Electronic Thesis and Dissertation Repository

4-22-2022 10:00 AM

Early-Warning Alert Systems for Financial-Instability Detection: An HMM-Driven Approach

Xing Gu, *The University of Western Ontario*

Supervisor: Mamon, Rogemar, *The University of Western Ontario*
Co-Supervisor: Yu, Hao, *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
© Xing Gu 2022

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

Part of the Applied Statistics Commons, Data Science Commons, Longitudinal Data Analysis and Time Series Commons, Probability Commons, Statistical Models Commons, and the Statistical Theory Commons

Recommended Citation

Abstract

Regulators' early intervention is crucial when the financial system is experiencing difficulties. Financial stability must be preserved to avert banks' bailouts, which hugely drain government's financial resources. Detecting in advance periods of financial crisis entails the development and customisation of accurate and robust quantitative techniques. The goal of this thesis is to construct automated systems via the interplay of various mathematical and statistical methodologies to signal financial instability episodes in the near-term horizon. These signal alerts could provide regulatory bodies with the capacity to initiate appropriate response that will thwart or at least minimise the occurrence of a financial crisis. This thesis presents three self-contained but related research undertakings on the subject of inventing early-warning alert systems described as follows.

Our first research study puts forward a generalised multivariate version of a hidden Markov model (HMM) that modulates the regime-switching framework. In particular, the bivariate dynamics of the Financial Stability Index (FSI) and Industrial Production Index (IPI) exhibiting salient features of stochasticity, mean reversion, seasonality, spikes and memory are accurately and simultaneously captured by the resulting HMM filters. An integrated early-warning device is constructed, where the FSI and IPI are taken as inputs, to capture both the financial and business cycles.

In our second research investigation, two different stochastic models are fused together to describe adequately the behaviours of four financial-market indices: Treasury bill yield-Eurodollar spread (TED), US Dollar Index (DXY), Volatility Index (VIX) and S&P 500 bid-ask spread, which are all deemed to mirror the liquidity levels in the financial markets. A blended multivariate HMM, which drives the regime-switching characteristics of market-liquidity risk, is proposed to capture the dynamics of four time series. An early-warning signal extraction method along with its validation diagnostics is devised to generate alerts prior to or at a relatively early stage of the crisis events.

The third research work in this thesis focuses on the determination of signs for possible crisis episodes that may wreak havoc to financial market or economic stability. Synthesising stochastic-process modelling, hidden Markov filtering, Random Forest and XGBoost, we create a hybrid supervised-learning system to detect anomalies in a multivariate time-series index data. Our methodology is capable of efficiently and accurately tracing concomitantly the FSIs of multiple countries and more importantly detecting anomalous FSIs' behaviour portending a possible financial instability. Our proposed model is able to generate dynam-

ically 6-step-ahead binary anomalous-normal classification predictions in a probabilistic sense. Two projected anomaly-warning signals are constructed to forecast two types of extremely anomalous events in the near future with a good accuracy.

Summary for Lay Audience

The wheels of the economy, via a financial system, provide crucial services to meet the needs of households and businesses. Financial mechanisms under this system enable money borrowing for house or car purchase, having protected savings and investments for retirement, and getting valuable job trainings, amongst others. Similarly, businesses require capital to strengthen their production, expand operations, and pay their new and existing workers' remuneration. Thus, a stable financial system is so desired as it foments and beefs up conditions for the prevention of major market-transaction disruptions. Economic participants could then raise and operate funds. However, when weaknesses in the system begin to manifest, problems can pop up and snowball if not controlled, thereby disrupting the supply of goods and services to the society.

The theme of this research is the development of an early-warning alert system (EWS) that generates short-term likelihood forecasts of financial-instability episodes. This thesis, therefore, extends the literature on dynamic modelling that identifies the economic regime characterised by the presence of threats to financial stability. Through a hidden Markov model approach in conjunction with the multivariate stochastic processes, filtering-based calibration, and machine learning techniques, three alert systems are put forward. The first EWS detects financial-stress occurrences using data related to financial and business cycles. In the second EWS, we determine the illiquidity regime, where liquidity refers to a company's ability to pay its short-term debts and cash out its assets quickly. The market-liquidity risk is assessed by examining the joint behaviours of four financial-market indices: Treasury bill yield-Eurodollar spread, US Dollar Index, Volatility Index, and S&P 500 bid-ask spread. A hybrid methodology is proposed in the third EWS to ferret out anomalies in the joint evolution of multiple countries' financial-stability indices.

Our EWSs support the monitoring of financial markets and structures as well as the implementation of regulators' policy frameworks to mitigate the impact of financial fragility. This thesis aids in achieving financial stability for the efficient allocation of resources, financial-risk management, maintenance of employment rate within the neighbourhood of the economy's natural rate, and subduing of relative real or financial asset's price movements.

To my father in Heaven.
To my beloved mother.
To my son, Terrance.

Co-Authorship Statement

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

The results of Chapter 2 were the basis of an article that appeared in the *European Journal of Control*, co-authored with Dr. Rogemar Mamon, Dr. Thibaut Duprey and Dr. Heng Xiong; see [40].

The bulk of the scientific content of Chapter 3 was published , with co-authorship that includes Dr. Rogemar Mamon, Dr. Matt Davison and Dr. Hao Yu, in the *International Journal of Control*; see [39].

Chapter 4 was converted recently to a manuscript (co-authored with Dr. Rogemar Mamon and Dr. Thibaut Duprey), and it is presently under review in the *IEEE Transactions on Neural Networks and Learning Systems*.

This dissertation employed an integrated-article format following Western's thesis guidelines. Each chapter is self-contained and could be read independently. I am responsible for the development of modelling frameworks, the implementation of algorithms, and the completion of the manuscripts. Dr. Rogemar Mamon guided me conscientiously on the research plan, general approaches and heuristics to address problems that were formulated in statistical or mathematical terms, interpretation involved in the empirical analysis of all the aforementioned three research studies, and in addressing competently the referees' concerns for successful journal publications. Many specific computing insights along with certain considerations related to model development were provided by Dr. Hao Yu. Dr. Thibaut Duprey contributed into my practical understanding of the methodologies of financial-stability modelling.

I certify that this document is a full product of my own work. It was conducted from September 2017 to present under the supervision of Dr. Rogemar Mamon and Dr. Hao Yu at *The* University *of* Western Ontario.

London, Ontario

Acknowledgements

I owe a debt of gratitude to my discerning supervisor, Dr. Rogemar Mamon, for his resolute guidance, undiminished motivation and unflinching support throughout my entire graduate education. His considerable experience and meticulous supervision have both ensured my research accomplishments to date and helped me hone many of my transferable skills. I am proud that the steady progress of my doctoral research is building momentum for potential impact, which could be ascribed to his thoughtful mentoring and detailed directions. More importantly, his enduring fortitude and distinct sensibility have carried me through the numerous tough times of my professional life at Western.

I would also like to express my special appreciation and thanks to my co-supervisor, Dr. Hao Yu, from whom I gained an extensive exposure to statistical computing. He is always generous with his time to provide me with many scientific knowhow and constructive advice in dealing with challenges in my research work. I am extremely grateful for his judicious suggestions and unwavering encouragement throughout my graduate studies.

I am deeply grateful to Dr. Thibaut Duprey for providing me with practical insights on various aspects of financial-stability analyses and for his generosity in facilitating access to relevant data sets in the validation of my proposed methodologies. As well, I wish to convey my sincere gratitude to Dr. Heng Xiong for his invaluable support, and in particular for guiding me develop crucial skills that enabled me quickly to get a handle on the thrilling and fascinating world of HMMs/HOHMMs.

I would also like to thank my thesis committee members: Dr. Maciej Augustyniak, Dr. Katarina Grolinger, Dr. Wenqing He and Dr. Jiandong Ren. Their valuable comments and feedbacks certainly strengthen this dissertation. Moreover, I acknowledge the outstanding faculty and staff of the Department of Statistical and Actuarial Sciences (DSAS) for their support and assistance in various ways. I sincerely appreciate the financial support provided by my supervisor's NSERC Discovery grant (RGPIN-2017-04235), the DSAS, Ontario Graduate Scholarship/Queen Elizabeth II Scholarship Program and Ontario Student Assistance Program. I take this opportunity to thank all of my colleagues and friends who never hesitated to extend their moral support and camaraderie whenever I had to surmount certain difficulties. The joyful moments shared together with them will be cherished forever and indelibly etched as happy memories.

Finally, I cannot thank my mother enough for her purest love and for being a constant

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research motivation and objectives

A financial crisis, as a result of financial system instability, generates two broad types of costs: (a) excess investment in some sectors and undue indebtedness and leverage; and (b) severe recessions induced and exacerbated by financial stress. The global financial crisis that started in 2007 was an episode of severe financial market stress, which spilled over to the real economy causing the Great Recession [21]. Considering the great uncertainty of our times, the quest for systematic and reproducible methods that could detect periods of financial crisis has become even more pressing.

In the literature of financial crisis analyses, many studies were based simply on one-state stochastic processes. However, such processes may not be able to capture precisely the behaviours of the underlying data series, especially during the periods of market uncertainty and abrupt fluctuations. We approach this problem by employing a regime-switching paradigm to improve model performance not only in achieving very good model fitting to the data but also in making reliable short-term predictions. The latter consideration plays an important role in pre-crisis warning signal detection.

The core motivation of this thesis is to build automated systems with the fusion of various mathematical and statistical models to detect financial instability episodes in advance. This is for the purpose of providing quantitative insights for practitioners, who could utilise the models' results and implications to build concrete regulatory policies that beef up the

resilience of the financial system. The main research objectives are detailed as follows:

(a) Develop stochastic modelling frameworks, with regime switching, to characterise specific time series related to various financial-stablity problems.

    a.1 Embed a multivariate HMM into an Ornstein-Uhlenbeck (OU) model or an OU hybridised with Geometric Brownian Motion (GBM) modelling frameworks to capture the dynamics of multidimensional data series.

    a.2 Develop self-calibrating recursive filtering algorithms to extract latent information (state of the hidden Markov chain and other HMM-related quantities) from observed multivariate time series.

    a.3 Obtain optimal model parameter estimates utilising the Expectation-Maximisation (EM) algorithm.

(b) Demonstrate the applicability of integrating HMM filtering and other modern machine learning algorithms to provide dynamic estimated levels of a specific country's exposure to financial instability.

    b.1 Create a novel ensemble supervised learning model combining HMM-modulated stochastic model with Random Forests (RF) and XGBoost to delineate multivariate time series (Financial Stress Indices).

    b.2 Develop training and tuning algorithms to calibrate the hybrid supervised learning model to generate multiple-step-ahead predictions to pinpoint the potential anomalous episodes in the future.

    b.3 Assess model's forecasting performance with various diagnostic tools.

    b.4 Define tailored metrics to features' importance analysis to verify the contribution of HMM filtering in terms of improving the model's predictive power.

(c) Construct various early warning alarm signals (EWAS) to advance further the proposed models' capacity for detecting early-stage episodes of financial instability.

    c.1 Build specific warning signals primarily based on HMM-state estimates to gives out alert before or at the early stage of the crisis events by classifying financial stress and liquidity regimes.

    c.2 Create various extreme alert signals utilising outcomes from HMM-related quantities and machine learning algorithms to forecast extreme anomalous episodes that lead to financial instability.

c.3 Assess the EWAS's performance with multiple statistical tools.

1.2 Literature review

This section provides a brief overview of modelling approaches for FSI, market illiquidity and other considerations for the analyses of financial stability. The aim is to provide an overview highlighting the current state of the art in financial-stability modelling and implications on the possible directions of this research field.

1.2.1 Background of Financial Stress Index (FSI)

Measuring the degree of financial stability or instability is a complex endeavour considering the interdependence of internal and external forces acting on the financial system and economy [32]. Financial stress is defined as the pressure endured by economic agents due to the uncertainty and changing expectations about possible losses arising from activities of the financial markets and institutions; see Illing and Liu [49]. The expert-based approach is the prevalent way of determining periods of systemic financial stress. Although there are reproducible methods for the detection of financial-stress events, these methods remain inadequate. There are, nevertheless, growing efforts to come up with a single composite quantitative index that could signify the extent of financial stress.

To buttress with evidence the time periods that are regarded to have undergone systemic financial stress, an FSI is developed in [49] for the Canadian financial system. The FSI development consolidates a survey of responses from senior Bank of Canada's policy makers and economists. The aim is to establish a common agreement on which events are causing financial stress on the Canadian financial market and economy in general. Aided by an FSI analysis, some key features of the relationships between financial stress and the real economy are given in Chapter 2.

In [49], it is asserted that there is a manifestation for FSI to be more or less concomitant of the level of past and current stresses, and the extent of financial system's stress could be gauged from it directly. Its predictive power for future stresses is open for consideration in as much as it could be a useful means in financial-stress assessment and portend early

stresses in the short-term future as FSI tracks down changes in economic variables. These would typically align with tight liquidity conditions and asset-price instability.

Duprey et al. [21] argued that the real economic dimension of systemic stress should be considered in financial-stress modelling. Systemic financial stress episodes are those events that qualify both as periods of financial market stress and periods of real economic stress. It was found in [21] that systemic financial stress events have the following recurrent patterns: (i) financial stress usually occurs first and is followed by real economic stress; and (ii) when associated with high financial market stress, recessionary periods last on average seven months longer, and real economic output declines on average by three additional percentage points.

Duprey et al. [21] constructed a new multi-country level FSI dependent on three core financial market segments: (i) equity markets, (ii) bond markets, and (iii) foreign exchange markets. The details of the FSI's inputs and construction procedure are elaborated in [21]. In the literature, it is possible to identify those systemic financial stress episodes, which are consistent with the expert-categorised stress periods. This is attained by using a Markov-switching and threshold vector autoregressive model to combine FSI with elements quantifying financial stress on business cycle. More specifically, 83% of systemic financial-stress periods detected in [21] are also identified as crisis states by the experts.

1.2.2 Background of market illiquidity

A liquid asset has the characteristic of being able to be traded quickly at a low cost [36]. In [53], the "low cost" condition is change to "without loss". Tenyakov et al. [76] defined liquidity in terms of large quantities of an asset that is realisable with its price impacted only slightly. Furthermore, liquidity relies on funding availability of traders and such funding depends on the market liquidity of assets [9].

The literature typically links an asset's market liquidity and traders' funding liquidity (i.e., the ease with which they can obtain funding) via a model. This model explains the properties of market liquidity, such as (i) drying up suddenly (fragility), (ii) co-movement across assets, (iii) correlation with volatility, (iv) subjected to "flight to quality", and (v) movement in tandem with the market.

The connection between market liquidity and the capital of financial intermediaries and

their funding constraints was examined by Gromb and Vayanos [38]. Through trading via intermediaries, who take advantage of price discrepancies when supplying liquidity, trading gains across segmented markets may materialise to investors. Shocks to asset prices causing capital losses will increase funding constraints and compel intermediaries to decrease their positions. Consequently, this further amplifies the shocks and intensifies market illiquidity.

The TED spread, which is the difference between the risky LIBOR rate and the risk-free US Treasury bill rate, could serve as a measure of market-liquidity level [8]. When the market is illiquid, banks want to get first-rate collateral, which makes holding Treasury bonds more attractive and this then decreases the Treasury bond rate. On the other hand, banks charge higher interest for unsecured loans to compensate for the credit risk, which pushes up the LIBOR rate. Therefore, TED spread widens in times of liquidity crises.

Bid-ask spread is another intuitive and popular measure of illiquidity; it is defined as the difference between the quoted ask and bid prices. Tenyakov et al. [76] pointed out that the S&P 500 bid-ask spread metric is able to capture the illiquidity episodes where TED spread failed to do so. In this study, we investigate the aggregate spread index based on the evolution of the S&P 500, which as stated in [76] contains at least in part the liquidity level of the stock market.

A positive correlation between illiquidity and volatility is another finding in many studies in the literature. Kyle [55] and Karpoff and Walkling [52] measured the illiquidity of stocks assuming that it has a positive correlation to price volatility and a negative one to the price level, market capitalisation, and number of shares. Vayanos and Wang [80] explained that liquidity suppliers who trade high-volatility assets are exposed to more risk and possibly to situations with more asymmetric information. Therefore, they require a large price movements to absorb liquidity shocks, which means that these shocks have huge price impact and cause wide transitory deviations between price and fundamental value. Hauser and Kedar [44] showed that liquidity facilitates a large share redistribution across agents causing changes in average risk aversion, which increases Sharpe-ratio variability, and hence, stock return volatility.

The VIX [1] provides an instantaneous measure of the future degree of volatility and market

---

[1] The Chicago Board Options Exchange (CBOE), defines the volatility index (VIX) as an up-to-the-minute market estimate of expected volatility that is calculated based on real-time S&P 500 index option bid-ask quotes.

uncertainty. The index has been utilised to gauge the level of investors' risk aversion or market sentiment; see Brunnermeier et al. [10] and Bekaert et al. [4], and it has also a negative relation with stock returns as documented by Giot [35] and Whaley [85]. Tenyakov et al. [76] found that VIX is capable of capturing some periods of market illiquidity that were not picked up by the TED spread.

Smales and Kininmonth [71] provided empirical evidence on the existence of a nexus between stock market returns, which are related to investors' sentiment, and the FX markets. It is pointed out in [71] that currency investments tend to depreciate when there is an increase in investors' fear. This disquieting response arises from the financial system's liquidity severely affected for instance by the 2008 collapse of Lehman Brothers and the 2012 European sovereign debt crisis. Zapatero [93] illustrated that in fully integrated markets, the volatility of the exchange rate is explained by the volatility of the stock markets of the two countries concerned. Thus, understanding the behaviour in the FX market is important to measure market illiquidity. An important quantity is the DXY, which is a measure of the value of the US dollar relative to the value of a basket of currencies of the majority of the US's most significant trading partners. Therefore, the DXY could mirror market illiquidity as it contains information regarding investors' expectation of the FX market. For additional discussion of DXY, see ICE: U.S. Dollar Index Contracts.

1.2.3   Implementations of HMM in finance and economics

Shifting structural regimes is a widely observed phenomenon in financial economics. Brunnermeier and Pedersen [9] described the relationship between volatility and speculators' positions through the so-called "margin spiral", which is a characterisation of a market that switches between two equilibriums (or regimes). For these regimes, one is the low-liquidity equilibrium with high volatility and reduced positions, and the other is the high-liquidity equilibrium with low volatility and increased positions.

A regime-switching-based technique is suited for modelling an evolving economic environment assumed to switch amongst different regimes. Such a technique has the capability of clinching changes in economic states by allowing model parameters to change stochastically. Ichiue and Koyama [48] proposed a regime-switching model to probe how exchange rate volatility and depreciation of low-interest-rate currencies are related to each other. In Afonso et al. [1], a time-varying parameter modelling approach is applied to determine the shifts in the pricing regime in the sovereign bond markets of the eurozone area. A regime-

switching mechanism could be embedded in models as shown in Duprey et al. [21]. The intent is to identify the states in the individual or joint dynamics of the data series in conjunction with the detection of financial and business-cycle turning points.

An HMM is a doubly-embedded stochastic process comprised of an observation series and an underlying hidden process delineated by its number of states and transition probabilities or intensities. Moreover, an HMM is the building block of a regime-switching-based technique. This technique pins down some underlying processes in finance or economics where they primarily evolve with the random shifts of their statistics (e.g., mean or variance) amongst different states. Thus, the HMM framework offers flexibility that allows structural regimes, which are governed by the location, scale and shape parameters of a distribution, to shift stochastically over time. Such a framework is suited in capturing widely observed phenomena in finance and economics. The Markov-switching methodology in economics could be traced back to Hamilton's work [41], with the static estimation of the model addressed. The Markov-switching model designed mainly to describe structural changes in time series aims to (i) differentiate regimes of the economy and (ii) estimate the probability of being in an expansion state or contraction state.

The parameters of an HMM could be estimated based on the observed data series unveiling the dynamics of the driving but unobserved Markov chain. Thus, it is necessary to devise optimal, efficient and self-updating estimation techniques. This kind of estimation for the model parameters and Markov chain's unobservable state, as expected, presents some challenges both from the practical and mathematical standpoints. In a comprehensive work, Elliott et al. [25] made significant landmarks in the estimation methodology under an HMM-based modeling framework via the change of measure technique for model identification with the processing of data by batches. The HMM-based technique provides directly immediate recursive filters for the estimates of model parameters without stipulating a priori the dynamics of the observation series other than to say that the observed process is governed by a Markov chain. The recursive filtering, which gives rise to a self-calibrating model, is an innovation vis-à-vis models in the past that are heavily dependent on the static model fitting approach of maximum likelihood estimation [59]. Therefore, the capacity to extract information in order to provide optimal model parameter estimates, by filtering out the noise from the data set, is the dominantly superior feature of HMM-based methods.

Recent research progress in this HMM filtering framework highlights various implementations in the areas of finance and economics. Elliott et al. [26] introduced an HMM approach

in describing the short-rate process and filtering methods were applied to obtain optimal estimates of the parameters. In Elliott and Mamon [24] a Vasiček model is presented in which the mean-reverting level depends on a continuous-time Markov chain. The optimal filtering of log returns of commodity prices was proposed in Mamon et al. [59] whereby both the mean and volatility are modulated by an HMM with finite state space. Erlwein et al. [28] built investment allocation strategies relying on HMM filtering approaches. A model for the evolution of arbitrage-free heating oil futures prices under a HMM regime-switching framework is given in Date et al. [17]. Xiong and Mamon [92] illustrated that an Ornstein–Uhlenbeck (OU) process, driven by a hidden Markov chain (HMC) to model both the mean-reversion and stochasticity, could be utilised to model the evolution of daily average temperatures for the analysis of weather derivatives.

A blended Kalman and HMM multi-regime dynamic filtering approach was proposed in Tenyakov et al. [74] to provide a powerful method for pairs-trading actualisation. Xi and Mamon [88] put forward a mean-reverting interest rate model whose mean-reverting level, speed of mean-reversion, and volatility are all modulated by a higher-order HMM (HOHMM). In Xiong and Mamon [91], an Ornstein–Uhlenbeck (OU) process modulated by a higher-order hidden Markov chain (HOHMM) was used to model the evolution of daily average temperatures (DATs). Erlwein et al. [29] employed an OU process with that HMM driven parameters as well to model the dynamics of electricity spot prices. To assess the levels of market and funding liquidity risks, Tenyakov et al. [76] introduced a multi-variate HMM-embedded OU process. In Siu and Elliot [70], the valuation of American options was discussed in an HMM with jump-diffusion framework, where the dynamics of a latent economic state process over time is captured by a continuous-time hidden Markov chain with finite-state. Mamon et al. [60] proposed an integrated pricing framework for a GMMB focusing on segregated fund contracts where the stock index, interest rate, and mortality rate are driven by HMMs. In Chen et al. [14], A Markov-switching vector autoregression (MS-VAR) model with parameters modulated by a hidden Markov chain was put foward to delineate the causal relationship between renewable-energy prices and economic growth. Xiong and Mamon [90] proposed a multi-dimensional HOHMM online filtering framework to capture the dynamics of salmon futures prices. In Xiang et al. [89], a Markov-switching model (MSM) capturing different economic regimes to model the spot price in the fishing industry was compared with a multi-factor model (MFM) with three correlated stochastic factors. The MSM gives more accurate results in terms of parameter estimation whilst the MFM outperforms the MSM under some metrics related to goodness of fit and model complexity. Mamplata et al. [61] developed a generalised multivariate

HMM model whose parameters are estimated recursively then utilised to test several dynamic investment strategies on four precious metals.

1.2.4 Implementations of machine learning in financial instability detection

Machine learning algorithms could be divided into two categories: supervised learning and unsupervised learning. The aim of supervised learning is to predict the value of a target variable based on some given input features. The supervised learning process is guided by the target variable. In unsupervised learning, the target variables are either unavailable or unobservable and the task is to describe particular associations and patterns, i.e., how the data are organised or clustered amongst a set of input features [43].

In machine learning, an ensemble method combines multiple predictors to obtain predictive modelling performance that is better than simply using a single-component modelling algorithm. Random forests (RFs) [7] are an ensemble machine learning framework for classification and regression via the construction of multiple decision trees. The RF's have become one of the most favoured supervised-learning models due to its considerable advantages, which are: (i) notable accuracy, (ii) robustness to outliers and noise, (iii) ease of use, and (iv) availability of internal estimates of error and variable importance [7]. In addition, the model has strong immunity from overfitting (see [7] [43]), especially for classification problems [43].

Owing to its superiority as pointed out above, an RF is widely employed as a predictive model to solve various regression and classifciation problems in time series analysis. Tyralis and Papacharalampous [78], Karasu and Altan [51] and Hao et al. [42] found that the predictive error could be reduced by selecting features using an RF on different time-series data sets.

Boosting is another powerful learning-ensemble algorithm that aims to convert a set of weak learners, which could only extract a small amount of information, into a strong learner [68]. The weak learners are weighted in some way related to their accuracy. The weights are updated after a weak learner is added. Eventually, the final model is established by weighting all the component learners based on their performance. The final result of the boosting model is obtained by averaging and counting votes in the solutions to regression and classification problems, respectively.

XGBoost (Extreme Gradient Boosting) is based on a gradient-tree boosting [31] and entails a numerical optimisation approach to minimise the loss function via the addition of trees in the gradient-descent algorithm. It is a scalable end-to-end boosting system, which is widely recognised by data scientists to achieve state-of-the-art results and outperforms many other competing models tackling various machine-learning challenges [13]. The fundamental idea of XGBoost [13] is to predict the sum of scores from multiple classification and regression trees (CARTs) considering that one tree is usually not capable of capturing sufficient information from the data. XGBoost uses a tree ensemble model that consists of a group of CARTs [6]. More specifically, new trees are built and trained on the errors of that previous model so that more importance are given to the observations in which existing learners were misclassified. A strong learner is updated by adding a new trained tree that generates a new prediction, whereby each tree's contribution is determined by minimising the overall error of the strong learner.

XGBoost became popular in tackling various finance and economics related problems. In Basak et al. [3], it was found that models based on XGBoost and RFs are able to achieve high-accuracy predictions of the stock-price movements' direction in the medium to long term. Nobre and Neves [64] constructed an automated stock-trading system which combines Principal Component Analysis (PCA), Discrete Wavelet Transform (DWT), XGBoost and a Multi-Objective Optimisation Genetic Algorithm (MOO-GA).

Some researches focusing on financial stability use machine-learning algorithms to build early-warning models to predict the occurrence of a financial crisis. Duttagupta and Cashin [22] utilised a Binary Classification Tree (BCT) model to investigate banking crises in 50 emerging markets and developing countries. An early-warning model is constructed with Random Forests in [2] to predict systemic banking crisis to financial stability. Ward [83] found that the out-of-sample classification performance of the banking crisis indicator constructed with the RF outperforms that of the logit models when tested on a long-run multiple countries data set. Casabianca et al. [12] developed an early-warning system to detect banking-crisis episodes with Adaptive Boosting (AdaBoost) which is able to obtain better out-of-sample performance than the logit models. An early-warning system, which is created in Fioramanti [30] using artificial neural network, surpasses the traditional parametric models in terms of prediction performance in identifying sovereign-debt-crisis episodes.

When dealing with real life-data, HMM could also be employed as a statistical learning model to make classifications on time series according to its estimated parameters. In Li et al. [56] for instance, an HMM-embedded device was developed to detect anomalies

in multivariate time series. In financial economics, where the manifestations of irregular or anomalous events are not apparently visible, an HMM-based model is a beneficial tool in the analysis of pertinent data. In Cao et al. [11], for example, an HMM with wavelet transformations and gradients was utilised to detect price manipulation activities in the stock markets.

1.3 Structure of the thesis

The remaining parts of this thesis are organised as follows. Chapter 2 presents the results of a research study, where a hybrid bivariate OU-GBM regime-switching model is pioneered with parameters driven by a hidden Markov chain to describe the features of the FSI and Industrial Production Index (IPI). We create an early warning device for financial crisis based on the HMM filtering approach. In Chapter 3, we employ a 4-dimensional blended HMM to evaluate the market liquidity by capturing the dynamics of 4 indices of the financial markets. An early-warning system (EWS) relying on the HMM-state estimates and the related statistical assessment on its capability of detecting financial instability are demonstrated. Chapter 4 introduces a hybrid supervised learning device which integrated stochastic modelling, hidden Markov model, Random Forest and XGBoost to detect anomaly episodes of multiple financial stress indices. We propose two early alert signals with outcomes from HMM filters and XGBoost classier to identify extreme anomalous episodes. Lastly, a summary of research contributions and directions of future work are provided in Chapter 5. The contents of Chapters 2-4 are briefly presented below.

1.3.1 Online estimation for a predictive analytics platform with a financial-stability-analysis application

An online parameter estimation via filtering recursions is constructed to support a data-analytics scheme in the predictive domain. Multivariate financial market indices or signals revealed in real time are used in our numerical implementation. This work contributes to the analysis and forecasting of financial crises in an environment that evolves dynamically. In particular, we capture the regime-switching characteristics of the FSI and IPI, designed to detect periods of financial crisis. We integrate two different stochastic models for FSI and IPI deemed to mirror the systemic financial stress levels in the financial and business

cycles, respectively. The joint dynamics of the FSI and IPI, exhibit stochasticity, mean reversion, seasonality, and occasional jumps are identified in the most efficient way. All parameters are modulated by a discrete-time hidden Markov chain that switches between economic regimes reflecting various interacting economic forces. Through change of reference probability technique, adaptive multivariate filters are derived which in turn provides online optimal parameter estimates. Historical Canadian economic-based FSI and IPI are examined and an early-warning signal extraction method is put forward to generate alerts at the early stage of some crisis events. Our modelling approach captures the empirical characteristics of FSI and IPI as well as provides auspiciously early warnings for episodes of systemic financial crisis.

1.3.2 An analysis and forecasting of financial market liquidity regimes

A multivariate HMM-based approach is developed to capture simultaneously the regime-switching dynamics of four financial market indicators: Treasury-Euro Dollar rate spread, US dollar index, volatility index and S&P 500 bid-ask spread. These indicators exhibit stochasticity, mean reversion, spikes and state memory, and they are deemed to drive the main characteristics of liquidity risk and regarded to mirror financial markets' liquidity levels. In this chapter, an online system is proposed in which observed indicators are processed and the results are then interfaced with an advanced alert mechanism that gives out appropriate measures. In particular, two stochastic models, with HMM-modulated parameters switching between liquidity regimes, are integrated to capture the evolutions of the four time series or their transformations. Parameter estimation is accomplished by deriving adaptive multivariate filters. Indicators' joint empirical characteristics are captured well and useful early warnings are obtained for occurrence prediction of illiquidity episodes.

1.3.3 A multivariate-index-driven anomaly detection system with supervised learning

We develop a hybrid supervised learning system to detect anomalies in a multivariate time-series index data. Our focus of application is the determination of signs for possible crisis episodes that may wreak havoc to financial market or economic stability. Our proposed statistical-computing approach synthesises stochastic process modelling, hidden Markov filtering, Random Forest and XGBoost. Such an approach is capable of efficiently and

accurately tracing simultaneously the financial stress indices (FSIs) of multiple countries and more importantly identifying anomalous FSIs' behaviour that signals an impending financial instability. We show that our method is capable of dynamically making 6-step-ahead binary anomalous-normal classification predictions in a probabilistic sense for the benefit of industry practitioners and regulators. Our method, which also gives rise to an early-warning system, is benchmarked with other alternative methods and its advantage is highlighted via various model validation measures.

Chapter 2

Online estimation with predictive analytics for financial stability analysis

2.1 Introduction

2.1.1 Motivation

The prediction of financial crises has garnered much attention from researchers and interdisciplinary methodologies have been employed in an attempt to explain, forecast, or suggest remedies to deal with the presence of financial stress. Our motivation is the early detection of financial-stress episodes, which is crucial for bankers and regulators to alleviate the shock brought about by financial crises. We aim to provide continuity and respond to such a research theme of developing approaches and strategies geared towards strengthening global financial stability. In this chapter, an early-warning predictive analytics system is developed to assist in detecting future periods of financial instability; in so doing, certain mitigation measures could be put in place in advance. It is also our intention to showcase the considerable potential impact and influence of powerful techniques in systems and control theory to deal with contemporary challenges in economics and finance such as those highlighted in Imanov [50] and Li and Zhang [57]

In the context of our exposition, we define financial stress as the force borne by economic agents due to uncertainty and changing expectations concerning possible losses on activities occurring in the financial markets and institutions; see Illing and Liu [49]. The global

financial crisis that started in 2007 was a period of "systemic" financial market stress, which spilled over to the real economy causing the Great Recession. A discrete-event dynamic system is one paradigm in mitigating financial instability, and the pursuit of a solution platform that incorporates information streaming in real time, similar to the concept in Murao et al. [63], is paramount. Early detection of financial-stress episodes is crucial for bankers and regulators to alleviate the shock brought about by financial crises. The expert-based approach is the currently prevalent way of identifying systemic financial-stress episodes; there are reproducible methods for the detection of financial-stress periods but they remain inadequate; refer to Duprey et al. [21] for further discussion. Considering the huge uncertainty of our times, the quest for objective and quantitative methods that could identify occurrences of systemic financial stress, which is usually difficult to pinpoint and classify, has become even more pressing (cf Liang [58]).

2.1.2 Modelling crisis episodes via FSI and IPI

To model and forecast crisis episodes, we consider the Financial Stress Index (FSI) and Industrial Production Index (IPI). Both the FSI and IPI are the primary components of our data series as previous research suggest they have the capability to link real economic turbulence and the instability in financial cycle. An FSI is developed in Illing and Liu [49] to capture epochs of systemic financial stress for the Canadian financial system. The development of FSI comes from a survey of senior Bank of Canada policy-makers and economists aimed to establish a consensus on which events have been the most stressful for Canadian markets over the past 25 years. The literature also indicate some key features of the relationships between financial stress and real economy. These include the findings that (i) real economic behaviour can be altered sufficiently to have adverse effects on the real economy if financial stress is systemic; (ii) some measures of confidence for business cycle demonstrates stronger relationships with the FSI during recessions, but do not necessarily appear to be strongly coincident with major financial stress events; and (iii) the sampling frequency and timing of data of business confidence may affect these results.

It is also emphasised in [49] that the FSI pins down the contemporaneous level of stress and provides an ordinal measure of stress in the financial system. However, it is not expected to have a strong predictive power for future stresses or crises. Hence, FSI is viewed as a preliminary attempt to quantify the stress spectrum. Yet, it could be more informative in developing an early warning indicator model using the FSI to explain changes in real

economic variables. Financial stress impairs not only the financial system but also lead to significant losses in the real economy. For example, it could result to tight liquidity conditions and asset-price instability, both of which cause an increase in the cost of capital and reduce private investment and consumption.

Duprey et al. [21] argued that the real economic dimension of systemic stress should be considered in financial-stress modelling. Systemic financial stress episodes are those events that qualify both as periods of financial market stress and periods of real economic stress. It is asserted in [21] that systemic financial stress events have the following recurrent patterns: (i) financial stress usually occurs first and is followed by real economic stress; and (ii) when associated with high financial market stress, recessionary periods last on average seven months longer, and real economic output declines on average by three additional percentage points.

2.1.3 Capturing regime-switching features using a hidden Markov model

A one-state stochastic model may not be adequate to describe the behaviour of financial stress with sufficient accuracy, especially during occurrences of some financial crises. The introduction of a regime-switching mechanism via an HMM yields modelling flexibility. A bivariate Markov-switching vector autoregressive (MSVAR) framework is proposed here to model jointly a newly constructed monthly country-specific FSI and IPI. This is aimed to identify regimes of systemic financial stress characterised by significant jumps in the joint dynamics of real economic and financial market stress data proxied by appropriate indices. The outcome is consistent with many expert-based stress periods in which 83% of the model-based systemic financial stress periods are also identified as crises by experts [21].

Methodologies have been developed in order to capture the dynamic behaviour of financial and economic variables. Typically, Markov chain-driven stochastic models are employed as framework for pricing financial derivatives under a regime-switching framework; see for example, Song et al. [72]. But, a regime-switching mechanism could also be embedded in models to identify states in the individual or joint dynamics of data series to detect financial and business cycle turning points [21]. A regime-switching-based technique assumes an economic environment being modelled evolves and shifts between different regimes. Such a technique has the flexibility and capability of handling changes in economic states by allowing model parameters to change stochastically. Davig and Gerlach [18] proposed

a test of the response of stock prices to Federal Reserve policy shocks using a Markov-switching framework. Gerdrup et al. [34] introduced a financial stress model with a shock component driven by a Markov chain. To ensure that reality is modelled more suitably, a regime-switching model is encapsulated in a hidden Markov model (HMM). That is, an HMM is a doubly-embedded stochastic process composed by an observation series and an underlying unseen process defined by its number of states and transition probabilities. Thus, an HMM has the parameters that change over time in accordance with the dynamics of an unobserved Markov chain but could be estimated. The estimation of HMM is via filtering methods, whose applications are quite popular in electronics and electrical engineering, physics and statistics (Mamon et al. [59]). Our HMM application to the area of financial stability modelling is rather new. Markov-switching modelling began with Hamilton's research [41], with static estimation of the model addressed, concentrating on a structural approach to (i) distinguish states of the economy and (ii) infer the probability of being in a state of expansion or contraction using a hidden Markov chain.

In practice, the data-fitting performance of Markov-switching models is of prime importance. As the Markov chain is unobserved, it is necessary to devise optimal, efficient and self-updating estimation techniques. As can be expected, this kind of estimation for the model parameters and Markov chain's unobservable state presents some challenges both from the practical and mathematical standpoints. In a comprehensive work, Elliott et al. [25] made significant landmarks in the estimation methodology under a HMM-based modeling framework via the change of measure technique for model identification after processing batches of data. The HMM-based technique provides directly immediate recursive filters for the estimates of model parameters without stipulating a priori the dynamics of the observation series other than to say that the observed process is governed by a Markov chain. The recursive filtering, which gives rise to a self-calibrating model, is an innovation vis-à-vis models in the past that are heavily dependent on the static model fitting approach of maximum likelihood estimation [59].

Developments in the regime-switching literature typically concentrate on the extension of the HMM framework to address various problems in quantitative finance, insurance, economics, epidemiology, and other branches of the sciences and engineering. Elliott et al. [26] introduced an HMM approach in describing the short-rate process and filtering methods were applied to obtain optimal estimates of the parameters. In Elliott and Mamon [24] a Vasiček model is presented in which the mean-reverting level depends on a continuous-time Markov chain. The optimal filtering of log returns of commodity prices is proposed

in [59], whereby both the mean and volatility are modulated by an HMM with finite state space. Erlwein et al. [28] built investment allocation strategies relying on HMM filtering approaches. A model for the evolution of arbitrage-free heating oil futures prices under a HMM regime-switching framework is given in Date et al. [17]. Xiong and Mamon [92] illustrated that an Ornstein–Uhlenbeck (OU) process, driven by a hidden Markov chain (HMC) to model both the mean-reversion and stochasticity, could be utilised to model the evolution of daily average temperatures for the analysis of weather derivatives. A blended Kalman and HMM multi-regime dynamic filtering approach was proposed in Tenyakov et al. [74] to provide a powerful method for pairs-trading actualisation. Xi and Mamon [88] put forward a mean-reverting interest rate model whose mean-reverting level, speed of mean-reversion, and volatility are all modulated by a higher-order HMM (HOHMM). In Xiong and Mamon [91], an Ornstein–Uhlenbeck (OU) process modulated by a higher-order hidden Markov chain (HOHMM) is used to model the evolution of daily average temperatures (DATs). Erlwein et al. [29] employed an OU process with that HMM driven parameters as well to model the dynamics of electricity spot prices. To assess the levels of market and funding liquidity risks, Tenyakov et al. [76] introduced a multivariate HMM-embedded OU process.

2.1.4 Methodology

The overview of our methodology and its rationale are important elements for this research investigation. This work puts forward a self-calibrating bivariate HMM governing the regime-switching framework, in an effort to extract early-warning signals for possible future-crisis occurrences. It reinforces research progress in the areas studied in [21] and [49] and promote similar and related objectives. We shall make use of both the FSI and IPI to capture, with improved accuracy, the behaviour and features arising from the impact of systemic financial stress on financial and business cycles. This is important for regulators who must prepare in advance to maintain financial stability in the regional and global scale. Noting that the FSI's constructed in [21] and [49] both have mean-reverting feature, it is, therefore, reasonable to consider the OU process as a natural candidate model for the FSI movement. We use a geometric Brownian motion (GBM) to describe IPI's dynamics as they are relatively smooth and tend to increase in the long run. In essence, we propose a hybrid bivariate stochastic process aptly to pin down the main features (e.g., seasonality, mean-reversion and jumps) of the joint movement of two indices. The FSI and IPI observations will be used to infer the presence or lack thereof of systemic financial stress, which

is referred to as financial stress widespread across different financial markets that can have particularly serious implications on the real side of the economy. Informational content is 'filtered out' from the observation process to find the 'true' state of the hidden Markov chain, which drives the random switching of economic regimes in discrete time.

In this chapter, we shall demonstrate the change of probability measure technique in estimating optimal filters of various quantities pertinent to the calculation of the bivariate model's parameters . This is done through the adaptive processing of market signals (treated as the FSI and IPI series) that will unconceal information about the states of the financial economy. A self-tuning algorithm is generated, which updates parameters with the continual arrival of new observed signals (i.e., indices). Our methodology extends the HMM multivariate results of Elliott et al. [25] and Tenyakov et al. [76] with each univariate series constituting the entire multivariate set up being different from each other. This more flexible modelling set up avoids the forward-backward algorithm typical in most filtering techniques, thereby entailing much less memory during computation. Other filters (e.g., Hamilton-type filters) are computationally intensive to implement because they are based on static algorithms requiring full reruns involving the original data set every time there is addition of a few data points, we successfully circumvented the issue of recalculation involving old data set that keeps getting larger as data collection continues.

2.1.5 Major contributions

To delineate our contributions from the current state of the art in systemic financial stress modelling, we highlight the following accomplishments in this chapter: (i) Our proposed approach combining the OU and GBM in an integrated HMM framework is a new attempt to capture jointly the important stylised properties of bivariate data series, such as mean reversion, seasonality, and stochasticity, in identifying systemic financial stress episodes. (ii) Although our methodology is adopted from the literature in Markovian regime-switching models supporting derivative pricing and risk management, we tailor a particular application to financial-stress modelling and analysis. (iii) Theoretical formulation is accompanied by detailed empirical implementation with model validation diagnostics and some aspects of inference adressed. (iv) Finally, our empirical work provides an impetus for an HMM-based early warning system (EWS). This work's significance hinges to its benefit to central bankers and regulatory authorities who are on the look out to avert, or at least mitigate, the effects of future financial crises.

We note that Hubrich and Tetlow [46] also empirically modelled the interaction between financial stress and economic dynamics in a Markov-switching framework using the United States' FSI, albeit utilising a Markov-switching vector autoregression set up. Our work is distinct but complements [46] considering our: (a) richly parameterised bivariate model blending GBM and OU specifications customised to pin down important properties of our time series data, (b) natural choice of model estimation via the HMM-based online filtering scheme providing regime-switching parameters as well, and (c) focus on bivariate data consistent with the principle of parsimony (i.e., choice of simple model specification but with greatest explanatory power). Additionally, it is observed in [46] that FSI modelling by some of the twelve Reserve Banks in the US employ principal component analysis to choose only the relevant variables. Lastly, we leave out other economic variables related to monetary policy. This is because as the result in [46] shows, conventional monetary policy is not particularly effective in financial-stress periods; instead, a much more potent means is to induce a switch in the economy from a high-stress regime to a financially stable regime.

The structure of this chapter is as follows. Section 2 presents the formulation of the joint modelling of FSI and IPI with a discrete-time hidden Markov chain governing the model parameters. In Section 3, we introduce the change of probability measure technique to derive recursive filtering equations for quantities that are functions of HMM, and carry out an online parameter estimation. In Details of the numerical implementation of our proposed model for the bivariate FSI and IPI data sets are given in Section 4. The selection of the most appropriate model setting is demonstrated in Section 4 by comparing prediction performance and penalised log-likelihood of different competing set ups. Furthermore, an empirical early warning signal extraction method is proposed in Section 5. The last section summarises and includes some concluding remarks.

2.2 Model construction

2.2.1 The OU process

Suppose X_t is an FSI following the OU process, i.e.,

$$\mathrm{d}X_t = \theta\left(\mu - X_t\right)\mathrm{d}t + \sigma\mathrm{d}W_t, \tag{2.1}$$

where μ is the mean level; θ is the speed of mean reversion; and σ is the volatility. In (2.1), W_t is a standard Brownian motion defined on a probability space $\left(\Omega, \mathscr{F}^X, P\right)$, where $\mathscr{F}^X$ is the filtration generated by X_t.

It is assumed that θ, μ and σ are all positive constants. By Itô's lemma,

$$X_t = X_0 e^{-\theta t} + (1 - e^{-\theta t})\mu + \sigma e^{-\theta t} \int_0^t e^{\theta s} \mathrm{d}W_s. \tag{2.2}$$

Discretising the solution in Eq. (2.2), we get

$$X_{t_{k+1}} = X_{t_k} e^{-\theta \cdot \Delta t} + (1 - e^{-\theta \cdot \Delta t})\mu + \sigma \sqrt{\frac{1}{2\theta}(1 - e^{-2\theta \cdot \Delta t})} \, w_{k+1}, \tag{2.3}$$

where the $\Delta t = t_{k+1} - t_k$ and $w_{k+1} \sim IID\ N(0, 1)$. The derivation of the second term in (2.3) is justified by the property of a normal distribution and the Itô's isometry.

2.2.2 The GBM

Let Q_t be an IPI following the GBM dynamics. That is,

$$\mathrm{d}Q_t = \eta Q_t \mathrm{d}t + \xi Q_t \mathrm{d}B_t, \tag{2.4}$$

where η is the percentage drift and ξ is the percentage volatility. In (2.4), B_t is a standard Brownian motion defined on a probability space $\left(\Omega, \mathscr{F}^Q, P\right)$, where $\mathscr{F}^Q$ is the filtration generated by Q_t. The parameters η and ξ are positive constants. By Itô's lemma, Eq. (2.4) has the solution

$$\ln Q_t - \ln Q_0 = \left(\eta - \frac{\xi^2}{2}\right)t + \xi B_t. \tag{2.5}$$

Let $Y_t = \ln Q_t$, so that

$$Y_t = Y_0 + \left(\eta - \frac{\xi^2}{2}\right)t + \xi B_t \tag{2.6}$$

Discretising Eq. (2.6), by applying the Euler approximation, we get

$$Y_{t_{k+1}} = Y_{t_k} + \left(\eta - \frac{\xi^2}{2}\right)\Delta t + \xi \sqrt{\Delta t} \cdot b_{k+1}, \tag{2.7}$$

where the $\Delta t = t_{k+1} - t_k$ and $b_{k+1} \sim IID\ N(0, 1)$. The third term is obtained by invoking the properties of a standard Brownian motion.

2.2.3 Markov chain-governed parameters

For both models of indices, it is realistic to think that their parameter values are time-dependent. Similar to Zhou and Mamon [95], we regard the states of a Markov chain as regimes of an economy, or more specifically, financial stress regimes that are dependent on certain factors causing economic turbulence. Thus, to capture economic regime-switching, the parameters θ, μ, σ, η and ξ are modulated by a discrete-time Markov chain $\mathbf{z}_k$, for $k = 0, 1, \ldots$. Its state space is finite and it is isomorphic to the canonical basis of $\mathbb{R}^N$, which is the set $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_N\}$. The vector $\boldsymbol{e}_i = (0, ..., 0, 1, 0, ..., 0)^\top$, where $\top$ denotes the transpose of a vector, is a unit vector with 1 in its ith component and helps simplify a lot of algebra in the succeeding calculations. The Markov chain is dictated by the representation

$$\mathbf{z}_{k+1} = \mathbf{\Pi}\mathbf{z}_k + \mathbf{v}_{k+1}. \tag{2.8}$$

Note that in (2.8), $\mathbf{\Pi}$ is a transition matrix and $\mathbf{v}_{k+1}$ is a martingale increment with $E[\mathbf{v}_{k+1}|\mathscr{F}^z] = \mathbf{0}$, where $\mathscr{F}^z$ is the filtration generated by $\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2, \ldots$.

In particular, the dependence of the parameters on the Markov chain is reflected in the notations $\theta(\mathbf{z}_k)$, $\mu(\mathbf{z}_k)$, $\sigma(\mathbf{z}_k)$, $\eta(\mathbf{z}_k)$, and $\xi(\mathbf{z}_k)$. Under the one-state setting, the parameters in the distribution of $X_{t_{k+1}}$ are constants over the time interval $(t_k, t_{k+1}]$. For the $N-$state setting, we assume that the parameters under the one-state setting depend on a Markov chain. Hence, the two-dimensional process $(X_t, Y_t)^\top$ can be expressed as

$$\begin{cases} X_{t_{k+1}} = X_{t_k} e^{-\theta(\mathbf{z}_k)\Delta t} + (1 - e^{-\theta(\mathbf{z}_k)\Delta t})\mu(\mathbf{z}_k) + \sigma(\mathbf{z}_k)\sqrt{\frac{1}{2\theta(\mathbf{z}_k)}(1 - e^{-2\theta(\mathbf{z}_k)\Delta t})}\, w_{k+1} \\ Y_{t_{k+1}} = Y_{t_k} + \left(\eta(\mathbf{z}_k) - \frac{\xi^2(\mathbf{z}_k)}{2}\right)\Delta t + \xi(\mathbf{z}_k)\sqrt{\Delta t}\, b_{k+1}. \end{cases} \tag{2.9}$$

Note that $\mu(\mathbf{z}_k) = \langle \boldsymbol{\mu}_k, \mathbf{z}_k \rangle$, $\theta(\mathbf{z}_k) = \langle \boldsymbol{\theta}_k, \mathbf{z}_k \rangle$, $\sigma(\mathbf{z}_k) = \langle \boldsymbol{\sigma}_k, \mathbf{z}_k \rangle$, $\eta(\mathbf{z}_k) = \langle \boldsymbol{\eta}_k, \mathbf{z}_k \rangle$ and $\xi(\mathbf{z}_k) = \langle \boldsymbol{\xi}_k, \mathbf{z}_k \rangle$, where $\boldsymbol{\mu}_k = (\mu_{k,1}, \mu_{k,2}, ..., \mu_{k,N})^\top$, $\boldsymbol{\sigma}_k = (\sigma_{k,1}, \sigma_{k,2}, ..., \sigma_{k,N})^\top$, $\boldsymbol{\theta}_k = (\theta_{k,1}, \theta_{k,2}, ..., \theta_{k,N})^\top$, $\boldsymbol{\eta}_k = (\eta_{k,1}, \eta_{k,2}, ..., \eta_{k,N})^\top$ and $\boldsymbol{\xi}_k = (\xi_{k,1}, \xi_{k,2}, ..., \xi_{k,N})^\top$ are all in $\mathbb{R}^{\mathbb{N}}$; and $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^N$. Write $\mathbf{w}_k := (w_k, b_k)^\top$ with w_k and b_k are independent for all $k = 0, 1, 2, \ldots$. Note that all processes in our modelling set up are supported by a complete probability space $(\Omega, \mathscr{F}, P)$, where $\mathscr{F} = \mathscr{F}^X \vee \mathscr{F}^Q \vee \mathscr{F}^z$.

2.2.4 Probability density of observation process

We derive the conditional probability density function of the process $(X_{t_{k+1}}, Y_{t_{k+1}})^\top$. X_{t_k} in Eqs. (2.9) can be written as

$$X_{t_{k+1}} = \alpha(\mathbf{z}_k)X_{t_k} + \beta(\mathbf{z}_k) + \kappa(\mathbf{z}_k)\, w_{k+1} \tag{2.10}$$

with

$$
\begin{cases}
\alpha\left(\mathbf{z}_k\right) = e^{-\theta(\mathbf{z}_k)\Delta t} \\[2mm]
\beta\left(\mathbf{z}_k\right) = (1 - e^{-\theta(\mathbf{z}_k)\Delta t})\mu(\mathbf{z}_k) \\[2mm]
\kappa\left(\mathbf{z}_k\right) = \sigma(\mathbf{z}_k)\sqrt{(2\theta(\mathbf{z}_k))^{-1}\left(1 - e^{-2\theta(\mathbf{z}_k)\cdot\Delta t}\right)}
\end{cases}
\tag{2.11}
$$

Eq. (2.10) implies

$$
X_{t_{k+1}}\Big|X_{t_k} \sim N\left(\alpha(\mathbf{z}_k)X_{t_k} + \beta\left(\mathbf{z}_k\right),\ \kappa\left(\mathbf{z}_k\right)\right).
\tag{2.12}
$$

For the distribution of Y_{k+1}, we consider

$$
\begin{aligned}
Y_{t_{k+1}} - Y_{t_k} &= \left(\eta(\mathbf{z}_k) - \frac{\xi^2(\mathbf{z}_k)}{2}\right)\Delta t + \xi(\mathbf{z}_k)\sqrt{\Delta t}\,b_{k+1} \\
&= \zeta\left(\mathbf{z}_k\right) + \nu\left(\mathbf{z}_k\right)b_{k+1}
\end{aligned}
\tag{2.13}
$$

with

$$
\begin{cases}
\zeta\left(\mathbf{z}_k\right) = \left(\eta(\mathbf{z}_k) - \frac{\xi^2(\mathbf{z}_k)}{2}\right)\Delta t \\[2mm]
\nu\left(\mathbf{z}_k\right) = \xi(\mathbf{z}_k)\sqrt{\Delta t}.
\end{cases}
\tag{2.14}
$$

This implies

$$
Y_{t_{k+1}}\Big|Y_{t_k} \sim N\left(Y_{t_k} + \zeta\left(\mathbf{z}_k\right),\ \nu\left(\mathbf{z}_k\right)\right).
\tag{2.15}
$$

2.3 Filters and parameter estimation

2.3.1 Change of measures

Under the real-world probability measure P, the true state of the underlying Markov chain $\mathbf{z}_k$ is neither observed nor measured directly since it is "latent" in the noisy observation process with "real world" dynamics given by Eqs. (2.10) and (2.13). Our objective is to "filter" the noise out of the observation process in the best possible way. Unfortunately, the derivation of filters under P is not straightforward. Inspired by the approach described in Elliott et al. [25], we perform a change of probability measure to introduce the the ideal-world measure $\bar{P}$ from the real-world probability measure P by invoking the discrete-time version of the Girsanov's theorem. Under this ideal measure, the observations are independent and identically distributed random variables which makes the calculations of conditional expectations manageable. The filters, which are conditional expectations, are then related back

to the real-world by the use of the Bayes' theorem for conditional expectation. Following section 3.4, page 62 of [25], the real-world measure P equivalent to an ideal measure $\bar{P}$ is constructed through the Radon-Nikodym derivative

$$\bar{\Lambda}_k = \left.\frac{\mathrm{d}P}{\mathrm{d}\bar{P}}\right|_{\mathscr{F}_k} = \prod_{l=1}^{k} \bar{\lambda}_l^{(X)}\bar{\lambda}_l^{(Y)} \tag{2.16}$$

with

$$\bar{\lambda}_l^{(X)} = (\kappa(\mathbf{z}_{l-1})\phi(X_l))^{-1}\, \phi\left(\frac{X_l - \alpha(\mathbf{z}_{l-1})X_{l-1} - \beta(\mathbf{z}_{l-1})}{\kappa(\mathbf{z}_{l-1})}\right)$$

$$= (\kappa(\mathbf{z}_{l-1}))^{-1} \exp\left\{-\frac{1}{2}\left[\left(\frac{X_l - \alpha(\mathbf{z}_{l-1})X_{l-1} - \beta(\mathbf{z}_{l-1})}{\kappa(\mathbf{z}_{l-1})}\right)^2 - (X_l)^2\right]\right\}$$

$$\bar{\lambda}_l^{(Y)} = (\nu(\mathbf{z}_{l-1})\phi(Y_l))^{-1}\, \phi\left(\frac{Y_l - Y_{l-1} - \zeta(\mathbf{z}_{l-1})}{\nu(\mathbf{z}_{l-1})}\right)$$

$$= (\nu(\mathbf{z}_{l-1}))^{-1} \exp\left\{-\frac{1}{2}\left[\left(\frac{Y_l - Y_{l-1} - \zeta(\mathbf{z}_{l-1})}{\nu(\mathbf{z}_{l-1})}\right)^2 - (Y_l)^2\right]\right\},$$

where $\phi(\cdot)$ is the density function of a standard normal random variable.

2.3.2 Filtering

Let $\hat{\mathbf{z}}_k$ be the conditional expectation of $\mathbf{z}_k$ given $\mathscr{F}_k$ under probability measure P, i.e.,

$$\hat{\mathbf{z}}_k := E\left[\mathbf{z}_k \middle| \mathscr{F}_k\right] = \left(\hat{z}_k^{(1)}, \hat{z}_k^{(2)}, \ldots, \hat{z}_k^{(N)}\right)^{\top} \in \mathbb{R}^N, \tag{2.17}$$

$$\hat{z}_k^{(i)} = P\left(\mathbf{z}_k = \mathbf{e}_i \middle| \mathscr{F}_k\right) = E\left[\langle \mathbf{z}_k, \mathbf{e}_i \rangle \middle| \mathscr{F}_k\right], \tag{2.18}$$

where $\mathscr{F}_k$ is the filtration generated by the bivariate observation reflecting all information available up to time k. By the Bayes' theorem,

$$\hat{\mathbf{z}}_k = E\left[\mathbf{z}_k \middle| \mathscr{F}_k\right] = \frac{\overline{E}\left[\bar{\Lambda}_k \mathbf{z}_k \middle| \mathscr{F}_k\right]}{\overline{E}\left[\bar{\Lambda}_k \middle| \mathscr{F}_k\right]}. \tag{2.19}$$

Write $\mathbf{p}_k := \overline{E}\left[\bar{\Lambda}_k \mathbf{z}_k \middle| \mathscr{F}_k\right]$ so that

$$\overline{E}\left[\bar{\Lambda}_k \middle| \mathscr{F}_k\right] = \overline{E}\left[\bar{\Lambda}_k\left(\sum_{i=1}^{N}\langle \mathbf{z}_k, \mathbf{e}_i \rangle\right)\middle| \mathscr{F}_k\right] = \sum_{i=1}^{N}\overline{E}\left[\langle \bar{\Lambda}_k \mathbf{z}_k, \mathbf{e}_i \rangle \middle| \mathscr{F}_k\right]$$

$$= \sum_{i=1}^{N}\left\langle\overline{E}\left[\bar{\Lambda}_k \mathbf{z}_k \middle| \mathscr{F}_k\right], \mathbf{e}_i\right\rangle = \sum_{i=1}^{N}\langle \mathbf{p}_k, \mathbf{e}_i \rangle. \tag{2.20}$$

Note that the second expression in (2.20) follows from the fact that $\sum_{i=1}^{N}\langle \mathbf{z}_k, \mathbf{e}_i\rangle = 1$. There-fore, the conditional expectation of $\mathbf{z}_k$ has the form

$$\hat{\mathbf{z}}_k = \frac{\mathbf{p}_k}{\sum_{i=1}^{N}\langle \mathbf{p}_k, \mathbf{e}_i\rangle}. \tag{2.21}$$

Following similar principles in [59], define the diagonal matrix

$$\mathbf{D}(X_k, Y_k) := \begin{pmatrix} d_1\,(X_k,\,Y_k) & & & O \\ & d_2\,(X_k,\,Y_k) & & \\ & & \ddots & \\ O & & & d_N\,(X_k,\,Y_k) \end{pmatrix} \tag{2.22}$$

with entries

$$\begin{aligned} d_i\,(X_k,\,Y_k) &= \bar{\lambda}_{k,i}^{(X)}\bar{\lambda}_{k,i}^{(Y)} \\ &= \frac{\exp\left\{-\frac{1}{2}\left[A_{k,i}^2 + B_{k,i}^2 - \left(X_k^2 + Y_k^2\right)\right]\right\}}{\kappa_{k-1,i}\,\nu_{k-1,i}}, \end{aligned} \tag{2.23}$$

where $A_{k,i} = \dfrac{X_k - \alpha_{k-1,i}X_{k-1} - \beta_{k-1,i}}{\kappa_{k-1,i}}$ and $B_{k,i} = \dfrac{Y_k - Y_{k-1} - \zeta_{k-1,i}}{\nu_{k-1,i}}$.

For $j = 1, 2 \ldots, N$, we define the following quantities:

$$\mathcal{G}_k^{s,j} = \sum_{n=1}^{k}\langle \mathbf{z}_{n-1}, \mathbf{e}_j\rangle\langle \mathbf{z}_n, \mathbf{e}_s\rangle \tag{2.24}$$

$$O_k^j = \sum_{n=1}^{k}\langle \mathbf{z}_{n-1}, \mathbf{e}_j\rangle \tag{2.25}$$

$$\mathcal{T}_k^{j,(X)} = \sum_{n=1}^{k}\langle \mathbf{z}_{n-1}, \mathbf{e}_j\rangle f(X) \tag{2.26}$$

$$\mathcal{T}_k^{j,(Y)} = \sum_{n=1}^{k}\langle \mathbf{z}_{n-1}, \mathbf{e}_j\rangle g(Y). \tag{2.27}$$

Equations (2.24) and (2.25) denote the respective number of jumps from state j to s and the amount of time that the process $\{\mathbf{z}_n\}$ occupies the state j up to time k. The quantities $\mathcal{T}_k^{j,(X)}$ and $\mathcal{T}_k^{j,(Y)}$ in (2.26) and (2.27) are auxiliary processes that depend on functions $f(\cdot)$ and $g(\cdot)$, respectively, of the observation process. In our empirical application, $f(\cdot)$ takes the form X_k, X_k^2 and $X_{k-1}X_k$ whilst $g(\cdot)$ takes the form Y_k, Y_k^2 and $Y_{k-1}Y_k$.

Let H_k be any scalar $\mathscr{F}_k$-adapted process; H_0 is $\mathscr{F}_0$ measurable. A filter for H_k is defined

as $E\left[H_k\middle|\mathscr{F}_k\right]$ and by the Bayes' theorem,

$$E\left[H_k\middle|\mathscr{F}_k\right] = \frac{\overline{E}\left[H_k\bar{\Lambda}_k\middle|\mathscr{F}_k\right]}{\overline{E}\left[\bar{\Lambda}_k\middle|\mathscr{F}_k\right]} = \frac{\overline{E}\left[H_k\bar{\Lambda}_k\middle|\mathscr{F}_k\right]}{\sum_{i=1}^{N}\langle\mathbf{p}_k,\,\mathbf{e}_i\rangle}. \tag{2.28}$$

Write $\gamma(H_k) := \overline{E}\left[H_k\bar{\Lambda}_k\middle|\mathscr{F}_k\right]$ and note that $\langle\overline{E}\left[H_k\mathbf{z}_k\bar{\Lambda}_k\middle|\mathscr{F}_k\right],\,\mathbf{1}\rangle = \langle\gamma(H_k\mathbf{z}_k),\,\mathbf{1}\rangle = \gamma(H_k\langle\mathbf{z}_k,\,\mathbf{1}\rangle) = \gamma(H_k)$, where $\mathbf{1}$ is a vector of 1's. Therefore, equation (2.28) becomes

$$E\left[H_k\middle|\mathscr{F}_k\right] = \frac{\gamma(H_k)}{\sum_{i=1}^{N}\langle\mathbf{p}_k,\,\mathbf{e}_i\rangle} = \frac{\langle\gamma(H_k\mathbf{z}_k),\,\mathbf{1}\rangle}{\langle\mathbf{p}_k,\,\mathbf{1}\rangle}. \tag{2.29}$$

Thus, $\forall 1 \leq s,\,j \leq N,\ k = 1, 2, 3\ldots$, we obtain the filters for the state of the Markov chain, number of jumps $\mathcal{G}$, occupation time O, and auxiliary process $\mathcal{T}$ as follows:

$$\mathbf{p}_k = \mathbf{\Pi}\,\mathbf{D}(X_k,\,Y_k)\,\mathbf{p}_{k-1} \tag{2.30}$$

$$\gamma(\mathcal{G}_k^{s,j}\mathbf{z}_k) = \mathbf{\Pi}\mathbf{D}(X_k,\,Y_k)\,\gamma(\mathcal{G}_{k-1}^{s,j}\mathbf{z}_{k-1}) + d_j\,(X_k,\,Y_k)\,\langle\mathbf{p}_k,\,\mathbf{e}_j\rangle\pi_{sj}\mathbf{e}_s \tag{2.31}$$

$$\gamma(O_k^j\mathbf{z}_k) = \mathbf{\Pi}\mathbf{D}(X_k,\,Y_k)\,\gamma(O_{k-1}^j\mathbf{z}_{k-1}) + d_j\,(X_k,\,Y_k)\,\langle\mathbf{p}_k,\,\mathbf{e}_j\rangle\mathbf{\Pi}\mathbf{e}_j \tag{2.32}$$

$$\gamma(\mathcal{T}_k^{j,(X)}(f)\mathbf{z}_k) = \mathbf{\Pi}\mathbf{D}(X_k,\,Y_k)\,\gamma(\mathcal{T}_{k-1}^{j,(X)}(f)\,\mathbf{z}_{k-1}) + f(X_k)\,d_j\,(X_k,\,Y_k)\,\langle\mathbf{p}_k,\,\mathbf{e}_j\rangle\mathbf{\Pi}\mathbf{e}_j \tag{2.33}$$

$$\gamma(\mathcal{T}_k^{j,(Y)}(g)\mathbf{z}_k) = \mathbf{\Pi}\mathbf{D}(X_k,\,Y_k)\,\gamma(\mathcal{T}_{k-1}^{j,(Y)}(g)\,\mathbf{z}_{k-1}) + g(Y_k)\,d_j\,(X_k,\,Y_k)\,\langle\mathbf{p}_k,\,\mathbf{e}_j\rangle\mathbf{\Pi}\mathbf{e}_j. \tag{2.34}$$

The derivations of Eqs. (2.30) – (2.34) follow similar steps and techniques applied in [59] or [25]. It is immediate from Eq. (2.29) that the normalised filter estimates of $\gamma(\mathcal{G}_k^{s,j})$, $\gamma(O_k^j)$, $\gamma(\mathcal{T}_k^{j,(X)}(f))$ and $\gamma(\mathcal{T}_k^{j,(Y)}(g))$ can be determined by summing the components of the vector expressions given in Eqs. (2.31)–(2.34) and then dividing each of them by the expression in Eq. (2.30).

2.3.3 Optimal parameter estimates

In this section, we present the optimal estimates for the parameters of our proposed integrated model using a dynamic maximum-likelihood approach. Since maximising the likelihood or log-likelihood functions is cumbersome for richer models such as the one in this chapter, the Expectation-Maximisation (EM) algorithm is utilised; see Elliott and Krisnamurthy [23] and Wu [86] for a review.

Here, we combine and extend the one-dimensional EM estimations in [59] and Tenyakov et al. [75] to our multivariate HMM setting, thereby obtaining the optimal model parameter

estimates below:

$$\hat{\pi}_{ji} = \frac{\hat{\mathcal{G}}_k^{j,i}}{\hat{O}_k^i} \tag{2.35}$$

$$\hat{\alpha}_i = \frac{\hat{\mathcal{T}}_k^{i,(X)}(X_l \cdot X_{l-1}) - \beta_i \cdot \hat{\mathcal{T}}_k^{i,(X)}(X_{l-1})}{\hat{\mathcal{T}}_k^{i,(X)}\left(X_{l-1}^2\right)} \tag{2.36}$$

$$\hat{\beta}_i = \frac{\hat{\mathcal{T}}_k^{i,(X)}(X_l) - \alpha_i \cdot \hat{\mathcal{T}}_k^{i,(X)}(X_{l-1})}{\hat{O}_k^i} \tag{2.37}$$

$$\hat{\zeta}_i = \frac{\hat{\mathcal{T}}_k^{i,(Y)}(Y_l) - \hat{\mathcal{T}}_k^{i,(Y)}(Y_{l-1})}{\hat{O}_k^i} \tag{2.38}$$

$$\hat{\kappa}_i^2 = \frac{\hat{\mathcal{T}}_k^{i,(X)}\left(X_l^2\right) + \alpha_i \cdot \hat{\mathcal{T}}_k^{i,(X)}\left(X_{l-1}^2\right) + \beta_i^2 \hat{O}_k^i - 2\alpha_i \hat{\mathcal{T}}_k^{i,(X)}(X_l X_{l-1})}{\hat{O}_k^i}$$
$$- \frac{2\beta_i \hat{\mathcal{T}}_k^{i,(X)}(X_l) - 2\alpha_i \beta_i \hat{\mathcal{T}}_k^{i,(X)}(X_{l-1})}{\hat{O}_k^i} \tag{2.39}$$

$$\hat{\nu}_i^2 = \frac{\hat{\mathcal{T}}_k^{i,(Y)}\left(Y_l^2\right) + \hat{\mathcal{T}}_k^{i,(Y)}\left(Y_{l-1}^2\right) + \zeta_i^2 \hat{O}_k^i - 2\hat{\mathcal{T}}_k^{i,(Y)}(Y_l Y_{l-1})}{\hat{O}_k^i}$$
$$- \frac{2\zeta_i \hat{\mathcal{T}}_k^{i,(Y)}(Y_l) - 2\zeta_i \hat{\mathcal{T}}_k^{i,(Y)}(Y_{l-1})}{\hat{O}_k^i}. \tag{2.40}$$

The proofs of Eqs. (2.35)–(2.40) are given in the Appendix (Supplementary Material). By Eqs. (2.11) and (2.14), we get

$$\begin{cases} \mu_i & = \dfrac{\beta_i}{1 - \alpha_i}, \\[2mm] \theta_i & = -\dfrac{\ln(\alpha_i)}{\Delta t}, \\[2mm] \sigma_i^2 & = \dfrac{2\theta_i \cdot \kappa_i^2}{1 - e^{-2\theta_i \cdot \Delta t}}, \\[2mm] \xi_i^2 & = \dfrac{\nu_i^2}{\Delta t}, \\[2mm] \eta_i & = \dfrac{\zeta_i + \frac{1}{2}\nu_i^2}{\Delta t}. \end{cases} \tag{2.41}$$

Now, we consider the statistical inference of evaluating the variability of the parameter estimates. Related works on this issue could be found in the results in Xi and Mamon [88] and Xiong and Mamon [91] where explicit formulae of the Fisher information for each parameter was derived. The inverse of the Fisher information is used to calculate the variance associated with the maximum-likelihood estimates. The sampling distribution of a maximum likelihood estimator is asymptotically normal; see Garthwaite et al. [33], for

example.

$$\mathcal{I}\left(\hat{\pi}_{ji}\right) = \frac{\hat{\mathcal{G}}_{k}^{j,i}}{\pi_{ji}^{2}} \tag{2.42}$$

$$\mathcal{I}\left(\hat{\alpha}_{i}\right) = \frac{\hat{\mathcal{T}}_{k}^{i,(X)}\left(X_{l-1}^{2}\right)}{\kappa_{i}^{2}} \tag{2.43}$$

$$\mathcal{I}\left(\hat{\beta}_{i}\right) = \frac{\hat{O}_{k}^{i}}{\kappa_{i}^{2}} \tag{2.44}$$

$$\mathcal{I}\left(\hat{\kappa}_{i}^{2}\right) = \frac{\hat{\mathcal{T}}_{k}^{i,(X)}\left(X_{k}^{2}\right) + \alpha_{i}^{2}\hat{\mathcal{T}}_{k}^{i,(X)}\left(X_{k-1}^{2}\right) + \hat{O}_{k}^{i}\beta_{i}^{2}}{\left(\kappa_{i}^{2}\right)^{3}}$$

$$+ \frac{2\alpha_{i}\beta_{i}\hat{\mathcal{T}}_{k}^{i,(X)}\left(X_{k-1}\right) - 2\alpha_{i}\hat{\mathcal{T}}_{k}^{i,(X)}\left(X_{k}X_{k-1}\right) - 2\beta_{i}\hat{\mathcal{T}}_{k}^{i,(X)}\left(X_{k}\right)}{\left(\kappa_{i}^{2}\right)^{3}}$$

$$- \frac{\hat{O}_{k}^{i}}{2\left(\kappa_{i}^{2}\right)^{2}} \tag{2.45}$$

$$\mathcal{I}\left(\hat{\zeta}_{i}\right) = \frac{\hat{O}_{k}^{i}}{\nu_{i}^{2}} \tag{2.46}$$

$$\mathcal{I}\left(\hat{\nu}_{i}^{2}\right) = \frac{\hat{\mathcal{T}}_{k}^{i,(Y)}\left(Y_{k}^{2}\right) + \hat{\mathcal{T}}_{k}^{i,(Y)}\left(Y_{k-1}^{2}\right) + \hat{O}_{k}^{i}\zeta_{i}^{2}}{\left(\nu_{i}^{2}\right)^{3}}$$

$$+ \frac{2\zeta_{i}\hat{\mathcal{T}}_{k}^{i,(Y)}\left(Y_{k-1}\right) - 2\hat{\mathcal{T}}_{k}^{i,(Y)}\left(Y_{k}Y_{k-1}\right) - 2\zeta_{i}\hat{\mathcal{T}}_{k}^{i,(Y)}\left(Y_{k}\right)}{\left(\nu_{i}^{2}\right)^{3}}$$

$$- \frac{\hat{O}_{k}^{i}}{2\left(\nu_{i}^{2}\right)^{2}}. \tag{2.47}$$

Suppose H and Θ are two vectors corresponding to the vectors of new and original model parameters, respectively. The Fisher information for H can be computed using the equation $\mathcal{I}(H) := -E\left[\frac{d^{2}}{dH^{2}}\log\left(\mathcal{L}(H)\right)\big|H\right]$. The derivation of the Fisher information involved in each estimator is straightforward; refer to the Supplementary Material. The results are summarised in Eqs. (2.42)–(2.47). The Fisher information for the original parameters can be calculated as $\mathcal{I}(\Theta) = \mathbf{J}^{\top}\mathcal{I}(H)\mathbf{J}$, where $\mathbf{J}$ is the Jacobian matrix with $(i, j)^{\text{th}}$ entry $J_{ij} = \frac{\partial H_{i}}{\partial\Theta_{j}}$. With the explicit Fisher-information expressions the variance for the estimates Θ readily follows, more specifically, $\mathcal{I}^{-1}(\Theta)$.

2.4 Numerical implementation

2.4.1 Preliminary analysis of data for implementation

The model is tested on a bivariate monthly time series covering the period Jan 1980 – Dec 2017. The Canadian FSI metric and data are based on [21] whilst the IPI data are compiled by Statistics Canada. There are 456×2 data points. The evolution of the data and the summary descriptive statistics are presented in Figure 2.1 and Table 2.1, respectively. The FSI is skewed to the right and has a large coefficient of variation of 1.003. The logarithm of IPI's is skewed to the left with a coefficient of variation of 0.045. The plot of the FSI shows a cyclical behaviour of the economy, shifting from high financial stress level to lower level frequently. This phenomenon provides support for using an OU process to describe the mean reverting feature of the underlying data series.
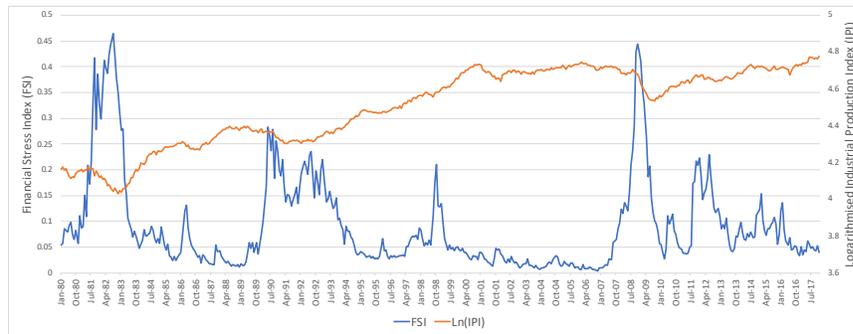
Figure 2.1: Evolution of FSI and logarithm of IPI.

Table 2.1: Descriptive statistics of FSI and log IPI

	Mean	Median	Minimum	Maximum	SD	CV	Kurtosis	Skewness
FSI	0.0909	0.0579	0.0042	0.4650	0.0912	1.003	3.8743	1.9548
Log(IPI)	4.4917	4.5486	4.0307	4.7441	0.2040	0.0454	-1.0316	-0.4954

Given its smaller variation range, the IPI is less volatile than the FSI. It is worth noting that it takes longer to transit from crisis state to a normal state. As both the mean and volatility of the IPI are dependent on the value of the index, but without apparent trends and presence of mean reversion, the geometric Brownian motion is an appropriate model for this index process. Equations (2.5)–(2.7) show that a Brownian motion with drift is obtained after applying a log transform on a GBM. The kurtosis of the FSI exceeds the normal distribution's by 3.8743; this signifies that the probability mass function of FSI is

Table 2.2: Least-squares parameter estimation on a particular data segregation.

FSI	Overall	Jan 82 – Dec 83	Jan 87 – Dec 88	Jan 08 – Dec 09	Jan 05 – Dec 06
		crisis	regular	crisis	regular
sample mean	0.0909	0.2563	0.0234	0.2152	0.0114
sample variance	0.0083	0.0222	0.0001	0.0153	0.0001
μ_{ls}	0.0931	0.2307	0.0998	0.0558	0.0355
θ_{ls}	2.1590	1.6771	2.1844	2.1654	2.2897
σ_{ls}^2	0.0022	0.0053	0.0047	0.0007	0.0006
Difference of log IPI	Overall	Jan 82 – Dec 83	Jan 87 – Dec 88	Jan 08 – Dec 09	Jan 05 – Dec 06
		crisis	regular	crisis	regular
sample mean	0.0014	0.002	0.0044	-0.0059	-0.0004
sample variance	0.0001	0.0003	0.0001	0.0002	0.0001
η_{ls}	0.0293	-0.0044	0.0413	0.0213	0.0347
ξ_{ls}^2	0.0017	0.0029	0.002	0.0013	0.0013

either concentrated around the mean with several values far from the mean or concentrated in the tails of the distribution; see Moors [62]. On the other hand, the logarithm of IPI has a negative excess kurtosis, which indicates that its probability mass function concentrates around the mean with fewer outliers relative to the normal distribution; see Westfall [84]. These facts inspire the use of a regime-switching model for the underlying bivariate data series; virtually, all distributional shapes can be reproduced by mixtures of distributions (normal in our case), which is intrinsic to the regime-switching approach.

Table 2.2 displays the descriptive statistics and possible segregations into two states based on the FSI values. To make the discussion more practical and the mathematics tractable, we assume 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 could persist over some time due to the weighted combination of volatilities under the above two regimes. The evolutions of FSI and IPI undergo several regimes characterised by different parameter values. This is supported by the estimates for parameters $\alpha, \beta, \kappa^2, \zeta$ and v^2 in the possible grouping periods obtained by using a least-squares method in each designated interval. Then, we recover the model parameters $\mu, \theta, \sigma^2, \eta$ and ξ^2 by the set of equations in Eq. (2.41) with $\Delta t = 1/12$.

The preliminary analysis results in Table 2.2 demonstrate possible segregation of the actual

data into different states in accordance with the values of the mean-reverting level, mean reversion rate and volatility for the FSI, and the values of percentage drift and percentage volatility for IPI. In particular, we see that the FSI has a low mean-reverting speed, high mean-reverting level and high volatility when the sample mean is high. When the sample mean of the FSI is low, the estimated mean reversion speed is high, and both the values of mean-reverting level and the volatility are low.

The logarithm of the IPI has a large percentage drift value when the sample mean of log returns grows. A decreasing percentage volatility in IPI could be observed due to IPI's increasing long-run trend, which makes the ratio between volatility and IPI fall in general even when the IPI plummets during an economic crisis.

2.4.2   Initial values for the parameter estimation

The implementation procedure starts by selecting initial values for the parameters. Several approaches have been employed by researchers to find initial parameters for the filtering algorithms (e.g., Xi and Mamon [88] and Erlwein et al. [29]).

In this chapter, we follow a method based on maximising the log-likelihood function, which was characteristically found to be efficient when searching for initial values by Tenyakov et al. [75] and Xiong and Mamon [91]. Model parameter estimation via the EM algorithm, producing expressions involving the HMM filters, extends the static maximum log-likelihood method. As noted in Tenyakov et al. [75], the usual log-likelihood maximisation approach could generate ideal initial parameters to fit the data well; consequently, this leads to an appreciable stability of the HMM filters.

Xi and Mamon [87] pointed out that, given a data set, if the initial values satisfy the model's assumptions, the estimated parameters will converge eventually to some stable level. However, the speed of convergence could be affected by the initial values. Based on our experiment, the initial values determined by the MLE method provide stability to our parameter estimation in less than 6 algorithm steps. This is very crucial for a small data set such as the data series in this chapter. On the other hand, if the initial values are invalid (i.e., inconsistent with model's supposition) ab initio or they are too far from the neighbourhood of a local maximiser, the implementation will fail or give results that are not interpretable.

In this chapter, we shall apply the maximum-likelihood method on the first 12 points of the

2-dimensional data set. The aim is to have reasonable initial estimates to achieve a stable model performance and relative ease of implementation.

When obtaining initial values, it is assumed that the estimates for the true parameter set $\Theta = \left\{ \pi_{ji}, \, \alpha, \, \beta, \, \kappa^2, \, \zeta, \, v^2 \right\}$ is homogeneous (i.e., the true model parameters stay the same for data subsets of any size) at least at the beginning. Based on the discussion in Subsection 2.2.4, the log-likelihood function is given by

$$L_{ini} = \sum_{i=1}^{N} \left(L_i^{(X)} + L_i^{(Y)} \right), \qquad (2.48)$$

where

$$L_i^{(X)} = \sum_{l=2}^{12} \left[-\log\left(\sqrt{2\pi}\kappa_i \right) - \frac{1}{2} \left(\frac{X_l - \alpha_i \cdot X_{l-1} - \beta_i}{\kappa_i} \right)^2 \right]$$

$$L_i^{(Y)} = \sum_{l=2}^{12} \left[-\log\left(\sqrt{2\pi}v_i \right) - \frac{1}{2} \left(\frac{Y_l - Y_{l-1} - \zeta_i}{v_i} \right)^2 \right].$$

Table 2.3: Initial parameter values for the 1-, 2-, and 3-state HMM-based models.

FSI	1-state HMM	2-state HMM		3-state HMM		
	-	State 1	State 2	State 1	State 3	State 2
μ_{init}	0.07683	0.10379	0.05401	0.09844	0.06327	0.05936
θ_{init}	2.27332	2.16877	2.36186	2.1895	2.32594	2.34111
σ^2_{init}	0.00013	0.00074	0.00009	0.00055	0.00015	0.00013
Difference of log IPI	1-state HMM	2-state HMM		3-state HMM		
	-	State 1	State 2	State 1	State 3	State 2
η_{init}	0.02276	-0.00241	0.03025	-0.00635	0.0196	0.0449
ξ^2_{init}	0.00263	0.00294	0.00254	0.00288	0.00287	0.00247

We first find the estimates of the parameters by simply maximising the log-likelihood function (2.48) under the assumption that the system always operates under a one-state setting (i.e., $N = 1$). We employ the function 'optim' in the statistical software R to solve the optimisation problem. The initial values for parameters of the single-state model are treated as benchmarks in generating the initial values for frameworks with more than one regime. All non-zero entries in the transition matrix are set to $1/N$ in the 2- or 3-state set ups. The results of the optimisation for the 1-, 2-, and 3-state HMM based models are exhibited in Table 2.3. In our case, all the initial values of $\left\{ \mu, \, \theta, \, \sigma^2, \, \eta, \, \xi^2 \right\}$ for the single-state model lie between the corresponding estimates for the two-regime-switching model.

2.4.3   Filtering procedure and results

The filters obtained in Subsection 2.3.3 were implemented under a data-point processing rather than a batch-processing scheme. Such a processing scheme is a generalisation of the typical procedure in the previous HMM filtering literature, namely the parameters are updated sub-optimally using group-data processing. More specifically, one complete algorithm step or algorithm run processes one bivariate data point. At the end of each algorithm step, new parameter estimates are calculated and they are utilised iteratively as starting values for the next algorithm step. Each data point in the bivariate data set is fed into the recursive filtering equations to obtain the best estimates, in the sense of conditional expectation, of various quantities of interest. The dynamics of the estimates are computed by first generating the estimates for parameters α, β, κ^2, ζ and ν^2. Then, the succeeding estimates are obtained through the set of equations in Eq. (2.41).

We have a couple of reasons why we adopt the point-by-point filtering method. First, both the FSI and IPI are monthly data. Thus, it is natural and rational to make the frequency of the filtering procedure compatible with the frequency of the data compilation. This compatibility is essential, without introducing any extra noise, to pin down the evolving major structural breaks. Second, the components of our bivariate time series have different dynamic features which makes it difficult to generate filtering results by processing a batch of data points obtained by a moving window without information loss. Although the data-point processing for our filtering method results to some variations in the model parameters estimates, the outputs contain all information about the FSI and IPI fluctuations. Our method captures well the joint effects of the bivariate time series accurately.

Our filtering procedure makes use of the initial parameters in Table 2.3. The data set contains 2 columns covering the FSI and log IPI. There are 432 time points considered in our filtering application. As pointed out in in Subsection 2.4.2, the first 12 bivariate data points (January – December 1980) are used to come up with the initial parameter values. The model performance is assessed using the remaining 444 monthly observations from January 1981 to December 2017. So, there are 444 algorithm steps in total with an algorithm step of 1 month.

The outcomes of the filtering-based parameter estimation under the one-state HMM framework are illustrated in Figure 2.2. All the parameter estimates converge to some definite levels at the end of the filtering process. The evolutions of the OU parameter estimates show some significant variations over time, especially at time windows of regarded eco-

nomic or financial-crisis episodes whilst the GBM parameters behave much smoother than the OU parameters do. The FSI's volatility, in particular, converges to the level of 0.02 with a periodic and fading amplitude patterns, which are consistent to FSI's unsettled nature. Note that the variation of parameters from 1981 to 1984 is ascribed mainly to the FSI's high volatility as well as to IPI's dynamics, tracing a locally 'valley-like' behaviour and manifesting an unambiguous trough during this period. The severe fluctuation of FSI drives the variation of the mean level and volatility of the OU process. The GBM's mean level follows the evolution of IPI, falling continuously from Sep 1981 to Jan 1983. As mentioned in Subsection 2.4.2, the parameters' initial values under the one-state parameter estimation can be used as benchmarks for assigning initial parameters when implementing the 2- and 3- state HMMs. We examined the movement of the parameter estimates under the 2-state HMM-based model. The behavioural patterns of the OU parameter estimates under a 2-regime setting are significantly different in each of the two states; in particular, there is no tendency for the two evolutions in any of the parameters to even coincide. In addition, the state-estimate outcomes for states 1 and 2 converge to different levels, and the filtering outputs for state 1 have inherently more fluctuations than those of state 2 in almost every crisis event.

The evolution of the GBM parameters' estimates converge to different levels with a behaviour that is more stable than the OU parameters'. This situation is in agreement with the result of the preliminary analysis on the possible data segregations illustrated in Table 2.2. Our results show that each of the parameter estimates in the 1-state HMM lies between the two estimates produced under the 2-state HMM; the estimates under both settings, as they unfold with time, exhibit similar patterns. This could be taken as indication of the data set's excellent fit with the 2-regime HMM. Furthermore, in comparison to the 1-regime HMM filtering, the 2-regime HMM filtering captures more effectively the major jumps in the FSI and log IPI. Therefore, a 2-state HMM model is worthy of consideration by regulators in the quest to mitigate potential financial crisis occurrences.
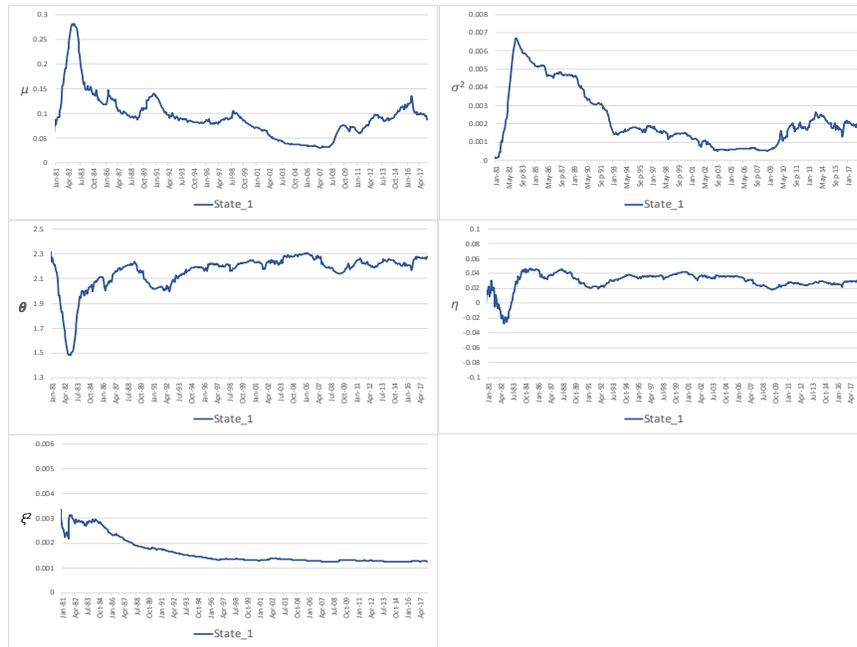
Figure 2.2: Evolution of parameter estimates for μ, θ, σ^2, η and ξ^2 under a 1-state HMM-based model.
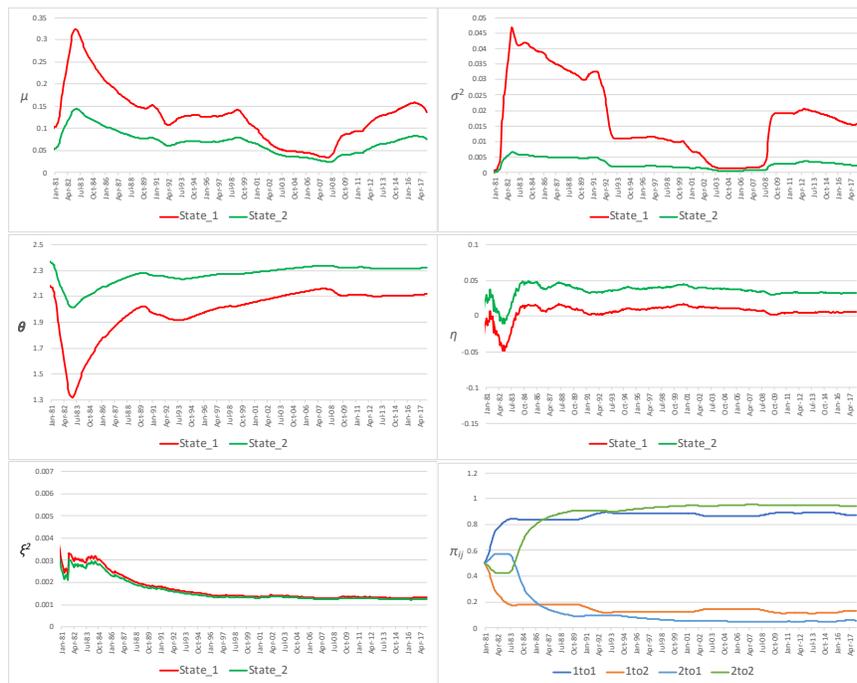
Figure 2.3: Evolution of parameter estimates for μ, θ, σ^2, η and ξ^2 under a 2-state HMM-based model.
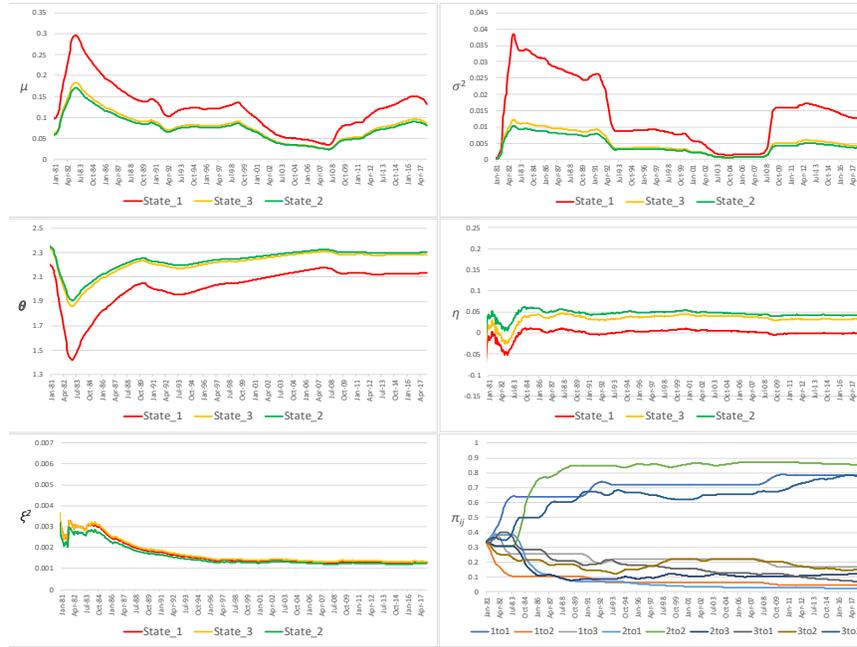
Figure 2.4: Evolution of parameter estimates for μ, θ, σ^2, η, ξ^2 and π_{ji} under a 3-state HMM-based model.

Filtered estimates under the 3-state HMM-based model are depicted in Figure 2.4. Noticeably, the evolution of the parameters in state 1 under the 3-regime setting is similar to that under the 2-state HMM. The parameters under the new added regime, state 3, evolve almost identically to the ones associated with state 2 that jibes with the state 2 under a two-regime setting. This suggests that a 2-state HMM is adequate to model the bivariate time series, which is in accord with the further statistical-analysis results from the 2-state HMM filtering in the next section.

2.4.4 Model selection and post-modelling diagnostics

As per the modelling formulations drawn up in Subsections 2.2.3 and 2.2.4, one-step ahead forecasts for X_k and Y_k under the 1-, 2- and 3- state HMMs can be expressed as the conditional expectation of the observation process at time $k + 1$ given the information set at time k, i.e.,

$$\begin{cases} E\left[X_{k+1} \mid \mathscr{F}_k\right] = E\left[\alpha(\mathbf{z}_k)X_k + \beta(\mathbf{z}_k) + \kappa(\mathbf{z}_k)\,w_{k+1} \mid \mathscr{F}_k\right] = \langle \alpha, \hat{\mathbf{z}}_k \rangle X_k + \langle \beta, \hat{\mathbf{z}}_k \rangle \\ E\left[Y_{k+1} \mid \mathscr{F}_k\right] = E\left[Y_k + \zeta(\mathbf{z}_k) + \nu(\mathbf{z}_k)\,b_{k+1} \mid \mathscr{F}_k\right] = Y_k + \langle \zeta, \hat{\mathbf{z}}_k \rangle. \end{cases} \qquad (2.49)$$
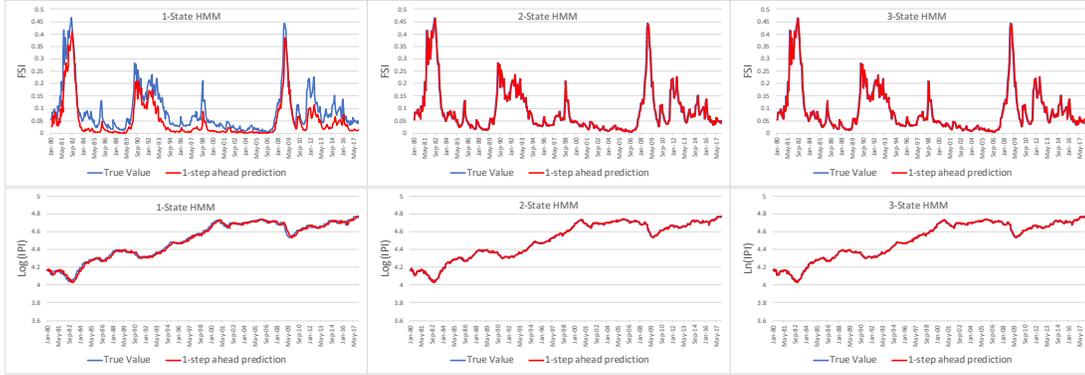
Figure 2.5: One-step ahead forecasts under the 1-, 2-, and 3-state HMMs.

Figure 2.5 plots the one-step ahead forecasts for the bivariate FSI and IPI series under the 1-, 2- and 3- state HMM settings. The 1-step forecast curves resemble superbly the dynamics of both the FSI and log IPI. In general, the prediction for log IPI outperforms (i.e., fits the original data points better) the one for the FSI. Moreover, in terms of the predicted values following the graph of the actual data, the 2- and 3-state HMMs are better than the 1-state HMM. This finding will be justified as well by the error analysis presented in the latter part of this section.

We perform an error analysis to evaluate the goodness of fit of different HMM settings. The criteria put forward in [29] and [92] are utilised, which include the root mean square error (RMSE), absolute mean error (MAE), relative absolute error (RAE) and mean absolute percentage error (MAPE). The notation h_j denotes the true value of our observational process at time j; $\hat{h}_j$ symbolises the one-step ahead prediction at time j, $\bar{h}$ stands for the sample mean of the underlying process, and n is the sample size. The RMSE, AME, RAE and MAPE for the one-step ahead predictions are calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{h}_j - h_j\right)^2}, \quad \text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|\hat{h}_j - h_j|,$$

$$\text{RAE} = \frac{\sum_{j=1}^{n}|\hat{h}_j - h_j|}{\sum_{j=1}^{n}|h_j - \bar{h}|}, \quad MAPE = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{\hat{h}_j - h_j}{h_j}\right|.$$

Table 2.4 displays the error-analysis results involving the 1-, 2- and 3-HMM settings. Prediction errors of log IPI under all HMM settings are generally much smaller than those of the FSI, which are expected considering that, in index level scale, FSI is less volatile than IPI. The error metrics showed that the 2-state HMM-based model outperforms the 1- and 3-state models in predictive accuracy for both FSI and log IPI even though the error

differences are very small.

Table 2.4: Results of error analysis.

	FSI			log IPI		
	$N = 1$	$N = 2$	$N = 3$	$N = 1$	$N = 2$	$N = 3$
RMSE	0.05155	0.02653	0.02653	0.01923	0.01066	0.01068
MAE	0.03990	0.01647	0.01649	0.01492	0.00824	0.00822
RAE	0.60414	0.24939	0.24966	0.08122	0.04487	0.04485
MAPE	0.59100	0.20643	0.20688	0.00334	0.00184	0.00184

Table 2.5: p-values for paired t-tests applied to RMSE, MAE, RAE, and APE values.

	FSI		
	2-state vs 1-state	3-state vs 1-state	2-state vs 3-state
RMSE	2.2829×10^{-06}	2.3908×10^{-06}	0.1504
MAE	1.7881×10^{-06}	1.4786×10^{-06}	0.2349
RAE	1.7783×10^{-06}	1.9797×10^{-06}	0.1783
MAPE	1.9168×10^{-06}	1.7256×10^{-06}	0.0962
	log(IPI)		
	2-state vs 1-state	3-state vs 1-state	2-state vs 3-state
RMSE	1.6498×10^{-07}	1.9827×10^{-07}	0.0827
MAE	1.0294×10^{-07}	1.6041×10^{-07}	0.0909
RAE	1.4086×10^{-07}	1.1196×10^{-07}	0.0662
MAPE	1.6662×10^{-07}	1.5956×10^{-07}	0.0797

We wish to determine whether the error mean differences are statistically significant in each pairwise HMM setting. A t-test, by applying the bootstrap method, is conducted on the 4 different error metrics spanning all possible paired HMM settings. In this chapter, the bootstrapped sample size is 10000. The outcomes of our pairwise t-test are reported in Table 2.5. For the 2-state HMM versus the 1-state HMM and 3-state versus 1-state HMM, all the p-values are smaller than 10^{-5} so that we can conclude the difference in the error metrics, after adding one or more regimes into the one-state model, is highly significant. For the 2-state HMM versus the 3-state HMM, all the p-values are greater

than 0.05, signifying that we cannot reject the null hypothesis of no difference. However, for a 5% significance level, both the 1- and 2-state HMM pair and 1- and 3-state HMM pair are statistically different in terms of their prediction errors. This result suggests that there is a recognisable benefit to using a regime-switching model whilst the 2- and 3-state HMM settings have similar forecasting capacity in our implementation. Hence, our 2-state HMM-based model is sufficient to capture the dynamics of the observed FSI and IPI.

Table 2.6: Number of estimated parameters under various HMM settings.

	HMM settings				
	1-state	2-state	3-state	$\cdots$	N-state
No. of parameters	5	12	21	$\cdots$	$3N + 2N + N(N-1)$

In choosing the best HMM set ups, we accompany our error analysis by a likelihood-based model selection analysis, such as the Akaike Information Criterion (AIC), as argued in [88]. This estimates the Kullback–Leibler information under the ML paradigm. The AIC metric is given by $AIC = -2L(\Theta) + 2c$, where c is the number of parameters to be estimated in the model as presented in Table 2.6. With $\Theta = \left\{ \pi_{ji}, \ \alpha, \ \beta, \ \kappa^2, \ \zeta, \ \nu^2 \right\}$, $L(\Theta)$ denotes the log-likelihood function that is custom-made for our HMM filterting procedure. In particular,

$$L(\Theta) = \sum_{i=1}^{N} \left[\langle \mathbf{z}_k, \ \mathbf{e}_i \rangle \left(L_i^{(X)} + L_i^{(Y)} \right) \right], \tag{2.50}$$

where

$$L_i^{(X)} = -\frac{1}{2} \log\left(2\pi\kappa_i^2\right) - \frac{(X_{k+1} - \alpha_i \cdot X_k - \beta_i)^2}{2\kappa_i^2}$$

$$\tag{2.51}$$

$$L_i^{(Y)} = -\frac{1}{2} \log\left(2\pi\nu_i^2\right) - \frac{(Y_{k+1} - Y_k - \zeta_i)^2}{2\nu_i^2}.$$

The AIC for each model is computed using the parameter estimates given at the end of each algorithm step. This means that we also obtain an AIC value after each algorithm run. The evolution of AIC values for the 1-, 2- and 3-state models after each algorithm step is portrayed in Figure 2.7. Given the form of the metric in (3.35), the model deemed the best and should be selected is the one that yields the lowest AIC value. In our case, the AIC increases considerably as the regime size grows; this arises from the substantial increase in the number of parameters especially due to the enlargement of the transition probability matrix's dimension. Even though the 1-state model brings in the smallest AIC
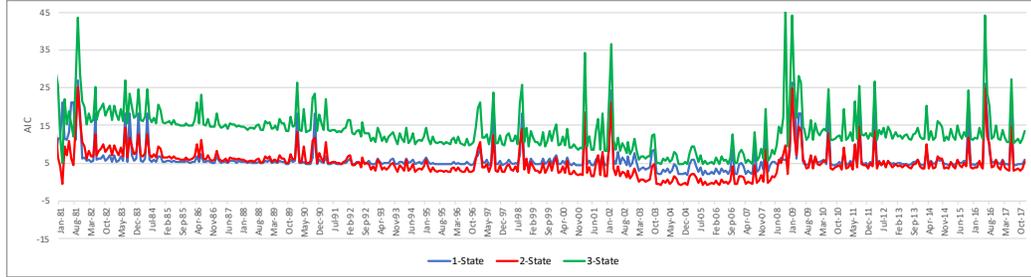
Figure 2.6: Evolution of AIC values for the 1-, 2-, and 3-state HMMs.

values for some of the steps, it is evident that a 2-state HMM continually sustains a low AIC value with a more stable pattern for the entire data set even during turbulent economic conditions. Therefore, the 2-regime HMM is the most suitable for modelling the bivariate data series under investigation having attained the greatest balance between model's fitness and complexity. This conclusion confirms and reinforces the previous results in our error analysis on the most appropriate modelling framework.

Table 2.7: AIC for all algorithm steps.

	$N = 1$	$N = 2$	$N = 3$
AIC Mean	6.2279	5.0032	13.3932
AIC Summation	2833.6962	2276.4655	6093.9106

2.4.5 Model diagnostics

We are interested on the statistical computing stability of our approach and the accuracy of parameter estimates. The bootstrap technique applied to the original data series will aid us in assessing these two points of interest. In particular, with the use of bootstrapped samples, we calculate the standard deviation (SD) for each parameter estimate through Eqs. (2.42)–(2.47). Ranges of tabulated SD over the entire algorithm steps for each parameter under the 2-state HMM are presented in Table 2.8 (see columns 4 and 5).

Following Visser et al. [82] concerning the inference on HMM parameter estimation, we employ the moving block bootstrapping method. The optimal moving block lengths for both the FSI and IPI time series are determined in accordance with Politis and Romano [66], Politis and White [67] and Patton et al.[65] using the R function "b.star" from"np"

Table 2.8: Parameter estimates' SD ranges: Bootstrap versus EM method.

	SD (Bootstrap)		SD (EM)	
	min	max	min	max
μ_1	0.06122	0.06322	0.06312	0.06447
μ_2	0.02618	0.02687	0.02668	0.02718
θ_1	0.07838	0.08658	0.08494	0.09154
θ_2	0.07294	0.07595	0.07543	0.07800
σ_1^2	0.01265	0.01288	0.01290	0.01315
σ_2^2	0.00166	0.00169	0.00167	0.00170
η_1	0.01583	0.01859	0.01576	0.01852
η_2	0.01423	0.01687	0.01412	0.01676
ξ_1^2	0.00070	0.00076	0.00072	0.00077
ξ_2^2	0.00059	0.00063	0.00061	0.00064
π_{11}	0.07109	0.07871	0.07048	0.07778
π_{12}	0.07109	0.07871	0.07048	0.07778
π_{21}	0.07508	0.07779	0.07707	0.07938
π_{22}	0.07508	0.07779	0.07707	0.07938

package. The respective optimal block lengths for FSI and IPI series turn out to be 30 and 41. We take the maximum of the two (i.e., 41) as the optimal value in bootstrapping the bivariate time series, which aims to capture the dependence and correlation properties.

We generate 10,000 bivariate time series samples from the original data using the R function "tsboot" from "boot" package. The 2-state EM-based HMM estimation is then performed on the bootstrapped samples. As well, the optimal parameter estimates utilising the HMM filtering scheme on the original data series are obtained, and their SDs using the Fisher information are generated; similar idea was adopted in Xi and Mamon [88]. For the bootstrapped samples, the ranges of the SDs over the algorithm steps are reported in second and third columns of Table 2.8. We observe that the ranges' upper and lower bounds of the bootstrap-based SDs under the HMM filtering method are approximately close to those computed under the EM algorithm. This is a strong evidence supporting that our approach's statistical computing aspect is stable.

Table 2.9: t-test results for prediction performance metrics in bootstrapped samples: HMM versus DCC-GARCH and HMM versus VAR.

Data	Benchmark models	Metric	Mean (HMM)	Mean (Benchmark)	Mean difference (HMM-Benchmark)	95% Confidence interval	
FSI	DCC-GARCH	MAE	0.01644	0.01814	-0.00170	-0.00175	-0.00164
	DCC-GARCH	MAPE	0.20618	0.22733	-0.02115	-0.02172	-0.02059
	DCC-GARCH	RAE	0.24919	0.27466	-0.02547	-0.02616	-0.02479
	DCC-GARCH	RMSE	0.02402	0.02865	-0.00463	-0.00472	-0.00454
	VAR	MAE	0.01644	0.02025	-0.00381	-0.00388	-0.00373
	VAR	MAPE	0.20618	0.25423	-0.04806	-0.04892	-0.04719
	VAR	RAE	0.24919	0.30828	-0.05909	-0.06014	-0.05805
	VAR	RMSE	0.02402	0.03236	-0.00833	-0.00846	-0.00821
Log(IPI)	DCC-GARCH	MAE	0.00793	0.00905	-0.00112	-0.00116	-0.00109
	DCC-GARCH	MAPE	0.00184	0.00200	-0.00016	-0.00016	-0.00015
	DCC-GARCH	RAE	0.04488	0.04863	-0.00375	-0.00388	-0.00362
	DCC-GARCH	RMSE	0.01036	0.01147	-0.00112	-0.00115	-0.00109
	VAR	MAE	0.00793	0.01014	-0.00221	-0.00225	-0.00217
	VAR	MAPE	0.00184	0.00217	-0.00033	-0.00034	-0.00033
	VAR	RAE	0.04488	0.05269	-0.00781	-0.00798	-0.00764
	VAR	RMSE	0.01036	0.01218	-0.00182	-0.00186	-0.00179

It has to be noted that our EWAS was constructed specifically to accommodate the HMM-filtered state estimate as a signal. Thus, it is not feasible to compare the EWAS's prediction performance under our HMM-based framework (with two states referring to crisis and non-crisis regimes) and those under the one-state models used as benchmarks. However, we compare the forecasting of FSI and IPI levels under our proposed model and two benchmark models. For each bootstrapped sample, one-step ahead predictions were generated using the HMM filtering process on all algorithm steps. We asses prediction performance, in terms of four error metrics, between HMM and two other benchmarked multivariate time series models: Dynamic Conditional Correlation-Generalised Autoregressive Conditional Heteroscedasticity (DCC-GARCH)(1,1) and Vector Autoregressive(VAR)(1). We make use of the R functions "dccfit" and "VAR" to build the two benchmark models. For each comparison (i.e., HMM versus DCC-GARCH(1,1) and HMM versus HMM versus VAR(1)), a t-test is conducted to ascertain the significance of the mean difference in the prediction error measurements between HMM and the model for benchmarking.

The estimated means, mean differences and corresponding 95% confidence intervals are reported in Table 2.9. We observe that the 95% confidence intervals are bounded away

from zero. This imply evidently that, based on the bootstrap experiment, the one-step-ahead prediction errors of the HMM framework are significantly smaller than those of other two benchmark models.

2.4.6 Early-warning detection capacity for financial crisis

To draw attention to the practical utility of this work, we use dated historical economic and financial crisis episodes to probe the proposed model's capacity of spawning an early-warning system under the 2-state HMM setting. In this research work, the pre-crisis stage is assumed to be totally captured by the HMM-state process only, and not by any other model parameters. We shall then delve into the relationship between historical events designated as financial crises and the HMM state estimate in an effort to devise some meaningful pre-crisis signalling system. We set forth an intuitive but effective approach that seizes the key features of a variable designed to emit warning signals before or at least at the early stage of the crisis events.

Let ϵ_k be the component related to "crisis" regime of the Markov chain $\hat{\mathbf{z}}_k$; that is, $\epsilon_k = \langle \hat{\mathbf{z}}_k, e_1 \rangle$. Write $\Delta \epsilon_k := \epsilon_k - \epsilon_{k-1}$. Define $a_k = \epsilon_{k-1} - z_r s_k / \sqrt{p}$ and $b_k = \epsilon_{k-1} + z_r s_k / \sqrt{p}$ as alarm thresholds for ϵ_k, where s_k is the moving sample SD of the data set $\{\epsilon_{k-p+1}, \epsilon_{k-p+2}, \ldots \epsilon_k\}$. In this study, we set $p = 12$; that is, the sample SD was calculated employing 12 data points; this choice is mainly due to economic data that are typically of low frequency and possessing long memory. Let $\{A_k\}$, $\{B_k\}$ and $\{C_k\}$ be sequences of events such that $C_k = A_k \cap B_k$ with $A_k = \{\epsilon_k \notin [a_k, b_k] \text{ or } \epsilon_k \geq 0.5\}$ and $B_k = \{\Delta \epsilon_k \geq 0\}$.

In our set up, we have dynamic thresholds. Such a set up is a substantial extension to that in Tenykov et al. [75] in which a static criterion to detect liquidity crisis was introduced. The manner of defining the indicators A_k and B_k is akin to capturing the stochastic features of crisis regimes, modelled by a Markov chain, before or at the beginning of historical-crisis events. In other words, the indicators capture not only the current but also the moving trend of the economy. Of course, these could be adjusted by practitioners according to varying objectives in getting early alerts effectively. The conditions instituted for an EWAS hinge on clinching the synchrony between its outcome and the historical dated crisis data. From the perspective of translating the indicators into economic intuition, they could be viewed as monitoring quantities that detect the latent variable's anomalous behaviour attributed to certain crisis episodes.

The setting of A_k is based on Tenykov et al.'s work [75] that keeps track of the probability of a Markov chain being in a crisis regime. In this work, the alert is also triggered when the probability of a Markov chain being in a crisis regime is greater than 0.5. When $\epsilon_k > 0.5$, the HMM indicates that the economy is in a crisis state. The 0.5 threshold alludes to the fact that the 2-state HMM filter is regarded as a binary classifier, which usually sets 0.5 as a default but adjustable benchmark to determine whether a subject is in one category or not.

We define our early warning alarm signal (EWAS) at time k as

$$\text{EWAS}_k = \frac{1}{q} \sum_{i=k-q+1}^{k} m_i, \tag{2.52}$$

where

$$m_i = \begin{cases} 1 & \text{if } C_i \neq \emptyset \\ 0 & \text{if } C_i = \emptyset \end{cases}, \quad i = k - (q-1), \ k - (q-2), \dots, k.$$

Clearly, EWAS as defined in (2.52) relies on HMM state estimates, which incorporate the filtered information from the bivariate time series and measures optimally the intensity (via the Markov chain's transition matrix) and frequency (via the Markov chain's occupation time) of certain stressful events over a specific time window. Note that r and q are two adjustable variables in the warning-signal setting. The quantity r corresponds to the width of the rolling confidence intervals for ϵ_k given $\mathscr{F}_{k-1}$. Specifically, it is the probability level corresponding to the standard score z. Assuming other parameters are fixed, small r values will make the warning process more sensitive so that it probably generates more false alarms (Type-II error) whilst large r values may lead to failure in signalling some crisis (Type-I error). The quantity q equals the number of data points covered by the moving window. Large q values will make the signal process long memory-dependent, which will possibly cause false warning alarm (Type-II error). Small q values will lead to a failure of generating alarm for a crisis at its early stage (Type-I error). In fact, q is the memory length for which the system has the sufficient count of the number of events triggering an alert. It is a hyper-parameter of the alarm system which needs to be set before fitting model to the data set instead of being learnt from the given data series. Since our purpose is to detect the anomaly related to crisis events, one natural consideration is to calibrate the value q against the historical crisis event. In practice, one could calibrate this quantity depending on one's preference for the trade off between Type I and II errors.

Suitable conditions instituted for EWAS hinge on its ability to clinch the synchrony between its outcome and the historical dated crisis data. When $\epsilon_k > 0.5$, the HMM indicates that the economy is in a crisis state. The lower bound a_k serves to identify the starting point

of a crisis as well as the ending point or point of recovery from a crisis.

We demonstrate our early warning detection method by using a set of dated Canadian historical systemic financial stress events. This set was obtained from a survey that sought inputs from senior Bank of Canada policy makers and economists to establish a consensus on which events have been the most stressful for the Canadian markets over the past 25 years and for what reasons. Respondents included a former governor, 3 governing council members, 8 senior bank officers, 12 bank officers, and 3 analysts; see [49] for the details of survey methodology. Note that the survey data in [49] only cover the period Jan 1980 – December 2001, and this work provides an updated set of dated financial-stress events. In particular, we enlarge the events set by adding other commonly viewed systemic financial stress events viz. sub-prime crisis, Bear Stearns scandal, Lehman Brothers collapse, 2008 recession, second bailout in Greece, Taper Tantrum 2013, NYMEX (WCC) oil price below $40 CAD for the first time, and the 2015 recession.

In our implementation, we first set $r = 0.999$ to minimise the chance of triggering false alarms due to small fluctuations in the HMM state ϵ_k. Second, we consider $q = 1, 2, \ldots, 10$. For each fixed q, we then compute the average values of EWAS in the neighbourhood (± 6 months) of every crisis event and for all regular (non-crisis) intervals. In the context of aiding the identification of crisis episodes, the purpose of such an average calculation is to associate the EWAS with a q value that produces the least number of false alarms. In other words, we look for a q that brings forth a large average signal intensity in the crisis-episode neighbourhoods and a small average signal intensity in the non-crisis intervals.
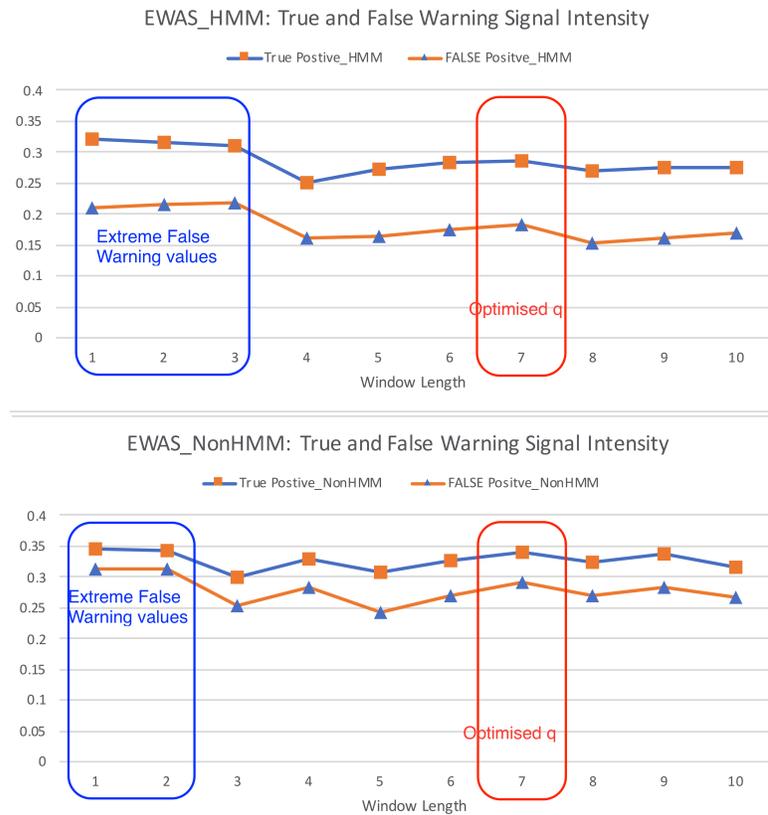
Figure 2.7: EWAS True/False warning intensity comparison: HMM vs. non-HMM

To delineate the advantage of EWAS gained from our HMM-based approach, we also apply exactly the same method to extract EWAS directly from the FSI and IPI data with the best window length value q. Figures 2.7 to 2.8 exhibit the EWAS computed with different q values. From Figure 2.7, it is evident that the True/False warning intensity of HMM-generated EWAS is significantly greater than that of the non-HMM-generated EWAS, which is in turn confirmed by the box plot in Figure 2.8. The non-HMM-generated EWAS, derived directly from FSI and IPI, are simply too noisy and ineffective in differentiating between crisis and regular regimes.

Our aim is to find an appropriate memory window length q such that the EWAS renders the best performance in distinguishing Crisis/Regular regimes. The ideal EWAS should have higher signal intensity on crisis events and relatively lower value over normal time periods. Our procedure in carrying out the determination of optimal q is as follows. (i) We assess the frequency distributions of False warning signal intensities generated from HMM and non-HMM EWAS (e.g., Figures 2.8 and 2.9). (ii) The extreme values are excluded;
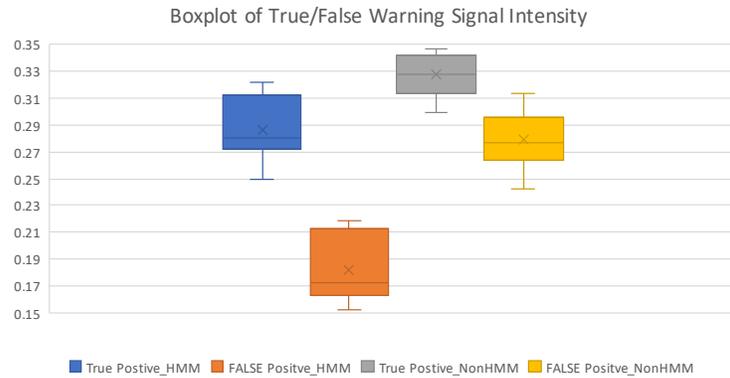
Figure 2.8: Box plot of HMM versus non-HMM EWAS True/False warning intensities

Figure 2.9: Histograms of HMM and non-HMM EWAS False warning intensities

these are the outliers, which are highlighted in orange in Figure 2.9. (iii) Finally, from the remaining data, we find q that bears the largest True-warning signal intensity. Based on the above procedure, we find that both the HMM EWAS and non-HMM EWAS have the commonly optimised window length of $q = 7$.

Figure 2.10 displays the HMM and non-HMM EWAS based on FSI and and IPI data. The non-HMM EWAS generates more noise making crisis-event detection unreliable. The time periods of historical real-crisis episodes are represented by light-red coloured rectangular areas whilst the financial-stress events are denoted by light-green coloured rectangular areas. We observe that the $EWAS_k$ metric (plotted as red curves) is capable of generating different alarm levels at a reasonably early stage of these dated prior crisis incidents and making very few Type-II errors. This supports that our EWAS under the 2-state HMM in conjunction with the filtering technique has good predictive performance in crisis forecasting.
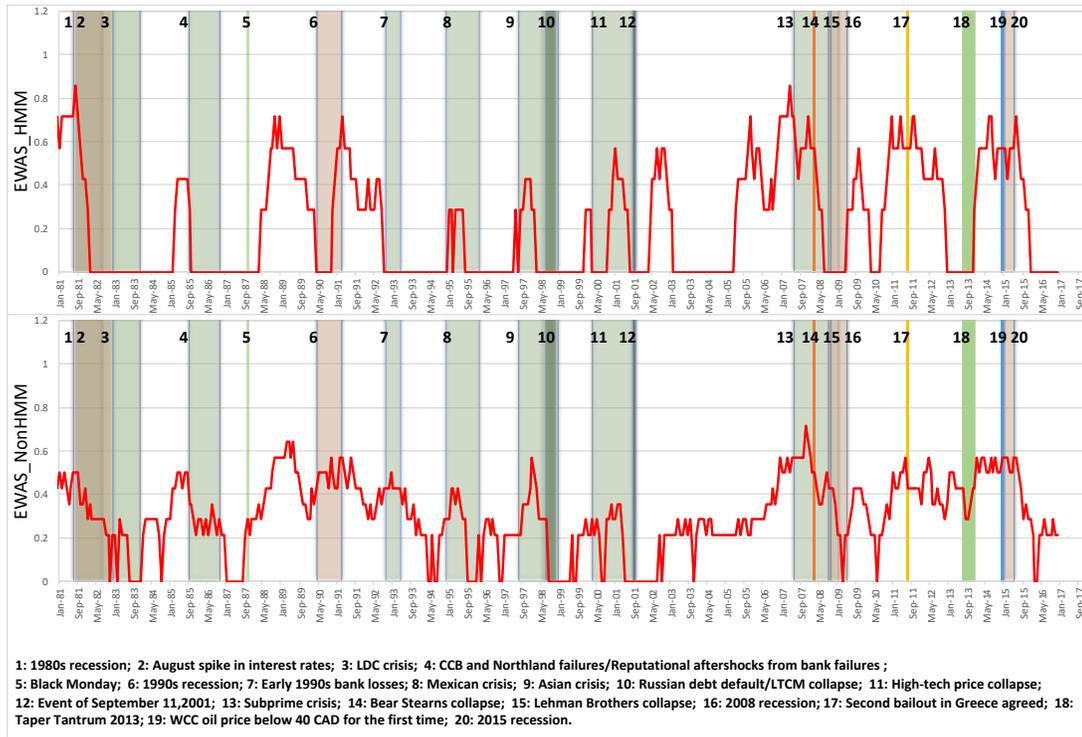
Figure 2.10: Outcome comparison between HMM and non-HMM EWAS

2.5 Conclusion

The development of a hybrid bivariate regime-switching HMM-based model to evaluate the level of systemic financial stress is the major contribution of this chapter. Our proposed model, which came out as a synthesis of an OU process and a GBM, is capable of capturing the main characteristics of the FSI and IPI. The parameters of this sysnthesised model are simultaneously governed by a hidden Markov chain in discrete time. The extraction of beneficial information from the observed bivariate stochastic process was elaborated in our methodology together with the model's one-step ahead predictions. Both the EM algorithm and change of reference probability measures smoothly link to produce a new set of multivariate HMM online recursive filters.

The actual bivariate time series data on Canadian monthly FSI and IPI compiled during a 37-year period were considered to examine the model's implementability and forecasting capability. To capture the dynamics of the underlying data set continually, we execute the application of the recursive filtering algorithm with data-point processing. Under the

1-, 2- and 3- state HMM settings, the one-step ahead predictions are calculated. Multi-regime HMMs have substantially smaller forecasting errors than those produced by the 1-state HMM in our post-modelling assessment. Our AIC comparison in conjunction with the one-step ahead forecasts' goodness-of-fit analysis justifies the choice of the optimal number of states. It is worth nothing that a 2-regime model surpasses the 1-regime model significantly in terms of performance under some criteria. This tells us that the 2-state HMM provides the most reliable framework for the data that we investigated.

By the block bootstrapping technique, statistical inference involving parameter estimates was carried out to assess the stability performance of our approach. The comparison with two benchmark multivariate time series models showed that our proposed model has better prediction accuracy in so far as obtaining the one-step ahead predictions. In addition to the statistical-inference-based validation, we closely examined the consistency of our model's results and performance against actual financial-stress events. An empirical pre-crisis detection method was established and tested on dated crisis episodes in economic history. We found that the proposed 2-regime HMM filters enabled fairly well the generation of early-warning signals for financial crises.

In several ways, this work reaffirms and boosts the merits of the HMM modelling as follows: (i) a new model (i.e., HMM-embedded OU-GBM framework) for bivariate data series that captures and describes occurrence of systemic financial stress episodes, (ii) an extended self-tuning estimation procedure that flexibly permits each univariate series component to have different dynamics, (iii) implementation with ample details of the newly derived filters together with suitable model validation metrics, and (iv) an early-warning system primarily relying on the HMM-state estimates calibrated from empirical data.

We provided a quantitative methodology for financial-stability modelling in which the estimation and implementation are designed to capture the effects and identification of financial crisis-events location in the context of drafting regulatory measures and oversights. Evolving parameter estimates of models tailored for systemic financial risk management are computed using our filtering recursions. The regular/crisis-regime classification that was elaborated by our regime-switching modelling approach accurately correspond to those already established and accepted by practitioners. The ease of interpretation, with direct relevance to bankers, economists, regulators and policy makers, is indeed another attractive hallmark of the regime-switching results.

Chapter 3

An analysis and forecasting of financial market liquidity regimes

3.1 Introduction

3.1.1 Background of market illiquidity

Keynes [53] characterises an asset as *liquid* "if it is more certainly realisable at short notice without loss"; that is, a liquid asset can be traded quickly and at a low cost [36]. By extension, if large quantities of an asset can be traded with a small impact on its price, the market for that asset is said to be liquid [76]. Thus, market liquidity reflects the trade-off between the speed for which some underlying asset can be traded and the variation in its price.

Brunnermeier and Pedersen [9] argued that market liquidity depends on trader's availability of funding. Conversely, traders' funding liquidity relies on assets' market liquidity. Gromb and Vayanos [38] linked market liquidity to the capital of financial intermediaries and their funding constraints. The general literature indicates that investors are subject to liquidity shocks and could realise gains from trade across segmented markets by trading with intermediaries who exploit price discrepancies to supply liquidity to investors. As per [38], shocks to asset prices that trigger capital losses tighten funding constraints and force intermediaries to reduce their positions. This then lowers market liquidity and amplifies the shocks. In [69], it is pointed out that the liquidity characteristics of the stock market,

all of which show the areas where the market has placed its current money, are considered the most important that affect the probability of a big gain or loss in the stock price.

Brunnermeier and Pedersen [9] proposed a model that links an asset's market liquidity and traders' funding liquidity (i.e., the ease with which they can obtain funding). The model explains the properties of market liquidity: (i) can suddenly dry up (fragility), (ii) co-moves across assets, (iii) is correlated related to volatility, (iv) is subject to "flight to quality", and (v) moves with the market.

Historically, TED spread, the difference between the risky LIBOR rate and the risk-free US Treasury bill rate, is viewed as a measure of market-liquidity level. When the market is illiquid, banks want to get first-rate collateral, which makes holding Treasury bonds more attractive and this then decreases the Treasury bond rate. On the other hand, banks charge higher interest for unsecured loans to compensate for the credit risk, which pushes up the LIBOR rate. Therefore, TED spread widens in times of liquidity crises; see Brunnermeier [8].

Illiquidity can be linked to a number of aggregate variables [80]. Huberman and Halka [45] illustrated that negative market returns and increased market volatility are followed by an increased level of illiquidity. Bid-ask spread is an intuitive and popular measure of illiquidity; it is defined as the difference between the quoted ask and bid prices. Tenyakov et al. [76] pointed out that the S&P 500 bid-ask spread metric is able to capture the illiquidity episodes where TED spread failed to do so. In this study, we investigate the aggregate spread index based on the evolution of the S&P 500, which as stated in [76] contains at least in part the liquidity level of the stock market.

The second key variable that we investigate is the volatility index (VIX). The Chicago Board Options Exchange (CBOE), defines it as an up-to-the-minute market estimate of expected volatility that is calculated based on real-time S&P 500 index option bid-ask quotes. More specifically, the VIX provides an instantaneous measure of the future degree of volatility and market uncertainty. The VIX has been utilised to gauge the level of investors' risk aversion or market sentiment; see Brunnermeier et al. [10] and Bekaert et al. [4], and it has also a negative relation with stock returns as documented by Giot [35] and Whaley [85].

A positive correlation between illiquidity and volatility is another finding in many studies in the literature. Kyle [55] and Karpoff and Walkling [52] measured the illiquidity of stocks

assuming that it has a positive correlation to price volatility and a negative one to the price level, market capitalisation, and number of shares. Vayanos and Wang [80] explained that liquidity suppliers who trade high-volatility assets are exposed to more risk and possibly to situations with more asymmetric information. Therefore, they require a larger price movement to absorb liquidity shocks, which means that these shocks have larger price impact and cause larger transitory deviations between price and fundamental value. Hauser and Kedar [44] showed that liquidity facilitates a large share redistribution across agents causing changes in average risk aversion, which increases Sharpe-ratio variability, and hence, stock return volatility. Tenyakov et al. [76] found that VIX is capable of capturing some periods of market illiquidity that were not picked up by the TED spread.

VIX contains market-sentiment information which is a joint effect from the stock and FX markets. For instance, Smales and Kininmonth [71] provided empirical evidence on the existence of of a relationship that links stock market returns, which are related to investor sentiment, with the FX markets. It is pointed out in [71] that currency investments tend to depreciate when there is an increased in investors' fear, which is linked to the financial system's liquidity severely affected for instance by the 2008 collapse of Lehman Brothers and the 2012 European sovereign debt crisis. Zapatero [93] illustrated that in fully integrated markets, the volatility of the exchange rate is explained by the volatility of the stock markets of the two countries concerned. Thus, understanding the behaviour in the FX market is important to measure market illiquidity. Therefore, the US Dollar Index (DXY) – a measure of the value of the US dollar relative to the value of a basket of currencies of the majority of the US's most significant trading partners – is examined and included as an important indicator in our financial-stability analysis; for a further discussion of DXY, see ICE: U.S. Dollar Index Contracts. The evolution of DXY is able to mirror market illiquidity as it contains information regarding investors' expectation of the FX market.

3.1.2   HMM-based regime-switching models

Shifting structural regimes is a widely observed phenomenon in financial economics. Brunnermeier and Pedersen [9] described the relationship between volatility and speculators' positions through the so-called "margin spiral", which is a characterisation of a market that switches between two equilibriums (or regimes): one is the low-liquidity equilibrium with high volatility and reduced positions, and the other is the high-liquidity equilibrium with low volatility and increased positions. In [8], it is shown that the economy has a

"self-stabilising effect" liquidity feature that is either a "loss" or a "margin".

A regime-switching-based technique is suited for modelling an evolving economic environment assumed to switch amongst different regimes. Such a technique has the capability of pinning down changes in economic states by allowing model parameters to change stochastically. Ichiue and Koyama [48] proposed a regime-switching model to probe how exchange rate volatility and depreciation of low-interest-rate currencies are related to each other. Afonso et al. [1] applied a time-varying parameter modelling approach to determine the shifts in the pricing regime in the sovereign bond markets of the eurozone area. A regime-switching mechanism could be embedded in models as shown in Duprey et al. [21] for identifying states in the individual or joint dynamics of the data series in conjunction with the detection of financial and business-cycle turning points. In regime-switching models, the use of a Markov chain is appropriate because of the ease in interpreting results and the capacity to model adequately the random-time occurrence of the switching. A Markov-switching methodology was popularised in Hamilton's research [41], where model estimation is addressed, albeit in a static sense, and a structural approach is put forward to (i) distinguish states of the economy and (ii) infer the probability of of an expansion state or contraction state. Employing a Markov-switching framework, Davig and Gerlach [18] studied the response of stock prices to Federal Reserve policy shocks . Gerdrup et al. [34] developed a financial model with a shock component driven by a Markov chain.

The efficient and accurate recovery of parameters for Markov-switching models using data is of utmost consideration for a successful implementation. Within the HMM-based framework, Elliott et al. [25] pioneered the change-of-measure method that yields recursive filters leading to a self-calibrating model; see [59] for a pedagogical introduction of this method. In particular, an HMM is a doubly-embedded stochastic process consisting of a series of observed values deemed to contain the underlying unobserved Markov chain typically with an undetermined number of states and transition probabilities to be estimated. In the current literature, the HMM filtering is tailored to certain applications in quantitative finance, insurance, economics, epidemiology, and other branches of the sciences and engineering. In finance specifically, these include short-rate modelling (e.g., Elliott et al. [26], Elliott and Mamon [24], Erlwein and Mamon. [27], and Xi and Mamon [88]); investment strategies (Tenyakov et al. [74] and Erlwein et al. [28]; commodity price forecasting (Date et al. [17]); weather derivative pricing (Xiong and Mamon [91] and Xiong and Mamon [92]); and modelling electricity spot prices (Erlwein et al. [29]).

3.1.3   Motivation

In Tenyakov et al. [76], a multivariate HMM framework is introduced to assess and fore-cast the market liquidity regime. Augmenting similar and related objectives in [76] and extending their theoretical developments, this work puts forward a modified multivariate regime-switching HMM framework to extract early-warning signals for possible liquidity crisis occurrences using certain indicators and new metrics. Considering that significant variation in illiquidity arises not only within asset classes but also across classes [80], we shall include four proxies: Treasury-Eurodollar spread (TED), Volatility Index (VIX), US Dollar Index (DXY), and S&P 500 bid-ask spread. The state of the liquidity characteristics could be inferred from the money, stock, and foreign-exchange markets with the use of these indices. Consequently, the outcome of this study is beneficial for the regulators who must adjust policies as well as for the investors who strategise their their investment positions taking into account the occurrence of market illiquidity episodes.

Noting that TED, VIX and DXY have a common mean-reverting feature, their movements are then modelled by an OU process. We use a geometric Brownian motion (GBM) to describe the S&P 500 bid-ask spread's dynamics on the basis the long-term evolution of its sample path. In essence, we propose a hybrid multivariate stochastic process apt to simultaneously describe the main features (e.g., seasonality, randomness, mean-reversion and jumps) of the joint dynamics of the four indices. Optimal estimates of various parameters will be 'filtered out' from the market information; that is the 'true' state of the Markov chain, probabilities of transition and all other parameters of our proposed multivariate model for 4 indices that mirror how liquidity regimes evolve in discrete time.

In this chapter, we shall demonstrate the change of probability measure technique in obtaining optimal filters of various quantities pertinent to the calculation of the multivariate model's parameters. This constitutes the adaptive processing of four-dimensional market signals revealing information about the states of the market liquidity. A self-tuning algorithm is generated, which updates parameters with the continual arrival of new observed signals coming from the time series of four indices.

Our methodology expands the HMM multivariate results of [25] and [76] in the sense that each univariate series that forms the entire multivariate set up has different dynamics specification; thus, the modelling framework allows for more flexibility. It retains the advantage of avoiding the forward-backward algorithm that is prevalent in most filtering techniques, thereby entailing much less memory during computation. In [76], it is remarked that other

filters (e.g., Hamilton-type filters) are computationally intensive to implement because they are based on static algorithms requiring full reruns involving the original data set every time a few data points are added. In our case, we successfully bypass this issue through the recursive filtering equations that keep updating estimates in real time.

Relative to the current state of market-liquidity modelling and analysis, the highlights of our research contributions are as follows. (i) Our proposed approach combines OU and GBM processes in an integrated HMM framework. This is a new attempt to capture jointly the important stylised properties of a multivariate data series. (ii) Estimation and inference under the new formulated setting are addressed with empirical illustrations. (iii) Finally, we come up with a new metric suited for an HMM-based early warning alert system.

The remaining parts of this chapter are structured in the following way. Section 2 presents the formulation of the 4-dimensional hybridised stochastic process with a discrete-time hidden Markov chain governing the model parameters. In Section 3, we perform a change of probability measure in establishing recursive filters for quantities that are functions of HMM, and then carry out an online parameter estimation. Details of the numerical implementation on actual data are given in Section 4 along with the selection of the most appropriate model setting via prediction performance and penalised log-likelihood of different competing set ups. Section 5 proposed an empirical early-warning signal extraction method and associated diagnostics. Lastly, Section 6 provides some concluding remarks.

3.2 Model Construction

3.2.1 Ornstein-Uhlenbeck (OU) process

Suppose X_t is an OU process whose stochastic evolution is

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

where μ is the mean level, θ is the speed of mean reversion, and σ is the volatility. The parameters θ, μ and σ are assumed positive. In (3.1), W_t is a standard Brownian motion defined on a probability space $\left(\Omega, \mathscr{F}^X, P\right)$, where $\mathscr{F}^{X_t}$ is the filtration generated by X_t.

By Itô's lemma, the solution to (3.1) is

$$X_t = X_0 e^{-\theta t} + (1 - e^{-\theta t})\mu + \sigma e^{-\theta t} \int_0^t e^{\theta s} dW_s. \tag{3.2}$$

Discretising the solution in Eq. (3.2), we get

$$X_{t_{k+1}} = X_{t_k} e^{-\theta \cdot \Delta t} + (1 - e^{-\theta \cdot \Delta t})\mu + \sigma \sqrt{\frac{1}{2\theta}(1 - e^{-2\theta \cdot \Delta t})} \, w_{k+1}, \tag{3.3}$$

where the $\Delta t = t_{k+1} - t_k$ and $\{w_{k+1}\}$ is a sequence of independent and identically distributed (IID) standard normals, i.e., $w_{k+1} \sim$ IID $N(0, 1)$. The derivation of the second term in (3.3) is justified by the Gaussian property and the Itô's isometry.

3.2.2   Geometric Brownian motion (GBM)

Let Q_t be a GBM with stochastic dynamics

$$dQ_t = \eta Q_t dt + \xi Q_t dB_t, \tag{3.4}$$

where η is the percentage drift and ξ is the percentage volatility. In (3.4), B_t is a standard Brownian motion defined on a probability space $\left(\Omega, \mathscr{F}^Q, P\right)$, where $\mathscr{F}^{Q_t}$ is the filtration generated by Q_t. The parameters η and ξ are positive constants. By Itô's lemma, Eq. (3.4) has the solution

$$\ln(Q_t) - \ln(Q_0) = \left(\eta - \frac{\xi^2}{2}\right)t + \xi B_t. \tag{3.5}$$

Write $Y_t := \ln(Q_t)$, so that

$$Y_t = Y_0 + \left(\eta - \frac{\xi^2}{2}\right)t + \xi B_t \tag{3.6}$$

Discretising Eq. (3.6) with the the Euler approximation, we get

$$Y_{t_{k+1}} = Y_{t_k} + \left(\eta - \frac{\xi^2}{2}\right)\Delta t + \xi \sqrt{\Delta t} \cdot b_{k+1}, \tag{3.7}$$

where the $\Delta t = t_{k+1} - t_k$ and $b_{k+1} \sim$ IID $N(0, 1)$.

3.2.3   Markov chain-governed parameters

For both models of indices, having parameter values that are time-dependent would be realistic. Adopting a similar idea from Zhou and Mamon [95], we regard the states of a Markov chain as regimes of financial market liquidity; more specifically, liquidity regimes are dependent on certain factors that cause market turbulence. Thus, for an economic state to be regime-switching, we posit that the parameters θ, μ, σ, η and ξ are modulated by a discrete-time Markov chain $\mathbf{z}_k$, for $k = 0, 1, \ldots$. The state space of $\mathbf{z}_k$ is finite and it is isomorphic to the canonical basis of $\mathbb{R}^N$, which is the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$. The vector

$e_i = (0, ..., 0, 1, 0, ..., 0)^\top$, where $\top$ denotes the transpose of a vector, is a unit vector with 1 in its ith component; this formulation simplifies a lot of algebra in the succeeding calculations. The semi-martingale representation of $\mathbf{z}_k$ is

$$\mathbf{z}_{k+1} = \mathbf{\Pi}\mathbf{z}_k + \mathbf{v}_{k+1}, \tag{3.8}$$

where $\mathbf{\Pi}$ is a transition matrix; $\mathbf{v}_{k+1}$ is a martingale increment with $E[\mathbf{v}_{k+1}|\mathscr{F}^{\mathbf{z}}] = \mathbf{0}$; and $\mathscr{F}^{\mathbf{z}}$ is the filtration generated by $\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2, \dots$.

The dependence of the parameters on the Markov chain is reflected in the notations $\theta(\mathbf{z}_k)$, $\mu(\mathbf{z}_k)$, $\sigma(\mathbf{z}_k)$, $\eta(\mathbf{z}_k)$, and $\xi(\mathbf{z}_k)$. Under the one-state setting, the parameters in the distribution of $X_{t_{k+1}}$ are constants over the time interval $(t_k, t_{k+1}]$. For the $N-$state setting, we assume that the parameters depend on a Markov chain.

Write $M := M_{OU} + M_{GBM}$ for the dimension of the underlying time series; M_{OU} and M_{GBM} represent the dimension of the data series assumed to follow the OU and GBM processes, respectively. In this study, we have $M_{OU} = 3$ and $M_{GBM} = 1$ based on the the assumed features exhibited by the data. Our modelling approach supposes that all the parameters of each component of the M-dimensional observation process are Markov-driven, i.e.,

$$\begin{cases} X_{t_{k+1}}^{(g)} &= X_{t_k}^{(g)} e^{-\theta^{(g)}(\mathbf{z}_k)\Delta t} + (1 - e^{-\theta^{(g)}(\mathbf{z}_k)\Delta t})\mu^{(\mathbf{g})}(\mathbf{z}_k) + \sigma^{(g)}(\mathbf{z}_k)\sqrt{\frac{1}{2\theta^{(g)}(\mathbf{z}_k)}(1 - e^{-2\theta^{(g)}(\mathbf{z}_k)\Delta t})}\, w_{k+1}^{(g)}, \\[3em] Y_{t_{k+1}} &= Y_{t_k} + \left(\eta(\mathbf{z}_k) - \frac{\xi^2(\mathbf{z}_k)}{2}\right)\Delta t + \xi(\mathbf{z}_k)\sqrt{\Delta t}\, b_{k+1}, \end{cases}$$

$$\tag{3.9}$$

where $g = 1, 2, ..., M_{OU}$. Note that $\mu^{(g)}(\mathbf{z}_k) = \langle \boldsymbol{\mu}_k^{(g)}, \mathbf{z}_k \rangle$, $\theta^{(g)}(\mathbf{z}_k) = \langle \boldsymbol{\theta}_k^{(g)}, \mathbf{z}_k \rangle$, $\sigma^{(g)}(\mathbf{z}_k) = \langle \boldsymbol{\sigma}_k^{(g)}, \mathbf{z}_k \rangle$, $\eta^{(h)}(\mathbf{z}_k) = \langle \boldsymbol{\eta}_k^{(h)}, \mathbf{z}_k \rangle$ and $\xi^{(h)}(\mathbf{z}_k) = \langle \boldsymbol{\xi}_k^{(h)}, \mathbf{z}_k \rangle$, where $\boldsymbol{\mu}_k^{(\mathbf{g})} = (\mu_{k,1}^{(g)}, \mu_{k,2}^{(g)}, ..., \mu_{k,N}^{(g)})^\top$, $\boldsymbol{\sigma}_k^{(\mathbf{g})} = (\sigma_{k,1}^{(g)}, \sigma_{k,2}^{(g)}, ..., \sigma_{k,N}^{(g)})^\top$, $\boldsymbol{\theta}_k^{(\mathbf{g})} = (\theta_{k,1}^{(g)}, \theta_{k,2}^{(g)}, ..., \theta_{k,N}^{(g)})^\top$, $\boldsymbol{\eta}_k^{(\mathbf{h})} = (\eta_{k,1}^{(h)}, \eta_{k,2}^{(h)}, ..., \eta_{k,N}^{(h)})^\top$ and $\boldsymbol{\xi}_k^{(\mathbf{h})} = (\xi_{k,1}^{(h)}, \xi_{k,2}^{(h)}, ..., \xi_{k,N}^{(h)})^\top$ are all in $\mathbb{R}^{\mathbb{N}}$; and $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^N$. Write $\mathbf{N}_{\mathbf{k}} := (w_k^{(1)}, w_k^{(2)}, w_k^{(3)}, b_k^{(1)}, b_k^{(2)})^\top$ whose components are independent for all $k = 0, 1, 2, \dots$.

Let $(\Omega, \mathscr{F}, P)$, where $\mathscr{F} = \mathscr{F}^X \vee \mathscr{F}^Q \vee \mathscr{F}^{\mathbf{z}}$ be a complete probability space that supports all processes in our modelling set up. Note that $\mathscr{F}^X$ and $\mathscr{F}^Q$ are the filtrations generated by all OU-processes and all the GBMs, respectively.

3.2.4   Probability density of observation process

Eq. (3.9) can be re-written as

$$X_{t_{k+1}}^{(g)} = \alpha^{(g)}(\mathbf{z}_k)X_{t_k}^{(g)} + \beta^{(g)}(\mathbf{z}_k) + \kappa^{(g)}(\mathbf{z}_k)\ w_{k+1}^{(g)} \tag{3.10}$$

with

$$\begin{cases} \alpha^{(g)}(\mathbf{z}_k) = e^{-\theta^{(g)}(\mathbf{z}_k)\Delta t} \\[2mm] \beta^{(g)}(\mathbf{z}_k) = (1 - e^{-\theta^{(g)}(\mathbf{z}_k)\Delta t})\mu^{(g)}(\mathbf{z}_k) \\[2mm] \kappa^{(g)}(\mathbf{z}_k) = \sigma^{(g)}(\mathbf{z}_k)\sqrt{(2\theta^{(g)}(\mathbf{z}_k))^{-1}\left(1 - e^{-2\theta^{(g)}(\mathbf{z}_k)\cdot\Delta t}\right)} \end{cases} \tag{3.11}$$

Eq. (3.10) implies

$$X_{t_{k+1}}^{(g)}\Big| X_{t_k}^{(g)} \sim N\left(\alpha^{(g)}(\mathbf{z}_k)X_{t_k}^{(g)} + \beta^{(g)}(\mathbf{z}_k),\ \kappa^{(g)}(\mathbf{z}_k)\right). \tag{3.12}$$

For the distribution of Y_{k+1}, we consider

$$Y_{t_{k+1}} - Y_{t_k} = \left(\eta(\mathbf{z}_k) - \frac{\xi^2(\mathbf{z}_k)}{2}\right)\Delta t + \xi(\mathbf{z}_k)\sqrt{\Delta t}b_{k+1} = \zeta(\mathbf{z}_k) + \nu(\mathbf{z}_k)b_{k+1} \tag{3.13}$$

with

$$\begin{cases} \zeta(\mathbf{z}_k) = \left(\eta(\mathbf{z}_k) - \frac{\xi^2(\mathbf{z}_k)}{2}\right)\Delta t \\[2mm] \nu(\mathbf{z}_k) = \xi(\mathbf{z}_k)\sqrt{\Delta t}. \end{cases} \tag{3.14}$$

Consequently,

$$Y_{t_{k+1}}\Big| Y_{t_k} \sim N\left(Y_{t_k} + \zeta(\mathbf{z}_k),\ \nu(\mathbf{z}_k)\right). \tag{3.15}$$

3.2.5   Filters and parameter estimation

We present the optimal estimates for the parameters of our proposed integrated model using a dynamic maximum-likelihood approach using the Expectation-Maximisation (EM) algorithm is utilised; see Elliott and Krisnamurthy [23], and Wu [86] for a review of the EM algorithm. Here, we combine the one-dimensional EM estimations in [59] and Tenyakov et al. [75]. Then, we extend the multivariate HMM setting, by considering not necessarily the same dynamic specification for each component series, and obtain optimal estimates of model parameters.

For $j = 1, 2 \ldots, N$, we define the following quantities:

$$\mathcal{G}_k^{s,j} = \sum_{n=1}^{k} \langle \mathbf{z}_{n-1}, \mathbf{e}_j \rangle \langle \mathbf{z}_n, \mathbf{e}_s \rangle \tag{3.16}$$

$$\mathcal{O}_k^j = \sum_{n=1}^{k} \langle \mathbf{z}_{n-1}, \mathbf{e}_j \rangle \tag{3.17}$$

$$\mathcal{T}_k^j (f_k) = \sum_{n=1}^{k} \langle \mathbf{z}_{n-1}, \mathbf{e}_j \rangle f_n(\cdot). \tag{3.18}$$

Equations (3.16) and (3.17) denote the respective number of jumps from state j to s and the amount of time that the process $\{\mathbf{z}_n\}$ occupies the state j up to time k. The quantities in $\mathcal{T}_k^j$ in Eq. (3.18) is an auxiliary process that depends on the function $f_n(\cdot)$, of the observation process. In our empirical application, $f_n(\cdot)$ takes the form x_n, x_n^2 and $x_{n-1}x_n$, where x_n is any component from the M-dimensional observation process. The details of filters' construction procedure along with the derivations of Eqs. (3.19) – (3.24) are given in Appendices B and C, respectively. The estimation results are given below.

$$\hat{\pi}_{ji} = \frac{\hat{\mathcal{G}}_k^{j,i}}{\hat{O}_k^i} \tag{3.19}$$

$$\widehat{\alpha_i^{(g)}} = \frac{\hat{\mathcal{T}}_k^i \left(X_k^{(g)} X_{k-1}^{(g)} \right) - \beta_i^{(g)} \cdot \hat{\mathcal{T}}_k^i \left(X_{k-1}^{(g)} \right)}{\hat{\mathcal{T}}_k^i \left(X_{k-1}^{(g)\,2} \right)} \tag{3.20}$$

$$\widehat{\beta_i^{(g)}} = \frac{\hat{\mathcal{T}}_k^i \left(X_k^{(g)} \right) - \alpha_i^{(g)} \cdot \hat{\mathcal{T}}_k^i \left(X_{k-1}^{(g)} \right)}{\hat{O}_k^i} \tag{3.21}$$

$$\widehat{\kappa_i^{2(g)}} = \frac{\hat{\mathcal{T}}_k^i \left(X_k^{(g)2} \right) + \alpha_i \cdot \hat{\mathcal{T}}_k^i \left(X_{k-1}^{(g)\,2} \right) + \beta_i^{(g)2} \hat{O}_k^i - 2\alpha_i^{(g)} \hat{\mathcal{T}}_k^i \left(X_k^{(g)} X_{k-1}^{(g)} \right) - 2\beta_i^{(g)} \hat{\mathcal{T}}_k^i \left(X_k^{(g)} \right)}{\hat{O}_k^i}$$
$$+ \frac{+2\alpha_i^{(g)} \beta_i^{(g)} \hat{\mathcal{T}}_k^i \left(X_{k-1}^{(g)} \right)}{\hat{O}_k^i} \tag{3.22}$$

$$\widehat{\zeta_i} = \frac{\hat{\mathcal{T}}_k^i (Y_k) - \hat{\mathcal{T}}_k^i (Y_{k-1})}{\hat{O}_k^i} \tag{3.23}$$

$$\widehat{v_i^2} = \frac{\hat{\mathcal{T}}_k^i \left(Y_k^2 \right) + \hat{\mathcal{T}}_k \left(Y_{k-1}^2 \right) + \zeta_i^2 \hat{O}_k - 2\hat{\mathcal{T}}_k^i (Y_k Y_{k-1}) - 2\zeta_i \hat{\mathcal{T}}_k^i (Y_k) + 2\zeta_i \hat{\mathcal{T}}_k^i (Y_{k-1})}{\hat{O}_k^i}. \tag{3.24}$$

From Xiong and Mamon's idea [91], we can compute the variance of estimators in our blended multivariate HMM filtering via the Fisher information $\mathcal{I}(\Theta) = -E\left[\frac{d^2}{d\Theta^2} \log\left(\mathcal{L}(\Theta)\right) \Big| \Theta \right]$. The derivation of the Fisher information involved in each estimator is straight forward and

the results are given below; see also Appendices F.1–C.12.

$$I\left(\pi_{ji}\right) = \frac{\hat{\mathcal{G}}_k^{j,i}}{\pi_{ji}^2} \tag{3.25}$$

$$I\left(\alpha_i^{(g)}\right) = \frac{\hat{\mathcal{T}}_k^i\left(X_{k-1}^{(g)\,2}\right)}{\kappa_i^{2(g)}} \tag{3.26}$$

$$I\left(\beta_i^{(g)}\right) = \frac{\hat{O}_k^i}{\kappa_i^{2(g)}} \tag{3.27}$$

$$I\left(\kappa_i^{2(g)}\right) = \frac{\hat{\mathcal{T}}_k^i\left(X_k^2\right) + \alpha_i^2\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) + \hat{O}_k^i\beta_i^2 - 2\alpha_i\hat{\mathcal{T}}_k^i\left(X_kX_{k-1}\right) - 2\beta_i\hat{\mathcal{T}}_k^i\left(X_k\right)}{\left(\kappa_i^{2(g)}\right)^3}$$
$$+ \frac{2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(X_{k-1}\right) - \hat{O}_k^i}{2\left(\kappa_i^{2(g)}\right)^2} \tag{3.28}$$

$$I\left(\zeta_i\right) = \frac{\hat{O}_k^i}{\nu_i^2} \tag{3.29}$$

$$I\left(\nu_i^2\right) = \frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right) + \hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right) + \hat{O}_k^i\zeta_i^2 - 2\hat{\mathcal{T}}_k^i\left(Y_kY_{k-1}\right) - 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_k\right) + 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{\left(\nu_i^2\right)^3}$$
$$- \frac{\hat{O}_k^i}{2\left(\nu_i^2\right)^2}. \tag{3.30}$$

From Eqs. (3.19) – (3.30), we can obtain the optimal values for the original model parameters specified in Eqs. (3.11) and (3.14):

$$\widehat{\mu_i^{(g)}} = \frac{\widehat{\beta_i^{(g)}}}{1 - \widehat{\alpha_i^{(g)}}}, \qquad \widehat{\theta_i^{(g)}} = -\frac{1}{\Delta t}\ln\left(\widehat{\alpha_i^{(g)}}\right), \qquad \widehat{\sigma_i^{2(g)}} = \frac{2\widehat{\theta_i^{(g)}} \cdot \widehat{\kappa_i^{2(g)}}}{1 - \exp\left(-2\widehat{\theta_i^{(g)}} \cdot \Delta t\right)};$$

$$\widehat{\eta_i} = \frac{2\widehat{\zeta_i} + \widehat{\nu_i^2}}{2\Delta t}, \qquad \widehat{\xi_i^2} = \frac{\widehat{\nu_i^2}}{\Delta t}. \tag{3.31}$$

Let Ξ and Θ be the vectors related to the new and original parameters, respectively. The Fisher information for the original parameters can be calculated as $I(\Theta) = \mathbf{J}^\top I(\Xi)\mathbf{J}$, where

$\mathbf{J}$ is the Jacobian matrix with $(i, j)^{\text{th}}$ component as $J_{ij} = \frac{\partial \Xi_i}{\partial \Theta_j}$. The MLE is consistent and has an asymptotically normal sampling distribution (Van de Varrt [79]). With the explicit Fisher-information expressions in (3.25) – (3.30), we can calculate the 95% confidence interval for the estimates Θ with $\hat{\Theta} \pm \frac{1.96}{\sqrt{\mathcal{I}(\Theta)}}$.

3.3   Numerical implementation

3.3.1   Preliminary analysis of data for implementation
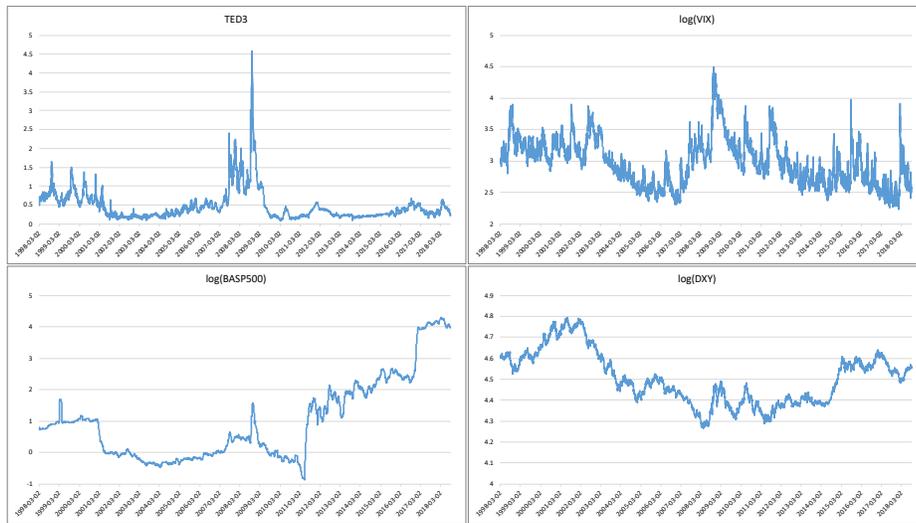
Figure 3.1: Evolution of input data series

The proposed multivariate model is tested on a 4-dimensional daily time series covering the period Mar 1998–Aug 2018 (i.e., 5159×4 data points) which includes the bid-ask spread of S&P 500 (BASP500), US DXY and CBOE VIX all obtained from Bloomberg and the TED spread compiled by the Federal Reserve Bank of St. Louis. The choice of daily (high-frequency) data is justified in Vayanos and Wang [80] from which it is argued that low-frequency measures (less frequent than daily) are imperfect proxies of illiquidity.

We made two transformations on the original data set: (i) To simplify the data processing and visualisation, we take the logarithm of BASP500, DXY and VIX. ii) We take the 21-day moving averages of the logarithm of BASP500 to make the input data smoother for the HMM filters without affecting the ability to capture the evolution of the illiquidity regime.

The evolution of the data and descriptive statistics are presented in Figure 3.1 and Table 3.1, respectively.

Table 3.1: Descriptive statistics of the input data series

Index	Mean	Median	Minimum	Maximum	SD	CV	Kurtosis	Skewness
log(BASP500)	0.999	0.751	-0.866	4.297	1.303	1.305	0.082	0.965
TED	0.472	0.340	0.090	4.580	0.416	0.881	18.132	3.388
log(VIX)	2.976	2.958	2.231	4.495	0.372	0.125	0.205	0.550
log(DXY)	4.503	4.490	4.267	4.795	0.125	0.028	-0.721	0.378

The log(BASP500) has the highest sample standard deviation and coefficient of variation due to its increasing behaviour, lack of mean reversion, and a large range which can be observed in Figure 3.1. The mean level mainly depends on the data's volatility. This makes a Brownian motion with some drift and volatility an ideal candidate model for log(BASP500). Hence, the BASP500 level will be modelled as per (3.5)–(3.7).

The distribution of TED is highly skewed to the right. The kurtosis of the TED exceeds that of the normal distribution by 18.132; this signifies that TED's probability mass function is concentrated around the mean with several values concentrated in the tails of the distribution (Moors [62]). The data set for log(DXY) has a negative excess kurtosis, which indicates that its probability mass function concentrates around the mean with fewer outliers (Westfall [84]). The log(DXY) exhibits similar numerical values for its mean, median, maximum and minimum; it evolves with small variation around the mean; see the lower right panel of Figure 3.1. The plot of log(VIX) shows cyclical dynamics, shifting more often from a higher to a lower level; this gives a standard deviation comparable to other series but with a small coefficient of variation. The evolutions of TED, log(VIX) and log(DXY) have mean-reverting features supporting the suitability of the OU process for describing their underlying characteristics.

Table 3.2 displays a possible data segregation into different state classification based on two historical illiquidity-crisis events, which are the "Dotcom Bubble" and "2007 Global Financial Crisis". To make the mathematics tractable and outcomes interpretable, we assume a two-state market having an "illiquid" regime associated with abnormally high indicator values and a "regular" regime. A transitional state may be devised and could persist over some time due to the weighted combination of volatilities under the above two regimes. The evolutions of the multivariate series undergo several regimes arising from the different

parameter values recovered from the data.

Such regime classification is supported by the estimates of α, β, κ^2, ζ and ν^2 in the possible grouping periods using a least-squares method for each designated interval. The parameters μ, θ, σ^2, η and ξ^2 are computed utilising the set of equations in Eqs. (3.31) with $\Delta t = 1/253$; the NYSE and NASDAQ have an average of about 253 trading days a year.

Table 3.2: Least-squares parameter estimation on certain data subsets

Index	Parameters	Normal	Dotcom Bubble	Financial Crisis	Overall
		Jun,03 to May, 05	Mar, 00 to Oct,02	Dec, 07 to Jun, 09	Mar,98 to Aug,18
log(BASP500)	η_{ls}	-0.00073	-0.0001	-0.00036	-0.000071
	ξ^2_{ls}	0.036	0.055	0.056	0.045
TED	μ_{ls}	0.230	0.540	1.390	0.479
	θ_{ls}	2.148	2.049	2.301	2.240
	σ^2_{ls}	0.640	0.827	1.108	0.859
log(VIX)	μ_{ls}	2.937	3.204	3.314	2.984
	θ_{ls}	1.549	1.656	1.625	1.658
	σ^2_{ls}	0.744	0.881	0.805	0.879
log(DXY)	μ_{ls}	4.518	4.714	4.354	4.505
	θ_{ls}	1.278	1.426	1.309	1.408
	σ^2_{ls}	0.005	0.007	0.008	0.005

The preliminary results in Table 3.2 demonstrate a possible segregation of the actual data into different state groupings. The basis of this grouping is the combined assessment of values for the mean-reverting level, mean reversion rate and volatility of the TED, log(VIX) and log(DXY) and the values of log(BASP500)'s percentage drift and percentage volatility. Indices fitted to the OU-process have high mean-reverting levels during the illiquidity period whilst the GBM's percentage drift and volatility under the illiquidity regimes are higher than those in the regular regime. In general, the OU volatilities are relative lower in regular states. The only exception is the log(DXY) having almost identical volatility levels even in different states as can be gleaned from Table 3.2. A plausible explanation for this is that the original data series has a smaller variation range, which leads to least-square estimates that are close to each other for these different periods.

The mean-reverting speed has a more complicated behaviour than the mean level's. For log(BASP500), log(VIX) and log(DXY), we observe higher mean-reverting rates at illiq-

uidity episodes. However, this is not the case for TED which has a higher mean-reverting rate in regular regime but it has lower ones during illquidity events. This is not surprising as Figure 3.1 shows that it has several extreme values during 2007-2009 but attains lower levels with small variation in other time periods.

Table 3.2 though offers only partial information since least-square estimates are neither accurate nor efficient to capture the dynamic features of data under varying states. In order to extract the characteristics of a data set under different regimes, we employ an HMM-based regime-switching model with an online-estimation scheme. These characteristics are encapsulated by the data's changing distribution as time progresses. Virtually all distributional shapes could be reproduced by a mixture of normal distributions, which is intrinsic to our regime-switching approach.

3.3.2 Initial values for the parameter estimation

The implementation procedure requires initial parameter values for Θ which is defined as $\Theta = \left\{\pi_{ji}, \alpha^{(g)}, \beta^{(g)}, \kappa^{2(g)}, \zeta, \nu^2\right\}$, where $g = 1, 2, 3$. Several ways of setting initial values for HMMs are tackled in Erlwein and Mamon [27], Erlwein et al. [29], Date and Ponomareva [16], Tenyakov et al. [76], and Xiong and Mamon [91]. We shall apply the maximum-likelihood method on the first 251 points of the 4-dimensional data set spanning the period Mar 1998–Feb 1999. The aim is to come up with reasonable initial estimates that could facilitate a stable implementation.

From Subsection 3.2.4, we consider, under the one-state setting (i.e., $N = 1$), the log-likelihood function

$$\mathcal{L}_{ini} = \sum_{i=1}^{N} \left(\sum_{g=1}^{3} \left(\mathcal{L}_i^{(X^{(g)})} \right) + \mathcal{L}_i^{(Y)} \right), \tag{3.32}$$

where

$$\mathcal{L}_i^{(X^{(g)})} = \sum_{l=2}^{251} \left[-\frac{1}{2} \log\left(2\pi\kappa_i^{2(g)}\right) - \frac{\left(X_l^{(g)} - \alpha_i^{(g)} \cdot X_{l-1}^{(g)} - \beta_i^{(g)}\right)^2}{2\kappa_i^{2(g)}} \right]$$

$$\tag{3.33}$$

$$\mathcal{L}_i^{(Y)} = \sum_{l=2}^{251} \left[-\frac{1}{2} \log\left(2\pi\nu_i^2\right) - \frac{(Y_l - Y_{l-1} - \zeta_i)^2}{2\nu_i^2} \right].$$

With the aid of the function 'optim' in the statistical software R to solve the optimisation

problem (3.32). These one-state initial values serve as guides to systematically generate initial values for frameworks with more than 1 regime. All non-zero entries in the transition matrix are set to $1/N$ for $N = 2,\ 3$. The initialisation results are exhibited in Table D.1 of Appendix D.1

3.3.3   Filtering procedure and results

The filters established in Appendix B make use of a data-point processing instead of the usual data-batch processing implementation. This type of processing generalises the conventional HMM sub-optimal batchwise data processing. Here, one algorithm step is completed with the input of one multivariate data point and new parameter estimates are generated, which are then used for the next algorithm step. It is also natural to have a processing frequency for the filtering procedure that is compatible with the frequency of the data compilation (daily in our case). Such compatibility avoids the introduction of extra "noise" in the filtering results.

Additionally, the components of our 4-dimensional time series have different dynamic features that filtering a batch of data points may result to loss of salient information. We found that our one-data point filter processing reflects well the original data's fluctuations.

Our filtering procedure relies on the initial parameters in Table D.1; see Appendices D.1. The first 251 four-dimensional data points (Mar 1998– Feb 1999) are used to for setting initial parameter values. The model performance is assessed with the remaining 4908 daily observations from Mar 1999 to Aug 2018.

The outcomes of the filtering-based parameter estimation under the one-state HMM are illustrated in Figure D.1 of Appendix D.2. Each parameter estimate converges to a certain level at the end of the filtering process.The GBM parameters' evolution is relatively smooth whilst the evolutions of OU parameters' estimates show some significant variations over time. All filtered parameters have different increasing trends, right after some illiquidity incidence (e.g., 2008 global financial crisis).

The dynamics of $\widehat{\mu}$ for the three OU processes coincide with that of the original data series.The mean reversion rates' estimates $\widehat{\theta}$ are less variable signifying stability for all three indices. The squared volatility $\widehat{\sigma}^2$'s exhibit diverse patterns for the three indicators; both logarithm of VIX and DXY's $\widehat{\sigma}^2$ have relative flat curves whilst the TED's has a big spike

at the 2008's financial crisis.
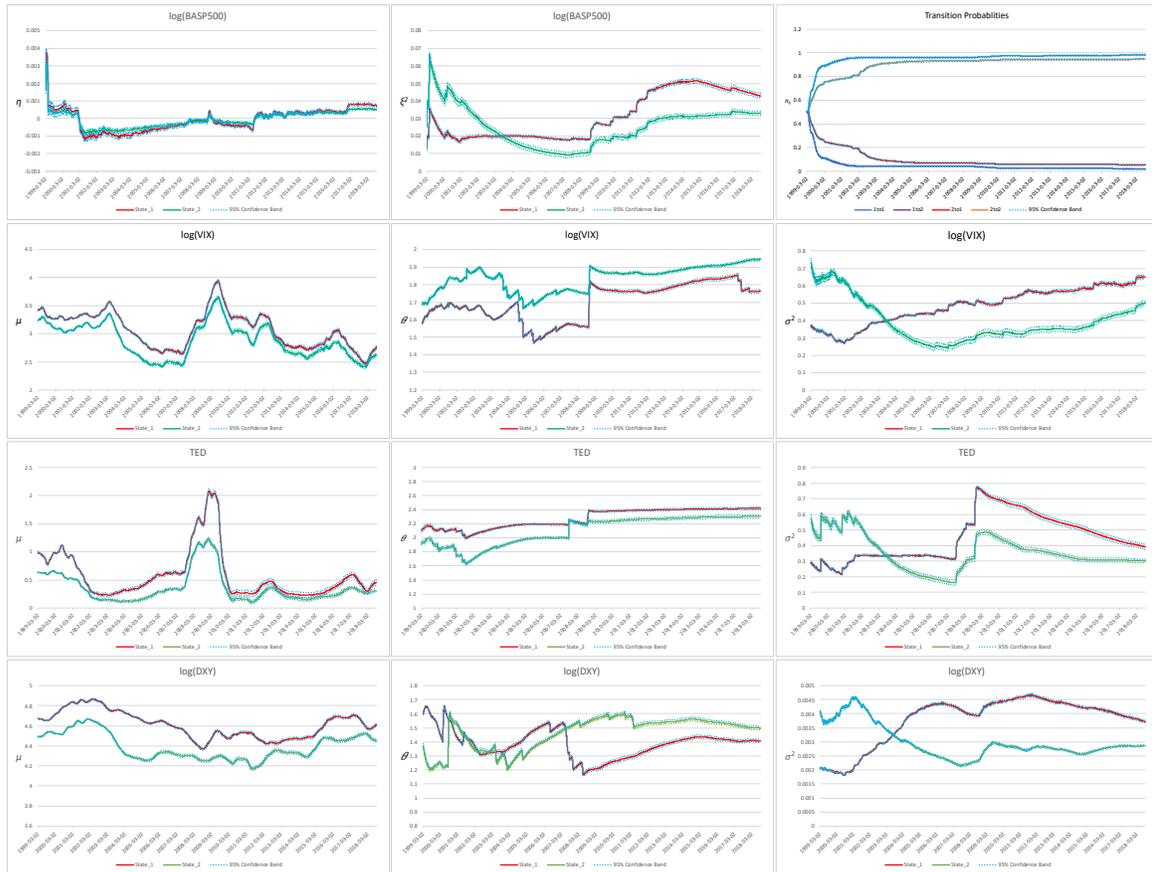
Figure 3.2: Evolution of parameter estimates for μ, θ, σ^2, η, ξ^2 and π_{ji} under a 2-state HMM-based model.

Figure 3.2 displays the movement of the parameter estimates under the 2-state HMM. Note that getting transition probabilities converge to 0 does not bespeak absence of regime-switching feature in the original data. The filtered probabilities' lack of variation is attributed to data's high frequency (daily) for which a significant market regime switch is rarely and not visibly seen. The OU parameter estimates under a 2-regime setting have different dynamics in each of the two states; there is no tendency for the two parameter-estimates evolutions to even coincide. Outcomes for states 1 and 2 bifurcate into two different levels whereas the GBM parameter estimates settle to different levels but they their behaviours are somewhat stable. These findings reconcile with the preliminary analysis on the possible data segregations in Table 3.2. The data set's excellent fit with the the 2-regime HMM is supported by the behaviour of parameter estimates that jibe with dynamics of the actual time series. Furthermore, in comparison to the 1-regime HMM filtering, the

2-regime HMM filtering captures more effectively the major spikes in the 4 indicators.
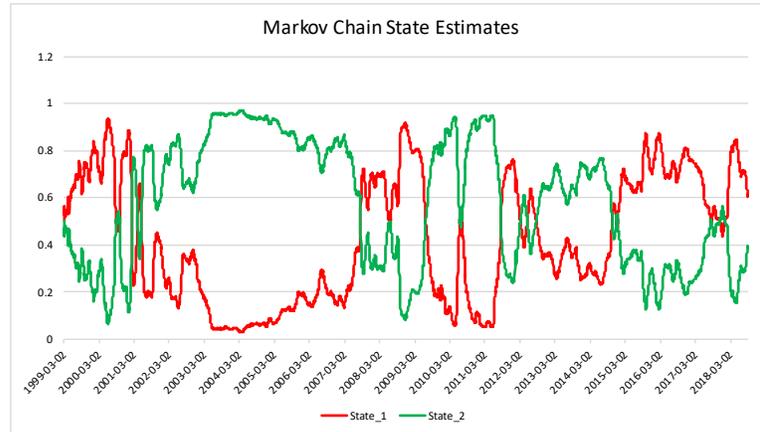
Figure 3.3: Evolution of HMM state estimate under a 2-state model

Figure 3.3 depicts the state estimate of the HMM. The red curve corresponds to the illiquid state whilst the green one refers to the liquid regime. The filtering of HMM states covered the period of historical illiquidity events mentioned in Table 3.2. We shall construct an early-warning alarm signal driven by the HMM-state estimate in Section 3.4. Filtered estimates for the 3-state HMM-based model are plotted in Figure D.2. The evolution of the parameters in state 1 under the 3-regime setting is similar to that in state 1 under the 2-state HMM. For some parameters (e.g, OU-parameter μ and GBM-parameter η), the parameters under the new added regime, state 3, possess similar patterns with those of one of the two states of a 2-regime HMM. It is true that the mean-reversion rate estimate $\widehat{\theta}$ and squared volatility estimate $\widehat{\sigma}^2$ for the added state (state 3) provide some new information given its different evolution feature. Nonetheless, these information do not impact prediction results as shown in the next section.

3.3.4 Model selection and diagnostic

3.3.4.1 Prediction and error analysis

From the model formulations in Subsections 3.2.3 and 3.2.4, the respective one-step ahead forecasts of X_k and Y_k are

$$\begin{cases} E\left[X_{k+1}^{(g)} \mid \mathscr{F}_k\right] = E\left[\alpha^{(g)}(\mathbf{z}_k)X_k + \beta^{(g)}(\mathbf{z}_k) + \kappa^{(g)}(\mathbf{z}_k) w_{k+1}^{(g)} \mid \mathscr{F}_k\right] = \langle \alpha^{(g)}, \hat{\mathbf{z}}_k \rangle X_k + \langle \beta^{(g)}, \hat{\mathbf{z}}_k \rangle, \\ E\left[Y_{k+1} \mid \mathscr{F}_k\right] = E\left[Y_k + \zeta(\mathbf{z}_k) + \nu(\mathbf{z}_k) b_{k+1} \mid \mathscr{F}_k\right] = Y_k + \langle \zeta, \hat{\mathbf{z}}_k \rangle, \end{cases}$$

(3.34)

where $g = 1, 2, 3$. Figure 3.4 gives the plots of the one-step ahead forecasts. The 1-step ahead forecasts trace superbly the dynamics of all four indices. The predictions under log(DXY) and log(BASP500) are better than those under log(VIX) and TED, i.e., the predictions fit the actual data better. In terms of the predicted values, the 2- and 3-state HMMs are better than the 1-state HMM especially for time periods when spikes occur. This is justified by an error analysis (similar to [29] as well as [92]). In particular, the forecasts' goodness of fit under different HMM settings is performed using the root mean square error (RMSE), absolute mean error (MAE), relative absolute error (RAE) and mean absolute percentage error (MAPE). Suppose h_j refers to the observed data value at time j and $\hat{h}_j$ symbolises the one-step ahead prediction at time j. Moreover, let $\bar{h}$ stand for the sample mean of the underlying process and n denote the sample size. With the above-mentioned notations, the RMSE, AME, RAE and MAPE are calculated as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left(\hat{h}_j - h_j\right)^2}, \quad \text{MAE} = \frac{1}{n} \sum_{j=1}^{n} |\hat{h}_j - h_j|,$$

$$\text{RAE} = \frac{\sum_{j=1}^{n} |\hat{h}_j - h_j|}{\sum_{j=1}^{n} |h_j - \bar{h}|}, \quad \text{MAPE} = \frac{1}{n} \sum_{j=1}^{n} \left|\frac{\hat{h}_j - h_j}{h_j}\right|.$$

Table 3.3 displays the error-analysis results and the prediction errors under the log(DXY) and log(BASP500) for all HMM settings are generally much smaller than those under the log(VIX) and TED, which are more volatile than the logarithms of DXY and BASP500. The chosen error metrics showed that the 2- and 3-state HMM-based models outperform the 1-state model in prediction accuracy for all 4 data series even though the errors are very small.

In order to determine whether the error mean differences are statistically significant in each pairwise HMM setting, a t-test is conducted and facilitated by the bootstrap method using
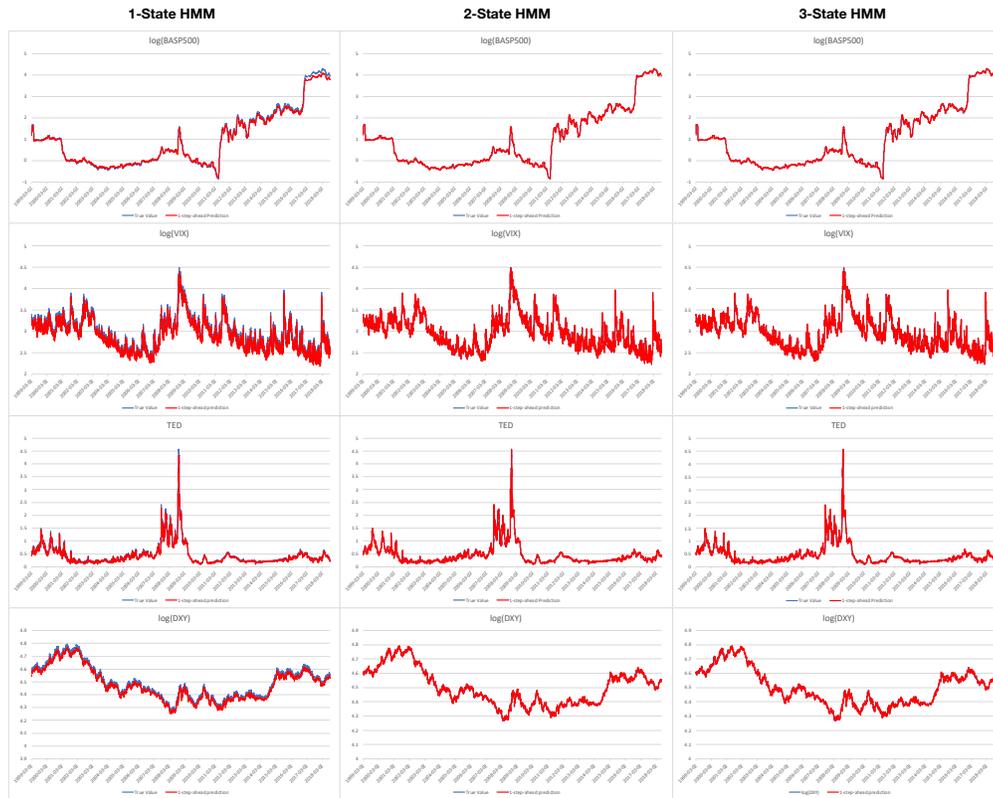
Figure 3.4: One-step ahead forecasts under the 1-, 2-, and 3-state HMMs

Table 3.3: Results of error analysis

| | log(BASP500) | | | | log(VIX) | | |
	1-state HMM	2-state HMM	3-state HMM		1-state HMM	2-state HMM	3-state HMM
RMSE	7.4115×10^{-2}	1.7338×10^{-2}	1.7339×10^{-2}	RMSE	8.9696×10^{-2}	6.7345×10^{-2}	6.7338×10^{-2}
MAE	5.4689×10^{-2}	9.3656×10^{-3}	9.3648×10^{-3}	MAE	6.9938×10^{-2}	4.7756×10^{-2}	4.7752×10^{-2}
RAE	5.4038×10^{-2}	8.8407×10^{-3}	8.8399×10^{-3}	RAE	2.3188×10^{-1}	1.5797×10^{-1}	1.5799×10^{-1}
MAPE	3.2005×10^{-1}	8.7599×10^{-2}	8.7622×10^{-2}	MAPE	2.3318×10^{-2}	1.6013×10^{-2}	1.6011×10^{-2}
	TED				log(DXY)		
	1-state	2-state	3-state		1-state	2-state	3-state
RMSE	6.0258×10^{-2}	5.2172×10^{-2}	5.2150×10^{-2}	RMSE	2.0834×10^{-2}	5.0960×10^{-3}	5.0938×10^{-3}
MAE	3.2725×10^{-2}	2.3927×10^{-2}	2.3907×10^{-2}	MAE	2.0061×10^{-2}	3.8330×10^{-3}	3.8308×10^{-3}
RAE	1.2340×10^{-1}	8.8434×10^{-2}	8.8412×10^{-2}	RAE	1.8911×10^{-1}	3.6045×10^{-2}	3.6047×10^{-2}
MAPE	6.9177×10^{-2}	5.2097×10^{-2}	5.2068×10^{-2}	MAPE	4.4422×10^{-3}	8.5180×10^{-4}	8.5127×10^{-4}

a bootstrapped sample size of 100,000. The outcomes of our pairwise t-test are reported in Table 3.4. For the 2-state HMM versus the 1-state HMM and 3-state versus 1-state HMM, all the p-values are smaller than 1% so that we can conclude that the difference in the error metrics, after adding one or more regimes into the one-state model, is statistically significant. For the 2-state HMM versus the 3-state HMM, most p-values are not significantly less than 5%, signifying that we cannot reject the null hypothesis of no difference. However, for a 5% significance level, the 1- and 2-state and 1-and 3-state HMM models are statistically different in terms of the prediction errors. This suggests that there is merit to using a regime-switching model whilst the 2- and 3-state HMM settings have similar forecasting capacity in our implementation. When various considerations are taken, the 2-state HMM-based model is sufficient to capture the dynamics of data set.

Table 3.4: p-values for paired t-tests applied to RMSE, MAE, RAE, and APE values

	log(BASP500)				log(VIX)		
	2-state HMM vs 1-state HMM	3-state HMM vs 1-state HMM	2-state HMM vs 3-state HMM		2-state HMM vs 1-state HMM	3-state HMM vs 1-state HMM	2-state HMM vs 3-state HMM
RMSE	5.2873×10^{-5}	3.3789×10^{-5}	1.7699×10^{-2}	RMSE	8.2564×10^{-4}	8.6343×10^{-4}	5.1127×10^{-2}
MAE	1.5286×10^{-6}	2.4510×10^{-6}	9.2244×10^{-2}	MAE	9.8775×10^{-4}	6.2133×10^{-4}	9.3636×10^{-2}
RAE	7.7296×10^{-6}	5.4001×10^{-6}	4.2437×10^{-2}	RAE	2.4689×10^{-3}	4.5646×10^{-3}	5.4729×10^{-2}
MAPE	4.9509×10^{-6}	9.8931×10^{-6}	7.0021×10^{-2}	MAPE	6.5236×10^{-3}	2.9940×10^{-3}	3.9241×10^{-2}
	TED				log(DXY)		
	2-state HMM vs 1-state HMM	3-state HMM vs 1-state HMM	2-state HMM vs 3-state HMM		2-state HMM vs 1-state HMM	3-state HMM vs 1-state HMM	2-state HMM vs 3-state HMM
RMSE	8.2065×10^{-3}	7.5948×10^{-3}	4.2852×10^{-2}	RMSE	4.2833×10^{-6}	5.6844×10^{-6}	2.6254×10^{-2}
MAE	4.1235×10^{-3}	6.8042×10^{-3}	9.8906×10^{-2}	MAE	7.0769×10^{-6}	1.0683×10^{-6}	6.4139×10^{-2}
RAE	1.2670×10^{-4}	7.9439×10^{-4}	7.1553×10^{-2}	RAE	6.9683×10^{-6}	9.8990×10^{-6}	4.7058×10^{-2}
MAPE	1.1153×10^{-3}	1.6865×10^{-3}	5.5724×10^{-2}	MAPE	9.5421×10^{-6}	1.9270×10^{-6}	7.0059×10^{-2}

3.3.4.2 Model selection

In choosing the best HMM set up, we complement our error analysis by a likelihood-based model selection analysis (e.g., Akaike Information Criterion (AIC) tailored to HMM similar to the one utilised in Xi and Mamon [88]). The AIC metric estimates the Kull-

back–Leibler information under the ML paradigm, and for each algorithm step this is given by $AIC = -2\mathcal{L}(\Theta) + 2c$, where c is the number of parameters to be estimated (cf Table 3.5). With $\Theta = \left\{ \pi_{ji},\ \alpha^{(g)},\ \beta^{(g)},\ \kappa^{2(g)},\ \zeta,\ v^2 \right\}$, for $g = 1, 2, 3$, let $\mathcal{L}(\Theta)$ denote the log-likelihood function customised for our HMM filtering procedure. In particular,

$$\mathcal{L}(\Theta) = \sum_{i=1}^{N} \left[\langle \mathbf{z}_k,\ \mathbf{e}_i \rangle \left(\sum_{g=1}^{3} \left(\mathcal{L}_i^{(X^{(g)})} \right) + \mathcal{L}_i^{(Y)} \right) \right], \tag{3.35}$$

where

$$\mathcal{L}_i^{(X^{(g)})} = -\frac{1}{2} \log \left(2\pi \kappa_i^{2(g)} \right) - \frac{\left(X_{k+1}^{(g)} - \alpha_i^{(g)} \cdot X_k^{(g)} - \beta_i^{(g)} \right)^2}{2\kappa_i^{2(g)}}$$

$$\tag{3.36}$$

$$\mathcal{L}_i^{(Y)} = -\frac{1}{2} \log \left(2\pi v_i^2 \right) - \frac{(Y_{k+1} - Y_k - \zeta_i)^2}{2v_i^2}.$$

The AIC for each model setting is computed using the parameter estimates given at the end of each algorithm run or step. The evolution of AIC values for the 1-, 2- and 3-state models after each algorithm run is portrayed in Figure 3.5 From (3.35), the model setting is deemed the best for the data set and should be selected if it yields the lowest AIC value. As an offshoot of growing regime size and time-series dimension, there will be an upsurge in the AIC's magnitude. This is due to the substantial increase in the number of parameters especially the enlargement of the transition probability matrix. Whilst the 1-state model produces a sequence (as a function of algorithm steps) of smallest AIC values, evidently the 2-state HMM generally produces a set of more stable AIC values which are the lowest during turbulent market periods. Table 3.6 contains the statistics of the AIC and log-likelihood values covering all algorithm steps. The likelihood statistic of the 2-state HMM is smaller than the 1-state model's. The standard deviation of the AIC and likelihood values under the 2-regime set-up are less than those under the 1-regime framework.

Table 3.5: Number of estimated parameters under various HMM settings

	HMM settings				
	1-state	2-state	3-state	$\cdots$	N-state
Number of parameters	11	24	39	$\cdots$	$(3 \cdot 3N) + (1 \cdot 2N) + N(N-1)$

Therefore, the 2-regime HMM is the most suitable for modelling the 4-dimensional data series for having attained the greatest balance between model's fitness and complexity.

This conclusion, which confirms and reinforces the previous results in our error analysis, is similar to the findings on the most appropriate modelling framework obtained in [76].
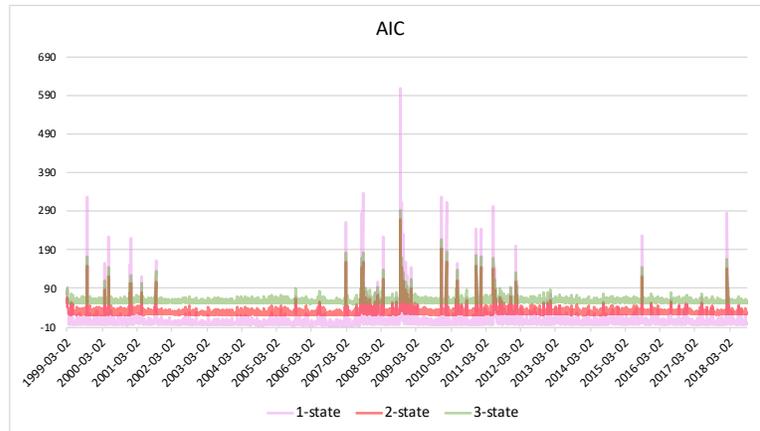
Figure 3.5: Evolution of AIC values for the 1-, 2-, and 3-state HMMs

Table 3.6: AIC statistics covering all algorithm steps

AIC	1-state	2-state	3-state
Mean	5.387	28.790	57.770
Standard Deviation	24.573	11.034	11.470
Log-likelihood	1-state	2-state	3-state
Mean	8.306	9.605	10.115
Standard Deviation	12.286	5.517	5.735

3.4 Early-warning detection of market illiquidity

To augment the benefits of this work, an early-warning alert signal (EWAS) is generated under the 2-state HMM setting. We use dated historical economic and illiquidity episodes to probe the proposed alert system's capacity to unmask signals for predicting market illiquidity. In this specific implementation, the pre-crisis stage is assumed to be totally captured by the HMM-state process alone, and not by any other model parameters. We shall then delve into the relationship between historical illiquidity events and the HMM state estimate in an effort to devise some meaningful pre-crisis signalling system. We set forth an intuitive

but effective approach that seizes the key features of a variable designed to emit warning signals before or at least at the early stage of the crisis events.

3.4.1 Market liquidity regime forecasting

From Eqs. (B.2)–(B.5), $\hat{\mathbf{z}}_t$ is the filtered-regime estimate of a hidden Markov chain that assesses the market-liquidity state. Based on Eq. (3.19), $\mathbf{\Pi}_t$ is the filtered transition probability matrix at time t. According to the set up in Eq. (3.8) from subsection 3.2.3, the n-step ahead predicted regime is computed as

$$\widetilde{\mathbf{z}}_{t+n} = E\left[\hat{\mathbf{z}}_{t+n}|\mathscr{F}_t\right] = \mathbf{\Pi}_t^n \cdot E\left[\mathbf{z}_t|\mathscr{F}_t\right] = \mathbf{\Pi}_t^n \cdot \hat{\mathbf{z}}_t \tag{3.37}$$

We further write $\hat{\mathbf{z}}_t^{(1)} = \langle \hat{\mathbf{z}}_t, e_1 \rangle$ and $\widetilde{\mathbf{z}}_{t+n}^{(1)} = \langle \widetilde{\mathbf{z}}_{t+n}, e_1 \rangle$ which are the first components of filtered and n-step forward predicted regimes, respectively. Both of them are related to the illiquidity regimes in this implementation. To make the filtered results interpretable, we denote by $\hat{S}_t$ and $\widetilde{S}_{t+n}$ the metrics for the current and n-step forward predicted market illquidity severity, respectively, i.e.,

$$\hat{S}_t = \mathbf{1}_{\{\hat{\mathbf{z}}_t^{(1)} \geq 0.5\}} \tag{3.38}$$

$$\widetilde{S}_{t+n} = \mathbf{1}_{\{\widetilde{\mathbf{z}}_{t+n}^{(1)} \geq 0.5\}}, \tag{3.39}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function.

For a given prediction step n, we shall test the performance of the proposed n-step forward regime forecasting method by measuring the predicted clustering error involving the Markov-chain regimes. More precisely, we compute Type I and II mis-clustering error rates (MCER$^{(I)}$ and MCER$^{(II)}$) defined as follows:

$$\begin{cases} \text{MCER}^{(I)} = \dfrac{1}{J-n} \sum_{t=n+1}^{J} \mathbf{1}_{\{\widetilde{S}_{t+n}=1, \hat{S}_{t+n}=0\}} \\ \text{MCER}^{(II)} = \dfrac{1}{J-n} \sum_{t=n+1}^{J} \mathbf{1}_{\{\widetilde{S}_{t+n}=0, \hat{S}_{t+n}=1\}}, \end{cases} \tag{3.40}$$

where J is the total number of observations. Figure 3.6 presents the results for the MCER$^{(I)}$ and MCER$^{(II)}$ yielding 21 steps ahead regime predictions. Both error rates grow when the number of steps increases. The level of Type II Error rate (false negative) is lower than 4% when $n < 12$ (trading days). In general, within 21 (trading days) prediction step length, both error rates are below 0.12, which make $\hat{\mathbf{z}}_t^{(1)}$ and $\widetilde{\mathbf{z}}_{t+n}$ two reliable candidates for the component of the early warning alarm generating system.
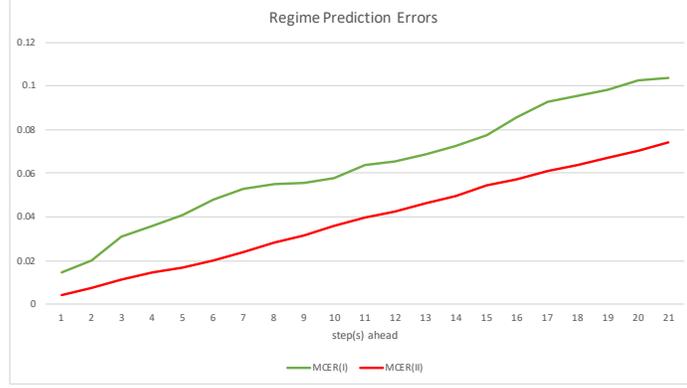
Figure 3.6: Type I and II mis-clustering error rates for the n-step ahead regime prediction

3.4.2   Construction early warning alarm signal

Let x_k be some time series with finite values. We construct a dynamic percentage rank function

$$R(x_k, s, t) = \frac{1}{t - s} \sum_{k=s}^{t-1} \mathbf{1}_{\{x_k < x_t\}}, \tag{3.41}$$

where s is a parameter to determine memory length. The function (3.41) has a short (long) memory if the value of s is large (small). Note that s can be chosen based on different criteria, and its impact on the early warning signal's performance is highly dependent on the data set.

In order to extract information in evaluating possibility of future market illiquidity episodes, we shall examine the current and predicted Markov chain state vis-à-vis historical dynamics of the HMM filtered regimes. Eq. (3.41) quantifies the severity of market illiquidity at time t by appraising the percentage rank of the current value of the Markov chain state relative to all past data points. We first define the Past Regime Factor (PRF) P_t, which emits alarm signals based on the filtered regime $\hat{\mathbf{z}}_k$, as

$$P_t = \frac{1}{2} \cdot \mathbf{1}_{\left\{R(\hat{\mathbf{z}}_k^{(1)}, s, t) > a_P\right\}} \cdot \mathbf{1}_{\left\{\left(R(\Delta \hat{\mathbf{z}}_k^{(1)}, s, t) > b_P\right) \cup \left(\hat{\mathbf{z}}_t^{(1)} > c_P\right)\right\}}, \tag{3.42}$$

where a_P, b_P, $c_P \in (0, 1)$ are threshold parameters, and $\Delta \hat{\mathbf{z}}_k^{(1)} = \hat{\mathbf{z}}_k^{(1)} - \hat{\mathbf{z}}_{k-1}^{(1)}$. The PRF's alarm signals are obtained by tracing the ranks of filtered regimes and the trend of their change. An alarm is triggered when two conditions are both satisfied: (i) The rank of filtered regime $\hat{\mathbf{z}}_k^{(1)}$ exceeds the upper bound a_P, and (ii) Either the rank of the differenced filtered regime $\Delta \hat{\mathbf{z}}_k^{(1)}$ is above some threshold b_P or $\hat{\mathbf{z}}_k^{(1)}$ is greater than the corresponding boundary c_P.

Recall the definition of $\widetilde{\mathbf{z}}_{t+n}$ and write $\dot{\widetilde{\mathbf{z}}}_k := \max\left(\widetilde{\mathbf{z}}_{k+1}^{(1)}, \widetilde{\mathbf{z}}_{k+2}^{(1)}..., \widetilde{\mathbf{z}}_{k+n}^{(1)}\right)$. We set $n = 21$ (trading days) in this implementation. Similar to PRF, we define the Forecast Regime Factor (FRF) F_t as

$$F_t = \frac{1}{2} \cdot \mathbf{1}_{\{R(\dot{\widetilde{\mathbf{z}}}_k, s, t) > a_F\}} \cdot \mathbf{1}_{\{(R(\Delta\dot{\widetilde{\mathbf{z}}}_k, s, t) > b_F)\cup(\dot{\widetilde{\mathbf{z}}}_k > c_F)\}}, \tag{3.43}$$

where a_F, b_F, $c_F \in (0, 1)$ threshold parameters; and $\Delta\dot{\widetilde{\mathbf{z}}}_k = \dot{\widetilde{\mathbf{z}}}_{k+1} - \dot{\widetilde{\mathbf{z}}}_k$. The FRF also generates alarm signals via the tracing of the ranks of forecast regimes and the trend of their change. More specifically, the alarm is triggered when two conditions are both met: (i) The rank of the maximum of 21 steps ahead predicted regimes $\dot{\widetilde{\mathbf{z}}}_k$'s exceeds the upper bound a_F and; (ii) Either the rank of the maximum differenced $\Delta\dot{\widetilde{\mathbf{z}}}_k$ 21 steps ahead predicted regimes is above some threshold b_F or $\dot{\widetilde{\mathbf{z}}}_k$ is greater than the corresponding boundary c_F.

The early-warning alarm signal (EWAS) at time t is then constructed as

$$\text{EWAS}_t = P_t + F_t. \tag{3.44}$$

EWAS (3.44) evaluates dynamically the status of liquidity way by tracing the backward as well as forward HMM evolution at time t. The range of EWAS output is the interval $[0, 1]$. An outcome of 1 means that the anticipated severity of illiquidity is high whilst 0 means that the future illiquidity risk is low. For ease of interpretation, we set to 0.5 all the boundary parameters in EWAS, i.e., a_P, b_P, c_P, a_F, b_F, and c_F. That is, we let the PRF and FRF capture the event that value of the underlying sequence's percentage rank exceeds 50%. The memory length parameter s is set to 0 in order to include all historical information. All EWAS parameters can be adjusted by practitioners based on certain characteristics of the input data to minimise the error rate.

3.4.3 EWAS outcomes and diagnostics

In this subsection, we shall examine the output of the EWAS_t and related error analysis.

3.4.3.1 Output of EWAS

Figure 3.7 provides the plots of EWAS (red curve) as well as dated illquidity risk episodes (cloured rectangular areas) We observe that the EWAS_k metric (plotted as red curves) is capable of generating different alarm levels before or sometimes at least at the early stage of these dated prior illiquidity incidences with very fewer Type-I Errors. This outcome

supports the 2-state HMM filters giving good predictive performance in market illquidity forecasting.
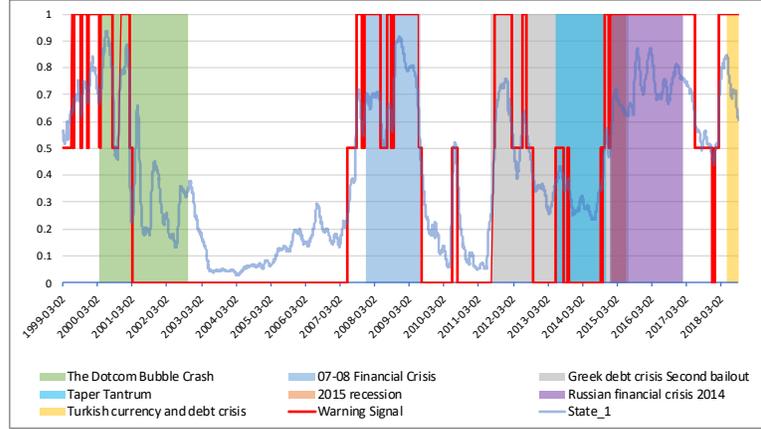
Figure 3.7: Outcome of EWAS

3.4.3.2 Type I and II errors

We construct a time series V_t based on dated illiquidity events as follows:

$$V_t := \mathbf{1}_{\{\text{Illquidity event happens at time } t\}}. \tag{3.45}$$

For a given prediction step number n, we shall assess the performance of EWAS_t by measuring the clustering error. More specifically, we will compute Type-I and Type-II errors. Note that, in this implementation, the Type-I error is related to false positive (warning) alarm whilst Type II Error (false negative) means failure of generating alarm for an illiquidity episode at its early stage. Let $A_t^{(FP)}$ be a triggered false positive alarm (Type-I error); in this case, a non-zero EWAS_t results in the following periods: (i) 21 (trading days) earlier than an illiquidity incident when $\hat{\mathbf{z}}_t^{(1)} < 0.5$ or (ii) in any interval where no illiquidity events happen when $\hat{\mathbf{z}}_t^{(1)} < 0.5$. Similarly, the false negative event $A_t^{(FN)}$ (Type-II error) occurs when $\text{EWAS}_t = 0$ within the 21 (trading days) intervals for each illiquidity event when $\hat{\mathbf{z}}_t^{(1)} \geq 0.5$. We define the mean Type-I and Type-II errors as

$$\begin{cases} \text{ME}^{(FP)} = \dfrac{1}{J} \sum_{t=1}^{J} \mathbf{1}_{\{A_t^{(FP)} \neq \emptyset\}} \\[4mm] \text{ME}^{(FN)} = \dfrac{1}{J} \sum_{t=1}^{J} \mathbf{1}_{\{A_t^{(FN)} \neq \emptyset\}}, \end{cases} \tag{3.46}$$

where J is the total number of observations. As shown in Table 3.7, both mean Type-I

and II errors are below 6% indicating that the proposed warning system has the ability to identify the upcoming and current illiquidity events.

Table 3.7: EWAS: Mean Type-I and II errors

	$\mathrm{ME}^{(FP)}$ (Type I)	$\mathrm{ME}^{(FN)}$ (Type II)	Total
Value	0.0515	0.0346	0.0861

3.4.3.3   Sensitivity and Specificity

Treating EWAS as a binary classifier, we evaluate the sensitivity and specificity by computing the True Positive Rate (TPR) and False Positive Rate (FPR). A perfect classifier, which has the highest performance in the identification of illiquidity episodes, leads to TPR = 1 and FPR = 0. The calculation of TPR and FTR is based on

$$
\begin{cases}
\mathrm{TPR} & = \dfrac{J_{\mathrm{TP}}}{J_{\mathrm{POS}}} = \dfrac{\sum_{t=1}^{J}\left(\mathbf{1}_{\{\mathrm{EWAS}_t>0\}} - \mathbf{1}_{\{A_t^{(FP)}\}}\right)}{\sum_{t=1}^{J}\left(\mathbf{1}_{\{\mathrm{EWAS}_t>0\}} - \mathbf{1}_{\{A_t^{(FP)}\}} + \mathbf{1}_{\{A_t^{(FN)}\}}\right)} \\[20pt]
\mathrm{FPR} & = \dfrac{J_{\mathrm{FP}}}{J_{\mathrm{NEG}}} = \dfrac{\sum_{t=1}^{J}\mathbf{1}_{\{A_t^{(FP)}\}}}{\sum_{t=1}^{J}\left(\mathbf{1}_{\{\mathrm{EWAS}_t=0\}} - \mathbf{1}_{\{A_t^{(FN)}\}} + \mathbf{1}_{\{A_t^{(FP)}\}}\right)},
\end{cases}
\tag{3.47}
$$

where J_{POS} (or J_{NEG}) is the number of condition positive (or negative) data points whilst J_{TP} (or J_{FP}) is the number related to true (or false) positive alarm. Note that $J = J_{\mathrm{POS}} + J_{\mathrm{NEG}}$ is the total observation number. The event $A_t^{(FP)}$ signifies a false positive alarm defined in 3.4.3.2. As shown in Table 3.8, the TPR is close to 1 and FPR is close to 0, which is a strong evidence for EWAS$_t$'s capability in pinpointing illiquidity events.

Table 3.8: True and false positive rates

	TPR	FPR
Value	0.9285	0.0998

3.4.4 Granger's causality test

In addition to the visualisation illustrated by Figure 3.7, we shall use some statistical tools to test whether or not the EWAS generates warning signal earlier than the illiquidity incidence. Granger (1969) [37] argued that causality in economics could be tested for by measuring the ability to predict the future values of a time series using prior values of another time series. The Granger's causality test has been widely used as an effective technique to determine the existence of predictive causality between two time series. In this study, we shall apply Granger's test to justify the effectiveness of the EWAS function in accordance with the succeeding descriptions.

3.4.4.1 Construct stationary time series

The Granger's test requires the time-series input to be stationary. Thus, we obtain a differenced time series $EWAS_t$ and V_t, and conduct the Augmented Dickey–Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test to check stationarity. Further information regarding these tests are detailed in [20] and [54].

Table 3.9: Stationarity: Results of ADF and KPSS tests

Tested Time Series	Test	H_0	p-values	Decision (5% significance)
$\Delta EWAS_t$	ADF	$\Delta EWAS_t$ is not stationary	< 0.01	Reject the Null Hypothesis
	KPSS	$\Delta EWAS_t$ is stationary	> 0.1	Do not reject the Null Hypothesis
ΔV_t	ADF	ΔV_t is not stationary	< 0.01	Reject the Null Hypothesis
	KPSS	ΔV_t is stationary	> 0.1	Do not reject the Null Hypothesis

Table 3.9 implies that $\Delta EWAS_t$ and ΔV_t are stationary time series at a 5% significance level. This result validates the assumption of the Granger's causality test on the bivariate data series that we are investigating.

3.4.4.2 Tests set up

Two Granger's causality tests are conducted on time series models formulated as

$$
T_1 : \begin{cases} \text{Restricted Model}\,(H_0) : & \Delta V_t = a_0 + \sum_{i=t-1}^{t-\tau} a_i \cdot \Delta V_i + \epsilon_t, \\[2mm] \text{Full Model}\,(H_a) : & \Delta V_t = a_0 + \sum_{i=t-1}^{t-\tau} a_i \cdot \Delta V_i + \sum_{j=t-1}^{t-\tau} b_j \cdot \Delta EWAS_j + \epsilon_t, \end{cases}
$$

(3.48)

$$
T_2 : \begin{cases} \text{Restricted Model}\,(H_0) : & \Delta EWAS_t = c_0 + \sum_{i=t-1}^{t-\tau} c_i \cdot \Delta EWAS_i + \upsilon_t, \\[2mm] \text{Full Model}\,(H_a) : & \Delta V_t = c_0 + \sum_{i=t-1}^{t-\tau} c_i \cdot \Delta EWAS_i + \sum_{j=t-1}^{t-\tau} d_j \cdot \Delta V_j + \upsilon_t, \end{cases}
$$

(3.49)

where a_i, b_i, c_i and d_i are unknown constants, and ϵ_t and υ_t are white-noise error terms. T_1 is designed to determine whether there is significant statistical evidence to support that $\Delta EWAS_t$ causes ΔV_t whilst T_2 aims to verify statistical significance of whether ΔV_t causes $\Delta EWAS_t$.

Note that τ is the lag parameter for the test. Similar to the method in [77], we use AIC, BIC and F-test to search the appropriate values for τ. Assuming an early-warning alarm signal corresponding to an illiquidity event will not be generated 21 more trading days prior it happens, we set an upper bound $\tau = 21$. Table 3.10 provides the optimal τ value based on different criteria for assessing the regression in Eqs. (3.48) and (3.49).

Table 3.10: Granger's causality test: selection of the lag τ

Criteria	AIC	BIC	F-test
Selected lag τ for T_1	2	1	8
Selected lag τ for T_2	14	1	19

3.4.4.3 Tests' result

We employ the R function "grangertest" in carrying out the Granger's test for T_1 and T_2 with 6 lag parameters in Table 3.10. The resulting p–values are exhibited in Table 3.11. We observe that all the p–values generated from T_1 are less than 5%; thus, we reject, at

a 5% significance level, all null hypotheses (Restricted Model) under T_1 for all 6 different lag parameters. This implies that $\Delta EWAS_t$ causes ΔV_t at a highly significant level. On the other hand, all the 6 p–values produced by T_2 are relatively large (> 0.5), which indicates not having enough evidence to reject the null hypothesis at 5% significance level. So, there is insufficient basis to say that ΔV_t causes $\Delta EWAS_t$. Consequently, with the data set analysed in this study, we conclude with empirically and statistical support that $EWAS_t$ is capable of generating early warning alarm signals that aid in detecting forthcoming market-illiquidity events .

Table 3.11: p–values for the Granger's causality test

Test	Lag for T_1			Lag for T_2		
	$\tau = 1$	$\tau = 2$	$\tau = 8$	$\tau = 1$	$\tau = 14$	$\tau = 19$
T_1	0.0083	0.0074	0.0062	0.0083	0.0377	0.0146
T_2	0.9996	0.9968	0.9274	0.9996	0.6589	0.5951

3.5 Conclusion

The major contribution of this work is the development of a hybrid multivariate regime-switching HMM-based model to assess the level of market illiquidity. Three OU processes and a GBM, whose parameters are simultaneously governed by a hidden Markov chain in discrete time, were synthesised. It was shown that our proposed model is capable of capturing various time-series characteristics simultaneously. We detailed a methodology in extracting essential information from the observed multi-dimensional stochastic process as well as linking the empirical results to the model's one-step ahead predictions. Taking advantage of the EM algorithm and change of reference probability measures, a new multivariate HMM online recursive filtering method is obtained.

The model's implementability and forecasting performance were tested on actual daily multi-dimensional time series of log(BASP500), TED, log(VIX) and log(DXY) compiled during a 15-year period. The recursive filtering algorithm, with data-point processing, was applied. The one-step ahead predictions are calculated under the 1-, 2- and 3- state HMM set-ups. The post-diagnostic modelling analysis reveals that forecasting errors generated

by the multi-regime HMMs are significantly smaller than those produced by the 1-state HMM. More specifically, the 1-regime model significantly underperforms compared to the 2-regime model. The choice of the optimal number of states is justified by the AIC comparison in conjunction with the goodness-of-fit numerical exercise for the one-step ahead forecasts. We found that the 2-state HMM provides the most reliable framework for the data that we investigated.

We explored the utility of our model's results to gauge occurrence of actual past illiquidity events. An empirical pre-crisis detection method was established and tested on historical illiquidity episodes. Our findings indicate that the proposed 2-regime HMM filters could generate early-warning signals before or at least at the early stage of the illiquidity risk incidences.

This work further elevates progress in HMM modelling in the following respects: (i) blending of the OU and GBM frameworks within HMM, which is new in an attempt to capture important stylised joint properties of series relevant for financial-market illiquidity analysis, (ii) development of an extended self-tuning estimation procedure under a multivariate HMM set up in which each univariate series component has different dynamics, (iii) detailed implementation of filtering method with model validation, (iv) creation of an early-warning system primarily relying on the HMM-state estimates of a regime-switching model, and (v) statistical assessment of EWAS's performance.

We provide a quantitative methodology for market illiquidity modelling in which the estimation and implementation are designed to capture the effects and identify illiquidity-events location in the context of drafting regulatory measures and policies. Our filtering equations could be employed to provide dynamic parameter estimates of models tailored for financial-market liquidity risk management. The liquid/illiquidity-regime classification in our regime-switching modelling approach accurately correspond to those identified by practitioners. The regime-switching results are also easy to interpret with direct relevance to bankers, economists, regulators and traders.

Chapter 4

A multivariate anomaly detection system with supervised learning

4.1 Introduction

The intent of this chapter is to construct an algorithmic platform that integrates machine learning algorithm and statistical models to tackle some challenges in the artificial intelligence's task of classification. We demonstrate how our methodology works in addressing one primary concern in economics and finance although with the generality of its principle our approach can be applied to other multivariate modelling problems in the natural sciences and engineering. In our case, an automated predictive mechanism to detect future periods of financial instability will be constructed. The aim is to widen available tools and techniques that could rein in situations that weaken global financial stability.

Since the financial crisis of 2007, researchers, market participants, and regulators alike have explored ways to examine and forecast the potential for instances of financial instability in the future. As there are various characterisations of financial stability, we rely on the World Bank of its portrayal, namely, (i) the absence of system-wide episodes, i.e., financial crisis, in which the financial system fails to function; and (ii) the resilience of financial systems to stress. In other words, financial stability is a state where the financial system is immune to systemic financial crisis and is able to smoothly perform its basic functions. Thus, the detection in advance of financial-stress episodes is pivotal for bankers and regulators in mitigating the shock attributed to the occurrence of financial stress.

Measuring the degree of financial stability or instability is a complex endeavour considering the interdependence of internal and external forces acting on the financial system and economy [32]. Yet, there growing efforts to come up with a single composite quantitative index that could signify the extent of financial stress. Duprey et al. [21] constructed a new country-level Financial Stress Index (FSI) covering three core financial market segments: (i) equity markets, (ii) bond markets, and (iii) foreign exchange markets. The details of the FSI's inputs and construction procedure are elaborated in [21]. The literature, using Markov-switching and one threshold vector autoregressive model to combine FSI with elements quantifying financial stress on business cycle, successfully identified those systemic financial stress episodes that are consistent with the expert-categorised stress periods. More specifically, 83% of systemic financial-stress periods detected in [21] are also identified as crisis states by the experts.

The machine learning algorithms could be divided into two categories: supervised learning and unsupervised learning. The aim of supervised learning is to predict the value of a target variable based on some given input features. The term "supervised" is due to the fact that the learning process is guided by the target variable. In unsupervised learning, the target variables are either unavailable or unobservable and the task is to describe particular associations and patterns, i.e., how the data are organised or clustered amongst a set of input features.[43]

Some researches focusing on financial stability use machine-learning algorithms to build early-warning models to predict the occurrence of a financial crisis. Duttagupta and Cashin [22] utilised a Binary Classification Tree (BCT) model to investigate banking crises in 50 emerging markets and developing countries. An early-warning model is constructed with Random Forests in [2] to predict systemic banking crisis to financial stability. Ward [83] found that the out-of-sample classification performance of the banking crisis indicator constructed with Random Forest outperforms that of the logit models when tested on a long-run multiple countries dataset. Casabianca et al. [12] developed an early-warning system to detect banking-crisis episodes with Adaptive Boosting (AdaBoost) which is able to obtain better out-of-sample performance than the logit models. In [30], an early-warning system, which is created using artificial neural network, surpasses the traditional parametric models in terms of prediction performance in identifying sovereign-debt-crisis episodes.

Instead of making predictions by incorporating multiple different types of indices as the input variables, our approach, in contrast to the prevalent methodologies in the literature, assumes a scenario that we only have partial information consisting of the FSI and an ag-

gregated index covering various component indices in our model development process. Our methodology – a composite of algorithms leading to a supervised-learning classification approach – is designed to 'filter out' essential information from the country-level FSI in [21] that will help identify high financial-stress episodes. The main point of interest is to capture the evolution of multiple countries' FSI altogether and then generate multiple step-ahead predictions that give out signals concerning the FSI levels in the future. The characterisation of the FSI's dynamics, whether they are anomalous or not, is carried out by introducing a discriminative signal. Such a signal is defined as the lag-τ differenced of the original time series. HMM-modulated OU processes are utilised to model 17 different countries' discriminative signals simultaneously. The HMM recursive online filters are then established to extract signal information from the input time series with the implementation of the change of probability measure technique. In conjunction with the Expectation-Maximisation (EM) algorithm, the HMM-driven multivariate OU model parameters are obtained. A detailed methodology construction and design is a particular feature in this chapter, wherein predictive-analytics performance is strengthened and improved through the employment of Random Forest and XGBoost classifier. Random Forest enables variable selections amongst the HMM and Non-HMM predictors concomitantly. XGBoost, whose hyper-parameters are tuned using time series cross-validation combined with grid search, is deployed as the final-stage classifier. The integrated modelling algorithm is trained when the arrival of new data; and model parameters are updated to capture the dynamics of the input-data series. We apply a direct-prediction approach , which avoids error propagation in a multi-step-ahead forecasting [73], in making 6-step ahead predictions on the financial-stress status of all countries considered in the study, thereby producing early-warning financial-crisis signals.

Using the actual multi-dimensional time series of 17 countries' FSIs, the predictive performance of our algorithmic modelling approach is assessed under different diagnostic measures and through benchmarking with five well-known competing algorithms. A tailored analysis on the features' importance indicates that the HMM-related variables have high impacts on the accuracy of the anomaly detection. Furthermore, two anomaly-warning signal systems are constructed to identify two different types of extreme anomalous episodes six months ahead. Our findings indicate that our proposed modelling algorithm has a stronger capability to detect in advance the FSIs' abberations with fewer false alarms. Our model could be utilised to obtain dynamic quantitative assessments of FSI status tailored to country-level financial stability management for economists and regulators.

The structure of the remaining parts of this chapter is as follows. Section 2 provides a brief overview of the underlying principles of the major components of our modelling algorithm. In Section 3, we illustrate the formulation of our methodology consisting of HMM filtering derivation, Random-Forests predictors' selection module, and XGBoost tuning and training. The details of a comparative examination involving the algorithm put forward in this work vis-à-vis existing approaches are given in Section 4. The construction and diagnostics of two signal-projection systems are laid out in Section 5. Lastly, Section 6 provides a summary and implications of this study.

4.2   Related work

4.2.1   HMM-based regime-switching models

HMM is a doubly-embedded stochastic process comprised of an observation series and an underlying hidden process delineated by its number of states and transition probabilities or intensities. The parameters of an HMM can be estimated based on the observed data series unveiling the dynamics of the driving but unobserved Markov chain. Thus, the capacity to extract information in order to provide optimal model parameter estimates, by filtering out the noise from the data set, is the dominantly superior feature of HMM-based methods.

When dealing with real life-data, HMM could be employed as a statistical learning model to make classifications on time series according to its estimated parameters. See for instance, Li et al. [56]) who constructed an HMM-embedded device to detect anomalies in multivariate time series. In financial economics, where the manifestations of irregular or anomalous events are not apparently visible, an HMM-based model is an advantageous tool in the analysis of pertinent data. In Cao et al. [11], for example, an HMM with wavelet transformations and gradients was utilised to detect price manipulation activities in the stock markets.

Moreover, an HMM is the building block of a regime-switching-based technique in pinning down some underlying processes in finance or economics where they primarily evolve with the random shifts of their statistics (e.g., mean or variance) amongst different states. Thus, the HMM framework offers the flexibility that allows structural regimes, which are governed by the location, scale and shape parameters of a distribution, to shift stochastically over time. Such a framework is suited in capturing widely observed phenomena in finan-

cial economics, natural sciences, and engineering. The Markov-switching methodology in economics could be traced back to Hamilton's work [41], with the static estimation of the model addressed. The Markov-switching model designed mainly to describe structural changes in time series aims to (i) differentiate regimes of the economy and (ii) estimate the probability of being in an expansion or contraction.

In this work, the HMM modulates a mean-reverting OU process to capture the dynamic behaviour of an FSI. The estimation of HMM is performed via filtering methods, which are quite common in electronics and electrical engineering. The recursive filtering, which gives rise to a self-calibrating model, is an innovation relative to the models created in the past that are heavily dependent on static-model fitting approach of maximum likelihood estimation [59]. A change of reference probability measure, which is a dominant feature in our HMM-filtering procedure and pioneered by Elliott et al. [25], facilitates the development of a self-calibrating model. An accessible introduction of this method can be found in Mamon et al.[59].

Recent research progress in this HMM filtering framework highlights various implementations in the areas of quantitative finance, insurance, economics, epidemiology, and other branches of the sciences and engineering. In finance specifically, these include short-rate modelling (e.g., Elliott et al. [26], Elliott and Mamon [24], Erlwein and Mamon. [27], and Xi and Mamon [88]); investment strategies (Tenyakov et al. [74] and Erlwein et al. [28]; commodity price forecasting (Date et al. [17]); weather derivative pricing (Xiong and Mamon [91] and Xiong and Mamon [92]); liquidity risk forecasting(Gu et al.[40]) modelling electricity spot prices (Erlwein et al. [29]).

The major motivation in this work in the incorporation of the HMM-based model into our modelling process is to augment the capacities of learning the FSI's evolution. The turning points and pertinent evolution patterns could indicate possible financial-instability episodes. Therefore, numerical outcomes from our HMM filters could enable the detection of anomalies related to impending financial stress.

4.2.2 Random Forests

In statistical learning, an ensemble method combines multiple predictors to obtain predictive modelling performance that is better than simply using a single-component modelling algorithm. Random forests (RFs) [7] are an ensemble machine-learning framework for

classification and regression via the construction of multiple decision trees. The RF have become one of the most favoured supervised-learning models due to its considerable advantages, which are: (i) notable accuracy; (ii) robustness to outliers and noise; (iii) ease of use; and (iv) availability of internal estimates of error and variable importance [7]. In addition, the model has strong immunity from overfitting (see [7] [43]), especially for classification problems [43]. The algorithm in RFs used for classification is summarised in Algorithm 1.

The RFs are also aimed at enhancing the variance reduction of bagging (short for bootstrap aggregating) via building a large collection of de-correlated trees. This is attained by selecting variables randomly in the tree-growing procedure [43]. More specifically, all trees are grown to some bootstrap samples drawn from a training set. It can be shown that , when the sample size is large, approximately 63.2% of the instances in the training set are used to build one tree [1]. As a consequence, the remaining 36.8% of the data are used to evaluate the out-of-bag (OOB) error which is a good metric to measure the predictive power for each tree.

Algorithm 1: Random Forests for Classification.[29]

   Input: Labelled dataset
   for $i = 1$ to N **do**
      1. Draw a bootstrap sample $\mathbf{B}_i$ of size n from the training set.
      2. Grow a random forest tree T_i to the sample $\mathbf{B}_i$:
      for each terminal node **do**
         do
            (1) Select m variables randomly from the p variables.
            (2) Find the best variable and split-point.
            (3) Split the node into two daughter nodes.
         until the minimum node size is reached;
      end
   end
   Output: $\hat{P}(x \in A) = \frac{1}{N} \sum_i^N \mathbf{1}_{C_i(x)=A}$,
        where $C_i(x)$ is class determined by tree T_i for some data point x.

The OOB samples are also used to construct a measure for variable importance by evaluating the prediction power of each predictor. Once a tree is built, the OOB samples are processed by the tree, and the prediction accuracy is obtained. The algorithm randomly permutes the values for each single variable in the OOB samples to calculate the accuracy.

---

[1]Brief proof: Let n be the sample size of the training set. The probability of an instance not selected to build a tree is $1 - (1 - \frac{1}{n})^n \to 1 - e^{-1} \approx 0.632$ as n goes to infinity.

The difference between the accuracy values before and after permutation is averaged over all trees, which is treated as a measure for the variables' importance in the RF.

Owing to its superiority as pointed out above, an RF is widely employed as a predictive model to solve various regression and classifciation problems in time series analysis. Tyralis and Papacharalampous [78], Karasu and Altan [51] and Hao et al. [42] found that the predictive error could be reduced by selecting features using an RF on different time-series data sets. In this research, we shall employ RFs to identify variables that contribute more to the prediction of our target variable. These selected features will be passed on to the next-stage classifier to make the final prediction.

4.2.3   Extreme Gradient Boosting (XGBoost)

Boosting is another powerful learning-ensemble algorithm that aims to convert a set of weak learners, which could only extract a small amount information, into a strong learner [68]. The weak learners are weighted in some way related to their accuracy. The weights are updated after a weak learner is added. Input samples with poor learning result lead to a higher weight, which is then emphasised more and reflected when adding the next weak learner. Eventually, the final model is established by weighting all the component learners based on their performance. The final result of the boosting model is obtained by averaging and counting votes in the solutions to regression and classification problems, respectively.

The fundamental idea of XGBoost [13] is to predict the sum of scores from multiple classification and regression trees (CARTs) considering that one tree is usually not capable of capturing sufficient information from the data. XGBoost uses a tree ensemble model that consists of a group of CARTs [6]. It is a scalable end-to-end boosting system, which is widely recognised by data scientists to achieve state-of-the-art results and outperforms many other competing models tackling various machine-learning challenges [13]. XGBoost is based on a gradient-tree boosting [31] and entails a numerical optimisation approach to minimise the loss function via the addition of trees in the gradient-descent algorithm. More specifically, new trees are built and trained on the errors of that previous model so that more importance are given to the observations in which existing learners were misclassified. A strong learner is updated by adding a new trained tree that generates a new prediction, whereby each tree's contribution is determined by minimising the overall error of the strong learner.

Following Chen and Guestrin [13], XGBoost could be described in the following way. For a given data set with n instances and m features, a tree-ensemble model uses K additive functions to predict the target variable y. That is,

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \tag{4.1}$$

where $f_k \in \mathcal{F}$ and $\mathcal{F} = \{f(x) = w_{q(x)} \mid q : R^m \rightarrow T, w \in R^T\}$ is a functional space that includes all possible CARTs. Note that $q(\cdot)$ is the structure of each tree that maps an instance to the corresponding leaf index. Also, T represents the number of leaves in the tree. Each f_k is related to an independent tree structure q and the leaf weight w.

Instead of growing multiple trees at once, the model is trained in an additive manner. To clarify this further, let $y_i^{(t)}$ be the prediction of the ith instance at the tth step, and f_t is added to minimise the objective function of the form given by (4.2). In other words, f_t is added in a greedy way that provides the optimal improvement to the model. In order to optimise the loss function efficiently, Chen and Guestrin [13] use a Taylor approximation through the first and second-order gradients g_i and h_i in Eq. (4.3); in turn, more information concerning the gradients' directions are gained. Taking into account the above-mentioned descriptions, the loss function is given by

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{4.2}$$

$$\approx \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t), \tag{4.3}$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$; $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$; and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$.

The function l in (4.2) is convex and continuously twice differentiable; it quantifies the difference between the true value y_i and the prediction $\hat{y}_i$. In order to avoid over-fitting, there is a regularisation term Ω in (4.2) that effectively penalises the complexity of the model. Thus, this regularisation incentivises in selecting a simple model with strong predictive power.

The major advantages and features of XGBoost are: (i) gradient-boosting-method implementation with the utility of the second-order partial derivatives, which provides more accurate approximations of the loss function; (ii) presence of regularisation that enhances model's predictive abilities; and (iii) capacity to incorporate parallel and distributed computing which makes learning faster than existing algorithms [13].

XGBoost became popular in tackling various real-life data-analysis problems. Zhang et al. [94] proposed an XGBoost-based classifier to detect indoor-human activities based on time series collected via smart-phone sensors. In Basak et al. [3], it was found that models based on XGBoost and RFs are able to achieve high-accuracy predictions of the stock-price movements' direction in the medium to long term. Nobre and Neves [64] constructed an automated stock-trading system which combines Principal Component Analysis (PCA), Discrete Wavelet Transform (DWT), XGBoost and a Multi-Objective Optimisation Genetic Algorithm (MOO-GA).

In this work, we shall employ XGBoost as a final classifier to detect anomalies related to that task of coming up with multiple-step-ahead prediction regarding the movements of all FSI time series for each country in the data set.

4.3 The integrated modelling algorithm construction

4.3.1 Model structure

This work presents an automated early-warning system designed to detect the anomaly in the FSI. This is accomplished by predicting the probability of occurrence of a rising edge in the next 6 months. A supervised learning framework consisting of HMM filters, RF and XGBoost binary classifier is proposed. Figure 4.1 illustrates a schematic layout of our modelling approach, which is implemented in the R programming language.

4.3.2 Input layer

4.3.2.1 Dataset

The data set in this work is taken from Duprey et al. [21] comprising the monthly FSIs of 17 countries for the period Jan 1968–Sep 2019 (i.e., a matrix of 621×17 data points). The descriptive statistics and the graphical evolution of the multi-dimensional time series are illustrated in Table 4.1 and Figure 4.2, respectively. Note that the range of FSI values is from 0 (no financial stress) to 1 (extremely high financial stress).
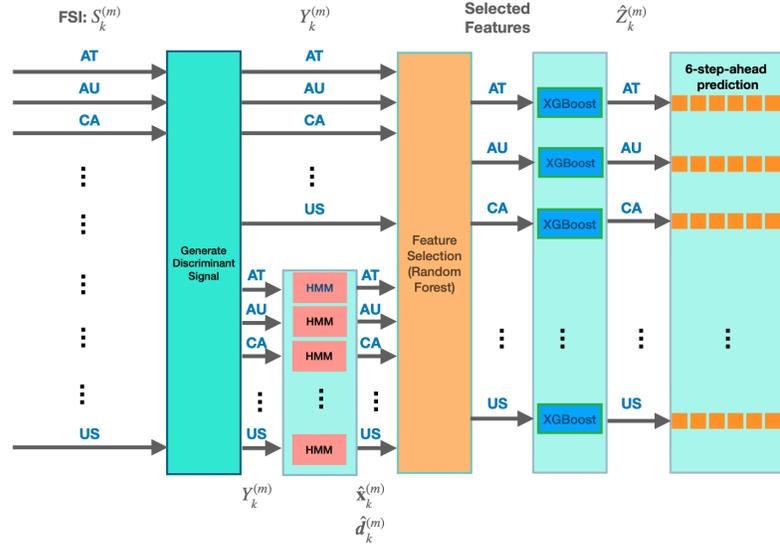
Figure 4.1: The architecture of the proposed modelling approach

4.3.2.2 Discriminative signal

Let S_t denote the FSI process. We shall forecast S_t's movement which may signal financial stress in the future. More precisely, at a given time t, we need to determine whether or not the time series will have a rising edge between $t + 1$ and $t + h$ inclusive. To carry this out, the original input undergoes the transformation

$$Y_t = \int_{t-\tau}^{t} dS_u = S_t - S_{t-\tau}, \tag{4.4}$$

where τ is the length of the moving window. In Eq. (4.4), Y_t has two interpretations: (i) the difference in the S_t values between time t and time $t - \tau$; and (ii) the cumulative sum of S_t changes over the interval $[t - \tau, t]$.

Since FSI is a monthly data series, we set $\tau = 12$. Thus, Y_t measures the aggregate elevation of the index in the last 12 months; it could be viewed as the FSI's year-over-year net growth. Thus, Y_t's progression also traces dynamically the movement of S_t. The evolution of Y_t is exhibited (red curve) in Figure 4.2.

4.3.2.3 Target variable formulation
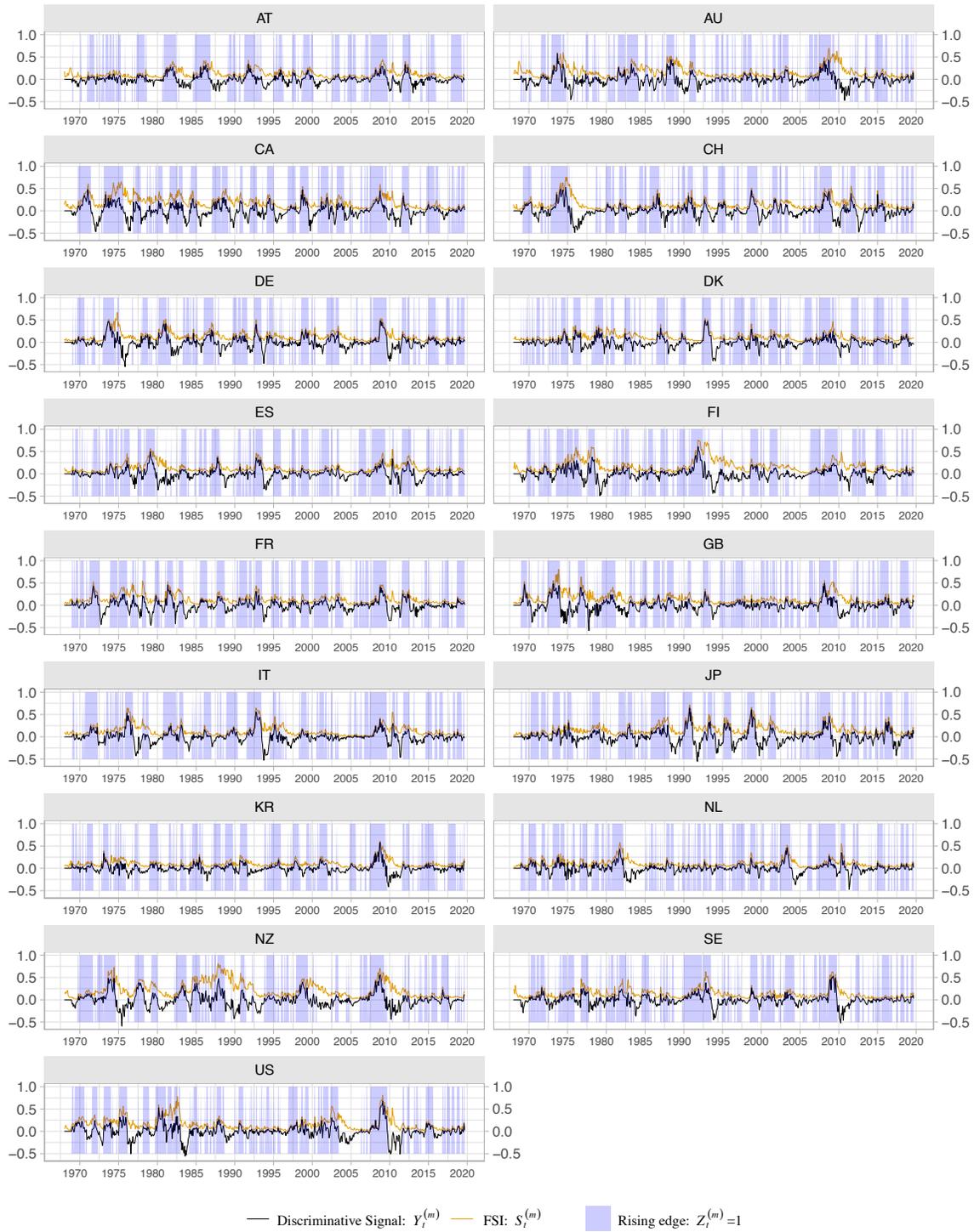
Figure 4.2: The FSIs and discriminative signals

Table 4.1: Descriptive statistics of the FSI data series

Code	Country	Mean	SD	Median	Minimum	Maximum	Skewness	Kurtosis
AT	Australia	1.209×10^{-1}	8.520×10^{-2}	9.457×10^{-2}	7.414×10^{-3}	4.411×10^{-1}	1.398×10^{0}	1.572×10^{0}
AU	Austria	1.643×10^{-1}	1.271×10^{-1}	1.219×10^{-1}	1.175×10^{-2}	6.990×10^{-1}	1.476×10^{0}	1.826×10^{0}
CA	Canada	1.933×10^{-1}	1.288×10^{-1}	1.588×10^{-1}	1.803×10^{-2}	6.526×10^{-1}	9.497×10^{-1}	5.072×10^{-1}
CH	Switzerland	1.475×10^{-1}	1.226×10^{-1}	1.099×10^{-1}	2.114×10^{-2}	7.528×10^{-1}	2.055×10^{0}	4.710×10^{0}
DE	Germany	1.354×10^{-1}	1.056×10^{-1}	1.024×10^{-1}	1.938×10^{-2}	6.697×10^{-1}	1.749×10^{0}	3.309×10^{0}
DK	Denmark	1.188×10^{-1}	8.288×10^{-2}	9.205×10^{-2}	1.455×10^{-2}	5.339×10^{-1}	1.644×10^{0}	3.401×10^{0}
ES	Spain	1.305×10^{-1}	9.901×10^{-2}	9.561×10^{-2}	1.966×10^{-2}	5.669×10^{-1}	1.734×10^{0}	3.089×10^{0}
FI	Finland	1.735×10^{-1}	1.382×10^{-1}	1.334×10^{-1}	1.962×10^{-2}	7.565×10^{-1}	1.713×10^{0}	3.191×10^{0}
FR	France	1.424×10^{-1}	9.978×10^{-2}	1.123×10^{-1}	2.052×10^{-2}	5.475×10^{-1}	1.414×10^{0}	1.806×10^{0}
GB	United Kindom	1.487×10^{-1}	1.164×10^{-1}	1.137×10^{-1}	1.167×10^{-2}	8.050×10^{-1}	1.869×10^{0}	4.421×10^{0}
IT	Italy	1.318×10^{-1}	1.139×10^{-1}	9.333×10^{-2}	8.979×10^{-3}	6.472×10^{-1}	1.995×10^{0}	4.745×10^{0}
JP	Japan	1.586×10^{-1}	1.152×10^{-1}	1.214×10^{-1}	1.842×10^{-2}	7.142×10^{-1}	1.528×10^{0}	2.557×10^{0}
KR	Korea	1.064×10^{-1}	7.808×10^{-2}	8.500×10^{-2}	1.228×10^{-2}	6.095×10^{-1}	2.368×10^{0}	8.225×10^{0}
NL	Netherlands	1.111×10^{-1}	8.624×10^{-2}	8.700×10^{-2}	1.654×10^{-2}	5.814×10^{-1}	2.186×10^{0}	6.203×10^{0}
NZ	New Zealand	2.251×10^{-1}	1.720×10^{-1}	1.657×10^{-1}	1.728×10^{-2}	8.081×10^{-1}	1.040×10^{0}	2.668×10^{-1}
SE	Sweden	1.334×10^{-1}	1.040×10^{-1}	1.023×10^{-1}	2.117×10^{-2}	6.292×10^{-1}	2.069×10^{0}	5.236×10^{0}
US	United States	1.640×10^{-1}	1.400×10^{-1}	1.195×10^{-1}	1.140×10^{-2}	8.093×10^{-1}	1.635×10^{0}	2.828×10^{0}

Define the binary target variable Z_t as

$$Z_t = \begin{cases} 1 \,(\text{Positive}), & \text{if } Y_t > 0, \\ 0 \,(\text{Negative}), & \text{if } Y_t \leq 0, \end{cases} \tag{4.5}$$

From Eq. (4.5), $Z_t = 1$ means that, at time t, FSI is growing relative to its value at time $t - \tau$ or, equivalently, it has positive net growth in the time interval $[t - \tau, t]$ which indicates that the rising possibility of FSI is high. The dynamics of Z_t are visualised as shaded areas (light purple) in Figure 4.2. It is apparent that the major increasing edges of FSI are captured by the target variable. That is, the FSI is increasing or going to rise when $Z_t = 1$. The main purpose of this work is to propose a model to forecast the value of Z_t in the next h time periods, i.e., Z_{t+h}. Thus, for regulators, this quantity could serve as an indicator for both the direction of the FSI in the short run and a country's exposure to financial instability. The distributions of Z_t for each country are presented in Figure 4.3 as a percent-stacked bar chart. It is worth noting that all the 17 data series are nearly balanced. That is each of them contains equal or almost equal number of samples from the positive and negative class corresponding to the value of the target variable Z_t.
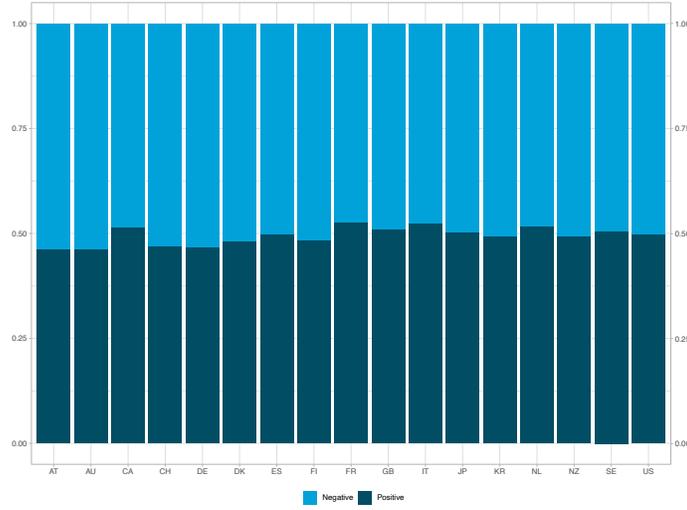
Figure 4.3: Distribution of the Z_t values

4.3.3 HMM filtering module

4.3.3.1 Ornstein-Uhlenbeck (OU) process

Figure 4.2, showing the discriminative signals Y_t derived from all the country level FSIs, portrays a mean-reverting behaviour, dynamically and randomly moving from high to low level and vice-versa. This fact provides support for an OU process as a means to capture the signals' mean-reverting attribute.

Suppose the discriminative signal Y_t in Eq. (4.4) Y_t follows an OU process in accordance with the stochastic differential equation

$$\mathrm{d}Y_t = \theta\,(\mu - Y_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t, \tag{4.6}$$

where μ represents the mean-reverting level, θ is the speed of mean reversion, and σ is the volatility. We assume that the parameters θ, μ and σ are positive. The standard Brownian motion B_t is defined on some probability space $\left(\Omega, \mathscr{F}^{Y_t}, P\right)$, where $\mathscr{F}^{Y_t}$ is the filtration generated by Y_t.

By Itô's lemma, it may be verified that the solution to Eq. (4.6) is

$$Y_t = Y_0 e^{-\theta t} + (1 - e^{-\theta t})\mu + \sigma e^{-\theta t} \int_0^t e^{\theta s}\mathrm{d}B_s. \tag{4.7}$$

To implement Eq. (4.7) empirically, we require its discretisation given by

$$Y_{k+1} = Y_k e^{-\theta \cdot \Delta t} + (1 - e^{-\theta \cdot \Delta t})\mu + \sqrt{\frac{\sigma^2}{2\theta}(1 - e^{-2\theta \cdot \Delta t})} \, B_{k+1}, \tag{4.8}$$

where $\Delta t = (k + 1) - k = 1$ and $\{B_{k+1}\}$ is a sequence of independent and identically distributed standard normal random variables. The third term in Eq. (4.8) is justified by the Gaussian property and the Itô's isometry.

4.3.3.2 HMM-governed OU process

To equip our modelling approach with the capability for time-varying or stochastic parameters, a hidden Markov-chain model (HMM) modulates the OU process in Eq. (4.6). In this work, we treat the states of the HMM as the country's financial-stress regime, i.e., normal or anomalous, which is an actualisation of interacting factors that cause financial instability.

To make the FSI regime-switching, the parameters θ, μ and σ are governed by a discrete-time Markov chain $\mathbf{x}_k$, for $k = 0, 1, \ldots$. The state space of $\mathbf{x}_k$ is finite and it is isomorphic to the canonical basis of $\mathbb{R}^N$, which is the set $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_N\}$. The vector $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ is a unit vector with 1 in its ith component. The semi-martingale representation of $\mathbf{x}_k$ is

$$\mathbf{x}_{k+1} = \boldsymbol{\Pi}_k \mathbf{x}_k + \mathbf{v}_{k+1}, \tag{4.9}$$

where $\boldsymbol{\Pi}_k$ is a transition matrix; $\mathbf{v}_{k+1}$ is a martingale increment with $E[\mathbf{v}_{k+1}|\mathscr{F}^{\mathbf{x}_k}] = \mathbf{0}$; and $\mathscr{F}^{\mathbf{x}_k}$ is the filtration generated by $\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_k$. The dependence of the model parameters on the HMM should be clear from the notations $\theta(\mathbf{x}_k)$, $\mu(\mathbf{x}_k)$ and $\sigma(\mathbf{x}_k)$.

With M as the dimension of the multivariate time series, we have $M = 17$ for the data series with summary statistics in Table 4.1. For the mth component of the M-dimensional discriminative signal, i.e., $Y_t^{(m)}$, its parameters are driven by some HMM $\mathbf{x}_k^{(m)}$ that evolves as per Eq. (4.9) so that

$$Y_{k+1}^{(m)} = Y_k^{(m)} e^{-\theta(\mathbf{x}_k^{(m)})\Delta t} + (1 - e^{-\theta(\mathbf{x}_k^{(m)})\Delta t})\mu(\mathbf{x}_k^{(m)}) + \sqrt{\sigma^2(\mathbf{x}_k^{(m)}) \cdot \frac{1 - e^{-2\theta(\mathbf{x}_k^{(m)})\Delta t}}{2\theta(\mathbf{x}_k^{(m)})}} \, b_{k+1}^{(m)}. \tag{4.10}$$

In Eq. (4.10), $\forall m = 1, 2, \ldots, M$,

$$\begin{cases} \mu(\mathbf{x}_k^{(m)}) = \langle \boldsymbol{\mu}_k^{(m)}, \mathbf{x}_k^{(m)} \rangle \\ \theta(\mathbf{x}_k^{(m)}) = \langle \boldsymbol{\theta}_k^{(m)}, \mathbf{x}_k^{(m)} \rangle \\ \sigma^2(\mathbf{x}_k^{(m)}) = \langle \boldsymbol{\sigma}_k^{2(m)}, \mathbf{x}_k^{(m)} \rangle, \end{cases} \tag{4.11}$$

with

$$
\begin{cases}
\boldsymbol{\mu}_k^{(m)} = (\mu_{k,1}^{(m)}, \mu_{k,2}^{(m)}, ..., \mu_{k,N}^{(m)})^\top \\
\boldsymbol{\theta}_k^{(m)} = (\theta_{k,1}^{(m)}, \theta_{k,2}^{(m)}, ..., \theta_{k,N}^{(m)})^\top \\
\boldsymbol{\sigma}_k^{2(m)} = (\sigma_{k,1}^{2(m)}, \sigma_{k,2}^{2(m)}, ..., \sigma_{k,N}^{2(m)})^\top
\end{cases}
\tag{4.12}
$$

where $\boldsymbol{\mu}_k^{(m)}$, $\boldsymbol{\theta}_k^{(m)}$ and $\boldsymbol{\sigma}_k^{2(m)}$ are all in $\mathbb{R}^{\mathbb{N}}$; and $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^N$.

4.3.3.3 Parameter estimation

Eq. (4.10) is re-parameterised as

$$
Y_{k+1}^{(m)} = \alpha(\mathbf{x}_k^{(m)})Y_k^{(m)} + \beta(\mathbf{x}_k^{(m)}) + \zeta(\mathbf{x}_k^{(m)}) \, b_{k+1}^{(m)},
\tag{4.13}
$$

where

$$
\begin{cases}
\alpha(\mathbf{x}_k^{(m)}) = e^{-\theta(\mathbf{x}_k^{(m)})\Delta t} \\
\beta(\mathbf{x}_k^{(m)}) = (1 - e^{-\theta(\mathbf{x}_k^{(m)})\Delta t})\mu(\mathbf{x}_k^{(m)}) \\
\zeta^2(\mathbf{x}_k^{(m)}) = \sigma^2(\mathbf{x}_k^{(m)}) \cdot \dfrac{1 - e^{-2\theta(\mathbf{x}_k^{(m)})\cdot\Delta t}}{2\theta(\mathbf{x}_k^{(m)})}.
\end{cases}
\tag{4.14}
$$

Eq. (4.13) implies that, given $Y_k^{(m)}$,

$$
Y_{k+1}^{(m)} \sim N\left(\alpha(\mathbf{x}_k^{(m)})Y_k^{(m)} + \beta(\mathbf{x}_k^{(m)}), \ \zeta^2(\mathbf{x}_k^{(m)})\right).
\tag{4.15}
$$

Let $\hat{\mathbf{x}}_k^{(m)}$ be the conditional expectation of the state of hidden Markov chain $\mathbf{x}_k^{(m)}$ given $\mathscr{F}_k^{(m)}$ under probability measure $P^{(m)}$. Thus,

$$
\hat{\mathbf{x}}_k^{(m)} := E\left[\mathbf{x}_k^{(m)} \middle| \mathscr{F}_k^{(m)}\right] = \left(\hat{x}_{k,1}^{(m)}, \hat{x}_{k,2}^{(m)}, ..., \hat{x}_{k,N}^{(m)}\right)^\top,
\tag{4.16}
$$

where

$$
\hat{x}_{k,i}^{(m)} = P\left(\mathbf{x}_k^{(m)} = \mathbf{e}_i \middle| \mathscr{F}_k^{(m)}\right) = E\left[\langle \mathbf{x}_k^{(m)}, \mathbf{e}_i \rangle \middle| \mathscr{F}_k^{(m)}\right].
\tag{4.17}
$$

We shall utilise the EM algorithm to find the estimate of the parameters of the HMM-driven model, which is a component of our hybridised modelling algorithm. See Elliott and Krisnamurthy [23] as well as Wu [86] for the details of the EM algorithm.

We further define the following quantities for states $j = 1, 2 ..., N$:

$$
\begin{cases}
\mathcal{G}_k^{s,j,(m)} = \sum_{n=1}^k \langle \mathbf{x}_{n-1}^{(m)}, \mathbf{e}_j \rangle \langle \mathbf{x}_n^{(m)}, \mathbf{e}_s \rangle \\
\mathcal{O}_k^{j,(m)} = \sum_{n=1}^k \langle \mathbf{x}_{n-1}^{(m)}, \mathbf{e}_j \rangle \\
\mathcal{T}_k^{j,(m)}(f) = \sum_{n=1}^k \langle \mathbf{x}_{n-1}^{(m)}, \mathbf{e}_j \rangle f(\cdot).
\end{cases}
\tag{4.18}
$$

The quantities $\mathcal{G}_k^{s,j,(m)}$ and $\mathcal{O}_k^{j,(m)}$ represent the respective number of jumps from state j to s and the amount of time that the Markov chain $\{\mathbf{x}_n^{(m)}\}$ occupies the state j up to time k, respectively. Also, $\mathcal{T}_k^{j,(m)}$ is an auxiliary process that depends on the function $f(\cdot)$. In our empirical application, $f(\cdot)$ takes the form $Y_n^{(m)}$, $(Y_n^{(m)})^2$ and $Y_n^{(m)}Y_{n-1}^{(m)}$. Following the procedure introduced in [59] and [75], we obtain the optimal parameter estimates below.

$$\hat{\mathbf{x}}_k^{(m)} = \frac{\mathbf{u}_k^{(m)}}{\sum_{i=1}^{N}\langle \mathbf{u}_k^{(m)}, \mathbf{e}_i \rangle}. \tag{4.19}$$

$$\hat{\pi}^{j,i,(m)} = \frac{\hat{\mathcal{G}}_k^{j,i,(m)}}{\hat{\mathcal{O}}_k^{i,(m)}} \tag{4.20}$$

$$\hat{\alpha}_{k,i}^{(m)} = \frac{\hat{\mathcal{T}}_k^{i,(m)}\left(Y_k^{(m)}Y_{k-1}^{(m)}\right) - \hat{\beta}_{k-1,i}^{(m)} \cdot \hat{\mathcal{T}}_k^{i,(m)}\left(Y_{k-1}^{(m)}\right)}{\hat{\mathcal{T}}_k^{i,(m)}\left(Y_{k-1}^{(m)\,2}\right)} \tag{4.21}$$

$$\hat{\beta}_{k,i}^{(m)} = \frac{\hat{\mathcal{T}}_k^{i,(m)}\left(Y_k^{(m)}\right) - \hat{\alpha}_{k-1,i}^{(m)} \cdot \hat{\mathcal{T}}_k^{i,(m)}\left(Y_{k-1}^{(m)}\right)}{\hat{\mathcal{O}}_k^{i,(m)}} \tag{4.22}$$

$$\widetilde{\zeta_{k,i}^{2}}^{(m)} = \frac{\hat{\mathcal{T}}_k^{i,(m)}\left(Y_k^{(m)2}\right) + \hat{\alpha}_{k-1,i}^{(m)} \cdot \hat{\mathcal{T}}_k^{i,(m)}\left(Y_{k-1}^{(m)\,2}\right)}{\hat{\mathcal{O}}_k^{i,(m)}}$$

$$+ \frac{\left(\hat{\beta}_{k-1,i}^{(m)}\right)^2 \hat{\mathcal{O}}_k^{i,(m)} + 2\alpha_{k-1,i}^{(m)}\hat{\beta}_{k-1,i}^{(m)}\hat{\mathcal{T}}_k^{i,(m)}\left(Y_{k-1}^{(m)}\right)}{\hat{\mathcal{O}}_k^{i,(m)}}$$

$$- \frac{2\hat{\alpha}_{k-1,i}^{(m)}\hat{\mathcal{T}}_k^{i,(m)}\left(Y_k^{(m)}Y_{k-1}^{(m)}\right)}{\hat{\mathcal{O}}_k^{i,(m)}}$$

$$- \frac{2\hat{\beta}_{k-1,i}^{(m)}\hat{\mathcal{T}}_k^{i,(m)}\left(Y_k^{(m)}\right)}{\hat{\mathcal{O}}_k^{i,(m)}} \tag{4.23}$$

The details of how we develop the HMM filtering and estimation are presented in Appendices E and F. For the original model parameters specified in Eq. (4.10), we use Eqs. (4.20) – (4.23) to get the following optimal values:

$$\begin{cases} \hat{\mu}_{k,i}^{(m)} = \dfrac{\hat{\beta}_{k,i}^{(m)}}{1 - \hat{\alpha}_{k,i}^{(m)}} \\[4mm] \hat{\theta}_{k,i}^{(m)} = \dfrac{\ln(\hat{\alpha}_{k,i}^{(m)})^{-1}}{\Delta t} \\[6mm] \widetilde{\sigma_{k,i}^{2}}^{(m)} = 2\hat{\theta}_{k,i}^{(m)} \cdot \widetilde{\zeta_{k,i}^{2}}^{(m)}\left(1 - e^{-2\hat{\theta}_{k,i}^{(m)}\cdot\Delta t}\right). \end{cases} \tag{4.24}$$

4.3.3.4 Feature construction: Markov-chain state estimate

Tenyakov et al. [76] and Gu et al. [40] find the state of the hidden Markov chain filtered from the given data series carries essential information that can be treated as an early warning signal of financial instability events. In this work, our underlying process is $Y_t^{(m)}$ which dynamically trace the trend of corresponding FSI. It's intuitive to consider that the HMM state estimate $\mathbf{x}_k^{(m)}$, which drives the OU process $Y_t^{(m)}$, is able to provide salient guidance to forecast the movement direction of FSI.

Equation (4.17) defines the estimate of hidden Markov chain $\mathbf{x}_k^{(m)}$ state at time k. The calculation formula is given in Eq. (E.6) in Appendix E. According to the set up in Eq. (4.16), we can write the ith component of $\hat{\mathbf{x}}_k^{(m)}$ as follows:

$$\hat{x}_{k,i}^{(m)} = \langle \hat{\mathbf{x}}_k^{(m)}, \, e_i \rangle, \quad \forall i \in \{1, 2, \cdots, N\} \tag{4.25}$$

Findings obtained by previous researches such as Erlwein et al. [29], Tenyakov et al. [76] and Xiong and Mamon [91] indicate that a 2-state HMM set up is sufficient to express the evolution of the underlying data series. Thus we set numbers of state $N = 2$ for all Markov chains in this work. We shall include the Markov-chain state estimates which related to anomalous state, i.e., $\hat{x}_{k,1}^{(m)}$, in the input set for next stage modelling. The evolutions of $\hat{x}_{k,1}^{(m)}$ are illustrated in Figure G.1 in Appendix G.

4.3.3.5 Feature construction: Deviation from the reverting mean

The evolution of $Y_t^{(m)}$ is captured by the outcomes of the HMM filters, which are used to predict its future dynamics. From Eq. (4.10),

$$\hat{Y}_{k+1}^{(m)} - \mu\left(\mathbf{x}_k^{(m)}\right) = E\left[Y_{k+1}^{(m)} \big| \mathscr{F}_k^{(m)}\right] - \mu\left(\mathbf{x}_k^{(m)}\right)$$
$$= \left(Y_k^{(m)} - \mu(\mathbf{x}_k^{(m)})\right) \cdot e^{-\theta(\mathbf{x}_k^{(m)})\Delta t}. \tag{4.26}$$

Equation (4.26) implies that the difference between $Y_k^{(m)}$ and its mean could be predicted using the HMM filtered parameters.

Write

$$d\left(\mathbf{x}_k^{(m)}\right) := \left(Y_{t_k}^{(m)} - \mu\left(\mathbf{x}_k^{(m)}\right)\right) \cdot e^{-\theta(\mathbf{x}_k^{(m)})\Delta t}. \tag{4.27}$$

With the formulation of notation in Eq. (4.11),

$$d\left(\mathbf{x}_k^{(m)}\right) = \langle \boldsymbol{d}_k^{(m)}, \, \mathbf{x}_k^{(m)} \rangle, \tag{4.28}$$

where

$$\boldsymbol{d}_k^{(m)} = (d_{k,1}^{(m)}, \, d_{k,2}^{(m)}, \, ..., \, d_{k,N}^{(m)})^{\top} \tag{4.29}$$

and

$$d_{k,i}^{(m)} = \left(Y_k^{(m)} - \mu_{k,i}^{(m)}\right) \cdot e^{-\theta_{k,i}^{(m)}\Delta t}, \quad \forall i \in \{1, 2, \cdots, N\}. \tag{4.30}$$

At time k, we obtain

$$\hat{d}_{k,i}^{(m)} = E\left[d_{k,i}^{(m)}\big|\mathscr{F}_k^{(m)}\right] = \left(Y_k^{(m)} - \hat{\mu}_{k,i}^{(m)}\right) \cdot e^{-\hat{\theta}_{k,i}^{(m)}\Delta t}, \quad \forall i \in \{1, \cdots, N\}. \tag{4.31}$$

In addition to the discriminative signal, we incorporate $\hat{d}_{k,i}^{(m)}$ along with the Markov-chain state estimate $\hat{x}_{k,1}^{(m)}$ as the initial inputs in our proposed modelling approach. The evolutions of $\hat{d}_{k,i}^{(m)}$ are presented in Figure G.2 in Appendix G.

4.3.4   Random-Forest-based feature-selection module

4.3.4.1   Model set up

The modelling-component set ups in Subsections 4.3.2 and 4.3.3 signify that the input features to the RF module can be divided into three categories: (i) discriminative signals, (ii) deviation from the reverting mean derived from the HMM filters, and (iii) Markov-chain state estimates. Each category constitutes a set of features of different countries with a range of time lags.

Our hybridised model is defined as follows:

$$Z_{k+h}^{(m)} = G_h^{(m)}\left(\mathbf{Y}_k, \mathbf{D}_k, \mathbf{X}_k|\mathscr{F}_k\right) \tag{4.32}$$

with

$$\begin{cases} \mathbf{Y}_k = \left\{Y_{k-q}^{(m)}\big|m = 1, 2, ..., 17; q = 0, 1, ..., (T-1).\right\} \\ \mathbf{D}_k = \left\{\hat{d}_{k-q,i}^{(m)}\big|i = 1, 2; m = 1, 2, ..., 17; q = 0, 1, ..., (T-1).\right\} \\ \mathbf{X}_k = \left\{\hat{x}_{k-q,1}^{(m)}\big|m = 1, 2, ..., 17; q = 0, 1, ..., (T-1).\right\}. \end{cases} \tag{4.33}$$

In Eq. (4.32), $Z_{k+h}^{(m)}$, as previously defined in equation (4.5), is the binary target variable corresponding to $Y_k^{(m)}$. The function $G_h^{(m)}(\cdot)$ is unknown and does not have a closed form although it could be estimated. Furthermore, q represents the memory length of the model with boundary T. Considering the monthly frequency of the data series, we set $T = 12$; that is, the proposed model has a memory length of 12 months. More specifically, for each country in the data set, we let the corresponding target variable $Z_k^{(m)}$ depends on the those three groups of predictors of itself as well as others with the maximum lag of 12 months. In this work, we use the first n ($n = 90$) rows of sample points to initialise our model.

The model for $G_h^{(m)}(\cdot)$ is essential in making multi-step predictions on the binary target variable Z_t. A recursive-filtering strategy is typically utilised in obtaining multi-step-ahead predictions. For a

time series x_t, the procedure in making the h-step-ahead prediction at time k is as follows: (i) Train the model on $\{x_1, x_2, \ldots,, x_k\}$; (ii) Predict x_{k+1} using the previously trained model; (iii) Predict x_{k+2} with the same trained model using a data set including the previously predicted x_{k+1}. The procedure continues until x_{k+h} is predicted.

Under a recursive prediction framework, the model is only trained once based on a 1-step-ahead optimisation criterion, which could make the model fail in capturing the temporal dynamics [73]. We note that incorporating the estimated value (i.e., prediction for last time point) in the inputs inevitably introduces cumulative errors in the long-horizon forecasting [81]. This recursive prediction's limitation is further confirmed in [15] in which propagation errors are sustained when tested on real-world data set.
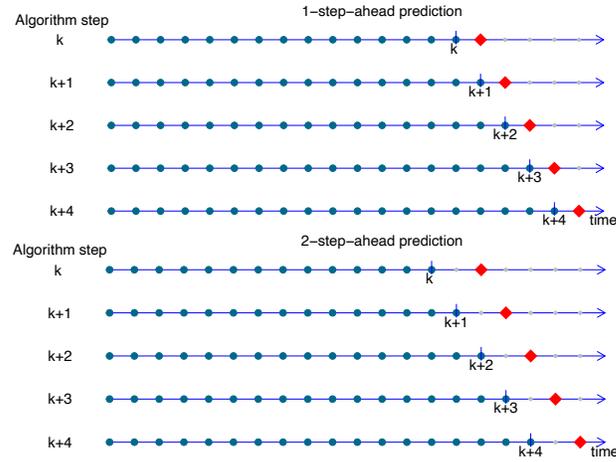
Figure 4.4: A demonstration of model training and prediction: 1 and 2-step-ahead predictions

In this chapter, we apply a direct prediction method for multi-step forecasting. Figure 4.4 diagrams the training and test schemes for 1-step-ahead and 2-step-ahead predictions. At each time point, we build a specific predictive model to generate a single output for an $h-$step-ahead prediction. More precisely, at time k, for all m data series and all h-step-ahead ($h = 1, 2, \ldots, 6$) forecasting tasks, a particular model $G_h^{(m)}(\cdot)$ is calibrated on a training set (marked by dark cyan dots in Figure 4.4) with a sample size of $(n + k - h)$ to generate a prediction $\hat{Z}_{k+h}^{(m)}$ (marked by large red diamonds in Figure 4.4). Such a prediction is a probabilistic value classifying the current data point as either positive (i.e., anomalous) or not.

4.3.4.2 Selection procedure

Tyralis and Papacharalampous [78] employed an RF to efficiently select the lag feature of a time series. In our case, we also use an RF in the selection of lag variables as well as country predictors.

More specifically, at time k, for every $Z_{k+h}^{(m)}$ with $m = 1, 2, ..., 17$, $h = 1, 2, \ldots, 6$, we first fit the model defined in Eq. (4.32) to the data with all the variables included in Eq. (4.33). After each round, the features' importance are calculated. We select the predictors with top-20 highest scores. The mean decrease of accuracy (i.e., OOB) is chosen as the measure of feature importance. The R function randomForest is utilised to implement the selection process.

We assign the number of trees to grow (ntree) to a value of 500. The number of variables selected as candidates at each split (mtry) is set to $p/2$, where p is the total number of predictors. To minimise the computation time of the feature selection, all the other parameters are set as the default value of the function. The predictors selected by RF are then passed to the XGBoost classifier to produce a final prediction.

4.3.5   Extreme Gradient Boosting (XGBoost) classification module

The purpose of the XGBoost classification module is the generation of predictions $\hat{Z}_{k+h}^{(m)}$ at time k for $h = 1, 2, \ldots, 6$. The input of the module includes the past target-variable series $Z_k^{(m)}$ and the features selected by the RF module. The output $\hat{Z}_{k+h}^{(m)}$, lying between 0 and 1, is the estimated probability that $Z_{k+h}^{(m)} = 1$, which could be viewed as the country's FSI having a rising edge at time $k + h$ as well as the country's estimated level of exposure to financial instability. The XGBoost is implemented with the R functions "xgb.train" and "xgb.cv" from "xgboost" package.

4.3.5.1   XGBoost model set up

Before utilising the XGBoost classifier, it is necessary to set the parameters that will direct the model's learning process. In Table 4.2, the model parameters' values or specified subsets of parameter space are presented; these information are based on empirical experience concerning XGBoost implementation in practice. Those parameters with potential values will be tuned via cross validation.

4.3.5.2   Hyper-parameter tuning

Five parameters in Table 4.2 are hyper-parameters: nrounds, eta, max_depth, gamma and alpha. Since these parameters could not be estimated directly from the data, the optimal values are searched in such a way that the pre-defined loss function is minimised, which consequently yield good out-of-sample prediction results.

Table 4.2: XGBoost model parameters' setting

Parameters	Value	Description
booster	gbtree	Booster to use
objective	binary:logistic	Objective function
eval_metric	logloss	Cross validation metric
nrounds	100	The number of boosting iteration
eta	{0.02, 0.06, 0.08, 0.1}	Learning rate
max_depth	{4, 5, 8}	Maximum depth of a tree
gamma	{0, 0.1, 0.3, 0.5}	Minimum loss reduction for a new node
alpha	{0.01, 0.02, 0.03, 0.04}	L1 regularisation term
others	Default	–

A gridsearch is a typical method in machine learning for tuning hyper-parameters. The basic procedure consists of the following steps: (i) Generate a list that incorporates all the combinations of hyper-parameters that are candidates for tuning; (ii) Define J training and validation partitions of the original data; (iii) For each $j = 1, 2, \ldots, J$, fit the models with all hyper-parameters setting defined in (i) on the training set; (iv) Make predictions on the validation set and calculate all the error metrics for all models; (v) Select the hyper-parameters for which the model gives the best performance, i.e., having the smallest log loss.

Note that the error metric used in the grid search algorithm is typically measured by cross validation. Bergmeir and Benítez [5] found that applying the traditional cross-validation methods on time series data will lead to both theoretical and practical problems , and as a consequence, fail to provide appropriate guidance to select the optimal parameters. Hyndman and Athanasopoulos [47] introduced a cross-validation technique tailored to time series and this is known as "evaluation on a rolling forecasting origin". In the literature, we implement this cross-validation method to make the training and validation partitions in step (ii) of the grid-search procedure. The process is illustrated in the Figure 4.5 with the examples of 1-step-ahead and 2-step-ahead prediction cases.

More specifically, our model makes an h-step-ahead prediction $\hat{Z}_{k+h}^{(m)}$ all data available up to time k as inputs. The dark cyan dots along the horizontal data axis in Figure 4.5 represent all the data available at time k whilst the large red diamonds stand for the corresponding prediction data point. To search for the optimal hyper-parameters we need to create J training and validation partitions of the data set known at time k. We set $J = 5$ considering the balance of search efficiency and effectiveness. As exhibited in Figure 4.5, each one of the short blue axes cv1 to cv5 shows a training-validation partition in the cross validation. For each $j = 1, 2, \ldots, J$, (i) the data points with time index from 1 to $k - J + j - h$ are in the training set (small light blue dots on the short axes) for cross validation; (ii) those with time index from $k - J + j$ to k are in the validation set (small red triangles along the
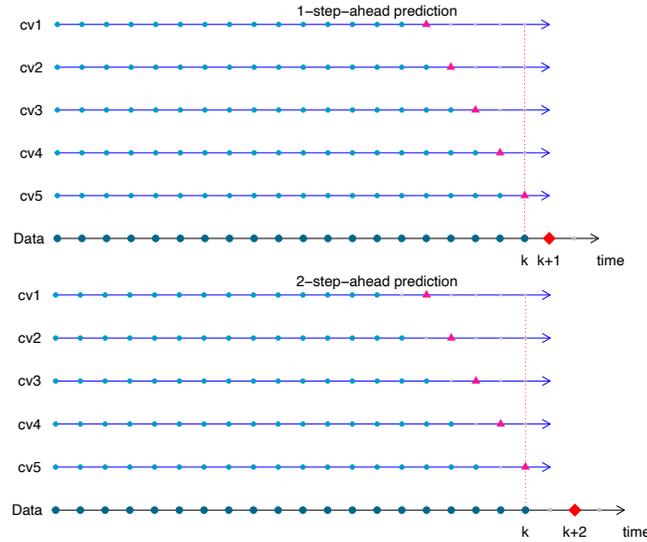
Figure 4.5: A demonstration of cross-validation set ups: 1-step-ahead and 2-step-ahead predictions

blue axes). Figure 4.5 demonstrates the cross-validation set ups exemplifying the 1-step-ahead and 2-step-ahead predictions. The procedure could be generalised to higher values of h.

4.3.5.3 Training and prediction

The XGBoost classifier with optimal parameter are trained on all data sets with selected features. More specifically, at time k, for each $h = 1, 2, \ldots, 6$, we fit the model with the best parameter obtained using the grid-search process with cross validation on the available data set with the corresponding features selected by the aid of the RF module.

In order to fully capture the dynamics of the multivariate data series, we set the length of each algorithm step to 1 time unit (month). So, for each algorithm step k, we continuously tune the XGBoost classifier's hyper-parameters and then train the model using all available data up to time k. The main principles behind our modelling algorithmic approach is summarised in Algorithm 1.

Algorithm 1: General Algorithm for step-h prediction

Input: FSI of 17 countries
initialization;
for $k = n$ to $621 - h$ **do**
    for $m = 1$ to 17 **do**
        1. Calculate discrimative signal $Y_k^{(m)}$ and evaluate $Z_k^{(m)}$;
        2. Calculate HMM filters parameters and evaluate
        $\hat{x}_{k,1}^{(m)}$, $\hat{d}_{k,1}^{(m)}$, and $\hat{d}_{k,2}^{(m)}$;
        3. Generate HMM features: $\{\mathbf{D}_{k-h}, \mathbf{X}_{k-h}\}$;
        4. Random Forest: Select features from $\{\mathbf{Y}_{k-h}, \mathbf{D}_{k-h}, \mathbf{X}_{k-h}\}$;
        5. XGBoost: Tune hyperparameters via grid search cross
        validation on features selected in Step 4;
        6. XGBoost: Train tuned model and generate a prediciton;
        Output: $\hat{Z}_{k+h}^{(m)} = \hat{P}\left(Z_{k+h}^{(m)} = 1\right)$.
    end
end

4.4 Results and diagnostics

Our proposed modelling approach is tested on 17 countries' FSI data series covering the period Jan 1968–Sep 2019. We benchmark our hybridised model with 5 different alternative models. As presented in Table 4.3, Model_1 is our proposed model. Model_2 uses RF instead of XGBoost as the final-stage classifier. Thus, the purpose of Model_2 is to test the significance of XGBoost's contribution on system's performance. Model_3 and Model_4 are parts of Model_1 and Model_2, respectively. The motivation of introducing these two benchmark models is to verify the importance of HMM features. Model_5 is Vector Autoregressive Model with a memory length of 12. The first four models are embedded with classifiers at their final-stage component. Therefore, their outcomes are $\hat{Z}_{k+h}^{(m)}$ which could be interpreted as the model's implied probabilities of $Z_{k+h}^{(m)} = 1$.

The Naive model makes classifications based on the h-step-ahead predictions using the value at time k whilst Model_5 generates the predictions in a recursive way. The original outcomes of Model_5 and Naive method are $\hat{x}_{k+h}^{(m)}$ and the h-step-ahead predicted values of $Y_k^{(m)}$. The binary forecasting values for target variable $Z_{k+h}^{(m)}$ are obtained based on the definition of $Z_t^{(m)}$ in equation (4.5).

Table 4.3: Models compared in this chapter

Model	Stage 1	Stage 2	Stage 3
Model_1	**HMM feature generation**	**RF feature selection**	**XGBoost classifier**
Model_2	HMM feature generation	RF feature selection	RF classifier
Model_3	RF feature selection	XGBoost classifier	-
Model_4	RF feature selection	RF classifier	-
Model_5	VAR(12)	-	-
Naive	$\hat{x}_{k+h} = Y_k, \ \forall h = 1, \ 2, \ ...$	-	-

Since this research is actually solving a supervised learning problem, the model's predicted class is defined through

$$\tilde{Z}_{k+h}^{(m)} = \begin{cases} 1 \ (\text{Predicted as Positive}), & \text{if } \hat{Z}_{k+h}^{(m)} \geq p_0, \\ 0 \ (\text{Predicted as Negative}), & \text{if } \hat{Z}_{k+h}^{(m)} < p_0, \end{cases} \tag{4.34}$$

where p_0 is a discrimination threshold. Equation (4.34) defines a decision rule concerning a data point at time $k + h$, whether to classify it as an anomalous point or not based on $\hat{Z}_{k+h}^{(m)}$ given all information at time k. In this work, we set $p_0 = 0.5$ for all classification models in Table 4.3 for simplicity and consistency of interpretation. After training the models on each individual time series $Y_t^{(m)}$, the models' forecasting power is evaluated vis-à-vis their stepwise out-of-sample prediction performance under 3 different metrics, which are discussed in the succeeding subsections.

4.4.1 Performance metrics derived from confusion matrix

The classification accuracy is a straightforward measure, which calculates the percentage of correct classification of a supervised learning model. The accuracy's evaluation formula is

$$
\text{ACC}_h^{(m)} = \frac{\sum\limits_{k} \left(\mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=1, \, Z_{k+h}^{(m)}=1 \right\}} + \mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=0, \, Z_{k+h}^{(m)}=0 \right\}} \right)}{\sum\limits_{k} \left(\mathbf{1}_{\left\{ Z_{k+h}^{(m)}=1 \right\}} + \mathbf{1}_{\left\{ Z_{k+h}^{(m)}=0 \right\}} \right)}, \tag{4.35}
$$

where $\mathbf{1}$ is an indicator function, and for $h = 1, \, 2, \ldots, \, 6$.

Although accuracy gives a direct assessment of the model, it does not tell the full story of classification performance. The metric will provide misleading outcomes when the data set is unbalanced (e.g., the distribution of observations across all categories is biased). A model's classification power in all known classes cannot definitely be assessed by the accuracy metric.

Our target is making binary classifications on multiple countries' FSI. We introduce a confusion matrix to gauge further the out-of-sample classification performance of our proposed model. More specifically, we trace the classification outcomes of all models' stepwise forecasts and compare them with the true values in the data set. For each step-ahead prediction, we aim to determine the numbers of True Positives, True Negatives, False Positives and False Negatives. For each step-ahead forecasting, the True Positive Rate (TPR) or Sensitivity, True Negative Rate (TNR) or Specificity, Positive Prediction Value (PPV) or Precision, and Negative Prediction Value (NPV).

Figure 4.6: Illustrating the concept of confusion matrix

$$
\begin{aligned}
\text{TPR}_h &= \frac{\sum_k \mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=1, \, Z_{k+h}^{(m)}=1 \right\}}}{\sum_k \mathbf{1}_{\left\{ Z_{k+h}^{(m)}=1 \right\}}}, \quad
\text{TNR}_h = \frac{\sum_k \mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=0, \, Z_{k+h}^{(m)}=0 \right\}}}{\sum_k \mathbf{1}_{\left\{ Z_{k+h}^{(m)}=0 \right\}}} \\[2ex]
\text{PPV}_h &= \frac{\sum_k \mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=1, \, Z_{k+h}^{(m)}=1 \right\}}}{\sum_k \mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=1 \right\}}}, \quad
\text{NPV}_h = \frac{\sum_k \mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=0, \, Z_{k+h}^{(m)}=0 \right\}}}{\sum_k \mathbf{1}_{\left\{ \tilde{Z}_{k+h}^{(m)}=0 \right\}}}
\end{aligned} \tag{4.36}
$$

The metrics in the set of Eqs. (4.36) are equal to 1 for perfect classification models. The numerical results of such metrics are displayed in Figure 4.7. It can be observed that even though each model's forecasting power declines as the prediction horizon increases, Model_1 achieves high scores in all five metrics when applied on all countries' FSI data series.

This result indicates that the proposed model has better out-of-sample classification capability than other benchmark models. More specifically, attaining high accuracy score indicates that Model_1, in general, has a lower chance than other benchmark model to misclassify the data points. Furthermore, getting higher TPR and TNR scores implies that Model_1 has a better capacity in differentiating anomalous/normal episodes with relative lower false positive/negative errors. In addition, obtaining higher PPV and NPV scores suggests that Model_1 has a higher credibility on the correctness of its predictions. Last but not least, the above advantages of Model_1 is robust with respect to different time series within our data set.
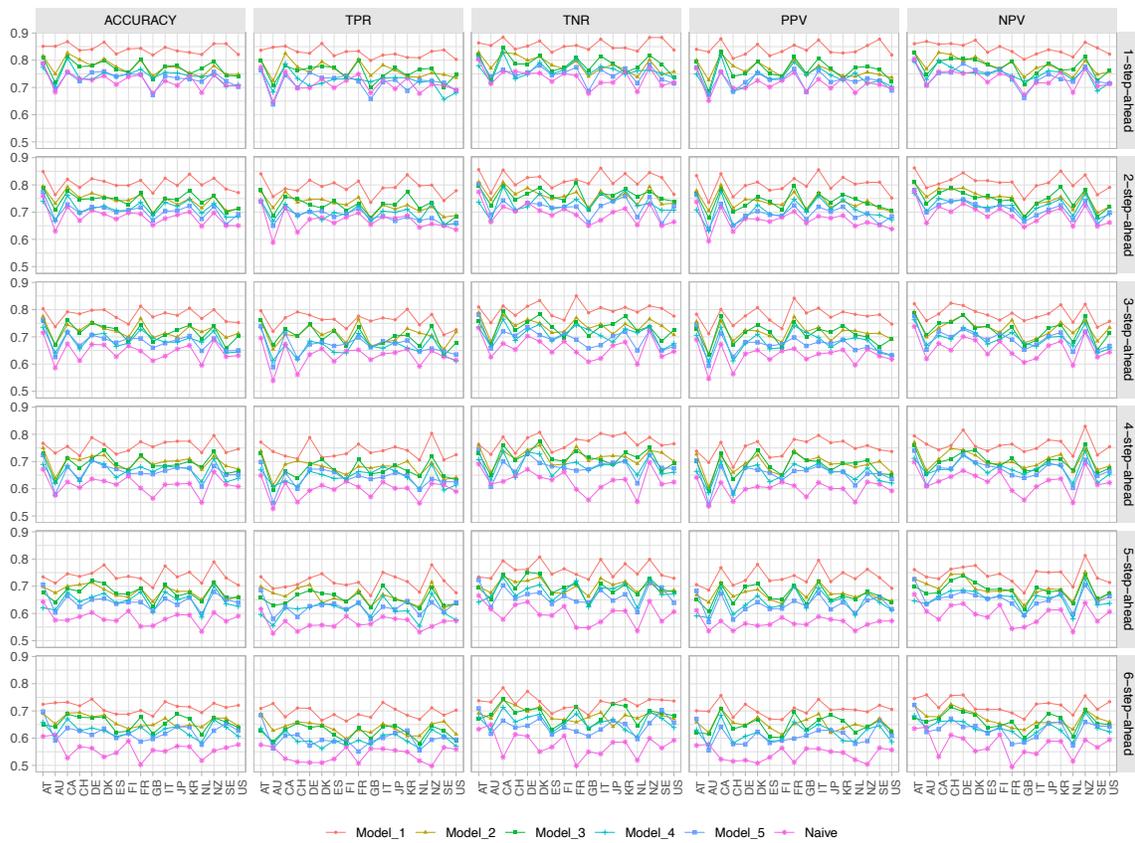
Figure 4.7: Comparison of models in terms of stepwise forecasting classification performance for various h's
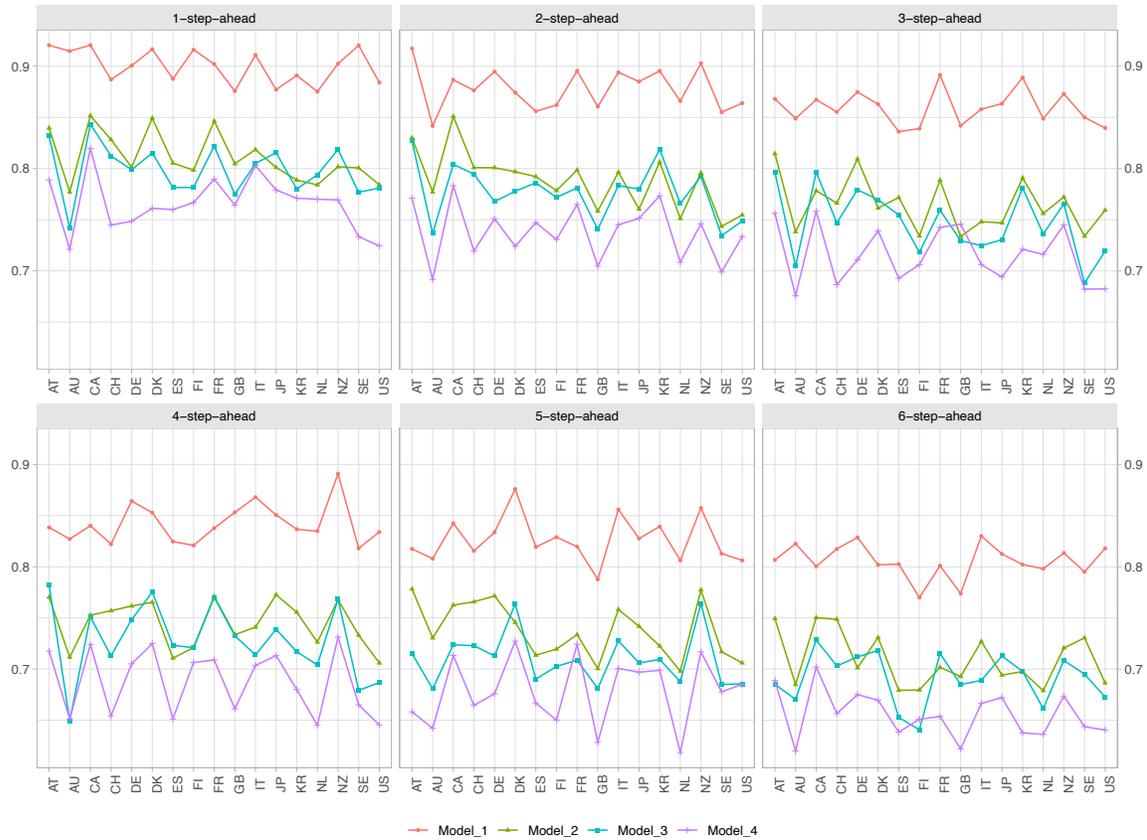
Figure 4.8: A comparison of models in terms of their stepwise forecasting classification performance under the AUC metric

4.4.2   Area under ROC curve (AUC)

The metrics derived from the confusion matrix in Subsection 4.4.1 assesses the model performance based on a fixed discrimination threshold $p_0 = 0.5$. Additional insights could be gained amongst different models with the investigation of each model's classification performance at different values of p_0.

A receiver operating characteristic (ROC) curve is a graphical tool showing the performance of a binary classification model at all discrimination thresholds. In general, the AUC considers the predicted probabilities as the criteria to judge the model's performance. The procedure to construct an ROC can be summarised as follows: (i) Order models' predicated probabilities $\hat{Z}_{k+h}^{(m)}$. (ii) Calculate $\tilde{Z}_{k+h}^{(m)}$ at multiple p_0's and obtain the corresponding TPR's and FPR's. (iii) Plot the TPRs against the FPRs. An important summary metric is the area under the curve (AUC) representing the probability that a binary classification model ranks a randomly chosen positive example higher than a random negative instance. Note that AUC's value lies in the interval $[0, 1]$. The AUC of a worst model that makes 100% wrong predictions is 0 whilst that of a model that makes 100% right forecasting is 1.

Figure 4.8 gives a visualisation of the AUC values for the 4 different classification models under several step-ahead predictions. Model_1 achieves the highest AUC score in all 6-step-ahead predictions when applied on each single countries' FSI time series. Note that Model_5 and the Naive method in Table 4.3 are not included in the AUC comparison because the probabilistic outcomes are not available for these two models.

In order to ascertain the statistical significance of the results in Figure 4.8, we conduct a pairwise one-tailed *t*-test facilitated by the bootstrap method. A sample size of 10,000 is chosen from each model's predicted probabilities on all countries' FSI data. A significance test of Model_1 versus Model_*j* ($j = 2, 3, 4$) on AUC at *h*-step-ahead predictions is formulated below.

$$\begin{cases} H_0 : AUC_{1,h}^{(m)} \leq AUC_{j,h}^{(m)} \\ H_a : AUC_{1,h}^{(m)} > AUC_{j,h}^{(m)}, \end{cases} \tag{4.37}$$

where *m* is the country index. The outcomes of the tests for each paired models are presented in Tables 4.4 - 4.6. In each pairwise test, for every single FSI time series, the *p*-values are less than 5% on all prediction horizons. This implies that H_0's are rejected and our proposed model has more robust out-of-sample classification power than the other 3 models.

In addition to the *t*-test, a DeLong test [19] is performed to validate our model's significant superiority with regard to AUC. The null and alternative hypotheses are defined in a similar manner to (4.37). Using the R function "roc.test" from package "pROC", the outcomes are obtained and displayed in Tables 4.7 - 4.9. All *p*-values are less than 5%, and so the null hypothesis is rejected. This justifies, in terms of the AUC score, that our proposed model dominates the other benchmark models.

4.4.3 Log loss

As pointed out in Subsection 4.4.2, the AUC determines the model's capability to differentiate two classes by measuring the probability that a positive instance is ranked higher than a negative one. In this subsection, we shall compare the models' classification power based on a logistic loss, which is also known as log-loss or cross-entropy loss. This entails the evaluation of the negative log-likelihood for the predicted probabilities generated by the classification model.

In this work, the logistic loss is calculated

$$L_h^{(m)} = -\frac{1}{N_h} \sum_k \left\{ Z_{k+h}^{(m)} \log\left(\hat{Z}_{k+h}^{(m)}\right) + \left(1 - Z_{k+h}^{(m)}\right) \log\left(\hat{Z}_{k+h}^{(m)}\right) \right\}. \tag{4.38}$$

From Eq. (4.38), the log loss could be computed using the actual class indicators and the logarithm of the predicted probabilities for each class. A classifier achieves a high log loss score if it yields

Table 4.4: *p*-values of bootstrapped t-test on AUC: Model_1 vs Model_2

Country	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead
AT	6.710×10^{-7}	5.392×10^{-7}	3.141×10^{-3}	5.558×10^{-4}	2.007×10^{-2}	2.937×10^{-3}
AU	4.391×10^{-11}	1.198×10^{-3}	3.178×10^{-7}	1.097×10^{-8}	2.377×10^{-4}	1.399×10^{-9}
CA	1.113×10^{-4}	2.302×10^{-2}	1.492×10^{-5}	2.958×10^{-6}	5.484×10^{-5}	1.367×10^{-2}
CH	4.170×10^{-4}	2.147×10^{-5}	7.994×10^{-6}	7.845×10^{-4}	5.961×10^{-3}	7.123×10^{-4}
DE	6.526×10^{-7}	3.580×10^{-7}	5.472×10^{-4}	1.246×10^{-6}	1.515×10^{-3}	3.711×10^{-8}
DK	1.612×10^{-4}	1.110×10^{-5}	5.708×10^{-7}	1.061×10^{-5}	5.704×10^{-10}	4.026×10^{-4}
ES	4.342×10^{-5}	2.837×10^{-4}	7.662×10^{-4}	1.601×10^{-7}	5.926×10^{-7}	5.088×10^{-8}
FI	1.117×10^{-9}	1.302×10^{-5}	5.145×10^{-8}	2.432×10^{-6}	4.652×10^{-7}	4.705×10^{-5}
FR	5.303×10^{-4}	4.828×10^{-7}	6.534×10^{-9}	6.192×10^{-4}	4.009×10^{-5}	1.996×10^{-6}
GB	3.022×10^{-4}	4.081×10^{-7}	3.258×10^{-7}	1.854×10^{-8}	9.359×10^{-5}	1.721×10^{-4}
IT	5.220×10^{-7}	3.704×10^{-8}	1.181×10^{-7}	9.993×10^{-11}	4.002×10^{-6}	4.728×10^{-7}
JP	1.483×10^{-4}	1.219×10^{-9}	6.282×10^{-9}	3.089×10^{-5}	3.808×10^{-5}	1.480×10^{-7}
KR	5.330×10^{-7}	1.199×10^{-6}	1.619×10^{-7}	2.517×10^{-5}	2.592×10^{-7}	1.027×10^{-5}
NL	4.072×10^{-6}	2.286×10^{-8}	1.619×10^{-6}	4.581×10^{-8}	1.870×10^{-7}	5.303×10^{-8}
NZ	4.999×10^{-8}	2.380×10^{-8}	6.802×10^{-7}	4.660×10^{-11}	2.188×10^{-5}	1.086×10^{-5}
SE	2.309×10^{-11}	5.048×10^{-8}	9.964×10^{-8}	1.280×10^{-4}	1.692×10^{-5}	1.478×10^{-3}
US	2.377×10^{-7}	4.291×10^{-7}	1.173×10^{-4}	7.635×10^{-10}	1.104×10^{-5}	2.902×10^{-9}

Table 4.5: *p*-values of bootstrapped t-test on AUC: Model_1 vs Model_3

Country	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead
AT	2.724×10^{-7}	3.712×10^{-8}	1.962×10^{-4}	4.511×10^{-3}	1.201×10^{-6}	1.860×10^{-7}
AU	7.749×10^{-18}	2.259×10^{-6}	9.266×10^{-11}	4.046×10^{-17}	2.289×10^{-8}	6.682×10^{-11}
CA	4.579×10^{-6}	5.753×10^{-6}	4.126×10^{-4}	1.319×10^{-5}	7.607×10^{-8}	7.213×10^{-4}
CH	8.005×10^{-5}	2.151×10^{-6}	5.954×10^{-7}	5.833×10^{-7}	4.387×10^{-5}	4.738×10^{-7}
DE	1.058×10^{-7}	4.518×10^{-11}	2.020×10^{-6}	9.052×10^{-8}	1.422×10^{-7}	1.962×10^{-7}
DK	3.040×10^{-7}	2.182×10^{-6}	3.196×10^{-6}	9.675×10^{-5}	7.921×10^{-8}	1.685×10^{-4}
ES	4.323×10^{-8}	1.889×10^{-4}	3.288×10^{-5}	2.062×10^{-6}	4.273×10^{-8}	4.669×10^{-11}
FI	1.585×10^{-11}	2.118×10^{-6}	1.361×10^{-8}	4.915×10^{-7}	6.913×10^{-9}	5.447×10^{-8}
FR	1.495×10^{-5}	4.942×10^{-10}	4.273×10^{-11}	7.734×10^{-4}	3.250×10^{-7}	1.342×10^{-4}
GB	4.070×10^{-6}	8.874×10^{-10}	3.475×10^{-7}	5.351×10^{-9}	4.270×10^{-7}	1.238×10^{-4}
IT	1.253×10^{-8}	1.545×10^{-8}	2.086×10^{-9}	5.414×10^{-14}	6.401×10^{-10}	9.986×10^{-11}
JP	1.634×10^{-3}	2.740×10^{-8}	1.555×10^{-10}	1.045×10^{-7}	2.009×10^{-7}	2.225×10^{-6}
KR	2.188×10^{-8}	2.108×10^{-5}	3.066×10^{-9}	1.562×10^{-8}	3.970×10^{-9}	1.368×10^{-5}
NL	2.395×10^{-5}	1.510×10^{-7}	3.173×10^{-8}	3.036×10^{-9}	6.759×10^{-8}	1.569×10^{-9}
NZ	9.642×10^{-6}	1.389×10^{-8}	1.202×10^{-7}	1.600×10^{-11}	5.182×10^{-6}	3.280×10^{-6}
SE	5.118×10^{-12}	1.160×10^{-8}	1.680×10^{-13}	9.289×10^{-10}	3.508×10^{-9}	2.932×10^{-6}
US	1.352×10^{-7}	1.031×10^{-8}	6.999×10^{-8}	9.143×10^{-12}	1.013×10^{-7}	5.704×10^{-10}

Table 4.6: *p*-values of bootstrapped t-test on AUC: Model_1 vs Model_4

Country	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead
AT	4.623×10^{-12}	3.535×10^{-14}	1.773×10^{-7}	4.284×10^{-7}	4.392×10^{-10}	2.840×10^{-7}
AU	5.568×10^{-18}	7.808×10^{-10}	3.261×10^{-13}	1.158×10^{-12}	1.020×10^{-11}	7.084×10^{-17}
CA	1.101×10^{-7}	1.250×10^{-7}	9.609×10^{-7}	7.595×10^{-8}	3.094×10^{-9}	3.676×10^{-5}
CH	6.903×10^{-12}	6.690×10^{-13}	1.736×10^{-11}	3.900×10^{-11}	2.339×10^{-9}	9.617×10^{-12}
DE	2.986×10^{-13}	1.110×10^{-13}	1.970×10^{-12}	3.508×10^{-12}	1.115×10^{-11}	1.235×10^{-9}
DK	4.253×10^{-13}	2.130×10^{-12}	1.203×10^{-7}	1.498×10^{-8}	1.521×10^{-11}	2.920×10^{-7}
ES	7.852×10^{-10}	7.292×10^{-7}	2.755×10^{-10}	9.898×10^{-13}	2.351×10^{-10}	2.097×10^{-12}
FI	7.926×10^{-14}	2.169×10^{-9}	4.723×10^{-10}	2.967×10^{-7}	4.860×10^{-13}	3.779×10^{-7}
FR	1.602×10^{-7}	6.806×10^{-11}	9.198×10^{-13}	4.455×10^{-8}	9.502×10^{-6}	1.983×10^{-10}
GB	2.959×10^{-7}	8.216×10^{-11}	6.099×10^{-6}	4.273×10^{-15}	6.390×10^{-10}	8.172×10^{-9}
IT	4.098×10^{-7}	3.028×10^{-12}	6.572×10^{-12}	3.651×10^{-13}	1.326×10^{-11}	2.935×10^{-11}
JP	9.162×10^{-7}	8.646×10^{-11}	8.310×10^{-14}	1.577×10^{-10}	3.349×10^{-8}	4.176×10^{-9}
KR	1.415×10^{-8}	4.483×10^{-9}	1.393×10^{-14}	5.524×10^{-11}	4.848×10^{-10}	1.087×10^{-10}
NL	1.931×10^{-7}	2.077×10^{-13}	1.394×10^{-8}	1.377×10^{-17}	9.825×10^{-15}	1.492×10^{-12}
NZ	7.162×10^{-11}	2.322×10^{-13}	1.259×10^{-8}	6.490×10^{-13}	3.662×10^{-9}	2.398×10^{-8}
SE	1.012×10^{-17}	5.027×10^{-12}	2.043×10^{-12}	1.451×10^{-11}	3.783×10^{-9}	5.072×10^{-9}
US	1.292×10^{-12}	4.583×10^{-10}	1.595×10^{-11}	2.010×10^{-14}	1.018×10^{-7}	4.113×10^{-12}

Table 4.7: *p*-values of Delong tests on AUC: Model_1 vs Model_2

Country	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead
AT	7.907×10^{-7}	4.020×10^{-7}	3.506×10^{-3}	4.967×10^{-4}	2.404×10^{-2}	3.220×10^{-3}
AU	6.707×10^{-12}	1.858×10^{-3}	1.864×10^{-7}	1.338×10^{-8}	2.364×10^{-4}	1.367×10^{-9}
CA	1.823×10^{-4}	2.011×10^{-2}	7.301×10^{-6}	3.556×10^{-6}	5.038×10^{-5}	1.304×10^{-2}
CH	5.268×10^{-4}	2.363×10^{-5}	8.169×10^{-6}	1.018×10^{-3}	5.704×10^{-3}	9.891×10^{-4}
DE	2.295×10^{-7}	1.896×10^{-7}	4.362×10^{-4}	5.771×10^{-7}	1.745×10^{-3}	2.222×10^{-8}
DK	1.651×10^{-4}	1.179×10^{-5}	5.361×10^{-7}	8.983×10^{-6}	3.798×10^{-10}	6.839×10^{-4}
ES	4.752×10^{-5}	3.180×10^{-4}	6.839×10^{-4}	1.195×10^{-7}	8.497×10^{-7}	2.976×10^{-8}
FI	1.536×10^{-9}	1.137×10^{-5}	1.763×10^{-7}	2.088×10^{-6}	5.660×10^{-7}	4.232×10^{-5}
FR	4.423×10^{-4}	3.303×10^{-7}	1.041×10^{-8}	6.101×10^{-4}	3.162×10^{-5}	2.612×10^{-6}
GB	2.658×10^{-4}	4.676×10^{-7}	5.968×10^{-7}	3.183×10^{-8}	5.583×10^{-5}	2.040×10^{-4}
IT	3.763×10^{-7}	6.099×10^{-8}	8.665×10^{-8}	1.240×10^{-10}	1.675×10^{-6}	7.450×10^{-7}
JP	1.358×10^{-4}	5.648×10^{-10}	8.501×10^{-9}	4.894×10^{-5}	4.363×10^{-5}	2.150×10^{-7}
KR	1.013×10^{-6}	9.121×10^{-7}	2.035×10^{-7}	5.119×10^{-5}	1.379×10^{-7}	5.490×10^{-6}
NL	5.639×10^{-6}	4.121×10^{-9}	2.659×10^{-6}	1.297×10^{-7}	1.915×10^{-7}	6.864×10^{-8}
NZ	6.968×10^{-8}	1.473×10^{-8}	7.359×10^{-7}	7.888×10^{-11}	3.054×10^{-5}	1.495×10^{-5}
SE	4.028×10^{-11}	5.782×10^{-8}	1.915×10^{-7}	1.142×10^{-4}	1.009×10^{-5}	1.666×10^{-3}
US	4.661×10^{-7}	2.783×10^{-7}	1.100×10^{-4}	2.233×10^{-9}	1.357×10^{-5}	6.296×10^{-9}

Table 4.8: *p*-values of Delong tests on AUC: Model_1 vs Model_3

Country	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead
AT	2.011×10^{-7}	1.346×10^{-7}	2.265×10^{-4}	4.003×10^{-3}	1.231×10^{-6}	1.806×10^{-7}
AU	2.400×10^{-17}	1.124×10^{-6}	1.989×10^{-10}	2.895×10^{-16}	2.652×10^{-8}	4.711×10^{-11}
CA	2.600×10^{-6}	2.975×10^{-6}	3.302×10^{-4}	1.447×10^{-5}	9.782×10^{-8}	6.123×10^{-4}
CH	4.928×10^{-5}	4.572×10^{-6}	9.888×10^{-7}	6.242×10^{-7}	4.829×10^{-5}	1.121×10^{-6}
DE	1.587×10^{-7}	5.070×10^{-11}	3.214×10^{-6}	2.976×10^{-7}	2.778×10^{-7}	2.540×10^{-7}
DK	4.465×10^{-7}	1.469×10^{-6}	2.905×10^{-6}	7.424×10^{-5}	3.393×10^{-8}	1.403×10^{-4}
ES	1.505×10^{-7}	1.742×10^{-4}	4.225×10^{-5}	2.032×10^{-6}	9.111×10^{-9}	6.952×10^{-11}
FI	5.836×10^{-12}	2.559×10^{-6}	1.283×10^{-8}	8.740×10^{-7}	3.819×10^{-9}	6.095×10^{-8}
FR	1.187×10^{-5}	4.704×10^{-10}	9.823×10^{-11}	6.502×10^{-4}	4.627×10^{-7}	8.758×10^{-5}
GB	2.407×10^{-6}	1.900×10^{-9}	2.271×10^{-7}	9.012×10^{-9}	3.924×10^{-7}	1.555×10^{-4}
IT	5.217×10^{-9}	5.009×10^{-9}	1.781×10^{-10}	8.638×10^{-14}	1.953×10^{-9}	2.562×10^{-10}
JP	1.539×10^{-3}	4.055×10^{-8}	4.318×10^{-10}	1.367×10^{-7}	1.570×10^{-7}	2.813×10^{-6}
KR	2.206×10^{-8}	3.269×10^{-5}	1.341×10^{-8}	2.249×10^{-8}	9.874×10^{-9}	9.874×10^{-6}
NL	3.348×10^{-5}	1.047×10^{-7}	2.729×10^{-8}	1.618×10^{-9}	4.517×10^{-8}	2.064×10^{-9}
NZ	7.561×10^{-6}	1.122×10^{-8}	6.679×10^{-8}	1.539×10^{-11}	8.053×10^{-6}	5.058×10^{-6}
SE	2.581×10^{-12}	8.289×10^{-9}	9.528×10^{-13}	3.266×10^{-9}	2.333×10^{-9}	5.976×10^{-6}
US	1.420×10^{-7}	2.455×10^{-8}	2.615×10^{-8}	1.504×10^{-11}	1.360×10^{-7}	3.097×10^{-10}

Table 4.9: *p*-values of Delong tests on AUC: Model_1 vs Model_4

Country	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead
AT	2.597×10^{-12}	2.675×10^{-14}	1.450×10^{-7}	2.483×10^{-7}	1.868×10^{-10}	2.669×10^{-7}
AU	8.075×10^{-19}	3.660×10^{-10}	2.933×10^{-13}	3.421×10^{-13}	3.562×10^{-11}	2.015×10^{-16}
CA	9.730×10^{-8}	1.455×10^{-7}	1.028×10^{-6}	5.239×10^{-8}	8.614×10^{-9}	2.881×10^{-5}
CH	1.350×10^{-11}	1.306×10^{-14}	4.233×10^{-12}	4.452×10^{-12}	1.528×10^{-9}	2.559×10^{-11}
DE	1.202×10^{-13}	6.366×10^{-13}	1.552×10^{-11}	9.884×10^{-12}	5.299×10^{-12}	1.819×10^{-9}
DK	3.443×10^{-13}	3.082×10^{-12}	9.693×10^{-8}	1.521×10^{-8}	2.456×10^{-11}	3.010×10^{-7}
ES	1.443×10^{-9}	5.681×10^{-7}	4.808×10^{-10}	1.414×10^{-12}	1.261×10^{-10}	1.331×10^{-11}
FI	1.962×10^{-14}	5.120×10^{-9}	3.628×10^{-10}	2.908×10^{-7}	5.424×10^{-13}	6.533×10^{-7}
FR	7.163×10^{-8}	3.029×10^{-10}	1.328×10^{-12}	2.350×10^{-8}	1.211×10^{-5}	4.790×10^{-10}
GB	3.248×10^{-7}	1.489×10^{-11}	6.499×10^{-6}	6.220×10^{-16}	1.787×10^{-10}	9.475×10^{-9}
IT	7.050×10^{-8}	5.000×10^{-12}	2.819×10^{-11}	1.200×10^{-13}	3.831×10^{-11}	1.412×10^{-11}
JP	1.138×10^{-6}	8.750×10^{-11}	6.450×10^{-14}	7.320×10^{-10}	3.613×10^{-8}	1.244×10^{-8}
KR	1.110×10^{-8}	3.900×10^{-9}	1.669×10^{-15}	3.916×10^{-11}	1.570×10^{-9}	6.657×10^{-11}
NL	3.184×10^{-7}	8.840×10^{-13}	5.378×10^{-9}	1.744×10^{-16}	4.996×10^{-15}	8.112×10^{-12}
NZ	1.158×10^{-10}	2.716×10^{-13}	8.010×10^{-9}	7.786×10^{-13}	4.280×10^{-9}	2.637×10^{-8}
SE	2.899×10^{-17}	2.977×10^{-12}	2.999×10^{-12}	4.140×10^{-11}	4.765×10^{-9}	8.968×10^{-9}
US	5.252×10^{-12}	1.063×10^{-9}	9.583×10^{-12}	6.889×10^{-15}	1.309×10^{-7}	5.491×10^{-12}

high predicted probabilities for true positive example and low values for true negative instances. Theoretically, a perfect classification model's log-loss score is 0 whilst a bad one has a log-loss score is some relatively large positive value up to infinity.

In Table 4.10, we report the logistic loss for the proposed model and 3 other benchmark classification models. It could be seen that Model_1 has the lowest score in all 6 prediction horizons. Hence, it has higher predicted probabilities than the other models.

Table 4.10: A log-loss comparison of four competing models

Step-ahead	Model	AT	AU	CA	CH	DE	DK	ES	FI	FR	GB	IT	JP	KR	NL	NZ	SE	US
1	**Model_1**	**0.3687**	**0.3808**	**0.3859**	**0.4302**	**0.4106**	**0.3923**	**0.4313**	**0.3794**	**0.4019**	**0.4543**	**0.4030**	**0.4439**	**0.4277**	**0.4450**	**0.4029**	**0.3756**	**0.4357**
	Model_2	0.5055	0.5914	0.4860	0.5276	0.5741	0.4880	0.5575	0.5614	0.4970	0.5550	0.5429	0.5694	0.5794	0.5907	0.5634	0.5629	0.5950
	Model_3	0.5192	0.6475	0.4962	0.5603	0.5721	0.5414	0.5966	0.5936	0.5410	0.6097	0.5638	0.5395	0.6005	0.5865	0.5342	0.5973	0.5905
	Model_4	0.5830	0.6928	0.5409	0.6581	0.6392	0.6209	0.6320	0.6210	0.5923	0.6122	0.5597	0.5989	0.6167	0.6162	0.5966	0.6581	0.6802
2	**Model_1**	**0.3856**	**0.5060**	**0.4308**	**0.4413**	**0.4171**	**0.4481**	**0.4817**	**0.4668**	**0.4134**	**0.4683**	**0.4260**	**0.4294**	**0.4187**	**0.4590**	**0.4021**	**0.4800**	**0.4635**
	Model_2	0.5188	0.5916	0.4816	0.5615	0.5688	0.5648	0.5814	0.5961	0.5738	0.6363	0.5765	0.6299	0.5498	0.6393	0.5657	0.6493	0.6421
	Model_3	0.5296	0.6543	0.5690	0.5694	0.6103	0.6065	0.6077	0.6055	0.6034	0.6627	0.6008	0.5908	0.5452	0.6204	0.5758	0.6720	0.6505
	Model_4	0.6302	0.7256	0.6017	0.6834	0.6250	0.6788	0.6378	0.6619	0.6199	0.7088	0.6663	0.6325	0.6131	0.6880	0.6421	0.7033	0.6581
3	**Model_1**	**0.4622**	**0.4858**	**0.4647**	**0.4832**	**0.4491**	**0.4646**	**0.5082**	**0.4990**	**0.4282**	**0.4984**	**0.4805**	**0.4619**	**0.4235**	**0.4875**	**0.4499**	**0.4851**	**0.5030**
	Model_2	0.5495	0.6422	0.5994	0.6067	0.5597	0.6231	0.6086	0.6587	0.5732	0.6596	0.6540	0.6455	0.5833	0.6293	0.6009	0.6644	0.6308
	Model_3	0.5869	0.7060	0.5744	0.6412	0.6065	0.6111	0.6356	0.6805	0.6282	0.6687	0.6872	0.6944	0.6047	0.6598	0.6151	0.7105	0.6824
	Model_4	0.6272	0.7474	0.6158	0.7234	0.7022	0.6650	0.7118	0.7018	0.6485	0.6335	0.7033	0.7322	0.6791	0.6871	0.6384	0.7195	0.7217
4	**Model_1**	**0.5145**	**0.5148**	**0.4978**	**0.5245**	**0.4670**	**0.4806**	**0.5252**	**0.5288**	**0.5090**	**0.4764**	**0.4565**	**0.4843**	**0.5041**	**0.4972**	**0.4165**	**0.5319**	**0.5082**
	Model_2	0.6251	0.6898	0.6414	0.6332	0.6306	0.6139	0.6959	0.6803	0.6073	0.6785	0.6504	0.6007	0.6440	0.6818	0.6130	0.6619	0.7128
	Model_3	0.5963	0.7700	0.6324	0.6855	0.6360	0.5927	0.6778	0.6830	0.6225	0.6526	0.6809	0.6442	0.6995	0.7019	0.6163	0.7476	0.7345
	Model_4	0.6873	0.7724	0.6710	0.7741	0.6905	0.6755	0.7909	0.6941	0.6890	0.7588	0.7123	0.6774	0.7332	0.7897	0.6634	0.7565	0.7702
5	**Model_1**	**0.5318**	**0.5324**	**0.5005**	**0.5403**	**0.5114**	**0.4454**	**0.5280**	**0.5158**	**0.5246**	**0.5719**	**0.4839**	**0.5208**	**0.5040**	**0.5396**	**0.4790**	**0.5350**	**0.5399**
	Model_2	0.5976	0.6552	0.6167	0.6185	0.6146	0.6603	0.7030	0.6772	0.6721	0.7232	0.6362	0.6482	0.6783	0.7233	0.5983	0.6709	0.6991
	Model_3	0.7003	0.7571	0.6803	0.6695	0.6959	0.6190	0.7321	0.7144	0.7150	0.7491	0.6867	0.7009	0.6951	0.7302	0.6243	0.7368	0.7420
	Model_4	0.7631	0.7814	0.7072	0.7551	0.7502	0.6707	0.7647	0.7871	0.6662	0.8150	0.7269	0.7011	0.7214	0.8294	0.6919	0.7367	0.7189
6	**Model_1**	**0.5477**	**0.5305**	**0.5602**	**0.5301**	**0.5160**	**0.5633**	**0.5481**	**0.5966**	**0.5490**	**0.5992**	**0.5163**	**0.5436**	**0.5554**	**0.5489**	**0.5357**	**0.5633**	**0.5284**
	Model_2	0.6601	0.7251	0.6502	0.6320	0.7253	0.6578	0.7432	0.7460	0.7163	0.7345	0.6805	0.7228	0.7338	0.7652	0.6864	0.6612	0.7474
	Model_3	0.7318	0.7486	0.6639	0.7011	0.6773	0.6759	0.7779	0.8066	0.6893	0.7108	0.7231	0.7044	0.7224	0.7489	0.6907	0.7065	0.7588
	Model_4	0.7358	0.8267	0.7145	0.7718	0.7471	0.7540	0.7875	0.7800	0.7675	0.8369	0.7609	0.7691	0.7740	0.7970	0.7263	0.7876	0.7806

4.4.4 Kolmogorov-Smirnov test

As highlighted in Subsection 4.4.3, we assess the classification models' good of fit by analysing the logistic loss, which measures the log-likelihood of the predicted probabilities. However, a logistic loss does not capture the difference in the models' outcome for each class. Hence, we need to determine the divergence, in terms of the distribution, of a model's responses for different classes. More specifically, in our case, we shall apply the two-sample Kolmogorv-Smirnov (KS) test on the predicted probabilities' distributions covering the anomalous (True Positive) and normal (True Negative) episodes. This enables the comparison of the models' capacity to separate the actual anomalous intervals of FSI from the normal ones. The KS statistic is

$$D_h^{(m)} = \sup \left| F_{0,h}^{(m)} - F_{1,h}^{(m)} \right|, \tag{4.39}$$

where $F_{0,h}^{(m)}$ and $F_{1,h}^{(m)}$ are the empirical cumulative distribution functions (ECDF) of $\hat{Z}_{k+h}^{(m)}$ for the true positive and negative classes, respectively. The statistic quantifies the supremum of the set of the distances between the points on ECDFs of the two classes.

Note that a perfect model that generates two mutually exclusive categories with each one having a separate class label of observations has a KS statistic score of 100%. On the other hand, the worst model that fails to differentiate two groups has a score of 0%. In general, a high KS-statistic value s suggests that the model is good in distinguishing the two classes.

Table 4.11 presents the KS statistics calculated from the predicted probabilities, which were produced by the proposed and benchmark models. It is noticeable that Model_1 attains the highest scores in all forecast horizons when implemented on all FSI time series. The higher the divergence between the distributions of predicted probabilities, the stronger the capacity to identify anomalous periods from regular episodes.

Table 4.11: Kolmogorov–Smirnov test results

Step-ahead	Model	AT	AU	CA	CH	DE	DK	ES	FI	FR	GB	IT	JP	KR	NL	NZ	SE	US
1	**Model_1**	**71.20**	**70.31**	**73.90**	**68.79**	**66.69**	**73.21**	**65.29**	**70.00**	**68.76**	**66.66**	**69.39**	**68.41**	**68.71**	**67.01**	**73.99**	**74.44**	**66.82**
	Model_2	62.65	50.50	67.83	59.77	55.14	60.98	54.50	53.24	59.67	49.93	54.95	56.74	53.36	50.47	58.19	49.51	50.79
	Model_3	61.59	44.20	63.65	54.87	57.25	60.52	53.50	54.14	61.18	45.94	55.41	58.70	51.04	53.83	62.19	49.49	50.51
	Model_4	54.07	41.33	60.70	46.49	44.65	50.16	47.32	50.87	53.47	46.78	53.02	53.21	51.50	49.46	50.43	42.58	43.39
2	**Model_1**	**70.24**	**55.35**	**64.47**	**60.91**	**65.84**	**64.56**	**63.71**	**63.41**	**66.22**	**59.23**	**64.67**	**63.35**	**70.23**	**62.02**	**64.92**	**60.57**	**62.80**
	Model_2	59.41	50.80	60.18	52.63	53.32	50.96	47.74	49.98	54.12	40.81	51.82	46.91	53.42	46.54	54.00	43.99	44.98
	Model_3	58.68	43.16	55.48	52.36	50.43	51.07	49.44	46.43	55.37	40.82	53.62	51.22	57.57	48.16	52.66	40.50	45.56
	Model_4	47.57	39.30	53.15	41.12	45.09	43.83	40.98	42.84	47.89	37.33	48.75	47.30	52.60	40.41	47.82	36.94	38.99
3	**Model_1**	**61.14**	**55.54**	**60.62**	**60.30**	**62.58**	**59.62**	**57.65**	**53.60**	**63.80**	**56.66**	**62.78**	**57.04**	**61.92**	**58.27**	**61.31**	**54.88**	**57.53**
	Model_2	55.61	39.99	49.85	47.66	52.39	45.09	45.64	41.82	54.76	40.36	44.67	42.96	48.95	45.64	50.37	41.09	47.53
	Model_3	50.88	38.44	52.97	47.16	51.46	46.37	45.37	39.01	48.95	36.55	44.63	46.05	48.37	38.71	50.48	32.35	45.02
	Model_4	49.04	36.56	45.28	32.92	43.32	41.34	32.59	36.20	46.22	39.51	40.73	38.11	38.77	37.53	46.98	29.83	31.03
4	**Model_1**	**54.31**	**51.19**	**53.47**	**49.16**	**57.91**	**54.90**	**49.62**	**56.41**	**60.04**	**53.93**	**55.13**	**58.70**	**56.05**	**52.54**	**61.34**	**51.60**	**54.71**
	Model_2	48.04	33.33	44.21	43.19	44.25	43.24	37.73	39.08	44.75	40.14	41.12	42.21	43.10	37.54	46.33	39.28	37.67
	Model_3	45.67	29.48	44.35	37.36	44.59	50.81	38.10	37.73	45.35	39.44	42.52	38.97	39.41	37.01	50.42	33.02	36.81
	Model_4	37.23	29.63	37.22	28.80	42.99	38.11	28.23	36.44	38.80	33.60	39.03	38.21	35.99	27.13	43.14	26.43	31.64
5	**Model_1**	**47.95**	**48.48**	**58.48**	**53.20**	**51.84**	**60.13**	**49.32**	**51.22**	**49.83**	**46.59**	**56.90**	**54.86**	**55.69**	**52.02**	**61.69**	**49.40**	**49.28**
	Model_2	41.00	36.51	41.63	43.76	43.64	37.65	32.99	33.63	42.03	32.03	40.18	39.07	37.90	33.68	44.75	33.73	36.42
	Model_3	34.30	33.54	40.96	39.79	45.75	42.92	34.63	34.55	40.16	25.29	42.87	34.95	34.16	29.88	44.53	32.82	33.40
	Model_4	27.25	25.40	35.96	31.78	34.48	35.62	28.60	33.24	38.20	24.35	35.85	32.22	33.40	22.16	42.52	29.70	28.79
6	**Model_1**	**50.00**	**49.24**	**49.02**	**48.63**	**50.08**	**47.55**	**44.56**	**43.42**	**46.21**	**46.87**	**51.63**	**51.30**	**48.61**	**45.67**	**50.62**	**46.31**	**48.38**
	Model_2	39.49	32.67	41.33	39.39	38.58	36.76	32.46	29.58	31.30	37.84	36.01	32.59	32.73	31.39	38.38	36.13	28.61
	Model_3	31.74	34.63	40.06	36.86	35.53	34.81	28.32	27.77	37.66	29.05	31.79	38.80	38.80	26.87	37.47	33.21	27.59
	Model_4	31.19	25.84	35.50	27.68	28.97	24.81	23.36	28.57	30.23	26.70	29.38	30.28	23.38	24.10	37.18	30.25	25.14

4.4.5   Analysis on features' importance

In addition to building a system with the capability to yield accurate predictions, it is also imperative to examine the model's interpretability. More specifically, we need to understand which features or input variables are most important for our system to detect the anomalous episodes in advance.

Thus, in this subsection, we shall diagnose the proposed model by investigating the features' importance.

As discussed in Subsection 4.3.4, we apply the RF to select the top-20 features on the basis of their importance which are being measured by the mean decrease of the OOB accuracy. The selected features are then used to train the XGBoost classifier for the generation of the out-of-sample (h-step-ahead) classifications. Our analysis mainly focuses on two types of quantities related to the features' importance: (i) frequencies of the features selected by the RF module, and (ii) gain score [2]) calculated by the XGBoost classifier.

At each algorithm step t, we trace the frequencies of all the features that are selected by the RF module and their importance scores calculated by the XGBoost classifier when making every h-step-ahead prediction ($h = 1, 2, ..., 6$). Furthermore, we calculate the marginal total frequencies of the features selected by the RF module and their marginal averaging importance scores output via the XGBoost module with respect to m (country index) and q (time lag). These quantities serve as comprehensive measurements of the contributions coming from different types of features in detecting anomalous episodes.

Let Ψ_t be the set of features selected by the RF module when performing an h-step-ahead prediction for one of the data series $Z_k^{(m)}$ at algorithm step t. We define the marginal total frequencies of these features, i.e., $\overline{D}_i(m)$, $\widetilde{D}_i(q)$ ($i = 1, 2$), $\overline{X}(m)$, $\widetilde{X}(q)$, $\overline{Y}(m)$ and $\widetilde{Y}(q)$ as follows:

$$
\begin{cases}
\overline{D}_i(m) := \sum_t \sum_{q=0}^{11} \mathbf{1}_{\{\hat{d}_{k-q,i}^{(m)} \in \Psi_t\}}, \ m \in \{1, 2, ..., 17\}, i = 1, 2 \\[2mm]
\overline{X}(m) := \sum_t \sum_{q=0}^{11} \mathbf{1}_{\{\hat{x}_{k-q,1}^{(m)} \in \Psi_t\}}, \ m \in \{1, 2, ..., 17\}, \\[2mm]
\overline{Y}(m) := \sum_t \sum_{q=0}^{11} \mathbf{1}_{\{Y_{k-q}^{(m)} \in \Psi_t\}}, \ m \in \{1, 2, ..., 17\}.
\end{cases}
\tag{4.40}
$$

$$
\begin{cases}
\widetilde{D}_i(q) := \sum_t \sum_{m=1}^{17} \mathbf{1}_{\{\hat{d}_{k-q,i}^{(m)} \in \Psi_t\}}, \ q \in \{0, 1, 2, ..., 11\}, i = 1, 2, \\[2mm]
\widetilde{X}(q) := \sum_t \sum_{m=1}^{17} \mathbf{1}_{\{\hat{x}_{k-q,1}^{(m)} \in \Psi_t\}}, \ q \in \{0, 1, 2, ..., 11\}, \\[2mm]
\widetilde{Y}(q) := \sum_t \sum_{m=1}^{17} \mathbf{1}_{\{Y_{k-q}^{(m)} \in \Psi_t\}}, \ q \in \{0, 1, 2, ..., 11\}.
\end{cases}
\tag{4.41}
$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function. The quantities in the sets of Eqs. (4.40) and (4.41) act as the preliminary measures for evaluating the association between the target variable (i.e., Anomalous/Normal) and the different types of input features. In general, a higher marginal total frequency

---

[2] According to the XGBoost R package document, the gain score is defined as the improvement in accuracy attained by a feature to the branches which it is on. The details of the Gain score's calculation could be found in the XGBoost Tutorials.

indicates that the related feature is treated as an important variable, with more recurrence, by the RF module in identifying the anomalies.

Let $\mathfrak{I}_t(\cdot)$ be the importance score of some feature calculated by the XGBoost classifier at each algorithm step. The marginal averaging XGBoost importance scores $\bar{d}_i(m)$, $\tilde{d}_i(q)$ ($i = 1, 2$), $\bar{x}(m)$, $\tilde{x}(q)$, $\bar{y}(m)$ and $\tilde{y}(q)$ are calculated as follows:

$$
\begin{cases}
\bar{d}_i(m) := \dfrac{1}{\overline{D}_i(m)} \displaystyle\sum_t \sum_{q=0}^{11} \mathfrak{I}_t\left(\hat{d}_{k-q,i}^{(m)}\right)\mathbf{1}_{\{\hat{d}_{k-q,i}^{(m)} \in \Psi_t\}}, \quad m \in \{1, 2, ..., 17\}, i = 1, 2 \\[2ex]
\bar{x}(m) := \dfrac{1}{\overline{X}(m)} \displaystyle\sum_t \sum_{q=0}^{11} \mathfrak{I}_t\left(\hat{x}_{k-q,1}^{(m)}\right)\mathbf{1}_{\{\hat{x}_{k-q,1}^{(m)} \in \Psi_t\}}, \quad m \in \{1, 2, ..., 17\}, \\[2ex]
\bar{x}(m) := \dfrac{1}{\overline{Y}(m)} \displaystyle\sum_t \sum_{q=0}^{11} \mathfrak{I}_t\left(Y_{k-q}^{(m)}\right)\mathbf{1}_{\{Y_{k-q}^{(m)} \in \Psi_t\}}, \quad m \in \{1, 2, ..., 17\}.
\end{cases}
\tag{4.42}
$$

$$
\begin{cases}
\tilde{d}_i(q) := \dfrac{1}{\widetilde{D}_i(q)} \displaystyle\sum_t \sum_{m=1}^{17} \mathfrak{I}_t\left(\hat{d}_{k-q,i}^{(m)}\right)\mathbf{1}_{\{\hat{d}_{k-q,i}^{(m)} \in \Psi_t\}}, \quad q \in \{0, 1, 2, ..., 11\}, i = 1, 2, \\[2ex]
\tilde{x}(q) := \dfrac{1}{\widetilde{X}(q)} \displaystyle\sum_t \sum_{m=1}^{17} \mathfrak{I}_t\left(\hat{x}_{k-q,1}^{(m)}\right)\mathbf{1}_{\{\hat{x}_{k-q,1}^{(m)} \in \Psi_t\}}, \quad q \in \{0, 1, 2, ..., 11\}, \\[2ex]
\tilde{y}(q) := \dfrac{1}{\widetilde{Y}(q)} \displaystyle\sum_t \sum_{m=1}^{17} \mathfrak{I}_t\left(Y_{k-q}^{(m)}\right)\mathbf{1}_{\{Y_{k-q}^{(m)} \in \Psi_t\}}, \quad q \in \{0, 1, 2, ..., 11\}.
\end{cases}
\tag{4.43}
$$

Note that the marginal averaging importance scores in the sets of Eqs. (4.42) and (4.43) serve as metrics for the dependence of the target variable on different features. In general, a higher value of the marginal averaging importance score implies that the corresponding feature is more important for the XGBoost classifier in the detection of the anomalous episodes.

The marginal total frequencies (with respect to a country index m) of the features selected through the RF module are presented as stacked bar charts in Figure 4.9. The plot shows that when predicting the FSI status $Z_k^{(m)}$ of a specific country, the RF module would more frequently select those features belonging to that specific country under consideration rather than the features belonging to other countries. Figure 4.10 is a visualisation of the marginal averaging importance scores (with respect to a country index m) of those selected features. It is noticeable that the average importance scores of the HMM-related features are higher than those of the non-HMM features. This result, along with the previous error analysis, further confirms that the features generated by the HMM have stronger impacts on forecasting the upcoming anomalous episodes. It could also be observed that in the process of detecting anomalies in a specific country's FSI, the features belonging to that country in question would have the highest importance scores relative to the features' scores of other countries. It could also be seen that, when the forecasting horizon h grows, the leading position of the importance score for each country's individual feature relegates to lower ranks. In other words, the predicted anomalous/normal state of a country's FSI tends to be increasingly affected by other countries' FSI status as the prediction horizon h increases. This phenomenon could be treated as

a numerical realisation of a country's idiosyncratic FSI and an aberration could propagate across countries to precipitate some systemic anomalous episodes.

Figure 4.11 exhibits the marginal total frequencies (with respect to a time lag q) of the features selected by the RF module. The chart shows that the features with $q = 0$ value are more likely to be selected by the RF. The marginal averaging importance scores (with respect to a time lag q) are illustrated in Figure 4.12. It could be observed that the HMM-related features get notably higher scores than the non-HMM features do, specifically, the original discriminative signals. We could also notice that the features with a time lag $q = 0$ achieve the highest importance scores. Nevertheless, the importance scores of features with a time lag $q > 0$ climb slightly as the forecasting horizon h rises. This indicates that the projected financial stress anomalous/nomal status is characterised by a strong short-memory dependence when the prediction step is small and by a weak long-memory dependence when the forecasting step increases. This serves as a numerical confirmation that a country's financial stress anomalies could accumulate to trigger a future crisis event.

Figure 4.9: The marginal total frequencies (with respect to a country index *m*) of the selected features

Figure 4.10: The marginal averaging importance scores (with respect to a country index m) of the selected features
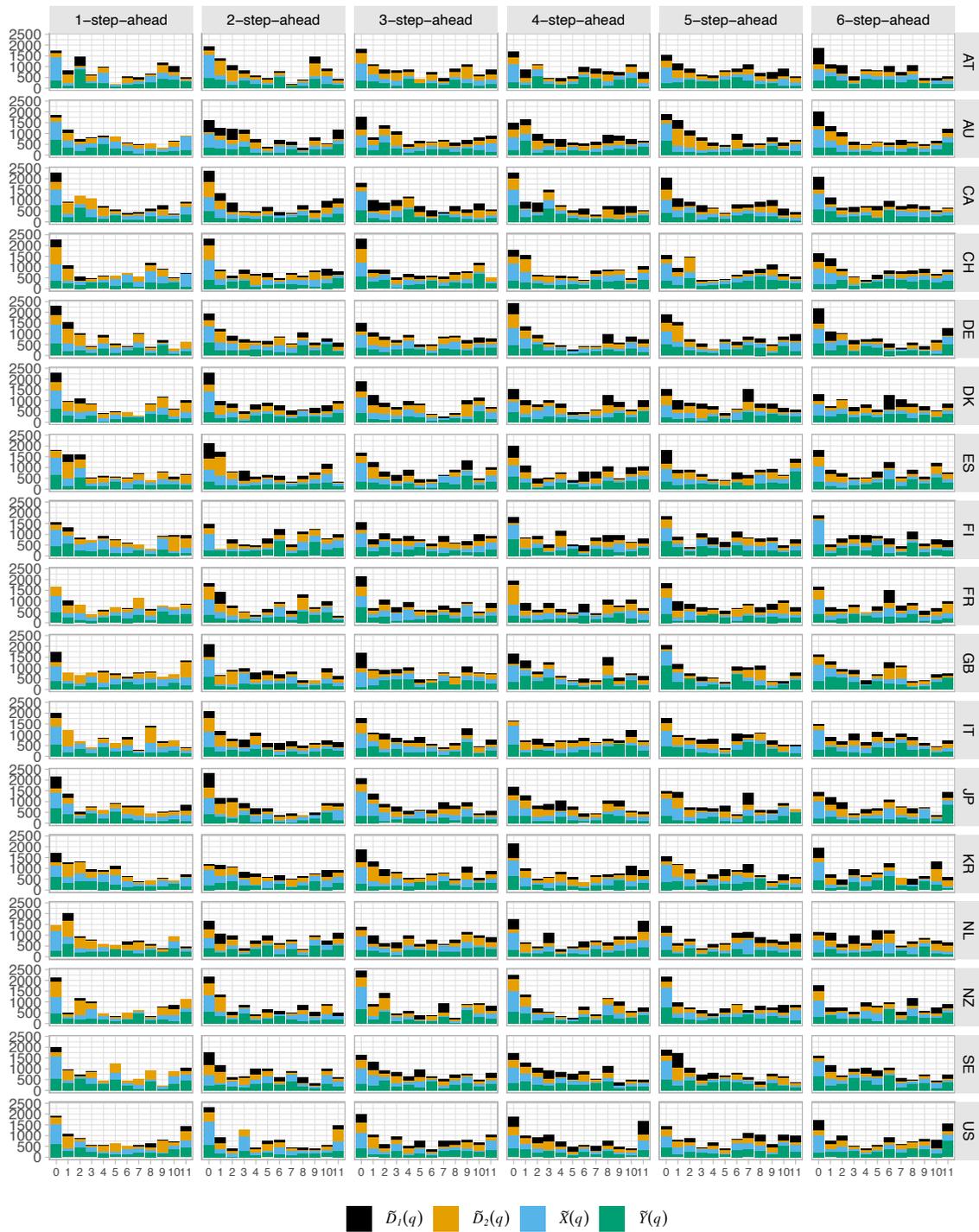
Figure 4.11: The marginal total frequencies (with respect to a time lag q) of the selected features

Figure 4.12: The marginal averaging importance scores (with respect to a time lag q) of the selected features

4.5 Forecasting with anomaly-warning signals

As defined in Eq. (4.34), $\tilde{Z}^{(m)}_{k+h}$, which is obtained from the probabilistic outcomes generated by the XGBoost classifier, provides a set of alert signals that quantify the FSI's exposure to aberrations in the future in a point-wise manner. In this section, we shall construct two warning signals to forecast the extent of the FSI anomaly in the near future. More specifically, given a fixed time horizon, we are going to forecast whether some extreme abnormality will happen which, in consequence, could precipitate high financial stress episodes in the short term.

We define two types of extreme abnormality episodes for all 17 countries' FSI as follows: (i) Multiple-Anomaly Episodes (MAE): $A^{(m)}_h$; and (ii) Consecutive-Anomaly Episodes (CAE): $H^{(m)}_h$. We say that the FSI enters the MAE at time t if the discriminative signal $Y^{(m)}_t \geq 0$ ($Z^{(m)}_t = 1$) for more than 50% of the time in the interval $[t,\ t + h - 1]$. Thus, we define the actual and projected Multiple-Anomaly-Episodes as follows:

$$
\begin{cases}
A^{(m)}_h := \left\{ t \;\middle|\; \frac{1}{h} \sum_{i=0}^{h-1} Z^{(m)}_{t+i} \geq 0.5 \right\}. \\[2mm]
\tilde{A}^{(m)}_h := \left\{ t \;\middle|\; \frac{1}{h} \sum_{i=0}^{h-1} \tilde{Z}^{(m)}_{t+i} \geq 0.5 \right\}.
\end{cases}
\tag{4.44}
$$

The FSI is said to move into a Consecutive-Anomaly Episodes (CAE) at time t if they exists, in $[t,\ t + h - 1]$, whenever at least one sub-interval with a length greater than 1 where $Y^{(m)}_t \geq 0$ ($Z^{(m)}_t = 1$) for all the time points in it. In other words, there exists at least one run of 1 for $Z^{(m)}_t$ in $[t,\ t + h - 1]$. Therefore, the actual and projected CAE could be expressed as

$$
\begin{cases}
H^{(m)}_h := \left\{ t \;\middle|\; \exists\ j, l \in \{0, 1, 2, \ldots h - 1\},\ j < l \text{ s.t. } \prod_{i=j}^{l} Z^{(m)}_{t+i} = 1 \right\} \\[2mm]
\tilde{H}^{(m)}_h := \left\{ t \;\middle|\; \exists\ j, l \in \{0, 1, 2, \ldots h - 1\},\ j < l \text{ s.t. } \prod_{i=j}^{l} \tilde{Z}^{(m)}_{t+i} = 1 \right\}.
\end{cases}
\tag{4.45}
$$

Note that in Eqs. (4.44) and (4.45), h is a prediction-step parameter which takes integer values in $[2, 6]$.

According to Eq. (4.44), $A^{(m)}_h$ incorporates the time points where $Z^{(m)}_t$ is positive in no less than 50% percent of the next h time steps which indicates that the FSI has rising edges multiple times within the next $h-$step-ahead points. Eq. (4.45) comprises those time points, where $Z^{(m)}_t$ is positive successively within an $h-$step-ahead horizon; this implies that the FSI is likely to surge on the next h time points. Since the actual MAE and CAE, $\tilde{A}^{(m)}_h$ and $\tilde{H}^{(m)}_h$, respectively, are related to $\tilde{Z}^{(m)}_t$ which is the predicted value of $Z^{(m)}_t$, the projected MAE and CAE could serve as the respective predictions of $A^{(m)}_h$ and $H^{(m)}_h$.

In order to capture the extremely anomalous episodes, we construct two alert signals utilising the

MAE and CAE: (i) Multiple-Anomaly Indicators (MAI); and (ii) Consecutive-Anomaly Indicators (CAI). The formulation for these two collections of warning signals are as follows:

$$
\begin{cases}
\text{MAI:} & U_{k,h}^{(m)} = \mathbf{1}_{\left\{k+1 \in \tilde{A}_h^{(m)}\right\}} \\
\text{CAI:} & V_{k,h}^{(m)} = \mathbf{1}_{\left\{k+1 \in \tilde{H}_h^{(m)}\right\}},
\end{cases}
\tag{4.46}
$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function and $h \in \{2, 3, \ldots, 6\}$. The MAI ($U_{k,h}^{(m)}$) takes the value 1 at time k when $\frac{1}{h}\sum_{i=0}^{h-1} \tilde{Z}_{k+1+i}^{(m)} \geq 0.5$ and 0 otherwise. Thus, it provides an aggregate value to project the mean intensity of the anomalies of the FSI in the upcoming h time points. The CAI ($V_{k,h}^{(m)}$) signals a positive at time k if there exists some integers j and l with $0 \leq j < l \leq h-1$ such that $\prod_{i=j}^{l} \tilde{Z}_{k+1+i}^{(m)} = 1$. Therefore, the CAI yields an aggregate value to predict the extent of the propagation of anomalies of FSI over the $h-$step ahead time horizon.

Figures 4.13 and 4.14 depict the MAI and CAI. Note that the binary signals are plotted in a descending order of the parameter h, from top to bottom in all graphs. We see that the signal captures the FSIs' rising edges (light purple shaded areas) for all the 17 countries included in our data set. Evidently, the MAI and CAI are capable of predicting the extent of the anomalies of FSI in the upcoming $h-$step-ahead horizon since all the signals are generated in an adapted way.

The error analysis for MAI is displayed in Table 4.12. The medians of the TPR for 3 and 6-step-ahead predictions are 80.97% and 76.89%, respectively. The medians of the ACC are 81.29% and 77.38% associated with the 3 and 6 forecasting horizons, respectively. According to the minimums of TPR and ACC of $U_{k,6}^{(m)}$, MAI is able to detect more than 71% of the MAE in the upcoming 2-6 months amongst all countries' FSI with false discovery rate (1-PPV) less than 29%.

The performance of the CAI is reported in Table 4.13. The TPR medians for the 3 and 6 prediction horizons are 78.95% and 77.15%, respectively. The ACC medians are 81.47% and 77.38% for the 3 and 6-step-ahead forecasting, respectively. The minimums of TPR and ACC of $V_{k,6}^{(m)}$ indicate that the CAI is capable of identifying at least 71% of the CAE in the next 2-6 months amongst all countries' FSI with false discovery rate less than 23%.

The performance measurements provided in Tables 4.12 and 4.13 justify that the prediction power of the proposed model could be further extended via the introduction of two anomaly-warning signals: MAI and CAI. The major advantages of MAI and CAI are the (i) capacity to detect and differentiate different types of extreme anomalies of FSI in advance; and (ii) specification of the time via the parameter h to capture extreme abnormalities. The two types of indicators, along with the point-wise predicted value $Z_k^{(m)}$, certainly provide a comprehensive quantitative framework to identify extreme abnormalities of FSI in the near future for practitioners.

Figure 4.13: The Multiple Anomaly Indicators (MAI): $U_{t,h}^{(m)}$

Figure 4.14: The Consecutive Anomaly Indicators (CAI): $V_{t,h}^{(m)}$

Table 4.12: Classification performance evaluation for the MAIs: $U_{k,h}^{(m)}$

h		AT	AU	CA	CH	DE	DK	ES	FI	FR	GB	IT	JP	KR	NL	NZ	SE	US	Min	Med	Max
	ACC	0.8434	0.8057	0.8377	0.8132	0.8264	0.8434	0.7981	0.8302	0.8075	0.7962	0.8472	0.7962	0.8226	0.8094	0.8434	0.8113	0.7943	0.7943	0.8132	0.8472
	TPR	0.8367	0.8188	0.8182	0.8293	0.8194	0.8459	0.8127	0.8185	0.8043	0.7953	0.8212	0.8051	0.8170	0.8000	0.8182	0.7915	0.8031	0.7915	0.8182	0.8459
2	TNR	0.8517	0.7873	0.8649	0.7942	0.8355	0.8400	0.7767	0.8472	0.8125	0.7979	0.8900	0.7834	0.8310	0.8256	0.8755	0.8442	0.7810	0.7767	0.8310	0.8900
	PPV	0.8754	0.8433	0.8936	0.8264	0.8657	0.8776	0.8421	0.8862	0.8691	0.8730	0.9249	0.8428	0.8780	0.8874	0.8934	0.8942	0.8482	0.8264	0.8754	0.9249
	NPV	0.8072	0.7565	0.7742	0.7975	0.7814	0.8008	0.7389	0.7625	0.7284	0.6906	0.7511	0.7359	0.7532	0.7061	0.7907	0.7089	0.7225	0.6906	0.7532	0.8072
	ACC	0.8412	0.7807	0.8166	0.8129	0.8355	0.8166	0.8072	0.8034	0.8412	0.7977	0.8091	0.7977	0.8393	0.8204	0.8318	0.7940	0.7788	0.7788	0.8129	0.8412
	TPR	0.8340	0.7712	0.7782	0.8233	0.8285	0.8097	0.8112	0.7903	0.8315	0.7790	0.7918	0.7967	0.8388	0.8106	0.8122	0.7395	0.7698	0.7395	0.8097	0.8388
3	TNR	0.8475	0.7884	0.8555	0.8047	0.8414	0.8227	0.8036	0.8149	0.8511	0.8168	0.8269	0.7986	0.8397	0.8302	0.8467	0.8470	0.7870	0.7870	0.8269	0.8555
	PPV	0.8273	0.7459	0.8449	0.7671	0.8115	0.8000	0.7860	0.7903	0.8506	0.8125	0.8256	0.7747	0.8153	0.8263	0.8017	0.8248	0.7668	0.7459	0.8115	0.8506
	NPV	0.8536	0.8105	0.7923	0.8536	0.8561	0.8315	0.8272	0.8149	0.8321	0.7839	0.7934	0.8188	0.8607	0.8148	0.8552	0.7695	0.7899	0.7695	0.8188	0.8607
	ACC	0.8030	0.7557	0.7898	0.7727	0.8011	0.7973	0.7936	0.7841	0.8239	0.7860	0.8011	0.7992	0.8182	0.7917	0.8163	0.7689	0.7670	0.7557	0.7936	0.8239
	TPR	0.8216	0.7821	0.7600	0.7868	0.8007	0.7840	0.7979	0.7708	0.8271	0.7778	0.7834	0.8007	0.8156	0.7799	0.8127	0.7533	0.7617	0.7533	0.7840	0.8271
4	TNR	0.7838	0.7258	0.8289	0.7578	0.8016	0.8133	0.7881	0.8000	0.8197	0.7981	0.8271	0.7975	0.8211	0.8082	0.8199	0.7902	0.7739	0.7258	0.8000	0.8289
	PPV	0.7978	0.7631	0.8539	0.7754	0.8155	0.8333	0.8233	0.8222	0.8531	0.8507	0.8693	0.8237	0.8394	0.8516	0.8220	0.8297	0.8136	0.7631	0.8237	0.8693
	NPV	0.8088	0.7469	0.7241	0.7698	0.7860	0.7597	0.7592	0.7442	0.7893	0.7083	0.7224	0.7720	0.7953	0.7224	0.8106	0.7024	0.7149	0.7024	0.7592	0.8106
	ACC	0.8197	0.7173	0.8083	0.7609	0.8197	0.8046	0.7932	0.7932	0.8216	0.7552	0.7761	0.7951	0.8216	0.7818	0.8178	0.7476	0.7552	0.7173	0.7932	0.8216
	TPR	0.8333	0.7156	0.7739	0.7585	0.8095	0.7860	0.7927	0.7875	0.7940	0.7280	0.7463	0.7886	0.8101	0.7689	0.8062	0.7043	0.7276	0.7043	0.7860	0.8333
5	TNR	0.8084	0.7185	0.8421	0.7629	0.8277	0.8204	0.7936	0.7979	0.8500	0.7820	0.8069	0.8007	0.8310	0.7947	0.8267	0.7889	0.7794	0.7185	0.8007	0.8500
	PPV	0.7843	0.6545	0.8279	0.7218	0.7857	0.7893	0.7708	0.7652	0.8446	0.7661	0.8000	0.7760	0.7967	0.7899	0.7787	0.7605	0.7427	0.6545	0.7787	0.8446
	NPV	0.8529	0.7722	0.7915	0.7957	0.8478	0.8175	0.8139	0.8179	0.8007	0.7455	0.7545	0.8123	0.8427	0.7741	0.8493	0.7370	0.7657	0.7370	0.8007	0.8529
	ACC	0.8042	0.7205	0.7909	0.7662	0.7947	0.7776	0.7719	0.7681	0.8004	0.7738	0.7548	0.7776	0.8042	0.7605	0.7795	0.7452	0.7357	0.7205	0.7738	0.8042
	TPR	0.8346	0.7366	0.7587	0.7744	0.7913	0.7656	0.7690	0.7727	0.7884	0.7641	0.7318	0.7818	0.8060	0.7492	0.7689	0.7220	0.7188	0.7188	0.7689	0.8346
6	TNR	0.7744	0.7045	0.8292	0.7577	0.7978	0.7905	0.7751	0.7634	0.8155	0.7867	0.7857	0.7729	0.8023	0.7753	0.7901	0.7749	0.7563	0.7045	0.7753	0.8292
	PPV	0.7834	0.7122	0.8411	0.7658	0.7852	0.7977	0.7918	0.7669	0.8431	0.8273	0.8216	0.7904	0.8090	0.8145	0.7868	0.8038	0.7811	0.7122	0.7918	0.8431
	NPV	0.8273	0.7294	0.7425	0.7665	0.8037	0.7576	0.7510	0.7692	0.7540	0.7137	0.6848	0.7638	0.7992	0.7012	0.7724	0.6858	0.6897	0.6848	0.7540	0.8273

Table 4.13: Classification performance evaluation for the CAIs: $V_{k,h}^{(m)}$

h		AT	AU	CA	CH	DE	DK	ES	FI	FR	GB	IT	JP	KR	NL	NZ	SE	US	Min	Med	Max
	ACC	0.8642	0.8245	0.8547	0.8283	0.8509	0.8509	0.8377	0.8208	0.8528	0.8151	0.8358	0.8472	0.8491	0.8226	0.8491	0.8415	0.8132	0.8132	0.8415	0.8642
	TPR	0.8520	0.7955	0.8241	0.7869	0.8462	0.8209	0.8316	0.8043	0.8519	0.7500	0.7915	0.8238	0.8396	0.8223	0.8222	0.7970	0.7903	0.7500	0.8222	0.8520
2	TNR	0.8713	0.8390	0.8758	0.8501	0.8534	0.8693	0.8413	0.8295	0.8535	0.8545	0.8652	0.8605	0.8542	0.8228	0.8629	0.8679	0.8256	0.8228	0.8542	0.8758
	PPV	0.7952	0.7107	0.8203	0.7347	0.7512	0.7933	0.7546	0.7150	0.8000	0.7576	0.7952	0.7718	0.7585	0.7330	0.7551	0.7811	0.7101	0.7101	0.7576	0.8203
	NPV	0.9094	0.8919	0.8786	0.8832	0.9138	0.8882	0.8949	0.8885	0.8933	0.8494	0.8625	0.8951	0.9071	0.8867	0.9042	0.8784	0.8793	0.8494	0.8885	0.9138
	ACC	0.8526	0.7732	0.8185	0.8053	0.8318	0.8147	0.8147	0.7940	0.8336	0.7902	0.8072	0.8091	0.8393	0.8166	0.8185	0.7940	0.7713	0.7713	0.8147	0.8526
	TPR	0.8489	0.7535	0.7720	0.7917	0.8203	0.7931	0.8139	0.7798	0.8347	0.7427	0.7724	0.7895	0.8465	0.8109	0.7895	0.7393	0.7723	0.7393	0.7895	0.8489
3	TNR	0.8553	0.7866	0.8602	0.8147	0.8397	0.8316	0.8154	0.8039	0.8327	0.8299	0.8375	0.8239	0.8344	0.8213	0.8375	0.8373	0.7705	0.7705	0.8316	0.8602
	PPV	0.8128	0.7074	0.8319	0.7467	0.7807	0.7863	0.7737	0.7359	0.8150	0.7851	0.8051	0.7725	0.7778	0.7878	0.7604	0.7828	0.7119	0.7074	0.7807	0.8319
	NPV	0.8844	0.8233	0.8081	0.8500	0.8704	0.8373	0.8497	0.8389	0.8509	0.7940	0.8089	0.8378	0.8881	0.8415	0.8590	0.8019	0.8217	0.7940	0.8389	0.8881
	ACC	0.8390	0.7386	0.7784	0.7708	0.8182	0.7898	0.8030	0.7727	0.8201	0.7860	0.7973	0.7803	0.8182	0.7955	0.8049	0.7576	0.7595	0.7386	0.7898	0.8390
	TPR	0.8419	0.7362	0.7394	0.7671	0.8135	0.7643	0.8075	0.7698	0.8107	0.7553	0.7857	0.7719	0.8230	0.7849	0.8025	0.7159	0.7557	0.7159	0.7719	0.8419
4	TNR	0.8364	0.7409	0.8238	0.7742	0.8225	0.8151	0.7985	0.7754	0.8306	0.8211	0.8105	0.7887	0.8140	0.8072	0.8069	0.8016	0.7632	0.7409	0.8072	0.8364
	PPV	0.8256	0.7248	0.8300	0.7520	0.8071	0.8040	0.8015	0.7578	0.8439	0.8288	0.8240	0.7838	0.7905	0.8202	0.7733	0.7918	0.7586	0.7248	0.8015	0.8439
	NPV	0.8519	0.7519	0.7309	0.7883	0.8285	0.7770	0.8046	0.7868	0.7954	0.7454	0.7701	0.7770	0.8436	0.7701	0.8327	0.7279	0.7603	0.7279	0.7770	0.8519
	ACC	0.8046	0.7400	0.7571	0.7571	0.7970	0.7780	0.7837	0.7476	0.8065	0.7837	0.7799	0.7818	0.7913	0.7666	0.7913	0.7324	0.7533	0.7324	0.7799	0.8065
	TPR	0.8333	0.7657	0.7184	0.7706	0.7978	0.7552	0.7904	0.7545	0.8140	0.7588	0.7818	0.7751	0.7955	0.7630	0.8092	0.7209	0.7543	0.7184	0.7706	0.8333
5	TNR	0.7743	0.7095	0.8119	0.7419	0.7960	0.8059	0.7754	0.7400	0.7965	0.8194	0.7773	0.7899	0.7871	0.7717	0.7736	0.7478	0.7521	0.7095	0.7754	0.8194
	PPV	0.7951	0.7578	0.8441	0.7706	0.8125	0.8264	0.8127	0.7628	0.8419	0.8582	0.8304	0.8175	0.7895	0.8246	0.7794	0.7920	0.7921	0.7578	0.8125	0.8582
	NPV	0.8156	0.7185	0.6705	0.7419	0.7804	0.7290	0.7500	0.7312	0.7627	0.7024	0.7185	0.7431	0.7931	0.6983	0.8039	0.6680	0.7097	0.6680	0.7312	0.8156
	ACC	0.7966	0.7452	0.7433	0.7567	0.7776	0.7757	0.7776	0.7338	0.7776	0.7833	0.7738	0.7510	0.7738	0.7567	0.7757	0.7357	0.7510	0.7338	0.7738	0.7966
	TPR	0.8182	0.7683	0.7100	0.7738	0.7900	0.7692	0.7917	0.7450	0.7875	0.7715	0.7879	0.7613	0.7965	0.7522	0.7986	0.7455	0.7563	0.7100	0.7715	0.8182
6	TNR	0.7708	0.7109	0.8000	0.7330	0.7611	0.7850	0.7570	0.7193	0.7621	0.8042	0.7500	0.7361	0.7469	0.7644	0.7490	0.7194	0.7427	0.7109	0.7500	0.8042
	PPV	0.8097	0.7987	0.8577	0.8000	0.8144	0.8392	0.8261	0.7762	0.8372	0.8754	0.8414	0.8055	0.7882	0.8485	0.7875	0.8173	0.8203	0.7762	0.8173	0.8754
	NPV	0.7806	0.6726	0.6190	0.7013	0.7319	0.7000	0.7137	0.6833	0.6978	0.6638	0.6774	0.6824	0.7563	0.6376	0.7615	0.6267	0.6623	0.6190	0.6833	0.7806

4.6   Conclusion

The key contribution of this work is the development of a hybrid multivariate supervised learning device. Our algorithmic approach integrated stochastic process modelling, hidden Markov model,

random forest and XGBoost. The application in this research focused on tracing the movement of multiple countries' FSIs simultaneously. The ultimate goal was the detection in advance of signs for possible crisis episodes that will disturb the financial stability.

A discriminative signal, which is evaluated as the lag-τ difference of the original data series, is introduced to characterise the anomalous behaviour of all FSI time series. The OU processes, with parameters governed by the HMM, are employed to model 17 different countries' discriminative signals altogether. Utilising the EM algorithm and change of reference probability measures, the HMM online recursive filters were constructed. This in turn, unveils the information content of the observed multiple stochastic processes. Two new features, ie., estimate of the Markov-chain state and deviation from the mean, were created to describe the characteristics of the underlying time series under the HMM filtering framework. A feature-selection module was established that consequently improves the out-of-sample predictive performance at a lower the computing cost of modelling. Random Forest was utilised to select the HMM and Non-HMM predictors concurrently based on the OOB accuracy.

XGBoost is the final stage classifier in our modelling approach. Time series cross-validation combined with grid search was applied to tune the hyper-parameters. In order to fully capture the evolution of the input data series, the classification model is tuned and trained once new data is available. To avoid error propagation in multi-step-ahead forecasting, a direct prediction method was applied to generate a multivariate 6-step ahead predictions for the countries' financial stress status, which served as an early-warning signal for the future occurrence of financial-crisis episodes.

Benchmarking with other five models, our model's implementability and predictive performance were tested on the actual multi-dimensional time series of 17 countries' FSI. Each models' predictive power was assessed though four different model-diagnostic tools: (i) confusion matrix, (ii) area under ROC, (iii) logistic loss, and (iv) Kolmogorov-Smirnov test. The confusion-matrix analysis showed that our model had the highest TPR, TNR, PPV and NPV scores in all forecasting horizons. This means that our proposed model is able to detect positive (i.e., anomalous) episodes more accurately with lower false positive error than the benchmark models when the discrimination threshold is set to 0.5. The comparison of AUCs showed that the proposed model scored higher than the other models at all prediction steps. The outcomes of the pairwise t-test and DeLong test supported the statistical significance of our model's AUC score. The logistic loss of the models revealed that our model has better capability in determining anomalous periods; this is based on higher predicted probabilities for positive instances than for negative ones. The result of the Kolmogorov-Smirnov test attested to the robustness of our proposed modelling approach in differentiating anomalous events from regular ones. Our modelling methodology for multivariate prediction substantially benefited from the fusion of RF, HMM and XGBoost.

A customised feature-importance analysis was conducted to measure the features' contribution to detect anomalous episodes. The results suggested that the HMM-features have significant impacts on the prediction accuracy. Our findings also indicate that, as the forecasting horizon increases, the prediction of anomalies of each country's FSI could have a long-memory effect on other countries' features.

Employing the anomalous/normal classification results, two projected anomaly-warning signals (MAI and CAI) are constructed to predict two types of extremely anomalous events (MAE and CAE). The results show that our proposed modelling approach is able to detect the specified extremely anomalous episodes in the near future (2-6 months ahead) with a good accuracy.

This research work contributed to supervised learning, multivariate time series, and financial-stability modelling in the following ways: (i) Development of a new framework integrating HMM with the supervised learning models: RF and XGBoost. This enabled us to trace with accuracy the dynamics of multiple FSI time series in an attempt to detect future financial stability; (ii) Utility of a multiple self-tuning estimation procedure under the HMM set up to generate new features for classification; (iii) Implementation of a 2-dimensional feature selection using RF; (iv) Application of time series cross validation and grid search for hyper-parameter tuning for XGBoost; (v) Independent training framework for different prediction steps; (vi) A comprehensive model performance diagnostics using multiple statistical tools. (vii)A customised analysis of the features' importance to assess and interpret different types of input variables' impact on the anomaly detection; and (viii) Construction of two early warning alert signals MAI and CAI to detect extremely anomalous episodes in the future.

This work laid out a new ensemble supervised learning classification model which is designed to detect future anomalous episode location for the benefit of financial-stability research and pertinent policy implications. Our proposed model provides predictive analytics for country-level FSI status which is relevant to financial-stability management. Our approach formulated a framework that allows us to recast our solution to anomalous/normal classification outcomes. Since these quantities are also given in a probabilistic form, they have natural interpretations as an estimated level of exposure to financial instability which is relevant to central bankers and economists. In addition, we conducted a tailored features' importance analysis to quantify the relationship between the predicted FSI's anomalous/normal status and different types of input variables; this in turn provides accessible interpretation of our proposed modelling to practitioners. Lastly, we enhanced further our model by constructing two projected warning signals, MAI and CAI, to forecast some specified extremely anomalous episodes in the near future. The said enhancement gives additional quantitative insights for the policy makers in mitigating possible financial-crisis event.

Chapter 5

Concluding remarks

5.1   Summary and commentaries

In this thesis, we designed various HMM-based EWAS to detect instances of financial instability based on signal alerts extracted from multiple financial and economic indices. We briefly described these *particular* EWAS. First, a theoretical framework that incorporates HMM into a multivariate setting was constructed to capture stylised characteristics of multivariate economic-indicator time series. Second, taking advantage of the EM algorithm and change of reference probability measures, we derived new multivariate HMM online recursive filtering algorithms to underpin the models' dynamic parameter estimation. Third, utilising HMM's valuable capabilities in extracting essential information from the observed data series, we built a hybrid supervised learning device by integrating stochastic modelling and HMM with Random Forest and XGBoost. This facilitates the identification of signs in multiple FSIs for possible episodes that will disturb financial stability. As a result, early warning signals arising from specific input data series were obtained when various anomalous behaviour were detected.

More *specifically*, the accomplishments in this research commenced with the construction of a hybrid bivariate OU-GBM regime-switching model with parameters driven by a hidden Markov chain to describe the features of FSI and IPI in Chapter 2. We created an early warning device for financial crisis based on the ensuing HMM filtering recursions. In Chapter 3, a 4-dimensional blended HMM-based framework was proposed to measure the market liquidity by capturing the joint dynamics of 4 financial indices. In addition, we tested the capacity of new EWAS in detecting financial instability with the use of HMM-state estimates along with a statistical assessment. In Chapter 4, we introduced a supervised learning method with the amalgamation of stochastic modelling and hidden

Markov model with Random Forest and XGBoost to detect anomalies in multiple financial stress indices. Two early warning signals were obtained through the HMM filters and XGBoost classier, thereby identifying extreme anomalous episodes.

Generally, this research work contributed thinking paradigms to HMM modelling, supervised learning, multivariate time series, and financial-stability modelling in the following ways: (i) A fusion-information-based design of a framework integrating different stochastic processes within HMM, which is a new approach in capturing jointly the important properties of series associated with tracking down financial instability; (ii) Construction of a new framework integrating HMM with the supervised learning models to trace with accuracy the dynamics of multiple FSI time series in an attempt to detect future financial stability; (iii) Development of an extended self-tuning estimation procedure under a multivariate HMM set up bringing forth new features for other machine learning models to detect anomalies; (iv) Implementation of independent model training strategy for different prediction horizons; and (v) Creation of various early-warning signals employing outputs from the interplay of HMM filters and some machine learning algorithms.

This thesis offers insights and research advances by developing a set of quantitative methodologies to assess the level of financial stress and market illiquidity. The time points of financial instability occurrences that were pinpointed out by our modelling approaches accurately correspond to those already established and accepted by practitioners. The outputs from our proposed models are straightforward to interpret with direct relevance to bankers, economists, regulators and policy makers.

5.2   Future research directions

Although we made significant progress with our proposed modelling setups for financial-stability analysis, certain limitations open opportunities. The contributions of this work are designed to stimulate further research activities with both theoretical and practical considerations. Potential enhancements or alternative quantitative methodologies are continually sought in the areas of financial stability management. The application aspect of this work would be of value to central bankers and regulatory authorities whose tasks are to mitigate the effects or the occurrence of future financial crises. We outline below several ramifications emanating from the various results of this research work.

- In Chapters 2 and 3, additional research investigations are needed to determine accurately the lag between the EWAS and the crisis taking place. It is desirable to come up with certain statistical models supporting a rigorous procedure that connects a pre-crisis alert from the

HMM filtering methodology with the actual outcomes.

- So far, our models are all based on regular HMM. However, it may be challenging for a HMM-driven to capture adequately some long-memory characteristics of financial indices. It is, therefore, ideal to introduce a higher-order HMM (HOHMM) that would drive the underlying process. In turn, as the state memory is recovered, filters could provide more informative estimates generating EWAS with better forecasting performance.

- All of our proposed models are formulated as stochastic processes whose parameters are governed by an HMM only. When dealing with real-world financial data, it is not uncommon that drift and volatility are influenced by different factors. Thus, it is rational to build modelling frameworks supported by stochastic processes whose parameters are modulated by different HMMs or HOHMMs to strengthen the framework's flexibility.

- The multivariate stochastic models considered in this research study are based on the assumption that all component processes have independent noise term; such an assumption may not be realistic with every financial time series encountered in practice. Henceforth, there is merit in putting forward multivariate stochastic processes with non-zero correlation, which could also be dependent on HMM or HOHMM. The benefits of HOHMM are elaborated in [90] and the references therein.

- In Chapter 4, we created a hybrid model combining the HMM with RF and XGBoost. The result demonstrated that HMMs could be utilised as an ideal feature-generating device in a machine learning framework. Implementation of the HMM in combination with a certain type of Neural Network is also anticipated to bring high predictive power in the context of big-data processing environment.

- In the field of machine learning, both the HMM and HOHMM are deemed unsupervised learning models. As demonstrated in this work, the HMM could be applied with a supervised learning model to achieve good forecasting results. In this thesis, setting the number of regimes to 2 definitely simplifies the mechanics of the HMM development. A generalisation, to pin down data's stylised facts, could be explored by constructing a methodology that ensures the the HMM or HOHMM satisfies certain objectives, i.e.,by, for example, minimising the MSE or classification error. More specifically, we could treat the number of states of the HMM and the memory length of HOHMM as hyper-parameters. The optimal values could be searched via cross validation.

Bibliography

[1] A. Afonso, M. Arghyrou, M. Gadea, and A. Kontonikas. "Whatever it takes" to resolve the european sovereign debt crisis? Bond pricing regime switches and monetary policy effects. *Journal of International Money and Finance*, 86:1–30, 2018.

[2] L. Alessi and C. Detken. Identifying excessive credit growth and leverage. *Journal of Financial Stability*, 35:215–225, 2018.

[3] S. Basak, S. Kar, S. Saha, L. Khaidem, and S. Dey. Predicting the direction of stock market prices using tree-based classifiers. *North American Journal of Economics and Finance*, 47: 552–567, 2019.

[4] G. Bekaert, M. Hoerova, and M. Duca. Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788, 2013.

[5] C. Bergmeir and J. Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012.

[6] L. Breiman. *Classification and Regression Trees*. Chapman and Hall/CRC, London, 1984.

[7] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[8] M. Brunnermeier. Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives*, 23(1):77–100, 2009.

[9] M. Brunnermeier and L. Pedersen. Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238, 2009.

[10] M. Brunnermeier, S. Nagel, and L. Pedersen. Carry trades and currency crashes. *NBER Macroeconomics Annual*, 23:313–348, 2008.

[11] Y. Cao, Y. Li, S. Coleman, A. Belatreche, and T. McGinnity. Adaptive hidden Markov model with anomaly states for price manipulation detection. *IEEE Transactions on Neural Networks and Learning Systems*, 26(2):318–330, 2014.

[12] E. Casabianca, M. Catalano, L. Forni, E. Giarda, and S. Passeri. An early warning system for banking crises: From regression-based analysis to machine learning techniques. *Marco Fanno Working Papers 235*, 2019.

[13] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[14] Y. Chen, R. Mamon, F. Spagnolo, and N. Spagnolo. Renewable energy and economic growth: A markov-switching approach. *Energy*, 244:123089, 2022.

[15] H. Cheng, P. Tan, J. Gao, and J. Scripps. Multistep-ahead time series prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 765–774. Springer, 2006.

[16] P. Date and K. Ponomareva. Linear and non-linear filtering in mathematical finance: A review. *IMA Journal of Management Mathematics*, 22(3):195–211, 2011.

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

[18] T. Daviga and J. Gerlach. State-dependent stock market reactions to monetary policy. *International Journal of Central Banking*, 2(4):65–83, 2006.

[19] E. DeLong, D. DeLong, and D. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845, 1988.

[20] D. Dickey and W. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 2(4):65–83, 1979.

[21] T. Duprey, B. Klaus, and T. Peltonenc. Dating systemic financial stress episodes in the EU countries. *Journal of Financial Stability*, 32:357–384, 2017.

[22] R. Duttagupta and P. Cashin. Anatomy of banking crises in developing and emerging market countries. *Journal of International Money and Finance*, 30(2):354–376, 2011.

[23] R. Elliott and V. Krishnamurthy. New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models. *IEEE Transactions on Automatic Control*, 44(5):938–951, 1999.

[24] R. Elliott and R. Mamon. An interest rate model with a Markovian mean-reverting level. *Quantitative Finance*, 2(6):454–458, 2002.

[25] R. Elliott, L. Aggoun, and J. Moore. *Hidden Markov Models: Estimation and Control*. Springer, New York, 1995.

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

[27] C. Erlwein and R. Mamon. An online estimation scheme for a Hull-White model with HMM-driven parameters. *Statistical Methods and Applications*, 18(1):87–107, 2009.

[28] C. Erlwein, R. Mamon, and M. Davison. An examination of HMM-based investment strategies for asset allocation. *Applied Stochastic Models in Business and Industry*, 27(3):204–221, 2009.

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

[30] M. Fioramanti. Predicting sovereign debt crises using artificial neural networks: A comparative approach. *Journal of Financial Stability*, 4(2):149–164, 2008.

[31] J. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001.

[32] B. Gadanecz and K. Jayaram. Measures of financial stability-a review. *Irving Fisher Committee Bulletin*, 31(1):365–383, 2008.

[33] P. Garthwaite, I. Jolliffe, and B. Jones. *Statistical Inference (2nd Edition)*. Oxford Science Publication, Oxford, 2002.

[34] K. Gerdrup, F. Hansen, T. Krogh, and J. Maih. Leaning against the wind when credit bites back. *International Journal of Central Banking*, 13(3):287–320, 2017.

[35] P. Giot. Relationships between implied volatility indexes and stock index returns. *Journal of Portfolio Management*, 31(3):92–100, 2005.

[36] R. Gopalan, O. Kadan, and M. Pevzner. Asset liquidity and stock liquidity. *Journal of Financial and Quantitative Analysis*, 47(2):333–364, 2012.

[37] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[38] D. Gromba and D. Vayanos. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics*, 66(2–3):36–407, 2002.

[39] X. Gu, R. Mamon, M. Davison, and H. Yu. An automated financial indices-processing scheme for classifying market liquidity regimes. *International Journal of Control*, 94(3):735–756, 2021.

[40] X. Gu, R. Mamon, T. Duprey, and H. Xiong. Online estimation for a predictive analytics platform with a financial-stability-analysis application. *European Journal of Control*, 57:205–221, 2021.

[41] J. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

[42] P. Hao, Y. Zhan, L. Wang, Z. Niu, and M. Shakir. Feature selection of time series modis data for early crop classification using random forest: A case study in Kansas, USA. *Remote Sensing*, 7(5):5347–5369, 2015.

[43] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition*. Springer, New York, 2017.

[44] S. Hauser and H. K. Levy. Liquidity might come at cost: The role of heterogeneous preferences. *Journal of Financial Markets*, 39:1–23, 2018.

[45] G. Huberman and H. Dominika. Systemic liquidity. *Journal of Financial Research*, 24(2):161–178, 2001.

[46] K. Hubrich and R. Tetlow. Financial stress and economic dynamics: The transmission of crises. *Journal of Monetary Economics*, 70:100–115, 2015.

[47] R. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2018.

[48] H. Ichiue and K. Koyama. Regime switches in exchange rate volatility and uncovered interest parity. *Journal of International Money and Finance*, 30(7):1436–1450, 2011.

[49] M. Illing and Y. Liu. Measuring financial stress in a developed country: An application to Canada. *Journal of Financial Stability*, 2:243–265, 2006.

[50] G. Imanov. Early warning fuzzy system of the financial stability. In *International Conference on Application of Information and Communication Technologies*, pages 100–115, 2009. URL http://dx.doi.org/10.1109/ICAICT.2009.5372596.

[51] S. Karasu and A. Altan. Recognition model for solar radiation time series based on random forest with feature selection approach. In *11th International Conference on Electrical and Electronics Engineering*, pages 8–11. IEEE, 2019.

[52] J. Karpoff and R. Walkling. Short-term trading around ex-dividend days. *Journal of Financial Economics*, 21(2):291–298, 1988.

[53] J. M. Keynes. *A Treatise on Money*. Macmillan, London, 1930.

[54] D. Kwiatkowski, P. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, J54(1–3):159–178, 1992.

[55] A. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1336, 1985.

[56] J. Li, W. Pedrycz, and I. Jamal. Multivariate time series anomaly detection: A framework of hidden Markov models. *Applied Soft Computing*, 60:229–240, 2017.

[57] L. Li and Q. Zhang. XGboost: A scalable tree boosting system. In *International Conference on E-Product E-Service and E-Entertainment*, pages 100–115, 2010. URL http://dx.doi.org/10.1109/ICEEE.2010.5660151.

[58] N. Liang. Systemic risk monitoring and financial stability. *Journal of Money, Credit and Banking*, 45(s1):129–135, 2013.

[59] R. Mamon, C. Erlwein, and B. Gopaluni. Adaptive signal processing of asset price dynamics with predictability analysis. *Information Sciences*, 178(1):203–219, 2008.

[60] R. Mamon, H. Xiong, and Y. Zhao. The valuation of a guaranteed minimum maturity benefit under a regime-switching framework. *North American Actuarial Journal*, 25(3):334–359, 2021.

[61] J. Mamplata and R. M. abd G. David. Modelling and filtering for dynamic investment in the precious-metals market. *International Journal of Computer Mathematics*.

[62] J. Moors. The meaning of kurtosis: Darlington reexamined. *The American Statistician*, 40(4): 283–284, 1986.

[63] T. Murao, K. Hirata, Y. Okajima, and K. Uchida. Real-time pricing for lqg power networks with independent types: A dynamic mechanism design approach. *European Journal of Control*, 39:95–105, 2018.

[64] J. Nobre and R. Neves. Combining principal component analysis, discrete wavelet transform and xgboost to trade in the financial markets. *Expert Systems with Applications*, 125:181–194, 2019.

[65] A. Patton, D. Politis, and H. White. Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. *Econometric Reviews*, 28(4):372–375, 2009.

[66] D. Politis and J. Romano. Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*, 4(2):461–476, 1994.

[67] D. Politis and H. White. Automatic block-length selection for the dependent bootstrap. *Econometric reviews*, 23(1):53–70, 2004.

[68] R. Schapire. A Brief Introduction to Boosting. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, volume 2, pages 1401–1406, 1999.

[69] S. Shieh, C. Lin, and P. Ho. Large changes in stock prices: Market, liquidity, and momentum effect. *Quarterly Review of Economics and Finance*, 52(2):183–197, 2012.

[70] T. Siu and R. Elliott. American option pricing and filtering with a hidden regime-switching jump diffusion. *The Journal of Derivatives*, 2022.

[71] L. Smales and J. Kininmonth. FX market returns and their relationship to investor fear. *International Review of Finance*, 16(4):659–675, 2016.

[72] Y. Song, N. Cai, and S. Kou. Computable error bounds of Laplace inversion for pricing Asian options. *INFORMS Journal on Computing*, 30(4):372–375, 2018.

[73] S. Taieb, A. Sorjamaa, and G. Bontempi. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, 73(10-12):1950–1957, 2010.

[74] A. Tenyakov and R. Mamon. A computing platform for pairs-trading online implementation via a blended Kalman-HMM filtering approach. *Journal of Big Data*, 4:46, 2017.

[75] A. Tenyakov, R. Mamon, and M. Davison. Modelling high-frequency FX rate dynamics: A zero-delay multi-dimensional HMM-based approach. *Knowledge-Based Systems*, 101:142–155, 2016.

[76] A. Tenyakov, R. Mamon, and M. Davison. Filtering of a discrete-time HMM-driven multivariate Ornstein-Uhlenbeck model with application to forecasting market liquidity regimes. *IEEE Journal of Selected Topics in Signal Processing*, 10(6):994–1005, 2016.

[77] D. Thornton and D. Batten. Lag-length selection and tests of Granger causality between money and income. *Journal of Money, Credit and Banking*, 17(2):164–178, 1985.

[78] H. Tyralis and G. Papacharalampous. Variable selection in time series forecasting using random forests. *Algorithms*, 10(4):114, 2017.

[79] A. van der Vaart. *Asymptotic Statistics*. Cambridge Univerity Press, Cambridge, 1998.

[80] D. Vayanos and J. Wang. Market liquidity–theory and empirical evidence. *Handbook of the Economics of Finance*, 2(B):1289–1361, 2013.

[81] A. Venkatraman, M. Hebert, and J. Bagnell. Improving multi-step prediction of learned time series models. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[82] I. Visser, M. Raijmakers, and P. Molenaar. Confidence intervals for hidden Markov model parameters. *British Journal of Mathematical and Statistical Psychology*, 53:317–327, 2000.

[83] F. Ward. Spotting the danger zone: Forecasting financial crises with classification tree ensembles and many predictors. *Journal of Applied Econometrics*, 32(2):359–378, 2017.

[84] P. Westfall. Kurtosis as peakedness, 1905-2014. R.I.P. *The American Statistician*, 68(3):191–195, 2014.

[85] R. Whaley. Understanding the VIX. *Journal of Portfolio Management*, 35(3):98–105, 2009.

[86] C. Wu. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11(1):95–103, 1983.

[87] X. Xi and R. Mamon. Parameter estimation of an asset price model driven by a weak hidden Markov chain. *Economic Modelling*, 28(1-2):36–46, 2011.

[88] X. Xi and R. Mamon. Capturing the regime-switching and memory properties of interest rates. *Computational Economics*, 44(3):307–337, 2013.

[89] R. Xiang, C. Jones, R. Mamon, and M. Chavez. Modelling exchange-driven fish price dynamics. *Journal of Modelling in Management*, 2021.

[90] H. Xiong and R. Mamon. An enabling framework for automated extraction of signals from market information in real time. *Knowledge-Based Systems*.

[91] H. Xiong and R. Mamon. A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics. *Journal of Computational Science*, 17(1):47–61, 2016.

[92] H. Xiong and R. Mamon. Putting a price tag on temperature. *Computational Management Science*, 15(2):1–38, 2017.

[93] F. Zapatero. Equilibrium asset prices and exchange rates. *Journal of Economic Dynamics and Control*, 19(4):787–811, 1995.

[94] W. Zhang, X. Zhao, and Z. Li. A comprehensive study of smartphone-based indoor activity recognition via xgboost. *IEEE Access*, 7:80027–80042, 2019.

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

Appendix A

Derivation of optimal parameter estimates in Chapter 2

A.1   Optimal estimate for π_{ji}

The idea of the proof is similar to that in [59]. To perform the measure change mentioned in Subsection E.1, define a new measure $P^{\hat{\pi}_{ji}}$ via $\left.\dfrac{dP^{\hat{\pi}_{ji}}}{dP^{\pi_{ji}}}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k} \lambda_l^*$, where

$$\lambda_l^* = \sum_{j,i=1}^{N} \left(\frac{\hat{\pi}_{ji}}{\pi_{ji}}\right)^{\langle \mathbf{z}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{z}_l, \mathbf{e}_j \rangle}.$$

So,

$$\log \frac{dP^{\hat{\pi}_{ji}}}{dP^{\pi_{ji}}} = \sum_{l=1}^{k} \sum_{j,i=1}^{N} \left(\log\left(\hat{\pi}_{ji}\right) - \log\left(\pi_{ji}\right)\right) \langle \mathbf{z}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{z}_l, \mathbf{e}_j \rangle = \sum_{j,i=1}^{N} \mathcal{G}_k^{j,i} \log\left(\hat{\pi}_{ji}\right) + R,$$

where R does not contain $\hat{\alpha}$.

It must be noted that $\sum_{j=1}^{N} \mathcal{G}_k^{j,i} = O_k^i$, $\sum_{j=1}^{N} \hat{\mathcal{G}}_k^{j,i} = \hat{O}_k^i$. The optimal estimate for π_{ji} is the value that maximises the log-likelihood $\log \dfrac{dP^{\hat{\pi}_{ji}}}{dP^{\pi_{ji}}}$ constrained to $\sum_{j=1}^{N} \hat{\pi}_{ji} = 1$. Introducing the Lagrange multiplier and solving the associated optimisation problem yield the solution

$$\hat{\pi}_{ji} = \frac{\hat{\mathcal{G}}_k^{j,i}}{\hat{O}_k^i}.$$

∎

A.2 Optimal estimate for α

Let $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)^\top \in \mathbb{R}^N$. To calculate $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N)^\top \in \mathbb{R}^N$, consider a new measure $P^{\hat{\alpha}}$ defined by $\left. \dfrac{dP^{\hat{\alpha}}}{dP^\alpha} \right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*$, where

$$\lambda_l^* = \exp\left\{ \frac{(X_l - \alpha\,(\mathbf{z}_{l-1})\,X_{l-1} - \beta\,(\mathbf{z}_{l-1}))^2 - (X_l - \hat{\alpha}\,(\mathbf{z}_{l-1})\,X_{l-1} - \beta\,(\mathbf{z}_{l-1}))^2}{2\kappa^2\,(\mathbf{z}_{l-1})} \right\}.$$

This implies

$$
\begin{aligned}
LR\,(\hat{\alpha}_i) &= E\left[\log \left. \frac{dP^{\hat{\alpha}}}{dP^\alpha} \right| \mathscr{F}_k \right] \\
&= E\left[\sum_{l=1}^k \frac{\left(\alpha^2\,(\mathbf{z}_{l-1}) - \hat{\alpha}^2\,(\mathbf{z}_{l-1})\right) X_{l-1}^2 + 2\,(\alpha\,(\mathbf{z}_{l-1}) - \hat{\alpha}\,(\mathbf{z}_{l-1}))\,(\beta\,(\mathbf{z}_{l-1})\,X_{l-1} - X_l X_{l-1})}{2\kappa^2\,(\mathbf{z}_{l-1})} \left| \mathscr{F}_k \right. \right] \\
&= E\left[\sum_{l=1}^k \sum_{i=1}^N \langle \mathbf{z}_{l-1},\, \mathbf{e}_i \rangle \frac{-\hat{\alpha}_i^2 X_{l-1}^2 - 2\hat{\alpha}_i\,(\beta_i X_{l-1} - X_l X_{l-1})}{2\kappa_i^2} + R \left| \mathscr{F}_k \right. \right] \\
&= \sum_{i=1}^N \frac{-\hat{\alpha}_i^2 \hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) - 2\hat{\alpha}_i\left(\beta_i \hat{\mathcal{T}}_k^i\,(X_{k-1}) - \hat{\mathcal{T}}_k^i\,(X_k X_{k-1})\right)}{2\kappa_i^2} + R,
\end{aligned}
$$

where R does not contain $\hat{\alpha}_i$.

We differentiate $LR\,(\hat{\alpha}_i)$ with respect to $\hat{\alpha}_i$ and then equate the result to zero. Consequently,

$$\hat{\alpha}_i = \frac{\hat{\mathcal{T}}_k^i\,(X_k \cdot X_{k-1}) - \beta_i \cdot \hat{\mathcal{T}}_k^i\,(X_{k-1})}{\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right)}.$$

∎

A.3 Optimal estimate for β

Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)^\top \in \mathbb{R}^N$. To calculate $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N\right)^T \in \mathbb{R}^N$, consider a new measure $P^{\hat{\beta}}$ defined by $\left.\dfrac{dP^{\hat{\beta}}}{dP^\beta}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*$, where

$$\lambda_l^* = \exp\left\{ \frac{(X_l - \alpha(\mathbf{z}_{l-1}) X_{l-1} - \beta(\mathbf{z}_{l-1}))^2 - \left(X_l - \alpha(\mathbf{z}_{l-1}) X_{l-1} - \hat{\beta}(\mathbf{z}_{l-1})\right)^2}{2\kappa^2(\mathbf{z}_{l-1})} \right\}.$$

This implies

$$
\begin{aligned}
LR\left(\hat{\beta}_i\right) &= E\left[\log \frac{dP^{\hat{\beta}}}{dP^\beta} \middle| \mathscr{F}_k \right] \\
&= E\left[\sum_{l=1}^k \frac{\beta^2(\mathbf{z}_{l-1}) - \hat{\beta}^2(\mathbf{z}_{l-1}) + 2(\alpha(\mathbf{z}_{l-1}) X_{l-1} - X_l)\left[\beta(\mathbf{z}_{l-1}) - \hat{\beta}(\mathbf{z}_{l-1})\right]}{2\kappa^2(\mathbf{z}_{l-1})} \middle| \mathscr{F}_k \right] \\
&= E\left[\sum_{l=1}^k \sum_{i=1}^N \langle \mathbf{z}_{l-1}, \mathbf{e}_i \rangle \frac{-\hat{\beta}_i^2 - 2(\alpha_i X_{l-1} - X_l)\hat{\beta}_i}{2\kappa_i^2} + R \middle| \mathscr{F}_k \right] \\
&= \sum_{i=1}^N \frac{2\hat{\beta}_i \hat{\mathcal{T}}_k^i(X_k) - 2\alpha_i \hat{\beta}_i \hat{\mathcal{T}}_k^i(X_{k-1}) - \hat{O}_k^i \hat{\beta}_i^2}{2\kappa_i^2} + R,
\end{aligned}
$$

where R does not contain $\hat{\beta}_i$.

By differentiating $LR\left(\hat{\beta}_i\right)$ with respect to $\hat{\beta}_i$ and setting the result to zero, we get

$$\hat{\beta}_i = \frac{\hat{\mathcal{T}}_k^i(X_l) - \alpha_i \hat{\mathcal{T}}_k^i(X_{l-1})}{\hat{O}_k^i}.$$

■

A.4 Optimal estimate for κ^2

Let $\boldsymbol{\kappa}^2 = \left(\kappa_1^2, \kappa_2^2, \ldots, \kappa_N^2\right)^\top \in \mathbb{R}^N$. To compute $\hat{\boldsymbol{\kappa}}^2 = \left(\hat{\kappa}_1^2, \hat{\kappa}_2^2, \ldots, \hat{\kappa}_N^2\right)^T \in \mathbb{R}^N$, consider a new measure $P^{\hat{\kappa}^2}$ defined by $\left.\dfrac{dP^{\hat{\kappa}^2}}{dP^{\kappa^2}}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*$, where

$$\lambda_l^* = \sqrt{\frac{\kappa^2(\mathbf{z}_{l-1})}{\hat{\kappa}^2(\mathbf{z}_{l-1})}} \exp\left\{ (X_l - \alpha(\mathbf{z}_{l-1}) X_{l-1} - \beta(\mathbf{z}_{l-1}))^2 \left(\frac{1}{2\kappa^2(\mathbf{z}_{l-1})} - \frac{1}{2\hat{\kappa}^2(\mathbf{z}_{l-1})} \right) \right\}.$$

This implies

$$
\begin{aligned}
LR\left(\hat{\kappa_i^2}\right) &= E\left[\log \frac{dP^{\hat{\kappa}^2}}{dP^{\kappa^2}}\middle| \mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^{k}\left[-\frac{1}{2}\log\left(\hat{\kappa}^2\left(\mathbf{z}_{l-1}\right)\right) - \frac{\left(X_l - \alpha\left(\mathbf{z}_{l-1}\right)X_{l-1} - \beta\left(\mathbf{z}_{l-1}\right)\right)^2}{2\hat{\kappa}^2\left(\mathbf{z}_{l-1}\right)}\right] + R\middle|\mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^{k}\sum_{i=1}^{N}\langle\mathbf{z}_{l-1}, \mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(\hat{\kappa_i^2}\right) - \frac{X_l^2 + \alpha_i^2 X_{l-1}^2 + \beta_i^2 - 2\alpha_i X_l X_{l-1} - 2\beta_i X_l + 2\alpha_i\beta_i X_{l-1}}{2\hat{\kappa_i^2}}\right) + R\middle|\mathscr{F}_k\right] \\
&= \sum_{i=1}^{N}\left(-\frac{1}{2}\hat{O}_k^i\log\left(\hat{\kappa_i^2}\right) - \frac{\hat{\mathcal{T}}_k^i\left(X_k^2\right) + \alpha_i^2\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) + \beta_i^2\hat{O}_k^i - 2\alpha_i\hat{\mathcal{T}}_k^i\left(X_k X_{k-1}\right) - 2\beta_i\hat{\mathcal{T}}_k^i\left(X_k\right) + 2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(X_{k-1}\right)}{2\hat{\kappa_i^2}}\right) \\
&\quad + R,
\end{aligned}
$$

where R does not contain $\hat{\kappa_i^2}$.

By differentiating $LR\left(\hat{\kappa_i^2}\right)$ with respect to $\hat{\kappa_i^2}$ and equating the resulting mathematical derivative to zero, we get

$$
\hat{\kappa_i^2} = \frac{\hat{\mathcal{T}}_k^i\left(X_k^2\right) + \alpha_i\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) + \beta_i^2\hat{O}_k^i - 2\alpha_i\hat{\mathcal{T}}_k^i\left(X_k X_{k-1}\right) - 2\beta_i\hat{\mathcal{T}}_k^i\left(X_k\right) + 2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(X_{k-1}\right)}{\hat{O}_k^i}.
$$

■

A.5 Optimal estimate for ζ

Let $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_N)^\top \in \mathbb{R}^N$. To get the estimate $\hat{\zeta} = \left(\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_N\right)^\top \in \mathbb{R}^N$, consider a new measure $P^{\hat{\zeta}}$ defined by $\left.\frac{dP^{\hat{\zeta}}}{dP^{\zeta}}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k}\lambda_l^*$, where

$$
\lambda_l^* = \exp\left\{\frac{\left(Y_l - Y_{l-1} - \zeta\left(\mathbf{z}_{l-1}\right)\right)^2 - \left(Y_l - Y_{l-1} - \hat{\zeta}\left(\mathbf{z}_{l-1}\right)\right)^2}{2\nu^2\left(\mathbf{z}_{l-1}\right)}\right\}.
$$

This tells us that

$$
\begin{aligned}
LR\left(\hat{\zeta}_i\right) &= E\left[\log\frac{dP^{\hat{\zeta}}}{dP^{\zeta}}\bigg|\mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^{k}\frac{\zeta^2\left(\mathbf{z}_{l-1}\right)-\hat{\zeta}^2\left(\mathbf{z}_{l-1}\right)+2\left(Y_{l-1}-Y_l\right)\left(\zeta\left(\mathbf{z}_{l-1}\right)-\hat{\zeta}\left(\mathbf{z}_{l-1}\right)\right)}{2v^2\left(\mathbf{z}_{l-1}\right)}\bigg|\mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^{k}\sum_{i=1}^{N}\langle\mathbf{z}_{l-1},\,\mathbf{e}_i\rangle\frac{-\left(\hat{\zeta}_i\right)^2-2\left(Y_{l-1}-Y_l\right)\left(\hat{\zeta}_i\right)}{2v_i^2}+R\bigg|\mathscr{F}_k\right] \\
&= \sum_{i=1}^{N}\left\{\frac{2\hat{\mathcal{T}}_k^i\left(Y_k\right)\hat{\zeta}_i-2\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)\hat{\zeta}_i-\hat{O}_k^i\hat{\zeta}_i^2}{2v_i^2}\right\}+R,
\end{aligned}
$$

where R does not contain $\hat{\zeta}_i$.

By differentiating $LR\left(\hat{\zeta}_i\right)$ with respect to $\hat{\zeta}_i$ and equating the result to zero, we get

$$
\hat{\zeta}_i = \frac{\hat{\mathcal{T}}_k^i\left(Y_l\right)-\hat{\mathcal{T}}_k^i\left(Y_{l-1}\right)}{\hat{O}_k^i}.
$$

∎

A.6   Optimal estimate for v^2

Write $v^2 := \left(v_1^2,\,v_2^2,\ldots,\,v_N^2\right)^\top\in\mathbb{R}^N$. We calculate $\hat{v}_i^{\,2} := \left(\hat{v_1^2},\,\hat{v_2^2},\ldots,\,\hat{v_N^2}\right)^\top\in\mathbb{R}^N$, by constructing a new measure $P^{\hat{v^2}}$ defined by $\dfrac{dP^{\hat{v^2}}}{dP^{v^2}}\bigg|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k}\lambda_l^*$, where

$$
\lambda_l^* = \sqrt{\frac{v^2\left(\mathbf{z}_{l-1}\right)}{\hat{v}^2\left(\mathbf{z}_{l-1}\right)}}\exp\left\{\left(Y_l-Y_{l-1}-\zeta\left(\mathbf{z}_{l-1}\right)\right)^2\left(\frac{1}{2v^2\left(\mathbf{z}_{l-1}\right)}-\frac{1}{2\hat{v}^2\left(\mathbf{z}_{l-1}\right)}\right)\right\}.
$$

So,

$$
\begin{aligned}
LR\left(\hat{v_i^2}\right) &= E\left[\log\frac{dP^{\hat{v^2}}}{dP^{v^2}}\middle|\mathscr{F}_k\right]\\
&= E\left[\sum_{l=1}^{k}\left[-\frac{1}{2}\log\left(\hat{v^2}\left(\mathbf{z}_{l-1}\right)\right) - \frac{\left(Y_l - Y_{l-1} - \zeta\left(\mathbf{z}_{l-1}\right)\right)^2}{2\hat{v^2}\left(\mathbf{z}_{l-1}\right)}\right]\middle|\mathscr{F}_k\right]\\
&= E\left[\sum_{l=1}^{k}\sum_{i=1}^{N}\langle\mathbf{z}_{l-1},\mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(\hat{v_i^2}\right) - \frac{Y_l^2 + Y_{l-1}^2 + \zeta_i^2 - 2Y_lY_{l-1} - 2\zeta_iY_l + 2\zeta_iY_{l-1}}{2\hat{v_i^2}}\right) + R\middle|\mathscr{F}_k\right]\\
&= \sum_{i=1}^{N}\left(-\frac{1}{2}\hat{O}_k^i\log\left(\hat{v_i^2}\right) - \frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right) + \hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right) + \zeta_i^2\hat{O}_k^i - 2\hat{\mathcal{T}}_k^i\left(Y_kY_{k-1}\right) - 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_k\right) + 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{2\hat{v_i^2}}\right)\\
&\quad + R,
\end{aligned}
$$

where R does not contain $\hat{v_i^2}$.

By differentiating $LR\left(\hat{v_i^2}\right)$ with respect to $\hat{v_i^2}$ and equate the result to zero, we get

$$
\hat{v_i^2} = \frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right) + \hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right) + \zeta_i^2\hat{O}_k^i - 2\hat{\mathcal{T}}_k^i\left(Y_kY_{k-1}\right) - 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_k\right) + 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{\hat{O}_k^i}.
$$

■

Appendix B

Procedure for filters' construction in Chapter 3

B.1 Change of measures

Under the real-world probability measure P, the true state of the underlying Markov chain $\mathbf{z}_k$ is neither observed nor measured directly since it is "latent" in the noisy observation process with "real world" dynamics given by Eqs. (3.10) and (3.13). Our objective is to "filter" the noise out of the observation process in the best possible way. Unfortunately, the derivation of filters under P is not straightforward.

Inspired by the approach described in [25], we perform a change of probability measure to introduce the the ideal-world measure $\bar{P}$ from the real-world probability measure P by invoking the discrete-time version of the Girsanov's theorem. Under this ideal measure, the observations are independent and identically distributed random variables which makes the calculations of conditional expectations manageable.

The filters, which are conditional expectations, are then related back to the real-world by the use of the Bayes' theorem for conditional expectation. Following [25] (chapter 3.4, page 62), the real-world measure P equivalent to an ideal measure $\bar{P}$ is constructed through the Radon-Nikodym derivative

$$\bar{\Lambda}_k = \left.\frac{dP}{d\bar{P}}\right|_{\mathscr{F}_k} = \prod_{l=1}^{k}\left[\left(\bar{\lambda}_l^Y(\mathbf{z}_{l-1})\right)\cdot\left(\prod_{g=1}^{3}\bar{\lambda}_l^{X^{(g)}}(\mathbf{z}_{l-1})\right)\right] \tag{B.1}$$

with

$$\bar{\lambda}_l^{X^{(g)}}(\mathbf{z}_{l-1}) = \left(\kappa^{(g)}(\mathbf{z}_{l-1})\phi(X_l^{(g)})\right)^{-1}\phi\left(\frac{X_l^{(g)} - \alpha^{(g)}(\mathbf{z}_{l-1})X_{l-1}^{(g)} - \beta^{(g)}(\mathbf{z}_{l-1})}{\kappa^{(g)}(\mathbf{z}_{l-1})}\right)$$

$$= \left(\kappa^{(g)}(\mathbf{z}_{l-1})\right)^{-1}\exp\left\{-\frac{1}{2}\left[\left(\frac{X_l^{(g)} - \alpha^{(g)}(\mathbf{z}_{l-1})X_{l-1}^{(g)} - \beta^{(g)}(\mathbf{z}_{l-1})}{\kappa^{(g)}(\mathbf{z}_{l-1})}\right)^2 - \left(X_l^{(g)}\right)^2\right]\right\}$$

$$\bar{\lambda}_l^Y(\mathbf{z}_{l-1}) = (\nu(\mathbf{z}_{l-1})\phi(Y_l))^{-1}\phi\left(\frac{Y_l - Y_{l-1} - \zeta(\mathbf{z}_{l-1})}{\nu(\mathbf{z}_{l-1})}\right)$$

$$= (\nu(\mathbf{z}_{l-1}))^{-1}\exp\left\{-\frac{1}{2}\left[\left(\frac{Y_l - Y_{l-1} - \zeta(\mathbf{z}_{l-1})}{\nu(\mathbf{z}_{l-1})}\right)^2 - (Y_l)^2\right]\right\},$$

where $\phi(\cdot)$ is the density function of a standard normal random variable.

B.2 Filtering

Let $\hat{\mathbf{z}}_k$ be the conditional expectation of $\mathbf{z}_k$ given $\mathscr{F}_k$ under probability measure P, i.e.,

$$\hat{\mathbf{z}}_k := E\left[\mathbf{z}_k\big|\mathscr{F}_k\right] = \left(\hat{z}_k^{(1)}, \hat{z}_k^{(2)}, \ldots, \hat{z}_k^{(N)}\right)^\top \in \mathbb{R}^N, \ \hat{z}_k^{(i)} = P\left(\mathbf{z}_k = \mathbf{e}_i\big|\mathscr{F}_k\right) = E\left[\langle\mathbf{z}_k, \mathbf{e}_i\rangle\big|\mathscr{F}_k\right], \quad \text{(B.2)}$$

where $\mathscr{F}_k$ is the filtration generated by the bivariate observation reflecting all information available up to time k. By the Bayes' theorem,

$$\hat{\mathbf{z}}_k = E\left[\mathbf{z}_k\big|\mathscr{F}_k\right] = \frac{\overline{E}\left[\bar{\Lambda}_k\mathbf{z}_k\big|\mathscr{F}_k\right]}{\overline{E}\left[\bar{\Lambda}_k\big|\mathscr{F}_k\right]}. \tag{B.3}$$

Write $\mathbf{p}_k := \overline{E}\left[\bar{\Lambda}_k\mathbf{z}_k\big|\mathscr{F}_k\right]$ so that

$$\overline{E}\left[\bar{\Lambda}_k\big|\mathscr{F}_k\right] = \overline{E}\left[\bar{\Lambda}_k\left(\sum_{i=1}^N\langle\mathbf{z}_k, \mathbf{e}_i\rangle\right)\Big|\mathscr{F}_k\right] = \sum_{i=1}^N\overline{E}\left[\langle\bar{\Lambda}_k\mathbf{z}_k, \mathbf{e}_i\rangle\big|\mathscr{F}_k\right]$$

$$= \sum_{i=1}^N\left\langle\overline{E}\left[\bar{\Lambda}_k\mathbf{z}_k\big|\mathscr{F}_k\right], \mathbf{e}_i\right\rangle = \sum_{i=1}^N\langle\mathbf{p}_k, \mathbf{e}_i\rangle. \tag{B.4}$$

For the second expression in (B.4) we rely on the fact that $\sum_{i=1}^N\langle\mathbf{z}_k, \mathbf{e}_i\rangle = 1$. Therefore, the conditional expectation of $\mathbf{z}_k$ has the form

$$\hat{\mathbf{z}}_k = \frac{\mathbf{p}_k}{\sum_{i=1}^N\langle\mathbf{p}_k, \mathbf{e}_i\rangle}. \tag{B.5}$$

Following similar principles in [59], define the diagonal matrix

$$
\mathbf{D}_k := \begin{pmatrix} d_{k,1} & & & O \\ & d_{k,2} & & \\ & & \ddots & \\ O & & & d_{k,N} \end{pmatrix},
\tag{B.6}
$$

with entries

$$
d_{k,j} = \left(\bar{\lambda}_{k,j}^{Y} \right) \left(\prod_{g=1}^{3} \bar{\lambda}_{k,j}^{X^{(g)}} \right);
\tag{B.7}
$$

where

$$
\bar{\lambda}_{k,j}^{X^{(g)}} = \left(\kappa_{k,j}^{(g)} \right)^{-1} \exp \left\{ -\frac{1}{2} \left[\left(\frac{X_k^{(g)} - \alpha_{k,j}^{(g)} X_{k-1}^{(g)} - \beta_{k,j}^{(g)}}{\kappa_{k,j}^{(g)}} \right)^2 - \left(X_k^{(g)} \right)^2 \right] \right\},
\tag{B.8}
$$

$$
\bar{\lambda}_{k,j}^{Y} = \left(\nu_{k,j} \right)^{-1} \exp \left\{ -\frac{1}{2} \left[\left(\frac{Y_k - Y_{k-1} - \zeta_{k,j}}{\nu_{k,j}} \right)^2 - (Y_k)^2 \right] \right\}.
\tag{B.9}
$$

Let H_k be any scalar $\mathscr{F}_k$-adapted process; H_0 is $\mathscr{F}_0$ measurable. A filter for H_k is defined as $E\left[H_k \middle| \mathscr{F}_k\right]$ and by the Bayes' theorem,

$$
E\left[H_k \middle| \mathscr{F}_k\right] = \frac{\overline{E}\left[H_k \bar{\Lambda}_k \middle| \mathscr{F}_k\right]}{\overline{E}\left[\bar{\Lambda}_k \middle| \mathscr{F}_k\right]} = \frac{\overline{E}\left[H_k \bar{\Lambda}_k \middle| \mathscr{F}_k\right]}{\sum_{i=1}^{N} \langle \mathbf{p}_k, \, \mathbf{e}_i \rangle}.
\tag{B.10}
$$

Write $\gamma(H_k) := \overline{E}\left[H_k \bar{\Lambda}_k \middle| \mathscr{F}_k\right]$, and so $\langle \overline{E}\left[H_k \mathbf{z}_k \bar{\Lambda}_k \middle| \mathscr{F}_k\right], \mathbf{1} \rangle = \langle \gamma(H_k \mathbf{z}_k), \mathbf{1} \rangle = \gamma(H_k \langle \mathbf{z}_k, \mathbf{1} \rangle) = \gamma(H_k)$, where $\mathbf{1}$ is a vector of 1's. Therefore, Eq. (E.7) becomes

$$
E\left[H_k \middle| \mathscr{F}_k\right] = \frac{\gamma(H_k)}{\sum_{i=1}^{N} \langle \mathbf{p}_k, \, \mathbf{e}_i \rangle} = \frac{\langle \gamma(H_k \mathbf{z}_k), \mathbf{1} \rangle}{\langle \mathbf{p}_k, \mathbf{1} \rangle}.
\tag{B.11}
$$

Thus, we obtain the filters for the state of the Markov chain, number of jumps $\mathcal{G}$, occupation time O, and auxiliary process $\mathcal{T}$.

$$\mathbf{p}_k = \mathbf{\Pi}\, \mathbf{D}_k\, \mathbf{p}_{k-1} \tag{B.12}$$

$$\gamma(\mathcal{G}_k^{s,j}\mathbf{z}_k) = \mathbf{\Pi}\mathbf{D}_k\, \gamma(\mathcal{G}_{k-1}^{s,j}\mathbf{z}_{k-1}) + d_{k,j}\langle \mathbf{p}_k,\, \mathbf{e}_j\rangle \pi_{sj}\mathbf{e}_s \tag{B.13}$$

$$\gamma(O_k^j\mathbf{z}_k) = \mathbf{\Pi}\mathbf{D}_k\, \gamma(O_{k-1}^j\mathbf{z}_{k-1}) + d_{k,j}\langle \mathbf{p}_k,\, \mathbf{e}_j\rangle \mathbf{\Pi}\mathbf{e}_j \tag{B.14}$$

$$\gamma(\mathcal{T}_k^j(f)\mathbf{z}_k) = \mathbf{\Pi}\mathbf{D}_k\, \gamma(\mathcal{T}_{k-1}^j(f)\,\mathbf{z}_{k-1}) + f(\cdot)\, d_{k,j}\langle \mathbf{p}_k,\, \mathbf{e}_j\rangle \mathbf{\Pi}\mathbf{e}_j \tag{B.15}$$

The derivations of Eqs. (B.12) – (B.15) are similar to those given in [59] or [25]. It is immediate, by Eq. (B.11), to determine the normalised filter estimates of $\gamma(\mathcal{G}_k^{s,j})$, $\gamma(O_k^j)$ and $\gamma(\mathcal{T}_k^j(f))$ by summing the components of the vector expressions given in Eqs. (B.13) – (B.15), and then dividing each by the expression in Eq. (B.12).

Appendix C

Derivation of optimal parameter estimates and Fisher information in Chapter 3

C.1 Optimal estimate for π_{ji}

The idea of the proof is similar to that in [59]. To perform the measure change mentioned in Subsection E.1, define a new measure $P^{\hat{\pi}_{ji}}$ via $\left.\dfrac{dP^{\hat{\pi}_{ji}}}{dP^{\pi_{ji}}}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k} \lambda_l^*$, where

$$\lambda_l^* = \sum_{j,i=1}^{N} \left(\frac{\hat{\pi}_{ji}}{\pi_{ji}}\right)^{\langle \mathbf{z}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{z}_l, \mathbf{e}_j \rangle}.$$

So,

$$\log \frac{dP^{\hat{\pi}_{ji}}}{dP^{\pi_{ji}}} = \sum_{l=1}^{k} \sum_{j,i=1}^{N} \left(\log\left(\hat{\pi}_{ji}\right) - \log\left(\pi_{ji}\right)\right) \langle \mathbf{z}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{z}_l, \mathbf{e}_j \rangle = \sum_{j,i=1}^{N} \mathcal{G}_k^{j,i} \log\left(\hat{\pi}_{ji}\right) + R,$$

where R does not contain $\hat{\alpha}$.

It must be noted that $\sum_{j=1}^{N} \mathcal{G}_k^{j,i} = O_k^i$, $\sum_{j=1}^{N} \hat{\mathcal{G}}_k^{j,i} = \hat{O}_k^i$. The optimal estimate for π_{ji} is the value that maximises the log-likelihood $\log \dfrac{dP^{\hat{\pi}_{ji}}}{dP^{\pi_{ji}}}$ constrained to $\sum_{j=1}^{N} \hat{\pi}_{ji} = 1$. Introducing the Lagrange

multiplier and solving the associated optimisation problem yield the solution

$$\hat{\pi}_{ji} = \frac{\hat{\mathcal{G}}_k^{j,i}}{\hat{\mathcal{O}}_k^i}.$$

∎

C.2 Fisher information for π_{ji}

We write the log-likelihood of π_{ji} as

$$\mathcal{L}\left(\pi_{ji}\right) = \sum_{l=1}^{k} \left(\log\left(\pi_{ji}\right)\right) \langle \mathbf{z}_{l-1},\, \mathbf{e}_i \rangle \langle \mathbf{z}_l,\, \mathbf{e}_j \rangle$$

Thus, the Fisher information of π_{ji} is

$$\mathcal{I}\left(\pi_{ji}\right) = -E\left[\frac{d^2}{d\pi_{ji}^2}\mathcal{L}\left(\pi_{ji}\right)\middle|\pi_{ji}\right] = \frac{\hat{\mathcal{G}}_k^{j,i}}{\pi_{ji}^2}$$

∎

C.3 Optimal estimate for α

Let $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)^\top \in \mathbb{R}^N$. To calculate $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N)^\top \in \mathbb{R}^N$, consider a new measure $P^{\hat{\alpha}}$ defined by $\left.\frac{dP^{\hat{\alpha}}}{dP^\alpha}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*$, where

$$\lambda_l^* = \exp\left\{\frac{(X_l - \alpha(\mathbf{z}_{l-1})X_{l-1} - \beta(\mathbf{z}_{l-1}))^2 - (X_l - \hat{\alpha}(\mathbf{z}_{l-1})X_{l-1} - \beta(\mathbf{z}_{l-1}))^2}{2\kappa^2(\mathbf{z}_{l-1})}\right\}.$$

This implies

$$
\begin{aligned}
LR\left(\hat{\alpha}_i\right) &= E\left[\log \frac{dP^{\hat{\alpha}}}{dP^{\alpha}}\middle|\mathscr{F}_k\right]\\
&= E\left[\sum_{l=1}^{k} \frac{\left(\alpha^2\left(\mathbf{z}_{l-1}\right) - \hat{\alpha}^2\left(\mathbf{z}_{l-1}\right)\right)X_{l-1}^2 + 2\left(\alpha\left(\mathbf{z}_{l-1}\right) - \hat{\alpha}\left(\mathbf{z}_{l-1}\right)\right)\left(\beta\left(\mathbf{z}_{l-1}\right)X_{l-1} - X_l X_{l-1}\right)}{2\kappa^2\left(\mathbf{z}_{l-1}\right)}\middle|\mathscr{F}_k\right]\\
&= E\left[\sum_{l=1}^{k}\sum_{i=1}^{N}\langle\mathbf{z}_{l-1}, \mathbf{e}_i\rangle \frac{-\hat{\alpha}_i^2 X_{l-1}^2 - 2\hat{\alpha}_i\left(\beta_i X_{l-1} - X_l X_{l-1}\right)}{2\kappa_i^2} + R\middle|\mathscr{F}_k\right]\\
&= \sum_{i=1}^{N} \frac{-\hat{\alpha}_i^2 \hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) - 2\hat{\alpha}_i\left(\beta_i \hat{\mathcal{T}}_k^i\left(X_{k-1}\right) - \hat{\mathcal{T}}_k^i\left(X_k X_{k-1}\right)\right)}{2\kappa_i^2} + R,
\end{aligned}
$$

where R does not contain $\hat{\alpha}_i$.

We differentiate $LR\left(\hat{\alpha}_i\right)$ with respect to $\hat{\alpha}_i$ and then equate the result to zero. Consequently,

$$
\hat{\alpha}_i = \frac{\hat{\mathcal{T}}_k^i\left(X_k \cdot X_{k-1}\right) - \beta_i \cdot \hat{\mathcal{T}}_k^i\left(X_{k-1}\right)}{\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right)}.
$$

∎

C.4   Fisher information for α_i

We write the log-likelihood of α_i as

$$
\mathcal{L}\left(\alpha_i\right) = \sum_{l=1}^{k}\left[\langle\mathbf{z}_{l-1}, \mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(2\pi\right) - \log\kappa_i - \frac{\left(X_l - \alpha_i X_{l-1} - \beta_i\right)^2}{2\kappa_i^2}\right)\right].
$$

Hence, the Fisher information of α_i is

$$
\mathcal{I}\left(\alpha_i\right) = -E\left[\frac{d^2}{d\alpha_i^2}\mathcal{L}\left(\alpha_i\right)\middle|\alpha_i\right] = E\left[\sum_{l=1}^{k}\langle\mathbf{z}_{l-1}, \mathbf{e}_i\rangle\left(\frac{X_{l-1}^2}{\kappa_i^2}\right)\middle|\alpha_i\right] = \frac{\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right)}{\kappa_i^2}.
$$

∎

C.5   Optimal estimate for β

Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)^\top \in \mathbb{R}^N$. To calculate $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N\right)^T \in \mathbb{R}^N$, consider a new measure $P^{\hat{\beta}}$ defined by $\left.\dfrac{dP^{\hat{\beta}}}{dP^\beta}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*$, where

$$\lambda_l^* = \exp\left\{\frac{(X_l - \alpha(\mathbf{z}_{l-1})X_{l-1} - \beta(\mathbf{z}_{l-1}))^2 - \left(X_l - \alpha(\mathbf{z}_{l-1})X_{l-1} - \hat{\beta}(\mathbf{z}_{l-1})\right)^2}{2\kappa^2(\mathbf{z}_{l-1})}\right\}.$$

This implies

$$
\begin{aligned}
LR\left(\hat{\beta}_i\right) &= E\left[\log \frac{dP^{\hat{\beta}}}{dP^\beta}\middle|\mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^k \frac{\beta^2(\mathbf{z}_{l-1}) - \hat{\beta}^2(\mathbf{z}_{l-1}) + 2(\alpha(\mathbf{z}_{l-1})X_{l-1} - X_l)\left[\beta(\mathbf{z}_{l-1}) - \hat{\beta}(\mathbf{z}_{l-1})\right]}{2\kappa^2(\mathbf{z}_{l-1})}\middle|\mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^k \sum_{i=1}^N \langle \mathbf{z}_{l-1}, \mathbf{e}_i\rangle \frac{-\hat{\beta}_i^2 - 2(\alpha_i X_{l-1} - X_l)\hat{\beta}_i}{2\kappa_i^2} + R\middle|\mathscr{F}_k\right] \\
&= \sum_{i=1}^N \frac{2\hat{\beta}_i \hat{\mathcal{T}}_k^i(X_k) - 2\alpha_i \hat{\beta}_i \hat{\mathcal{T}}_k^i(X_{k-1}) - \hat{O}_k^i \hat{\beta}_i^2}{2\kappa_i^2} + R,
\end{aligned}
$$

where R does not contain $\hat{\beta}_i$.

By differentiating $LR\left(\hat{\beta}_i\right)$ with respect to $\hat{\beta}_i$ and setting the result to zero, we get

$$\hat{\beta}_i = \frac{\hat{\mathcal{T}}_k^i(X_l) - \alpha_i \hat{\mathcal{T}}_k^i(X_{l-1})}{\hat{O}_k^i}.$$

$\blacksquare$

C.6   Fisher information for β_i

We write the log-likelihood of β_i as

$$\mathcal{L}(\beta_i) = \sum_{l=1}^k \left[\langle \mathbf{z}_{l-1}, \mathbf{e}_i\rangle\left(-\frac{1}{2}\log(2\pi) - \log \kappa_i - \frac{(X_l - \alpha_i X_{l-1} - \beta_i)^2}{2\kappa_i^2}\right)\right].$$

Ergo, the Fisher information of β_i is

$$\mathcal{I}\left(\beta_i\right) = -E\left[\frac{d^2}{d\beta_i^2}\mathcal{L}\left(\mathcal{F}_i\right)\Big|\beta_i\right] = E\left[\sum_{l=1}^{k}\frac{\langle \mathbf{z}_{l-1},\, \mathbf{e}_i\rangle}{\kappa_i^2}\Big|\beta_i\right] = \frac{\hat{O}_k^i}{\kappa_i^2}.$$

∎

C.7 Optimal estimate for κ^2

Let $\boldsymbol{\kappa}^2 = \left(\kappa_1^2,\, \kappa_2^2,\, \ldots,\, \kappa_N^2\right)^{\top} \in \mathbb{R}^N$. To compute $\hat{\boldsymbol{\kappa}}^2 = \left(\hat{\kappa_1^2},\, \hat{\kappa_2^2},\, \ldots,\, \hat{\kappa_N^2}\right)^{T} \in \mathbb{R}^N$, consider a new measure $P^{\hat{\kappa}^2}$ defined by $\dfrac{dP^{\hat{\kappa}^2}}{dP^{\kappa^2}}\Big|_{\mathcal{F}_k} = \Lambda_k^* = \prod_{l=1}^{k}\lambda_l^*$, where

$$\lambda_l^* = \sqrt{\frac{\kappa^2\left(\mathbf{z}_{l-1}\right)}{\hat{\kappa}^2\left(\mathbf{z}_{l-1}\right)}}\,\exp\left\{\left(X_l - \alpha\left(\mathbf{z}_{l-1}\right)X_{l-1} - \beta\left(\mathbf{z}_{l-1}\right)\right)^2\left(\frac{1}{2\kappa^2\left(\mathbf{z}_{l-1}\right)} - \frac{1}{2\hat{\kappa}^2\left(\mathbf{z}_{l-1}\right)}\right)\right\}.$$

This implies

$$
\begin{aligned}
LR\left(\hat{\kappa_i^2}\right) &= E\left[\log\frac{dP^{\hat{\kappa}^2}}{dP^{\kappa^2}}\Big|\mathcal{F}_k\right] \\
&= E\left[\sum_{l=1}^{k}\left[-\frac{1}{2}\log\left(\hat{\kappa}^2\left(\mathbf{z}_{l-1}\right)\right) - \frac{\left(X_l - \alpha\left(\mathbf{z}_{l-1}\right)X_{l-1} - \beta\left(\mathbf{z}_{l-1}\right)\right)^2}{2\hat{\kappa}^2\left(\mathbf{z}_{l-1}\right)}\right] + R\Big|\mathcal{F}_k\right] \\
&= E\left[\sum_{l=1}^{k}\sum_{i=1}^{N}\langle \mathbf{z}_{l-1},\, \mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(\hat{\kappa_i^2}\right) - \frac{X_l^2 + \alpha_i^2 X_{l-1}^2 + \beta_i^2 - 2\alpha_i X_l X_{l-1} - 2\beta_i X_l + 2\alpha_i\beta_i X_{l-1}}{2\hat{\kappa_i^2}}\right) + R\Big|\mathcal{F}_k\right] \\
&= \sum_{i=1}^{N}\left(-\frac{1}{2}\hat{O}_k^i\log\left(\hat{\kappa_i^2}\right) - \frac{\hat{\mathcal{T}}_k^i\left(X_k^2\right) + \alpha_i^2\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) + \beta_i^2\hat{O}_k^i - 2\alpha_i\hat{\mathcal{T}}_k^i\left(X_k X_{k-1}\right) - 2\beta_i\hat{\mathcal{T}}_k^i\left(X_k\right) + 2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(X_{k-1}\right)}{2\hat{\kappa_i^2}}\right) \\
&\quad + R,
\end{aligned}
$$

where R does not contain $\hat{\kappa_i^2}$.

By differentiating $LR\left(\hat{\kappa_i^2}\right)$ with respect to $\hat{\kappa_i^2}$ and equating the resulting mathematical derivative to

zero, we get

$$\hat{\kappa_i^2} = \frac{\hat{\mathcal{T}}_k^i\left(X_k^2\right) + \alpha_i\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) + \beta_i^2\hat{O}_k^i - 2\alpha_i\hat{\mathcal{T}}_k^i\left(X_kX_{k-1}\right) - 2\beta_i\hat{\mathcal{T}}_k^i\left(X_k\right) + 2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(X_{k-1}\right)}{\hat{O}_k^i}.$$

$\blacksquare$

C.8 Fisher information for κ_i^2

We write the log-likelihood of κ_i^2 as

$$\mathcal{L}\left(\kappa_i^2\right) = \sum_{l=1}^{k}\left[\langle \mathbf{z}_{l-1},\ \mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(2\pi\right) - \frac{1}{2}\log\kappa_i^2 - \frac{(X_l - \alpha_iX_{l-1} - \beta_i)^2}{2\kappa_i^2}\right)\right].$$

Consequently, the Fisher information of κ_i^2 is

$$\begin{aligned}
\mathcal{I}\left(\kappa_i^2\right) &= -E\left[\frac{d^2}{d\kappa_i^{2^2}}\mathcal{L}\left(\kappa_i^2\right)\Big|\kappa_i^2\right] \\
&= E\left[\sum_{l=1}^{k}\langle \mathbf{z}_{l-1},\ \mathbf{e}_i\rangle\left(-\frac{1}{2\kappa_i^4} + \frac{X_l^2 + \alpha_i^2X_{l-1}^2 + \beta_i^2 - 2\alpha_iX_lX_{l-1} - 2\beta_iX_l + 2\alpha_i\beta_iX_{l-1}}{\kappa_i^6}\right)\Big|\kappa_i^2\right] \\
&= \frac{\hat{\mathcal{T}}_k^i\left(X_k^2\right) + \alpha_i^2\hat{\mathcal{T}}_k^i\left(X_{k-1}^2\right) + \hat{O}_k^i\beta_i^2 - 2\alpha_i\hat{\mathcal{T}}_k^i\left(X_kX_{k-1}\right) - 2\beta_i\hat{\mathcal{T}}_k^i\left(X_k\right) + 2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(X_{k-1}\right)}{\kappa_i^6} - \frac{\hat{O}_k^i}{2\kappa_i^4}.
\end{aligned}$$

$\blacksquare$

C.9 Optimal estimate for ζ

Let $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_N)^\top \in \mathbb{R}^N$. To get the estimate $\hat{\zeta} = \left(\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_N\right)^\top \in \mathbb{R}^N$, consider a new measure $P^{\hat{\zeta}}$ defined by $\dfrac{dP^{\hat{\zeta}}}{dP^\zeta}\Big|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k}\lambda_l^*$, where

$$\lambda_l^* = \exp\left\{\frac{(Y_l - Y_{l-1} - \zeta(\mathbf{z}_{l-1}))^2 - \left(Y_l - Y_{l-1} - \hat{\zeta}(\mathbf{z}_{l-1})\right)^2}{2\nu^2(\mathbf{z}_{l-1})}\right\}.$$

This tells us that

$$
\begin{aligned}
LR\left(\hat{\zeta_i}\right) &= E\left[\log \frac{dP^{\hat{\zeta}}}{dP^{\zeta}} \middle| \mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^{k} \frac{\zeta^2\left(\mathbf{z}_{l-1}\right) - \hat{\zeta}^2\left(\mathbf{z}_{l-1}\right) + 2\left(Y_{l-1} - Y_l\right)\left(\zeta\left(\mathbf{z}_{l-1}\right) - \hat{\zeta}\left(\mathbf{z}_{l-1}\right)\right)}{2v^2\left(\mathbf{z}_{l-1}\right)} \middle| \mathscr{F}_k\right] \\
&= E\left[\sum_{l=1}^{k}\sum_{i=1}^{N}\langle \mathbf{z}_{l-1},\, \mathbf{e}_i\rangle \frac{-\left(\hat{\zeta}_i\right)^2 - 2\left(Y_{l-1} - Y_l\right)\left(\hat{\zeta}_i\right)}{2v_i^2} + R \middle| \mathscr{F}_k\right] \\
&= \sum_{i=1}^{N}\left\{\frac{2\hat{\mathcal{T}}_k^i\left(Y_k\right)\hat{\zeta}_i - 2\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)\hat{\zeta}_i - \hat{O}_k^i\hat{\zeta}_i^2}{2v_i^2}\right\} + R,
\end{aligned}
$$

where R does not contain $\hat{\zeta}_i$.

Taking the derivative of $LR\left(\hat{\zeta}_i\right)$ with respect to $\hat{\zeta}_i$ and equating the result to zero, we have

$$
\hat{\zeta}_i = \frac{\hat{\mathcal{T}}_k^i\left(Y_l\right) - \hat{\mathcal{T}}_k^i\left(Y_{l-1}\right)}{\hat{O}_k^i}.
$$

■

C.10 Fisher information for ζ_i

Consider the log-likelihood of ζ_i given by

$$
\mathcal{L}\left(\zeta_i\right) = \sum_{l=1}^{k}\left[\langle \mathbf{z}_{l-1},\, \mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(2\pi\right) - \log v_i - \frac{\left(Y_l - Y_{l-1} - \zeta_i\right)^2}{2v_i^2}\right)\right].
$$

This tells us that the Fisher information of ζ_i is

$$
\mathcal{I}\left(\zeta_i\right) = -E\left[\frac{d^2}{d\zeta_i^2}\mathcal{L}\left(\zeta_i\right) \middle| \zeta_i\right] = E\left[\sum_{l=1}^{k}\frac{\langle \mathbf{z}_{l-1},\, \mathbf{e}_i\rangle}{v_i^2} \middle| \zeta_i\right] = \frac{\hat{O}_k^i}{v_i^2}.
$$

■

C.11 Optimal estimate for v^2

Write $v^2 := \left(v_1^2, v_2^2, \ldots, v_N^2\right)^\top \in \mathbb{R}^N$. We calculate $\hat{v_i}^2 := \left(\hat{v_1^2}, \hat{v_2^2}, \ldots, \hat{v_N^2}\right)^\top \in \mathbb{R}^N$, by constructing a new measure $P^{\hat{v}^2}$ defined by $\left.\dfrac{dP^{\hat{v}^2}}{dP^{v^2}}\right|_{\mathcal{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*$, where

$$\lambda_l^* = \sqrt{\frac{v^2\left(\mathbf{z}_{l-1}\right)}{\hat{v}^2\left(\mathbf{z}_{l-1}\right)}} \exp\left\{\left(Y_l - Y_{l-1} - \zeta\left(\mathbf{z}_{l-1}\right)\right)^2\left(\frac{1}{2v^2\left(\mathbf{z}_{l-1}\right)} - \frac{1}{2\hat{v}^2\left(\mathbf{z}_{l-1}\right)}\right)\right\}.$$

So,

$$
\begin{aligned}
LR\left(\hat{v_i}^2\right) &= E\left[\log \left.\frac{dP^{\hat{v}^2}}{dP^{v^2}}\right|\mathcal{F}_k\right] \\
&= E\left[\sum_{l=1}^k \left[-\frac{1}{2}\log\left(\hat{v}^2\left(\mathbf{z}_{l-1}\right)\right) - \frac{\left(Y_l - Y_{l-1} - \zeta\left(\mathbf{z}_{l-1}\right)\right)^2}{2\hat{v}^2\left(\mathbf{z}_{l-1}\right)}\right] \middle|\mathcal{F}_k\right] \\
&= E\left[\sum_{l=1}^k \sum_{i=1}^N \langle \mathbf{z}_{l-1}, \mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(\hat{v_i^2}\right) - \frac{Y_l^2 + Y_{l-1}^2 + \zeta_i^2 - 2Y_lY_{l-1} - 2\zeta_iY_l + 2\zeta_iY_{l-1}}{2\hat{v_i^2}}\right) + R \middle|\mathcal{F}_k\right] \\
&= \sum_{i=1}^N \left(-\frac{1}{2}\hat{O}_k^i\log\left(\hat{v_i^2}\right) - \frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right) + \hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right) + \zeta_i^2\hat{O}_k^i - 2\hat{\mathcal{T}}_k^i\left(Y_kY_{k-1}\right) - 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_k\right) + 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{2\hat{v_i^2}}\right) \\
&\quad + R,
\end{aligned}
$$

where R does not contain $\hat{v_i}^2$.

By differentiating $LR\left(\hat{v_i}^2\right)$ with respect to $\hat{v_i}^2$ and equate the result to zero, we get

$$\hat{v_i^2} = \frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right) + \hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right) + \zeta_i^2\hat{O}_k^i - 2\hat{\mathcal{T}}_k^i\left(Y_kY_{k-1}\right) - 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_k\right) + 2\zeta_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{\hat{O}_k^i}.$$

$\blacksquare$

C.12 Fisher information for v_i^2

The representation of the log-likelihood of v_i^2 is

$$\mathcal{L}\left(v_i^2\right) = \sum_{l=1}^{k} \left[\langle \mathbf{z}_{l-1}, \, \mathbf{e}_i \rangle \left(-\frac{1}{2} \log\left(2\pi\right) - \log v_i - \frac{\left(Y_l - Y_{l-1} - \zeta_i\right)^2}{2v_i^2} \right) \right].$$

This implies that the Fisher information of v_i^2 is

$$\mathcal{I}\left(v_i^2\right) = -E\left[\frac{d^2}{dv_i^2} \mathcal{L}\left(v_i^2\right) \middle| v_i^2 \right]$$

$$= E\left[\sum_{l=1}^{k} \langle \mathbf{z}_{l-1}, \, \mathbf{e}_i \rangle \left(-\frac{1}{2v_i^4} + \frac{Y_l^2 + Y_{l-1}^2 + \zeta_i^2 - 2Y_l Y_{l-1} - 2\zeta_i Y_l + 2\zeta_i Y_{l-1}}{v_i^6} \right) \middle| v_i^2 \right]$$

$$= \frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right) + \hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right) + \hat{O}_k^i \zeta_i^2 - 2\hat{\mathcal{T}}_k^i\left(Y_k Y_{k-1}\right) - 2\zeta_i \hat{\mathcal{T}}_k^i\left(Y_k\right) + 2\zeta_i \hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{v_i^6} - \frac{\hat{O}_k^i}{2v_i^4}$$

∎

Appendix D

Additional details and results for HMM filtering in Chapter 3

D.1 Initial values for HMM-based models

Table D.1: Initial parameter values for the 1-, 2-, and 3-state HMMs

Index	Parameters	1-State HMM	2-State HMM		3-State HMM		
			State 1	State 2	State 1	State 3	State 2
log(BASP500)	η_{ini}	1.6×10^{-3}	1.9×10^{-3}	1.3×10^{-3}	1.8×10^{-3}	1.5×10^{-3}	1.3×10^{-3}
	ξ_{ini}^2	0.034	0.011	0.022	0.006	0.013	0.014
TED	μ_{ini}	0.814	0.992	0.636	0.952	0.745	0.676
	θ_{ini}	2.097	2.104	1.922	2.104	2.019	1.937
	σ_{ini}^2	0.857	0.290	0.575	0.165	0.346	0.354
log(VIX)	μ_{ini}	3.329	3.421	3.238	3.403	3.302	3.256
	θ_{ini}	1.522	1.583	1.693	1.545	1.732	1.734
	σ_{ini}^2	1.092	0.371	0.736	0.211	0.443	0.453
log(DXY)	μ_{ini}	4.581	4.672	4.489	4.655	4.531	4.506
	θ_{ini}	1.227	1.597	1.377	1.679	1.231	1.157
	σ_{ini}^2	0.006	0.002	0.004	0.001	0.002	0.003

D.2   1-state HMM filtering outcomes
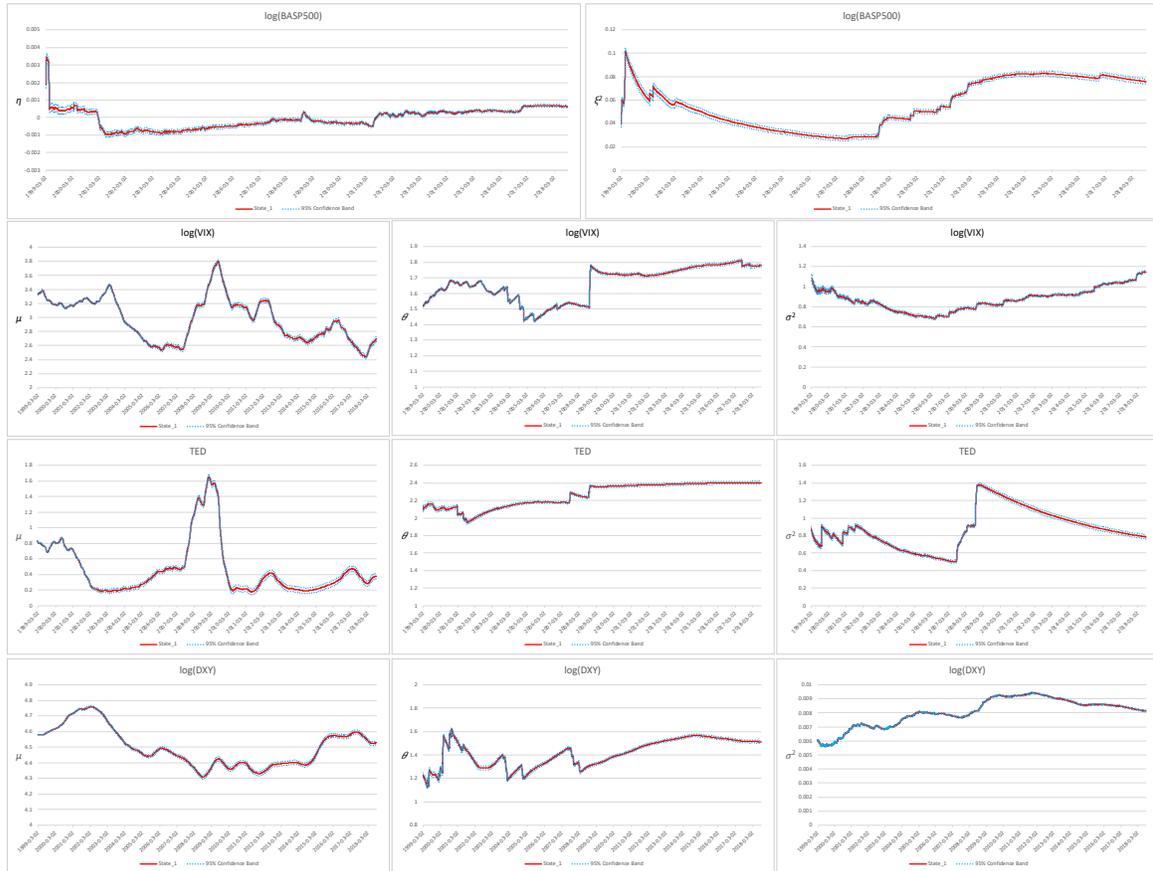
Figure D.1: Evolution of parameter estimates for μ, θ, σ^2, η and ξ^2 under a 1-state HMM

D.3 3-state HMM filtering outcomes
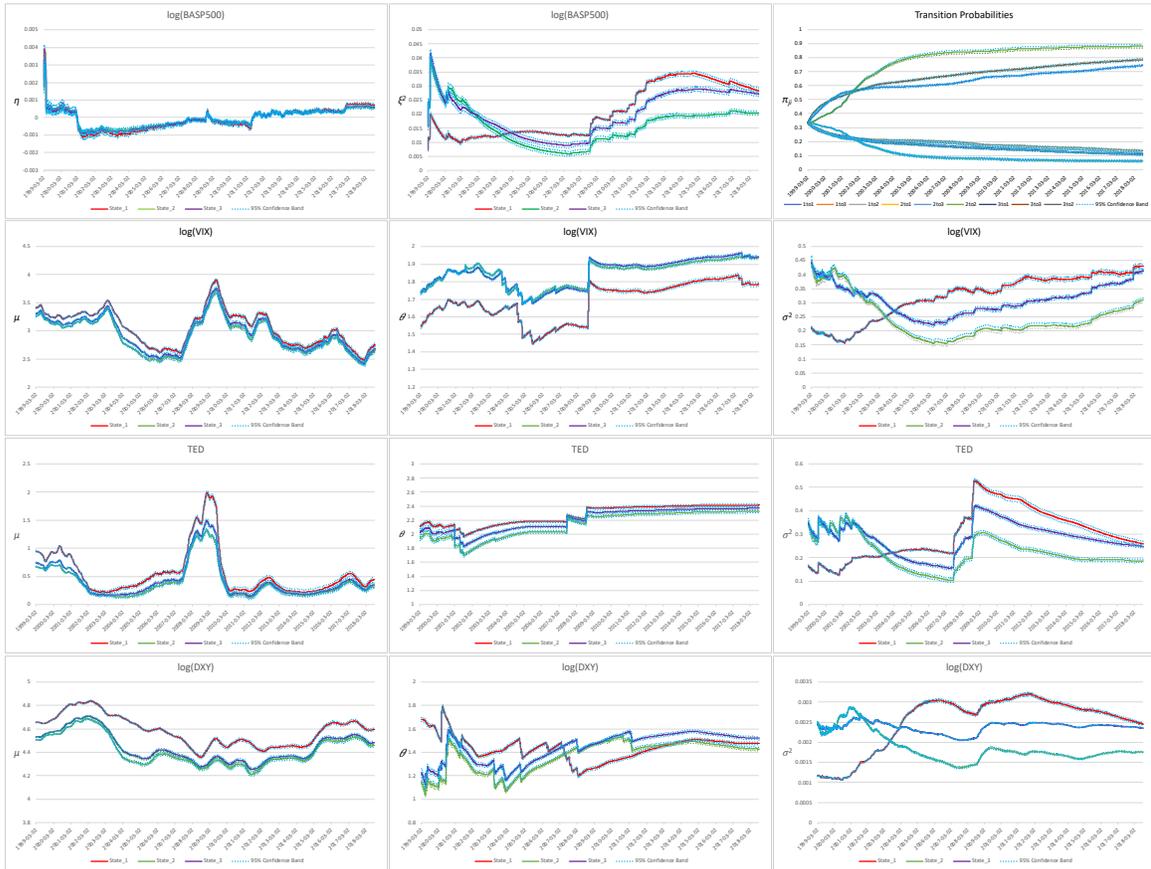
Figure D.2: Evolution of parameter estimates for μ, θ, σ^2, η, ξ^2 and π_{ji} under a 3-state HMM

Appendix E

Procedure for filters' construction in Chapter 4

E.1   Change of measures

Under the probability measure P, which describes our stochastic model in the real-world, the true state of the Markov chain $\mathbf{x}_k^{(m)}$ is "hidden" in the noisy observation process given by equation (4.10). Therefore, the true state could neither be observed nor be measured directly. Instead, we could construct filters that extract the latent information out of the observation process. Nonetheless, filtering is not straightforward under the probability measure P due to the evaluation of expectation of products of random variables that are not independent.

Adopting a change-of-probability-measure method [25], an ideal reference measure $\overline{P}$ equivalent to the real-world probability measure P is constructed. This is made possible by the discrete-time version of the Girsanov's theorem. Under $\overline{P}$, the observations are independent and identically distributed random variables facilitating the computations of conditional expectations involved in the filters. Applying the Bayes' theorem for conditional expectation, the filters are then related back to the real world. Alternatively, we could back out P from $\overline{P}$ through the Radon-Nikodym derivative

$$\bar{\Lambda}_k^{(m)} = \left.\frac{\mathrm{d}P^{(m)}}{\mathrm{d}\overline{P}^{(m)}}\right|_{\mathscr{F}_k^{(m)}} = \prod_{l=1}^{k} \bar{\lambda}_l^{(m)}(\mathbf{x}_{l-1}), \tag{E.1}$$

where

$$\bar{\lambda}_l^{(m)}(\mathbf{x}_{l-1}^{(m)}) = \frac{\phi(v_l^{(m)})}{\zeta(\mathbf{x}_{l-1}^{(m)})\phi(Y_l^{(m)})} = \frac{\exp\left\{-\frac{1}{2}\left[(v_l^{(m)})^2 - (Y_l^{(m)})^2\right]\right\}}{\zeta(\mathbf{x}_{l-1}^{(m)})} \tag{E.2}$$

and

$$\nu_l^{(m)} = \frac{Y_l^{(m)} - \alpha(\mathbf{x}_{l-1}^{(m)})Y_{l-1}^{(m)} - \beta(\mathbf{x}_{l-1}^{(m)})}{\zeta(\mathbf{x}_{l-1}^{(m)})}. \tag{E.3}$$

In Eq. (E.2), $\phi(\cdot)$ is the density function of a standard normal random variable.

E.2 Filtering

By the Bayes' theorem,

$$\hat{\mathbf{x}}_k^{(m)} = E\left[\mathbf{x}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right] = \frac{\overline{E}\left[\bar{\Lambda}_k^{(m)}\mathbf{x}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]}{\overline{E}\left[\bar{\Lambda}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]}. \tag{E.4}$$

Write $\mathbf{u}_k^{(m)} := \overline{E}\left[\bar{\Lambda}_k^{(m)}\mathbf{x}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]$ so that

$$\overline{E}\left[\bar{\Lambda}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right] = \overline{E}\left[\bar{\Lambda}_k^{(m)}\left(\sum_{i=1}^N \langle\mathbf{x}_k^{(m)}, \mathbf{e}_i\rangle\right)\big|\mathscr{F}_k^{(m)}\right] = \sum_{i=1}^N \overline{E}\left[\langle\bar{\Lambda}_k^{(m)}\mathbf{x}_k^{(m)}, \mathbf{e}_i\rangle\big|\mathscr{F}_k^{(m)}\right]$$

$$= \sum_{i=1}^N \langle\overline{E}\left[\bar{\Lambda}_k^{(m)}\mathbf{x}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right], \mathbf{e}_i\rangle = \sum_{i=1}^N \langle\mathbf{u}_k^{(m)}, \mathbf{e}_i\rangle. \tag{E.5}$$

The second equation in Eq. (E.5) comes from the fact that $\sum_{i=1}^N \langle\mathbf{x}_k^{(m)}, \mathbf{e}_i\rangle = 1$. Thus, the estimate of $\mathbf{x}_k^{(m)}$ is

$$\hat{\mathbf{x}}_k^{(m)} = \frac{\mathbf{u}_k^{(m)}}{\sum_{i=1}^N \langle\mathbf{u}_k^{(m)}, \mathbf{e}_i\rangle}. \tag{E.6}$$

Let $\xi_k^{(m)}$ be any scalar and $\mathscr{F}_k^{(m)}$-adapted process; ξ_0 is $\mathscr{F}_0^{(m)}$ measurable. A filter for $\xi_k^{(m)}$ is defined as the $E\left[\xi_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]$. So, from the Bayes' theorem,

$$E\left[\xi_k^{(m)}\big|\mathscr{F}_k^{(m)}\right] = \frac{\overline{E}\left[\xi_k^{(m)}\bar{\Lambda}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]}{\overline{E}\left[\bar{\Lambda}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]} = \frac{\overline{E}\left[\xi_k^{(m)}\bar{\Lambda}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]}{\sum_{i=1}^N \langle\mathbf{u}_k^{(m)}, \mathbf{e}_i\rangle}. \tag{E.7}$$

Write

$$\gamma(\xi_k^{(m)}) := \overline{E}\left[\xi_k^{(m)}\bar{\Lambda}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right]. \tag{E.8}$$

So, we get

$$\left\langle \overline{E}\left[\xi_k^{(m)}\mathbf{x}_k^{(m)}\bar{\Lambda}_k^{(m)}\big|\mathscr{F}_k^{(m)}\right], \mathbf{1}\right\rangle = \langle\gamma(\xi_k^{(m)}\mathbf{x}_k^{(m)}), \mathbf{1}\rangle = \gamma(\xi_k^{(m)}\langle\mathbf{x}_k^{(m)}, \mathbf{1}\rangle) = \gamma(\xi_k^{(m)}), \tag{E.9}$$

where $\mathbf{1}$ is a vector of 1's. Therefore, Eq. (E.7) becomes

$$E\left[\xi_k^{(m)}\big|\mathscr{F}_k^{(m)}\right] = \frac{\gamma(\xi_k^{(m)})}{\sum_{i=1}^N \langle\mathbf{u}_k^{(m)}, \mathbf{e}_i\rangle} = \frac{\langle\gamma(\xi_k^{(m)}\mathbf{x}_k^{(m)}), \mathbf{1}\rangle}{\langle\mathbf{u}_k^{(m)}, \mathbf{1}\rangle}. \tag{E.10}$$

Following similar formulation in [59], we define the diagonal matrix

$$\mathbf{C}_k^{(m)} := \begin{pmatrix} \bar{\lambda}_{k,1}^{(m)} & & & O \\ & \bar{\lambda}_{k,2}^{(m)} & & \\ & & \ddots & \\ O & & & \bar{\lambda}_{k,N}^{(m)} \end{pmatrix}, \tag{E.11}$$

with diagonal entries

$$\bar{\lambda}_{k,i}^{(m)} = \frac{\exp\left\{-\frac{1}{2}\left[(v_{k,i}^{(m)})^2 - (Y_k^{(m)})^2\right]\right\}}{\zeta_{k-1,i}^{(m)}}, \tag{E.12}$$

where

$$v_{k,i}^{(m)} = \frac{Y_k^{(m)} - \alpha_{k-1,i}^{(m)}Y_{k-1}^{(m)} - \beta_{k-1,i}^{(m)}}{\zeta_{k-1,i}^{(m)}}. \tag{E.13}$$

Consequently, the filters for the state of the Markov chain, number of jumps $\mathcal{G}_k^{s,j,(m)}$, occupation time $O_k^{j,(m)}$, and auxiliary process $\mathcal{T}_k^{j,(m)}$ are

$$\mathbf{u}_k^{(m)} = \mathbf{\Pi}_{k-1}^{(m)}\,\mathbf{C}_k^{(m)}\,\mathbf{u}_{k-1}^{(m)} \tag{E.14}$$

$$\gamma(\mathcal{G}_k^{s,j,(m)}\mathbf{x}_k^{(m)}) = \mathbf{\Pi}_{k-1}^{(m)}\mathbf{C}_k^{(m)}\,\gamma(\mathcal{G}_{k-1}^{s,j,(m)}\mathbf{x}_{k-1}^{(m)}) + \bar{\lambda}_{k,j}^{(m)}\langle\mathbf{u}_k^{(m)}, \mathbf{e}_j\rangle\pi_{k-1}^{s,j,(m)}\mathbf{e}_s \tag{E.15}$$

$$\gamma(O_k^{j,(m)}\mathbf{x}_k^{(m)}) = \mathbf{\Pi}_{k-1}^{(m)}\mathbf{C}_k^{(m)}\,\gamma(O_{k-1}^{j,(m)}\mathbf{x}_{k-1}^{(m)}) + \bar{\lambda}_{k,j}^{(m)}\langle\mathbf{u}_k^{(m)}, \mathbf{e}_j\rangle\mathbf{\Pi}_{k-1}^{(m)}\mathbf{e}_j \tag{E.16}$$

$$\gamma(\mathcal{T}_k^{j,(m)}(f)\mathbf{x}_k^{(m)}) = \mathbf{\Pi}_{k-1}^{(m)}\mathbf{C}_k^{(m)}\,\gamma(\mathcal{T}_{k-1}^{j,(m)}(f)\,\mathbf{x}_{k-1}^{(m)}) + f(\cdot)\,\bar{\lambda}_{k,j}^{(m)}\langle\mathbf{u}_k^{(m)}, \mathbf{e}_j\rangle\mathbf{\Pi}_{k-1}^{(m)}\mathbf{e}_j. \tag{E.17}$$

Equations (E.14) – (E.17) could be derived in a similar manner to those analogous filters given in [59] or [25]. From Eq. (E.10), the normalised filter estimates of $\gamma(\mathcal{G}_k^{s,j,(m)})$, $\gamma(O_k^{j,(m)})$ and $\gamma(\mathcal{T}_k^{j,(m)}(f))$ are determined by summing the components of the vector expressions in Eqs. (E.15) – (E.17), and then dividing each by the expression in Eq. (E.14).

Appendix F

Derivation of optimal parameter estimates in Chapter 4

In this Appendix, we shall derive the optimal estimates of the parameters for the HMM-driven OU process. Without loss of generality, the superscript (m) in all model parameters which is used to denote the FSI of different countries' is omitted in the following derivation.

F.1 Optimal estimate for π_{ji}

The proof is similar to that in [59]. To perform the measure change, we define a new measure $P^{\hat{\pi}_{ji}}$ via $\left.\dfrac{\mathrm{d}P^{\hat{\pi}_{ji}}}{\mathrm{d}P^{\pi_{ji}}}\right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k} \lambda_l^*$, where

$$\lambda_l^* = \sum_{j,i=1}^{N} \left(\frac{\hat{\pi}_{ji}}{\pi_{ji}}\right)^{\langle \mathbf{x}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{x}_l, \mathbf{e}_j \rangle}.$$

So,

$$\log \frac{\mathrm{d}P^{\hat{\pi}_{ji}}}{\mathrm{d}P^{\pi_{ji}}} = \sum_{l=1}^{k} \sum_{j,i=1}^{N} \left(\log\left(\hat{\pi}_{ji}\right) - \log\left(\pi_{ji}\right)\right) \langle \mathbf{x}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{x}_l, \mathbf{e}_j \rangle = \sum_{j,i=1}^{N} \mathcal{G}_k^{j,i} \log\left(\hat{\pi}_{ji}\right) + R,$$

where R does not contain $\hat{\alpha}$.

164

It must be noted that $\sum_{j=1}^{N} G_k^{j,i} = O_k^i$, $\sum_{j=1}^{N} \hat{G}_k^{j,i} = \hat{O}_k^i$. The optimal estimate for π_{ji} is the value that maximises the log-likelihood $\log \dfrac{\mathrm{d}P^{\hat{\pi}_{ji}}}{\mathrm{d}P^{\pi_{ji}}}$ constrained to $\sum_{j=1}^{N} \hat{\pi}_{ji} = 1$. Introducing the Lagrange multiplier and solving the associated optimisation problem yield the solution

$$\hat{\pi}_{ji} = \frac{\hat{G}_k^{j,i}}{\hat{O}_k^i}.$$

∎

F.2   Optimal estimate for α

Let $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)^\top \in \mathbb{R}^N$. To calculate $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N)^\top \in \mathbb{R}^N$, consider a new measure $P^{\hat{\alpha}}$ defined by $\left. \dfrac{\mathrm{d}P^{\hat{\alpha}}}{\mathrm{d}P^{\alpha}} \right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k} \lambda_l^*$, where

$$\lambda_l^* = \exp\left\{ \frac{(Y_l - \alpha(\mathbf{x}_{l-1}) Y_{l-1} - \beta(\mathbf{x}_{l-1}))^2 - (Y_l - \hat{\alpha}(\mathbf{x}_{l-1}) Y_{l-1} - \beta(\mathbf{x}_{l-1}))^2}{2\zeta^2(\mathbf{x}_{l-1})} \right\}.$$

This implies

$$LR(\hat{\alpha}_i) = E\left[\log \left. \frac{\mathrm{d}P^{\hat{\alpha}}}{\mathrm{d}P^{\alpha}} \right| \mathscr{F}_k \right] = E\left[\sum_{l=1}^{k} \sum_{i=1}^{N} \langle \mathbf{x}_{l-1}, \mathbf{e}_i \rangle \frac{-\hat{\alpha}_i^2 Y_{l-1}^2 - 2\hat{\alpha}_i(\beta_i Y_{l-1} - Y_l Y_{l-1})}{2\zeta_i^2} + R \middle| \mathscr{F}_k \right]$$

$$= \sum_{i=1}^{N} \frac{-\hat{\alpha}_i^2 \hat{\mathcal{T}}_k^i(Y_{k-1}^2) - 2\hat{\alpha}_i(\beta_i \hat{\mathcal{T}}_k^i(Y_{k-1}) - \hat{\mathcal{T}}_k^i(Y_k Y_{k-1}))}{2\zeta_i^2} + R,$$

where R does not contain $\hat{\alpha}_i$.

We differentiate $LR(\hat{\alpha}_i)$ with respect to $\hat{\alpha}_i$ and then equate the result to zero, yielding

$$\hat{\alpha}_i = \frac{\hat{\mathcal{T}}_k^i(Y_k \cdot Y_{k-1}) - \beta_i \cdot \hat{\mathcal{T}}_k^i(Y_{k-1})}{\hat{\mathcal{T}}_k^i(Y_{k-1}^2)}.$$

∎

F.3   Optimal estimate for β

Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)^\top \in \mathbb{R}^N$. To calculate $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N \right)^\top \in \mathbb{R}^N$, consider a new measure $P^{\hat{\beta}}$ defined by $\left. \dfrac{\mathrm{d}P^{\hat{\beta}}}{\mathrm{d}P^{\beta}} \right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k} \lambda_l^*$, where

$$\lambda_l^* = \exp\left\{ \frac{(Y_l - \alpha(\mathbf{x}_{l-1}) Y_{l-1} - \beta(\mathbf{x}_{l-1}))^2 - \left(Y_l - \alpha(\mathbf{x}_{l-1}) Y_{l-1} - \hat{\beta}(\mathbf{x}_{l-1}) \right)^2}{2\zeta^2(\mathbf{x}_{l-1})} \right\}.$$

This means that

$$
\begin{aligned}
LR\left(\hat{\beta}_i \right) = E\left[\log \frac{\mathrm{d}P^{\hat{\beta}}}{\mathrm{d}P^{\beta}} \middle| \mathscr{F}_k \right] &= E\left[\sum_{l=1}^{k} \sum_{i=1}^{N} \langle \mathbf{x}_{l-1}, \mathbf{e}_i \rangle \frac{-\hat{\beta}_i^2 - 2(\alpha_i Y_{l-1} - Y_l)\hat{\beta}_i}{2\zeta_i^2} + R \middle| \mathscr{F}_k \right] \\
&= \sum_{i=1}^{N} \frac{2\hat{\beta}_i \hat{\mathcal{T}}_k^i(Y_k) - 2\alpha_i \hat{\beta}_i \hat{\mathcal{T}}_k^i(Y_{k-1}) - \hat{O}_k^i \hat{\beta}_i^2}{2\zeta_i^2} + R,
\end{aligned}
$$

where R does not contain $\hat{\beta}_i$.

We get

$$\hat{\beta}_i = \frac{\hat{\mathcal{T}}_k^i(Y_l) - \alpha_i \hat{\mathcal{T}}_k^i(Y_{l-1})}{\hat{O}_k^i}$$

through the differentiation of $LR\left(\hat{\beta}_i \right)$ with respect to $\hat{\beta}_i$ and setting the result to zero.

$\blacksquare$

F.4   Optimal estimate for ζ^2

Let $\boldsymbol{\zeta}^2 = \left(\zeta_1^2, \zeta_2^2, \ldots, \zeta_N^2 \right)^\top \in \mathbb{R}^N$. To compute $\hat{\boldsymbol{\zeta}}^2 = \left(\hat{\zeta}_1^2, \hat{\zeta}_2^2, \ldots, \hat{\zeta}_N^2 \right)^T \in \mathbb{R}^N$, consider a new measure $P^{\hat{\zeta}^2}$ defined by $\left. \dfrac{\mathrm{d}P^{\hat{\zeta}^2}}{\mathrm{d}P^{\zeta^2}} \right|_{\mathscr{F}_k} = \Lambda_k^* = \prod_{l=1}^{k} \lambda_l^*$, where

$$\lambda_l^* = \sqrt{\frac{\zeta^2(\mathbf{x}_{l-1})}{\hat{\zeta}^2(\mathbf{x}_{l-1})}} \exp\left\{ (Y_l - \alpha(\mathbf{x}_{l-1}) Y_{l-1} - \beta(\mathbf{x}_{l-1}))^2 \left(\frac{1}{2\zeta^2(\mathbf{x}_{l-1})} - \frac{1}{2\hat{\zeta}^2(\mathbf{x}_{l-1})} \right) \right\}.$$

Hence,

$$
\begin{aligned}
LR\left(\hat{\zeta_i^2}\right) &= E\left[\log\frac{dP^{\hat{\zeta^2}}}{dP^{\zeta^2}}\bigg|\mathscr{F}_k\right] = E\left[\sum_{l=1}^{k}\sum_{i=1}^{N}\langle\mathbf{x}_{l-1},\,\mathbf{e}_i\rangle\left(-\frac{1}{2}\log\left(\hat{\zeta_i^2}\right)\right.\right.\\
&\qquad\left.\left.-\frac{Y_l^2+\alpha_i^2 Y_{l-1}^2+\beta_i^2-2\alpha_i Y_l Y_{l-1}-2\beta_i Y_l+2\alpha_i\beta_i Y_{l-1}}{2\hat{\zeta_i^2}}\right)+R\bigg|\mathscr{F}_k\right]\\
&= \sum_{i=1}^{N}\left(-\frac{1}{2}\hat{O}_k^i\log\left(\hat{\zeta_i^2}\right)\quad-\frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right)+\alpha_i^2\hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right)+\beta_i^2\hat{O}_k^i}{2\hat{\zeta_i^2}}\right.\\
&\qquad\left.-\frac{2\alpha_i\hat{\mathcal{T}}_k^i\left(Y_k Y_{k-1}\right)+2\beta_i\hat{\mathcal{T}}_k^i\left(Y_k\right)-2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{2\hat{\zeta_i^2}}\right)+R,
\end{aligned}
$$

where R does not contain $\hat{\zeta_i^2}$.

By differentiating $LR\left(\hat{\zeta_i^2}\right)$ with respect to $\hat{\zeta_i^2}$ and equating the resulting mathematical derivative to zero, we get

$$
\hat{\zeta_i^2} = \frac{\hat{\mathcal{T}}_k^i\left(Y_k^2\right)+\alpha_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}^2\right)+\beta_i^2\hat{O}_k^i-2\alpha_i\hat{\mathcal{T}}_k^i\left(Y_k Y_{k-1}\right)-2\beta_i\hat{\mathcal{T}}_k^i\left(Y_k\right)}{\hat{O}_k^i}\quad+\frac{2\alpha_i\beta_i\hat{\mathcal{T}}_k^i\left(Y_{k-1}\right)}{\hat{O}_k^i}.
$$

∎

Appendix G

Visualisations of HMM features in Chapter 4

This Appendix presents the evolution of HMM's regimes and the deviations from the mean-reverting level.
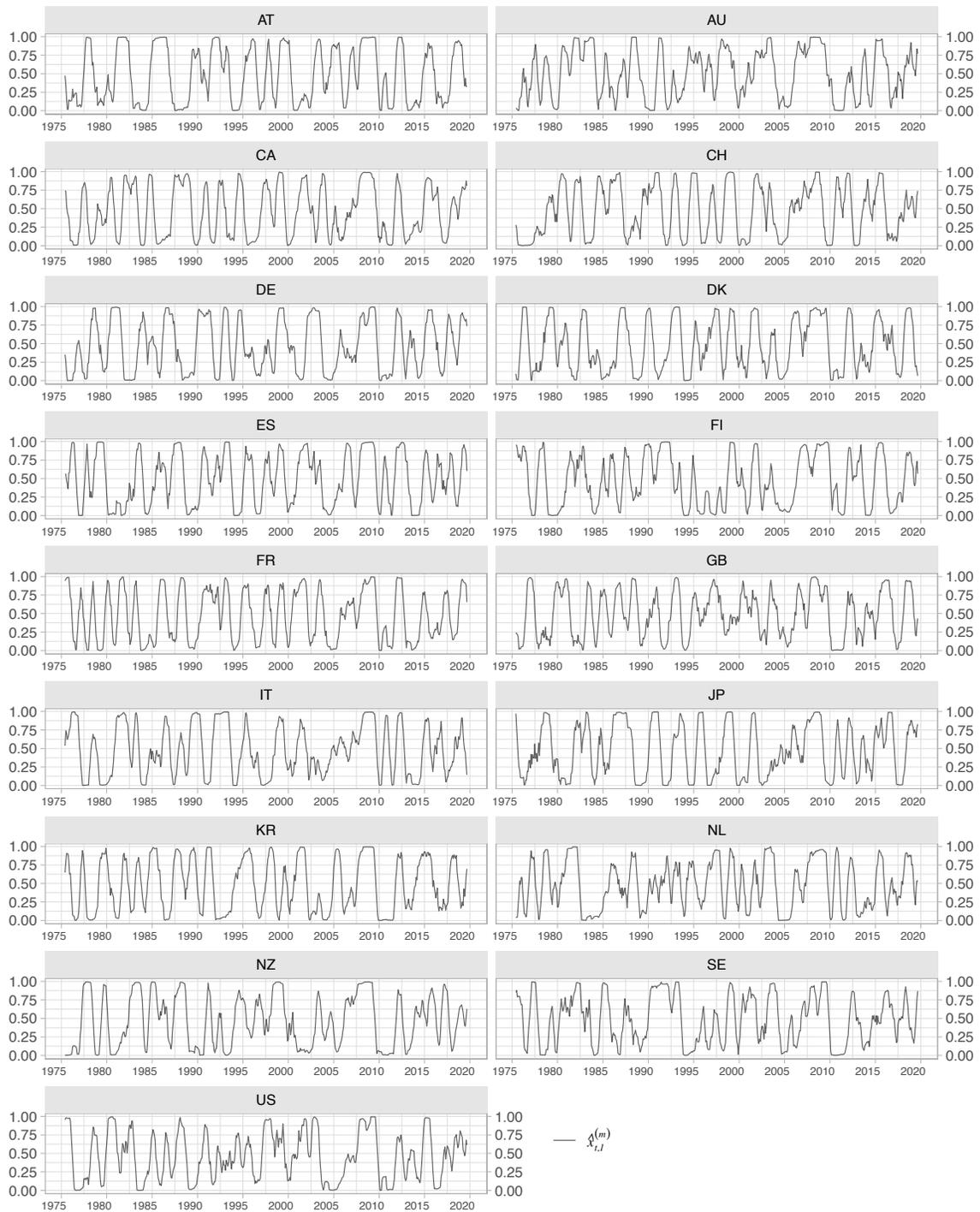
Figure G.1: The HMM's state estimate $\hat{x}_{t,1}^{(m)}$
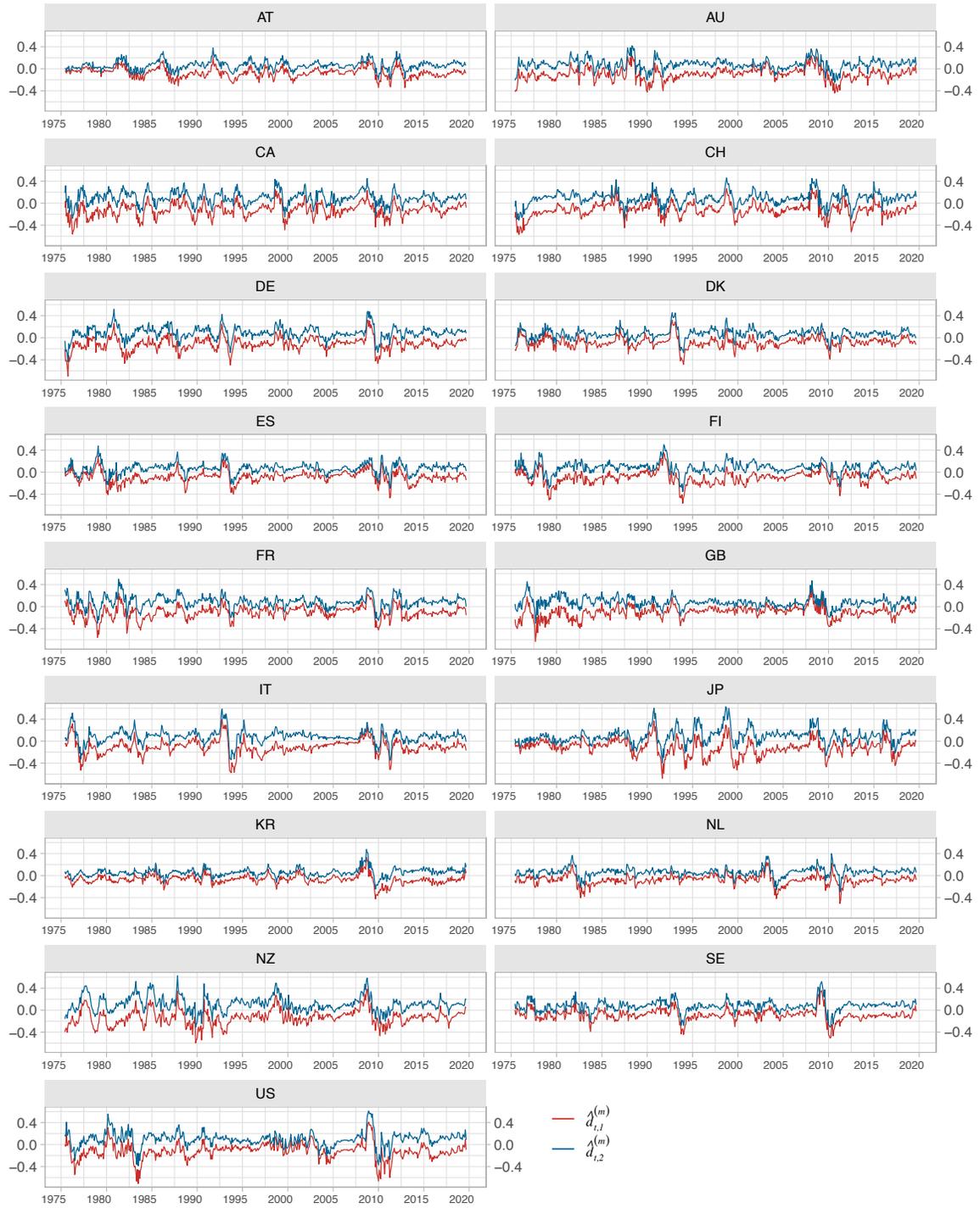
Figure G.2: Deviations from the reverting mean: $d_{t,1}^{(m)}$ and $d_{t,2}^{(m)}$

Curriculum Vitae

Name: Xing Gu

Post-Secondary Education and Degrees:

Ph.D. in Statistics (Financial Modelling), 2016-2022
The University *of* Western Ontario, London, ON

M.Sc. in Statistics, 2015-2016
The University *of* Western Ontario, London, ON

B.A. in Statistics, 2012-2015
Concordia University, Montréal, QC

B.E. in Telecommunication Engineering, 2001-2005
Shanghai University, Shanghai, China

Honours and Awards:

Western Graduate Research Scholarship, 2015 - 2020
Ontario Graduate Scholarship, 2015-2016, 2018-2019 and 2019-2020

Related Work Experience:

Research and Teaching Assistant, 2015-2021
The University *of* Western Ontario

Sessional Lecturer, 2018
The University *of* Western Ontario

Publications:

[1] X. Gu, R. Mamon, M. Davison and H. Yu An automated financial indices-processing scheme for classifying market liquidity regimes. *International Journal of Control*, 94(3):735 –756, 2021.

[2] X. Gu, R. Mamon, T. Duprey and H. Xiong Online estimation for a predictive analytics platform with a financial-stability-analysis application. *European Journal of Control*, 57:205–221, 2021.

[3] X. Gu, R. Mamon and T. Duprey A multivariate-index-driven anomaly detection system with supervised learning. submitted to *IEEE Transactions on Neural Networks and Learning Systems*.