Gaussian process regression methods and extensions for stock market prediction

Thesis submitted for the degree of

Doctor of Philosophy

at the

University of Leicester

by

Zexun Chen

Department of Mathematics

University of Leicester

August 2017

UNIVERSITY OF LEICESTER

"Your time is limited, so don't waste it living someone else's life. Don't be trapped by dogma - which is living with the results of other people's thinking. Don't let the noise of others' opinions drown out your own inner voice. And most important, have the courage to follow your heart and intuition."

Steve Jobs

"Mathematics is a game played according to certain simple rules with meaningless marks on paper."

David Hilbert

"It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment."

Carl Friedrich Gauss

Abstract

Gaussian process regression (GPR) is a kernel-based nonparametric method that has been proved to be effective and powerful in many areas, including time series prediction. In this thesis, we focus on GPR and its extensions and then apply them to financial time series prediction. We first review GPR, followed by a detailed discussion about model structure, mean functions, kernels and hyper-parameter estimations. After that, we study the sensitivity of hyper-parameter and performance of GPR to the prior distribution for the initial values, and find that the initial hyper-parameters' estimates depend on the choice of the specific kernels, with the priors having little influence on the performance of GPR in terms of predictability. Furthermore, GPR with Student$-t$ process (GPRT) and Student$-t$ process regression (TPR), are introduced. All the above models as well as autoregressive moving average (ARMA) model are applied to predict equity indices. We find that GPR and TPR shows relatively considerable capability of predicting equity indices so that both of them are extended to state-space GPR (SSGPR) and state-space TPR (SSTPR) models, respectively. The overall results are that SSTPR outperforms SSGPR for the equity index prediction. Based on the detailed results, a brief market efficiency analysis confirms that the developed markets are unpredictable on the whole. Finally, we propose and test the multivariate GPR (MV-GPR) and multivariate TPR (MV-TPR) for multi-output prediction, where the model settings, derivations and computations are all directly performed in matrix form, rather than vectorising the matrices involved in the existing method of GPR for multi-output prediction. The effectiveness of the proposed methods is illustrated through a simulated example. The proposed methods are then applied to stock market modelling in which the Buy&Sell strategies generated by our proposed methods are shown to be profitable in the equity investment.

Acknowledgements

It is pleasure to acknowledge the help and support I have received from many people during the four-year PhD study at the University of Leicester. I would like to take this opportunity to acknowledge a few.

Firstly, I would like to express my utmost gratitude to my supervisors, Professor Alexander N. Gorban and Dr Bo Wang, for their supports, time and patience, who also have been my inspiration throughout my studies. I must say thanks to them for all of the help, encouragement and suggestions which are by no means limited to my research but have also proven valuable in my life. This thesis draws on their talents, knowledge and contribution.

I would like to thank all of the staff in the Department of Mathematics for their help and making so many things easier. I am grateful to Dr Evgeny Mirkes and Prof. Sergei Petrovskii.

On a personal level, I would like to thank all of my friends in Michael Atiyah building, Juxi, Ruhao, Peter etc. A special thanks goes to Yanshan, Wenyan and Isaudin Ismail who provided all the ultimate care, support and enjoyable time.

Finally, it gives me the most pleasure to thank my parents for the love, faith and support as well as my girlfriend Yujin. Without their help, this thesis would not be possible.

Contents

List of Figures

List of Tables

Abbreviations

ARD	Automatic Relevance Determination
ARMA	Autoregressive Moving Average
ARMSE	Average Root Mean Square Error
CG	Conjugate Gradient
EM	Expectation Maximization
EMH	Efficient Market Hypothesis
GMV-GP	Matrix-variate Gaussian Process
GMV-GPR	Matrix-variate Gaussian Process Regression
GMV-TP	Matrix-variate Student$-t$ Process
GMV-TPR	Matrix-variate Student$-t$ Process Regression
GP	Gaussian Process
GPR	Gaussian Process Regression
GPRT	Gaussian Process Regression with Student$-t$ Likelihood
LIN	Linear
LL	Log Loss
LOO-CV	Leave-one-out Cross Validation
LP	Local Periodic
MAE	Mean Absolute Error
MAER	Mean Absolute Error Ratio
MSE	Mean Squared Error
MSER	Mean Squared Error Ratio
MSLL	Mean Standardized Log Loss
MV-GP	Matrix-variate Gaussian Process
MV-TP	Matrix-variate Student$-t$ Process

PCA	Principal Component Analysis
PER	Periodic
PSD	Positive semi-definite
RMSE	Root Mean Square Error
RQ	Rational Quadratic
SE	Squared Exponential
SM	Spectral Mixture
SRMSE	Standardized Root Mean Square Error
SSGPR	State-Space Gaussian Process Regression
SSTPR	State-Space Student$-t$ Process Regression
SVM	Support Vector Machine
TP	Student$-t$ process
TPR	Student$-t$ Process Regression
VB	Variational Bayesian

Ticker Symbols

Ticker	Index (Exchange)
DAX	Deutscher Aktienindex (Frankfurt Stock Exchange, Germany)
HSI	Hang Seng Index (Hong Kong Stock Exchange, Hong Kong)
INDU	Dow Jones Industrial Average (New York & NASADAQ Stock Exchange, US)
NDX	NASDAQ 100 Index (NASDAQ Stock Exchange, US)
NKY	Nikkei 225 (Tokyo Stock Exchange, Japan)
SENSEX	Standard & Poor's Bombay Stock Exchange Sensitive Index (Bombay Stock Exchange, India)
SHSZ300	China Securities Index (Shanghai&Shenzhen Stock Exchange, China)
SPX	Standard & Poor's 500 (New York & NASADAQ Stock Exchange, US)
UKX	Financial Times Stock Exchange 100 Index (London Stock Exchange, UK)
XU100	Borsa Istanbul 100 Index (Istanbul Stock Exchange, Turkey)

Ticker	Company (Exchange)
BIDU	Baidu, Inc. (NASDAQ)
CTRP	Ctrip.com International, Ltd. (NASDAQ)
NTES	NetEase, Inc. (NASDAQ)

MMM	3M (NYSE)
AXP	American Express (NYSE)
AAPL	Apple (NASDAQ)
BA	Boeing (NYSE)
CAT	Caterpillar (NYSE)
CVX	Chevron (NYSE)
CSCO	Cisco Systems (NASDAQ)
KO	Coca-Cola (NYSE)
DD	DuPont (NYSE)
XOM	ExxonMobil (NYSE)
GE	General Electric (NYSE)
GS	Goldman Sachs (NYSE)
HD	The Home Depot (NYSE)
IBM	IBM (NYSE)
INTC	Intel (NASDAQ)
JNJ	Johnson & Johnson (NYSE)
JPM	JPMorgan Chase (NYSE)
MCD	McDonald's (NYSE)
MRK	Merck (NYSE)
MSFT	Microsoft (NASDAQ)
NKE	Nike (NYSE)
PFE	Pfizer (NYSE)
PG	Procter & Gamble (NYSE)
TRV	Travelers (NYSE)
UNH	UnitedHealth Group (NYSE)
UTX	United Technologies (NYSE)
VZ	Verizon (NYSE)
V	Visa (NYSE)
WMT	Wal-Mart (NYSE)
DIS	Walt Disney (NYSE)

Notations

Generalities

$\mathbb{R}$	real number set
$\mathbb{N}$	nature number set
$\mathcal{X}$	Input space
p	dimensionality of the input space
d	number of outputs
n	number of data points per output
m	number of test (predictive) points
ℓ	input scale of the kernels based on SE and PER
s_f^2	output-scale amplitude of the kernels
p	period term of the kernels with periodicity
x(or $\boldsymbol{x}$)	input point, 1-dimension (p-dimension)
y(or y)	target point, 1-dimension (d-dimension)
z(or $\boldsymbol{z}$)	input point, 1-dimension (p-dimension)
$\mathbb{X}$	set of training input data, $X = \{x_i\}_{i=1}^n$ or $\{\boldsymbol{x}_i\}_{i=1}^n$
$\mathbb{Z}$	set of test (predictive) inputs, $Z = \{z_i\}_{i=1}^m$ or $\{\boldsymbol{z}_i\}_{i=1}^m$
$\mathbb{Y}$	set of training target data, $\mathbb{Y} = \{y_i\}_{i=1}^n$ or $\{\mathrm{y}_i\}_{i=1}^n$
$\mathcal{D}$	training set, $\mathcal{D} = \{\mathbb{X}, \mathbb{Y}\}$

Operators

$\mathbb{E}[\cdot]$	expected value
$\mathrm{cov}[\cdot, \cdot]$	covariance operator
$\mathrm{tr}(\cdot)$	trace of a matrix
$\mathrm{etr}(\cdot)$	exponential trace of a matrix

L	Lag operator
$\langle \cdot, \cdot \rangle$	inner product
$\lfloor \cdot \rfloor$	largest integer toward negative infinity
$\det(\cdot)$	determinant of a square matrix.
$\log(\cdot)$	natural logarithm of a positive number
$\propto$	proportional to
$\mathrm{vec}(\cdot)$	vectorization of a matrix
$\mathbf{A}^{\mathrm{T}}$(or $\mathbf{A}^*$)	transpose (conjugate transpose) of matrix $\mathbf{A}$
$\mathbf{A} \otimes \boldsymbol{B}$	Kronecker product between matrices $\mathbf{A}$ and $\mathbf{B}$

Functions

$f(x)$ or $f(\boldsymbol{x})$	real-valued function evaluated at x or $\boldsymbol{x}$
$\boldsymbol{f}(x)$ or $\boldsymbol{f}(\boldsymbol{x})$	vector-valued function, $\boldsymbol{f}(x) = [f_1(x), \ldots, f_d(x)]$ or $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), \ldots, f_d(\boldsymbol{x})]$
$\mu(x)$ or $\mu(\boldsymbol{x})$	real-valued mean function evaluated at x or $\boldsymbol{x}$
$\boldsymbol{u}(x)$ or $\boldsymbol{u}(\boldsymbol{x})$	vector-valued mean function, $\boldsymbol{u}(x) = [\mu_1(x), \ldots, \mu_d(x)]$ or $\boldsymbol{u}(\boldsymbol{x}) = [\mu_1(\boldsymbol{x}), \ldots, \mu_d(\boldsymbol{x})]$
$k(x, x')$ or $k(\boldsymbol{x}, \boldsymbol{x}')$	covariance (kernel) function evaluated at (x, x') or $(\boldsymbol{x}, \boldsymbol{x}')$
δ_{ij}	Kronecker delta, $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$

Matrices

X	training inputs, $X = [x_1, \cdots, x_n]^{\mathrm{T}}$ or $[\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]^{\mathrm{T}}$
$\boldsymbol{y}$ or Y	training targets, $\boldsymbol{y} = [y_1, \cdots, y_n]^{\mathrm{T}}$ or $Y = [\mathrm{y}_1^{\mathrm{T}}, \cdots, \mathrm{y}_n^{\mathrm{T}}]^{\mathrm{T}}$
Z	test (predictive) inputs, $Z = [z_1, \cdots, z_n]^{\mathrm{T}}$ or $[\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n]^{\mathrm{T}}$
$\boldsymbol{\mu}$	mean vector, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^{\mathrm{T}}$
M	$n \times d$ mean matrix with $M_{ij} = \mu_j(x_i)$ or $\mu_j(\boldsymbol{x}_i)$
Σ	covariance matrix (inputs), $\Sigma_{ij} = k(x_i, x_j)$ or $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$
Ω	scale covariance matrix (outputs)
$\mathbf{I}_n$	identity matrix of size n

Distributions

$\mathrm{Uniform}(a, b)$	uniform distribution in the range of (a,b)

$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$	(multivariate) normal distribution with parameter $\boldsymbol{\mu}, \Sigma$
$\mathcal{T}(\nu, \boldsymbol{\mu}, \Sigma)$	(multivariate) Student$-t$ distribution with parameter ν, $\boldsymbol{\mu}, \Sigma$
$\mathcal{MN}(M, \Sigma, \Omega)$	matrix-variate normal distribution with parameter M, Σ, Ω
$\mathcal{MT}(\nu, M, \Sigma, \Omega)$	matrix-variate Student$-t$ distribution with parameter ν, M, Σ, Ω

Processes

$\mathcal{GP}$	Gaussian process
$\mathcal{TP}$	Student$-t$ process
$\mathcal{MGP}$	matrix-variate Gaussian process
$\mathcal{MTP}$	matrix-variate Student$-t$ process

To my family

Chapter 1

Introduction

1.1 The predictability of financial time series

A time series is a series of data points indexed in time order. Time series analysis, which comprises methods for analysing time series in order to capture meaningful statistics and other information of the data [1], has many applications in economics and finance. The analysis of financial time series is involved with the theory and the practice of financial market valuation over time [2]. Unlike other time series analysis, uncertainty is a key feature, with statistical theory and methods playing an essential role in financial time series [2]. One of the most noteworthy problems is forecasting. The predictability of financial time series has to be explicit before prediction. Indeed, the effectiveness of forecasting financial time series in various markets leads to a heated debate [3].

Undoubtedly, investors are particularly eager to find a useful method to forecast the future and generate extraordinary profits. However, the efficient market hypothesis (EMH) given by Eugene Fama in 1970, significantly hits their endeavours. In the last half century, EMH, which describes the random walk behaviour of price in financial markets, has been accepted by many economics and finance scholars [4]. The EMH asserts that financial markets are "informationally efficient", meaning that information will be immediately incorporated into the price when news appears because of the efficient spread of information [4]. As a result, investors cannot consistently obtain excess returns on a risk-adjusted basis, given the information available at the time that the investment is made.

In summary, the EMH briefly contains the following three main points:

(i) In the market, everyone is rational and every stock market company is under strict monitoring by rational people. Investors can utilise available information to generate higher rates of remuneration.

(ii) The reaction of security markets to new market information is rapid and accurate, with stock prices fully reflecting all the available information.

(iii) Competition in the market leads the stock price from the old equilibrium transition to a new equilibrium, while the new information corresponding to the price change is independent and random.

Besides, the EMH can be divided into three forms (weakly, semi-strongly and strongly efficient markets) depending on the degree of information efficiency. The details of these three forms are presented in the following.

(i) In a weakly efficient market, stock prices already reflect all past information, which can be derived by examining trading such as historical prices, volumes, short interest and so on. That is to say, the future price cannot be predicted by analysing the past, with even technical analysis useless in a weak efficient market [5]. However, a fundamental analysis is still valid.

(ii) In a semi-strongly efficient market, stock prices already reflect all publicly available information, including price-earnings ratios, cash flows earnings forecasts, company management and so on. As a result, neither fundamental analysis nor technical analysis can reliably produce excess returns. However, insider information can produce abnormal profits [5, 6].

(iii) In a strongly efficient market, stock prices already reflect all relevant information to a firm, and even includes insider information. That is, the stock price is immediate self-adjusting and reflects all available private and public information. As a result, no one can earn excess returns, including fund managers or investors who have a monopolistic access to information [5].

The technical analysis, fundamental analysis and insider information have to be verified by testing whether the market is weakly, semi-strongly or strongly efficient, respectively. Technical analysis consists of a variety of forecasting techniques based on historical trading data, such as chart analysis, pattern recognition analysis and computerised technical trading systems. Fundamental analysis, in accounting and finance, is the analysis of a business's financial statements (usually to analyze the business's assets, liabilities, and earnings) and its competitors and markets.

As mentioned, technical analysis contradicts weak form efficiency but it is widely used by many investors. Over last decades, the dominance of the EMH in the research field has become far less prominent [4] at the beginning of the 21st century. An increasing number of finance and economics scholars have started to believe that financial markets are partially predictable [4]. The technical analysis researchers even suggest that a suitable strategy could be built into financial markets to generate extra profits (see, e.g., [6, 7, 8]).

The psychological and behavioural elements are gradually being considered as the increasing prominence in the determination of stock price, with future stock prices somewhat more predictable based on historical price patterns and certain fundamental valuations [4]. In 2004, Lo offered an alternative market theory to EMH from a behavioural perspective, according to which markets are adaptable and switch between efficiency and inefficiency at different periods [9].

Of course, the predictability of time series strongly depends on the efficiency of the market, with a large volume of literature existing on the predictability of financial time series using diverse methods in various markets. For instance, Ankit Agarwal tested the weak form of Indian stock market by using simple technical trading rules to find high predictability of Indian stock markets considering transaction costs [5]. Yanshan Shi discusses the predictability of kNN based on four equity indices (FTSE100, DAX, HANGSENG and NASDAQ), highlighting that the future index can be predicted by using historical information in the HANGSENG market. Using a technical analysis probably leads to generating excess profits in this case [10]. Other prevailing methods, such as dividend yields [11], support vector machines [12], Bayesian models [13, 14], are also widely used to analyse the predictability of financial time series and have the capacity of making considerable predictions in several major equity markets over the world. Actually, the debate about EMH and technical analysis is located in whether historical trading data can help investors consistently generate excess profits. An interesting study of 20 new equity markets in emerging economies indicates that developing markets can gain more expected returns and have higher volatility when compared with developed countries' markets [15]. That is to say, more opportunities in emerging economies can be achieved by investors taking advantage of technical analysis based on historical data because these markets are more inefficient.

The predictability of financial time series is also associated with test methods including in-sample and out-of-sample. Empirical studies show that in-sample tests have relatively strong predictability, while out-of-sample tests are weak. This is because an out-of-sample (extrapolation) test may fail to discover predictability in a population, while the in-sample (interpolation) test can correctly discover it (see, e.g., [4, 13, 16, 17]).

As a result, it is also essential to analyse the predictability of financial time series and to investigate prediction problems when predictability is positive because traders can take advantage of effective predictions, such as signalling of changes and trends, to discover the trading points or detect crises so that they can make critical preparations before an extreme situation occurs [18]. For instance, correlation and variance will be gradually higher before a crisis, making it a valid signal of an impending financial crisis [19, 20, 21].

1.2 Gaussian process for machine learning and its extensions for financial time series prediction

Over the last few decades, Gaussian processes regression (GPR) has been proven to be a powerful and effective method for non-linear regression problems due to many desirable properties, such as ease of obtaining and expressing uncertainty in predictions, the ability to capture a wide variety of behaviour through a simple parameterisation and a natural Bayesian interpretation [22]. Neal [23] reveals that many Bayesian regression models based on neural networks converge to GPs in the limit of an infinite number of hidden units [24]. GPs have been suggested as a replacement of supervised neural networks in non-linear regression [25, 26] and classification [25].

In particular, GPR has an excellent capability of forecasting time series [27, 28, 29]. As a powerful non-parameter tool, it has been widely used in financial market prediction and has shown a superior ability in forecasting [29, 30, 31, 32]. Forecasting financial time series is an attractive topic for investors and scholars since a financial market is a complicated dynamic system with a huge volume of time series data. An effective prediction can help investors generate excess profits. However, the EMH developed by Fama asserts that one cannot constantly obtain excess returns from an efficient market [33]. In other words, historical data cannot be used to predict and beat the efficient market. Nevertheless, the predictability of financial markets remains a hot topic [3].

With the development of financial market theory, many researchers have already found evidence that the distribution of financial time series is not Gaussian [34, 35]. Specifically, the empirical distributions of financial data have heavier tails in the two sides than those from a Gaussian distribution [34]. As a result, some fat-tailed distributions, such as Student$-t$ distribution, Pareto distribution, Lévy distribution [36], and the family of stable distributions, are applied in various financial time series models [34]. Some heavy-tailed distributions are used in the extension of GPR. For instance, in 2009, Vanhatalo et al. pointed out that Gaussian likelihood of GPR models can be replaced by Student$-t$ likelihood to adapt the heavily-tails of financial data, namely GPR with the Student$-t$ likelihood (GPRT) model. Although the Student$-t$ likelihood is considered in GPRT, which can reduce the influence of outlying observations and improve the prediction, the latent process is still GP [37]. If the latent GP is replaced by Student$-t$ process (TP), GPR model is extended to a Student$-t$ process regression (TPR) model, which is used to capture the fat-tails. Shah et al.[38] show that TP can be an alternative to GP as a non-parametric method in prediction problems because TP can retain the desirable properties of a GP model, such as non-parametric representation given a known kernel, analytic marginal and predictive distribution, and easy model choice based on covariance functions [38]. For example, in 2015, Arno Solin and Simo Särkkä predicted the share price of Apple Inc. using both GPR and TPR, and the result is that TPR could have a comparatively better performance than GPR.

Although EMH claims that the current price reflects all the past information on a stock in an efficient market. However, not all the markets are efficient. The current price cannot be independent of historical prices for all the markets [3]. In other words, historical prices are able to particularly influence the current price. The prediction of price tomorrow can be regarded as a function of previous prices and today's price; that is, it is state-space model, where the state consists of the historical prices. It is a natural and direct idea for traders and investors because they all want to generate excess profits from the analysis of historical data, including trading information and operation records of firms. As a result, both GPR and TPR are extended to state-space GPR (SSGPR) and state-space TPR (SSTPR) models given by [39]. For stock market prediction using SSGPR and SSTPR, the historical prices can be treated as the state in the state-space model to predict the next few days' prices using the iterative multi-step-ahead method.

Despite the popularity of GPR in various modelling tasks, there still exists a conspicuous imperfection, that is, the majority of GPR models are implemented for single response variables or considered independently for multiple responses variables without consideration of their correlation [40, 41]. However, some correlations between financial time series cannot be ignored. For example, "Chinese concept stocks", which refer to the stock issued by firms whose asset or earning have essential activities in Mainland China, are heavily influenced by the political and economic environment of China together. For this reason, all these stocks have the potential and unneglectable correlation theoretically, which is probably reflected in the movement of the stock price. Furthermore, the diverse industrial sector, which includes some stocks in a similar industrial group, usually has a joint price trend or distribution during a period. The correlations between the stocks in the same industrial sector should not be ignored and the predictions for the whole industrial sector together are desirable. In order words, multi-output problems have to be considered in financial time series prediction.

In order to resolve the multi-output prediction problem, Gaussian process regression for vector-valued functions regarded as a pragmatic and straightforward method, is proposed. The core of this method is to vectorise the multi-response variables and to construct a "big" covariance, which describes the correlations between the inputs as well as between the outputs [40, 41, 42, 43]. This modelling strategy is feasible due to that the matrix-variate Gaussian distributions can be reformulated as multivariate Gaussian distributions [42, 44]. Intrinsically, Gaussian process regression for vector-valued functions is still a conventional Gaussian process regression model since it merely vectorises multi-response variables of which are assumed to follow a developed case of GP with a "big" kernel.

One important issue is how to apply the results of prediction models to the financial market investments. For example, it is known that the accurate prediction of the future for an equity market is almost impossible. Admittedly, the more realistic idea is to make a strategy based on the Buy&Sell signals in different prediction models [45]. The Buy&Sell strategy based on an effective prediction models should be required to have more profits than Buy&Hold strategy, which means buying shares of stock without selling.

1.3 Contributions and outline of thesis

The main topic of this thesis is to develop Gaussian process regression methods and then apply them to stock market prediction. The main contributions of this thesis are covered in Chapter 4, 5, 6 and Appendix A, B, C, D, E. Two papers including one revised manuscripts and one submission have been produced during the course of this thesis:

- **Zexun Chen** and Bo Wang. How priors of initial hyperparameters affect Gaussian process regression models, *Neurocomputing*, 2016, submitted; E-print: arXiv:1605.07906 [stat.ML].
- **Zexun Chen**, Bo Wang and Alexander N. Gorban. Multivariate Gaussian and Student$-t$ Process Regression for Multi-output Prediction, *Entropy*, 2017, revised; E-print: arXiv:1703.04455 [stat.ML].

The rest of the thesis is organised as follows.

Chapter 2 contains some useful preliminaries about Gaussian and Student$-t$ distribution, Gaussian and Student$-t$ process, matrix-variate distribution and matrix algebra, and classical financial time series models.

Chapter 3 reviews Gaussian process regression from weight-space and function-space perspectives in detail, including all the assumptions and derivations. In addition, extra attentions are given to several important parts of GPR, including kernel, mean function and hyper-parameter estimation.

In Chapter 4 we study the sensitivity of the hyper-parameter estimation and the performance of GPR to the prior distribution for the initial values. The vague and data-dominated priors are taken for the initial values of hyper-parameters over several commonly used kernels and then the influence of the priors on the performance of GPR model is investigated. The results show that the sensitivity of the the hyper-parameter estimation depends on the choice of kernels, but the priors have little influence on the performance of GPR models in terms of predictability.

In Chapter 5, several Gaussian process regression extensions, including Gaussian process regression with Student$-t$ process (GPRT) and Student$-t$ process regression (TPR), are introduced, with all the above models then applied to predict 10 main equity indices spread throughout the world. The experiments include a comparison of GPR, GPRT and TPR, and the classical time series model ARMA. We find that GPR and TPR shows relatively considerable capability of predicting equity indices so that leave-one-out cross-validation (LOO-CV), $k-$fold cross validation and sliding window methods are used to make a further model evaluation of GPR and TPR. Furthermore, GPR and TPR are extended to state-space Gaussian process regression (SSGPR) and state-space Student$-t$ process regression (SSTPR) models in order to make an effective prediction for the stock markets based on the historical trading data in dynamic system. The overall results are that SSTPR outperforms SSGPR for the equity index prediction. Based on the detailed results, a brief market efficiency analysis confirms that the developed markets are unpredictable on the whole.

In Chapter 6, we propose the multivariate Gaussian process regression (MV-GPR) and Student$-t$ process regression (MV-TPR) for multi-output prediction, where the model settings, derivations and computations are all directly performed in matrix form, rather than vectorizing the matrices involved in the existing methods of Gaussian process for vector-valued function model. The effectiveness of the proposed methods is illustrated through a simulated example. The proposed methods are then applied to stock market modelling in which the Buy&Sell strategies generated by our proposed methods are shown to be profitable in the equity investment.

Some discussions and future work are presented in Chapter 7.

Appendix A shows the graphs of predictions by GPR, GPRT, TPR models for INDU, NDX, SPX,and UKX.

Appendix B presents the graphs of predictions by GPR, GPRT, TPR and ARMA(1,1) models for DAX, HSI, INDU, NDX, NKY, SENSEX, SPX, and UKX.

The details of negative log marginal likelihood and gradient evaluation for MV-GPR and MV-TPR are described in Appendix C.

Appendix D contains the details of investment for three Chinese stocks listed in NASDAQ.

Appendix E contains the details of investment for the stocks listed in Dow 30.

Chapter 2

Preliminaries

2.1 Gaussian process and Student$-t$ process

2.1.1 Gaussian distribution and process

A Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distribution. Mathematically, for any set $\mathcal{X}$ [1], a Gaussian process (GP) on $\mathcal{X}$ is a set of random variables $(f(x), x \in \mathcal{X})$ such that, for any $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in \mathcal{X}, (f(x_1), \ldots, f(x_n))$ is (multivariate) Gaussian.

As a Gaussian distribution is specified by a mean vector and a covariance matrix, a GP is also fully determined by a mean function and a covariance function. In other words, we have [46, 47]:

Theorem 2.1 (Gaussian Processes)**.** *For any set $\mathcal{X}$, any mean function $\mu : \mathcal{X} \mapsto \mathbb{R}$ and any covariance function (also called kernel) $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, there exists a GP $f(x)$ on $\mathcal{X}$, s.t. $\mathbb{E}[f(x)] = \mu(x)$, $\operatorname{cov}(f(x_s), f(x_t)) = k(x_s, x_t), \forall x, x_s, x_t \in \mathcal{X}$. It is denoted by $f \sim \mathcal{GP}(\mu, k)$.*

Additionally, a Gaussian distribution also has the Gaussian marginal distribution and conditional distribution.

[1] Although $\mathcal{X}$ can be any set, it usually is $\mathbb{R}$ or $\mathbb{R}^n$.

Theorem 2.2 (Marginalization and conditional distribution)**.** *Let $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, and partition $\boldsymbol{y}, \boldsymbol{\mu}$ and Σ as*

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \begin{array}{l} n_1 \\ n_2 \end{array}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \begin{array}{l} n_1 \\ n_2 \end{array}$$

$$\Sigma = \begin{array}{cl} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} & \begin{array}{l} n_1 \\ n_2 \end{array} \\ \begin{array}{cc} \quad n_1 & \quad n_2 \end{array} & \end{array}$$

where $\boldsymbol{y}, \boldsymbol{\mu} \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$. Then, $\boldsymbol{y}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_{11})$ and,

$$\boldsymbol{y}_2 | \boldsymbol{y}_1 \sim \mathcal{N}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

2.1.2 Fat-tailed distribution

In mathematics, a fat-tailed distribution is a probability distribution that has the property, along with the other heavy-tailed distributions that it exhibits large extremely large kurtosis particularly relative to the ubiquitous normal which itself is an example of an exceptionally thin tail distribution. Fat-tailed distributions have been empirically encountered in a variety of areas: economics, physics, and earth sciences. The definition in probability theory is in the following.

Definition 2.3. The definition of a random variable X is said to have a fat tail if

$$Pr(X > x) \sim x^{-\alpha} \text{ as } x \to \infty, \alpha > 0,$$

where $Pr(\cdot)$ is the probability function. That is, if X has probability density function $f_X(x)$,

$$f_X(x) \sim x^{-(\alpha+1)} \text{ as } x \to \infty, \alpha > 0.$$

Here the notation "$\sim$" refers to the asymptotic equivalence of functions.

2.1.3 Student$-t$ distribution and process

It is known that the Student$-t$ distribution is an extension of the Gaussian distribution, which is also symmetric and bell-shaped, but has heavier tails. The probability density function of Student$-t$ distribution is

$$p(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

According to the definition of fat-tailed distribution, it's easy to verify that Student$-t$ distribution is a fat-tailed distribution since $(1+x^2)^{-\frac{\nu+1}{2}} \sim x^{-(\nu+1)}$ as $x \to \infty$.

Compared with Gaussian distribution, there is an important parameter ν in the probability density function. In fact, ν is the degree of freedom in the Student$-t$ distribution, which control how fat the tails are. The Figure 2.1 shows the density of the Student$-t$ distribution for increasing values of ν . The normal distribution is shown as a red line for comparison. Note that the t-distribution (red line) becomes closer to the normal distribution as ν increases. In fact, the Student$-t$ distribution goes to Gaussian distribution as the degree of freedom tends to infinity. For demonstration, Figure 2.1 shows the density density of the Student$-t$ distribution for 1, 2, 5 and 10 degrees of freedom compared to the standard Gaussian distribution.

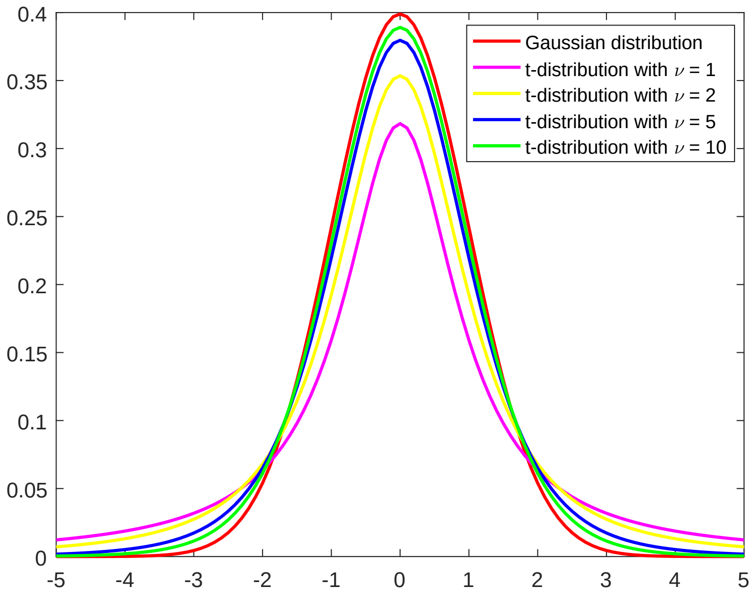


FIGURE 2.1: Density of the Student$-t$ distribution for 1, 2, 5 and 10 degrees of freedom compared to the standard Gaussian distribution.

With the popularity of GPR over the last decade, it is natural to consider more general families of elliptical process, such as Student$-t$ process (TP), where any collection of function values has a desirable extension distribution of the Gaussian distribution [48]. Some basic knowledge related to the Student$-t$ process described in [38] are introduced below.

Definition 2.4 (Multivariate Student$-t$ distribution)**.** A random vector $\boldsymbol{y} \in \mathbb{R}^n$ is said to have a multivariate Student$-t$ distribution with parameters $\nu \in \mathbb{R}^+ \setminus [0, 2]$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ if and only if its probability density function is given by

$$p(\boldsymbol{y}|\nu, \mu, \Sigma) = \frac{\Gamma(\frac{\nu+n}{2})}{((\nu-2)\pi)^{\frac{n}{2}}\Gamma(\frac{n}{2})}(\det\Sigma)^{-\frac{1}{2}} \times \left(1 + \frac{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\nu-2}\right)^{-\frac{\nu+n}{2}}. \tag{2.1}$$

We denote this by $\boldsymbol{y} \sim \mathcal{T}(\nu, \boldsymbol{\mu}, \Sigma)$.

Lemma 2.5. *If* $\boldsymbol{y} \sim \mathcal{T}(\nu, \boldsymbol{\mu}, \Sigma)$*, then,* $\mathbb{E}[\boldsymbol{y}] = \boldsymbol{\mu}, \mathrm{cov}[\boldsymbol{y}] = \Sigma$.

Similar to multivariate Gaussian distributions, multivariate Student$-t$ distributions are also consistent with marginalization and conditional distribution.

Theorem 2.6 (Marginalization and conditional distribution)**.** *Let* $\boldsymbol{y} \sim \mathcal{T}(\nu, \boldsymbol{\mu}, \Sigma)$*, and partition* $\boldsymbol{y}, \boldsymbol{\mu}$ *and* Σ *as*

$$\boldsymbol{y} = \begin{array}{c} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \end{array} \begin{array}{c} n_1 \\ n_2 \end{array}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \begin{array}{c} n_1 \\ n_2 \end{array}$$

$$\Sigma = \begin{array}{cc} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \\ \begin{array}{cc} n_1 \quad & \quad n_2 \end{array} & \end{array}$$

where $\boldsymbol{y}, \boldsymbol{\mu} \in \mathbb{R}^n$*,* $\nu \in \mathbb{R}^+ \setminus [0, 2]$ *and* $\Sigma \in \mathbb{R}^{n\times n}$*. Then,*

$$\boldsymbol{y}_1 \sim \mathcal{T}(\nu, \boldsymbol{\mu}_1, \Sigma_{11}), \quad \boldsymbol{y}_2|\boldsymbol{y}_1 \sim \mathcal{T}(\hat{\nu}, \hat{\boldsymbol{\mu}}_2, \frac{\nu+\beta-2}{\nu+n_1-2} \times \hat{\Sigma}_{22}),$$

where $\hat{\nu} = \nu + n_1, \hat{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1), \beta = (\boldsymbol{y}_1 - \boldsymbol{\mu}_1)^{\mathrm{T}}\Sigma_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1), \hat{\Sigma}_{22} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Given the definition of multivariate Student$-t$ distribution, a Student$-t$ process is naturally defined as a collection of random variables which have the joint multivariate Student$-t$ distribution. Like GP, a Student$-t$ process with a specific degree of freedom is fully specified by a mean function and a covariance function [38].

Theorem 2.7 (Student$-t$ process)**.** *For any set $\mathcal{X}$, any mean function $\mu : \mathcal{X} \mapsto \mathbb{R}$ and any covariance function (also called kernel) $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, there exists a TP $f(x)$ with the degree of freedom $\nu \in \mathbb{R}^+ \setminus [0, 2]$ on $\mathcal{X}$,s.t. $\mathbb{E}[f(x)] = \mu(x)$, $\mathrm{cov}(f(x_s), f(x_t)) = k(x_s, x_t), \forall x, x_s, x_t \in \mathcal{X}$. It denotes $f \sim \mathcal{TP}(\nu, \mu, k)$.*

2.2 Matrix-variate distributions

Matrix-variate distributions have many useful properties, as discussed in the literature [44, 49, 50]. Below we list some of them which are used in this thesis.

2.2.1 Matrix-variate Gaussian distribution

Definition 2.8 (Matrix-variate Gaussian distribution)**.** The random matrix $X \in \mathbb{R}^{n\times d}$ is said to have a matrix-variate Gaussian distribution with mean matrix $M \in \mathbb{R}^{n\times d}$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and $\Omega \in \mathbb{R}^{d\times d}$ if and only if its probability density function is given by

$$p(X|M, \Sigma, \Omega) = (2\pi)^{-\frac{dn}{2}} \det(\Sigma)^{-\frac{d}{2}} \det(\Omega)^{-\frac{n}{2}} \mathrm{etr}(-\frac{1}{2}\Omega^{-1}(X - M)^{\mathrm{T}}\Sigma^{-1}(X - M)), \tag{2.2}$$

where Ω and Σ are positive semi-definite. It denotes $X \sim \mathcal{MN}_{n,d}(M, \Sigma, \Omega)$.

Like multivariate Gaussian distributions, matrix-variate Gaussian distributions also hold several important properties as follows.

Theorem 2.9 (Transposable)**.** *If $X \sim \mathcal{MN}_{n,d}(M, \Sigma, \Omega)$, then*

$$X^{\mathrm{T}} \sim \mathcal{MN}_{d,n}(M^{\mathrm{T}}, \Omega, \Sigma).$$

Matrix-variate Gaussian distributions are related to the multivariate Gaussian distributions in the following way.

Theorem 2.10 (Vectorizable)**.** *$X \sim \mathcal{MN}_{n,d}(M, \Sigma, \Omega)$ if and only if*

$$\mathrm{vec}(X^{\mathrm{T}}) \sim \mathcal{N}_{nd}(\mathrm{vec}(M^{\mathrm{T}}), \Sigma \otimes \Omega).$$

Furthermore, matrix-variate Gaussian distributions are consistent under the marginalization and conditional distribution.

Theorem 2.11 (Marginalization and conditional distribution)**.** *Let* $X \sim \mathcal{MN}_{n,d}(M, \Sigma, \Omega)$*, and partition* X, M, Σ *and* Ω *as*

$$X = \begin{array}{cc} \begin{bmatrix} X_{1r} \\ X_{2r} \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \end{array} = \begin{array}{c} \begin{bmatrix} X_{1c} & X_{2c} \end{bmatrix} \\ \begin{array}{cc} d_1 & d_2 \end{array} \end{array}, \quad X = \begin{array}{cc} \begin{bmatrix} M_{1r} \\ M_{2r} \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \end{array} = \begin{array}{c} \begin{bmatrix} M_{1c} & M_{2c} \end{bmatrix} \\ \begin{array}{cc} d_1 & d_2 \end{array} \end{array}$$

$$\Sigma = \begin{array}{cc} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \\ \begin{array}{cc} n_1 & n_2 \end{array} & \end{array} \quad \textit{and} \quad \Omega = \begin{array}{cc} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} & \begin{array}{c} d_1 \\ d_2 \end{array} \\ \begin{array}{cc} d_1 & d_2 \end{array} & \end{array}$$

Then,

(i) $X_{1r} \sim \mathcal{MN}_{n_1,d}\,(M_{1r}, \Sigma_{11}, \Omega)$,

$$X_{2r}|X_{1r} \sim \mathcal{MN}_{n_2,d}\left(M_{2r} + \Sigma_{21}\Sigma_{11}^{-1}(X_{1r} - M_{1r}), \Sigma_{22\cdot1}, \Omega\right);$$

(ii) $X_{1c} \sim \mathcal{MN}_{n,d_1}\,(M_{1c}, \Sigma, \Omega_{11})$,

$$X_{2c}|X_{1c} \sim \mathcal{MN}_{n,d_2}\left(M_{2c} + (X_{1c} - M_{1c})\Omega_{11}^{-1}\Omega_{12}, \Sigma, \Omega_{22\cdot1}\right);$$

where $\Sigma_{22\cdot1}$*and* $\Omega_{22\cdot1}$ *are the Schur complements [51] of* Σ_{11} *and* Ω_{11}*, respectively,*

$$\Sigma_{22\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \quad \Omega_{22\cdot1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}.$$

If we assume $d = 1$ and $\Omega = 1$, the matrix-variate Gaussian distribution degenerate to multivariate Gaussian distribution. As a result, multivariate Gaussian distribution is a special case of matrix-variate Gaussian distribution.

2.2.2 Matrix-variate Student$-t$ distribution

Definition 2.12 (Matrix-variate Student$-t$ distribution)**.** The random matrix $X \in \mathbb{R}^{n\times d}$ is said to have a matrix-variate Student t distribution with the mean matrix $M \in \mathbb{R}^{n\times d}$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}, \Omega \in \mathbb{R}^{d\times d}$ and the degree of freedom ν if and only if its probability density function is given by

$$p(X|\nu, M, \Sigma, \Omega) = \frac{\Gamma_n[\frac{1}{2}(\nu + d + n - 1)]}{\pi^{\frac{1}{2}dn}\Gamma_n[\frac{1}{2}(\nu + n - 1)]} \det(\Sigma)^{-\frac{d}{2}} \det(\Omega)^{-\frac{n}{2}} \times$$
$$\det(\mathbf{I}_n + \Sigma^{-1}(X - M)\Omega^{-1}(X - M)^{\mathrm{T}})^{-\frac{1}{2}(\nu+d+n-1)}, \quad (2.3)$$

where Ω and Σ are positive semi-definite, and

$$\Gamma_n(\lambda) = \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma(\lambda + \frac{1}{2} - \frac{i}{2}).$$

We denote this by $X \sim \mathcal{MT}_{n,d}(\nu, M, \Sigma, \Omega)$.

Theorem 2.13 (Expectation and covariance). *Let* $X \sim \mathcal{MT}(\nu, M, \Sigma, \Omega)$*, then*

$$\mathbb{E}(X) = M, \quad \mathrm{cov}(\mathrm{vec}(X^{\mathrm{T}})) = \frac{1}{\nu - 2}\Sigma \otimes \Omega, \nu > 2.$$

Theorem 2.14 (Transposable). *If* $X \sim \mathcal{MT}_{n,d}(\nu, M, \Sigma, \Omega)$*, then*

$$X^{\mathrm{T}} \sim \mathcal{MT}_{n,d}(\nu, M^{\mathrm{T}}, \Omega, \Sigma).$$

Theorem 2.15 (Asymptotics). *Let* $X \sim \mathcal{MT}_{n,d}(\nu, M, \Sigma, \Omega)$*,then*

$$X \xrightarrow{d} \mathcal{MN}_{n,d}(M, \Sigma, \Omega) \text{ as } \nu \to \infty,$$

where "$\xrightarrow{d}$" *denotes convergence in distribution.*

Theorem 2.16 (Marginalization and conditional distribution). *Let* $X \sim \mathcal{MT}_{n,d}(\nu, M, \Sigma, \Omega)$*, and partition* X, M, Σ *and* Ω *as*

$$X = \begin{bmatrix} X_{1r} \\ X_{2r} \end{bmatrix} \begin{array}{c} n_1 \\ n_2 \end{array} = \begin{array}{c} \begin{bmatrix} X_{1c} & X_{2c} \end{bmatrix} \\ \begin{array}{cc} d_1 & \quad d_2 \end{array} \end{array}, \quad X = \begin{bmatrix} M_{1r} \\ M_{2r} \end{bmatrix} \begin{array}{c} n_1 \\ n_2 \end{array} = \begin{array}{c} \begin{bmatrix} M_{1c} & M_{2c} \end{bmatrix} \\ \begin{array}{cc} d_1 & \quad d_2 \end{array} \end{array}$$

$$\Sigma = \begin{array}{cc} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \\ \begin{array}{cc} n_1 & \quad n_2 \end{array} & \end{array} \quad \text{and} \quad \Omega = \begin{array}{cc} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} & \begin{array}{c} d_1 \\ d_2 \end{array} \\ \begin{array}{cc} d_1 & \quad d_2 \end{array} & \end{array}$$

Then,

(i) $X_{1r} \sim \mathcal{MT}_{n_1,d}\,(\nu, M_{1r}, \Sigma_{11}, \Omega)$,

$$\begin{aligned} X_{2r}|X_{1r} \sim \quad & \mathcal{MT}_{n_2,d}\Big(\nu + n_1, M_{2r} + \Sigma_{21}\Sigma_{11}^{-1}(X_{1r} - M_{1r}), \Sigma_{22\cdot 1}, \\ & \qquad \Omega + (X_{1r} - M_{1r})^{\mathrm{T}}\Sigma_{11}^{-1}(X_{1r} - M_{1r})\Big); \end{aligned}$$

(ii) $X_{1c} \sim \mathcal{MT}_{n,d_1}\,(, \nu, M_{1c}, \Sigma, \Omega_{11})$,

$$\begin{aligned} X_{2c}|X_{1c} \sim \quad & \mathcal{MT}_{n,d_2}\Big(\nu + d_1, M_{2c} + (X_{1c} - M_{1c})\Omega_{11}^{-1}\Omega_{12}, \\ & \qquad \Sigma + (X_{1c} - M_{1c})\Omega_{11}^{-1}(X_{1c} - M_{1c})^{\mathrm{T}}, \Omega_{22\cdot 1}\Big); \end{aligned}$$

where $\Sigma_{22\cdot 1}$*and* $\Omega_{22\cdot 1}$ *are the Schur complements of* Σ_{11} *and* Ω_{11}*, respectively,*

$$\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \quad \Omega_{22\cdot 1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}.$$

If we assume $d = 1$ and $\Omega = \nu - 2$, the matrix-variate Student$-t$ distribution degenerate to multivariate Student$-t$ distribution. As a result, multivariate Student$-t$ distribution is a special case of matrix-variate Student$-t$ distribution.

2.3 Matrix algebra theory

Matrix algebra is one of the most important areas of mathematics for data analysis and for statistical theory. Several useful properties are listed below.

2.3.1 Matrix identities

The matrix inversion theorem, also called the Woodbury, Sherman & Morrison formula, (see, e.g. [52]).

Theorem 2.17 (Matrix Inversion)**.** *For matrices* $\mathbf{G}, \mathbf{W}, \mathbf{U}$*, and* $\mathbf{V}$*, assuming* $\mathbf{G}$ *and* $\mathbf{W}$ *are invertible,* $\mathbf{G}$ *is* $n \times n$*,* $\mathbf{W}$ *is* $m \times m$*,* $\mathbf{U}$ *and* $\mathbf{V}$ *are* $n \times m$*, there exists,*

$$(\mathbf{G} + \mathbf{U}\mathbf{W}\mathbf{V}^{\mathrm{T}})^{-1} = \mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{U}(\mathbf{W}^{-1} + \mathbf{V}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{U})^{-1}\mathbf{V}^{\mathrm{T}}\mathbf{G}^{-1}.$$

The matrix derivative with respect to a parameter involved in the matrix can be shown as follow [53].

Theorem 2.18 (Matrix derivative with respect to a parameter)**.** *For matrix* $\mathbf{A}$*, there exists,*

$$\frac{\partial}{\partial \theta}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \theta}\mathbf{A}^{-1},$$

where $\partial \mathbf{A}/\partial \theta$ *is a matrix whose entries are the derivatives of entries in* $\mathbf{A}$*. If* $\mathbf{A}$ *is positive definite, there exists,*

$$\frac{\partial}{\partial \theta}\log\det\mathbf{A} = \mathrm{tr}(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \theta}).$$

According to the chain rule of derivatives of a matrix, there exists [53]:

Theorem 2.19 (Matrix derivative with respect to a matrix)**.** *Let* $\mathbf{U} = f(X)$*, then the derivative of the function* $g(\mathbf{U})$ *with respect to* X*,*

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \mathrm{tr}\left[\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\right)^{\mathrm{T}} \frac{\partial \mathbf{U}}{\partial X_{ij}}\right].$$

At the same time, there are another two useful formulas of derivatives with respect to X*,*

$$\frac{\partial \log\det(X)}{\partial X} = (X^{\mathrm{T}})^{-1}, \quad \frac{\partial}{\partial X}\mathrm{tr}(\mathbf{A}X^{-1}\mathbf{B}) = -(X^{-1}\mathbf{B}\mathbf{A}X^{-1})^{\mathrm{T}},$$

where $\mathbf{A}$ *and* $\mathbf{B}$ *are constant matrices.*

2.3.2 Matrix decompositions

In linear algebra, a matrix decomposition is a factorization of a matrix into a product of matrices. In numerical analysis, matrix decompositions are used to implement efficient matrix algorithms. Two important matrix decompositions are introduced as following.

2.3.2.1 Cholesky decomposition

The Cholesky (Chol) decomposition of a Hermitian positive definite matrix $\mathbf{A}$ is a decomposition of the form,

$$\mathbf{A} = \mathbf{L}\mathbf{L}^*,$$

where $\mathbf{L}$ denotes a lower triangular matrix with real positive diagonal entries, and $\mathbf{L}^*$ is the conjugate transpose of $\mathbf{L}$. If all entries of $\mathbf{A}$ are real, conjugate transpose means only transpose and Hermitian positive definite is only symmetric positive definite. The result is also true if $\mathbf{A}$ is (Hermitian) positive semi-definite where the diagonal entries of $\mathbf{L}$ are allowed to be zero.

Cholesky decomposition is widely used in solving linear system. For example, if $\boldsymbol{Ax} = \boldsymbol{b}$, then the following steps are applied:

Step 1: Solving triangular system $\mathbf{L}\boldsymbol{y} = \boldsymbol{b}$;

Step 2: Solving another triangular system $\mathbf{L}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$.

Besides, the determinant computation is efficient using Cholesky decomposition,

$$\det \mathbf{A} = \prod_{i=1}^{n} \mathbf{L}_{ii}^2,$$

$$\log\det \mathbf{A} = 2\sum_{i=1}^{n} \log \mathbf{L}_{ii},$$

where $\mathbf{L}_{ii}, i = 1, 2, \cdots, n$ are the diagonal elements of $\mathbf{L}$.

2.3.2.2 Singular value decomposition

The Singular Value Decomposition (SVD) is a factorization of a real or complex matrix. It is actually the generalization of the eigendecomposition of a positive semi-definite normal matrix [2]. It has many useful applications in signal processing and statistics. The SVD form of $m \times n$ matrix $\mathbf{A}$ is,

$$\mathbf{A} = \mathbf{UDV}^*,$$

where $\mathbf{U}$ is a $m \times m$ unitary matrix, $\mathbf{D}$ is a diagonal $m \times n$ matrix with non-negative real number on the diagonal, and $\mathbf{V}$ is a $n \times n$ unitary matrix.

The SVD is very general since it can be applied to any $m \times n$ matrix. As we know, eigendecomposition is another similar matrix decomposition method, though it can only be applied to certain classes of square matrices. Nevertheless, the two decompositions are related. Given an SVD of $\mathbf{A}$, as described above, the following

[2] A complex square matrix A is normal if $\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$

two relations hold [54]:

$$\begin{aligned}\mathbf{A}^*\mathbf{A} &= \mathbf{V}\mathbf{D}^*\mathbf{U}^*\mathbf{U}\mathbf{D}\mathbf{V}^* = \mathbf{V}(\mathbf{D}^*\mathbf{D})\mathbf{V}^* \\ \mathbf{A}\mathbf{A}^* &= \mathbf{U}\mathbf{D}\mathbf{V}^*\mathbf{V}\mathbf{D}^*\mathbf{U}^* = \mathbf{U}(\mathbf{D}\mathbf{D}^*)\mathbf{U}^*\end{aligned}$$

If $\mathbf{A}$ is a normal matrix, which by definition must be square, the spectral theorem says that it can be unitarily diagonalized using a basis of eigenvectors, so that it can be written as $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$, where $\mathbf{U}$ is unitary matrix and $\mathbf{D}$ now is a diagonal matrix. In a further special case, if $\mathbf{A}$ is also positive semi-definite, the decomposition $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ is also a singular value decomposition [54].

2.4 Autoregressive moving average model

In the statistical analysis of time series, autoregressive moving average (ARMA) is a classical model constructed by two polynomials: one for the autoregression and the other for the moving average. The first description of general ARMA model was introduced in 1951 [55], and was then popularised in 1971 [56].

The notation ARMA(p, q) refers to the model with p autoregressive terms and q moving average terms. It can be specified in terms of the lag operator L, where $L^i X_t = X_{t-i}$. The AR(p) model is given by

$$\epsilon_t = (1 - \sum_{i=1}^{p} \varphi_i L^i) X_t = \varphi(L) X_t$$

where φ represents the polynomial $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i$, the random variable ϵ_t is white noise, $\{\varphi_i\}_{i=1}^{p}$ are parameters. Then the MA(q) model is given by

$$X_t = (1 + \sum_{i=1}^{q} \theta_i L^i) \epsilon_t = \theta(L) \epsilon_t,$$

where θ represents the polynomial $\theta(L) = 1 + \sum_{i=1}^{1} \theta_i L^i$ and $\{\theta_i\}_{i=1}^{q}$ are the parameters. Finally, the combined ARMA(p, q) model is given by,

$$(1 - \sum_{i=1}^{p} \varphi_i L^i) X_t = (1 + \sum_{i=1}^{q} \theta_i L^i) \epsilon_t.$$

It can be rewritten as,

$$\varphi(L)X_t = \theta(L)\epsilon_t \quad \text{or} \quad \frac{\varphi(L)}{\theta(L)}X_t = \epsilon_t.$$

If $\varphi(L) = 1$, ARMA process becomes MA(q) while the ARMA process degenerates to AR(p) if $\theta(L) = 1$.

The ARMA(p, q) process is widely used in the analysis of time series and is one of the classical methods for prediction. The estimation of parameters in ARMA(p, q) process is facilitated by plotting the partial autocorrelation functions for a rough estimate of p, and similarly using the autocorrelation functions for an approximate estimate of q. Furthermore, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are also usually recommended to find appropriate values of p and q [57, 58].

Chapter 3

Gaussian process regression

3.1 Introduction

GP provides a practical probabilistic approach to supervised learning, which can be divided into regression and classification problems. It is known that the responses for classification are the discrete class labels whereas regressions consider outputs as continues variables [59]. There are several ways to derive GPR models, mainly including *weight-space view* and *function-space view*. The former interpretation is from the view of the Bayesian framework while the latter considers GP as a defining a distribution over functions, and inference performing directly in the space of functions [59]. We introduce GPR models from these two ways.

3.1.1 Weight-space view

The simple linear regression models have been well studied and used extensively. Our introduction for GPR starts from Bayesian linear regression and then we make a simple enhancement to this class of models by projecting the inputs into a high-dimensional feature space and using the linear model there.

3.1.1.1 Bayesian linear regression

We review the Bayesian analysis of standard linear regression model with Gaussian noise

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}, \qquad y = f(\boldsymbol{x}) + \varepsilon, \tag{3.1}$$

where $\boldsymbol{x}$ is the multi-dimensional input vector, $\boldsymbol{w}$ is a vector of weights of this model (the size equals the dimension of the input $\boldsymbol{x}$), and f is the function value and y is the target value. Let $\mathcal{D}$ be a training set $\mathcal{D} = (X, \boldsymbol{y}) = \{(\boldsymbol{x}_i, y_i)|\boldsymbol{x}_i \in \mathbb{R}^p, y_i \in \mathbb{R}, i = 1, \cdots, n\}$. We assume that the noise follows an independent, identically distributed Gaussian with zero mean and variance σ_n^2,

$$\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma_n^2).$$

Given the model assumptions, the likelihood function is obtained by

$$\begin{aligned} p(\boldsymbol{y}|X, \boldsymbol{w}) &= \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) = \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_n^2}(\boldsymbol{y} - X^{\mathrm{T}}\boldsymbol{w})^2\right) \\ &= \mathcal{N}(X^{\mathrm{T}}\boldsymbol{w}, \sigma_n^2 \mathbf{I}_n). \end{aligned} \tag{3.2}$$

Inference in this linear regression model is based on the posterior distribution over the weights, computed by Bayes' rule [59]

$$p(\boldsymbol{w}|\mathcal{D}) = \frac{p(\boldsymbol{y}|X, \boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{y}|X)} = \frac{p(\boldsymbol{y}|X, \boldsymbol{w})p(\boldsymbol{w})}{\int p(\boldsymbol{y}|X, \boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}} \propto p(\boldsymbol{y}|X, \boldsymbol{w})p(\boldsymbol{w}), \tag{3.3}$$

where $\propto$ means "proportional to". Assuming that the prior of $\boldsymbol{w}$ is a Gaussian distribution with zero mean and covariance matrix Σ_p, $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_p)$, the distribution of the weights given the training set $\mathcal{D}$ are computed by,

$$\begin{aligned} p(\boldsymbol{w}|\mathcal{D}) &\propto \exp\left(-\frac{1}{2\sigma_n^2}(\boldsymbol{y} - X^{\mathrm{T}}\boldsymbol{w})^{\mathrm{T}}(\boldsymbol{y} - X^{\mathrm{T}}\boldsymbol{w})\right) \exp\left(-\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\Sigma_p^{-1}\boldsymbol{w}\right) \\ &\propto \exp\left(-\frac{1}{2}(\boldsymbol{w} - \boldsymbol{\mu}_w)^{\mathrm{T}}\Sigma_w^{-1}(\boldsymbol{w} - \boldsymbol{\mu}_w)\right), \end{aligned} \tag{3.4}$$

where $\boldsymbol{\mu}_w = \sigma_n^{-2}(\sigma_n^{-2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} + \Sigma_p^{-1})^{-1}\boldsymbol{X}\boldsymbol{y}, \Sigma_w = (\sigma_n^{-2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} + \Sigma_p^{-1})^{-1}$. In fact, it is Gaussian distribution with mean $\boldsymbol{\mu}_w$ and covariance matrix Σ_w.

$$p(\boldsymbol{w}|\mathcal{D}) \sim \mathcal{N}(\boldsymbol{\mu}_w, \Sigma_w).$$

To make predictions for test points, we integrate over all the possible parameters, weighted by their posterior probability. The predictive distribution for $f_* = f(\boldsymbol{z})$ given by averaging the output of all possible linear models with respect to the

Gaussian posterior

$$\begin{aligned} p(f_*|\boldsymbol{z},\mathcal{D}) &= \int p(f_*|\boldsymbol{z},\boldsymbol{w})p(\boldsymbol{w}|\mathcal{D})d\boldsymbol{w} \\ &= \mathcal{N}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\mu}_w, \boldsymbol{z}^{\mathrm{T}}\Sigma_w\boldsymbol{z}). \end{aligned} \tag{3.5}$$

According to Eq.(3.5), the predictive distribution is again Gaussian, with a mean given by the posterior mean of the weights from (3.4) multiplied by the test input. The predictive variance is a quadratic form of the test input with the posterior covariance matrix, which means that predictive uncertainties increase with magnitude of the test input [59].

Finally, considering noise part, y_* is given by

$$p(y_*|\boldsymbol{z},\mathcal{D}) = \mathcal{N}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\mu}_w, \boldsymbol{z}^{\mathrm{T}}\Sigma_w\boldsymbol{z} + \sigma_n^2\mathbf{I}). \tag{3.6}$$

3.1.1.2 Projections of inputs into feature space

The main drawback of standard linear regression model is that the output is limited to be a linear combination of the inputs. That is to say, if the relationship between input and output cannot be approximated by a linear function, the model performance is poor [59]. A simple approach to overcome this problem is projections of inputs into feature space, which is to initially project the inputs into some high dimensional space using a set of basis functions and then apply Bayesian linear regression in this space [59].

Now we consider a basis function of form $\phi(\|\boldsymbol{x}-\boldsymbol{x}_i\|)$, where ϕ is a non-linear function and $\|\boldsymbol{x}-\boldsymbol{x}_i\|$ is the distance of the vector $\boldsymbol{x}$ from the prototype vector $\boldsymbol{x}_i$, where the distance could be defined in all Hilbert spaces. For the case of n training points, where each point is presented as a prototype, the mapping can be defined by,

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} w_i\phi(\|\boldsymbol{x}-\boldsymbol{x}_i\|) = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{w}, \tag{3.7}$$

where $\boldsymbol{\phi}(\boldsymbol{x}) = [\phi(\|\boldsymbol{x}-\boldsymbol{x}_1\|),\cdots,\phi(\|\boldsymbol{x}-\boldsymbol{x}_n\|)]^{\mathrm{T}}$. The model considering noise is

$$y = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{w} + \varepsilon. \tag{3.8}$$

Let the matrix $\Phi(X)$ be the aggregation of columns $\boldsymbol{\phi}(\boldsymbol{x})$ for all the training inputs,

$$\Phi(X) = \begin{bmatrix} \phi(\|\boldsymbol{x}_1 - \boldsymbol{x}_1\|) & \cdots & \phi(\|\boldsymbol{x}_1 - \boldsymbol{x}_n\|) \\ \vdots & \ddots & \vdots \\ \phi(\|\boldsymbol{x}_n - \boldsymbol{x}_1\|) & \cdots & \phi(\|\boldsymbol{x}_n - \boldsymbol{x}_n\|) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}(\boldsymbol{x}_1)^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\phi}(\boldsymbol{x}_n)^{\mathrm{T}} \end{bmatrix}.$$

The analysis for this model is similar to the standard Bayesian linear model except that everywhere $\Phi = \Phi(X)$ is substituted for X and $\boldsymbol{z}$ is replaced by $\boldsymbol{\phi}(\boldsymbol{z})$. Therefore, the predictive distribution in Eq.(3.5) becomes

$$p(f_*|\boldsymbol{z}, \mathcal{D}) = \mathcal{N}(\boldsymbol{\phi}_*^{\mathrm{T}} \boldsymbol{\mu}'_w, \boldsymbol{\phi}_*^{\mathrm{T}} \Sigma'_w \boldsymbol{\phi}_*), \tag{3.9}$$

where $\boldsymbol{\phi}_* = \boldsymbol{\phi}(\boldsymbol{z}), \boldsymbol{\mu}'_w = \sigma_n^{-2}(\sigma_n^{-2}\Phi\Phi^{\mathrm{T}} + \Sigma_p^{-1})^{-1}\Phi\boldsymbol{y}, \Sigma'_w = (\sigma_n^{-2}\Phi\Phi^{\mathrm{T}} + \Sigma_p^{-1})^{-1}$. Conveniently, Eq.(3.9) could be rewritten as follow,

$$\begin{aligned} p(f_*|\boldsymbol{z}, \mathcal{D}) = \mathcal{N}(&\boldsymbol{\phi}_*^{\mathrm{T}} \Sigma_p \Phi (K + \sigma_n^2 \mathbf{I})^{-1} \boldsymbol{y}, \\ &\boldsymbol{\phi}_*^{\mathrm{T}} \Sigma_p \boldsymbol{\phi}_* - \boldsymbol{\phi}_*^{\mathrm{T}} \Sigma_p \Phi (K + \sigma_n^2 \mathbf{I})^{-1} \Phi^T \Sigma_p \boldsymbol{\phi}_*). \end{aligned} \tag{3.10}$$

where K is defined by $K = \Phi^{\mathrm{T}} \Sigma_p \Phi$.

The equivalence between Eq.(3.9) and Eq.(3.10) is shown in the below. From the view of mean

$$\begin{aligned} \sigma_n^{-2}\Phi(K + \sigma_n^2\mathbf{I}) &= \sigma_n^{-2}\Phi(\Phi^{\mathrm{T}}\Sigma_p\Phi + \sigma_n^2\mathbf{I}) \\ &= \sigma_n^{-2}\Phi\Phi^{\mathrm{T}}\Sigma_p\Phi + \Phi \\ &= (\sigma_n^{-2}\Phi\Phi^{\mathrm{T}} + \Sigma_p^{-1})\Sigma_p\Phi. \end{aligned} \tag{3.11}$$

Therefore, $\sigma_n^{-2}(\sigma_n^{-2}\Phi\Phi^{\mathrm{T}} + \Sigma_p^{-1})^{-1}\Phi = \Sigma_p\Phi(K + \sigma_n^2 I)^{-1}$ (assuming all the matrices here are invertible) and the equation is

$$\begin{aligned} \boldsymbol{\phi}_*^{\mathrm{T}} \boldsymbol{\mu}'_w &= \boldsymbol{\phi}_*^{\mathrm{T}} \sigma_n^{-2} (\sigma_n^{-2}\Phi\Phi^{\mathrm{T}} + \Sigma_p^{-1})^{-1} \Phi \boldsymbol{y} \\ &= \boldsymbol{\phi}_*^{\mathrm{T}} \Sigma_p \Phi (K + \sigma_n^2 I)^{-1} \boldsymbol{y}. \end{aligned} \tag{3.12}$$

From the view of variance, it can be proved using matrix inversion Theorem 2.17 by setting $G^{-1} = \Sigma_p, W^{-1} = \sigma_n^2\mathbf{I}$, and $V = U = \Phi$.

It is noted that the feature space usually enters in the form of $\boldsymbol{\phi}_*^{\mathrm{T}}\Sigma_p\Phi$, $\boldsymbol{\phi}_*^{\mathrm{T}}\Sigma_p\boldsymbol{\phi}_*$, $\Phi^T\Sigma_p\boldsymbol{\phi}_*$ in Eq.(3.10). The entries of these matrices are invariably of form

$$\boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\Sigma_p\boldsymbol{\phi}(\boldsymbol{x}'),$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are either the training or the test sets.

According to the definition in [59], $k(\boldsymbol{x},\boldsymbol{x}') = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\Sigma_p\boldsymbol{\phi}(\boldsymbol{x}')$, where $k(\cdot,\cdot)$ is called kernel or covariance function. In fact, the kernel can be rewritten in inner product form using Theorem 3.2 (in next section),

$$\begin{aligned} k(\boldsymbol{x},\boldsymbol{x}') &= \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\Sigma_p\boldsymbol{\phi}(\boldsymbol{x}') \\ &= \langle \psi(\boldsymbol{x}), \psi(\boldsymbol{x}')\rangle. \end{aligned} \tag{3.13}$$

In particular, if we consider dot product as inner product, there is a simple expression for $k(\cdot,\cdot)$ using $\psi(\boldsymbol{x})$. Since Σ_p is positive semi-definite we can define $\Sigma_p^{1/2}$ and thus $\Sigma_p = (\Sigma_p^{1/2})^2$. According to Singular Value Decomposition (SVD), $\Sigma_p = UDU^{\mathrm{T}}$, where D is diagonal, then $\Sigma_p^{1/2} = UD^{1/2}U^{\mathrm{T}}$. Hence, defining $\psi(\boldsymbol{x}) = \Sigma_p^{1/2}\phi(\boldsymbol{x})$ and it is a simple dot product representation $k(\boldsymbol{x},\boldsymbol{x}') = \psi(\boldsymbol{x})\cdot\psi(\boldsymbol{x}')$ [59].

Based on the discussions above, we can find that the kernel contains all the information of feature vectors so that it is a convenient alternative as the feature vectors. That is also a reason why the discussions of kernel of GP are primary.

3.1.2 Function-space view

An alternative approach of obtaining the identical results is derived in the function-space directly [59]. As we know, GP is a collection of random variables, any finite number of which have a joint Gaussian distribution.

According to Theorem 2.1, a Gaussian process is completely specified by its mean function and kernel, that is

$$\begin{aligned} \mu(x) &= \mathbb{E}[f(x)] \\ k(x,x') &= \mathrm{cov}(f(x), f(x')). \end{aligned}$$

It can be denoted $f(x) \sim \mathcal{GP}(\mu, k)$. For example, Bayesian linear regression model $f(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{w}$ with prior $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_p)$, thus the mean and kernel function is obtained

$$\begin{aligned}\mu(\boldsymbol{x}) &= \mathbb{E}[f(\boldsymbol{x})] = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\mathbb{E}[\boldsymbol{w}] = 0,\\ k(\boldsymbol{x},\boldsymbol{x}') &= \mathbb{E}[f(\boldsymbol{x})f(\boldsymbol{x}')] = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\mathbb{E}[\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}]\boldsymbol{\phi}(\boldsymbol{x}') = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\Sigma_p\boldsymbol{\phi}(\boldsymbol{x}'),\end{aligned}$$

so that $f(\boldsymbol{x}) \sim \mathcal{GP}(0, \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\Sigma_p\boldsymbol{\phi}(\boldsymbol{x}'))$.

Now we consider a general regression model $y = f(\boldsymbol{x}) + \varepsilon$, where $f(\boldsymbol{x}) \sim \mathcal{GP}(\mu, k)$ and $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$. Given n pairs of observations $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n, \boldsymbol{x}_i \in \mathbb{R}^p, y_i \in \mathbb{R}$, it yields that $[f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n)]$ follow a multivariate Gaussian distribution

$$[f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \dots, f(\boldsymbol{x}_n)]^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, K),$$

where $\boldsymbol{\mu} = [\mu(\boldsymbol{x}_1), \dots, \mu(\boldsymbol{x}_n)]^{\mathrm{T}}$ is the mean vector and K is the $n \times n$ covariance matrix of which the (i,j)-th element $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. In order to predict $f_* = f(Z)$ at the test locations $Z = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m]^{\mathrm{T}}$, the joint distribution of the training observations $\boldsymbol{y}$ and the predictive targets f_* are given by

$$\begin{bmatrix}\boldsymbol{y}\\ f_*\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}(X)\\ \boldsymbol{\mu}(Z)\end{bmatrix}, \begin{bmatrix}K(X,X)+\sigma_n^2\mathbf{I} & K(Z,X)^{\mathrm{T}}\\ K(Z,X) & K(Z,Z)\end{bmatrix}\right), \tag{3.14}$$

where $\boldsymbol{\mu}(X) = \boldsymbol{\mu}$, $\boldsymbol{\mu}(Z) = [\mu(\boldsymbol{z}_1), \dots, \mu(\boldsymbol{z}_m)]^{\mathrm{T}}$, $K(X,X) = K$, $K(Z,X)$ is an $m \times n$ matrix of which the (i,j)-th element $[K(Z,X)]_{ij} = k(\boldsymbol{z}_i, \boldsymbol{x}_j)$, and $K(Z,Z)$ is an $m \times m$ matrix with the (i,j)-th element $[K(Z,Z)]_{ij} = k(\boldsymbol{z}_i, \boldsymbol{z}_j)$. Thus, taking advantage of Theorem 2.2, the predictive distribution is

$$p(f_*|X, \boldsymbol{y}, Z) = \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\Sigma}), \tag{3.15}$$

$$\hat{\boldsymbol{\mu}} = K(Z,X)^{\mathrm{T}}(K(X,X)+\sigma_n^2\mathbf{I})^{-1}(\boldsymbol{y} - \boldsymbol{\mu}(X)) + \boldsymbol{\mu}(Z), \tag{3.16}$$

$$\hat{\Sigma} = K(Z,Z) - K(Z,X)^{\mathrm{T}}(K(X,X)+\sigma_n^2\mathbf{I})^{-1}K(Z,X). \tag{3.17}$$

Taking noise part into consideration, the predictive distribution of targets $\boldsymbol{y}_*$ given the training set and the test locations are finally written by

$$p(\boldsymbol{y}_*|X, \boldsymbol{y}, Z) = \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\Sigma} + \sigma_n^2\mathbf{I}), \tag{3.18}$$

3.2 Kernels

It can be seen from Eq.(3.16) and Eq.(3.17) that the kernel $k(\cdot,\cdot)$ plays a crucial role in the predictive mean and variance. As discussed in [59], kernels contain our presumptions about the function we wish to learn and define the closeness and similarity between data points. As a result, the choice of kernel has a profound impact on the performance of a GPR model, just as in activation function, learning rate can affect the result of a neural network [60].

It is known that a symmetric $n \times n$ matrix C is said to be a positive semi-definite (PSD) if for any non-zero column vector $\lambda \in \mathbb{R}^n, \lambda^{\mathrm{T}} C \lambda \geq 0$. Before listing several useful kernels, we introduce the definition of positive semi-definite kernels.

Definition 3.1. A positive semi-definite kernel (also called covariance function) on $\mathcal{X}$ is a function $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}, s.t. \forall n \in \mathbb{N}, \forall x_1, \ldots, x_n \in \mathcal{X}$, the matrix C is positive semi-definite, where $C_{ij} = k(x_i, x_j)$.

For example, $\mathcal{X} = \mathbb{R}^d$, $k(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}$, hence $C = \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$. Letting $\boldsymbol{a} \in \mathbb{R}^n$, then $\boldsymbol{a}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{a} = \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \boldsymbol{a} = (\boldsymbol{a}\boldsymbol{x}^{\mathrm{T}})^2 \geq 0$. Therefore, the bivariate function $k(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}$ is a PSD kernel. A general method of reproducing a proper kernel is introduced in the following and more details can be found in [61, 62].

Theorem 3.2. *A function $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ can be written as $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$, where $\Phi(x)$ is a feature map: $x \mapsto \Phi(x) \in \mathbb{H}$ (Hilbert space) and $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{H}$, if and only if $k(x, y)$ is PSD kernel.*

Some commonly-used kernels are listed as follows.

3.2.1 Squared exponential

The most widely-used kernel in GPR is Squared Exponential (SE), which is defined as

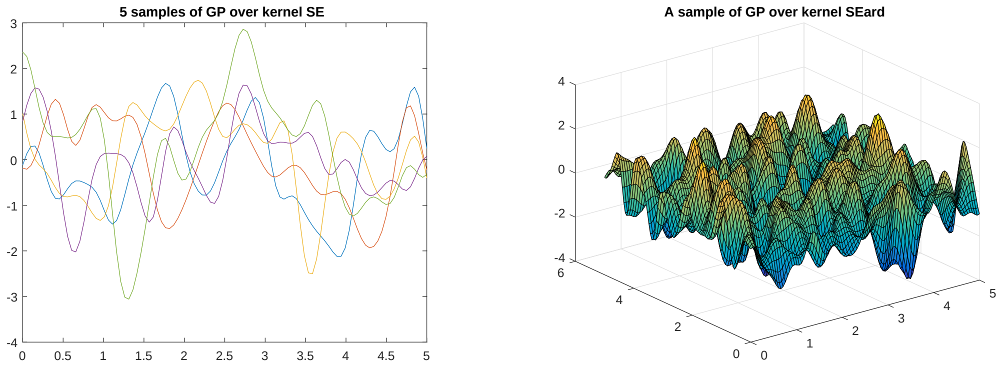
$$k_{SE}(x, x') = s_f^2 \exp(-\frac{\|x - x'\|^2}{2\ell^2}),$$

where $\|\cdot\|$ is L^2-norm (Euclidean norm), s_f^2 is the signal variance and is also considered as an output-scale amplitude [63] and the parameter ℓ is the input (length or time) scale [63]. The kernel can also be defined by Automatic Relevance

Determination (ARD).

$$k_{SEard}(\boldsymbol{x}, \boldsymbol{x}') = s_f^2 \exp(-\frac{(\boldsymbol{x}-\boldsymbol{x}')^{\mathrm{T}}\Theta^{-1}(\boldsymbol{x}-\boldsymbol{x}')}{2}),$$

where Θ is a diagonal matrix with the element components $\{\ell_i^2\}_{i=1}^p$, which are the length scales for each corresponding input dimension. Some samples of GP over kernel SE and SEard are demonstrated in Figure 3.1.



(a) (b)

FIGURE 3.1: The samples of GP over kernel SE and SEard. (a): 5 samples of GP over kernel SE with parameters $[\ell, s_f] = [0.2, 1]$. (b): A sample of 2-dimensional GP over kernel SEard with parameters $[\ell_1, \ell_2, s_f] = [0.2, 0.2, 1]$.

3.2.2 Periodic

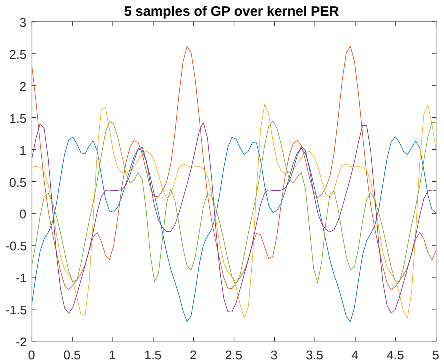
The Periodic (PER) kernel is used to model functions which exhibit a periodic pattern. An effective method is the warping production (see, e.g. [64]) where the 1-dimensional input variable x is mapped to the 2-dimension to make a periodic function of x

$$k_{PER}(x, x') = k_{SE}(\varpi(x), \varpi(x')) = s_f^2 \exp(-\frac{2\sin^2(\pi\frac{(x-x')}{\mathrm{p}})}{\ell^2}),$$

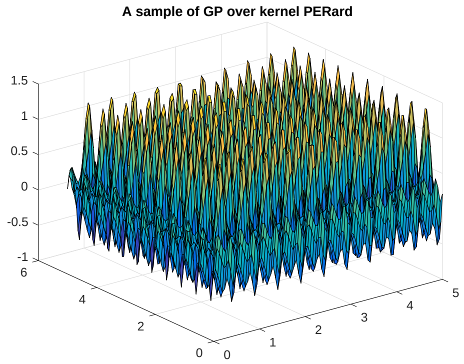
where $\varpi(x) = [\sin(\pi x/\mathrm{p}), \cos(\pi x/\mathrm{p})]^{\mathrm{T}}$ and p is the period. The second equation is due to $(\sin(\pi x/\mathrm{p})-\sin(\pi x'/\mathrm{p}))^2+(\cos(\pi x/\mathrm{p})-\cos(\pi x'/\mathrm{p}))^2 = 4\sin^2(\pi(x-x')/2\mathrm{p})$. Like SEard kernel, the ARD kernel for PER is defined by

$$\begin{aligned} k_{PERard}(\boldsymbol{x}, \boldsymbol{x}') &= k_{SEard}(\varpi(\boldsymbol{x}), \varpi(\boldsymbol{x}')) \\ &= s_f^2 \exp(-2\sin(\frac{\pi(\boldsymbol{x}-\boldsymbol{x}')}{\boldsymbol{p}})^{\mathrm{T}}\Theta^{-1}\sin(\frac{\pi(\boldsymbol{x}-\boldsymbol{x}')}{\boldsymbol{p}})), \end{aligned}$$

where $\sin(\boldsymbol{x}/\boldsymbol{p}) = [\sin(x_1/\mathrm{p}_1), \ldots, \sin(x_p/\mathrm{p}_p)]$ and $\boldsymbol{p} = \{\mathrm{p}_i\}_{i=1}^{p}$ are the periods for each corresponding input dimension. Some samples of GP over kernel PER and PERard are demonstrated in Figure 3.2.



(a)



(b)

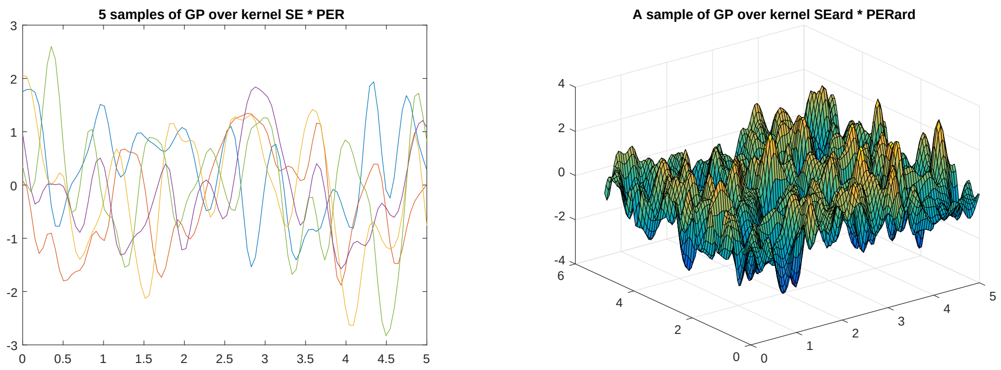
FIGURE 3.2: The samples of GP over kernel PER and PERard. (a): 5 samples of GP over kernel PER with parameters $[\ell, \mathrm{p}, s_f] = [0.5, 2, 1]$. (b): A sample of 2-dimension GP over kernel PERard with parameters $[\ell_1, \ell_2, \mathrm{p}_1, \mathrm{p}_2, s_f] = [0.5, 0.5, 2, 2, 1]$.

3.2.3 Local periodic

As shown in [65], the positive semi-definite kernels are closed under addition and multiplication. Local Periodic (LP) is such a composite kernel which is obtained by multiplying SE and PER [65]. That is,

$$k_{LP}(x, x') = k_{SE}(x, x') \times k_{PER}(x, x'),$$
$$k_{LPard}(\boldsymbol{x}, \boldsymbol{x}') = k_{SEard}(\boldsymbol{x}, \boldsymbol{x}') \times k_{PERard}(\boldsymbol{x}, \boldsymbol{x}').$$

It is a well-known kernel to capture locally periodic structure of data hence can be applied to many kernel-based models. Some samples of GP over kernel LP and LPard are shown in Figure 3.3.



(a) (b)

FIGURE 3.3: The samples of GP over kernel LP and LPard. (a): 5 samples of GP over kernel LP with parameters $[\ell, s_f] = [0.5, 1]$ for PER and $[\ell, \mathrm{p}, s_f] = [0.5, 2, 1]$ for PER. (b): A sample of 2-dimension GP over kernel LPard with parameters $[\ell_1, \ell_2, s_f] = [0.5, 0.5, 1]$ for SEard and $[\ell_1, \ell_2, \mathrm{p}_1, \mathrm{p}_2, s_f] = [0.5, 0.5, 2, 2, 1]$ for PERard.

3.2.4 Spectral mixture

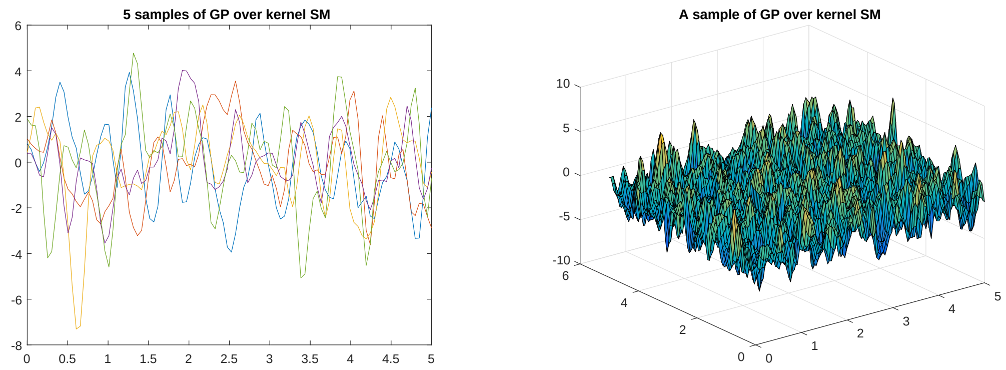
The Spectral Mixture (SM) kernel was introduced by Wilson [66] and is defined as a scaled mixture of Q Gaussians:

$$k_{SM}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{q=1}^{Q} w_q \prod_{i=1}^{n} \exp(-2\pi^2 \tau_i^2 \nu_q^{(i)}) \cos(2\pi\tau_i \mu_q^{(i)}),$$

where τ_i is the ith component of the p dimensional vector $\tau = \boldsymbol{x} - \boldsymbol{x}'$, $\{w_q\}_{q=1}^{Q}$ are the weights, the inverse means for ith component $\{1/\mu_q^{(i)}\}_{q=1}^{Q}$ represent the component periods and each inverse standard deviations for ith component $\{1/\sqrt{\nu_q^{(i)}}\}_{q=1}^{Q}$ represents the length scales [66]. Some samples of GP over kernel SM with component $Q = 2$ are shown in Figure 3.4.

3.3 Mean function

As a GP is specified by the mean function and the kernel, there is no doubt that the mean function has an influence on the performance of GP, so that it must be selected with this in mind [63]. In the majority of studies, the zero-offset mean function is usually used. Of course, simplifying the model is another important reason. All the data can be centralized so that the data can satisfy the zero mean assumption. Additionally, the mean function only dominate the predictions in



(a) (b)

FIGURE 3.4: The samples of GP over kernel SM. (a): 5 samples of GP over kernel SM with parameters $[w_1, \mu_1, \sqrt{\nu}_1, w_2, \mu_2, \sqrt{\nu}_2] = [1, 2, 1, 1, 2, 1]$. (b): A sample of 2-dimension GP over kernel SM with parameters $[w_1, \mu_1^{(1)}, \sqrt{\nu_1^{(1)}}, \mu_1^{(2)}, \sqrt{\nu_1^{(2)}}, w_2, \mu_2^{(1)}, \sqrt{\nu_2^{(1)}}, \mu_2^{(2)}, \sqrt{\nu_2^{(2)}}] = [1, 2, 1, 2, 1, 1, 2, 1, 2, 1]$.

region far from the training data [63]. For the forecasting problems, long term forecasting is influenced by the mean function more significantly than the short term forecasts.

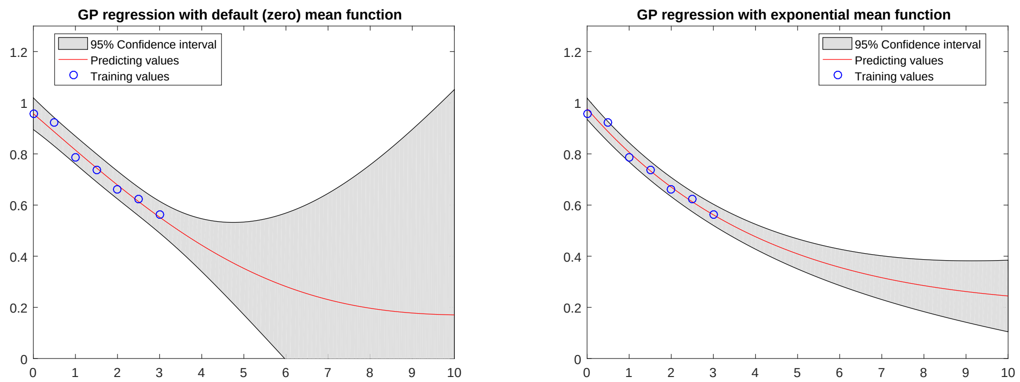


FIGURE 3.5: The effect of GPR models with the simple mean functions. The flat(zero-offset) mean function is used in the left panel while the exponential decay mean function is used in the right panel

For example, we consider the case in which we know that the observed points consist of a deterministic part with noise. That is, the training data are generated by

$$y = f + \varepsilon, \text{with } f = \exp(-0.2x),$$

where $\varepsilon \sim \mathcal{N}(0, 0.02^2)$ and the 7 points of x are equally spaced in $[0, 3]$. Therefore, our GPR model is considered as

$$y = f + \varepsilon, f \sim \mathcal{GP}\left(\mu(x; \theta_\mu), k(x, x; \theta_k)\right), \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

where the mean function μ has hyper-parameters θ_μ that incorporate the information of the deterministic part and the kernel k has hyper-parameters θ_k [63]. In this case, our observations are generated by an exponential decay function with noise. Hence, the mean function we choose is in form of $\mu(x) = a\exp(-bx)$, where a, b are unknown hyper-parameters of the mean function. Figure 3.5 (left panel) shows the GPR with SE kernel and flat mean function are used to model the 7 noisy data drawn from an exponential decay function. The GPR performs well on the training data but long term predictions are actually dominated by the flat mean function. In the right panel of Figure 3.5, the GPR with the same kernel is used, but the mean function is an exponential decay with unknown hyper-parameters. Admittedly, 7 points are sufficient for inferring the hyper-parameters in the exponential function, however, it results in long-term forecasts are dominated by the specific mean function (exponential decay) [63].

According to the simple example, an appropriate mean function significantly improve the performance of prediction of GPR. However, the effective mean function is not easy to pick up for complicated data, especially financial data. Thus the zero-offset mean functions are usually adopted in this thesis.

Given the mean function $\mu(x)$ being zero, the predictive mean in Eq.(3.16) and variance in Eq.(3.17) are given by

$$\hat{\boldsymbol{\mu}} = K(Z, X)^{\mathrm{T}}(K(X, X) + \sigma_n^2\mathbf{I})^{-1}\boldsymbol{y}, \tag{3.19}$$

$$\hat{\Sigma} = K(Z, Z) - K(Z, X)^{\mathrm{T}}(K(X, X) + \sigma_n^2\mathbf{I})^{-1}K(Z, X). \tag{3.20}$$

3.4 Model parameters estimation

In the previous sections, we have seen how to construct a Gaussian process regression model using a given kernel and zero mean function. The predictive mean and variances can be obtained as long as all the undetermined hyper-parameters are learned from the data. By the Bayesian method, we need to define a prior distribution on the hyper-parameter and integrate over them in order to make predictions, this is to say, we need to find

$$p(y_*|Z,\mathcal{D}) = \int p(y_*|Z,\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})\mathrm{d}\boldsymbol{\theta}, \tag{3.21}$$

where y_* is the sum of f_* and the noise and $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots\}$ contains all the hyper-parameters. However, this integral is usually analytically intractable. There are two useful methods to overcome this weakness [64]:

1. Approximating the integral by using the most reasonable values of the hyper-parameters $\boldsymbol{\theta}_R$, that is, $p(y_*|Z,\mathcal{D}) = p(y_*|Z,\mathcal{D},\boldsymbol{\theta}_R)$.
2. Performing the integration over $\boldsymbol{\theta}$ numerically using Monte Carlo methods [26, 23].

Despite Monte Carlo methods can perform GPR without the need of estimating hyper-parameters [26, 28, 67, 64], the common approach is approximating the integral by using the most reasonable values of the hyper-parameters due to the high computational cost of Monte Carlo methods. If the most reasonable value means the most probable value, this method is called maximum marginal likelihood. We compute the marginal likelihood in the below.

For a noisy regression, the marginal likelihood function $p(\boldsymbol{y}|X,\boldsymbol{\theta})$ is represented as

$$p(\boldsymbol{y}|X,\boldsymbol{\theta}) = \int p(\boldsymbol{y}|f,X,\boldsymbol{\theta})p(f|X,\boldsymbol{\theta})\mathrm{d}f. \tag{3.22}$$

In GPR models, the prior is Gaussian and the likelihood is also Gaussian

$$p(f|X,\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0},K), \tag{3.23}$$

$$p(\boldsymbol{y}|f,X,\boldsymbol{\theta}) = \mathcal{N}(f,\sigma_n^2\mathrm{I}). \tag{3.24}$$

Given Eq.(3.23) and Eq.(3.24), the marginal likelihood is still Gaussian

$$p(\boldsymbol{y}|X,\boldsymbol{\theta}) = \int \mathcal{N}(f,\sigma_n^2\mathrm{I})\mathcal{N}(\mathbf{0},K)\mathrm{d}f = \mathcal{N}(\mathbf{0},K+\sigma_n^2\mathrm{I}) = \mathcal{N}(\mathbf{0},\Sigma_\theta), \tag{3.25}$$

where we denote $\Sigma_\theta = K_\theta + \sigma_n^2\mathbf{I} = K + \sigma_n^2\mathbf{I}$ since $\boldsymbol{\theta}$ is involved in the covariance matrix K. The commonly used negative log marginal likelihood (NLML) is denoted by

$$\mathcal{L} = -\log p(\boldsymbol{y}|X,\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\Sigma_\theta^{-1}\boldsymbol{y} + \frac{1}{2}\log\det\Sigma_\theta + \frac{n}{2}\log 2\pi. \tag{3.26}$$

The partial derivatives of NLML with respect to the hyper-parameters are given by

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{1}{2}\mathrm{tr}(\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_i}) - \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_i}\Sigma_\theta^{-1}\boldsymbol{y}. \tag{3.27}$$

In fact, the noisy GPR models can also be considered as the noise-free regression with a noisy kernel,

$$y = f, \quad f \sim \mathcal{GP}(0, k'),$$

where $k' = k'(x_i, x_j) = k(x_i, x_j) + \delta_{ij}\sigma_n^2$ and $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$. The marginal likelihood is regarded as the Gaussian prior multiplying the identity likelihood, and of course, it remains Gaussian. It is the same expression as the noisy regression model since

$$p(\boldsymbol{y}|X, \boldsymbol{\theta}) = p(f|X, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{0}, K') = \mathcal{N}(\mathbf{0}, K + \sigma_n^2\mathrm{I}) = \mathcal{N}(\mathbf{0}, \Sigma_\theta), \tag{3.28}$$

where $K' = K'(X, X) = K(X, X) + \sigma_n^2\mathrm{I}$. The third equality is due to the definition of the kernel and the covariance matrix. In this case, the estimation of hyper-parameters must contain the noise level and then the NLML is rewritten as $\mathcal{L}(\boldsymbol{\theta}, \sigma_n^2)$. The partial derivatives of NLML with respect to σ_n^2 are given by

$$\frac{\partial}{\partial \theta_i}\mathcal{L}(\boldsymbol{\theta}, \sigma_n^2) = \frac{1}{2}\mathrm{tr}(\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_i}) - \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_i}\Sigma_\theta^{-1}\boldsymbol{y}, \tag{3.29}$$

$$\frac{\partial}{\partial \sigma_n^2}\mathcal{L}(\boldsymbol{\theta}, \sigma_n^2) = \frac{1}{2}\mathrm{tr}(\Sigma_\theta^{-1}) - \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\Sigma_\theta^{-1}\Sigma_\theta^{-1}\boldsymbol{y}. \tag{3.30}$$

It is noted that the disadvantages of maximum marginal likelihood cannot be ignored. Firstly, for many kernels the marginal likelihood function is not convex with respect to the hyper-parameters, thus the optimisation algorithm may converge to a local optimum whereas the global one may provide better results [65]. Consequently, the optimised hyperparameters achieved by maximum likelihood estimation and the performance of GPR may depend on the initial values of the optimisation algorithm [26, 28, 66, 64]. Secondly, the computation cost is also a problem. The evaluation of the gradient of the log likelihood requires the computation of the inverse matrix, which has associated computational cost that is of order n^3 and thus computing gradients is a time-consuming task for large data sets [64].

In addition to Monte Carlo and maximum marginal likelihood, cross-validation can be considered as an effective method to select a kernel and estimate the parameters, which compares various models and then choose the one with the least error [68]. It has been shown that cross-validation can outperform maximum likelihood estimation when the kernel is misspecified [68].

Admittedly, maximum marginal likelihood is still a mainstream approach in implementing GPR models and we adopt this method in the further study. Faced with the sensitivity of initial value, a common strategy adopted by most GPR practitioners is a heuristic method. That is, the optimisation is repeated using several initial values generated randomly from a simple prior distribution, which is often selected based on their expert opinions and experiences. The final estimates of the hyper-parameters are the ones with the largest likelihood values after convergence [26, 28, 66]. Further discussions about how priors of initial hyper-parameters affect the GPR models are shown in Chapter 4.

3.5 Model evaluation

One question of the essence in modelling is how to evaluate a model. We need to make sure that the predictions made by our models actually make sense and that we can depend on the models for further analysis or decision making. The principle of model evaluation is to compare the predicted values with the actual values. Of course, one asset of the Gaussian process regression model consists predictive mean and variance and both of them should be analysed in model evaluation.

Firstly, there are several ways to evaluate the accuracy of the mean predictions, including mean squared error (MSE), and mean absolute error (MAE), which are defined by

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2, \quad \text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i|,$$

where $\{\hat{y}_i\}$ and $\{y_i\}, i = 1, 2, \ldots, m$, are the predicted mean values and the actual test values, respectively. Sometimes, MSE is used as root mean squared error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}.$$

The RMSE can be affected seriously by the overall scale of the output values, thus we utilize the standardized root mean squared error (SRMSE) which is normalized by the standard deviation of $\{y_i\}$, i.e.

$$\text{SRMSE} = \frac{\text{RMSE}}{\sigma_y},$$

where σ_y^2 is the standard deviation of $\{y_i\}, i = 1, 2, \ldots, m$. This implies any model which can provide the prediction close to the sample mean of the test targets to have an SRMSE of approximately one [59]. In other words, any prediction model with the SRMSE around 1 is satisfactory.

Another measure considering both predictive mean and predictive variance is log loss. As the predictive distribution for each test point is Gaussian or Student$-t$ (it is useful in Student$-t$ related models), this log loss is defined by

$$\text{LL} = \begin{cases} \frac{1}{2}\log(2\pi\hat{\sigma}_i^2) + \frac{(y_i - \hat{y}_i)^2}{2\hat{\sigma}_i^2}, \ i = 1, 2, \ldots, m, & \text{if model is GPR} \\ \frac{1}{2}\log((\hat{\nu} - 2)\pi\hat{\sigma}_i^2) + \frac{(y_i - \hat{y}_i)^2}{(\hat{\nu} - 2)\hat{\sigma}_i^2}, \ i = 1, 2, \ldots, m, & \text{if model is TPR} \end{cases},$$

where $\{\hat{\sigma}_i\}_{i=1}^m$ are the predictive variances and $\hat{\nu}$ is the predictive degree of freedom in TPR model. This loss can be standardized by subtracting the loss that would be obtained by the null model which predicts using a Gaussian with the sample mean and sample variance of the training outputs [59]. And the mean standardized log loss (MSLL) is the average of the standardized log loss for $i = 1, 2, \ldots, m$. Therefore, the MSLL is zero for null model, and the smaller value means better model in terms of loss [59].

When it comes to compare two predictions, the mean squared error ratio (MSER) and mean absolute error ratio (MAER) are defined by

$$\text{MSER}_{1-2} = \frac{\text{MSE}_1}{\text{MSE}_2}, \quad \text{MAER}_{1-2} = \frac{\text{MAE}_1}{\text{MAE}_2},$$

where $\text{MSE}_i, \text{MAE}_i, i = 1, 2$ are the MSE and MAE for predictor 1 and predictor 2, respectively. If MSER or MAER is smaller than 1, it indicates that the predictor 1 outperforms predictor 2, and vice versa.

3.6 Summary

In this chapter, we introduce Gaussian process regression from weight-space view and function-space view in details, including all the assumptions and derivations.

In addition, the extra attentions are paid on the several important parts of GPR, including model structure, kernel, mean function and parameters estimation.

The kernel contains our presumptions about the function we wish to learn and define the closeness and similarity between data points while mean function dominates the predictions in the region far from training data. When it comes to parameter estimation, it is essential that the predictive mean and variances can be obtained only if all the undetermined parameters are learned from the data. At last, several model evaluation approaches are introduced and they will be widely used in the rest of the thesis.

Chapter 4

Initial hyper-parameters selection

4.1 Introduction

As discussed in the previous chapters, GPR is a kernel-based nonparametric method, which relies on the appropriate selection of kernel [60] and the hyper-parameters involved. The choice of the kernel has a profound impact on the performance of a GPR model, just as activation function, learning rate can affect the result of a neural network [60]. Once a kernel is selected, the unknown hyper-parameters involved in the kernel need to be estimated from the training data. Although Monte Carlo methods can perform GPR without the need of estimating hyper-parameters [26, 67, 28, 64], the common approach is to estimate the hyper-parameters by means of maximum marginal likelihood [25] due to the high computational cost of Monte Carlo methods. Unfortunately, marginal likelihood functions are not usually convex with respect to the hyper-parameters, which means local optima may exist [65] and the optimised hyper-parameters, which depend on the initial values, may not be the global optima [26, 28, 64, 66].

A common approach to tackle this issue is to use multiple starting points randomly selected from a specific prior distribution and after convergence chooses the optimised values with the largest marginal likelihood as the estimates. Therefore, the choice of prior distribution may play a vital role in the performance of GPR model. However, there exists little research in the literature to study the impact of the prior distributions on the hyper-parameter estimation and the performance of GPR. Most researchers using GPR tend to choose a simple prior distribution based on their expert opinions and experiences, such as the uniform distribution in the range of (0,1) [26, 28, 66].

In this chapter, we study the sensitivity of the hyper-parameter estimation and the performance of GPR to the prior distributions for the initial values. We consider different types of priors, including vague and data-dominated priors, for the initial values of hyper-parameters over some commonly-used kernels and investigate the influence of the priors on the performance of GPR model.

4.2 Sensitivity of prior distributions for initial hyper-parameters

According to Eq.(3.26), we find that the likelihood functions for many kernels are not always convex with respect to the hyper-parameters, therefore the optimisation algorithm may converge to a local optimum whereas the global one may provide better results [65]. As a result, the optimised hyper-parameters achieved by maximum likelihood estimation and the performance of GPR may depend on the initial values of the optimisation algorithm [26, 28, 64, 66].

A common strategy adopted by most GPR practitioners is a heuristic method. That is, the optimisation is repeated using several initial values generated randomly from a simple prior distribution, which is often selected based on their expert opinions and experiences. The final estimates of the hyper-parameters are the ones with the largest likelihood values after convergence [26, 28, 66]. It is, therefore, interesting to know how prior distributions affect the performance of GPR since the above strategy can not guarantee a global maximum of the likelihood function is found, or the sensitivity of prior distributions to the performance of GPR, which, to the best of our knowledge, has not been studied in the literature. In this thesis, we consider several different priors and study their influences to the estimates of the hyper-parameters and the performance of GPR models for some commonly used kernels.

4.3 Prior distributions of initial hyper-parameters

The prior distributions considered include non-informative [60] and data-dominated [66], which are briefly introduced as follows.

4.3.1 Vague priors

In the cases when there is little information about the data, vague prior distributions are often selected with the intention that they should have slight or no influence on the inferences [69, 70]. Many justifications and interpretations of non-informative priors have been proposed over the years, including invariance [71] and maximum entropy [72]. However, with small amounts of data, the use of non-informative prior may be problematic and a vague prior distribution may lead to significant influence on any inference made because the results are easily sensitive to the selection of prior distributions [69].

Let θ_i be the notation for a hyper-parameter in a given kernel and the uniform distribution be denoted as $\mathrm{Uniform}(a, b)$. Below we list the weakly informative prior distributions which are discussed in this thesis.

Prior 1

$$\theta_i \sim \mathrm{Uniform}(0, 1).$$

This is probably the most common prior distribution. Actually, it is not strictly a 'vague' prior since the range of the distribution is restricted. However, this prior is widely used for the estimation of the unknown parameters in GPR models.

Prior 2

$$\log(\theta_i) \sim \mathrm{Uniform}(-1, 1).$$

This prior distribution is uniform on the log hyper-parameters in $(-1, 1)$, so the range of the hyper-parameters is $(1/\mathrm{e}, \mathrm{e})$.

Prior 3

$$\log(\theta_i) \sim \mathrm{Uniform}(-10, 10).$$

This prior is similar to Prior 2 with much larger range. So the range of the hyper-parameters is approximately $(0, \mathrm{e}^{10})$.

Prior 4

$$\theta_i \sim \mathcal{N}(0, 1).$$

The standard normal prior is a popular and simple choice. It is not strictly a 'vague' prior either, and cannot be used for positive parameters.

Prior 5

$$\frac{\pi}{\theta_i} \sim \text{Uniform}(0, 1).$$

This prior is specified for the period parameter for kernels that contain periodic part. The range of the parameter is $(\pi, +\infty)$.

Prior 6

$$\log(\frac{\pi}{\theta_i}) \sim \text{Uniform}(-5, 5).$$

This prior is also specified for the period parameter. It is similar to Prior 5 but with a range $(\pi \mathrm{e}^{-5}, \pi \mathrm{e}^{5})$.

4.3.2 Data-dominated priors

Data-dominated priors are incorporated with some information inferred from training data, such as the possible range of the initial hyper-parameters. The following data-dominated priors are used in this study.

Prior 7

$$\theta_i \sim \text{Uniform}(0, \text{Nyq}).$$

This prior is also specified for the period parameter and is based on Nyquist frequency [73], where Nyq equals half the sampling rate of the data, or the half of the largest interval between input points if the data are not regularly sampled [66]. Nyquist frequency can be used to find the approximate period of data in signal processing and spectral analysis. For example, Wilson [66] used this prior to initialise the SM kernel.

Prior 8

$$\frac{1}{\theta_i} \sim \mathcal{TN}(\text{MaxI}),$$

where $\mathcal{TN}(\text{MaxI})$ is the truncated normal distribution with mean proportional to the maximal range of the inputs (MaxI) [66]. It is an improved version of Prior 4 and is used by Wilson [66] for the length scale in the SM kernel.

Prior 9

$$\frac{\pi}{\theta_i} \sim \text{Uniform}\left(\frac{\pi}{\text{MaxI}}, \pi\text{Nyq}\right).$$

This prior is also specified for the period parameters and has the range $(1/\mathrm{Nyq}, \mathrm{MaxI})$. It was first used in [66] to find a suitable range for the initial hyper-parameters.

4.4 Experiments

4.4.1 Samples from Gaussian process

In this section, we study how the priors of initial hyper-parameters affect the estimates of the hyper-parameters and the performance of GPR models using data generated from specified Gaussian processes. Since the true models are known, the accuracy of the estimates can be compared.

Letting $x_i = i$ for $i = 1, 2, \ldots, 400$, we generate samples $\{y_i\}$ from GPs with zero mean and SE and PER kernels, respectively. These two kernels are used as the demonstration because SE is the most widely-used kernel in GPR while PER is the simplest kernel which may suffer from the problem of local optima in optimisation procedure because integer multiples of the true period, such as harmonics, are often local optima [65].

To evaluate the influences of the prior distributions on the hyper-parameter estimation, ten values randomly generated from each prior distribution discussed in Section 4.3 (where applicable) are used as the starting values for the maximum likelihood procedure, implemented by the conjugate gradient algorithm. Among the ten estimates after the procedure converges the one with the largest maximum likelihood is chosen as the optimal estimate, denoted by $\boldsymbol{\theta}_{final}$, and is compared with $\boldsymbol{\theta}_{act}$.

To study the impact of the priors on the predictability of GPR, we consider two types of prediction: interpolation and extrapolation. Denote the whole data set by $\Omega = \{(i, y_i); i = 1, 2, \ldots, 400\}$. For interpolation, the test set is given by $\mathcal{D}_{I2} = \{(i, y_i); i = 5j + 1, j = 0, 1, \ldots, 79\}$ and the training set is $\mathcal{D}_{I1} = \Omega - \mathcal{D}_{I2}$. For extrapolation, the training set is $\mathcal{D}_{E1} = \{(i, y_i); i = 1, 2, \ldots, 320\}$ and the test set is $\mathcal{D}_{E2} = \Omega - \mathcal{D}_{E1}$.

4.4.1.1 Squared exponential kernel

As can be seen in Section 4.3, not all of the priors are suitable for every hyper-parameter. For the SE kernel, we use **Prior** 1, **Prior** 2, **Prior** 3 for both hyper-parameters ℓ and s_f. The data are generated by using $\boldsymbol{\theta}_{act} = [\ell, s_f] = [5, 2]$.

To compare $\boldsymbol{\theta}_{act}$ and $\boldsymbol{\theta}_{final}$, Figure 4.1 illustrates their visual positions, where "□" represents $\boldsymbol{\theta}_{act}$, "$\star$" represents $\boldsymbol{\theta}_{final}$, the "+"s are the intermediate values during the process of optimisation and the colour of the symbols stands for the value of the negative log marginal likelihood (NLML).

Apparently, regardless of the priors, the optimisation converges very fast and the estimated hyper-parameter $\boldsymbol{\theta}_{final}$ is always very close to $\boldsymbol{\theta}_{act}$.

We now test the prediction performance by the GPR with SE kernel. Only **Prior 1** is used since the estimated hyper-parameters from different priors are almost the same. The samples are generated using two GP models with two different hyper-parameters: $\boldsymbol{\theta}_{act} = [5, 2]$ and $\boldsymbol{\theta}_{act} = [15, 7]$, respectively. And the above experiment is repeated 20 times and the average results are reported in Table 4.1. As demonstration, Figure 4.2 shows typical predictions of GP and the corresponding SRMSEs and MSLLs.

It is obvious that the mean estimate of $\boldsymbol{\theta}_{final}$ is very close to $\boldsymbol{\theta}_{act}$ with small standard errors for both cases, and the GPR model performs well and stably for both interpolation and extrapolation predictions.

TABLE 4.1: Results of GP predictions with SE kernel (the standard errors are given in the brackets)

		Interpolation		
$\boldsymbol{\theta}_{act}$		$\boldsymbol{\theta}_{final}$	SRMSE	MSLL
ℓ	5	4.97 (0.133)	0.03 (0.004)	-3.48 (0.130)
s_f	2	2.00 (0.187)		
ℓ	15	14.95 (0.467)	0.01 (0.002)	-4.74 (0.229)
s_f	7	6.94 (1.110)		
		Extrapolation		
$\boldsymbol{\theta}_{act}$		$\boldsymbol{\theta}_{final}$	SRMSE	MSLL
ℓ	5	4.98 (0.165)	1.02 (0.141)	-0.14 (0.114)
s_f	2	1.97 (0.210)		
ℓ	15	15.01 (0.495)	1.19 (0.535)	-0.56 (0.322)
s_f	7	7.01 (1.242)		

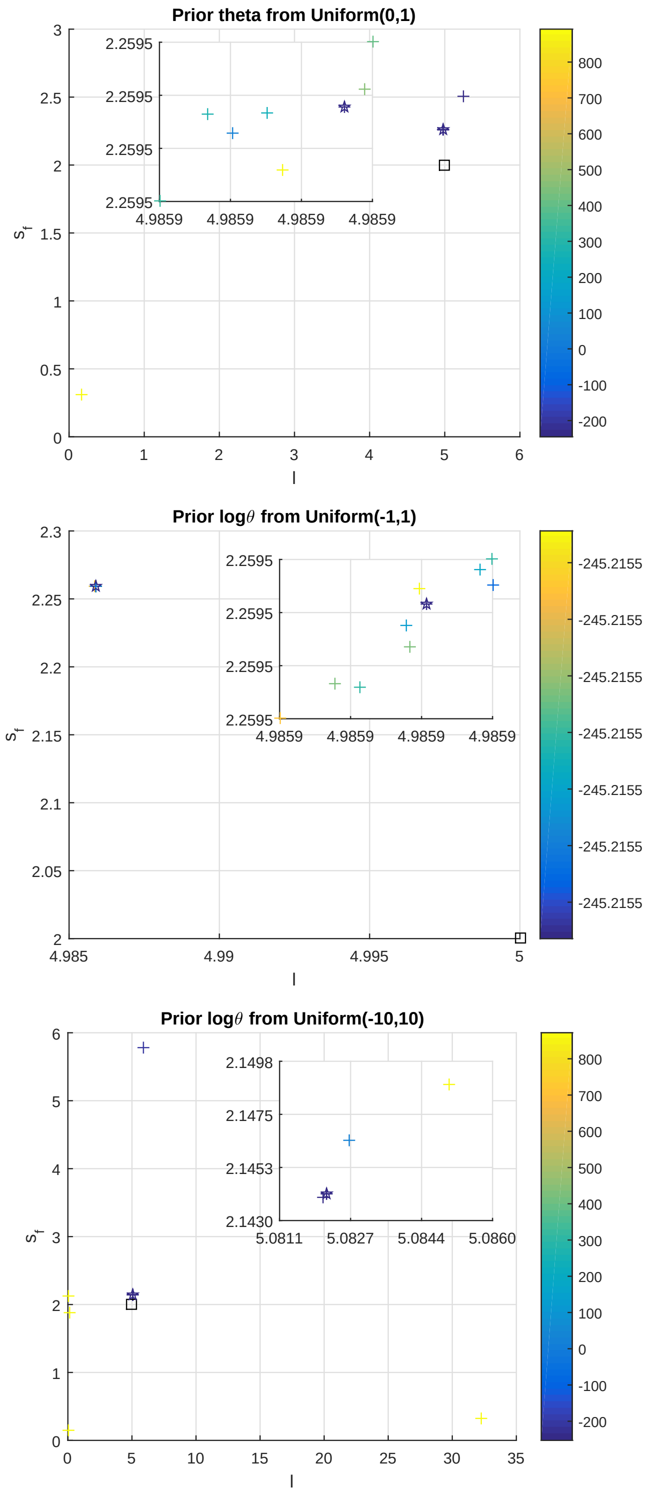


FIGURE 4.1: Positions of the estimated hyper-parameters for the SE kernel. Top to bottom: Priors 1, 2 and 3.

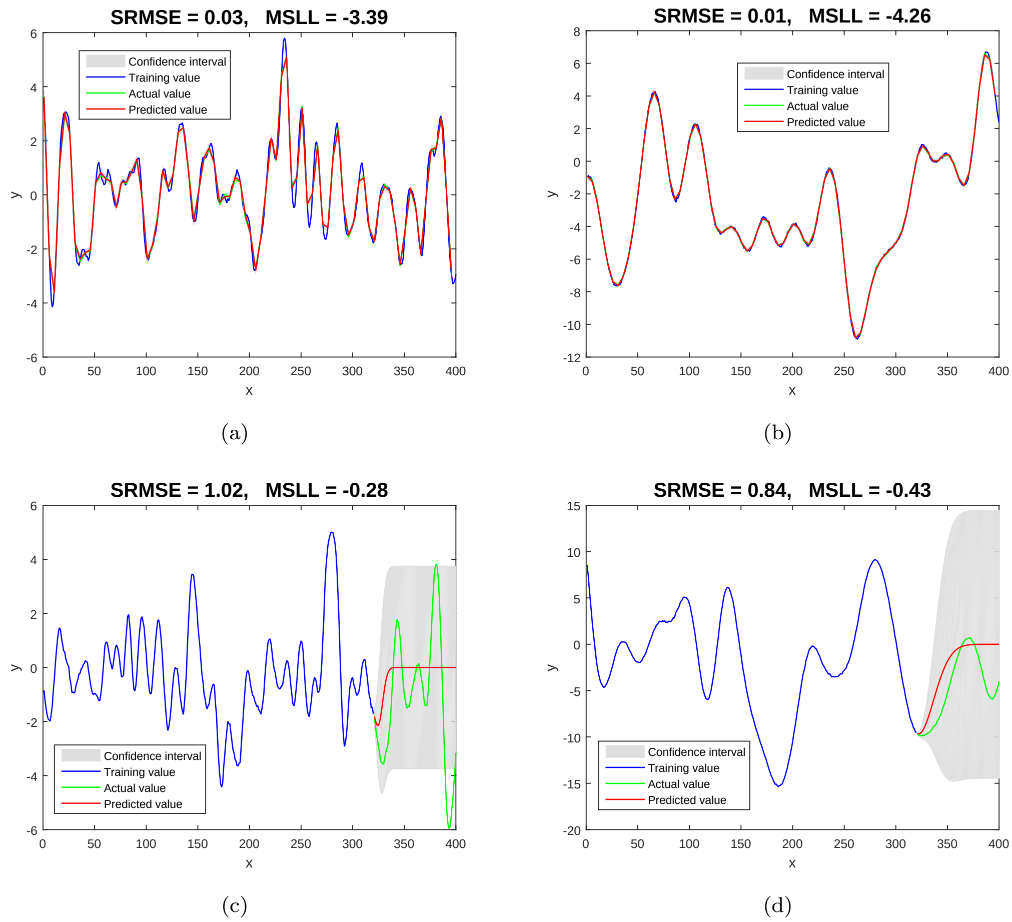


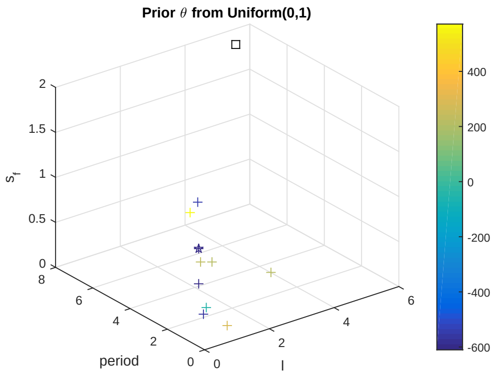
FIGURE 4.2: GP predictions with SE kernel using Prior 1. (a) and (b): interpolations; (c) and (d): extrapolations. (a) and (c): $\boldsymbol{\theta}_{act} = [5, 2]$; (b) and (d): $\boldsymbol{\theta}_{act} = [15, 7]$.

4.4.1.2 Periodic kernel

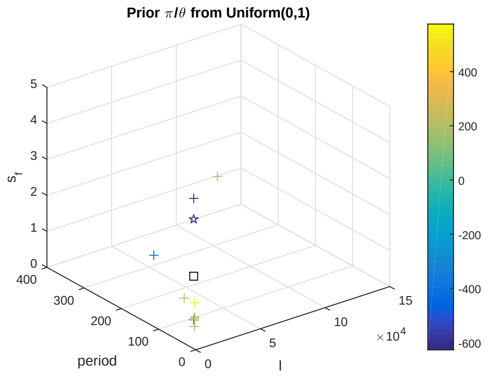
Three parameters $[\ell, \mathrm{p}, s_f]$ are involved in the PER kernel. We consider five priors (**Prior** 1, **Prior** 5, **Prior** 6, **Prior** 7 and **Prior** 9) for the p term and **Prior 1** for the parameters ℓ and s_f. In the following experiment, the data are generated using the true parameters $\boldsymbol{\theta}_{act} = [5, 7, 2]$.

Figure 4.3 shows the visual positions of $\boldsymbol{\theta}_{act}$ and $\boldsymbol{\theta}_{final}$, where the symbols have the same meanings as in Figure 4.1. It can be seen that, for all the priors considered, the estimates $\boldsymbol{\theta}_{final}$ are always far away from the true value $\boldsymbol{\theta}_{act}$. Therefore, it is difficult to achieve the global maximum by the maximum marginal likelihood method for the PER kernel, and the estimates are very sensitive to prior distributions of the initial hyper-parameters.

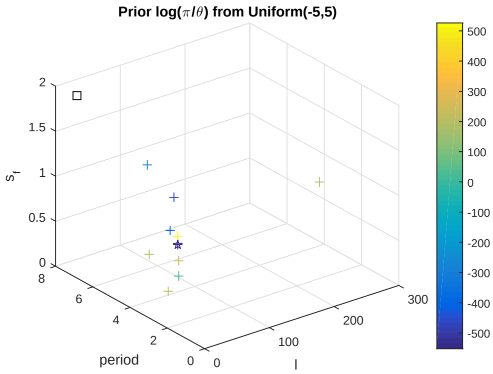
The same strategy as for the SE kernel is used to test the prediction performance by the GPR with PER kernel, and the results are reported in Table 4.2. It can be



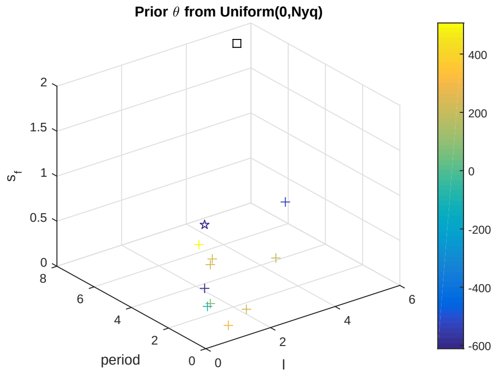
(a)



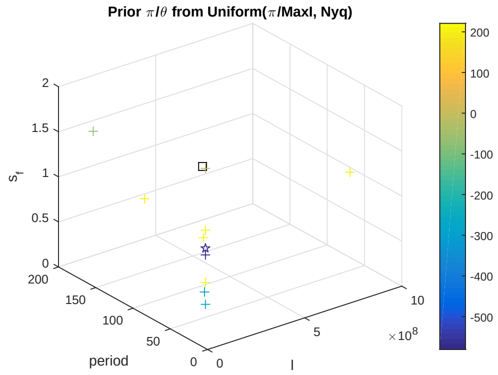
(b)



(c)



(d)



(e)

FIGURE 4.3: Positions of the estimated hyper-parameters for the PER kernel. The priors for the p term are: (a) **Prior** 1, (b) **Prior** 5, (c) **Prior** 6, (d) **Prior** 7 and (e) **Prior** 9.

seen that the averages of the estimated hyper-parameters are very different than the true values either, which confirms that the estimates obtained by numerical optimisation of likelihood function are biased. However, both the means and standard deviations of SRMSE and MSLL are very small, which indicates that the GPR models perform very well and stably for both interpolation and extrapolation, despite the poor estimates of the hyper-parameters. Therefore, although the parameter estimation for the PER kernel is sensitive to prior distributions, the GPRs still provide good results and the performance is hardly influenced by the choice of priors.

4.4.2 Samples from time series

It is of interest to investigate how prior distributions of the hyper-parameters influence the predictability of GPR if the data are generated from other models.

We consider a simple time series model ARMA(2,1) with autoregressive coefficient $[0.8, -0.45]$ and moving average coefficient -0.5, and generate 400 samples $\{y_i, i = 1, 2, \ldots, 400\}$ with $x_i = i$ and the starting values $y_1 = y_2 = 1$. Here we consider extrapolation only as this type of prediction is more meaningful in time series modelling. We select the first 320 data points as the training data and the rest as the test data. The GPR models are applied using two composite kernels: local periodic (LP) and spectral mixture (SM) with 4 components, both of which are known as useful kernels for data with complex pattern [66]. For LP kernel, different priors are used for the p parameter while Prior 1 is used for all the remaining parameters. For SM kernel, the three parameters $[w_q, \mu_q, \nu_q]$ are involved. However, w_q can be initialised as constants proportional to the standard deviation of the data [66]. Therefore we only focus on the remaining hyper-parameters μ_q and ν_q (since only 1-dimension input, $\mu_q^{(1)} = \mu_q$ and $\nu_q^{(1)} = \nu_q$). We denote

$$[\mu_q, \sqrt{\nu_q}] \sim PS_{ij}, \text{ if } \mu_q \sim \textbf{ Prior } i \text{ and } \sqrt{\nu_q} \sim \textbf{ Prior } j,$$

where $i = 5, 6, 7, 9$ and $j = 1, 8$. Here PS_{78} is the priors used by Wilson [66].

For comparison of the performance, the prediction is also performed using the true model ARMA(2,1) with the true parameters. The experiment is repeated 20 times and the averages and the standard deviations are reported in Tables 4.3 and 4.4, respectively.

Table 4.2: Results of GP predictions with PER kernel (the standard errors are given in the brackets)

Interpolation					
Prior	$\boldsymbol{\theta}_{act}$		$\boldsymbol{\theta}_{final}$	SRMSE	MSLL
Prior 1	ℓ	5	0.25 (0.170)	0.35 (0.454)	-1.44 (1.205)
	p	7	1.98 (2.807)		
	s_f	2	2.41 (2.453)		
Prior 5	ℓ	5	4.24 (8.968)	0.48 (0.815)	-1.63 (1.365)
	p	7	4.81 (1.293)		
	s_f	2	95.98 (205.308)		
Prior 6	ℓ	5	1.19 (0.713)	0.28 (0.252)	-1.50 (0.705)
	p	7	2.98 (2.288)		
	s_f	2	3.58 (6.246)		
Prior 7	ℓ	5	1.45 (1.289)	0.28 (0.253)	-1.48 (0.696)
	p	7	0.34 (0.143)		
	s_f	2	1.51 (0.838)		
Prior 9	ℓ	5	1.67 (1.978)	0.28 (0.252)	-1.51 (0.712)
	p	7	13.54 (16.614)		
	s_f	2	39.26 (80.484)		
Extrapolation					
Prior	$\boldsymbol{\theta}_{act}$		$\boldsymbol{\theta}_{final}$	SRMSE	MSLL
Prior 1	ℓ	5	1.23 (1.048)	0.14 (0.041)	-1.98 (0.287)
	p	7	0.40 (0.254)		
	s_f	2	1.43 (1.230)		
Prior 5	ℓ	5	17.87 (47.950)	0.24 (0.119)	-1.52 (0.492)
	p	7	7.73 (3.126)		
	s_f	2	56.50 (214.170)		
Prior 6	ℓ	5	2.01 (2.023)	0.24 (0.120)	-1.51 (0.493)
	p	7	2.20 (1.323)		
	s_f	2	2.38 (3.984)		
Prior 7	ℓ	5	24.18 (87.495)	0.17 (0.103)	-1.91 (0.538)
	p	7	0.25 (0.117)		
	s_f	2	139.47 (612.532)		
Prior 9	ℓ	5	4.07 (6.696)	0.18 (0.106)	-1.82 (0.563)
	p	7	5.80 (2.898)		
	s_f	2	4.71 (9.551)		

Table 4.3: Results of GP predictions with LP kernel for ARMA data (the standard errors are given in the brackets)

	GPR with LP kernel		ARMA(2,1)	
Priors	SRMSE	MSLL	SRMSE	MSLL
Prior 1	1.006 (0.0221)	-0.001 (0.0151)		
Prior 5	1.007 (0.0223)	-0.001 (0.0154)		
Prior 6	1.006 (0.0219)	-0.001 (0.0150)	1.006 (0.0143)	-0.002(0.0077)
Prior 7	1.006 (0.0219)	-0.001 (0.0150)		
Prior 9	1.005 (0.0225)	-0.002 (0.0149)		

Table 4.4: Results of GP predictions with SM kernel for ARMA data (the standard errors are given in the brackets)

	GPR with SM kernel		ARMA(2,1)	
Priors	SRMSE	MSLL	SRMSE	MSLL
PS_{51}	1.008 (0.0251)	0.001 (0.0207)		
PS_{61}	1.007 (0.0236)	-0.001 (0.0187)		
PS_{71}	1.009 (0.0255)	0.001 (0.0210)		
PS_{91}	1.006 (0.0252)	-0.002 (0.0188)	1.006 (0.0143)	-0.002 (0.0077)
PS_{58}	1.036 (0.0498)	0.038 (0.0415)		
PS_{68}	1.043 (0.0509)	0.036 (0.0450)		
PS_{78}	1.019 (0.0350)	0.012 (0.0351)		
PS_{98}	1.032 (0.0490)	0.028 (0.0441)		

The results show that for both LP and SM kernels, the performance of the GPR models has no significant differences using different prior distributions, and is comparable to that by the true model. In other words, the performance of GPR models is not sensitive to the choice of prior distributions and is as good as the true model as far as this experiment concerns.

4.5 Summary

In this chapter, we conduct the simulation studies to investigate the influences of various prior distributions of the initial hyper-parameters in GPR models to the parameter estimation and the predictability of the models when numerical optimisation of likelihood function was utilised. Nine commonly used priors and four kernels, including two basic kernels (SE and PER) and two composite kernels (LP and SM), are considered.

The numerical results show that the sensitivity of the hyper-parameter estimation depends on the choice of kernels. The estimates for SE kernel are robust regardless of the prior distributions, whilst they are very different using different priors for PER kernel which implies that the prior distributions have a huge impact on the estimates of the parameters. However, it is interesting to see that the GPR models always perform well in terms of predictability, despite the poor estimates of the hyper-parameters in some cases. Particularly the performances of the GPR models using various priors are consistently comparable with that of the true time series model in terms of prediction. Overall, the prior distributions of the hyper-parameters have little impact on the performance of GPR models.

Chapter 5

Financial time series prediction using Gaussian process regression and its extensions

5.1 Introduction

The financial market is a complicated dynamic system with a massive amount of time series trading data, including price (opening, closing, high, low and adjusted closing price) and volume. Therefore, it is natural to study how to forecast the financial time series using the trading data. It is an attractive topic for investors and scholars since a successful prediction can help the investors make excess profits.

As discussed in the previous chapters, GPR has been widely used in the financial market prediction, and as a powerful non-parameter tool, shows the superior ability in forecasting [29, 30, 31, 32]. It is known that Gaussian distribution is an important assumption of GPR. A Gaussian distribution assumes that all values in the sample will be distributed equally around the mean given enough observations. Approximate 99.73% of all variations falls within three standard deviations of the mean and hence there is the only 0.27% chance of an extreme event occurring. This property is essential because nearly all the classical models in economics assume normality. However, the financial market is a dynamic complex system, far less than perfect and heavily influenced by unpredictable human behaviours leading to the fat tails' risk. Briefly, according to these uncontrolled factors to the financial market, the real market has far more than 0.27% chance to occur an extreme event. Indeed, the conventional financial wisdom was faced with a huge challenge, especially after the 2008 Financial Crisis. Even when everything is in normal operation, unexpected and uncontrolled events can still pose a threat. These potentially catastrophic events lay stress on the ongoing relevance of heavy tails throughout the finance industry [35].

With the development of financial market theory, many researchers have already found evidence that the distribution of financial time series is not Gaussian [34, 35]. Specifically, the empirical distributions of financial data have heavier tails in the two sides than those from a Gaussian distribution [34]. As a result, some fat-tailed distributions, such as Student$-t$ distribution, Pareto distribution, Lévy distribution, and the family of stable distribution, are applied in various financial time series models [34]. Some heavy-tailed distributions can be used in the extension of GPR. As a result, some extensions of Gaussian process, such as Gaussian process regression with Student$-t$ likelihood and Student$-t$ process regression, can be applied to financial time series to capture the heavy tails of financial time series.

5.2 Gaussian process regression with Student$-t$ likelihood

Considering the fat tails, the simple approach is to substitute Gaussian noise in the regression model for Student$-t$ noise, that is, Gaussian process regression with Student$-t$ likelihood (GPRT). The likelihood satisfies Student$-t$ distribution,

$$p(\boldsymbol{y}|f) = \mathcal{T}(\nu, \mathbf{0}, \sigma_n^2 \mathrm{I}). \tag{5.1}$$

where ν is the degree of freedom and σ_n^2 is the scale parameter, which has the similar meaning to Gaussian noise [74].

Unfortunately, the integral of Gaussian multiplying Student$-t$ likelihood is analytically intractable, resulting in that the marginal likelihood has to be approximated by some numerical methods, including Expectation Maximisation (EM), Variational Bayesian(VB) and Laplace approximation [37]. More details can be found in [37].

5.3 Student$-t$ process regression

Although GPRT adapts the fat-tailed time series to some degree, the latent process is still GP [37]. If the latent GP is replaced by Student$-t$ process (TP), GPR model is extended to Student$-t$ process regression (TPR) model, which is used to capture the fat-tails. Shah et al.[38] point out that TP can be an alternatives to GP as a non-parametric method in prediction problems because TP can retain the desirable properties of GP model, such as non-parametric representation given a known kernel, analytic marginal and predictive distribution, and easy model choice based on covariance functions [38]. A brief introduction of Student$-t$ process regression(TPR) model is reviewed as follows.

Analogous to GPR model, our discussions are from function-space view since the Student$-t$ distribution also has the analytic marginal and predictive properties as same as normal distribution. It is noted that TPR model is only discussed in noise-free regression with a noisy kernel since the sum of two independent Student$-t$ distribution is analytically intractable [38].

5.3.1 Predictive model

Given n pairs of observations $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n, \boldsymbol{x}_i \in \mathbb{R}^p, y_i \in \mathbb{R}$, we assume the following TPR model

$$f \sim \mathcal{TP}(\nu, \mu, k), y_i = f(\boldsymbol{x}_i), i = 1, \ldots, n. \tag{5.2}$$

Similarly, in order to predict $f_* = f(Z)$ at the test locations $Z = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m]$, the joint distribution of the training observations $\boldsymbol{y}$ and the predictive targets f_* are given by

$$\begin{bmatrix} \boldsymbol{y} \\ f_* \end{bmatrix} \sim \mathcal{T}\left(\nu, \begin{bmatrix} \boldsymbol{\mu}(X) \\ \boldsymbol{\mu}(Z) \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma_n^2 \mathbf{I} & K(Z,X)^{\mathrm{T}} \\ K(Z,X) & K(Z,Z) \end{bmatrix}\right), \tag{5.3}$$

where ν is the degree of freedom and other symbols are the same in Eq.(3.14). Therefore, taking advantage of Theorem 2.6, the predictive distribution is

$$p(f_*|X, \boldsymbol{y}, Z) \sim \mathcal{T}\left(\hat{\nu}, \hat{\boldsymbol{\mu}}, \hat{\Sigma}\right),$$

where $\hat{\nu} = \nu + n, \beta = (\boldsymbol{y} - \boldsymbol{\mu}(X))^{\mathrm{T}} K(X,X)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}(X))$ and,

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= K(Z,X)^{\mathrm{T}} K(X,X)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}(X)) - \boldsymbol{\mu}(Z), & (5.4)\\ \hat{\Sigma} &= \frac{\nu + \beta - 2}{\nu + n - 2} \left(K(Z,Z) - K(Z,X)^{\mathrm{T}} K(X,X)^{-1} K(Z,X)\right). & (5.5) \end{aligned}$$

The predictive mean and covariance given the test locations and training set are

$$\mathbb{E}[f_*|X, \boldsymbol{y}, Z] = \hat{\boldsymbol{\mu}}, \quad \operatorname{cov}[f_*|X, \boldsymbol{y}, Z] = \hat{\Sigma}. \tag{5.6}$$

Similarly, the zero mean function is usually used in TPR model and the expressions of the predictive mean and covariance are

$$\begin{aligned} \mathbb{E}[f_*|X, \boldsymbol{y}, Z] &= K(Z,X)^{\mathrm{T}} K(X,X)^{-1} \boldsymbol{y}, & (5.7)\\ \operatorname{cov}[f_*|X, \boldsymbol{y}, Z] &= \frac{\nu + \beta - 2}{\nu + n - 2} \left(K(Z,Z) - K(Z,X)^{\mathrm{T}} K(X,X)^{-1} K(Z,X)\right), & (5.8) \end{aligned}$$

where $\beta = \boldsymbol{y}^{\mathrm{T}} K(X,X)^{-1} \boldsymbol{y}$.

5.3.2 Parameters estimation

Analogue to GPR model, all the undetermined parameters, containing hyper-parameters in the kernel and the degree of freedom, have to be estimated by maximum likelihood. The numerical method for optimisation is also the conjugate gradient method. As the definition of TPR model, it is noise-free model, and thus the negative log marginal likelihood given training inputs X, hyper-parameters $\boldsymbol{\theta}$, and the degree of freedom, is

$$\begin{aligned} \mathcal{L} &= -\log p(\boldsymbol{y}|X, \boldsymbol{\theta}, \nu)\\ &= \frac{\nu + n}{2} \log(1 + \frac{\beta}{\nu - 2}) + \frac{1}{2} \log\det K_\theta + \frac{n}{2} \log((\nu - 2)\pi)\\ &\quad + \log\Gamma(\frac{\nu}{2}) - \log\Gamma(\frac{\nu + n}{2}), & (5.9) \end{aligned}$$

where $\beta = \boldsymbol{y}^{\mathrm{T}} K_\theta^{-1} \boldsymbol{y}$ and $K_\theta = K(X,X)$. Using the matrix calculus equalities, we obtain the partial derivatives of NLML with respect to the different undetermined hyper-parameters and the degree of freedom,

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{1}{2}\mathrm{tr}(K_\theta^{-1}\frac{\partial K_\theta}{\partial \theta_i}) - \frac{\nu+n}{2(\beta+\nu-2)} \cdot \boldsymbol{y}^{\mathrm{T}} K_\theta^{-1}\frac{\partial K_\theta}{\partial \theta_i}K_\theta^{-1}\boldsymbol{y}, \tag{5.10}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \nu} &= \frac{1}{2}\log\left(1+\frac{\beta}{\nu-2}\right) - \frac{\beta(\nu+n)}{2(\nu-2)(\beta+\nu-2)} \\ &\quad + \frac{n}{2(\nu-2)} + \frac{1}{2}\psi(\frac{\nu}{2}) - \frac{1}{2}\psi(\frac{\nu+n}{2}),\end{aligned} \tag{5.11}$$

where $\psi(x)$ is the derivative of $\Gamma(x)$.

In Chapter 3.4, we have discussed the noise-free regression model with a noisy kernel. This method is reasonable for TPR, namely, adding the noise in the kernel. It is necessary and useful since noise is unavoidable in financial time series [38]. Adding the noise in the kernel, the negative log marginal likelihood is $\mathcal{L}(\nu, \boldsymbol{\theta}, \sigma_n^2)$ and the partial derivatives of the negative log marginal likelihood with respect to three parameters are

$$\begin{aligned}\frac{\partial}{\partial \nu}\mathcal{L}(\nu, \boldsymbol{\theta}, \sigma_n^2) &= \frac{1}{2}\log\left(1+\frac{\beta}{\nu-2}\right) - \frac{\beta(\nu+n)}{2(\nu-2)(\beta+\nu-2)} \\ &\quad + \frac{n}{2(\nu-2)} + \frac{1}{2}\psi(\frac{\nu}{2}) - \frac{1}{2}\psi(\frac{\nu+n}{2}),\end{aligned} \tag{5.12}$$

$$\frac{\partial}{\partial \theta_i}\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\theta}, \sigma_n^2) = \frac{1}{2}\mathrm{tr}(\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_i}) - \frac{\nu+n}{2(\beta+\nu-2)} \cdot \boldsymbol{y}^{\mathrm{T}}\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_i}\Sigma_\theta^{-1}\boldsymbol{y}, \tag{5.13}$$

$$\frac{\partial}{\partial \sigma_n^2}\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\theta}, \sigma_n^2) = \frac{1}{2}\mathrm{tr}(\Sigma_\theta^{-1}) - \frac{\nu+n}{2(\beta+\nu-2)} \cdot \boldsymbol{y}^{\mathrm{T}}\Sigma_\theta^{-1}\Sigma_\theta^{-1}\boldsymbol{y}, \tag{5.14}$$

where $\Sigma_\theta = K_\theta + \sigma_n^2\mathrm{I}$ and now $\beta = \boldsymbol{y}^{\mathrm{T}}\Sigma_\theta^{-1}\boldsymbol{y}$.

The non-convexity of the marginal likelihood function is inescapable so that TPR is faced with the sensitivity of initial hyper-parameters. As a consequence, we take the same heuristic method as GPR to address this problem in the further studies.

5.3.3 Relation to Gaussian process

Apparently, the model representations of GPR and TPR are similar. In particular, the predictive mean of TPR is the same as GPR explicitly conditioned on the same kernel with the same hyper-parameters [38], by comparison of Eq.(3.19) and Eq.(5.7). However, these two models do not have the same result because the marginal likelihoods of GPR and TPR are different, resulting in different estimations of the undetermined parameters. When it comes to predictive covariance, there is an essential difference. From the view of Eq.(3.20) and Eq.(5.8), the predictive covariance of TPR explicitly relies on the training observations while GPR's predictive covariance is independent of training set. Obviously, the predictive covariance of TPR definitely differs from that of GPR, after learning kernel hyper-parameters [38].

It is known that the degree of freedom of Student$-t$ distribution controls how heavy-tailed the distribution is. A smaller value of the degree of freedom corresponds to the fatter tails. If the degree of freedom goes to infinity, the Student$-t$ distribution converges to Gaussian distribution as well as the relationship between GP and TP.

Theorem 5.1. *Let $f \sim \mathcal{TP}(\nu, \mu, k)$ and $g \sim \mathcal{GP}(\mu, k)$. Then f tends to g in distribution as $\nu \to \infty$.*

In addition to Theorem 5.1, Shah et al. [38] point out that, perhaps less intuitively, the predictive distribution converges to a GP as n goes to infinity. In fact, GPR is a special case of TPR with the infinite degree of freedom. Therefore, TPR can resolve some problems that GPR model cannot do, without loss the ability of prediction performance at the expense of one more parameter estimation, and may capture more characteristics of data.

5.4 Experiments of model comparisons

In this section, we apply all the models discussed above to several real financial time series prediction. The 10 major equity indices we selectively use are listed in Table 5.1 and all the data is collected from Bloomberg Terminal. The outline of model comparison experiments are described in the following.

Firstly, GPR, GPRT, and TPR models are used to predict four equity indices, including INDU, NDX, SPX, and UKX, over the 2013 - 2014 period, called Experiment 5.4.2. It is a glance at the overall perspective of the performances of these models.

Secondly, we further consider GPR, GPRT, TPR, and ARMA models in the predictions of eight equity indices, including DAX, HSI, INDU, NDX, NKY, SENSEX, SPX, and UKX, over the 2013 - 2014 period, called Experiment 5.4.3. Compared with the first comparison experiment, one more classical time series model,

Table 5.1: 10 main equity indices in the world

Country/Region	Ticker	Equity Index
US	NDX	NASDAQ 100 Index
US	INDU	Dow Jones Industrial Average
US	SPX	S&P 500 Index
Germany	DAX	DAX Index
UK	UKX	FTSE 100 Index
Japan	NKY	Nikkei 225 Index
Hong Kong	HSI	Hong Kong Hang Seng Index
India	SENSEX	S&P BSE SENSEX Index
Turkey	XU100	Borsa Istanbul 100 Index
China	SHSZ300	Shanghai Shenzhen CSI Index

ARMA, is considered and four more equity indices, namely, DAX, HSI, NKY, and SENSEX, are used for predictions.

Thirdly, the model validations of GPR and TPR predictions, including leave-one-out cross validation (LOO-CV), k-fold, and sliding window, are extensively studied for all the listed equity indices over the 2013- 2014 period. These experiments are called Experiment 5.4.4.1, Experiment 5.4.4.2, and Experiment 5.4.4.3, respectively.

5.4.1 Data pre-processing

Before starting the discussion of model comparison experiments, the data pre-processing has to be discussed at first. The historical trading data of the stock market mainly contains closing prices, opening prices, high prices, low prices and volumes over the specific period, e.g. day, week, month, quarter and year. Many investors and traders pay more attentions on the closing price since it has been the value of a stock on a trading period until it changes on the next trading period.

Compared with the closing price, the adjusted closing price is more widely considered. The adjusted closing price is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time prior to the next period's opening, including the dividends, stock splits, and new stock offerings. The adjusted closing price is commonly used in data analysis because it can ensure all the data more smooth and accurate without too much excessive fluctuation. Of course, if we just consider the equity indices, there are no adjusted closing index, hence we just take closing, opening, high, low index and volume into account.

Besides, further data pre-processing has to been operated before modelling since the trading data is quite complicated and rambling. Briefly, there are two approaches as follows.

5.4.1.1 Zero-mean normalisation

Zero-mean normalisation is a common tool to do data pre-processing in machine learning. The zero-normalized data $\{\tilde{y}_i\}_{i=1}^n$ is defined by original data $\{y_i\}_{i=1}^n$ as

$$\tilde{y}_i = \frac{y_i - \mu}{\sigma},$$

where μ and σ are the mean and standard deviation of the sample data $\{y_i\}_{i=1}^n$ respectively.

5.4.1.2 Simple return and logarithmic return

Simple return and logarithmic return(log-return) are widely-used in finance, especially in quantitative finance.

Let p_t be the price (volume) at time t, then the simple return at time t is defined by,

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}.$$

The advantage of using simple return rather than prices is normalisation, which measures all variables in a comparable metric. Thus it can ensure evaluation of analytic relationship amongst two or more variables despite originating from price series of unequal values, which is essential in many multi-dimensional problems in machine learning[75]. For example, interpreting an equity covariance matrix is performed reasonably when the variables are both measured in percentage [75].

The log-return R_t at time t can also be defined by,

$$R_t = \log(1 + r_t) = \log \frac{p_t}{p_{t-1}}.$$

The strength of using Log-return is also apparent. According to the logarithmic identity, we can obtain

$$\sum_{i=1}^{n} R_i = \sum_{i=1}^{n} \log \frac{p_i}{p_{i-1}} = \log \prod_{i=1}^{n} \frac{p_i}{p_{i-1}} = \log \frac{p_n}{p_0}$$

That is to say, compounding returns over n periods is the same as the log return between initial and final periods. From the view of algorithmic complexity, this result reduces $O(n)$ multiplications to $O(1)$ additions [75]. Moveover, this sum is powerful for cases in which returns diverge from normal since the central limit theorem tells us that the sample average of that sum can converge to normality [75].

It is noted that log-return is called continuous return occasionally and the length of return series is one less than the original series.

5.4.2 The comparison of Gaussian process regression, Gaussian process regression with Student$-t$ likelihood and Student$-t$ process regression for equity index series prediction

We firstly consider a simple example (called Experiment 5.4.2) to show the performances of GPR, GPRT and TPR models for the equity index predictions. The selected indices are INDU, NDX, SPX and UKX over 2013-2014 period. For each of the indices, we denote the whole data set as $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n+m}$, where $N = n+m$ is the number of trading days, y_i is the closing index after zero-normalization and x_i is the time. To simplify the models, we assume $x_i = i$ and the training set and test set are denoted by $\mathcal{D}_1 = \{(x_i, y_i)\}_{i=1}^{n}, \mathcal{D}_2 = \{(x_i, y_i)\}_{i=n+1}^{n+m}$, respectively. We fix 40 days[1] in the test set, with Table 5.2 presenting the training and test sets for different indices. The data pre-processing method for the closing index is zero-normalisation.

In this experiment, we attempt to make the satisfying predictions based on the simple models so that we use a linear mean function. In addition, we use SM

[1]Generally speaking, 40 days include eight trading weeks (five trading days in a week), which is actually two trading months (four weeks in a month). It is a suitable forecasting length for the prediction based on nearly two years' data.

TABLE 5.2: The training and test sets for the four equity indices in Experiment 5.4.2

	N	$\mathcal{D}_1$	$\mathcal{D}_2$
INDU	504	$\{(i, y_i)\}_{i=1}^{464}$	$\{(i, y_i)\}_{i=465}^{504}$
NDX	504	$\{(i, y_i)\}_{i=1}^{464}$	$\{(i, y_i)\}_{i=465}^{504}$
SPX	504	$\{(i, y_i)\}_{i=1}^{464}$	$\{(i, y_i)\}_{i=465}^{504}$
UKX	506	$\{(i, y_i)\}_{i=1}^{466}$	$\{(i, y_i)\}_{i=467}^{506}$

with 10 components [2] as Wilson used in [66], which can discover complex pattern effectively. The performance of these predictions are shown in Appendix A and Table 5.3 presents the detailed statistics.

TABLE 5.3: Results of the four equity indices predictions using GPR, GPRT and TPR models

		GPR	GPRT	TPR
INDU	SRMSE	1.719	1.942	**1.326**
	MSLL	-2.477	-	**-2.627**
NDX	SRMSE	2.165	2.232	**2.151**
	MSLL	-2.259	-	**-2.573**
SPX	SRMSE	1.169	2.177	**1.093**
	MSLL	-2.696	-	**-3.918**
UKX	SRMSE	1.389	1.417	**1.330**
	MSLL	**0.320**	-	0.804

For the four equity indices predictions, it can be seen that TPR outperforms GPRT and GPR in terms of SRMSE. The result is confirmed by comparing the MSLLs of these models, where TPR predictions in INDU, NDX, and SPX have smaller MSLLs.

[2] In fact, the number of components should be determined by the analysis of spectral density. However, it is too difficult to do that in analysing the financial data. Hence, we use 10 components as default, which are considered as the sufficient number of components in [66]

5.4.3 Further comparison of Gaussian process regression, Gaussian process regression with Student$-t$ likelihood, Student$-t$ process regression and ARMA model for equity index series prediction

A further experiment (called Experiment 5.4.3) using more equity indices is carried out by comparing GPR, GPRT, TPR and classical time series models for index series prediction. The selected indices are DAX, HSI, INDU, NDX, NKY, SENSEX, SPX and UKX over the 2013-2014 period. In this experiment, we determine 60 days[3] in the test set, with other details presented in Table 5.4. The data pre-processing method for closing index is zero-normalisation.

TABLE 5.4: The training and test sets for different equity indices in Experiment 5.4.3

	N	$\mathcal{D}_1$	$\mathcal{D}_2$
DAX	505	$\{(i, y_i)\}_{i=1}^{445}$	$\{(i, y_i)\}_{i=446}^{505}$
HSI	491	$\{(i, y_i)\}_{i=1}^{431}$	$\{(i, y_i)\}_{i=432}^{491}$
INDU	504	$\{(i, y_i)\}_{i=1}^{464}$	$\{(i, y_i)\}_{i=465}^{504}$
NDX	504	$\{(i, y_i)\}_{i=1}^{464}$	$\{(i, y_i)\}_{i=465}^{504}$
NKY	489	$\{(i, y_i)\}_{i=1}^{429}$	$\{(i, y_i)\}_{i=430}^{489}$
SENSEX	494	$\{(i, y_i)\}_{i=1}^{434}$	$\{(i, y_i)\}_{i=435}^{494}$
SPX	504	$\{(i, y_i)\}_{i=1}^{464}$	$\{(i, y_i)\}_{i=465}^{504}$
UKX	506	$\{(i, y_i)\}_{i=1}^{466}$	$\{(i, y_i)\}_{i=467}^{506}$

ARMA(1,1) is used for comparison since ARMA is the classical time series model, where ARMA(1,1) is the simplest one. We use GPR, GPRT, TPR and ARMA(1,1) to do the eight equity indices predictions based on the training and test sets. It is worth noting that the mean functions in GPR, GPRT and TPR are zero-offset because we cannot find significant effect of composite mean function in the predictions through the last experiment. At the same time, admittedly, SM kernel is still used, the number of components decreases from 10 to 2 because we find that the prediction curves using SM with 10 components vibrate too heavily (just like white noise) and are time consuming during computation. The performance of these predictions are shown in Appendix B, with Table 5.5 presenting the detailed statistics.

[3]In fact, 60 days = 12 trading weeks = 3 trading month = 1 trading quarter. Compared with 40 day forecasting, 60 day test set is better to be considered.

TABLE 5.5: Results of eight equity indices predictions using GPR, GPRT, TPR and ARMA(1,1) model

		GPR	GPRT	TPR	ARMA(1,1)
DAX	SRMSE	1.252	**0.844**	0.894	1.001
	MSLL	-0.286	-	**-0.632**	-
HSI	SRMSE	1.888	2.042	**1.493**	1.558
	MSLL	-0.067	-	**-0.336**	-
INDU	SRMSE	1.078	0.954	**0.948**	1.312
	MSLL	-1.452	-	**-1.577**	-
NDX	SRMSE	1.650	0.799	**0.745**	0.957
	MSLL	-1.560	-	**-1.677**	-
NKY	SRMSE	1.659	2.389	**1.350**	1.449
	MSLL	**-0.758**	-	-0.683	-
SENSEX	SRMSE	1.025	1.047	0.865	0.959
	MSLL	-0.720	-	**-3.591**	-
SPX	SRMSE	1.064	0.950	**0.854**	1.137
	MSLL	**-1.482**	-	-1.203	-
UKX	SRMSE	0.938	1.108	**0.891**	1.046
	MSLL	-0.105	-	**-0.160**	-

As the demonstration shows in Appendix B, TPR performs better than GPRT, GPR and ARMA on the whole. This result is also confirmed by Table 5.5 in terms of SRMSE. Additionally, TPR predictions are also proven better in terms of MSLL because of the six equity indices predictions with smaller MSLLs.

5.4.4 Model validation of Gaussian process regression and Student$-t$ process regression for equity index series prediction

The two experiments we have conducted above have a fixed forecasting horizon (40 days and 60 days) in the predictions and the results are obvious and intuitive, even though the statistics might be unconvincing because each experiment for each index prediction has only been conducted once. Therefore, model validation is necessary in the further studies. According to the results of the previous two experiments, GPRT and ARMA model do not provide an outstanding performance, thus we just select GPR and TPR to compare the model validation. The methods of model validation we have chosen are leave-one-out cross-validation (LOO-CV), k-fold and sliding window.

5.4.4.1 Leave-one-out cross-validation

In statistics, cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalise to an independent data set. It is a useful approach to combine average prediction error measures, which is defined as a different loss function, to derive a more accurate evaluation of the model's predictive performance [76]. In a prediction problem, the whole data set contains two parts: a training set and test set. The purpose of cross-validation is to choose a data set to "test" the model based on the training set, i.e. the validation set. In particular, LOO-CV is useful and it is one of the most widely used cross-validation, taking one observation as the validation set and the remaining observations as the training set. This is repeated for all the observations picked in the validation set.

TABLE 5.6: The training and validation sets (j^{th} validation) for 10 equity indices in Experiment 5.4.4.1

	N	$\mathcal{D}_1$	$\mathcal{D}_3$	j
DAX	505	$\{(i,y_i)\}_{i=1,i\neq j}^{505}$	(j,y_j)	$\overline{1,505}$ [4]
HSI	491	$\{(i,y_i)\}_{i=1,i\neq j}^{491}$	(j,y_j)	$\overline{1,491}$
INDU	504	$\{(i,y_i)\}_{i=1,i\neq j}^{504}$	(j,y_j)	$\overline{1,504}$
NDX	504	$\{(i,y_i)\}_{i=1,i\neq j}^{504}$	(j,y_j)	$\overline{1,504}$
NKY	489	$\{(i,y_i)\}_{i=1,i\neq j}^{489}$	(j,y_j)	$\overline{1,489}$
SENSEX	494	$\{(i,y_i)\}_{i=1,i\neq j}^{494}$	(j,y_j)	$\overline{1,494}$
SHSZ300	483	$\{(i,y_i)\}_{i=1,i\neq j}^{483}$	(j,y_j)	$\overline{1,483}$
SPX	504	$\{(i,y_i)\}_{i=1,i\neq j}^{504}$	(j,y_j)	$\overline{1,504}$
UKX	506	$\{(i,y_i)\}_{i=1,i\neq j}^{506}$	(j,y_j)	$\overline{1,506}$
XU100	501	$\{(i,y_i)\}_{i=1,i\neq j}^{501}$	(j,y_j)	$\overline{1,501}$

Currently, we apply LOO-CV to the equity indices predictions using GPR and TPR (called Experiment 5.4.4.1). The selected indices are 10 equity indices in Table 5.1 over the 2013-2014 period. The training set and the validation set denoted by $\mathcal{D}_3$ of j^{th} validation are shown in Table 5.6. We take advantage of GPR and TPR models when undertaking these indices predictions based on these

[4] $\overline{n,n+m}, n,m \in \mathbb{N}$ means $n, n+1, \cdots, n+m$.

training and validation sets. The mean functions of GPR and TPR are all zero-offset and the kernel is SE. Based on the results of the previous experiments, we find that SM kernel have little improvement of GPR, GPRT and TPR models with much computing time so that we use SE kernel, which only needs to estimate two undetermined parameters. The RMSEs [5] of LOO-CV experiment using GPR and TPR models are reported in Table 5.7(a).

Table 5.7: The RMSE of LOO-CV using GPR and TPR models

(a) Index prediction

	GPR	TPR
DAX	0.107	0.107
HSI	0.166	0.166
INDU	0.084	0.084
NDX	0.055	0.055
NKY	0.122	0.122
SENSEX	0.063	0.063
SHSZ300	0.104	**0.101**
SPX	0.067	0.067
UKX	0.147	0.147
XU100	0.190	0.190

(b) Log-return prediction

	GPR	TPR
DAX	1.004	**1.003**
HSI	0.999	**0.998**
INDU	1.004	**1.003**
NDX	1.002	1.002
NKY	1.002	1.002
SENSEX	0.989	**0.986**
SHSZ300	0.999	**0.997**
SPX	1.002	**1.001**
UKX	**1.002**	1.003
XU100	**1.000**	1.002

Actually, only zero-normalization of the closing index is used in all the experiments above and it needs to consider log-return series prediction [6] using GPR and TPR. That is, all the targets in GPR and TPR models are indices' log-return series rather than the closing indices. The results are reported in Table 5.7(b). According to both Table 5.7(a) and Table 5.7(b), TPR perform as well as GPR, even outperform in some cases, especially in terms of log-return series. This empirical result is consistent with the theoretical conclusion in Section 5.3.3. Furthermore, linear predictor is taken into comparison with GPR and TPR. Given a time series $\{y_i\}_{i=1}^n$ with equal width increment, the linear predictor $\check{y}_i$ is defined by

$$\check{y}_i = \begin{cases} y_{i+1}, & \text{if } i = 1 \\ \frac{y_{i-1}+y_{i+1}}{2}, & \text{if } i = 2, \dots, n-1 \\ y_{i-1}, & \text{if } i = n. \end{cases}$$

[5] Here we use RMSE rather than SRMSE because the SRMSE is the RMSE divided by the standard deviation of test points where only one test point in Experiment experiment:LOOCV

[6] The length of log-return series is one less than the original index series

In order to measure the predictive performance, we use mean absolute error (MAE) and root mean square error (RMSE). Figure 5.1 shows that GPR and TPR have the same predictive performance, of which both perform as well as a linear predictor, and in particular, a little better in log-return series prediction.

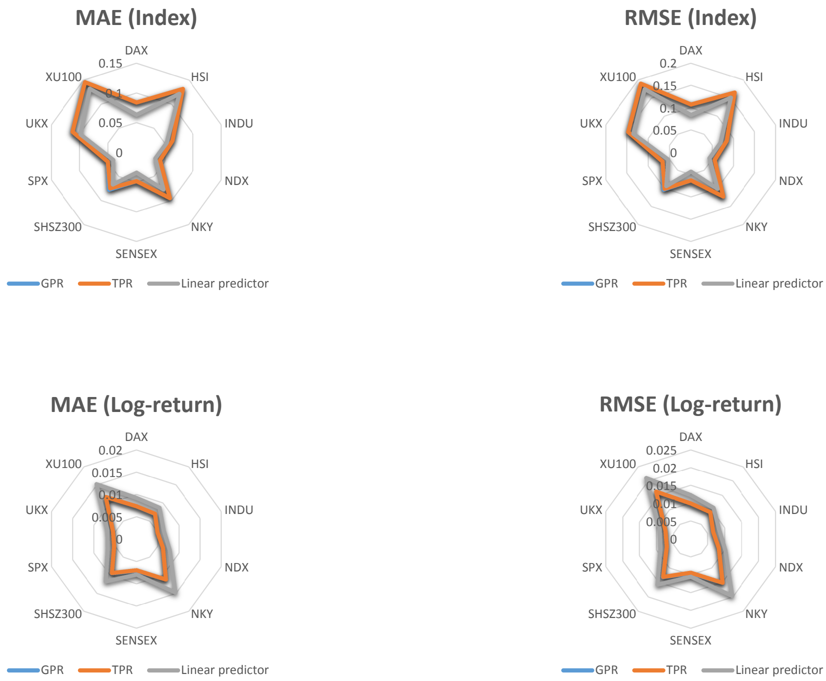


FIGURE 5.1: Graph of LOO-CV for the index and log-return prediction in terms of MAE and RMSE using GPR, TPR and liner predictor. Blue line stands for GPR model, orange stands for TPR model and gray line stands for linear predictor (Blue line is almost covered by orange line).

5.4.4.2 k-fold cross-validation

It is known that the daily fluctuation of a stock index is usually between ±3%, maximum ±10%. This may be the reason why the difference between the models is very small based on LOO-CV. As a result, another cross-validation, named k-fold cross-validation, is under consideration. In k-fold cross-validation, the original sample is randomly partitioned into k equal sized sub-samples. Among the k sub-samples, a single sub-sample is considered as the validation data and the remaining $k-1$ sub-samples regarded as training data. The cross-validation process is then repeated k times, with each of the k sub-samples used only once as the validation set. The strength of this approach is that all observations are used for both training and validation, with each observation only used for validation once. Among k-fold, 10-fold cross-validation is widely used [77].

In particular, one approach applies the k-fold cross-validation to the equity indices predictions using GPR and TPR (called Experiment 5.4.4.2) in a special case. The special case of k-fold means that the original sample is specifically partitioned into k continuous equal sized sub-samples [7] and $k = 10$ is used in this experiment. The detailed j^{th} validation set of 10 major equity indices are shown in Table 5.8 and the training set is the remaining observations in the original data set. The number of validation are all 10 since it is 10-fold cross-validation ($j = \overline{1, 10}$).

TABLE 5.8: The j^{th} validation set for 10 equity indices in Experiment 4

	N	$\mathcal{D}_3$
DAX	505	$\{(i + 50(j-1), y_{i+50(j-1)})\}_{i=6}^{55}$
HSI	491	$\{(i + 49(j-1), y_{i+49(j-1)})\}_{i=2}^{50}$
INDU	504	$\{(i + 50(j-1), y_{i+50(j-1)})\}_{i=5}^{54}$
NDX	504	$\{(i + 50(j-1), y_{i+50(j-1)})\}_{i=5}^{54}$
NKY	489	$\{(i + 48(j-1), y_{i+48(j-1)})\}_{i=10}^{57}$
SENSEX	494	$\{(i + 49(j-1), y_{i+49(j-1)})\}_{i=5}^{53}$
SHSZ300	483	$\{(i + 48(j-1), y_{i+48(j-1)})\}_{i=4}^{51}$
SPX	504	$\{(i + 50(j-1), y_{i+50(j-1)})\}_{i=5}^{54}$
UKX	506	$\{(i + 50(j-1), y_{i+50(j-1)})\}_{i=7}^{56}$
XU100	501	$\{(i + 50(j-1), y_{i+50(j-1)})\}_{i=2}^{51}$

Consequently, we predict these equity indices using GPR and TPR. The mean function and the kernel we select are the same as in Experiment 5.4.4.1. The mean and standard error (given in the brackets) of 10 validations' SRMSE are reported in Table 5.9(a). The experiment is also carried out for the log-return series, with the results presented in Table 5.9(b).

Visually, we draw Figure 5.2 to show the result of 10-fold cross-validation for both index and log-return predictions in terms of MSER and MAER[8]. Figure 5.2 shows that the TPR make the considerable predictions in the equity markets as well as GPR because both MSER and MAER are almost all around one (Only SENSEX

[7] For an ordinary k-fold, all the points, including continuous data set and discontinuous data set are considered in sub-samples. However, the discontinuous prediction of the stock market time series is useless. Hence we use the special case of k-fold

[8] Without special definitions, the MSER is $\text{MSER}_{GPR-TPR}$ and MAER is $\text{MAER}_{GPR-TPR}$

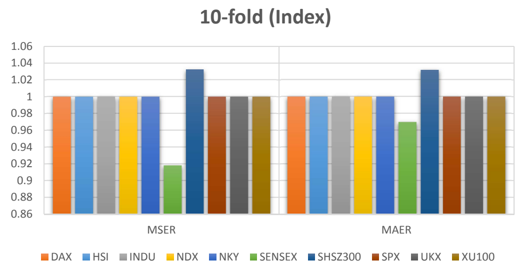
TABLE 5.9: The SRMSE of 10-fold cross-validation for the index and log-return predictions using GPR and TPR models

(a) Index prediction

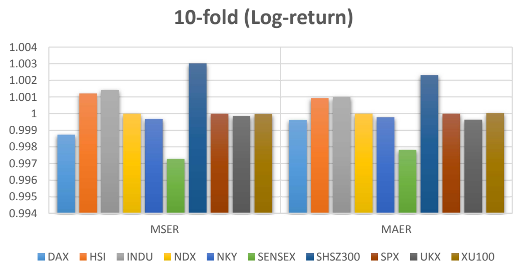
	GPR	TPR
DAX	2.727(1.541)	2.727(1.541)
HSI	1.498(0.455)	1.498(0.455)
INDU	2.565(1.239)	2.565(1.239)
NDX	4.557(4.122)	4.557(4.122)
NKY	1.905(1.229)	1.905(1.229)
SENSEX	4.545(2.318)	4.545(2.318)
SHSZ300	2.171(1.354)	**2.163**(1.370)
SPX	3.375(1.976)	3.375(1.976)
UKX	1.387(0.542)	1.387(0.542)
XU100	1.813(1.336)	1.813(1.336)

(b) Log-return prediction

	GPR	TPR
DAX	1.003(0.011)	1.003(0.011)
HSI	0.995(0.006)	**0.994**(0.006)
INDU	0.999(0.012)	1.001(0.015)
NDX	0.993(0.003)	0.993(0.003)
NKY	0.998(0.009)	0.998(0.009)
SENSEX	0.999(0.013)	1.000(0.012)
SHSZ300	1.023(0.049)	1.023(0.049)
SPX	0.994(0.005)	0.994(0.005)
UKX	0.995(0.006)	0.995(0.006)
XU100	1.001(0.014)	1.001(0.014)



(a) Index prediction



(b) Log-return prediction

FIGURE 5.2: Graph of 10-fold cross-validation for the index and log-return predictions in terms of MSER and MAER using GPR and TPR models.

is smaller than one and SHSZ300 is larger than one obviously). This also serves as another piece of evidence for the theoretical conclusion in Section 5.3.3.

5.4.4.3 Sliding window

Compared with LOO-CV and k-fold, the sliding window is a more reliable and preferable approach for evaluating a prediction model in the stock market because not only traders but also researchers focus considerably on dynamic future forecasting, rather than a historical interpolation or a static prediction. The sliding window is widely used for backtesting a statistical model using historical data to dynamically evaluate stability and the predictive accuracy of a time series model [78]. Backtesting works in the following way on the whole. The historical data is firstly divided into a training sample and a test sample. The model is then estimated using the training sample and some days' predictions made for the test sample. The training sample is then rolled ahead using a given increment, with the training and test exercises repeated until it is not possible to make any prediction. The algorithm of the sliding window analysis for the predictive performance is presented in Algorithm 1.

Algorithm 1 Sliding Window Analysis for Predictive Performance

1: Select a sliding window size, w, i.e., the number of consecutive observation in each sliding window.
2: Select a forecast horizon (the number of test points), h. The forecast horizon depends on the application and periodicity of the data. The Figure 5.3 shows how the sliding window partitions the data set.
3: Select the increments between successive rolling windows, τ. If τ is 1 period, then partition the entire data set into $n = N - \mathrm{w} + 1$ subsamples. The first sliding window contains observations for a period 1 through w, the second sliding window contains observations for period 2 through w + 1, and so on. The Figure 5.3 shows the partitions.
4: For each sliding window sub-sample, make predictions using GPR and TPR, then compute the SRMSE of these two models.
5: Compute the mean and standard derivation of all the sliding window subsamples' SRMSE, and also compute MSER and MAER.
6: Compare the SRMSE between GPR and TPR. The model with the smallest SRMSE has the best predictive performance. Besides, the predictive performance of GPR and TPR are also directly reflected by MSER and MAER

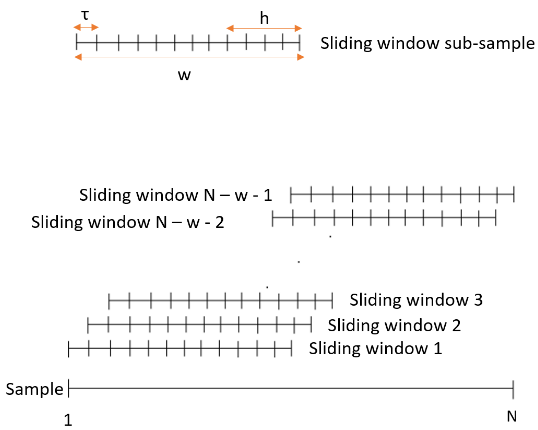


FIGURE 5.3: Graph of how sliding window partitions the data set

In our new experiment (called Experiment 5.4.4.3), all the 10 indices' trading data are used by setting $\tau = 1$ and $h = 5$, as well as $h = 20$. Since there are different trading days in the indices, we set $\mathrm{w} = \lfloor N/2 \rfloor$, where $\lfloor \cdot \rfloor$ means the nearest integers towards minus infinity and N the number of trading days. The details of j^{th} training and test sets for 10 equity indices in Experiment 5.4.4.3 are reported in Table 5.10 (in each sub-experiment, $h = 5$ and $h = 20$, respectively). In each sliding window sub-sample, the zero-offset mean function, and SE kernel

TABLE 5.10: The j^{th} training and test sets for 10 equity indices in Experiment 5

	N	$\mathcal{D}_1$	$\mathcal{D}_2$	j
DAX	505	$\{(i, y_i)\}_{i=j}^{253+j-h}$	$\{(i, y_i)\}_{i=254+j-h}^{253+j}$	$\overline{1, 252}$
HSI	491	$\{(i, y_i)\}_{i=j}^{246+j-h}$	$\{(i, y_i)\}_{i=247+j-h}^{246+j}$	$\overline{1, 245}$
INDU	504	$\{(i, y_i)\}_{i=j}^{252+j-h}$	$\{(i, y_i)\}_{i=253+j-h}^{252+j}$	$\overline{1, 252}$
NDX	504	$\{(i, y_i)\}_{i=j}^{252+j-h}$	$\{(i, y_i)\}_{i=253+j-h}^{252+j}$	$\overline{1, 252}$
NKY	489	$\{(i, y_i)\}_{i=j}^{245+j-h}$	$\{(i, y_i)\}_{i=246+j-h}^{245+j}$	$\overline{1, 244}$
SENSEX	494	$\{(i, y_i)\}_{i=j}^{247+j-h}$	$\{(i, y_i)\}_{i=248+j-h}^{247+j}$	$\overline{1, 247}$
SHSZ300	483	$\{(i, y_i)\}_{i=j}^{240+j-h}$	$\{(i, y_i)\}_{i=241+j-h}^{242+j}$	$\overline{1, 241}$
SPX	504	$\{(i, y_i)\}_{i=j}^{252+j-h}$	$\{(i, y_i)\}_{i=253+j-h}^{252+j}$	$\overline{1, 252}$
UKX	506	$\{(i, y_i)\}_{i=j}^{254+j-h}$	$\{(i, y_i)\}_{i=255+j-h}^{254+j}$	$\overline{1, 253}$
XU100	501	$\{(i, y_i)\}_{i=j}^{251+j-h}$	$\{(i, y_i)\}_{i=252+j-h}^{251+j}$	$\overline{1, 250}$

are used in both GPR and TPR models.

All the results of the sliding window analysis of the 5-day and 20-day-ahead predictions, including index and log-return, are reported in Table 5.11. The results, with respect to the fluctuation of MSER and MAER, are visualized in Figure 5.4.

For the 5-day-ahead prediction, TPR outperforms GPR in the HSI and UKX equity markets in terms of index prediction, while TPR performs better in the DAX, INDU, NDX, SENSEX and SHSZ300 markets in terms of log-return series prediction according to Table 5.11(a) and 5.11(c). For the 20-day-ahead prediction, GPR outperforms TPR in the SENSEX, SHSZ300, SPX, and UKX markets in terms of index prediction, while GPR nearly loses the weak superiority of prediction in these four markets according to Table 5.11(b) and 5.11(d). These results are also found in Figure 5.4.

When reading Table 5.11 from left to right (compare Table 5.11(a) and Table 5.11(b), and comparing Table 5.11(c) and Table 5.11(d), respectively), there is an interesting fact that, with the increase in the number of forecasting days, TPR no longer performs better in any of the markets when compared with GPR (compare

Table 5.11: The SRMSE of sliding window analysis of h-day-ahead index and log-return predictions using GPR and TPR models

(a) $h = 5$ (Index)

	GPR	TPR
DAX	**2.922(2.351)**	2.950(2.431)
HSI	2.848(5.199)	**2.833(5.198)**
INDU	**3.381(3.256)**	3.406(3.271)
NDX	3.903(3.992)	3.904(3.992)
NKY	**3.851(3.820)**	3.891(4.021)
SENSEX	**3.519(3.550)**	3.538(3.565)
SHSZ300	**4.260(3.305)**	4.282(3.332)
SPX	3.982(4.485)	3.972(4.489)
UKX	2.655(2.088)	**2.624(2.028)**
XU100	2.880(2.821)	2.880(2.821)

(b) $h = 20$ (Index)

	GPR	TPR
DAX	3.410(2.929)	3.416(2.925)
HSI	2.119(1.548)	2.120(1.545)
INDU	4.353(3.088)	4.364(3.080)
NDX	5.637(4.087)	5.637(4.087)
NKY	3.220(2.429)	3.220(2.441)
SENSEX	**5.450(3.530)**	5.458(3.535)
SHSZ300	**4.698(3.891)**	4.707(3.906)
SPX	**5.468(3.992)**	5.498(3.993)
UKX	**2.397(2.103)**	2.411(2.161)
XU100	2.945(2.579)	2.945(2.579)

(c) $h = 5$ (Log-return)

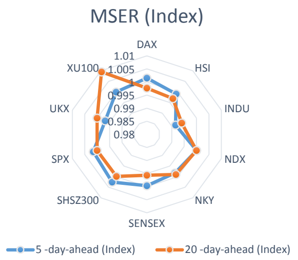
	GPR	TPR
DAX	1.048(0.190)	**1.046(0.187)**
HSI	1.066(0.270)	1.065(0.270)
INDU	1.041(0.169)	**1.040(0.168)**
NDX	1.040(0.234)	**1.039(0.233)**
NKY	1.057(0.222)	1.057(0.222)
SENSEX	1.082(0.275)	**1.077(0.272)**
SHSZ300	1.083(0.385)	**1.082(0.382)**
SPX	1.037(0.173)	1.036(0.173)
UKX	1.054(0.197)	**1.051(0.191)**
XU100	1.034(0.194)	1.034(0.194)

(d) $h = 20$ (Log-return)

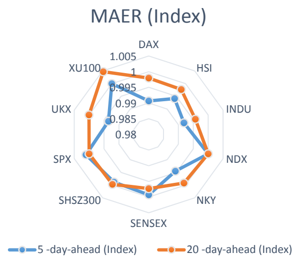
	GPR	TPR
DAX	1.003(0.029)	1.003(0.029)
HSI	1.001(0.031)	1.001(0.030)
INDU	1.007(0.050)	1.007(0.051)
NDX	1.005(0.042)	1.005(0.041)
NKY	0.998(0.029)	0.998(0.029)
SENSEX	**1.014**(0.047)	1.015(0.047)
SHSZ300	1.010(0.050)	1.010(0.049)
SPX	1.004(0.046)	1.004(0.046)
UKX	1.004(0.034)	1.004(0.036)
XU100	1.003(0.033)	1.003(0.034)

Table 5.11(a) and Table 5.11(b)), even though the difference is not apparent actually. A similar result is shown in Figure 5.4(a); that is, 5 equity (DAX, HSI, NDX, NKY, SENSEX, SHSZ300, and SPX) markets' MSER for the 5-day-ahead prediction are larger than those of the 20-day-ahead prediction, with 3 equity (INDU, UKX and XU100) markets showing opposite results. Hence, TPR gradually loses potential superiority in equity index prediction when the prediction horizon is increased to some extent.

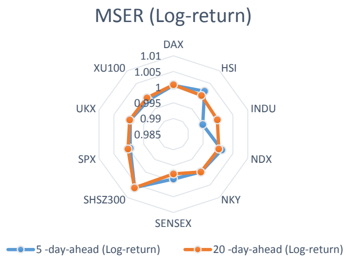
However, Figure 5.4(b) related to the other indicator, MAER, shows a completely opposite result, with there being no apparent result based on the log-return subexperiment according to Figure 5.4(c) and Figure 5.4(d). Therefore, the above conclusion remains to be verified later.



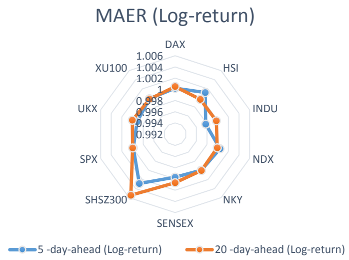
(a) $h = 5$ (Index)



(b) $h = 20$ (Index)



(c) $h = 5$ (Log-return)



(d) $h = 20$ (Log-return)

FIGURE 5.4: Graph of sliding window analysis of h-day-ahead index and log-return prediction in terms of MSER and MAER using GPR and TPR models. Blue line stands for 5-day-ahead prediction ($h = 5$) and orange line stands for 20-day-ahead prediction ($h = 20$).

5.5 State space model

As discussed previously, EMH claims that the current price reflects all past information on a stock in the efficient market. However, not all markets are efficient, so there is no doubt that the current price cannot be independent of historical prices in all markets [3]. In other words, the historical prices can determine the current price to some degree. As a result, the prediction of price tomorrow can be regarded as a function of previous prices and today's price, that is, the state-space model and historical prices construct the state. This is a natural and direct idea for traders and investors because they all want to obtain excess profits from explicit analysis of historical data, including trading information and the operation records of firms.

Although GPR has a noisy regression model and TPR does not have an analytically tractable noisy representation, both GPR and TPR models can be extended to the state-space Gaussian process regression (SSGPR) and state-space Student$-t$ process regression (SSTPR) models. A brief market efficiency analysis can be conducted based on the results of the SSGPR and SSTPR prediction models for several markets because no technical model offers a satisfactory prediction ability to obtain excess profits in an even weakly-form efficient market, even though some methods may exist that are able to offer outstanding performance in inefficient markets.

5.5.1 Model of state space

Recalling the regression model in Chapter 3.1.2, $\boldsymbol{x}_i$ is considered in $\mathbb{R}^p$ as the function of previous outputs y_i. Briefly, we take advantage of the discussion of the state-space model in [79] as follows. Consider the time series $y_{t_1}, \ldots, y_t$ and the state-space model

$$\begin{cases} \boldsymbol{x}_{t_i} = [y_{t_i-1}, \ldots, y_{t_i-L}]^{\mathrm{T}}, \\ y_{t_i} = f(\boldsymbol{x}_{t_i}) + \varepsilon_{t_i}, \end{cases} \tag{5.15}$$

where the state x at time t_i consists of previous outputs,up to a given lags L (now the dimension of input space $p = L$) and ε_{t_i} is the noise as the same in Chapter 3.1.2 with zero mean and variance σ_n^2. In addition, $f \sim \mathcal{GP}(0, k_{ard})$, where k_{ard} is the ARD kernel.

In fact, this is a one-step-ahead prediction, with the multi-step-ahead being our focus. The simplest method is to repeat the one-step-ahead prediction many times, which is a multi-step-ahead prediction, namely, the iterative method. The detailed iterative h-step-ahead forecasting method is illustrated as follows: it forecasts only one-step-ahead, using the estimate of the output of the current forecasting, and the previous outputs (up to the lag L), as the input for forecasting the next time step, until the forecasting h-step-ahead is made [79].

By means of the model (5.15) and assuming all the data is known up to time step t, the forecasting of y at $t+h$ is obtained by

$$\begin{aligned}
\boldsymbol{x}_{t+1} = [y_t, y_{t-1}, \ldots, y_{t+1-L}]^{\mathrm{T}} \;&\Rightarrow\; f(\boldsymbol{x}_{t+1}) \sim \mathcal{N}\left(\mu(\boldsymbol{x}_{t+1}), \sigma^2(\boldsymbol{x}_{t+1})\right), \\
&\quad\;\; \hat{y}_{t+1} = \mu(\boldsymbol{x}_{t+1}) \\
\boldsymbol{x}_{t+2} = [\hat{y}_{t+1}, y_t, \ldots, y_{t+2-L}]^{\mathrm{T}} \;&\Rightarrow\; f(\boldsymbol{x}_{t+2}) \sim \mathcal{N}\left(\mu(\boldsymbol{x}_{t+2}), \sigma^2(\boldsymbol{x}_{t+2})\right), \\
&\quad\;\; \hat{y}_{t+2} = \mu(\boldsymbol{x}_{t+2}) \\
&\vdots \\
\boldsymbol{x}_{t+h} = [\hat{y}_{t+h-1}, \hat{y}_{t+h-2}, \ldots, \hat{y}_{t+h-L}]^{\mathrm{T}} \;&\Rightarrow\; f(\boldsymbol{x}_{t+h}) \sim \mathcal{N}\left(\mu(\boldsymbol{x}_{t+h}), \sigma^2(\boldsymbol{x}_{t+h})\right), \\
&\quad\;\; \hat{y}_{t+h} = \mu(\boldsymbol{x}_{t+h}),
\end{aligned}$$

where the one point predictive mean $\mu(\boldsymbol{x}_{t+h-i})$ and variance $\sigma^2(\boldsymbol{x}_{t+h-i})$ are computed using Eq.(3.19) and Eq.(3.20). The above steps do not consider uncertainty information since $\hat{y}_{t+i}$ equals the predictive mean, rather than following a normal distribution.

Girard et al. point out in [79] that the predictive mean is consistent after propagating uncertainty information, while the predictive variance will be more complicated but more realistic. In order to simplify the model, we consider the model without uncertainty information only since the predictive mean is more meaningful to us.

Similarly, we attempt to achieve SSTPR model. The state-space model is now

$$\begin{cases} \boldsymbol{x}_{t_i} = [y_{t_i-1}, \ldots, y_{t_i-L}]^{\mathrm{T}}, \\ y_{t_i} = f(\boldsymbol{x}_{t_i}), \end{cases} \tag{5.16}$$

where $f \sim \mathcal{TP}(\nu, 0, k_{ard})$, where ν is the degree of freedom and k_{ard} is the noisy ARD kernel, and other parameters have the same meaning in Eq.(5.15). Therefore, the prediction of y at $t+h$ is computed via

$$\begin{aligned}
\boldsymbol{x}_{t+1} = [y_t, y_{t-1}, \ldots, y_{t+1-L}]^{\mathrm{T}} \;&\Rightarrow\; f(\boldsymbol{x}_{t+1}) \sim \mathcal{T}\left(\nu_1, \mu(\boldsymbol{x}_{t+1}), \sigma^2(\boldsymbol{x}_{t+1})\right), \\
&\quad\;\; \hat{y}_{t+1} = \mu(\boldsymbol{x}_{t+1}) \\
\boldsymbol{x}_{t+2} = [\hat{y}_{t+1}, y_t, \ldots, y_{t+2-L}]^{\mathrm{T}} \;&\Rightarrow\; f(\boldsymbol{x}_{t+2}) \sim \mathcal{T}\left(\nu_2, \mu(\boldsymbol{x}_{t+2}), \sigma^2(\boldsymbol{x}_{t+2})\right), \\
&\quad\;\; \hat{y}_{t+2} = \mu(\boldsymbol{x}_{t+2}) \\
&\vdots \\
\boldsymbol{x}_{t+h} = [\hat{y}_{t+h-1}, \hat{y}_{t+h-2}, \ldots, \hat{y}_{t+h-L}]^{\mathrm{T}} \;&\Rightarrow\; f(\boldsymbol{x}_{t+h}) \sim \mathcal{T}\left(\nu_h, \mu(\boldsymbol{x}_{t+h}), \sigma^2(\boldsymbol{x}_{t+h})\right), \\
&\quad\;\; \hat{y}_{t+h} = \mu(\boldsymbol{x}_{t+h}),
\end{aligned}$$

where the one point predictive mean $\mu(\boldsymbol{x}_{t+h-i})$ and variance $\sigma^2(\boldsymbol{x}_{t+h-i})$ are computed using Eq.(5.7) and Eq.(5.8). Of course, the predictive degree of freedom and variance are also inaccurate because we do not consider the uncertainty of inputs. Similar to SSGPR, we focus on the predictive mean and consider the model without uncertainty information.

5.5.2 Experiments for equity index series prediction

According to the models above, we consider the L days' historical closing prices of equity index as the state in the SSGPR and SSTPR to predict the next few days prices using iterative h-step-ahead method. This is Experiment 5.5.2.

The data used in this experiment is all the closing trading prices of the 10 equity indices. Due to the expensive computation, we initially do the 10-fold cross-validation for the whole data set of each index, and only selected the last fold's data as the test set. The dynamic analysis is necessary and thus the sliding window analysis is used in this experiment, fixing the prediction horizon at $h = 5$ and the increments at $\tau = 5$. Furthermore, we consider 10 and 20 historical days' prices as the state in the sub-experiments respectively so that the lag is $L = 10$ and $L = 20$. Similarly, due to the different trading days in the indices and the different lags, the sliding window size of each equity index is computed as

$$\mathrm{w} = 9 \times \lfloor N/10 \rfloor - L + 1.$$

In each sliding window sub-sample, the zero-offset mean function, and SE kernel are used in both GPR and TPR models. Table 5.12 shows the results of 5-day-ahead prediction using SSGPR and SSTPR with $L = 10$ and $L = 20$, respectively.

It can be seen that SSTPR clearly outperforms SSGPR in the major equity markets when the model considers $L = 10$ days' historical indices as the state to forecast the following 5 days' indices. The result is less apparent when the number of historical days' indices increases to $L = 20$ days. In other words, SSTPR gradually lose superiority in the prediction of equity indices after comparison with SSGPR when the historical data information becomes sufficient according to this experiment.

To further analyse SSGPR and SSTPR, we repeated Experiment 5.5.2 by increasing L from 5 to 20, denoted as Experiment 5.6. The SRMSEs of the 10 indices' performances are demonstrated in Figure 5.5. Roughly speaking, DAX predic-

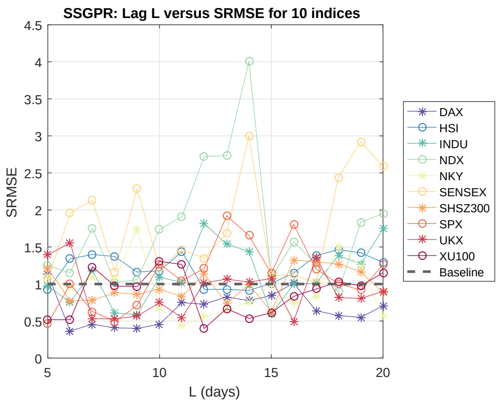
TABLE 5.12: The SRMSE of sliding window analysis of 5-day-ahead predictions using SSGPR and SSTPR models

(a) $L = 10$

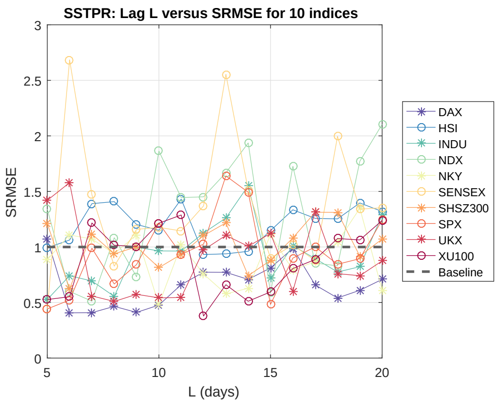
	SSGPR	SSTPR
DAX	1.805(0.961)	**1.651(0.774)**
HSI	1.708(1.478)	**1.631(0.958)**
INDU	11.087(17.576)	**5.417(7.496)**
NDX	8.182(8.691)	**4.374(3.554)**
NKY	2.942(2.395)	**2.661(1.383)**
SENSEX	9.458(21.797)	**5.001(5.894)**
SHSZ300	4.742(4.593)	**4.701(4.413)**
SPX	**3.616(2.919)**	4.528(3.757)
UKX	1.973(1.415)	**1.938(1.037)**
XU100	2.383(1.347)	2.108(1.637)

(b) $L = 20$

	SSGPR	SSTPR
DAX	**1.871(1.109)**	2.135(1.388)
HSI	1.826(0.938)	**1.595(0.545)**
INDU	**9.652(14.517)**	12.98(24.272)
NDX	7.942(6.550)	**7.367(5.129)**
NKY	**4.145(4.521)**	5.134(4.880)
SENSEX	8.871(11.590)	**8.182(12.837)**
SHSZ300	5.212(5.091)	5.662(4.594)
SPX	6.478(8.194)	**4.657(4.478)**
UKX	**1.529(0.742)**	1.889(1.428)
XU100	3.005(3.591)	**2.564(1.545)**



(a) SSGPR



(b) SSTPR

FIGURE 5.5: The SRMSE of prediction with the increasing L from 5 to 20.

tions by both SSGPR and SSTPR models are always satisfying since the SRMSEs are all smaller than 1. However, INDU, SENSEX and SPX predictions are not stable and sometimes they have very bad performances because the SRMSEs are sometimes greater than 1.5. From the view of the increasing of lag L, the median SRMSE of 10 indices' predictions by SSGPR and SSTPR are shown in Figure 5.6.

Overall, SSTPR outperforms SSGPR throughout the period, even though both experience some fluctuation. SSTPR is at a great advantage when it has a significantly lower median SRMSE value before the lag reaches 8. As the lags increase, the gap between SSTPR and SSGPR narrows, with SSGPR beating SSTPR twice when the lag arrives at 9 and 13, respectively. When the lag continues to rise to 14, a drastic drop is seen in SSTPR prediction, showing that SSTPR regains its advantage at this point, with this positive feature remaining stable until the end when the lag climbs to 20, in which there is a similar performance for SSGPR.

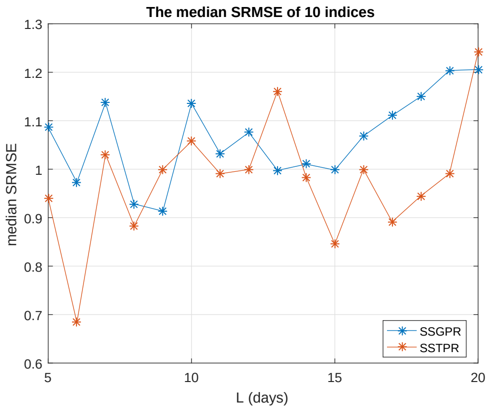


Figure 5.6: The median SRMSE of 10 indices' predictions

5.6 Stock market efficiency analysis

As previously discussed, the technical analysis is based on historical trading data, including a variety of technical prediction models and heavily depending on market efficiency. For example, in a weakly-formed efficient market, no technical model can make an outstanding prediction. As a consequence, the market efficiency is reflected by the usefulness of technical forecasting models. Therefore, we undertake a further market efficiency analysis based on the results of our SSGPR and SSTPR models.

(a) $L = 10$ (b) $L = 20$

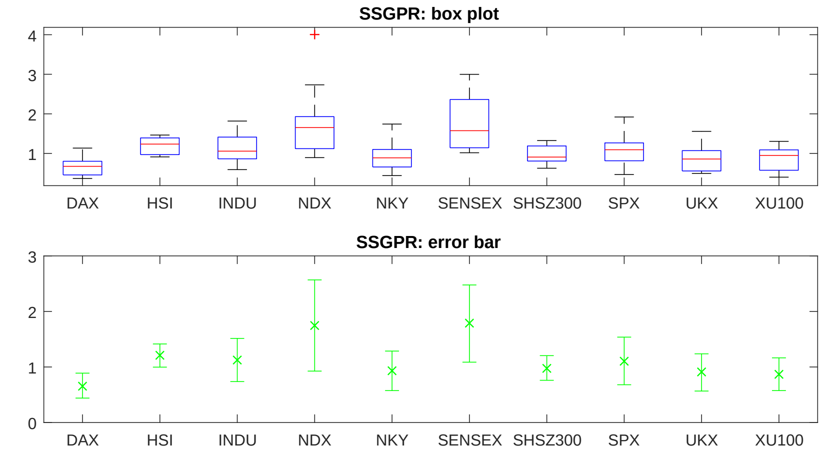
FIGURE 5.7: The SRMSE of sliding window analysis using SSGPR and SSTPR (5-day-ahead prediction, L days lagged).

Initially, in order to show the results of the performance of SSGPR and SSTPR predictions in different markets visually, we draw a bar chart in Figure 5.7 based on the average value of SRMSEs using SSGPR and SSTPR in Table 5.12(a) and Table 5.12(b).

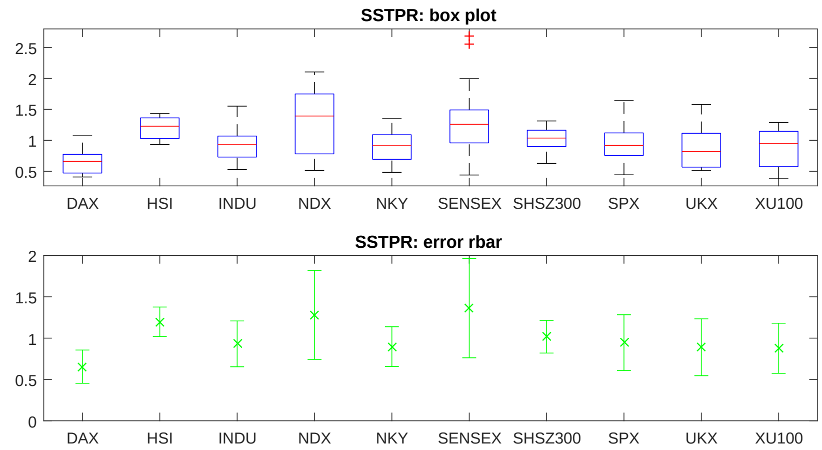
It can be seen that INDU and SENSEX have the largest SRMSE among the 10 equity indices, no matter what state-space model we choose and no matter how many days historical indices we consider as the state. Therefore, INDU (US) undoubtedly has strong efficiency in our market list. Of course, as our know, the Dow Jones Industrial Average (INDU) is a stock market index that shows how 30 large publicly owned companies based in the United States have traded during a standard trading session in the stock market and that it is the one of the most developed stock markets throughout the world. Concerning SENSEX, it looks abnormal because the Indian market is emerging and should be less inefficient, something which will be discussed later.

Scientifically, the explicit results should be tested using an adequate experimental analysis. We further study the results of Experiment 5.6, which is actually Experiment 5.5.2 repeated 16 times from lag $L = 5$ to $L = 20$, using a box plot and error bar to graphically depicting the 16 times SRMSEs of the 10 indices. The box plot is useful for identifying outliers and for comparing distributions of data, with the error bar a graphical representation of the variability of data. The box plot and error bar analysis of the 16 times experiments is presented in Figure 5.8.

According to Figure 5.8, NDX and SENSEX have the worst predictability since both have a comparatively tall box plot with several outliers and long error bars with comparatively large means and standard deviations, of which both are mainly greater than one. This suggests that the NDX and SENSEX markets are the most



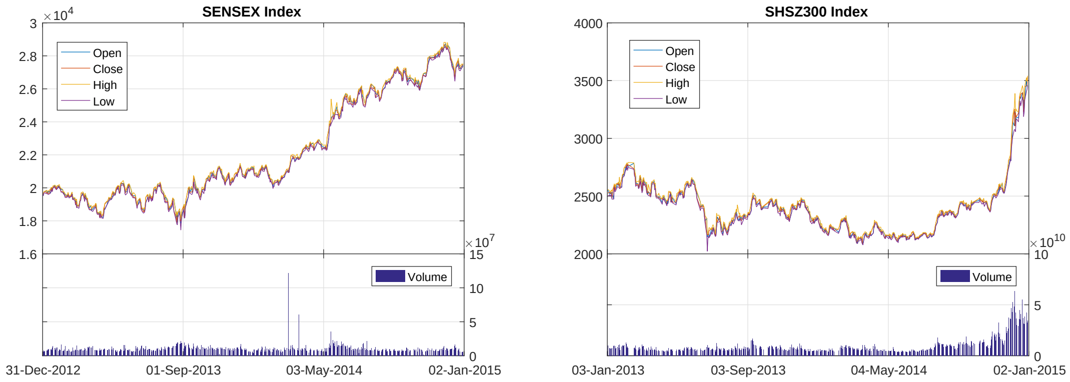
(a) SSGPR



(b) SSTPR

FIGURE 5.8: Box plot and error plot of 16 times experiments for 10 indices.

efficient during this period when compared to the other markets. NDX comprises 100 of the largest domestic and international non-financial securities listed on The NASDAQ stock exchange and based on market capitalization. The index reflects companies across major industry groups of the US stock market, with there being no doubt that NDX is proven to be efficient. An interesting result is from SENSEX and shown in Figure 5.9(a). Ankit Agarwal (2006) tested the weak form of the Indian stock market using the simple technical trading rules, finding high predictability of Indian stock markets, even after regarding transaction costs [5]. However, Gourishankar and Jyoti (2014) point out that the Indian stock market is becoming efficient [80]. Compared to these two conclusions, our interesting result from SENSEX can serve as evidence supporting the latter inference. In addition, it can be seen that HSI has a comparatively short box plot and error bar, of which both are great than one. Concerning DAX, both the box plot and error bar are under one. These results reveal that the HSI market is more efficient and that DAX is less efficient. For other markets, we cannot conclude any explicit result because the means and medians of SRMSES for the remaining indices are all around one, with a moderate box plot and error bar.



(a) SENSEX (India) (b) SHSZ300 (China)

FIGURE 5.9: Graph of two market indices in the two years period between 2013 and 2014.

Another confusing result concerns the Chinese SHSZ300. Compared with NKY, SHSZ300 seems to have more efficiency, but actually the Chinese market is an emerging market, while the stock market in Japan is comparatively developed. The candle chart for SHSZ300 is shown in Figure 5.9(b), highlighting that SHSZ300 rises perpendicularly all of a sudden at the end of 2014. It is an abnormal rise influenced by complicated economic and political efforts, which are impossible to reflect in previous trading data. Maybe these special influences can account for abnormal efficiency in the Chinese market.

5.7 Summary

This chapter mainly introduce GPR and its extensions, including GPRT and TPR, showing TPR as an alternative to GPR when undertaking financial time series prediction. We have taken advantage of 10 equity indices from 1 January 2013 to 31 December 2014, to test and compare these models using various methods, containing leave-one-out cross-validation (LOO-CV), k-fold cross-validation and sliding window. The results can be concluded as follows.

By simply comparing GPR, TPR and Gaussian Process Regression with Student$-t$ likelihood (GPRT) using INDU, NDX, SPX and UKX indices with 40 test days, we have found that TPR completely outperform GPR. This conclusion is consistent after another simple comparison of GPR, GPRT, TPR and the classical time series model ARMA(1,1) using many more indices, DAX, HSI, INDU, NDX, NKY, SENSEX, SPX and UKX with 60 test days.

However, the conclusion is not so apparent in the more statistic based experiments. After comparing GPR and TPR based on 10 major equity indices from around the world using LOO-CV and k-fold cross-validation, the performance of TPR is the same as GPR. Specifically, the performance of GPR and TPR are not good in terms of index prediction in LOO-CV, with even GPR and TPR not have better index prediction than simple linear predictor.

When our discussion consider sliding window analyses, the TPR model has a slightly better predictive performance than GPR, especially when making short-term predictions, e.g. a one-week-ahead prediction in specific markets. To conclude, GPR and TPR can make a considerable prediction of equity indices. The frameworks of GPR and TPR are flexible and easy to extend, so we can take other historical trading information, such as opening price, highest price, lowest price and volume, into consideration when using GPR and TPR models.

Furthermore, we introduce all the GPR and TPR models in state-space, namely, SSGPR and SSTPR. They are more realistic models because all the historical prices are taken into account when we predict the next day's or next few days' prices. This is attractive for traders and investors because they all want to obtain excess profits from the explicit analysis of historical data. Of course, the premise is the inefficient market. We apply SSGPR and SSTPR models to the dynamic predictions of 10 indices using sliding windows. The overall results are that SSTPR outperforms SSGPR for the equity index prediction. Based on the detailed results, a brief market efficiency analysis confirms that the developed markets are unpredictable on the whole. Admittedly, there are several outliers and it needs to be considered that these results are just a naive glimpse into market efficiency analysis.

Chapter 6

Multivariate Gaussian and Student$-t$ process regression for multi-output prediction

6.1 Introduction

It has been shown that GPR and its extensions are proved to be useful in financial time series prediction. Despite the popularity of GPR in various modelling tasks, there still exists a conspicuous imperfection, that is, the majority of GPR models are implemented for single response variables or considered independently for multiple responses variables without consideration of their correlation [40, 41]. In order to resolve multi-output prediction problem, Gaussian process regression for vector-valued function is proposed and regarded as a pragmatic and straightforward method. The core of this method is to vectorise the multi-response variables and construct a "big" covariance, which describes the correlations between the inputs as well as between the outputs (see, e.g. [40, 41, 42, 43]). This modelling strategy is feasible due to that the matrix-variate Gaussian distributions can be reformulated as multivariate Gaussian distributions [42, 44]. Intrinsically, Gaussian process regression for vector-valued function is still a conventional Gaussian process regression model since it merely vectorises multi-response variables of which are assumed to follow a developed case of GP with a reproduced kernel. Although this vectorisation method of conventional GPR model has been proven to be a useful tool to deal with multiple response variable prediction, it cannot be extended to more general elliptical processes model such as Student$-t$ process regression

(TPR) for vector-valued function, because the equivalence between matrix-variate and multivariate Student$-t$ distributions does not exist under the vector operator [44].

To overcome this drawback, we propose another derivation of dependent Gaussian process regression, named as multivariate Gaussian process regression (MV-GPR), where the model settings, derivations and computations are all directly performed in matrix form, rather than vectorizing the matrices as done in the existing methods. MV-GPR is a more straightforward method, and can be implemented in the same way as the conventional GPR. Based on the derivation of MV-GPR, we further introduce the multivariate Student$-t$ process and then derive a new considerable method, multivariate Student$-t$ process regression (MV-TPR) for multi-output prediction. The usefulness of proposed methods are illustrated through several simulated examples. Furthermore, we also verify empirically that MV-TPR has superiority in the prediction based on some widely-used datasets, including air quality prediction and bike rent prediction. The proposed methods are then applied to stock market modelling and shown to make the profitable stock investment strategies.

In this chapter, (1), We propose a concise and straightforward derivation of dependent Gaussian process regression, MV-GPR. (2), Based on the derivation of MV-GPR, we can easily extend to MV-TPR. (3), The effectiveness of the proposed MV-GPR and MV-TPR are illustrated through several simulated examples. (4), We apply MV-GPR and MV-TPR to produce profitable investment strategies in the stock markets. (5), MV-TPR shows its superiority in the air quality prediction and the bike rent prediction.

6.2 Multivariate process definitions

Following the definition of Gaussian process, a multivariate Gaussian process should be a collection of random vector-valued variables, any finite number of which have matrix-variate Gaussian distribution. Therefore, we define a multivariate Gaussian process as follows.

Definition 6.1 (MV-GP)**.** $\boldsymbol{f}$ is a multivariate Gaussian process on $\mathcal{X}$ with vector-valued mean function $\boldsymbol{u} : \mathcal{X} \mapsto \mathbb{R}^d$, covariance function (also called kernel) $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ and positive semi-definite parameter matrix $\Omega \in \mathbb{R}^{d\times d}$ if any finite collection of vector-valued variables have a joint matrix-variate Gaussian distribution,

$$[\boldsymbol{f}(x_1)^{\mathrm{T}}, \ldots, \boldsymbol{f}(x_n)^{\mathrm{T}}]^{\mathrm{T}} \sim \mathcal{MN}(M, \Sigma, \Omega), n \in \mathbb{N},$$

where $\boldsymbol{f}, \boldsymbol{u} \in \mathbb{R}^d$ is a row vector which components are the functions $\{f_i\}_{i=1}^d$ and $\{\mu_i\}_{i=1}^d$ respectively. Furthermore, $M \in \mathbb{R}^{n\times d}$ with $M_{ij} = \mu_j(x_i)$, and $\Sigma \in \mathbb{R}^{n\times n}$ with $\Sigma_{ij} = k(x_i, x_j)$. Sometimes Σ is called the column covariance matrix while Ω is the row covariance matrix. We denote $\boldsymbol{f} \sim \mathcal{MGP}(\boldsymbol{u}, k, \Omega)$.

Furthermore, we define the multivariate Student−t process based on the definition of MV-GP and Student−t process proposed by [38].

Definition 6.2 (MV-TP)**.** $\boldsymbol{f}$ is a multivariate Student−t process on $\mathcal{X}$ with parameter $\nu > 2$, vector-valued mean function $\boldsymbol{u} : \mathcal{X} \mapsto \mathbb{R}^d$, covariance function (also called kernel) $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ and positive semi-definite parameter matrix $\Omega \in \mathbb{R}^{d\times d}$ if any finite collection of vector-valued variables have a joint matrix-variate Student−t distribution,

$$[\boldsymbol{f}(x_1)^{\mathrm{T}}, \ldots, \boldsymbol{f}(x_n)^{\mathrm{T}}]^{\mathrm{T}} \sim \mathcal{MT}(\nu, M, \Sigma, \Omega), n \in \mathbb{N},$$

where $\boldsymbol{f}, \boldsymbol{u} \in \mathbb{R}^d$ is a row vector which components are the functions $\{f_i\}_{i=1}^d$ and $\{\mu_i\}_{i=1}^d$ respectively. Furthermore, $M \in \mathbb{R}^{n\times d}$ with $M_{ij} = \mu_j(x_i)$, and $\Sigma \in \mathbb{R}^{n\times n}$ with $\Sigma_{ij} = k(x_i, x_j)$. We denote $\boldsymbol{f} \sim \mathcal{MTP}(\nu, \boldsymbol{u}, k, \Omega)$.

6.3 Multivariate process regression models

6.3.1 Multivariate Gaussian process regression

For a Gaussian process regression, the noisy model, $y = f(x) + \varepsilon$ is usually considered. However, for a MV-TPR, there may exist the same analytically intractable problems. In order to unify the model derivation, we adopt the same method used in [38] for both MV-GPR and MV-TPR and consider a noise free regression model with noise-incorporated kernel.

Given n pairs of observations $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n, \boldsymbol{x}_i \in \mathbb{R}^p, \boldsymbol{y}_i \in \mathbb{R}^d$, we assume the following model

$$\begin{aligned} \boldsymbol{f} &\sim \mathcal{MGP}(\boldsymbol{u}, k', \Omega), \\ \boldsymbol{y}_i &= \boldsymbol{f}(\boldsymbol{x}_i), \text{for } i = 1, \cdots, n, \end{aligned}$$

where $k' = k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \delta_{ij}\sigma_n^2$, $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$. Similarly, we assume $\boldsymbol{u} = \mathbf{0}$ as commonly done in GPR.

By the definition of MV-GP, the collection of functions $[\boldsymbol{f}(\boldsymbol{x}_1), \ldots, \boldsymbol{f}(\boldsymbol{x}_n)]$ follow a matrix-variate Gaussian distribution

$$[\boldsymbol{f}(\boldsymbol{x}_1)^{\mathrm{T}}, \ldots, \boldsymbol{f}(\boldsymbol{x}_n)^{\mathrm{T}}]^{\mathrm{T}} \sim \mathcal{MN}(\mathbf{0}, K', \Omega),$$

where K' is the $n \times n$ covariance matrix of which the (i, j)-th element $[K']_{ij} = k'(\boldsymbol{x}_i, \boldsymbol{x}_j)$. In order to predict $\boldsymbol{f}_* = \boldsymbol{f}(Z)$ at the test locations $Z = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m]^{\mathrm{T}}$, the joint distribution of the training observations observations $Y = [\boldsymbol{y}_1^{\mathrm{T}}, \cdots, \boldsymbol{y}_n^{\mathrm{T}}]^{\mathrm{T}}$ and the predictive targets $\boldsymbol{f}_*$ are given by

$$\begin{bmatrix} Y \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{MN}\left(\mathbf{0}, \begin{bmatrix} K'(X,X) & K'(Z,X)^{\mathrm{T}} \\ K'(Z,X) & K'(Z,Z) \end{bmatrix}, \Omega\right), \tag{6.1}$$

where $K'(X, X)$ is an $n \times n$ matrix of which the (i, j)-th element $[K'(X, X)]_{ij} = k'(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $K'(Z, X)$ is an $m \times n$ matrix of which $[K'(Z, X)]_{ij} = k'(\boldsymbol{x}_{n+i}, \boldsymbol{x}_j)$, and $K'(Z, Z)$ is an $m \times m$ matrix with $[K'(Z, Z)]_{ij} = k'(\boldsymbol{x}_{n+i}, \boldsymbol{x}_{n+j})$. Thus, taking advantage of conditional distribution of MV-GP, the predictive distribution is

$$p(\boldsymbol{f}_*|X, Y, Z) = \mathcal{MN}(\hat{M}, \hat{\Sigma}, \hat{\Omega}), \tag{6.2}$$

where

$$\begin{aligned} \hat{M} &= K'(Z,X)^{\mathrm{T}} K'(X,X)^{-1} Y, & (6.3) \\ \hat{\Sigma} &= K'(Z,Z) - K'(Z,X)^{\mathrm{T}} K'(X,X)^{-1} K'(Z,X), & (6.4) \\ \hat{\Omega} &= \Omega. & (6.5) \end{aligned}$$

As a result, the expectation and the covariance are obtained

$$\begin{aligned}\mathbb{E}[\boldsymbol{f}_*] &= \hat{M} = K'(Z,X)^{\mathrm{T}}K'(X,X)^{-1}Y, \quad (6.6)\\ \mathrm{cov}(\mathrm{vec}(\boldsymbol{f}_*^{\mathrm{T}})) &= \hat{\Sigma}\otimes\hat{\Omega}\\ &= [K'(Z,Z)-K'(Z,X)^{\mathrm{T}}K'(X,X)^{-1}K'(Z,X)]\otimes\Omega. \quad (6.7)\end{aligned}$$

6.3.2 Multivariate Student−t process regression

Multivariate Student−t process regression model can be formulated along the same line as MV-GPR and it briefly presented below.

Given n pairs of observations $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n, \boldsymbol{x}_i \in \mathbb{R}^p, \boldsymbol{y}_i \in \mathbb{R}^d$, we assume

$$\begin{aligned}\boldsymbol{f} &\sim \mathcal{MTP}(\nu, \boldsymbol{u}, k', \Omega), \nu > 2,\\ \boldsymbol{y}_i &= \boldsymbol{f}(\boldsymbol{x}_i), \text{for } i = 1, \cdots, n,\end{aligned}$$

where ν is the degree of freedom of Student−t process and the remaining parameters have the same meaning of MV-GP regression model. Consequently, the predictive distribution is obtained as

$$p(\boldsymbol{f}_*|X,Y,Z) = \mathcal{MT}(\hat{\nu}, \hat{M}, \hat{\Sigma}, \hat{\Omega}), \quad (6.8)$$

where

$$\begin{aligned}\hat{\nu} &= \nu + n, \quad (6.9)\\ \hat{M} &= K'(Z,X)^{\mathrm{T}}K'(X,X)^{-1}\boldsymbol{y}, \quad (6.10)\\ \hat{\Sigma} &= K'(Z,Z)-K'(Z,X)^{\mathrm{T}}K'(X,X)^{-1}K'(Z,X), \quad (6.11)\\ \hat{\Omega} &= \Omega + Y^{\mathrm{T}}K'(X,X)^{-1}Y. \quad (6.12)\end{aligned}$$

According to the expectation and the covariance of matrix-variate Student−t distribution, the predictive mean and covariance are given by

$$\begin{aligned}\mathbb{E}[\boldsymbol{f}_*] &= \hat{M} = K'(X_*,X)^{\mathrm{T}}K'(X,X)^{-1}Y, \quad (6.13)\\ \mathrm{cov}(\mathrm{vec}(\boldsymbol{f}_*^{\mathrm{T}})) &= \frac{1}{\nu+n-2}\hat{\Sigma}\otimes\hat{\Omega}\\ &= \frac{1}{\nu+n-2}[K'(X_*,X_*)-K'(X_*,X)^{\mathrm{T}}K'(X,X)^{-1}K'(X_*,X)]\\ &\quad \otimes(\Omega+Y^{\mathrm{T}}K'(X,X)^{-1}Y). \quad (6.14)\end{aligned}$$

6.4 Covariance functions

Despite there are two covariance matrices, column covariance and row covariance, in both MV-GPR and MV-TPR, only the column covariance depended on inputs is considered as kernel since it contains our presumptions about the function we wish to learn and define the closeness and similarity between data points [59]. Of course, the choice of kernel also has a profound impact on the performance of MV-GP as well as MV-TP. Several samples over two typical kernels, Squared Exponential (SE) and Periodic (PER) are listed as follows.

The samples of MV-GP and MV-TP over SE kernel are shown in Figure 6.1, where the input x has 100 equally spaced values in [0, 5] and the row covariance matrix $\Omega = \left(\begin{smallmatrix} 1.5 & 0.5 \\ 0.5 & 2.5 \end{smallmatrix}\right)$.

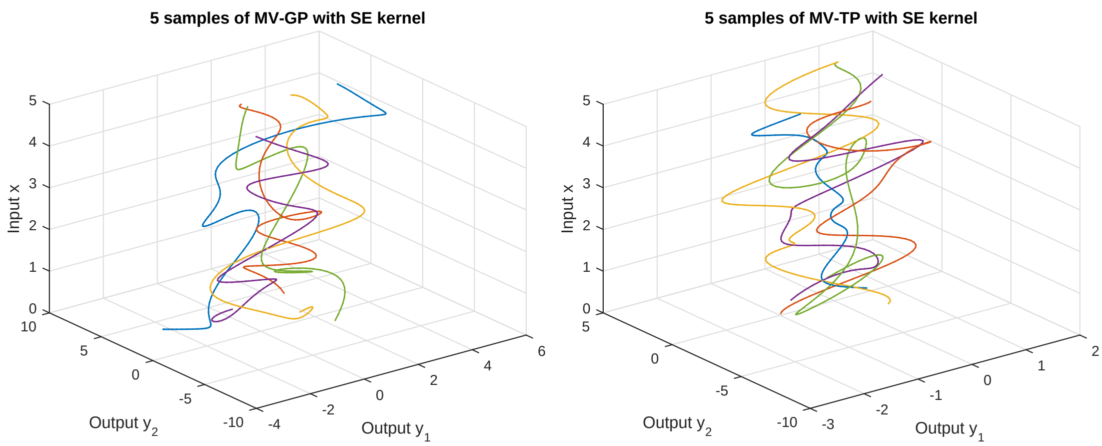


FIGURE 6.1: 5 samples of matrix-variate process over SE kernel. Left panel: MV-GP with parameter $[\ell, s_f^2] = [\log(0.5), \log(1.5)]$; Right panel: MV-TP with parameter $[\nu, \ell, s_f^2] = [3, \log(0.5), \log(1.5)]$

The samples of MV-GP and MV-TP over PER kernel are shown in Figure 6.2, where the input x has 100 equally spaced values in [0, 5] and the row covariance matrix $\Omega = \left(\begin{smallmatrix} 1.5 & 0.5 \\ 0.5 & 2.5 \end{smallmatrix}\right)$.

6.5 Parameters estimation

Similar to GPR models, the hyper-parameters involved in the kernel of MV-GPR as well as MV-TPR need to be estimated from the training data. Although Monte Carlo methods can perform GPR without the need of estimating hyper-parameters

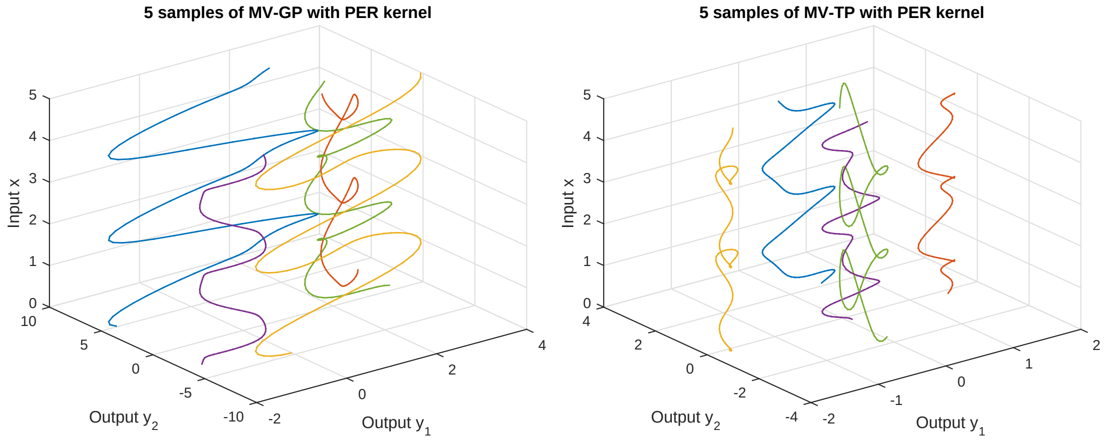


FIGURE 6.2: 5 samples of matrix-variate process over PER kernel. Left panel: MV-GP with parameter $[\ell, \mathrm{p}, s_f^2] = [\log(2), \log(2), \log(2)]$; Right panel: MV-TP with parameter $[\nu, \ell, s_f^2] = [3, \log(2), \log(2), \log(2)]$

[26, 28, 67, 64], the common approach is to estimate them by means of maximum marginal likelihood due to the high computational cost of Monte Carlo methods.

The undetermined parameters contain the hyper-parameters in the kernel, noisy level σ_n^2 and the row covariance parameter matrix Ω. Due to positive semi-definite, Ω can be denoted $\Omega = \Phi\Phi^{\mathrm{T}}$, where

$$\Phi = \begin{bmatrix} \phi_{11} & 0 & \cdots & 0 \\ \phi_{21} & \phi_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{d1} & \phi_{d2} & \cdots & \phi_{dd} \end{bmatrix}.$$

To guarantee the uniqueness of Φ, the diagonal elements are restricted to be positive and denote $\varphi_{ii} = \log(\phi_{ii})$ for $i = 1, 2, \cdots, d$.

6.5.1 Estimation of parameters in multivariate Gaussian process regression

In the MV-GPR model, the observations are followed by a matrix-variate Gaussian distribution $Y \sim \mathcal{MN}_{n,d}(\mathbf{0}, K', \Omega)$ where K' is the noisy column covariance matrix with element $[K']_{ij} = k'(x_i, x_j)$ so that $K' = K + \sigma_n^2 \mathbf{I}$ where K is noise-free column covariance matrix with element $[K]_{ij} = k(x_i, x_j)$. As we know there are hyper-parameters in the kernel k so that we can denote $K = K_\theta$. The hyper-parameter set denotes $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots\}$, thus

$$\frac{\partial K'}{\sigma_n^2} = \mathbf{I}_n, \quad \frac{\partial K'}{\partial \theta_i} = \frac{\partial K_\theta}{\partial \theta_i}.$$

According to the matrix-variate distribution, the negative log marginal likelihood of observations is

$$\mathcal{L} = \frac{nd}{2}\log(2\pi) + \frac{d}{2}\log\det(K') + \frac{n}{2}\log\det(\Omega) + \frac{1}{2}\mathrm{tr}((K')^{-1}Y\Omega^{-1}Y^{\mathrm{T}}). \quad (6.15)$$

The derivatives of the negative log marginal likelihood with respect to parameter σ_n^2, θ_i, ϕ_{ij} and φ_{ii} are as follows

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \sigma_n^2} &= \frac{d}{2}\mathrm{tr}((K')^{-1}) - \frac{1}{2}\mathrm{tr}(\alpha_{K'}\Omega^{-1}\alpha_{K'}^{\mathrm{T}}), \\
\frac{\partial \mathcal{L}}{\partial \theta_i} &= \frac{d}{2}\mathrm{tr}\left((K')^{-1}\frac{\partial K_\theta}{\partial \theta_i}\right) - \frac{1}{2}\mathrm{tr}\left(\alpha_{K'}\Omega^{-1}\alpha_{K'}^{\mathrm{T}}\frac{\partial K_\theta}{\partial \theta_i}\right), \\
\frac{\partial \mathcal{L}}{\partial \phi_{ij}} &= \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})] - \frac{1}{2}\mathrm{tr}[\alpha_\Omega(K')^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})], \\
\frac{\partial \mathcal{L}}{\partial \varphi_{ii}} &= \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})] - \frac{1}{2}\mathrm{tr}[\alpha_\Omega(K')^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})],
\end{aligned}$$

where $\alpha_{K'} = (K')^{-1}Y$, $\alpha_\Omega = \Omega^{-1}Y^{\mathrm{T}}$, $\mathbf{E}_{ij}$ is the $d \times d$ elementary matrix having unity in the (i,j)-th element and zeros elsewhere, and $\mathbf{J}_{ii}$ is the same as $\mathbf{E}_{ij}$ but with the unity being replaced by $e^{\varphi_{ii}}$. The details can be found in C.1.

6.5.2 Estimation of parameters in multivariate Student $-t$ process regression

In the MV-TPR model, the observations are followed by a matrix-variate Student$-t$ distribution $Y \sim \mathcal{MT}_{n,d}(\nu, \mathbf{0}, K', \Omega)$. The negative log marginal likelihood is

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2}(\nu + d + n - 1)\log\det(\mathbf{I}_n + (K')^{-1}Y\Omega^{-1}Y^{\mathrm{T}}) + \frac{d}{2}\log\det(K') + \frac{n}{2}\log\det(\Omega) \\
&\quad + \log\Gamma_n\left(\frac{1}{2}(\nu + n - 1)\right) - \log\Gamma_n\left(\frac{1}{2}(\nu + d + n - 1)\right) + \frac{1}{2}dn\log\pi \\
&= \frac{1}{2}(\nu + d + n - 1)\log\det(K' + Y\Omega^{-1}Y^{\mathrm{T}}) - \frac{\nu + n - 1}{2}\log\det(K') + \frac{n}{2}\log\det(\Omega) \\
&\quad + \log\Gamma_n\left(\frac{1}{2}(\nu + n - 1)\right) - \log\Gamma_n\left(\frac{1}{2}(\nu + d + n - 1)\right) + \frac{1}{2}dn\log\pi.
\end{aligned}$$

Therefore the parameters of MV-TPR contains the all the parameters in MV-GPR and one more parameter: the degree of freedom ν. The derivatives of the negative log marginal likelihood with respect to parameter ν,σ_n^2, θ_i, ϕ_{ij} and φ_{ii} are as follows

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \nu} &= \frac{1}{2}\log\det(U) - \frac{1}{2}\log\det(K') + \frac{1}{2}\psi_n(\frac{1}{2}\tau) - \frac{1}{2}\psi_n\left(\frac{1}{2}(\tau+d)\right),\\
\frac{\partial \mathcal{L}}{\partial \sigma_n^2} &= \frac{(\tau+d)}{2}\mathrm{tr}(U^{-1}) - \frac{\tau}{2}\mathrm{tr}((K')^{-1}),\\
\frac{\partial \mathcal{L}}{\partial \theta_i} &= \frac{(\tau+d)}{2}\mathrm{tr}\left(U^{-1}\frac{\partial K_\theta}{\partial \theta_i}\right) - \frac{\tau}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial K_\theta}{\partial \theta_i}\right),\\
\frac{\partial \mathcal{L}}{\partial \phi_{ij}} &= -\frac{(\tau+d)}{2}\mathrm{tr}[U^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})\alpha_\Omega] + \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})],\\
\frac{\partial \mathcal{L}}{\partial \varphi_{ii}} &= -\frac{(\tau+d)}{2}\mathrm{tr}[U^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})\alpha_\Omega] + \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})],
\end{aligned}
$$

where $U = K' + Y\Omega^{-1}Y^{\mathrm{T}}$, $\tau = \nu + n - 1$ and $\psi_n(\cdot)$ is the derivative of the function $\log\Gamma_n(\cdot)$ with respect to ν. The details can be found in C.2.

In fact, the sensitivity of initial hyper-parameter also exists because the negative log marginal likelihood is not convex and there may exists local optima, like conventional GPR and TPR. The same heuristic methods resolving the problems in GPR and TPR, is adopted for MV-GPR and MV-TPR.

6.6 Experiments and applications

In this section, we demonstrate the usefulness of the matrix-variate process regression models using some numerical examples, including simulated data and real data.

6.6.1 Simulated example

We first consider a simulated data from two specific functions. The true model used to generate data is given by,

$$
\begin{aligned}
\boldsymbol{y} &= [f_1(x), f_2(x)] + [\varepsilon^{(1)}, \varepsilon^{(2)}],\\
f_1(x) &= 2x\cdot\cos(x), \quad f_2(x) = 1.5x\cdot\cos(x+\pi/5),
\end{aligned}
$$

where the vector noise produced from a sample of multivariate Gaussian process $[\varepsilon^{(1)}, \varepsilon^{(2)}] \sim \mathcal{MGP}(0, k_{SE}, \Omega)$. We select k_{SE} with $[\ell, s_f^2] = [\log(1.001), \log(5)]$ and $\Omega = \left(\begin{smallmatrix} 1 & 0.25 \\ 0.25 & 1 \end{smallmatrix}\right)$. The covariate x has 100 equally spaced values in [-10, 10] so that a sample of 100 observations for y_1 and y_2 are obtained.

For model training, we try to use less points with one part missing so that the zth data points where $z = \{3r+1\}_{r=1}^{12} \cup \{3r+2\}_{r=22}^{32}$ are selected for both y_1 and y_2. The prediction is performed at all 100 covariate values equally spaced in [-10, 10]. The RMSEs between the predicted values and the true ones from $f_1(x)$ and $f_2(x)$ are calculated. At the same time, the conventional GPR and TPR models are conducted for the two outputs independently and the ARMSEs are compared with the proposed models. The process above is repeated 1000 times and the results are reported in Table 6.1 and an example of prediction is given in Figure 6.3. The ARMSE (Average Root Mean Square Error) for 100 points predictions repeated 1000 times is defined by

$$\text{ARMSE} = \frac{1}{1000} \sum_{i=1}^{1000} \left(\frac{1}{100} \sum_{j=1}^{100} (\hat{y}_{ij} - y_{ij})^2 \right)^{\frac{1}{2}},$$

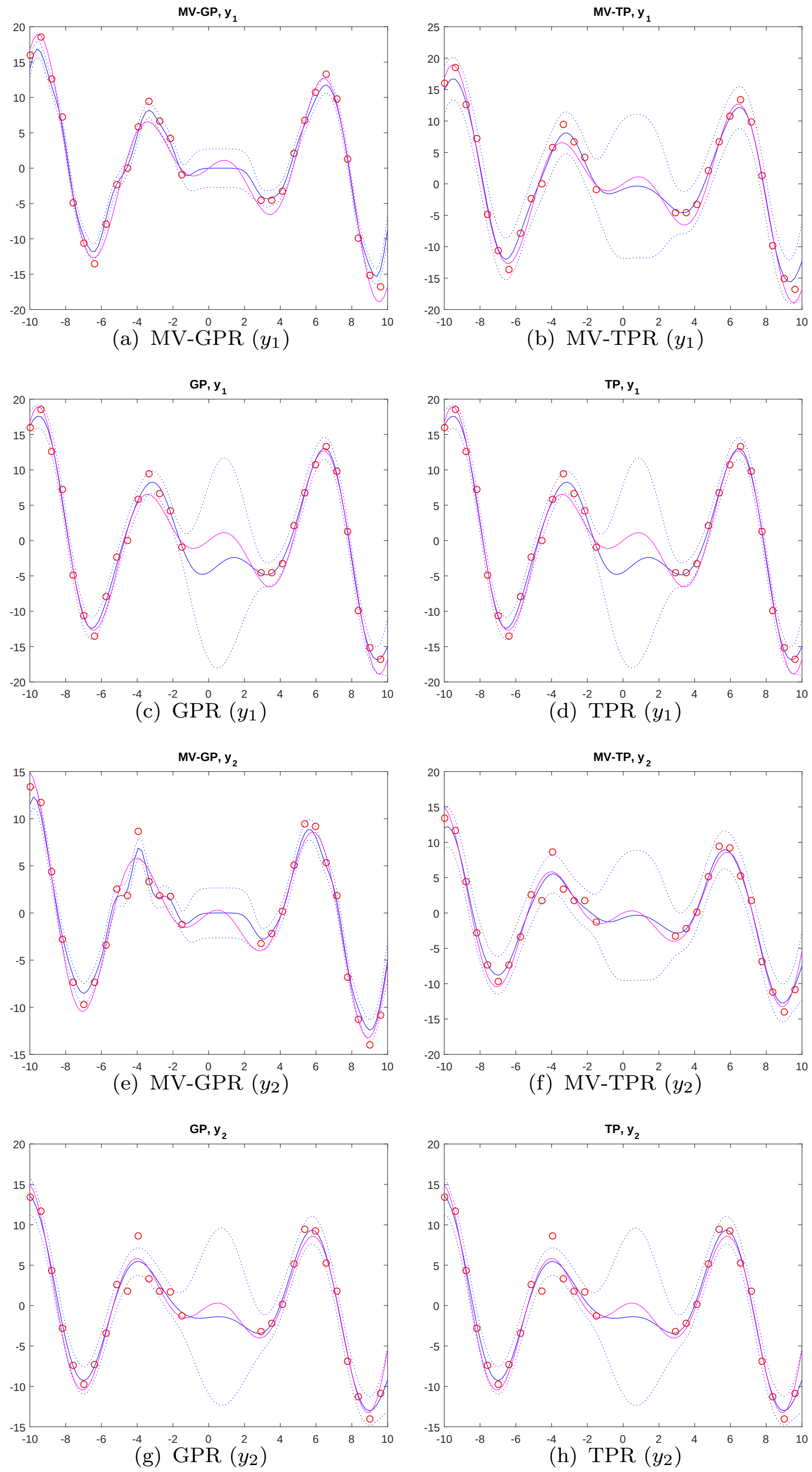
where y_{ij} is the jth observation in ith experiment while $\hat{y}_{ij}$ is the jth prediction in ith experiment.

TABLE 6.1: The ARMSE by the different models (multivariate Gaussian noisy data)

Output 1 (y_1)				Output 2 (y_2)			
MV-GPR	GPR	MV-TPR	TPR	MV-GPR	GPR	MV-TPR	TPR
1.540	1.594	**1.258**	1.585	1.749	2.018	**1.518**	2.017

The same experiment is conducted for the case where the vector noise is a sample from multivariate Student−t process $[\varepsilon^{(1)}, \varepsilon^{(2)}] \sim \mathcal{MTP}(3, 0, k_{SE}, \Omega)$. We select k_{SE} with parameter $[\ell, s_f^2] = [\log(1.001), \log(5)]$ and $\Omega = \left(\begin{smallmatrix} 1 & 0.25 \\ 0.25 & 1 \end{smallmatrix}\right)$. The result of ARMSEs are presented in Table 6.2 and an example of prediction is demonstrated in Figure 6.4.

According to the tables and figures above, it can be seen that the multivariate process regression models are able to discover a more desirable pattern in the gap than using the conventional GPR and TPR model independently. It also reveals that taking correlations between the two outputs into consideration improves the accuracy of prediction compared with the methods of modelling each



(a) MV-GPR (y_1) (b) MV-TPR (y_1)
(c) GPR (y_1) (d) TPR (y_1)
(e) MV-GPR (y_2) (f) MV-TPR (y_2)
(g) GPR (y_2) (h) TPR (y_2)

FIGURE 6.3: Predictions for MV-GP noise data using different models. From panels (a) to (d): predictions for y_1 by MV-GPR, MV-TPR, GPR and TPR. From panels (e) to (h): predictions for y_2 by MV-GPR, MV-TPR, GPR and TPR. The solid blue lines are predictions, the solid red lines are the true functions and the circles are the observations. The dash lines represent the 95% confidence intervals

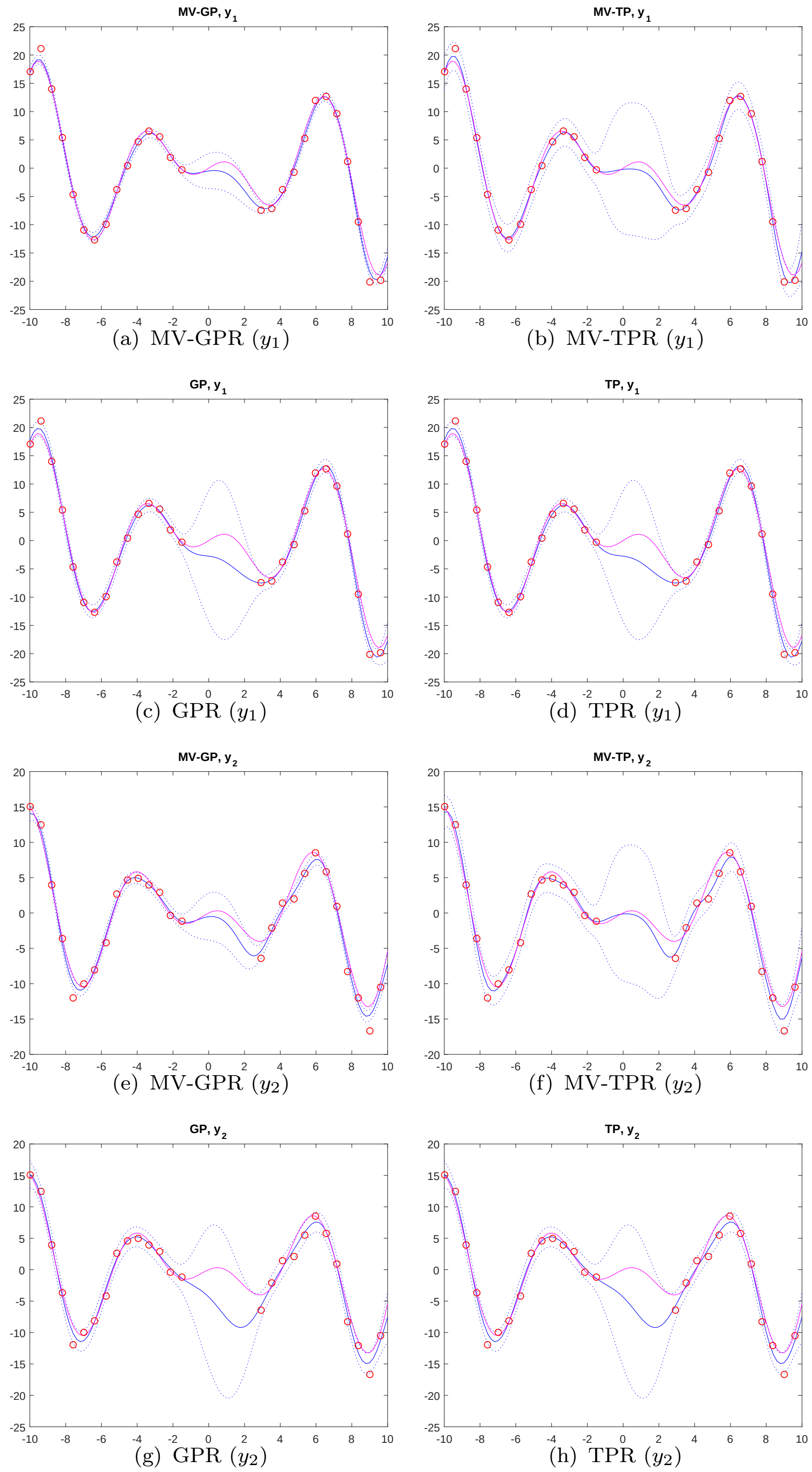


FIGURE 6.4: Predictions for MV-TP noise data using different models. From panels (a) to (d): predictions for y_1 by MV-GPR, MV-TPR, GPR and TPR. From panels (e) to (h): predictions for y_2 by MV-GPR, MV-TPR, GPR and TPR. The solid blue lines are predictions, the solid red lines are the true functions and the circles are the observations. The dash lines represent the 95% confidence intervals

Table 6.2: The ARMSE by the different models (multivariate Student−t noisy data)

Output 1 (y_1)				Output 2 (y_2)			
MV-GPR	GPR	MV-TPR	TPR	MV-GPR	GPR	MV-TPR	TPR
1.441	1.505	**1.238**	1.503	1.636	1.941	**1.464**	1.940

output independently. In particular, MV-TPR performs better than MV-GPR in the predictions of both types of noisy data. This may be explained by the fact that MV-TPR has a better modelling flexibility with one more parameter which can capture the degree of freedom of the data and take the correlations between two responses into account. The reason will be further studied in the next real data experiments.

It is notable that the predictive variance of MV-GPR is much smaller than the independent GPR model. This is likely caused by the loss of information in the independent model. As discussed in [81], the prediction uncertainty of GPR is useful in building the predicting model by ensemble learning.

6.6.2 Real Data Examples

We further test our proposed methods on two real datasets [1]. The selected mean function is zero-offset and the selected kernel is SEard. Before the experiments, all the data have been done zero-normalisation.

6.6.2.1 Air quality prediction

The dataset contains 9358 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device with 15 attributes [82]. We delete all the points with missing attributes (887 points remaining). The first 864 points are considered in our experiment because the data is hourly observed (1 day = 24 hours) and the whole data set is divided into 9 subsets (each subset has 4-days' data points, totally 864 data points). In the experiment, there are 9 attributes' input, including time, true hourly averaged concentration CO in mg/m^3 (COGT), true hourly averaged overall Non Metanic HydroCarbons concentration in $microg/m^3$ (NMHCGT), true

[1]These data sets are from the UC Irvine Machine Learning Repository: `https://archive.ics.uci.edu/ml/index.php`.

hourly averaged Benzene concentration in $microg/m^3$ (C6H6GT), true hourly averaged NOx concentration in ppb (NOx), true hourly averaged NO2 concentration in $microg/m^3$ (NO2), absolute humidity (AH), temperature (T) and relative humidity (RH). The output consists of 5 attributes, including PT08.S1 (tin oxide) hourly averaged sensor response, PT08.S2 (titania) hourly averaged sensor response, PT08.S3 (tungsten oxide) hourly averaged sensor response, PT08.S4 (tungsten oxide) hourly averaged sensor response and PT08.S5 (indium oxide) hourly averaged sensor response.

The cross-validation method is taken as $k-$fold, where $k = 9$. Each subset is considered as test set and the remaining subsets are considered as training set. Four models, containing MV-GPR, MV-TPR, GPR (predict each output independently) and TPR (predict each output independently) are applied to make multi-output prediction based on the divided training and test sets. The process is repeated for 9 times.

For each subset's prediction, MSE (mean square error) and MAE (mean absolute error) are calculated and the median of the 9 MSEs and MAEs is used to evaluate each output. Finally, the maximum median of all the outputs (MMO) is used to evaluate the multi-dimensional prediction. The results are shown in Table 6.3 and we can verify empirically that MV- TPR performs the best since it has smallest maximum error in terms of MSE and MAE.

Table 6.3: Air quality prediction results based on MSEs and MAEs

(a) MSE

		MV-GPR	MV-TPR	GPR	TPR
Outputs (Median of 9 subsets' MSEs)	PT08S1CO	0.091	0.065	0.079	0.074
	PT08S2NMHC	8.16×10^{-5}	3.42×10^{-5}	1.91×10^{-7}	7.32×10^{-8}
	PT08S3NOx	0.036	0.027	0.022	0.025
	PT08S4NO2	0.015	0.014	0.010	0.009
	PT08S5O3	0.092	0.073	0.060	0.067
MMO		0.092	**0.073**	0.079	0.074

(b) MAE

		MV-GPR	MV-TPR	GPR	TPR
Outputs (Median of 9 subsets' MAEs)	PT08S1CO	0.240	0.204	0.212	0.223
	PT08S2NMHC	6.39×10^{-3}	1.15×10^{-2}	1.80×10^{-4}	9.26×10^{-5}
	PT08S3NOx	0.141	0.122	0.115	0.120
	PT08S4NO2	0.095	0.089	0.079	0.073
	PT08S5O3	0.231	0.210	0.199	0.205
MMO		0.240	**0.210**	0.212	0.223

6.6.2.2 Bike rent prediction

This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information [83]. There are 16 attributes. We test our proposed methods for multi-output prediction based on daily count dataset. After deleting all the points with missing attributes, we use the first 168 data points in the season Autumn because the data is daily observed (1 week = 7 days) and the whole dataset is divided into 8 subsets (each subset has 3 weeks' data points). In the experiment, there are 8 attributes' input, including normalized temperature, normalized feeling temperature, normalized humidity, normalized wind speed, whether day is holiday or not, day of the week, working day or not and weathersit. The output consists of 2 attributes, including the count of casual users (Casual) and the count of registered users (Registered).

The cross-validation method is taken as $k-$fold, where $k = 8$. All the remaining steps are the same as the air quality prediction experiment, except k is 8 so that the process is repeated 8 times.

Table 6.4: Bike rent prediction results based on MSEs and MAEs

(a) MSE

		MV-GPR	MV-TPR	GPR	TPR
Outputs (median of 8 subsets' MSEs)	Casual	0.411	**0.334**	0.424	0.397
	Registered	0.982	**0.903**	1.134	1.111
MMO		0.982	**0.903**	1.134	1.111

(b) MAE

		MV-GPR	MV-TPR	GPR	TPR
Outputs (median of 8 subsets' MAEs)	Casual	0.558	**0.488**	0.540	0.546
	Registered	0.897	**0.855**	0.916	0.907
MMO		0.897	**0.855**	0.916	0.907

The results are shown in Table 6.4 and we can also verify empirically that MV-TPR performs the best in terms of MSE and MAE.

6.6.3 Application to stock market investment

In the previous subsections, the examples show the usefulness of our proposed methods in terms of more accurate prediction. Furthermore, our proposed methods can be applied to produce trading strategies in the stock market investment.

It is known that the accurate prediction of future for an equity market is almost impossible. Admittedly, the more realistic idea is to make a strategy based on the Buy&Sell signal in the different prediction models [45]. In this section, we consider a developed Dollar 100 (dD100) as a criterion of the prediction models. The dD100 criterion is able to reflect the theoretical future value of \$100 invested at the beginning, and traded according to the signals constructed by predicted value and the reality. The details of dD100 criterion are described in Section 6.6.3.2.

Furthermore, the equity index is an important measurement of the value of a stock market and is used by many investors making trades and scholars studying stock markets. The index is computed from the weighted average of the selected stocks' prices, thus it is able to describe how the whole stock market in the consideration performs in a period and thus many trading strategies of a stock or a portfolio have to take the information of the index into account. As a result, our experimental predictions for specific stocks are based on the indices as well.

6.6.3.1 Data preparation

We collect daily price data, containing opening, closing, and adjusted closing for the stocks (the details are shown in Section 6.6.3.3 and Section 6.6.3.4) and three main indices in the US, Dow Jones Industrial Average (INDU), S&P500 (SPX), and NASDAQ(NDX) from Yahoo Finance in the period of 2013 – 2014. The log returns of adjusted closing price and inter-day log returns are consequently achieved by definitions

$$\text{Log return: } LR_i = \log \frac{ACP_i}{ACP_{i-1}},$$
$$\text{Inter-day log return: } ILR_i = \log \frac{CP_i}{OP_i},$$

where ACP_i is the adjusted closing price of ith day ($i > 1$), CP_i is the closing price of ith day, and OP_i is the openning price of ith day. Therefore, there are totally 503 daily log returns and log inter-day returns for all the stocks and indices from 2013 to 2014.

6.6.3.2 Prediction model and strategy

The sliding windows method is used throughout our prediction models, including GPR, TPR, MV-GPR, and MV-TPR, based on the indices, INDU, SPX, and NDX. The size of training sample is 303, which is used to forecast the next 10 days, and the training set is updated by dropping off the earliest 10 days and added on the latest 10 days. The sliding-forward process ran 20 times, resulting in a total 200 prediction days, in groups of 10. The updated training set allows all the models and parameters to adapt the dynamic structure of the equity market [45]. Specifically, the inputs consist of the log returns of 3 indices. The targets are multiple stocks' log returns. Due to the multi-dimensional inputs, Standard Exponential with automatic relevance determination (SEard) is used kernel for all of these prediction models.

Admittedly, we forecast the log returns of the specifically grouped stocks based on the 3 main indices in the US. It is noteworthy that the predicted log returns of stocks are used to produce a buy or sell signal for trading rather than to discover an exact pattern in the future. The signal BS produced by the predicted log returns of the stocks is defined by

$$BS_i = \hat{LR_i} - LR_i + ILR_i, i = 1, \cdots, 200,$$

where $\{\hat{LR_i}\}_{i=1}^{200}$ are the predicted log returns of a specific stock, $\{LR_i\}_{i=1}^{200}$ are the true log returns while $\{ILR_i\}_{i=1}^{200}$ are the inter-day log returns. The Buy&Sell strategy relying on the signal BS is described in Table 6.5.

TABLE 6.5: Buy&Sell strategy of dD100 investment

Decision	Condition
Buy	$\hat{LR_i} > 0$, & $BS_i > 0$ & we have the position of cash
Sell	$\hat{LR_i} < 0$, & $BS_i < 0$ & we have the position of share
Keep	No action is taken for the rest of the option

It has to mention that the stock in our experiment is counted in Dollar rather than the number of shares, which means we can precisely buy or sell a specific Dollar valued stock theoretically. For example, if the stock price is \$37 when we only have \$20, we can still buy \$20 valued stock theoretically rather than borrow \$17 and then buy 1 share. Furthermore, it is also necessary to explain why we choose the signal BS. By definition of signal, we rewrite it as

$$\begin{aligned} BS_i &= \log(\frac{A\hat{C}P_i}{ACP_{i-1}}) - \log(\frac{ACP_i}{ACP_{i-1}}) + \log(\frac{CP_i}{OP_i}) \\ &= \log(\frac{A\hat{C}P_i}{ACP_{i-1}}) - \log(\frac{ACP_i}{ACP_{i-1}}) + \log(\frac{ACP_i}{AOP_i}) \\ &= \log(\frac{A\hat{C}P_i}{AOP_i}), \end{aligned}$$

where $\{ACP_i\}_{i=0}^{200}$ ar the last 201 adjusted closing prices for a stock, $\{CP\}_{i=1}^{200}$ are the last 200 closing prices, and $\{AOP\}_{i=1}^{200}$ are the adjusted opening prices. If $BS_i > 0$, the predicted closing price is higher than adjusted opening price, which means we can obtain the inter-day profit by buying the shares at the opening price [2] as long as the signal based on our predictions is no problem. Meanwhile, the opposite performance based on BS strategy means that we can avoid the inter-day loss by selling decisively at the opening price. Furthermore, the reasonable transaction fee 0.025% is considered since the strategy might trade frequently [3]. As a result, this is a reasonable strategy since we can definitely obtain a profit by buying the shares and cut the loss by selling the shares in time only if our prediction has no serious problem. It is also an feasible strategy because the decision is given by the next day's reality and our prediction models.

At last, BS signal varies in different prediction models so that we denote these Buy&Sell strategies based on MV-GPR, MV-TPR, GPR, and TPR model as MV-GPR strategy, MV-TPR strategy, GPR strategy, and TPR strategy, respectively.

6.6.3.3 Chinese companies in NASDAQ

In recent years, the "Chinese concepts stock" has received extensive attention among international investors owing to the fast development of Chinese economy and an increasing number of the Chinese firms have been traded in the international stock markets [84]. The "Chinese concepts stock" refers to the stock issued

[2] Actually, the value has to be considered as adjusted opening price since all the shares counted as Dollar. The adjusted opening price is also easily to compute based on the real opening price and the dividend information

[3] The figure 0.025% is comprehensive consideration referred to the NASDAQ website:`http://nasdaq.cchwallstreet.com/`

by firms whose asset or earning have essential activities in Mainland China. Undoubtedly, all these "Chinese concept stocks" are heavily influenced by the political and economic environment of China together. For this reason, all these stocks have the potential and unneglectable correlation theoretically, which is probably reflected in the movement of stock price. The performance of multiple targets prediction, which takes the potential relationship into consideration, should be better. Therefore, the first real data example is based on three biggest Chinese companies described in Table 6.6.

TABLE 6.6: Three biggest "Chinese concept" stocks

Ticker	Exchange	Company
BIDU	NASDAQ	Baidu, Inc.
CTRP	NASDAQ	Ctrip.com International, Ltd.
NTES	NASDAQ	NetEase, Inc.

We apply MV-GPR, MV-TPR, GPR and TPR models to Buy&Sell strategies, with the results demonstrated in Figure 6.5. Furthermore, Table D.1, D.2 and D.3 summarize the results by the period for each stock respectively. In particular, the Buy&Sell signal examples for each stock are shown in Table D.4, D.5 and D.6 respectively, along with other relevant details.

According to Figure 6.5, there is no doubt that a \$100 investment for each stock has sharply increased over 200 days period using Buy&Sell strategies regardless of the stock trend price during this period. In particular, the stock price of BIDU and NTES rose gradually while CTRP hit the peak and then decreased on a large scale. Anyway, the Buy&Sell strategies based on different prediction models have still achieved more considerable profits than the Buy&Hold strategies for the corresponding stock investment. However, the different prediction models have diverse performances for each stock. For BIDU, GPR-based models, including MV-GPR and GPR, outperform TPR-based models, including MV-TPR and TPR. For NTES, all the models for Buy&Sell strategy have the similar performance. Admittedly, TPR-based models, especially MV-TPR, have an outstanding performance for stock CTRP. The results indicate that multivariate process regression models, including MV-GPR and MV-TPR, have a better prediction in some particular cases. To explore further, the degree of freedom of Student−t distribution for each stock is estimated by MATLAB build-in function 'fitdist' with parameter 'tLocationScale' and the computation results are 5.87, 2.99 and 3.67 for BIDU, CTRP, and NTES, respectively. As we know, the degree of freedom of Student−t

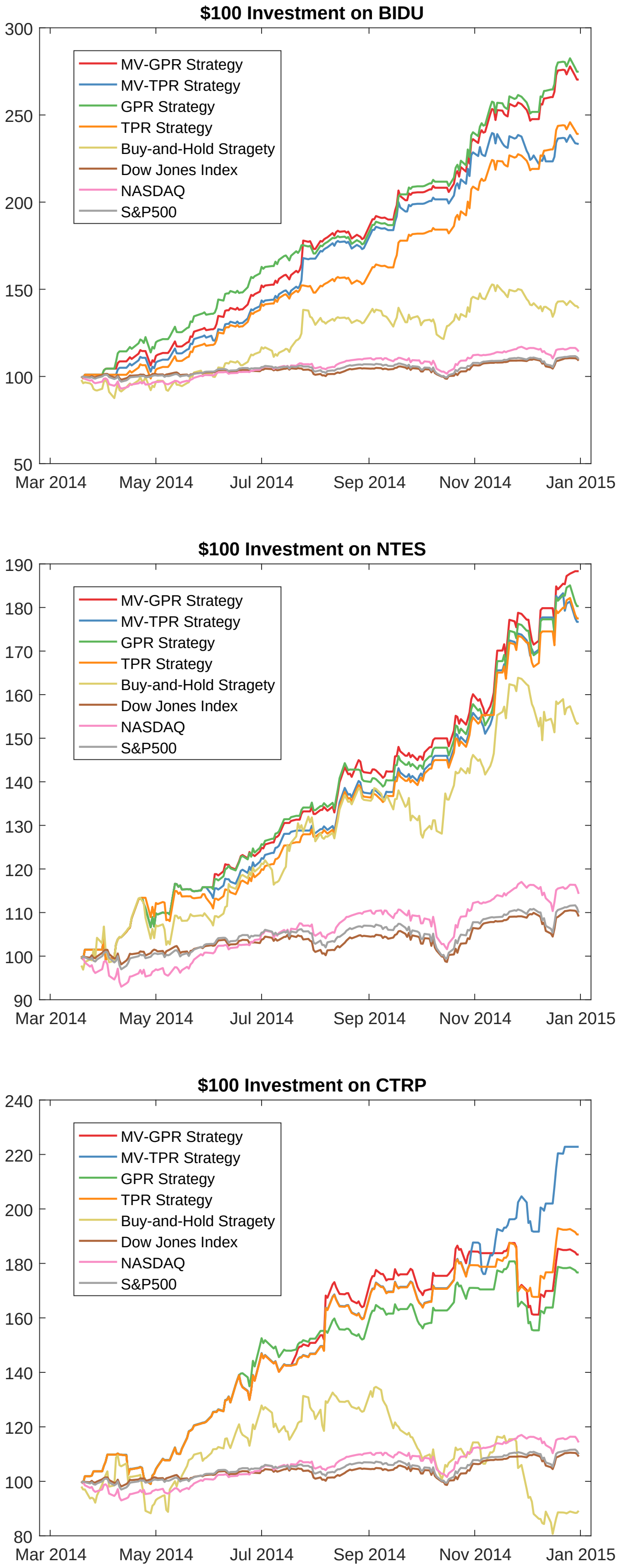


FIGURE 6.5: The movement of invested $100 in 200 days for 3 Chinese stocks in the US market. The top 4 lines in legend are Buy&Sell strategies based on 4 prediction models, MV-GPR, MV-TPR, GPR, and TPR, respectively. The last 4 lines are Buy&Hold strategies for the stock and 3 indices, INDU, NASDAQ, and NDX, respectively

distribution controls the fatness of tails, and thus it shows that the data of CTRP is the heaviest tailed while BIDU is the lightest tailed. This may be the reason why TPR-based model have an outstanding performance for CTRP while GPR-based model outperforms for BIDU. Nevertheless, the more accurate inference remains to be studied.

6.6.3.4 Diverse sectors in Dow 30

Owing to the globalisation of capital, there has been a significant shift in the relative importance of national and economic influences in the world's largest equity markets and the impact of the industrial sector is now gradually replacing that of country effects in these markets [85]. Therefore, a further example is carried out under the diverse industrial sectors in Dow 30 from New York Stock Exchange (NYSE) and NASDAQ.

Initially, the classification of stocks based on diverse industrial sectors in Dow 30 has to be done. There are two main industry classification taxonomies, including Industry Classification Benchmark (ICB) and Global Industry Classification Standard (GICS). In our research, ICB is used to segregate markets into sectors within the macro economy. The stocks in Dow 30 are classified in Table 6.7. Due to the multivariate process models considering at least two related stocks in one group, the first (Basic Materials), as well as last industrial sector (Telecommunications), consisting of only one stock, are excluded. Our experiments are performed 7 times from 7 grouped industrial sector stocks, including Oil&Gas, Industrial, Consumer Goods, Health Care, Consumer Services, Financials and Technology, respectively.

Secondly, the four models, MV-GPR, MV-TPR, GPR and TPR, are similarly applied and the stock investment ranking is listed in Table 6.8 (The detailed results are summarised in Table E.1. All the figures of $100 stock investment on the diverse industrial sectors are listed in Appendix E.2 and the tables are shown in Appendix E.3 summarise the results by the period for each stock under the diverse industrial sectors, respectively). On the whole, for each stock, there is no doubt that using the Buy&Sell strategy is much better than using the Buy&Hold strategy regardless of the industrial sector. Specifically, MV-GPR makes a satisfactory performance overall in the industrial sector Industrials, Consumer Services and Financials while MV-TPR has a higher ranking in Health Care in general.

[4]Note that the terms "industry" and "sector" are reversed from the Global Industry Classification Standard (GICS) taxonomy.

TABLE 6.7: Stock components of Dow 30

Ticker	Company	Exchange	Industry	Industry [4](ICB)
DD	DuPont	NYSE	Chemical industry	Basic Materials
KO	Coca-Cola	NYSE	Beverages	Consumer Goods
PG	Procter & Gamble	NYSE	Consumer goods	Consumer Goods
MCD	McDonald's	NYSE	Fast food	Consumer Goods
NKE	Nike	NYSE	Apparel	Consumer Services
DIS	Walt Disney	NYSE	Broadcasting and entertainment	Consumer Services
HD	The Home Depot	NYSE	Home improvement retailer	Consumer Services
WMT	Wal-Mart	NYSE	Retail	Consumer Services
JPM	JPMorgan Chase	NYSE	Banking	Financials
GS	Goldman Sachs	NYSE	Banking, Financial services	Financials
V	Visa	NYSE	Consumer banking	Financials
AXP	American Express	NYSE	Consumer finance	Financials
TRV	Travelers	NYSE	Insurance	Financials
UNH	UnitedHealth Group	NYSE	Managed health care	Health Care
JNJ	Johnson & Johnson	NYSE	Pharmaceuticals	Health Care
MRK	Merck	NYSE	Pharmaceuticals	Health Care
PFE	Pfizer	NYSE	Pharmaceuticals	Health Care
BA	Boeing	NYSE	Aerospace and defense	Industrials
MMM	3M	NYSE	Conglomerate	Industrials
GE	General Electric	NYSE	Conglomerate	Industrials
UTX	United Technologies	NYSE	Conglomerate	Industrials
CAT	Caterpillar	NYSE	Construction and mining equipment	Industrials
CVX	Chevron	NYSE	Oil & gas	Oil & Gas
XOM	ExxonMobil	NYSE	Oil & gas	Oil & Gas
CSCO	Cisco Systems	NASDAQ	Computer networking	Technology
IBM	IBM	NYSE	Computers and technology	Technology
AAPL	Apple	NASDAQ	Consumer electronics	Technology
INTC	Intel	NASDAQ	Semiconductors	Technology
MSFT	Microsoft	NASDAQ	Software	Technology
VZ	Verizon	NYSE	Telecommunication	Telecommunications

TABLE 6.8: Stock investment ranking under different strategies

Ticker	Industry	Buy&Sell Strategy				Buy&Hold Stragegy			
		MV-GPR	MV-TPR	GPR	TPR	Stock	INDU	NDX	SPX
CVX	Oil & Gas	3rd	4th	2nd	***1st***	8th	7th	5th	6th
XOM	Oil & Gas	4th	2nd	3rd	***1st***	8th	7th	5th	6th
MMM	Industrials	2nd	3rd	***1st***	4th	5th	8th	6th	7th
BA	Industrials	***1st***	2nd	3rd	4th	8th	7th	5th	6th
CAT	Industrials	3rd	4th	2nd	***1st***	8th	7th	5th	6th
GE	Industrials	2nd	4th	3rd	***1st***	8th	7th	5th	6th
UTX	Industrials	2nd	4th	3rd	***1st***	8th	7th	5th	6th
KO	Consumer Goods	2nd	***1st***	3rd	4th	6th	8th	5th	7th
MCD	Consumer Goods	2nd	4th	***1st***	3rd	8th	7th	5th	6th
PG	Consumer Goods	3rd	4th	***1st***	2nd	5th	8th	6th	7th
JNJ	Health Care	3rd	2nd	***1st***	4th	6th	8th	5th	7th
MRK	Health Care	3rd	2nd	4th	***1st***	8th	7th	5th	6th
PFE	Health Care	4th	***1st***	3rd	2nd	8th	7th	5th	6th
UNH	Health Care	2nd	3rd	***1st***	4th	5th	8th	6th	7th
HD	Consumer Services	***1st***	4th	3rd	2nd	5th	8th	6th	7th
NKE	Consumer Services	2nd	3rd	4th	***1st***	5th	8th	6th	7th
WMT	Consumer Services	***1st***	4th	3rd	2nd	5th	8th	6th	7th
DIS	Consumer Services	3rd	2nd	***1st***	4th	5th	8th	6th	7th
AXP	Financials	2nd	4th	***1st***	3rd	8th	7th	5th	6th
GS	Financials	2nd	***1st***	3rd	4th	5th	8th	6th	7th
JPM	Financials	2nd	4th	***1st***	3rd	6th	8th	5th	7th
TRV	Financials	2nd	3rd	***1st***	4th	5th	8th	6th	7th
V	Financials	***1st***	4th	3rd	2nd	5th	8th	6th	7th
AAPL	Technology	4th	2nd	3rd	***1st***	5th	8th	6th	7th
CSCO	Technology	2nd	***1st***	3rd	4th	5th	8th	6th	7th
IBM	Technology	4th	***1st***	2nd	3rd	8th	7th	5th	6th
INTC	Technology	3rd	4th	2nd	***1st***	5th	8th	6th	7th
MSFT	Technology	2nd	4th	***1st***	3rd	5th	8th	6th	7th

As we discussed before, the degree of freedom of Student−t distribution describes how heavy the tails are and thus the estimated degree of freedom is probably one of the determinative factors for the difference of 4 Buy&Sell strategies. The estimated degrees of freedom for the given stocks under the 7 industries are shown in Figure 6.6, where the degree of freedom of Student−t distribution for each stock' log returns is also computed by MATLAB build-in function 'fitdist' with parameter 'tLocationScale'.

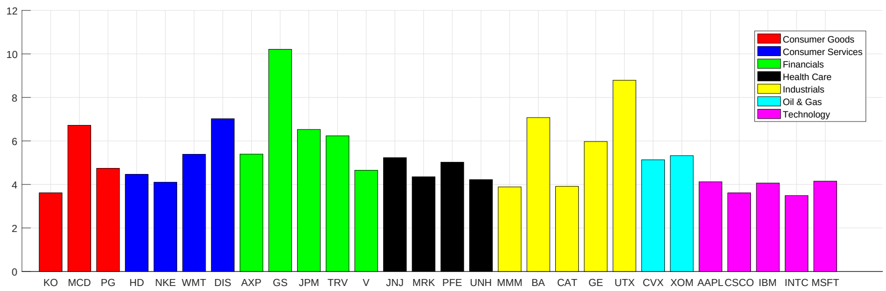
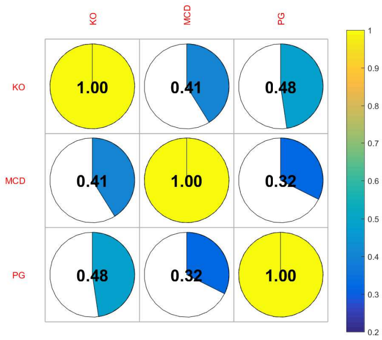


FIGURE 6.6: The estimated degrees of freedom for the given stocks under the 7 industries

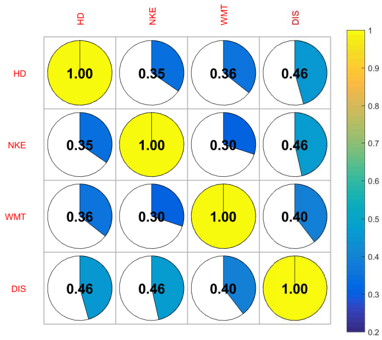
Additionally, the correlations among the multiple outputs, as the main supplementary information, are taken into account in our multivariate process regression models and hence the correlations with other stocks in the same industries are deserved to be another important factor. The correlation matrices of the 7 industries are visualised in Figure 6.7.

To further study the influence of the estimated degrees of freedom and correlations with other stocks in the same industry on the performance of \$100 investment, the stock distribution under the 7 industries based on the estimated degree of freedom and average correlation with other stocks in the same industry, is demonstrated in Figure 6.8 and Figure 6.9.

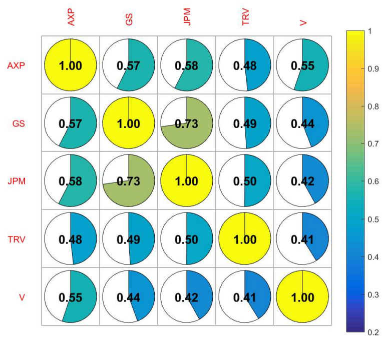
Then a colour distinct classification of 28 stocks for the different Buy&Sell strategies is in the scatter plot, Figure 6.10, where MV-based strategies contain MV-GPR and TPR while Ind-based strategies contain GPR and TPR, and GPR-based strategies contain MV-GPR and GPR while TPR-based strategies contain MV-TPR and TPR. The coloured point of stock means that the coloured strategy (or strategy group) for this stock performs better than the others. For example, the point at the top of Figure 6.10(a) is blue (blue stands for MV-based strategies here), which means that MV-GPR outperforms GPR and MV-TPR outperforms TPR.



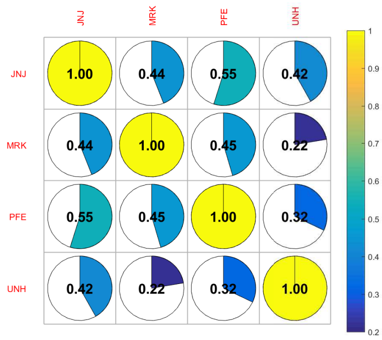
(a) Consumer Goods



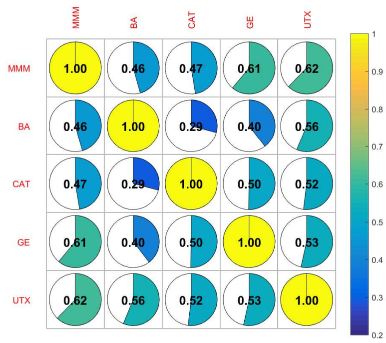
(b) Consumer Services



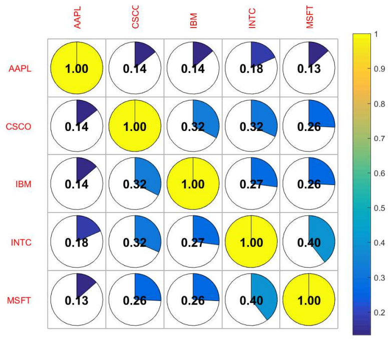
(c) Financials



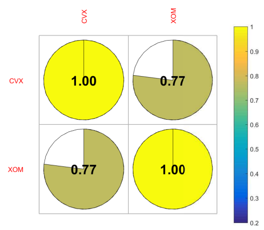
(d) Health Care



(e) Industrials



(f) Technology



(g) Oil & Gas

Figure 6.7: Correlation matrix plots of the 7 industries

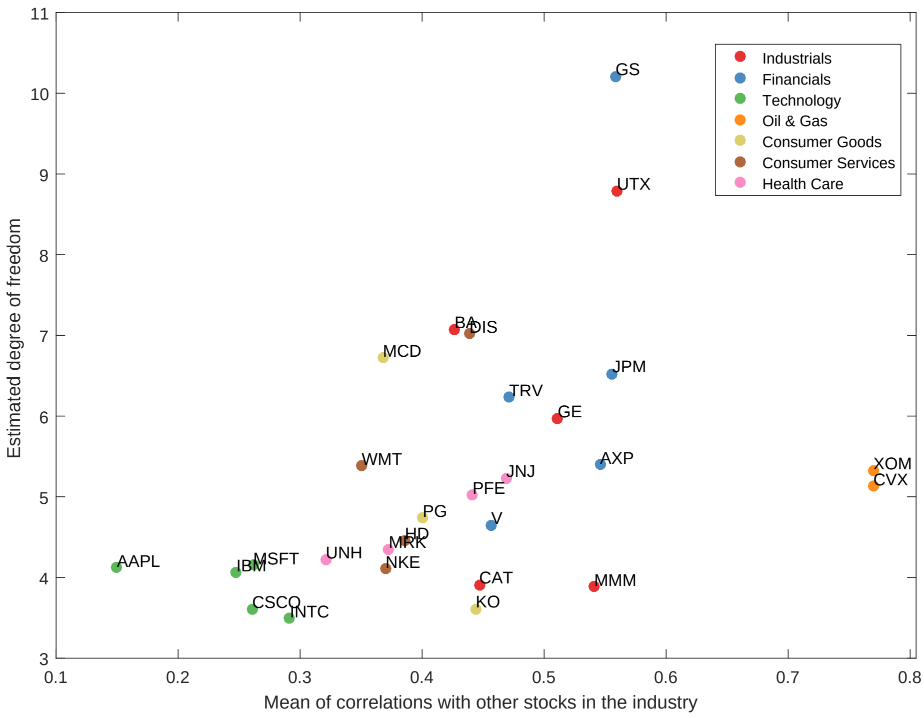


FIGURE 6.8: Stock distribution under the 7 industries based on the estimated degree of freedom and mean of correlations with other stocks in the same industry.

According to Figure 6.10(a) and 6.10(b), we find that MV-based strategies perform more satisfactory in the middle of average correlation (from 0.25 to 0.60) with the other stocks in the same industry and TPR-based strategies might have a more considerable result only if the estimated degree of freedom is located in the interval $(4, 5.5)$. Specifically, the small estimated degree of freedom (nearly less than 5) and medium-small average correlation (approximately from 0.25 to 0.5) are the essential requirements for better performance of MV-TPR while MV-GPR strategy is likely to obtain the profitable investment only if both the estimated degree of freedom and the mean of correlations with others in the same industry are moderate.

Furthermore, the industrial sector portfolio is taken into account, which consists of these grouped stocks by the same weight investment on each stock. For example, the Oil & Gas portfolio investment is \$100 with \$50 shares CVX and \$50 shares XOM while the Technology portfolio investment is \$100 with the same \$20 investment on each stock in the industrial sector Technology. The diverse industry portfolio investments' ranking lists in Table 6.9 (The details are described in

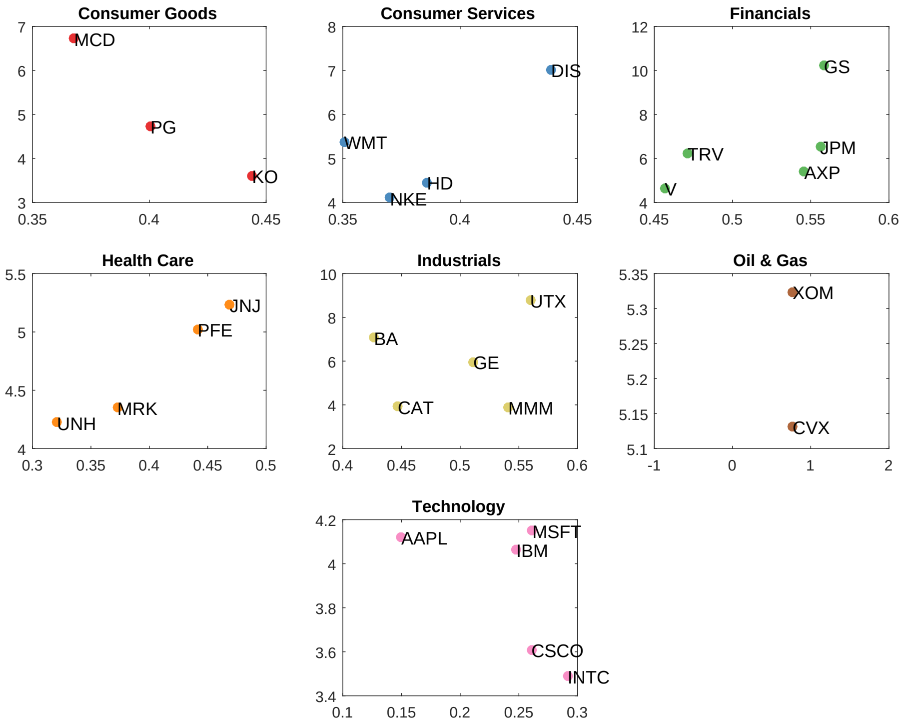
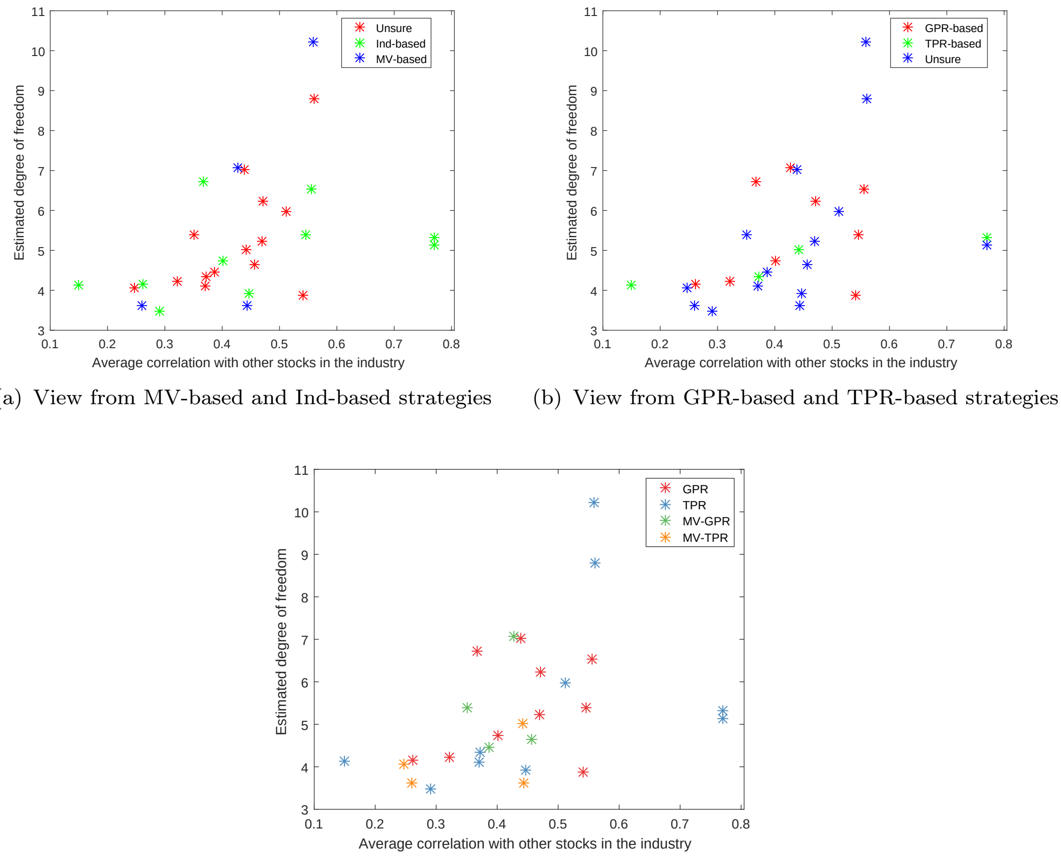


FIGURE 6.9: Stock distribution in the 7 different industries. For all sub-figures, y-label: the estimated degree of freedom. x-label: the mean of correlations with other stocks in the same industry.

Table E.2, all the figures of diverse industrial sector portfolio investment are presented in Appendix E.4). Apparently, the Buy&Sell strategies performed better than the Buy&Hold strategies. MV-GPR suits better in three industries, including Consumer Goods, Consumer Services, and Financials, followed by TPR which performed best in Oil&Gas and Industrials. The optimum investment strategy in Health Care is MV-TPR while in Technology industry, using GPR seems to be the most profitable.

To sum up, despite the fact that the multivariate process regression models, including MV-GPR and MV-TPR. cannot make the best performances in all the cases compared with independent GPR and TPR, undeniably, the proposed models are still applicable to stock market investment.



(a) View from MV-based and Ind-based strategies (b) View from GPR-based and TPR-based strategies

(c) View from the detailed strategies

FIGURE 6.10: Scatter plot for the stocks using optimum investment strategy (or strategy group)

TABLE 6.9: Industry portfolio investment ranking under different strategies

Industry Portfolio	Buy&Sell Strategy				Buy&Hold Stragegy			
	MV-GPR	MV-TPR	GPR	TPR	Stock	INDU	NDX	SPX
Oil & Gas	4th	3rd	2nd	***1st***	8th	7th	5th	6th
Industrials	2nd	4th	3rd	***1st***	8th	7th	5th	6th
Consumer Goods	***1st***	4th	2nd	3rd	7th	8th	5th	6th
Health Care	4th	***1st***	3rd	2nd	6th	8th	5th	7th
Consumer Services	***1st***	4th	3rd	2nd	5th	8th	6th	7th
Financials	***1st***	4th	2nd	3rd	5th	8th	6th	7th
Technology	4th	3rd	***1st***	2nd	5th	8th	6th	7th

6.7 Summary

In this chapter, we propose an alternative derivation of dependent Gaussian process regression for multi-output prediction, where the model settings, derivations and computations are all directly performed in matrix form. MV-GPR is a more straightforward method and can be implemented in the same way as conventional GPR. Like the conventional Gaussian process for vector-valued function, our models are also able to learn the correlations between inputs and outputs. In comparison to the conventional Gaussian process for vector-valued function, our formulations are more convenient and flexible.

Furthermore, we define the multivariate Student−t process and then derive a new considerable method, MV-TPR for multi-output prediction. Both MV-GPR and MV-TPR have closed form expressions for the marginal likelihoods and predictive distributions. The usefulness of the proposed methods are illustrated through several numerical examples.

The proposed methods are also applied to stock market modelling and are shown to make profitable stock investment. Firstly, MV-GPR and MV-TPR are applied to predict three "Chinese concept stocks" together as 3-dimension outputs. The Buy&Sell strategies based on the proposed models have more satisfactory performances compared with Buy&Hold strategies for the corresponding stocks and three main indices in the US, especially the strategy based on MV-TPR has outstanding returns for NetEase among three Chinese stocks. Secondly, our proposed methods are applied to make Buy&Sell strategies from view of industrial sectors in Dow 30 and their results indicate that the strategies based on MV-GPR generally has considerable performances in Industrials, Consumer Goods, Consumer Services, and Financials sectors while the strategies based on MV-TPR can make maximum profit in Health Care sector.

Chapter 7

Conclusions and future work

7.1 Conclusions

In this thesis, we introduce the framework of Bayesian non-parametric Gaussian process regression and its extensions, including Gaussian process regression with Student$-t$ likelihood (GPRT), Student$-t$ process regression (TPR), state-space Gaussian process regression (SSGPR), state-space Student$-t$ process regression (SSTPR), multivariate Gaussian process regression (MV-GPR), and multivariate Student$-t$ process regression (MV-TPR). By applying all these models to stock markets, GPR and its extensions show the powerful ability and usefulness in financial time series prediction. This thesis is divided into 3 main parts.

In the first part, we carefully review Gaussian process regression from both weight-space view and function-space view followed by a detailed introduction and discussion of mean function, kernel, and hyper-parameter estimation. In particular, we study the sensitivity of prior distribution for initial value to the hyper-parameter estimation and the performance of GPR. The results of several numerical experiments show that the sensitivity of the initial hyper-parameters depends on the choice of the specific kernel, but the priors have little influence on the performance of the GPR models in terms of predictability.

The second part introduce several Gaussian process regression extensions, including Gaussian process regression with Student$-t$ process (GPRT) and Student$-t$ process regression (TPR), with all the above models applied to predict 10 main equity indices from all over the world. According to the experimental results, both GPR and TPR show a considerable capability of predicting equity indices. Both

GPR and TPR are extended to state-space Gaussian process regression (SSGPR) and state-space Student$-t$ process regression (SSTPR) models in order to make an reasonable prediction of the selected stock markets based on historical trading data in a dynamic system. In addition, a brief market efficiency analysis is also conducted by taking advantage of the results of the SSGPR and SSTPR prediction models for the equity indices.

The final part focuses on the multi-output prediction using Gaussian process. We proposed a multivariate Gaussian process regression (MV-GPR) and a multivariate Student$-t$ process regression (MV-TPR) for multi-output prediction, where the model settings, derivations and computations are all performed in matrix form directly, rather than vectorising the matrices involved. Compared with the independent Gaussian process regression and Student$-t$ process regression models, both MV-GPR and MV-TPR significantly show outstanding performances in the simulated examples. The proposed methods are then applied to stock market modelling. The Buy&Sell strategies, which are generated by our proposed methods, are shown to be profitable when making stock market investments.

7.2 Future work

Short-term future work will be carried out on the remaining problems identified in this thesis. Firstly, in Chapter 4, we have studied the sensitivity of the hyper-parameter estimation and the performance of GPR on the prior distribution for the initial value. It is noted that in terms of evaluating the influences of prior distributions on the performance of GPR models, the study in this thesis is far from comprehensive. More priors and kernels should be considered, as well as complex data sets, including real data. The theoretical analysis might also be of importance because it is not feasible for numerical examples to cover all scenarios. In addition, according to Chapter 6, we assume that different outputs would be observed for the same covariate values. In practice, different responses are sometimes observed at several different locations. This is, however, difficult for the proposed method for multi-output prediction since all the outputs have to be considered as a matrix rather than a vector with adjustable length, leaving each response reliant on the same input. Furthermore, the kernel in our model is squared exponentially and is the same for each output, with it potentially being better to use different kernels for different outputs [41]. All these problems remain to explore in future work.

The further future work consists of three directions, including model extension and specification, efficient implementation, and extensive application. From the view of model extension and specification, GP has to be extended further in order to learn more powerful representation of the data in this deep learning age. For example, deep Gaussian process, of which the data is modelled as the output of a multivariate GP and the inputs of that GP are then governed by another GP [86]. Of course, traditional GP models also deserve to be further studied, especially the kernel selection. Although several sophisticated covariance functions have been widely used [65, 66], the relationship between data characteristics and kernels still need to be explored. For example, we will attempt to construct a complicated kernel to capture the fluctuation of the stock market efficiently. In fact, though GP is a rigorous model theoretically, the applications in industry are limited owing to the high computational complexity of its inference method, especially when the size of the training set is increasingly huge [87, 88, 89]. In order to resolve this problem, many approximate inferences are proposed to reduce the running time at the cost of accuracy. For instance, re-sampling the data sets to eliminate those less informative points [90] and using sparse techniques to reduced the rank of the kernel [91]. Therefore, it is necessary to study the efficient implementation of GP at the expense of less accuracy loss. At last, GP models will have many useful and worthy applications in other fields, such as biology, geography, and physics.

Appendix A

Graphs of predictions by Gaussian process regression, Gaussian process regression with Student$-t$ likelihood and Student$-t$ process regression models for INDU, NDX, SPX, and UKX

[1] Two-sigma confidence interval is the 95% confidence interval for Gaussian ,but it is not the same as 95% confidence interval for Student$-t$ distribution.

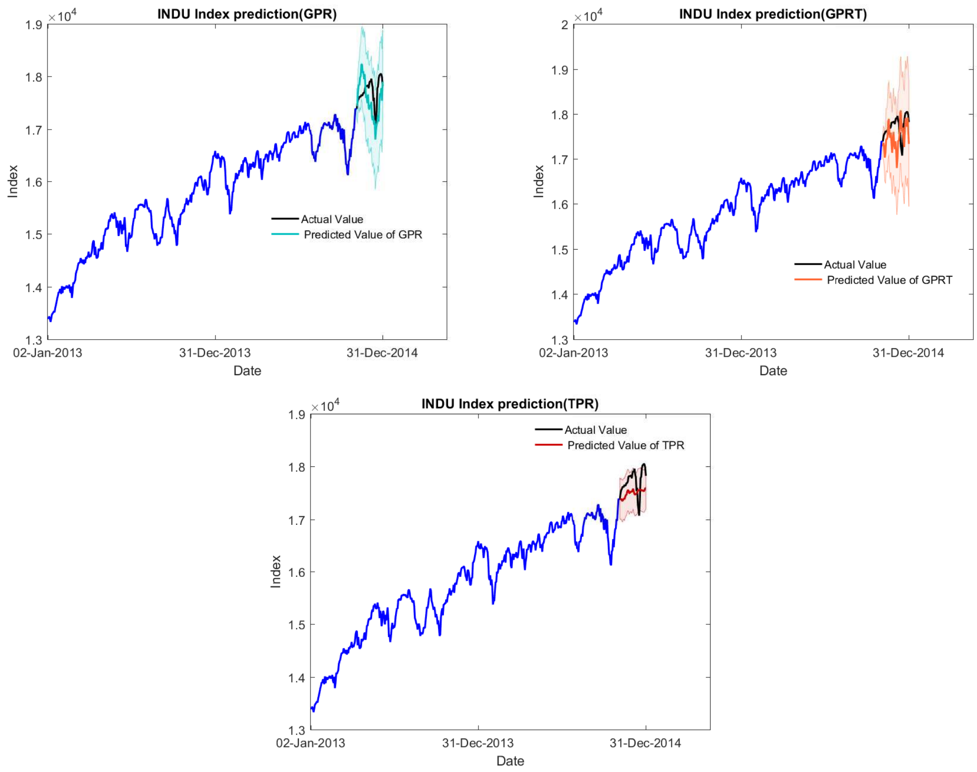


FIGURE A.1: INDU predictions by GPR, GPRT and TPR respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma [1] confidence intervals.

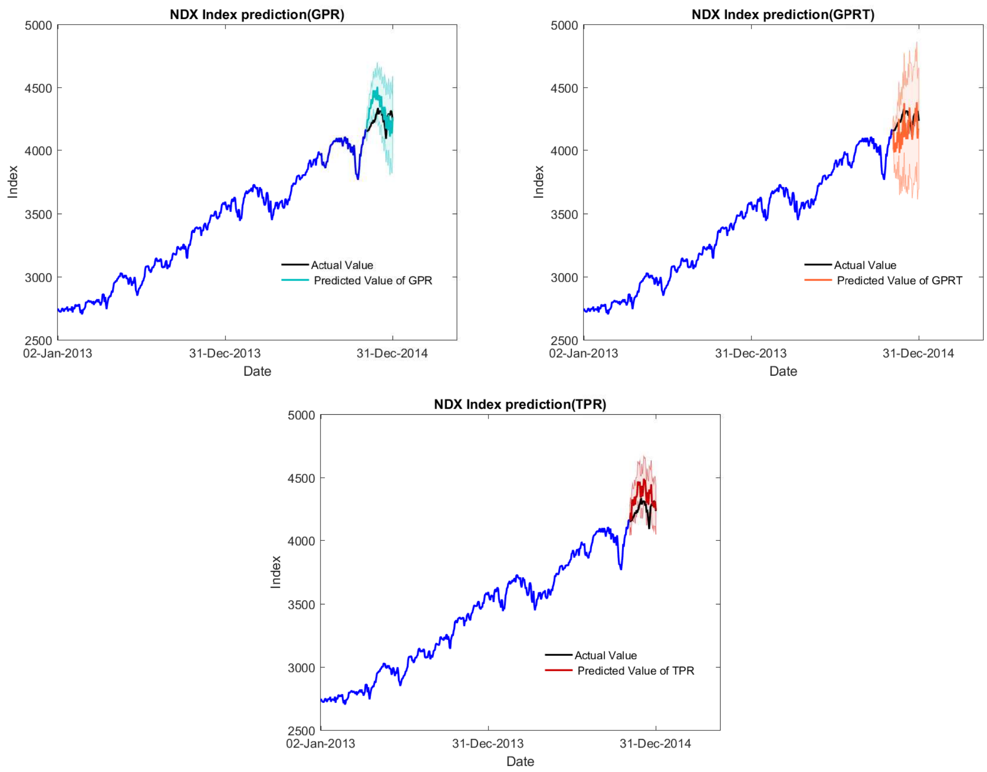


FIGURE A.2: NDX predictions by GPR, GPRT and TPR respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

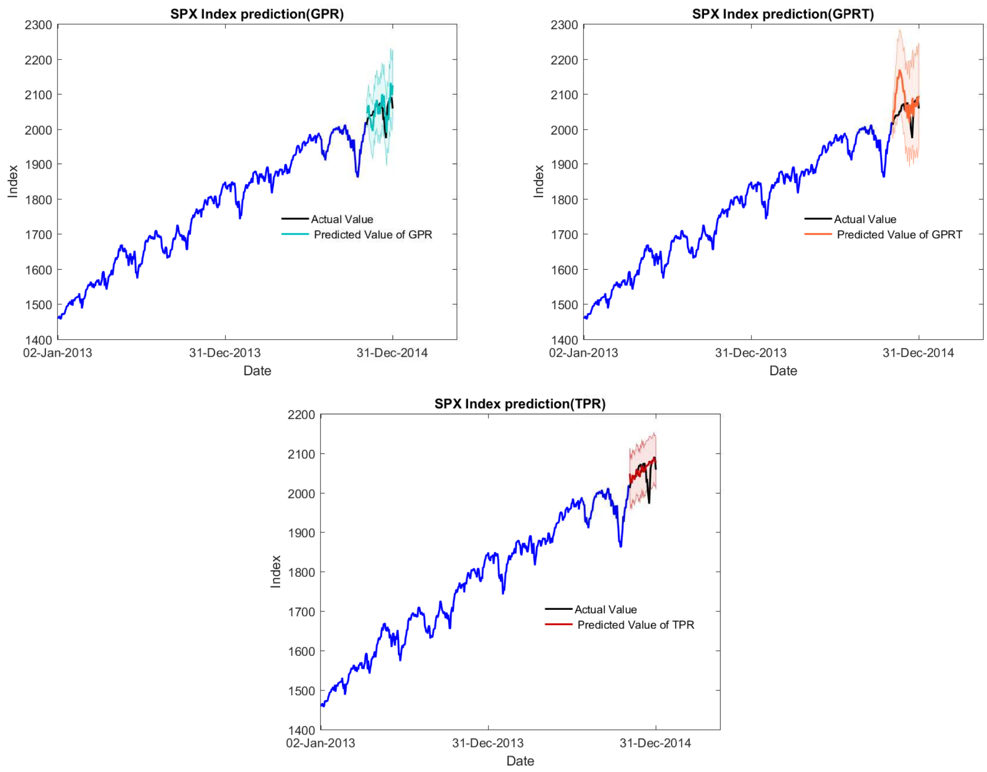


FIGURE A.3: SPX predictions by GPR, GPRT and TPR respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

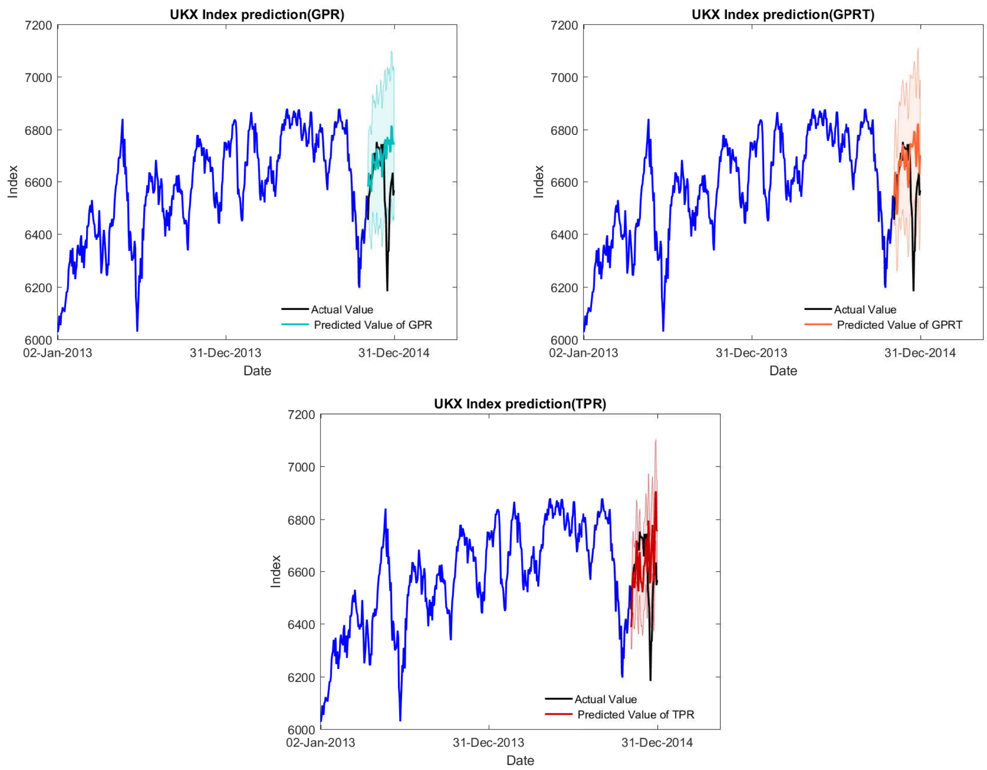


FIGURE A.4: UKX predictions by GPR, GPRT and TPR respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

Appendix B

Graphs of predictions by Gaussian process regression, Gaussian process regression with Student$-t$ likelihood, Student$-t$ process regression and ARMA(1,1) models for DAX, HSI, INDU, NDX, NKY, SENSEX, SPX, and UKX

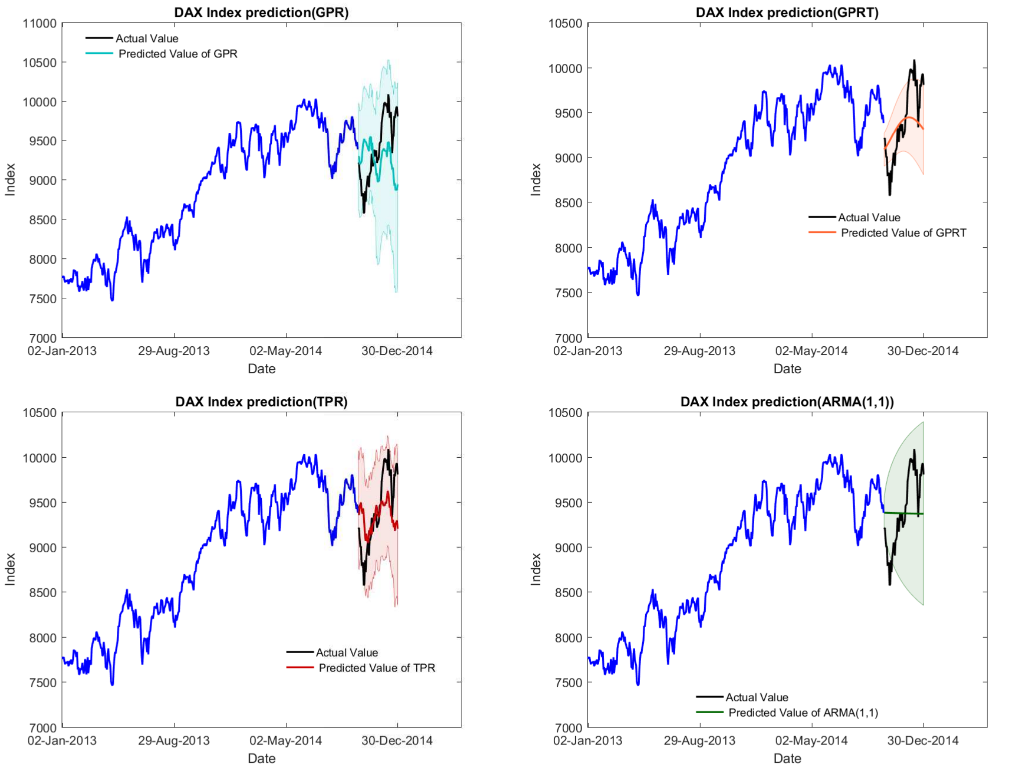


Figure B.1: DAX predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

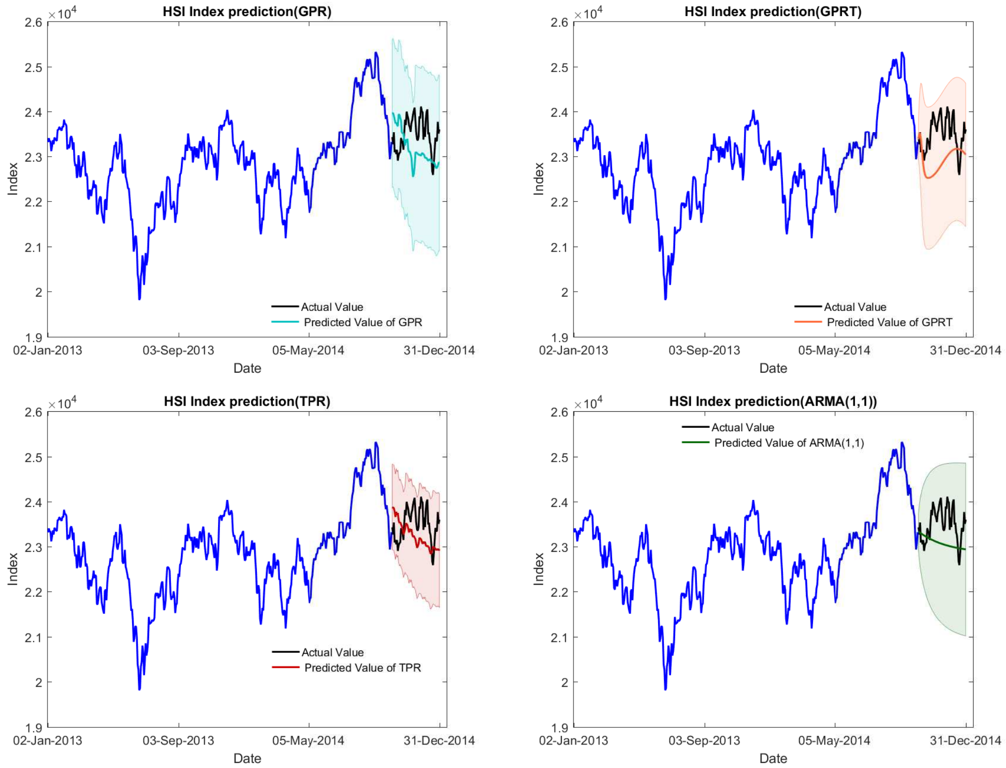


Figure B.2: HSI predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

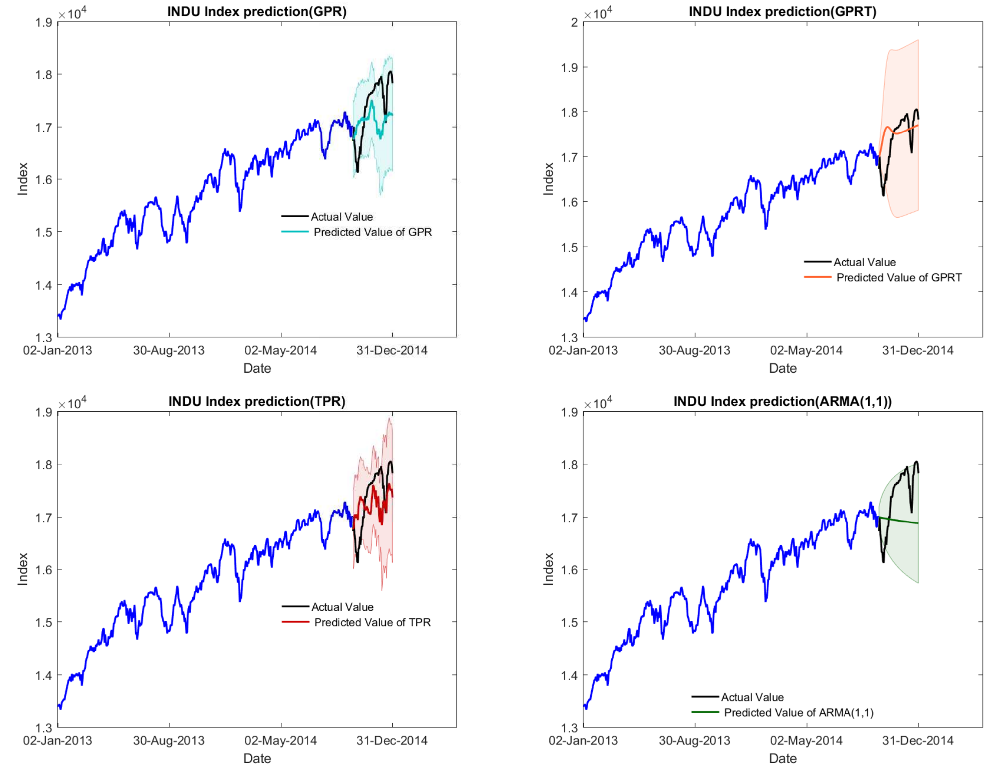


FIGURE B.3: INDU predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

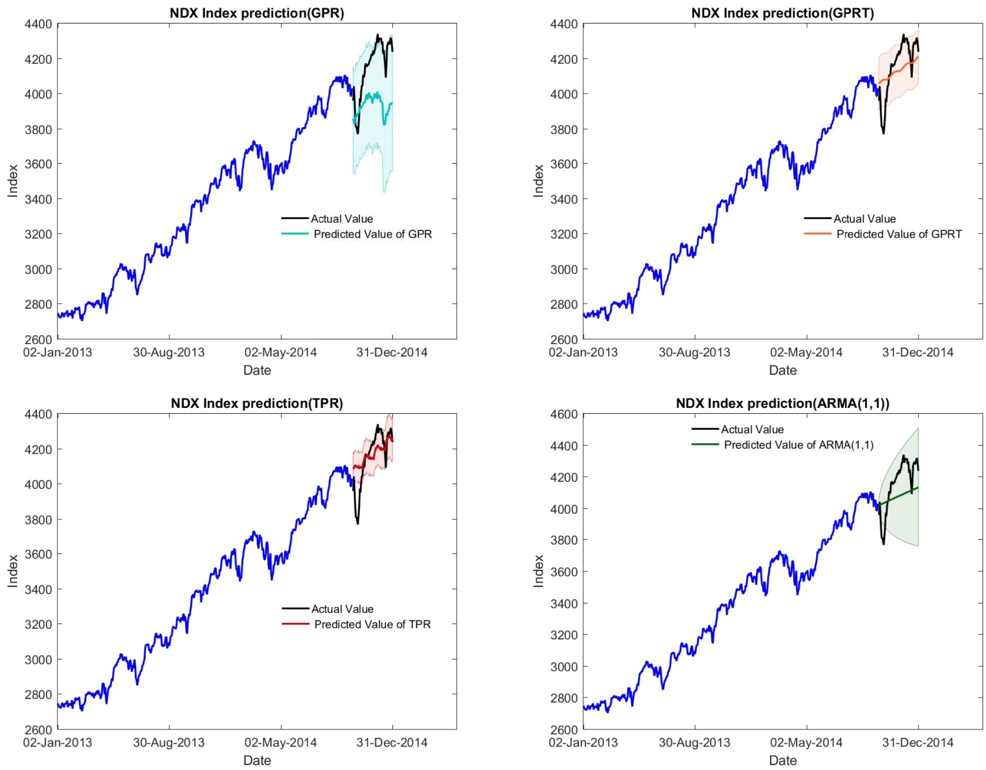


FIGURE B.4: NDX predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

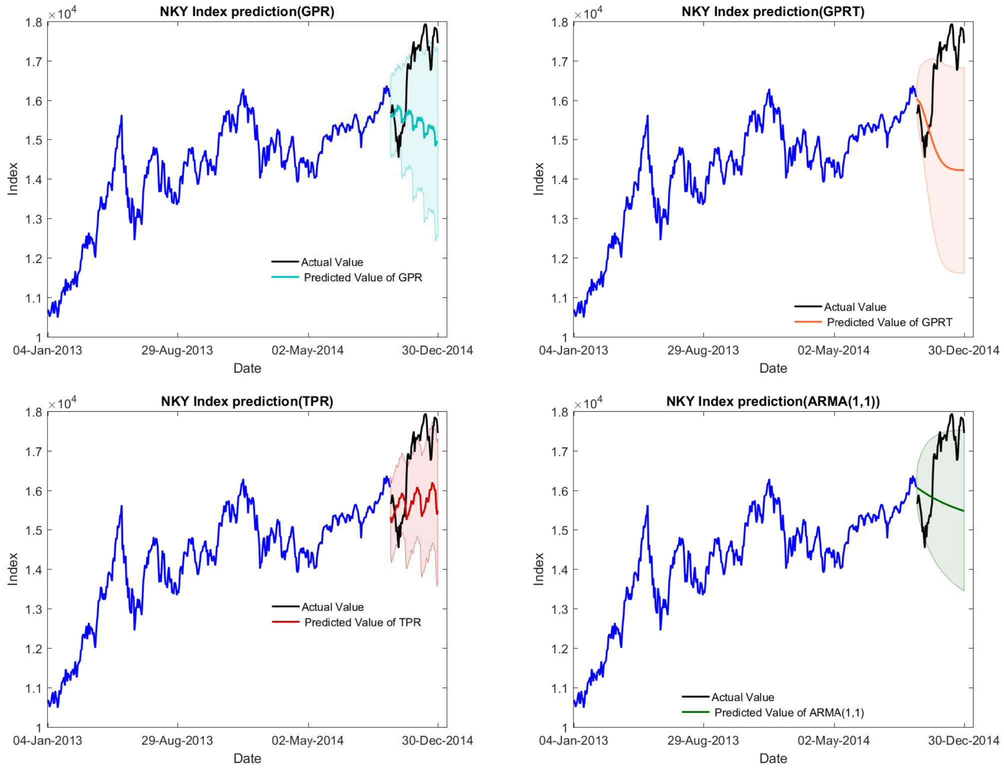


FIGURE B.5: NKY predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

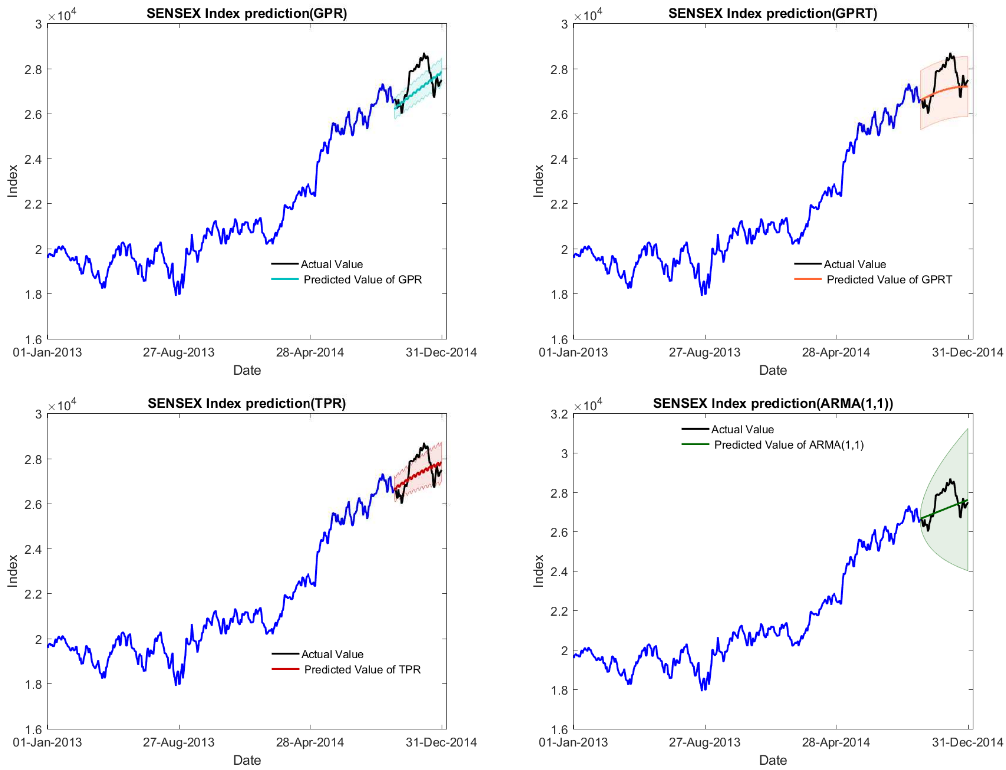


FIGURE B.6: SENSEX predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

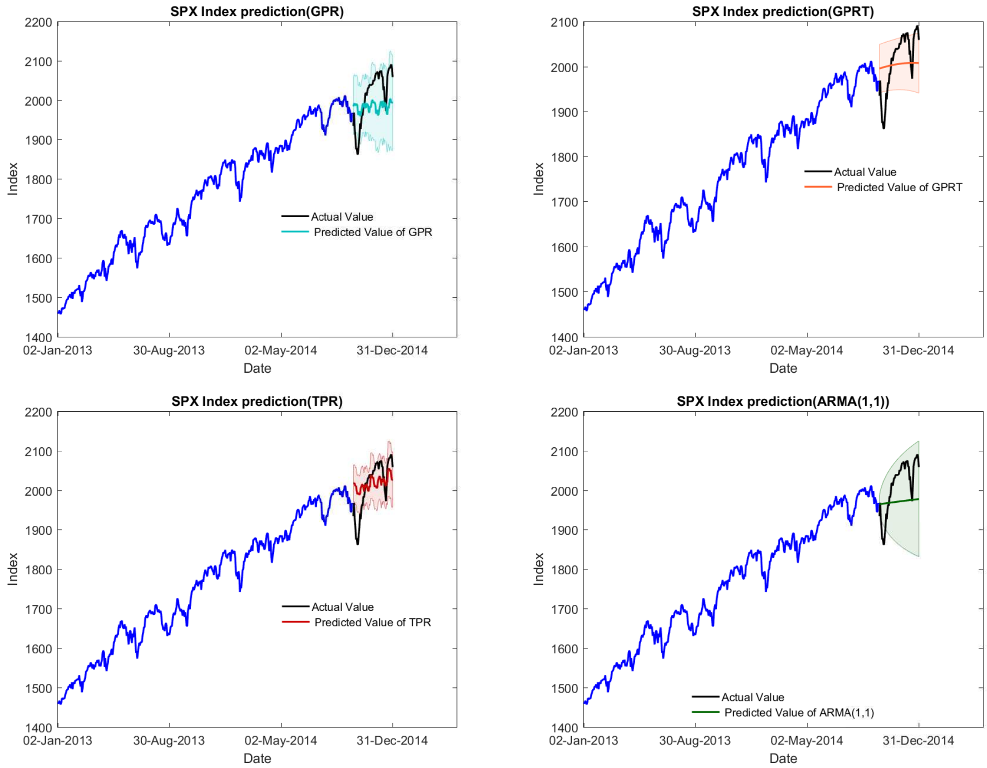


FIGURE B.7: SPX predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

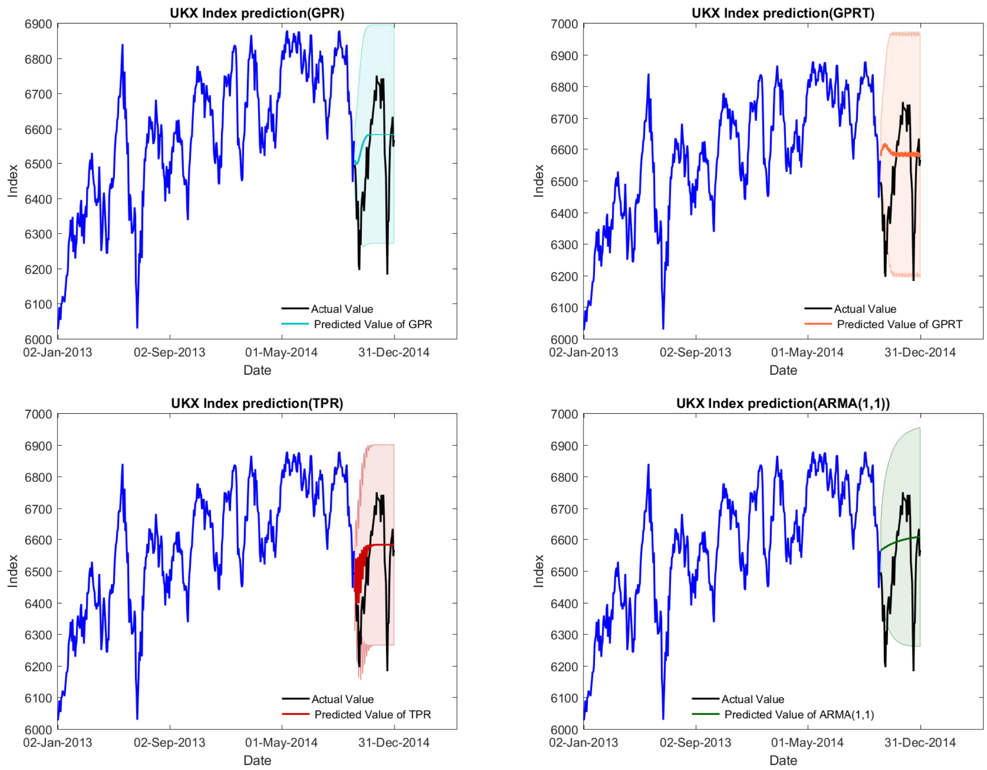


FIGURE B.8: UKX predictions by GPR, GPRT, TPR and ARMA(1,1) respectively. Solid colourful dark lines are predicted values and light colourful regions are two-sigma confidence intervals.

Appendix C

Negative log marginal likelihood and gradient evaluation for multivariate Gaussian and Student$-t$ process

C.1 Multivariate Gaussian process regression

For a matrix-variate observations $Y \sim \mathcal{MN}_{n,d}(M, \Sigma, \Omega)$ where $M \in \mathbb{R}^{n\times d}, \Sigma \in \mathbb{R}^{n\times n}, \Omega \in \mathbb{R}^{d\times d}$, the negative log likelihood is

$$\mathcal{L} = \frac{nd}{2}\log(2\pi) + \frac{d}{2}\log\det(\Sigma) + \frac{n}{2}\log\det(\Omega) + \frac{1}{2}\mathrm{tr}(\Sigma^{-1}(Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}), \tag{C.1}$$

where actually $\Sigma = K + \sigma_n^2\mathbf{I}$ As we know there are several parameters in the kernel k so that we can denote $K = K_\theta$. The parameter set denotes $\Theta = \{\theta_1, \theta_2, \ldots\}$. Besides, we denote the parameter matrix $\Omega = \Phi\Phi^{\mathrm{T}}$ since Ω is positive semi-definite, where

$$\Phi = \begin{bmatrix} \phi_{11} & 0 & \cdots & 0 \\ \phi_{21} & \phi_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{d1} & \phi_{d2} & \cdots & \phi_{dd} \end{bmatrix}.$$

To guarantee the uniqueness of Φ, the diagonal elements are restricted to be positive and denote $\varphi_{ii} = \log(\phi_{ii})$ for $i = 1, 2, \cdots, d$. Therefore,

$$\frac{\partial \Sigma}{\sigma_n^2} = \mathbf{I}_n, \quad \frac{\partial \Sigma}{\partial \theta_i} = \frac{\partial K'_\theta}{\partial \theta_i}, \qquad \frac{\partial \Omega}{\partial \phi_{ij}} = \mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij}, \quad \frac{\partial \Omega}{\partial \varphi_{ii}} = \mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii},$$

where $\mathbf{E}_{ij}$ is the $d \times d$ elementary matrix having unity in the (i,j)-th element and zeros elsewhere, and $\mathbf{J}_{ii}$ is the same as $\mathbf{E}_{ij}$ but with the unity being replaced by $e^{\varphi_{ii}}$.

The derivatives of the negative log likelihood with respect to σ_n^2, θ_i, ϕ_{ij} and φ_{ii} are as follows. The derivative with respect to θ_i is

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \theta_i} &= \frac{d}{2}\frac{\partial \log\det(\Sigma)}{\partial \theta_i} + \frac{1}{2}\frac{\partial}{\partial \theta_i}\mathrm{tr}(\Sigma^{-1}(Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}) \\
&= \frac{d}{2}\mathrm{tr}\left[\left(\frac{\partial \log\det(\Sigma)}{\partial \Sigma}\right)^{\mathrm{T}}\frac{\partial \Sigma}{\partial \theta_i}\right] + \frac{1}{2}\mathrm{tr}\left[\left(\frac{\partial \mathrm{tr}(\Sigma^{-1}G)}{\partial \Sigma}\right)^{\mathrm{T}}\frac{\partial \Sigma}{\partial \theta_i}\right] \\
&= \frac{d}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial K'_\theta}{\partial \theta_i}\right) - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}G\Sigma^{-1}\frac{\partial K'_\theta}{\partial \theta_i}\right) \\
&= \frac{d}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial K'_\theta}{\partial \theta_i}\right) - \frac{1}{2}\mathrm{tr}\left(\alpha_\Sigma\Omega^{-1}\alpha_\Sigma^{\mathrm{T}}\frac{\partial K'_\theta}{\partial \theta_i}\right),
\end{aligned} \tag{C.2}$$

where $G = (Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}$ and $\alpha_\Sigma = \Sigma^{-1}(Y-M)$.The fourth equality is due to the symmetry of Σ.

Due to $\partial\Sigma/\partial\sigma_n^2 = \mathbf{I}_n$, the derivative with respect to σ_n^2 is:

$$\frac{\partial \mathcal{L}}{\partial \sigma_n^2} = \frac{d}{2}\mathrm{tr}(\Sigma^{-1}) - \frac{1}{2}\mathrm{tr}(\alpha_\Sigma\Omega^{-1}\alpha_\Sigma^{\mathrm{T}}). \tag{C.3}$$

Letting $\alpha_\Omega = \Omega^{-1}(Y-M)^{\mathrm{T}}$, the derivative with respect to ϕ_{ij} is

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \phi_{ij}} &= \frac{n}{2}\frac{\partial \log\det(\Omega)}{\partial \phi_{ij}} + \frac{1}{2}\frac{\partial}{\partial \phi_{ij}}\mathrm{tr}(\Sigma^{-1}(Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}) \\
&= \frac{n}{2}\mathrm{tr}\left(\Omega^{-1}\frac{\partial \Omega}{\partial \phi_{ij}}\right) - \frac{1}{2}\mathrm{tr}\left[((\Omega^{-1}(Y-M)^{\mathrm{T}}\Sigma^{-1}(Y-M)\Omega^{-1})^{\mathrm{T}})^{\mathrm{T}}\frac{\partial \Omega}{\partial \phi_{ij}}\right] \\
&= \frac{n}{2}\mathrm{tr}\left(\Omega^{-1}\frac{\partial \Omega}{\partial \phi_{ij}}\right) - \frac{1}{2}\mathrm{tr}\left(\alpha_\Omega\Sigma^{-1}\alpha_\Omega^{\mathrm{T}}\frac{\partial \Omega}{\partial \phi_{ij}}\right) \\
&= \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})] - \frac{1}{2}\mathrm{tr}[\alpha_\Omega\Sigma^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})],
\end{aligned} \tag{C.4}$$

where the third equation is due to the symmetry of Ω. Similarly, the derivative with respect to φ_{ii} is

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \varphi_{ii}} &= \frac{n}{2}\frac{\partial \log\det(\Omega)}{\partial \varphi_{ii}} + \frac{1}{2}\frac{\partial}{\partial \varphi_{ii}}\mathrm{tr}(\Sigma^{-1}(Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}) \\
&= \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})] - \frac{1}{2}\mathrm{tr}[\alpha_\Omega\Sigma^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})].
\end{aligned} \tag{C.5}$$

C.2 Multivariate Student$-t$ process regression

The negative log likelihood of observations $Y \sim \mathcal{MT}_{n,d}(\nu, M, \Sigma, \Omega)$ where $M \in \mathbb{R}^{n\times d}, \Sigma \in \mathbb{R}^{n\times n}, \Omega \in \mathbb{R}^{d\times d}$, is

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2}(\nu+d+n-1)\log\det(\mathbf{I}_n + \Sigma^{-1}(Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}) \\
&\quad + \frac{d}{2}\log\det(\Sigma) + \frac{n}{2}\log\det(\Omega) + \log\Gamma_n\left(\frac{1}{2}(\nu+n-1)\right) + \frac{1}{2}dn\log\pi \\
&\quad - \log\Gamma_n\left(\frac{1}{2}(\nu+d+n-1)\right) \\
&= \frac{1}{2}(\nu+d+n-1)\log\det(\Sigma + (Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}) - \frac{\nu+n-1}{2}\log\det(\Sigma) \\
&\quad + \log\Gamma_n\left(\frac{1}{2}(\nu+n-1)\right) - \log\Gamma_n\left(\frac{1}{2}(\nu+d+n-1)\right) \\
&\quad + \frac{n}{2}\log\det(\Omega) + \frac{1}{2}dn\log\pi.
\end{aligned}$$

Letting $U = \Sigma + (Y-M)\Omega^{-1}(Y-M)^{\mathrm{T}}$ and $\alpha_\Omega = \Omega^{-1}(Y-M)^{\mathrm{T}}$, the derivative of U with respect to σ_n^2, θ_i,ν, ϕ_{ij} and φ_{ii} are

$$\frac{\partial U}{\partial \sigma_n^2} = \mathbf{I}_n, \quad \frac{\partial U}{\partial \theta_i} = \frac{\partial K'_\theta}{\partial \theta_i}, \quad \frac{\partial U}{\partial \nu} = 0, \tag{C.6}$$

$$\frac{\partial U}{\partial \phi_{ij}} = -(Y-M)\Omega^{-1}\frac{\partial \Omega}{\partial \phi_{ij}}\Omega^{-1}(Y-M)^{\mathrm{T}} = -\alpha_\Omega^{\mathrm{T}}\frac{\partial \Omega}{\partial \phi_{ij}}\alpha_\Omega, \tag{C.7}$$

$$\frac{\partial U}{\partial \varphi_{ii}} = -(Y-M)\Omega^{-1}\frac{\partial \Omega}{\partial \varphi_{ii}}\Omega^{-1}(Y-M)^{\mathrm{T}} = -\alpha_\Omega^{\mathrm{T}}\frac{\partial \Omega}{\partial \varphi_{ii}}\alpha_\Omega. \tag{C.8}$$

Therefore, the derivative of negative log marginal likelihood with respect to θ_i is

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \theta_i} &= \frac{(\tau+d)}{2}\frac{\partial \log\det(U)}{\partial \theta_i} - \frac{\tau}{2}\frac{\partial \log\det(\Sigma)}{\partial \theta_i} \\
&= \frac{(\tau+d)}{2}\mathrm{tr}\left(U^{-1}\frac{\partial K'_\theta}{\partial \theta_i}\right) - \frac{\tau}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial K'_\theta}{\partial \theta_i}\right),
\end{aligned} \tag{C.9}$$

where the constant $\tau = \nu + n - 1$.

The derivative with respect to σ_n^2 is

$$\frac{\partial \mathcal{L}}{\partial \sigma_n^2} = \frac{(\tau+d)}{2}\mathrm{tr}(U^{-1}) - \frac{\tau}{2}\mathrm{tr}(\Sigma^{-1}). \tag{C.10}$$

The derivative with respect to ν is

$$\frac{\partial \mathcal{L}}{\partial \nu} = \frac{1}{2}\log\det(U) - \frac{1}{2}\log\det(\Sigma) + \frac{1}{2}\psi_n(\frac{1}{2}\tau) - \frac{1}{2}\psi_n[\frac{1}{2}(\tau+d)] \tag{C.11}$$

where $\psi_n(\cdot)$ is the derivative of the function $\log\Gamma_n(\cdot)$ with respect to ν.

The derivative of $\mathcal{L}$ with respect to ϕ_{ij} is

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \phi_{ij}} &= \frac{(\tau+d)}{2}\frac{\partial \log\det(U)}{\partial \phi_{ij}} + \frac{n}{2}\frac{\partial \log\det(\Omega)}{\partial \phi_{ij}} \\ &= -\frac{(\tau+d)}{2}\mathrm{tr}[U^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})\alpha_\Omega] + \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{E}_{ij}\Phi^{\mathrm{T}} + \Phi\mathbf{E}_{ij})].\end{aligned} \tag{C.12}$$

Similarly, the derivative with respect to φ_{ii} is

$$\frac{\partial \mathcal{L}}{\partial \varphi_{ii}} = -\frac{(\tau+d)}{2}\mathrm{tr}[U^{-1}\alpha_\Omega^{\mathrm{T}}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})\alpha_\Omega] + \frac{n}{2}\mathrm{tr}[\Omega^{-1}(\mathbf{J}_{ii}\Phi^{\mathrm{T}} + \Phi\mathbf{J}_{ii})]. \tag{C.13}$$

Appendix D

Investment details of three Chinese stocks listed in NASDAQ

Table D.1: The movement of invested $100 for 200 days split in to 20 periods (Stock: BIDU)

Forecast terms	Buy&Sell decisions by prediction models				Buy&Hold stock/index			
	MV-GPR	MV-TPR	GPR	TPR	BIDU	INDU	NDX	SPX
Beginning ($)	100				100			
Period 1	103.53	100.97	103.53	100.97	97.14	101.20	98.70	100.71
Period 2	109.80	106.09	115.60	102.03	94.87	99.55	94.10	98.44
Period 3	108.37	104.71	115.96	100.71	93.89	101.50	96.64	100.62
Period 4	117.17	113.22	125.37	108.89	95.21	101.70	96.94	100.87
Period 5	127.84	123.52	136.79	118.80	102.22	102.22	100.79	102.55
Period 6	137.94	130.14	147.59	128.18	107.39	102.44	101.54	103.09
Period 7	146.20	137.93	156.43	135.86	112.11	103.12	103.25	104.54
Period 8	155.97	147.15	166.88	144.94	113.57	103.72	105.34	105.09
Period 9	177.94	167.88	175.27	152.21	138.22	103.82	106.98	105.67
Period 10	179.83	174.07	177.13	153.83	131.26	101.33	104.90	103.17
Period 11	179.08	173.35	176.05	153.19	130.71	104.07	109.34	106.20
Period 12	190.96	184.85	187.73	163.35	137.58	104.75	110.49	106.91
Period 13	201.08	194.63	204.44	177.89	131.12	105.12	109.57	106.52
Period 14	207.28	200.64	210.75	183.38	132.54	104.01	108.35	104.94
Period 15	210.54	203.80	214.06	186.26	132.19	100.39	104.41	101.70
Period 16	233.92	226.43	237.83	206.95	144.35	106.31	112.48	107.77
Period 17	252.47	233.71	256.69	223.36	148.99	108.03	113.68	109.03
Period 18	250.38	227.57	254.56	221.51	143.25	109.45	116.17	110.38
Period 19	260.24	223.38	264.59	230.23	134.23	104.49	110.33	105.37
Period 20	270.29	233.29	274.81	239.13	139.12	109.10	114.29	109.97

Table D.2: The movement of invested $100 for 200 days split in to 20 periods (Stock: CTRP)

Forecast terms	Buy&Sell decisions by prediction models				Buy&Hold stock/index			
	MV-GPR	MV-TPR	GPR	TPR	CTRP	INDU	NDX	SPX
Beginning ($)	100				100			
Period 1	105.67	105.67	105.67	105.67	100.78	101.20	98.70	100.71
Period 2	102.39	102.52	102.39	102.39	99.65	99.55	94.10	98.44
Period 3	102.77	102.90	102.77	102.77	91.63	101.50	96.64	100.62
Period 4	110.05	110.19	110.05	110.05	100.39	101.70	96.94	100.87
Period 5	121.51	121.66	121.51	121.51	108.33	102.22	100.79	102.55
Period 6	131.02	131.19	131.02	131.02	113.59	102.44	101.54	103.09
Period 7	138.90	139.08	144.25	138.90	120.90	103.12	103.25	104.54
Period 8	140.19	140.37	145.58	140.19	118.02	103.72	105.34	105.09
Period 9	150.29	146.38	151.82	146.20	131.35	103.82	106.98	105.67
Period 10	167.38	163.03	154.49	162.82	128.82	101.33	104.90	103.17
Period 11	166.33	162.01	154.28	161.80	127.37	104.07	109.34	106.20
Period 12	176.07	171.50	163.32	171.28	133.52	104.75	110.49	106.91
Period 13	176.00	171.43	163.25	171.21	117.53	105.12	109.57	106.52
Period 14	170.50	166.08	158.15	165.86	109.88	104.01	108.35	104.94
Period 15	178.68	174.04	165.73	173.82	108.35	100.39	104.41	101.70
Period 16	184.31	187.64	170.96	179.30	114.27	106.31	112.48	107.77
Period 17	183.77	191.85	176.72	180.87	115.96	108.03	113.68	109.03
Period 18	163.88	194.84	158.00	169.76	93.94	109.45	116.17	110.38
Period 19	169.89	201.99	163.80	176.73	80.69	104.49	110.33	105.37
Period 20	183.25	222.82	176.67	190.63	89.20	109.10	114.29	109.97

Table D.3: The movement of invested $100 for 200 days split in to 20 periods (Stock: NTES)

Forecast terms	Buy&Sell decisions by prediction models				Buy&Hold stock/index			
	MV-GPR	MV-TPR	GPR	TPR	NTES	INDU	NDX	SPX
Beginning ($)	100				100			
Period 1	104.51	104.51	104.51	104.51	106.79	101.20	98.70	100.71
Period 2	106.19	106.19	106.19	106.19	106.35	99.55	94.10	98.44
Period 3	106.80	106.80	106.80	109.12	104.14	101.50	96.64	100.62
Period 4	115.90	115.90	115.90	114.28	108.66	101.70	96.94	100.87
Period 5	115.82	115.82	115.82	114.21	109.84	102.22	100.79	102.55
Period 6	120.96	117.65	120.73	115.20	116.55	102.44	101.54	103.09
Period 7	123.27	121.54	124.72	119.01	120.32	103.12	103.25	104.54
Period 8	127.77	125.33	128.62	122.73	117.34	103.72	105.34	105.09
Period 9	133.21	128.81	134.09	127.95	128.53	103.82	106.98	105.67
Period 10	133.36	128.96	134.24	128.09	127.40	101.33	104.90	103.17
Period 11	141.13	136.47	142.80	135.56	134.83	104.07	109.34	106.20
Period 12	141.45	136.78	139.44	135.87	137.23	104.75	110.49	106.91
Period 13	145.98	141.16	143.90	140.22	134.27	105.12	109.57	106.52
Period 14	147.95	144.00	145.84	143.04	129.90	104.01	108.35	104.94
Period 15	151.75	147.70	149.59	146.71	139.19	100.39	104.41	101.70
Period 16	158.59	154.36	156.33	153.33	144.80	106.31	112.48	107.77
Period 17	170.12	165.58	167.70	165.07	156.05	108.03	113.68	109.03
Period 18	177.19	171.24	174.67	171.93	162.04	109.45	116.17	110.38
Period 19	179.84	177.72	177.28	174.50	153.20	104.49	110.33	105.37
Period 20	188.32	176.70	180.31	177.48	153.52	109.10	114.29	109.97

Table D.4: The detailed movement of invested \$100 for last 10 days period (Period 20, Stock: BIDU)

Day	Buy&Sell decisions by prediction models								Buy&Hold stock/index			
	MV-GPR		MV-TPR		GPR		TPR		BIDU	INDU	NDX	SPX
	Decision	Dollar	Decision	Dollar	Decision	Dollar	Decision	Dollar				
190	260.24		223.38		264.59		230.23		134.23	104.49	110.33	105.37
191	Buy	263.29	Buy	226.00	Buy	267.69	Buy	232.93	136.60	106.25	112.37	107.51
192	Keep	272.74	Keep	234.11	Keep	277.29	Keep	241.29	141.50	108.83	115.14	110.09
193	Keep	275.50	Keep	236.48	Keep	280.10	Keep	243.73	142.94	108.99	115.52	110.60
194	Keep	275.94	Keep	236.85	Keep	280.55	Keep	244.12	143.16	109.94	115.84	111.02
195	Sell	275.70	Sell	236.65	Sell	280.31	Sell	243.91	142.28	110.33	115.45	111.21
196	Buy	273.26	Buy	234.55	Buy	277.83	Buy	241.75	140.95	110.37	115.55	111.20
197	Keep	277.87	Keep	238.52	Keep	282.52	Keep	245.83	143.33	110.51	116.39	111.56
198	Keep	272.35	Keep	233.77	Keep	276.90	Keep	240.94	140.48	110.42	116.35	111.66
199	Sell	270.29	Keep	233.57	Sell	274.81	Sell	239.13	140.36	110.08	115.53	111.11
200	Keep	270.29	Sell	233.29	Keep	274.81	Keep	239.13	139.12	109.10	114.29	109.97

TABLE D.5: The detailed movement of invested $100 for last 10 days period (Period 20, Stock: CTRP)

	Buy&Sell decisions by prediction models								Buy&Hold stock/index			
Day	MV-GPR		MV-TPR		GPR		TPR		CTRP	INDU	NDX	SPX
	Decision	Dollar	Decision	Dollar	Decision	Dollar	Decision	Dollar				
190	169.89		201.99		163.80		176.73		80.69	104.49	110.33	105.37
191	Buy	175.02	Buy	208.09	Buy	168.75	Buy	182.07	83.65	106.25	112.37	107.51
192	Keep	180.77	Keep	214.92	Keep	174.28	Keep	188.05	86.39	108.83	115.14	110.09
193	Keep	185.40	Keep	220.43	Keep	178.75	Keep	192.87	88.61	108.99	115.52	110.60
194	Sell	184.91	Sell	220.28	Sell	178.28	Sell	192.36	88.55	109.94	115.84	111.02
195	Keep	184.91	Buy	222.82	Keep	178.28	Keep	192.36	88.45	110.33	115.45	111.21
196	Keep	184.91	Keep	222.82	Keep	178.28	Keep	192.36	88.22	110.37	115.55	111.20
197	Buy	185.16	Keep	222.82	Buy	178.52	Buy	192.62	88.92	110.51	116.39	111.56
198	Keep	184.06	Keep	222.82	Keep	177.46	Keep	191.47	88.39	110.42	116.35	111.66
199	Sell	183.25	Keep	222.82	Sell	176.67	Sell	190.63	88.45	110.08	115.53	111.11
200	Keep	183.25	Keep	222.82	Keep	176.67	Keep	190.63	89.20	109.10	114.29	109.97

TABLE D.6: The detailed movement of invested $100 for last 10 days period (Period 20, Stock: NTES)

	Buy&Sell decisions by prediction models								Buy&Hold stock/index			
Day	MV-GPR		MV-TPR		GPR		TPR		NTES	INDU	NDX	SPX
	Decision	Dollar	Decision	Dollar	Decision	Dollar	Decision	Dollar				
190	179.84		177.72		177.28		174.50		153.20	104.49	110.33	105.37
191	Buy	176.57	Buy	174.49	Buy	174.06	Buy	171.33	151.35	106.25	112.37	107.51
192	Keep	184.84	Keep	182.66	Keep	182.21	Keep	179.36	158.44	108.83	115.14	110.09
193	Keep	184.16	Keep	181.98	Keep	181.54	Keep	178.69	157.86	108.99	115.52	110.60
194	Keep	185.46	Keep	183.27	Keep	182.82	Keep	179.95	158.97	109.94	115.84	111.02
195	Sell	185.33	Keep	179.25	Sell	182.69	Sell	179.83	155.49	110.33	115.45	111.21
196	Buy	187.22	keep	180.86	Buy	184.55	Buy	181.66	156.88	110.37	115.55	111.20
197	Keep	187.75	Keep	181.38	Keep	185.08	Keep	182.18	157.33	110.51	116.39	111.56
198	Sell	188.32	Keep	177.52	Keep	181.15	Keep	178.30	153.98	110.42	116.35	111.66
199	Keep	188.32	Sell	176.70	Sell	180.31	Sell	177.48	153.29	110.08	115.53	111.11
200	Keep	188.32	Keep	176.70	Keep	180.31	Keep	177.48	153.52	109.10	114.29	109.97

Appendix E

Investment details of the stocks listed in Dow 30

E.1 Final stock investment details

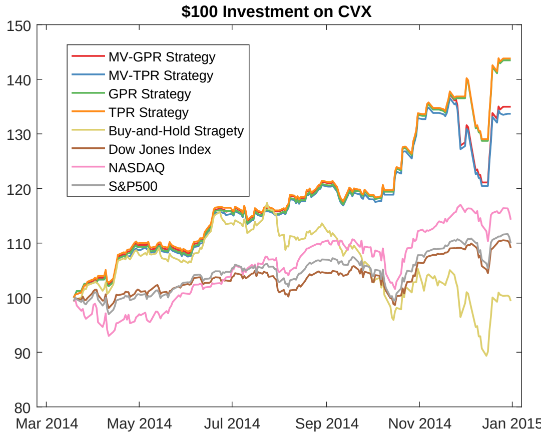
Table E.1: The detailed stock investment results under different strategies

Ticker	Industry	Buy&Sell Strategy				Buy&Hold Stragegy			
		MV-GPR	MV-TPR	GPR	TPR	Stock	INDU	NDX	SPX
CVX	Oil & Gas	134.97	133.69	143.47	143.81	99.38	109.10	114.29	109.97
XOM	Oil & Gas	128.39	132.72	131.31	136.02	99.74			
MMM	Industrials	166.76	162.96	167.12	162.65	125.96			
BA	Industrials	160.12	159.98	158.60	157.38	106.39			
CAT	Industrials	142.58	138.45	146.13	151.75	97.16			
GE	Industrials	137.51	134.63	135.35	139.72	101.15			
UTX	Industrials	144.29	139.47	143.29	145.18	101.94			
KO	Consumer Goods	128.11	128.59	124.88	124.52	112.47			
MCD	Consumer Goods	120.69	117.09	122.19	119.59	98.81			
PG	Consumer Goods	126.62	123.32	127.14	127.10	117.04			
JNJ	Health Care	146.00	146.70	147.42	145.16	113.65			
MRK	Health Care	129.40	134.48	129.36	135.05	102.45			
PFE	Health Care	128.60	136.53	130.26	134.48	100.16			
UNH	Health Care	164.98	164.63	166.14	162.79	131.14			
HD	Consumer Services	171.46	165.74	169.55	170.18	133.33			
NKE	Consumer Services	147.17	146.13	142.36	148.26	122.27			
WMT	Consumer Services	136.50	132.59	133.77	135.67	117.31			
DIS	Consumer Services	168.19	168.43	168.51	168.12	115.97			
AXP	Financials	160.39	158.52	160.73	160.12	102.34			
GS	Financials	170.46	171.29	167.71	165.72	116.16			
JPM	Financials	174.90	170.07	176.48	172.12	110.09			
TRV	Financials	149.81	145.88	150.70	145.71	128.18			
V	Financials	161.50	153.48	157.04	158.70	116.37			
AAPL	Technology	201.82	206.64	203.45	208.07	147.34			
CSCO	Technology	159.13	164.88	158.61	155.92	131.34			
IBM	Technology	116.10	128.79	124.92	123.74	88.06			
INTC	Technology	183.80	179.45	185.52	188.22	149.24			
MSFT	Technology	173.61	166.09	176.57	172.76	120.01			

Table E.2: The detailed industry portfolio investment results under different strategies

Industry Portfolio	Buy&Sell Strategy				Buy&Hold Stragegy			
	MV-GPR	MV-TPR	GPR	TPR	Stock	INDU	NDX	SPX
Oil & Gas	131.68	133.20	137.39	139.92	99.56	109.10	114.29	109.97
Industrials	150.25	147.10	150.10	151.34	106.52			
Consumer Goods	125.14	123.00	124.73	123.73	109.44			
Health Care	142.24	145.59	143.30	144.37	111.85			
Consumer Services	155.83	153.22	153.55	155.56	122.22			
Financials	163.41	159.85	162.53	160.47	114.63			
Technology	166.89	169.17	169.81	169.74	127.20			

E.2 Stock investment on diverse sectors



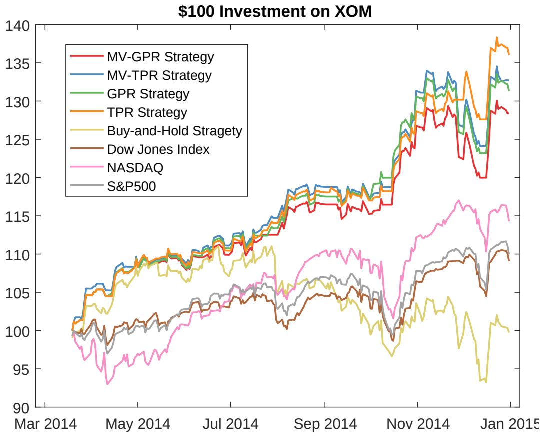


Figure E.1: Stock investment in Oil & Gas sector

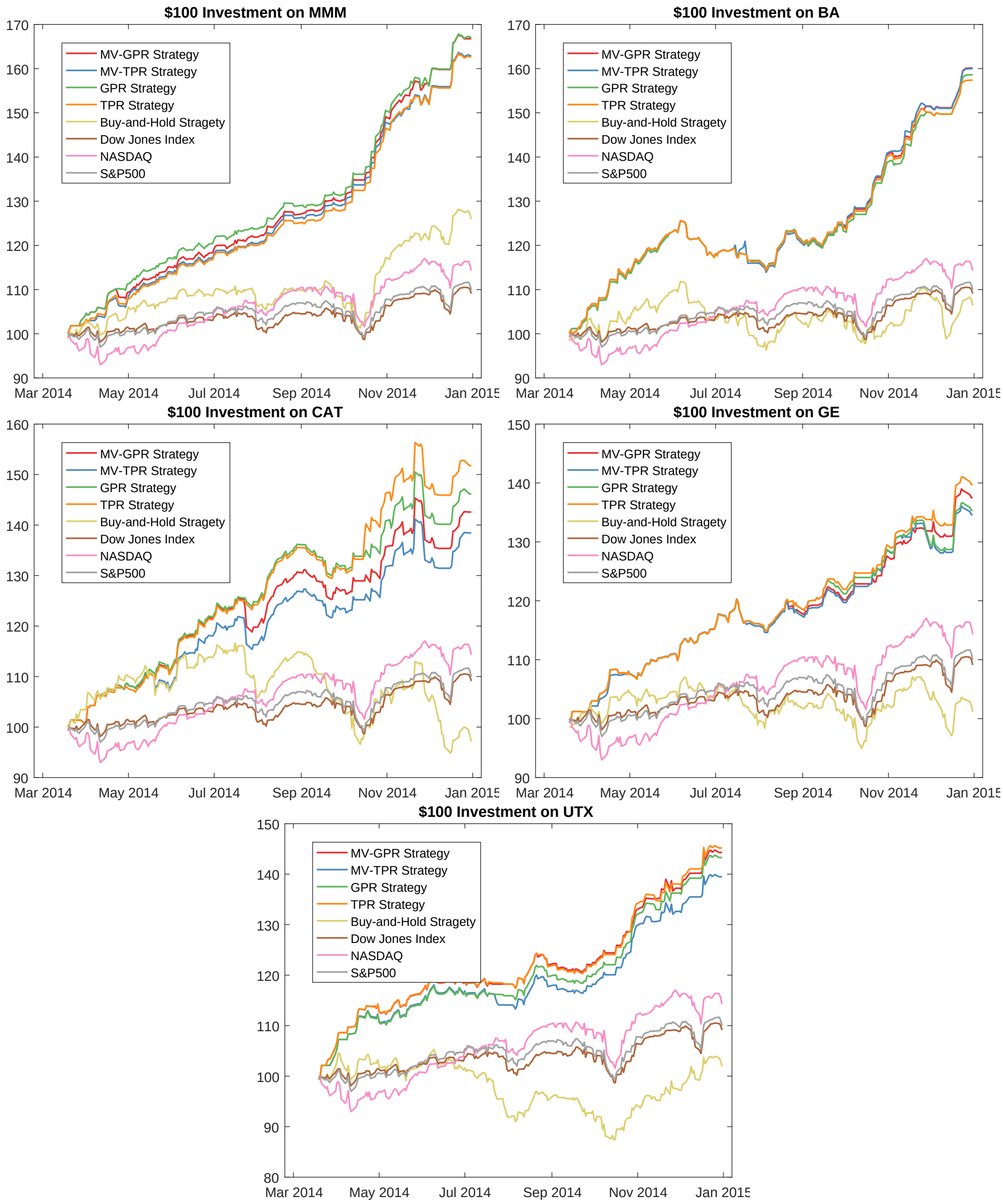


FIGURE E.2: Stock investment in Industrials sector

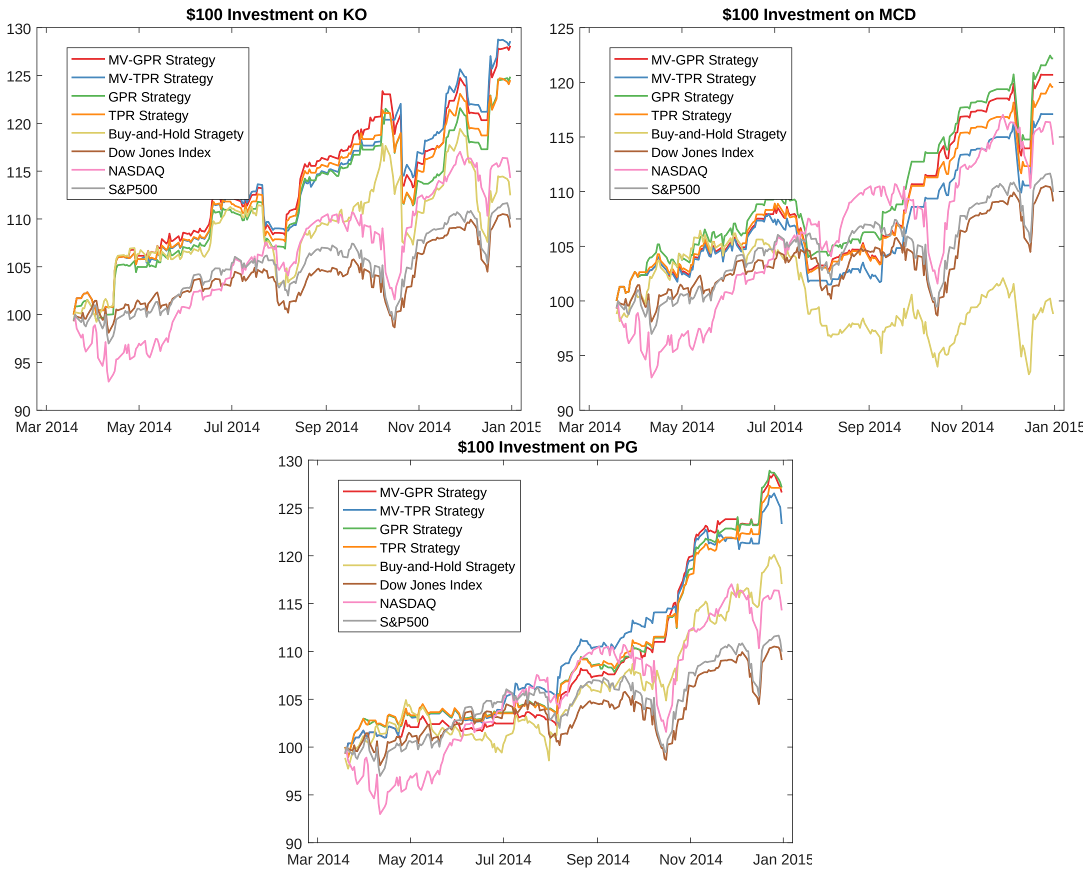


Figure E.3: Stock investment in Consumer Goods sector

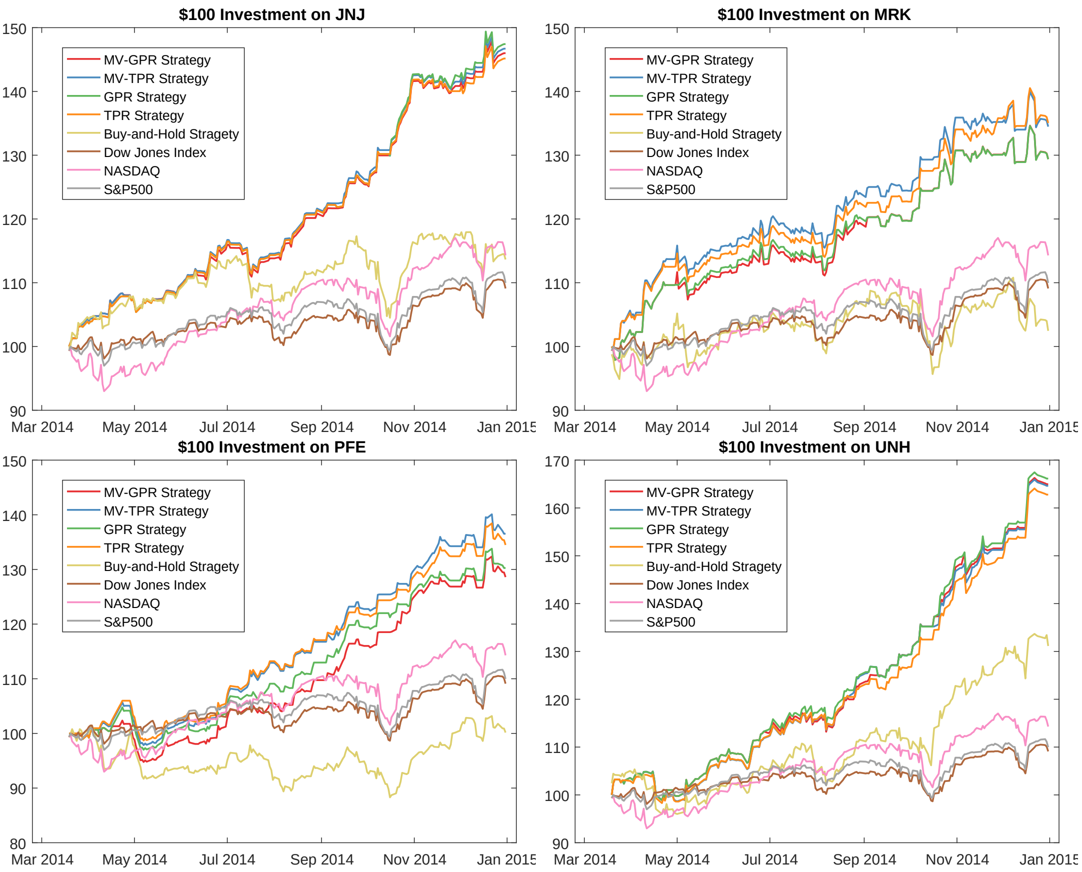


FIGURE E.4: Stock investment in Health Care sector

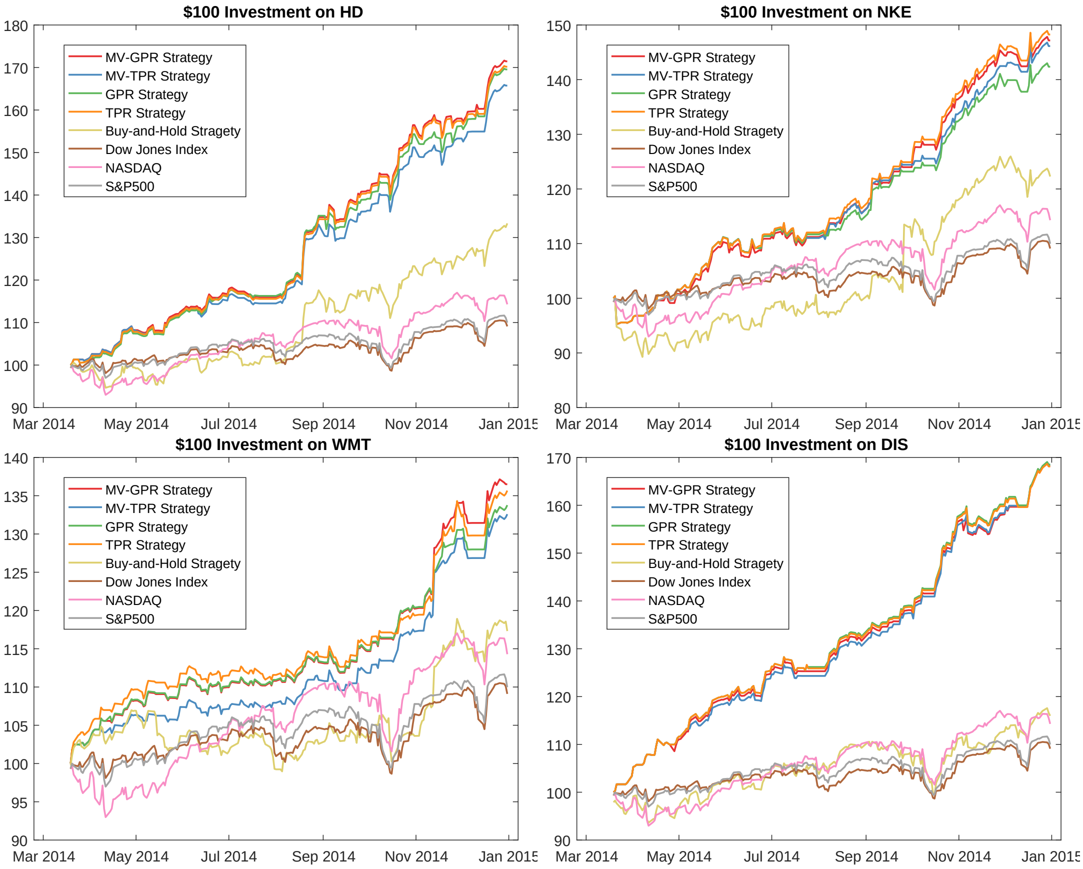


Figure E.5: Stock investment in Consumer Services sector

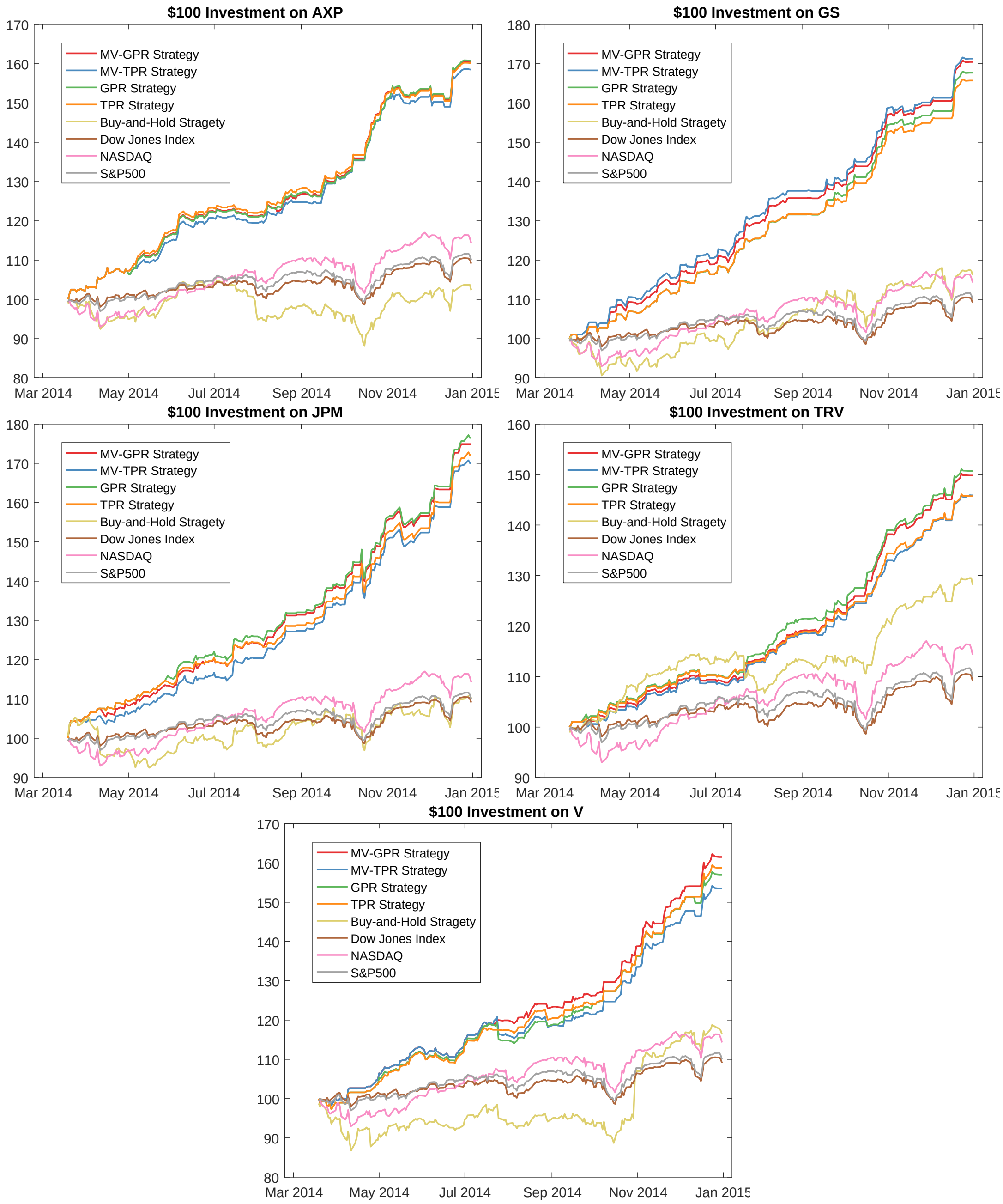


FIGURE E.6: Stock investment in Financials sector

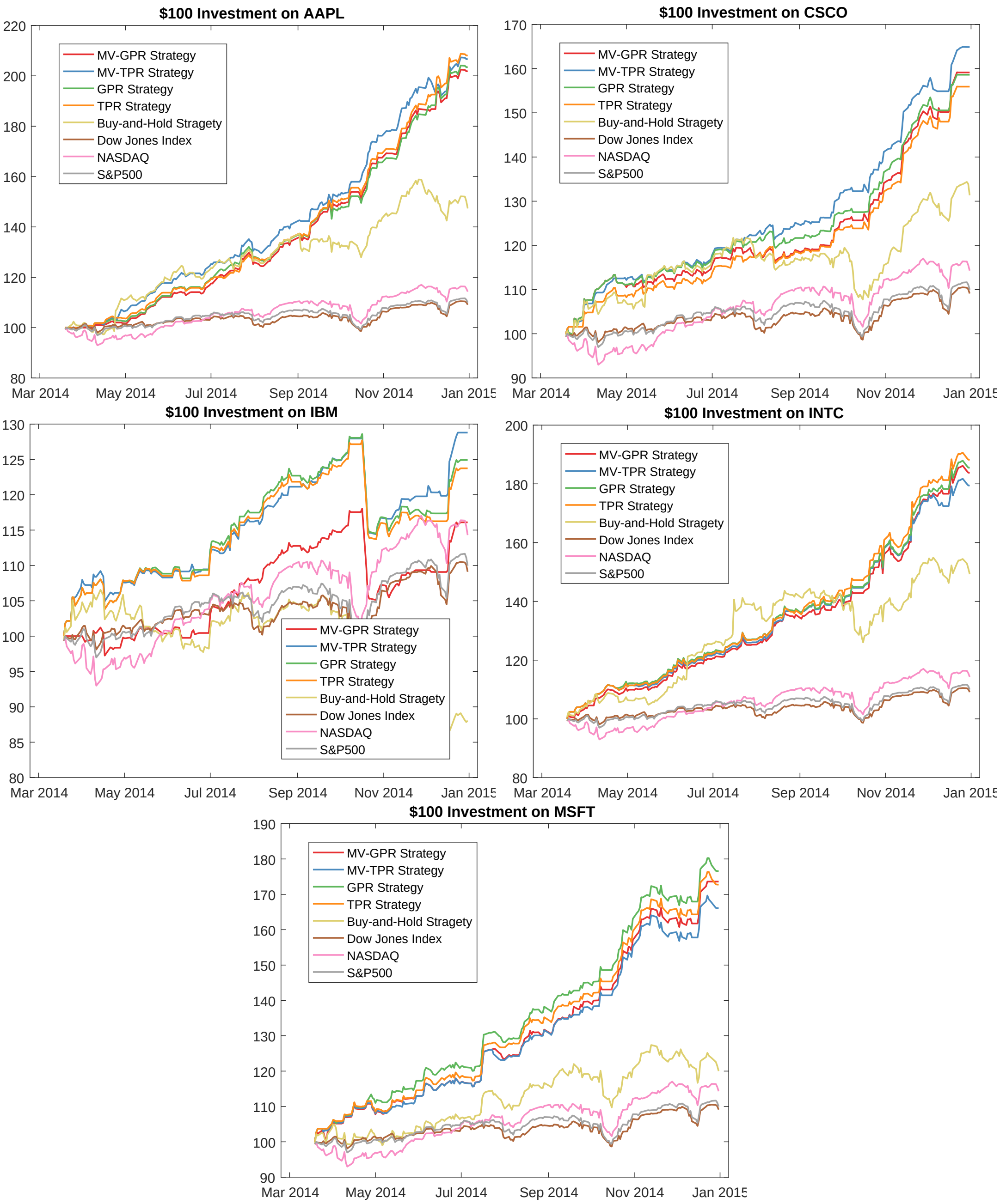


FIGURE E.7: Stock investment in Technology sector

E.3 The details of stock investments in the period

Table E.3: The movement of invested $100 for 200 days split in to 20 periods (Stock: CVX; Industry: Oil & Gas)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	CVX	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	103.02	102.70	102.70	103.35	102.69	101.20	98.70	100.71
Period 2	104.55	104.23	104.23	104.89	103.81	99.55	94.10	98.44
Period 3	109.55	108.65	109.22	109.91	108.32	101.50	96.64	100.62
Period 4	109.78	109.20	109.44	110.13	109.10	101.70	96.94	100.87
Period 5	108.18	107.61	107.85	108.53	106.46	102.22	100.79	102.55
Period 6	110.53	109.95	110.32	110.89	109.76	102.44	101.54	103.09
Period 7	116.10	115.49	115.88	116.47	113.94	103.12	103.25	104.54
Period 8	114.07	113.47	113.67	114.25	111.81	103.72	105.34	105.09
Period 9	116.42	115.80	116.01	116.60	116.25	103.82	106.98	105.67
Period 10	118.64	118.02	118.23	118.84	111.28	101.33	104.90	103.17
Period 11	119.04	118.42	118.63	119.24	111.57	104.07	109.34	106.20
Period 12	120.06	119.43	119.64	120.25	110.78	104.75	110.49	106.91
Period 13	119.61	118.98	119.52	119.80	108.39	105.12	109.57	106.52
Period 14	118.33	117.71	118.24	118.52	103.65	104.01	108.35	104.94
Period 15	123.38	122.73	123.29	123.58	97.86	100.39	104.41	101.70
Period 16	133.41	132.71	133.31	133.62	102.50	106.31	112.48	107.77
Period 17	134.28	133.57	134.18	134.50	102.54	108.03	113.68	109.03
Period 18	131.59	130.90	139.87	140.21	101.01	109.45	116.17	110.38
Period 19	121.10	120.47	128.73	129.03	90.09	104.49	110.33	105.37
Period 20	134.97	133.69	143.47	143.81	99.38	109.10	114.29	109.97

Table E.4: The movement of invested $100 for 200 days split in to 20 periods (Stock: XOM; Industry: Oil & Gas)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	104.59	105.54	104.59	104.59	103.25	101.20	98.70	100.71
Period 2	105.66	106.42	105.75	105.66	104.26	99.55	94.10	98.44
Period 3	109.34	109.69	109.45	109.34	108.20	101.50	96.64	100.62
Period 4	108.87	109.27	109.03	109.41	108.81	101.70	96.94	100.87
Period 5	109.36	109.70	109.53	109.91	107.72	102.22	100.79	102.55
Period 6	109.57	110.08	109.73	110.12	108.14	102.44	101.54	103.09
Period 7	110.81	112.03	111.67	111.37	108.53	103.12	103.25	104.54
Period 8	109.80	111.00	110.65	110.35	108.22	103.72	105.34	105.09
Period 9	112.53	113.77	112.41	113.09	109.75	103.82	106.98	105.67
Period 10	116.15	118.43	117.02	117.73	106.09	101.33	104.90	103.17
Period 11	115.42	117.68	116.44	116.98	105.50	104.07	109.34	106.20
Period 12	116.48	118.76	117.51	118.23	104.72	104.75	110.49	106.91
Period 13	115.89	118.16	117.84	117.63	103.40	105.12	109.57	106.52
Period 14	115.73	118.00	119.22	117.47	101.24	104.01	108.35	104.94
Period 15	120.63	123.00	124.28	122.44	98.30	100.39	104.41	101.70
Period 16	126.54	131.10	130.37	128.44	102.03	106.31	112.48	107.77
Period 17	127.07	131.90	130.91	128.98	102.61	108.03	113.68	109.03
Period 18	124.86	129.15	128.18	132.78	101.62	109.45	116.17	110.38
Period 19	119.98	124.11	123.18	127.60	93.22	104.49	110.33	105.37
Period 20	128.39	132.72	131.31	136.02	99.74	109.10	114.29	109.97

Table E.5: The movement of invested $100 for 200 days split in to 20 periods (Stock: MMM; Industry: Industrials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	104.64	103.24	104.64	103.24	102.85	101.20	98.70	100.71
Period 2	106.35	104.41	106.35	104.93	101.01	99.55	94.10	98.44
Period 3	109.17	107.19	111.07	107.72	104.78	101.50	96.64	100.62
Period 4	112.25	111.18	114.21	110.76	106.67	101.70	96.94	100.87
Period 5	115.04	113.94	117.04	113.50	107.92	102.22	100.79	102.55
Period 6	116.85	115.73	118.88	115.29	108.44	102.44	101.54	103.09
Period 7	117.79	116.66	119.84	116.22	108.94	103.12	103.25	104.54
Period 8	119.61	118.46	121.69	118.01	109.37	103.72	105.34	105.09
Period 9	121.67	120.23	123.52	119.79	109.99	103.82	106.98	105.67
Period 10	124.33	122.85	126.22	122.40	106.75	101.33	104.90	103.17
Period 11	127.57	126.76	129.51	125.60	109.89	104.07	109.34	106.20
Period 12	127.99	126.95	129.23	125.78	110.47	104.75	110.49	106.91
Period 13	130.16	129.10	131.43	127.91	111.09	105.12	109.57	106.52
Period 14	132.06	130.99	133.35	129.78	107.00	104.01	108.35	104.94
Period 15	136.59	135.48	137.92	134.23	104.91	100.39	104.41	101.70
Period 16	148.64	147.43	150.09	146.07	116.88	106.31	112.48	107.77
Period 17	153.86	150.83	155.73	151.57	120.99	108.03	113.68	109.03
Period 18	158.35	154.50	158.45	154.21	123.10	109.45	116.17	110.38
Period 19	159.79	155.92	159.90	155.62	121.92	104.49	110.33	105.37
Period 20	166.76	162.96	167.12	162.65	125.96	109.10	114.29	109.97

Table E.6: The movement of invested $100 for 200 days split in to 20 periods (Stock: BA; Industry: Industrials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	105.88	105.41	105.41	105.88	103.13	101.20	98.70	100.71
Period 2	109.08	108.60	108.60	109.08	99.96	99.55	94.10	98.44
Period 3	114.65	114.14	114.14	114.65	103.78	101.50	96.64	100.62
Period 4	119.39	118.86	118.86	119.39	107.57	101.70	96.94	100.87
Period 5	122.74	122.73	122.73	122.74	109.31	102.22	100.79	102.55
Period 6	121.55	121.55	121.55	121.55	106.93	102.44	101.54	103.09
Period 7	118.08	118.08	118.08	118.08	103.55	103.12	103.25	104.54
Period 8	118.79	118.79	118.79	118.79	103.61	103.72	105.34	105.09
Period 9	116.59	115.99	116.58	116.59	99.66	103.82	106.98	105.67
Period 10	115.87	115.27	115.87	115.70	98.17	101.33	104.90	103.17
Period 11	123.14	122.51	123.14	122.96	103.73	104.07	109.34	106.20
Period 12	120.98	121.24	120.99	121.69	104.15	104.75	110.49	106.91
Period 13	122.27	122.54	122.29	122.99	104.66	105.12	109.57	106.52
Period 14	126.96	127.24	125.80	126.51	102.75	104.01	108.35	104.94
Period 15	130.54	130.83	129.34	130.08	101.16	100.39	104.41	101.70
Period 16	141.01	141.32	139.24	140.51	102.56	106.31	112.48	107.77
Period 17	144.35	145.51	142.53	143.84	105.12	108.03	113.68	109.03
Period 18	151.24	151.11	149.80	149.83	108.28	109.45	116.17	110.38
Period 19	151.14	151.01	149.70	149.73	101.70	104.49	110.33	105.37
Period 20	160.12	159.98	158.60	157.38	106.39	109.10	114.29	109.97

Table E.7: The movement of invested $100 for 200 days split in to 20 periods (Stock: CAT; Industry: Industrials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	101.17	101.17	101.17	101.17	103.87	101.20	98.70	100.71
Period 2	106.53	106.53	106.53	106.06	106.67	99.55	94.10	98.44
Period 3	108.65	108.65	108.65	107.92	110.33	101.50	96.64	100.62
Period 4	109.95	110.43	110.43	109.70	111.52	101.70	96.94	100.87
Period 5	112.13	108.88	112.39	111.87	108.45	102.22	100.79	102.55
Period 6	117.88	114.46	118.15	117.60	111.64	102.44	101.54	103.09
Period 7	121.20	117.69	121.48	120.91	113.60	103.12	103.25	104.54
Period 8	123.41	119.83	123.69	123.12	115.11	103.72	105.34	105.09
Period 9	119.63	116.16	124.63	124.08	110.45	103.82	106.98	105.67
Period 10	124.35	120.75	129.55	128.98	108.80	101.33	104.90	103.17
Period 11	128.60	124.87	133.98	133.39	113.04	104.07	109.34	106.20
Period 12	129.80	126.04	134.71	134.11	113.68	104.75	110.49	106.91
Period 13	125.31	121.68	130.05	129.47	106.29	105.12	109.57	106.52
Period 14	126.81	123.13	131.60	131.02	103.24	104.01	108.35	104.94
Period 15	128.68	124.96	133.55	138.75	99.81	100.39	104.41	101.70
Period 16	135.91	131.97	141.05	146.54	106.38	106.31	112.48	107.77
Period 17	138.87	134.85	143.77	149.37	107.95	108.03	113.68	109.03
Period 18	136.56	132.61	141.39	147.21	105.63	109.45	116.17	110.38
Period 19	135.37	131.45	140.15	145.93	94.83	104.49	110.33	105.37
Period 20	142.58	138.45	146.13	151.75	97.16	109.10	114.29	109.97

Table E.8: The movement of invested $100 for 200 days split in to 20 periods (Stock: GE; Industry: Industrials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	101.08	101.08	101.08	101.08	100.90	101.20	98.70	100.71
Period 2	105.25	104.37	105.25	105.25	100.70	99.55	94.10	98.44
Period 3	107.98	108.05	107.98	107.98	104.88	101.50	96.64	100.62
Period 4	109.26	109.33	109.26	109.26	104.37	101.70	96.94	100.87
Period 5	110.64	110.71	110.64	110.64	104.29	102.22	100.79	102.55
Period 6	113.02	113.09	113.02	113.02	105.15	102.44	101.54	103.09
Period 7	114.54	114.61	114.54	114.54	103.38	103.12	103.25	104.54
Period 8	117.93	118.01	117.93	117.93	104.40	103.72	105.34	105.09
Period 9	116.59	116.22	116.59	116.59	101.41	103.82	106.98	105.67
Period 10	115.87	115.50	115.87	115.87	100.90	101.33	104.90	103.17
Period 11	118.96	118.57	119.80	119.80	102.83	104.07	109.34	106.20
Period 12	119.25	118.82	120.07	120.07	102.56	104.75	110.49	106.91
Period 13	121.88	121.45	122.96	123.71	103.42	105.12	109.57	106.52
Period 14	121.56	121.13	122.63	123.38	100.01	104.01	108.35	104.94
Period 15	122.87	123.00	123.96	124.71	99.26	100.39	104.41	101.70
Period 16	127.04	128.07	128.16	128.95	101.91	106.31	112.48	107.77
Period 17	130.43	130.82	131.73	132.33	105.52	108.03	113.68	109.03
Period 18	131.74	128.98	129.48	133.66	103.30	109.45	116.17	110.38
Period 19	130.96	128.22	128.71	132.86	97.12	104.49	110.33	105.37
Period 20	137.51	134.63	135.35	139.72	101.15	109.10	114.29	109.97

Table E.9: The movement of invested $100 for 200 days split in to 20 periods (Stock: UTX; Industry: Industrials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	106.77	105.45	105.45	106.77	102.86	101.20	98.70	100.71
Period 2	110.74	109.37	109.37	110.74	101.05	99.55	94.10	98.44
Period 3	114.19	112.29	112.07	114.29	103.22	101.50	96.64	100.62
Period 4	114.98	113.06	112.84	115.07	102.85	101.70	96.94	100.87
Period 5	116.28	114.34	114.12	116.38	102.00	102.22	100.79	102.55
Period 6	118.56	116.58	116.36	118.66	102.44	102.44	101.54	103.09
Period 7	118.66	116.68	116.46	118.76	101.47	103.12	103.25	104.54
Period 8	118.24	116.27	116.05	118.34	100.05	103.72	105.34	105.09
Period 9	118.23	114.10	116.19	118.49	95.45	103.82	106.98	105.67
Period 10	119.48	115.30	117.16	119.47	92.89	101.33	104.90	103.17
Period 11	123.59	119.27	121.60	124.01	96.44	104.07	109.34	106.20
Period 12	121.56	117.31	119.27	121.23	95.74	104.75	110.49	106.91
Period 13	120.67	116.45	118.39	120.34	93.87	105.12	109.57	106.52
Period 14	123.80	119.47	121.46	123.46	91.88	104.01	108.35	104.94
Period 15	125.85	121.46	123.48	125.52	89.47	100.39	104.41	101.70
Period 16	133.37	130.22	132.39	134.57	93.72	106.31	112.48	107.77
Period 17	135.29	130.77	132.92	134.67	95.18	108.03	113.68	109.03
Period 18	137.00	132.41	136.04	137.84	97.22	109.45	116.17	110.38
Period 19	140.18	135.49	139.20	141.04	100.37	104.49	110.33	105.37
Period 20	144.29	139.47	143.29	145.18	101.94	109.10	114.29	109.97

Table E.10: The movement of invested $100 for 200 days split in to 20 periods (Stock: KO; Industry: Consumer Goods)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	100.92	100.92	100.14	100.92	100.13	101.20	98.70	100.71
Period 2	104.58	104.58	103.78	104.58	104.74	99.55	94.10	98.44
Period 3	106.17	105.82	105.00	105.82	106.33	101.50	96.64	100.62
Period 4	106.94	106.20	105.46	106.28	106.60	101.70	96.94	100.87
Period 5	107.91	107.16	106.41	107.24	106.00	102.22	100.79	102.55
Period 6	108.74	107.98	107.23	108.07	106.16	102.44	101.54	103.09
Period 7	112.52	111.74	110.96	111.83	110.39	103.12	103.25	104.54
Period 8	111.38	111.73	109.93	110.70	110.23	103.72	105.34	105.09
Period 9	109.23	109.70	107.80	108.55	107.69	103.82	106.98	105.67
Period 10	110.85	109.06	109.40	110.16	103.61	101.33	104.90	103.17
Period 11	115.55	113.35	114.04	114.83	108.00	104.07	109.34	106.20
Period 12	116.34	114.96	114.49	115.62	109.73	104.75	110.49	106.91
Period 13	119.24	117.13	116.71	117.81	111.70	105.12	109.57	106.52
Period 14	120.70	118.09	117.67	118.78	115.35	104.01	108.35	104.94
Period 15	120.92	122.04	118.90	118.99	114.53	100.39	104.41	101.70
Period 16	115.74	116.82	113.81	114.92	110.61	106.31	112.48	107.77
Period 17	118.01	119.48	115.03	117.97	113.55	108.03	113.68	109.03
Period 18	123.93	124.84	120.80	122.29	118.65	109.45	116.17	110.38
Period 19	120.33	121.21	117.29	118.73	107.60	104.49	110.33	105.37
Period 20	128.11	128.59	124.88	124.52	112.47	109.10	114.29	109.97

Table E.11: The movement of invested $100 for 200 days split in to 20 periods (Stock: MCD; Industry: Consumer Goods)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	102.52	102.52	102.52	102.52	100.65	101.20	98.70	100.71
Period 2	104.00	104.00	105.18	104.28	103.66	99.55	94.10	98.44
Period 3	103.01	102.84	104.18	103.29	104.23	101.50	96.64	100.62
Period 4	104.77	104.59	105.96	105.05	105.92	101.70	96.94	100.87
Period 5	104.65	104.47	105.93	105.02	105.02	102.22	100.79	102.55
Period 6	104.91	104.73	106.19	105.28	103.39	102.44	101.54	103.09
Period 7	107.94	107.76	109.26	108.32	105.20	103.12	103.25	104.54
Period 8	107.85	106.80	109.17	107.64	104.02	103.72	105.34	105.09
Period 9	102.87	101.86	104.13	102.67	99.20	103.82	106.98	105.67
Period 10	103.82	102.00	105.09	103.62	96.95	101.33	104.90	103.17
Period 11	104.70	102.92	105.64	104.55	97.88	104.07	109.34	106.20
Period 12	103.46	101.71	105.58	103.32	96.69	104.75	110.49	106.91
Period 13	106.84	104.84	108.83	106.70	98.18	105.12	109.57	106.52
Period 14	110.68	108.60	112.74	110.53	98.09	104.01	108.35	104.94
Period 15	114.17	110.77	114.99	112.73	95.74	100.39	104.41	101.70
Period 16	116.89	113.40	117.72	115.41	97.85	106.31	112.48	107.77
Period 17	117.30	113.80	118.14	115.82	100.32	108.03	113.68	109.03
Period 18	118.37	114.85	119.22	116.68	100.30	109.45	116.17	110.38
Period 19	113.96	110.56	114.77	112.33	93.56	104.49	110.33	105.37
Period 20	120.69	117.09	122.19	119.59	98.81	109.10	114.29	109.97

Table E.12: The movement of invested $100 for 200 days split in to 20 periods (Stock: PG; Industry: Consumer Goods)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	101.40	101.40	102.95	102.64	100.83	101.20	98.70	100.71
Period 2	101.01	101.01	102.15	102.24	101.46	99.55	94.10	98.44
Period 3	102.30	103.50	103.45	103.55	104.43	101.50	96.64	100.62
Period 4	102.41	103.50	103.56	103.65	102.68	101.70	96.94	100.87
Period 5	102.33	103.43	103.48	103.58	101.71	102.22	100.79	102.55
Period 6	101.87	102.77	103.01	102.92	100.90	102.44	101.54	103.09
Period 7	102.46	103.38	103.62	103.53	99.46	103.12	103.25	104.54
Period 8	103.08	105.99	104.24	104.15	102.67	103.72	105.34	105.09
Period 9	103.34	106.26	104.50	104.41	101.45	103.82	106.98	105.67
Period 10	105.43	108.41	106.61	106.52	103.22	101.33	104.90	103.17
Period 11	108.04	111.09	109.25	109.16	106.33	104.07	109.34	106.20
Period 12	107.60	110.33	108.32	108.72	106.25	104.75	110.49	106.91
Period 13	109.44	112.06	109.46	110.01	108.15	105.12	109.57	106.52
Period 14	110.17	113.23	110.60	110.71	106.56	104.01	108.35	104.94
Period 15	114.42	114.45	113.58	113.69	107.34	100.39	104.41	101.70
Period 16	120.01	119.60	118.69	118.15	112.28	106.31	112.48	107.77
Period 17	122.40	121.14	121.44	120.51	112.87	108.03	113.68	109.03
Period 18	123.83	121.84	124.05	123.11	117.02	109.45	116.17	110.38
Period 19	123.26	121.27	123.47	122.25	114.82	104.49	110.33	105.37
Period 20	126.62	123.32	127.14	127.10	117.04	109.10	114.29	109.97

Table E.13: The movement of invested $100 for 200 days split in to 20 periods (Stock: JNJ; Industry: Health Care)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	104.06	104.06	103.60	103.60	104.31	101.20	98.70	100.71
Period 2	107.25	107.25	106.77	106.77	105.66	99.55	94.10	98.44
Period 3	107.67	107.63	107.49	107.49	107.88	101.50	96.64	100.62
Period 4	107.43	107.39	107.25	107.25	107.43	101.70	96.94	100.87
Period 5	108.55	108.58	108.36	108.36	108.07	102.22	100.79	102.55
Period 6	111.13	111.84	111.62	111.62	109.96	102.44	101.54	103.09
Period 7	115.09	115.82	115.59	115.59	113.31	103.12	103.25	104.54
Period 8	114.45	115.18	114.95	114.95	112.72	103.72	105.34	105.09
Period 9	113.25	113.97	113.74	113.74	109.51	103.82	106.98	105.67
Period 10	116.15	116.89	116.65	116.65	108.41	101.33	104.90	103.17
Period 11	120.15	120.92	120.68	120.68	111.32	104.07	109.34	106.20
Period 12	121.67	122.45	122.21	122.21	112.35	104.75	110.49	106.91
Period 13	125.59	126.39	125.80	125.80	116.48	105.12	109.57	106.52
Period 14	127.29	128.10	127.50	127.50	113.22	104.01	108.35	104.94
Period 15	132.15	133.00	133.15	132.37	107.11	100.39	104.41	101.70
Period 16	141.65	142.56	142.72	141.89	116.03	106.31	112.48	107.77
Period 17	140.65	141.55	141.71	140.88	116.94	108.03	113.68	109.03
Period 18	141.53	142.21	142.91	140.72	117.93	109.45	116.17	110.38
Period 19	143.08	143.77	144.48	142.26	111.68	104.49	110.33	105.37
Period 20	146.00	146.70	147.42	145.16	113.65	109.10	114.29	109.97

Table E.14: The movement of invested $100 for 200 days split in to 20 periods (Stock: MRK; Industry: Health Care)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	101.86	104.90	101.86	104.53	99.45	101.20	98.70	100.71
Period 2	106.83	110.02	106.83	109.63	98.89	99.55	94.10	98.44
Period 3	109.70	113.78	109.70	112.57	103.32	101.50	96.64	100.62
Period 4	108.89	112.94	109.67	111.75	99.45	101.70	96.94	100.87
Period 5	110.89	115.02	111.69	113.80	101.80	102.22	100.79	102.55
Period 6	112.78	116.97	113.59	115.73	103.63	102.44	101.54	103.09
Period 7	114.80	119.07	115.63	117.81	104.04	103.12	103.25	104.54
Period 8	114.38	118.86	115.20	117.37	103.88	103.72	105.34	105.09
Period 9	113.60	118.06	114.42	116.58	103.37	103.82	106.98	105.67
Period 10	112.99	117.43	113.80	115.96	100.52	101.33	104.90	103.17
Period 11	117.90	122.53	118.74	120.99	105.18	104.07	109.34	106.20
Period 12	120.29	125.01	120.26	122.53	108.52	104.75	110.49	106.91
Period 13	120.52	125.25	120.49	123.54	108.47	105.12	109.57	106.52
Period 14	121.70	126.47	121.66	124.74	106.68	104.01	108.35	104.94
Period 15	124.87	129.77	124.83	127.99	96.78	100.39	104.41	101.70
Period 16	130.77	135.90	130.73	134.04	105.50	106.31	112.48	107.77
Period 17	130.43	135.55	130.39	133.70	106.47	108.03	113.68	109.03
Period 18	130.40	135.51	130.36	136.09	108.83	109.45	116.17	110.38
Period 19	128.96	134.02	128.92	134.59	102.47	104.49	110.33	105.37
Period 20	129.40	134.48	129.36	135.05	102.45	109.10	114.29	109.97

Table E.15: The movement of invested $100 for 200 days split in to 20 periods (Stock: PFE; Industry: Health Care)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	99.20	99.20	99.20	99.20	100.06	101.20	98.70	100.71
Period 2	100.46	101.34	100.46	101.62	93.61	99.55	94.10	98.44
Period 3	99.26	102.49	101.60	103.38	97.96	101.50	96.64	100.62
Period 4	95.17	98.26	97.42	99.12	91.95	101.70	96.94	100.87
Period 5	98.35	101.55	100.68	102.44	93.53	102.22	100.79	102.55
Period 6	98.07	101.68	100.39	102.14	93.06	102.44	101.54	103.09
Period 7	100.67	104.37	103.05	104.86	93.50	103.12	103.25	104.54
Period 8	104.00	107.83	106.46	108.33	95.01	103.72	105.34	105.09
Period 9	103.67	110.96	107.20	111.24	95.39	103.82	106.98	105.67
Period 10	105.54	112.44	108.63	112.73	90.34	101.33	104.90	103.17
Period 11	107.78	114.83	110.94	115.13	92.19	104.07	109.34	106.20
Period 12	110.75	117.14	114.00	118.14	93.72	104.75	110.49	106.91
Period 13	116.42	123.20	119.84	122.15	96.20	105.12	109.57	106.52
Period 14	116.20	122.96	119.60	121.91	92.98	104.01	108.35	104.94
Period 15	119.02	125.95	122.24	124.87	89.03	100.39	104.41	101.70
Period 16	125.72	130.66	126.82	129.55	96.23	106.31	112.48	107.77
Period 17	127.35	134.77	128.47	132.92	97.49	108.03	113.68	109.03
Period 18	126.86	134.25	128.48	132.93	101.51	109.45	116.17	110.38
Period 19	126.67	134.05	128.03	132.46	98.61	104.49	110.33	105.37
Period 20	128.60	136.53	130.26	134.48	100.16	109.10	114.29	109.97

Table E.16: The movement of invested $100 for 200 days split in to 20 periods (Stock: UNH; Industry: Health Care)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	103.37	103.37	103.37	102.74	104.83	101.20	98.70	100.71
Period 2	104.90	104.90	104.90	104.26	101.84	99.55	94.10	98.44
Period 3	99.77	98.78	99.77	98.62	96.12	101.50	96.64	100.62
Period 4	103.34	102.06	103.34	102.15	98.85	101.70	96.94	100.87
Period 5	108.03	106.69	108.03	106.79	101.67	102.22	100.79	102.55
Period 6	108.54	107.19	108.54	107.29	102.23	102.44	101.54	103.09
Period 7	112.86	111.46	112.86	111.56	105.90	103.12	103.25	104.54
Period 8	114.48	113.76	115.19	113.86	106.65	103.72	105.34	105.09
Period 9	115.26	115.64	117.10	115.75	108.98	103.82	106.98	105.67
Period 10	114.43	114.81	116.25	114.91	103.11	101.33	104.90	103.17
Period 11	119.97	120.65	121.88	120.48	107.82	104.07	109.34	106.20
Period 12	125.01	126.75	126.72	124.18	113.13	104.75	110.49	106.91
Period 13	127.13	127.10	127.07	124.53	113.43	105.12	109.57	106.52
Period 14	132.08	132.04	132.02	129.38	109.97	104.01	108.35	104.94
Period 15	137.36	135.65	137.79	134.55	114.43	100.39	104.41	101.70
Period 16	148.22	147.43	149.76	145.19	122.60	106.31	112.48	107.77
Period 17	150.74	150.01	151.38	147.73	124.68	108.03	113.68	109.03
Period 18	152.78	152.46	153.86	150.76	129.02	109.45	116.17	110.38
Period 19	155.84	155.52	156.94	153.78	124.34	104.49	110.33	105.37
Period 20	164.98	164.63	166.14	162.79	131.14	109.10	114.29	109.97

Table E.17: The movement of invested $100 for 200 days split in to 20 periods (Stock: HD; Industry: Consumer Services)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	101.87	101.87	101.49	101.49	99.26	101.20	98.70	100.71
Period 2	102.91	102.91	102.13	102.53	94.90	99.55	94.10	98.44
Period 3	108.24	108.24	107.42	107.83	99.42	101.50	96.64	100.62
Period 4	108.13	107.70	107.31	107.72	95.42	101.70	96.94	100.87
Period 5	112.05	111.60	111.19	111.63	99.91	102.22	100.79	102.55
Period 6	113.47	111.99	112.61	113.04	98.65	102.44	101.54	103.09
Period 7	116.37	114.86	115.81	115.94	101.57	103.12	103.25	104.54
Period 8	117.32	115.79	116.75	116.88	100.13	103.72	105.34	105.09
Period 9	116.01	114.50	116.27	115.57	101.92	103.82	106.98	105.67
Period 10	119.15	117.60	119.42	118.70	103.68	101.33	104.90	103.17
Period 11	131.26	129.56	131.56	130.77	114.50	104.07	109.34	106.20
Period 12	136.53	131.96	134.66	136.02	114.81	104.75	110.49	106.91
Period 13	138.36	133.73	136.47	137.84	116.16	105.12	109.57	106.52
Period 14	142.83	138.05	140.88	142.30	117.89	104.01	108.35	104.94
Period 15	147.18	142.26	145.17	146.63	116.11	100.39	104.41	101.70
Period 16	154.16	149.01	152.06	153.59	121.47	106.31	112.48	107.77
Period 17	157.27	150.22	153.29	156.69	123.92	108.03	113.68	109.03
Period 18	157.23	152.56	155.48	156.64	124.68	109.45	116.17	110.38
Period 19	160.28	154.93	158.50	159.09	123.28	104.49	110.33	105.37
Period 20	171.46	165.74	169.55	170.18	133.33	109.10	114.29	109.97

Table E.18: The movement of invested $100 for 200 days split in to 20 periods (Stock: NKE; Industry: Consumer Services)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	96.60	96.60	96.60	96.60	93.75	101.20	98.70	100.71
Period 2	98.20	98.20	98.20	98.20	91.09	99.55	94.10	98.44
Period 3	100.81	101.64	101.64	101.64	91.93	101.50	96.64	100.62
Period 4	103.36	104.21	104.21	104.21	92.73	101.70	96.94	100.87
Period 5	109.54	110.44	110.44	110.35	96.56	102.22	100.79	102.55
Period 6	107.81	108.70	108.70	108.61	94.53	102.44	101.54	103.09
Period 7	110.90	111.37	111.37	111.72	97.17	103.12	103.25	104.54
Period 8	110.95	111.75	111.43	111.77	97.71	103.72	105.34	105.09
Period 9	111.21	111.03	111.69	112.03	98.25	103.82	106.98	105.67
Period 10	113.76	113.57	112.54	114.60	97.42	101.33	104.90	103.17
Period 11	116.80	116.71	115.55	117.66	100.24	104.07	109.34	106.20
Period 12	120.85	121.49	120.28	121.74	104.49	104.75	110.49	106.91
Period 13	123.21	123.64	122.41	124.12	102.34	105.12	109.57	106.52
Period 14	127.64	125.11	123.86	128.59	113.13	104.01	108.35	104.94
Period 15	129.68	127.10	125.84	130.64	112.73	100.39	104.41	101.70
Period 16	137.04	134.33	132.99	138.06	118.50	106.31	112.48	107.77
Period 17	140.87	138.08	136.70	141.91	121.81	108.03	113.68	109.03
Period 18	144.36	142.43	139.29	145.43	124.32	109.45	116.17	110.38
Period 19	142.44	141.44	137.79	143.49	118.53	104.49	110.33	105.37
Period 20	147.17	146.13	142.36	148.26	122.27	109.10	114.29	109.97

Table E.19: The movement of invested $100 for 200 days split in to 20 periods (Stock: WMT; Industry: Consumer Services)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	103.38	103.38	103.38	104.77	102.96	101.20	98.70	100.71
Period 2	105.34	103.97	105.47	106.76	103.11	99.55	94.10	98.44
Period 3	108.22	106.33	108.35	109.68	106.91	101.50	96.64	100.62
Period 4	109.07	106.25	109.20	110.58	106.26	101.70	96.94	100.87
Period 5	109.11	106.29	109.23	110.62	102.54	102.22	100.79	102.55
Period 6	110.06	107.21	110.19	111.59	102.20	102.44	101.54	103.09
Period 7	109.28	106.45	109.40	110.79	101.09	103.12	103.25	104.54
Period 8	110.59	107.73	110.72	112.12	103.67	103.72	105.34	105.09
Period 9	110.19	107.34	110.32	110.96	102.52	103.82	106.98	105.67
Period 10	111.54	108.66	111.67	112.32	101.43	101.33	104.90	103.17
Period 11	113.90	110.95	114.03	114.69	102.87	104.07	109.34	106.20
Period 12	113.08	110.77	113.21	113.87	103.96	104.75	110.49	106.91
Period 13	113.29	110.49	113.43	114.09	103.66	105.12	109.57	106.52
Period 14	115.82	112.96	115.96	116.64	105.07	104.01	108.35	104.94
Period 15	117.80	114.88	117.94	117.02	102.07	100.39	104.41	101.70
Period 16	120.33	117.35	120.47	119.45	103.62	106.31	112.48	107.77
Period 17	129.16	125.97	126.73	128.22	113.52	108.03	113.68	109.03
Period 18	134.23	129.55	130.70	132.56	117.37	109.45	116.17	110.38
Period 19	131.44	126.85	127.98	129.80	113.32	104.49	110.33	105.37
Period 20	136.50	132.59	133.77	135.67	117.31	109.10	114.29	109.97

Table E.20: The movement of invested $100 for 200 days split in to 20 periods (Stock: DIS; Industry: Consumer Services)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	105.31	105.31	105.31	105.31	99.17	101.20	98.70	100.71
Period 2	107.82	107.82	107.82	107.82	94.42	99.55	94.10	98.44
Period 3	110.75	111.33	111.33	111.33	96.46	101.50	96.64	100.62
Period 4	114.83	113.93	115.43	115.43	98.38	101.70	96.94	100.87
Period 5	119.36	118.42	119.99	119.99	102.16	102.22	100.79	102.55
Period 6	120.15	119.20	120.78	120.78	100.67	102.44	101.54	103.09
Period 7	122.87	121.90	123.51	123.51	102.67	103.12	103.25	104.54
Period 8	127.12	126.12	127.78	127.78	105.64	103.72	105.34	105.09
Period 9	125.30	124.31	126.18	125.89	104.84	103.82	106.98	105.67
Period 10	128.17	127.16	129.08	128.78	105.59	101.33	104.90	103.17
Period 11	132.54	131.50	133.47	133.16	110.02	104.07	109.34	106.20
Period 12	133.94	132.89	134.89	134.57	110.10	104.75	110.49	106.91
Period 13	135.73	135.12	136.68	136.37	108.56	105.12	109.57	106.52
Period 14	140.05	139.42	141.04	140.71	107.67	104.01	108.35	104.94
Period 15	146.39	145.73	147.42	147.08	103.98	100.39	104.41	101.70
Period 16	157.08	156.37	158.18	157.82	111.50	106.31	112.48	107.77
Period 17	154.85	155.24	157.04	156.68	109.92	108.03	113.68	109.03
Period 18	159.48	159.71	161.56	161.18	113.64	109.45	116.17	110.38
Period 19	159.68	159.92	159.99	159.62	111.02	104.49	110.33	105.37
Period 20	168.19	168.43	168.51	168.12	115.97	109.10	114.29	109.97

Table E.21: The movement of invested \$100 for 200 days split in to 20 periods (Stock: AXP; Industry: Financials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning(\$)	100.00				100.00			
Period 1	103.45	103.45	103.45	103.45	99.45	101.20	98.70	100.71
Period 2	106.42	106.42	106.42	106.42	94.10	99.55	94.10	98.44
Period 3	107.23	107.23	107.23	107.23	95.62	101.50	96.64	100.62
Period 4	111.01	109.44	110.82	111.79	96.74	101.70	96.94	100.87
Period 5	116.13	114.49	115.93	116.94	99.83	102.22	100.79	102.55
Period 6	120.76	119.06	120.55	121.61	103.63	102.44	101.54	103.09
Period 7	121.57	119.86	121.36	122.43	103.13	103.12	103.25	104.54
Period 8	122.78	121.05	122.57	123.64	103.60	103.72	105.34	105.09
Period 9	121.29	119.58	121.08	122.14	100.81	103.82	106.98	105.67
Period 10	124.12	122.36	123.90	124.98	95.92	101.33	104.90	103.17
Period 11	125.43	124.30	125.87	126.97	97.47	104.07	109.34	106.20
Period 12	126.46	124.41	126.25	127.36	97.52	104.75	110.49	106.91
Period 13	129.98	129.45	129.63	130.76	97.59	105.12	109.57	106.52
Period 14	132.98	132.44	132.62	133.78	95.73	104.01	108.35	104.94
Period 15	142.18	140.52	140.71	141.94	92.40	100.39	104.41	101.70
Period 16	152.98	151.19	151.40	152.72	99.93	106.31	112.48	107.77
Period 17	151.56	149.78	151.87	151.30	99.13	108.03	113.68	109.03
Period 18	153.96	152.16	154.28	153.70	102.29	109.45	116.17	110.38
Period 19	150.80	149.04	151.12	150.55	97.06	104.49	110.33	105.37
Period 20	160.39	158.52	160.73	160.12	102.34	109.10	114.29	109.97

Table E.22: The movement of invested \$100 for 200 days split in to 20 periods (Stock: GS; Industry: Financials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning(\$)	100.00				100.00			
Period 1	102.14	103.33	102.14	102.14	98.47	101.20	98.70	100.71
Period 2	102.67	103.87	102.67	102.67	91.95	99.55	94.10	98.44
Period 3	109.34	110.61	106.99	106.99	94.85	101.50	96.64	100.62
Period 4	110.54	112.07	108.16	108.16	94.63	101.70	96.94	100.87
Period 5	114.53	116.12	112.07	112.07	95.73	102.22	100.79	102.55
Period 6	116.88	118.49	114.36	114.43	98.83	102.44	101.54	103.09
Period 7	119.76	121.42	117.18	117.25	100.06	103.12	103.25	104.54
Period 8	120.40	122.07	117.81	117.88	98.14	103.72	105.34	105.09
Period 9	128.62	130.40	124.67	124.74	104.46	103.82	106.98	105.67
Period 10	133.77	135.62	129.66	129.74	102.59	101.33	104.90	103.17
Period 11	135.77	137.64	131.60	131.68	104.50	104.07	109.34	106.20
Period 12	135.71	137.59	131.55	131.62	107.59	104.75	110.49	106.91
Period 13	137.95	139.10	135.34	133.79	110.69	105.12	109.57	106.52
Period 14	142.53	143.72	139.83	138.23	112.00	104.01	108.35	104.94
Period 15	145.15	146.71	142.74	141.11	106.24	100.39	104.41	101.70
Period 16	157.24	158.93	154.63	152.87	114.00	106.31	112.48	107.77
Period 17	157.39	158.17	154.86	153.02	113.46	108.03	113.68	109.03
Period 18	160.07	160.85	157.49	155.62	113.98	109.45	116.17	110.38
Period 19	160.53	161.32	157.94	156.07	109.86	104.49	110.33	105.37
Period 20	170.46	171.29	167.71	165.72	116.16	109.10	114.29	109.97

Table E.23: The movement of invested $100 for 200 days split in to 20 periods (Stock: JPM; Industry: Financials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	105.37	104.09	105.37	105.37	104.62	101.20	98.70	100.71
Period 2	105.54	103.62	106.99	106.99	95.09	99.55	94.10	98.44
Period 3	108.51	106.66	109.59	109.59	97.14	101.50	96.64	100.62
Period 4	110.74	108.67	111.83	111.83	94.33	101.70	96.94	100.87
Period 5	113.41	111.29	115.86	114.45	96.69	102.22	100.79	102.55
Period 6	117.22	114.23	119.48	118.02	98.98	102.44	101.54	103.09
Period 7	119.25	115.52	120.83	119.36	99.59	103.12	103.25	104.54
Period 8	118.82	115.11	120.40	118.92	97.51	103.72	105.34	105.09
Period 9	123.94	120.07	125.58	124.05	103.12	103.82	106.98	105.67
Period 10	125.75	122.91	127.42	124.27	98.45	101.33	104.90	103.17
Period 11	131.23	127.17	131.84	128.58	102.21	104.07	109.34	106.20
Period 12	131.88	127.80	132.49	129.21	104.65	104.75	110.49	106.91
Period 13	137.59	133.33	138.23	134.81	106.44	105.12	109.57	106.52
Period 14	141.85	137.46	142.51	138.98	105.87	104.01	108.35	104.94
Period 15	143.93	139.48	144.60	141.02	99.62	100.39	104.41	101.70
Period 16	155.91	151.09	156.64	152.76	107.10	106.31	112.48	107.77
Period 17	154.67	149.89	155.39	151.55	106.22	108.03	113.68	109.03
Period 18	159.40	155.05	160.14	156.18	107.45	109.45	116.17	110.38
Period 19	163.34	158.88	164.10	160.04	102.79	104.49	110.33	105.37
Period 20	174.90	170.07	176.48	172.12	110.09	109.10	114.29	109.97

Table E.24: The movement of invested $100 for 200 days split in to 20 periods (Stock: TRV; Industry: Financials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	100.73	100.10	101.66	100.73	100.48	101.20	98.70	100.71
Period 2	102.35	101.71	103.30	103.05	102.26	99.55	94.10	98.44
Period 3	104.67	104.02	105.64	105.38	107.85	101.50	96.64	100.62
Period 4	107.16	106.49	108.15	107.89	110.31	101.70	96.94	100.87
Period 5	108.34	107.66	109.35	109.08	112.04	102.22	100.79	102.55
Period 6	109.63	109.09	110.65	110.52	113.77	102.44	101.54	103.09
Period 7	109.40	108.86	110.42	110.29	113.06	103.12	103.25	104.54
Period 8	109.13	108.59	110.15	110.02	113.38	103.72	105.34	105.09
Period 9	112.90	112.34	113.95	112.62	110.29	103.82	106.98	105.67
Period 10	115.19	114.62	116.26	114.91	108.35	101.33	104.90	103.17
Period 11	118.15	117.56	120.46	117.86	112.27	104.07	109.34	106.20
Period 12	119.17	118.58	121.50	118.89	112.35	104.75	110.49	106.91
Period 13	121.27	119.85	122.80	120.98	113.95	105.12	109.57	106.52
Period 14	125.25	123.79	126.83	124.12	114.07	104.01	108.35	104.94
Period 15	128.95	125.91	130.81	126.47	112.27	100.39	104.41	101.70
Period 16	138.18	132.97	139.00	134.39	120.40	106.31	112.48	107.77
Period 17	139.74	135.65	140.57	135.91	123.77	108.03	113.68	109.03
Period 18	143.94	139.82	144.80	140.00	125.76	109.45	116.17	110.38
Period 19	145.08	140.93	145.95	141.11	124.82	104.49	110.33	105.37
Period 20	149.81	145.88	150.70	145.71	128.18	109.10	114.29	109.97

Table E.25: The movement of invested \$100 for 200 days split in to 20 periods (Stock: V; Industry: Financials)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning(\$)	100.00				100.00			
Period 1	99.53	99.53	98.49	98.49	94.74	101.20	98.70	100.71
Period 2	102.66	102.66	101.58	101.58	90.04	99.55	94.10	98.44
Period 3	104.67	104.67	103.57	103.57	89.41	101.50	96.64	100.62
Period 4	108.45	108.45	107.31	107.09	92.78	101.70	96.94	100.87
Period 5	113.09	113.09	111.90	111.67	94.91	102.22	100.79	102.55
Period 6	111.86	111.86	111.00	110.46	93.63	102.44	101.54	103.09
Period 7	111.74	111.74	110.88	110.35	92.39	103.12	103.25	104.54
Period 8	116.95	116.95	116.06	115.49	95.94	103.72	105.34	105.09
Period 9	119.98	116.45	115.05	117.52	94.95	103.82	106.98	105.67
Period 10	120.63	116.74	115.50	118.15	93.05	101.33	104.90	103.17
Period 11	124.13	120.88	119.60	122.34	95.72	104.07	109.34	106.20
Period 12	123.16	118.48	119.66	121.32	95.58	104.75	110.49	106.91
Period 13	125.78	120.99	122.51	123.59	94.74	105.12	109.57	106.52
Period 14	127.13	122.30	124.82	124.93	93.82	104.01	108.35	104.94
Period 15	131.52	126.52	129.12	129.23	92.07	100.39	104.41	101.70
Period 16	138.83	133.55	136.30	136.42	106.99	106.31	112.48	107.77
Period 17	144.60	139.76	141.97	142.09	110.87	108.03	113.68	109.03
Period 18	152.06	145.73	149.30	149.43	115.40	109.45	116.17	110.38
Period 19	154.08	146.42	149.82	151.40	112.47	104.49	110.33	105.37
Period 20	161.50	153.48	157.04	158.70	116.37	109.10	114.29	109.97

Table E.26: The movement of invested $100 for 200 days split in to 20 periods (Stock: AAPL; Industry: Technology)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	100.00	100.70	100.70	100.70	101.76	101.20	98.70	100.71
Period 2	101.00	101.17	101.83	101.83	97.31	99.55	94.10	98.44
Period 3	101.79	106.42	102.63	103.72	110.86	101.50	96.64	100.62
Period 4	105.41	110.59	105.92	107.13	112.20	101.70	96.94	100.87
Period 5	112.20	117.70	112.61	113.89	120.04	102.22	100.79	102.55
Period 6	114.02	121.50	116.06	115.73	122.05	102.44	101.54	103.09
Period 7	114.27	121.75	116.32	115.99	120.21	103.12	103.25	104.54
Period 8	120.70	125.75	122.86	119.80	125.93	103.72	105.34	105.09
Period 9	127.96	133.32	130.25	127.01	129.17	103.82	106.98	105.67
Period 10	124.73	130.83	126.31	126.80	125.92	101.33	104.90	103.17
Period 11	133.34	139.92	134.45	134.77	134.66	104.07	109.34	106.20
Period 12	134.95	142.35	136.08	136.40	130.73	104.75	110.49	106.91
Period 13	146.01	149.82	147.23	147.58	134.32	105.12	109.57	106.52
Period 14	149.77	153.68	148.10	151.38	132.40	104.01	108.35	104.94
Period 15	156.71	164.87	154.96	158.39	132.59	100.39	104.41	101.70
Period 16	169.29	178.11	167.40	171.10	145.40	106.31	112.48	107.77
Period 17	177.27	187.22	175.29	179.17	152.16	108.03	113.68	109.03
Period 18	186.52	197.05	186.24	190.36	153.01	109.45	116.17	110.38
Period 19	191.11	193.82	192.65	197.02	142.49	104.49	110.33	105.37
Period 20	201.82	206.64	203.45	208.07	147.34	109.10	114.29	109.97

Table E.27: The movement of invested $100 for 200 days split in to 20 periods (Stock: CSCO; Industry: Technology)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	107.89	107.11	107.89	105.06	107.41	101.20	98.70	100.71
Period 2	111.35	109.53	111.35	107.44	106.43	99.55	94.10	98.44
Period 3	111.04	112.65	111.04	108.72	107.46	101.50	96.64	100.62
Period 4	110.73	112.05	111.76	109.28	106.06	101.70	96.94	100.87
Period 5	112.28	114.67	114.38	110.81	114.76	102.22	100.79	102.55
Period 6	113.90	116.00	115.70	112.09	114.66	102.44	101.54	103.09
Period 7	113.70	115.88	115.49	111.89	114.62	103.12	103.25	104.54
Period 8	117.12	119.40	118.99	115.26	119.57	103.72	105.34	105.09
Period 9	118.56	121.59	120.10	116.68	121.68	103.82	106.98	105.67
Period 10	118.38	123.70	122.19	118.71	117.27	101.33	104.90	103.17
Period 11	117.35	123.69	120.48	117.04	115.49	104.07	109.34	106.20
Period 12	119.11	125.27	122.09	118.80	116.90	104.75	110.49	106.91
Period 13	119.94	126.25	123.19	119.63	116.99	105.12	109.57	106.52
Period 14	126.01	132.64	127.89	124.19	118.07	104.01	108.35	104.94
Period 15	125.35	131.94	127.66	123.54	108.25	100.39	104.41	101.70
Period 16	134.72	141.80	137.20	132.77	116.04	106.31	112.48	107.77
Period 17	143.55	151.10	144.93	141.48	124.97	108.03	113.68	109.03
Period 18	150.72	157.17	152.78	148.54	131.34	109.45	116.17	110.38
Period 19	150.17	154.88	150.55	148.00	125.53	104.49	110.33	105.37
Period 20	159.13	164.88	158.61	155.92	131.34	109.10	114.29	109.97

Table E.28: The movement of invested \$100 for 200 days split in to 20 periods (Stock: IBM; Industry: Technology)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning(\$)	100.00				100.00			
Period 1	100.00	107.98	106.87	106.87	104.80	101.20	98.70	100.71
Period 2	100.84	108.80	107.64	107.64	106.15	99.55	94.10	98.44
Period 3	99.81	107.69	107.91	107.91	105.86	101.50	96.64	100.62
Period 4	100.91	108.88	109.22	109.11	102.27	101.70	96.94	100.87
Period 5	100.37	108.29	108.84	108.51	99.59	102.22	100.79	102.55
Period 6	99.76	109.15	108.17	107.85	98.21	102.44	101.54	103.09
Period 7	100.45	109.45	109.37	108.60	97.75	103.12	103.25	104.54
Period 8	103.85	111.85	113.07	112.28	101.88	103.72	105.34	105.09
Period 9	107.38	115.66	116.92	116.10	105.35	103.82	106.98	105.67
Period 10	109.50	117.93	119.64	118.81	101.74	101.33	104.90	103.17
Period 11	111.69	119.92	122.04	121.19	103.80	104.07	109.34	106.20
Period 12	111.83	120.86	121.68	120.83	103.65	104.75	110.49	106.91
Period 13	113.75	123.82	123.74	122.87	105.27	105.12	109.57	106.52
Period 14	115.63	125.86	125.96	125.08	103.05	104.01	108.35	104.94
Period 15	109.65	119.35	119.45	118.61	92.18	100.39	104.41	101.70
Period 16	107.13	116.61	116.70	115.88	89.60	106.31	112.48	107.77
Period 17	108.62	119.38	118.30	117.47	90.10	108.03	113.68	109.03
Period 18	109.33	119.77	117.80	116.98	89.28	109.45	116.17	110.38
Period 19	109.07	119.85	117.36	116.25	83.10	104.49	110.33	105.37
Period 20	116.10	128.79	124.92	123.74	88.06	109.10	114.29	109.97

Table E.29: The movement of invested \$100 for 200 days split in to 20 periods (Stock: INTC; Industry: Technology)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning(\$)	100.00				100.00			
Period 1	103.94	105.34	105.34	105.34	104.59	101.20	98.70	100.71
Period 2	108.94	110.41	110.41	110.41	107.73	99.55	94.10	98.44
Period 3	110.48	111.97	112.69	111.97	107.40	101.50	96.64	100.62
Period 4	110.80	111.82	112.75	112.29	106.87	101.70	96.94	100.87
Period 5	113.31	114.36	115.31	114.84	109.42	102.22	100.79	102.55
Period 6	117.07	118.15	119.13	118.65	113.48	102.44	101.54	103.09
Period 7	119.90	121.01	122.01	121.51	124.93	103.12	103.25	104.54
Period 8	122.67	123.44	124.47	124.32	126.84	103.72	105.34	105.09
Period 9	125.17	125.96	127.01	126.86	139.01	103.82	106.98	105.67
Period 10	126.44	127.24	128.30	128.14	133.19	101.33	104.90	103.17
Period 11	135.51	136.66	137.51	137.63	142.76	104.07	109.34	106.20
Period 12	136.58	138.30	138.59	139.66	144.35	104.75	110.49	106.91
Period 13	136.96	138.69	138.98	140.05	141.82	105.12	109.57	106.52
Period 14	140.10	141.87	142.16	144.46	139.36	104.01	108.35	104.94
Period 15	144.51	146.34	146.64	149.01	129.03	100.39	104.41	101.70
Period 16	157.43	159.42	159.75	162.33	140.18	106.31	112.48	107.77
Period 17	158.14	160.14	160.47	163.06	140.81	108.03	113.68	109.03
Period 18	176.46	175.92	178.12	181.11	154.63	109.45	116.17	110.38
Period 19	176.63	172.45	178.29	181.29	146.24	104.49	110.33	105.37
Period 20	183.80	179.45	185.52	188.22	149.24	109.10	114.29	109.97

TABLE E.30: The movement of invested $100 for 200 days split in to 20 periods (Stock: MSFT; Industry: Technology)

Forecast term	Buy&Sell				Buy&Hold			
	MV-GPR	MV-TPR	GPR	TPR	XOM	INDU	NDX	SPX
Beginning($)	100.00				100.00			
Period 1	105.51	105.71	106.22	106.22	104.94	101.20	98.70	100.71
Period 2	108.67	108.87	109.30	109.40	100.71	99.55	94.10	98.44
Period 3	108.86	109.07	112.11	109.60	102.36	101.50	96.64	100.62
Period 4	111.29	109.89	114.03	111.48	102.67	101.70	96.94	100.87
Period 5	112.69	111.27	115.54	112.88	102.93	102.22	100.79	102.55
Period 6	114.61	114.57	118.97	116.22	103.54	102.44	101.54	103.09
Period 7	116.49	116.45	120.91	118.13	106.45	103.12	103.25	104.54
Period 8	116.83	116.79	121.27	118.47	107.39	103.72	105.34	105.09
Period 9	126.28	124.47	131.08	128.06	113.54	103.82	106.98	105.67
Period 10	124.50	124.21	129.23	127.79	110.22	101.33	104.90	103.17
Period 11	130.93	130.08	137.37	134.40	115.92	104.07	109.34	106.20
Period 12	134.76	134.60	141.38	138.33	119.31	104.75	110.49	106.91
Period 13	137.53	135.94	142.79	139.71	120.82	105.12	109.57	106.52
Period 14	140.01	138.39	145.36	142.22	118.33	104.01	108.35	104.94
Period 15	145.97	144.29	151.56	148.28	113.17	100.39	104.41	101.70
Period 16	158.72	156.89	164.80	161.23	121.80	106.31	112.48	107.77
Period 17	165.48	163.57	171.81	168.10	126.98	108.03	113.68	109.03
Period 18	162.02	158.04	168.22	164.58	125.20	109.45	116.17	110.38
Period 19	161.76	157.79	167.95	164.32	116.67	104.49	110.33	105.37
Period 20	173.61	166.09	176.57	172.76	120.01	109.10	114.29	109.97

E.4 Industrial sector portfolio investment

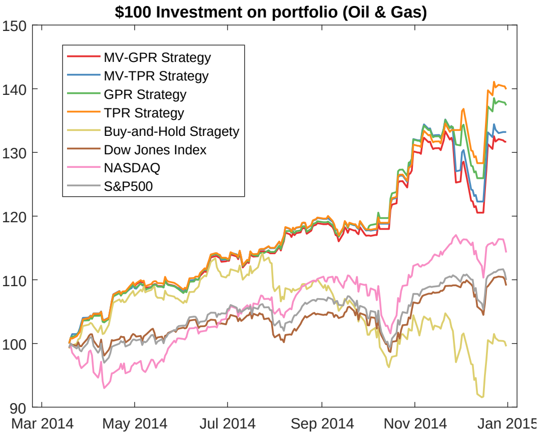


FIGURE E.8: Oil & Gas portfolio investment

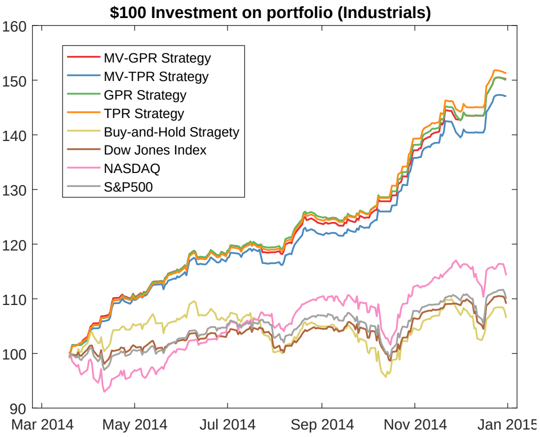


FIGURE E.9: Industrials portfolio investment

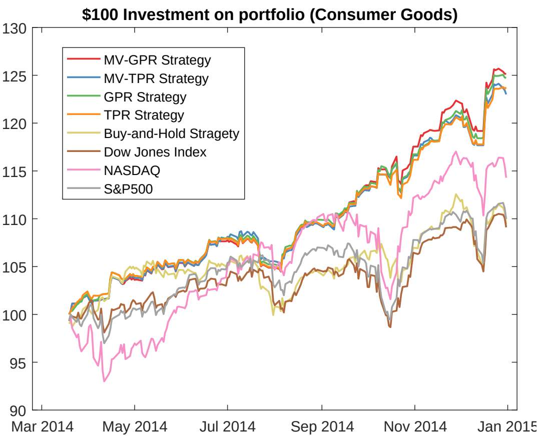


FIGURE E.10: Consumer Goods portfolio investment

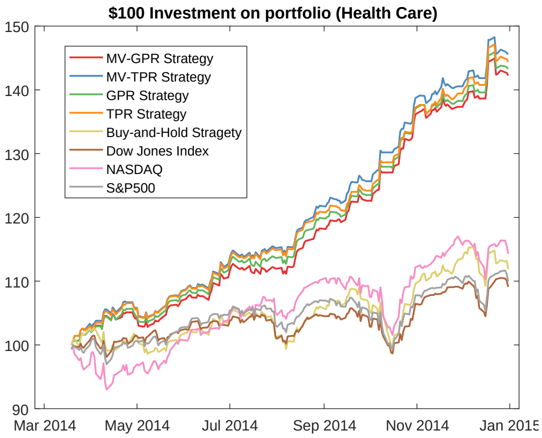


FIGURE E.11: Health Care portfolio investment

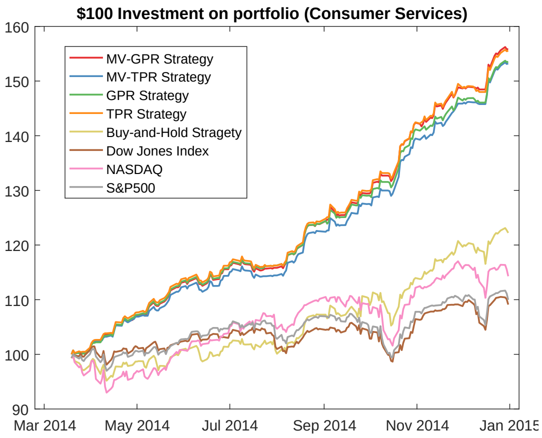


FIGURE E.12: Consumer Services portfolio investment

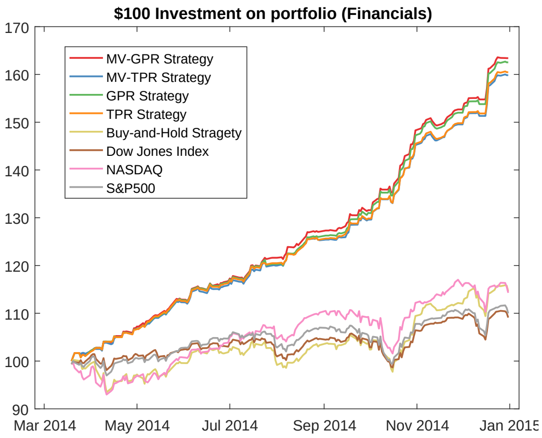


FIGURE E.13: Financials portfolio investment

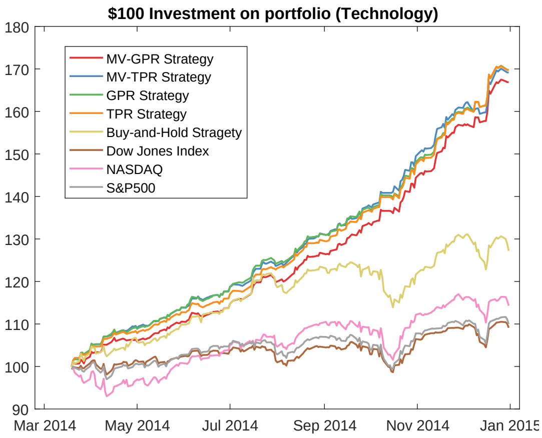


FIGURE E.14: Technology portfolio investment

Bibliography

[1] M. Imdadullah, Time series analysis, Basic Statistics and Data Analysis.

[2] R. S. Tsay, Analysis of financial time series, Vol. 543, John Wiley & Sons, 2005.

[3] C. W. Granger, Forecasting stock market prices: Lessons for forecasters, International Journal of Forecasting 8 (1) (1992) 3–13.

[4] B. G. Malkiel, The efficient market hypothesis and its critics, The Journal of Economic Perspectives 17 (1) (2003) 59–82.

[5] A. Agarwal, Testing weak form of market efficiency by application of simple technical trading rules on the Indian stock market, Ph.D. thesis, University of Nottingham (2006).

[6] M. Beechey, D. W. Gruen, J. Vickery, et al., The efficient market hypothesis: a survey, Reserve Bank of Australia, Economic Research Department, 2000.

[7] S. H. Irwin, J. W. Uhrig, Do technical analysts have holes in their shoes?, Review of Research in Futures Markets 3 (3) (1984) 264–277.

[8] M. P. Taylor, H. Allen, The use of technical analysis in the foreign exchange market, Journal of international Money and Finance 11 (3) (1992) 304–314.

[9] A. W. Lo, The adaptive markets hypothesis, The Journal of Portfolio Management 30 (5) (2004) 15–29.

[10] Y. Shi, A. Gorban, T. Yang, Is it possible to predict long-term success with k-NN? case study of four market indices (FTSE100, DAX, HANGSENG, NASDAQ), in: Journal of Physics: Conference Series, Vol. 490, IOP Publishing, 2014, p. 012082.

[11] A. Ang, G. Bekaert, Stock return predictability: Is it there?, Review of Financial studies 20 (3) (2007) 651–707.

[12] K.-j. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (1) (2003) 307–319.

[13] D. Avramov, Stock return predictability and model uncertainty, Journal of Financial Economics 64 (3) (2002) 423–458.

[14] K. M. Cremers, Stock return predictability: A Bayesian model selection perspective, Review of Financial Studies 15 (4) (2002) 1223–1249.

[15] C. R. Harvey, Predictable risk and returns in emerging markets, Review of Financial studies 8 (3) (1995) 773–816.

[16] A. Inoue, L. Kilian, In-sample or out-of-sample tests of predictability: Which one should we use?, Econometric Reviews 23 (4) (2005) 371–402.

[17] D. E. Rapach, M. E. Wohar, In-sample vs. out-of-sample tests of stock return predictability in the context of data mining, Journal of Empirical Finance 13 (2) (2006) 231–247.

[18] Y. Shi, knn predictability analysis of stock and share closing prices, Ph.D. thesis, Department of Mathematics (2016).

[19] F. Longin, B. Solnik, Is the correlation in international equity returns constant: 1960–1990?, Journal of international money and finance 14 (1) (1995) 3–26.

[20] A. N. Gorban, E. V. Smirnova, T. A. Tyukina, Correlations, risk and crisis: From physiology to finance, Physica A: Statistical Mechanics and its Applications 389 (16) (2010) 3193–3217.

[21] E. Haven, P. Molyneux, J. Wilson, S. Fedotov, M. Duygun, The Handbook of Post Crisis Financial Modelling, Springer, 2016.

[22] C. E. Rasmussen, Evaluation of Gaussian processes and other methods for non-linear regression, Ph.D. thesis, Citeseer (1996).

[23] R. M. Neal, Bayesian learning for neural networks, Vol. 118, Springer Science & Business Media, 2012.

[24] C. K. Williams, Computing with infinite networks, Advances in neural information processing systems (1997) 295–301.

[25] D. MacKay, Gaussian processes-a replacement for supervised neural networks? Lecture notes for a tutorial at NIPS 1997 (1997).

[26] C. K. Williams, C. E. Rasmussen, Gaussian processes for regression, in: Advances in Neural Information Processing Systems, 1996, pp. 514–520.

[27] S. Brahim-Belhouari, J. Vesin, Bayesian learning using Gaussian process for time series prediction (2001).

[28] S. Brahim-Belhouari, A. Bermak, Gaussian process for nonstationary time series prediction, Computational Statistics and Data Analysis 47 (4) (2004) 705–712.

[29] M. T. Farrell, A. Correa, Gaussian process regression models for predicting stock trends, Relation 10 (2007) 3414.

[30] M. Mojaddady, M. Nabi, S. Khadivi, Stock market prediction using twin Gaussian process regression, International Journal for Advances in Computer Research (JACR) preprint.

[31] H. Mori, M. Ohmi, Probabilistic short-term load forecasting with Gaussian processes, in: Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems, IEEE, 2005, pp. 6–pp.

[32] D. Petelin, J. Sindelar, J. Prikryl, J. Kocijan, Financial modeling using Gaussian process models., in: IDAACS (2), 2011, pp. 672–677.

[33] E. F. Fama, L. Fisher, M. C. Jensen, R. Roll, The adjustment of stock prices to new information, International economic review 10 (1) (1969) 1–21.

[34] E. Jondeau, S.-H. Poon, M. Rockinger, Financial modeling under non-Gaussian distributions, Springer Science & Business Media, 2007.

[35] I. Bremmer, P. Keat, The fat tail: the power of political knowledge for strategic investing, Oxford University Press, 2010.

[36] S. Fedotov, N. Korabel, Emergence of Lévy walks in systems of interacting individuals, Physical Review E 95 (3) (2017) 030107.

[37] J. Vanhatalo, P. Jylänki, A. Vehtari, Gaussian process regression with Student-t likelihood, in: Advances in Neural Information Processing Systems, 2009, pp. 1910–1918.

[38] A. Shah, A. G. Wilson, Z. Ghahramani, Student-t processes as alternatives to Gaussian processes, arXiv preprint arXiv:1402.4306.

[39] A. Solin, S. Särkkä, State space methods for efficient inference in Student-t process regression, in: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, 2015, pp. 885–893.

[40] P. Boyle, M. Frean, Multiple output Gaussian process regression.

[41] B. Wang, T. Chen, Gaussian process regression with multiple response variables, Chemometrics and Intelligent Laboratory Systems 142 (2015) 159–165.

[42] D. Chakrabarty, M. Biswas, S. Bhattacharya, et al., Bayesian nonparametric estimation of Milky Way parameters using matrix-variate data, in a new Gaussian process based method, Electronic Journal of Statistics 9 (1) (2015) 1378–1403.

[43] M. A. Alvarez, L. Rosasco, N. D. Lawrence, Kernels for vector-valued functions: A review, arXiv preprint arXiv:1106.6251.

[44] A. K. Gupta, D. K. Nagar, Matrix variate distributions, Vol. 104, CRC Press, 1999.

[45] O. Akbilgic, H. Bozdogan, M. E. Balaban, A novel hybrid RBF neural networks model as a forecaster, Statistics and Computing 24 (3) (2014) 365–375.

[46] H. Cramér, M. R. Leadbetter, Stationary and related stochastic processes: Sample function properties and their applications, Courier Corporation, 2013.

[47] I. A. Ibragimov, Y. A. Rozanov, Gaussian random processes, Vol. 9, Springer Science & Business Media, 2012.

[48] K.-T. Fang, S. Kotz, K. W. Ng, Symmetric multivariate and related distributions, Chapman and Hall, 1990.

[49] A. P. Dawid, Some matrix-variate distribution theory: notational considerations and a Bayesian application, Biometrika 68 (1) (1981) 265–274.

[50] S. Zhu, K. Yu, Y. Gong, Predictive matrix-variate t models, in: Advances in Neural Information Processing Systems, 2008, pp. 1721–1728.

[51] F. Zhang, The Schur complement and its applications, Vol. 4, Springer Science & Business Media, 2006.

[52] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, P. B. Kramer, Numerical recipes: the art of scientific computing (1987).

[53] J. E. Gentle, Matrix algebra: theory, computations, and applications in statistics, Springer Science & Business Media, 2007.

[54] G. H. Golub, C. F. Van Loan, Matrix computations, Vol. 3, JHU Press, 2012.

[55] P. Whitle, Hypothesis testing in time series analysis, Vol. 4, Almqvist & Wiksells, 1951.

[56] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[57] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Selected Papers of Hirotugu Akaike, Springer, 1998, pp. 199–213.

[58] G. Schwarz, et al., Estimating the dimension of a model, The annals of statistics 6 (2) (1978) 461–464.

[59] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, Adaptive computation and machine learning, MIT Press, 2006.
URL `https://books.google.co.uk/books?id=GhoSngEACAAJ`

[60] A. G. Wilson, R. P. Adams, Gaussian process kernels for pattern discovery and extrapolation, arXiv preprint arXiv:1302.4245.

[61] J. Mercer, Functions of positive and negative type, and their connection with the theory of integral equations, Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character 209 (1909) 415–446.

[62] H. Q. Minh, P. Niyogi, Y. Yao, MercerâĂŹs theorem, feature maps, and smoothing, in: International Conference on Computational Learning Theory, Springer, 2006, pp. 154–168.

[63] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, S. Aigrain, Gaussian processes for time-series modelling, Phil. Trans. R. Soc. A 371 (1984) (2013) 20110550.

[64] D. J. MacKay, Introduction to Gaussian processes, NATO ASI Series F Computer and Systems Sciences 168 (1998) 133–166.

[65] D. K. Duvenaud, J. R. Lloyd, R. B. Grosse, J. B. Tenenbaum, Z. Ghahramani, Structure discovery in nonparametric regression through compositional kernel search., in: ICML (3), 2013, pp. 1166–1174.

[66] A. G. Wilson, Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes, Ph.D. thesis, PhD thesis, University of Cambridge (2014).

[67] R. M. Neal, Monte Carlo implementation of Gaussian process models for Bayesian regression and classification, arXiv preprint physics/9701026.

[68] F. Bachoc, Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification, Computational Statistics & Data Analysis 66 (2013) 55–69.

[69] P. C. Lambert, A. J. Sutton, P. R. Burton, K. R. Abrams, D. R. Jones, How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS, Statistics in Medicine 24 (15) (2005) 2401.

[70] A. Gelman, et al., Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper), Bayesian Analysis 1 (3) (2006) 515–534.

[71] H. Jeffreys, The theory of probability, OUP Oxford, 1998.

[72] R. D. Rosenkrantz, ET Jaynes: Papers on probability, statistics and statistical physics, Vol. 158, Springer Science & Business Media, 2012.

[73] H. Cramér, U. Grenander, Probability and statistics: the Harald Cramér volume, Almqvist & Wiksell, 1959.

[74] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian data analysis, Vol. 2, Taylor & Francis, 2014.

[75] R. S. Hudson, A. Gregoriou, Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns, International Review of Financial Analysis 38 (2015) 151–162.

[76] G. Seni, J. F. Elder, Ensemble methods in data mining: improving accuracy through combining predictions, Synthesis Lectures on Data Mining and Knowledge Discovery 2 (1) (2010) 1–126.

[77] G. McLachlan, K.-A. Do, C. Ambroise, Analyzing microarray gene expression data, Vol. 422, John Wiley & Sons, 2005.

[78] E. Zivot, J. Wang, Modeling financial time series with S-Plus®, Vol. 191, Springer Science & Business Media, 2007.

[79] A. Girard, C. Rasmussen, J. Quiñonero Candela, R. Murray Smith, Gaussian process priors with uncertain inputs-application to multiple-step ahead time series forecasting, Advances in Neural Information Processing Systems (2003) 545–552.

[80] G. S. Hiremath, J. Kumari, Stock returns predictability and the adaptive market hypothesis in emerging markets: evidence from India, SpringerPlus 3 (1) (2014) 428.

[81] Y. Liu, Z. Gao, Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models, Journal of Applied Polymer Science 132 (6).

[82] S. De Vito, E. Massera, M. Piga, L. Martinotto, G. Di Francia, On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, Sensors and Actuators B: Chemical 129 (2) (2008) 750–757.

[83] H. Fanaee-T, J. Gama, Event labeling combining ensemble detectors and background knowledge., Progress in AI 2 (2-3) (2014) 113–127.

[84] Y. Luo, F. Fang, O. A. Esqueda, The overseas listing puzzle: Post-IPO performance of Chinese stocks and ADRs in the US market, Journal of multinational financial management 22 (5) (2012) 193–211.

[85] S. P. Baca, B. L. Garbe, R. A. Weiss, The rise of sector effects in major equity markets, Financial Analysts Journal 56 (5) (2000) 34–40.

[86] A. C. Damianou, N. D. Lawrence, Deep Gaussian processes., in: AISTATS, 2013, pp. 207–215.

[87] Y. Zhang, G. Luo, Recursive prediction algorithm for non-stationary Gaussian process, Journal of Systems and Software.

[88] Y. Zhang, G. Luo, Fast algorithm for non-stationary Gaussian process prediction, Power 800 (2014) 900.

[89] A. Svensson, A. Solin, S. Särkkä, T. B. Schön, Computationally efficient Bayesian learning of Gaussian process state space models, arXiv preprint arXiv: 1506: 02267.

[90] Y. Cao, M. A. Brubaker, D. J. Fleet, A. Hertzmann, Efficient optimization for sparse Gaussian process regression, in: Advances in Neural Information Processing Systems, 2013, pp. 1097–1105.

[91] A. Gittens, M. W. Mahoney, Revisiting the Nyström method for improved large-scale machine learning, J. Mach. Learn. Res 28 (3) (2013) 567–575.