previous study, Engel and Hamilton [16], for instance, suggested the suitability of an HMM in modelling FX rates based on raw data taken as an arithmetic average of the bid-and-ask prices for the exchange rate (in dollars per unit of foreign currency) for the last day of the quarter, beginning with the third quarter of 1973 and ending with the first quarter of 1988. Cheung and Erlandsson [7] contributed to this idea by incorporating a rigorous Monte-Carlo approach to test for multi-regime dynamics of quarterly and monthly data. In Yuan [41], FX rates are modelled using a static HMM with a special smoothing parameter. Since many successful trading strategies are dependent on fast FX rate movements (cf. Aldridge [2]) and given the peculiar characteristics of the FX data collected recently, we regard the zero-delay HMM to be more appropriate than the commonly used one-step-delay HMM. Evidence will also show that the new resulting dynamic filtered estimators for model parameters are able to capture better the trends of the FX rate process.

We adapt the technique of Elliott et al [13] in establishing expressions for zero-delay filters. Such an approach uses the change of probability measure method to evaluate filters under some ideal measure and relate the calculations back to the real-world measure through the Bayes' theorem for conditional expectation. In the derivation of the filter expressions, we also incorporate the idea explored in Erwein et al. [17] concerning

the dependencies of several processes evolving in parallel on the same discrete-time Markov chain. This approach performs well in delineating the amount of fluctuations of the FX caused by volatilities from those due to regime changes. In arguing that a multi-regime behaviour is present in a given data set, we employ a sequential algorithm following closely the testing procedure of Rodionov [33]. The procedure tailored to our HMM filtering-based estimation technique provides clear support for the existence of regimes. In establishing a statistically correct quantity of regimes needed to model the data, we use the Bayesian Information Criterion (BIC), which is a penalised version of the Akaike Information Criterion (AIC). The justification based on BIC is further reinforced by the CHull criterion.

We examine the fitting and prediction performance of our proposed model and estimation approach against popular models using FX rate data on two major currency pairs, namely the Japanese yen (JPY) vesus US dollar (USD) and JPY versus UK sterling pound (GBP). The JPY is chosen as the base currency in our study owing to its reputation as one of the most volatile currencies in the world, and hence, regime-switching model is a reasonable model to investigate the characterisation of its dynamics.

This research develops and examines the performance of a zero-delay model in capturing the stylised characteristics of FX rates recorded with greater frequencies. The zero-delay model may appear to be a slight modification of the usual one-step-delay model, but such modification could produce different empirical results favouring the former model. Furthermore, whilst zero-delay and one-step-delay models may at times yield similar conclusions, the examination of data on leading currencies demonstrates that the former model could outperform its competitors with respect to better fit and prediction accuracy for out-of sample forecasts. We show the robustness of the zero-delay model by testing it on a random subset of data and analysing its performance vis-a-vis those of the alternative models on the same data set.

To achieve our objectives, this chapter is structured as follows. Section 5.2 describes the modelling framework and filtering recursions of a zero-delay model. The multi-dimensional filtering scheme is then considered and estimates of the parameters for the proposed model are determined using the EM algorithm. In section 5.3, we provide numerical evidence that favours the proposed model over other competing models with

respect to goodness-of-fit measures and log likelihood-driven criteria. This chapter culminates with some concluding remarks in section 5.4.

5.2 Model formulation

If s_k is the FX rate at time t_k, we posit that over a very short-time period, i.e., over several seconds or minutes, the dynamics of its log returns evolve according to the equation

$$y_k := \ln \frac{s_k}{s_{k-1}} = \mu\ \Delta t_k + \sigma\sqrt{\Delta t_k}\ \epsilon_{t_k}. \tag{5.1}$$

In equation (5.1), the parameters μ and σ are the respective drift and volatility parameters under the real-world measure P, where $\epsilon_{t_k} \sim N(0,1)$, i.e., ϵ_{t_k} is a standard normal random variable and $\Delta_{t_k} = t_k - t_{k-1}$.

Given the stochastic nature of the FX process, μ and σ do not only change with time but also depend on the state of world according to a finite-state discrete-time Markov chain $\boldsymbol{x}_{t_k}$ incorporating the over-all interaction of various factors. Therefore, equation (5.1) can be rewritten as

$$\begin{aligned} y_k := \ln \frac{s_k}{s_{k-1}} &= \mu\ (\boldsymbol{x}_{t_k})\Delta t_k + \sigma(\boldsymbol{x}_{t_k})\sqrt{\Delta t_k}\ \epsilon_{t_k}, & (5.2)\\ \boldsymbol{x}_{t_k} &= \mathbf{A}\boldsymbol{x}_{t_{k-1}} + \boldsymbol{v}_{t_k}, & (5.3) \end{aligned}$$

where $\mathbf{A}$ is a transition probability matrix and $\boldsymbol{v}_{t_k}$ is a martingale increment. This formulation is extended to the multivariate setting below. For brevity, we shall write $\boldsymbol{x}_k$ instead of $\boldsymbol{x}_{t_k}$ and similar notational adjustments will made to other variables.

Let $(\Omega, \mathcal{F}, P)$ be a probability space under which $\boldsymbol{x}_k$ is a homogeneous Markov chain with a finite-state space. Without loss of generality, let $\Delta t_k = 1$. The dynamics of the $d-$dimensional observation process is then

$$y_k^g = \mu^g(\boldsymbol{x}_k) + \sigma^g(\boldsymbol{x}_k)\epsilon_k^g\,, \quad 1 \le g \le d, \tag{5.4}$$

where $\{\epsilon_k^g\}$ is a sequence of independent standard Gaussian random variables. In particular, they are independent for each component of the row vector y_k. To considerably simplify the algebra involved in the filtering equations, we associate the state space of $\boldsymbol{x}_k$ with the standard basis of $\mathbb{R}^N$, which is the set of unit vectors $\boldsymbol{e}_h,\ h = 1, 2, \ldots, N$ and $\boldsymbol{e}_h$ is a vector having 1 in its h^{th} entry and 0 elsewhere. So, in equation (5.4),

$\mu^g(\boldsymbol{x}_k) = \langle \boldsymbol{\mu}_k^g, \boldsymbol{x}_k \rangle$ and $\sigma^g(\boldsymbol{x}_k) = \langle \boldsymbol{\sigma}_k^g, \boldsymbol{x}_k \rangle$. The notation $\langle \cdot, \cdot \rangle$ is the usual scalar product and $\boldsymbol{\mu}^g = (\mu^g(1), \mu^g(2), \ldots, \mu^g(N))^\top, \boldsymbol{\sigma}^g = (\sigma^g(1), \sigma^g(2), \ldots, \sigma^g(N))^\top \in I\!R^N$, where $\top$ denotes the transpose of a vector.

The use of mixture of normals in our filtering methodology is well supported. Virtually, every distribution can be approximated by a mixture of normals with any required precision. Applications to finance under this setting can be found in Kamaruzzaman et al. [26], Wirjanto and Xu [40], or Jung [21], amongst others. Some works on mixtures of normal distributions with HMM to deal with FX modelling for option (e.g. Kaehler and Marnet [25]) found that normal mixtures capture well the leptokurtosis of the data, whereas the Markov-switching models capture both leptokurtosis and the heteroskedasticity. In this work, we verify the assumption of relative normality. The term "relative" refers to the fact that it is not possible and, in certain application, is not feasible to check normality on every subset of available data. Modelling assumptions in this paper both from the intuitive and rigorous levels are discussed in the succeeding sections. As an alternative to our proposed method, some heavy-tailed noise drivers, such as t-distributed Levy processes, for example, could be used. Even though the model might produce better fit, the calculation becomes less tractable, and hence such alternative is avoided by practitioners. We work under mixture of normals assumption as the main goal of the project is to produce estimation schemes that can be easily implemented in practice. Of course, it may be necessary to employ non-normal distributions when the data strongly suggest but in that case it is also possible to increase the number of normal distributions so that the mixture of normals is able to provide a better fit.

To find the optimal estimate for the state of $\boldsymbol{x}_k$, we use a change of measure approach described in Elliott et al. [13]. We perform the calculations under a new, ideal measure $\widetilde{P}$ under which all observations y_k's are IID standard normal random variables and y_k's are independent from $\boldsymbol{x}_k$. Using a discrete-time version of the Girsanov's theorem, the real-world measure P can be recovered via the Radon-Nikodým derivative

$$\Lambda_k := \frac{dP}{d\widetilde{P}}\bigg|_{\mathcal{F}_k} = \prod_{g=1}^{d} \prod_{l=1}^{k} \lambda_l^g, \ k \geq 1,$$

$$\Lambda_0 = 1 \ \text{ and } \ \lambda_l^g = \frac{\phi[\sigma^g(\boldsymbol{x}_l)^{-1}(y_l^g - \mu^g(\boldsymbol{x}_l))]}{\sigma^g(\boldsymbol{x}_l)\phi(y_l^d)},$$

where $\phi(\cdot)$ is the density function of an $N(0,1)$ random variable. The filtration $\mathcal{F}_k$ is given by $\mathcal{F}_k := \mathcal{F}_k^{\boldsymbol{x}} \vee \mathcal{F}_k^y$, where $\mathcal{F}_k^{\boldsymbol{x}}$ and $\mathcal{F}_k^y$ are the filtrations generated by the Markov chain $\boldsymbol{x}_k$ and observation process y_k, respectively.

We construct the filters for relevant quantities related to the Markov chain under the multivariate setting. Model parameter estimates will then be obtained in terms of these filters. Define the conditional probabilities of $\boldsymbol{x}_k$ given $\mathcal{F}_k$ under P as

$$p_k^i := P(\boldsymbol{x}_k = \boldsymbol{e}_h | \mathcal{F}_k^y) = E[\langle \boldsymbol{x}_k, \boldsymbol{e}_h \rangle \, | \mathcal{F}_k^y].$$

Write $\widehat{\mathbf{p}}_k := (\widehat{p}_k(1), \widehat{p}_k(2), \ldots, \widehat{p}_k(N))^\top \in I\!R^N$ and so the optimal estimate for $\boldsymbol{x}_k$ given the available information up to time k is

$$\widehat{\mathbf{p}}_k = E[\boldsymbol{x}_k | \mathcal{F}_k^y] = \frac{\widetilde{E}[\Lambda_k \boldsymbol{x}_k | \mathcal{F}_k^y]}{\widetilde{E}[\Lambda_k | \mathcal{F}_k^y]} \tag{5.5}$$

by the Bayes' theorem for conditional expectation. Define $\mathbf{c}_k := \widetilde{E}[\Lambda_k \boldsymbol{x}_k | \mathcal{F}_k^y]$ and note that

$$\sum_{h=1}^{N} \langle \boldsymbol{x}_k, \boldsymbol{e}_h \rangle = 1.$$

Thus,

$$\sum_{h=1}^{N} \langle \Lambda_k, \boldsymbol{e}_h \rangle = \sum_{h=1}^{N} \left\langle \widetilde{E}[\Lambda_k \boldsymbol{x}_k | \mathcal{F}_k^y], \boldsymbol{e}_h \right\rangle = \widetilde{E}\left[\Lambda_k \sum_{h=1}^{N} \langle \boldsymbol{x}_k, \boldsymbol{e}_h \rangle \middle| \mathcal{F}_k^y \right] = \widetilde{E}[\Lambda_k | \mathcal{F}_k^y]. \tag{5.6}$$

The construction of $\mathbf{c}_k$ along with equation (5.6) yields

$$\widehat{\mathbf{p}}_k = \frac{\mathbf{c}_k}{\sum_{h=1}^{N} \langle \mathbf{c}_k, \boldsymbol{e}_h \rangle}.$$

In addition to the state's optimal estimate, we also consider the following quantities:

$$\mathcal{J}^{sr}{}_{k+1} = \sum_{n=1}^{k+1} \langle \boldsymbol{x}_{n-1}, \boldsymbol{e}_r \rangle \langle \boldsymbol{x}_n, \boldsymbol{e}_s \rangle \tag{5.7}$$

$$\mathcal{O}^r{}_{k+1} = \sum_{n=1}^{k+1} \langle \boldsymbol{x}_{n-1}, \boldsymbol{e}_r \rangle \tag{5.8}$$

$$\mathcal{T}^r_{k+1}(f(y^g_{k+1})) = \sum_{n=1}^{k+1} \langle \boldsymbol{x}_{n-1}, \boldsymbol{e}_r \rangle f(y_n^g) \ , \ 1 \le r \le N \ \ 1 \le g \le d \tag{5.9}$$

with $f(y_n^g) = y_n^g$ or $f(y_n^g) = (y_n^g)^2$.

Equation (5.7) counts the number of jumps from $\boldsymbol{e}_r$ to $\boldsymbol{e}_s$ at time t_{k+1}. The amount of time up to t_{k+1} that $\boldsymbol{x}$ occupies state $\boldsymbol{e}_r$ is given by $\mathcal{O}_{k+1}^r$ in (5.8). The function $\mathcal{T}_{k+1}^r(f)$ in (5.9) is an auxiliary quantity that occurs in the estimation of model parameters.

For any process G_k, we denote the conditional expectation under $\widetilde{P}$ of $\Lambda_k G_k$ by $\gamma(G_k) := \widetilde{E}[\Lambda_k G_k | \mathcal{F}_k^y]$. The adaptive filters enable the updating of model parameters that will incorporate past and current market conditions. Taking advantage of the semi-martingale representation of $\boldsymbol{x}_k$, we obtain the recursive filters for $\gamma(\mathcal{J}_k^{ji}\boldsymbol{x}_k)$, $\gamma(\mathcal{O}_k^i\boldsymbol{x}_k)$ and $\gamma(\mathcal{T}_k^i(f(y_k^g))\boldsymbol{x}_k)$, and they are presented in the results that follow.

Proposition 1: *Suppose* $\mathbf{A} = (a_{ji})$ *is the transition matrix,* $\boldsymbol{a}_i = \mathbf{A}\boldsymbol{e}_i$*, and*

$$\Gamma^i = \Gamma^i(\underline{y_{k+1}}) = \Gamma^i(y_{k+1}^1, ..., y_{k+1}^d) = \prod_{g=1}^{d} \frac{\phi\left(\frac{y_{k+1}^g - \mu^g(i)}{\sigma^g(i)}\right)}{\sigma^g(i)\phi\left(y_{k+1}^g\right)}.$$

Then

$$\gamma(\boldsymbol{x}_{k+1}) = \sum_{i,j=1}^{N} \langle \gamma(\boldsymbol{x}_k), \boldsymbol{e}_i \rangle \, a_{ji} \Gamma^j \boldsymbol{a}_i \tag{5.10}$$

$$\begin{aligned} \gamma(\mathcal{J}_{k+1}^{ji}\boldsymbol{x}_{k+1}) &= \sum_{m,l=1}^{N} \left\langle \gamma(\mathcal{J}_k^{ji}\boldsymbol{x}_k), \boldsymbol{e}_m \right\rangle a_{lm}\Gamma^l \boldsymbol{a}_m \\ &+ \langle \gamma(\boldsymbol{x}_k), \boldsymbol{e}_i \rangle \, a_{ji}\Gamma^j \boldsymbol{a}_i \end{aligned} \tag{5.11}$$

$$\begin{aligned} \gamma(\mathcal{O}_{k+1}^{i}\boldsymbol{x}_{k+1}) &= \sum_{m,l=1}^{N} \left\langle \gamma(\mathcal{O}_k^{i}\boldsymbol{x}_k), \boldsymbol{e}_m \right\rangle a_{lm}\Gamma^l \boldsymbol{a}_m \\ &+ \sum_{l=1}^{N} \langle \gamma(\boldsymbol{x}_k), \boldsymbol{e}_i \rangle \, a_{li}\Gamma^l \boldsymbol{a}_i \end{aligned} \tag{5.12}$$

$$\begin{aligned} \gamma(\mathcal{T}_{k+1}^{i}\left(f\left(y_{k+1}^g\right)\right)\boldsymbol{x}_{k+1}) &= \sum_{m,l=1}^{N} \left\langle \gamma(\mathcal{T}_k^{i}\left(f\left(y_k^g\right)\right)\boldsymbol{x}_k), \boldsymbol{e}_m \right\rangle a_{lm}\Gamma^l \boldsymbol{a}_m \\ &+ f(y_{k+1}^g) \sum_{l=1}^{N} \langle \gamma(\boldsymbol{x}_k), \boldsymbol{e}_i \rangle \, a_{li}\Gamma^l \boldsymbol{a}_i. \end{aligned} \tag{5.13}$$

Proof: See Appendix.

It is important to note that even though the above recursive formulae look similar to corresponding equations in a one-step-delay model, the derivation is based on the new model. In this research project, the filtering algorithms developed by Elliott [13] are extended under a zero-delay modelling framework. Compared to the one-step delay filtering equations, there is the presence of multiplier terms in the filtering equations of the current framework; namely, the terms a_{lm}, a_{ji}, etc in equations (5.11)-(5.13). Although this extension was alluded to in the second edition of Elliott et al. [13], this was not comprehensively examined in the succeeding reprints, and more importantly, applications to data were not investigated.

Another difference between the one-step-delay and zero-delay sets of filters lies in the processing of data sets that is backshifted in time for the former set of filters through the transition matrix. Both filtering methods yield recursive formulae with similar forms. However, it is observed that with a greater number of time steps in the processing of data sets in combination with an appropriate smoother for the one-step-delay modelling framework, both filters will provide close parameter estimates.

We apply the Expectation-Maximisation (EM) algorithm (cf. Dempster et al. [12]) to find the optimal estimates of the model parameters. As the next result shows, the transition probabilities a_{ji}, and the levels of the drift $\mu(i)$ and volatility $\sigma(i)$ are expressed in terms of the quantities in (5.7)-(5.9). The following results are stated without proof. It has to be noted that although we work under the zero-delay modelling set-up, the reasoning behind the proof of the following results are similar to those in the one-step delay model case (see Erlwein et al. [17]), and hence it is easily reproducible.

Proposition 2: *Consider a multivariate dataset* $y_1^g,\ y_2^g, \ldots, y_k^g,\ 1 \leq g \leq d$ *observed up to time* k. *The respective EM estimates of parameters* $\{a_{ji},\ \mu^g(i),\ \sigma(i)\}$

$$\widehat{a}_{ji} = \frac{\gamma\left(\mathcal{J}_k^{ji}\right)}{\gamma\left(\mathcal{O}_k^i\right)} \tag{5.14}$$

$$\widehat{\mu}^g(i) = \frac{\gamma\left(\mathcal{T}_k^i\left(y_k^g\right)\right)}{\gamma\left(\mathcal{O}_k^i\right)} \tag{5.15}$$

$$\widehat{\sigma}^g(i) = \sqrt{\frac{\gamma\left(\mathcal{T}_k^i\left(\left(y_k^g\right)^2\right)\right) - 2y_k^g\gamma\left(\mathcal{T}_k^i\left(y_k^g\right)\right) + \left(y_k^g\right)^2\gamma\left(\mathcal{O}_k^i\right)}{\gamma\left(\mathcal{O}_k^i\right)}} \tag{5.16}$$

Now, in terms of the recursive filters given in Proposition 1, the optimal estimate of the Markov chain is then

$$\widehat{\mathbf{p}}_k = \frac{\mathbf{c}_k}{\sum_{h=1}^{N} \langle \mathbf{c}_k, \boldsymbol{e}_h \rangle} = \frac{\gamma(\boldsymbol{x}_k)}{\sum_{h=1}^{N} \langle \gamma(\boldsymbol{x}_k), \boldsymbol{e}_h \rangle}.$$

Furthermore, for any process G_k, we have

$$\gamma(G_k) = \gamma(G_k \langle \boldsymbol{x}_k, \mathbf{1} \rangle) = \langle \gamma(G_k \boldsymbol{x}_k), \mathbf{1} \rangle .$$

Hence, when $G_k = \mathcal{J}_k$, $\mathcal{O}_k$ or $\mathcal{T}_k$, equations (5.14)-(5.16) in Proposition 2 are fully determined, and Proposition 1 gives the dynamic updates of the model parameters.

5.3 Numerical case study

In this section, we analyse the performance of our proposed model and estimation method. An implementation to FX rate data compiled by Bloomberg is conducted. Two data sets are considered, namely, the JPY/USD and JPY/GBP, spanning the period of 0935 HRS, 06 July 2012 – 1840 HRS, 11 July 2012 with a five-minute interval between each observation. The JPY was selected as the base currency as it is known to possess random wild fluctuations in log returns or increments. This characteristic is deemed suitable for an HMM to capture.

The distributional structure of Japan's natural resources and electricity supply combined with its geographical location implies that the JPY movement is mostly affected by internal factors (e.g., occurrence of natural disasters, supply and demand of raw materials, political climate, institutional policies, manufacturing sector stability, amongst others). Data sourced out from the World Bank in 2012 [38] reveals that Japan is the third largest economy in terms of total GDP and it is also the third largest automobile manufacturing country. The share of automotive exports in Japan is 17 percent and the industry actively participates in generation of FX earnings (cf. Klink et al. [28]); this is direct consequence due to automobile manufacturers reaping benefits of globalisation by means of exports. The Honda Motor Corporation's 2012 Report explained that the variations in Honda's stock price volatility are caused by various factors including fierce and increasing competition, short-term fluctuations in demand, changes in tariffs, import regulations and other taxes, and shortages of certain materials, amongst others.

Since Honda is one of the biggest contributors to the national GDP, it is recognised that its stock price is highly correlated with the JPY currency. Therefore any major price movements in the value of major car companies have strong association with large price movements of JPY relative to the USD, GBP and other major currencies trading in the FX market. Apparently, the above-mentioned circumstances surrounding Japan could contribute to the volatile nature of its currency against major currencies such as the GBP and USD.

In this project, we take FX rates as inputs to a model and assume that they contain latent information. These FX rates are "filtered" to extract the information essential in the characterisation of the "best" estimates of parameters for our proposed model. These "best" estimates can in turn be utilised for pricing derivatives, risk management and forecasting over a short horizon. From the actual data, we generate the observed log returns y_k^g for $g = 1, 2$ ($1 \equiv$ JPY/GBP, $2 \equiv$ JPY/USD), $k = 1, \ldots, 1000$. It is assumed that the log return process is decomposed into two parts: the mean μ and volatility σ, which are both driven by an unobserved HMM $\boldsymbol{x}_k$. The goal is to estimate μ, σ, $\boldsymbol{x}_k$ and the matrix $\mathbf{A}$ in an optimal way.

5.3.1 Regime-switching assumption in the data

Before starting the model implementation, we first validate the important modelling assumption of multi-regime behaviour in our data. In addition to simple visual check of spikes, which shows instances of regime shifts in the mean and volatility of log returns, there are a few formal statistical tests described in the literature that could be used to determine the presence of regime switches; see for example Rodinov [34]. We choose a sequential algorithm adapted from the works of Rodionov [33] and apply the testing procedure with a 99.9 percent confidence level on every data point. This test essentially checks if the difference between the mean values of two consecutive regimes are statistically significant according to the principles of the Student's t-test. This test of significance for each of the mean and variance levels is a two-step test. In the first step, the size of a sample window encompassing the first few data points is chosen, and the mean and variance of the data in this sample window are calculated. These statistics (sample mean and variance) are employed in establishing a "test interval". In the second step, data points that belong outside the "test interval" are compared to a computed regime shift index (RSI). An inference of whether a possible regime switch

occurred is made on the basis of this RSI.

For the detection of regime shifts in the variance of the log returns, the data points are first de-meaned, i.e., the data is transformed to have a zero mean. The test for the shifts in variance is conducted in a manner similar to that of the regime-shift determination for the mean of the log returns. Needless to say, the size of the sample window used to construct the RSI and the "test interval" have direct impact on the detection of mean and variance shifts of the underlying process; see Rodionov [34]. For the purpose of providing support to the Markovian assumption in this paper, we heuristically find that a window of size 6 data points (30-minute interval) is adequate.

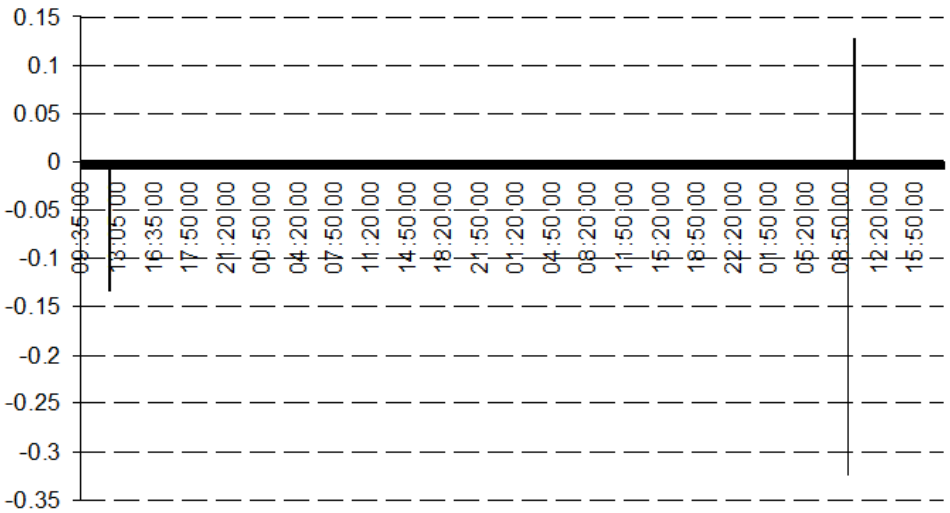


Figure 5.1: Illustrating the occurrence of regime switches in the mean of log returns with a 99.9 % confidence level for JPY/GBR covering the period 09:35, 10 July – 19:20, 11 July 2012

Figures 5.1 and 5.2 depict episodes of regime switches in the mean of log returns. Regime changes occur at times where the vertical bars are drawn. The orientation of the vertical bars (above or below the horizontal axis) indicate whether the regime change in the mean that occurred was going up or down. A salient point to notice is that within the time interval [09:15, 11:15] on 11 July 2012 (right-end portion of Figures 5.1 and 5.2), there were noticeable simultaneous changes in the mean of FX

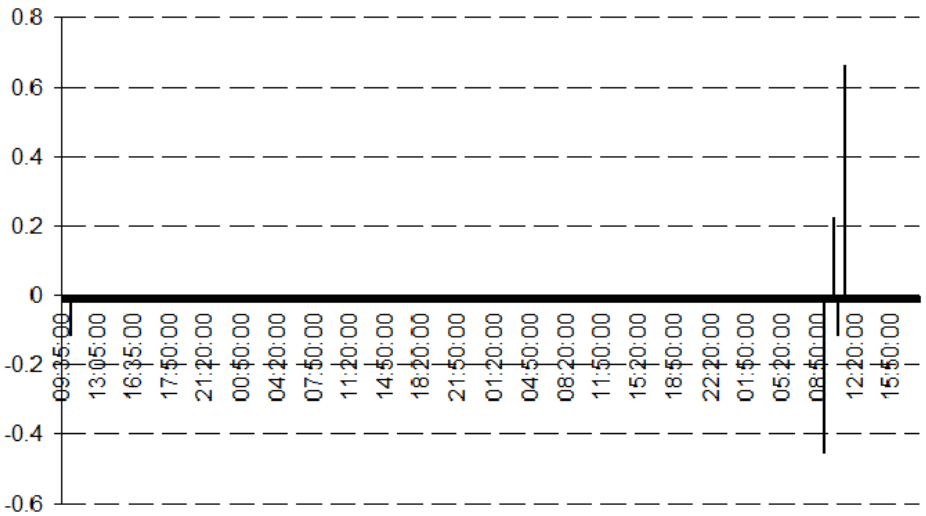


Figure 5.2: Illustrating the occurrence of regime switches in the mean of log returns with a 99.9 % confidence level for JPY/USD covering the period 09:35, 10 July – 19:20, 11 July 2012

rates' log returns for both currency pairs. This strongly supports our assumption that the joint FX rates' behaviour in the mean is driven by the same Markov chain.

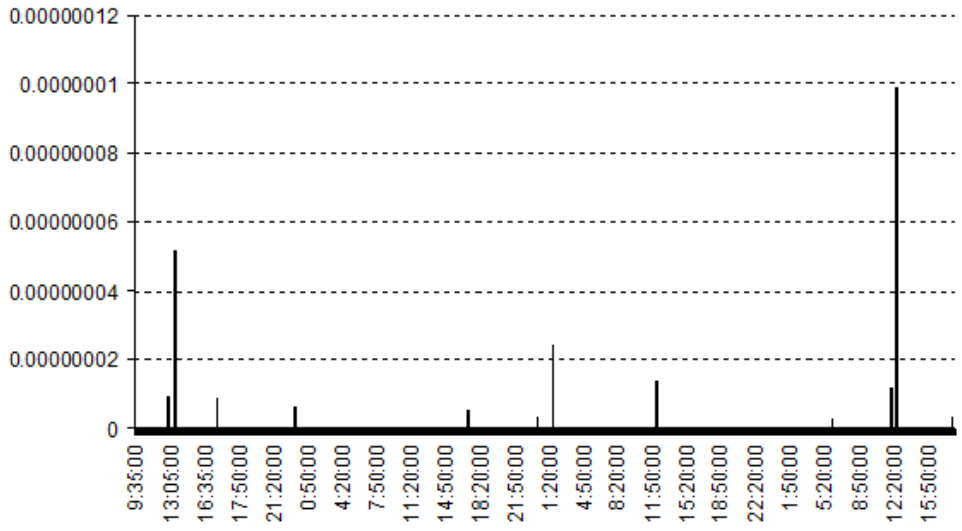


Figure 5.3: Illustrating the occurrence of regime switches in the volatility with a 99.9 % confidence level for JPY/GBR covering the period 09:35, 10 July – 19:20, 11 July 2012

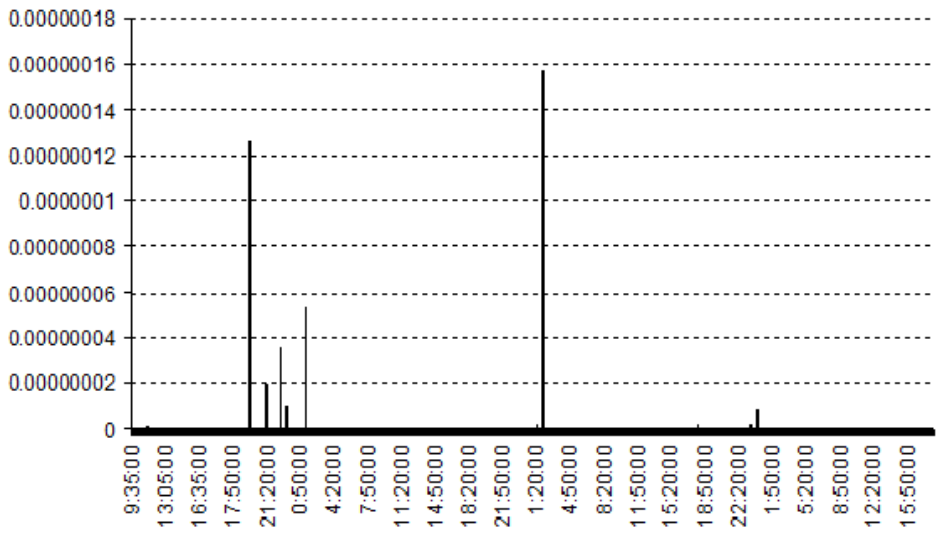


Figure 5.4: Illustrating the occurrence of regime switches in the volatility with a 99.9 % confidence level for JPY/USD covering the period 09:35, 10 July – 19:20, 11 July 2012

As portrayed in Figures 5.3 and 5.4, there is also evidence of regime switches in volatility happening closely in tandem, at 01:40 on 10 July 2012 and at 01:45 on 10 July 2012 (middle section of Figures 5.3 and5.4) for the JPY/GBR and JPY/USD, respectively. This instance serves another indication that a multidimensional HMM may work well to model our bivariate FX rate data.

In general, the result of any statistical test in establishing the presence of regime switches is a function of confidence level, size of "test interval", and the data possessing extreme volatility movement (or the lack of it). It has to be noted though that when the data is recorded with very short frequency, such as in our case, it may not be that straightforward to differentiate between the shifts in the volatility parameter and movements of volatility itself. Sudden increases in the variance are normally adjudged as regime switches although such assessment result is still test-specifc. In certain sequential tests, some spikes in the data are simply treated as outliers even though these may be considered as triggers for a regime switch in other detection tests for regime switching. Relying only on Rodionov's regime-switching test procedure [33], our aim is not to pursue other test procedures in detecting multi-regime dynamics but rather on showing enough evidence that our HMM approach is suitable for the FX-rate data sets that we are examining in this paper.

The time period for the data set in our study was chosen randomly. We note, however, that major waves of trades usually happens closer to the end of the month and markets are relatively calm during the first couple of weeks of the month. We also included a weekend/non-trading day to our sample data set to see if the parameters of the model respond to a weekend/business day transition. Roughly a week of data was picked, and this is deemed reasonably a long period of time for frequent trading.

In the next subsections, we describe several algorithms to model the distinguishing distributional characteristics of FX rates, propose approaches for model calibration, and perform model validation via goodness-of-fit tests and forecasting exercise.

5.3.2 Benchmarking the zero-delay HMM

We evaluate the performance of several competing HMMs based on several statistical criteria. We also make comparisons, through the assessment of fit and prediction performance, of our proposed modelling approach with the widely used generalised autoregressive conditionally heteroskedastic (GARCH), Bollerslev [5] and the RW model. In addition to the zero-delay filters developed in section 6.2, we consider two more HMM filtering algorithms, viz.: the static estimation approach presented in Hamilton [23] and the dynamic filters for the one-step delay model derived in Erlwein et al. [17]. The models included in the benchmarking of our zero-delay HMM are as follows.

1. Static independent log-normal model (ILN). It is assumed that the log returns process is lognormally distributed with constant mean and variance, i.e.,

$$y_k = \mu + \sigma\epsilon_k, \quad \text{where } \epsilon_k \sim N(0,1),$$

   where $N(0,1)$ stands for a standard normal random variable.

2. Static regime-switching one-step delay log-normal model with N regimes (RSLN-N). This extends the the ILN model to multi regimes. The mean and variance of log returns are not constant. Rather, they are driven by discrete-time Markov chain with finite state space. Thus,

$$y_{k+1} = \mu_S(\boldsymbol{x}_k) + \sigma_S(\boldsymbol{x}_k)\epsilon_{k+1}, \quad \text{where } \epsilon_{k+1} \sim N(0,1).$$

Under this model, a static estimation is performed by processing the entire data set, and only one set of model estimates are obtained. The notation μ_S and σ_S signify estimates calculated from this static estimation procedure. *Similar explanation goes for the subscript notation in the next two models.*

3. Dynamic regime-switching one-step delay log-normal model with N regimes (DRSLN-N). Under the DRSLN-N,

$$y_{k+1} = \mu_D(\boldsymbol{x}_k) + \sigma_D(\boldsymbol{x}_k)\epsilon_{k+1}, \text{ where } \epsilon_{k+1} \sim N(0,1).$$

This model features a dynamic estimation procedure that produces a sequence of parameter estimates evolving through filtering-algorithm steps.

4. Dynamic regime-switching zero-delay log-normal model with N regimes (ZDRSLN-N). This model has all the properties of the DRSLN-N with the assumption that the distribution of y_k depends without delay on $\boldsymbol{x}_k$, i.e., a zero-time lag dependence. That is,

$$y_k = \mu_Z(\boldsymbol{x}_k) + \sigma_Z(\boldsymbol{x}_k)\epsilon_k, \text{ where } \epsilon_k \sim N(0,1).$$

Remark 1: *We recognise that we do not explicitly model the correlation amongst the currency pairs. Nevertheless, the currency pairs are governed by the same HMM, and are therefore implicitly correlated. Filters with appropriate correlation structure for the various white noise governing the log returns of currency pairs will most likely be better. Notwithstanding the limitation of the proposed model, this research investigation can be treated as a lower bound for the credibility of a study that takes explicitly into account correlated white noise (or correlated Brownian motions in the case of continuous-time modelling set up).*

Remark 2: *Under the one-step delay model (3^{rd} model), the reaction to $\boldsymbol{x}_k$ is not instantaneous. With the zero-delay model (4^{th} model), there is no delay in the reaction to $\boldsymbol{x}_k$. Under the latter framework, this is tantamount to relabelling the observation process. So, in the calculation of filters under the zero-delay model in Proposition 1, if $\{\mathcal{F}_l{}^{y*}\}$ is the complete filtration generated by the observations then $\mathcal{F}_l{}^{y*} = \mathcal{F}_{l+1}^{y}$, where $\mathcal{F}_{l+1}^{y}$ is the filtration generated by the observations under the one-step delay model. This implies that $\widetilde{E}[\boldsymbol{x}_{k+1}|\mathcal{F}_l{}^{y*}] = \widetilde{E}[\boldsymbol{x}_{k+1}|\mathcal{F}_{l+1}^{y}]$. See further the Appendix.*

5.3.3 Numerical implementation

5.3.3.1 Optimal number of states in HMM

Our choice of the optimal number of states for the HMM is based on the maximisation of log likelihood with a penalty for increasing the complexity of the model, i.e., adding more parameters. This is further complemented by a few error analyses choosing the number of states that yield the lowest value for a given error metric. Such evaluation of log likelihood-maximisation with penalty criteria and error metrics assume that we are able to successfully obtain models' parameter estimates. This is further discussed in subsection 5.3.4.

5.3.3.2 Initial parameter estimates

Various approaches may be used to find initial parameters in the implementation of filtering algorithms. Least-square method (see, for example, Erlwein and Mamon [19]), likelihood maximisation if it can be accomplished (cf. Date and Ponomareva [11]) and use of first two sample moments are the most popular.

In our case, we first assume that the time-series data are stationary. Then under the one-state setting, we find the estimates $\widehat{\mu}^g$ and $\widehat{\sigma}^g$ using (i) the sample moments (sample mean and sample standard deviation) and (ii) maximum likelihood. For both sample moment-based and maximum-likelihood methods, we identify the maximum and minimum of the data set and assign the starting values as $\mu^g(1) = \frac{\max(y_k^g)-\widehat{\mu}^g}{2}$ and $\mu^g(2) = \frac{\min(y_k^g)+\widehat{\mu}^g}{2}$. This procedure may be generalised for the $N-$regime case by "spreading" starting values of $\mu^g(i)$ evenly around $\widehat{\mu}^g$. Initial values of $\sigma^g(i)$ are typically taken to be $\widehat{\sigma}^g$ in either regime and as the data are filtered, two volatility estimates emerge, which may or may not be different for each of the two regimes. The EM algorithm's rate of convergence is very fast so long as the starting parameters are close to the "true" parameter values. But as these "true" values are unknown to begin with, we rely on data processing's stability and speed with the aid of the filtering equations attaining convergence in each algorithm step.

We choose $1/N$ for the initial value of each entry in the transition probability matrix, and this yields a stable convergence. An alternative approach described in Hamilton [23] and Hardy [24] is useful for the two-state RSLN model. Such approach works very fast because it is computationally easy to maximise a smooth function (exponential

Method	$\mu(1)$	$\mu(2)$	$\sigma(1)$	$\sigma(2)$	a_{12}	a_{21}
Sample	8.0×10^{-4}	-7.0×10^{-4}	3.0×10^{-4}	3.0×10^{-4}	5.0×10^{-1}	5.0×10^{-1}
Likelihood	2.0×10^{-6}	7.0×10^{-7}	3.8×10^{-4}	1.8×10^{-4}	3.6×10^{-2}	5.1×10^{-2}

Table 5.1: Initial values for all filtering algorithms (JPY/GBP data)

Method	$\mu(1)$	$\mu(2)$	$\sigma(1)$	$\sigma(2)$	a_{12}	a_{21}
Sample	6.7×10^{-4}	-4.0×10^{-4}	2.2×10^{-4}	2.2×10^{-4}	5.0×10^{-1}	5.0×10^{-1}
Likelihood	2.0×10^{-6}	-3.2×10^{-6}	3.8×10^{-4}	2.0×10^{-4}	9.0×10^{-2}	1.1×10^{-2}

Table 5.2: Initial values for all filtering algorithms (JPY/USD data)

function in this case) over the six parameters. It is worth noting that whilst the results of sample moment-based and likelihood maximisation approaches in obtaining initial values (Tables 5.1 and 5.2) are different, they still make the filtering algorithms converge and are able to replicate identical dynamics of a_{ij}, and $\mu(i)$ and $\sigma(i)$ for each pair of currencies.

By construction, the dynamic filters, with a single underlying unobserved HMM, take into account the behaviour of two FX rates jointly. Hence, the starting values for a_{12} and a_{21} were taken as the average of the corresponding values in Tables 5.1 and 5.2 in each method. Although the likelihood maximisation (ML) is an ideal method to initialise parameter values given its statistical formality and grounding, its implementation in practice may pose one insurmountable challenge. The maximum likelihood value found using numerical methods may not necessarily be a global maximum, and proving that is not clear-cut if not impossible. Such a problem akin to ML estimation becomes more pronounced under HMM with four or more states where computational time increases with dimensionality.

We further remark that instead of employing the maximal and minimal data values in the above assignment of initial value estimation, appropriate quantiles could be considered instead. This is because outliers can substantially affect this procedure. In spite of the fact that the use of sample moments to select initial estimates is rather heuristics, they enable the filtering algorithms to yield parameter estimates that have identical evolution to those produced by the ML method in the 2- and 3-state settings.

5.3.3.3 Filters, data processing and estimation

In the implementation of the ILN and RSLN-N models, we employ the estimation method proposed in Hardy [24]. This method, however, was formulated not taking into account the joint dynamics of two FX rates. Thus, parameter estimates are obtained by applying the method separately to each data set. This implies a simplifying assumption that each FX rate data series is independent of each other. The same approach and assumption are taken in the implementation of the GARCH model, where we utilise the standard-fitting and forecasting procedures from Matlab's Econometrics Toolbox.

The dynamic filtering algorithms put forward in this research on the other hand were designed to work on two-dimensional data of FX rates collected at same time points. Dynamic filters process a moving window of several univariate or multivariate data points. This gives rise to certain filtered quantities related to the Markov chain that subsequently produce a sequence of model parameter estimates using Proposition 2. The processing of a window of data points is termed as one complete pass or algorithm step. The initial values are used to generate parameter estimates in the first algorithm step, which in turn serve as the initial values to obtain the parameter estimates after the second pass through the data, and this process continues until the last algorithm step. The size of the moving window is chosen based on some criterion, which is the maximisation of the log-likelihood function in our case.

Our findings show that the numerical values of log-likelihood function are not monotonic with respect to the window size, and the models with the best fit (in the context of penalty-based information criterion) do not always produce the best out-of-sample forecasts. This issue is linked to an associated instability of the log-likelihood maximisation due to various factors, most notably the numerical error from the division of quantities close to zero. Whilst scaling (e.g., working in terms of percentage returns, log, etc) may help in a few occasions, we note that this issue has no simple quick fix because it is a data-dependent problem. The selection of the appropriate window size in our case is made on the basis of (i) maximising log-likelihood values and (ii) reproducing the nonlinear evolution of the probabilities $\widehat{\mathbf{p}}$; refer to equation (5.5). We display the plots of $\widehat{\mathbf{p}}$ in Figures 5.5 and 5.6 through the algorithm steps. The behaviour of $\widehat{\mathbf{p}}$ under the zero-delay HMM (Figure 5.6) looks very similar to the behaviour of $\widehat{\mathbf{p}}$ under the one-step delay model (Figure 5.5).

From the above window-size selection guidelines, our numerical experiments demonstrate that the most suitable moving window must contain 7 bivariate data points, which correspond to a 35-minute interval. A wider moving window smooths out the prediction process. That is, with more data points in the moving window, the graphs of $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\mathbf{p}}$ would look like straight lines, which are contrary to the essence of nonlinear dynamics of parameters. Moreover, a large sample window decreases the quality of the out-of-sample forecasts as FX rate dynamics are not also fully captured. On the other hand, smaller window size magnifies the effect of volatility, which markedly dominates the drift component. Consequently, this results to large numerical errors making the predictions meaningless.

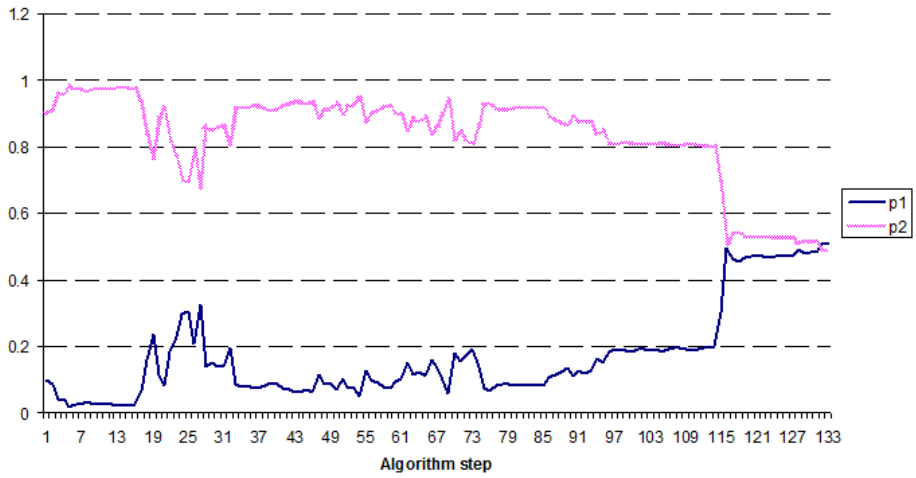


Figure 5.5: Evolution of $\widehat{\mathbf{p}}$ under the one-step delay HMM

5.3.4 Comparison of numerical results

In Tables 5.3 and 5.4, we present various statistical measures to gauge the goodness of fit of the proposed model in comparison with other popular existing models. We evaluate the Akaike information [1] and Bayesian information [35] criteria, abbreviated as AIC and BIC, respectively, that highlights the maximisation of log likelihood and penalty for model complexity. The model with the highest AIC and BIC is most preferred. The AIC and BIC metrics are given by

$$AIC = \ln L - b$$

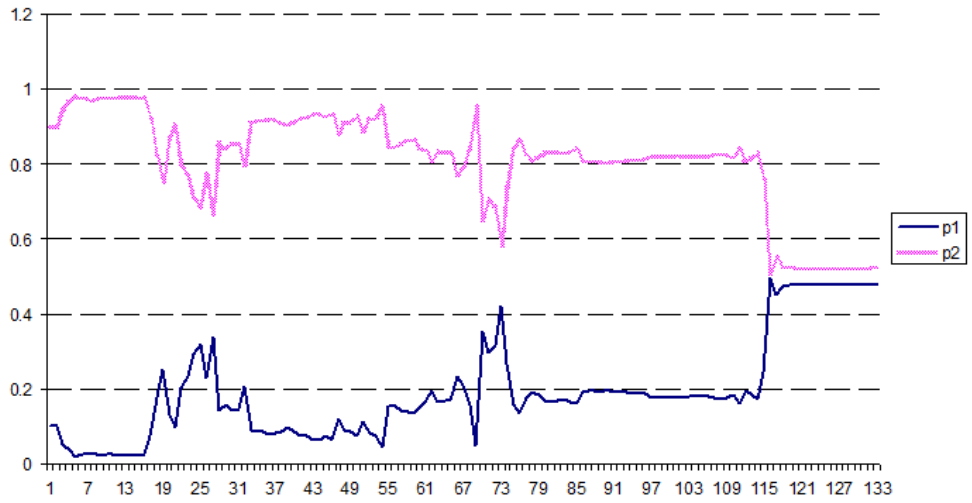


Figure 5.6: Evolution of $\widehat{\mathbf{p}}$ under the zero-delay HMM

$$BIC = \ln L - 0.5k \ln m,$$

where L is a likelihood or log-likelihood function, m is the number of observations and b is the number of parameters in a model.

Furthermore, we also evaluate the fitting performance of the models by analyzing the errors when prediction values are compared to actual values. The prediction is an all out-of-sample forecasting because only data in the past are continually processed to generate forecasts as move through time. With FX rates treated as a 2-dimensional observation process, the one-step ahead forecasts using equation (5.1) is

$$E[y_{k+1}^g | \mathcal{F}_k^g] = y_k^g \sum_{j=1}^{N} \langle \widehat{\boldsymbol{x}}_k, \boldsymbol{e}_j \rangle \exp\left(\widehat{\mu}^g(j) + \frac{\widehat{\sigma}^g(j)^2}{2}\right), \tag{5.17}$$

where $\widehat{\mathbf{x}}_k$ is the estimate for the unconditional distribution of the Markov chain and $\mathcal{F}_k^g$ is the filtration generated by y_k^g.

One way to examine the model's goodness of fit is to use the root-mean-square-error (RMSE) metric given by

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{m}(s_k^g - \widehat{s}_k^g)^2}{m}},$$

Model	Log-likelihood	BIC	AIC	RMSE	Param.
RW	n/a	n/a	n/a	2.6510×10^{-4}	n/a
ILN	6603.99	6597.08	6601.99	n/a	2
RSLN-2	6680.68	6659.96	6674.68	n/a	6
RSLN-3	6702.21	6660.76	6690.21	n/a	12
DILN	6781.19	6774.28	6779.19	8.0811×10^{-6}	2
DRSLN-2	6877.03	**6856.30**	6871.03	6.5976×10^{-6}	6
DRSLN-3	**6890.12**	6848.67	**6878.12**	6.5897×10^{-6}	12
ZDILN	6804.96	6798.05	6802.96	8.2164×10^{-6}	2
ZDRSLN-2	6862.17	**6841.44**	6856.17	$\mathbf{6.5533 \times 10^{-6}}$	6
ZDRSLN-3	**6878.29**	6836.84	**6866.29**	6.5832×10^{-6}	12
GARCH(1,1)	6667.86	6654.05	6663.86	2.4720×10^{-4}	4

Table 5.3: Results of likelihood- and error-based fitting measures covering JPY/GBP data collected between 09:35, 06 July 2012 and 18:40, 11 July 2012

where s_k^g, $\widehat{s}_k^g$ and m are the actual value, predicted value and length of the data series, respectively. The model producing the lowest RMSE is most preferred. For modelling multivariate data, the corresponding appropriate RMSE metric is given, for example, in Date et al. [9].

Remark 3: *We do not consider the RMSE for multivariate modelling here because the implementation of ILN and GARCH models was performed using univariate-based algorithms. Thus, to make our model comparison valid, the univariate version of the RMSE is employed on each time series of FX rates.*

Under both the AIC and BIC, our empirical results reveal that for the JPY/GBP data, the multi-regime models with dynamic filters (one-step and zero-delay models) significantly outperform other competing models. From Table 5.3, the DRSLN-N and ZDRSLN-N models fit the data very well judging from the chosen criteria. The ZDRSLN-N models are marginally better than the DRSLN-N models as the former have slightly smaller RMSEs than the latter. In the case of the JPY/USD data, the RW model outperforms all other models. This is because the log increments of the JPY/USD spot-rate process are so small that numerical errors muddled the calculation of the RMSE. But, under the log likelihood-driven criteria (AIC and BIC), it is clear

Model	Log likelihood	BIC	AIC	RMSE	Param.
RW	n/a	n/a	n/a	$\mathbf{2.883 \times 10^{-6}}$	n/a
ILN	6999.81	6992.90	6997.81	n/a	2
RSLN-2	7065.49	7044.76	7059.48	n/a	6
RSLN-3	7078.31	7036.87	7066.31	n/a	12
DILN	7102.50	7095.60	7100.50	4.0598×10^{-6}	2
DRSLN-2	7230.05	**7209.32**	**7224.05**	4.0021×10^{-6}	6
DRSLN-3	7235.12	7193.67	7223.12	7.9566×10^{-6}	12
ZDILN	7200.95	7193.62	7198.53	4.0678×10^{-6}	2
ZDRSLN-2	**7239.74**	**7218.35**	**7233.07**	$\mathbf{3.1228 \times 10^{-6}}$	6
ZDRSLN-3	**7244.13**	7202.69	7232.13	6.5533×10^{-6}	12
GARCH(1,1)	7027.22	7013.41	7023.22	2.1801×10^{-4}	4

Table 5.4: Results of loglikelihood- and error-based fitting measures covering JPY/USD data collected between 09:35, 06 July 2012 and 18:40, 11 July 2012

that the ZDRSLN-2 won this comparison race of FX-rate models. The ZDRSLN-2 also ranked second in terms of the RMSE criterion. Thus, we still place a great deal of confidence in the performance of the ZDRSLN-2 model. We note that the likelihood statistics for the random walk are not available. The main advantage of choosing a ZDRSLN-2 model over a random walk model is flexibility and adjustability. It is a huge advantage to have a model that can be adjusted for the data dynamics; a random walk model cannot offer such flexibility.

For both data sets, the fit of the ILN and GARCH(1,1) models is poor relative to those of the other models across all criteria. We observe significant improvement in the log-likelihood values of the dynamic models over those of the static RSLN.

It appears logical that within the context of a multi-regime modelling comparison, the Markov-switching (MS) GARCH model [20, 27] would be the more appropriate choice over the usual GARCH model. However, the calibration procedure for the MS-GARCH model involves either the Markov chain Monte Carlo (MCMC) algorithms or maximisation of a quasi-likelihood function. Therefore, the calibration procedures are not readily accessible to practitioners; in addition, such estimation algorithms are not fast enough to be considered for frequent-trading models. Trader requires rolling over their

positions quickly and would tend to rely on the simpler version of the GARCH model than endeavour in dealing with the implementation issues of the MS GARCH model.

The curse of dimensionality is another significant issue. For example, there are already twelve parameters to estimate in the two regime MS-GARCH(1,1) model. But, we note that a simple GARCH model is the standard for hedge fund managers working with volatility in discrete time. We do not provide a detailed analysis of how parameters in the simple GARCH (1,1) are chosen for the purpose of benchmarking. The GARCH (1,1)'s AIC and BIC statistics were calculated using the standard Matlab Econometrics Toolbox.

5.3.5 CHull criterion

In any modelling endeavour, the main goal is to select a model that optimally balances model's goodness of fit/misfit and complexity/simplicity. To reinforce the AIC, BIC and RMSE criteria in our analyses above, we consider the model-fitting criterion called CHull (based on the concept of convex hull) in the context of mixtures of factor analyses developed by Ceulmans and Kiers [6]. We tailor this criterion to our regime-switching methodology aided by the guidelines provided in Buteel et al. [4]. A generic but succinct synopsis of CHull is described in Wilderjans et al. [39] comprising of two main steps: (i) determining the convex hull of the fit-measure-versus-complexity-measure plot of the models under comparison and (ii) identifying the model on the boundary of the convex hull such that increasing complexity (i.e., adding more parameters) has only a small effect on the fit measure, whereas lowering complexity (e.g., having less parameters) changes the goodness of fit/misfit significantly.

In our CHull implementation, we keep track and compare the changes in the log-likelihood (simple fitting measure) when the dimensions of the models (measure of complexity) are increased or decreased. In particular, the change of the log-likelihood of the models on the convex hull is given by

$$\left(\frac{\log L_n - \log L_{n-1}}{f_n - f_{n-1}}\right) / \left(\frac{\log L_{n+1} - \log L_n}{f_{n+1} - f_n}\right), \tag{5.18}$$

where L_n is the log-likelihood value of the nth model, with models being ordered according to the number of free parameters. Here, f_n stands for the number of free parameters in the nth model. In our case, $n = 2,\ 4,\ 6$ and 12. In summary, the above procedure

can be described as simple as plotting the values of the log-likelihood for all models versus the number of free parameters and choosing the points where the log-likelihood values are the biggest and building a convex hull using these chosen points. We pick the model that produces the greatest statistic value in equation (5.18).

The points corresponding to the models lying on the convex hull are connected by a line. From Figures 5.7 and 5.8, the models chosen by the BIC is on the upper boundary of the convex hull. As seen in Figures 5.7 and 5.8, the models whose log-likelihood values lie on the convex hull are the same models which are selected in accordance with the BIC criterion. This provides extra support on the appropriateness of the dynamic regime-switching models (especially the zero-delay model) in capturing the trend of the bivariate FX data.

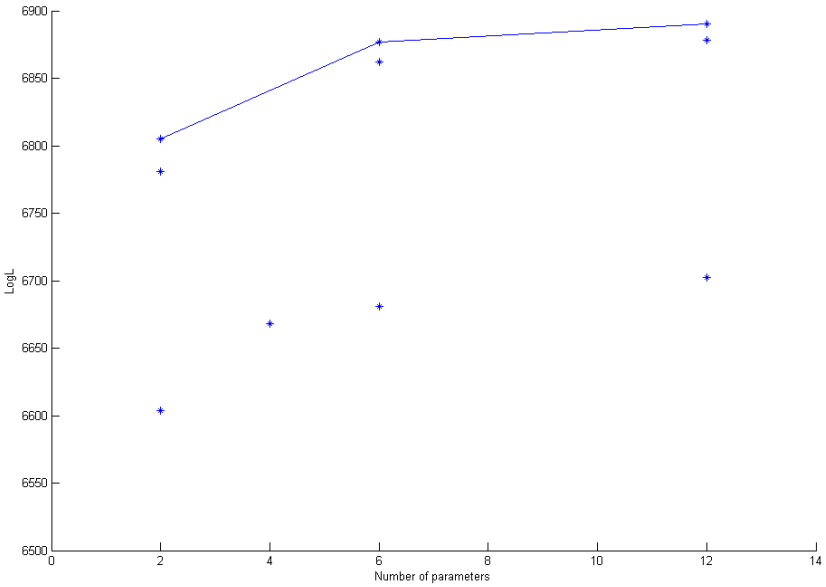


Figure 5.7: Chull for the JPY/GBP data

5.3.6 Parameter estimation results and further model validation

5.3.6.1 Dynamics of parameter estimates

Figures 5.9 - 5.13 depict the estimation outputs of the filtering experiment for the zero-delay 2-state model. The plots for the outputs of the one-step delay 2-state model look

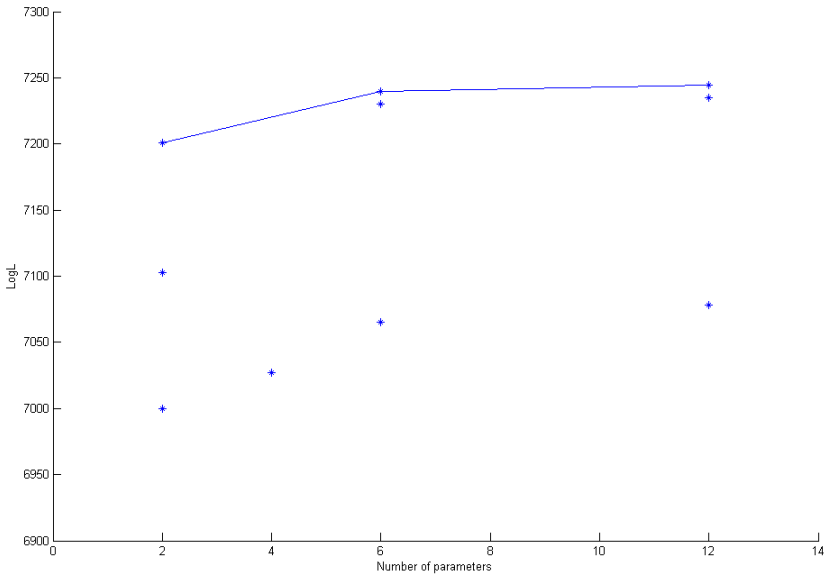


Figure 5.8: Chull for the JPY/USD data

and behave in a similar manner, and are therefore omitted. The estimated drift and volatility dynamics clearly show the multi-regime pattern of the FX-rate process. This empirically illustrates the unpredictability of the parameters' behaviour in the models under consideration. Such unpredictability is consistent with Nailiu and King's findings [32] regarding extreme difficulty in estimating parameters and modelling the FX-rate process in general.

5.3.6.2 Validating the white-noise assumption

The multivariate extension of a regular Q-Q plot was comprehensively tackled in Liang and Ng [29] and the result in our case is shown in Figure 5.14. Since the data has several dimensions, it is important to understand that producing two different Q-Q plots will not capture properly the noise structure of the data. Visually, the output of the method is interpreted as a simple one dimensional Q-Q plot; the points have to be as close as possible to the straight line. The graph supports the normality assumption of the residuals whilst the general dynamics of the process is explained by the model. We argue that for extremely noisy data the line in Figure 5.14 is very straight and the normality hypothesis is validated for the particular data set.

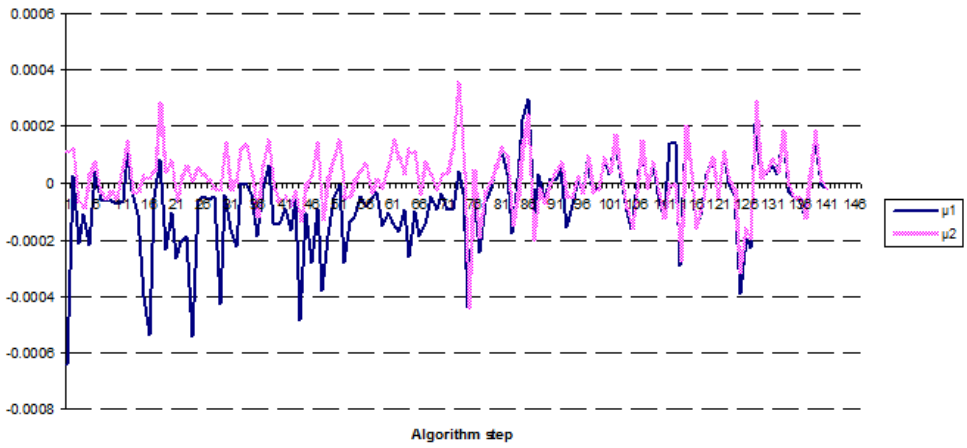


Figure 5.9: Evolution of $\widehat{\mu}_Z(i)$ for the JPY/GBP data under the 2-state HMM

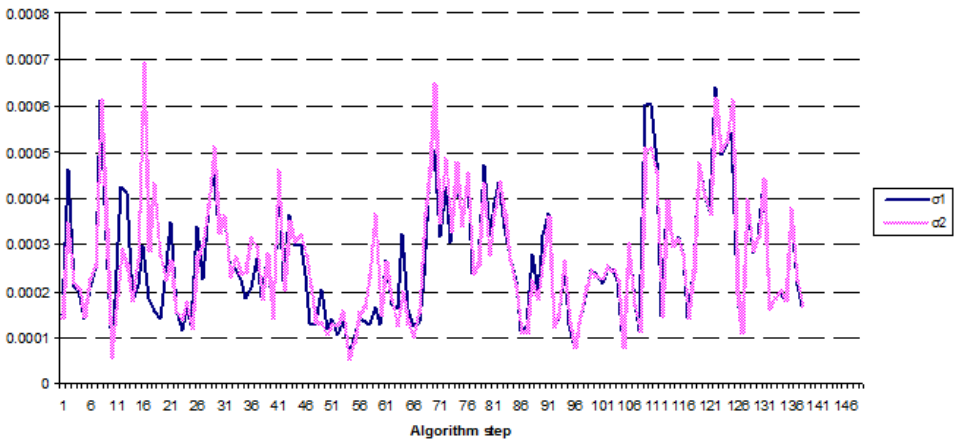


Figure 5.10: Evolution of $\widehat{\sigma}_Z(i)$ for the JPY/GBP data under the 2-state HMM

5.3.7 Frequent trading and the ZDRSLN-N modelling set-up

Following the discussion in subsection 5.3.4, we focus our comparison on several dynamic regime-switching models if no delay in the reaction time of y_k to $\boldsymbol{x}_k$ in frequent trading is a pertinent assumption. We implemented the ZDRSLN-N and DRSLN-N filters using the most popular computing platforms in the industry. These include the Microsoft Visual C++ and Visual Basic for Applications (VBA) in Microsoft Excel. The codes written in Excel are more instructive than those in C++, and without a doubt they are more useful for practitioners keen on implementation. However, it takes

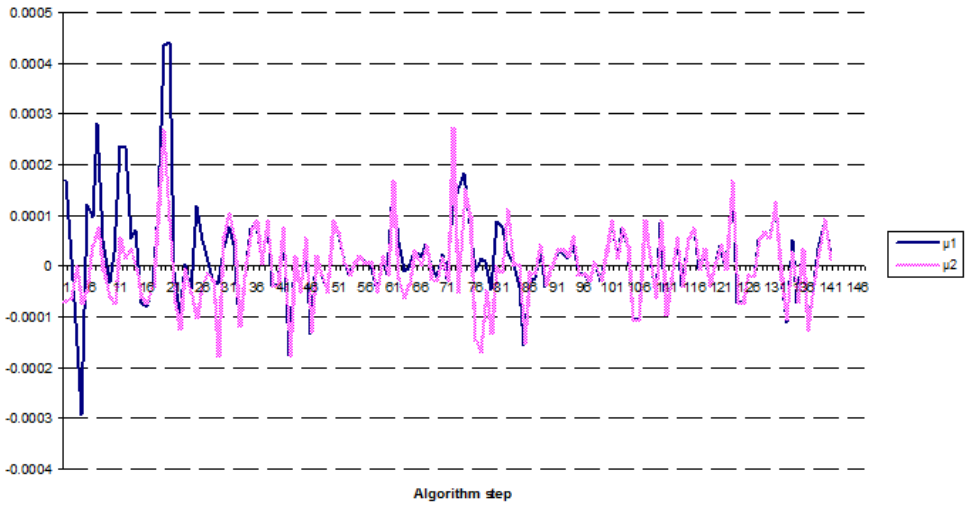


Figure 5.11: Evolution of $\widehat{\mu}_Z(i)$ for the JPY/USD data under the 2-state HMM

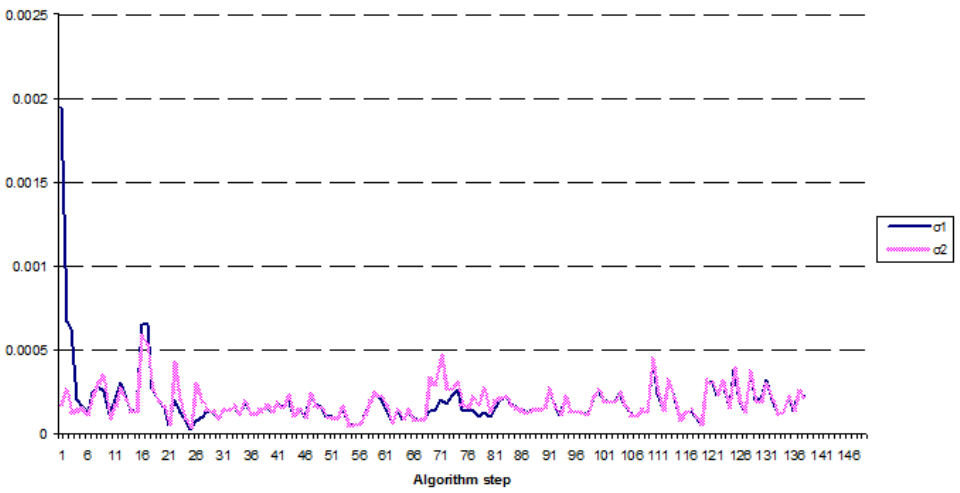


Figure 5.12: Evolution of $\widehat{\sigma}_Z(i)$ for the JPY/USD data under the 2-state HMM

a very long time to generate results using Excel. It would only take 8.29 seconds to run the filtering together with the parameter estimation codes and produce the needed numerical outputs (e.g., plots of parameters' evolution) in C++ whilst it would take 7 minutes to perform the same tasks in VBA.

In the financial industry the design of many algorithms is compatible with Excel work-

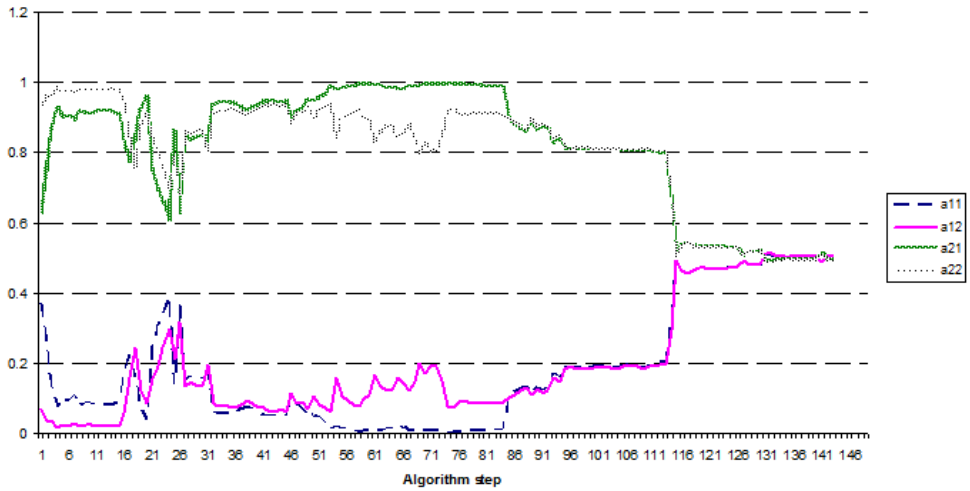


Figure 5.13: Evolution of transition probabilities under the zero-delay 2-state HMM

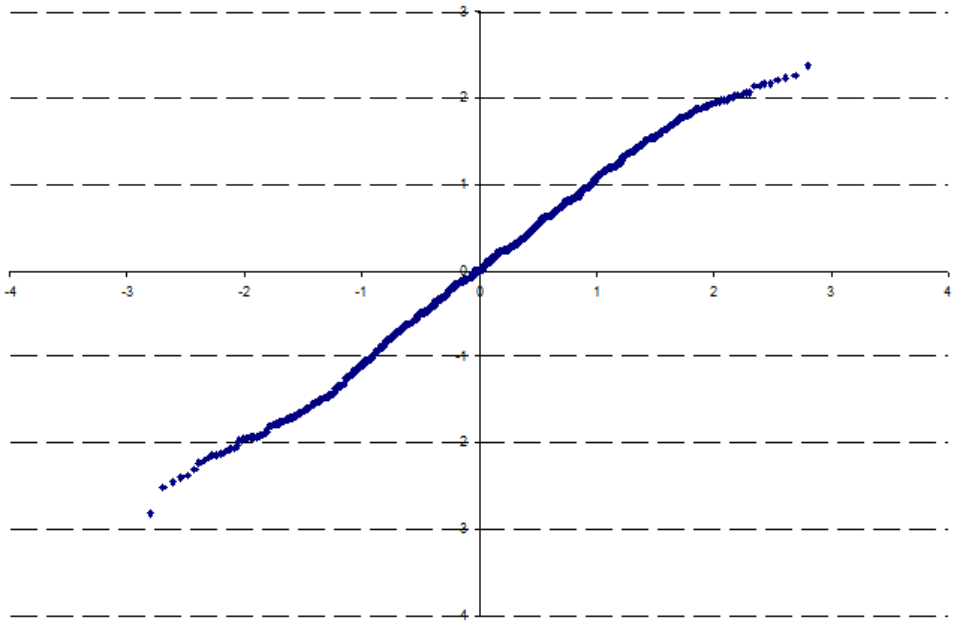


Figure 5.14: Q-Q plot for the bivariate FX rate data

books. We therefore implemented a C++ code to interface with Excel sheets. This endeavour imposes additional limitation on the performance of the code as Excel is used not only as a database, but also as a graph-plotting engine. If C++ is run for the purpose of calculating the parameter values alone, the computing time involved is only

2 seconds. This shows that C++ implementation is convincingly more efficient than VBA.

Given the appropriateness and seeming dependability of the zero-delay models in outperforming other models under a frequent trading environment, we now concentrate on the analysis of our complete JPY/GBP and JPY/USD FX bivariate data set. That is, all data points that were recorded, some of which occurred in time intervals of less than 5 minutes, are considered. Thus, we wish to examine time series data with a much higher frequency than the ones studied in the previous subsections. The period covered is the same as that in the previous exercise, i.e., from 09:35, 06 July 2012 to 18:40, 11 July 2012.

Even with higher frequency, the trades on our two currency pairs are still at irregular intervals and do not occur simultaneously. Moreover, they do not necessarily happen every second or even within a minute. Sometimes, there were several trades in a one-minute interval, at other times, there was none at all. So, from the original data sets, we construct two new "derived" FX data sets where data points are "observed" at synchronised time points (but finer intervals of uniform lengths) to be able to apply our filtering algorithms in conjunction with our zero-delay models.

The new-data-generation procedure combines simple heuristics of data imputation and smoothing in the following way. The time series data is binned into two-minute intervals. With the two-minute interval binning, we ended up with 35 and 63 empty bins for JPY/USD and JPY/GBP currency pairs, respectively. The bin length is justified on the basis of minimising the number of empty bins whilst attempting to retain the effect of the model's drift.

The data values are averaged in each two-minute bin interval and this average represents the data point of the modified new data set assigned at the beginning of the bin interval. If there were no trades in a given time interval, the previous average value is used, i.e., it becomes a new trading value for the next empty bin.

We concentrate on the performance analysis of the adjudged best two models from the previous subsection given the newly obtained 2-minute time-series bivariate data

set. We include the RW model in the analysis. The resulting goodness-of-fit statistics from the application of both the dynamic one-step and zero-step delay filtering algorithms on the new data set are depicted in Tables 5.5 and 5.6. The results ostensibly favours the ZDRSLN-2 model in terms of BIC and AIC. The RMSEs show that the ZDRSLN-2 model is able to outperform significantly the RW model with approximately 43% decrease in RMSE for the JPY/USD data and approximately 18% decrease in RMSE for the JPY/GBP data.

Model	Log likelihood	BIC	AIC	RMSE	Parameters
RW	n/a	n/a	n/a	3.692×10^{-4}	n/a
DRSLN-2	6671.307	6650.584	6665.307	8.472×10^{-4}	6
ZDRSLN-2	**6728.407**	**6707.684**	**6722.407**	$\mathbf{3.031 \times 10^{-4}}$	6

Table 5.5: Loglikelihood- and error-based goodness-of-fit measures for the new JPY/GBP data set with 2-minute frequency covering the period 09:35, 06 July 2012 - 18:40, 11 July 2012

Model	Log-likelihood	BIC	AIC	RMSE	Parameters
RW	n/a	n/a	n/a	1.010×10^{-5}	n/a
DRSLN-2	7115.221	71022.19	7109.221	1.323×10^{-5}	6
ZDRSLN-2	**7204.691**	**7191.66**	**7198.691**	$\mathbf{5.780 \times 10^{-6}}$	6

Table 5.6: Loglikelihood- and error-based goodness-of-fit measures for the new JPY/USD data set with 2-minute-frequency covering the period 09:35, 06 July 2012 - 18:40, 11 July 2012

The 2-minute interval was chosen heuristically and it is completely data-dependent. The new data was then obtained from the original data set, and we correspondingly perform our analysis.

5.4 Conclusion

In this chapter, we put forward an alternative approach to jointly model the behaviour of high-frequency multivariate FX rates using an HMM. New filtering recursive equations are derived assuming a zero-delay modelling paradigm. These recursions yield a self-calibrating multi-dimensional model that practitioners may employ for various financial modelling endeavours. To demonstrate the applicability and performance of our proposed model together with its parameter estimation methodology, we consider its implementation on the JPY/USD and JPY/GBP FX rates. A comparative analysis was conducted examining various HMM competing models with increasing complexity in terms of regime dimension and lag order as well as commonly used models for simple benchmarking. Our selection criteria in assessing model performance are in adherence to balancing the goodness of fit via log-likelihood maximisation and model complexity penalty (AIC, BIC, Chull). This was complemented by evaluating the model's RMSE and choosing the model with the least forecasting error.

Whilst we tend to emphasise model development for data under the HFT's framework, this investigation can be viewed as bridging the gap between the usual modelling of low-frequency data (daily, weekly monthly or quarterly) and the modelling of the currently emerging data sets resulting from rapid trading (minutes or seconds). More specifically, some of the challenges unique to high-frequency data were elaborated, and certain ways to formally or heuristically rectify them were given. Common issues in the implementation of our proposed approach were detailed and addressed.

The empirical evidence of our extensive study using various filtering estimation algorithms in conjunction with model validation diagnostics shows that the proposed dynamic models, in particular the ZDRSLN-N, outperform other competing models (GARCH, RSLN, DRSLN). These zero-delay models could be potentially beneficial in forecasting FX rates to aid practitioners in setting their trading positions. Certainly, further examinations need to be pursued such as performance analysis of the ZDRSLN-N model with respect to a few benchmark models on several other FX rates at different time periods. Armed with a better model in terms of fitting high-frequency data, a natural direction of this study would be the pricing of FX derivatives as well as testing the accuracy of risk measures for portfolios with exposures to FX-rate movements.

Two extensions could be pursued to probably improve further the model performance and make the filtering algorithms more flexible. (i) Smoothers could be incorporated that will allow parameters to change quickly without having the need to filter a quite large subset of data. This procedure will decrease the RMSEs and simultaneously shorten the time of parameter estimation. (ii) A correlation structure between the white noise drivers of the data could be introduced. This would lead to a collection of enhanced filtering algorithms but most likely, a set of new challenges in the numerical implementation have to be tackled.

Finally, an alternative approach could be explored by considering instead the modelling of the price dynamics of currency or FX futures. From the estimates of FX futures prices, forecasts of FX spot rates and other parameters of interest can be recovered. This methodology is similar to the idea worked out in Date et al. [10] in capturing the evolution of arbitrage-free futures prices on commodities. The methodology of modelling FX futures prices directly is ideal but only if backing out FX rates for one particular currency pair. Nonetheless, when a joint evolution of FX rates is needed, the approach in this research prochect is still deemed more appropriate and relevant.

5.5 References

[1] Akaike, H., 1974. A new look at the statistical model identification, IEEE Transactions on Automatic Control 19(6), 716 – 723. 141

[2] Aldridge, I., 2010. High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading, John Wiley & Sons, Inc., Hoboken, New Jersey. 124

[3] Bacharoglou, A., 2010. Approximation of probability distributions by convex mixtures of Gaussian measures, Proceedings of the AMS, 138 (7), 2619–2628.

[4] Bulteel, K., Wilderjans, T., Tuerlinckx, F., Ceulemans, E., 2013. CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers, Behavior Research Methods 45(3), 782–791. 145

[5] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307 –327. 136

[6] Ceulemans, E., Kiers, H., 2006. Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method, British Journal of Mathematical and Statistical Psychology 59, 133 – 150. 145

[7] Cheung, Y., Erlandsson, R., 2005. Exchange rates and Markov switching dynamics, Journal of Business and Economic Statistics 23(3), 314–320. 124

[8] Coppes, R.C., 1995. Are exchange rate changes normally distributed?, Economic Letters 47(2), 117–121.

[9] Date, P., Jalen, L., Mamon, R., 2008. A new algorithm for latent state estimation in nonlinear time series models, Applied Mathematics and Computation 203(1), 224–232. 143

[10] Date, P., Mamon, R., Tenyakov, A., 2013. Filtering and forecasting commodity futures prices under an HMM framework, Energy Economics 40, 1001–1013. 154

[11] Date, P., Ponomareva, K., 2011. Linear and nonlinear filtering in mathematical finance: A review, IMA Journal of Management Mathematics 22, 195–211. 138

[12] Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society-Series B (Methodological) 39(1), 1–38. 130

[13] Elliott, R., Aggoun, L., Moore, J., 1995. Hidden Markov models: Estimation and control, Springer, New York. 124, 127, 130

[14] Elliott, R., Hunter, W., Jamieson, B., 2001. Financial signal processing: A self-calibrating model, International Journal of Theoretical and Applied Finance 4, 567– 584.

[15] Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, Econometrica 50, 987–1008. 124

[16] Engel, C., Hamilton, J., 1990. Long swings in the dollar: Are they in the data and do markets know it?, American Economic Review 80, 689–713. 124

[17] Erlwein, C., Mamon, R., Davison, M., 2011. An examination of HMM-based investment strategies for asset allocation, Applied Stochastic Models in Business and Industry 27, 204–221. 124, 130, 136

[18] Erlwein, C., Benth, F., Mamon, R., 2010. HMM filtering and parameter estimation of electricity spot price model, Energy Economics 32, 1034–1043.

[19] Erlwein, C., Mamon, R., 2009. An on-line estimation scheme for a Hull-White model with HMM-driven parameters, Statistical Methods and Applications 18(1), 87–107 138

[20] Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime - switching process, Journal of Financial Economics 42, 27–62. 144

[21] Jung, C., 1995. Forecasting of foreign exchange rate by normal mixture models, Journal of Economic Studies 22(1), 45–57 127

[22] Haldane, A., 2011. Race to zero. Speech at the International Economic Association Sixteenth World Congress, Beijing, 08 July 2011. URL: http://www.bankofengland.co.uk/publications/Documents/speeches/2011/speech509.pdf 124

[23] Hamilton, J., 1994. Time Series Analysis, Princeton University Press, New Jersey. 136, 138

[24] Hardy, M., 2002. A regime-switching model of long-term stock returns, North American Actuarial Journal 6(1), 171–173. 138, 140

[25] Kaehler, J., Marnet, V., 1994. Markov-switching models for exchange rate dynamics and the pricing of foreign currency options, Econometrics Analysis of Financial Markets Studies in Economics, 203–230. 127

[26] Kamaruzzaman, Z., A., Isa, Z., Ismail, M., T., 2012. Mixtures of normal distributions: Application to Bursa Malaysia stock market indices, World Applied Sciences Journal 16 (6), 781–790. 127

[27] Klaassen, F., 2002. Improving GARCH volatility forecasts with regimeswitching GARCH. Empirical Economics 27, 363–394. 144

[28] Klink, G., Mathur, M., Kidambi, R., Sen, K., 2013. The contribution of the automotive industry to technology and value creation. A.T. Kearney, Inc. 131

[29] Liang, J., Ng, K., 2009. A multivariate normal plot to detect nonnormality, Journal of Computational and Graphical Statistics 18(1), 52–72. 147

[30] Maheu, J., McCurdy, T., 2007. Modeling foreign exchange rates with jumps, Working Paper, Department of Economics, University of Toronto. http://www.economics.utoronto.ca/public/workingPapers/tecipa-279-1.pdf. 124

[31] Meese, R., Rogoff, K., 1982. The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification?, International Finance Discussion Papers, Board of Governors of the Federal Reserve System, 204. http://www.federalreserve.gov/pubs/ifdp/1982/204/ifdp204.pdf . 123

[32] Nailliu, J., King, M., 2005. What drives movements in exchange rates?, Bank of Canada Review, Autumn, 27. http://faculty.haas.berkeley.edu/lyons/ Bailliu_King_what%20drives%20movements.pdf . 123, 147

[33] Rodionov, S., 2004. A sequential algorithm for testing climate regime shifts, Geophysical Research Letters, 31:L09204. doi:10.1029/2004GL019448. 125, 132, 135

[34] Rodionov, S., 2005. A brief overview of the regime shift detection methods, In: Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability, V. Velikova and N. Chipev (Eds.), UNESCO-ROSTE/BAS Workshop on Regime Shifts, 14–16 June, Varna, Bulgaria, 17–24. URL: http://www.beringclimate.noaa.gov/regimes/ 132, 133

[35] Schwarz, G., 1978. Estimating the dimension of a model, Annals of Statistics 6(2), 461–464. 141

[36] Shah, N., 2011. High-frequency trading's new frontier: Currency derivatives, Wall Street Journal, Markets Section, 18 October 2011. URL: http://online.wsj.com/article/SB10001424052970204479504576639023900658918.html . 122

[37] Taylor, S., 1986. Modeling Financial Time Series, Wiley, Chichester. 124

[38] The World Bank Goup: GDP (current US$). 131
URL: http://data.worldbank.org/indicator/NY.GDP.MKTP.CD

[39] Wilderjans, T., Ceulemans, E., Meers, K., 2013. CHull: A generic convex-hull-based model selection method, Behaviour Research Methods 45(1), 1–15. 145

[40] Wirjanto, T., S., Xu, D., 2009. The applications of mixture of normal distributions in finance: A selected survey, working paper, University of Waterloo, 2009. http://economics.uwaterloo.ca/documents/mn-review-paper-CES.pdf . 127

[41] Yuan, C., 2011. Forecasting exchange rates: The multi-state Markov-switching model with smoothing, International Review of Economics and Finance 20(2), 342–362.

5.6 Appendix

Proof of recursive filters in Proposition 1

We provide the derivation of the filters given in Proposition 1 of section 5.2.

Filter for the state of $\boldsymbol{x}_k$

$$\begin{aligned}\gamma(\boldsymbol{x}_{k+1}) &= \widetilde{E}[\Lambda_{k+1}\boldsymbol{x}_{k+1}|\mathcal{F}^y_{k+1}] = \sum_{j=1}^{N}\widetilde{E}\left[\Pi\boldsymbol{x}_k\Lambda_k\left\langle\boldsymbol{x}_{k+1},\boldsymbol{e}_j\right\rangle|\,\mathcal{F}^y_{k+1}\right]\Gamma^j \qquad (5.19)\\ &= \sum_{j=1}^{N}\widetilde{E}\left[\Pi\boldsymbol{x}_k\Lambda_k\left\langle\Pi\boldsymbol{x}_k+\boldsymbol{v}_{k+1},\boldsymbol{e}_j\right\rangle|\,\mathcal{F}^y_{k+1}\right]\Gamma^j\\ &= \sum_{j=1}^{N}\widetilde{E}\left[\Pi\boldsymbol{x}_k\Lambda_k\left\langle\Pi\boldsymbol{x}_k,\boldsymbol{e}_j\right\rangle|\,\mathcal{F}^y_{k}\right]\Gamma^j.\end{aligned}$$

By noting that $\sum_{j=1}^{N}\langle\boldsymbol{x}_k,\boldsymbol{e}_i\rangle=1$, it follows that

$$\gamma(\boldsymbol{x}_{k+1})=\sum_{i,j=1}^{N}\langle\gamma(\boldsymbol{x}_k),\boldsymbol{e}_i\rangle\,a_{ji}\Gamma^j\mathbf{a}_i.$$

Filter for the jump process $\mathcal{J}$

$$\begin{aligned}\gamma(\mathcal{J}^{ji}_{k+1}\boldsymbol{x}_{k+1}) &= \widetilde{E}[\Lambda_{k+1}\mathcal{J}^{ji}_{k+1}\boldsymbol{x}_{k+1}|\mathcal{F}^y_{k+1}] \qquad (5.20)\\ &= \sum_{l=1}^{N}\widetilde{E}\left[\Lambda_k\Pi\boldsymbol{x}_k\left\langle\boldsymbol{x}_{k+1},\boldsymbol{e}_l\right\rangle\left(\mathcal{J}^{ji}_k+\left\langle\boldsymbol{x}_k,\boldsymbol{e}_i\right\rangle\left\langle\boldsymbol{x}_{k+1},\boldsymbol{e}_j\right\rangle\right)|\mathcal{F}^y_{k+1}\right]\Gamma^l\\ &= \sum_{l=1}^{N}\widetilde{E}\left[\Lambda_k\Pi\boldsymbol{x}_k\left\langle\Pi\boldsymbol{x}_k+\boldsymbol{v}_{k+1},\boldsymbol{e}_l\right\rangle\left(\mathcal{J}^{ji}_k+\left\langle\boldsymbol{x}_k,\boldsymbol{e}_i\right\rangle\left\langle\boldsymbol{x}_{k+1},\boldsymbol{e}_j\right\rangle\right)|\mathcal{F}^y_{k+1}\right]\Gamma^l\\ &= \sum_{l=1}^{N}\widetilde{E}\left[\Lambda_k\Pi\boldsymbol{x}_k\left\langle\Pi\boldsymbol{x}_k,\boldsymbol{e}_l\right\rangle\left(\mathcal{J}^{ji}_k+\left\langle\boldsymbol{x}_k,\boldsymbol{e}_i\right\rangle\left\langle\boldsymbol{x}_{k+1},\boldsymbol{e}_j\right\rangle\right)|\mathcal{F}^y_{k+1}\right]\Gamma^l\\ &= \sum_{l=1}^{N}\widetilde{E}\left[\Pi\left\langle\Pi\boldsymbol{x}_k,\boldsymbol{e}_l\right\rangle\Lambda_k\boldsymbol{x}_k\mathcal{J}^{ji}_k|\mathcal{F}^y_k\right]\Gamma^l\\ &+ \widetilde{E}\left[\Lambda_k\left\langle\Pi\boldsymbol{x}_k,\boldsymbol{e}_j\right\rangle\left\langle\boldsymbol{x}_k,\boldsymbol{e}_i\right\rangle\Pi\boldsymbol{x}_k|\mathcal{Y}^y_k\right]\Gamma^j\\ &= \sum_{m,l=1}^{N}\left\langle\gamma(\mathcal{J}^{ji}_k\boldsymbol{x}_k),\boldsymbol{e}_m\right\rangle a_{lm}\Gamma^l\mathbf{a}_m+\widetilde{E}\left[\Pi\boldsymbol{x}_k\Lambda_k\left\langle\Pi\boldsymbol{x}_k,\boldsymbol{e}_j\right\rangle\left\langle\boldsymbol{x}_k,\boldsymbol{e}_i\right\rangle|\mathcal{F}^y_k\right]\Gamma^j\\ &= \sum_{m,l=1}^{N}\left\langle\gamma(\mathcal{J}^{ji}_k\boldsymbol{x}_k),\boldsymbol{e}_m\right\rangle a_{lm}\Gamma^l\mathbf{a}_m+\langle\gamma(\boldsymbol{x}_k),\boldsymbol{e}_i\rangle\,a_{ji}\Gamma^j\mathbf{a}_i.\end{aligned}$$

Filter for the auxiliary process $\mathcal{T}$

$$
\begin{aligned}
&\gamma(\mathcal{T}^i_{k+1}(f(y^g_{k+1}))\boldsymbol{x}_{k+1}) = \widetilde{E}[\Lambda_{k+1}\mathcal{T}^i_{k+1}(f(y^g_{k+1}))\boldsymbol{x}_{k+1}|\mathcal{F}^y_{k+1}] \qquad (5.21)\\
=\ & \sum_{l=1}^{N} \widetilde{E}\left[\Lambda_k \Pi \boldsymbol{x}_k \left\langle \boldsymbol{x}_{k+1}, \boldsymbol{e}_l \right\rangle \left(\mathcal{T}^i_k(y^g_k) + f(y^g_{k+1})(y_{k+1}) \left\langle \boldsymbol{x}_k, \boldsymbol{e}_i \right\rangle\right) |\mathcal{F}^y_{k+1}\right] \Gamma^l\\
=\ & \sum_{l=1}^{N} \widetilde{E}\left[\Lambda_k \Pi \boldsymbol{x}_k \left\langle \Pi\boldsymbol{x}_k + \boldsymbol{v}_{k+1}, \boldsymbol{e}_l \right\rangle \left(\mathcal{T}^i_k(f(y^g_k)) + f(y^g_{k+1}) \left\langle \boldsymbol{x}_k, \boldsymbol{e}_i \right\rangle\right) |\mathcal{F}^y_{k+1}\right] \Gamma^l\\
=\ & \sum_{l=1}^{N} \widetilde{E}\left[\Lambda_k \Pi \boldsymbol{x}_k \left\langle \Pi\boldsymbol{x}_k, \boldsymbol{e}_l \right\rangle \left(\mathcal{T}^i_k f(y^g_k) + f(y^g_{k+1}) \left\langle \boldsymbol{x}_k, \boldsymbol{e}_i \right\rangle\right) |\mathcal{F}^y_{k+1}\right] \Gamma^l\\
=\ & \sum_{l=1}^{N} \widetilde{E}\left[\Pi \left\langle \Pi\boldsymbol{x}_k, \boldsymbol{e}_l \right\rangle \Lambda_k \boldsymbol{x}_k \mathcal{T}^i_k(f(y^g_k))|\mathcal{F}^y_k\right] \Gamma^l\\
& + \sum_{l=1}^{N} \widetilde{E}\left[\Lambda_k \left\langle \Pi\boldsymbol{x}_k, \mathbf{e}_l \right\rangle \left\langle \boldsymbol{x}_k, \boldsymbol{e}_i \right\rangle \Pi\boldsymbol{x}_k|\mathcal{F}^y_k\right] f(y^g_{k+1})\Gamma^l\\
=\ & \sum_{m,l=1}^{N} \left\langle \gamma(\mathcal{T}^i_k(f(y^g_k))\boldsymbol{x}_k), \boldsymbol{e}_m \right\rangle a_{lm}\Gamma^l \mathbf{a}_m\\
& + \sum_{l=1}^{N} \widetilde{E}\left[\Lambda_k \left\langle \Pi\boldsymbol{x}_k, \boldsymbol{e}_j \right\rangle \left\langle \boldsymbol{x}_k, \boldsymbol{e}_i \right\rangle \Pi\boldsymbol{x}_k|\mathcal{Y}^y_k\right] f(y^g_{k+1})\Gamma^l\\
=\ & \sum_{m,l=1}^{N} \left\langle \gamma(\mathcal{T}^i_k(f(y^g_k))\boldsymbol{x}_k), \boldsymbol{e}_m \right\rangle a_{lm}\Gamma^l \mathbf{a}_m\\
& + f(y^g_{k+1}) \sum_{l=1}^{N} \left\langle \gamma(\boldsymbol{x})_k, \boldsymbol{e}_i \right\rangle a_{li}\Gamma^l \mathbf{a}_i.
\end{aligned}
$$

Filter for the auxiliary process $\mathcal{O}$

The proof follows from the derivation of the filter for the auxiliary process $\mathcal{T}$ by noting that when $f^g(y) \equiv 1$ in equation (5.9), we obtain equation (5.8).

■

6

An estimation algorithm for a Markov-switching model with any number of states

6.1 Introduction

The economy exhibits cyclical patterns alternating between stability and growth. Thus, the parameters of models describing the financial or economic process must also change over time. Such phenomenon in the economy can be effectively captured by embedding a hidden Markov chain to modulate model parameters. Consequently, this popularised the use of hidden Markov models (HMMs) to support various financial modelling objectives such as valuation, risk management and asset allocation. In the estimation of risk measures, HMM-based models are ideal especially for long-maturity insurance contracts such as segregated funds because they have the ability to adapt to dynamic changes in asset log returns over a long period of time.

Advanced mathematical and statistical machineries were developed to enhance and support the applications of HMMs to finance; see Date and Ponamareva [1] and Erlwein et al. [5], amongst others. Model calibration, which is obtaining reliable parameter estimates using observed values from the financial market, is a prime consideration and needs to be put in place. Otherwise there is no way to make the model implementation a success.

The pioneering works of Hamilton [6], Hardy [8] and Elliott et al. [2] offered comprehensive procedures in the estimation of parameters for Markov-switching models. Whilst these works described the underlying mathematical principles with augmented insightful discussions on a few implementation issues, details of implementation to specific data sets are left to the readers to hurdle such as the issues of initialisation and algorithm stability. Concentrating on statistic estimation, Hardy's work [8] is well-acknowledged in the actuarial community; it gives practically detailed procedure that can easily be implemented in Excel with VBA, which is the main platform currently used in the industry. However, to the best of our knowledge, the systematic procedure on finding appropriate parameters of Markov-switching models using data under three or more regimes is left unaddressed.

In the sequel below, we explain how to calibrate an HMM-driven financial model assuming the number of regimes is more than two. As in Hardy [8] and the general Black-Scholes framework that is still popular with practitioners, we assume that the underlying process follows a geometric Brownian motion with regime-switching parameters. From the programming point of view, our procedure is compatible with an object-oriented interface as it directly encapsulates already known results and methodology.

The organisation of this chapter is as follows. In section 6.2, the modelling framework is set up. Section 6.3 explains how to proceed with a general calibration procedure if number of the regimes is N, any integer number greater than two. Section 6.4 presents some conclusions.

6.2 Modelling setup

We focus on a regime-switching log-normal model in discrete time and extend the estimation procedure when there are at least 3 regimes. We assume log-normal increments for asset prices over short time intervals. That is, log returns are normally distributed with constant mean and variance which depend on the state $R_{t_i} = 1, 2, .., N$, $i = 1, 2, .., n$, where R_{t_i} denotes a regime where the process is over the period $[t_i, t_{i+1})$.

So, if S_t stands for the stock price process,

$$\ln \frac{S_{t_{i+1}}}{S_{t_i}} | R_{t_i} \sim \phi(\mu_{t_i}, \sigma_{t_i}), \tag{6.1}$$

where ϕ is a Gaussian density with mean μ_{t_i} and standard deviation σ_{t_i} over the interval $[t_i, t_{i+1})$. The distributions of log-increments conditional on R_{t_i} are independent. For simplicity, we set $S_0 = 1$ in future calculations.

The transition matrix of the underlying hidden Markov process, $\mathbf{\Pi} = (p_{kj})$, is defined as

$$p_{k,j} = P(R_{t_{i+1}} = j | R_{t_i} = k), \quad j = 1, 2, ..N, \quad k = 1, 2, \ldots, N. \tag{6.2}$$

For a two-regime model, we have only six parameters to estimate. Unfortunately, the curse of dimensionality is extreme in this case. For a three-regime model, it is necessary to estimate twelve parameters, i.e., $\Theta = (\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3, p_{11}, p_{12}, p_{21}, p_{22}, p_{31}, p_{32})$.

The procedure for calculating maximum likelihood estimates under a two-regime HMM is known; see Hardy [8], Hamilton [6] and Elliott et al. [2]. It consists of maximising the likelihood function over six parameters. In this case, after using the optimisation constraints on p_{12} and p_{21} $(0 \leq p_{ij} \leq 1)$ and calculating $\pi_1 = \dfrac{p_{21}}{p_{12} + p_{21}}$, the maximisation procedure can be easily done with standard optimisation algorithms. Unfortunately, that is not the case if the number of regimes is greater than two. In the succeeding discussions, we describe how to overcome the challenges in working with models having three or more regimes.

In financial applications, knowing the probability density function of S_{t_n}, $f_{S_{t_n}}(x)$, is essential for finding different risk measures of a portfolio such as value at risk (VaR) and expected shortfall also known as CVaR and conditional tail expectation. To find $f_{S_{t_n}}(x)$, we need to know the joint distribution function for the total sojourn in regime $1, \ldots, N-1$, i.e., we require $P(R(N-1) = k_{N-1}, R(N-2) = k_{N-2}, \ldots, P(1) = k_1)$, for $k_i = 0, 1, 2, \ldots, N$; $\quad i = 1, 2, \ldots, n$. The algorithm for finding $P(R(1) = i)$ in a two-regime model is elaborated in Hardy [8]. We provide a generalisation of the two-regime algorithm in this chapter.

Following the discussion in Hardy [8], we generalise f_{S_n} as follows:

$$S_{t_n} | R \sim \text{LN}(\mu^*(\zeta), \sigma^*(\zeta)), \text{ where} \tag{6.3}$$

$$\mu^*(\zeta) = \sum_{\zeta:k_i\in\{0,1,2,...,N\},\ \sum_{i=0}^{N} k_i=n} k_i\mu_i, \tag{6.4}$$

$$\sigma^*(\zeta) = \sqrt{\sum_{\zeta:k_i\in\{0,1,2,...,N\},\ \sum_{i=0}^{N} k_i=n} k_i\sigma_i^2}, \tag{6.5}$$

where LN stands for the log-normal distribution.

Therefore, the density function for S_{t_i} is

$$f_{S_{t_n}}(x) = \sum_{R=0}^{N} \phi\left(\frac{\ln x - \mu^*(\zeta)}{\sigma^*(\zeta)}\right) P(R(N-1) = k_{N-1}, R(N-2) = k_{N-2}, \ldots, P(1) = k_1). \tag{6.6}$$

Formulae (6.3)-(6.6) are simple generalisations of the algorithm for two regimes. As long as the joint density for staying in different regimes, μ_i, and σ_i are known, it is easy to calculate the values of $f_{S_{t_i}}(x), \forall x > 0$.

6.3 Estimation of model parameters under an N-regime model

This section summarises the results of the parameter estimation for a 2-regime model. We then explain how to generalise the results on an N-regime model

6.3.1 Maximum likelihood estimation in 2-regime model.

Write $y_{t_{n+1}} := \ln \frac{S_{t_{n+1}}}{S_{t_n}}$. The likelihood function for $y = (y_1, y_2, .., y_m)$ can be written as

$$L(\Theta) = f(y_{t_1}|\Theta) f(y_{t_2}|\Theta, y_{t_1}) \cdots f(y_{t_n}|\Theta, y_1, ..., y_{t_{n-1}}). \tag{6.7}$$

The calculation of $L(\Theta)$ is performed recursively, and this is described by Hamilton and Susmmel [7]. The starting values of the recursion are obtained using the invariant distribution $\boldsymbol{\pi} = (\pi_1, \pi_2)$. In the case of a two-regime HMM $\boldsymbol{\pi}$ is known explicitly, and given by

$$\boldsymbol{\pi} = \left(\frac{p_{2,1}}{p_{1,2} + p_{2,1}}, \frac{p_{1,2}}{p_{1,2} + p_{2,1}}\right). \tag{6.8}$$

The maximisation of the likelihood function is with respect to six parameters. It can easily be done using standard optimisation routines, such as Solver in Excel/VBA or any optimisation routine in Matlab.

6.3.2 Maximum likelihood estimation in an N-regime

model. The procedure for finding the parameter estimates for an N-regime model is somewhat similar to the one for the two-regime HMM. The key difference is the estimation of the invariant distribution $\boldsymbol{\pi}$. To the best of our knowledge, there is no explicit formula for the stationary distribution when $K > 2$. This problem is resolved numerically under a weak constraint that the transition matrix $\mathbf{\Pi}$ must be regular, i.e. $\forall\ 1 \leq i, j \leq N,\ \ 0 < p_{ij} < 1$. This works even in the case when $p_{ij} = 1$ for some i and j as during numerical optimisation there is always presence of numerical errors and such an estimate $\hat{p}_{ij}$ would never equal to one. Under the above assumption a normalised eigenvector corresponding to an eigenvalue 1 provides values of stationary distribution $\boldsymbol{\pi}$. The existence of such an eigenvalue is a direct consequence of the Perron-Frobenius theorem.

Consequently, the algorithm for maximising the likelihood function is with respect to $N^2 + N$ parameters. The difference between this approach and the method for $N = 2$ lies in the function g for $\boldsymbol{\pi} = (g(p_{i,j}), 1 \leq i, j \leq N - 1)$. For $N = 2$, g is described by (6.8). The function g in general does not have a closed form for a number of regimes exceeding two and therefore, it has to be estimated numerically.

As the data sample size $n \to \infty$, the time to complete the maximisation procedure increases dramatically. Working with a large n is not really a problem in real-life applications. In our framework, it is assumed that the quantity of regimes does not change although this is not necessarily true in practice. From studies employing more complex dynamic filters, as described in in Elliott et al. [2]; Xi and Mamon[9]; Erlwein et al. [4], the number N of regimes may change after some periods of time for a given data set. When n is very large, we apply the algorithm only to the previous $m < n$ time points. In this argument, we assume that the most recent information about the process is more relevant than its overall performance.

6.3.3 Joint probability function for occupation time in different regimes in an N-regime model.

We start by generalising the procedure for finding the marginal probability mass function $P(R(k) = i)$ for the total sojourn in regime k for $k = 0, 1, 2, \ldots, N;\ \ i = 1, 2, \ldots, n$. Using the same approach one can find a distribution of staying in regime-1 under the

$N-$state HMM, which is $P(R(1) = i)$. However, the case of finding the distribution of being in regime k is still unexplained. There is always a trade-off between the time consumed on writing a code and the performance of the algorithm. From the practical point of view, it does not help if it takes a long time to write a very complicated code that only marginally increases the precision of the method. In typical applications, parameters are calculated by imposing the number of regimes i.e., $N = 2, 3, \ldots$. It is rather time consuming writing a code for each regime case especially if starting from scratch.

We shall propose an alternative way of modelling the probability function for staying in regime K whilst utilising the code for the finding probability function under a 2-regime model. We also show how to calculate the joint probability function for the amount of time spent in each regime, i.e $P(R(1) = k_1,\ R(2) = k_2\ , \ldots,\ P(N-1) = k_{N-1})$.

Suppose the underlying process has N states and its transition matrix is given by

$$\Pi = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}. \tag{6.9}$$

We "transform" our N-regime Markov process into a two-regime process by considering the process as either being in regime one or being in any other regime. Therefore if we define $\mathbf{\Pi}'$ as the transition matrix for the proposed 2-regime process, we have

$$\Pi' = \begin{pmatrix} p_{11} & 1 - p_{11} \\ \sum_{i=2}^{K} p_{i1} / \sum_{i=2,j=1}^{K} p_{ij} & 1 - \sum_{i=2}^{K} p_{i1} / \sum_{i=2,j=1}^{K} p_{ij} \end{pmatrix} \tag{6.10}$$

Following the already established algorithm for the two regime-case, it is easy to find $P'(R(1) = i)$ for the "transformed" model. From our formulation of the transformed model, $P'(R(1) = i) = P(R(1) = i)$. Now, consider a new Markov process which has $N - 1$ states and a transition matrix $\mathbf{\Pi}_2$,

$$\Pi_2 = \begin{pmatrix} p_{22} / \sum_{j=2}^{N} p_{2j} & p_{23} / \sum_{j=2}^{N} p_{2j} & \cdots & p_{2N} / \sum_{j=2}^{N} p_{2j} \\ p_{32} / \sum_{j=2}^{N} p_{3j} & p_{33} / \sum_{j=2}^{N} p_{3j} & \cdots & p_{3N} / \sum_{j=2}^{N} p_{3j} \\ \vdots & \ddots & \vdots & \\ p_{N2} / \sum_{j=2}^{N} p_{Nj} & p_{N3} / \sum_{j=2}^{N} p_{Nj} & \cdots & p_{NN} / \sum_{j=2}^{N} p_{Nj} \end{pmatrix}. \tag{6.11}$$

Applying the previous argument to Π_2, we can find the probability function of being in regime 1 under a Markov chain with a transition matrix $\mathbf{\Pi}_2$. Unfortunately, this procedure will not give us a distribution of being in regime 2, but it will help us calculate the required distribution after several steps of the proposed algorithm.

By definition of $R(i)$,

$$\sum_{i=1}^{N} R(i) = n,$$

and so,

$$P(R(i) = k) = P(R(i^c) = n - k),$$

where i^c stands for the regime not equal to i. To calculate the joint probability function we recall that

$$\begin{aligned} P(R(N-1) &= k_{N-1}, R(N-2) = k_{N-2}, ..., P(1) = k_1) \\ &= P(R(N-1) = k_{N-1} | R(N-2) = k_{N-2}, ..., R(1) = k_1) \\ &\cdot \quad P(R(N-2) = k_{N-2} | R(N-3) = k_{N-3}, ..., R(1) = k_1) \\ &\cdot \quad \ldots\ldots\ldots\ldots \\ &\cdot \quad P(R(3) = k_3 | R(2) = k_2, R(1) = k_1) \\ &\cdot \quad P(R(2) = k_2 | R(1) = k_1) P(R(1) = k_1) \end{aligned}$$

We already showed the calculation of $P(R(1) = k_1)$. To calculate $P(R(2) = k_2 | R(1) = k_1)$, it is important to understand the procedure described earlier when the 2-step procedure was applied to $\mathbf{\Pi}_2$ In short, it consists of working recursively backwards using conditional probabilities when the conditioning is based on the knowledge about the previous regime (cf. section 6.1 of Hardy [8]). By assuming a Markov model being driven by $\mathbf{\Pi}_2$, we constrain the original Markov chain with transition matrix $\mathbf{\Pi}$ going through all of its states except state 1. This means $P(R(2) = k_2 | R(1) = k_1)$ equals to the probability function of being in regime 1 the under Markov chain with transition matrix Π_2.

Given the background above, the general algorithm for finding $P(R(N-1) = k_{N-1}, R(N-2) = k_{N-2}, \ldots, P(1) = k_1)$ is as follows:

Algorithm for finding joint probability mass function for
$R(N-1) = k_{N-1}, R(N-2) = k_{N-2}, ..., R(1) = k_1$, j^{th}*-step*

- Delete the first row and first column of the matrix Π_{j-1} and normalize the resulting matrix as shown in (6.11).
- Construct the matrix Π'_j using the approach described in (6.10).
- Apply the algorithm in finding the probability function of staying in regime 1 as it is depicted in Hardy [8].

After $P(R(N-1) = k_{N-1}, R(N-2) = k_{N-2}, ..., P(1) = k_1)$ is determined, it is easy to construct a joint distribution function for the occupation time spent in different regimes. As shown in section 6.2, this distribution is an essential component to find the density and cumulative distribution function of S_{t_n}.
This method is straightforward and it can be applied even to a large data set with thousands of data points. There is, of course, an associated computer memory requirement that must be allocated to the efficient calculation of the final distribution which is used in equation (6.6).

This approach in conjunction with calculating $f_{S_{t_n}}$ in equation (6.6) yields a set of risk measures, which is the primary objective in risk management. Preliminary results demonstrate that for the distribution of S_{t_n} when the quantity of regimes is small (e.g., $N < 5$) and n is a reasonable value, this approach is preferable to the Monte-Carlo method.

6.4 Conclusion

We put forward a new procedure in estimating parameters of the Markov-switching model by extending the methodology proposed in Hamilton [6] and Hardy [8]. The proposed algorithm is suitable for implementation in an uncomplicated manner and it is based on ideas and procedures that were previously tested on a 2-regime framework. In other words, the approach can be easily adopted to practice once the 2-regime approach was already implemented. Elements of algorithm's structural design is of prime consideration. This approach offers the advantage of systematic implementation over rewriting the code from scratch, which may only lead to negligible additional computing speed for the user. The approach could be employed by practitioners for the calculation of risk measures, and the modelling of financial and economic variables underlying many long-term contracts. Academic and industry users can implement the approach with major financial platforms such as Matlab, Excel/VBA or C++.

6.5 References

[1] Date, P., Ponomareva, K., 2011. Linear and non-linear filtering in mathematical finance: A review, IMA Journal of Management Mathematics 22, 195–211. 161

[2] Elliott, R., Aggoun, L., Moore, J., 1995. Hidden Markov Models: Estimation and Control, Springer, New York. 162, 163, 165

[3] Elliott, R., Chan, L., Siu, T., 2005. Option pricing and Esscher transform under regime switching, Annals of Finance 1, 423–432.

[4] Erlwein, C., Mamon, R., 2009. An online estimation scheme for Hull-White model with HMM-driven parameters, Statistical Methods and Applications 18(1), 87–107. 165

[5] Erlwein, C., Mamon, R., Davison, M., 2011. An examination of HMM-based investment strategies for asset allocation, Applied Stochastic Models in Business and Industry 27, 204–221. 161

[6] Hamilton, J., 1994. Time Series Analysis, Princeton University Press, New Jersey. 162, 163, 168

[7] Hamilton, J., Susmel, R. 1993. Autoregressive conditional heteroskedasticity and changes in regime, Journal of Econometrics 64, 307–333. 164

[8] Hardy, M., 2002. A regime-switching model of long-term stock returns, North American Actuarial Journal 6(1), 171–173. 162, 163, 167, 168

[9] Xi, X., Mamon, R., 2011. Parameter estimation of an asset price model driven by a weak hidden Markov chain, International Journal of Theoretical and Applied Papers on Economic Modelling 28, 36–46. 165

7

Conclusions and further extensions of research

7.1 Contributions

In this thesis, we developed new regime-switching models and the corresponding implementation algorithms tailored to filtering and forecasting of commodity futures prices, pairs trading strategy, liquidity risk forecasting, and FX rates behaviour modelling. Demonstrations on applying our proposed approaches to data sets were conducted, and related statistical inference questions were addressed. The performance of the models put forward in this thesis was assessed on the basis of some statistical metric or by comparing them to current standard models in the literature.

We emphasised the numerical works performed in each model that could be translated to financial practice. The results outlined in this thesis were obtained under the assumption that model parameters are driven by a dynamic HMM and therefore filtering techniques are naturally necessary for parameter estimation. This setting is more realistic than most of the HMM frameworks for modelling that make use of static parameter estimation. To facilitate the parameter estimation procedure, the change of measure methodology was invaluable in conjunction with the EM-algorithm. Our multi-dimensional filtering extensions of the approach previously inspired by Elliott, Mamon, and their collaborators (cf. Mamon and Elliott [3] and [4]; Elliott et al. [2]; Aggoun and Elliott [1]). Our results enriched previous economic interpretations and insights on various financial application areas.

Our implicit goal in this thesis is to bridge the gap between academic and practical approaches by proposing methods and procedures that are theoretically sound but at the same time easily accessible for quants, and some major industry hurdles are reasonably resolved. In addition to an extensive statistical and mathematical formulation, we offered some insights and interpretations in the context of financial market participants' behaviour, convention, and regulation. One distinct feature of this thesis is the comparison exercise involving several competing models with respect to profit targetting and loss level measurements, which are of special interest to hedgers and traders as well as government bodies with oversight functions.

Specifically, we showcased four major contributions: (a) filtering of commodity futures prices and prediction focusing on the analysis of heating oil futures contracts, which were tackled in chapter 2; (ii) forecasting illiquidity via an OU-multivariate model, which was the main theme of chapter 3, and it was found that the crisis of 2008-2009 was not the trigger for creating an illiquidity regime but such a regime already existed a long time before the subprime crisis–it just did not surface with apparent clarity until the crisis event. (iii) implementation of pairs trading strategy backed up by the integration of two filtering algorithms in chapter 4, where it was shown that for a financial sector it is possible to attain a successful algorithmic trading; and finally (iv) modelling FX rate evolution for high-frequency trading in chapter 5 using an HMM under a zero-delay framework.

To complement the above big four studies, we developed static estimation procedure given in chapter 6. The procedure can be used in the estimation for any number of regimes for a Markov-switching model. A central concern that was addressed is the justification of various ways in effectively initialising the algorithms.

7.2 Further extensions of research

Whilst providing solutions to various financial problems, we came to realise the limitations and many imperfections of of our approaches. Rather than viewing these as drawbacks and deficiencies, they can be turned into opportunities that could stimulate more research activities. The list below outlines several research directions that are offshoots of the results in this thesis.

- The extent of the effect of correlation between the driving noise factors and of correlation between Markov chains modulating various model parents is unknown. For example, in our modelling framework for liquidity forecasting in chapter 3, we assume that the three data series correlate only through the underlying Markov chain. In contrast, a simple Kalman filter is built under the assumption that its noise terms are correlated and this is specified by a variance-covariance matrix. It would be appropriate to investigate a new type of dynamic HMM filters where a covariance structure is taken into account for different Markov chains driving different time series.

- A natural extension of the liquidity forecasting problem is to link its applications to trading strategies. For example, it is known that S&P 500 futures are one of the most liquid contracts in the world. Even during uncertain times, bid-ask spread usually makes a very significant correction to contract's price.Buy-side organisations such as insurance companies, hedge funds, pension management funds usually get the largest hit in correcting their positions. By accurately predicting the state of liquidity, these institutions could gain insights on how and when to change their strategies to avoid paying unnecessary fees.

- In the pairs trading strategy in chapter 4 involving the OU-process, the model parameters can only be stable under suitable conditions. For instance, the speed of mean reversion, mean level and variance have all to be positive. However, for automatic trading adhering to these conditions implies full reset of parameters. This in turn, creates loss of valuable historical information as implied by the filters Perhaps, there is a transformation, and its construction is left to further research, that can train the algorithm to effect the parameters attain some form of stability. Alternatively, constraint optimisation with applied to filtering equations construction must be considered.

- For the pairs trading problem, an extension to our approach would be the construction of strategy where the noise is not necessarily Gaussian. From our experience, as long as it is possible to develop filters for the one-dimensional process, the multi-state extension is theoretically available in principle.

- With the frequent trading discussed in chapter 5, we worked with a two-minute bin to deal with uniformising the frequency of data points. Clearly, a standard

procedure with mathematical underpinnings is sought for "cleaning" the data set with a view towards using it for frequent trading model calibration.

7.3 References

[1] Aggoun, L., Elliott, R., 2004. Measure Theory and Filtering: Introduction with Applications, Cambridge University Press, New York. 170

[2] Elliott, R., Aggoun, L., Moore, J., 1995. Hidden Markov Models: Estimation and Control, Springer, New York. 170

[3] Mamon, R., Elliott, R., 2007. Hidden Markov Models in Finance, Springer, New York. 170

[4] Mamon, R., Elliott, R., 2014. Hidden Markov Models in Finance: Further Developments and Applications, Volume II, Springer, New York. 170

Curriculum Vitae

Name:	Anton Tenyakov
Post-Secondary Education and Degrees:	University of Western Ontario London, ON 2010 - 2014 PhD Statistics (Financial Modelling) York University Toronto, ON 2009 - 2010 MA, Probability York University Toronto, ON 2008 - 2009 BSc Hon., Applied Mathematics
Honours and Awards:	Queen Elizabeth II Graduate Scholarship 2013-2014 Ontario Graduate Scholarship 2011-2013
Related Work Experience:	Teaching Assistant University of Western Ontario 2010 - 2014

Publications:

P.Date, R. Mamon and A. Tenyakov, 2013. Filtering and forecasting futures market prices under an HMM framework, Energy Economics, 40, pp 1001-1013