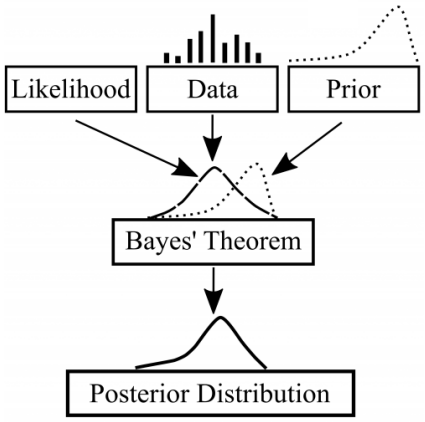
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
CHENNAI – 600036

Bayesian Statistics in Finance



A Thesis

Submitted by

SHUBHANGI SIKARIA

For the award of the degree

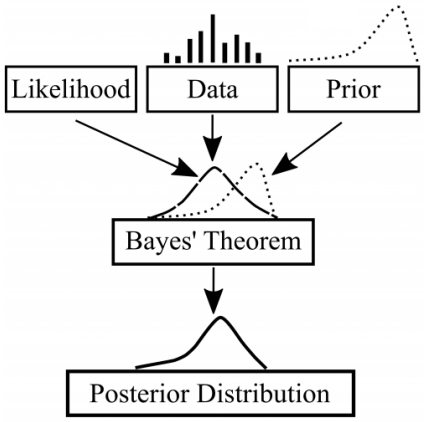
Of

DOCTOR OF PHILOSOPHY

August 2022

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
CHENNAI – 600036

Bayesian Statistics in Finance



A Thesis

Submitted by

SHUBHANGI SIKARIA

For the award of the degree

Of

DOCTOR OF PHILOSOPHY

August 2022

© 2022 Indian Institute of Technology Madras

THESIS CERTIFICATE

This is to undertake that the Thesis titled **BAYESIAN STATISTICS IN FINANCE**, submitted by me to the Indian Institute of Technology Madras, for the award of **Doctor of Philosophy**, is a bonafide record of the research work done by me under the supervision of **Dr. Neelesh S Upadhye and Dr. Rituparna Sen**. The contents of this Thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Place: Chennai 600036 **Shubhangi Sikaria**

Date: August 2022

Dr. Neelesh S Upadhye
Research advisor
Associate Professor
Department of Mathematics
IIT Madras

Dr. Rituparna Sen
Research co-advisor
Associate Professor
Applied Statistics Unit
ISI Bangalore

© 2022 Indian Institute of Technology Madras

LIST OF PUBLICATIONS

I. REFEREED JOURNAL ARTICLES BASED ON THESIS

S. Sikaria, R. Sen and N.S. Upadhye: **Bayesian Filtering for Multiperiod Mean-Variance Portfolio Selection**, Journal of Statistical Theory and Practice 15, 40 (2021).

R. Sen, A. Majumdar and S. Sikaria: **Bayesian Testing of Granger Causality in Functional Time Series**, Journal of Quantitative Economics (2022), https://doi.org/10.1007/s40953-022-00306-x

II. PRESENTATIONS IN CONFERENCES

S. Sikaria, R. Sen and N.S. Upadhye: **Bayesian Filtering for Multi-period Mean-Variance Portfolio Selection**, Research Workshop and Conference on Statistical methods in finance from 16- 21 Dec 2019, organized by Chennai Mathematical Institute and Indian Statistical Institute, Chennai.

S. Sikaria and R.Sen: **Granger Causality Analysis for Functional Time Series Data**, Virtual ISI World Statistics Congress 2021 from 11-16 July 2021, organised by International Statistical Institute.

S. Sikaria and R. Sen: **Bayesian Filtering for Generalized Mean-Variance Portfolio Selection**, Advanced Statistical Techniques in Business and Industry from 28-30 Dec 2021, organised by International Society for Business and Industrial Statistics (ISBIS).

S. Sikaria and R.Sen: **Option Pricing using Hawkes Process**, International Workshop on Statistical Modelling, 18-22 July 2022, organized by Statistical Modelling Society.

III. PUBLICATIONS IN CONFERENCE PROCEEDINGS

S. Sikaria and R.Sen: **Granger Causality Analysis for Functional Time Series Data**, Proceedings 63rd ISI World Statistics Congress, 11 - 16 July 2021, Virtual, pp 396-400.

S. Sikaria and R.Sen: **Option Pricing using Hawkes Process**, Proceedings of the 36th International Workshop on Statistical Modelling, 18-22 July 2022, Trieste, Italy, pp 577-581.

ACKNOWLEDGEMENTS

I am highly grateful to my supervisor, Dr. Neelesh Upadhye, for his constant support and patience during my Ph.D. study. I express my deepest gratitude to my esteemed co-guide, Dr. Rituparna Sen, Associate Professor, ISI Bangalore, for her invaluable supervision and teaching. Their ample knowledge and experience have encouraged me throughout my academic research. I am incredibly thankful to my Doctoral Committee members, Prof. Sundar S, Prof. V. Vetrivel, Dr. Sarang S Sane, and Dr. John Augustine, for their recommendations to enhance my research work.

I want to thank all the professors who taught me in the Department of Mathematics and helped me to complete my course work. I thank the Indian Institute of Technology, Madras (IIT Madras) for giving me an opportunity to pursue an M.Tech and Ph.D. from the most prestigious institute in India and for providing me with the fellowship and financial means to complete my Ph.D. I am thankful to the Department of Mathematics for providing me with all the facilities to conduct my research.

Special thanks to the co-author of one of my papers, Dr. Anandamayee Majumdar, Senior Statistician, Stock Assessment Program, Inter-American Tropical Tuna Commission, San Diego, CA, USA, for her valuable suggestions and continuous guidance throughout the project.

My wholehearted thanks to my supportive friends from the Department of Mathematics, especially Sayani Das and Ayushi Chauhan, and from the Finance Club of IIT Madras, mainly Srijan Upadhyay. Their presence at IIT, Madras, made the research life easy and delightful.

Finally, my family deserves endless appreciation: my parents for fulfilling all my needs and teaching me to fight against all odds, my brother and sister for unconditional love and care, and my husband for being my pillar of strength.

ABSTRACT

KEYWORDS Portfolio optimization; Prediction distribution; Dynamic programming; Dynamic linear model; Functional data analysis; Functional autoregression; Granger causality; Bayes factor; Option pricing; Hedging; Hawkes process

In this thesis, we provide the applications of Bayesian methods in finance, explaining the techniques that can be applied in the real world. Bayesian statistics is an approach of statistical inference that allows one to combine prior information about a population parameter with the data from external sources and update the current knowledge base. In recent years, Bayesian methods have seen impressive growth in the field of finance, as it provides a theoretically sound framework and robust estimation setting.

We studied four applications of Bayesian methods in finance—first, the mean-variance portfolio selection problem for multi-period. We sequentially update the returns of risky assets by employing the dynamic linear model (DLM). These updated returns are then supplied to dynamic programming to fetch optimal portfolio strategy at all intermediate times. In the second application, we use Bayesian DLM to simultaneously predict the yield curves of two economies to account for the cross-correlation between them. We work on functional data analysis techniques to study the yield curve data. Next, we perform the comparative test for two competing models by utilizing one of the Bayesian testing methods called the Bayes factor. We investigate the causal relationship between the yield curves of two countries. Lastly, we employ Bayesian filtration along with the Markov process to hedge an option.

CONTENTS

Page

ACKNOWLEDGEMENTS **i**

ABSTRACT **iii**

LIST OF TABLES **ix**

LIST OF FIGURES **xi**

GLOSSARY **xiii**

ABBREVIATIONS **xv**

NOTATION **xvii**

CHAPTER 1 INTRODUCTION **1**

CHAPTER 2 PRELIMINARIES **7**

2.1 Portfolio Optimization 7
2.1.1 Multi-Period Portfolio Optimization 8
2.2 Dynamic Linear Models 9
2.2.1 Prediction Distributions 11
2.3 Hawkes Process 12
2.3.1 Self-Exciting Process 13
2.4 The Fundamental Theorem Of Arbitrage Pricing 14
2.4.1 Hedging 16

CHAPTER 3 BAYESIAN FILTERING FOR MULTI-PERIOD MEAN-VARIANCE PORTFOLIO SELECTION **17**

3.1 Bayesian Forecasting 20
3.2 Dynamic Programming 22
3.3 Generalized Model for Multiple Risky Assets 25
3.4 Simulations 27
3.5 Empirical Illustration 30
3.5.1 Data 30
3.5.2 Results 30
3.5.3 Credible Interval Prediction 31
3.5.4 Results for Multiple Risky Assets 34

CHAPTER 4 BAYESIAN MODELLING AND FORECASTING FOR YIELD CURVES 37

4.1 The Model . . . 40
4.1.1 Multivariate Functional Autoregressive Model . . . 40
4.1.2 Dynamic Linear Models for MFAR (p) . . . 41
4.1.3 Theoretical Results . . . 42
4.2 Forecasting MFAR(p) . . . 42
4.2.1 Initialization . . . 43
4.2.2 Sampling Distribution . . . 44
4.3 Simulation study . . . 45
4.4 An Empirical Illustration: Yield Curves . . . 47

CHAPTER 5 BAYESIAN TESTING OF GRANGER CAUSALITY IN YIELD CURVES 49

5.1 Granger Causality in Functional Autoregressive Model . . . 50
5.2 Methodology . . . 51
5.2.1 F Test . . . 51
5.2.2 Bayesian Information Criteria . . . 52
5.2.2.1 Prior Distribution . . . 53
5.2.2.2 Likelihood Function . . . 55
5.2.2.3 Posterior Sampling Distribution . . . 55
5.2.2.4 Algorithm and Interpretation of Bayes Factor . . . 55
5.3 Simulation Study . . . 57
5.4 Application to Real Data . . . 57
5.4.1 Testing causality between yield curves of two countries . . . 58
5.4.2 Testing whether meteorological factors cause air pollution level . . . 59

CHAPTER 6 PRICING AND HEDGING OF OPTIONS USING HAWKES PROCESS 61

6.1 Hawkes Process . . . 62
6.2 Parameter estimation . . . 63
6.3 Option Pricing . . . 64
6.3.1 Option pricing with two state intensity . . . 65
6.4 Hedging using Bayesian Filtration . . . 66
6.4.1 Hedging with constant intensity rate . . . 66
6.4.2 Hedging with stochastic intensity rate . . . 68
6.5 Empirical Results . . . 70

CHAPTER 7 CONCLUDING REMARKS 73

APPENDIX A MULTI-PERIOD MV PORTFOLIO SELECTION 75

APPENDIX B MEAN SIGNATURE PLOT 77

REFERENCES 79

CURRICULUM VITAE **87**

DOCTORAL COMMITTEE **89**

LIST OF TABLES

Table	Caption	Page
3.1	Probability that Bayesian updates perform better for different values of ϕ (Fix $\mu = 0.01$ and $\sigma^2 = 0.2$)	29
3.2	Probability that Bayesian updates perform better for different values of μ (Fix $\phi = 0.1$ and $\sigma^2 = 1$)	29
3.3	Probability that Bayesian updates perform better for different values of σ^2 (Fix $\phi = 0.1$ and $\mu = 0.01$)	29
4.1	Simulation Results showing h-step ahead RMSFEs. Lower RMSFE for MFAR suggests it performs better than FAR.	47
4.2	h-step ahead RMSFEs of USA and UK yield curves.	47
5.1	Interpretation of Bayes Factor and Logarithm of Bayes Factor values	56
5.2	Bayes Factor Values of Simulated Series. Large positive values of $ln(B_{12})$ in first two cases depicts strong causal relationships. In third case, low value of $ln(B_{12})$ suggest no causal relation exists.	57
5.3	Bayes Factor Values of Yield Curves Example. Large negative values of $ln(B_{12})$ shows no causal relationship exits between USA and UK yield curves.	59
6.1	Nifty option prices on 1-Sep-2021 expiring on 9-Sep-2021 compared to purposed model and Black Scholes prices.	70
6.2	Nifty option prices on 1-Sep-2021 expiring on 9-Sep-2021 compared to purposed model and Black Scholes prices.	71

LIST OF FIGURES

Figure	Caption	Page
3.1	Efficient Frontiers of updated and non-updated model	32
3.2	Efficient frontiers with different exit time distribution	32
3.3	Efficient frontiers with different exit time	33
3.4	Histogram of empirical distribution of optimal portfolio returns for $\omega = 3$	34
3.5	Central 50% credible intervals for the return of optimal portfolios	34
3.6	Efficient frontiers for multiple assets of Bayesian updated and non-updated model	36
4.1	Raw data of interest rates (USA). Year: 2019. x-axis: Day, y-axis: Maturity in months, z-axis: Interest rate.	37
5.1	Boxplot of the logarithm of the Bayes factor for the dependent and independent cases depicting well-separated distributions of both cases.	58
5.2	The surface plot displays the evolution of the Pollution level in Delhi between the years 2015 and 2020, observed at every hour of a day.	60

GLOSSARY

Bayes Factor	Ratio of the marginal data densities of competing statistical models, usually a null and an alternative.
Bayesian Filtration	An algorithm to estimate the unknown probability density function recursively over time using incoming data set.
Dynamic Programming	Technique to compute optimal strategy efficiently by evolving data state.
Efficient Frontier	Set of optimal portfolios that provide the highest expected return for a given level of risk.
Functional Data	The data is functional if either it comes from a smooth function or is modeled by a smooth function.
Gibbs Sampling Algorithm	A Markov chain Monte Carlo (MCMC) algorithm that iteratively draws an instance from a specified multivariate probability distribution, when joint distribution is unknown.
Granger Causality	A time series X is said to have a causal influence on the time series Y if the prediction of series Y could be improved by the inclusion of past measurements from the time series X.
Hedging	An investment position that mitigates potential losses due to changes in asset price.
Markov Process	A stochastic process whose future probabilities depend only on the most recent values.
Yield Curve	Term structure of interest rates representing the relationship between spot rates of zero-coupon securities and their term to maturity.

ABBREVIATIONS

ACD Autoregressive Conditional Duration.

AIC Akaike Information Criterion.

AR(p) Autoregressive Model of order p.

ARMA Autoregressive Moving Average.

DLM Dynamic Linear Model.

DP Dynamic Programming.

FAR(p) Functional Autoregressive Model of order p.

FARMA Functional Autoregressive Moving Average.

FDLM Functional Dynamic Linear Model.

FLC Factor Loading Curves.

fMRI functional Magnetic Resonance Imaging.

FPCA Functional Principal Component Analysis.

GARCH Generalized Autoregressive Conditional Heteroscedasticity.

GcGMC Granger causality generalized measures of correlation.

HP Hawkes Process.

KFAS Kalman Filter and Smoother for Exponential Family State Space Models.

LQ Linear Quadratic.

MFAR Multivariate Functional Autoregressive Model.

mFPCA Multivariate Functional Principal Component Analysis.

MHM Modified Harmonic Mean.

MPT Modern Portfolio Theory.

MVCM Multivariate Varying Coefficient Model.

MVPS Mean-Variance Portfolio Selection.

NSE National Stock Exchange of India Limited.

RMSFE Root Mean Square Forecast Error.

S&P 500 Standard and Poor's 500.

UK United Kingdom.

USA United States of America.

USD United States Dollar.

VaR Value-at-Risk.

VAR(p) Vector Autoregressive Model of order p.

VARMA Vector Autoregressive Moving Average.

VIX Chicago Board Options Exchange's CBOE Volatility Index.

NOTATION

$\mathbb{E}[\cdot]$ Expectation operator

$\mathbb{P}(\mathrm{A})$ Probability of event A

$\mathbb{R}$ Set of real numbers

$\mathbb{R}^n$ Real coordinate space of dimension n

$\mathcal{N}(a, b)$ Normal distribution with mean a and variance b

$\mathrm{Var}(\cdot)$ Variance operator

max Maximum of

min Minimum of

$\sum$ Summation of

$\mathrm{X} \sim \mathrm{Y}$ X has the distribution of Y

CHAPTER 1

INTRODUCTION

Financial markets evolve continuously, leading to higher volume and complexity. This leads to random noise in financial data, causing uncertainty in prediction. Quantitative financial modeling extracts the relevant information and reduces uncertainty using statistical analysis. Such statistical models help in determining economic parameters through time and the relationships among multiple assets. Bayesian statistics incorporate the data from external sources and update the current knowledge domain based on the new data. It helps to evaluate the systematic changes in parameter values and revise the probabilities as further information arrives.

The two main philosophical approaches to statistics are based on their interpretation of the concept of probability. The first one is frequentist interpretation: The probability of an event is obtained from the frequency of its occurrence among all possible random samples. The alternative approach is called Bayesian statistics or 'inverse probability,' which directly applies the laws of probabilities to make inferences about the parameters. The unknown parameters are fixed in the frequentist method, and the information obtained from the sampling data is relevant for inference. However, the Bayesian inference considers the parameters as random variables with a certain probability distribution, subject to modification as new information is acquired.

The concept of Bayes' theorem was introduced by Reverend Thomas Bayes in 1763. However, recent advancements in computational power have fostered its implementation in applied statistics in the past fifty years. The Bayesian workflow begins with a prior distribution that incorporates the experimenter's subjective beliefs about the parameters before observing the data. Subsequently, it gathers the data and obtain the likelihood

function of the data given the parameters. Later, revise the prior distribution with the likelihood function of the data using Bayes' theorem. The resultant distribution is the posterior distribution that comprises updated beliefs about the parameters and the observed evidence.

The central issue of the market is the degree of predictability of financial time series. Bayesian techniques deliver the exact posterior density for parameters and produce highly efficient estimation. In the famous work by Escobar and West (1995), they utilized mixtures of a normal distribution to model the data and illustrate the usage of Bayesian inference for density estimation. Karatzas and Zhao (2001) employed Bayesian learning to compute unknown drift and Gaussian asset returns for an optimal portfolio allocation. Bayesian perspectives are beneficial in analyzing the time series (see: Durbin and Koopman (2000), McCulloch and Tsay (1993), Pai and Ravishanker (1998)). We applied a linear Bayesian filtering approach to update time-series parameters sequentially and obtain posterior predictive distribution. We also proposed a dynamic linear model for multivariate functional time series and employed the Bayesian approach along with Gibbs sampling to perform inference and prediction for multivariate time series.

In financial modeling, we typically strive to compare the competitive models. Bayesian testing is more efficient and convenient than the significance test as it assigns a probability to the hypothesis. Bayes factor, introduced by Kass and Raftery (1995), is a ratio of probabilities and can be interpreted intuitively. Significance tests such as p-value rely on hypothetical repeated samples, while the Bayes factor derives its value from the observed data and thus leads to a more objective understanding. Assaf and Tsionas (2018) provides a detailed comparison of the Bayes factor and p-values. The Bayes factor has been studied broadly and has considerable extensions like the posterior Bayes factor (Aitkin (1991)), fractional Bayes Factor (O'Hagan (1995)), and intrinsic Bayes factor (Berger and Pericchi (1996)). One can find numerous applications of the Bayes factor in different fields. Raggi and Bordignon used it to compare stochastic volatility

models. Heck *et al.* (2022) presents a brief review of applications of the Bayes factor in psychology. Singer *et al.* (2015) leveraged the sensitivity of the Bayes factor on the prior to compare hypotheses about human trails on the web. Our study utilizes the Bayes factor to address the Granger causality for functional time series.

The presence of volatility in the financial market exposes the investors to the market risk of stocks, exchange rates, or interest rates. Institutions or corporations apply modern financial techniques to mitigate or reduce the downside risk. Financial derivatives, including options, forwards, etc., are an adequate tool to avoid risk. Hedging is an investment position taken to mitigate the risk of losses due to changes in the price of the security. Hedging with options can prevent the downside risk in undesirable scenarios while allowing positive incomes. There have been several innovations in hedging strategies. Anderson and Danthine (1980) obtain the optimal hedge ratio by assuming explicit consideration of price expectation production possibilities and the number of the available future market. Later, Lence and Hayes (1994) applied the Bayes criterion to solve for mean-variance hedge ratio in two scenarios based on the multivariate normal distribution. Following Lence and Hayes (1994), Foster and Whiteman (2002) postulated a time series model for spot and futures prices and applied a numerical Bayesian procedure to compute predictive densities and optimal hedges. Recently, Billio *et al.* (2018) built a system of simultaneous equations for dynamic hedging in energy futures markets by applying the Bayesian multi-chain Markov-switching GARCH model. In our study, we take up the idea of using the Hawkes process with stochastic jump intensities as stock price models and hedging an option using the Markov process embedded with Bayesian filtering.

In this thesis, we consider four applications of Bayesian statistics: (1) Multi-period mean-variance portfolio optimization, (2) Multivariate functional data analysis, (3) Granger causality testing in functional time series, and (4) Option pricing and hedging.

Chapter 2

We explain the terms we use throughout the thesis and also state the known results that would give us a better insight into the later developments.

Chapter 3

We optimize the portfolio for the multi-period model applying dynamic programming techniques which allows to rebalance the portfolio weights at intermediate times. We consider the situation where the distribution of parameters is unknown and is estimated directly from the dynamically arriving data. We implement the Bayesian filtering method through dynamic linear models to sequentially update the parameters. We also acknowledge the uncertain investment lifetime to make the model more adaptive to the market conditions. These updated parameters are put into the dynamic mean-variance problem to arrive at optimal efficient portfolios. We implement this model to the S&P500 illustrating that the data strongly favor the Bayesian updating and is practically implementable.

Chapter 4

We propose a two-level hierarchical Gaussian process that is dynamically modeled as a multivariate functional autoregressive model (MFAR). We develop a multivariate functional autoregressive model (MFAR), which captures the cross-correlation among multiple functional time series and thus improves forecast accuracy. We estimate the parameters under the Bayesian dynamic linear models (DLM) framework. The performance of this model is illustrated through a series of simulation studies and applications to yield curve prediction of various countries. Our model provides improved forecast accuracy in most cases.

Chapter 5

We investigate the causal connection between functional time series through the application of Granger causality. In order to capture Granger causality from one FAR series to another we employ Bayes Factor. Motivated by the broad application of functional data in finance, we investigate the causality between the yield curves of USA

and UK. Bayes factor values shows that no causal relation exists among the interest rates of these two countries. Furthermore, we illustrate a climatology example, suggesting that the meteorological factors Granger cause pollutant daily levels in Delhi.

Chapter 6

We propose a methodology for European options pricing in which the Hawkes process drives variations in asset prices. We associate two point processes corresponding to the sum of the asset's positive and negative jumps. The point processes have a self and mutually exciting stochastic intensities with an exponential kernel. We employ a mean signature plot to estimate the parameters. We also demonstrate the hedging strategy for constant and stochastic intensities. We use Bayesian filtering in the Markov process to obtain the optimal hedge ratios for the stochastic intensities case. We examine the model's implementation in real data applications and compare it with the Black-Scholes formula.

CHAPTER 2

PRELIMINARIES

In this chapter, we give a detailed description of the basic terminology we will be using further. We also state some fundamental theorems. These theorems will not be proven as they are substantial and well-known.

2.1 PORTFOLIO OPTIMIZATION

Investing in the stock market exposes the investor to risk, which can be reduced by investment in an assortment or range of securities. The major challenge faced by the investors is how to allocate their wealth among multiple assets within the market. This issue can be resolved by using the portfolio optimization technique. It is a quantitative tool that allows investors to select the best portfolio out of all portfolios based on their preferences. Harry Markowitz, in 1952, put forward a practical approach to allocate wealth by considering the trade-off between risk and return and identified the portfolio's risk with the variance of its return. His work was the first to formalize the concept of diversification in investing, which forms a cornerstone of the Modern Portfolio Theory (MPT). This mean-variance analysis by Markowitz forms the basis for portfolio optimization in today's financial sector. The key insight is that the risk and return of an asset should be assessed by its contribution to the portfolio's overall risk and return. Let $\mathrm{X} = \mathrm{X}(x, \omega)$ be a payoff function that depends on decision vector x and random element $\omega \in \Omega$, where Ω is a set of random elements. Traditionally, Markowitz's mean-variance model can be stated as the minimization of risk while the mean payoff exceeds some predetermined level r_0. The variance of the decision's payoff represents the risk. Mathematically, the model can be expressed as:

$$\min_{x \in \mathrm{S}} \{\mathrm{Var}(\mathrm{X}(x, \omega)) | \mathbb{E}[\mathrm{X}(x, \omega)] \geq r_0\}, \tag{2.1}$$

where $S \subset \mathbb{R}^n$ is the set of feasible decisions x.

Based on the investors objective, alternative forms are frequently employed:

(i) Maximize the return $\mathbb{E}[X(x, \omega)]$ while the risk is below certain threshold ϕ_0:

$$\max_{x \in S} \{\mathbb{E}[X(x, \omega)] | Var(X(x, \omega)) \leq \phi_0\}. \tag{2.2}$$

(ii) Weighted combination of mean and variance is optimzed:

$$\max_{x \in S} \{\mathbb{E}[X(x, \omega)] - \lambda Var(X(x, \omega)) | \lambda \geq \lambda_0\}. \tag{2.3}$$

2.1.1 Multi-Period Portfolio Optimization

The single period model described above is static in nature. Hence the decision about the investment policy can only be made at the beginning of the investment period. Numerous work on continuous and multi-period mean-variance portfolio optimization was found in the literature. However, the main obstacle in solving dynamic mean-variance optimization problem is the non-linearity of conditional variance, that is, $Var[Var[x_T|\mathcal{F}_t]|\mathcal{F}_s] \neq Var[x_T|\mathcal{F}_s]$, $s \leq t$, where x_T is the wealth at time T. To tackle this problem, Li and Ng (2000) transformed the original mean-variance equations into a tractable auxiliary stochastic linear-quadratic (LQ) problem.

Lets assume that the financial market is defined on a probability space $(\Omega, \mathcal{F}; \{\mathcal{F}_t\}_{0 \leq t \leq T}; \mathbb{P})$ with finite time horizon $[0, T]$. Let r^0 be a deterministic risk free return in one time step and r_t denote the return of risky asset at time t. Consider the excess return of a risky asset at period $t + 1$ as $e_t = r_t - r^0$. Let x_t denote the investor's wealth in the risky asset in the period $t (t = 0, 1, \cdots, T)$. Define investment series over T periods, $\{u_0, u_1, ..., u_{T-1}\}$, where u_t is the amount invested in the risky security at time t. Formally, the investor's problem is given by:

$$\begin{aligned} \max_{\{u_s\}_{s=t}^{T}} \quad & \mathbb{E}(x_T) - \omega Var(x_T) \\ \text{s.t.} \quad & x_{t+1} = r_t^0 x_t + e_t u_t, \end{aligned} \tag{2.4}$$

where $\omega > 0$ represents the investor attitude towards risk. The following theorem illustrates the transformation of problem 2.4 to LQ problem.

Theorem 2.1. *If $\{u_s^*\}_{s=t}^{T}$ is the optimal control for the problem defined in equation 2.4, then $\{u_s^*\}_{s=t}^{T}$ is also the optimal control for the following problem:*

$$\min_{\{u_s\}_{s=t}^{T}} \quad \mathbb{E}\left[\left(x_T - \frac{\gamma}{2}\right)^2 | x_t\right], \tag{2.5}$$

where $\gamma = \frac{1}{\omega} + 2\mathbb{E}_{u^*}[x_T|x_t]$. Here the operator $\mathbb{E}_{u^*}[\cdot]$ denotes the expectation of the investor's terminal wealth if she invests according to the optimal strategy $\{u_s^*\}_{s=t}^{T}$.

Proof. See Li and Ng (2000). ■

In chapter 3, we extend the multi-period portfolio optimization for uncertain exit time and update the parameters dynamically using Bayesian tools. Also, we reconsider theorem 2.1 for the extension mentioned earlier.

2.2 DYNAMIC LINEAR MODELS

Dynamic linear models (DLM) are a particular case of state-space models, which provide a generic framework to investigate non-stationary time series models. It offers a sequential learning approach for system state updating and prediction. It uses Bayesian techniques for routine learning and organizing uncertain information. Harrisons and Stevens (1971) extended the basic structural models to a class of dynamic linear models. Later in 1976, Harrisons and Stevens introduced the generalized dynamic linear model along with Kalman filter recurrence relations. DLM provides a parametric representation of a process, leading to better interpretation and forecast modeling.

Here, we present the general DLM definition and prediction distributions as delivered by West and Harrison (2006). The DLM is employed in chapter 3 for Bayesian forecasting of returns of risky assets. In chapter 4, we will convert the two-level hierarchical model

to DLM form for functional data modeling and forecasting. Firstly, we show the general normal DLM for a $(r \times 1)$ vector observation $\mathbf{Y}_t$ defined over times $t = 1, 2, \cdots$.

Definition 2.1 (**Normal Dynamic Linear Model**)**.** The general normal dynamic linear model (DLM) is characterised by a set of quadruples

$$\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t = \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$$

for each time t, where

(a) $\mathbf{F}_t$ is a known $(n \times r)$ matrix;

(b) $\mathbf{G}_t$ is a known $(n \times n)$ matrix;

(c) $\mathbf{V}_t$ is a known $(r \times r)$ variance matrix;

(d) $\mathbf{W}_t$ is a known $(n \times n)$ variance matrix.

This quadruple defines the model relating $\mathbf{Y}_t$ to the $(n \times 1)$ parameter vector $\boldsymbol{\theta}_t$ at time t, and the $\boldsymbol{\theta}_t$ sequence through time, via the sequentially specified distributions

$$\begin{aligned} (\mathbf{Y}_t|\boldsymbol{\theta}_t) &\sim \mathcal{N}[\mathbf{F}_t'\boldsymbol{\theta}_t|\mathbf{V}_t] \qquad \text{and} \\ (\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) &\sim \mathcal{N}[\mathbf{G}_t\boldsymbol{\theta}_{t-1}|\mathbf{W}_t]. \end{aligned} \tag{2.6}$$

An alternative representation of the equations 2.6 along with the initial prior at time $t = 0$, provide the complete definition of dynamic linear model as follows:

Definition 2.2 (**Univariate Dynamic Linear Model**)**.** For each t, the general univariate DLM is defined by:

$$\begin{aligned} &\text{Observation Equation:} && \mathbf{Y}_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \boldsymbol{v}_t, && \boldsymbol{v}_t \sim \mathcal{N}[0|\mathbf{V}_t], \\ &\text{System equation:} && \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, && \boldsymbol{w}_t \sim \mathcal{N}[0|\mathbf{W}_t], \\ &\text{Initial information:} && (\boldsymbol{\theta}_0|\mathrm{D}_0) \sim \mathcal{N}[\boldsymbol{m}_0, \mathbf{C}_0], \end{aligned} \tag{2.7}$$

for some prior moments $\boldsymbol{m}_0$ and $\mathbf{C}_0$ and D_{t-1} is the information set available prior to time t. The observational and evolution error sequences are assumed to be internally and

mutually independent, and are independent of $(\boldsymbol{\theta}_0|\mathrm{D}_0)$.

The observation equation defines the sampling distribution for $\mathbf{Y}_t$ conditioned on $\boldsymbol{\theta}_t$. It implies that given $\boldsymbol{\theta}_t$, $\mathbf{Y}_t$ is independent of past observations and other parameter values. The evolution equation states the time evolution of $\boldsymbol{\theta}_t$, that is, given $\boldsymbol{\theta}_{t-1}$, the distribution of $\boldsymbol{\theta}_t$ is independent of $\mathbf{Y}_t$ and other observations prior to time $t-1$.

2.2.1 Prediction Distributions

Theorem 2.2. *In the univariate DLM of equation 2.7, one-step forecast and posterior distributions are given, for each t, as follows:*

(a) Posterior at $t-1$: For some mean $\boldsymbol{m}_{t-1}$ and variance matrix $\mathbf{C}_{t-1}$,

$$(\boldsymbol{\theta}_{t-1}|\mathrm{D}_{t-1}) \sim \mathcal{N}[\boldsymbol{m}_{t-1}, \mathbf{C}_{t-1}].$$

(b) Prior at t:

$$(\boldsymbol{\theta}_t|\mathrm{D}_{t-1}) \sim \mathcal{N}[\boldsymbol{a}_t, \mathbf{R}_t],$$

where

$$\boldsymbol{a}_t = \mathbf{G}_t\boldsymbol{m}_{t-1} \qquad \textit{and} \qquad \mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t.$$

(c) One-step forecast:

$$(\mathbf{Y}_t|\mathrm{D}_{t-1}) \sim \mathcal{N}[f_t, \mathrm{Q}_t],$$

where

$$f_t = \mathbf{F}_t'\boldsymbol{a}_t \qquad \textit{and} \qquad \mathrm{Q}_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + \mathbf{V}_t.$$

(d) Posterior at t:

$$(\boldsymbol{\theta}_t|\mathrm{D}_t) \sim \mathcal{N}[\boldsymbol{m}_t, \mathbf{C}_t],$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \mathbf{A}_t e_t \qquad \textit{and} \qquad \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t\mathrm{Q}_t\mathbf{A}_t',$$

where

$$\mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t\mathrm{Q}_t^{-1} \qquad \textit{and} \qquad e_t = \mathbf{Y}_t - f_t.$$

Proof. See West and Harrison (2006). ∎

2.3 HAWKES PROCESS

Natural clustering is frequently observed in the events that have occured over a period. In insurance, an event can be a claim arrival by a person, increasing the chance of subsequent claim arrivals (Swishchuk and Kazmerchuk (2001)). In geophysics, an earthquake is an event indicative of the likelihood of aftershocks to follow in the near future (Bacry and Muzy (2016)). In epidemiology, events can be an infected person that can spread the infections and result in an outbreak (Chiang *et al.* (2022)). The property in which each occurrence of an event excites the process and increases the probability of succeeding events is called the self-exciting property. A similar phenomenon is observed in financial markets, where trading activities often cluster in time.

Hawkes (1971) introduced the Hawkes process (HP), reflecting endogenously induced clustering. The HP is a mathematical tool to model self and mutually exciting processes. They can be viewed as stochastic versions of popular point processes. HP is used in machine learning for event clustering and causal inference. We have employed the Hawkes process to price and hedge the options in chapter 6.

To understand the Hawkes process, we first need to study basic definitions provided by Laub *et al.* (2015) related to the point process.

Definition 2.3 (**Point Process).** If a sequence of random variables $\mathrm{T} = \{\mathrm{T}_1, \mathrm{T}_2, \cdots\}$, taking values in $[0, \infty)$, has $\mathbb{P}(0 \leq \mathrm{T}_1 \leq \mathrm{T}_2 \leq \cdots) = 1$, and the number of points in a bounded region is a.s. finite, then T is a (simple) point process.

Definition 2.4 (**Counting Process).** A counting process is a stochastic process $(\mathrm{N}(t) : t \geq 0)$ taking values in N_0 that satisfies $\mathrm{N}(0) = 0$, is almost surely (a.s.) finite, and is a right-continuous step function with increments of size +1.

A counting process is basically cumulative sum of the number of arrivals upto current time. While the point process is a sequence of random arrival times where the counting process has jumped.

Definition 2.5 (**Conditional Intensity Function).** Consider a counting process $\mathrm{N}(\cdot)$ with associated filtration $\mathcal{F}(\cdot)$, defined as $(\mathcal{F}(u), u \geq 0)$ an increasing sequence of $\sigma-$algebras. If a (non-negative) function $\lambda^*(t)$ exists such that

$$\lambda^*(t) = \lim_{h \downarrow 0} \frac{\mathbb{E}[\mathrm{N}(t+h) - \mathrm{N}(t) | \mathcal{F}(t)]}{h} \tag{2.8}$$

which only relies on information of $\mathrm{N}(\cdot)$ in the past (that is, $\lambda^*(t)$ is $\mathcal{F}(t)$-measurable), then it is called the conditional intensity function of $\mathrm{N}(\cdot)$.

2.3.1 Self-Exciting Process

The process is self-exciting if the conditional intensity function increases after an arrival. Originally, self-exciting process is defined by an intensity which is linearly dependent on past events. This is also termed as linear self-exciting process. For a linear self-exciting process $\mathrm{N}(\cdot)$ the intensity function is defined as:

$$\lambda(t) = \mu + \int_0^t \phi(t-u) d\mathrm{N}(u) = \mu + \sum_{\mathrm{T}_i < t} \phi(t - \mathrm{T}_i), \tag{2.9}$$

where $0 < \mathrm{T}_1 < \mathrm{T}_2 < \cdots < \mathrm{T}_n < \cdots$ are times at which arrival occurs. $\mu > 0$ is deterministic base intensity and the function $\phi(u) \geq 0, u > 0$ called the excitation function of the process. $\phi(\cdot)$ expresses the contribution of past events T_i on the current value of intensity.

Definition 2.6 (**Hawkes Process).** Consider $(\mathrm{N}(t) : t \geq 0)$ a counting process, with associated filtration $(\mathcal{F}(t) : t \geq 0)$, that satisfies

$$\mathbb{P}(\mathrm{N}(t+h) - \mathrm{N}(t) = m | \mathcal{F}(t)) = \begin{cases} \lambda^*(t)h + o(h) & m = 1 \\ o(h) & m > 1 \\ 1 - \lambda^*(t)h + o(h) & m = 0. \end{cases} \tag{2.10}$$

Suppose the process' conditional intensity function is of the form

$$\lambda(t) = \mu + \int_0^t \phi(t-u) d\mathrm{N}(u) \tag{2.11}$$

for some $\mu > 0$ and $\phi : (0, \infty) \to [0, \infty)$ such that $\phi(\cdot) \neq 0$ to avoid the trivial case, that is, a homogeneous Poisson process. Such a process $N(\cdot)$ is a Hawkes process.

The simplest choice of excitation function is exponential kernel, given as:

$$\phi(u) = \alpha e^{-\beta u}, \quad u > 0,$$

which is parameterised by $\alpha, \beta > 0$.

Definition 2.7 (**Mutually Exciting Hawkes Process).** Consider a collection of m counting processes $\{N_1(\cdot), \cdots, N_m(\cdot)\}$ denoted **N**. Say $\{T_{i,j} : i \in 1, ..., m, j \in N\}$ are the random arrival times for each counting process (and $t_{i,j}$ for observed arrivals). If for each $i = 1, \cdots, m$, $N_i(\cdot)$ has conditional intensity of the form

$$\lambda_i^*(t) = \mu_i + \sum_{j=1}^{m} \int_{-\infty}^{t} \phi_j(t-u) dN_j(u) \tag{2.12}$$

for some $\mu_i > 0$ and $\phi_i : (0, \infty) \to [0, \infty)$, then N is called a mutually exciting Hawkes process.

2.4 THE FUNDAMENTAL THEOREM OF ARBITRAGE PRICING

The fundamental theorem of arbitrage pricing establishes the necessary and sufficient conditions for the absence of arbitrage in market and completeness of market. An arbitrage is a trading strategy to make profit from the difference in asset prices in two or more markets. Completeness of market is a property which states that every contingent claim is attainable. We will consider the case of single period market.

Assume the market has K + 1 assets of which K assets are risky with share price at time t is given as $S_t^1, S_t^2, \cdots, S_t^K$ and one riskless asset, that is, it's value is independent of market scenario. The price of riskless asset at time t is assumed to be S_t^0. To account for the uncertainity in the market, we consider N possible market scenario, denoted as $\omega_1, \omega_2, \cdots, \omega_N$. Thus, the price of asset j at time t in market scenarios ω_i is $S_t^j(\omega_i)$ for

$j = 1, 2, \cdots, \mathrm{K}$. Let r be the riskless rate of return such that the share price of riskless asset at $t = 1$ is

$$\mathrm{S}_1^0(\omega_i) = e^r, \quad \forall\, i = 1, 2, \cdots, \mathrm{N}.$$

Definition 2.8 (**Portfolios).** A portfolio is vector of K real numbers denoted as:

$$\theta := \{\theta^1, \theta^2, \cdots, \theta^{\mathrm{K}}\} \in \mathbb{R}^{\mathrm{K}}$$

where, θ^j is the number of shares of j-th asset owned by the investor. The value of the portfolio θ at time $t = 0$ is

$$\mathrm{V}_0(\theta) = \sum_{j=1}^{\mathrm{K}} \theta^j \mathrm{S}_0^j$$

and the value of portfolio at time $t = 1$ in market scenario ω_i is

$$\mathrm{V}_1(\theta; \omega_i) = \sum_{j=1}^{\mathrm{K}} \theta^j \mathrm{S}_1^j(\omega_i).$$

Definition 2.9 (**Arbitrage).** An arbitrage is a portfolio that makes risk free profit, that is, earning profit from nothing. An arbitrage portfolio θ is formally written as:

$$\mathrm{V}_0(\theta) \leq 0 \quad \text{and} \quad \mathrm{V}_1(\theta; \omega_i) > 0 \quad \forall\, i = 1, 2, \cdots, \mathrm{N},$$

or

$$\mathrm{V}_0(\theta) < 0 \quad \text{and} \quad \mathrm{V}_1(\theta; \omega_i) \geq 0 \quad \forall\, i = 1, 2, \cdots, \mathrm{N}.$$

Definition 2.10 (**Equilibrium Measure).** An equilibrium measure, also called as risk-neutral measure is a probability distribution $\pi_i = \pi(\omega_i)$ on the set of market scenarios such that the share price of every asset at time $t = 0$ is the discounted expectation, under π, of the share price at time $t = 1$. Formally, a probability distribution π is an equilibrium measure if

$$\mathrm{S}_0^j = e^{-r} \sum_{i=1}^{\mathrm{N}} \pi(\omega_i) \mathrm{S}_1^j(\omega_i) \quad \forall\, j = 1, 2, \cdots, \mathrm{K}.$$

Theorem 2.3 (**Fundamental Theorem of Arbitrage Pricing).** *There exists an*

equilibrium measure if and only if arbitrages do not exist.

2.4.1 Hedging

Hedging is an investment position in one or more assets with an intent to reduce the risk of loss from another asset. It involves forming a portfolio that has the same market value as the market value of the asset that is to be hedged. In chapter 6, we hedge an option using a stock, a bond, and another option of the same asset. We employed the fundamental theorem of arbitrage pricing to form the hedging strategy.

Consider there are K + 1 freely traded securities in the market $\mathcal{M}$, denoted as B and $A^1, A^2, \cdots, A^K$. The share price at time t in market scenario ω_i is given as $S_t^B(\omega_i)$ and $S_t^j(\omega_i)$ for $j = 1, 2, \cdots, K$ and $i = 1, 2, \cdots, N$.

Definition 2.11 (**Replicating Portfolio).** A replicating portfolio for asset B is a portfolio $\theta = (\theta_1, \theta_2, \cdots, \theta_K)$ whose market value for each market scenario is same as the market value of asset B in that market scenario. Mathematically, a portfolio θ is a replicating portfolio for asset B if

$$S_1^B(\omega_i) = \sum_{j=1}^{K} \theta_j S_1^j(\omega_i) \quad \forall\, i = 1, 2, \cdots, N.$$

Let's assume there exists a derivative security in the market $\mathcal{M}$ which is a tradeable asset.

Theorem 2.4 (**Completeness Theorem).** *Let $\mathcal{M}$ be an arbitrage-free market with a riskless asset. If for every derivative security there is a replicating portfolio in the assets $A^1, A^2, \cdots, A^K$, then the market $\mathcal{M}$ is complete. Conversely, if the market $\mathcal{M}$ is complete, and if the unique equilibrium measure π gives positive probability to every market scenario ω_i, then for every derivative security there is a replicating portfolio in the assets $A^1, A^2, \cdots, A^K$.*

CHAPTER 3

BAYESIAN FILTERING FOR MULTI-PERIOD MEAN-VARIANCE PORTFOLIO SELECTION

Investing in the stock market exposes the investors to the risk, which can be reduced by investment in an assortment or range of securities. The major challenge faced by investors is how to allocate their capital over many financial assets. Thus, the intention is to ascertain an optimal portfolio that gives the best allocation of wealth by generating a high return along with a low risk. Markowitz (1952) paved the foundation of the modern portfolio theory, which is modeled as a return-risk bi-criteria optimization problem, characterizing portfolio return with a mean rate of return and risk with variance. The analytical solution of mean-variance portfolio selection (MVPS) in a single-period was derived by Merton (1972). However, the single-period model is static, which implies that the decision can be made only at the beginning of the investment period, and one needs to wait for results until the investment horizon ends. Due to the long investment horizon, it is preferable to rebalance the portfolio weights at intermediate time points after observing current values. A natural extension to a multi-period model allows investors to make financing decisions at distinct time points, and account for the immediate market scenario. Among multi-period investment models, one can have the time horizon to be continuous and discrete. Zhou and Li (2000) formulated a closed-form solution in a multi-period model for a continuous-time setting by embedding nonstandard problems into a class of auxiliary stochastic linear-quadratic (LQ) problems. On a similar line, Li and Ng (2000) proposed an efficient algorithm for determining an optimal analytical solution to maximize utility function in the multi-period discrete-time portfolio selection problem.

In work mentioned above, the stock prices are assumed to follow a vector-valued

Geometric Brownian motion implying that the stock returns are time-independent. The empirical evidence shows that the returns of risky assets always exhibit serial correlations that are captured by ARMA models. Balvers and Mitchell (1997) were first to derive an explicit analytical solution to the dynamic portfolio problem by employing normal ARMA(1,1) to incorporate the autocorrelation among returns of risky assets. Xu and Li (2008) investigated dynamic portfolio selection for serially correlated returns by embedding the mean-variance model into a quadratic utility model. Later, they applied dynamic programming for one risky asset and one risk-free asset to obtain the explicit optimal investment strategy. General forms of correlation structure for returns based on stochastic market for single risky asset was assumed by Celikyurt and Özekici (2007) and Dokuchaev (2007). Gao *et al.* (2015) and Chiu and Wong (2014) investigated a dynamic MVPS for multiple risky assets and one riskless asset with a general correlation for a market. He *et al.* (2015) established the optimal investment policy for the multi-period model by solving the stationary equation directly without using the embedding technique. Meanwhile, to make the MVPS model more realistic the multi-period discrete-time model has been studied extensively by incorporating various real features in recent years (see the ref. Gao and Li (2014), Wei and Gao (2015), Skaf and Boyd (2009)).

Generally the parameters like drift and volatility of returns for the portfolio selection models are estimated from the past data and are assumed to remain constant for later periods. However, this solution is not realistic as the parameters are not adaptive according to market conditions. To circumvent this issue, Mao and Särndal (1966) used Bayesian inference for a discrete-time single period portfolio selection model. Later an extensive literature incorporating Bayesian statistics have emerged, see Frost and Savarino (1986) and Aguilar and West (2000). Karatzas and Zhao (2001) did the most recognizable work on Bayesian learning to compute optimal portfolio allocation for an unknown drift and Gaussian asset returns. Recently, Bismuth *et al.* (2019) extended the previous results for optimal portfolio liquidation and transition problems in continuous time in which expected returns of risky assets are estimated online. De Franco *et al.* (2018) adapted the

methodology of Zhou and Li (2000) to the Bayesian learning framework and embedded the Bayesian-Markowitz problem into an auxiliary standard control problem and then applied dynamic programming approach. Bodnar *et al.* (2017) deal with the global minimum variance portfolio problem where the prior distribution of logarithmic returns are assumed to be normal and independent. They utilized various standard priors for mean vector and covariance matrix to derive posterior distribution for the weights. Recently, Bauder *et al.* (2018) derived posterior prediction distributions of returns to obtain optimal portfolio weights by assuming returns to be infinitely exchangeable and multivariate centered spherically symmetric for unknown mean vector and covariance matrix. In the literature mentioned above, asset returns are independent of past observations. However, empirical evidence has shown a serial correlation among the financial time series, which we captured in this paper using the vector autoregressive (VAR) model. Later, we applied linear Bayesian filtering approach to update the VAR model parameter sequentially and to obtain posterior prediction distribution.

The literature mentioned above makes an implication that the investment time horizon is deterministic, that is investor operates the investment strategy until the explicit exit time. However, due to market situations or an investor's personal reasons, he/she may be forced to leave the financial market before the exit time. In that case, the investment lifetime is uncertain and considered as a random variable. Martellini and Urošević (2006) maximize the quadratic expected utility function with uncertain exit time in which exit time depends on asset price behavior. Guo and Hu (2005) analyzed a multi-period mean-variance investment problem with an uncertain time of exiting. Later, some researchers Huang *et al.* (2008), Blanchet-Scalliet *et al.* (2008), Yi *et al.* (2008), Huang *et al.* (2010) extended the portfolio selection problem with stochastic time horizon by adopting efficient methodologies to make it more practical/realistic. Zhang and Li (2012) derived an analytical solution for a multi-period optimization problem with serially correlated returns and considered uncertain exit time as an exogenous random variable with a discrete probability distribution.

We consider the mean-variance portfolio optimization problem with uncertain exit time and serially correlated returns of multiple risky assets whose parameters are updated dynamically using Bayesian forecasting for dynamic linear models. We assume that the distribution of exit time is discrete and known, as it was in Zhang and Li (2012), and the returns of the risky asset are autocorrelated subject to normal VAR(p) process. Firstly, we convert the autoregressive model into a dynamic linear model (DLM) whose parameters are assumed to follow a normal distribution with mean and variance estimated from the past data. DLM has been used previously in financial optimization problems by Irie *et al.* (2019), where the returns are still considered to be independent over time. To sequentially update the distribution and to obtain a one-step-ahead forecast of returns and posterior distribution of VAR model parameters, we apply Bayesian techniques by West and Harrison (2006). We implement a dynamic programming approach for an uncertain time horizon to the updated returns to obtain the optimal strategy for our investment.

3.1 BAYESIAN FORECASTING

Prior research has found that the estimation of parameters using a Bayesian framework can improve the constructed portfolio's performance. In this work, we also take the Bayesian approach to account for estimating the risk of a portfolio in predicted stock returns. For details and proofs, refer to section 2.2. Here, we consider that the returns are serially correlated and fit autoregressive models to estimate the underlying time series of returns. The returns of the risky asset at time period $t (t = 1, ..., \mathrm{T})$ within the investment horizon is denoted by r_t. Suppose that r_t is a weakly stationary AR(p) series given by

$$r_t = \mu + \sum_{i=1}^{p} \phi_i (r_{t-i} - \mu) + \epsilon_t, \tag{3.1}$$

for some sequence of coefficients $\phi_1, \phi_2, ..., \phi_p$ where $\mathbb{E}[r_t|\mathcal{F}_t] = \mu$ and the ϵ_t are zero-mean, uncorrelated random quantities with constant variance σ^2.

Consider a dynamic linear model with r_t a T-vector based on T-vector θ_t via:

$$\mathbf{r}_t = \mathbf{F}_t'\theta_t + \nu_t, \qquad \nu_t \sim \mathrm{N}[0, \mathbf{V}_t]$$

$$\theta_t = \mathbf{G}_t\theta_{t-1} + \omega_t, \qquad \omega_t \sim \mathrm{N}[0, \mathbf{W}_t]$$

$$(\theta_0|\mathrm{D}_0) \sim \mathrm{N}[\mathbf{m}_0, \mathbf{C}_0]$$

for some prior moments $\mathbf{m}_0$ and $\mathbf{C}_0$ and where the ν_t, ω_t are independent and mutually independent innovations sequences. Initial information D_0 is the probabilistic representation about the level μ_0 at time $t = 0$. Each information set D_t comprises of all the information available at time t.

All autoregressive models can be written in dynamic linear form in a variety of ways for various purposes. Here, we are considering regression model form to write AR(p) model of equation (3.1) as a simple, static regression sequentially defined over time

$$r_t = \mathbf{F}_t'\theta + \nu_t,$$

where $\mathbf{F}_t' = (r_{t-1}, ..., r_{t-p})$, $\theta' = (\mu, \phi_1, ..., \phi_p)$ and $\nu_t = \epsilon_t$. Now the posterior distributions for AR parameters can be obtained by employing standard DLM results. Initial information to implement DLM form is as follows:

$$(\theta|\mathrm{D}_0) \sim \mathrm{N}[\mathbf{m}_0, \mathbf{C}_0].$$

where $\mathbf{m}_0 = \mathbb{E}[\theta|\mathrm{D}_0]$ and $\mathbf{C}_0 = \mathrm{Var}[\theta|\mathrm{D}_0]$.

Sequential learning of AR parameters can be obtained by employing standard DLM results. As the AR parameters θ is constant for a time period, posterior distribution of θ at time $t-1$ given by $(\theta|\mathrm{D}_{t-1}) \sim \mathrm{N}(m_{t-1}, \mathrm{C}_{t-1})$ concides with prior distribution at time t. Thus, prior distribution of θ at time t is $(\theta|\mathrm{D}_t) \sim \mathrm{N}(m_{t-1}, \mathrm{C}_{t-1})$.

One step ahead prediction distribution of the return series can be obtained from the prior

distribution of θ which would lead to

$$(r_t|\mathrm{D}_{t-1}) \sim \mathrm{N}(f_t, \mathrm{Q}_t)$$
$$\text{where, } f_t = \mathbf{F}_t'\mathbf{m}_{t-1},$$
$$\mathrm{Q}_t = \mathbf{F}_t'\mathbf{m}_{t-1}\mathbf{F}_t + \sigma^2.$$

Posterior distribution of θ at time t can be derived from the prediction distribution of the returns, which is:

$$(\theta_{t-1}|\mathrm{D}_{t-1}) \sim \mathrm{N}(m_t, \mathrm{C}_t)$$
$$\text{where, } \mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{A}_t(r_t - f_t),$$
$$\mathbf{C}_t = \mathbf{C}_{t-1} - \mathbf{A}_t\mathrm{Q}_t\mathbf{A}_t', \quad \mathbf{A}_t = \mathbf{C}_{t-1}\mathbf{F}_t\mathrm{Q}_t^{-1}.$$

As θ_t is constant for a time period so this posterior distribution can be used as prior to forecast the returns.

3.2 DYNAMIC PROGRAMMING

In a multi-period setting, an investor can reallocate the portfolio's weights at intermediate times after observing the values until that time. Due to unpredictability in the financial market, it is always more profitable to evolve the portfolio after observing the current market scenario. As defined by Iyengar (2005), Dynamic programming (DP) is the mathematical framework that permits the decision-maker to compute a decent overall strategy efficiently by evolving data state. In discrete-time portfolio optimization, the investor can rebalance his wealth at distinct points in time with the most recent information.

Consider that a capital market consisted of a risky security with random returns and a riskless security with deterministic returns. An investor with an initial wealth of x_0 joins the market at time 0 and allocates his/her wealth among these two securities. The

investor can reallocate his/her wealth invested in the risky asset at the beginning of each consecutive time period. The updated returns of the risky asset obtained after employing Bayesian forecasting at time period $t (t = 0, 1, 2, ..., \mathrm{T})$ are denoted by a vector $\tilde{r}_t$, which are considered as serially correlated, that is, the value of $\tilde{r}_t$ depends upon the past realized returns $\tilde{r}_s$, $s < t$. Let r_t^0 be the given deterministic return of the riskless security at period t. The uncertain exit time τ is considered as a discrete random variable and the actual exit time is $\mathrm{T} \wedge \tau = min\{\mathrm{T}, \tau\}$ with probability distribution $p_t = \mathbb{P}\{\mathrm{T} \wedge \tau = t\}, t = 1, 2, ...$

Define the excess return of a risky asset at period t as $e_t = \tilde{r}_t - r_t^0$. We use the updated return series to calculate the mean and mean square for excess returns for $t = 1, 2, ..., \mathrm{T}$

$$\begin{aligned}
\mathbb{E}[e_t] &= \mathbb{E}[\tilde{r}_t - r_t^0] = \mathbb{E}[\tilde{r}_t] - r_t^0 = f_t - r_t^0, \\
\mathbb{E}[e_t^2] &= \mathbb{E}[(\tilde{r}_t - r_t^0)^2] \\
&= \mathbb{E}[\tilde{r}_t^2] + (r_t^0)^2 - 2r_t^0\mathbb{E}[\tilde{r}_t] \\
&= \mathrm{Q}_t + f_t^2 + (r_t^0)^2 - 2r_t^0 f_t.
\end{aligned}$$

where $f_t = \mathbb{E}[e_t]$ is mean of excess return and $\mathrm{Q}_t = \mathbb{E}[e_t^2]$ is mean square of excess return. Let x_t be the wealth of the investor at the beginning of each time period $t (t = 0, 1, ..., \mathrm{T})$. Define investment series over T periods, $\mathrm{U} := \{u_0, u_1, ..., u_{\mathrm{T}-1}\}$, where u_t is the amount invested in the risky security at time t. The investment strategy U is assumed to be self-financing, that is, there is no exogenous inclusion or exclusion of money, which can be described mathematically as:

$$x_{t+1} = \tilde{r}_t^0 x_t + e_t u_t, \quad t = 0, 1, ..., \mathrm{T} - 1. \tag{3.2}$$

The multi-period MVPS for uncertain exit time can be written as follows

$$\mathrm{P}(\omega) \begin{cases} \max\limits_{\mathrm{U}} & \mathbb{E}_0(x_{\mathrm{T}\wedge\tau}) - \omega \mathrm{Var}_0(x_{\mathrm{T}\wedge\tau}) \\ \text{s.t.} & x_{t+1} = r_t^0 x_t + e_t u_t, \quad t = 0, 1, ..., \mathrm{T} - 1, \end{cases}$$

where ω is a given positive constant, representing the investor's risk aversion and illustrates the attitude of investors towards risk.

Due to nonlinearity of conditional variance, employing dynamic programming to this multi-period portfolio optimization is not straight forward. More precisely, for the expected value operator, dynamic programming is applicable because of the smoothing property: $\mathbb{E}[\mathbb{E}(\cdot|\mathcal{F}_j)|\mathcal{F}_k] = \mathbb{E}(\cdot|\mathcal{F}_k) \forall j > k$, where $\mathcal{F}_t$ denotes an information set available at time t and $\mathcal{F}_{t-1} \subset \mathcal{F}_t, \forall t$, while the variance operator does not satisfy the smoothing property: $\text{Var}[\text{Var}(\cdot|\mathcal{F}_j)|\mathcal{F}_k] \neq \text{Var}(\cdot|\mathcal{F}_k)\ \forall j > k$. Thus, Li and Ng (2000) transformed the original mean-variance equations into an auxiliary framework to find an analytical optimal solution which is described in the appendix. Here we present the optimal portfolio allocation strategy and efficient frontier for the original problem $\text{P}(\omega)$.

Theorem 3.1. *The optimal strategy of the mean-variance problem* $\text{P}(\omega)$ *is given by*

$$u_t^* = \frac{1 + 2\omega\lambda_0 x_0}{2\omega(1-\Theta)} \frac{\mathbb{E}_t(\lambda_{t+1} e_t)}{\mathbb{E}_t(\omega_{t+1} e_t^2)} - \frac{\mathbb{E}_t(\omega_{t+1} e_t)}{\mathbb{E}_t(\omega_{t+1} e_t^2)} r_t^0 x_t,$$

where

$$\begin{aligned}
\theta_t &= \frac{\mathbb{E}_t^2(\lambda_{t+1} e_t)}{\mathbb{E}_t(\omega_{t+1} e_t^2)}, \qquad \Theta = \sum_{t=1}^{\text{T}} \mathbb{E}_0(\theta_{t-1}), \\
\omega_t &= p_t + (r_t^0)^2 \left[\mathbb{E}_t(\omega_{t+1}) - \frac{\mathbb{E}_t^2(\omega_{t+1} e_t)}{\mathbb{E}_t(\omega_{t+1} e_t^2)} \right], \quad \omega_{\text{T}} = p_{\text{T}}, \\
\lambda_t &= p_t + r_t^0 \left[\mathbb{E}_t(\lambda_{t+1}) - \frac{\mathbb{E}_t(\omega_{t+1} e_t)\mathbb{E}_t(\lambda_{t+1} e_t)}{\mathbb{E}_t(\omega_{t+1} e_t^2)} \right], \quad \lambda_{\text{T}} = p_{\text{T}},
\end{aligned}$$

for $t = 0, 1, ..., \text{T} - 1$.

The efficient frontier of the original problem $\text{P}(\omega)$ *is given by*

$$\text{Var}_0(x_{\text{T}\wedge\tau}) = \frac{(1-\Theta)}{\Theta} \left[\mathbb{E}_0(x_{\text{T}\wedge\tau}) - \frac{\lambda_0 x_0}{1-\Theta} \right]^2 + \left[\omega_0 - \frac{\lambda_0^2}{1-\Theta} \right]^2 x_0^2, \tag{3.3}$$

for $\mathbb{E}_0(x_{\text{T}\wedge\tau}) \in \left[\frac{\lambda_0 x_0}{1-\Theta}, +\infty \right)$.

Note that ω_{t+1} and $\lambda_{t+1}$is known at time t, for $t = 0, 1, ..., \mathrm{T}-1$, so

$$\mathbb{E}_t^2(\omega_{t+1}e_t) = \omega_{t+1}\mathbb{E}_t^2(e_t) \text{ and } \mathbb{E}_t(\omega_{t+1}e_t^2) = \omega_{t+1}\mathbb{E}_t(e_t^2),$$

$$\mathbb{E}_t^2(\lambda_{t+1}e_t) = \lambda_{t+1}\mathbb{E}_t^2(e_t) \text{ and } \mathbb{E}_t(\lambda_{t+1}e_t^2) = \lambda_{t+1}\mathbb{E}_t(e_t^2).$$

3.3 GENERALIZED MODEL FOR MULTIPLE RISKY ASSETS

We extend the previous portfolio optimization problem to the general situation with multiple risky assets. We model the returns of the risky assets using a vector autoregressive model of order p (VAR(p)) to account for the serial correlation within the returns of a risky asset and the cross-correlation among the returns of different risky assets. Here, we consider n securities with random returns r_t^i $(i = 1, 2, \cdots, n)$ and $t = (1, 2, \cdots, \mathrm{T})$ and a riskless securities with deterministic returns r_t^0. The VAR(p) model series for the $\mathbf{R}_t = (r_t^1, \cdots, r_t^n)'$ is given by:

$$\mathbf{R}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\Phi}_i(\mathbf{R}_{t-j} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t \quad t = 1, \cdots, \mathrm{T},$$

for sequence of $(n \times n)$ coefficient matrices $\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_p$, where $\mathbb{E}[\mathbf{R}_t|\mathcal{F}_t] = \boldsymbol{\mu}$ and the $\boldsymbol{\epsilon}_t$ is a $(n \times 1)$ zero-mean white noise process with invariant covariance matrix $\not{\Sigma}$.

We write $\mathbf{R}_t = (r_t^1, \cdots, r_t^n)'$ an $(n \times 1)$ vector of time series into multivariate normal dynamic linear model (DLM) as:

$$\mathbf{R}_t = \boldsymbol{\Pi}'\mathbf{Z}_t + \boldsymbol{\epsilon}_t$$

where $\mathbf{Z}_t' = (1, \mathbf{R}_{t-1}', \cdots, \mathbf{R}_{t-p}')$, $\boldsymbol{\Pi} = [\pi_1, \cdots, \pi_n]$ and π_i $(i = 1, \cdots, n)$ is a $(k \times 1)$ vector of parameters where $k = np + 1$. Consider $\mathbf{Z}$ as a $(\mathrm{T} \times k)$ matrix with t^{th} row given by $\mathbf{Z}_t'$. We represent $\boldsymbol{\Pi}$ into a $(nk \times 1)$ vector denoted by $vec(\boldsymbol{\Pi})$. For stationary and ergodic VAR models, $vec(\boldsymbol{\Pi})$ is consistent and asymptotically normally distributed with asymptotic covariance matrix is kronecker product $(\otimes)$ of $\boldsymbol{\Sigma}$ and $(\mathbf{Z}'\mathbf{Z})^{-1}$, given as:

$$\widehat{avar}(vec(\boldsymbol{\Pi})) = \boldsymbol{\Sigma} \otimes (\mathbf{Z}'\mathbf{Z})^{-1}$$

where

$$\mathbf{\Sigma} = \frac{1}{\mathrm{T}-k}\sum_{t=1}^{\mathrm{T}} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \quad \text{and} \quad \boldsymbol{\epsilon}_t = \mathbf{R}_t - \mathbf{\Pi}'\mathbf{Z}_t.$$

Consider the initial information to implement DLM as follows:

$$(\mathbf{\Pi}|\mathrm{D}_0) \sim \mathcal{N}[\mathbf{m}_0, \mathbf{C}_0].$$

where $\mathbf{m}_0 = \mathbb{E}[\mathbf{\Pi}|\mathrm{D}_0] = \mathbf{\Pi}$ and $\mathbf{C}_0 = \mathrm{Var}[\mathbf{\Pi}|\mathrm{D}_0] = \widehat{avar}(vec(\mathbf{\Pi}))$.

Prediction distribution of the vector return series is:

$$(\mathbf{R}_t|\mathrm{D}_{t-1}) \sim \mathcal{N}[\mathbf{f}_t, \mathbf{Q}_t]$$

$$\text{where, } \mathbf{f}_t = \mathbf{Z}_t'\mathbf{m}_{t-1},$$

$$\mathbf{Q}_t = \tilde{\mathbf{Z}}_t'\mathbf{m}_{t-1}\tilde{\mathbf{Z}}_t + \mathbf{\Sigma} \quad \text{and} \quad \tilde{\mathbf{Z}}_t = \begin{bmatrix} \mathbf{Z}_t' & 0 \\ 0 & \mathbf{Z}_t' \end{bmatrix}.$$

Posterior distribution of $\mathbf{\Pi}$ at time t is:

$$(\mathbf{\Pi}|\mathrm{D}_t) \sim \mathcal{N}[\mathbf{m}_t, \mathbf{C}_t]$$

$$\text{where, } \mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{A}_t(\mathbf{R}_t - \mathbf{f}_t),$$

$$\mathbf{C}_t = \mathbf{C}_{t-1} - \mathbf{A}_t\mathbf{Q}_t\mathbf{A}_t' \quad \text{and} \quad \mathbf{A}_t = \mathbf{C}_{t-1}\mathbf{Z}_t'\mathbf{Q}_t^{-1}.$$

Let $\tilde{\mathbf{R}}_t = (\tilde{r}_t^1, \cdots, \tilde{r}_t^n)'$ denote the sequentially updated return series obtained from the prediction distribution. Define the excess return for i^{th} security at period $t+1$ as $e_t^i = \tilde{r}_t^i - r_t^0$, $\mathbf{e}_t = (e_t^1, \cdots, e_t^n)'$ and the investment strategy $\mathrm{U}_t = (u_t^1, \cdots, u_t^n)$ for $i = 1, 2, \cdots, n$ and $t = 0, 1, \cdots, \mathrm{T}-1$, where u_t^i is the amount invested in the i^{th} risky asset at time t.

The multi-period mean-variance portfolio selection for uncertain exit time with serial and cross correlation among the risky asset can be formulated as:

$$\mathrm{P}(\omega)\begin{cases} \max\limits_{\mathbf{U}} & \mathbb{E}_0(x_{\mathrm{T}\wedge\tau}) - \omega\mathrm{Var}_0(x_{\mathrm{T}\wedge\tau}) \\ \text{s.t.} & x_{t+1} = r_t^0 x_t + \mathbf{e}_t\mathrm{U}_t, \quad t = 0, 1, ..., \mathrm{T}-1, \end{cases}$$

where ω is a given positive constant, representing the investor's risk aversion. To obtain the optimal portfolio allocation strategy we have extension of theorem 1 to the situation with multiple assets.

Theorem 3.2. *The optimal strategy of the mean-variance problem* $\mathrm{P}(\omega)$ *is given by*

$$\mathrm{U}_t^* = \frac{1 + 2\omega\lambda_0 x_0}{2\omega(1-\Theta)}\mathbb{E}_t^{-1}(\omega_{t+1}\mathbf{e}_t\mathbf{e}_t')\mathbb{E}_t(\lambda_{t+1}\mathbf{e}_t) - \mathbb{E}_t^{-1}(\omega_{t+1}\mathbf{e}_t\mathbf{e}_t')\mathbb{E}_t(\omega_{t+1}\mathbf{e}_t)r_t^0 x_t,$$

where

$$\xi_t = \mathbb{E}_t(\lambda_{t+1}\mathbf{e}_t)\mathbb{E}_t^{-1}(\omega_{t+1}\mathbf{e}_t\mathbf{e}_t')\mathbb{E}_t(\lambda_{t+1}\mathbf{e}_t), \qquad \Theta = \sum_{t=1}^{\mathrm{T}}\mathbb{E}_0(\xi_{t-1}),$$

$$\omega_t = p_t + (r_t^0)^2\left[\mathbb{E}_t(\omega_{t+1}) - \mathbb{E}_t(\omega_{t+1}\mathbf{e}_t)\mathbb{E}_t^{-1}(\omega_{t+1}\mathbf{e}_t\mathbf{e}_t')\mathbb{E}_t(\omega_{t+1}\mathbf{e}_t)\right], \quad \omega_{\mathrm{T}} = p_{\mathrm{T}},$$

$$\lambda_t = p_t + r_t^0\left[\mathbb{E}_t(\lambda_{t+1}) - \mathbb{E}_t(\lambda_{t+1}\mathbf{e}_t)\mathbb{E}_t^{-1}(\omega_{t+1}\mathbf{e}_t\mathbf{e}_t')\mathbb{E}_t(\omega_{t+1}\mathbf{e}_t)\right], \quad \lambda_{\mathrm{T}} = p_{\mathrm{T}},$$

for $t = 0, 1, ..., \mathrm{T}-1$.

The efficient frontier of the original problem $\mathrm{P}(\omega)$ *is given by*

$$\mathrm{Var}_0(x_{\mathrm{T}\wedge\tau}) = \frac{(1-\Theta)}{\Theta}\left[\mathbb{E}_0(x_{\mathrm{T}\wedge\tau}) - \frac{\lambda_0 x_0}{1-\Theta}\right]^2 + \left[\omega_0 - \frac{\lambda_0^2}{1-\Theta}\right]^2 x_0^2,$$

for $\mathbb{E}_0(x_{\mathrm{T}\wedge\tau}) \in \left[\frac{\lambda_0 x_0}{1-\Theta}, +\infty\right)$.

3.4 SIMULATIONS

In this section, we consider a fixed time series model for asset returns to study the effect of model parameters on the efficient frontier. We have developed the theory for VAR(p) models for multiple assets. Here we restrict to $p = 1$ for a single asset to better interpret the results. Consider the AR(1) model for returns:

$$(r_t - \mu) = \phi(r_{t-1} - \mu) + \epsilon_t \tag{3.4}$$

Here, ϕ captures the autocorrelation in the return series, μ is the unconditional expectation of r_t and ϵ_t is a random variable having normal distribution with mean 0, variance σ^2

and ϵ_t is independent of $\epsilon_s (s < t)$. We investigated the effect of each of these three AR model parameters on the return series.

Keeping deterministic and random factors constant, we simulated 100 samples of the model mentioned above for ϕ ranging from -0.9 to 0.9. Table 3.1 presents the probability when sequentially updating using the Bayesian technique performs better for the range of values of ϕ. For $|\phi|$ close to 1, the autoregressive model breaks down due to non-stationarity. In this case, to capture the dependence, we need to take into account the integrating effect. Bayesian updating performs best for $\phi \in [-0.4, 0.4]$ as the serial correlation among asset returns is adequately captured. However, when $\phi = 0$, the probability is quite low as there is no autocorrelation among asset returns and both the methods perform equivalently.

Another model parameter is μ, which is deterministic in nature. Table 3.2 shows the probability when dynamic programming with updates perform better for different values of deterministic factor μ and fixing other parameters. As the absolute value of deterministic factor $|\mu|$ increases, the relative contribution of the dependent part decreases, and both the methods are reasonably comparable. But for small values of $|\mu|$, the effect of dependence factor ϕ dominates and leads to better performance of the proposed method at the same level of variance.

Finally, we fix deterministic and dependence factors and vary the variance of random part i.e., ϵ_t. From Table 3.3, it is clear that both the models perform adequately for a lower variance of ϵ_t, as in lesser variance, there is not much gain in updating the parameters. However, for higher values of variance our algorithm always performs better.

Table 3.1: Probability that Bayesian updates perform better for different values of ϕ (Fix $\mu = 0.01$ and $\sigma^2 = 0.2$)

Value of ϕ	Probability	Value of ϕ	Probability
-0.9	0.06	0.1	1
-0.8	0.13	0.2	1
-0.7	0.17	0.3	0.99
-0.6	0.36	0.4	0.82
-0.5	0.54	0.5	0.53
-0.4	0.83	0.6	0.38
-0.3	0.98	0.7	0.25
-0.2	0.99	0.8	0.05
-0.1	0.99	0.9	0.01
0	0.17		

Table 3.2: Probability that Bayesian updates perform better for different values of μ (Fix $\phi = 0.1$ and $\sigma^2 = 1$)

Value of μ	Probability	Value of μ	Probability
-0.5	0.07	0.01	1
-0.4	0.32	0.03	1
-0.3	0.43	0.05	0.95
-0.2	0.54	0.1	0.83
-0.1	0.78	0.2	0.80
-0.05	0.99	0.3	0.51
-0.03	0.99	0.4	0.43
-0.01	1	0.5	0.09
0	1		

Table 3.3: Probability that Bayesian updates perform better for different values of σ^2 (Fix $\phi = 0.1$ and $\mu = 0.01$)

Value of σ^2	Probability
0.05	0.74
0.1	0.80
0.15	0.96
0.2	0.99
0.25	1

3.5 EMPIRICAL ILLUSTRATION

3.5.1 Data

For an empirical illustration, we apply the modeling and optimization framework to the US stock market index S&P 500 and five stocks from it based on market capitalization which include Apple Inc.(AAPL), Microsoft Corporation(MFST), Amazon.com Inc.(AMZN), Facebook Inc. Class A(FB) and Alphabet Inc. Class A(GOOGL), corresponding to one year up to two years. An asset weekly returns r_t is given by (P_t/P_{t-1}) where P_t denotes the asset price at time t. We downloaded our data from Yahoo finance in USD. The time period of the return data is 156 weeks from July 2017 to June 2020. We used an initial period of 130 weeks of this data for estimating the parameters of the autoregressive process, followed by sequential updating of returns using the remaining data set by applying the dynamic linear model.

3.5.2 Results

In this section, we provide a numerical example to demonstrate the impacts of the uncertainty of exit time and the serial correlations of returns on the efficient frontier. To illustrate the results, we have considered a single risky asset S&P 500. We assume that the return rate r_t is subject to AR(p) model. The order of the AR model is estimated by taking different combinations for the lag values, ranging from one to five, and selecting the most appropriate based on the AIC values. The results show that the return rate is well-captured by AR(1) model given in equation (3.4).

To illustrate the effect of Bayesian updating in multiperiod, we compare the efficient frontiers of mean-variance portfolio optimization problem with and without the dynamically updating the returns. For the above mentioned data, AR parameters estimate for initial time $t = 0$ are obtained as: $\mu_0 = 0.00207$, $\phi_0 = -0.13366$ and $\sigma_0^2 = 0.00085$. Other parameters are set as: initial wealth $x_0 = 1$, length of investment time T = 26, i.e. 26 weeks, riskless asset return $r_t^0 = 0.0057$ (risk-free rate for USA) for

$t = 1, 2, ..., \mathrm{T}$ and exit time distribution is given as:

$$\mathbb{P}\{\mathrm{T} \wedge \tau = t\} = \begin{cases} 0.001 & t = 1, ..., \mathrm{T}-1 \\ 1 - 0.001(\mathrm{T}-1) & t = \mathrm{T} \end{cases}$$

Figure 3.1 illustrates that it is beneficial for the investor to rebalance the portfolio by looking at the current prices and subsequently update the model parameters. We also compared the efficient frontiers of a dynamic mean-variance optimization problem with Bayesian updates under different exit time and different exit time distribution. Probability distributions of uncertain exit time $t = \mathrm{T} \wedge \tau$ by varying spread considered for the study are

$$\mathbb{P}_1\{\mathrm{T} \wedge \tau = t\} = \begin{cases} 0.001 & t = 1, ..., 25 \\ 0.975 & t = 26 \end{cases}$$

$$\mathbb{P}_2\{\mathrm{T} \wedge \tau = t\} = \begin{cases} 0.001 & t = 1, ..., 24 \\ 0.3 & t = 25 \\ 0.676 & t = 26 \end{cases}$$

$$\mathbb{P}_3\{\mathrm{T} \wedge \tau = t\} = \begin{cases} 0.001 & t = 1, ..., 15 \\ 0.05 & t = 16, ..., 25 \\ 0.485 & t = 26 \end{cases}$$

The broader spread of exit time distribution offers investors more number of opportunities to exit the market and earn the maximum gain when the market is in his favor, as illustrated in Figure 3.2. Figure 3.3 compares the efficient frontier with different exit time. Figure 3.3 implies that the investor who spends more time in the market, enjoys more expected wealth at the same level of risk than the one with a higher probability of leaving the market at an earlier stage.

3.5.3 Credible Interval Prediction

In the case of the Bayesian approach, we have the whole prediction distribution of the asset returns instead of just point estimates. Using this prediction distribution, we can

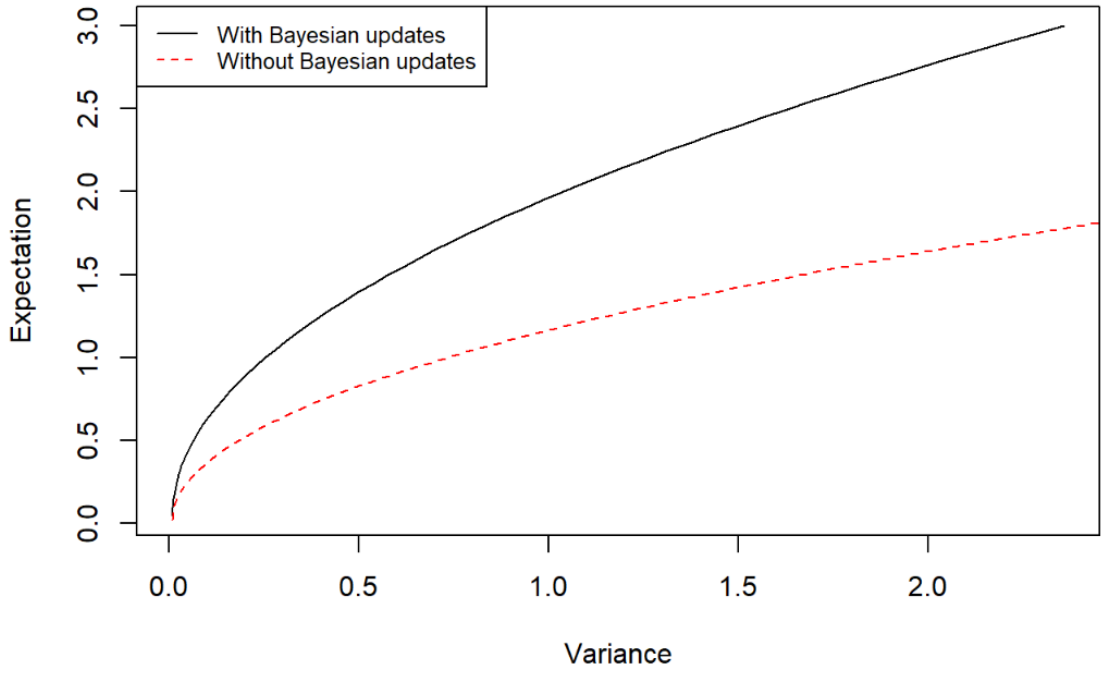


Figure 3.1: Efficient Frontiers of updated and non-updated model

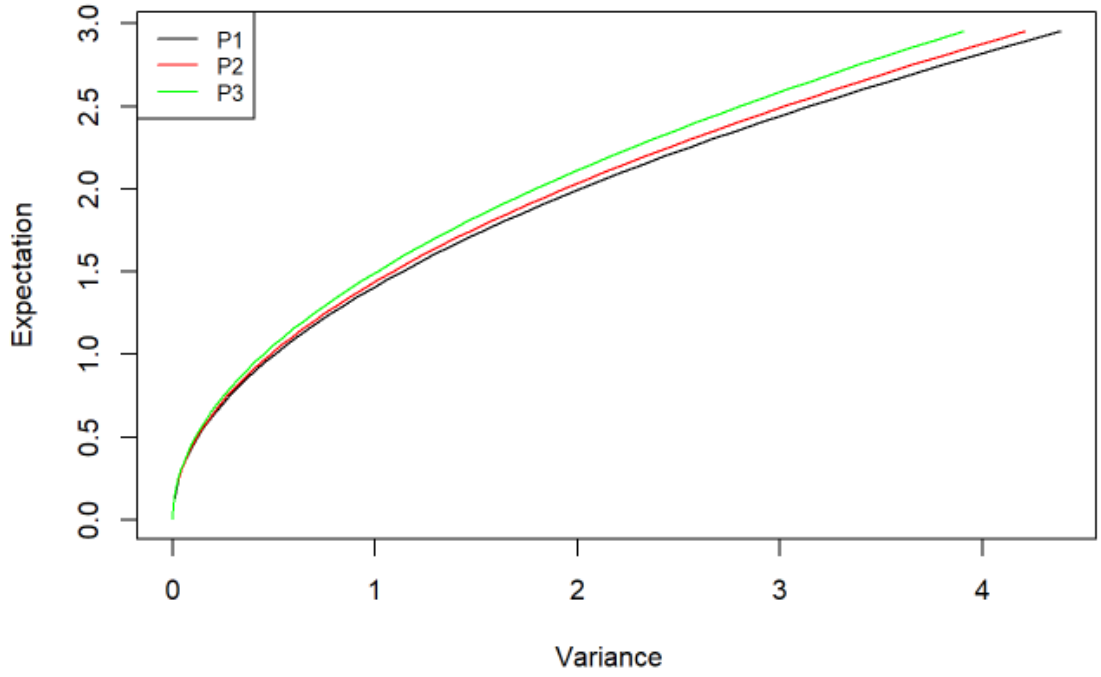


Figure 3.2: Efficient frontiers with different exit time distribution

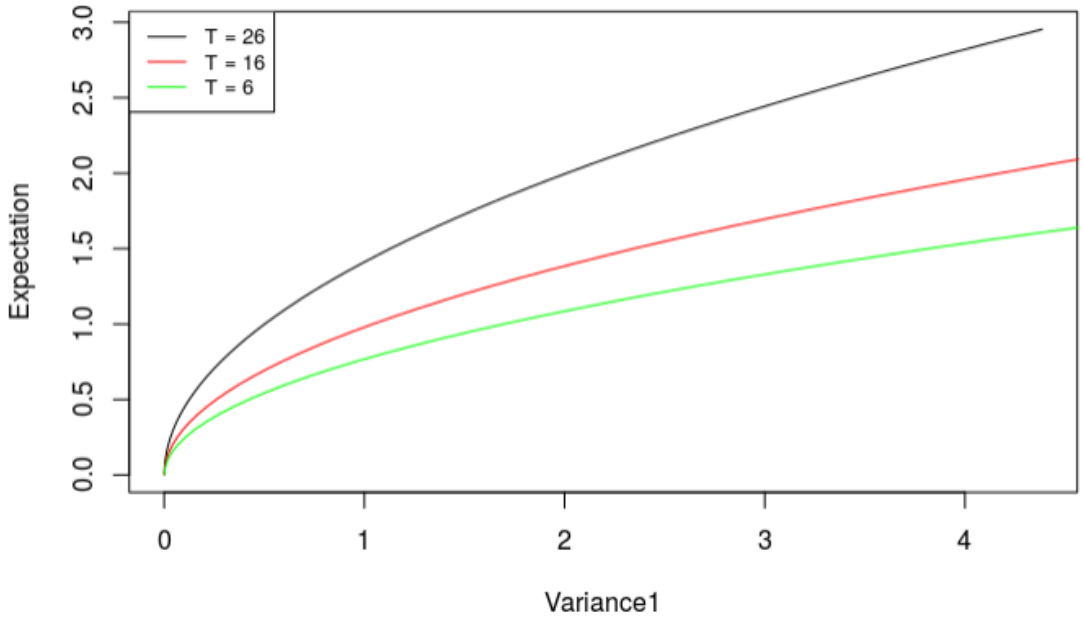


Figure 3.3: Efficient frontiers with different exit time

form prediction intervals for the expected terminal wealth. To obtain the prediction intervals, we simulate the value of optimal portfolio return for several values of risk tolerance factor $\omega \in (0, 3)$ (see Figure 3.5).

The prediction intervals for the whole range of ω are obtained as follows:

(i) Firstly, obtain the initial information for the AR parameters from the given data set. Using these initial parameters and the recent data, generate the AR(p) model for asset returns r_t^i as described in (3.1) for $t = 1, 2, ..., \mathrm{T}$ and $i = 1, ..., 1000$.

(ii) Fix ω and for $k \in \{1, ..., 1000\}$ calculate:

(a) the mean and mean square for excess returns for $t = 1, 2, ..., \mathrm{T}$ as follows

$$\mathbb{E}[e_t^{(k)}] = m_0 r_t^{(k)} - r_t^0$$

$$\mathbb{E}[(e_t^{(k)})^2] = (r_t^0)^2 + m_0^2 (r_t^{(k)})^2 - 2m_0 r_t^{(k)} r_t^0 + \sigma_t^2$$

(b) the parameters $p_t, \omega_t, \lambda_t, \theta_t$ and Θ for $t = 1, 2, ..., \mathrm{T}$ using theorem 1.

(c) λ^* as given (A.3) and substitute it in (A.2) to obtain optimal portfolio return $r^{*(k)}$.

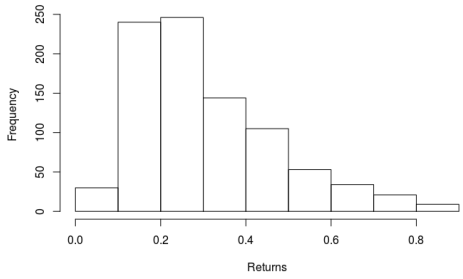


Figure 3.4: Histogram of empirical distribution of optimal portfolio returns for $\omega = 3$

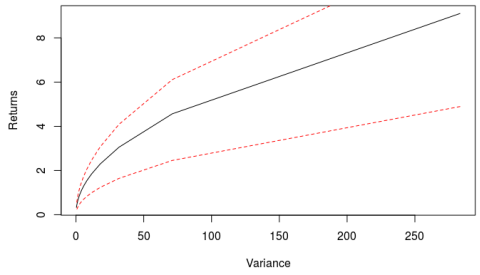


Figure 3.5: Central 50% credible intervals for the return of optimal portfolios

(iii) Obtain the empirical distribution of the optimal portfolio return using these simulated values $r^{*(k)}$, $k = 1, \cdots, 1000$, which is represented by the histogram (see Figure 3.4).

(iv) Fix the significance level of the prediction interval as α and compute the expectation $\mathrm{E}_\omega[r^*]$, $\alpha/2-$ and $(1-\alpha/2)-$ quantiles from the empirical distribution.

(v) Use the expectation of optimal portfolio returns to calculate the variance of terminal wealth $\mathrm{Var}_\omega[r^*]$ using (3.3).

(vi) Repeat steps (ii)-(v) for $\omega \in (0, 3]$ to obtain the expectation, variance and intervals for the whole range of ω. Plot the expectation along with prediction intervals from (iv) for the computed values of variance in part (v).

A portfolio with a higher risk aversion coefficient is less risky and therefore lies further left on the efficient frontier. Thus, the smaller ω will have larger credible interval.

3.5.4 Results for Multiple Risky Assets

To demonstrate the impacts of the serial and cross-correlations among returns of various assets on the efficient frontier, we consider five stocks mentioned in Sect. 3.5.1 for our portfolio. We assume that the return rate matrix $\mathbf{R}_t$ is subject to VAR(p) model. Based on the AIC values VAR(1) model well-capture the interdependencies among the chosen stocks. Thus, the model we obtained is as follows:

$$\mathbf{R}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{R}_{t-j} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t \quad t = 1, \cdots, \mathrm{T},$$

where $\mathbb{E}[\mathbf{R}_t|\mathcal{F}_t] = \boldsymbol{\mu}$, $\boldsymbol{\Phi}$ is a (5×5) coefficient matrices and the $\boldsymbol{\epsilon}_t$ is a (5×1) zero-mean white noise process with invariant covariance matrix $\boldsymbol{\Sigma}$. VAR parameters estimate for initial time $t = 0$ are obtained as:

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 0.006632 \\ -0.01047 \\ 0.00704 \\ 0.00245 \\ 0.00299 \end{bmatrix} \text{ and } \boldsymbol{\Phi}_0 = \begin{bmatrix} 0.0399 & 0.2395 & -0.0185 & 0.0322 & 0.1183 \\ 0.0160 & -0.1943 & -0.0006 & -0.0020 & 0.0093 \\ 0.0684 & -1.1354 & 0.0137 & 0.1444 & -0.0398 \\ -0.0618 & 2.7002 & 0.0131 & -0.1892 & -0.0411 \\ -0.0882 & -2.1994 & 0.0667 & 0.0850 & -0.0589 \end{bmatrix}.$$

Other parameters are similar to the single risky asset case. Figure 3.6 illustrates that our model with Bayesian forecasting captures the cross-correlation among assets and reduces the risk for the investors. We also observe that the variance is quite notable in the case of multiple assets model. The reason for this significant variance is low diversification provided by only five stocks compared to S&P 500. Our portfolio consists of "Big Five" American technology companies belonging to the same industry leads to a further reduction in diversification.

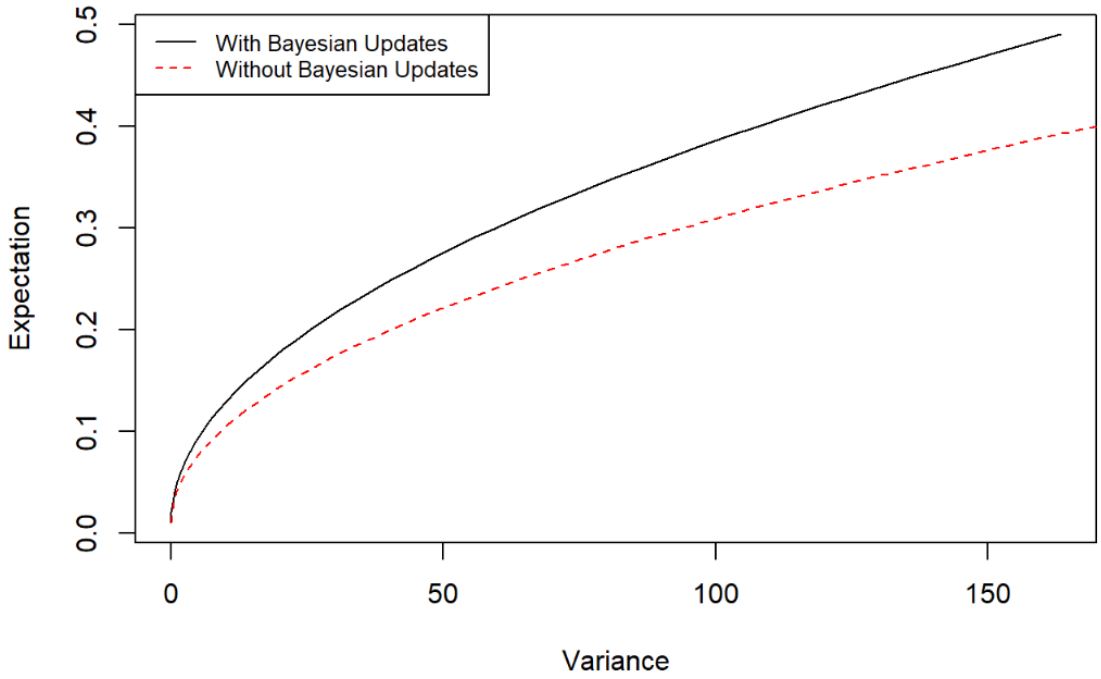


Figure 3.6: Efficient frontiers for multiple assets of Bayesian updated and non-updated model

CHAPTER 4

BAYESIAN MODELLING AND FORECASTING FOR YIELD CURVES

The yield curve is an instrument for understanding the state of the bond market. Yield curves also called the term structure of interest rates, represent the relationship between bond yield and their term to maturity. Investors closely watch the term structure of interest rates as it enables them to understand the dynamics of future economic activities and predict the inflation level. In recent decades, significant progress has been seen in modeling yield curves employing the dynamic Nelson-Siegel factor model and its extensions (see the ref. Diebold and Li (2006), Koopman *et al.* (2010), Carriero *et al.* (2012)). However, yield curve data can be viewed as functional data. This approach avoids putting a parametric form on the functions and relies only on smoothness. As a result, this paper proposes modeling yield curves using multivariate functional data analysis. Figure 4.1 represents the term structure of USA interest rates for the year 2019.

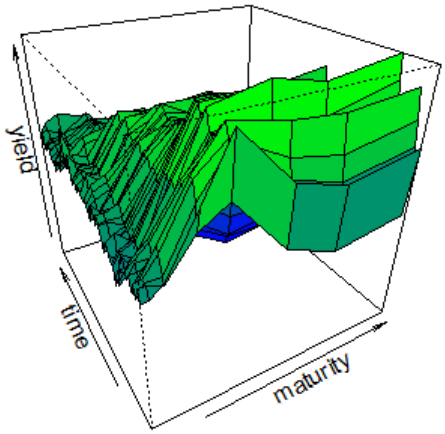


Figure 4.1: Raw data of interest rates (USA). Year: 2019. x-axis: Day, y-axis: Maturity in months, z-axis: Interest rate.

A functional time series arises when at each point in time we observe a random function,

generally at discrete points on a dense support. Examples of functions include satellite images (Guyet and Nicolas (2016)), fMRI data(Stoehr *et al.* (2021)) and yield curves (Sen and Klüppelberg (2019)). Another common practice that leads to functional time series is to slice high-frequency temporal sequence records into curves over logical consecutive time intervals. Examples of such curve time series include the daily price curve of financial transactions (Horváth *et al.* (2014), Li *et al.* (2020*b*)), electricity price curves as a function of day-of-year (Chen and Li (2017), Chen *et al.* (2019)), daily pollution levels as a function of the time of the day (Aue *et al.* (2015)), and daily patterns of environmental data (Besse *et al.* (2000), Ramsay and Dalzell (1991)). The functional autoregressive (FAR) model, developed by Bosq (2000), plays a central role in modeling and predicting functional time series. The FAR model extends the vector autoregressive (VAR) model to the infinite-dimensional space and considers the serial correlation between the functional observations.

Most of the existing literature relies on functional principal component analysis (FPCA) to treat multivariate functional data. In a multivariate setting, along with serial correlation, one needs to study the correlation among the data series. Ramsay and Silverman (2005) introduced a classical multivariate FPCA that concatenates the multiple functional data and performs univariate FPCA on a single long vector. However, this model is restricted to multivariate random functions on the same finite, one-dimensional interval. Chiou *et al.* (2014) discussed a normalized multivariate functional principal component analysis (mFPCA$_n$) which accounts for data measured in different units using normalized covariance operators. Although mFPCA$_n$ produces sound estimates on dense grids without calculation errors, it leads to inadmissible results with significant measurement errors in the case of sparse settings.

Apart from FPCA, there is abundant literature that models multivariate functional data using regression analysis. Zhu *et al.* (2012) developed a multivariate varying coefficients model (MVCM) to estimate covariance function. They also establish the uniform

convergence rate of this covariance function. Li *et al.* (2017) extended the previous varying-coefficient single index model to perform the regression on functional response data. Zhu *et al.* (2017) performed linear transformation and adopted the Bayesian approach to sample the model parameters in the transformed space. Zhu *et al.* (2016) introduced a notion of graphical models in multivariate functional data by constructing Bayesian inference of undirected, decomposable graphs based on Markov distribution. The literature mentioned above mainly focuses on dense functional data. Li *et al.* (2020*a*) adopted the fast covariance-based FPCA method for multivariate sparse functional data, approximating covariance functions using Tensor-product B-splines. The underlying assumption while using FPCA is that the observable process with measurement errors follows the multivariate functional autoregressive (MFAR) process. We later prove that such an observable process follows a multivariate functional autoregressive moving average (FARMA) process. Thus, existing approaches based on FPCA produce inefficient estimates.

This paper proposes a hierarchical model for multivariate functional time series, which is the extension of Kowal *et al.* (2019) to the multivariate setting. This model will overcome the issue of inconsistent solutions arising in the FPCA approach and produce efficient results in a sparse setting. To estimate the MFAR model, we firstly form a two-level hierarchical model using the observed values. For computation, we later represent the hierarchical model as a dynamic linear model (DLM). We use observation equations to discretize the functional data and obtain measurement error. The evolution equations take into account the correlation among the functional time series and define the process model. The latent process in our case is the multivariate functional autoregressive model of order p, written as MFAR(p). A dynamic functional factor model further specifies the dynamic innovation process. We employ an efficient Gibbs sampling algorithm to perform inference and prediction for multivariate time series.

4.1 THE MODEL

Suppose we have K functional time series defined as $\{\mathrm{Y}_t^n\}$, $n = 1, \cdots, \mathrm{K}$ on some compact index set $\mathcal{T} \subset \mathbb{R}^{\mathrm{D}}$, typically $\mathrm{D} = 1$. Consider $\{\mathbf{Y}_t\}$ as K−variate functional time series represented as: $\mathbf{Y}_t = [\mathrm{Y}_t^1, \cdots, \mathrm{Y}_t^{\mathrm{K}}]'$ where $\mathrm{Y}_t^n \in \mathrm{L}^2(\mathcal{T})$ for $n = 1, 2, \cdots, \mathrm{K}$.

4.1.1 Multivariate Functional Autoregressive Model

Multivariate functional autoregressive model of order p, written as MFAR(p):

$$\mathbf{Y}_t - \boldsymbol{\mu} = \sum_{l=1}^{p} \boldsymbol{\Psi}_l(\mathbf{Y}_{t-l} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t. \tag{4.1}$$

where $\boldsymbol{\mu} = [\mu^1, \mu^2, \cdots, \mu^{\mathrm{K}}]'$ is mean of $\mathbf{Y}_t$ under stationarity, $\boldsymbol{\epsilon}_t = [\epsilon_t^1, \epsilon_t^2, \cdots, \epsilon_t^{\mathrm{K}}]'$ is a vector of innovation functions with each ϵ_t^n being a sequence of independent mean zero random functions with $\mathbb{E} \mid\mid \epsilon_t^n \mid\mid^2 < \infty$ for $n = 1, \cdots, \mathrm{K}$ and $\boldsymbol{\Psi}_l = [\Psi_l^{i,j}]_{i,j=1}^{\mathrm{K}}$ where each $\Psi_l^{i,j}$ is a bounded linear operator on $\mathrm{L}^2(\mathcal{T})$ for $i, j = 1, \cdots, \mathrm{K}$ and $l = 1, 2, \cdots, p$.

We assume that each function $\mathrm{Y}_t^n(\tau^n)$, $n = 1, \cdots, \mathrm{K}$ is observed at unequally-spaced points $\tau^n \in \mathcal{T}$. We restrict to $p = 1$ and consider only integral operators for better interpretability and computational convenience. In practice, the functional observations $\{\mathrm{Y}_t\}$ are observed via discrete samples of each curve with some measurement error. Suppose we observe $y_{i,t}^n \in \mathbb{R}$ sampled with measurement error $v_{i,t}^n$ from $\{\mathrm{Y}_t^n\} \in \mathrm{L}^2(\mathcal{T})$, $n = 1, \cdots, \mathrm{K}$:

$$y_{i,t}^n = \mathrm{Y}_t^n(\tau_{i,t}^n) + v_{i,t}^n \tag{4.2}$$

where $\tau_{i,t}^n$ for $i = 1, \cdots, m_t^n$ are observation points of Y_t^n and $v_{i,t}^n$ is a mean zero measurement error with finite variance. Assume $\boldsymbol{\alpha}_t = \mathbf{Y}_t - \boldsymbol{\mu}$, then two level hierarchical model is given as:

$$\begin{aligned} y_{i,t}^n &= \mu^n(\tau_{i,t}^n) + \alpha_t^n(\tau_{i,t}^n) + v_{i,t}^n, && i = 1, \cdots, m_t^n, \\ \alpha_t^n &= \sum_{m=1}^{\mathrm{K}} \int \psi^{n,m}(\tau^n, u)\alpha_{t-1}^m(u)du + \epsilon_t^n(\tau^n), && \forall \tau \in \mathcal{T} \text{ and } n = 1, \cdots, \mathrm{K}, \end{aligned} \tag{4.3}$$

for $t = 2, \cdots, \mathrm{T}$, where $\psi^{n,m}$ are integral operators and we assume that $\{\boldsymbol{v}_{i,t}\}$ and $\{\boldsymbol{\epsilon}_t\}$ are mutually independent.

The presence of measurement error in model (4.3) increases the estimation error of $\mathbf{\Psi}$ and produces inefficient forecasts. The following proposition illustrates that MFAR(p) model for the observables is inappropriate.

Proposition 1: Let $\mathbf{Y}_t - \boldsymbol{\mu} = \sum_{l=1}^{p} \mathbf{\Psi}(\mathbf{Y}_{t-l} + \boldsymbol{\epsilon}_t)$ and suppose we observe $\boldsymbol{y}_t = \mathbf{Y}_t + \boldsymbol{v}_t$ are independent white noise processes. Then the observable process $\{\boldsymbol{y}_t\}$ follows a multivariate functional autoregressive moving average (FARMA) process of order (p, p).

Proof. See Kowal *et al.* (2019). ■

To overcome this model mis-specification, we separate the observed data into functional process and measurement errors in the dynamic linear model (DLM).

4.1.2 Dynamic Linear Models for MFAR (p)

For practical implementation of model (4.3), we select a finite set of evaluation points for each of functional time series, $\mathcal{T}_e^n = \{\tau_1^n, \cdots, \tau_{\mathrm{M}}^n\} \subset \mathcal{T}$ for $n = 1, \cdots, \mathrm{K}$. By selecting large M and $\mathcal{T}_e^k$ dense in $\mathcal{T}$, we can accurately approximate the integral in (4.3) using quadrature methods:

$$\int \psi^{n,m}(\tau^n, u)\alpha_{t-1}^n(u)du \approx (\psi^{n,m}(\tau^n, \tau_1^m), \cdots, \psi^{n,m}(\tau^n, \tau_{\mathrm{M}}^m))\mathrm{Q}^{n,m}(\alpha_{t-1}^n(\tau_1^m), \cdots, \alpha_{t-1}^n(\tau_{\mathrm{M}}^m))',$$
$$n, m = 1, \cdots, \mathrm{K},$$

where $\mathrm{Q}^{n,m}$ is a known quadrature weight matrix. Let Z_t^n be $m_t^n \times \mathrm{M}$, $n = 1, \cdots, \mathrm{K}$, incidence matrix that identifies the observation points observed at time t for n-th functional times series. We can formulate the dynamic linear model (DLM) corresponding to the hierarchical model (4.3) as:

$$\begin{aligned}
&\boldsymbol{y}_t = \mathbf{Z}_t\boldsymbol{\mu} + \mathbf{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{v}_t, \qquad [\boldsymbol{v}_t \mid \Sigma_v] \overset{indep}{\sim} N(\mathbf{0}, \Sigma_v) \text{ for } t = 1, \cdots, \mathrm{T},\\
&\boldsymbol{\alpha}_t = \boldsymbol{\psi}\mathbf{Q}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\epsilon}_t, \qquad [\boldsymbol{\epsilon}_t \mid \mathbf{K}_\epsilon] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{K}_\epsilon) \text{ for } t = 2, \cdots, \mathrm{T},\\
&\boldsymbol{\alpha}_1 \sim N(\mathbf{0}, \mathbf{K}_\epsilon)
\end{aligned} \tag{4.4}$$

where $\boldsymbol{y}_t = [y_{i,t}^n]$, is a discrete sample of K-variate functional time series $\{\mathbf{Y}_t\}$ for $i = 1, \cdots, m_t^n$ and $n = 1, \cdots, \mathrm{K}$. Here, $\boldsymbol{\mu} = [\mu^1(\tau^1), \cdots, \mu^{\mathrm{K}}(\tau^{\mathrm{K}})]'$ and $\boldsymbol{\alpha}_t = [\alpha_t^1(\tau^1), \cdots, \alpha_t^{\mathrm{K}}(\tau^{\mathrm{K}})]'$ where $\mu^i(\tau^i) = [\mu^i(\tau_1^i), \cdots, \mu^i(\tau_{\mathrm{M}}^i)]$ and $\alpha_t^i(\tau^i) = [\alpha_t^i(\tau_1^i), \cdots, \alpha_t^i(\tau_{\mathrm{M}}^i)]$ for $i = 1, \cdots, \mathrm{K}$. Other matrices are defined as:

$$\mathbf{Z}_t = \begin{bmatrix} \mathrm{Z}_t^1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathrm{Z}_t^{\mathrm{K}} \end{bmatrix}, \ \boldsymbol{\psi} = \begin{bmatrix} \{\psi^{1,1}(\tau_i, \tau_j)\}_{i,j=1}^{\mathrm{M}} & \cdots & \{\psi^{1,\mathrm{K}}(\tau_i, \tau_j)\}_{i,j=1}^{\mathrm{M}} \\ \vdots & \ddots & \vdots \\ \{\psi^{1,\mathrm{K}}(\tau_i, \tau_j)\}_{i,j=1}^{\mathrm{M}} & \cdots & \{\psi^{\mathrm{K},1}(\tau_i, \tau_j)\}_{i,j=1}^{\mathrm{M}} \end{bmatrix}$$

$$\text{and } \mathbf{Q} = \begin{bmatrix} \mathrm{Q}^{1,1} & \cdots & \mathrm{Q}^{1,\mathrm{K}} \\ \vdots & \ddots & \vdots \\ \mathrm{Q}^{1,\mathrm{K}} & \cdots & \mathrm{Q}^{\mathrm{K},1} \end{bmatrix}.$$

Model (4.4) can be extended to multiple lags to the MFAR(p) by replacing the evolution equation with $\boldsymbol{\alpha}_t = \sum_{l=1}^{p} \boldsymbol{\psi}_l \mathbf{Q} \boldsymbol{\alpha}_{t-l} + \boldsymbol{\epsilon}_t$.

4.1.3 Theoretical Results

Let $\mathcal{F}_t$ denotes the information set available at time t. Thus, $\mathcal{F}_t = \{\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \cdots, \boldsymbol{y}_1\} \cup \mathcal{F}_0$, where $\mathcal{F}_0$ is the information prior at $t = 1$.

Theorem 4.1. *For any finite set of evaluation points $\mathcal{T}_e \subset \mathcal{T}$, the unique best linear predictor of the conditional random matrix $\boldsymbol{\delta} \sim [\boldsymbol{\delta}|\mathbf{Y}, \boldsymbol{\Theta}]$, where $\boldsymbol{\delta}, \mathbf{Y} \subset \mathcal{F}_{\mathrm{T}} \cup \{\boldsymbol{\alpha}_t(\tau) : \tau \in \mathcal{T}_e, t = 1, \cdots, \mathrm{T}\}$ and $\Theta = \{\boldsymbol{\mu}, \Sigma_v, \boldsymbol{\psi}, \mathbf{K}_\epsilon\}$, under the risk $\mathcal{R}_e$ and conditional on model (4.3) with integral approximation, is the conditional expectation $\hat{\boldsymbol{\delta}}(\mathbf{Y}|\boldsymbol{\Theta}) = \mathbb{E}[\boldsymbol{\delta}|\mathbf{Y}, \boldsymbol{\Theta}]$ as computed in model (4.4).*

Corollary 4.1.1. *The unique best linear predictor of $[\boldsymbol{\alpha}_t(\tau^*)|\mathcal{F}_s]$ for any time t, s and any point $\tau^* \in \mathcal{T}$ is the corresponding expectation under model (4.4).*

4.2 FORECASTING MFAR(P)

Here, we will present the algorithm for forecasting multivariate functional autoregression model of order p. Firstly, we will initialize the DLM (4.4) parameters and later we

propose the Gibbs sampling algorithm for model (4.4), which is an extension of Gibbs sampling algorithm for univariate functional autoregression model of order p Kowal *et al.* (2019).

4.2.1 Initialization

Step 1: We firstly initialize $\mu = [\mu_Y \ \mu_X]'$ as smooth mean of $[\{y_t^n\} \ \{x_t^n\}]_{t=1}^{'T}$ and fit a spline to each $[Y_t - \mu_Y \, X_t - \mu_X]'$ to estimate $\boldsymbol{\alpha}_t$. Thus, we obtain the estimates of $\mathbb{E}[Y_t \, X_t]'$, which is further used in estimating the measurement error variance-covariance matrix $\Sigma_{\boldsymbol{v}}$:

$$\Sigma_{\boldsymbol{v}} = \begin{bmatrix} \sigma_{v_Y}^{-2} & 0 \\ 0 & \sigma_{v_X}^{-2} \end{bmatrix},$$

where $\sigma_{v_Y}^{-2} = \sum_{t=1}^{T}(y_t - \mathbb{E}[Y_t])$ and $\sigma_{v_X}^{-2} = \sum_{t=1}^{T}(x_t - \mathbb{E}[X_t])$.

Step 2: Now, we use $\boldsymbol{\alpha}_t$ to initialize the FAR kernels $\boldsymbol{\psi}$. For $\boldsymbol{\psi} = B_{\boldsymbol{\psi}}\theta_{\boldsymbol{\psi}}$, where $B_{\boldsymbol{\psi}} = (b_{\boldsymbol{\psi}}(\tau_1), \cdots, b_{\boldsymbol{\psi}}(\tau_M))$ for M number of evaluation points and $\theta_{\boldsymbol{\psi}}$ is sampled from $[\theta_{\boldsymbol{\psi}} \mid \cdots] \sim N(A_{\boldsymbol{\psi}} a_{\boldsymbol{\psi}}, A_{\boldsymbol{\psi}})$, where

$$A_{\boldsymbol{\psi}}^{-1} = \lambda_{\boldsymbol{\psi}}\Omega_{\boldsymbol{\psi}} + [B_{\boldsymbol{\psi}}'\{\sum_{t=p+1}^{T} \boldsymbol{\alpha}_{t-1}'\boldsymbol{\alpha}_{t-1}\}B_{\boldsymbol{\psi}}][B_{\boldsymbol{\psi}}'B_{\boldsymbol{\psi}}],$$

$$a_{\boldsymbol{\psi}} = vec\left(B_{\boldsymbol{\psi}}'\{\sum_{t=p+1}^{T} \boldsymbol{\alpha}_{t-1}'\boldsymbol{\alpha}_{t-1}\}B_{\boldsymbol{\psi}}'\right)$$

where, $B_{\boldsymbol{\psi}}$ is a cubic B-spline basis functions with equally spaced knots, $\Omega_{\boldsymbol{\psi}}$ is a prior precision matrix and $\lambda_{\boldsymbol{\psi}} = 1$, is a smoothing parameter.

Step 3: Using estimates for FAR kernels in second equation of model (4.4), we compute the innovations $\boldsymbol{\epsilon}_t$. We decompose the innovations into factor loading curves (FLCs) $\Phi = (\phi_1, \cdots, \phi_{J_{\boldsymbol{\epsilon}}})$, and time-dependent factors, $e_t = (e_{1,t}, \cdots, e_{J_{\varepsilon},t})'$ as described in equation (5.6). For simplicity, we assume $[e_t \mid \Sigma_e] \overset{indep}{\sim} N(0, \Sigma_e)$, with $\Sigma_e = diag(\{\sigma_j^2\}_{j=1}^{J_{\boldsymbol{\epsilon}}})$, where $J_{\boldsymbol{\epsilon}}$ is the number of factors. We initialize the FDLM parameters using the singular value decomposition (SVD) of $(\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T)' = U_e D_e V_e'$. We set Φ equal to the first $J_{\boldsymbol{\epsilon}}$ columns of V_e and the factors $(e_1, \cdots, e_T)'$ equal to the first $J_{\boldsymbol{\epsilon}}$ columns of $(U_e D_e)$. Further, we estimate $\{\sigma_j^2\}$ and σ_η^2 using the conditional maximum likelihood estimators and $\{\eta_t\}$ is the mean zero approximation error with $\eta_t \overset{iid}{\sim} GP(0, \sigma_\eta^2)$.

Step 4: The implied covariance matrix for $\boldsymbol{\epsilon}_t$ is $K_{\boldsymbol{\epsilon}} = \Phi\Sigma_e\Phi' + \sigma_\eta^2 I_M$, conditional on ϕ_j, σ_j^2 and σ_η^2. We obtain the inverse of $K_{\boldsymbol{\epsilon}}$ using the Woodbury identity:

$$K_{\boldsymbol{\epsilon}}^{-1} = \sigma_\eta^{-2} I_M - \sigma_\eta^{-2}\Phi\tilde{\Sigma_e}\Phi',$$

where $\tilde{\Sigma_e} = diag(\{\sigma_j^2/(\sigma_\eta^2 + \sigma_j^2)\}_{j=1}^{J_\epsilon})$.

4.2.2 Sampling Distribution

Step 1: We sample the FAR kernels $\boldsymbol{\psi}$, $\boldsymbol{\psi}_{=}B_\psi\theta_\psi$, , using parametrization of $\theta_\psi = \tilde{\xi}_\psi\tilde{\theta}_\psi$, where we sample $\tilde{\theta}_\psi$ from $[\tilde{\theta}_\psi \mid \cdots] \sim N(A_\psi a_\psi, A_\psi)$, where

$$A_\psi^{-1} = \lambda_\psi\Omega_\psi + s_l\tilde{\xi}_{\psi_l}^2[(B_\psi'Q)\{\sum_{t=p+1}^{T}\boldsymbol{\alpha}_{t-1}'\boldsymbol{\alpha}_{t-1}\}(B_\psi'Q)'] \otimes [B_\psi'K_\epsilon^{-1}B_\psi],$$

$$a_\psi = \tilde{\xi}_\psi vec\left(B_\psi'\{\sum_{t=p+1}^{T}\boldsymbol{\alpha}_{t-1}'\boldsymbol{\alpha}_{t-1}\}(B_\psi'Q)'\right).$$

Sample $[\tilde{\xi}_\psi \mid \cdots] \sim N(A_{\tilde{\xi}_\psi}a_{\tilde{\xi}_\psi}, A_{\tilde{\xi}_\psi})$, where

$$A_{\tilde{\xi}_\psi}^{-1} = 10^{-6} + \tilde{\theta}_\psi'\left(\left[(B_\psi'Q)\left(\sum_{t=p+1}^{T}\boldsymbol{\alpha}_{t-1}'\boldsymbol{\alpha}_{t-1}\right)(B_\psi'Q)'\right] \otimes \left[B_\psi'K_\epsilon^{-1}B_\psi\right]\right)\tilde{\theta}_\psi$$

$$a_\psi = \tilde{\theta}_\psi' vec\left(B_\psi'\{\sum_{t=p+1}^{T}\left[\boldsymbol{\alpha_t} - \sum_{k\neq 1}s_k\boldsymbol{\alpha}_{t-k}'\right]\boldsymbol{\alpha}_{t-1}\}(B_\psi'Q)'\right)$$

and sample $[\lambda_\psi \mid \cdots] \sim Gamma\left(1/2 + J_\psi^2/2, 1/2 + \theta_\psi'\Omega_\psi\theta_\psi/2\right)$.

Step 2: We compute the innovations using the estimate of FAR kernels. Under FDLM, we decompose the innovations into FLCs and time-dependent factors.

(a) The factors $\{e_t\}_{t=1}^{T}$: sample $[e_t \mid \cdots] \sim N(A_e a_{e_t}, A_e)$, where

$$A_e^{-1} = diag\left(\{\sigma_\eta^2 + \sigma_j^2)\}_{j=1}^{J_\epsilon}\right),$$

$$a_{e_t} = \sigma_\eta^{-2}\Phi'\boldsymbol{\epsilon}_t.$$

(b) The factor precision, σ_j^{-2}: sample

$$[\sigma_{J_\epsilon}^{-2} \mid \cdots] \sim Gamma\left(10^{-3} + \frac{T}{2}, 10^{-3} + \frac{1}{2}\sum_{t=1}^{T}e_{J_\epsilon,t}^2\right).$$

(c) The approximation error precision, σ_η^{-2}: sample

$$[\sigma_\eta^{-2} \mid \cdots] \sim Gamma\left(10^{-3} + \frac{TM}{2}, 10^{-3} + \frac{1}{2}\sum_{t=1}^{T}||\ \boldsymbol{\epsilon}_t - \Phi e_t\ ||^2\right),$$

where $||\cdot||^2$ denotes the Euclidean distance and M is the number of evaluation points .

(d) The factor loading curves $\Phi = B_\phi \Xi$, where B_ϕ is the matrix of basis functions and $\Xi = (\xi_1, \cdots, \xi_{J_\epsilon})$. For $j = 1, \cdots, J_\epsilon$, sample $\xi_j \sim N(A_{\xi_j} a_{\xi_j}, A_{\xi_j})$, where

$$A_{\xi_j}^{-1} = \sigma_\eta^{-2} \left(\sum_{t=1}^{T} e_{j,t}^2 \right) B_\phi' B\phi,$$

$$a_{\xi_j} = \sigma_\eta^{-2} B_\phi' \sum_{t=1}^{T} e_{j,t} \left(\epsilon_t - B_\phi \sum_{k \neq j} \xi_k e_{k,t} \right).$$

Step 3: We obtain the covariance matrix K_ϵ and its inverse K_ϵ^{-1} using the similar approach as in Step 4 of initialization.

Step 4: We sample measurement error precision, $\sigma_{\nu_Y}^{-2}$ and $\sigma_{\nu_X}^{-2}$ as follows:

$$[\sigma_{\nu_Y}^{-2} \mid \cdots] \sim Gamma \left(10^{-3} + \frac{1}{2} \sum_{t=1}^{T} m_t, 10^{-3} + \frac{1}{2} \sum_{t=1}^{T} (y_t - \mu_Y - \alpha_t)^2 \right),$$

$$[\sigma_{\nu_X}^{-2} \mid \cdots] \sim Gamma \left(10^{-3} + \frac{1}{2} \sum_{t=1}^{T} m_t, 10^{-3} + \frac{1}{2} \sum_{t=1}^{T} (x_t - \mu_X - \alpha_t)^2 \right).$$

Step 5: Form the DLM and sample $[\boldsymbol{\alpha}_{t=1}^{T} \mid \cdots]$, using KFAS package in R. We sample the mean function $\mu = B_\phi' \theta_\mu$, where $[\theta_\mu \mid \cdots] \sim N(A_\mu a_\mu, A_\mu)$ where

$$A_\mu^{-1} = \Lambda_\mu^{-1} + \sigma_\nu^{-2} \sum_{t=1}^{T} B_\phi' (Z_t)' Z_t B_\phi,$$

$$a_\mu = \sigma_\nu^{-2} \sum_{t=1}^{T} B_\phi' (Z_t)' (w_t - Z_t \alpha_t),$$

where, $w_t = [y_t \;\; x_t]'$, $\Lambda_\mu = diag(10^8, 10^8, \lambda_\mu^{-1}, \cdots, \lambda_\mu^{-1})$ and $[\lambda_\mu \mid \cdots] \sim Gamma \left(\frac{1}{2}(J_\mu - 3), \frac{1}{2} \sum_{j=3}^{J_\mu} \theta_{\mu,j}^2 \right)$.

4.3 SIMULATION STUDY

In this section, we compared the predicted results of simulated functional time series using the proposed method for MFAR relative to FAR. We are particularly interested in one-step and five-step forecasting and finding the causal relationship between two

functional time series. Throughout the experiments, we consider two scenarios based on model (M1) and model (M2) with varying sample size T. The FAR(1) kernel used for model (M2) is the Bimodal-Gaussian kernel, Ψ as:

$$\Psi(\tau,u) \propto \frac{0.75}{\pi(0.3)(0.4)} \exp -(\tau-0.2)^2/(0.3)^2 - (u-0.3)^2/(0.4)^2 + \frac{0.45}{\pi(0.3)(0.4)} \exp -(\tau-0.7)^2/(0.3)^2 - (u-0.8)^2/(0.4)^2,$$

and MFAR(1) kernels used for model (M1) are also Bimodal-Gaussian kernal, Ψ_Y and Ψ_X as:

$$\Psi_Y(\tau,u) = \Psi(\tau,u)$$
$$\Psi_X(\tau,u) \propto \frac{0.75}{\pi(0.35)(0.35)} \exp -(\tau-0.5)^2/(0.35)^2 - (u-0.3)^2/(0.35)^2 + \frac{0.45}{\pi(0.35)(0.35)} \exp -(\tau-0.75)^2/(0.35)^2 - (u-0.6)^2/(0.35)^2.$$

For stationarity, each kernel is rescaled to pre-specified squared norm, $C_\Phi = \int\int \Psi^2(\tau,u)d\tau du$, with $C_\Phi < 1$. For FAR(1) we select $C_\Psi = 0.5$ and for MFAR(1) we select $C_{\Psi_Y}, C_{\Psi_X} = (0.5, 0.1)$. Thus, functional time series $\{Y_t\}$ generated from MFAR(1) kernels is correlated to functional time series $\{X_t\}$ while $\{Y_t\}$ generated from FAR(1) kernel is independent of $\{X_t\}$. We use smooth Gaussian process for innovation process ϵ_t. In our experiment, we consider dense sampling design using $m_t = 30$ equally-spaced observation points on $[0, 1]$ for all t, however the model works well with sparse sampling design also. Data sets generated using FAR(1) kernel and MFAR(1) kernels are applied to proposed model (MFAR) as well as to univariate model (FAR) proposed by Kowal *et al.* (2019). Table 4.1 shows the one- and five-step root mean squared forecasting errors (RMSFEs) for both the models. When $\{Y_t\}$ is dependent on $\{X_t\}$,we obtain lower root mean squared forecasting errors for the MFAR model than the FAR model suggesting MFAR performs better than FAR in forecasting the functional time series.

Table 4.1: Simulation Results showing h-step ahead RMSFEs. Lower RMSFE for MFAR suggests it performs better than FAR.

	h=1		h=5	
Data Generation	RMSFE (MFAR)	RMSFE (FAR)	RMSFE (MFAR)	RMSFE (FAR)
$\{Y_t\}$ is dependent on $\{X_t\}$	0.009160474	0.009312481	0.01064169	0.0110255
$\{Y_t\}$ is independent of $\{X_t\}$	0.009157562	0.00964527	0.01069097	0.008994967

	h=1			h=5	
Country	MFAR(1)	FAR(1)	Country	MFAR(1)	FAR(1)
USA	0.06716552	0.1259286	USA	0.1579645	0.1463301
UK	0.05066788	0.05098459	UK	0.07166108	0.07252083

Table 4.2: h-step ahead RMSFEs of USA and UK yield curves.

4.4 AN EMPIRICAL ILLUSTRATION: YIELD CURVES

To assess the performance of the proposed model, we conducted an extensive forecasting study using yield curves of two countries: USA and UK. For our study we consider 5 month data starting from Jan 2019 till May 2019. Estimates of the yield curves are provided for maturities $\mathcal{T} = \{90, 180, 360, 720, 1080, 2520, 3600, 10800\}$ days. First four month are used for estimation to produce one-step ahead and five-step(one business week) ahead forecasts of next month. The one- and five-step root mean squared forecasting errors (RMSFEs) are calculated using the forecasted values and data for the fifth month. We compare the RMSFE values for USA and UK yield curves using MFAR(1) model and separately applied FAR(1) model to both the yield curves, presented in Table 4.2. Clearly, proposed model performs better than the univariate case as multivariate model accounts for cross correlation among the yield curves of various countries.

CHAPTER 5

BAYESIAN TESTING OF GRANGER CAUSALITY IN YIELD CURVES

The study of cross-correlation among the bivariate functional time series arises a natural question is whether there exists a directionality of information. We are interested in analyzing the causal relations among a set of variables. As discussed by Granger (1969), the first time series X is said to have a causal influence on the second time series Y if the prediction of second series Y could be improved by the inclusion of past measurements from the first time series X. Later, classical Granger causality was extended to different types of time series model. Boudjellaba *et al.* (1994) derived the necessary and sufficient condition for non-causality in stationary vector autoregressive (VAR) process. Comte and Lieberman (2000) investigated second-order non-causality in the framework of the VARMA process with generalized autoregressive conditional heteroscedasticity (GARCH) type errors. This unified treatment of first and second-order causality is widely used with some modification in various fields (see, e.g., Hafner and Herwartz (2008), Woźniak (2015), Chen *et al.* (2019)). For most of these models, restrictions on parameters are linear functions. Droumaguet and Woźniak (2012) applied Markov-switching VAR models for which the restrictions for non-causality in variance are a non-linear function of parameters and used the Bayesian approach to test non-causality. Allen and Hooper (2018) used Granger causality generalized measures of correlation (GcGMC) to analyze the causal relations between VIX and S&P 500.

Above mentioned literature has applied some modified versions of Granger causality to multivariate time series. Saumard (2017) investigated the causality of functional time series by testing the equality of covariance operators for dependent processes. Recently, Shang *et al.* (2021) extended the GcGMC to bivariate functional time series. This

paper will address the Granger causality for functional time series utilizing the Bayes factor technique introduced by Kass and Raftery (1995). Bayes factor provides a way of evaluating evidence in favor of the null hypothesis; in our case, it tests the non-causality. Following Geweke (1999), we use a sample of the posterior distribution to evaluate the model's marginal likelihood, estimate the integral over the posterior sample, and calculate the modified harmonic mean (MHM). Using the marginal density of data of the two models under consideration, we calculate and use the Bayes factor to choose the preferred model.

5.1 GRANGER CAUSALITY IN FUNCTIONAL AUTOREGRESSIVE MODEL

We investigate the causal relationship between two functional time series data Y_t and X_t, $t \in \mathrm{T}$ observed on unequally spaced points $\tau \in \mathcal{T}$ where $\mathcal{T} \subset \mathbf{R}$ is a compact index set. Firstly, we model Y_t as functional autoregressive model of order 1, written as FAR(1), and partition the linear projection of Y_t on Y_{t-1} and X_{t-1} to account for the Granger causality as

$$\begin{pmatrix} \mathrm{Y}_t - \mu_{\mathrm{Y}} \\ \mathrm{X}_t - \mu_{\mathrm{X}} \end{pmatrix} = \begin{bmatrix} \Psi_{\mathrm{YY}} & \Psi_{\mathrm{YX}} \\ 0 & \Psi_{\mathrm{XX}} \end{bmatrix} \begin{pmatrix} \mathrm{Y}_{t-1} - \mu_{\mathrm{Y}} \\ \mathrm{X}_{t-1} - \mu_{\mathrm{X}} \end{pmatrix} + \begin{pmatrix} \epsilon_{\mathrm{Y}_t} \\ \epsilon_{\mathrm{X}_t} \end{pmatrix} \tag{M1}$$

In second model, we model Y_t as linear projection of Y_{t-1} only as

$$\begin{pmatrix} \mathrm{Y}_t - \mu_{\mathrm{Y}} \\ \mathrm{X}_t - \mu_{\mathrm{X}} \end{pmatrix} = \begin{bmatrix} \Psi'_{\mathrm{YY}} & 0 \\ 0 & \Psi'_{\mathrm{XX}} \end{bmatrix} \begin{pmatrix} \mathrm{Y}_{t-1} - \mu_{\mathrm{Y}} \\ \mathrm{X}_{t-1} - \mu_{\mathrm{X}} \end{pmatrix} + \begin{pmatrix} \epsilon'_{\mathrm{Y}_t} \\ \epsilon'_{\mathrm{X}_t} \end{pmatrix} \tag{M2}$$

where Y_t and $\mathrm{X}_t \in \mathrm{L}^2(\mathcal{T})$ and μ_{Y} and μ_{X} are mean of $\{\mathrm{Y}_t\}$ and $\{\mathrm{X}_t\}$ respectively. Consider $\epsilon_t = (\epsilon_{\mathrm{Y}_t}\ \epsilon_{\mathrm{X}_t})'$ and $\epsilon'_t = (\epsilon'_{\mathrm{Y}_t}\ \epsilon'_{\mathrm{X}_t})'$ as zero mean independent random functions with covariance matrices K_ϵ and $\mathrm{K}_{\epsilon'}$ respectively. In model (M1), Y_t depends on both Y_{t-1} and X_{t-1} and is considered as unrestricted model while model (M2) is considered as restricted as Y_t depends only it's past values. In this setting, X is said to Granger cause Y if model (M1) is significantly better than model (M2) at forecasting future values of Y.

Model (M1) and (M2) are in MFAR form, but for the computation purpose, we need to

convert the MFAR models into DLM form as described in equation (4.4). Let Θ_1 and Θ_2 be the vector of parameters corresponding to model (M1) and (M2), respectively. Now, $\Theta_1 = (\mu, \Sigma_\nu, \Psi, \mathrm{K}_\epsilon)$ and $\Theta_2 = (\mu, \Sigma_\nu, \Psi', \mathrm{K}_{\epsilon'})$ where,

$$\mu = \begin{bmatrix} \mu_{\mathrm{Y}} \\ \mu_{\mathrm{X}} \end{bmatrix}, \ \Sigma_\nu = \begin{bmatrix} \sigma^2_{\nu_{\mathrm{Y}}} & 0 \\ 0 & \sigma^2_{\nu_{\mathrm{X}}} \end{bmatrix}, \ \Psi = \begin{bmatrix} \Psi_{\mathrm{YY}} & \Psi_{\mathrm{YX}} \\ 0 & \Psi_{\mathrm{XX}} \end{bmatrix} \text{ and } \Psi' = \begin{bmatrix} \Psi'_{\mathrm{YY}} & 0 \\ 0 & \Psi'_{\mathrm{XX}} \end{bmatrix}.$$

5.2 METHODOLOGY

We hypothesize that the unrestricted model (M1) yields better prediction in contrast to the restricted model (M2). The null hypothesis indicates no additional contribution from the past of X to the future evolution of Y once the past of Y has been take into account. The alternative is that there is an additional contribution. In the univariate setting the usual way to test for Granger causality is to look at the forecast error sum of squares from each model and compare them through an F test. For multivariate data, this can be extended by taking the trace or determinant of the covariance matrices under each model. We propose two approaches to test for Granger causality, firstly based on F test and second based on Bayesian Information Criteria.

5.2.1 F Test

We perform the sampling for both the models and estimate the one-step forecasts $[y_{\mathrm{T}+h}|y_{1:(\mathrm{T}+h-1)}], h = 1, \cdots, \mathrm{R}$ for both the models. Let $\hat{\mathrm{Y}}_{\mathrm{T}+h}$ and $\hat{\mathrm{Y}}^*_{\mathrm{T}+h}$ be the function of one-step forecast for first model (M1) and second model (M2) respectively for $h = 1, \cdots, \mathrm{R}$. We compute the sum of square errors for both models $\mathrm{SSE}_1 = \sum_{h=1}^{\mathrm{R}} ||\mathrm{Y}_{\mathrm{T}+h} - \hat{\mathrm{Y}}_{\mathrm{T}+h}||^2$ and $\mathrm{SSE}_2 = \sum_{h=1}^{\mathrm{R}} ||\mathrm{Y}_{\mathrm{T}+h} - \hat{\mathrm{Y}}^*_{\mathrm{T}+h}||^2$ where $\mathrm{Y}_{\mathrm{T}+h} = (\mathrm{Y}_{\mathrm{T}+h}(\tau_1), \cdots, \mathrm{Y}_{\mathrm{T}+h}(\tau_{\mathrm{M}})))'$ are the observed values used to measure the one-step forecasting performance at the evaluation points. To investigate the effect of Granger causality of X_t on Y_t, we performed permutation test described as follows.

1. Compute the observed F-statistic, denoted as F_{obs}, using the observed data as:

$$\mathrm{F} = \frac{(\mathrm{SSE}_2 - \mathrm{SSE}_1)/(k_1 - k_2 + 1)}{\mathrm{SSE}_1/(n - k_1 - 1)}. \tag{5.1}$$

2. Take a random sample of X_t, denoted as $X_t^{(l)}$. Subsitute X_t in model (M1) with $X_t^{(l)}$,

$$Y_t - \mu_Y = \Psi_1(Y_{t-1} - \mu_Y) + \Psi_2(X_{t-1}^{(l)} - \mu_X) + \epsilon_t. \tag{5.2}$$

3. Compute the F-statistic based on $X_t^{(l)}$, denote it by $F^{(l)}$.

4. Repeat step 2 to 3 M times to obtain $F^{(1)}, \cdots, F^{(M)}$.

5. The p-value of the text is approximated by

$$p_F = \frac{1}{M} \sum_{l=1}^{M} I[F^{(l)} < F_{obs}], \tag{5.3}$$

where $I[A]$ is an indicator function of event A, and reject the null hypothesis when $p_F < \alpha$, α is the level of significance.

5.2.2 Bayesian Information Criteria

As the dimension of the data becomes large of multivariate data, the F test becomes more unreliable as the estimation of the covariance matrix in higher dimensions is not efficient. For functional data, which is infinite dimensional, this procedure is unusable. Hence we look into Bayesian procedures which concentrate on the dimension of the parameters and not of the data.

The Bayes factor is a central quantity of interest in Bayesian hypothesis testing. The ratio of the marginal data densities for both hypothesis-specific models is known as the Bayes factor. Let $p(Y_t \mid M)$ be the marginal data density for model M; then the Bayes factor would be

$$B_{12} = \frac{p(Y_t \mid M1)}{p(Y_t \mid M2)}. \tag{5.4}$$

Bayesian hypothesis testing begins by specifying different prior distributions on the parameters involved in the models. We also have the likelihood for the data given the prior values of the parameter. Using these, the crucial step is the computation of marginal data densities, $p(Y_t \mid M)$ for both models. For computing this value, we apply the

Modified Harmonic Mean (MHM) method. Let Θ be a K-variate vector of parameters corresponding to the model. Firstly, obtain the sample of draws $\{\Theta^{(i)}\}_{i=1}^{S}$, where S is the number of samples, using the posterior distribution of the parameters $p(\Theta \mid Y_t)$ for each model. Now, the marginal density of data is computed as:

$$p(Y_t \mid M) = \left[S^{-1} \sum_{i=1}^{S} \frac{h(\Theta^{(i)})}{\mathcal{L}(Y_t; \Theta^{(i)}) p(\Theta^{(i)})} \right]^{-1} \tag{5.5}$$

where, $\mathcal{L}(Y_t; \Theta^{(i)})$ is the likelihood function and $p(\Theta^{(i)})$ is the prior density function of parameters obtained by substituting sample of draws $\{\Theta^{(i)}\}_{i=1}^{S}$. Also, $h(\Theta^{(i)})$ is a K-variate truncated Normal distribution with mean equal to posterior mean of Θ and variance set to the posterior covariance matrix of Θ.

In the rest of this section we elaborate the various steps involved in this procedure. In sections 5.2.2.1 and 5.2.2.2 we present the assumed prior distribution and the likelihood of the data under the two models. In section 5.2.2.3 we describe the method to draw samples from the posterior distribution of the parameter. Finally in section 5.2.2.4 we present the algorithm to compute the Bayes factor.

5.2.2.1 Prior Distribution

We specify the prior distributions for each of parameter in the vector $\Theta = (\mu, \Sigma_\nu, \Psi, \epsilon_t)$ below.

1. Prior distribution for mean function, μ:

   $$\mu(\tau) = b'_\mu(\tau)\theta_\mu, \qquad \text{for } \tau \in \mathcal{T},\ \mathcal{T} \text{ is the index set,}$$

   where, b'_μ is low rank thin plate spline basis and

   $$\theta_\mu \sim N(0, \Lambda_\mu)$$

   $$s.t.\ \Lambda_\mu = diag(10^8, 10^8, \lambda_\mu^{-1}, \lambda_\mu^{-1}, \cdots, \lambda_\mu^{-1}) \text{ and } \lambda_\mu^{-1/2} \sim \text{Uniform}(0, 10^4).$$

2. Prior distribution for observed covariance matrix, Σ_ν:

   $$\Sigma_\nu = \begin{bmatrix} \sigma^2_{\nu_Y} & 0 \\ 0 & \sigma^2_{\nu_X} \end{bmatrix}$$

   $$where,\ \sigma^2_{\nu_Y} \sim \text{Gamma}(10^{-3}, 10^{-3}) \text{ and } \sigma^2_{\nu_X} \sim \text{Gamma}(10^{-3}, 10^{-3}).$$

3. Prior distribution for FAR kernel, Ψ:

$$\Psi(\tau, u) = (b'_\Psi(u) \otimes b'_\Psi(\tau))\theta_\Psi, \qquad \text{for } \tau, u \in \mathcal{T},$$

where, b'_Ψ is cubic B-spline basis and

$$[\theta_\Psi \mid \lambda_\Psi] \sim N(0, \lambda_\Psi^{-1}\Omega_\Psi^{-1}) \qquad (\theta_\Psi \text{ induces a Gaussian prior on } \Psi)$$

$$s.t.\ \ \lambda_\Psi = \tilde{\zeta}_\Psi^{-2}\tilde{\lambda}_\Psi, \qquad \text{where, } \tilde{\zeta}_\Psi \sim N(0, 10^6) \text{ and } \tilde{\lambda}_\Psi \sim \text{Gamma}(1/2, 1/2) \text{ and}$$

$$\Omega_\Psi = \Omega_2 + \kappa\Omega_0, \quad \text{where, } \log(\kappa) \sim N(0, 4),\ \ \Omega_0 = \int\int \Psi(\tau, u)d\tau du \text{ and}$$

$$\Omega_2 = \int\int \left(\frac{d^2}{d\tau^2}\Psi(\tau, u) + \frac{d^2}{d\tau du}\Psi(\tau, u) + \frac{d^2}{du^2}\Psi(\tau, u)\right) d\tau du.$$

4. Prior distribution for evolution error, ϵ_t:
We use the functional dynamic linear model (FDLM) approach of Kowal et.al. (2016), in which they decomposes the evolution error ϵ_t into factor loading curves ϕ_j, and time-dependent factors $e_{j,t} \in \mathbf{R}$, for $j = 1, \cdots, \mathrm{J}_\epsilon$:

$$\epsilon_t(\tau) = \sum_{j=1}^{\mathrm{J}_\epsilon} e_{j,t}\phi_j(\tau) + \eta_t(\tau), \qquad \forall\tau \in \mathcal{T}, \tag{5.6}$$

where, J_ϵ is the number of factors and $\{\eta_t\}$ is mean zero approximation error. Priors for each of the parameters in equation (5.6) are as follows:

a) Prior distribution for $\{\eta_t\}$:

$$\eta_t \overset{iid}{\sim} \mathcal{GP}(0, \mathrm{K}_\eta) \quad \text{where, } \mathrm{K}_\eta(\tau, u) = \sigma_\eta^2 \mathbf{1}(\tau = u)$$
$$s.t.\ \ \sigma_\eta^{-2} \sim \mathit{Gamma}(10^{-3}, 10^{-3}) \text{ and } \mathbf{1}(.) \text{ is the indicator function.}$$

b) Prior distribution for e_t:

$$e_t \sim N(0, \Sigma_e) \qquad \text{where, } \Sigma_e = \mathit{diag}(\{\sigma_j^2\}_{j=1}^{\mathrm{J}_\epsilon}).$$

$$\text{Joint distribution of } [\sigma_1^{-2}, \cdots, \sigma_{\mathrm{J}_\epsilon}^{-2}] = [\sigma_{\mathrm{J}_\epsilon}^{-2}] \prod_{j=1}^{\mathrm{J}_\epsilon - 1} [\sigma_j^{-2} \mid \sigma_{j+1}^{-2}, \cdots, \sigma_{\mathrm{J}_\epsilon}^{-2}]$$

where, $\sigma_{\mathrm{J}_\epsilon}^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3})$ and
$[\sigma_j^{-2} \mid \sigma_{j+1}^{-2}, \cdots, \sigma_{\mathrm{J}_\epsilon}^{-2}] = [\sigma_j^{-2} \mid \sigma_{j+1}^{-2}] \sim \text{Uniform}(0, \sigma_{j+1}^{-2})$ for $j = 1, \cdots, \mathrm{J}_\epsilon - 1$.

c) Prior distribution for ϕ_j for $j = 1, \cdots, \mathrm{J}_\epsilon$:

$$\phi_j(\tau) = b'_\phi(\tau)\theta_\phi, \qquad \text{for } \tau \in \mathcal{T},$$

where, b'_ϕ is low rank thin plate spline basis and

$$\theta_\phi \sim N(0, \Lambda_\phi)$$

$$s.t.\ \Lambda_\phi = diag(10^8, 10^8, \lambda_\phi^{-1}, \lambda_\phi^{-1}, \cdots, \lambda_\phi^{-1}) \text{ and } \lambda_\phi^{-1/2} \sim \text{Uniform}(0, 10^4).$$

5.2.2.2 Likelihood Function

We use the dynamic linear model described in equation (4.4) to form the likelihood function for Y_t. At initial time $i = 1$, we have

$$\mathbf{Y}_1 \sim \mathrm{N}(\mathbf{Z}_1\boldsymbol{\mu}^{(1)} + \mathbf{Z}_1\boldsymbol{\alpha}_1, \Sigma_v^{(1)})$$

$$\boldsymbol{\alpha}_1 \sim \mathrm{N}(\mathbf{0}, \mathbf{K}_{\boldsymbol{\epsilon}}^{(1)})$$

Now, for $i = 2, \cdots, n$, one-step ahead distribution of Y_t is given as:

$$\mathbf{Y}_{i|i-1} \sim \mathrm{N}(\mathbf{Z}_i\boldsymbol{\mu}^{(i)} + \mathbf{Z}_i\boldsymbol{\alpha}_{i|i-1}, \Sigma_v^{(i)} + \mathbf{Z}_i\boldsymbol{\Xi}_{i|i-1}\mathbf{Z}_i^{\mathrm{T}})$$

$$\boldsymbol{\alpha}_{i|i-1} = \boldsymbol{\psi}^{(i)}\boldsymbol{\alpha}_{i-1|i-1}, \quad \text{and} \quad \boldsymbol{\Xi}_{i|i-1} = \boldsymbol{\psi}^{(i)}\boldsymbol{\Xi}_{i-1|i-1}[\boldsymbol{\psi}^{(i)}]^{\mathrm{T}} + \mathbf{K}_{\boldsymbol{\epsilon}}^{(i)})$$

Priors for each time is given as:

$$\boldsymbol{\alpha}_{i|i} = \boldsymbol{\alpha}_{i|i-1} + \mathbf{K}_i\boldsymbol{v}_i, \quad \text{and} \quad \boldsymbol{\Xi}_{i|i} = (\mathrm{I_M} - \mathbf{K}_i\mathbf{Z}_i)\boldsymbol{\Xi}_{i|i-1},$$

$$\text{where,} \quad \mathrm{K}_i = \boldsymbol{\Xi}_{i|i-1}\mathbf{Z}_i^{\mathrm{T}}(\mathbf{Z}_i\boldsymbol{\Xi}_{i|i-1}\mathbf{Z}_i^{\mathrm{T}} + \Sigma_v)^{-1}, \quad \text{and} \quad v_i = (\mathbf{Y}_{i|i-1} - \mathbf{Z}_i\boldsymbol{\mu}^{(i)}) - \mathbf{Z}_i\boldsymbol{\alpha}_{i|i-1}.$$

5.2.2.3 Posterior Sampling Distribution

We use Gibbs sampling algorithm to obtain the posterior sampling distribution. Firstly, we initialize the DLM (4.4) parameters and later we present the sampling distributions. Section 4.2 of chapter 3 provides a detailed steps of the Gibbs sampling algorithm for multivariate functional autoregressive model.

5.2.2.4 Algorithm and Interpretation of Bayes Factor

Algorithm to compute Bayes factor:

1. Translate the models M1 and M2 into DLM form as described in equation (4.4).

2. Consider model M1. Initialize the parameter space $\Theta_1 = (\mu, \Sigma_\nu, \Psi, \epsilon_t)$ using the steps mentioned in Section 5.2.2.3 under Initialization.

3. For $i = 1, \cdots, \mathrm{S}$, where, S is the number of simulations:

   a) Obtain the sample of $\{\Theta_1^{(i)}\}$ by applying the posterior sampling distribution given in Section 5.2.2.3 under Sampling.

Table 5.1: Interpretation of Bayes Factor and Logarithm of Bayes Factor values

$ln(B_{12})$	B_{12}	Evidence against H_0
0 to 1	1 to 3	Not worth more than a bare mention
1 to 3	3 to 20	Substantial
3 to 5	20 to 150	Strong
>5	>150	Decisive

b) Substitute the posterior sample $\{\Theta_1^{(i)}\}$ in prior distributions mentioned in Section 5.2.2.1 to obtain prior density function $p(\Theta_1^{(i)})$.

c) Compute the likelihood function $\mathcal{L}(Y_t; \Theta_1^{(i)})$ as described in Section 5.2.2.2 and obtain a truncated Normal distribution with mean equal to posterior mean of Θ_1 and variance set to the posterior covariance matrix of Θ_1, denoted by $h(\Theta_1^{(i)})$.

d) Substitute the values obtained from prior density function $p(\Theta_1^{(i)})$, likelihood function $\mathcal{L}(Y_t; \Theta_1^{(i)})$ and truncated Normal density function $h(\Theta_1^{(i)})$ in equation (5.5) to compute the marginal data density corresponding to model M1, $p(Y_t \mid M1)$.

4. Repeat step 2 and 3, considering model M2 and parameter space as $\Theta_2 = (\mu, \Sigma_\nu, \Psi', \epsilon_t')$ to obtain the marginal data density corresponding to model M2, $p(Y_t \mid M2)$.

5. Compute Bayes factor B_{12} using the marginal data densities of model M1 and M2 and equation (5.4).

Interpretation of Bayes factor:

Kass and Raftery (1995) provided the interpretation of the Bayes factor values in terms of weighing evidence against the null hypothesis, presented in Table 5.1. A positive logarithm of the Bayes factor is interpreted as evidence in favor of the unrestricted model. Symmetrically, a negative logarithm of the Bayes factor indicates that the data prefer the restricted model.

Table 5.2: Bayes Factor Values of Simulated Series. Large positive values of $ln(B_{12})$ in first two cases depicts strong causal relationships. In third case, low value of $ln(B_{12})$ suggest no causal relation exists.

	$\{X_t\}$ Cause $\{Y_t\}$	$\{Y_t\}$ cause $\{X_t\}$	$\{Y_t\}$ and $\{X_t\}$ are independent
B_{12}	8.70756E+33	1.35857E+52	0.007122615
$ln(B_{12})$	78.14949918	120.0408611	-4.94448041

5.3 SIMULATION STUDY

To acknowledge the causal relationship between $\{X_t\}$ and $\{Y_t\}$, we calculate the F test using F_{obs} statistics and Bayes factor as described in section 5.2.2.4. We consider three different simulated cases: (1) $\{Y_t\}$ is generated from MFAR and $\{X_t\}$ is generated from FAR, and we check whether the $\{X_t\}$ series causes $\{Y_t\}$ series, (2) $\{X_t\}$ is generated from MFAR and $\{Y_t\}$ is generated from FAR, and we check for causality in the opposite direction, and (3) $\{X_t\}$ and $\{Y_t\}$ both are generated from FAR. Table 5.2 presents the Bayes factor B_{12} and logarithm of Bayes factor $ln(B_{12})$ values. Based on $ln(B_{12})$, we conclude that there is statistical evidence in favor of a causal relationship in first two cases and no such relationship exists when $\{X_t\}$ and $\{Y_t\}$ are independent.

We now repeat the analysis presented in Table 5.2 forty times for each of the dependent and independent cases and compute the Bayes factor. The boxplot of the logarithm of these Bayes factors is presented fin Figure 5.1. It can be seen that the distributions are well-separated. Therefore the method is useful in detecting whether the data arises from the dependent or the independent model.

5.4 APPLICATION TO REAL DATA

We apply the proposed method to two real datasets. The first application is to daily data of yield curves over different maturities for two countries. For the second application, the response variable is still functional but the predictor variable is multivariate. The response is the daily curve of particulate matter in a city while the predictors are several meteorological variables.

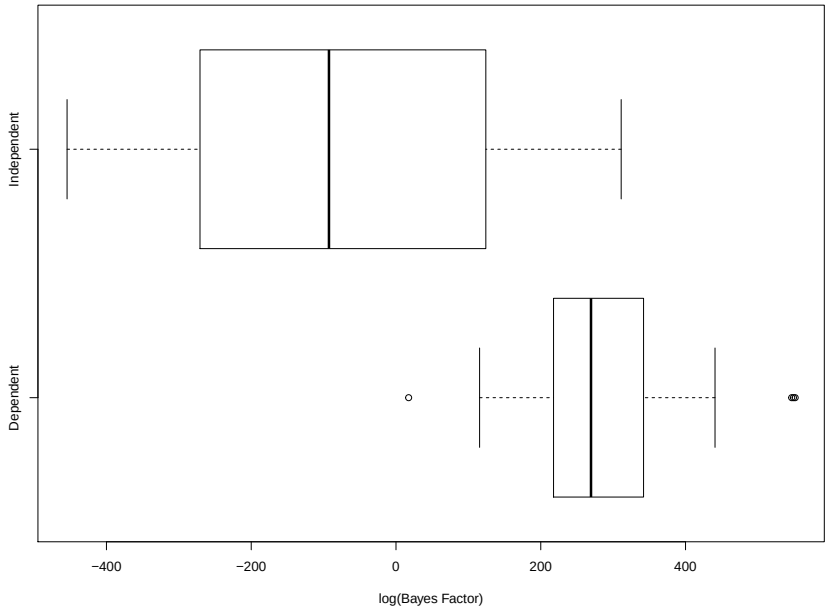


Figure 5.1: Boxplot of the logarithm of the Bayes factor for the dependent and independent cases depicting well-separated distributions of both cases.

5.4.1 Testing causality between yield curves of two countries

To assess the performance of the proposed model, we examine the causal relationship between the yield curves of two countries: USA and UK. The yield curve depicts the relationships of bond yields with different maturities. Yield curves have a significant impact on the money supply within the economy and thus give insights into the country's economic condition. Time series of yield curves have been studied by several authors. See Sen and Klüppelberg (2019) for an introduction and exploratory plots. For our study, we considered 5-month data starting from Jan 2021 till May 2021, which includes 82 working days data. Values of the yield curve are available for 9 maturities: 3, 6, 12, 24, 36, 60, 84, 120, 360 months. We obtain the USA bond yield data for T maturities from the U.S. Department of the Treasury website and the UK bond yield data from investing.com. The objective is to investigate whether UK yield curve Granger cause USA yield curve and vice-versa. We denote the USA yield curves as Y_t and the UK yield curve as X_t, substituting them in models (M1) and (M2). Once we have the unrestricted

Table 5.3: Bayes Factor Values of Yield Curves Example. Large negative values of $ln(B_{12})$ shows no causal relationship exits between USA and UK yield curves.

	UK Cause USA	USA cause UK
Bayes Factor (B_{12})	3.18731E-20	1.06315E-16
Logarithm(Bayes Factor) $ln(B_{12})$	-44.89252354	-36.78012961

and restricted models, we apply the Bayesian analysis to calculate the Bayes factor B_{12} as defined in the equation (5.4). We perform 1000 simulations to generate posterior samples of Θ of dimension 123. Looking at the logarithm of Bayes factor, $ln(B_{12})$ in table 5.3, we conclude that it is not statistically significant to say that the UK yield curve Granger cause USA yield curve. We also check whether the reverse relation exists or not, i.e., whether USA yield curve cause UK yield curve. In this case, we take UK yield curves as Y_t and the USA yield curve as X_t in model (M1) and (M2). Based on $ln(B_{12})$ (obtained from table 5.3), we conclude that USA yield curve does not Granger cause UK yield curve. Now we reverse the predictor and response variables and repeat the analysis. We observe that the Bayes Factor is low in this case also and hence conclude that the UK yield curve does not Granger cause the USA yield curve.

5.4.2 Testing whether meteorological factors cause air pollution level

Numerous pieces of literature have been forecasting air quality using meteorological features (see, e.g., Aneja *et al.* (2001), Kumar and Goyal (2011), Meng *et al.* (2019)). Here, we investigate the causal effect of meteorological factors such as temperature, humidity, wind speed, and solar radiation on air pollution concentration. To compute particle pollution levels, we use particulate matter with diameters 10 micrometers and smaller, denoted by PM10. For our study, we used daily data from Jan 2015 to June 2020 in Delhi, India, consisting of 2008 days. However, after data cleansing we are left with 1670 days of data. We downloaded the data from official Central Pollution Control Board of India website. We take Y_t as PM10 in Delhi for each hour of a day; that is, [0,23] is the support of each function. Figure 5.2 provides with the surface plot depicting the pollution

level in Delhi over the years 2015 and 2020. For X, we take the daily average of three meteorological factors: temperature, relative humidity, and wind speed obtained from meteoblue.com. In this case, X is a vector instead of a function. Using X_t and Y_t, we obtain unrestricted and restricted models and apply the algorithm to compute the Bayes factor as mentioned in section 5.2.2.4. In this case, the estimated parameter space Θ is of length 361. The logarithm of the Bayes factor is obtained as $\ln(B_{12}) = 78.61206652$. Based on the interpretation of $\ln(B_{12})$ from table (5.4), we conclude that there is strong evidence that meteorological factors Granger cause air pollution.

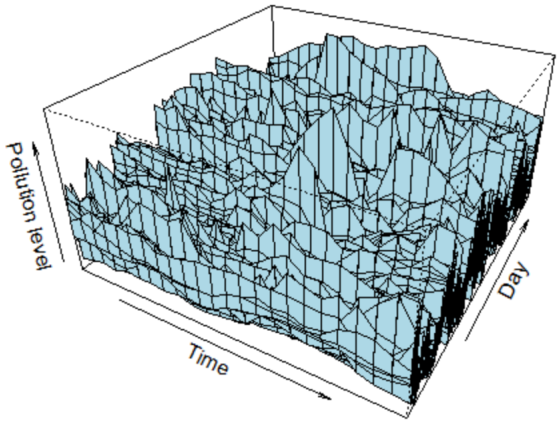


Figure 5.2: The surface plot displays the evolution of the Pollution level in Delhi between the years 2015 and 2020, observed at every hour of a day.

CHAPTER 6

PRICING AND HEDGING OF OPTIONS USING HAWKES PROCESS

The recent availability of high-frequency financial data leads to extensive research in modeling intraday asset prices. Such high-frequency data prevails the issue of controlling the market microstructure noise. Most approaches in finance for asset pricing rely on diffusion models such as the Brownian motion model and its variants. These diffusion models presume continuity in stock price dynamics; however, the stock price movement is discrete and characterized by tick size. The discrepancy between the assumed model and reality results in inconsistent volatility prediction and hence mispricing of options.

To circumvent the effect of microstructure noise, Engle and Russell Engle and Russell (1998) introduced the first point process with dependent arrival rates, namely Autoregressive Conditional Duration (ACD) model for irregular intervals. Later, Bowsher Bowsher (2007) modeled multivariate market events such as the timing of trades and mid-quote changes using the Hawkes process with vector conditional intensity. Hawkes process, introduced by A.G. Hawkes Hawkes (1971), is a class of multivariate point processes with self and mutually exciting stochastic intensities.

One can find wide applications of the Hawkes process in finance. Barcy Bacry *et al.* (2015) provides an excellent survey of Hawkes process in finance, focusing on high frequency and market microstructure noise. Barcy Bacry *et al.* (2013) introduced a tick-by-tick model using the multivariate Hawkes process, which deals with the effect of microstructure noise and also preserves the Brownian diffusion behavior on large scales. In this paper, we work on similar lines as Barcy Bacry *et al.* (2013) to price the options. The model is associated with two point process corresponding to positive and negative

jumps of the asset prices with appropriate stochastic intensities.

Options are widely used to hedge potential losses or reduce an investment's downside risk. A crucial issue in hedging is determining the optimal hedge ratio: the number of derivative contracts to buy or sell for each unit of the underlying asset on which the investor bears the risk. Most of the literature on pricing and hedging assumes continuous asset price (for reference see: Duffie and Richardson (1991), Pham (2000), Dolinsky and Soner (2014)). One can also find work on discontinuous price setting to find the optimal hedging strategy as in Hayashi and Mykland (2005) and Mykland (2000), but they don't give perfect hedging. Bardhan and Chao (1993) considered that asset prices are driven by a multidimensional point process of random jumps with stochastic intensity. They constructed hedging portfolios using the equivalent martingale measure approach. Arai (2005) provides mean-variance hedging strategy for discontinuous asset prices. Gao *et al.* (2019) implied Bayesian inference to form an optimal hedging strategy that minimizes the Value-at-Risk (VaR) of the hedged portfolio for the European put option. Sen (2009) modeled the stock price process as a birth and death process and considered the intensity rate as a two-state Markov process to acquire the hedging ratios. In this paper, we assume the asset prices are driven by a point process of random jumps with stochastic intensity and employ Markov process along with Bayesian techniques to obtain the hedge ratios.

6.1 HAWKES PROCESS

We have considered two counting process $N_1(t)$ and $N_2(t)$ over time horizon $t \in [0, T]$ representing the sum of positive and negative jumps of asset price respectively. Let $X(t)$ be the price of the asset for $t \in [0, T]$ such that:

$$X(t) = N_1(t) - N_2(t). \tag{6.1}$$

Let $N_1(t)$ and $N_2(t)$ are two independent Poisson processes with intensities μ_1 and μ_2. As defined by Barcy (2013), the bivariate process $\{N_1(t), N_2(t)\}$ is said to be linear

Hawkes process if $N_1(t)$ and $N_2(t)$ have no common jumps and if there exist non-negative functions $\{\phi_{ij}\}_{i,j=1,2}$ such that:

$$\lambda_i(t) = \mu_i + \int_{-\infty}^{t} \phi_{ii}(t-s)dN_i(s) + \int_{-\infty}^{t} \phi_{ij}(t-s)dN_j(s). \tag{6.2}$$

We have considered a simplified version of the intensity function with only mean reverting terms, given as:

$$\begin{aligned}\lambda_1(t) &= \mu + \int_{-\infty}^{t} \phi(t-s)dN_2(s)\\ \lambda_2(t) &= \mu + \int_{-\infty}^{t} \phi(t-s)dN_1(s)\end{aligned} \tag{6.3}$$

where μ is an exogenous intensity and $\phi(t)$ is a right-sided exponential kernel defined as:

$$\phi(t) = \omega \exp(-\delta t) 1_{R^+}(t) \tag{6.4}$$

where ω, $\delta > 0$ and $\phi(t)$ satisfy the stability condition:

$$||\phi||_1 = \frac{\omega}{\delta} < 1. \tag{6.5}$$

6.2 PARAMETER ESTIMATION

To estimate the parameters from real data we compute the signature plot. The signature plot is defined as quadratic variation of $X(t)$ over the time period of $[0, T]$. The estimation of the parameters can be done using the best fit of the realized signature plot to the mean signature plot. For exponential kernel defined in (6.4) with the stability condition (6.5), we have the mean signature plot as:

$$C(\tau) = \frac{2\mu\delta}{(\delta-\omega)(\omega+\delta)^3\tau}[\delta^2 + \omega(\omega+2\delta)(1-\exp\{-(\omega+\delta)\tau\}]. \tag{6.6}$$

In appendix B, we derive the mean signature plot defined in (6.6).

The realized signature plot is the realized volatility of $X(t)$ over $[0, T]$, defined as:

$$\hat{C}(\tau) = \frac{1}{T} \sum_{n=0}^{T/\tau} |X((n+1)\tau) - X(n\tau)|^2. \tag{6.7}$$

There are mainly 3 parameters which we have defined using the parameter space $\theta = (\mu, \omega, \delta)$. To estimate the parameter, we minimize the least square criterion given as:

$$F(\theta) = |\hat{C}(\tau) - C(\tau)|^2. \tag{6.8}$$

We applied Newton Rapson iterations on the regression estimator $F(\theta)$ to obtain the estimates of the parameter θ.

6.3 OPTION PRICING

In this section, we explain the procedure to price the options:

1. Choose M large enough such that:

$$\lambda_1(t) < M \quad \text{and} \quad \lambda_2(t) < M, \forall t \in [0, T].$$

2. Simulate on $[0, T]$ a standard Poisson process with an intensity M and apply thinning procedure to each jump of the obtained process as follows:

   a) Reject the point with probability $(M - \lambda_1(t) - \lambda_2(t))/M$

   b) Mark the point as jump of $N_1(t)$ with probability $\lambda_1(t)/M$

   c) Mark the point as jump of $N_2(t)$ with probability $\lambda_2(t)/M$

3. Obtain the cumulative sum of jumps corresponding to $N_1(t)$ and $N_2(t)$, written as $\tilde{N}_1(t)$ and $\tilde{N}_2(t)$ respectively. The simulated path $\tilde{X}(t)$ is then given as:

$$\tilde{X}(t) = \tilde{N}_1(t) - \tilde{N}_2(t), \quad \forall t \in [0, T]. \tag{6.9}$$

4. Using the simulated path of $\tilde{X}(t)$, the simulated option price $\tilde{O}_T$ is:

$$\tilde{O}_T = (\tilde{X}(T) - K)^+, \quad \text{where K is the strike price.} \tag{6.10}$$

5. Repeat step 2 to step 4 for a large number N simulations and take an average of the

simulated option prices.

6.3.1 Option pricing with two state intensity

Here, we present the Monte Carlo approach to obtain the terminal option price, O_T corresponding to two state Markov process with fixed transition matrix Q. Let $\theta_t = (\mu_t, \omega_t, \delta_t)$ is a Markov process taking two states defined as $\{\theta_1 = (\mu_1, \omega_1, \delta_1), \theta_2 = (\mu_2, \omega_2, \delta_2)\}$. Let ζ_t being the unobserved state process taking two values 1 and 2 corresponding to states θ_1 and θ_2 respectively. Given the filtration $\mathcal{F}_t = \sigma(S_u; 0 \leq u \leq t)$, where S_t is the stock price at time t, we define the probabilities as:

$$\begin{aligned} \pi_1(t) &= \mathbb{P}(\zeta_t = 1|\mathcal{F}_t) \\ \pi_2(t) &= \mathbb{P}(\zeta_t = 2|\mathcal{F}_t) \end{aligned} \tag{6.11}$$

. The algorithm to price the option is as follows:

1. Fix $\zeta_0 = 1$.
2. Generate the ζ process.
3. Use equation (6.4) to compute kernel $\phi(t)$.
4. Use equation (6.3) to obtain the stochastic intensities corresponding to $N_1(t)$ and $N_2(t)$ counting process.
5. Repeat the steps from 1-4 in section 6.3 to obtain the simulated option price, $\tilde{O}_T$.
6. Take the expectation of $\tilde{O}_T$ with respect to π_0.
7. Repeat step 5 to step 6 for a large number N and take an average of the simulated option prices, represented by O_T^1.
8. Now, fix $\zeta_0 = 2$. Repeat step 2 till step 7 to obtain an average of the simulated option prices corresponding to $\zeta_0 = 2$, represented by O_T^2.
9. Option price for the two state intensity is given as: $O_T = \pi_1(T)O_T^1 + \pi_2(T)O_T^2$.

6.4 HEDGING USING BAYESIAN FILTRATION

Hedging an option involves forming a portfolio that will offset the risk of holding single option. The main objective is to formulate a replicating and self financing portfolio against an option such that the portfolio value and the hedged option's value are always inline with each other and the portfolio is self sufficient. Thus, the investor doesn't need to pour additional sum of money after the initial investment and obtain the same terminal value as the payoff of the hedged option. To form such an portfolio, we will use the stock underlying the option, the bond and an another option derived from the stock.

6.4.1 Hedging with constant intensity rate

To hedge an option O_1 we trade in the stock, the bond and another option O_2. Let $F_0(x,t) = B_0 \exp - \int_0^t r_s ds$ be the price of bond, $F_1(x,t)$ and $F_2(x,t)$ be the expected payoff of the options corresponding to O_1 and O_2 respectively given the filtration $\mathcal{F}_t = \sigma(S_u; 0 \leq u \leq t)$. Let $F_3(x,t) = N_1(t) - N_2(t)$ be the price of the stock where $N_1(t)$ and $N_2(t)$ is the cumulative sum of positive and negative jumps respectively. Assume $F_i(x,t),\ i = 0,1,2,3$ are continuous with respect to $N_1(t)$ and $N_2(t)$ and the following partial derivatives exist:

$$\frac{\partial F_i}{\partial t} = \alpha_{F_i} \quad \frac{\partial F_i}{\partial N_1} = \beta_{F_i} \quad \frac{\partial F_i}{\partial N_2} = \gamma_{F_i} \tag{6.12}$$

Let $V(t)$ is self financing riskless portfolio. Let $\psi_i(t)$ is the amount of asset i invested to construct the portfolio $V(t)$. Thus, the portfolio $V(t)$ is defined as:

$$V(t) = \sum_{i=0}^{3} \psi_i(t) F_i(x,t)$$

Let $u_i(t) = \frac{\psi_i(t)F_i(x,t)}{V(t)}$ is the proportion of wealth invested in asset i, such that,

$$\sum_{i=0}^{3} u_i(t) = 1. \tag{6.13}$$

Since $\mathrm{V}(t)$ is self financing,

$$\begin{aligned}\frac{d\mathrm{V}(t)}{\mathrm{V}(t)} &= \sum_{i=0}^{3} \psi_i(t) \frac{d\mathrm{F}_i(x,t)}{\mathrm{F}_i(x,t)} \\ &= u_0(t) r_t dt + \sum_{i=1}^{2} u_i(t) \tilde{\alpha}_{\mathrm{F}_i} dt + \sum_{i=1}^{2} u_i(t) \tilde{\beta}_{\mathrm{F}_i} d\mathrm{N}_1(t) \\ &\quad + \sum_{i=1}^{2} u_i(t) \tilde{\gamma}_{\mathrm{F}_i} d\mathrm{N}_2(t) + u_3(t) \frac{d\mathrm{N}_1(t) - d\mathrm{N}_2(t)}{\mathrm{F}_3(x,t)}\end{aligned}$$

As $\mathrm{V}(t)$ is riskless, we have:

$$\frac{u_3(t)}{\mathrm{F}_3(x,t)} + u_1(t) \tilde{\beta}_{\mathrm{F}_1} + u_2(t) \tilde{\beta}_{\mathrm{F}_2} = 0 \tag{6.14}$$

$$-\frac{u_3(t)}{\mathrm{F}_3(x,t)} + u_1(t) \tilde{\gamma}_{\mathrm{F}_1} + u_2(t) \tilde{\gamma}_{\mathrm{F}_2} = 0 \tag{6.15}$$

The no arbitrage condition implies

$$u_0(t) r_t + u_1(t) \tilde{\alpha}_{\mathrm{F}_1} + u_2(t) \tilde{\alpha}_{\mathrm{F}_2} = r_t \tag{6.16}$$

Solving equations (6.13)-(6.16), we obtain the asset weights as:

$$u_0(t)) = -\frac{1}{r_t} \left(u_1(t) \tilde{\alpha}_{\mathrm{F}_1} + u_2(t) \tilde{\alpha}_{\mathrm{F}_2} \right) \tag{6.17}$$

$$u_1(t) = \left[\left(1 - \frac{\tilde{\alpha}_{\mathrm{F}_1}}{r_t} - \mathrm{F}_3(x,t) \tilde{\beta}_{\mathrm{F}_1} \right) - \left(\frac{\tilde{\gamma}_{\mathrm{F}_1} + \tilde{\beta}_{\mathrm{F}_1}}{\tilde{\gamma}_{\mathrm{F}_2} + \tilde{\beta}_{\mathrm{F}_2}} \right) \left(1 - \frac{\tilde{\alpha}_{\mathrm{F}_2}}{r_t} - \mathrm{F}_3(x,t) \tilde{\beta}_{\mathrm{F}_2} \right)^{-1} \right] \tag{6.18}$$

$$u_2(t) = \left[\left(1 - \frac{\tilde{\alpha}_{\mathrm{F}_2}}{r_t} - \mathrm{F}_3(x,t) \tilde{\beta}_{\mathrm{F}_2} \right) - \left(\frac{\tilde{\gamma}_{\mathrm{F}_2} + \tilde{\beta}_{\mathrm{F}_2}}{\tilde{\gamma}_{\mathrm{F}_1} + \tilde{\beta}_{\mathrm{F}_1}} \right) \left(1 - \frac{\tilde{\alpha}_{\mathrm{F}_1}}{r_t} - \mathrm{F}_3(x,t) \tilde{\beta}_{\mathrm{F}_1} \right)^{-1} \right] \tag{6.19}$$

$$u_3(t) = -\mathrm{F}_3(x,t) \left(u_1(t) \tilde{\beta}_{\mathrm{F}_1} + u_2(t) \tilde{\beta}_{\mathrm{F}_2} \right) \tag{6.20}$$

A self financing replicating portfolio for option O_2 is obtain by investing ψ_i / ψ_2, that is, $u_i \mathrm{F}_2 / u_2 \mathrm{F}_i$ of asset i.

6.4.2 Hedging with stochastic intensity rate

Let θ_t is a two state Markov process as defined in section 6.3.1 and ζ_t is the corresponding unobserved state process with the probabilities $\pi_i(t),\ i = 1, 2$.

Let $G_i(x,t) := \mathbb{E}(P|\mathcal{G}t)$ be the expected payoff given the complete filtration $\mathcal{G}_t = \sigma(S_u, \theta_u; 0 \leq u \leq t)$ for an option when the stock price is $F_3(x,t)$ and $\theta_t = \theta_i$.Now, $F_1(x,t)$ is defined as

$$F_1(x,t) := \mathbb{E}(P|\mathcal{F}_t) = \pi_1(t)G_1(x,t) + \pi_2(t)G_2(x,t). \tag{6.21}$$

As shown in Snyder (1973), the $\pi_i(t)$ process evolves as:

$$d\pi_1(t) = a(t)dt + b(t,1)dN_1(t) + b(t,2)dN_2(t) \tag{6.22}$$

where $a(t)$ and $b(t,i), i = 1, 2$ are $\mathcal{F}_t$ adapted processes. Note that $\pi_2(t) = 1 - \pi_1(t)$.

Theorem 6.1. $dF_1/F_1 = \tilde{\alpha}_{F_1}dt + \tilde{\beta}_{F_1}dN_1(t) + \tilde{\gamma}_{F_1}dN_2(t)$ *where,*

$$\tilde{\alpha}_{F_1} = \pi_1\alpha_{G_0} + \pi_2\alpha_{G_1} + G_0a(t) - G_1a(t)$$
$$\tilde{\beta}_{F_1} = \pi_1\beta_{G_0} + \pi_2\beta_{G_1} + G_0b(t,1) - G_1b(t,1) + \beta_{G_0}b(t,1) + \beta_{G_1}b(t,1)$$
$$\tilde{\gamma}_{F_1} = \pi_1\gamma_{G_0} + \pi_2\gamma_{G_1} + G_0b(t,2) - G_1b(t,2) + \gamma_{G_0}b(t,2) + \gamma_{G_1}b(t,2)$$

Proof: $dG = \pi_1dG_0 + \pi_2dG_1 + G_0d\pi_1 + G_1d\pi_2 + d[G_0 - G_1, \pi_1]$

Now,

$$dG_0 = \alpha_{G_0}dt + \beta_{G_0}dN_1(t) + \gamma_{G_0}dN_2(t),$$
$$dG_1 = \alpha_{G_1}dt + beta_{G_1}dN_1(t) + \gamma_{G_1}dN_2(t), \text{ and}$$
$$d(G_0 - G_1, \pi_1) = d(G_0 - G_1)d(\pi_1)$$

As we note here,

$$d[X,Y]_t = \lim_{\Delta t \longrightarrow 0}(X(t+\delta t) - X(t))(Y(t+\delta t) - Y(t)).$$

Thus,

$$d[\mathrm{N}_0, \mathrm{N}_0]_t = d\mathrm{N}_0(t)$$

$$d[\mathrm{N}_1, \mathrm{N}_1]_t = d\mathrm{N}_1(t)$$

$$d[\mathrm{N}_0, \mathrm{N}_1]_t = 0$$

As $d\mathrm{N}_0, d\mathrm{N}_1$ is zero or one, the product will also be zero or one. Also the cross product will always be zero, as N_0 and N_1 do not have jumps simultaneously.
Expand $d(\mathrm{G}_0 - \mathrm{G}_1, \pi_1)$:

$$\begin{aligned} d(\mathrm{G}_0 - \mathrm{G}_1, \pi_1) &= [(\alpha_{\mathrm{G}_0} - \alpha_{\mathrm{G}_1})dt + (\beta_{\mathrm{G}_0} + \beta_{\mathrm{G}_1})d\mathrm{N}_1(t) + (\gamma_{\mathrm{G}_0} + \gamma_{\mathrm{G}_1})d\mathrm{N}_2(t)]d\pi_1 \\ &= [(\alpha_{\mathrm{G}_0} - \alpha_{\mathrm{G}_1})dt + (\beta_{\mathrm{G}_0} + \beta_{\mathrm{G}_1})d\mathrm{N}_1(t) + (\gamma_{\mathrm{G}_0} + \gamma_{\mathrm{G}_1})d\mathrm{N}_2(t)] \\ &\qquad (a(t)dt + b(t,1)d\mathrm{N}_1(t) + b(t,2)d\mathrm{N}_2(t)) \\ &= (\beta_{\mathrm{G}_0} + \beta_{\mathrm{G}_1})b(t,1)d\mathrm{N}_1(t) + (\gamma_{\mathrm{G}_0} + \gamma_{\mathrm{G}_1})b(t,2)d\mathrm{N}_2(t) \end{aligned}$$

Thus,

$$\begin{aligned} d\mathrm{G} &= (\pi_1\alpha_{\mathrm{G}_0} + \pi_2\alpha_{\mathrm{G}_1})dt + (\pi_1\beta_{\mathrm{G}_0} + \pi_2\beta_{\mathrm{G}_1})d\mathrm{N}_1(t) + (\pi_1\gamma_{\mathrm{G}_0} + \pi_2\gamma_{\mathrm{G}_1})d\mathrm{N}_2(t) \\ &\quad + (\mathrm{G}_0 - \mathrm{G}_1)d\pi_1 + (\beta_{\mathrm{G}_0} + \beta_{\mathrm{G}_1})b(t,1)d\mathrm{N}_1(t) + (\gamma_{\mathrm{G}_0} + \gamma_{\mathrm{G}_1})b(t,2)d\mathrm{N}_2(t) \\ &= (\pi_1\alpha_{\mathrm{G}_0} + \pi_2\alpha_{\mathrm{G}_1})dt + (\pi_1\beta_{\mathrm{G}_0} + \pi_2\beta_{\mathrm{G}_1})d\mathrm{N}_1(t) + (\pi_1\gamma_{\mathrm{G}_0} + \pi_2\gamma_{\mathrm{G}_1})d\mathrm{N}_2(t) \\ &\quad + (\mathrm{G}_0 - \mathrm{G}_1)(a(t)dt + b(t,1)d\mathrm{N}_1(t) + b(t,2)d\mathrm{N}_2(t)) + (\beta_{\mathrm{G}_0} + \beta_{\mathrm{G}_1})b(t,1)d\mathrm{N}_1(t) \\ &\quad + (\gamma_{\mathrm{G}_0} + \gamma_{\mathrm{G}_1})b(t,2)d\mathrm{N}_2(t) \\ &= (\pi_1\alpha_{\mathrm{G}_0} + \pi_2\alpha_{\mathrm{G}_1} + \mathrm{G}_0a(t) - \mathrm{G}_1a(t))dt + (\pi_1\beta_{\mathrm{G}_0} + \pi_2\beta_{\mathrm{G}_1} + \mathrm{G}_0b(t,1) - \mathrm{G}_1b(t,1) \\ &\quad + \beta_{\mathrm{G}_0}b(t,1) + \beta_{\mathrm{G}_1}b(t,1))d\mathrm{N}_1(t) + (\pi_1\gamma_{\mathrm{G}_0} + \pi_2\gamma_{\mathrm{G}_1} + \mathrm{G}_0b(t,2) - \mathrm{G}_1b(t,2) \\ &\quad + \gamma_{\mathrm{G}_0}b(t,2) + \gamma_{\mathrm{G}_1}b(t,2))d\mathrm{N}_2(t) \\ &= \tilde{\alpha}_{\mathrm{F}_1}dt + \tilde{\beta}_{\mathrm{F}_1}d\mathrm{N}_1(t) + \tilde{\gamma}_{\mathrm{F}_1}d\mathrm{N}_2(t) \end{aligned}$$

Here, $\tilde{\alpha}_{\mathrm{F}_1}, \tilde{\beta}_{\mathrm{F}_1}$ and $\tilde{\gamma}_{\mathrm{F}_1}$ are the parameters corresponding to option O_1. Similarly, obtain

the parameters $\tilde{\alpha}_{F_2}, \tilde{\beta}_{F_2}$ and $\tilde{\gamma}_{F_2}$ corresponding to option O_2. Using the parameters corresponding to both the options in equation (6.12) form the self financing riskless portfolio as shown for constant intensity rate setting in Section 6.4.1.

6.5 EMPIRICAL RESULTS

We apply the modeling framework to the Indian stock market index Nifty50 corresponding to one-day and one-week option prices for an empirical illustration. For our study, we have downloaded the tick-by-tick asset price from Dukascopy Swiss Banking Group. The options data corresponding to the same asset is obtained from NSE India website. The time period of the data is 3.5 months, from Jun 2021 to Sep 2021. An initial period of 3 months of this data is a training sample, which we used for stable parameters estimation. The remaining dataset is used as a test sample.

We have shown Nifty option prices data on 1-Sep-2021 expiring on 2-Sep-2021 and 9-Sep-2021 corresponding to various strike prices. Model price is the price obtained from applying the methodology described in section 6.2. Black Scholes price is the price obtained from employing the famous Black Scholes model. Price of Nifty50 on 1-Sep-2021 is 17076.25. In Table 6.1 and 6.2, we also present the last traded prices to test our model. It is observed that our model gives better prediction compared Black Scholes model which is generally upward biased.

Table 6.1: Nifty option prices on 1-Sep-2021 expiring on 9-Sep-2021 compared to purposed model and Black Scholes prices.

Strike Price	Last Traded Price	Model Price	Black Scholes Price
15000	2132.8	2383.921	2502.19
15500	1572.60	1883.92	2153.53
16000	1080.8	1384.38	1836.41
16500	567.05	894.09	1551.68
17000	88	495.17	1299.31

Table 6.2: Nifty option prices on 1-Sep-2021 expiring on 9-Sep-2021 compared to purposed model and Black Scholes prices.

Strike Price	Last Traded Price	Model Price	Black Scholes Price
15500	1589.1	1559.58	1658.89
16000	1089.25	1059.58	1242.76
16500	595.45	559.58	883.42
17000	168.2	40.99	592.66
17500	14.05	3.63	373.83

CHAPTER 7

CONCLUDING REMARKS

We solve the mean-variance analysis in multidimensional, considering both the drift vector and covariance matrix of asset returns as uncertain and assumed to have prior distribution based on the past data. This prior information and Bayesian forecasting method are used to obtain the posterior distribution of asset returns at each time point. The dynamic programming approach and the embedding technique of Li and Ng (2000) are applied to these predicted returns. This approach of combining the Bayesian technique with dynamic programming allows us to find optimal portfolio weights based on historical as well as the latest available information of asset returns. In particular, we showed that the constructed Bayesian efficient frontier results in higher expected returns at a similar level of risk, presented by a numerical example. Another advantage of the Bayesian approach is that it allows us to construct the prediction interval of future realizations of the optimal portfolio returns. The numerical example also shows that the exit time have significant impacts on the optimal strategy and the efficient frontier. Simulation of the AR(1) process allows us to give more importance to the recent data and shows that the dynamic programming with the Bayesian approach efficiently captures the serial correlation of asset returns.

We introduces the two-level hierarchical Gaussian process for modeling time series of multivariate functional data. Our model considers not only auto-correlations in the functional responses, but also considers cross-correlations among the series. Simulation analysis indicates that the proposed methodology for estimating MFAR(p) produces improve prediction. We examine the Granger causality among the functional responses utilizing F-tests and Bayesian techniques. The positive F-statistics in the simulation analysis strongly indicate that such causal relationship exists in correlated series. To apply

Bayesian tests, we compare the marginal data densities of the restricted and unrestricted model to form a conclusion on the causal relationship among the functional time series. We implement the procedure on two real datasets. For the data on yield curves we find that the yield curve of USA is not Granger caused by that of UK. Nor is the yield curve of UK Granger caused by that of USA. In the pollution example we find that daily pollution level curves are indeed Granger caused by meteorological factors.

We have priced the options using a bivariate tick-by-tick asset price model with stochastic intensities, namely the Hawkes process. This model preserves the discrete nature of stock price movement and controls the market microstructure noise. We also derived the hedging strategy for an option using a stock, a bond, and another option. We considered two cases while deriving the hedging strategy, firstly, when the intensity rates are constant, and second, when the intensity rates are stochastic. We assumed that the parameter vector is a two-state Markov process for the stochastic intensities case and applied Bayesian filtration to obtain the optimal hedge ratios.

APPENDIX A

MULTI-PERIOD MV PORTFOLIO SELECTION

In chapter 3 instead of solving the orginial problem $\mathrm{P}(\omega)$ directly, we first consider the transformed auxiliary problem using the similar embedding technique as mentioned above:

$$\mathrm{A}(\lambda,\omega)\begin{cases}\max\limits_{\mathrm{U}} & \mathrm{E}_0(\lambda x_{\mathrm{T}\wedge\tau}-\omega x_{\mathrm{T}\wedge\tau}^2)\\ \text{s.t.} & x_{t+1}=r_t^0 x_t+e_t\mathrm{U}_t, \qquad t=0,1,...,\mathrm{T}-1,\end{cases}$$

for a given constant $\lambda>0$. The following two theorems can be proven by similar method described by Li and Ng (2000).

Theorem A.1. *Any optimal solution of* $\mathrm{P}(\omega)$ *will be optimal solution of* $\mathrm{A}(\lambda,\omega)$ *with* $\lambda^*=1+2\omega\mathrm{E}_0(x_{\mathrm{T}\wedge\tau})|_{\mathrm{U}^*}$.

Theorem A.2. *If* U^* *is the optimal solution of* $\mathrm{A}(\lambda,\omega)$ *then the necessary condition for* U^* *to be the optimal solution of* $\mathrm{P}(\omega)$ *is* $\lambda^*=1+2\omega\mathrm{E}_0(x_{\mathrm{T}\wedge\tau})|_{\mathrm{U}^*}$.

Zhang and Li (2012) derived the analytical solution of the auxiliary problem by transforming the auxiliary problem $\mathrm{A}(\lambda,\omega)$ into portfolio selection problem with certain exit time and applied dynamic programming approach to solve it, which is presented in the following theorem.

Theorem A.3. *The value function of problem* $\mathrm{A}(\lambda,\omega)$ *is*

$$\begin{aligned}f_t*(x_t)&=\max_{u_t} f_t(x_t)\\&=\max_{u_t}\mathrm{E}_t\left[\sum_{s=t}^{\mathrm{T}}(\lambda x_s-\omega x_s^2)p_s\right]\\&=-\omega\omega_t x_t^2+\lambda\lambda_t x_t+\Xi, \qquad t=0,1,...,\mathrm{T}-1,\end{aligned}$$

and the optimal strategy of problem $\mathrm{A}(\lambda, \omega)$ *is given by*

$$u_t* = \frac{\lambda}{2\omega}\frac{\mathrm{E}_t(\lambda_{t+1}e_t)}{\mathrm{E}_t(\omega_{t+1}e_t^2)} - \frac{\mathrm{E}_t(\omega_{t+1}e_t)}{\mathrm{E}_t(\omega_{t+1}e_t^2)}r_t^0 x_t, \qquad t = 0, 1, ..., \mathrm{T}-1,$$

where

$$\begin{aligned}
\Xi_t &= \mathrm{E}_t\left[\frac{\lambda^2}{4\omega}\sum_{s=t}^{\mathrm{T}-1}\theta_s\right], \\
\omega_t &= p_t + (r_t^0)^2\left[\mathrm{E}_t(\omega_{t+1}) - \frac{\mathrm{E}_t^2(\omega_{t+1}e_t)}{\mathrm{E}_t(\omega_{t+1}e_t^2)}\right], \quad \omega_{\mathrm{T}} = p_{\mathrm{T}}, \\
\lambda_t &= p_t + r_t^0\left[\mathrm{E}_t(\lambda_{t+1}) - \frac{\mathrm{E}_t(\omega_{t+1}e_t)\mathrm{E}_t(\lambda_{t+1}e_t)}{\mathrm{E}_t(\omega_{t+1}e_t^2)}\right], \quad \lambda_{\mathrm{T}} = p_{\mathrm{T}},
\end{aligned}$$

for $t = 0, 1, ..., \mathrm{T}-1$.

Now, insert the optimal strategy given in Theorem A.3 into wealth dynamics described in equation (4.1) for $t = \mathrm{T}$ and taking expectations on both sides based on the information available at time $\mathrm{T}-1$, we obtain

$$\mathrm{E}_{\mathrm{T}-1}(\lambda_{\mathrm{T}}x_{\mathrm{T}}) = \lambda_{\mathrm{T}-1}x_{\mathrm{T}-1} - p_{\mathrm{T}-1}x_{\mathrm{T}-1} + \frac{\lambda}{2\omega}\theta_{\mathrm{T}-1}. \tag{A.1}$$

Taking expectations recursively on both sides of equation A.1 at time $\mathrm{T}-2, ..., 1, 0$, we conclude that

$$\mathrm{E}_0(x_{\mathrm{T}\wedge\tau}) = \lambda_0 x_0 + \frac{\lambda}{2\omega}\Theta. \tag{A.2}$$

Substituting (A.2) in Theorem A.2 to satisfy the necessary condition of attaining the optimality of the original problem $\mathrm{P}(\omega)$ for the same optimal strategy of problem $\mathrm{A}(\lambda, \omega)$. We obtain

$$\lambda^* = \frac{1 + 2\omega\lambda_0 x_0}{1 - \Theta}. \tag{A.3}$$

Finally, substitute λ^* in Theorem A.3 to yield the optimal strategy for $\mathrm{P}(\omega)$ and efficient frontier, which is summarized in the Theorem 3.2.

APPENDIX B

MEAN SIGNATURE PLOT

For the right-sided exponential kernel as defined in equation (6.4), Bacry *et al.* (2013) obtained the closed form for mean signature plot. They presented the following proposition:

Proposition 1. For right-sided exponential kernel $\phi(t)$ with the stability condition (6.5), the mean signature plot is:

$$C(\tau) = \Lambda\left(\kappa^2 + (1-\kappa^2)\frac{1-e^{-\gamma\tau}}{\gamma\tau}\right) \tag{B.1}$$

where,

$$\Lambda = \frac{2\mu}{1-||\phi||_1}, \quad \kappa = \frac{1}{1+||\phi||_1}, \quad \text{and } \gamma = \omega + \delta. \tag{B.2}$$

To obtain the desired form of mean signature plot, we substitute the values of Λ, κ, and γ as follows:

$$\begin{aligned} C(\tau) &= \Lambda\left[\kappa^2 + (1-\kappa^2)\frac{1-e^{-\gamma\tau}}{\gamma\tau}\right] \\ &= \frac{2\mu}{1-||\phi||_1}\left[\left(\frac{1}{1+||\phi||_1}\right)^2 + \left(1-\left(\frac{1}{1+||\phi||_1}\right)^2\right)\frac{1-e^{-(\omega+\delta)\tau}}{(\omega+\delta)\tau}\right]. \end{aligned}$$

Now, substituting the stability condition ((6.5)), we obtain:

$$\begin{aligned} C(\tau) &= \frac{2\mu}{1-\omega/\delta}\left[\left(\frac{1}{1+\omega/\delta}\right)^2 + \left(1-\left(\frac{1}{1+\omega/\delta}\right)^2\right)\frac{1-e^{-(\omega+\delta)\tau}}{(\omega+\delta)\tau}\right] \\ &= \frac{2\mu\delta}{(\delta-\omega)(\omega+\delta)^3\tau}\left[\delta^2 + \omega(\omega+2\delta)e^{-(\omega+\delta)\tau}\right]. \end{aligned}$$

REFERENCES

1. **Aguilar, O.** and **M. West** (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, **18**(3), 338–357.

2. **Aitkin, M.** (1991). Posterior bayes factors. *Journal of the Royal Statistical Society: Series B (Methodological)*, **53**(1), 111–128.

3. **Allen, D. E.** and **V. Hooper** (2018). Generalized correlation measures of causality and forecasts of the vix using non-linear models. *Sustainability*, **10**(8), 2695.

4. **Anderson, R. W.** and **J.-P. Danthine** (1980). Hedging and joint production: Theory and illustrations. *The Journal of Finance*, **35**(2), 487–498.

5. **Aneja, V. P.**, **A. Agarwal**, **P. A. Roelle**, **S. B. Phillips**, **Q. Tong**, **N. Watkins**, and **R. Yablonsky** (2001). Measurements and analysis of criteria pollutants in new delhi, india. *Environment International*, **27**(1), 35–42.

6. **Arai, T.** (2005). An extension of mean-variance hedging to the discontinuous case. *Finance and Stochastics*, **9**(1), 129–139.

7. **Assaf, A. G.** and **M. Tsionas** (2018). Bayes factors vs. p-values. *Tourism Management*, **67**, 17–31.

8. **Aue, A.**, **D. D. Norinho**, and **S. Hörmann** (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, **110**(509), 378–392.

9. **Bacry, E.**, **S. Delattre**, **M. Hoffmann**, and **J.-F. Muzy** (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative finance*, **13**(1), 65–77.

10. **Bacry, E.**, **I. Mastromatteo**, and **J.-F. Muzy** (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, **1**(01), 1550005.

11. **Bacry, E.** and **J.-F. Muzy** (2016). First-and second-order statistics characterization of hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory*, **62**(4), 2184–2202.

12. **Balvers, R. J.** and **D. W. Mitchell** (1997). Autocorrelated returns and optimal intertemporal portfolio choice. *Management Science*, **43**(11), 1537–1551.

13. **Bardhan, I.** and **X. Chao** (1993). Pricing options on securities with discontinuous returns. *Stochastic processes and their applications*, **48**(1), 123–137.

14. **Bauder, D.**, **T. Bodnar**, **N. Parolya**, and **W. Schmid** (2018). Bayesian mean-variance analysis: Optimal portfolio selection under parameter uncertainty. *arXiv preprint arXiv:1803.03573*.

15. **Berger, J. O.** and **L. R. Pericchi** (1996). The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91**(433), 109–122.

16. **Besse, P. C.**, **H. Cardot**, and **D. B. Stephenson** (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, **27**(4), 673–687.

17. **Billio, M.**, **R. Casarin**, and **A. Osuntuyi** (2018). Markov switching garch models for bayesian hedging on energy futures markets. *Energy Economics*, **70**, 545–562.

18. **Bismuth, A.**, **O. Guéant**, and **J. Pu** (2019). Portfolio choice, portfolio liquidation, and portfolio transition under drift uncertainty. *Mathematics and Financial Economics*, **13**(4), 661–719.

19. **Blanchet-Scalliet, C.**, **N. El Karoui**, **M. Jeanblanc**, and **L. Martellini** (2008). Optimal investment decisions when time-horizon is uncertain. *Journal of Mathematical Economics*, **44**(11), 1100–1113.

20. **Bodnar, T.**, **S. Mazur**, and **Y. Okhrin** (2017). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research*, **256**(1), 292–307.

21. **Bosq, D.**, *Linear processes in function spaces: theory and applications*, volume 149. Springer Science & Business Media, Germany, 2000.

22. **Boudjellaba, H.**, **J.-M. Dufour**, and **R. Roy** (1994). Simplified conditions for noncausality between vectors in multivariate arma models. *Journal of Econometrics*, **63**(1), 271–287.

23. **Bowsher, C. G.** (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, **141**(2), 876–912.

24. **Carriero, A.**, **G. Kapetanios**, and **M. Marcellino** (2012). Forecasting government bond yields with large bayesian vector autoregressions. *Journal of Banking & Finance*, **36**(7), 2026–2047.

25. **Celikyurt, U.** and **S. Özekici** (2007). Multiperiod portfolio optimization models in stochastic markets using the mean–variance approach. *European Journal of Operational Research*, **179**(1), 186–202.

26. **Chen, Y.** and **B. Li** (2017). An adaptive functional autoregressive forecast model to predict electricity price curves. *Journal of Business & Economic Statistics*, **35**(3), 371–388.

27. **Chen, Y.**, **J. Marron**, and **J. Zhang** (2019). Modeling seasonality and serial dependence of electricity price curves with warping functional autoregressive dynamics. *The Annals of Applied Statistics*, **13**(3), 1590–1616.

28. **Chiang, W.-H.**, **X. Liu**, and **G. Mohler** (2022). Hawkes process modeling of covid-19 with mobility leading indicators and spatial covariates. *International journal of forecasting*, **38**(2), 505–520.

29. **Chiou, J.-M.**, **Y.-T. Chen**, and **Y.-F. Yang** (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 1571–1596.

30. **Chiu, M. C.** and **H. Y. Wong** (2014). Mean–variance portfolio selection with correlation risk. *Journal of Computational and Applied Mathematics*, **263**, 432–444.

31. **Comte, F.** and **O. Lieberman** (2000). Second-order noncausality in multivariate garch processes. *Journal of time series analysis*, **21**(5), 535–557.

32. **De Franco, C.**, **J. Nicolle**, and **H. Pham** (2018). Bayesian learning for the markowitz portfolio selection problem. *arXiv preprint arXiv:1811.06893*.

33. **Diebold, F. X.** and **C. Li** (2006). Forecasting the term structure of government bond yields. *Journal of econometrics*, **130**(2), 337–364.

34. **Dokuchaev, N.** (2007). Discrete time market with serial correlations and optimal myopic strategies. *European Journal of Operational Research*, **177**(2), 1090–1104.

35. **Dolinsky, Y.** and **H. M. Soner** (2014). Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, **160**(1), 391–427.

36. **Droumaguet, M.** and **T. Woźniak** (2012). Bayesian testing of granger causality in markov-switching vars.

37. **Duffie, D.** and **H. R. Richardson** (1991). Mean-variance hedging in continuous time. *The Annals of Applied Probability*, 1–15.

38. **Durbin, J.** and **S. J. Koopman** (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**(1), 3–56.

39. **Engle, R. F.** and **J. R. Russell** (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127–1162.

40. **Escobar, M. D.** and **M. West** (1995). Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, **90**(430), 577–588.

41. **Foster, F. D.** and **C. H. Whiteman** (2002). Bayesian cross hedging: an example from the soybean market. *Australian Journal of Management*, **27**(2), 95–122.

42. **Frost, P. A.** and **J. E. Savarino** (1986). An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, **21**(3), 293–305.

43. **Gao, J.** and **D. Li** (2014). Multiperiod mean-variance portfolio optimization with general correlated returns. *IFAC Proceedings Volumes*, **47**(3), 9007–9012.

44. **Gao, J.**, **D. Li**, **X. Cui**, and **S. Wang** (2015). Time cardinality constrained mean–variance dynamic portfolio selection and market timing: A stochastic control approach. *Automatica*, **54**, 91–99.

45. **Gao, R.**, **Y. Li**, **Y. Bai**, and **S. Hong** (2019). Bayesian inference for optimal risk hedging strategy using put options with stock liquidity. *IEEE Access*, **7**, 146046–146056.

46. **Geweke, J.** (1999). Using simulation methods for bayesian econometric models: inference, development, and communication. *Econometric reviews*, **18**(1), 1–73.

47. **Granger, C. W.** (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, **37**(3), 424–438.

48. **Guo, W.** and **Q. Hu** (2005). Multi-period portfolio optimization when exit time is uncertain. *Journal of Management Sciences in China*, **8**(2), 14–19.

49. **Guyet, T.** and **H. Nicolas** (2016). Long term analysis of time series of satellite images. *Pattern Recognition Letters*, **70**, 17–23.

50. **Hafner, C. M.** and **H. Herwartz** (2008). Testing for causality in variance using multivariate garch models. *Annales d'Economie et de Statistique*, **89**, 215–241.

51. **Hawkes, A. G.** (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**(1), 83–90.

52. **Hayashi, T.** and **P. A. Mykland** (2005). Evaluating hedging errors: an asymptotic approach. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, **15**(2), 309–343.

53. **He, J.**, **Q.-G. Wang**, **P. Cheng**, **J. Chen**, and **Y. Sun** (2015). Multi-period mean-variance portfolio optimization with high-order coupled asset dynamics. *IEEE Transactions on Automatic Control*, **60**(5), 1320–1335.

54. **Heck, D. W.**, **U. Boehm**, **F. Böing-Messing**, **P.-C. Bürkner**, **K. Derks**, **Z. Dienes**, **Q. Fu**, **X. Gu**, **D. Karimova**, **H. A. Kiers**, *et al.* (2022). A review of applications of the bayes factor in psychological research. *Psychological Methods*.

55. **Horváth, L.**, **P. Kokoszka**, and **G. Rice** (2014). Testing stationarity of functional time series. *Journal of Econometrics*, **179**(1), 66–82.

56. **Huang, D.**, **S. Zhu**, **F. J. Fabozzi**, and **M. Fukushima** (2010). Portfolio selection under distributional uncertainty: A relative robust cvar approach. *European Journal of Operational Research*, **203**(1), 185–194.

57. **Huang, D.**, **S.-S. Zhu**, **F. J. Fabozzi**, and **M. Fukushima** (2008). Portfolio selection with uncertain exit time: a robust cvar approach. *Journal of Economic Dynamics and Control*, **32**(2), 594–623.

58. **Irie, K.**, **M. West**, *et al.* (2019). Bayesian emulation for multi-step optimization in decision problems. *Bayesian Analysis*, **14**(1), 137–160.

59. **Iyengar, G. N.** (2005). Robust dynamic programming. *Mathematics of Operations Research*, **30**(2), 257–280.

60. **Karatzas, I.** and **X. Zhao** (2001). Bayesian adaptive portfolio optimization. *Option pricing, interest rates and risk management*, 632–669.

61. **Kass, R. E.** and **A. E. Raftery** (1995). Bayes factors. *Journal of the american statistical association*, **90**(430), 773–795.

62. **Koopman, S. J.**, **M. I. Mallee**, and **M. Van der Wel** (2010). Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, **28**(3), 329–343.

63. **Kowal, D. R.**, **D. S. Matteson**, and **D. Ruppert** (2019). Functional autoregression for sparsely sampled data. *Journal of Business & Economic Statistics*, **37**(1), 97–109.

64. **Kumar, A.** and **P. Goyal** (2011). Forecasting of air quality in delhi using principal component regression technique. *Atmospheric Pollution Research*, **2**(4), 436–444.

65. **Laub, P. J.**, **T. Taimre**, and **P. K. Pollett** (2015). Hawkes processes. *arXiv preprint arXiv:1507.02822*.

66. **Lence, S. H.** and **D. J. Hayes** (1994). The empirical minimum-variance hedge. *American Journal of Agricultural Economics*, **76**(1), 94–104.

67. **Li, C.**, **L. Xiao**, and **S. Luo** (2020*a*). Fast covariance estimation for multivariate sparse functional data. *Stat*, **9**(1), e245.

68. **Li, D.** and **W.-L. Ng** (2000). Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical finance*, **10**(3), 387–406.

69. **Li, D.**, **P. M. Robinson**, and **H. L. Shang** (2020*b*). Long-range dependent curve time series. *Journal of the American Statistical Association*, **115**(530), 957–971.

70. **Li, J.**, **C. Huang**, **Z. Hongtu**, and **A. D. N. Initiative** (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, **112**(519), 1169–1181.

71. **Mao, J. C.** and **C. E. Särndal** (1966). A decision theory approach to portfolio selection. *Management Science*, **12**(8), B–323.

72. **Markowitz, H.** (1952). Portfolio selection. *The journal of finance*, **7**(1), 77–91.

73. **Martellini, L.** and **B. Urošević** (2006). Static mean-variance analysis with uncertain time horizon. *Management Science*, **52**(6), 955–964.

74. **McCulloch, R. E.** and **R. S. Tsay** (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the american Statistical association*, **88**(423), 968–978.

75. **Meng, C.**, **T. Cheng**, **X. Gu**, **S. Shi**, **W. Wang**, **Y. Wu**, and **F. Bao** (2019). Contribution of meteorological factors to particulate pollution during winters in beijing. *Science of The Total Environment*, **656**, 977–985.

76. **Merton, R. C.** (1972). An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis*, **7**(4), 1851–1872.

77. **Mykland, P. A.** (2000). Conservative delta hedging. *Annals of Applied Probability*, 664–683.

78. **O'Hagan, A.** (1995). Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**(1), 99–118.

79. **Pai, J. S.** and **N. Ravishanker** (1998). Bayesian analysis of autoregressive fractionally integrated moving-average processes. *Journal of Time Series Analysis*, **19**(1), 99–112.

80. **Pham, H.** (2000). On quadratic hedging in continuous time. *Mathematical Methods of Operations Research*, **51**(2), 315–339.

81. **Raggi, D.** and **S. Bordignon** (). Comparing stochastic volatility models using the bayes factor.

82. **Ramsay, J.** and **B. Silverman**, *Functional data analysis*. Springer Series in Statistics, New York, 2005.

83. **Ramsay, J. O.** and **C. Dalzell** (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, **53**(3), 539–561.

84. **Saumard, M.**, Linear causality in the sense of granger with stationary functional time series. *In Functional Statistics and Related Fields*. Springer, Switzerland, 2017, 225–231.

85. **Sen, R.** (2009). Hedging options in the incomplete market with stochastic volatility. *Statistics and Its Interface*, **2**(4), 469–479.

86. **Sen, R.** and **C. Klüppelberg** (2019). Time series of functional data with application to yield curves. *Applied Stochastic Models in Business and Industry*, **35**(4), 1028–1043.

87. **Shang, H. L.**, **K. Ji**, and **U. Beyaztas** (2021). Granger causality of bivariate stationary curve time series. *Journal of Forecasting*, **40**(4), 626–635.

88. **Singer, P.**, **D. Helic**, **A. Hotho**, and **M. Strohmaier**, Hyptrails: A bayesian approach for comparing hypotheses about human trails on the web. *In Proceedings of the 24th International Conference on World Wide Web*. 2015.

89. **Skaf, J.** and **S. Boyd**, Multi-period portfolio optimization with constraints and transaction costs. *In Working Manuscript*. Citeseer, 2009.

90. **Stoehr, C.**, **J. A. Aston**, and **C. Kirch** (2021). Detecting changes in the covariance structure of functional time series with application to fmri data. *Econometrics and Statistics*, **18**, 44–62.

91. **Swishchuk, A. V.** and **Y. Kazmerchuk** (2001). Stability of stochastic differential delay itô's equations with poisson jumps and with markovian switchings. application to financial models. *Teor. Imovir. Mater. Stat*, **64**, 141–151.

92. **Wei, C.** and **J. Gao**, Dynamic mean-variance portfolio optimization with noshorting constraint and correlated returns. *In The 27th Chinese Control and Decision Conference (2015 CCDC)*. IEEE, 2015.

93. **West, M.** and **J. Harrison**, *Bayesian forecasting and dynamic models*. Springer Science & Business Media, 2006.

94. **Woźniak, T.** (2015). Testing causality between two vectors in multivariate garch models. *International Journal of Forecasting*, **31**(3), 876–894.

95. **Xu, Y.-h.** and **Z.-f. Li** (2008). Dynamic portfolio selection based on serially correlated return-dynamic mean-variance formulation. *System Engineering Theory and Practice*, **28**(8), 123–131.

96. **Yi, L.**, **Z. Li**, and **D. Li** (2008). Multi-period portfolio selection for asset-liability management with uncertain investment horizon. *Journal of Industrial & Management Optimization*, **4**(3), 535–552.

97. **Zhang, L.** and **Z. Li** (2012). Multi-period mean-variance portfolio selection with uncertain time horizon when returns are serially correlated. *Mathematical Problems in Engineering*, **2012**.

98. **Zhou, X. Y.** and **D. Li** (2000). Continuous-time mean-variance portfolio selection: A stochastic lq framework. *Applied Mathematics and Optimization*, **42**(1), 19–33.

99. **Zhu, H.**, **R. Li**, and **L. Kong** (2012). Multivariate varying coefficient model for functional responses. *Annals of statistics*, **40**(5), 2634.

100. **Zhu, H.**, **J. S. Morris**, **F. Wei**, and **D. D. Cox** (2017). Multivariate functional response regression, with application to fluorescence spectroscopy in a cervical pre-cancer study. *Computational statistics & data analysis*, **111**, 88–101.

101. **Zhu, H.**, **N. Strawn**, and **D. B. Dunson** (2016). Bayesian graphical models for multivariate functional data. *Journal of Machine Learning Research*, **17**, 1–27.

CURRICULUM VITAE

NAME Shubhangi Sikaria

DATE OF BIRTH 10 July 1994

EDUCATION QUALIFICATIONS

2016 **Master of Science**

Institution	BITS Pilani
Specialization	Mathematics

2018 **Master of Technology**

Institution	IIT Madras
Specialization	Industrial Mathematics and Scientific Computing

Doctor of Philosophy

Institution	IIT Madras
Specialization	Statistical Finance
Registration Date	28 July 2016

DOCTORAL COMMITTEE

Chairperson

Prof. V. Vetrivel
Head of the Department
Department of Mathematics
IIT Madras

Guide(s)

Dr. Neelesh S Upadhye
Associate Professor
Department of Mathematics
IIT Madras

Dr. Rituparna Sen
Associate Professor
Applied Statistics Unit
ISI Bangalore

Member(s)

Prof. Sundar S (Departmental Member)
Professor
Department of Mathematics
IIT Madras

Dr. Sarang S Sane (Departmental Member)
Assistant Professor
Department of Mathematics
IIT Madras

Dr. John Augustine (Member from other Department)
Associate Professor
Department of Computer Science & Engineering
IIT Madras