

Forecasting Financial Time Series using Linear Predictive Filters

Bin Li

A thesis submitted for the degree of

Doctor of Philosophy in Electrical and Electronic Engineering of

Imperial College London and the Diploma of

Imperial College London

Department of Electrical and Electronic Engineering

Imperial College London

July 2012

Abstract

Forecasting financial time series is regarded as one of the most challenging applications of time series prediction due to their dynamic nature. However, it is the fundamental element of most investment activities thus attracting the attention of practitioners and researchers for many decades.

The purpose of this research is to investigate and develop novel methods for the prediction of financial time series considering their dynamic nature. The predictive performance of asset prices time series themselves is exploited by applying digital signal processing methods to their historical observations. The novelty of the research lies in the design of predictive filters by maximising their spectrum flatness of forecast errors. The filters are then applied to forecast linear combinations of daily open, high, low and close prices of financial time series.

Given the assumption that there are no structural breaks or switching regimes in a time series, the sufficient and necessary conditions that a time series can be predicted with zero errors by linear filters are examined. It is concluded that a band-limited time series can be predicted with zero errors by a predictive filter that has a constant magnitude response and constant group delay over the bandwidth of the time series. Because real world time series are not band-limited thus cannot be forecasted without errors, statistical tests of spectrum flatness which evaluate the departure of the spectral density from a constant value are introduced as measures of the predictability of time series. Properties of a time series are then investigated in the frequency domain

using its spectrum flatness. A predictive filter is designed by maximising the error spectrum flatness that is equivalent to maximise the “whiteness” of forecast errors in the frequency domain.

The focus is then placed on forecasting real world financial time series. By applying spectrum flatness tests, it is found that the property of the spectrum of a linear combination of daily open, high, low and close prices, which is called target prices, is different from that of a random walk process as there are much more low frequency components than high frequency ones in its spectrum. Therefore, an objective function is proposed to derive the target price time series from the historical observations of daily open, high, low and close prices. A predictive filter is then applied to obtain the one-step ahead forecast of the target prices, while profitable trading strategies are designed based on the forecast of target prices series. As a result, more than 70% success ratio could be achieved in terms of one-step ahead out-of-sample forecast of direction changes of the target price time series by taking the S&P500 index for example.

Statement of Originality

As far as the author is aware, Chapter 3, Chapter 4, and Chapter 5 of this thesis contains substantial parts which are original contributions to the area of forecasting financial time series. All material in this thesis which is not my own work has been properly acknowledged.

The following aspects of the thesis are believed to be original, with the most significant contributions considered to be:

1. The proposal of sufficient and necessary conditions that a time series can be predicted with zero errors by linear filters given the assumption that there are no structural breaks or regime switching in the time series. It is concluded that a band-limited time series can be predicted with zero errors by a predictive filter that has a constant amplitude response and constant group delay over the bandwidth of the time series. (Chapter 3)
2. The design of highpass and lowpass negative group delay filters that have approximately constant negative group delay and magnitude response in the passband. These negative group delay filters can be used to forecast band-limited signals whose bandwidths are within the passband of the filters with zero errors. (Chapter 3)
3. The design of linear predictive filters which maximise the spectrum flatness of forecast errors. Especially, the design of linear predictive filters that maximise

the Exponentially Weighted Standard Deviation Weighted Error Spectrum Flatness that can be used to forecast time series which have structural breaks in the in-sample data set. (Chapter 4)

4. The design of adaptive filters and hybrid filters implemented as a series of successive linear predictive filters, which can be used to forecast time series that have structural breaks in the out-of-sample data set. (Chapter 4)
5. The proposal of a target price concept which is defined as a linear combination of open, high, low and close prices that has maximum predictability. The most important property of a target price time series is that there are more low frequency components than high frequency ones in its spectrum. Also, an objective function is proposed to derive the target price series from historical observations of daily open, high, low and close prices. (Chapter 5)
6. The design of forecast procedures for forecasting real world financial time series which take into consideration of possible structural breaks. (Chapter 5)
7. The proposal of profitable trading strategies based on the one-step ahead forecast of target prices time series. (Chapter 5)

Acknowledgments

First of all, I would like to thank my PhD supervisors, Dr. Christos-Savvas Bouganis and Prof. Anthony Constantinides, for their priceless advice, boundless support, patience and encouragement. My special thanks go to Dr. Gerry Salkin for providing me the funding opportunity and valuable advice. Finally, I wish to express my great appreciation to my family for their support and encouragement in all aspects of my daily life.

Contents

1	Introduction	1
1.1	Financial Time Series	1
1.2	Motivation, Objectives and Contributions	4
1.3	Thesis Outline	6
2	Background	8
2.1	Predictability of Asset Prices	8
2.1.1	Efficient Market Hypothesis	9
2.1.2	Portmanteau Test	10
2.1.3	Variance Ratio Test	11
2.2	Spectrum Flatness Test	13
2.3	Time Series Forecasting Methodology	16
2.3.1	ARMA Models	17
2.3.2	Time Varying Coefficients Models	19
2.3.3	Artificial Neural Network	21

2.3.4	Wavelet Transform	22
2.3.5	Interval Time Series	24
2.4	Forecast Evaluation	26
2.4.1	Standard Statistical Measure	26
2.4.2	Success Ratio Test	28
2.4.3	In-Sample vs Out-of-Sample Test	29
2.5	Summary	30
3	Predictability of Time Series	32
3.1	Predictability of a Time Series	32
3.1.1	Necessary Condition	32
3.1.2	Sufficient Condition	34
3.2	Negative Group Delay	35
3.2.1	Group Delay	35
3.2.2	Group Delay of a Band-limited Signal	36
3.2.3	Predictability of a Filter in terms of Group Delay	37
3.2.4	Minimum Phase Filter	38
3.3	Perfect Forecast of Band-limited Signal using Negative Group Delay Filters	39
3.3.1	Design of Negative Group Delay Filters	39
3.3.2	Forecast Band-limited Signal using Negative Group Delay Filters	43
3.4	Conclusion	45

4	Linear Predictive Filters	47
4.1	Predictive Filter Theory	47
4.1.1	Linear Predictive Filter	48
4.2	Coefficients Estimation Methods	51
4.2.1	Least Square Errors	51
4.2.2	Maximum Spectrum Flatness	52
4.3	Spectrum Flatness Evaluation Methods	55
4.3.1	Error Spectrum Flatness with Periodogram	56
4.3.2	Standard Deviation Weighted Error Spectrum Flatness	57
4.3.3	Exponentially Weighted Standard Deviation Weighted Error Spec- trum Flatness	60
4.3.4	Simulation	63
4.4	Adaptive and Hybrid Filter Method	66
4.4.1	Adaptive Filter	67
4.4.2	Hybrid Filter	68
4.4.3	Comparison to Time Varying Coefficients Model	69
4.4.4	Simulation	70
4.5	Bounds of Forecast Errors	75
4.5.1	One-Step Ahead Forecast Error Bound	75
4.5.2	Naive Forecast Error Bound	77

4.5.3	Theorem	77
4.6	Conclusion	79
5	Forecasting Financial Time Series	80
5.1	Target Price	80
5.2	Optimise the Price Combination	82
5.2.1	Open, High, Low and Close Price	82
5.2.2	Intraday Price	88
5.3	Forecast Procedure	92
5.3.1	Financial Time Series Forecasting Procedure	92
5.3.2	Procedure of Applying the Linear Predictive Filter	95
5.4	Trading Strategy	97
5.5	Empirical Results and Analysis	101
5.5.1	Data	101
5.5.2	Goodness-of-fit	101
5.5.3	Profitability of the Trading Strategy	111
5.6	Summary	115
6	Conclusion and Future Work	117
6.1	Summary and Conclusion	117
6.2	Future Work	119

6.2.1	Regime Switching Model	120
6.2.2	Higher-Order Spectra	121
6.2.3	Exogenous Variables	122
6.2.4	Forecast Combination	123
6.2.5	Financial Time Series Combination	123
Bibliography		124
A Bounds of Forecast Errors		139
A.1	One-Step Ahead Forecast Bound	139
A.2	In-Sample One-Step Ahead Forecast Bound	140
A.3	Out-of-Sample One-Step Ahead Forecast Bound	141
A.4	Proof of Theorem	143
B Trading Strategies Backtesting on the S&P500		145
C List of Key Functions		147
C.1	Negative Group Delay Filter	147
C.2	Linear Predictive Filter	147
C.3	Forecast Financial Time Series	149

List of Figures

1.1	A time series of daily prices and returns of the S&P500	2
1.2	Magnitude response of daily returns of the S&P500	3
3.1	Magnitude response and Pole/Zero plot of a lowpass filter $H(z)$	41
3.2	Magnitude response and Pole/Zero plot of the minimum phase part $H_{\min}(z)$ of $H(z)$	41
3.3	Group delay of the minimum phase filter $H_{\min}(z)$	41
3.4	Magnitude response and group delay of the predictive filter $F(z)$	42
3.5	Input low frequency band-limited signal $\{x_n\}$ and its magnitude response	43
3.6	Output signal $\{\hat{x}_n\}$ and input low frequency band-limited signal $\{x_n\}$.	43
3.7	Magnitude response and Pole/Zero plot of the lowpass filter $\bar{H}_{\min}(z)$.	44
3.8	Magnitude response and group delay of the predictive filter $\bar{F}(z)$	44
3.9	Input high frequency band-limited signal $\{x_n\}$ and its magnitude response	45
3.10	Output signal $\{\hat{x}_n\}$ and input high frequency band-limited signal $\{x_n\}$	45
4.1	Lowband band-limited time series and its magnitude response	63

4.2	Highband band-limited time series and its magnitude response	63
4.3	Combined band-limited time series and its magnitude response	64
4.4	$S^W(\omega)$ of SDWESF and $S^E(\omega)$ of EWSDWESF	65
4.5	Out-of-sample forecast by ARMA and LPF	66
4.6	Coefficients of the adaptive filter (left) and the hybrid filter (right) . . .	71
4.7	Out-of-sample forecast of the adaptive filter with overlapping segments of length $L = 64$ (top) and $L = 32$ (bottom)	72
4.8	Coefficients of the adaptive filter with overlapping segments of length $L = 64$ (left) and $L = 32$ (right)	72
4.10	Coefficients of the hybrid filter with overlapping segments of length $L = 64$ (left) and $L = 32$ (right)	73
4.9	Out-of-sample forecast of the hybrid filter with overlapping segments of length $L = 64$ (top) and $L = 32$ (bottom)	74
5.1	The first order difference of close prices of the S&P500 and its magnitude response	81
5.2	Magnitude response of the in-sample first order difference of open/high/low/close prices of the S&P500	81
5.3	Close prices, typical prices and target prices of the S&P500	85
5.4	Magnitude response of the in-sample first order difference of typical and target prices of the S&P500	85
5.5	Effective lowpass filter	87

5.6	Close prices and target prices of the S&P500	89
5.7	Magnitude response of the in-sample first order difference of close/target prices of the S&P500	90
5.8	Forecast procedure for a price time series	94
5.9	In-sample and out-of-sample data set	95
5.10	Real world forecast procedure of a LPF	96
5.11	Model configuration reset and coefficients estimation if structure breaks	97
5.12	Target price trading strategy	99
5.13	Magnitude response and group delay of the LPF for the OHLC target price series	105
5.14	Magnitude response and group delay of the LPF for the HLC target price series	106
5.15	Coefficients of the adaptive filter and the hybrid filter for OHLC target prices of the S&P500	107
5.16	Coefficients of the adaptive filter and the hybrid filter for HLC target prices of the S&P500	107
5.17	ACF and PACF of the out-of-sample data set of OHLC target prices of the S&P500	108
5.18	ACF and PACF of the residue of the out-of-sample forecast by ARMA and LPF	109
5.19	Simulated price path and corresponding daily segments	114

List of Tables

2.1	Time series forecast methodologies	25
2.2	Statistical accuracy measures	26
3.1	Steps to design a negative group delay predictive filter $F(z)$	42
4.1	Comparison of three error spectrum flatness evaluation methods	62
4.2	Comparison of the goodness-of-fit tests of out-of-sample forecast	65
4.3	Comparison of fixed coefficient model, adaptive filter and hybrid filter .	69
4.4	Goodness-of-fit tests of the out-of-sample forecast of the adaptive Filter and the hybrid Filter	70
4.5	Comparison of goodness-of-fit tests of out-of-sample forecast	74
5.1	Flatness tests of the PSD of the first order difference of close/typical/target prices of the S&P500	86
5.2	Flatness tests of the PSD of the first order difference of target prices of other Combinations	87
5.3	Flatness tests of the PSD of the first order difference of close/typical/target prices	90

5.4	Flatness tests of the PSD of the first order difference of target prices of other combinations	91
5.5	Out-of-sample goodness-of-fit of OHLC target prices of the S&P500 . . .	103
5.6	Out-of-sample goodness-of-fit of HLC target prices of the S&P500 . . .	103
5.7	Out-of-sample goodness-of-fit of close prices of the S&P500	103
5.8	Out-of-sample goodness-of-fit of OHLC target prices of FTSE	110
5.9	Out-of-sample goodness-of-fit of OHLC target prices of SHCOMP . . .	110
5.10	Out-of-sample goodness-of-fit of OHLC target prices of EUR/USD . . .	110
5.11	P&L of the S&P500 in 2011	112
5.12	P&L of the simulated price path	114
B.2	Trading strategies backtesting on the S&P500	146

List of Terms and Acronyms

ANN Artificial Neural Network

ACF Autocorrelation Function

AIC Akaike's Information Criterion

AR Autoregressive

ARMA Autoregressive Moving Average

ARIMA Autoregressive Integrated Moving Average

BIC Bayesian Information Criterion

DA Directional Accuracy

DFT Discrete Fourier Transform

EMH Efficient Market Hypothesis

ESF Error Spectrum Flatness

EWMA Exponentially Weighted Moving Average

EWSDWESF Exponentially Weighted Standard Deviation Weighted Error Spectrum Flatness

FCAR Functional Coefficient Autoregressive

FFT Fast Fourier Transform

FIR Finite Impulse Response

FPE Final Prediction Error

GBM Geometric Brownian Motion

IIR Infinite Impulse Response

ITS Interval Time Series

LMS Least Mean Square

LPF Linear Predictive Filter

LTI Linear, Time-Invariant

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MASE Mean Absolute Scaled Error

MRAE Mean Relative Absolute Error

MSE Mean Square Error

OHLC Open, High, Low and Close prices

PACF Partial Autocorrelation Function

P&L Profit and Loss

PSD Power Spectral Density

RLS Recursive Least Square

RMSE Root Mean Square Error

RWH Random Walk Hypothesis

SDWESF Standard Deviation Weighted Error Spectrum Flatness

SR Success Ratio

TVC Time Varying Coefficients

VR Variance Ratio

WOSA Weighted Overlapped Segment Averaging

WT Wavelet Transform

Chapter 1

Introduction

1.1 Financial Time Series

A time series is a sequence of observations taken at successive times. Time series forecasting¹ methods attempt to discover patterns in historical data series and extrapolate these patterns into the future. Forecasting financial time series involves projection of time series such as stock prices/returns², interest rates, inflation, exchange rates, etc, into the future based on their historical values. Among them, forecasting asset prices is one of the most important and most widely discussed topics in financial economics. Understanding and forecasting asset prices are important for asset pricing and management, portfolio selection and optimization, option pricing, and risk management, etc. Forecast accuracy is of great importance to investors as forecast is used for decision making. However, assets prices are influenced by economical, political and even psychological factors [2] [61], which introduce high randomness into asset prices thus making forecasting of an asset's

¹ Forecast is a prediction or estimation of an actual value in a future time period. Forecast and prediction are typically used interchangeably.

² Return is defined as the profit or loss of holding an asset in a particular period

price a difficult task.

Financial time series, or more specifically asset price/return time series, are characterised by their complex nature. Firstly, asset price series behave nearly like random-walk processes and asset return series behaves very much like white noise processes (Figure 1.1 shows a typical time series of stock index prices and returns; Figure 1.2 shows the magnitude response of the return series). It implies that under this condition the prediction is theoretically impossible [57]. Secondly, asset prices are characterised by such features as non-linearity (see [1] and [51] for non-linearity) and high non-stationarity. Thirdly, asset price/return time series have structural instability/structural breaks³, or in other words, “regime switching”, which implies the behaviour of the series changes associated with events such as economic and financial crises or abrupt changes in government policy [52] [111]. More detailed reviews of stylised facts of asset prices and returns are give in [38] and [120].

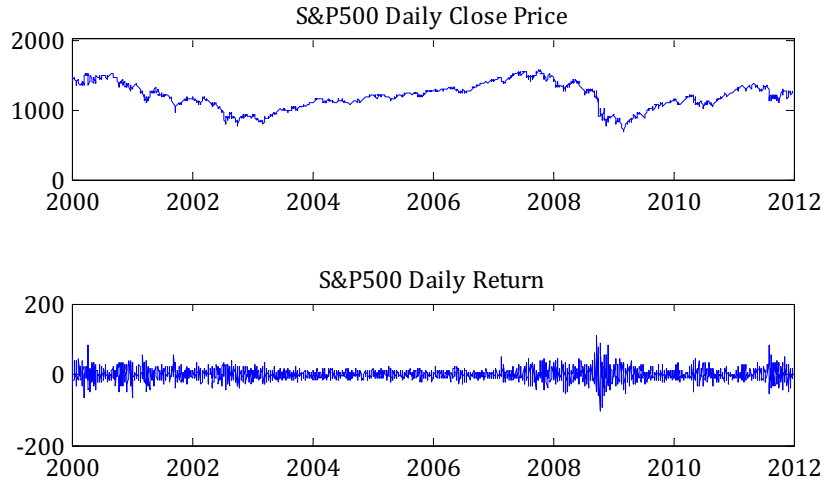


Figure 1.1: A time series of daily prices and returns of the S&P500

Because of aforementioned complex features of financial time series, forecasting financial time series has been regarded as one of the most challenging applications of

³ Structural break is defined in [83] as an unpredictable event in which the relationship among the variables in a model changes, and this change cannot be predicted in any sense from past data.

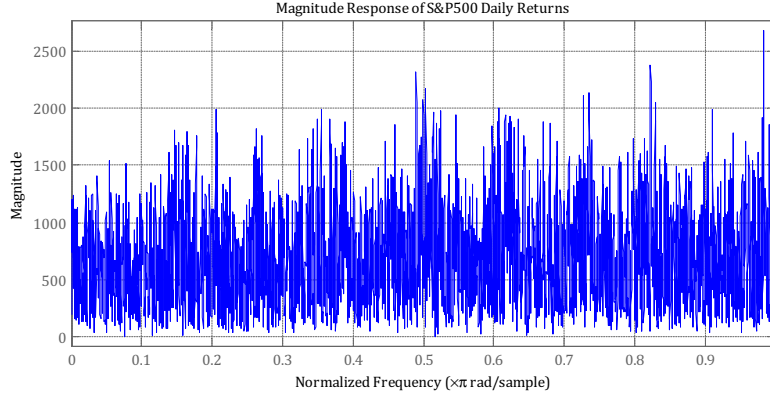


Figure 1.2: Magnitude response of daily returns of the S&P500

time series forecasting. For a long time, it was thought that asset prices could not be predicted by any methods better than the naive approach⁴ from historical observations, as current asset prices already reflect all the relevant information⁵. Thus the current price is the best predictor of the future price given only past prices. However, many studies have disputed the validity of the Efficient Market Hypothesis. It has been shown in numerous research papers that by applying suitable models can lead to somewhat successful predictions for asset returns, for example, the articles [24], [45] and more recently [124]. The predictability of financial time series is investigated and forecast methodologies are proposed for better prediction of asset prices than the naive approach in the thesis.

Most research has been focused on forecasting price levels of assets and most trading activities rely on price levels. However, recent articles show that the direction-of-change forecasts can be made with success [58] [101] [36] [122]. Moreover, some recent studies have suggested that trading strategies guided by forecasting direction changes of asset prices are more effective and may generate higher profits [73] [75] [94]. In this thesis not only the magnitude of asset prices/returns, but also the direction changes of the asset prices will be addressed in terms of forecasting financial time series.

⁴ The naive approach simply uses the asset price of today as the forecast for the price of tomorrow

⁵ Efficient Market Hypothesis (EMH) states that information is instantly and efficiently incorporated into asset prices, so it is impossible to use past information to foretell future price movements. The EMH is discussed in details in Section 2.1.1.

1.2 Motivation, Objectives and Contributions

Forecasting financial time series is regarded as one of the most challenging and important activities in finance and economics. It is the fundamental element of most investment activities as accurate forecast will significantly benefit investors by guiding decision making in terms of trading, asset management and risk management, etc., thus attracting attentions of practitioners and researchers for many decades.

Nevertheless, perfect asset price forecasts and easy gains from trading and asset management cannot and should not be expected because of the dynamic and complicated nature of financial time series. Two fundamental questions should be answered for any attempts to forecast a financial time series with a specific method and to benefit from forecasting the financial time series:

1. Under what circumstance a specific financial time series could be better predicted using a specific method than the naive approach based on its historical observations?
2. How to make profits by trading the underlying asset in the market taking advantage of the forecasts?

These two questions define two of the most important objectives of our research that we will be discussed in details in this thesis. To evaluate the prediction performance of a specific method in terms of forecasting financial time series, both asset price/return levels and direction changes of asset prices are considered. The most commonly used and natural threshold choice for the direction change is 50% [94], which implies that the proposed model is unable to forecast the future market directions correctly. Instead, the direction change predictability of proposed models is compared to that of the naive

forecast to evaluate their performance in this thesis⁶. If the success ratio⁷ of forecasting out-of-sample direction changes of the proposed method is higher than that of the naive forecast or other benchmark methods, the proposed method is considered to be better than the benchmark ones.

The purpose of this research is to investigate and develop novel methods for the prediction of financial time series considering their dynamic and complex nature. The predictive performance of asset prices time series themselves is exploited by applying digital signal processing methods to their historical observations. The novelty of the research lies in the design of predictive filters by maximising the flatness of forecast errors spectrum and applying the filters to forecast linear combinations of daily open, high, low and close prices of assets.

The main contributions of the thesis are as follows:

1. The proposal of sufficient and necessary conditions that a time series can be predicted with zero errors by linear filters given the assumption that there are no structural breaks or regime switching in the time series. It is concluded that a band-limited time series can be predicted with zero errors by a predictive filter that has a constant amplitude response and constant group delay over the bandwidth of the time series. (Chapter 3)
2. The design of highpass and lowpass negative group delay filters that have approximately constant negative group delay and magnitude response in the passband. These negative group delay filters can be used to forecast band-limited signals whose bandwidths are within the passband of the filters with zero errors. (Chapter 3)

⁶ For comparison, the best performing method in [94] on out-of-sample monthly data of the S&P500 from 01/1989 to 12/2006 could achieve 61% success ratio.

⁷ Success Ratio (SR) is defined as the proportion of times that the direction changes in a time series is correctly predicted by its forecast. SR is discussed in details in Section 2.4.2.

3. The design of linear predictive filters which maximise the spectrum flatness of forecast errors. Especially, the design of linear predictive filters that maximise the Exponentially Weighted Standard Deviation Weighted Error Spectrum Flatness that can be used to forecast time series with structural breaks in the in-sample data set. (Chapter 4)
4. The design of adaptive filters and hybrid filters implemented as a series of successive linear predictive filters, which can be used to forecast time series that have structural breaks in the out-of-sample data set. (Chapter 4)
5. The proposal of a target price concept which is defined as a linear combination of open, high, low and close prices that has maximum predictability. The most important property of a target price time series is that there are more low frequency components than high frequency ones in its spectrum. Also, an objective function is proposed to derive the target price series from historical observations of daily open, high, low and close prices. (Chapter 5)
6. The design of forecast procedures for forecasting real world financial time series which take into consideration of possible structural breaks. (Chapter 5)
7. The proposal of profitable trading strategies based on the one-step ahead forecast of target prices time series. (Chapter 5)

1.3 Thesis Outline

The remaining parts of the thesis consist of five chapters, organised as follows:

Chapter 2 provides background information on predictability of asset prices, which includes the popular efficient market hypothesis, market efficiency/random walk tests and spectrum flatness tests, literature reviews on current time series forecasting methodologies and model evaluation methods.

The sufficient and necessary conditions that a time series can be predicted with zero errors by linear filters are presented in Chapter 3. The predictability of a linear filter in terms of group delay is then exploited. Moreover, the methodology to design highpass and lowpass negative group delay filters that have approximately constant negative group delay and magnitude response in the passband is described.

Chapter 4 provides the predictive filter theory and coefficient estimation methods. The design of linear predictive filters by maximising the spectrum flatness of forecast errors, including Standard Deviation Weighted Spectrum Flatness and Exponentially Weighted Standard Deviation Weighted Spectrum Flatness are then presented, and simulation results are also given. In addition, the design of adaptive filters and hybrid filters is introduced for a time series that has structural breaks in the out-of-sample data set. Furthermore, a theorem that shows whether a proposed predictive filter is better than the naive forecast in terms of their upper bounds of forecast errors is also presented in this chapter.

A novel concept - target price time series is proposed in Chapter 5. An objective function is proposed to derive the target price series from historical observations of daily open, high, low and close prices which has maximum predictability. Chapter 5 also presents forecast procedures for forecasting real world financial time series. Furthermore, trading strategies are proposed based on the one-step ahead forecast of target price time series and applied to the S&P500 index. Empirical results and profitability analysis are also given.

Finally, Chapter 6 provides a summary of contributions of the thesis, as well as conclusions presented throughout the thesis. Also, it presents some ideas for further research, improvements and investigation, as a future work of the thesis.

Chapter 2

Background

2.1 Predictability of Asset Prices

Whether an asset's price can be predicted has been debated for decades. First defined in [44], the weak form *Efficient Market Hypothesis* (EMH) states that current asset prices already reflect all the relevant information and thus it is impossible to outperform the overall market by using past information. This implies that given only past prices, the current price is the best predictor of future prices, and the price change is expected to be zero. This is agreed with the *Random Walk Hypothesis* (RWH) [66], which suggests that the movement of stock prices follows a random walk¹ process. Therefore it is impossible to predict future asset price movements using past ones.

The EMH seems not to be completely convincing for many economists. Since stock prices clearly do not fully reflect a company's future performance in many cases, it is possible to make profits by predicting the short-term stock prices considering predictor variables such as the dividend yield, the price-earning ratio, or macroeconomic variables such as inflation and interest rates [18]. Moreover, recent work has disputed

¹ Here, random walk means that stock price changes are independent and identically distributed (i.i.d) which implies that the best prediction of future stock prices is the last available one.

the validity of EMH. Many studies analyse market efficiency by testing the stationarity of daily close prices or market values at given points of time [77]. It has been shown in numerous research papers that applying suitable models can lead to somewhat successful predictions. The early literature which gives evidences that stock prices/returns are predictable with the help of some financial variables including, e.g. [24], [26], [46], [45], [81] and more recently [74] and [124].

A widely used test of *weak-form efficiency* is to examine whether asset prices follow a random walk - a test that can be used with individual assets or stock indices. The RWH has two testable implications. First, asset prices are not predictable using past price information. Second, the variance of asset returns is linearly associated with the holding period. The former can be tested by examining the serial correlation structure of asset returns using Portmanteau tests (Section 2.1.2), while the latter using Variance Ratio tests (Section 2.1.3) [68].

2.1.1 Efficient Market Hypothesis

The EMH has been known as one of the cornerstones of modern financial economics. Fama [44] first defined an *efficient financial market* as “one in which security prices always fully reflect available information”. In a more general way, Jensen et al. [64] defined market efficiency as “a market is efficient with respect to information set θ_t if it is impossible to make economic profits by trading on the basis of information set θ_t ”. According to the EMH, information is immediately and efficiently incorporated into asset prices so that it is impossible to use past information to forecast future price movements. This implies that given only past prices and return data, the current price is the best predictor of future prices, and the price change is expected to be zero. Fama identified three levels of market efficiency [44]:

- *Weak form efficiency* states that prices of assets instantly and fully reflect all

information of past prices, which implies that future price movements are determined entirely by information not contained in the price series thus following a *random walk* process.

- *Semi-strong form efficiency* states that asset prices fully reflect all of the publicly available information. Therefore, only investors with additional inside information could have advantage on the market.
- The *strong form efficiency* states that asset prices fully reflect all of the public and inside information available. Therefore, no one can have advantage on the market in predicting prices since there is no data that would provide any additional value to the investors.

Kendall [66] was the first to suggest that the movement of stocks follows a random walk process, which led to the creation of the RWH, which is closely related to the weak form EMH.

2.1.2 Portmanteau Test

In time series analysis, the portmanteau test is a statistic test of autocorrelation in residuals of a forecast model. A common portmanteau test is the Box & Pierce Q Statistic [21],

$$Q = n \sum_{k=1}^h \gamma_k^2 \quad (2.1)$$

where γ_k is the autocorrelation coefficient for lag k (Equation 2.12), h the maximum lag being considered and n the number of observations in the series. Initially, the Box-Pierce Q test is designed for testing residuals from a forecast model. If residuals are white noise², the statistic Q has a chi-square (χ^2) distribution with $h - m$ degree

² white noise is a random signal with a flat PSD.

of freedom where m is the number of parameters in the model which has been fitted to the data. An alternative portmanteau test is the Ljung-Box Q^* Statistic [78],

$$Q^* = n(n+2) \sum_{k=1}^h (n-k)^{-1} \gamma_k^2 \quad (2.2)$$

Q^* has a distribution closer to the chi-square distribution than does the Q statistic.

The Q (or Q^*) test is designed for testing whether the first h autocorrelations of a time series (or residuals) are zero. However, when the series presents some kind of non-linear dependence, such as conditional heteroskedasticity³, the Q (or Q^*) test is no longer valid. Therefore, in order to verify a random walk process for a non-linear time series whose successive observations are possibly dependent even if uncorrelated, it is necessary to consider Variance Ratio tests.

2.1.3 Variance Ratio Test

The *Variance Ratio* (VR) test is widely used in literature to test market efficiency. It is based on the property that the variance of increments of a random walk process $\{x_n\}$ is linear in the sampling interval⁴. That is, the variance of $\{x_n - x_{n-2}\}$ is twice the variance of $\{x_n - x_{n-1}\}$. According to [79], the VR is defined as the ratio of $1/k$ times the variance of the k -period return of a price time series to the variance of the one period return, should be equal to 1 for all values of k . Suppose that x_n is an asset return at time n , where $n = 1, 2, \dots, N$. Following [133], VR, denoted as $VR(k)$, can be expressed as

$$VR(x; k) = \frac{\left\{ \frac{1}{Nk} \sum_{n=k+1}^N (x_n + x_{n-1} + \dots + x_{n-k} - k\hat{\mu})^2 \right\}}{\frac{1}{N} \sum_{n=1}^N (x_n - \hat{\mu})^2} \quad (2.3)$$

³ Heteroskedasticity means the variance of a series changes over time

⁴ $VR = 1$ is a necessary, but not a sufficient condition of random walk as time series recognised as non-random walk by VR test are only subset of non random walk processes.

where $\hat{\mu} = \frac{1}{N} \sum_{n=1}^N x_n$. The RWH requires that $VR = 1$. An estimated $VR < 1$ implies negative serial correlation (mean reversion), while $VR > 1$ implies positive serial correlation (mean aversion). In the following subsections, the notation used in [59] is adopted to formulate two conventional VR tests, i.e. Lo-MacKinlay VR test and Chow-Denning VR test.

2.1.3.1 Lo-MacKinlay VR Test

If $\{x_n\}$ is *independent and identically distributed* (i.i.d), then under the null hypothesis that $VR(k) = 1$, the test statistic

$$M_1(x; k) = (VR(x; k) - 1) \left(\frac{2(2k - 1)(k - 1)}{3kN} \right)^{-\frac{1}{2}} \quad (2.4)$$

follows the standard normal distribution asymptotically. If $\{x_n\}$ exhibits conditional heteroscedasticity, under null hypothesis that $VR(k) = 1$, the test statistic

$$M_2(x; k) = (VR(x; k) - 1) \left(\sum_{j=1}^{k-1} \left[\frac{2(k-j)}{k} \right]^2 \delta_j \right)^{-\frac{1}{2}} \quad (2.5)$$

where

$$\delta_j = \frac{\sum_{n=j+1}^N (x_n - \hat{\mu})^2 (x_{n-j} - \hat{\mu})^2}{\left[\sum_{n=1}^N (x_n - \hat{\mu})^2 \right]^2} \quad (2.6)$$

follows the standard normal distribution asymptotically as well. By applying the test, Lo & Mackinlay [79] rejected the random walk hypothesis for weekly stock market returns of U.S. Extensive Monte Carlo results reported in [80] suggest that, the M_2 test (Equation 2.5) performs better than the Box-Pierce test of serial correlation.

2.1.3.2 Chow-Denning VR Test

The Lo-MacKinlay test is an individual test where the null hypothesis is tested for an individual value of k . Its weakness lies in that it ignores the joint nature of testing for

the RWH thus the probability of incorrect rejection of the true null hypothesis can be much larger than the chosen level of significance [59]. To avoid this problem, Chow & Denning [35] proposed a joint test as follows.

Under the null hypothesis, $VR(k_i) = 1$ for $i = 1, \dots, l$ against the alternative hypothesis that $VR(k_i) \neq 1$ for some i . The Chow-Denning test statistic is

$$MV_1 = \sqrt{N} \max_{1 \leq i \leq l} |M_1(x; k)| \quad (2.7)$$

where $M_1(x; k)$ is defined in Equation 2.4. Similarly, the heteroskedasticity-robust version of the Chow-Denning test MV_2 can be written as

$$MV_2 = \sqrt{N} \max_{1 \leq i \leq l} |M_2(x; k)| \quad (2.8)$$

where $M_2(x; k)$ is defined in Equation 2.5. The Chow-Denning test is based on the idea that the decision regarding the null hypothesis can be made based on the maximum absolute value of the individual VR statistics. The null hypothesis is rejected at α level of significance if the MV_1 statistic is greater than the $(1 - (\alpha^*/2))$ th percentile of the standard normal distribution where $\alpha^* = 1 - (1 - \alpha)^{1/l}$.

2.2 Spectrum Flatness Test

The power spectrum is able to reveal repetitive patterns and correlation structures in a signal process. According to [128, Chapter 9], the more correlated or predictable a signal⁵ is, the more concentrated its power spectrum is and, conversely, the more random or unpredictable a signal, the wider the spread of its power spectrum is. Therefore the power spectrum of a signal can be used to exploit predictive patterns in the signal process, thus it is crucial in time series forecasting.

⁵ Signal is used interchangeable with time series in the thesis.

The spectrum flatness test is initially designed to test the departure of spectral density of estimation errors from a constant value. According to [43], if the time series $\{x_n\}$ is independent and identically distributed (i.i.d), its spectral density would be a constant, while the converse is not true except where $\{x_n\}$ is a Gaussian series. Therefore the test can be used to check for independence only in the Gaussian case while serial uncorrelation otherwise.

Denote $\varphi(\theta)$ the *power spectral density* (PSD), the distance of randomness [43] is defined as

$$D(\varphi) = \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\theta) d\theta \right) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \varphi(\theta) d\theta \quad (2.9)$$

as a measure of spectrum flatness.

In the discrete form, $D(\varphi)$ can be represented as

$$D(\varphi) = \log \left(\frac{1}{N} \sum_{k=0}^{N-1} \varphi_k \right) - \frac{1}{N} \sum_{k=0}^{N-1} \log \varphi_k \quad (2.10)$$

where φ_k is the sample power spectral density

$$\varphi_k = \sum_{n=-N}^N \gamma_n e^{-\frac{2\pi i}{2N+1} kn} \quad (2.11)$$

and γ_n is the sample *autocorrelation function* (ACF)

$$\gamma_n = \frac{1}{N - |n|} \sum_{m=0}^{N-1} x_m x_{m+n}, \quad n = 0, \pm 1, \pm 2, \dots, \pm (N-1) \quad (2.12)$$

$D(\varphi)$ has the following properties:

1. $D(\varphi) > 0$ for all $\varphi(\theta) > 0$, $\theta \in [0, 2\pi]$
2. $D(\lambda\varphi) = D(\varphi)$ for all $\lambda > 0$
3. $D(\varphi) = 0$ if and only if $\varphi(\theta) = c$, where c is some positive constant.

The first property is justified by Jensen's inequality⁶. The third property shows that the minimum $D(\varphi) = 0$ is reached if and only if the power spectral density is a positive constant over the whole bandwidth of its spectrum, which implies white noise⁷.

McElroy & Holan [88] improved $D(\varphi)$ by proposing a spectral variance of the logged spectral density to measure a spectral density's departure from constancy

$$\psi(\varphi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log^2 \varphi(\theta) d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \varphi(\theta) d\theta \right)^2 \quad (2.13)$$

$\psi(\varphi)$ can be represented in the discrete form as

$$\psi(\varphi) = \frac{1}{N} \sum_{k=0}^{N-1} \log^2 \varphi_k - \left(\frac{1}{N} \sum_{k=0}^{N-1} \log \varphi_k \right)^2 \quad (2.14)$$

$\psi(\varphi)$ has the same properties as $D(\varphi)$ while an exclusive property: $\psi(\varphi) = \psi(\frac{1}{\varphi})$, which has the benefit that if $\varphi(\theta)$ is the spectrum of model errors, peaks and valleys in $\varphi(\theta)$ are equally persuasive in indicating departures from whiteness.

Taking advantage of both spectrum flatness tests, we could determine the randomness of a time series in terms of its power spectral density's departure from constancy and exploit the predictability of a time series from its power spectrum.

⁶ if f is a convex function on $[a, b]$, for $\{x_n\}_{n=1}^N \in [a, b]$, and $\{p_n\}_{n=1}^N$ with $p_n \geq 0$ and $\sum_{n=1}^N p_n = 1$, then $f\left(\sum_{n=1}^N p_n x_n\right) \leq \sum_{n=1}^N p_n f(x_n)$ with equality iff $p_1 = p_2 = \dots = p_n$. The inequality is reversed if f is concave.

⁷ White noise is defined as a random process with equal power at all frequencies. In practice, we consider a band-limited noise process, with a flat spectrum covering the defined frequency band of a white noise process.

2.3 Time Series Forecasting Methodology

According to [87], there are two major types of forecasting models: time series and explanatory models. Time series models treat a model as a black box and make no attempt to discover factors affecting behavior of a time series, while explanatory models assume that a variable to be forecasted exhibits an explanatory relationship with one or more independent variables thus trying to discover the form of the relationship and use it to forecast future values of the variable. In this research, we only consider time series models by applying signal processing methods.

In the context of time series forecasting, linear time series prediction methods analyse historical data and attempt to approximate future values of a time series as a linear combination of historical data. Linear prediction models [126] and ARMA models are both linear models that can be used for forecasting asset prices. Since linear prediction model is equivalent to the *autoregressive* (AR) model, which is a special case of ARMA model, we only discuss ARMA models in this section. In addition, there are other existing models and techniques in literature that can be used for time series forecast, such as time-varying coefficients model, artificial neural network and wavelet transform. A selected literature review is listed in this section. A comprehensive review of time series forecasting is given in [41]. Moreover, more than 100 related published articles that focus on techniques applied to forecast stock markets are surveyed and summarised in [13].

George Box famously wrote that “essentially, all models are wrong, but some are useful” [19]. It implies that none of these models can completely represent the reality, but some of them are useful to interpret the reality. Accordingly, the different time series forecasting methodologies are able to explain different aspects of real world time series, or certain types of real world time series. But none of them can be applied universally to all the real world time series. Therefore, in reality, it is necessary to carefully select a forecast methodology for a certain time series.

2.3.1 ARMA Models

First proposed in [20], the *Autoregressive Moving Average* (ARMA) model provides a parsimonious representation of any stationary stochastic process. Given a time series $\{x_n\}$, the ARMA model consists of two parts, an *autoregressive* (AR) part and a *moving average* (MA) part. The model is usually referred to as the ARMA(p, q) model where p is the order of the *autoregressive* part and q the order of the *moving average* part.

AR(p) refers to the autoregressive model of order p .

$$x_n = \sum_{i=1}^p \phi_i x_{n-i} + e_n \quad (2.15)$$

where ϕ_i is a constant and e_n white noise error. If we define an autoregressive operator of order p by $\phi(B) = 1 - \sum_{i=1}^p \phi_i B^i$, the AR model can be written economically as $\phi(B)x_n = e_n$, where B is the *backward shift operator*, which is defined by $Bx_n = x_{n-1}$, then $B^m x_n = x_{n-m}$.

MA(q) refers to the moving average model of order q :

$$x_n = \sum_{i=1}^q \theta_i e_{n-i} + e_n \quad (2.16)$$

where θ_i is a constant and e_n the error term. If we defined a moving average operator of order q by $\theta(B) = 1 + \sum_{i=1}^q \theta_i B^i$, the MA model can be written economically as $x_n = \theta(B)e_n$.

ARMA(p, q) refers to the model with p autoregressive terms and q moving average terms. This model contains the AR(p) and MA(q) models,

$$x_n = \sum_{i=1}^p \phi_i x_{n-i} + \sum_{i=1}^q \theta_i e_{n-i} + e_n \quad (2.17)$$

or

$$\phi(B)x_n = \theta(B)e_n \quad (2.18)$$

Equation 2.17 can also be represented in z -domain as

$$\left(1 - \sum_{i=1}^p \phi_i z^{-i}\right) X(z) = \left(1 + \sum_{i=1}^q \theta_i z^{-i}\right) E(z) \quad (2.19)$$

Therefore, an ARMA series is the output of filtering a white noise process $\{e_n\}$ with a causal linear filter $H(z)$ with p poles and q zeros

$$H(z) = \frac{1 - \sum_{i=1}^p \phi_i z^{-i}}{1 + \sum_{i=1}^q \theta_i z^{-i}} \quad (2.20)$$

By allowing differencing of the time series, the ARMA model can be extended to an *Autoregressive Integrated Moving Average* (ARIMA) model that is able to handle non-stationary series with trends and seasonal components. The model is generally referred to as an ARIMA(p, d, q) model where d refers to the order of integrated parts of the model.

$$\phi(B)\nabla^d x_n = \theta(B)e_n \quad (2.21)$$

where the *backward differencing operator* ∇ is defined as:

$$\nabla x_n = x_n - x_{n-1} = (1 - B)x_n \quad (2.22)$$

Some well-known special cases arise naturally. For example, an ARIMA(0, 1, 0) model is given by $x_n = x_{n-1} + e_n$, which is a *random walk* process. Also, ARIMA(0, 0, 0) is simply a *white noise* process. If a seasonal effect is suspected in the model, a SARIMA (seasonal ARIMA) model in [20] can help capture the seasonality in a time series. The general form of SARIMA model is denoted as ARIMA(p, d, q)(P, D, Q) $_s$, where s is the number of periods per season.

For order identification and model selection, *Akaike's Information Criterion* (AIC) proposed in [5] is normally employed to overcome the problem of *overfitting*, which implies that the *Mean Square Error* (MSE) can be made smaller simply by increasing the number of parameters in the model in the ARIMA case. *Bayesian Information Criterion* (BIC) [112] and *Final Prediction Error* (FPE) [4] are also commonly used for model selection (See [22] for details). The ARMA model are applied in the early research for forecasting financial time series such as [48], [116], [25], and most recently for forecasting emerging market stocks by [91], [33], [113] and [28], etc.

2.3.2 Time Varying Coefficients Models

Time varying coefficients model is a particular class of functional coefficient models or more generally varying coefficients models. A selective overview on the major methodological and theoretical development on the varying coefficient models are given in [47]. In the simplest form, a *functional coefficient autoregressive* (FCAR) models is an AR model in which AR coefficients are allowed to vary as a function of another variable, such as lagged value of time series itself (e.g. Equation 2.24) or a variable exogenous to the time series. The functional form is usually left unspecified and estimated non-parametrically using kernel methods. These types of models are first introduced in [30] and have been further investigated, for example, in [31] and [23]. The simplest time-varying coefficients can be written as

$$x_n = \alpha_n x_{n-1} + \epsilon_n \quad (2.23)$$

$$\alpha_n = a_0 + a_1 \alpha_{n-1} + \nu_n \quad (2.24)$$

where α_n is time-varying AR coefficient, ϵ_n and ν_n are white noise terms. One of the advantages of time-varying coefficient model is that non-linear time series can be represented by a deterministic time-varying coefficient model without first specifying

the nature of the non-linearity.

The time variation in coefficients is tested in [6] in an ad-hoc way by splitting their entire sample into different sub-periods. The authors clearly document the time varying pattern of coefficients, and find stock returns are predictable by dividend yields and short rates at short horizons. In [40], the authors show that empirical evidence supports the existence of time variation in regression coefficients and identified in-sample stock return predictability but failed to unambiguously show the existence or non-existence of out-of-sample predictability. In term of forecast, the multi-step ahead forecasting using uni-variate and multivariate functional coefficient VFCAR model is presented in [54]. Their empirical results indicate that the bootstrap method appears to give slightly more accurate forecast results than naive plug-in predictor and multistage predictor. In contrast to conventional time-varying coefficient models, A F-ARMA model is proposed in [82] and the authors argue that under very weak conditions the behavior of the AR coefficient can be exactly represented by a sufficiently long Fourier series. To forecast non-stationary process, Korale & Constantinides [70] propose an Endomorphic model which is a multistage analysis where the parameters are decomposed into successive AR processes whose coefficients are updated via adaptive equations constrained on prediction errors.

One potential limitation of time varying coefficients models is that cumulative errors from multistage estimations, i.e., ϵ_n and ν_n , etc.. could be greater than errors from a constant coefficients model because it is more difficult to specify and estimate varying coefficients models than constant coefficients ones. Thus it is necessary to consider the tradeoff between the gain from forecasting using the model and the difficulty of specifying and estimating the model for a specific time series.

2.3.3 Artificial Neural Network

Application of *Artificial Neural Network* (ANN) in forecasting financial time series is dramatically increasing in recent years. The main idea of ANNs is that inputs get filtered through one or more hidden layers each of which consists of hidden units, or nodes, before they reach the output. The intermediate output is related to the final output [41].

The most important advantage of an ANN is its ability to learn from data through adaptively changing its structure based on external or internal information that flows through the network during the learning phase and generates output variables based on its learning. Another valuable quality is the non-linear nature of ANNs. ANNs are non-linear statistical data modeling tools which can be used to model complex relationships between inputs and outputs or to find patterns in data. Therefore, ANNs can adapt to irregularities and unusual features in a time series of interest in situations where an explicit model-based approach fails.

ANNs have been popularly applied for forecasting financial time series. A good early survey is given in [135] and a comprehensive review of the ANNs and their applications in various aspects in finance such as bonds, inflation, bank failures, etc., is given in [89]. Qi & Zhang [105] investigate how to best model trend using ANNs and apply ANNs to forecast the real gross national product (GNP) series. They find that differencing data first is the best practical approach to build an effective ANN forecasting model for most real-world time series. ANNs are used in [97] for one-step ahead prediction of weekly Indian rupee/US dollar exchange rate, and the authors find that ANNs have superior in-sample forecast than linear autoregressive and random walk models - ANNs outperform random walk by five out of six evaluation criteria and beat the linear autoregressive model by four out of six evaluation criteria in out-of-sample forecasting. A hybrid model combining ARIMA and ANN is used in [67] to forecast GBP/USD exchange rate, and they find that for short-term forecasting (1 month),

both ANN and hybrid models are much better in accuracy than the simple random walk model. The ANN model gives a comparable performance to the ARIMA model and the hybrid model outperforms both ARIMA and ANN models for longer time horizons (6 and 12 month). More hybrid forecasting models using ARIMA and ANNs have been proposed and applied to forecast financial time series with good prediction performance, for example, [9], [96], [29], [129], etc.

A disadvantage of ANNs is that they do not allow much understanding of data because there are not explicit models. They provide a “black box” approach to forecast. Another drawback of ANN methods is the danger of overfitting problem of in-sample training data [72]. The ANN trained on in-sample data performs reasonably well in terms of in-sample goodness-of-fit tests, for out-of-sample ones only if there are no structural breaks in the out-of-sample data set. Other drawbacks that have been criticised include excessive training time, and large number of parameters that must be experimentally selected to generate good forecast [15].

2.3.4 Wavelet Transform

Wavelet analysis is capable of revealing aspects of data that other signal analysis techniques miss, aspects like trends, breakdown points, discontinuities in higher derivatives and self-similarity. This makes it suitable for the analysis of non-linear and non-stationary financial time series. Moreover, the wavelet transform is able to decompose a time series into multiple resolution constituent time series. According to [49], wavelet methods provide insight into the dynamics of financial time series beyond that of standard time series methodologies. In addition, the authors of [76] argue that since the wavelet coefficients obtained can indicate local characteristics of a non-stationary time series at the time-scale space, to identify system states, in practice, one often extracts features based on wavelet coefficients. According to [106], the benefits of a wavelet approach to the analysis of economic and financial data include

the flexibility in handling very irregular data series, time-scale decomposition of data and a non-parametric representation of each individual time-series, and determining whether a time series can be forecasted at corresponding forecast horizon. A selective review on wavelet techniques is provided in [106] in these fields.

Early approaches of forecasting using wavelets are discussed in [8] and [14], respectively. The main idea of [8] is to decompose a signal into its time-scale components and then to treat each approximation at each time-scale as a separate series. Each component is forecasted using the ARIMA method. The final forecast for the complete series is obtained by adding up component forecasts. The research focuses on three components: trend, seasonal fluctuations and noise. The major innovation of [14] is to analyse individual time decompositions by ANNs and to base forecasts on neural-network estimates. In [108], the original time series is divided to multiresolution ones and then forecasted separately. These forecasts are then combined to achieve an aggregate forecast for the original time series. Very similar ideas are investigated in [92]. They apply a discrete wavelet transform to decompose a time series and then to forecast independently at each resolution level. The results indicate that the multiresolution approaches outperform the traditional single resolution approach in modeling and forecasting. There are a few other applications of wavelet transform to forecast financial time series in recent research, for example, [60], [110], [71], etc.

The major disadvantage of decomposition with wavelet transform is that cumulative errors from forecasts at each resolution level could be greater than those of forecasting the original time series directly. Therefore, it is necessary to investigate under which situation a decomposition method with wavelet transform outperform single resolution approaches.

2.3.5 Interval Time Series

The early research on time series forecasts has focused on single-valued time series, i.e. series where every observation at each time point is a single value, such as daily close price. Over recent years the modelling and forecasting interval time series (ITS) has received considerable attention. An interval of a financial time series is defined by its upper and lower bounds, i.e., the daily high H and low prices L , as $x_n = [L_n, H_n]$, or equivalently by its center $c_n = \frac{H_n + L_n}{2}$ and radius $r_n = \frac{H_n - L_n}{2}$, as $y_n = \langle c_n, r_n \rangle$. An interval time series $\{x_n\}$ is then a time series where the variable observed through time is an interval variable.

One of the main advantages of interval time series forecasting is that it avoids the major drawback of point forecast of single-value financial time series, which is the neglect of volatility and variability information reflected by daily high, low prices. For example, for a given asset, the historical volatility information could be used as explanatory variable to forecast the last interval variables. The variability information which is changes of prices in a daily session (i.e., intraday prices) would give more information than the daily close price that neglects the intraday variability to forecast a financial time series.

The primer approach to forecast interval time series is presented in [12]. “ITS should be expressed in terms of their upper and lower bounds, or of their center and radius series. Each of these series should be independently analysed using classical time series analysis methods to find the pattern (trend, cycle and seasonality) including possible non-linearity. Then, it should be determined which pair of series is going to be forecasted and the forecasting method for each one; moreover, if appropriate, a multivariate method can be applied. Then, the value of the parameters of the chosen method should be estimated minimizing an ITS error measure in the training set. Finally, the accuracy of the calibrated method has to be corroborated with the test set.”

Besides the primer approach, different approaches have been introduced to forecast interval financial time series. The linear interval methods are proposed in [53] and [56] to forecast annual and quarterly S&P500 index variability, and weekly GBP/USD spot rate, respectively, and both found that the forecasting accuracy is significantly higher than the point forecast. The uni-variate or multivariate forecasting methods are applied in [11] using classic forecasting methods include exponential smoothing, the k-NN algorithm and the multilayer perception to forecast the daily DJIA index and EUR/USD spot rate, and the forecast results agree with the results obtained in [56]. A further study of uni-variate methods can be found in [85], where ITS are forecasted using ARIMA models, hybrid ARIMA and neural network models in [134], and a non-linear threshold model in [109]. Additional information on multivariate models is available in [34] and [27].

The advantages and disadvantages of aforementioned methodologies can be briefly summarised in Table 2.1:

	Advantages	Disadvantages
ARMA	explicit model specification effectively an IIR filter	cannot catch dynamics/non-linearity
TVC	catch dynamics/nonstationarity/non-linearity explicit model specification analyse data at multistage	cumulative errors from the multistage estimation inherited disadvantages of applied models at multistage
ANN	catch dynamics/nonstationarity/non-linearity catch in-sample structural breaks	no explicit model specification overfitting excessive training time
WT	catch dynamics/nonstationarity/non-linearity analyse data at multiresolution	cumulative errors from the multiresolution estimation inherited disadvantages of applied models at multiresolution
ITS	include volatility and variability information	inherited disadvantages of applied models

Table 2.1: Time series forecast methodologies

2.4 Forecast Evaluation

2.4.1 Standard Statistical Measure

Forecasting accuracy refers to “goodness of fit”, which refers to how well the forecasting model fit a set of observations. Forecasting accuracy is regarded as an “optimist’s term for forecast errors” in [10]. A forecast error represents the difference between the forecast value and the actual value. If x_n is the actual observation at time n and \hat{x}_n is the forecast for the same period, then the forecast error is defined as

$$e_n = x_n - \hat{x}_n \quad (2.25)$$

Usually, \hat{x}_n is the *one-step* ahead forecast so that e_n is the one-step ahead forecast error.

According to [63], there are five categories of accuracy measures: scale-dependent measures, measures based on percentage errors, measures based on relative errors, relative measures and scaled errors. The commonly available statistical accuracy measures are summarised in category in Table 2.2:

Abbrev.	Name	Category
MSE	Mean Square Error	scale-dependent measures
RMSE	Root Mean Square Error	scale-dependent measures
MAE	Mean Absolute Error	scale-dependent measures
MAPE	Mean Absolute Percentage Error	measures based on percentage errors
MRAE	Mean Relative Absolute Error	measures based on relative errors
Theil’s U	Theil’s U -statistic	relative measures
MASE	Mean Absolute Scaled Error	scale error

Table 2.2: Statistical accuracy measures

The scale-dependent measures MSE, RMSE and MAE are commonly used when comparing different methods on the same set of data. However it is not applicable when comparing forecast performance across different data sets. MAPE has the advantage

of being scale-independent and thus being frequently used to compare forecast performance across different data sets. However, it is infinite or undefined if $x_n = 0$ for any n . The relative error measure MRAE has the same drawback as percentage error measure. Therefore, one can use relative measure Theil's U -statistic proposed in [121] rather than relative errors. The U -statistic essentially compares the performance of a forecast against a naive one-step ahead forecast. The problem of relative measures is that it can only be computed when there are several forecasts on the same series so that it cannot be used to measure out-of-sample forecast accuracy at a single forecast horizon [63]. To solve the problem, Hyndman & Koehler [63] proposed a scale error model by scaling the absolute error based on the in-sample MAE from a benchmark forecast method. Assuming the benchmark method is the random walk model, then a scale error is defined as

$$q_n = \frac{e_n}{\frac{1}{N-1} \sum_{i=2}^N |x_i - x_{i-1}|} \quad (2.26)$$

which is independent of the scale of data. A scaled error is less than one if it arises from a better forecast than the average one-step benchmark forecast computed in-sample. Conversely, it is greater than one if the forecast is worse than the average one-step benchmark forecast computed in-sample. The Mean Absolute Scaled Error is defined as

$$\text{Mean Absolute Scaled Error (MASE)} = \text{mean}(|q_n|) \quad (2.27)$$

The only circumstance under which the MASE would be infinite or undefined is when all historical observation $\{x_i\}$ are equal.

In addition to the standard statistical measures, the Diebold-Mariano test proposed in [42] compares the forecast accuracy of two forecast methods based on the null hypothesis of no difference in the accuracy of two alternative forecasts. The Diebold-Mariano statistic is a significance test on whether a forecast method significantly outperforms the benchmark with respect to out-of-sample prediction errors. However, the Diebold-Mariano statistic can only be used for model selection when a forecast

method is “significantly” better than the benchmark, thus it is not useful for method comparison when forecasting the magnitude of a financial time series using linear models which often have similar performance. It is found in [39] that basing the choice of prediction models on such significance tests is problematic, as this practice may favor the null model, usually a simple benchmark.

Considering the pros and cons of those statistical accuracy measures, MSE are chosen for performance comparison on the same set of data and MASE for comparison across data sets in the research.

2.4.2 Success Ratio Test

Distinct from other time series, direction changes of asset prices are of great interest. Sometimes financial practitioners are more interested in direction changes of future asset prices than their magnitudes as trading strategies guided by forecasting direction changes of asset prices are more effective and may generate higher profits [73] [75] [94]. So the *Success Ratio* (SR) test is extremely important to measure the performance of different forecast models. The SR test is based on the proportion of times that the direction of change in a price time series $\{x_n\}$ is correctly predicted by the forecast $\{\hat{x}_n\}$ ⁸

$$\text{SR} = \frac{1}{N} \sum_{n=1}^N \text{HS}((x_{n+1} - x_n) \cdot (\hat{x}_{n+1} - x_n)) \quad (2.28)$$

where N is the length of the price time series $\{x_n\}$, \hat{x}_n the prediction of x_n , and HS is the modified Heaviside function defined as:

$$\text{HS}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.29)$$

⁸ The definition of SR is conservative as it does not count the case when $x_{n+1} = \hat{x}_{n+1} = x_n$.

If $\{x_n\}$ is a return time series, SR is defined as⁹,

$$\text{SR} = \frac{1}{N} \sum_{n=1}^N \text{HS}(x_{n+1} \cdot \hat{x}_{n+1}) \quad (2.30)$$

The ultimate goal of financial time series forecasting is to gain a significant outcome from the viewpoint of profit maximisation rather than minimisation of statistical measures. Therefore, in the context of financial time series forecasting, the out-of-sample sign test that is tied to profitability of one's trading decision making is more important than statistical measures and significance tests. The *Directional Accuracy* (DA) Test is a nonparametric significance test that is proposed in [99] which focuses on the correct prediction of direction changes in a time series. The DA Test is further improved in [50] and [101] to exploit economic value of the forecast and deal with structural breaks in a data sample. The DA test is a popular test on direction changes in financial time series, e.g., see [100], [106], [104], [102], etc.

2.4.3 In-Sample vs Out-of-Sample Test

For a given forecasting method, the *in-sample* test which fits a model of interest using all available historical data is likely to understate forecasting errors thus overestimating its predictability. With in-sample test, model selection and estimation are designed to calibrate a forecasting procedure to the historical data, whose pattern may not persist into the future. Moreover, models selected by best in-sample fit may not best predict out-of-sample data. Therefore, *out-of-sample* tests are used by forecasters to assess the goodness-of-fit of proposed models. An out-of-sample evaluation of forecast accuracy begins with the division of the historical data series into a fit period (in-sample data set) and a test period (out-of-sample data set). The fit period is used to identify and

⁹ It is worth noticing that the naive forecast of a return time series has $\hat{x}_{n+1} = 0$ for all n , therefore $\text{SR} = 0$ according to this definition. In this thesis, $\text{SR} = \frac{1}{N} \sum_{n=1}^N \text{HS}(x_{n+1} \cdot x_{n+2})$ are used as SR of naive forecast

estimate a model while the test period is reserved to assess the model's forecasting accuracy. A good review of out-of-sample tests for forecasting accuracy is given in [119].

2.5 Summary

This chapter provided background information on predictability of asset prices, which includes the popular efficient market hypothesis, market efficiency/random walk tests and spectrum flatness tests. The literature review on current time series forecast methodologies and model evaluation methods is also given.

The EMH states that current asset prices already reflect all the relevant information and thus it is impossible to outperform the overall market using past information. Therefore, forecasting financial time series makes sense only if the series is not a random walk process. Portmanteau tests and variance ratio tests can be used to examine whether asset prices follow a random walk process.

The power spectrum of a signal can be used to exploit predictive patterns in the signal process. The more correlated or predictable a signal is, the more concentrated its power spectrum is and, conversely, the more random or unpredictable a signal is, the wider the spread of its power spectrum spreads. Therefore spectrum tests can be used to test the predictability of a time series in its frequency domain.

The advantages and disadvantages of current time series forecasting methodologies are summarised in Table 2.1. The proposed methodologies in the thesis will try to overcome problems such as inability to catch dynamics and non-linearity of a time series, no understanding of data, etc., while taking advantage of the explicit model specification and variability information of asset prices, etc.

The magnitude of asset prices/returns and direction changes of asset prices are both

important in terms of guiding trading strategies by forecasting financial time series. Therefore, MSE, RMSE and MASE are chosen for performance comparison on asset price/return levels forecast, while SR and DA are chosen for performance evaluation of direction changes forecast. These goodness-of-fit tests are applied to the out-of-sample data set.

Chapter 3

Predictability of Time Series

3.1 Predictability of a Time Series

The fundamental assumption of forecasting time series is that a pattern existing in historical observations will continue existing in the future. It implies that there must be no structural breaks or regime switching in the time series in order to be predicted. Based on this assumption, we derive the sufficient and necessary conditions under which a time series can be perfectly predicted¹ without errors.

3.1.1 Necessary Condition

In the z -domain, the predictive filter output, which is the one-step ahead forecast signal $\hat{X}(z)$, is the product of the input signal $X(z)$, and the filter frequency response $F(z)$

$$\hat{X}(z) = F(z)X(z) \tag{3.1}$$

¹ The term perfectly predict or completely predict refer to the situation where a time series can be predicted with no errors.

The one-step ahead forecast error signal $\bar{E}(z) = zE(z)$ is defined as the difference between the desired signal $\bar{X}(z) = zX(z)$ (i.e., $\{x_{n+1}\}$), and the filter output $\hat{X}(z)$,

$$\begin{aligned}\bar{E}(z) = zE(z) &= \bar{X}(z) - \hat{X}(z) \\ &= (z - F(z))X(z) \\ &= H(z)(zX(z)) \\ &= zH(z)X(z)\end{aligned}\tag{3.2}$$

If there is a filter with frequency response $zH(z) = z - F(z)$ that is able to eliminate the forecast error signal $\bar{E}(z)$, i.e. $\bar{E}(z) = 0$, the predictive filter output $\hat{X}(z)$ can completely reconstruct the one-step ahead shift of the input signal $\bar{X}(z)$, i.e. $\hat{x}_n = x_{n+1}$. Therefore, $\{x_{n+1}\}$ can be obtained by introducing a predictive filter $F(z) = z(1 - H(z))$ to filter input signal $\{x_n\}$. In other words, if a time series can be perfectly forecasted, the one-step ahead forecast error $\bar{E}(z)$ has to be zero, i.e. $E(z)$ has to be zero. In order for $E(z)$ to be zero, $H(z)X(z)$ has to be zero. Therefore, the bandwidth of $H(z)$ must be complimentary to that of $X(z)$, and both $H(z)$ and $X(z)$ are band-limited, if none of them are zero.

For example, assuming that the input signal $\{x_n\}$ is a band-limited² time series whose cutoff frequency is ω_c in the frequency domain. By introducing a highpass filter $H(z)$ whose cutoff frequency $\omega \geq \omega_c$, we can make the output signal $E(z) = 0$. Therefore, we are able to apply a predictive filter $F(z) = z(1 - H(z))$ to completely forecast $\{x_n\}$ with no errors.

Also, for a bandpass signal that has energy between ω_1 and ω_2 (bandwidth $\omega_2 - \omega_1$), by introducing a bandstop filter $H_p(z)$ with stop band between ω_1 and ω_2 , we could apply a predictive filter $F(z) = z(1 - H_p(z))$ to completely forecast the bandpass signal.

² A signal is said to be baseband band-limited if the power spectrum density goes to zero for all frequencies beyond the threshold called the cutoff frequency. A real world signal can never be truly band-limited as the law of Fourier transformations says that if a signal is finite in time, its spectrum extends to infinite frequency, and if its bandwidth is finite, its duration is infinite in time.

Generally speaking, if a time series $\{x_n\}$ can be perfectly forecasted by a predictive filter $F(z)$, where $F(z) = z(1 - H(z))$, it has to be a band-limited signal whose power spectrum density goes to zero at some frequencies, and $H(z)$ has to be a filter whose bandwidth is complimentary to the bandwidth of the signal.

3.1.2 Sufficient Condition

For a band-limited time series $\{x_n\}$ with frequency response $X(z)$ and baseband bandwidth ω_0 , if there exists a highpass filter $H(z)$ with cutoff frequency $\omega \geq \omega_0$, we could obtain

$$H(z)X(z) = 0 \quad (3.3)$$

If $H(z)$ is a FIR filter, by taking the Inverse Fourier transform of Equation 3.3 we have

$$x_n + a_1x_{n-1} + a_2x_{n-2} + \cdots = 0 \quad (3.4)$$

where a_1, a_2, \dots are coefficients of the highpass filter, thus x_n can be represented by

$$x_n = -a_1x_{n-1} - a_2x_{n-2} - \cdots \quad (3.5)$$

which is equivalent to $z^{-1}\bar{X}(z) = F(z)(z^{-1}X(z))$, i.e., $\bar{X}(z) = F(z)X(z)$, where $F(z) = z(1 - H(z))$. Other generalizations of band-limited signals, for example, signals occupying multiple non-contiguous bands, can also be completely forecasted by filter $F(z)$ as long as $H(z)$ is a filter whose bandwidth is complimentary to the bandwidth of the signal.

Consequently, we could say that if $\{x_n\}$ is a band-limited time series we are able to perfectly forecast it with a predictive filter $F(z)$, where $F(z) = z(1 - H(z))$ and $H(z)$ is a filter whose bandwidth is complimentary to the bandwidth of the signal.

3.2 Negative Group Delay

Group delay is a useful measure of time delay of a signal passing through a filter. Linear phase filter can guarantee that the group delay and phase delay of the filter are constant, and all frequency components have equal time delay thus no phase distortion, which implies the amplitude envelope of the output is exactly the same as that of the input. Minimum phase filter, however, has the minimum group delay among filters that have the same magnitude response.

3.2.1 Group Delay

The group delay of a filter is defined as the negative of the derivative of the phase response with respect to the frequency. Let $F(\omega)$ be the frequency response function of a predictive filter,

$$F(\omega) = A(\omega)e^{j\theta(\omega)} \quad (3.6)$$

where $A(\omega)$ is the magnitude response and $\theta(\omega) = \arg(F(\omega))$ the phase response function. The group delay $\tau(\omega)$ is then defined as the negative of the slope of the phase function $\theta(\omega)$ at a frequency ω [95],

$$\tau(\omega) = -\frac{d\theta(\omega)}{d\omega} \quad (3.7)$$

For a linear phase response, the group delay is constant.

3.2.2 Group Delay of a Band-limited Signal

The group delay of a filter measures the time delay of an input signal through the filter. If the input signal is band-limited between frequency band ω_0 and $\omega_0 + \omega_c$, and the filter has a constant magnitude response (≈ 1) and constant group delay over the bandwidth of the signal, then the output signal will be a replica of the input signal with time delay equaling the group delay $\tau(\omega)$.

It has been proved in [131] that if the transfer function of a filter is given by Equation 3.6 and 3.7, and

1. the magnitude response of the filter is approximately constant $\approx A(\omega_0)$ over the bandwidth of the signal, and
2. the group delay of the filter is approximately constant $\approx \tau(\omega_0)$ over the bandwidth of the signal,

then the output of the filter approximates a replica of the input signal, but time delayed by $\tau(\omega_0)$, scaled in magnitude by $A(\omega_0)$ and with a phase shift $\theta(\omega_0)$.

According to [131], we assume that an input band-limited signal $\{x_n\}$ with the Fourier Transform $X(\omega)$ whose energy concentrates with frequency band ω_0 and $\omega_0 + \omega_c$ (positive band) passes through a filter $F(\omega)$, the output is

$$\hat{X}(\omega) = X(\omega)F(\omega) \tag{3.8}$$

If both the magnitude response and group delay of the filter are approximately constant over the bandwidth of the signal, i.e.,

$$\begin{cases} A(\omega) \approx A(\omega_0) \\ \tau(\omega) \approx \tau(\omega_0) \end{cases} \quad \omega_0 \leq \omega \leq \omega_0 + \omega_c \tag{3.9}$$

where ω_0 is the fundamental frequency, by applying the first order Taylor expansion on the phase $\theta(\omega)$, we have

$$\theta(\omega) \approx \theta(\omega_0) + \frac{d\theta(\omega)}{d\omega} (\omega - \omega_0) = \theta(\omega_0) - \tau(\omega_0) (\omega - \omega_0) \quad (3.10)$$

Therefore, the Fourier transform of the output of the filter can be written as

$$\begin{aligned} \hat{X}(\omega) &= X(\omega)F(\omega) \\ &= X(\omega)A(\omega_0)e^{j(\theta(\omega_0)-\tau(\omega_0)(\omega-\omega_0))} \\ &= \left[X(\omega)e^{-j\tau(\omega_0)\omega} \right] \left[A(\omega_0)e^{j(\theta(\omega_0)+\tau(\omega_0)\omega_0)} \right], \quad \omega_0 \leq \omega \leq \omega_0 + \omega_c \end{aligned} \quad (3.11)$$

The first bracket of the output \hat{x}_n is the time shift of x_n , while the second one is the magnitude scale $A(\omega_0)$ and phase shift. If the input is a baseband band-limited signal with $\omega_0 = 0$, Equation 3.11 can be represented as

$$\hat{X}(\omega) = A(\omega_0)X(\omega)e^{-j\tau(\omega_0)\omega}, \quad \omega_0 \leq \omega \leq \omega_0 + \omega_c \quad (3.12)$$

Therefore if we could make the scale $A(\omega_0) = 1$ and the group delay $\tau(\omega_0) < 0$, the baseband band-limited time series $\{x_n\}$ can be perfectly forecasted.

3.2.3 Predictability of a Filter in terms of Group Delay

In the context of forecasting time series, the group delay $\tau(\omega)$ of a predictive filter measures the time delay of the original time series at frequency ω when it passes through the filter. If the group delay at the frequency $\bar{\omega}$ is negative, i.e., $\tau(\bar{\omega}) < 0$ and the magnitude scale $A(\bar{\omega}) = 1$, the filter is able to completely forecast the time series

at the specific frequency $\bar{\omega}$ from its past observations. However, if the group delay of a filter is negative at the whole frequency domain, the filter is not a causal system. Therefore, ideally, for a band-limited signal, if we could design a filter with constant negative group delay $\tau(\omega)$ and constant magnitude response $A(\omega)$ over the bandwidth of the signal, we could perfectly forecast³ the signal $\tau(\omega)$ steps ahead according to Equation 3.12.

For a naive forecast, group delays are always 1 sample ($\tau(\omega) = 1$) at all frequency. Therefore, we can say that a filter is able to provide some predictability if its group delays at the majority of frequencies where the signal concentrates are less than 1 sample. In this situation, the output signal of the filter is still $1 - \tau(\omega)$ sample ahead of the output of the naive forecast.

3.2.4 Minimum Phase Filter

A *linear, time-invariant* (LTI) system is said to be minimum-phase if all its zeros and poles are inside the unit circle [95]. The distinct property of a minimum phase filter in terms of time series forecasting is that for all causal and stable systems that have the same magnitude response, the minimum phase filter has the minimum group delay and minimum energy delay [95]. Therefore, it is necessary to convert the designed linear predictive filter to a minimum phase one in order to minimise the group delay of the filter while keeping the same magnitude response.

The naive approach to convert a linear phase filter to minimum phase one is to factorise the given transfer function and replace zeros and poles that are outside the unit circle with their reciprocal. However, factorisation is not reliable for very high degree polynomials⁴ [115], especially when some zeros and poles are located on the unit circle.

³ The forecasted signal has to be adjusted accordingly with the magnitude scale.

⁴ A lowpass FIR filter of order 80 is referred as an example of high degree polynomials in [115] and 128 in [114].

For high order linear phase FIR filters, a root moments based method is proposed in [115] to convert them to minimum phase ones.

3.3 Perfect Forecast of Band-limited Signal using Negative Group Delay Filters

A band-limited signal can be perfectly forecasted by a predictive filter that has constant negative group delay and constant magnitude response in its passband that matches the bandwidth of the signal.

3.3.1 Design of Negative Group Delay Filters

There are few methods available to design negative group delay filters in literature. For example, a frequency selective FIR filter is proposed in [127]. The filter could achieve $\tau(0) = -1$ at the zero frequency by manipulating the zero locations to convert a linear phase filter into a minimum phase one. However, this method can only be used to forecast signals with very narrow bandwidth as the negative group delay is only constant at the specified frequency. Moreover, a bandpass IIR amplifier with negative group delay in its stopband was designed in [90]. However, the magnitude response at the frequency region with negative group delay are not flat so that it cannot be used to filter signal without shape distortion.

By applying the root moments method proposed in [115], we are able to design high order filters (e.g. $n = 64$) with approximately constant negative group delay and magnitude response that can be used to approximately perfectly forecast band-limited signals.

For a linear phase FIR filter with the transfer function

$$H(z) = z^n + h_1 z^{n-1} + h_2 z^{n-2} + \cdots + h_n = \prod_{i=1}^n (z - r_i) \quad (3.13)$$

where r_i is a root of the polynomial $H(z)$, and can be divided in 3 categories based on its location with respect to the unit circle:

- $r_i = r_j^{\text{in}}$ if the root r_i is inside the unit circle
- $r_i = r_j^{\text{out}}$ if the root r_i is outside the unit circle
- $r_i = r_j^{\text{on}}$ if the root r_i is on the unit circle

Therefore, the transfer function can be written as

$$H(z) = \left[\prod_j (z - r_i^{\text{in}}) \right] \left[\prod_j (z - r_i^{\text{out}}) \right] \left[\prod_j (z - r_i^{\text{on}}) \right] \quad (3.14)$$

or

$$H(z) = H_{\min}(z) H_{\max}(z) H_o(z) \quad (3.15)$$

where $H_{\min}(z)$ is the minimum phase part of $H(z)$, $H_{\max}(z)$ the maximum phase part of $H(z)$, and $H_o(z)$ contains all the roots that lie on the unit circle.

To design a filter with approximately constant negative group delay at lower band, we design a lowpass filter $H(z)$ first, and extract its minimum phase part $H_{\min}(z)$ using the root moments method.

For example, we design a lowpass filter using the *Remez* method of Matlab with the normalised cutoff frequency at 0.9. The magnitude response and pole/zero plots are shown in Figure 3.1

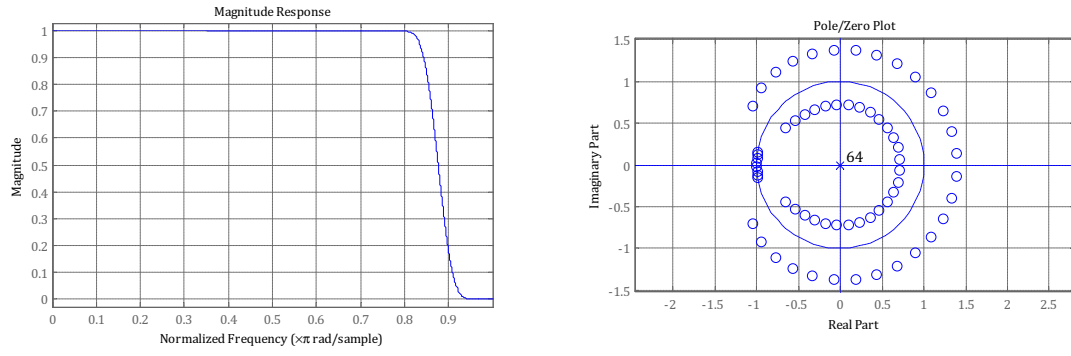


Figure 3.1: Magnitude response and Pole/Zero plot of a lowpass filter $H(z)$

The magnitude response and pole/zero plots of the minimum phase part $H_{\min}(z)$ are shown in Figure 3.2

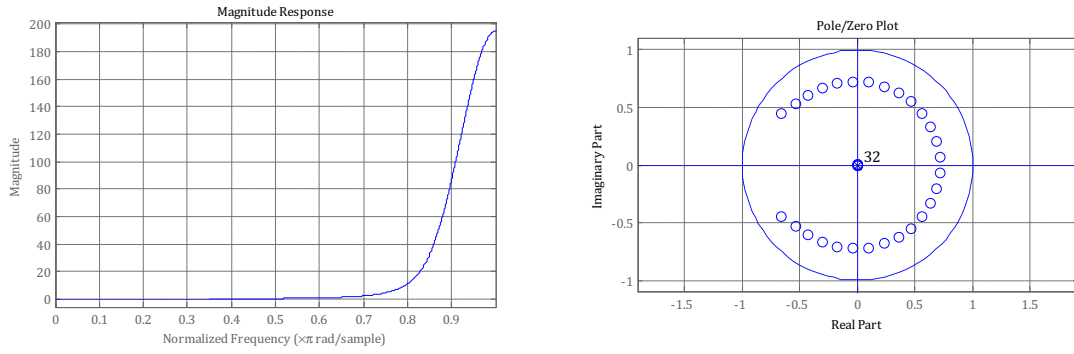


Figure 3.2: Magnitude response and Pole/Zero plot of the minimum phase part $H_{\min}(z)$ of $H(z)$

The group delay of the minimum phase filter $H_{\min}(z)$ is then shown in Figure 3.3

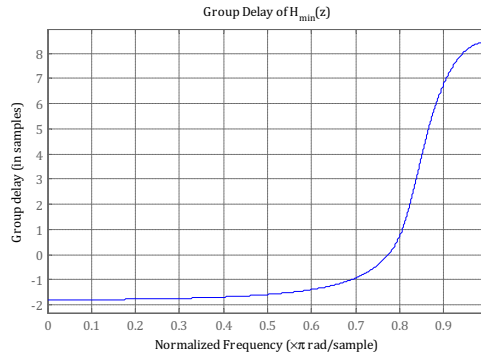


Figure 3.3: Group delay of the minimum phase filter $H_{\min}(z)$

The negative group delay predictive filter $F(z)$ can be derived as $F(z) = z(1 - H_{\min}(z))$. Figure 3.4 shows the magnitude response and group delay of the derived predictive filter $F(z)$.

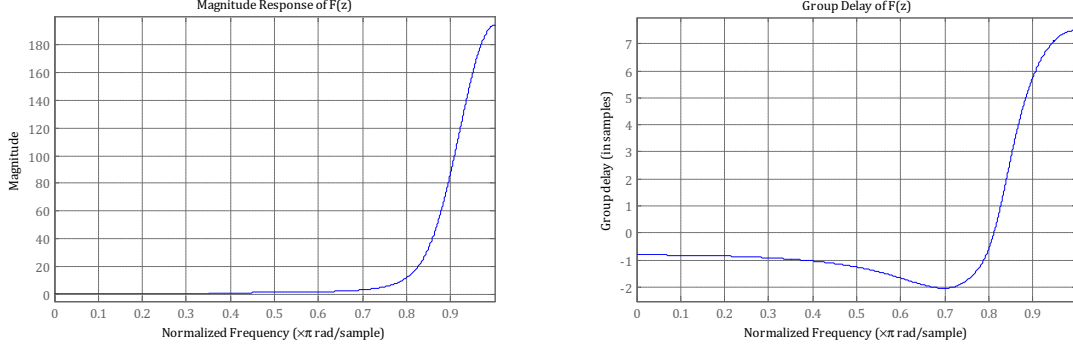


Figure 3.4: Magnitude response and group delay of the predictive filter $F(z)$

We can tell that the predictive filter $F(z)$ has approximately constant negative group delay in the low frequency band (e.g., the normalised frequency band 0 to 0.4 in Figure 3.4) and its magnitude response is smooth (approximately constant) and nearly zero. Essentially, the predictive filter $F(z)$ is a highpass filter with negative group delay in its stopband. However, the magnitude response of the stopband is not exactly zero but approximately a small constant. Therefore, we could use $F(z)$ to forecast baseband band-limited signals with bandwidth less than 0.4 in this example.

The methodology to design a negative group delay predictive filter $F(z)$ with approximately constant negative group delay and magnitude response in the low frequency band using the root moments method is summarised in Table 3.1:

Step	Method
1	to design a high order lowpass linear phase FIR filter (e.g. $n = 64$) $H(z)$ using the <i>Remez</i> method (Figure 3.1)
2	to decompose $H(z)$ to $H_{\min}(z)$, $H_{\max}(z)$ and $H_o(z)$ using the root moments method
3	to extract the minimum phase filter $H_{\min}(z)$ from $H(z)$ (Figure 3.2)
4	to derive the predictive filter as $F(z) = z(1 - H_{\min}(z))$ (Figure 3.4)

Table 3.1: Steps to design a negative group delay predictive filter $F(z)$

The negative group delay predictive filter can be implemented in Matlab using func-

tions listed in Section C.1.

3.3.2 Forecast Band-limited Signal using Negative Group Delay Filters

To demonstrate that a baseband band-limited signal can be perfectly forecasted with a negative group delay filter as discussed in Section 3.2.2, we generate a approximate baseband band-limited signal $\{x_n\}$ by passing a random signal through a lowpass filter. The input signal $\{x_n\}$ and its frequency response are shown in Figure 3.5

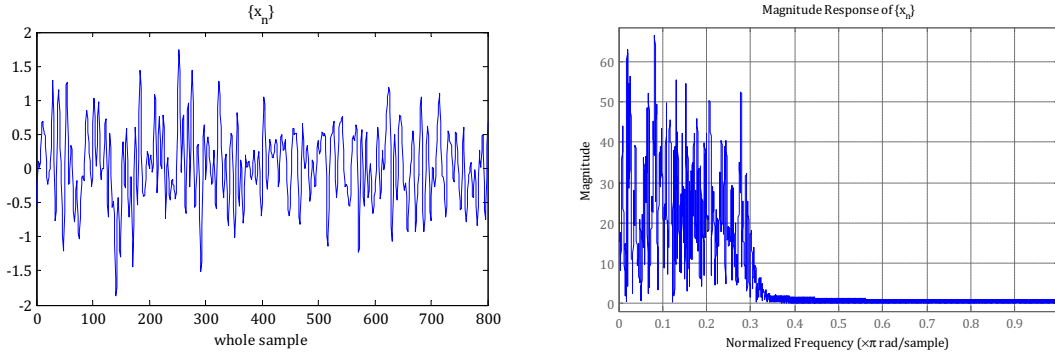


Figure 3.5: Input low frequency band-limited signal $\{x_n\}$ and its magnitude response

By passing $\{x_n\}$ through the negative group delay predictive filter $F(z)$, we obtain the output signal $\{\hat{x}_n\}$ as shown in Figure 3.6

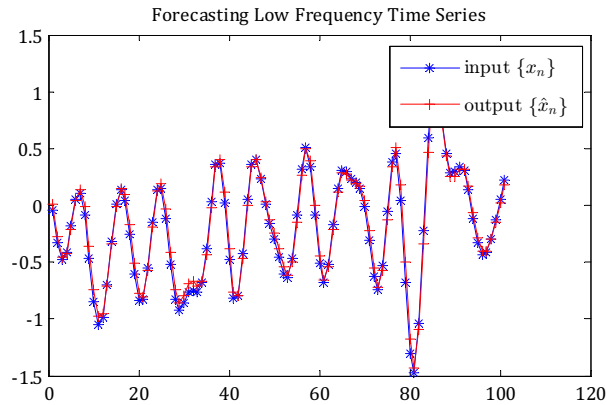


Figure 3.6: Output signal $\{\hat{x}_n\}$ and input low frequency band-limited signal $\{x_n\}$

We shall notice that there are slight waveform distortion of the output signal. The reason is that we cannot generate a truly band-limited signal $\{x_n\}$ in real world as a band-limited signal extends infinitely over time.

The method shown in Section 3.3.1 can also be applied to design a lowpass filter that has approximately constant negative group delay in the high frequency band. Also, we could apply the quadrature mirror filter [95] method to create a lowpass filter from the designed highpass filter $H_{\min}(z)$ and convert it to a minimum phase filter $\bar{H}_{\min}(z)$. The magnitude response and pole/zero plots of the lowpass filter $\bar{H}_{\min}(z)$ are illustrated in Figure 3.7.

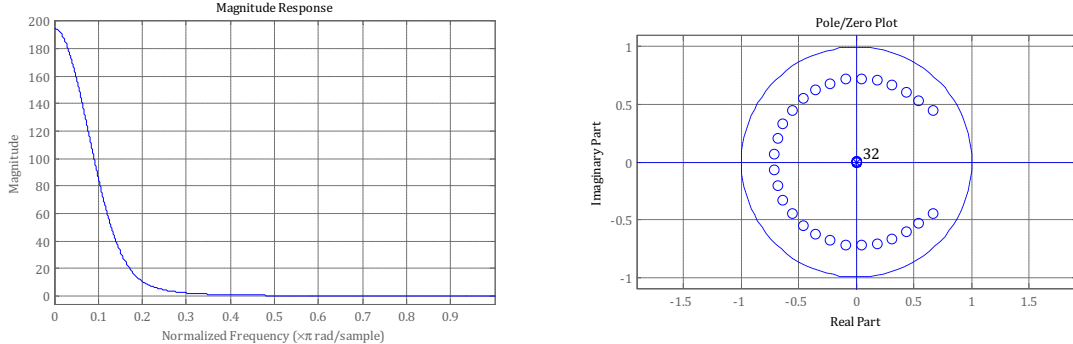


Figure 3.7: Magnitude response and Pole/Zero plot of the lowpass filter $\bar{H}_{\min}(z)$

The negative group delay predictive filter $\bar{F}(z)$ can be calculated as $\bar{F}(z) = z(1 - \bar{H}_{\min}(z))$.

Figure 3.8 shows the magnitude response and group delay of the derived predictive filter $\bar{F}(z)$.

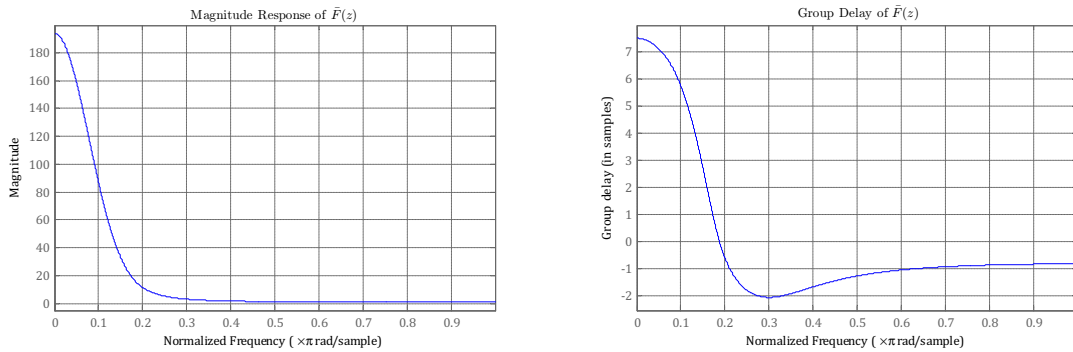


Figure 3.8: Magnitude response and group delay of the predictive filter $\bar{F}(z)$

By filtering a band-limited signal whose energy concentrates in the high frequency band as shown in Figure 3.9, we obtain the output signal that forecasts the input signal with slight waveform distortion (Figure 3.10).

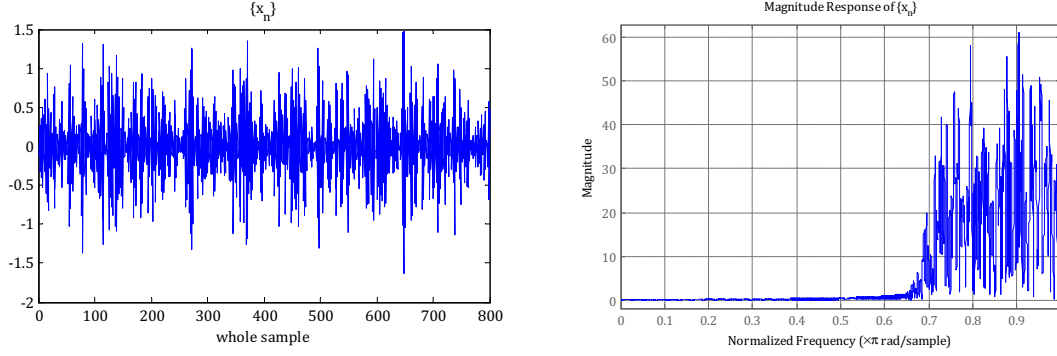


Figure 3.9: Input high frequency band-limited signal $\{x_n\}$ and its magnitude response

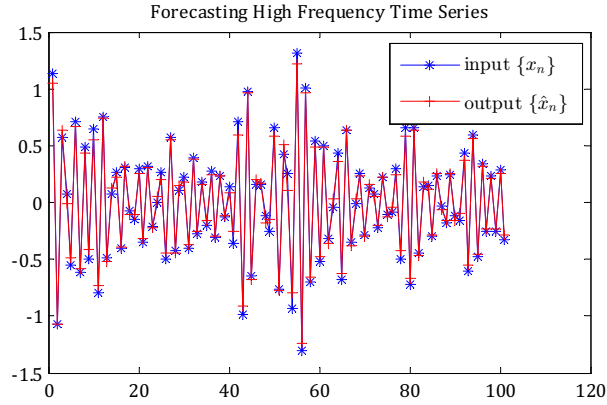


Figure 3.10: Output signal $\{\hat{x}_n\}$ and input high frequency band-limited signal $\{x_n\}$

3.4 Conclusion

This chapter presented the sufficient and necessary conditions that a time series can be predicted with zero errors by linear filters. It is concluded that if and only if a time series is band-limited it can be predicted with zero errors by linear filters, given the assumption that there are no structural breaks or regime switching in the time series.

In addition, we concluded that a band-limited time series can be predicted with zero errors by a predictive filter that has a constant magnitude response and a constant

group delay over the bandwidth of the signal. Generally speaking, if the group delay at the frequency $\bar{\omega}$ is negative, i.e., $\tau(\bar{\omega}) < 0$, and the magnitude response $A(\bar{\omega}) = 1$, the filter is able to completely forecast the time series at the specific frequency $\bar{\omega}$ from its past observations. Therefore, if we could design a filter with constant negative group delay $\tau(\omega)$ and constant magnitude response $A(\omega)$ over the bandwidth of the signal, we could perfectly forecast a signal $\tau(\omega)$ samples ahead.

A minimum phase filter has the minimum group delay for all causal and stable systems that have the same magnitude response. Therefore, it is necessary to convert a designed linear predictive filter to a minimum phase one in order to minimise the group delay of the filter while not changing the same magnitude response.

We also proposed a methodology to design highpass and lowpass negative group delay filters that have approximately constant negative group delay and magnitude response in the passband as summarised in Table 3.1. The designed filters can be used to forecast lowband and highband band-limited time series one-step ahead with slight waveform distortion.

Chapter 4

Linear Predictive Filters

As shown in Section 3.1, if a time series is band-limited it can be predicted with zero errors by linear filters, given the assumption that there are no structural breaks or regime switching in the time series. However, the real world time series can never be truly band-limited, thus cannot be forecasted without errors. Therefore, a generic predictive filter method is proposed for forecasting real world time series with white noise errors in this chapter.

4.1 Predictive Filter Theory

Let $\{x_n\}$ be a non-band-limited signal that cannot be forecasted without errors, a predictive filter is defined as a filter that takes signal $\{x_n\}$ as an input while outputs its one-step ahead forecast $\{\hat{x}_n\}$.

In z -domain, the one-step ahead forecast $\hat{X}(z)$, is the product of the input signal $X(z)$, and the predictive filter $F(z)$

$$\hat{X}(z) = F(z)X(z) \tag{4.1}$$

The one-step ahead forecast error signal $\bar{E}(z) = zE(z)$ is defined as the difference between the desired one-step ahead shifted signal $\bar{X}(z)$ (i.e., $\{x_{n+1}\}$) and the filter output $\hat{X}(z)$ as

$$\begin{aligned}\bar{E}(z) = zE(z) &= \bar{X}(z) - \hat{X}(z) \\ &= (z - F(z))X(z) \\ &= zH(z)X(z)\end{aligned}\tag{4.2}$$

As shown in Section 3.1, if there is a filter with transfer function $H(z) = 1 - z^{-1}F(z)$ that is able to eliminate the forecast error signal $E(z)$, i.e. $E(z) = 0$, the predictive filter output $\hat{X}(z)$ can completely reconstruct the one-step ahead shifted signal $\bar{X}(z)$, i.e. $\hat{x}_n = x_{n+1}$, with no errors. Therefore, $\{x_{n+1}\}$ can be obtained by introducing the predictive filter $F(z) = z(1 - H(z))$ to filter input signal $\{x_n\}$.

However, no such filter exists in reality as $\{x_n\}$ can never be a band-limited signal thus cannot be forecasted without errors. The solution is then to make the one-step ahead forecast $\{\hat{x}_n\}$ as closely approaching the desired one-step ahead shifted signal $\{x_{n+1}\}$ as possible either by minimising the forecast errors $E(z)$ (Section 4.2.1) or by maximising the spectrum flatness of forecast errors $E(z)$ (Section 4.2.2).

4.1.1 Linear Predictive Filter

To show how the predictive filter theory works, it is applied to forecast linear time series, for example, stochastic time series following AR or ARMA processes (Section 2.3.1), and deterministic time series in this section.

4.1.1.1 AR process

If a time series $\{x_n\}$ follows an AR process

$$x_n = \sum_{i=1}^p \phi_i x_{n-i} + e_n \quad (4.3)$$

where e_n is the error term that follows a white noise process.

It can be represented in z -domain as

$$\left(1 - \sum_{i=1}^p \phi_i z^{-i}\right) X(z) = E(z) \quad (4.4)$$

Let $H(z) = 1 - \sum_{i=1}^p \phi_i z^{-i}$, which is the transfer function of a FIR filter. $\{x_n\}$ can be forecasted by $\hat{x}_n = \sum_{i=1}^p \phi_i x_{n-i}$ without errors by designing a filter with transfer function $H(z)$ that is able to make $E(z) = 0$, thus $e_n = 0$. Therefore, by introducing a FIR predictive filter

$$F(z) = z \left(1 - \left(1 - \sum_{i=1}^p \phi_i z^{-i}\right)\right) = \sum_{i=1}^p \phi_i z^{-i+1} \quad (4.5)$$

the AR process $\{x_n\}$ can be forecasted one-step ahead without errors.

If the error term $e_n \neq 0$, $F(z)$ should be able to forecast one-step ahead of $\{x_n\}$ with white noise errors $\{e_n\}$.

4.1.1.2 ARMA process

If $\{x_n\}$ follows an ARMA process,

$$x_n = \sum_{i=1}^p \phi_i x_{n-i} + \sum_{j=1}^q \theta_j e_{n-j} + e_n \quad (4.6)$$

that can be represented in z -domain as

$$\left(1 - \sum_{i=1}^p \phi_i z^{-i}\right) X(z) = \left(1 + \sum_{j=1}^q \theta_j z^{-j}\right) E(z) \quad (4.7)$$

Let $H(z) = \frac{1 - \sum_{i=1}^P \phi_i z^{-i}}{1 + \sum_{j=1}^Q \theta_j z^{-j}}$, which is the transfer function of an IIR filter. In the same fashion as the FIR filter, $\{x_n\}$ can be forecasted by $F(z)$

$$F(z) = z \left(1 - \frac{1 - \sum_{i=1}^P \phi_i z^{-i}}{1 + \sum_{j=1}^Q \theta_j z^{-j}}\right) = \frac{\sum_{j=1}^Q \theta_j z^{-j+1} - \sum_{i=1}^P \phi_i z^{-i+1}}{1 + \sum_{j=1}^Q \theta_j z^{-j}} \quad (4.8)$$

with white noise errors $\{e_n\}$.

4.1.1.3 Deterministic Signal

Deterministic signals are those signals whose values are completely fixed according to specified functions for any given time. Deterministic signals carry no new information as all the new values are predetermined thus can be forecasted completely with no errors.

Let's take the following simple deterministic signal for example

$$x_n = \sin(\omega_0 n) \quad (4.9)$$

$\{x_n\}$ is represented in z -domain as

$$X(z) = \frac{\sin \omega_0 z^{-1}}{1 - 2 \cos \omega_0 z^{-1} + z^{-2}} \quad (4.10)$$

Therefore the linear predictive filter $F(z)$ can be obtained by applying Equation 4.5

$$F(z) = 2 \cos \omega_0 - z^{-1} \quad (4.11)$$

That is, $\{x_n\}$ can be forecasted with no errors by $\hat{X}(z) = F(z)X(z)$ as

$$x_{n+1} = 2 \cos \omega_0 x_n - x_{n-1} \quad (4.12)$$

4.2 Coefficients Estimation Methods

For a real world time series $\{x_n\}$ that is not band-limited, one widely used method to estimate coefficients of the predictive filters is to minimise forecast errors $E(\omega)$, and it can be implemented by the well-known Wiener filters proposed in [132]. Alternatively, we propose a coefficients estimation method that is to make the spectrum of the forecast error $E(\omega)$ as flat as that of a white noise process as possible. This can be implemented by maximising the flatness of the *power spectral density* (PSD) of forecast errors $E(\omega)$.

4.2.1 Least Square Errors

The Wiener filter is designed to minimise the mean square forecast error between desired signal $\{x_{n+1}\}$ and output signal $\{\hat{x}_n\}$ which can be stated concisely as follows:

$$\min \sum_{n=0}^{N-1} |e_n|^2 \quad (4.13)$$

It is equivalent to minimise the total energy of forecast errors

$$\min \sum_{z=0}^{N-1} |E(\omega)|^2 \quad (4.14)$$

where $E(\omega) = \sum_{n=0}^{N-1} e_n \cdot e^{-2j\omega n}$, ω is the normalized frequency, according to Parseval's theorem $\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{z=0}^{N-1} |X(\omega)|^2$, where $X(\omega)$ is the DFT of x_n , both of length N . That is, the total energy contained in a waveform $\{x_n\}$ summed across all of time n is equal to the total energy of the waveform's Fourier Transform $X(\omega)$ summed across all of its frequency components ω .

Therefore, the coefficient α_i of the Wiener filter can be obtained by

$$\alpha_i = \arg \min \sum_{z=0}^{N-1} |E(\omega)|^2 \quad (4.15)$$

It is worth noticing that fitting a FIR or IIR filter by $\min \sum_{z=0}^{N-1} |E(\omega)|^2$ is equivalent to fit an AR or ARMA model estimated by the least square method.

4.2.2 Maximum Spectrum Flatness

As discussed in Section 2.2, the power spectrum of a signal can be used to exploit predictive patterns in a signal process. The more correlated or predictable a signal is, the more concentrated its power spectrum is and, conversely, the more random or unpredictable a signal, the wider its power spectrum spreads. Therefore, by designing a filter with desired frequency response $H(\omega)$ that is able to maximise the spectrum flatness of output error signal, i.e. $E(\omega)$, we can obtain an output signal with maximum “whiteness”. Assuming the spectrum flatness can be quantified as F_{PSD} , the coefficients α_i of the filter can be obtained by

$$\alpha_i = \arg \max F_{\text{PSD}} \quad (4.16)$$

Let φ_k denote the sample PSD

$$\varphi_k = \sum_{n=-N}^N \gamma_n e^{-\frac{2\pi i}{2N+1} kn} \quad (4.17)$$

where γ_n is the sample ACF

$$\gamma_n = \frac{1}{N - |n|} \sum_{m=0}^{N-1} x_m x_{m+n}, \quad n = 0, \pm 1, \pm 2, \dots, \pm (N - 1) \quad (4.18)$$

As shown in Section 2.2, the spectrum flatness F_{PSD} can be measured by flatness test statistics $D(\varphi)$ and $\psi(\varphi)$. In addition, the variance of the spectrum $\text{Var}(\varphi)$ measures the dispersion of data points of spectrum around their mean value, thus can also be used to measure the spectrum flatness F_{PSD} . The following subsections show that the metrics $D(\varphi)$, $\psi(\varphi)$ and $\text{Var}(\varphi)$ reach their minimum values if the PSD is constant, in other words, the spectrum is flat.

It has been proved in Section 6.4.3 of [65] that maximising the error spectrum flatness of a minimum phase predictive filter is equivalent to minimising the prediction error power of the predictive filter, and to minimising the least square prediction errors. Therefore, the coefficients of a minimum phase predictive filter estimated from Equation 4.15 should be the same as those estimated from Equation 4.16.

4.2.2.1 $D(\varphi)$

In the discrete form, spectrum flatness test $D(\varphi)$ can be represented as

$$D(\varphi) = \log \left(\frac{1}{N} \sum_{i=0}^{N-1} \varphi_i \right) - \frac{1}{N} \sum_{i=0}^{N-1} \log \varphi_i \quad (4.19)$$

Taking the partial differential of $D(\varphi)$ with respect to the PSD at k th frequency φ_k

$$\frac{\partial D(\varphi)}{\partial \varphi_k} = \frac{1}{\sum_{i=0}^{N-1} \varphi_i} - \frac{1}{N\varphi_k} \quad (4.20)$$

To minimise Equation 4.19 by setting Equation 4.20 equal zero

$$\varphi_k = \frac{1}{N} \sum_{i=0}^{N-1} \varphi_i \quad (4.21)$$

Consequently, if the value of every φ_k equals the expected value of the sample power spectral density, which is a constant, $D(\varphi)$ reaches its minimum value.

4.2.2.2 $\psi(\varphi)$

For the spectrum flatness test $\psi(\varphi)$, by taking the partial differential of $\psi(\varphi)$ with respect to any φ_k

$$\psi(\varphi) = \frac{1}{N} \sum_{i=0}^{N-1} \log^2 \varphi_i - \left(\frac{1}{N} \sum_{i=0}^{N-1} \log \varphi_i \right)^2 \quad (4.22)$$

We get

$$\frac{\partial \psi(\varphi)}{\partial \varphi_k} = \frac{2 \log \varphi_k}{N\varphi_k} - \frac{2}{N^2\varphi_k} \sum_{i=1}^{N-1} \log \varphi_i \quad (4.23)$$

To minimise Equation 4.22 by making Equation 4.23 equal zero

$$\log \varphi_k = \frac{1}{N} \sum_{i=0}^{N-1} \log \varphi_i \quad (4.24)$$

Therefore, if the value of every $\log \varphi_k$ equals the expected log value of the sample power spectral density, which is a constant, $\psi(\varphi)$ reaches its minimum value.

4.2.2.3 $\text{Var}(\varphi)$

In addition, the variance of φ_k can be represented as

$$\text{Var}(\varphi) = \frac{1}{N} \sum_{i=0}^{N-1} \varphi_i^2 - \left(\frac{1}{N} \sum_{i=0}^{N-1} \varphi_i \right)^2 \quad (4.25)$$

By taking the partial differential of $\text{Var}(\varphi)$ with respect to any φ_k

$$\frac{\partial \text{Var}(\varphi)}{\partial \varphi_k} = \frac{2\varphi_k}{N} - \frac{2}{N^2} \sum_{i=0}^{N-1} \varphi_i \quad (4.26)$$

To minimise Equation 4.25 by making Equation 4.26 equals zero

$$\varphi_k = \frac{1}{N} \sum_{i=0}^{N-1} \varphi_i \quad (4.27)$$

Accordingly, if the value of every φ_k equals the expected value of the sample power spectral density, which is a constant, $\text{Var}(\varphi)$ reaches its minimum value.

4.2.2.4 Minimax

Another way is to minimise the maximum magnitude of the error spectrum considering the property of almost everywhere constancy of the power spectral density of a white noise

$$\min \left\{ \max |E(\omega)|^2 \right\} \quad (4.28)$$

4.3 Spectrum Flatness Evaluation Methods

In order to apply the aforementioned metrics to evaluate the spectrum flatness, it is necessary to estimate the spectrum of time series regarding the possible structural

breaks in a time series. One conventional way to estimate the spectrum of a time series is the periodogram method discussed in Section 4.3.1 which can be used to evaluate the spectrum flatness of time series with no structural breaks with the aforementioned metrics. For a time series that have structural breaks, two methods are proposed in Section 4.3.2 and 4.3.3 to evaluate the spectrum and spectrum flatness. By maximising the error spectrum flatness using these methods, coefficients of the specified model could be derived (Equation 4.16). The three methods are applied to a simulated time series with a structural break in the in-sample data set in Section 4.3.4.

4.3.1 Error Spectrum Flatness with Periodogram

A direct approach to estimate the PSD of an input signal $\{x_n\}$ is the sample periodogram, which utilises the whole data set of length N .

$$P(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x_n w_n e^{-j\omega n} \right|^2 \quad (4.29)$$

where w_n is a window function. Alternatively, the sample periodogram is given by the Fourier transform of the ACF γ_n as in Equation 4.18

$$P(\omega) = \sum_{n=-(N-1)}^{N-1} \gamma_n e^{-j\omega n} \quad (4.30)$$

For an equally weighted time series, w_n is effectively a rectangular window. The true PSD of a time series is obtained when its length N goes to infinity,

$$S(\omega) = E \left[\lim_{N \rightarrow \infty} P(\omega) \right] \quad (4.31)$$

In reality, it is impossible to get the true PSD but a sample one as the real world time series is effectively truncated by a rectangular window.

The PSD of errors $\{e_n\}$ that is the output of the input signal $\{x_n\}$ passing through a filter $H(\omega)$ is

$$E^P(\omega) = |H(\omega)|^2 P(\omega) \quad (4.32)$$

where $P(\omega)$ is the sample periodogram of the input signal $\{x_n\}$, and $H(\omega)$ is the filter's DFT.

As shown in Section 4.2.2, the coefficient α_i of the filter h_n can be obtained by maximising the spectrum flatness of $E^P(\omega)$ as shown in Equation 4.16.

The coefficient α_i of the predictive filter can also be obtained by minimising the total energy $\sum_{z=0}^{N-1} |E(\omega)|^2$ as shown in Equation 4.15.

4.3.2 Standard Deviation Weighted Error Spectrum Flatness

The raw periodogram is not a good spectral estimate because it is not consistent with high statistical variability which does not decrease as the number of samples increases [118].

One method to solve the variance problem is known as the Bartlett's method [16]. The idea is to divide the set of N samples into K sets of L non-overlapping samples, compute the power spectral density of each sample using the periodogram method and average them at the same frequency to get a smooth PSD estimation.

The Welch method [130] or *weighted overlapped segment averaging* (WOSA) method uses a modified version of Bartlett's method in which the segments of the series contributing to each periodogram are allowed to overlap and be windowed prior to computation of the periodogram.

Let $\{x_n\}$, $n = 1, 2, \dots, N$ be a sample of a stationary time series. Let $P(\omega)$ be the sample power spectrum of $\{x_n\}$. We take overlapping segments of length L with the starting points of these segments D unit apart. Let $\{x_n(1)\}$, $n = 1, 2, \dots, L$ be the first segment and K the number of segments. Then

$$\begin{aligned} x_n(1) &= x_n & n &= 0, 1, \dots, L-1 \\ x_n(2) &= x_{n+D} & n &= 0, 1, \dots, L-1 \\ &\vdots \\ x_n(K) &= x_{n+(K-1)D} & n &= 0, 1, \dots, L-1 \end{aligned} \quad (4.33)$$

For each segment of length L we calculate a modified periodogram. That is, we select a data window w_n , $n = 0, 1, \dots, L-1$, and form the sequences $x_n(1)w_n, \dots, x_n(K)w_n$. The periodogram of the m th segment is given by

$$P_m(\omega) = \frac{1}{L} \left| \sum_{n=0}^{L-1} x_n(m)w_n e^{-j\frac{2\pi nk}{L}} \right|^2 \quad (4.34)$$

The Welch estimate of the PSD is given by

$$S^W(\omega) = \frac{1}{K} \sum_{i=0}^{K-1} P_i(\omega) \quad (4.35)$$

The main drawback of the Welch method is that the use of several segments of reduced length makes it prone to severe leakage errors. According to [7], a typical solution to reduce the transients magnitude is to use a smooth data window, e.g. the Hanning window, rather than the rectangular window. Another intuitive solution is to increase overlap between adjacent segments so that transient affects get averaged out to some extent. Conventionally, a Hanning window with 50% overlap is used for estimating

the PSD using the Welch method. However, the half-sine window with 67% overlap is proven to be the optimal real symmetric window that globally minimises leakage errors in the case of stationary random noises [7].

The PSD of errors can be calculated in the same fashion as Equation 4.32

$$E^W(\omega) = |H(\omega)|^2 S^W(\omega) \quad (4.36)$$

The weighted PSD of errors is calculated by considering the standard deviation of $P_m(\omega)$ at each frequency ω , i.e., $\sigma(\omega)$.

$$\sigma(\omega) = \sqrt{\frac{1}{K} \sum_{i=0}^{K-1} (P_i(\omega) - S^W(\omega))^2} \quad (4.37)$$

The normalised standard deviation is then derived as the standard deviation divided by its mean

$$\sigma^N(\omega) = \frac{\sigma(\omega)}{S^W(\omega)} = \sqrt{\frac{1}{K} \sum_{i=0}^{K-1} \left(\frac{P_i(\omega)}{S^W(\omega)} - 1 \right)^2} \quad (4.38)$$

The purpose to use normalised standard deviation is to allow the appropriate comparison of a variable with a large mean and correspondingly large standard deviation with variables with smaller means and correspondingly smaller standard deviations.

The normalised *standard deviation weighted error spectrum* at each frequency ω is represented as

$$E'(\omega) = \frac{E^W(\omega)}{\sigma^N(\omega)} \quad (4.39)$$

The coefficient α_i of the filter $H(\omega)$ can be obtained by maximising the spectrum flatness of $E'(\omega)$ as shown in Equation 4.16. The intuition is to give more weights to the error spectrum of certain frequencies where the estimated spectrum of the input signal is less volatile. The reason is that the less volatile spectrum at certain frequencies implies more confidence in the PSD estimation at these frequencies.

The coefficient α_i of the predictive filter can also be obtained by minimising the total energy $\sum_{z=0}^{N-1} |E'(\omega)|^2$ as shown in Equation 4.15.

4.3.3 Exponentially Weighted Standard Deviation Weighted Error Spectrum Flatness

In the context of time series forecasting, it is desirable to use as many historical observations as possible to estimate predictive filters in order to maximise the accuracy of the estimation if there are no structural breaks or switching regime in historical patterns. Otherwise, it is reasonable to assign more weight to the most recent observations than the old ones for a time series that has structural breaks.

By assigning exponential weights to estimate the periodogram of each segment $P_m(\omega)$

$$P_m^E(\omega) = w(m)P_m(\omega) \quad (4.40)$$

where exponential weight $w_m = (1 - \lambda) \lambda^{K-m-1}$ and λ represents the degree of weighting decrease, which is a constant decay factor between 0 and 1. The smaller the λ , the faster the decay of the old observation. This approach is often referred as *Exponentially Weighted Moving Average* (EWMA).

The estimated PSD for the whole sample is then

$$S^E(\omega) = \sum_{i=0}^{K-1} w(i) P_i(\omega) \quad (4.41)$$

Contrary to an equally weighted model, the PSD estimation reacts faster to changes in the spectrum as recent data carry more weight than data from the distant past.

The PSD of the errors can be calculated as

$$E^E(\omega) = |H(\omega)|^2 S^E(\omega) \quad (4.42)$$

The exponentially weighted standard deviation of $P_m(\omega)$ at each frequency ω is

$$\sigma^E(\omega) = \sqrt{(1 - \lambda) \sum_{i=0}^{K-1} \lambda^{K-i-1} (P_i(\omega) - S^E(\omega))^2} \quad (4.43)$$

where $S^E(\omega)$ is the equally weighted PSD at frequency ω as shown in Equation 4.35.

Also, the standard deviation $\sigma^E(\omega)$ changes faster with the most recent data than with the older ones.

The normalised exponentially weighted standard deviation

$$\sigma^{EN}(\omega) = \frac{\sigma^E(\omega)}{S^E(\omega)} \sqrt{(1 - \lambda) \sum_{i=0}^{K-1} \lambda^{K-i-1} \left(\frac{P_i(\omega)}{S^E(\omega)} - 1 \right)^2} \quad (4.44)$$

The normalised *exponentially weighted standard deviation weighted error spectrum* at each frequency ω is represented as

$$E''(\omega) = \frac{E^E(\omega)}{\sigma^{EN}(\omega)} \quad (4.45)$$

The coefficient α_i of the filter $H(\omega)$ can be obtained by maximising the spectrum flatness of $E''(\omega)$ as shown in Equation 4.16.

The coefficient α_i of the predictive filter can also be obtained by minimising the total energy $\sum_{z=0}^{N-1} |E''(\omega)|^2$ as shown in Equation 4.15.

In summary, the periodogram method of spectrum flatness is applicable to design a predictive filter for a time series that has no structural breaks in the in-sample data set, while the *Standard Deviation Weighted Error Spectrum Flatness* (SDWESF) and the *Exponentially Weighted Standard Deviation Weighted Error Spectrum Flatness* (EWSDWESF) methods are able to deal with time series with structural breaks in the in-sample data set. Furthermore, The EWSDWESF method adjusts the estimated spectrum to structural breaks more quickly and effectively than the SDWESF method as it assigns more weights to the most recent observations than the old ones.

In addition, the SDWESF and EWSDWESF methods give more weights to the error spectrum at certain frequencies where the estimated spectrum of the input signal is less volatile. Thus we are more confident about the PSD estimation at these frequencies than those at the more volatile ones.

The comparison of the three proposed methods are shown in Table 4.1, and the simulation results of applying the three methods to an artificial time series with structural breaks in the in-sample data set is discussed in the next section.

	Periodogram	SDWESF	EWSDWESF
training set for spectrum estimation	whole	overlapped segments	overlapped segments
adjusted to structural breaks	No	in-sample	in-sample, fast
adjusted to volatility of spectrum	No	Yes	Yes

Table 4.1: Comparison of three error spectrum flatness evaluation methods

4.3.4 Simulation

In order to demonstrate the effectiveness of the EWSDWESF method to forecast time series with structural breaks, a time series (1600 observations, illustrated in Figure 4.3) is simulated which combines a baseband band-limited time series (800 observations, illustrated in Figure 4.1) and a band-limited time series whose energy concentrates in higher band (800 observations, Figure 4.2) over time.

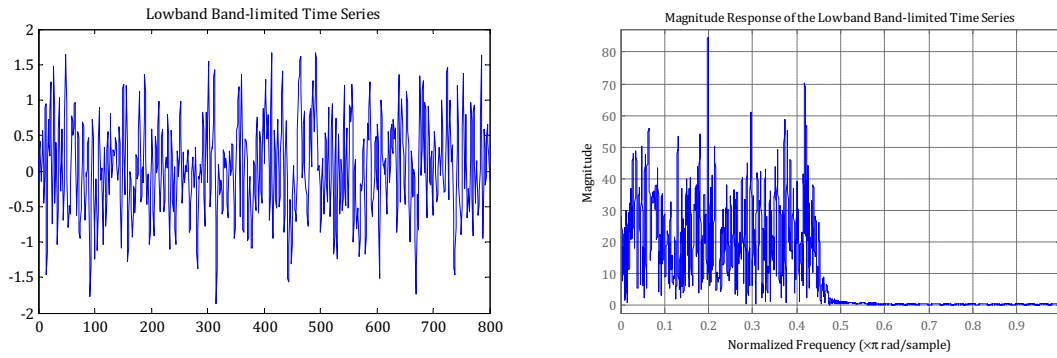


Figure 4.1: Lowband band-limited time series and its magnitude response

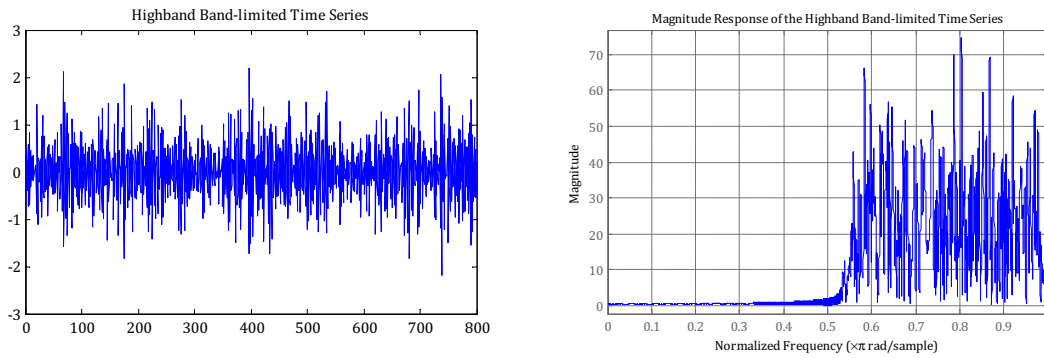


Figure 4.2: Highband band-limited time series and its magnitude response

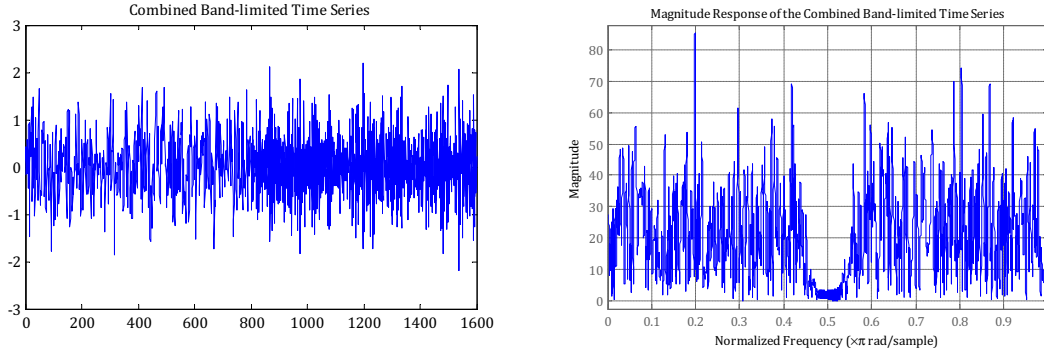


Figure 4.3: Combined band-limited time series and its magnitude response

The first 1000 observations are used for the in-sample model estimation, the next 300 ones for the model selection and the last 300 ones as the out-of-sample data set for the goodness-of-fit test. 5 methods are used to forecast the out-of-sample data set:

1. the ARMA model¹ estimated by minimising prediction errors;
2. the FIR filter estimated by minimising the variance of the error periodogram;
3. the FIR filter estimated by maximising the SDWESF;
4. the FIR filter estimated by maximising the EWSDWESF;
5. the naive forecast;

The implementation of the SDWESF method needs an input of segment size L , and the EWSDWESF method needs inputs of both the segment size L and the decay factor λ . One possible way to decide the segment size and decay factor is to use the segment size and decay factor selected from the in-sample testing set (Figure 5.8) which has the minimum MSE or maximum success ratio of direction change forecast.

¹ The *armasel* function of the Matlab toolbox ARMASA (<http://www.dsc.tudelft.nl/Research/Software/index.html>) by P.M.T. Broersen is used as a benchmark method of ARMA model in this thesis

The out-of-sample goodness-of-fit tests of these methods are shown in Table 4.2. Note that the ARMA (in-sample) means using the first 1300 observations as in-sample data set for the ARMA estimation.

	ARMA (in-sample)	ARMA	Periodogram	SDWESF	EWSDWESF	Naive
MSE	0.4679	0.2546	0.2435	0.2528	0.0103	1.5799
MASE	0.5356	0.4014	0.3922	0.4001	0.07991	0.9975
SR	61.33%	74%	74%	74%	95.67%	26%
DA	3.9186	8.346	8.346	8.346	15.8471	1

Table 4.2: Comparison of the goodness-of-fit tests of out-of-sample forecast

It is easy to tell that the EWSDWESF method greatly outperforms ARMA (in-sample) (97.6% better for MSE, 57.6% better for Success Ratio²) and ARMA (95.7% better for MSE, 31.8% better for Success Ratio).

The estimated PSD for the whole sample $S^W(\omega)$ of SDWESF (window size $L = 64$ and filter order 1) and $S^E(\omega)$ of EWSDWESF (window size $L = 64$, decay factor $\lambda = 0.85$ and filter order 5) are shown in Figure 4.4. The $S^E(\omega)$ of EWSDWESF eliminates effects of the lowband time series that present in the first half of the whole time series. This is the reason that it significantly outperforms the other methods when there is a structural break in the in-sample data set.

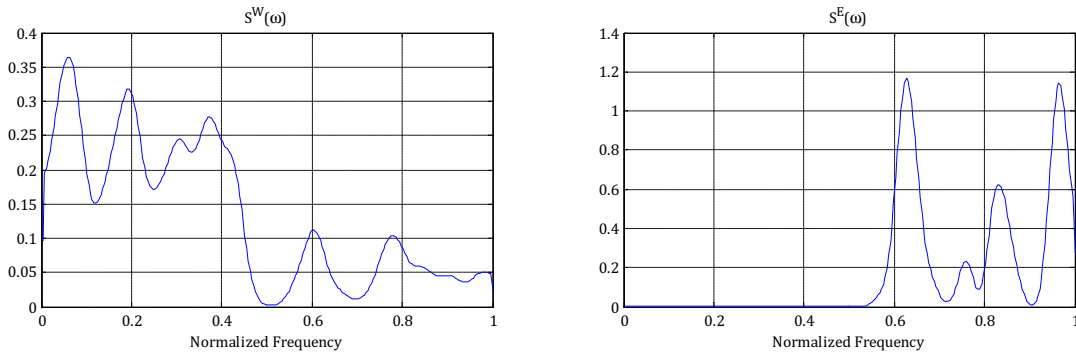


Figure 4.4: $S^W(\omega)$ of SDWESF and $S^E(\omega)$ of EWSDWESF

² The DA test is designed to assess the performance of sign predictions. As the limiting distribution of this test is $N(0, 1)$, its one-sided critical values at the 1%, 5%, 10% levels are 2.33, 1.645 and 1.282, respectively.

The out-of-sample forecast of ARMA (in-sample) and EWSDWESF method are shown in Figure 4.5 (only show the last 100 observations).

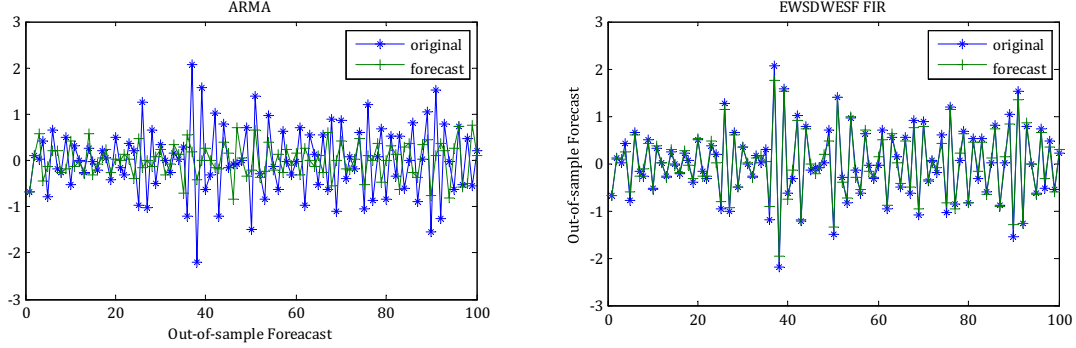


Figure 4.5: Out-of-sample forecast by ARMA and LPF

In summary, the EWSDWESF method greatly improves the ARMA model estimated by least square errors when dealing with time series that have structural breaks in the in-sample data set.

4.4 Adaptive and Hybrid Filter Method

As shown in Section 4.3.4, the EWSDWESF method is powerful to handle time series that have structural breaks within in-sample observations, and it fits a single global model for all future evolutions of a time series. In other words, there are structural breaks in its historical observations but not future ones. However, this is not true in the context of financial time series as they are dominated by uncertainty in the future which could cause large price movements. Therefore, an adaptive filter and a hybrid filter method are proposed to address time series with structural breaks in the out-of-sample data set in this section.

4.4.1 Adaptive Filter

One major difference between adaptive filtering and ARMA models of the Box-Jenkins methodology is that parameters of the latter are fixed, while those of adaptive filtering are not. This enables adaptive filter to deal with non-stationary data and to adapt to changes in the data pattern (by updating the model parameters as new data become available) much better than fixed parameter models. The adaptive filters were applied to forecast time series in literature such as [86], [32] and [70].

Different from conventional *Recursive Least Square* (RLS) or *Least Mean Square* (LMS) methods that recursively update coefficients based on estimation errors of the last iteration [55], the adaptive filter method is implemented by updating coefficients of the predictive filters in Section 4.3.2 and 4.3.3 estimated by maximising the error spectrum flatness within a moving rectangular window. Taking the exponentially weighted standard deviation weighted error spectrum flatness method for example, the error PSD of the m th moving window is represented as

$$E_m^E(\omega) = |H_m(\omega)|^2 S_m^E(\omega) \quad (4.46)$$

where $S_m^E(\omega)$ is the estimated PSD for the whole time series sample within the m th moving window, $|H_m(\omega)|^2$ is the PSD of the predictive filter whose coefficients are estimated by maximising the spectrum flatness of the normalised exponentially weighted standard deviation weighted errors at the same moving window, and is calculated as the same fashion as Equation 4.45

$$E_m''(\omega) = \frac{E_m^E(\omega)}{\sigma_m^{EN}(\omega)} \quad (4.47)$$

After each solution based on the m th window is calculated, a new sample is taken,

the oldest sample is removed and a new solution for the $(m + 1)$ th window is then calculated.

4.4.2 Hybrid Filter

The proposed adaptive filter recalculates its coefficients whenever new sample are taken into the moving window. It is suitable for time series that have a large amount of structural breaks, and time series that have small magnitude of structural breaks. However, for a time series which has few structural breaks, it is not necessary to recalculate coefficients for each moving window. Consequently, a hybrid filter is proposed which switches from a constant coefficients model as described in Section 4.3 to an adaptive filter in the presence of structural breaks, and keeps using the newly estimated filter as a constant coefficients model until the next structural break.

The hybrid filter is implemented by applying the coefficient set α_{m-1} estimated from the $(m - 1)$ th moving window to observations within the m th window and derive the out-of-sample forecast error e_m . If the forecast error e_m is less than or equal to the forecast error e_{m-1} of the previous window $m - 1$, i.e.,

$$|e_m|^2 \leq |e_{m-1}|^2 \quad (4.48)$$

the coefficient set α_{m-1} will continue being used to forecast the time series using the observations of the m th window, i.e., $\alpha_m = \alpha_{m-1}$. Otherwise, a new coefficient set $\bar{\alpha}_m$ is estimated using observations of the m th window and the out-of-sample forecast error \bar{e}_m is calculated and compared to e_m . If $e_m \geq \bar{e}_m$, the new coefficient set $\bar{\alpha}_m$ is used to replace α_{m-1} for forecasting the m th window, or $\alpha_m = \bar{\alpha}_m$.

The hybrid filter avoids recalculating the coefficients when there are no structural

breaks or only small magnitude structural breaks. The magnitude of structural breaks³ that a hybrid filter watches or ignores can be controlled by ε_m which relaxes the Equation 4.48 as

$$|e_m|^2 - |e_{m-1}|^2 \leq \varepsilon_m \quad (4.49)$$

The greater the ε_m , the larger magnitude of structural breaks the hybrid filter takes into consideration, the less recalculation the hybrid filter performs to forecast a time series. Specially, the hybrid filter turns to a fixed coefficient filter when $\varepsilon_m \rightarrow \infty$, and an adaptive filter when $\varepsilon_m \rightarrow -\infty$.

Consequently, the hybrid filter is more flexible than the adaptive filter in terms of controlling the magnitude of structural breaks. It is more suitable to forecast a time series that has few but large magnitude structural breaks.

Table 4.3 shows the comparison of the adaptive filter, hybrid filter and those fixed coefficient filters (Section 4.3) with respect to the number of coefficients calculations for the out-of-sample forecast and structural breaks that they are suitable to deal with.

	Fixed Coef Filter	Adaptive Filter	Hybrid Filter
coefficients calculation	once	available of new data	depends on forecast errors
adjusted to structural breaks	in-sample	all samples	all samples
magnitude of structural breaks	large	small	large
frequency of structural breaks	few	high	few

Table 4.3: Comparison of fixed coefficient model, adaptive filter and hybrid filter

4.4.3 Comparison to Time Varying Coefficients Model

The adaptive filter and hybrid filter are effectively time varying coefficients models (Section 2.3.2) whose coefficients are series of coefficients of successive LPFs. The essential difference between the proposed adaptive/hybrid filter and conventional time

³ The magnitude of structural breaks is defined in terms of the magnitude of forecast errors in presence of a structure break. Even if the magnitude of a time series has not changed, it is possible that the magnitude of structural breaks is large. For example, the time series shown in Figure 4.3 is said to have a large magnitude of structural break.

varying coefficients models is that coefficients of the filter are estimated by minimising forecast errors or maximising the ESF of each moving window of the original time series, while the latter treats each coefficient over time as an individual time series and estimates the next coefficients using the past coefficients. Consequently, the adaptive/hybrid filter does not have the inherited limitations of the time varying coefficients model.

4.4.4 Simulation

The same data set as in Section 4.3.4 is used to demonstrate the feasibility of the adaptive filter and hybrid filter to forecast a time series that has structural breaks in the in-sample data set but not the out-of-sample one. The goodness-of-fit tests of the out-of-sample forecast using adaptive filter and hybrid filter respectively are shown in Table 4.4. The goodness-of-fit test results of both adaptive filter and hybrid filter are slightly improved over that of the EWSDWESF method in Table 4.2 and much better than those of the other methods.

	EWSDWESF	Adaptive	Hybrid
MSE	0.0103	0.0101	0.0097
MASE	0.07991	0.0786	0.0775
SR	95.67%	97.00%	97.00%
DA	15.8471	16.3116	16.3083

Table 4.4: Goodness-of-fit tests of the out-of-sample forecast of the adaptive Filter and the hybrid Filter

The coefficient sets over time of the adaptive filter and hybrid filter are illustrated in Figure 4.6. We can observe that the coefficient set of the hybrid filter is constant over time while that of the adaptive filter changes smoothly over time. This matches the expectation as there are no structural breaks in the out-of-sample data set. In this case the hybrid filter will be selected over the adaptive filter as it needs less calculation of coefficients while provides equal predictability as the adaptive filter.

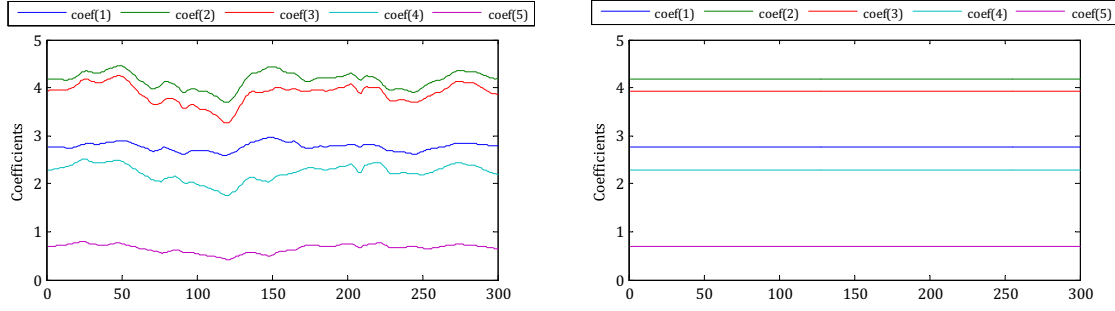


Figure 4.6: Coefficients of the adaptive filter (left) and the hybrid filter (right)

Furthermore, the adaptive filter and hybrid filter methods are evaluated when a time series has a structural break in the out-of-sample data set. The same data set as the one in Section 4.3.4 is used but the out-of-sample data set is now 200 observations from the 701st to the 900th which includes a structural break at the 801st observation.

The out-of-sample forecast of the adaptive filter with a moving rectangular window of size 512 are tested with overlapping segments of length $L = 64$ and $L = 32$ as shown in Equation 4.34. Figure 4.7 shows that the adaptive filter with $L = 32$ adapts to the structural break faster than that with $L = 64$. The reason is that the exponentially weighted PSD for the whole sample (Equation 4.41) adapts to changes of structure faster with a smaller segment L . However, there is a tradeoff between the speed of adaption and the resolution of spectrum, as the smaller the segment, the smaller the spectrum resolution, which would lead to a worse PSD estimation. As a result, it is necessary to choose a proper segment size for either the EWSDWESF method or the adaptive/hybrid filter that utilises the EWSDWESF method. One possible way to decide the size of moving window is to use the size of the moving window selected from the in-sample testing set (Figure 5.9) which has the minimum MSE or maximum success ratio of direction changes forecast.

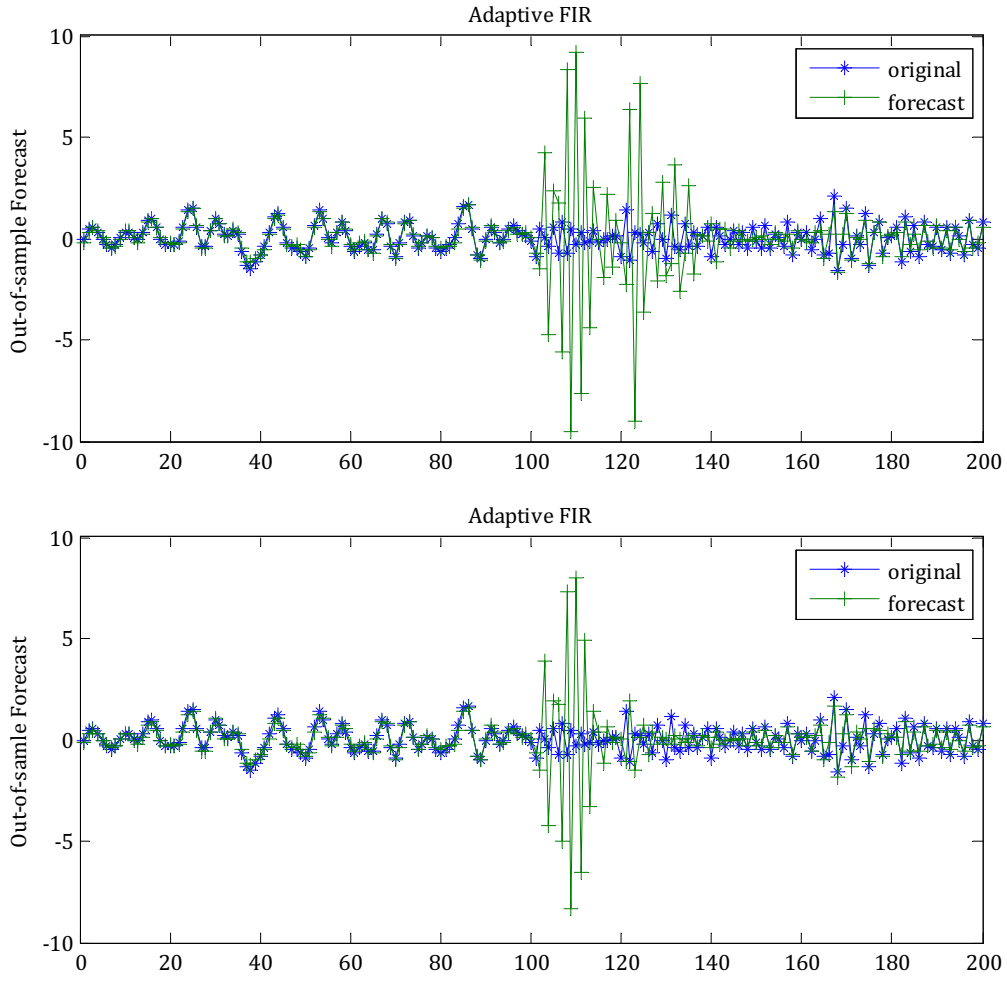


Figure 4.7: Out-of-sample forecast of the adaptive filter with overlapping segments of length $L = 64$ (top) and $L = 32$ (bottom)

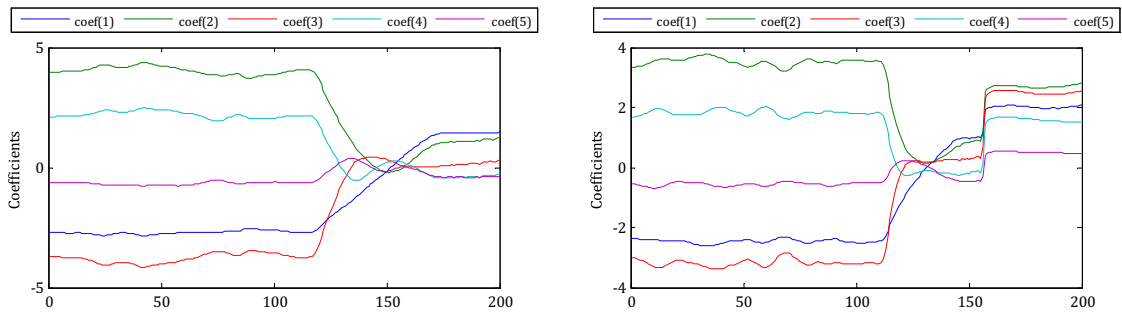


Figure 4.8: Coefficients of the adaptive filter with overlapping segments of length $L = 64$ (left) and $L = 32$ (right)

The Figure 4.8 illustrates the coefficient sets over time of the adaptive filters with

overlapping segments of length $L = 64$ and $L = 32$. As expected, the coefficients of the adaptive filter with $L = 32$ adapt to a structural break faster than those with $L = 64$. Moreover, coefficients tend to be stable after the structural break when the newly estimated adaptive filter starts well matching the new pattern.

The out-of-sample forecast and coefficient sets over time of a hybrid filter are shown in Figure 4.9 and 4.10, respectively. Analogically, both the out-of-sample forecast and coefficients of the hybrid filter with $L = 32$ adapt to the structural break faster than those with $L = 64$.

Comparing Figure 4.10 to Figure 4.8, it is clear that the coefficients of the hybrid filter are constant when there are no structural breaks while change gradually to another stable state after a structural break.

The Table 4.5 compares goodness-of-fit tests of the out-of-sample forecast using adaptive filter and hybrid filter with overlapping segments of length $L = 64$ and $L = 32$ (shown in the parenthesis), respectively. Both the adaptive filter and hybrid filter with $L = 32$ outperform those with $L = 64$ due to the faster adaption to the changes in structure.

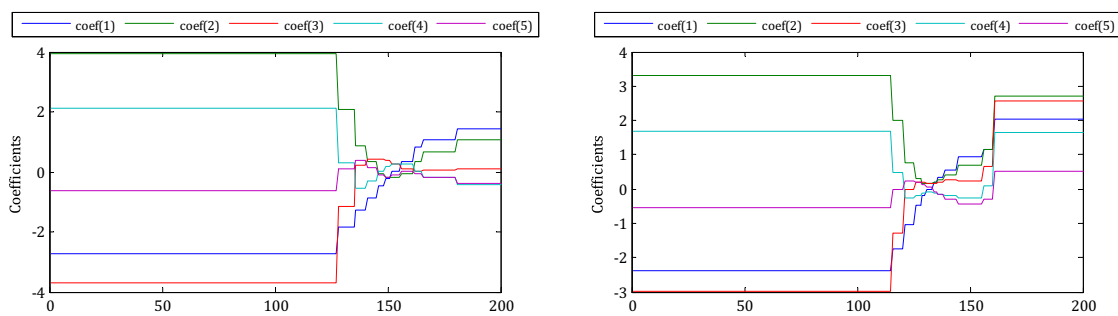


Figure 4.10: Coefficients of the hybrid filter with overlapping segments of length $L = 64$ (left) and $L = 32$ (right)

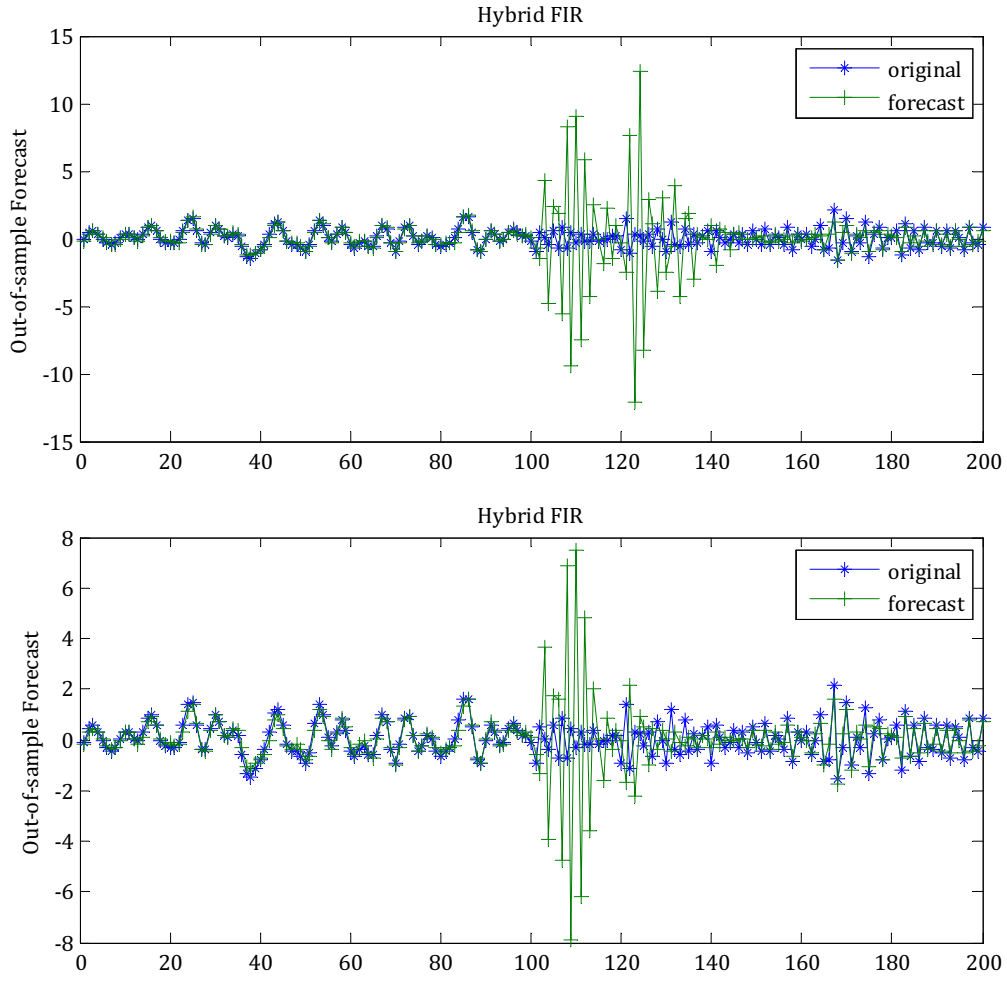


Figure 4.9: Out-of-sample forecast of the hybrid filter with overlapping segments of length $L = 64$ (top) and $L = 32$ (bottom)

	Adaptive ($L=64$)	Adaptive (32)	Hybrid (64)	Hybrid (32)
MSE	3.9989	2.0277	5.3383	1.9322
MASE	1.1293	0.73237	1.2992	0.75736
SR	77.5%	80%	77.5%	78.5%
DA	7.7968	8.5129	7.7948	8.0788

Table 4.5: Comparison of goodness-of-fit tests of out-of-sample forecast

To conclude, both the adaptive filter and hybrid filter provide a similar performance as the EWSDWESF method when forecasting a time series that has no structural breaks in the out-of-sample data set, while they could provide supreme performance over the fixed coefficient filters in the situation where a time series has structural breaks in the out-of-sample data set, that is, structural breaks in the future evolvement of the time

series.

The fixed coefficient filter, adaptive filter and hybrid filter can be implemented in Matlab using functions listed in Section C.2.

4.5 Bounds of Forecast Errors

The RWH suggests that movements of asset prices follow a random walk process, which implies that given only past prices, the current price is the best predictor of future prices. It is necessary to study under what conditions, a sophisticated forecast model could outperform the naive model⁴. Whether a forecast model could outperform the naive model in terms of the upper bounds of forecast errors is examined in this section.

4.5.1 One-Step Ahead Forecast Error Bound

For a predictive filter with transfer function $F(\omega)$ that gives the one-step ahead in-sample forecast \hat{x}_n^{IS} , the forecast error e_n^{IS} is bounded (Section A.2 for derivation) as

$$|e_n^{\text{IS}}| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + 1) \quad (4.50)$$

where F_k is the frequency response of the predictive filter, X_k the frequency response of the input time series $\{x_n\}$.

The MAE and MSE of the one-step ahead in-sample forecast are bounded as

$$\text{MAE}^{\text{IS}} \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + 1) \quad (4.51)$$

⁴ We use the naive forecast, naive model and random walk model interchangeable, which all refer to the model that use the present observation as the forecast of the future one.

$$\text{MSE}^{\text{IS}} \leq \frac{1}{N^2} \sum_{k=0}^{N-1} |X_k|^2 (|F_k|^2 + 1) \quad (4.52)$$

Let a real number α_n be the ratio of two successive observations, i.e. $\alpha_n = \frac{x_{n+1}}{x_n}$, the actual x_{n+1} can be represented using x_n as

$$x_{n+1} = \alpha_n x_n \quad (4.53)$$

For one-step ahead out-of-sample forecast \hat{x}_n^{OS} , the forecast error e_n^{OS} is also bounded (see Section A.3 for derivation) as

$$|e_n^{\text{OS}}| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + |\alpha_n|) \quad (4.54)$$

And the MAE and MSE of the out-of-sample forecast are bounded as

$$\text{MAE}^{\text{OS}} \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + A) \quad (4.55)$$

$$\text{MSE}^{\text{OS}} \leq \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} |X_k|^2 (|F_j|^2 + B) \quad (4.56)$$

where $A = \frac{1}{M} \sum_{n=0}^{M-1} |\alpha_n|$ and $B = \frac{1}{M} \sum_{n=0}^{M-1} \alpha_n^2$, and M is the length of the out-of-sample data set.

By comparing Equation 4.50 to Equation 4.54, it is derived that if $|\alpha_n| > 1$ the upper bounds the out-of-sample forecast errors is higher than that of in-sample. The greater the $|\alpha_n|$, the higher the upper bounds of out-of-sample forecast errors than that of in-sample forecast errors. For the same reason, the greater the A or B , the higher the upper bounds of MAE or MSE of the out-of-sample forecast errors.

4.5.2 Naive Forecast Error Bound

Particularly, if the naive model is used for forecasting, which implies the present value is used as the forecast for the future ones, i.e., $F_k = 1$, the in-sample forecast errors:

$$\left| e_n^{\text{RWIS}} \right| \leq \frac{2}{N} \sum_{k=0}^{N-1} |X_k| \quad (4.57)$$

while the out-of-sample forecast errors:

$$\left| e_n^{\text{RWOS}} \right| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (1 + |\alpha_n|) \quad (4.58)$$

For a signal $\{x_n\}$ that has energy over the whole frequency bands, that is, $|X_k| > 0$ when $k \in [0, N]$, it is derived by comparing Equation 4.54 to 4.58 that when $\frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| < \frac{1}{N} \sum_{k=0}^{N-1} |X_k|$, the upper bounds of out-of-sample forecast errors by the predictive filter $F(\omega)$ are lower than those of the naive forecast. Specially for a band-limited signal, the upper bounds of out-of-sample forecast errors by the predictive filter $F(\omega)$ are lower than those of the naive forecast when $|F_k| < 1$ for all k .

4.5.3 Theorem

From the analysis in Section 4.5.2, the following theorem can be derived:

Theorem 4.1. *The upper bounds of one-step ahead forecast errors using a predictive filter $F(\omega)$ are lower than those using the naive forecast if and only if the magnitude response $\frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| < \frac{1}{N} \sum_{k=0}^{N-1} |X_k|$, where $X_k > 0$.*

Proof. See Section A.4 □

According to the theorem, if $|F_k| < 1$ for all k of a predictive filter $F(\omega)$, the upper bounds of out-of-sample forecast errors of the filter will be lower than those of the naive forecast. Therefore, the theorem provides a method to estimate the forecast performance of a predictive filter over the naive forecast in terms of the upper bounds of forecast errors, without the need to calculate forecast errors of either the filter or the naive forecast, when $|F_k| < 1$ for all k .

Whether a sophisticated predictive filter outperforms the naive forecast can be decided by comparing its magnitude response $|F_k|$ to 1. If $|F_k| < 1$ for all k , the filter provides better forecast performance than the naive forecast in terms of the upper bound of forecast errors. However, if $\frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| > \frac{1}{N} \sum_{k=0}^{N-1} |X_k|$, it implies that the upper bounds of the forecast errors of the filter are higher than those of the naive forecast. It is worth noticing that $|F_k|$ does not provide any insight on how tight these bounds are, nor whether a model is better than the naive forecast for a specific signal in terms of specific goodness-of-fit metrics. It is possible that forecast errors are far less than the upper bounds when applying a predictive filter on a specific signal. Consequently, the goodness-of-fit metrics discussed in Section 2.4.1 should be used to evaluate the forecast performance of a predictive filter in terms of magnitude of errors and success ratio of direction changes, etc.

The real importance of the magnitude response $|F_k|$ lies in the fact that it provides a unique insight into the worst case scenario of the forecast performance of applying a predictive filter to forecast a time series. By calculating the upper bounds of in-sample and out-of-sample forecast errors using Equation 4.50 and 4.54, we could access to the worst case forecast errors, as well as the worst case MAE and MSE. One of the potential applications of the worst case forecast errors of a predictive filter is to decide the amount of capital we should assign to a trading strategy which adopts the filter to forecast a financial time series.

4.6 Conclusion

This chapter presented a predictive filter theory and coefficient estimation methods. The linear predictive filter can be applied to forecast real world time series which are not band-limited, with forecast errors following a white noise process. The coefficients of a predictive filter can be derived by maximising the spectrum flatness of forecast errors. Three methods are proposed to estimate the spectrum flatness of forecast errors. The periodogram method is applicable to time series with no structural breaks, while the SDWESF method and the EWSDWESF method are able to deal with time series with structural breaks in the in-sample data set, and the latter adjusts to structural breaks more quickly than the former.

In addition, the adaptive filter and hybrid filter were introduced for forecasting time series that have structural breaks not only in the in-sample data set but also out-of-sample one. The hybrid filter is more suitable to forecast time series that have few but large magnitude structural breaks, while the adaptive filter is more suitable for time series that have large amount of small structural breaks.

Furthermore, a theorem that estimates the forecast performance of a predictive filter over the naive forecast in terms of the upper bounds of forecast errors was presented. The magnitude response of the predictive filter provides a unique insight into the worst case scenario of the forecast performance of applying a predictive filter to forecast a time series.

Chapter 5

Forecasting Financial Time Series

5.1 Target Price

For a real world financial time series which is not band-limited, the methods mentioned in Chapter 3 cannot be applied to forecast it without errors. However, forecasting financial time series with errors can still be useful to guide decision making as long as a greater success ratio of forecasting direction changes or a smaller magnitude errors of price movements could be achieved by a predictive filter comparing to the naive forecast. Therefore, the methods proposed in Chapter 4 could be applied to forecast financial time series with errors that follow a white noise process.

In the context of financial time series, the spectrum of the first order difference of close prices of a financial time series is close to that of a white noise process in terms of spectrum flatness (Figure 5.1 and 5.2). That makes it difficult to forecast using linear models. It is consistent with the EMH which states that current price already contains all the information, so it is impossible to use past information to foretell future price movements.

The first order difference of close prices of the S&P500 (2011-01-01 \rightarrow 2011-12-30) and

its spectrum is shown in Figure 5.1.

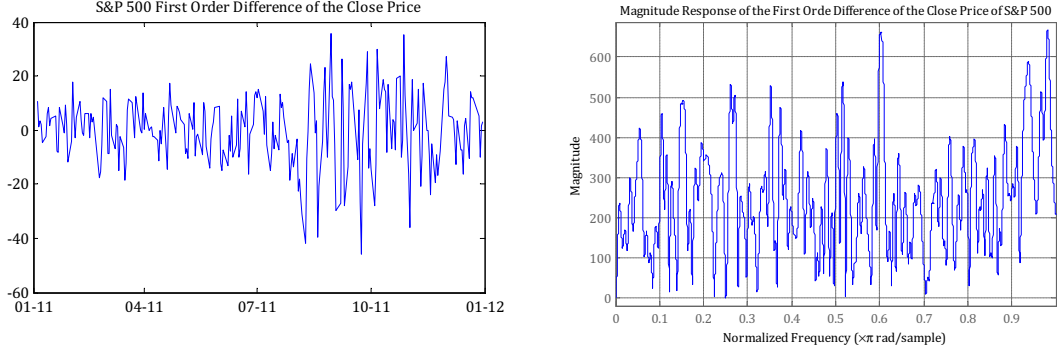


Figure 5.1: The first order difference of close prices of the S&P500 and its magnitude response

Figure 5.2 shows the spectrum of the in-sample data set of the first order difference of open, high, low and close prices of the S&P500 (2000 observations), respectively.

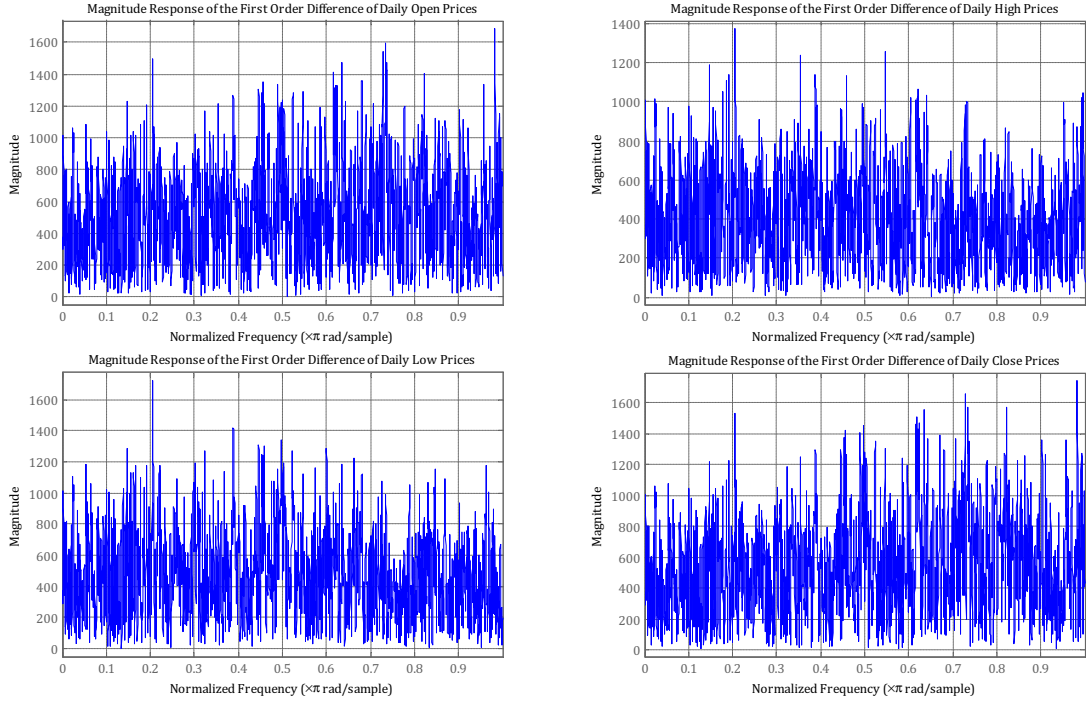


Figure 5.2: Magnitude response of the in-sample first order difference of open/high/low/close prices of the S&P500

Many previous studies such as those mentioned in Section 2.1 and 2.3 usually focused on close prices only. However, price data of the open, high, low, and close are all

available for any time frame in practice. The most recent studies such as [34] and [27], showed that daily high and low prices of some equities are predictable by using the High-Low range as an explanatory variable. Therefore, it is necessary to consider daily high, low, open and close prices when we study the behavior of a financial time series.

One of the contributions of this research is that the linear combination of daily open (or that of a certain time frame), high, low and close (OHLC) prices, which we call *target price*¹, is found to be able to provide more predictability than daily close prices only. Furthermore, daily open prices are proposed to be used in a trading strategy taking advantage of forecasting target prices.

5.2 Optimise the Price Combination

5.2.1 Open, High, Low and Close Price

The target prices of a financial time series² could be obtained by optimising linear weights assigned to daily open, high, low and close prices. In order for target prices to be useful for trading purposes, a constraint which forces target prices to be within the daily High-Low range should be considered to estimate their weights. Therefore, it is reasonable to forecast a target price time series which is a proxy of the daily price time series of an asset.

Let $y_n = \alpha x_n^O + \beta x_n^H + \gamma x_n^L + \delta x_n^C$ be a target price that is a linear combination of daily open price x_n^O , high price x_n^H , low price x_n^L and close price x_n^C . $\{y_n\}$ is then the target

¹ The daily target price of an asset which is a linear combination of daily open, high, low, and close prices is considered in the thesis. However, it is applicable to any time frame as long as the data of open, high, low and close prices are available.

² The target price of a financial time series is used interchangeable to the target price of the underlying asset of the time series in the thesis.

price time series that is the linear combination of daily open, high, low and close price time series, i.e., $\{x_n^O\}$, $\{x_n^H\}$, $\{x_n^L\}$, $\{x_n^C\}$. Since the individual price series $\{x_n^O\}$, $\{x_n^H\}$, $\{x_n^L\}$ and $\{x_n^C\}$ are close to random walk processes (Figure 5.2) that cannot be predicted better than using the naive forecast, we would wish to find a target price series $\{y_n\}$ which could give more predictability than each $\{x_n\}$ alone.

This can be done by using weights, α , β , γ and δ in this case, as unknown variables and defining an objective function which is the spectrum flatness of the artificial time series of weighted open, high, low and close price series. The idea is that the less flat the spectrum of the artificial time series is, the more predictable the time series is. Ideally we are able to perfectly predict a target price time series without errors if we could make it band-limited by assigning proper weights to the open, high, low and close price series.

The objective function is constructed in terms of spectrum flatness as:

$$\max_{\alpha, \beta, \gamma, \delta} \left(\sum \frac{|Y_1(k)|^2}{N+1} - \sum \frac{|Y_2(k)|^2}{L-N} \right) \quad (5.1)$$

where

$$Y(k) = \begin{cases} Y_1(k) & k = 0, 1, \dots, N \\ Y_2(k) & k = N+1, N+2, \dots, L \end{cases} \quad (5.2)$$

is the Fourier transform of y_n . $Y_1(k)$ is the low frequency component between 0 and N , which is a threshold that separates low frequency components from high frequency ones, while $Y_2(k)$ is the high frequency component between $N+1$ and L , which is the length of frequency range of the one-side spectrum.

The target price $\{y_n\}$ has to be within the daily High-Low range in order to act as a proxy of daily price to guide the trading decision making in practice. The following

constraints are added to the optimisation program to make the target price within the daily High-Low range:

$$\begin{cases} y_n \leq x_n^H \\ y_n \geq x_n^L \end{cases} \quad (5.3)$$

Nevertheless, the objective function (Equation 5.1) is not convex so that there are more than one y_n that satisfies the constraints and the optimisation program could end up with a local minimum³. In practice, we could accept the local minimum as long as the target price time series is more predictable than a random walk process and we could make profits by forecasting the target price time series.

A special case of target prices is that when $\alpha = 0$ and $\beta = \gamma = \delta = 1/3$, the target price

$$y_n = \frac{x_n^H + x_n^L + x_n^C}{3} \quad (5.4)$$

is called the *typical price* in the conventional technical analysis⁴.

³ It is possible to use a global search optimisation method, for example, genetic algorithm, to find the global minimum.

⁴ Refer to Calculation of Typical Price while Calculating Commodities Channel Index (CCI). Available at http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:commodity_channel_in

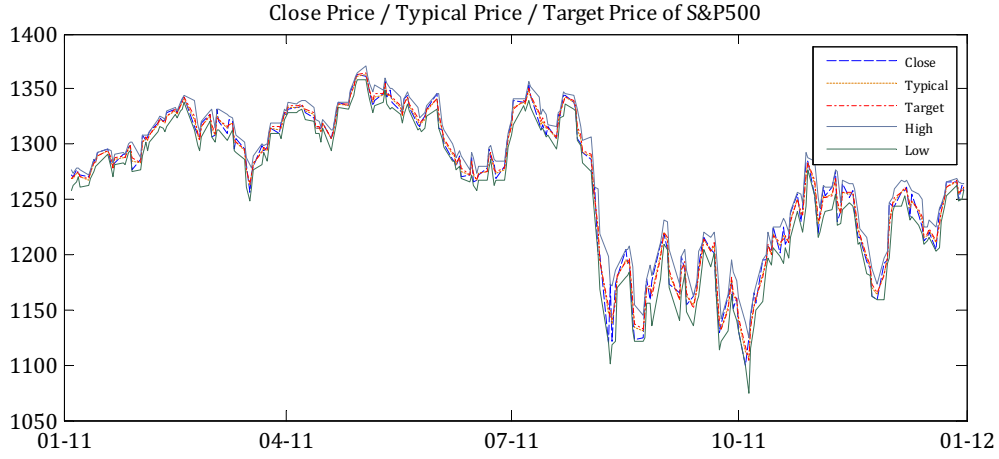


Figure 5.3: Close prices, typical prices and target prices of the S&P500

The close price, typical price and target price ($y_n = 0.5022x_n^O - 0.0317x_n^H + 0.0672x_n^L + 0.4626x_n^C$ in this case) of the S&P500 are illustrated in Figure 5.3

Figure 5.4 shows the spectrum of the first order difference of typical prices and target prices respectively. We can observe that the target price time series has more low frequency components than high frequency ones. Comparing to Figure 5.2, the target price time series has the least flat spectrum, while the spectrum of the typical price time series is less flat than any of those open, high, low and close price series whose spectrum are as flat as that of a random walk process.

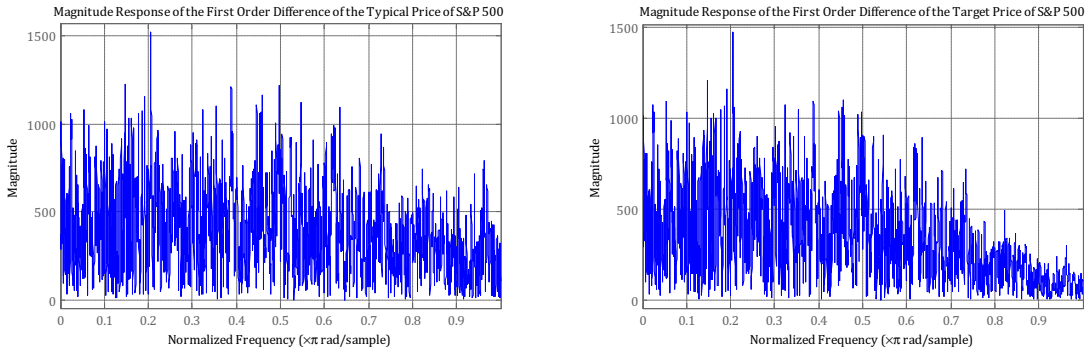


Figure 5.4: Magnitude response of the in-sample first order difference of typical and target prices of the S&P500

The flatness tests of the PSD of the first order difference of the close price, typical

price, and target price time series are shown in Table 5.1. The greater the $D(\varphi)$ and $\psi(\varphi)$ tests, the less flat the spectrum of the time series. Same as we observed, the target price time series has the least spectrum flatness, while the close price series has the greatest one⁵.

	Close Price	Typical Price	Target Price (OHLC)
$D(\varphi)$	0.6061	0.6605	0.9635
$\psi(\varphi)$	1.7963	1.8967	3.0718

Table 5.1: Flatness tests of the PSD of the first order difference of close/typical/target prices of the S&P500

The optimisation algorithm of the price combination that derives target prices effectively smooths the daily price time series, that is, it has more low frequency components than the high frequency ones as shown in Figure 5.4. Consequently, the target price time series derived from daily open, high, low and close prices has more predictability than any of the individual price time series alone, and it is feasible to apply the linear predictive filter methods as discussed in Chapter 4 to forecast it, and they would provide better performance than the naive forecast. The forecast procedures of real world financial time series are discussed in Section 5.3.

In practice, the OHLC target price time series should be used as a proxy of daily asset price time series for forecasting and trading, as the optimisation algorithm picks up the best combination of weights. In addition, for comparison, we also exploit the other combinations of daily open, high, low and close prices to derive target prices. The flatness tests of the PSD of the first order difference of target prices of the other combinations of daily open (O), high (H), low (L) and close (C) are shown in Table 5.2

⁵ For comparison, the average $D(\varphi)$ and $\psi(\varphi)$ test values for a random generated white noise process are about 0.58 and 1.66, respectively. Theoretically, the $D(\varphi)$ and $\psi(\varphi)$ tests are both 0 for ideal white noise.

	HLC	OHC	OLC	OHL	HL	OC	OH	OL	HC	LC
$D(\varphi)$	0.6697	0.9736	0.9625	0.6617	0.6503	0.9691	0.6554	0.6267	0.6397	0.6171
$\psi(\varphi)$	1.8956	3.0121	3.0786	1.8108	1.8464	2.9969	1.9700	1.7877	1.9372	1.7625

Table 5.2: Flatness tests of the PSD of the first order difference of target prices of other Combinations

We can tell that the combinations which include daily open and close prices give less flat spectrum thus more predictability. This phenomenon can be explained by the special nature of financial time series, that is, the daily close price is close in value to the open price of the next day. Let's consider a target price which is simply the average of daily open and close price (OC), that is, $y_n = \frac{x_n^O + x_n^C}{2}$. If the daily open price is exactly the same as the close price of the day before, i.e., $x_n^O = x_{n-1}^C$, then $y_n = \frac{1}{2}x_{n-1}^C + \frac{1}{2}x_n^C$. Therefore, the daily target price time series $\{y_n\}$ is effectively the daily close price series $\{x_n^C\}$ passing through a lowpass filter $H(z) = 0.5 + 0.5z^{-1}$ (Figure 5.5).

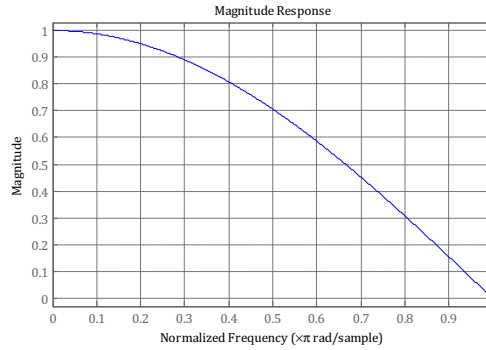


Figure 5.5: Effective lowpass filter

In essence, as long as the condition $x_n^O = x_{n-1}^C$ holds, $\{y_n\}$ can be represented as a lowpass filtered daily close price series $\{x_n^C\}$ by $H(z)$, and the magnitude response of $H(z)$ can be shaped by assigning different parameters to the filter and further shaped by introducing extra price information such as daily high, low, etc.

However, the daily open price is not identical to the close price of the day before due to after-hours trading and changes in investor valuations or expectations of the asset

occurring outside of trading hours.⁶ Consequently, the target price time series we get is not exactly band-limited but has more low frequency components than the high frequency ones as shown in Figure 5.4. This still give us advantage to forecast the target price time series using linear predictive filter methods.

5.2.2 Intraday Price

The concept of the target price can also be extended to the combination of intraday prices provided the high frequency price data are available. In the similar fashion as the weighted daily open, high, low and close prices, the broad concept of the target price of a financial time series is defined as the linear combination of higher frequency intraday prices, for example, intraday hourly prices or intraday 5-minute prices.

Taking the linear combination of intraday hourly prices for example, the target price on a certain day n can be represented as

$$y_n = w_0 x_n^{(0)} + w_1 x_n^{(1)} + \cdots + w_m x_n^{(m)} = \sum_{i=0}^m w_i x_n^{(i)} \quad (5.5)$$

where w_i denotes the weight of the i th hourly price $x_n^{(i)}$ on day n , and m is the total number of hourly prices within a day. $x_n^{(0)}$ is effectively the open price on day n . In the matrix form, the target price time series $\{y_n\}$ can be expressed as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_1^{(0)} & x_1^{(1)} & \cdots & x_1^{(m)} \\ x_2^{(0)} & x_2^{(1)} & \cdots & x_2^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{(0)} & x_n^{(1)} & \cdots & x_n^{(m)} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \quad (5.6)$$

⁶ Refer to the “*Opening Price*” explanation by Investopedia. Available at <http://www.investopedia.com/terms/o/openingprice.asp>

The weight w_i can be derived by applying the similar objective function as Equation 5.1

$$\max_w \left(\sum \frac{|Y_1(k)|^2}{N+1} - \sum \frac{|Y_2(k)|^2}{L-N} \right) \quad (5.7)$$

where the magnitude response $Y(k)$ of the target price time series $\{y_n\}$ is expressed by Equation 5.2 and the constraints in Equation 5.3 should be satisfied when applying the optimisation program.

It is worth noticing that the popular algorithm trading strategies *Time Weighted Average Price* (TWAP) and *Volume Weighted Average Price* (VWAP) [69] are special cases of the intraday target price.

The target price of daily open, high, low and close prices (OHLC), hourly price (1H) and 5-minute prices (5m) and the close price of the S&P500⁷ are illustrated in Figure 5.6.

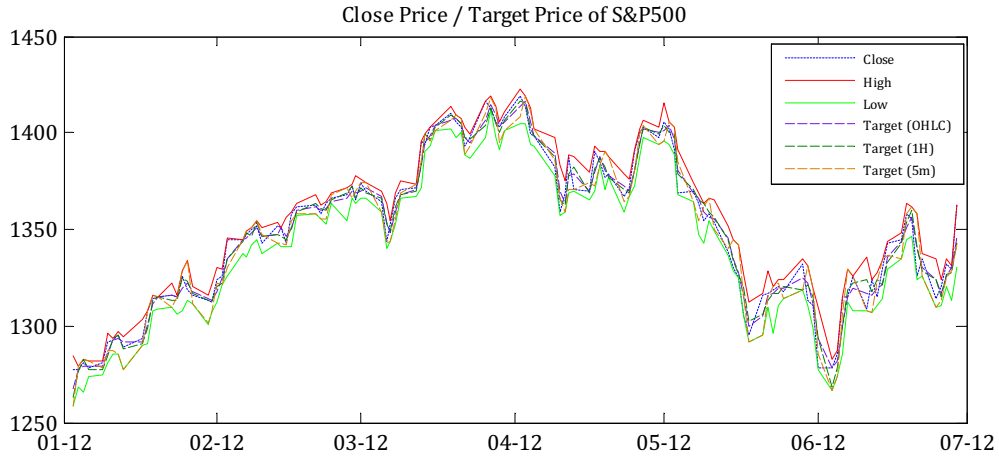


Figure 5.6: Close prices and target prices of the S&P500

Figure 5.7 shows the magnitude responses of the first order difference of close prices

⁷ The intraday hourly and 5-minute price data of the S&P500 are available from 2012-01-01 to 2012-06-30. There are 125 trading days within this period. There are 8 hourly prices and 80 5-minute prices within a day including the open prices.

and target prices HLC, OHLC, OC, hourly prices and 5-minute prices, respectively.

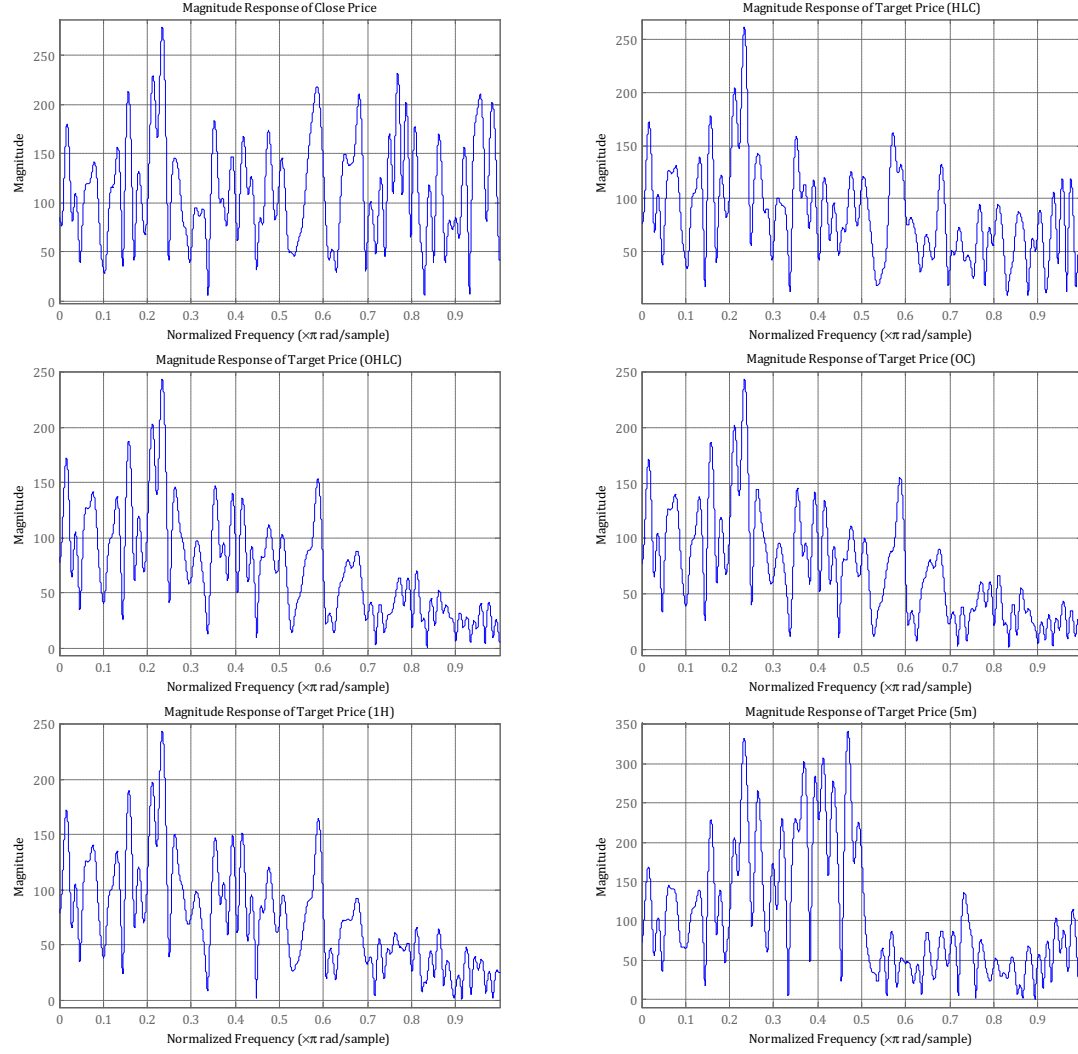


Figure 5.7: Magnitude response of the in-sample first order difference of close/target prices of the S&P500

The flatness tests of the PSD of the first order difference of close prices and target prices of daily open, high, low and close (OHLC), hourly prices and 5-minute prices, are given in Table 5.3.

	Close Price	Target Price (OHLC)	Target Price (1H)	Target Price (5m)
$D(\varphi)$	0.4245	0.8294	0.8231	1.0784
$\psi(\varphi)$	1.0711	2.4483	2.6167	2.8379

Table 5.3: Flatness tests of the PSD of the first order difference of close/typical/target prices

For comparison, the flatness tests of the PSD of the first order difference of target price of other combination of daily open (O), high (H), low (L) and close (C) prices are shown in Table 5.4

	HLC	OHC	OLC	OHL	HL	OC	OH	OL	HC	LC
$D(\varphi)$	0.4940	0.8520	0.8208	0.5641	0.4961	0.8445	0.5903	0.5044	0.5498	0.4212
$\psi(\varphi)$	1.2054	2.5106	2.3598	1.6745	1.3169	2.4381	1.9726	1.6471	1.5833	1.1670

Table 5.4: Flatness tests of the PSD of the first order difference of target prices of other combinations

The target price series derived from 5-minute prices has the least flat spectrum, which implies the target price (5m) time series is the most predictable one. The close price has the most flat spectrum as expected. Therefore it has the least predictability. The target price series derived from 1-hour prices has similar spectrum flatness as the OHLC target price series. Both of them are less predictable than the 5-minute target price series but more predictable than the close price series. The daily open, high, low and close prices are essentially the intraday prices at certain time points. Therefore, the high frequency price information provides more flexibility and possibility to obtain a more predictable target price time series than the low frequency ones.

From the perspective of asset price movements, the target price is effectively a price at a certain time point in theory⁸ within the daily Open-Close time range, and its magnitude is within the daily High-Low price range. In other words, the target price of day n is $y_n = x_n^{(k_n)}$, where $O_n \leq k_n \leq C_n$ and $x_n^{(L)} \leq x_n^{(k_n)} \leq x_n^{(H)}$, O_n and C_n denote the open and close time of day n of the market. Normally k_n is changing over day n depending on the actual price movements within the day n . The target price at time k_{n-1} of day $n-1$, i.e., $x_{n-1}^{(k_{n-1})}$ effectively contains price information from open time O_{n-1} to time k_{n-1} and from time k_{n-1} to close time C_{n-1} , rather than only the price information at fixed time k itself ($x_{n-1}^{(k)}$). Consequently, the extra price information from time k_{n-1} to close time C_{n-1} of day $n-1$ that embedded in the

⁸ It is possible for the market never to trade at the target price in practice if it gaps for example.

target price $x_{n-1}^{(k_{n-1})}$ provides more predictability to forecast the target price $x_n^{(k_n)}$ as long as $k_n < C_n$ because the trading time interval between C_{n-1} and k_n is less than one period, i.e., a day in this case. As a result, the target price time series $\{x_n^{(k_n)}\}$ has more predictability than time series of prices at fixed time k ($\{x_n^{(k)}\}$), for example, close price time series $\{x_n^C\}$.

In practice, if the intraday high frequency price data is not available, daily open, high, low and close prices can be used to derive a target price time series, which could provide similar predictability as the target price of intraday high frequency prices. It is worth noticing that the assumption behind the target price time series derived from daily close and open prices is that the daily open prices are identical or at least close to the close prices of prior days. In the situation that the daily close prices are always significantly different from the open prices of next days, alternative combinations of intraday high frequency or high, low prices should be used to derive the target price time series given that it has more predictability than the close prices time series.

5.3 Forecast Procedure

5.3.1 Financial Time Series Forecasting Procedure

For a real world financial time series, it is necessary to first check whether it follows a random walk process, or equivalently, whether the first order difference, or the return of the price time series follows a white noise process. This can be done by Variance Ratio tests on the price time series, or Spectrum Flatness tests on the first order difference or the return of the time series. It is reasonable to forecast a price time series only if it does not follow a random walk process, otherwise, the naive approach should be simply used to forecast it.

If a price time series does not follow a random walk process, a linear predictive filter

is then applied to the in-sample observations of the first order difference of the price time series . It is worth noticing that either fixed parameter models (Section 4.1.1), or adaptive/hybrid filters can be used for forecasting, but both of them will need to be updated periodically to adjust for the possible future structural breaks. The goodness-of-fit of the out-of-sample forecast has to be examined against that of the naive forecast. If the forecast method is better than the naive approach, it will be used in the real world forecast. The forecast procedure is illustrated in Figure 5.8.

For a financial time series whose daily open, high, low and close prices are available, the process to optimise the price combination to derive target prices will be added to the forecast procedure. The rest of the procedures shown in Figure 5.8 are applied to the target price time series. Moreover, the estimated forecast model will be applied to the target price time series to forecast one-step ahead.

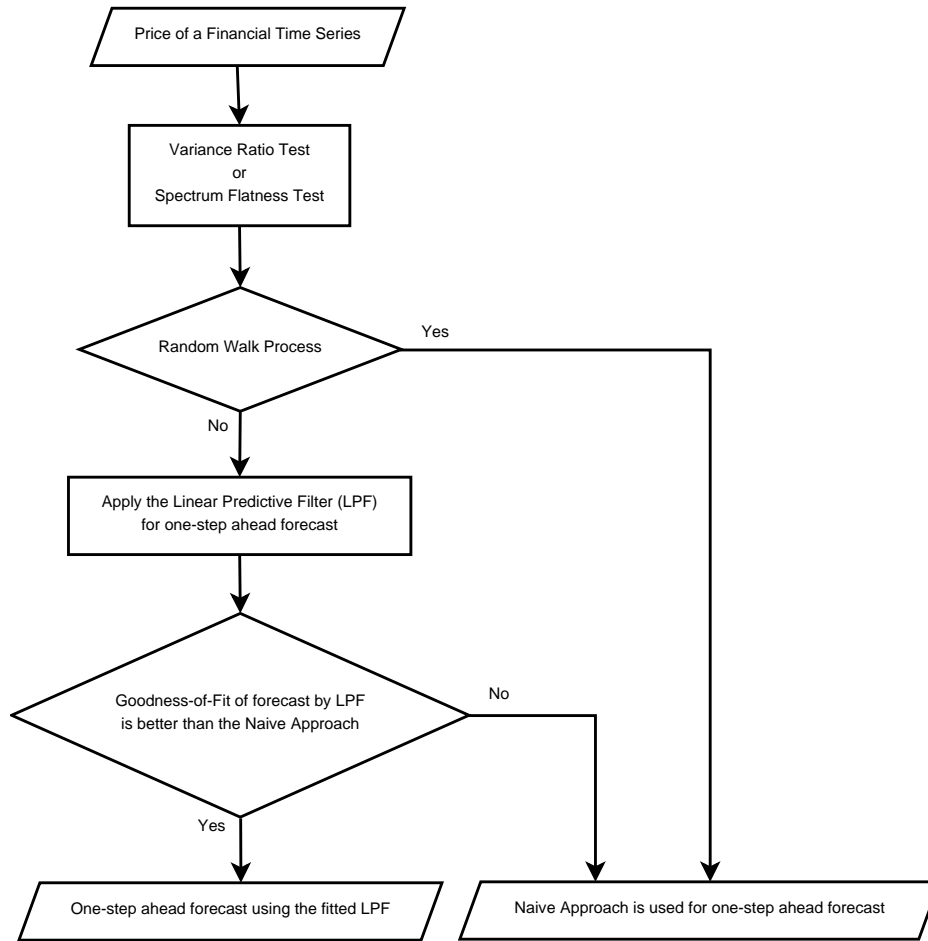


Figure 5.8: Forecast procedure for a price time series

To summarise, the forecast procedures for a financial time series with open, high, low and close prices are listed as follows:

1. To optimise the price combination of open, high, low and close prices to derive a target price time series;
2. To apply Variance Ratio tests or Spectrum Flatness tests to the target price series to check whether it follows a random walk process;
3. To estimate a linear predictive filter or an adaptive/hybrid filter using the in-sample data set and apply the fitted filter to the out-of-sample data set of the target price series only if it does not follow a random walk process, otherwise apply the naive approach;

4. To compare the goodness-of-fit test results of the out-of-sample forecast of the linear predictive filter or the adaptive/hybrid filter to those of the naive approach;
5. To use the estimated filter for real world forecast of the target price time series if the out-of-sample goodness-of-fit is better than that of the naive approach, otherwise apply the naive approach.

5.3.2 Procedure of Applying the Linear Predictive Filter

In order to apply the linear predictive filter to the out-of-sample data set, it is necessary to divide the whole data set into 3 distinct subsets at first, i.e., in-sample training set, in-sample testing set, and out-of-sample validation set (Figure 5.9). The in-sample training set is the largest set and is used by the linear predictive filter to learn the pattern presented in data. The in-sample testing set, whose length will be set as same as that of the out-of-sample validation set, is used to test the performance of the estimated model from the training set. The specified order of the filter, length of the moving window and weight for EWSDWESF that give the best performing model will be selected as the winning model configuration, and then applied to the whole in-sample data set to estimate the coefficients of the filter to obtain the estimated linear predictive filter. The out-of-sample validation set is then used for out-of-sample goodness-of-fit test of the estimated linear predictive filter.

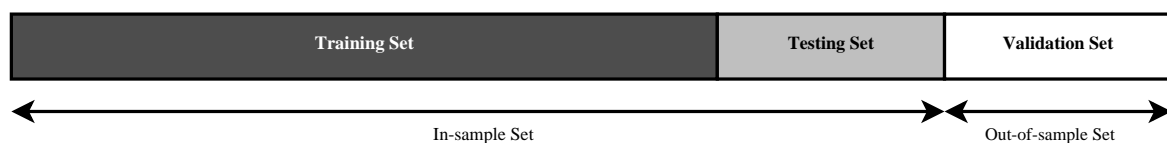


Figure 5.9: In-sample and out-of-sample data set

Once the model configuration and coefficients are obtained, we could use the estimated linear predictive filter for one-step ahead forecast of a real world financial time series. The linear predictive filter could be applied to the whole in-sample and out-of-sample

data set to forecast one-step ahead directly given there are no structural breaks. Or a new set of coefficients are estimated based on the whole in-sample and out-of-sample data set and the new linear predictive filter is then used for one-step ahead forecast. The procedure of applying the LPF to forecast real world time series is illustrated in Figure 5.10.

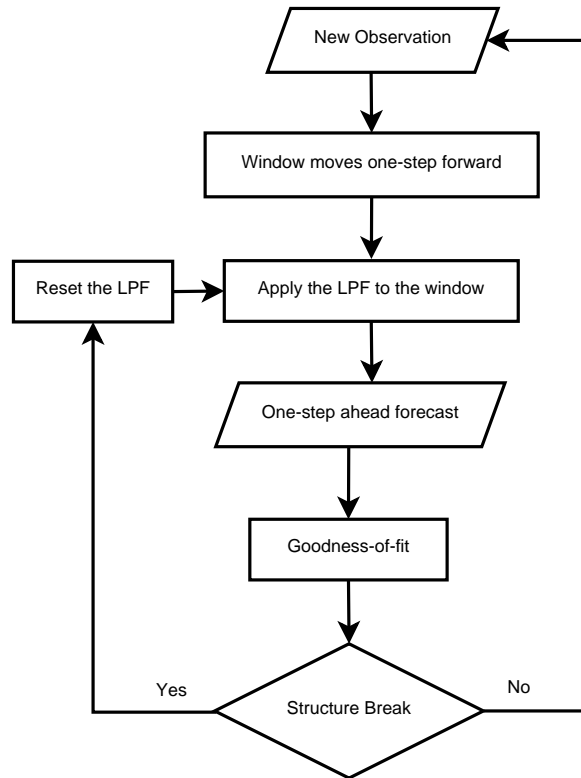


Figure 5.10: Real world forecast procedure of a LPF

When a new observation is available, the window whose length is the sum of the length of the in-sample data set and that of the out-of-sample data set (Figure 5.9) is then moved one-step forward to encompass the observation and the estimated LPF is applied to the window to forecast one-step ahead. The one-step ahead forecast is then added into the out-of-sample data set to derive the goodness-of-fit and compared to the previous tests. If the goodness-of-fit tests are significantly different from the previous ones, there are possible structural breaks in the time series. Therefore, the model configuration is reset and new coefficients are then estimated from the current window

using the aforementioned procedure. Figure 5.11 shows the procedure to forecast one-step ahead for the real world financial time series when there is a structural break.

It is worth noticing that the procedure of resetting model configuration and coefficients estimation is different from that of the adaptive filter or hybrid filter (Section 4.4). The latter only updates coefficients of the model which has the same set of model configuration with the same length of moving window. Therefore, the model configuration also needs to be reset given significant structural breaks when applying adaptive filter or hybrid filter for one-step ahead forecast of a real world financial time series.

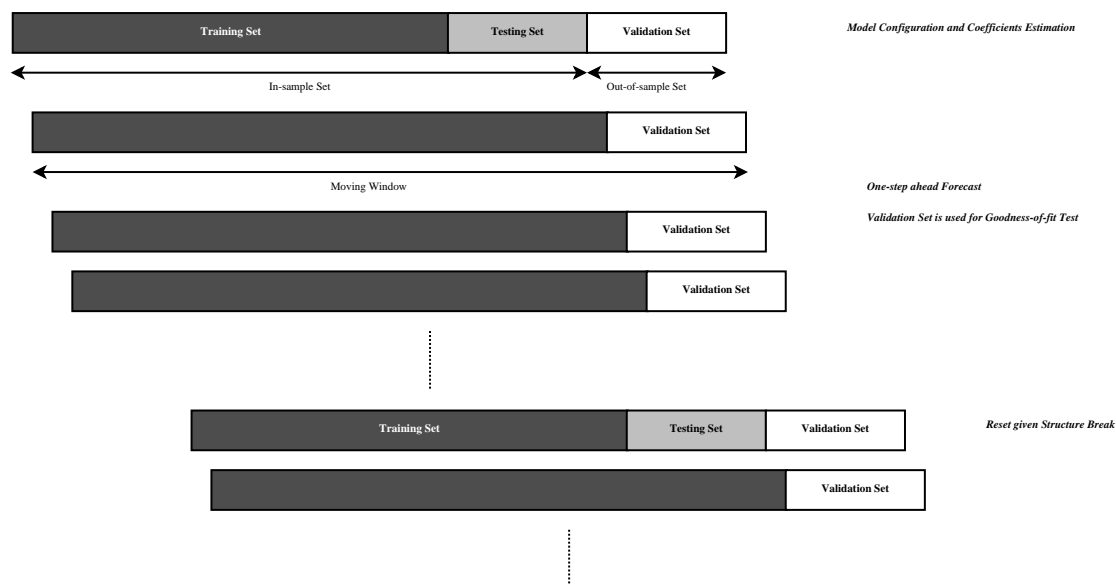


Figure 5.11: Model configuration reset and coefficients estimation if structure breaks

5.4 Trading Strategy

The one-step ahead forecast based on target prices deals with the probability of up or down movements of asset prices which are proxied by target prices rather than close prices. Hence it is necessary to develop trading strategies based on the target price time series. In other words, we need path-dependent trading strategies that enable us to make trading decision based on target prices rather than conventional technical indicators that based on close prices.

The proposed trading strategy is based on target prices of liquid assets, e.g., equity indices, foreign exchange, etc. The idea is to choose the most liquid assets, such as EUR/USD spot rate, S&P500 index, etc., which have the smallest bid/ask spreads, with which we could minimise transaction costs if we trade in high frequency. The other reasons include that liquid assets normally have more variable open, high, low and close prices, relatively greater daily High-Low ranges, more active trades than the less liquid ones, whose prices change less frequently. The target prices of a liquid asset can easily be reached thus triggering trading signals due to the active trading during a day. In addition, it is possible for the target price time series to have more low frequency components due to looser High-Low range constraints.

Take the daily trade on the S&P500 index as an example, we would go long S&P500 when the target price of yesterday is higher than the open price of today and the sign forecast shows that there is more than 50% possibility that the target price of today would be higher than that of yesterday, while we would go short S&P500 when the target price of yesterday is lower than the open price of today and the sign forecast shows that there is more than 50% possibility that the target price of today would be lower than that of yesterday. The flow chart of the strategy is shown in Figure 5.12.

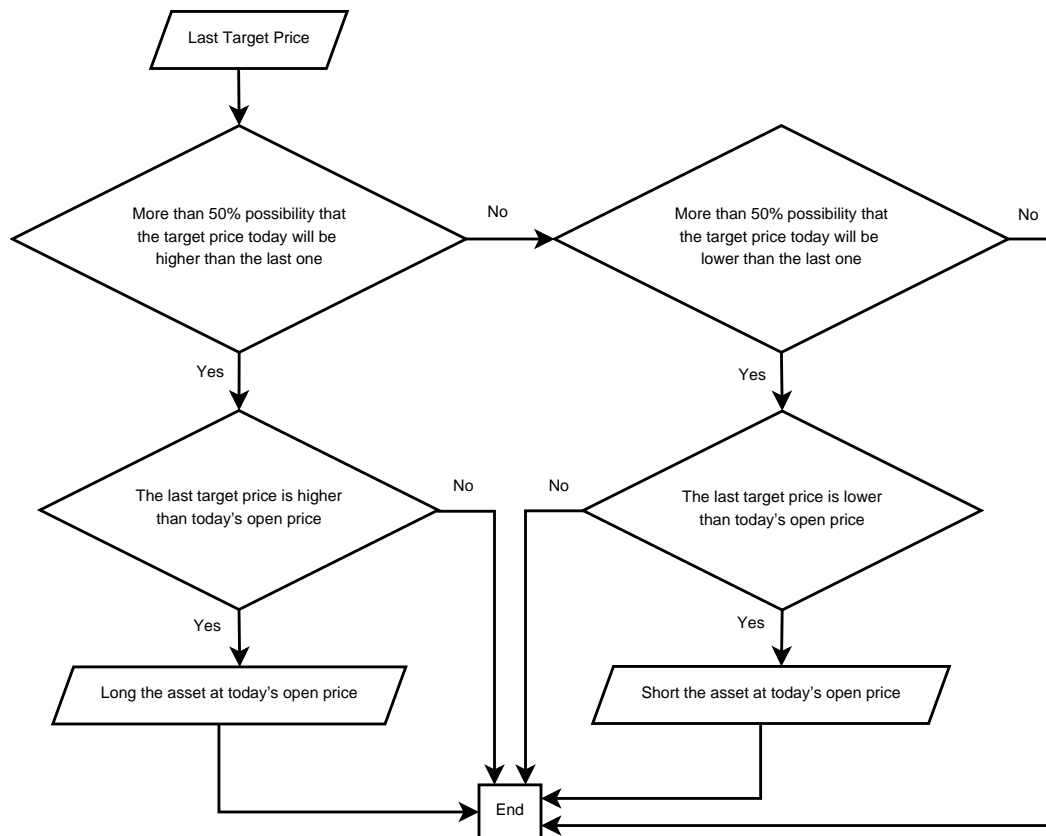


Figure 5.12: Target price trading strategy

This strategy ensures that the probability of upside or downside movements from today's open price are even more possible than estimated, which gives more certainty of making profits.

If the MSE of the one-step ahead forecast is small (smaller than the MSE of the naive forecast), which implies the magnitude forecast of target prices are accurate (more accurate than that of the naive forecast), a trading strategy could be designed by comparing the forecasted target price, instead of the target price of yesterday, to today's open price. If the forecasted target price is higher than today's open price, we would long S&P500, otherwise, we should short it. The direction forecast of the target price can also be considered into the trading strategy to add more confidence.

In order to prevent suffering large losses resulting from extreme unfavorable price movements, we should set up a stop loss level with every order we placed. This

ensures that current orders would exit automatically upon predefined losses in case that prices move to the unfavorable direction. In addition, we should also set up a take profit limit level with each order in order to realise the profits when prices reach the predefined levels automatically in case they will drop afterwards. The stop-limit levels give us precise control over the exit of an open position with respect to the desired minimum level of profit and maximum level of acceptable loss. The stop-limit levels could be created as proportional to target prices, high-low ranges or at absolute levels depending on backtest results. Furthermore, the trailing stop⁹ could be incorporated to generate more profits given the correct direction forecast while less losses when the price moves unfavorably. If we apply this trading strategy in a long run, statistically we would end up with positive profits.

One problem with this trading strategy is that we could suffer unexpected losses if the price moves to the unfavorable direction and triggers the stop loss level before it moves to the favorable direction as expected by the sign forecast. The reason is that the sign forecast is not path-dependent, as it only tells the possibility of the one-step ahead forecast of target price to be higher or lower than the current target price during a certain period but not the exact price movement path during this period. Therefore, we should properly set up stop-limit levels in order to make profits by applying this trading strategy. Moreover, it is reasonable to combine conventional momentum indicators or candle charting methodologies [103] with the one-step ahead forecast to refine the performance of the trading strategy.

⁹ Trailing stop orders are used to maximise and protect profit as an asset price moves to the favorable direction and limit losses otherwise. The trailing stop price is adjusted as the asset price fluctuates. Refer to “*Trailing-Stop Techniques*” by Investopedia. Available at <http://www.investopedia.com/articles/trading/03/080603.asp>

5.5 Empirical Results and Analysis

To further explain the linear predictive filter methodology and its effectivity of forecasting the target price of a real world financial time series, we apply the filters to some of the real world financial time series and report out-of-sample goodness-of-fit tests in this section. In addition, the proposed trading strategy is applied to the one-step ahead forecast of target prices in order to test its profitability.

5.5.1 Data

The daily S&P500 index data set (2252 observations, 2003-01-24 \rightarrow 2011-12-30) as shown in Section 5.1 and 5.2 is used to demonstrate the predictability of the linear predictive filter over the target price time series. The first 2000 observations (2003-01-24 \rightarrow 2010-12-31) are used as the in-sample data set, where the first 1748 observations (2003-01-24 \rightarrow 2009-12-31) are used as the training set and the next 252 observations (2010-01-04 \rightarrow 2010-12-31) the testing set. The last 252 observations (2011-01-01 \rightarrow 2011-12-30) are used as the out-of-sample validation set. The daily open, high, low and close prices of 2252 trading days are extracted from Bloomberg. Two target price time series, $y_n = 0.5022x_n^O - 0.0317x_n^H + 0.0672x_n^L + 0.4626x_n^C$ (OHLC) and $y_n = 0.4010x_n^H + 0.4507x_n^L + 0.1495x_n^C$ (HLC), and the close price time series are used as proxy price series of the S&P500. They are derived by optimising the objective function in Section 5.2.

5.5.2 Goodness-of-fit

5.5.2.1 Methodologies and Time Series

7 methods are used for one-step ahead forecast:

1. the ARMA model¹⁰ estimated by minimising the prediction error;
2. the FIR filter estimated by minimising the variance of error periodogram;
3. the FIR filter estimated by maximising the SDWESF;
4. the FIR filter estimated by maximising the EWSDWESF;
5. the adaptive filter;
6. the hybrid filter;
7. the naive forecast.

The naive forecast and ARMA model are used as benchmark of the out-of-sample forecast performance. The proposed methods are then compared to the naive forecast and ARMA model when they are applied to the financial time series.

In order to demonstrate the effectivity of each method and the predictability of target price time series, the 7 methods are applied to 3 price time series of the S&P500:

1. the OHLC target price time series ($y_n = 0.5022x_n^O - 0.0317x_n^H + 0.0672x_n^L + 0.4626x_n^C$)
2. the HLC target price time series ($y_n = 0.4010x_n^H + 0.4507x_n^L + 0.1495x_n^C$)
3. the Close price time series (x_n^C)

All three price time series reject the null hypothesis of random walk at 99% confidence level when applied the Chow-Denning VR test¹¹. Thus none of them follow random walk processes.

¹⁰The *armasel* function of the Matlab toolbox ARMASA (<http://www.dsc.tudelft.nl/Research/Software/index.html>) by P.M.T. Broersen is used as a benchmark method of ARMA model in this thesis

¹¹The Chow-Denning statistic is calculated by the *Chow.Denning* function of R package *vrtest*. The MV_2 statistics with holding period $k = [2, 4, 8, 16, 32]$ for the OHLC, HLC target price and close price time series are 8.32, 5.43, 3.46, respectively, comparing to the 1% critical value 3.08.

The target price calculation and proposed trading strategy can be implemented in Matlab using functions listed in Section C.3.

5.5.2.2 Goodness-of-fit Test Results and Analysis

Table 5.5 lists the goodness-of-fit test results of the out-of-sample forecast of the OHLC target price time series. The goodness-of-fit test results are obtained by applying each method to the out-of-sample validation set following the forecast procedure described in Section 5.3.2.

S&P500 (OHLC)	ARMA	Periodogram	SDWESF	EWSDWESF	Adaptive	Hybrid	Naive
MSE	92.74	93.08	91.96	93.10	94.32	93.38	151.65
MASE	0.7999	0.8004	0.7994	0.8037	0.8061	0.8024	1
SR	71.03%	70.64%	71.43%	72.62%	70.24%	71.03%	64.68%
DA	6.6836	6.5583	6.8118	7.2619	6.4302	6.6836	4.6667

Table 5.5: Out-of-sample goodness-of-fit of OHLC target prices of the S&P500

Table 5.6 shows the goodness-of-fit test results of the out-of-sample forecast of the HLC target price time series.

S&P500 (HLC)	ARMA	Periodogram	SDWESF	EWSDWESF	Adaptive	Hybrid	Naive
MSE	144.38	146.37	144.11	148.66	141.38	140.80	219.33
MASE	0.8093	0.8162	0.8080	0.8260	0.8050	0.8041	1
SR	64.68%	64.68%	61.11%	65.48%	64.29%	63.49%	60.71%
DA	4.6568	4.6568	3.5289	4.9300	4.5469	4.2868	3.3991

Table 5.6: Out-of-sample goodness-of-fit of HLC target prices of the S&P500

The goodness-of-fit test results of the out-of-sample forecast of daily close prices are listed in Table 5.7.

S&P500 (Close)	ARMA	Periodogram	SDWESF	EWSDWESF	Adaptive	Hybrid	Naive
MSE	314.78	324.55	317.13	322.02	302.87	294.14	700.77
MASE	0.6812	0.6872	0.6779	0.6779	0.6827	0.6783	1
SR	48.81%	49.21%	48.02%	52.38%	51.98%	48.81%	52.38%
DA	-0.2485	-0.18137	-0.50314	0.2680	0.6705	0.63469	0.61877

Table 5.7: Out-of-sample goodness-of-fit of close prices of the S&P500

We can tell that the predictability of the price time series of the S&P500 can be sorted as OHLC target price series > HLC target price series > Close price series. All 7 methods applied to the OHLC target price time series outperform those applied to the close price series in terms of the MSE, Success Ratio and DA tests.

The best SR is 72.62% when applying the linear predictive filter to the OHLC target price time series, while 65.48% for the HLC target price time series and 52.38% for the close price time series. Even the success ratio of the OHLC target price for the naive forecast could reach 64.68%, which implies we could achieve 64.68% correct hit rate by simply using the most recent observation as the one-step ahead forecast of the target price time series without any sophisticated models. By contrast, the best success ratio of close price forecast is only 52.38%. This is consistent with the EMH that states the close price of an asset cannot be predicted better than the naive forecast, especially for the highly liquid asset classes, e.g., the S&P500 index in this case.

If we consider OHLC target prices, HLC target prices and close prices as proxies of daily S&P500 prices, the MSE of forecasting S&P500 daily prices are improved significantly from 700.77 of the naive forecast of the close price series to 91.96 of the linear predictive filter of the OHLC target price series. Although the MSE differences are results of forecasting different price time series, the different prices, i.e, OHLC target price, HLC target price and close price, are all proxies of the daily price of the S&P500. We could actually place orders at any of these prices in the market. Therefore, the improvement of magnitude forecast of target price time series over the close price time series would be more helpful for trading decision making processes than simply using close prices as the daily S&P500 prices.

5.5.2.3 Group Delay

The estimated linear predictive filter using the EWSDWESF method for the OHLC target price of the S&P500 is $y_n = 0.8135y_{n-1} - 0.6574y_{n-2} + 0.5091y_{n-3} - 0.3605y_{n-4} +$

$0.2814y_{n-5} - 0.1579y_{n-6} + e_n$ and for the HLC target price is $y_n = 0.2291y_{n-1} - 0.1788y_{n-2} + e_n$. Figure 5.13 and 5.14 show the magnitude responses and group delay of the two LPFs. The group delay at some low frequency bands of both LPFs is negative, which implies the one-step ahead forecast of these low frequency components of the target price series leads the time series itself. However, the group delay of high frequency components of the LPF designed for the OHLC target price time series is slightly higher than that of the naive forecast (whose Group Delay = 1). We can tell from Figure 5.4 that the high frequency components of the target price time series where the LPF has group delay greater than 1 is far less than the low frequency components where the group delay of the LPF is less than 1. Consequently, the estimated LPF is more suitable for a time series that has more low frequency components than high frequency ones, e.g., the OHLC and HLC target price series whose magnitude responses are shown in Figure 5.4.

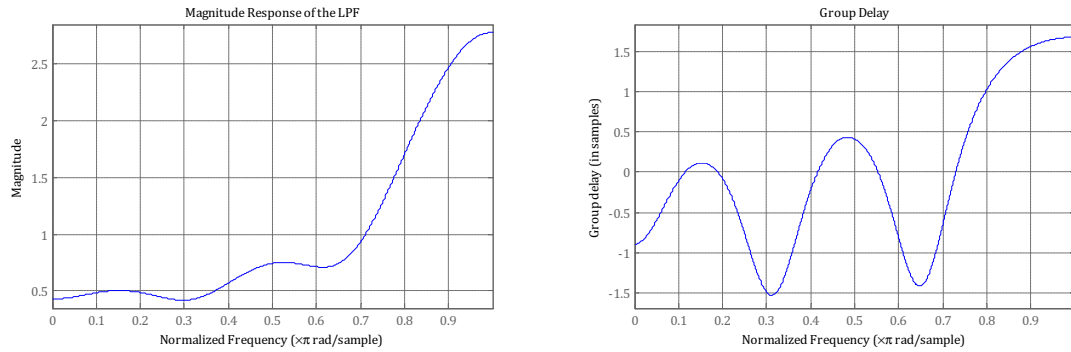


Figure 5.13: Magnitude response and group delay of the LPF for the OHLC target price series

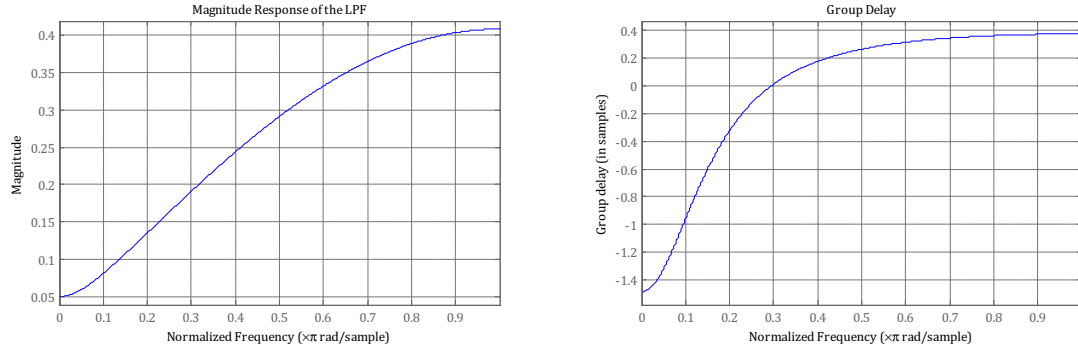


Figure 5.14: Magnitude response and group delay of the LPF for the HLC target price series

5.5.2.4 Magnitude Response $|F_k|$

According to the Theorem proposed in Section 4.5.3, the magnitude response of the LPF designed for the HLC target price time series $|F_k| < 1$ for all k (Figure 5.14). Therefore, the LPF provides better forecast performance than the naive approach in terms of the upper bound of forecast errors. It is confirmed by the MSE shown in Table 5.5. However, it is difficult to tell from Figure 5.13 whether the magnitude response of the LPF designed for the OHLC $\frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| < \frac{1}{N} \sum_{k=0}^{N-1} |X_k|$ as not $|F_k| < 1$ for all k . Due to the fact that the original OHLC target price time series has much less high frequency components where the magnitude response of the LPF is greater than 1, the actual upper bounds of forecast errors of the LPF are still less than the upper bounds of the naive forecast.

5.5.2.5 Coefficients of the Adaptive Filter and Hybrid Filter

The coefficients over time for the adaptive filter and hybrid filter designed for the OHLC target price time series and HLC target price time series are shown in Figure 5.16 and 5.16, respectively.

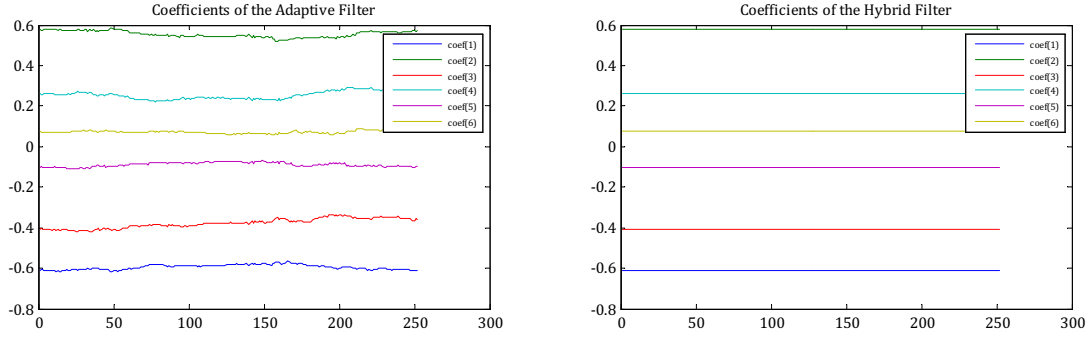


Figure 5.15: Coefficients of the adaptive filter and the hybrid filter for OHLC target prices of the S&P500

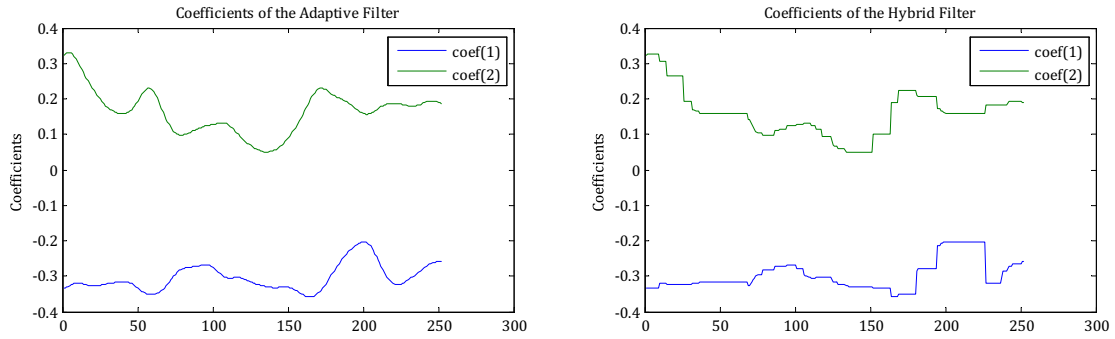


Figure 5.16: Coefficients of the adaptive filter and the hybrid filter for HLC target prices of the S&P500

We can observe from changing coefficients over time for both the adaptive filter and the hybrid filter in Figure 5.16 that there are some structural breaks in the out-of-sample data set of the HLC target price time series. The existence of structural breaks explains the reason why both filters deliver better MSE than fixed parameter models (Table 5.6). However, for the OHLC target price time series, the coefficients of the adaptive filter only change slightly while those of the hybrid filter have no changes (Figure 5.15). It implies that there are no structural breaks in the OHLC target price time series thus the adaptive/hybrid filter doesn't outperform those fixed parameter models in this case (Table 5.5).

Furthermore, the structural breaks illustrated in Figure 5.16 are not as significant as those in Figure 4.8 and 4.10. Therefore, we could use the same model configuration to

update the filters for the out-of-sample forecast until there is a significant structural break detected, when we should reset the model configuration and update coefficients of the new model.

5.5.2.6 Forecast Errors Analysis

At the end of the forecast procedure, errors of the out-of-sample forecast is checked against the white noise process. This can be done conventionally by the Ljung-Box Q^* Statistic¹² shown in Section 2.1.2, or plotting the *autocorrelation function* (ACF) and *partial autocorrelation function* (PACF), illustrated in Figure 5.18.

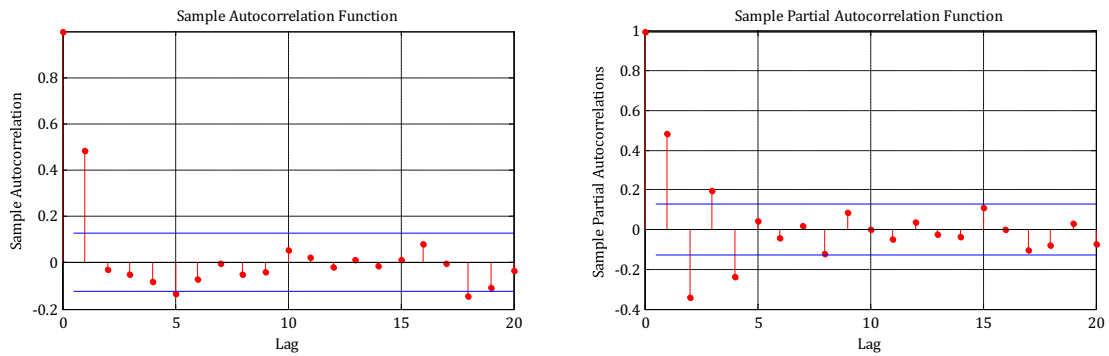


Figure 5.17: ACF and PACF of the out-of-sample data set of OHLC target prices of the S&P500

¹²The Ljung-Box Q^* Statistic is calculated by the *lbqtest* function of Matlab

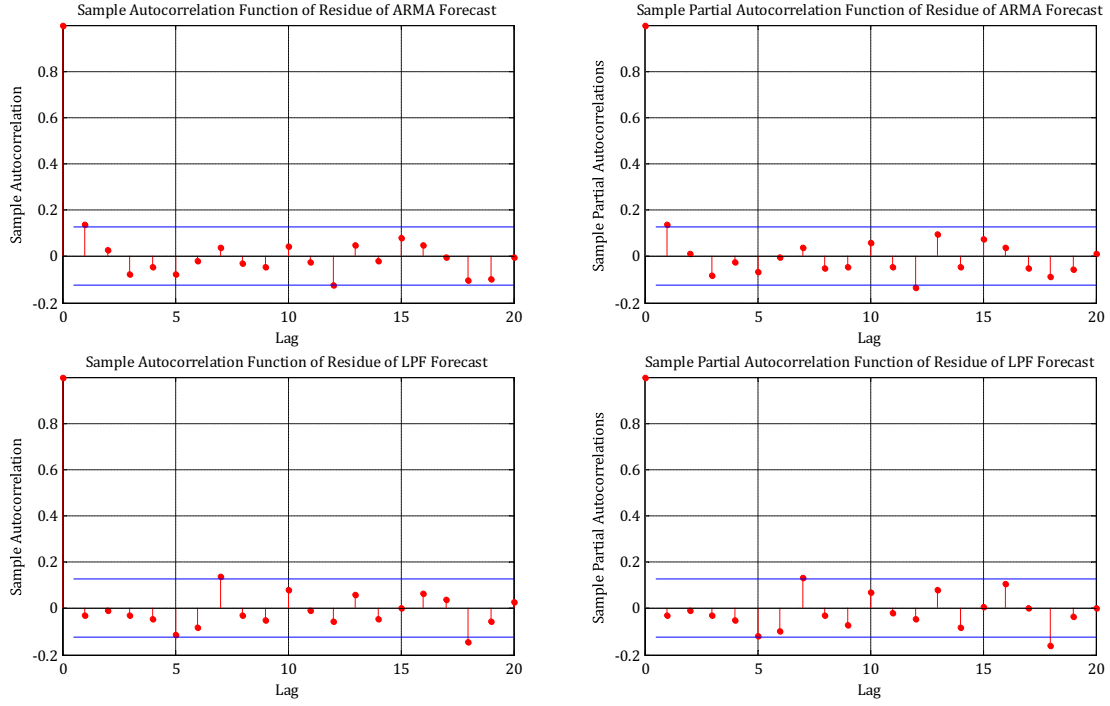


Figure 5.18: ACF and PACF of the residue of the out-of-sample forecast by ARMA and LPF

The Ljung-Box Q^* Statistic fails to reject the the null of no residual autocorrelation for either ARMA or LPF estimated by maximising EWSDWESF. Figure 5.17 shows that there is significant autocorrelation of the target price time series at lag 1. Figure 5.18 shows that the ACF and PACF of the residues at lag 1 are slightly higher than the confidence bound as a result of the ARMA forecast, and significantly reduced by applying the LPF. This also demonstrates that the LPF is better than the ARMA in terms of lag 1 autocorrelation reduction in this case. The Ljung-Box Q^* Statistic and ACF/PACF of the residues of the ARMA and LPF forecast show that the residues follow a white noise process thus we could not extract further information from the residues using linear methods.

In addition, the Chow-Denning VR test¹³ fails to reject the null hypothesis of white noise at 95% confidence level for forecast errors of both ARMA and LPF. It confirms

¹³The MV_2 statistics with holding period $k = [2, 4, 8, 16, 32]$ for the forecast errors of ARMA and LPF are 2.05 and 1.20, respectively, comparing to the 5% critical value 2.31.

that both forecast errors are white noise processes.

5.5.2.7 Other Financial Time Series

In addition, the 7 methods are applied to FTSE Index, Shanghai Composite Index (SHCOMP), EUR/USD OHLC target price time series, which represent the equity index, emerging market equity index, and exchange rate, respectively. The same length of data sets are used as that of the S&P500. The daily prices are also provided by Bloomberg.

FTSE	ARMA	Periodogram	SDWESF	EWSDWESF	Adaptive	Hybrid	Naive
MSE	1613.16	1681.51	1640.93	1828.63	1902.23	1620.21	2918.22
MASE	0.7689	0.7837	0.7788	0.8075	0.8102	0.7787	1
Success Ratio	71.43%	73.02%	72.62%	73.81%	72.22%	71.43%	68.65%
DA	6.8017	7.3082	7.1826	7.5782	7.0569	6.8017	5.9259

Table 5.8: Out-of-sample goodness-of-fit of OHLC target prices of FTSE

SHCOMP	ARMA	Periodogram	SDWESF	EWSDWESF	Adaptive	Hybrid	Naive
MSE	603.32	584.16	521.18	514.44	522.37	522.47	780.08
MASE	0.8860	0.8671	0.8299	0.8198	0.8242	0.8242	1
Success Ratio	55.95%	58.73%	59.52%	59.52%	61.11%	60.32%	59.52%
DA	1.8694	2.7640	3.0167	3.0167	3.5289	3.2839	3.0167

Table 5.9: Out-of-sample goodness-of-fit of OHLC target prices of SHCOMP

EURUSD	ARMA	Periodogram	SDWESF	EWSDWESF	Adaptive	Hybrid	Naive
MSE ($\times 10^{-5}$)	3.3590	3.7147	3.5570	3.5194	3.6849	3.7363	7.9163
MASE	0.7652	0.7895	0.7771	0.7736	0.7866	0.7908	1
Success Ratio	69.05%	70.24%	71.43%	69.84%	71.43%	71.03%	61.11%
DA	5.9420	6.3371	6.7569	6.1977	6.7569	6.5927	3.4016

Table 5.10: Out-of-sample goodness-of-fit of OHLC target prices of EUR/USD

In conclusion, the proposed linear predictive filter and adaptive/hybrid filter methods significantly outperform the naive approach in all cases and are better than the benchmark ARMA method in most of the cases.

5.5.3 Profitability of the Trading Strategy

5.5.3.1 Financial Spread Betting

The proposed trading strategy is applied to the one-step ahead forecast of the OHLC target price time series to test its profitability in the context of financial spread betting¹⁴, where the spread is defined as the difference between the price we can buy at (offer price) and the price we can sell at (bid price). We would buy at the offer price if we think the market will rise, or sell at the bid price if we think it will fall.

Suppose the S&P500 Index is trading on the market at 1344 bid, and 1345 offer (quoted 1344-1345). If we think the index price is going to go up, we might bet £1 a point (order size) at the offer price 1345. If the price goes up to 1355-1356, we could realise profits by closing the position at the bid price 1355. The profits are then £10 ($(1355 - 1345) \times 1 = 10$) with the spread bet. The transaction cost is effectively the bid-offer spread, £1 ($1356 - 1355 = 1$) in this case.

The benefits of spread betting are tax free profits and leverage trade. The latter implies trading on margin. For example, if we place a bet worth the equivalent of £1345 and margin rate for that product is 5%. The initial deposit you are asked to hold with the broker is only £67.25 ($1345 \times 0.05 = 67.25$) - much less than the £1345 actual bet. However, the losses can also be amplified as well as profits by the leverage. Therefore, proper stop loss and take profit limit levels should be set up when applying the proposed trading strategy to financial spread betting.

5.5.3.2 P&L on the S&P500

Table 5.11 shows the profit and loss (P&L) of applying the trading strategy to the out-of-sample data set using OHLC target price of the S&P500. The Total Net Profit

¹⁴Key features of spread betting. IG Index. <http://www.igindex.co.uk/spread-betting/benefits-of-spread-betting.html>

is calculated as $TotalNetProfit = TotalProfit - TotalTrade$, where the number of Total Trade is effectively the total transaction costs (assuming spread is 1-point). The Initial Investment is calculated as the sum of the total number of trading days (252 trading days in 2011) multiplies spread, maximum stop loss¹⁵ for each trading days, and 5% margin on the current asset price¹⁶. it is worth noticing that this is the maximum initial investment that guarantee no margin calls and cover all the possible losses in 2011. The actual initial investment could be much less than this one and could be adjusted according to the actual price movements. The annual return is derived by $AnnualReturn = TotalNetProfit/InitialInvestment$, which is the worst case annual return in 2011 as a result of the maximum initial investment. If we deposit less initial investment with the broker and dynamically adjust it with price movements, we could achieve a much greater annual return¹⁷.

	ARMA	Periodogram	SDWESF	EWSDWESF	Adaptive	Hybrid	Naive
Long Profit	131.24	116.77	160.57	158.33	241.48	153.50	0
Short Profit	254.03	262.54	232.49	244.27	234.87	277.06	0
Total Profit	385.27	379.31	393.07	402.60	476.35	430.56	0
Total Trade	110	108	112	136	99	103	0
Total Net Profit	275.27	271.31	281.06	266.60	377.35	327.56	0
Initial Inv	667	667	667	667	667	667	667
Annual Return	41.27%	40.68%	42.13%	39.97%	56.57%	49.11%	0
Sharpe Ratio	3.23	3.41	3.30	3.20	3.60	3.49	0

Table 5.11: P&L of the S&P500 in 2011

It is worth noticing that the profits shown in Table 5.11 is obtained by placing level order, which is the buy or sell order that has the same amount of order size per point. However, we could achieve much higher profits by placing orders using certain money management approaches, for example, dynamic adjusting the order size with respect to the total account balance, e.g., increasing the order size if the total account balance

¹⁵Take profit and stop loss levels of this strategy are set at 5% and 0.1% of the price level of index of the previous day, respectively.

¹⁶Initial Investment in Table 5.11 is calculated as (total number of trading days \times (spread + highest high price \times maximum stop loss) + 5% \times highest high price)

¹⁷The annualised Sharpe Ratio is calculated as daily Sharpe Ratio multiplies $\sqrt{252}$.

goes up while decreasing the order size when the total account balance goes down.

The P&L of the proposed trading strategy (56.57% annual return, even the worst 39.97% annual return) significantly outperforms the buy&hold strategy, which achieves 0% annual return ($ClosePrice_{2011-12-30} - OpenPrice_{2011-01-03} = -0.02$) in 2011, and the momentum indicator strategy¹⁸ which has total profit -135.64 (i.e., total loss 135.64) with 32 trades in 2011. Table B.2 shows the backtest results of applying 23 trading strategies on historical prices (in 2011) of the S&P500 from Bloomberg. We can tell that none of the 23 trading strategies outperforms the proposed strategy in terms of annual return and Sharpe ratio.

5.5.3.3 P&L on Simulated Price Paths

The backtest fails to consider the possible intraday price movements because of the limitation to access the historical tick data in 2011. Instead, the daily high, low prices are used to check against the stop loss and take profit limit levels to derive the P&L. Consequently, the backtest result does not exactly reflect the profitability of the trading strategy. To overcome the problem, we simulate price paths using the Geometric Brownian Motion (GBM) model as described in [62], $dx_t = \mu x_t dt + \sigma x_t dW_t$, where x_t is the price at time t , μ the mean return, σ the standard deviation of return, and W a Wiener process or Brownian motion. The price x_T at time T can be represented as $x_T = x_0 e^{(\mu - \frac{\sigma^2}{2})T + \sigma\sqrt{T}}$, where x_0 is the initial price. The simulated price process is then divided into segments, for example, 100 samples a segment, where the 100 samples represent the intraday price movements of a day. It enables us to further exploit the profitability of the trading strategy taking into account the intraday price

¹⁸The Momentum Indicator measures the net difference in price between two points on a chart. The Momentum line is the difference between the closing price and the closing price N (typically 10) periods ago. A simple moving average (typically 5 periods) of the Momentum line is shown as a second line. A buy signal is generated when the Momentum line crosses above the moving average. A sell signal is generated when the Momentum line crosses below the moving average line. (refer to Bloomberg)

paths. The maximum, minimum, first and last price of the segment represent the daily high, low, open and close price, respectively. One of the simulated price path and the corresponding segments are shown in Figure 5.19.

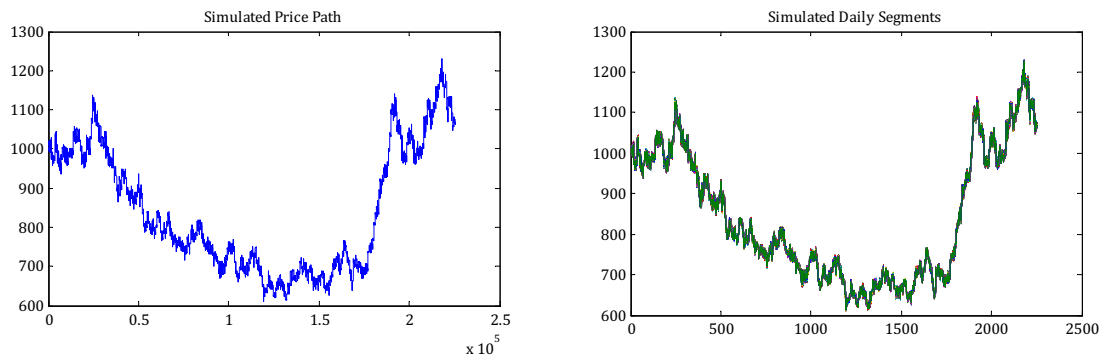


Figure 5.19: Simulated price path and corresponding daily segments

By applying the trading strategy to OHLC target prices of the simulated daily segments, we obtain the P&L shown in Table 5.12. The column High-Low Range shows the P&L calculated from the daily high-low range using the same method as in Table 5.11, while Full Path shows the P&L calculated by considering the full intraday price path, which is similar to the P&L of applying the trading strategy in practice. It shows the performance of the High-Low Range and Full Path is similar in terms of annual return. It justifies that the P&L shown in Table 5.11 can be used to represent the actual P&L of applying the trading strategy to the full price path of the S&P500.

	High-Low Range	Full Path
Long Profit	740.67	738.37
Short Profit	729.71	731.92
Total Profit	1470.38	1470.29
Total Trade	175	175
Total Net Profit	1295.38	1295.29
Initial Inv	3413	3413
Annual Return	37.95%	37.95%

Table 5.12: P&L of the simulated price path

The trading strategy is significantly better than the buy&hold strategy, which gives 4.24% annual return (total profit 43.33), and the momentum indicator strategy, which

has total profit 205.63 with 27 trades.

5.5.3.4 Limitations of the Trading Strategy

As discussed in Section 5.4, a limitation of this trading strategy is that we could suffer unexpected losses if the price moves to the unfavorable direction and triggers the stop loss level before it moves to the favorable direction as expected by the sign forecast. Therefore, the performance of the trading strategy is sensitive to the stop loss and take profit limit levels.

Theoretically, it is possible to set the stop loss and take profit limit at any level. However, in practice, the stop loss and take profit limit levels are subject to broker's limitation. Some brokers have tighter stop loss and take profit limit level requirements than others. Therefore, it is worth noticing that the actual performance of the proposed strategy in practice is subject to brokers' limitations, such as spread, stop loss and take profit limit level, margin, minimum and maximum size of bet, etc. It is necessary to choose a retail broker who would offer the tightest spread, stop loss and take profit limit levels, least margin requirements, and widest minimum and maximum size of bet range, if you are a personal investor, or institutional brokers in order to maximise the profitability and minimise the transaction costs.

5.6 Summary

This chapter focused on forecasting real world financial time series using proposed LPFs and adaptive/hybrid filters.

One of the contributions of the thesis is that the linear combination of daily open, high, low and close (OHLC) prices, which we call target price, is found to be able to provide more predictability than daily close prices alone. There are more low frequency

components than high frequency ones in the spectrum of the derived target price time series. This feature of the target price time series can be explained by the nature of financial time series, that is, the daily open price is close to the close price of the last day, thus a target price time series is effectively the close price time series passing through a lowpass filter.

An objective function was proposed to derive the target price series from historical observations of daily open, high, low and close prices, or intraday high frequency prices of an asset. In addition, the procedures of applying a LPF to forecast a target price time series and a trading strategy based on forecasting target price time series were also proposed in this chapter.

The empirical results were given which were obtained by applying proposed LPFs and adaptive/hybrid filters to forecast target price time series of the S&P500 following the proposed forecast procedures. We could achieve better forecast performance (72.62% accuracy in terms of success ratio of direction changes) by applying the proposed filters to target price time series of the S&P500 than the naive forecast. Moreover, by applying the trading strategy to the one-step ahead forecast of target price series of the S&P500, we could achieve more than 42.13% annual return (outperform 22 of 23 trade strategies) and 2.37 annualised Sharpe Ratio.

Chapter 6

Conclusion and Future Work

6.1 Summary and Conclusion

The purpose of this research is to investigate and forecast financial time series considering their dynamic and complex nature. The predictive performance of asset prices time series themselves is exploited by applying digital signal processing methods to their historical observations. The novelty of the research lies in the design of predictive filters by maximising the spectrum flatness of forecast errors and applying the filters to forecast target price time series which are linear combinations of daily open, high, low and close prices of assets. Two fundamental questions raised in Section 1.2 have been answered in this thesis in terms of forecasting financial time series with specific methods and benefiting from the forecasts of the financial time series.

We started from analysing a band-limited time series, and concluded that the sufficient and necessary conditions that a time series can be predicted with zero errors by linear filters given the assumption that there are no structural breaks or regime switching in the time series is that the time series is band-limited. We also concluded that a band-limited time series can be predicted with zero errors by a predictive filter that

has a constant magnitude response and a constant group delay over the bandwidth of the signal.

However, the real world time series are not band-limited thus cannot be forecasted without errors. To analyse and forecast the non-band-limited time series, we introduced the statistical tests of spectrum flatness, which evaluate the departure of the spectrum from a constant value, as measures of the predictability of a time series. The properties of a time series were then investigated in the frequency domain through its spectrum flatness.

The predictive filter theory was proposed and three linear predictive filters were designed for forecasting general time series by maximising the spectrum flatness of forecast errors. Among them, the LPF by maximising the ESF estimated by the periodogram method is useful to forecast a time series that has no structural breaks, while the LPF by maximising the SDWESF and EWSDWESF are powerful to forecast a time series that has structural breaks in the in-sample data set but not the out-of-sample data set. The latter adjusts to structural breaks more quickly than the former, and greatly outperforms the ARMA model estimated by least square errors when dealing with time series that have structural breaks in the in-sample data set. For a time series that has structural breaks in the out-of-sample data set, the adaptive filter and hybrid filter were proposed which utilise a series of successive LPFs. The hybrid filter is more suitable to forecast time series that have few but large magnitude structural breaks, while the adaptive filter is more suitable for time series that have large amount of small structural breaks.

In addition, a theorem that estimates the forecast performance of a predictive filter over the naive forecast in terms of upper bounds of forecast errors was presented. The magnitude response of a predictive filter provides a unique insight into the worst case scenario of the forecast performance of applying the predictive filter to forecast a time series

The focus was then placed on forecasting real world financial time series. By applying spectrum flatness tests, we found that the spectrum of a linear combination of daily open, high, low and close prices, which we call target prices, is significantly different from that of a random walk process, as it has more low frequency components than high frequency ones. This feature of the target price time series can be explained by the nature of financial time series, that is, the daily open price is close to the close price of the day before, thus a target price time series is effectively a close price time series passing through a lowpass filter.

An objective function was proposed to derive the target price series from historical observations of daily open, high, low and close prices, or intraday high frequency prices of an asset. In addition, the procedures of applying a LPF to forecast a target price time series and a trading strategy based on forecasting target price time series were also proposed.

The empirical results were given which were obtained by applying proposed LPFs and adaptive/hybrid filters to forecast target price time series of the S&P500 following proposed forecast procedures. We have achieved better forecast performance (72.62% accuracy in terms of success ratio of direction changes) by applying proposed filters to target price time series of the S&P500 than the naive forecast. Moreover, by applying the trading strategy to the one-step ahead forecast of target price series of the S&P500, we have achieved more than 42.13% annual return (outperform 22 of 23 trade strategies) and 2.37 annualised Sharpe Ratio.

6.2 Future Work

Some research questions over the thesis that may need further work are listed as follows:

1. **Optimisation Algorithm:** the objective functions that define the spectrum flatness for either LPF estimation or target price estimation are not convex, thus some optimisation algorithm may end up with a local minimum. Therefore, the convexity of these objective functions needs further investigation and a proper global searching optimisation algorithm should be adopted for better LPF estimation and target price estimation.
2. **Segment Size and Decay Factor:** the implementation of LPF using the SDWESF method needs an input of segment size, while that using the EWSDWESF method requires inputs of both segment size and decay factor. The current solution to decide the segment size and decay factor is to use those selected from the in-sample testing set which give the minimum MSE or maximum success ratio of direction change forecast. Methods to find the best segment size and decay factor need further investigation in order to obtain the best LPF.
3. **Moving Window:** the implementation of the adaptive filter and hybrid filter needs an input of size of the moving window, which is currently selected from the in-sample testing set which give the minimum MSE or maximum success ratio of direction change forecast. Methods to find the best size of moving window need further investigation in order to obtain the best adaptive filter or hybrid filter.

Besides, the following sections give a brief overview of several related future research ideas.

6.2.1 Regime Switching Model

Structural breaks in financial time series cannot be modeled implicitly using simple linear time series models. To overcome this problem, we proposed LPFs estimated by maximising the EWSDWESF and adaptive/hybrid filters to accommodate the non-linear features of financial time series. Also, the Markov switching model has been

widely applied to solve the regime switching problem in financial time series. For example, regime shifts in stock market returns was studied in [125] using the *Markov Switching Autoregressive* (MS-AR) model. Bull and bear US stock markets used a Markov switching model were identified in [84]. However, the classical approach treats the number of regimes as given which is not suitable for out-of-sample forecasting, as it does not account for new regimes occurring after the end of the estimate sample. A Bayesian estimation and prediction procedure was proposed in [98] that allows for the possibility of new breaks occurring over the forecast horizon by means of a hierarchical hidden Markov chain model. The Markov switching regime model is a useful approach to model the non-linearities in time series assuming different behavior (structural break) in one regime to another.

The difference between our LPF and adaptive/hybrid filter method and Markov switching model is that the former considers structural breaks in terms of spectrum changes of the original time series, while the latter considers structural breaks in the mean and volatility of the original series.

One potential research topic is to combine the Markov switching regime model with our LPF and adaptive/hybrid filter approach to identify major structural breaks in the mean and volatility of the original series, and structural breaks in the spectrum of each regime, or the other way around.

6.2.2 Higher-Order Spectra

The proposed LPFs estimated by maximising the error spectrum flatness are based on the distribution of power among its frequency components. However, the information contained in the power spectrum is essentially that which is presented in the autocorrelation function. The phase relations between frequency components are suppressed. Therefore, a time series $\{x_n\}$ is completely characterized by its spectrum only

if it is a Gaussian process [93]. For a non-Gaussian process, the higher order spectra, a.k.a., polyspectra, defined in terms of higher order statistics (“cumulants”) of a signal carry more information of non-linearities in the source of the time series than power spectrum. According to [93], the higher-order spectra¹ can be used to reconstruct non-minimum phase signals, extract information due to deviations from Gaussianity, and detect and characterise non-linear properties in signals.

The higher-order spectra provide a powerful tool to analyse a non-linear and non-Gaussian time series. A promising future work is to incorporate the higher-order spectra into the optimisation problem of the predictive filters in order to give better forecast performance to deal with the non-linearities of a financial time series and possible non-Gaussian forecast errors.

6.2.3 Exogenous Variables

In this research, we apply the digital signal processing method to historical observations of a financial time series itself rather than considering other exogenous variables. However, studies give evidence that stock prices/returns are predictable with the help of some exogenous financial variables such as interest rate, dividend-price ratio, dividend yield, retail sales, gold price, industrial production index, etc., for example, [24], [26], [46], [45], and more recently [74] and [124]. In addition, historical values of daily high-low range and mean of daily high-low are also identified to be helpful to forecast an interval financial time series (Section 2.3.5).

It is then necessary to take into consideration these exogenous variables such as those fundamental variables or interval time series variables when we forecast a financial time series in the future work. It is promising to further improve the performance of the predictive filter methods with the help of these explanatory variables. Furthermore,

¹ The bispectrum is the Fourier transform of the third-order statistics, and the trispectrum is the Fourier transform of the fourth-order statistics of a stationary signal.

multiple variate models such as vector autoregression (VAR) model or ANNs instead of the uni-variate autoregression can be applied to take advantage of the explanatory power of these exogenous variables.

6.2.4 Forecast Combination

Forecast combination is another promising future work. Forecast combination is considered a simple and effective way to improve the forecasting performance over that given by individual models. Timmermann [123] shows that simple equal-weighted average combination approach often give better performance than more sophisticated approaches. Forecast combinations have been used successfully in empirical work in finance and economics, for example, [37], [117], [107], [3], and [17], etc.

As a future work, the LPF and adaptive/hybrid filter could be pooled with other linear or non-linear models and assigned proper weights to further boost the forecast performance. A proper model selection mechanism should be developed to decide which model to be pooled, and also weights of these models should be carefully calculated and tested. Key questions to be answered in terms of weights are, for example, fixed weights or time varying weights, correlation of performance of models, etc.

6.2.5 Financial Time Series Combination

The target price time series considered in this thesis is the linear combination of daily high, low and close prices of a individual financial time series. Although close price cannot be better predicted using any approaches than the naive one. The predictability of combination of financial time series has not been exploited properly to our knowledge. The idea to create the target price time series from a single financial time series can be extended to multiple financial time series by examining the target price of the linear

combination of multiple financial time series. For example, we could investigate the predictability of the target price of the linear combination of 3 major US stock indices, i.e., DJIA, S&P500 and NASDAQ. The linear combination could eliminate the jump effects of events to the stock indices thus reducing structural breaks of the target price of the combination. Other categories of financial time series could be examined to further exploit the predictability over a single series. For example, we could examine a stock on the same company but trading over different exchanges, e.g., BHP traded on the UK and Australian stock exchanges.

The combination of financial time series could provide a powerful tool to make profits by trading in different markets or across different asset classes in the form of pair trading, basket of stocks trading and so on, and worth further research in the area of high frequency algorithm trading.

Bibliography

- [1] ABHYANKAR, A., COPELAND, L., AND WONG, W. Uncovering nonlinear structure in real-time stock-market indexes: the S&P 500, the DAX, the Nikkei 225, and the FTSE-100. *Journal of Business & Economic Statistics* (1997), 1–14.
- [2] ABU-MOSTAFA, Y., AND ATIYA, A. Introduction to financial forecasting. *Applied Intelligence* 6, 3 (1996), 205–213.
- [3] AIOLFI, M., CAPISTRÁN CARMONA, C., AND TIMMERMANN, A. Forecast combinations. *CREATES Research Paper No. 2010-21* (2010).
- [4] AKAIKE, H. Fitting autoregressive models for prediction. *Annals of the institute of statistical mathematics* 21, 1 (1969), 243–247.
- [5] AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control* 19, 6 (1974), 716–723.
- [6] ANG, A., AND BEKAERT, G. Stock return predictability: Is it there? *Review of Financial Studies* 20, 3 (2007), 651–707.
- [7] ANTONI, J., AND SCHOUKENS, J. A comprehensive study of the bias and variance of frequency-response-function measurements: Optimal window selection and overlapping strategies. *Automatica* 43, 10 (2007), 1723–1736.
- [8] ARINO, M. A. Forecasting time series via discrete wavelet transform. Working Paper, 1996.

- [9] ARMANO, G., MARCHESI, M., AND MURRU, A. A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences* 170, 1 (2005), 3–33.
- [10] ARMSTRONG, J. *Principles of forecasting: a handbook for researchers and practitioners*, vol. 30. Springer, 2001.
- [11] ARROYO, J., ESPÍNOLA, R., AND MATÉ, C. Different approaches to forecast interval time series: a comparison in finance. *Computational Economics* 37, 2 (2011), 169–191.
- [12] ARROYO, J., AND MATÉ, C. Introducing interval time series: Accuracy measures. *COMPSTAT 2006, proceedings in computational statistics* (2006), 1139–1146.
- [13] ATSALAKIS, G., AND VALAVANIS, K. Surveying stock market forecasting techniques—part ii: Soft computing methods. *Expert Systems with Applications* 36, 3 (2009), 5932–5941.
- [14] AUSSEM, A., COMBELL, J., AND MURTAGH, F. Wavelet-based feature extraction and decomposition strategies for financial forecasting. *Journal of Computational Intelligence in Finance* 6, 2 (1998), 5–12.
- [15] BALESTRASSI, P., POPOVA, E., PAIVA, A., AND MARANGON LIMA, J. Design of experiments on neural network’s training for nonlinear time series forecasting. *Neurocomputing* 72, 4 (2009), 1160–1178.
- [16] BARTLETT, M. Smoothing periodograms from time series with continuous spectra. *Nature* 161, 4096 (1948), 686–687.
- [17] BILLIO, M., CASARIN, R., RAVAZZOLO, F., AND VAN DIJK, H. Bayesian combinations of stock price predictions with an application to the amsterdam exchange index. *Tinbergen Institute Discussion Paper No. 2011-082/4* (2011).

- [18] BODIE, Z., KANE, A., AND MARCUS, A. *Investments*. McGraw-Hill, 2008.
- [19] BOX, G., AND DRAPER, N. *Empirical model-building and response surfaces*. John Wiley & Sons, 1987.
- [20] BOX, G., AND JENKINS, G. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970.
- [21] BOX, G., AND PIERCE, D. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 332 (1970), 1509–1526.
- [22] BROCKWELL, P., AND DAVIS, R. *Introduction to time series and forecasting*. Springer, 2002.
- [23] CAI, Z., FAN, J., AND YAO, Q. Functional coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95 (2000), 941–956.
- [24] CAMPBELL, J. Stock returns and the term structure. *Journal of financial economics* 18, 2 (1987), 373–399.
- [25] CAMPBELL, J. A variance decomposition for stock returns. *The Economic Journal* 101, 405 (1991), 157–179.
- [26] CAMPBELL, J., AND SHILLER, R. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* (1988), 195–228.
- [27] CAPORIN, M., RANALDO, A., AND DE MAGISTRIS, P. S. On the predictability of stock prices: A case for high and low prices.
- [28] CARVALHAL, A., AND DE MELO MENDES, B. Evaluating the forecast accuracy of emerging market stock returns. *Emerging Markets Finance and Trade* 44, 1 (2008), 21–40.

- [29] CHEN, K., AND WANG, C. A hybrid sarima and support vector machines in forecasting the production values of the machinery industry in taiwan. *Expert Systems with Applications* 32, 1 (2007), 254–264.
- [30] CHEN, R., AND LUI, L. Functional coefficient autoregressive models. *Journal of American Statistical Association* 421 (1993), 298–308.
- [31] CHEN, R., AND LUI, L. Functional coefficient autoregressive models: estimation and tests of hypotheses. *Journal of Time Series Analysis* 222 (2001), 151–173.
- [32] CHEN, Y., STATHAKI, T., AND CONSTANTINIDES, A. Nonlinear adaptive prediction of nonstationary signals with application to speech coding. In *Multimedia Signal Processing, 1997., IEEE First Workshop on* (1997), IEEE, pp. 125–130.
- [33] CHENG, H., AND GLASCOCK, J. Dynamic linkages between the greater china economic area stock markets—mainland china, hong kong, and taiwan. *Review of Quantitative Finance and Accounting* 24, 4 (2005), 343–357.
- [34] CHEUNG, Y. An empirical model of daily highs and lows. *International Journal of Finance & Economics* 12, 1 (2007), 1–20.
- [35] CHOW, K., AND DENNING, K. A simple multiple variance ratio test. *Journal of Econometrics* 58, 3 (August 1993), 385–401.
- [36] CHRISTOFFERSEN, P., AND DIEBOLD, F. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52, 8 (2006), 1273–1287.
- [37] CLEMEN, R. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5, 4 (1989), 559–583.
- [38] CONT, R. Empirical properties of asset returns: stylized facts and statistical issues.

- [39] COSTANTINI, M., AND KUNST, R. On the usefulness of the diebold-mariano test in the selection of prediction models: some monte carlo evidence. Tech. rep., Working paper, University of Vienna, 2011.
- [40] DANGL, T., HALLING, M., AND RANDL, O. Equity return prediction: Are coefficients time-varying? 2006.
- [41] DE GOOIJER, J., AND HYNDMAN, R. 25 years of time series forecasting. *International journal of forecasting* 22, 3 (2006), 443–473.
- [42] DIEBOLD, F., AND MARIANO, R. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 3 (1995).
- [43] DROUCHE, K. A Test for Spectrum Flatness. *Journal of Time Series Analysis* 28, 6 (2007), 793–806.
- [44] FAMA, E. Efficient capital markets: A review of theory and empirical work. *Journal of finance* (1970), 383–417.
- [45] FAMA, E., AND FRENCH, K. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 1 (1989), 23–49.
- [46] FAMA, E. F., AND FRENCH, K. R. Dividend yields and expected stock returns. *Journal of Financial Economics* 22 (October 1988), 3–25.
- [47] FAN, J., AND ZHANG, W. Statistical methods with varying coefficient models. *Stat Interface* 1, 1 (2008), 179–195.
- [48] FRENCH, K., SCHWERT, G., AND STAMBAUGH, R. Expected stock returns and volatility. *Journal of Financial Econometrics* 19 (1987), 3–30.
- [49] GENÇAY, R., SELÇUK, F., AND WHITCHER, B. *An introduction to wavelets and other filtering methods in finance and economics*. Academic Pr, 2002.

- [50] GRANGER, C., AND PESARAN, M. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19, 7 (2000), 537–560.
- [51] GUIDOLIN, M., HYDE, S., McMILLAN, D., AND ONO, S. Non-linear predictability in stock and bond returns: When and where is it exploitable? *International Journal of Forecasting* 25, 2 (2009), 373–399.
- [52] HAMILTON, J. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society* (1989), 357–384.
- [53] HAN, A., HONG, Y., LAI, K., AND WANG, S. Interval time series analysis with an application to the sterling-dollar exchange rate. *Journal of Systems Science and Complexity* 21, 4 (2008), 558–573.
- [54] HARVILL, J. L., AND RAY, B. K. A note on multi-step forecasting with functional coefficient autoregressive models. *International Journal of Forecasting* 20 (2005), 435–446.
- [55] HAYKIN, S. *Adaptive Filter Theory*. Prentice Hall, 2002.
- [56] HE, L., AND HU, C. Impacts of interval computing on stock market variability forecasting. *Computational Economics* 33, 3 (2009), 263–276.
- [57] HELLSTRÖM, T., AND HOLMSTRÖM, K. Predictable patterns in stock returns. Tech. rep., Malardalen University, 1998.
- [58] HONG, Y., AND CHUNG, J. Are the directions of stock price changes predictable? statistical theory and evidence. Tech. rep., Cornell University, 2003.
- [59] HOQUE, H. A., KIM, J. H., AND PYUN, C. S. A comparison of variance ratio tests of random walk: A case of asian emerging stock markets. *International Review of Economics & Finance* 16, 4 (2007), 488–502.

- [60] HSIEH, T., HSIAO, H., AND YEH, W. Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing* 11, 2 (2011), 2510–2525.
- [61] HUANG, W., NAKAMORI, Y., AND WANG, S. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research* 32, 10 (2005), 2513–2522.
- [62] HULL, J. *Options, futures and other derivatives*. Pearson Prentice Hall, 2009.
- [63] HYNDMAN, R., AND KOEHLER, A. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 4 (2006), 679–688.
- [64] JENSEN, M., FIELD, S., AND PARK, T. Some anomalous evidence regarding market efficiency. *Journal of Financial Economics* 6, 2/3 (1978), 95–101.
- [65] KAY, S. *Modern spectral estimation: theory and application*. No. v. 1 in Prentice-Hall signal processing series. Prentice Hall, 1988.
- [66] KENDALL, M. G. The analysis of economic time series, part I: prices. *Journal of the Royal Statistical Society* 96 (1953), 11–25.
- [67] KHASHEI, M., AND BIJARI, M. An artificial neural network (p,d,q) model for timeseries forecasting. *Expert Systems with Applications* 37, 1 (2010), 479–489.
- [68] KIM, J., SHAMSUDDIN, A., AND LIM, K. Stock return predictability and the adaptive markets hypothesis: Evidence from century-long us data. *Journal of Empirical Finance* (2011).
- [69] KISSELL, R., GLANTZ, M., MALAMUT, R., AND CHRISS, N. A. *Optimal trading strategies: quantitative approaches for managing market impact and trading risk*. Amacom, 2003.

- [70] KORALE, A., AND CONSTANTINIDES, A. The endomorphic model: An approach to adaptive prediction. In *EUPSICO 2000: European signal processing conference* (2000), pp. 2473–2476.
- [71] LANGI, A., PITARA, S., ET AL. Stock prices trends analysis using wavelet transform. In *Cloud Computing and Social Networking (ICCCSN), 2012 International Conference on* (2012), IEEE, pp. 1–4.
- [72] LEBARON, B., AND WEIGEND, A. A bootstrap evaluation of the effect of data splitting on financial time series. *Neural Networks, IEEE Transactions on* 9, 1 (1998), 213–220.
- [73] LEITCH, G., AND TANNER, J. Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review* (1991), 580–590.
- [74] LETTAU, M., AND VAN NIEUWERBURGH, S. Reconciling the return predictability evidence. *Review of Financial Studies* 21, 4 (2008), 1607–1652.
- [75] LEUNG, M., DAOUK, H., AND CHEN, A. Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* 16, 2 (2000), 173–190.
- [76] LI, J., SHI, Z., AND LI, X. Genetic programming with wavelet-based indicators for financial forecasting. *Transactions of the Institute of Measurement & Control* 28, 3 (2006), 285.
- [77] LIM, K., AND BROOKS, R. The evolution of stock market efficiency over time: a survey of the empirical literature. *Journal of Economic Surveys* 25, 1 (2011), 69–108.
- [78] LJUNG, G., AND BOX, G. On a measure of lack of fit in time series models. *Biometrika* 65, 2 (1978), 297–303.

- [79] LO, A., AND MACKINLAY, C. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1 (1988), 41–66.
- [80] LO, A., AND MACKINLAY, C. The size and power of the variance ratio test in finite samples. *Journal of Econometrics* 40 (1989), 203–238.
- [81] LO, A., AND MACKINLAY, C. *A non-random walk down Wall Street*. Princeton, NJ, 1999.
- [82] LUDLOW, J., AND ENDERS, W. Estimating non-linear ARMA models using Fourier coefficients. *International Journal of Forecasting* 16, 3 (2000), 333–347.
- [83] MAHEU, J., AND GORDON, S. Learning, forecasting and structural breaks. *Journal of Applied Econometrics* 23, 5 (2008), 553–583.
- [84] MAHEU, J., AND MCCURDY, T. Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics* (2000), 100–112.
- [85] MAIA, A., DE CARVALHO, F., AND LUDERMIR, T. Forecasting models for interval-valued time series. *Neurocomputing* 71, 16 (2008), 3344–3352.
- [86] MAKRIDAKIS, S., AND WHEELWRIGHT, S. Adaptive filtering: an integrated autoregressive/moving average filter for time series forecasting. *Operational Research Quarterly* (1977), 425–437.
- [87] MAKRIDAKIS, S., WHEELWRIGHT, S., AND HYNDMAN, R. *Forecasting: methods and applications*. Wiley New York, 1997.
- [88] MCELROY, T., AND HOLAN, S. Spectral domain diagnostics for testing model proximity and disparity in time series data. *Statistical Methodology* 6, 1 (2009), 1–20.
- [89] MCNELIS, P. *Neural networks in finance: gaining predictive edge in the market*. Academic Press, 2005.

- [90] MITCHELL, M., AND CHIAO, R. Causality and negative group delays in a simple bandpass amplifier. *American Journal of Physics* 66 (Jan. 1998), 14–19.
- [91] MOBAREK, A., AND KEASEY, K. Weak-form market efficiency of an emerging market: Evidence from dhaka stock market of bangladesh. In *ENBS Conference held on Oslo* (2000), pp. 1–30.
- [92] MURTAGH, F., STARCK, J., AND RENAUD, O. On neuro-wavelet modeling. *Decision Support Systems* 37, 4 (2004), 475–484.
- [93] NIKIAS, C., AND MENDEL, J. Signal processing with higher-order spectra. *Signal Processing Magazine, IEEE* 10, 3 (1993), 10–37.
- [94] NYBERG, H. Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting* 27, 2 (2011), 561–578.
- [95] OPPENHEIM, A., AND SCHAFER, R. *Discrete-time signal processing*. Pearson Education, 2010.
- [96] PAI, P., AND LIN, C. A hybrid arima and support vector machines model in stock price forecasting. *Omega* 33, 6 (2005), 497–505.
- [97] PANDA, C., AND NARASIMHAN, V. Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling* 29, 2 (2007), 227–236.
- [98] PESARAN, M., PETTENUZZO, D., AND TIMMERMANN, A. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* 73, 4 (2006), 1057–1084.
- [99] PESARAN, M., AND TIMMERMANN, A. A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics* 10, 4 (1992), 461–465.

- [100] PESARAN, M., AND TIMMERMAN, A. Predictability of stock returns: Robustness and economic significance. *Journal of Finance* (1995), 1201–1228.
- [101] PESARAN, M., AND TIMMERMAN, A. How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting* 20, 3 (2004), 411–425.
- [102] PONS, J. The rationality of price forecasts: a directional analysis. *Applied Financial Economics* 11, 3 (2001), 287–290.
- [103] PRING, M. *Technical Analysis Explained: The Successful Investor's Guide To Spotting Investment Trends And Turning Points*. McGraw-Hill, 2002.
- [104] QI, M. Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business & Economic Statistics* (1999), 419–429.
- [105] QI, M., AND ZHANG, G. Trend time-series modeling and forecasting with neural networks. *Neural Networks, IEEE Transactions on* 19, 5 (2008), 808–816.
- [106] RAMSEY, J. B. The contribution of wavelets to the analysis of economic and financial data. *Phil Trans R Soc Lond* 357 (1999), 2593–2606.
- [107] RAVAZZOLO, F., PAAP, R., VAN DIJK, D., AND FRANSES, P. *Bayesian model averaging in the presence of structural breaks*. Econometric Institute, Erasmus School of Economics, Erasmus University, 2006.
- [108] RENAUD, O., STARCK, J., AND MURTAGH, F. Prediction based on a multiscale decomposition. *International Journal of Wavelets Multiresolution and Information Processing* 1 (2003), 217–232.
- [109] RODRIGUES, P., AND SALISH, N. Modeling and forecasting interval time series with threshold models: an application to s&p500 index returns. *Working Papers* (2011).

- [110] SADAM, A., ISMAIL, M., ABDUL, K., AND ARIFFIN, S. Selecting wavelet transforms model in forecasting financial time series data based on arima model. *Applied Mathematical Sciences*, 5, 7 (2011), 315–326.
- [111] SCHALLER, H., AND VAN NORDEN, S. Regime switching in stock market returns. *Applied Financial Economics* 7, 2 (1997), 177–191.
- [112] SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics* (1978), 461–464.
- [113] SIMONS, D., AND LARYEA, S. Testing the efficiency of selected african stock markets. *Finance India* 20, 2 (2006), 553–571.
- [114] STATHAKI, T., AND FOTINOPOULOS, I. Equiripple minimum phase fir filter design from linear phase systems using root moments. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on* 48, 6 (2001), 580–587.
- [115] STATHAKI, T., FOTINOPOULOS, I., AND CONSTANTINIDES, A. Root moments: a nonlinear signal transformation for minimum fir filter design. In *Proc. of the NSIP* (1999), vol. 99, pp. 20–23.
- [116] STOCK, J., AND WATSON, M. Variable trends in economic time series. *The Journal of Economic Perspectives* 2, 3 (1988), 147–174.
- [117] STOCK, J., AND WATSON, M. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23, 6 (2004), 405–430.
- [118] STOICA, P., AND MOSES, R. *Spectral analysis of signals*. Pearson/Prentice Hall, 2005.
- [119] TASHMAN, L. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 16, 4 (2000), 437–450.

- [120] TAYLOR, S. *Modelling Financial Time Series*. World Scientific, 2008.
- [121] THEIL, H. *Applied economic forecasting*, vol. 4. North-Holland Pub. Co., 1966.
- [122] THOMAKOS, D., AND WANG, T. Optimal probabilistic and directional predictions of financial returns. *Journal of Empirical Finance* 17, 1 (2010), 102–119.
- [123] TIMMERMAN, A. Forecast combinations. *Handbook of Economic Forecasting* (2006).
- [124] TIMMERMAN, A. Elusive return predictability. *International Journal of Forecasting* 24, 1 (2008), 1–18.
- [125] TURNER, C., STARTZ, R., AND NELSON, C. A markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics* 25, 1 (1989), 3–22.
- [126] VAIDYANATHAN, P. P. *The Theory of Linear Prediction*. Morgan and Claypool Publishers, 2008.
- [127] VAINIO, O. Frequency-selective fir predictors. *Signal Processing Letters, IEEE* 8, 2 (2001), 33–35.
- [128] VASEGHI, S. *Advanced digital signal processing and noise reduction*. Wiley, 2005.
- [129] WANG, J., WANG, J., ZHANG, Z., AND GUO, S. Stock index forecasting based on a hybrid model. *Omega* (2011).
- [130] WELCH, P. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *Audio and Electroacoustics, IEEE Transactions on* 15, 2 (1967), 70–73.
- [131] WHITE, P. Group delay explanations and applications. Tech. rep., Applied Radio Labs, 1999.

- [132] WIENER, N. *Extrapolation, interpolation, and smoothing of stationary time series*. Wiley New York, 1950.
- [133] WRIGHT, J. H. Alternative variance-ratio tests using ranks and signs. *Journal of Business & Economic Statistics* 18, 1 (January 2000), 1–9.
- [134] ZHANG, G. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing* 50 (2003), 159–175.
- [135] ZHANG, G., PATUWO, B. E., AND HU, M. Y. Forecasting with artificial networks: the state of the art. *International Journal of Forecasting* 14 (1998), 35–62.

Appendix A

Bounds of Forecast Errors

A.1 One-Step Ahead Forecast Bound

For a filter with the transfer function $F(z)$ that gives a one-step ahead forecast $\hat{X}(z) = F(z)X(z)$, where $\hat{X}(z)$ is the output one-step ahead forecast of $X(z)$, the upper bounds of the one-step ahead forecast value can be represented as

$$\begin{aligned} |\hat{x}_n| &= \frac{1}{N} \left| \sum_{k=0}^{N-1} X_k F_k e^{\frac{2\pi i}{N} kn} \right| \\ &\leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k e^{\frac{2\pi i}{N} kn}| \\ &\leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| \left| e^{\frac{2\pi i}{N} kn} \right| \\ &\leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| |F_k| \end{aligned} \tag{A.1}$$

where F_k is the frequency response of the transfer function $F(z)$ that we used for prediction.

A.2 In-Sample One-Step Ahead Forecast Bound

The one-step ahead in-sample prediction of x_{n+1} is

$$\hat{x}_{n+1} = \frac{1}{N} \sum_{k=0}^{N-1} X_k F_k e^{\frac{2\pi i}{N} kn}$$

The actual value of x_{n+1} , $n+1 < N-1$ can be represented as

$$x_{n+1}^{\text{IS}} = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} k(n+1)}$$

Then their difference is

$$\begin{aligned} \hat{x}_n - x_{n+1}^{\text{IS}} &= \frac{1}{N} \sum_{k=0}^{N-1} X_k F_k e^{\frac{2\pi i}{N} kn} - \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} k(n+1)} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} \left(F_k - e^{\frac{2\pi i}{N} k} \right) \end{aligned}$$

The upper bounds of their difference are then

$$\begin{aligned} \left| \hat{x}_n - x_{n+1}^{\text{IS}} \right| &\leq \left| \hat{x}_n - x_{n+1}^{\text{IS}} \right| \\ &= \frac{1}{N} \left| \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} \left(F_k - e^{\frac{2\pi i}{N} k} \right) \right| \\ &\leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| \left| F_k - e^{\frac{2\pi i}{N} k} \right| \\ &\leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + 1) \end{aligned}$$

So the errors of one-step ahead in-sample forecast

$$|e_{n+1}| = \left| \hat{x}_n - x_{n+1}^{\text{IS}} \right| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + 1) \quad (\text{A.2})$$

The upper bounds of the MAE and MSE can be represented respectively as

$$\begin{aligned}
\text{MAE}^{\text{IS}} &= \frac{1}{M} \sum_{n=0}^{M-1} |e_{n+1}| \\
&\leq \frac{1}{MN} \sum_{n=0}^{M-1} \sum_{k=0}^{N-1} |X_k| (|F_k| + 1) \\
&= \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + 1)
\end{aligned}$$

$$\begin{aligned}
\text{MSE}^{\text{IS}} &= \frac{1}{M} \sum_{n=0}^{M-1} (e_{n+1})^2 \\
&= \frac{1}{M} \sum_{n=0}^{M-1} \left(\frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} \left(F_k - e^{\frac{2\pi i}{N} k} \right) \right)^2 \\
&= \frac{1}{M} \sum_{n=0}^{M-1} \left(\frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} \left(F_k - e^{\frac{2\pi i}{N} k} \right) \right) \left(\frac{1}{N} \sum_{j=0}^{N-1} \overline{X_j e^{\frac{2\pi i}{N} jn} \left(F_j - e^{\frac{2\pi i}{N} j} \right)} \right) \\
&= \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} \left(X_k \left(F_k - e^{\frac{2\pi i}{N} k} \right) \right) \overline{\left(X_j \left(F_j - e^{\frac{2\pi i}{N} j} \right) \right)} \frac{1}{M} \sum_{n=0}^{M-1} e^{\frac{2\pi i}{N} (k-j)n} \\
&= \frac{1}{N^2} \sum_{k=0}^{N-1} \left| X_k \left(F_k - e^{\frac{2\pi i}{N} k} \right) \right|^2 \\
&\leq \frac{1}{N^2} \sum_{k=0}^{N-1} |X_k|^2 (|F_k|^2 + 1)
\end{aligned}$$

A.3 Out-of-Sample One-Step Ahead Forecast Bound

Let α_n be the ratio of two successive observations, i.e. $\alpha_n = \frac{x_{n+1}}{x_n}$, then the actual x_{n+1} can be represented using x_n as

$$x_{n+1} = \alpha_n x_n$$

The one-step ahead forecast of x_{n+1} is

$$\hat{x}_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k F_k e^{\frac{2\pi i}{N} kn}$$

The actual value of x_{n+1} can be represented as

$$x_{n+1}^{\text{OS}} = \frac{\alpha_n}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn}$$

Then their difference is

$$\begin{aligned} \hat{x}_n - x_{n+1}^{\text{OS}} &= \frac{1}{N} \sum_{k=0}^{N-1} X_k F_k e^{\frac{2\pi i}{N} kn} - \frac{\alpha_n}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} (F_k - \alpha_n) \end{aligned}$$

The bound of their difference is then

$$\left| \hat{x}_n - x_{n+1}^{\text{IS}} \right| \leq \left| \hat{x}_n - x_{n+1}^{\text{OS}} \right| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| |F_k - \alpha_n| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + |\alpha_n|)$$

So the errors of one-step ahead out-of-sample forecast

$$|e_{n+1}| = \left| \hat{x}_n - x_{n+1}^{\text{OS}} \right| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + |\alpha_n|) \quad (\text{A.3})$$

The upper bounds of the MAE are then

$$\begin{aligned} \text{MAE}^{\text{OS}} &= \frac{1}{M} \sum_{n=0}^{M-1} |e_{n+1}| \\ &\leq \frac{1}{MN} \sum_{n=0}^{M-1} \sum_{k=0}^{N-1} |X_k| (|F_k| + |\alpha_n|) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} |X_k| |F_k| + \frac{1}{MN} \sum_{n=0}^{M-1} \sum_{k=0}^{N-1} |X_k| |\alpha_n| \end{aligned} \quad (\text{A.4})$$

Let $A = \frac{1}{M} \sum_{n=0}^{M-1} |\alpha_n|$, then Equation A.4 can be further represented as

$$\text{MAE}^{\text{OS}} \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| (|F_k| + A)$$

The upper bounds of the MSE

$$\begin{aligned}
\text{MSE}^{\text{OS}} &= \frac{1}{M} \sum_{n=0}^{M-1} (e_{n+1})^2 \\
&= \frac{1}{M} \sum_{n=0}^{M-1} \left| \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} (F_k - \alpha_n) \right|^2 \\
&\leq \frac{1}{MN^2} \sum_{n=0}^{M-1} \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} |X_k|^2 (|F_j|^2 + |\alpha_n|^2) \tag{A.5}
\end{aligned}$$

Let $B = \frac{1}{M} \sum_{n=0}^{M-1} \alpha_n^2$, then Equation A.5 can be represented as

$$\text{MSE}^{\text{OS}} \leq \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} |X_k|^2 (|F_j|^2 + B)$$

A.4 Proof of Theorem

Proof. The upper bounds for in-sample and out-of-sample one-step ahead forecast with predictive filter $F(\omega)$ are $\frac{1}{M} \sum_{k=0}^{M-1} |X_k| (|F_k| + 1)$ and $\frac{1}{M} \sum_{k=0}^{M-1} |X_k| (|F_k| + |\alpha_n|)$, respectively, while the upper bounds for in-sample and out-of-sample one-step ahead naive forecast are $\frac{2}{M} \sum_{k=0}^{M-1} |X_k|$, $\frac{1}{M} \sum_{k=0}^{M-1} |X_k| (1 + |\alpha_n|)$, where M is the length of frequency bands where the input signal X_k has energy, i.e., $|X_k| > 0$. For a baseband band-limited signal, M is the bandwidth. For a signal that is not band-limited, M is the length of the full frequency bands,

If $\frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| |F_k| < \frac{1}{N} \sum_{k=0}^{N-1} |X_k|$, the bound $\frac{1}{M} \sum_{k=0}^{M-1} |X_k| (|F_k| + 1) < \frac{2}{M} \sum_{k=0}^{M-1} |X_k|$ and $\frac{1}{M} \sum_{k=0}^{M-1} |X_k| (|F_k| + |\alpha_n|) < \frac{1}{M} \sum_{k=0}^{M-1} |X_k| (1 + |\alpha_n|)$ holds.

Conversely, if $\frac{1}{M} \sum_{k=0}^{M-1} |X_k| (|F_k| + 1) < \frac{2}{M} \sum_{k=0}^{M-1} |X_k|$ or $\frac{1}{M} \sum_{k=0}^{M-1} |X_k| (|F_k| + |\alpha_n|) < \frac{1}{M} \sum_{k=0}^{M-1} |X_k| (1 + |\alpha_n|)$, we can derive that $\frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| \leq \frac{1}{N} \sum_{k=0}^{N-1} |X_k| |F_k| < \frac{1}{N} \sum_{k=0}^{N-1} |X_k|$

$\frac{1}{N} \sum_{k=0}^{N-1} |X_k|$, given $|X_k| > 0$.

Therefore, $\frac{1}{N} \sum_{k=0}^{N-1} |X_k F_k| < \frac{1}{N} \sum_{k=0}^{N-1} |X_k|$ is the necessary and sufficient condition for the upper bounds of one-step ahead forecast errors using specific methods lower than naive forecast. □

Appendix B

Trading Strategies Backtesting on the S&P500

The results of trading strategies backtesting on the S&P500 are extracted from Bloomberg Terminal (the BTST function). The specification of the backtest is shown as follows:

- Security: SPX Index
- Start Date: 2011/01/01
- End Date: 2011/12/31
- Trading Approach: Long & Short
- Period: Daily

“Avg Dur” in the table is short for “Average Duration (in days)”.

Strategy	Number of Trades			P&L	Statistics	
	Long	Short	Total	%Total	Avg Dur	Sharpe
Bollinger Bands (Boll)	4	4	8	52.99	41.38	2.58
KBand	11	12	23	36.48	15.65	1.76
MA Envelopes (MAE)	4	4	8	29.63	36	1.49
Stochastics (TAS)	6	7	13	18.09	26.46	0.97
RSI	1	1	2	15.76	176.5	0.83
William's %R (Wm)	6	7	13	12.97	27.15	0.7
Trading Envelopes (TE)	3	3	6	12.43	57.67	0.68
Cmdty Channel Index (CMCI)	7	8	15	7.01	23.6	0.43
Rex Oscillator	45	45	90	-1.01	3.99	0.06
BuyAndHold	1	0	1	-1.2	360	0.06
MACD	11	10	21	-2.37	17.14	-0.01
Rate of Change (ROC)	59	59	118	-3.28	3.05	-0.04
Exponential MA (EMAvg)	10	10	20	-5.66	14.7	-0.2
MA Oscillator (MAO)	14	13	27	-8.7	13.22	-0.3
Triangular MA (TMAvg)	9	9	18	-11.28	16.33	-0.47
Fear & Greed	16	16	32	-11.9	11.03	-0.47
Accum/Distrib Osc (ADOsc)	64	65	129	-13.27	2.78	-0.5
Weighted MA (WMAvg)	11	11	22	-13.47	13.36	-0.58
Simple MA (SMAvg)	12	12	24	-15.05	12.25	-0.68
Ichimoku (GOC)	6	6	12	-18.1	24.5	-0.79
DMI	14	14	28	-21.39	11.04	-1
Variable MA (VMAvg)	21	21	42	-21.98	7.93	-0.98
Parabolic (PTPS)	13	13	26	-34.65	13.62	-1.53

Table B.2: Trading strategies backtesting on the S&P500

Appendix C

List of Key Functions

C.1 Negative Group Delay Filter

`rootMoments` Calculate root moments of a given polynomial.

`newtonIdentities` Calculate Newton identities of roots.

`tongrpdelayfilter` Convert a non-minimum phase filter to a minimum phase one using the method proposed in [115].

C.2 Linear Predictive Filter

`autofitfir` Automatically find a FIR LPF by maximising the forecast error spectrum flatness. The order of the filter is selected by specified criteria.

autofitfiroos Automatically find a FIR LPF by maximising the forecast error spectrum flatness. The coefficients of the LPF is estimated from the in-sample training set, while the order of the LPF is selected by applying a specified criteria to the in-sample testing set.

autofitfir_adaptive Fit an adaptive filter to the out-of-sample data set and output out-of-sample forecast.

autofitfir_hybrid Fit a hybrid filter to the out-of-sample data set and output out-of-sample forecast.

autoarmasa Automatically find the order of an ARMA model using the armasel function of the Matlab toolbox ARMASA by P.M.T. Broersen.
(<http://www.dsc.tudelft.nl/Research/Software/index.html>)

fftfir Fit a FIR LPF by maximising the error spectrum flatness using the Periodogram method. If the NAG toolbox is available, function **fftfir_NAG** and **fftfir_func** are used instead of **fftfir**.

fftfir_wpsd Fit a FIR LPF by maximising the error spectrum flatness using the SDWESF or EWSDWESF method. If the NAG toolbox is available, function **fftfir_wpsd_NAG** and **fftfir_func_wpsd** are used instead of **fftfir_wpsd**.

findbestmovwinweight Find the best moving window and weight from the in-sample training set and testing set by choosing the LPF with greater SR/least BIC/least MASE.

specwstd Calculate the standardised standard deviation as weights for the SDWESF method, or exponential weighted standard deviation as weights for the EWSDWESF method.

tominph Convert a non-minimum phase filter to a minimum phase one.

flatTest Calculate spectrum flatness tests $D(\varphi)$ and $\psi(\varphi)$

evaluation Calculate the goodness-of-fit tests of forecast time series against the original one, i.e., MSE, RMSE, MAPE, MASE, SR, DA, and DA p-value.

datest Calculate Pesaran Timmermann test of directional accuracy for out of sample forecasts

C.3 Forecast Financial Time Series

getTargetPrice Calculate the linear combination of a set of prices time series to derive the target price time series.

GBM_simulation Generate a price path using the GBM model

simCandle Simulate a candle time series (with open, high, low, close prices) using price path generated by the GBM model.

momentumStrategy Calculate the P&L of the momentum indicator strategy

backtestPrice Calculate the P&L using the proposed trading strategy which is applied to the open, high, low and close prices.

backtestReal Calculate the P&L using the proposed trading strategy which is applied to the intraday high frequency prices.