ON AUTOMATIC BOUNDARY CORRECTIONS USING EMPIRICAL MODE DECOMPOSITION

ABOBAKER MOHAMED JABER

UNIVERSITI SAINS MALAYSIA

2016

ON AUTOMATIC BOUNDARY CORRECTIONS USING EMPIRICAL MODE DECOMPOSITION

by

ABOBAKER MOHAMED JABER

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

March 2016

ACKNOWLEDGEMENTS

Thanks are due to Allah for granting me the ability to complete this study. It would have been impossible to complete this research project without the scholarship granted to me by Libyan Government and I will be extremely grateful to them for the rest of my life.

I would like to state my sincere thanks to my supervisor Dr. Mohd Tahir Ismail for unrestrained support, endurance and encouragement during this research project. I have benefited from his experience, knowledge and expertise in the field of Statistics. Many thanks to him for supporting me to attend meetings and conferences (nationally and internationally). My supervisor treated me as one of his family members during our formal and informal meetings. I would like to thank all post graduate students at Universiti Sains Malaysia, who took part in our study in particular, students at the School of Mathematical Sciences.

Many thanks also go to my daughter and my son, Saja and Al-Harth for their presence and their sense of humour, which encouraged me to go on. Many thanks again to my brothers and sister in my home country who have all the time encouraged me to complete this research project. I would also like to express my sincere thanks and appreciation to my father, mother and wife for their permanent support and encouragement and for the sacrifices they have made and the care they have continually granted. So, I would like to dedicate this work to them. I am grateful to everybody for their helpful and constructive comments in the writing of this thesis.

TABLE OF CONTENTS

Page

Acknowledgements ii

Table of Contents iii

List of Tables vii

List of Figures xi

List of Abbreviations xxx

List of Symbols xxxii

List of Publications xxxiv

Abstrak xxxvi

Abstract xxxviii

CHAPTER 1 – INTRODUCTION

1.1 Introduction 1

1.2 Problem Statement 2

1.3 Objectives 3

1.4 Scope of the Study 3

1.5 Significance of the study 4

1.6 Limitation of Study 4

1.7 Outlines 5

CHAPTER 2 – LITERATURE REVIEW

2.1 Introduction 6

2.2 Signal Analysis 6

2.2.1 Fourier Analysis 6

2.2.1(a) Fourier Transform 7

2.2.1(b) Short-Time Fourier Transform 8

2.2.2 Wavelet Transform 9
2.2.3 The Wigner-Ville Distribution (WVD) 9
2.2.4 Hilbert-Huang Transform 10
2.3 Empirical Mode Decomposition 11
2.3.1 Intrinsic Mode Function (IMF) 12
2.3.2 Sifting Process 13
2.3.3 Theoretical Developments 15
2.3.4 Issues of Hilbert Huang Transform 16
2.4 Statistical Empirical Mode Decomposition(SEMD) 17
2.5 Non-parametric Regression 19
2.6 Kernel density estimation (KDE) 19
2.6.1 Kernel Density Error Criteria 21
2.6.2 Local Polynomial Regression 25
2.7 Bandwidth Selection 26
2.7.1 Direct Plug-In 28
2.8 Boundary effects and local linear regression 29
2.8.1 Boundary behavior of local linear regression 30
2.9 Quantiles 32
2.10 Conditional quantiles 33
2.11 Nonparametric Quantile Regression 35
2.11.1 Local Linear Quantile Regression 35
2.11.2 Bandwidth Selection 36
2.12 The behavior of local linear quantile estimator at boundary region 36

CHAPTER 3 – PROPOSED METHODS

3.1 Introduction 39
3.2 Problem of End Effects 39

3.2.1 Interpolation Problems 40

3.2.2 Piecewise Cubic Hermite Interpolation 40

3.3 Problems of the Spline Interpolation of Empirical Mode Decomposition .. 41

3.4 Empirical mitigation boundary processing techniques 42

3.4.1 The extended-based boundary processing method 43

3.4.2 Cosine window-based boundary processing method 44

3.5 Methodology 45

3.5.1 Empirical Mode Decomposition Combined with Local Linear Quantile Regression for Automatic Boundary Correction (EMD-LLQ) 46

3.5.2 Empirical Mode Decomposition Combined with Local Linear Regression for Automatic Boundary Correction (EMD-LL) 48

CHAPTER 4 – SIMULATION STUDY

4.1 Simulation study for EMD.LLQ 51

4.1.1 The Numerical Summary Experiment EMD.LLQ 54

4.1.2 The graphical summary Experiment EMD.LLQ 57

4.2 Simulation study for EMD.LLR 61

4.2.1 The Numerical Summary Experiment EMD.LLR 63

4.2.2 The graphical summary Experiment EMD.LLR 66

4.3 Summary 71

CHAPTER 5 – APPLICATION OF EMPIRICAL MODE DECOMPOSITION WITH LOCAL LINEAR QUANTILE REGRESSION IN FINANCIAL TIME SERIES FORECASTING

5.1 Introduction 73

5.2 Why EMD-LLQ (Empirical Mode Decomposition Combined with Local Linear Quantile Regression)? 74

5.2.1 Holt-Winters 75

5.5 Summary 80

CHAPTER 6 – CONCLUSION

6.1 Introduction 81

6.2 Discussion of the Results 81

6.3 The EMD-LLQ case results 81

6.4 The EMD-LLR case results 82

6.5 Contribution of the thesis 82

6.6 Future Work 83

REFERENCES 84

APPENDICES 91

LIST OF PUBLICATIONS 159

LIST OF TABLES

		Page
Table 4.1	Test functions used in simulation	52
Table 4.2	A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary	54
Table 4.3	A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary.	55
Table 4.4	A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary.	56
Table 4.5	The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, different noise structures, and sample size of 100.	63
Table 4.6	The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, different noise structures, and sample size of 100.	64
Table 4.7	The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, different noise structures, and sample size of 100.	64
Table 4.8	The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under even odd boundary solution, different noise structures, and sample size of 100.	65
Table 5.1	Comparison of RMSE, MAE, and MASE values for KLSE using the Holt-Winter, LLQ, EMD, and EMD-LLQ methods.	80
Table 5.2	Comparison of RMSE, MAE, and MASE values for NZX50 using the Holt-Winter, LLQ, EMD, and EMD-LLQ methods.	80
Table A.1	A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary.	92
Table A.2	A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary.	93

Table A.3 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 93

Table A.4 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 94

Table A.5 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 94

Table A.6 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 95

Table A.7 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 95

Table A.8 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 96

Table A.9 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 96

Table A.10 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 97

Table A.11 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 97

Table A.12 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 98

Table A.13 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 98

Table A.14 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 99

Table A.15 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 99

Table A.16 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 100

Table A.17 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 101

Table A.18 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 102

Table A.19 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 102

Table A.20 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 103

Table A.21 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 103

Table A.22 A comparison of the EMD with wave boundary condition, SEMD and proposed methods in the presence of boundary. 104

Table C.1 The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, different noise structures, and sample size of 100. 135

Table C.2 The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, different noise structures, and sample size of 100. 136

Table C.3 The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, different noise structures, and sample size of 100. 136

Table C.4 The global and local MSE and MAE for test Function 1 of the LLR, classical EMD, proposed method and SEMD under even odd boundary solution, different noise structures, and sample size of 100. 137

Table C.5 The global and local MSE and MAE for test Function 6 of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, different noise structures, and sample size of 100. 137

Table C.6 The global and local MSE and MAE for test Function 6 of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, different noise structures, and sample size of 100. 138

Table C.7 The global and local MSE and MAE for test Function 6 of the LLR, classical EMD, proposed method and SEMD under even odd boundary solution, different noise structures, and sample size of 100. 138

Table C.8 The global and local MSE and MAE for test Function 7 of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, different noise structures, and sample size of 100. 139

Table C.9 The global and local MSE and MAE for test Function 7 of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, different noise structures, and sample size of 100. 139

Table C.10 The global and local MSE and MAE for test Function 7 of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, different noise structures, and sample size of 100. 140

Table C.11 The global and local MSE and MAE for test Function 7 of the LLR, classical EMD, proposed method and SEMD under even odd boundary solution, different noise structures, and sample size of 100. 140

Table C.12 The global and local MSE and MAE for test Function 8 of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, different noise structures, and sample size of 100. 141

Table C.13 The global and local MSE and MAE for test Function 8 of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, different noise structures, and sample size of 100. 141

Table C.14 The global and local MSE and MAE for test Function 8 of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, different noise structures, and sample size of 100. 142

Table C.15 The global and local MSE and MAE for test Function 8 of the LLR, classical EMD, proposed method and SEMD under even odd boundary solution, different noise structures, and sample size of 100. 142

LIST OF FIGURES

Page

Figure 2.1	An example of extracts of IMF from sifting process	13
Figure 2.2	An explain on how EMD is working	15
Figure 2.3	Different types of kernel density estimation	21
Figure 2.4	Comparing the fit generated by different bandwidths for the precipitation data	28
Figure 4.1	The eight test functions used in simulation	52
Figure 4.2	The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structures is normal, value of quantiles (0.25), and sample size of 100.	57
Figure 4.2(a)	Test Function one.	57
Figure 4.2(b)	Test Function two.	57
Figure 4.2(c)	Test Function three.	57
Figure 4.2(d)	Test Function four.	57
Figure 4.2(e)	Test Function five.	57
Figure 4.2(f)	Test Function six.	57
Figure 4.2(g)	Test Function seven.	57
Figure 4.2(h)	Test Function eight.	57
Figure 4.3	The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structures is t-distribution with 3 degrees of freedom, value of quantiles (0.25), and sample size of 100.	58
Figure 4.3(a)	Test Function one.	58
Figure 4.3(b)	Test Function two.	58
Figure 4.3(c)	Test Function three.	58
Figure 4.3(d)	Test Function four.	58

Figure 4.3(e) Test Function five. 58

Figure 4.3(f) Test Function six. 58

Figure 4.3(g) Test Function seven. 58

Figure 4.3(h) Test Function eight. 58

Figure 4.4 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structures is AR (0.5), value of quantiles (0.25), and sample size of 100. 59

Figure 4.4(a) Test Function one. 59

Figure 4.4(b) Test Function two. 59

Figure 4.4(c) Test Function three. 59

Figure 4.4(d) Test Function four. 59

Figure 4.4(e) Test Function five. 59

Figure 4.4(f) Test Function six. 59

Figure 4.4(g) Test Function seven. 59

Figure 4.4(h) Test Function eight. 59

Figure 4.5 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, three different noise structures, and sample size of 100. 66

Figure 4.5(a) Test Function one with Normal-distribution Error. 66

Figure 4.5(b) Test Function one with t-distribution Error. 66

Figure 4.5(c) Test Function one with AR(0.5)-distribution Error. 66

Figure 4.6 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLR, classical EMD, and proposed method and SEMD under periodic boundary solution, three different noise structures, and sample size 100. 67

Figure 4.6(a) Test Function one with Normal-distribution Error. 67

Figure 4.6(b) Test Function one with t-distribution Error. 67

Figure 4.6(c) Test Function one with AR(0.5)-distribution Error 67

Figure 4.7	The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100.	68
Figure 4.7(a)	Test Function one with Normal-distribution Error	68
Figure 4.7(b)	Test Function one with t-distribution Error	68
Figure 4.7(c)	Test Function one with AR(0.5)-distribution Error	68
Figure 4.8	The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLR, classical EMD, proposed method and SEMD under evenodd boundary solution, three different noise structures, and sample size of 100.	69
Figure 4.8(a)	Test Function one with Normal-distribution Error	69
Figure 4.8(b)	Test Function one with t-distribution Error.....	69
Figure 4.8(c)	Test Function one with AR(0.5)-distribution Error.	69
Figure 5.1	KLSE closing price index	77
Figure 5.2	NZX50 closing price index	77
Figure 5.3	Steps of the experiment analysis.	79
Figure B.1	The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structure is normal, value of quantiles (0.50), and sample size of 100.	105
Figure B.1(a)	Test Function one.....	105
Figure B.1(b)	Test Function two.....	105
Figure B.1(c)	Test Function three.	105
Figure B.1(d)	Test Function four.	105
Figure B.1(e)	Test Function five.....	105
Figure B.1(f)	Test Function six.....	105
Figure B.1(g)	Test Function seven.....	105
Figure B.1(h)	Test Function eight.	105

Figure B.2 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structure is t-distributions with 3 of degree freedom, value of quantiles (0.50), and sample size of 100. 106

Figure B.2(a) Test Function one. 106

Figure B.2(b) Test Function two. 106

Figure B.2(c) Test Function three. 106

Figure B.2(d) Test Function four. 106

Figure B.2(e) Test Function five. 106

Figure B.2(f) Test Function six. 106

Figure B.2(g) Test Function seven. 106

Figure B.2(h) Test Function eight. 106

Figure B.3 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structure is AR(0.5) , value of quantiles (0.50), and sample size of 100. 107

Figure B.3(a) Test Function one. 107

Figure B.3(b) Test Function two. 107

Figure B.3(c) Test Function three. 107

Figure B.3(d) Test Function four. 107

Figure B.3(e) Test Function five. 107

Figure B.3(f) Test Function six. 107

Figure B.3(g) Test Function seven. 107

Figure B.3(h) Test Function eight. 107

Figure B.4 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structure is normal distributions, value of quantiles (0.75), and sample size of 100. 108

Figure B.4(a) Test Function one. 108

Figure B.4(b) Test Function two. 108

Figure B.4(c) Test Function three. 108

Figure B.4(d) Test Function four. 108

Figure B.4(e) Test Function five. 108

Figure B.4(f) Test Function six. 108

Figure B.4(g) Test Function seven. 108

Figure B.4(h) Test Function eight. 108

Figure B.5 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structure is t - distributions with 3 of degree freedom, value of quantiles (0.75), and sample size of 100. 109

Figure B.5(a) Test Function one. 109

Figure B.5(b) Test Function two. 109

Figure B.5(c) Test Function three. 109

Figure B.5(d) Test Function four. 109

Figure B.5(e) Test Function five. 109

Figure B.5(f) Test Function six. 109

Figure B.5(g) Test Function seven. 109

Figure B.5(h) Test Function eight. 109

Figure B.6 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under wave boundary solution, noise structures is AR(0.5) , value of quantiles (0.75), and sample size of 100. 110

Figure B.6(a) Test Function one. 110

Figure B.6(b) Test Function two. 110

Figure B.6(c) Test Function three. 110

Figure B.6(d) Test Function four. 110

Figure B.6(e) Test Function five. 110

Figure B.6(f) Test Function six. 110

Figure B.6(g) Test Function seven. 110

Figure B.6(h) Test Function eight. 110

Figure B.7 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under symmetric boundary solution, noise structures is normal, value of quantiles (0.25), and sample size of 100. 111

Figure B.7(a) Test Function one. 111

Figure B.7(b) Test Function two. 111

Figure B.7(c) Test Function three. 111

Figure B.7(d) Test Function four. 111

Figure B.7(e) Test Function five. 111

Figure B.7(f) Test Function six. 111

Figure B.7(g) Test Function seven. 111

Figure B.7(h) Test Function eight. 111

Figure B.8 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under symmetric boundary solution, noise structure is t - distributions with 3 of degree freedom, value of quantiles (0.25), and sample size of 100. 112

Figure B.8(a) Test Function one. 112

Figure B.8(b) Test Function two. 112

Figure B.8(c) Test Function three. 112

Figure B.8(d) Test Function four. 112

Figure B.8(e) Test Function five. 112

Figure B.8(f) Test Function six. 112

Figure B.8(g) Test Function seven. 112

Figure B.8(h) Test Function eight. 112

Figure B.9 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under symmetric boundary solution, noise structures is AR(0.5), value of quantiles (0.25), and sample size of 100. 113

Figure B.9(a) Test Function one. 113

Figure B.9(b) Test Function two. 113

Figure B.9(c) Test Function three. 113

Figure B.9(d) Test Function four. 113

Figure B.9(e) Test Function five. 113

Figure B.9(f) Test Function six. 113

Figure B.9(g) Test Function seven. 113

Figure B.9(h) Test Function eight. 113

Figure B.10 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, and proposed method and SEMD under symmetric boundary solution, noise structures is normal, value of quantiles (0.50), and sample size 100. 114

Figure B.10(a) Test Function one. 114

Figure B.10(b) Test Function two. 114

Figure B.10(c) Test Function three. 114

Figure B.10(d) Test Function four. 114

Figure B.10(e) Test Function five. 114

Figure B.10(f) Test Function six. 114

Figure B.10(g) Test Function seven. 114

Figure B.10(h) Test Function eight. 114

Figure B.11 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under symmetric boundary solution, noise structures is t - distributions with 3 of degree freedom, value of quantiles (0.50), and sample size of 100. 115

Figure B.11(a) Test Function one. 115

Figure B.11(b) Test Function two. 115

Figure B.11(c) Test Function three. 115

Figure B.11(d) Test Function four. 115

Figure B.11(e) Test Function five. 115

Figure B.11(f) Test Function six. 115

Figure B.11(g) Test Function seven. 115

Figure B.11(h) Test Function eight. 115

Figure B.12 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under symmetric boundary solution, noise structure is AR(0.5), value of quantiles (0.50), and sample size of 100. 116

Figure B.12(a) Test Function one. 116

Figure B.12(b) Test Function two. 116

Figure B.12(c) Test Function three. 116

Figure B.12(d) Test Function four. 116

Figure B.12(e) Test Function five. 116

Figure B.12(f) Test Function six. 116

Figure B.12(g) Test Function seven. 116

Figure B.12(h) Test Function eight. 116

Figure B.13 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under symmetric boundary solution, noise structures is normal, value of quantiles (0.75), and sample size of 100. 117

Figure B.13(a) Test Function one. 117

Figure B.13(b) Test Function two. 117

Figure B.13(c) Test Function three. 117

Figure B.13(d) Test Function four. 117

Figure B.13(e) Test Function five. 117

Figure B.13(f) Test Function six. 117

Figure B.13(g) Test Function seven. 117

Figure B.13(h) Test Function eight. 117

Figure B.14 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, and proposed method and SEMD under symmetric boundary solution, noise structures is t-distrbution, value of quantiles (0.75), and sample size 100. 118

Figure B.14(a) Test Function one. 118

Figure B.14(b) Test Function two. 118

Figure B.14(c) Test Function three. 118

Figure B.14(d) Test Function four. 118

Figure B.14(e) Test Function five. 118

Figure B.14(f) Test Function six. 118

Figure B.14(g) Test Function seven. 118

Figure B.14(h) Test Function eight. 118

Figure B.15 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under symmetric boundary solution, noise structures is AR(0.5), value of quantiles (0.75), and sample size of 100. 119

Figure B.15(a) Test Function one. 119

Figure B.15(b) Test Function two. 119

Figure B.15(c) Test Function three. 119

Figure B.15(d) Test Function four. 119

Figure B.15(e) Test Function five. 119

Figure B.15(f) Test Function six. 119

Figure B.15(g) Test Function seven. 119

Figure B.15(h) Test Function eight. 119

Figure B.16 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structures is normal, value of quantiles (0.25), and sample size of 100. 120

Figure B.16(a) Test Function one. 120

Figure B.16(b) Test Function two. 120

Figure B.16(c) Test Function three. 120

Figure B.16(d) Test Function four. 120

Figure B.16(e) Test Function five. 120

Figure B.16(f) Test Function six. 120

Figure B.16(g) Test Function seven. 120

Figure B.16(h) Test Function eight. 120

Figure B.17 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structures is t - distributions with 3 of degree freedom, value of quantiles (0.25), and sample size of 100. 121

Figure B.17(a) Test Function one. 121

Figure B.17(b) Test Function two. 121

Figure B.17(c) Test Function three. 121

Figure B.17(d) Test Function four. 121

Figure B.17(e) Test Function five. 121

Figure B.17(f) Test Function six. 121

Figure B.17(g) Test Function seven. 121

Figure B.17(h) Test Function eight. 121

Figure B.18 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structure is AR(0.5), value of quantiles (0.25), and sample size of 100. 122

Figure B.18(a) Test Function one. 122

Figure B.18(b) Test Function two. 122

Figure B.18(c) Test Function three. 122

Figure B.18(d) Test Function four. 122

Figure B.18(e) Test Function five. 122

Figure B.18(f) Test Function six. 122

Figure B.18(g) Test Function seven. 122

Figure B.18(h) Test Function eight. 122

Figure B.19 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structure is normal, value of quantiles (0.50), and sample size of 100. 123

Figure B.19(a) Test Function one. 123

Figure B.19(b) Test Function two. 123

Figure B.19(c) Test Function three. 123

Figure B.19(d) Test Function four. 123

Figure B.19(e) Test Function five. 123

Figure B.19(f) Test Function six. 123

Figure B.19(g) Test Function seven. 123

Figure B.19(h) Test Function eight. 123

Figure B.20 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structure is t-distribution with 3 degrees of freedom, value of quantiles (0.50), and sample size of 100. 124

Figure B.20(a) Test Function one. 124

Figure B.20(b) Test Function two. 124

Figure B.20(c) Test Function three. 124

Figure B.20(d) Test Function four. 124

Figure B.20(e) Test Function five. 124

Figure B.20(f) Test Function six. 124

Figure B.20(g) Test Function seven. 124

Figure B.20(h) Test Function eight. 124

Figure B.21 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structure is AR(0.5), value of quantiles (0.50), and sample size of 100. 125

Figure B.21(a) Test Function one. 125

Figure B.21(b) Test Function two. 125

Figure B.21(c) Test Function three. 125

Figure B.21(d) Test Function four. 125

Figure B.21(e) Test Function five. 125

Figure B.21(f) Test Function six. 125

Figure B.21(g) Test Function seven. 125

Figure B.21(h) Test Function eight. 125

Figure B.22 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structure is normal, value of quantiles (0.75), and sample size of 100. 126

Figure B.22(a) Test Function one. 126

Figure B.22(b) Test Function two. 126

Figure B.22(c) Test Function three. 126

Figure B.22(d) Test Function four. 126

Figure B.22(e) Test Function five. 126

Figure B.22(f) Test Function six. 126

Figure B.22(g) Test Function seven. 126

Figure B.22(h) Test Function eight. 126

Figure B.23 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structure is t-distribution with 3 degrees of freedom, value of quantiles (0.75), and sample size of 100. 127

Figure B.23(a) Test Function one. 127

Figure B.23(b) Test Function two. 127

Figure B.23(c) Test Function three. 127

Figure B.23(d) Test Function four. 127

Figure B.23(e) Test Function five. 127

Figure B.23(f) Test Function six. 127

Figure B.23(g) Test Function seven. 127

Figure B.23(h) Test Function eight. 127

Figure B.24 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under periodic boundary solution, noise structure is AR(0.5), value of quantiles (0.75), and sample size of 100. 128

Figure B.24(a) Test Function one. 128

Figure B.24(b) Test Function two. 128

Figure B.24(c) Test Function three. 128

Figure B.24(d) Test Function four. 128

Figure B.24(e) Test Function five. 128

Figure B.24(f) Test Function six. 128

Figure B.24(g) Test Function seven. 128

Figure B.24(h) Test Function eight. 128

Figure B.25 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under even odd boundary solution, noise structure is normal, value of quantiles (0.50), and sample size of 100. 129

Figure B.25(a) Test Function one. 129

Figure B.25(b) Test Function two. 129

Figure B.25(c) Test Function three. 129

Figure B.25(d) Test Function four. 129

Figure B.25(e) Test Function five. 129

Figure B.25(f) Test Function six. 129

Figure B.25(g) Test Function seven. 129

Figure B.25(h) Test Function eight. 129

Figure B.26 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under even odd boundary solution, noise structure is t-distribution with 3 degrees of freedom, value of quantiles (0.50), and sample size of 100. 130

Figure B.26(a) Test Function one. 130

Figure B.26(b) Test Function two. 130

Figure B.26(c) Test Function three. 130

Figure B.26(d) Test Function four. 130

Figure B.26(e) Test Function five. 130

Figure B.26(f) Test Function six. 130

Figure B.26(g) Test Function seven. 130

Figure B.26(h) Test Function eight. 130

Figure B.27 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under even odd boundary solution, noise structure is AR(0.5), value of quantiles (0.50), and sample size of 100. 131

Figure B.27(a) Test Function one. 131

Figure B.27(b) Test Function two. 131

Figure B.27(c) Test Function three. 131

Figure B.27(d) Test Function four. 131

Figure B.27(e) Test Function five. 131

Figure B.27(f) Test Function six. 131

Figure B.27(g) Test Function seven. 131

Figure B.27(h) Test Function eight. 131

Figure B.28 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under even odd boundary solution, noise structure is normal, value of quantiles (0.75), and sample size of 100. 132

Figure B.28(a) Test Function one. 132

Figure B.28(b) Test Function two. 132

Figure B.28(c) Test Function three. 132

Figure B.28(d) Test Function four. 132

Figure B.28(e) Test Function five. 132

Figure B.28(f) Test Function six. 132

Figure B.28(g) Test Function seven. 132

Figure B.28(h) Test Function eight. 132

Figure B.29 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under even odd boundary solution, noise structure is t-distribution with 3 degrees of freedom, value of quantiles (0.75), and sample size of 100. 133

Figure B.29(a) Test Function one. 133

Figure B.29(b) Test Function two. 133

Figure B.29(c) Test Function three. 133

Figure B.29(d) Test Function four. 133

Figure B.29(e) Test Function five. 133

Figure B.29(f) Test Function six. 133

Figure B.29(g) Test Function seven. 133

Figure B.29(h) Test Function eight. 133

Figure B.30 The Box plots demonstrate the global and local MSE and MAE for eight test Functions of the LLQ, classical EMD, proposed method and SEMD under even odd boundary solution, noise structure is AR(0.5), value of quantiles (0.75), and sample size of 100. 134

Figure B.30(a) Test Function one. 134

Figure B.30(b) Test Function two. 134

Figure B.30(c) Test Function three. 134

Figure B.30(d) Test Function four. 134

Figure B.30(e) Test Function five. 134

Figure B.30(f) Test Function six. 134

Figure B.30(g) Test Function seven. 134

Figure B.30(h) Test Function eight. 134

Figure D.1 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, three different noise structures, and sample size of 100. 143

Figure D.1(a) Test Function one with Normal-distribution Error Type. 143

Figure D.1(b) Test Function one with T-distribution Error Type. 143

Figure D.1(c) Test Function one with AR(0.5)-distribution Error Type. 143

Figure D.2 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, three different noise structures, and sample size of 100. 144

Figure D.2(a) Test Function one with Normal-distribution Error Type. 144

Figure D.2(b) Test Function one with T-distribution Error Type. 144

Figure D.2(c) Test Function one with AR(0.5)-distribution Error Type. 144

Figure D.3 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 145

Figure D.3(a) Test Function one with Normal-distribution Error Type. 145

Figure D.3(b) Test Function one with T-distribution Error Type. 145

Figure D.3(c) Test Function one with AR(0.5)-distribution Error Type. 145

Figure D.4 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under evenodd boundary solution, three different noise structures, and sample size of 100. 146

Figure D.4(a) Test Function one with Normal-distribution Error Type. 146

Figure D.4(b) Test Function one with T-distribution Error Type. 146

Figure D.4(c) Test Function one with AR(0.5)-distribution Error Type. 146

Figure D.5 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, three different noise structures, and sample size of 100. 147

Figure D.5(a) Test Function six with Normal-distribution Error Type. 147

Figure D.5(b) Test Function six with T-distribution Error Type. 147

Figure D.5(c) Test Function six with AR(0.5)-distribution Error Type. 147

Figure D.6 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, three different noise structures, and sample size of 100. 148

Figure D.6(a) Test Function six with Normal-distribution Error Type. 148

Figure D.6(b) Test Function six with T-distribution Error Type. 148

Figure D.6(c) Test Function six with AR(0.5)-distribution Error Type. 148

Figure D.7 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 149

Figure D.7(a) Test Function six with Normal-distribution Error Type. 149

Figure D.7(b) Test Function six with T-distribution Error Type. 149

Figure D.7(c) Test Function six with AR(0.5)-distribution Error Type. 149

Figure D.8 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under evenodd boundary solution, three different noise structures, and sample size of 100. 150

Figure D.8(a) Test Function six with Normal-distribution Error Type. 150

Figure D.8(b) Test Function six with T-distribution Error Type..... 150

Figure D.8(c) Test Function six with AR(0.5)-distribution Error Type..... 150

Figure D.9 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under wave boundary solution, three different noise structures, and sample size of 100. 151

Figure D.9(a) Test Function seven with Normal-distribution Error Type..... 151

Figure D.9(b) Test Function seven with T-distribution Error Type..... 151

Figure D.9(c) Test Function seven with AR(0.5)-distribution Error Type..... 151

Figure D.10 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under periodic boundary solution, three different noise structures, and sample size of 100. 152

Figure D.10(a) Test Function seven with Normal-distribution Error Type..... 152

Figure D.10(b) Test Function seven with T-distribution Error Type..... 152

Figure D.10(c) Test Function seven with AR(0.5)-distribution Error Type..... 152

Figure D.11 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 153

Figure D.11(a) Test Function seven with Normal-distribution Error Type..... 153

Figure D.11(b) Test Function seven with T-distribution Error Type..... 153

Figure D.11(c) Test Function seven with AR(0.5)-distribution Error Type..... 153

Figure D.12 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 154

Figure D.12(a) Test Function seven with Normal-distribution Error Type..... 154

Figure D.12(b) Test Function seven with T-distribution Error Type..... 154

Figure D.12(c) Test Function seven with AR(0.5)-distribution Error Type..... 154

Figure D.13 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 155

Figure D.13(a) Test Function eight with Normal-distribution Error Type. 155

Figure D.13(b) Test Function eight with T-distribution Error Type. 155

Figure D.13(c) Test Function eight with AR(0.5)-distribution Error Type. 155

Figure D.14 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 156

Figure D.14(a) Test Function eight with Normal-distribution Error Type. 156

Figure D.14(b) Test Function eight with T-distribution Error Type. 156

Figure D.14(c) Test Function eight with AR(0.5)-distribution Error Type. 156

Figure D.15 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 157

Figure D.15(a) Test Function eight with Normal-distribution Error Type. 157

Figure D.15(b) Test Function eight with T-distribution Error Type. 157

Figure D.15(c) Test Function eight with AR(0.5)-distribution Error Type. 157

Figure D.16 The Box plots demonstrate the global and local MSE and MAE for four test Functions of the LLR, classical EMD, proposed method and SEMD under symmetric boundary solution, three different noise structures, and sample size of 100. 158

Figure D.16(a) Test Function eight with Normal-distribution Error Type. 158

Figure D.16(b) Test Function eight with T-distribution Error Type. 158

Figure D.16(c) Test Function eight with AR(0.5)-distribution Error Type. 158

LIST OF ABBREVIATIONS

CV Cross-Validation Criterion

CDF Cumulative Density Function

DPI Direct Plug-In Selection Strategy

EMD Empirical Mode Decomposition

EMD-LLQ Empirical Mode Decomposition Combined with Local Linear Quantile Regression

EMD-LLR Empirical Mode Decomposition Combined with Local Linear Regression

HHT Hilbert Huang Transform

HSA Hilbert Spectral Analysis

IMF Intrinsic Mode Function

LLR Local Linear Regression

LPK Local Polynomial Kernel Estimator

LOWESS locally weighted scatter plot smoothing

MISE Mean Integreated Square Error

MSE Mean Square Erorr

MAE Mean Absolute Error

NW Nadaraya Watson

ROT Role of Thumb

RSC Residual Squares Criterion

STFT Short-Time Fourier Transform

WLS Weighted Least Square

WVD The Wigner-Ville Distribution

LIST OF SYMBOLS

h Bandwidth or smoothing parameter

C Constant

$F(y/x)$ Conditional quantile of Y given x

$d(t)$ Details that satisfied IMF conditions

W_P Diagonal matrix

$d(t)$ Details that satisfied IMF conditions

$\hat{f}(x;h)$ Estimated of probability density function

F_{t+m} Forecast for m period

ω_0 fundamental angular frequency in radians per second

X_i Independently distributed observation

$\boldsymbol{F}$ Known continuous function

K Kernel function

S Length of Seasonality

L_t Level Series

$R(g)$ Parsimonious notation for any square integrable function

$p_i(x)$ Polynomial of order degree

τ/ρ_τ Quantile

$m_k(t)$ Residue

Φ **and** ϕ Standard density and distribution functions of h_{mean}

n Sample size

a_0, b_0 *sin* and *cosine* Coefficients

b_t Trend

e_i Vector

SUATU PEMBETULAN AUTOMATIK SEMPADAN MENGGUANKAN PENGURAIAN MOD EMPIRIK

ABSTRAK

Penghuraian mod empirik (EMD) sangat berguna dalam menganalisis siri masa tak pegun dan tak linear. Namun EMD mengalami masalah batasan yang disebabkan oleh penggunaan EMD untuk isyarat yang terhingga. Akibatnya, kepincangan besar terjadi dipinggir dan pergerakan buatan berlaku ketika andaian batasan klasik tidak terpenuhi. Kajian ini memperkenalkan dua kaedah baru dengan dua tahap untuk mengurangkan secara automatik kesan batasan yang hadir dalam EMD iaitu gabungan penghuraian mod empirik dengan regresi linear kuantil tempatan, EMD-LLQ dan gabungan peng-huraian mod empirik dengan dengan regresi linear tempatan, EMD-LL. Kejituan kae-dah ini ditunjukkan melalui kajian simulasi dan contoh sebenar dengan dibandingkan dengan kaedah imputasi lain yang sedia ada. Pada bahagian simulasi, pada tahap perta-ma, regresi linear kuantil tempatan (LLQ) dan regresi linear tempatan (LL) digunakan bertujuan untuk memberikan keperihalan yang cekap dari data yang tidak bagus dan hingar. Siri tersisa diandaikan tersembunyi dalam reja. Oleh itu, EMD digunakan pada residual pada tahap kedua. Anggaran terakhir adalah penghasiltambahan anta-ra anggaran penyuaian LLQ dan EMD serta LL dan EMD. Penemuan menarik telah dihasilkan dari kajian ini. Salah satu yang penting adalah kaedah yang dicadangkan EMD-LLQ dan EMD-LL lebih cekap daripada batasan tradisional lanjutan (gelom-bang, berkala, simetri, dan ganjil genap) EMD dan lanjutan EMD iaitu penghuraian mod empirik berstatistik (SEMD)

mod empirik berstatistik (SEMD) ketika batasan wujud. Untuk aplikasi data sebenar kaedah yang dicadangkan, EMD-LLQ dan EMD-LL digunakan untuk meramal dua indeks pasaran saham. Ujikaji terperinci dilakukan untuk kaedah yang dicadangkan, di mana kaedah EMD-LPQ, EMD, dan Holt-Winter dibandingkan. Dapatan menunjukkan kaedah cadangan EMD-LPQ dan EMD-LL adalah lebih unggul daripada EMD dan Holt-Winter dalam meramal harga penutupan saham.

ON AUTOMATIC BOUNDARY CORRECTIONS USING EMPIRICAL MODE DECOMPOSITION

ABSTRACT

Empirical mode decomposition (EMD) is particularly useful in analyzing nonstationary and nonlinear time series. Yet EMD struggle from boundary problems caused by the application of the EMD to a finite signal. Consequently, large bias at the edges and artificial wiggles happen when the classical boundary assumptions are not met. This study introduces two new two-stage methods to automatically decrease the boundary effects present in EMD namely Empirical Mode Decomposition combined with local linear quantile regression, EMD-LLQ and Empirical Mode Decomposition combined with local linear regression, EMD-LL. The accuracy of the method is shown by simulation studies and real example with comparison to other existing imputation methods. For simulation part: at the first stage, local linear quantile regression (LLQ) and local linear regression are applied to provide an efficient description of the corrupted and noisy data. The remaining series is assumed to be hidden in the residuals. Hence, EMD is applied to the residuals at the second stage. The final estimate is the summation of the fitting estimates from LLQ and EMD and LL and EMD. Many interesting features have been noted from the study. One of the most striking points is that the pro-posed methods EMD-LLQ and EMD-LL are more efficient than traditional boundary extensions (wave, periodic, symmetric, and even odd) of EMD and its extension sta-tistical empirical mode decomposition (SEMD) when the boundaries are present. For the real data application we apply the proposed techniques , EMD-LLQ

, and EMD-LL to forecast two closed stock market index. Detailed experiments are carried out for the proposed method, in which EMD-LLQ, EMD, and Holt-Winter methods are com-pared. The proposed EMD-LLQ and EMD-LL methods are determined to be superior to the EMD and Holt-Winter methods in predicting the stock closing prices.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Traditional and advanced nonparametric curve estimation methods do not require constraints on the functional form of the curves of interest; hence, these methods allow flexible modeling of the data. Unfortunately, most nonparametric curve estimation methods suffer from endpoint effect (boundary effect) which usually show a sharp increase in variance and bias when estimating a function $f(t)$ at points near the boundary of support of the function (e.g, $t < h$) (Kyung-Joon and Schucany, 1998). Boundary problem is a serious issue that significantly influences global achievement visually and negatively affects the rate of convergence in the traditional asymptotic analysis. Several mechanisms are needed to reduce boundary effects. The boundary effect in traditional nonparametric regression, such as kernel methods, has been extensively discussed by Gasser and Müller (1979), and Granovsky and Müller (1991). To solve such problem in kernel methods, Rice et al. (1984) proposed a linear combination of two kernel estimators with diverse bandwidths to diminish the bias. The traditional nonparametric regression named smoothing spline has also encountered boundary effects (Rice and Rosenblatt, 1981). Eubank and Speckman (1991) also put forward boundary adjustment methods. Although these methods enabled actual boundary adjustment, the more active methods are difficult to implement to some extent. An uncomplicated and more direct methodology to boundary adjustment based on local polynomial fitting for perceptive studies was suggested by Fan and Gijbels (1992). The ease of use of this

method may be ascribed to "automatic" boundary adaptation, that is, explicit adjustment is not required. Although local polynomial fitting is widely recognized because it is easy to apply, how well it could improve other nonparametric methods that suffer from boundary problem remains as the main question. Advanced nonparametric methods (time-frequency analyses), such as the wavelet method, struggle with boundary effect. Oh et al. (2001) applied a low-order polynomial to adjust the boundary problem in a wavelet method. Empirical mode decomposition (EMD) has recently attracted the attention of researchers especially when dealing with nonlinear and nonstationary data. Despite its advantages, EMD shares with other methods a similar problem that has been identified as end edges. The end edge problem has been discussed extensively since Huang first reported it in 1998. Such empirical solutions as wave, periodic, symmetric, even, and odd solutions have been applied to solve such a problem, see Kim and Oh (2009). Unfortunately, these solutions do not effectively solve the problem of end edges. This study adopts the method of Oh et al. (2001) by applying local linear quantile regression and local linear regression to mitigate the boundary effect of EMD automatically. Choosing a local linear smoother and its robust version, local linear quantile smoother, ensures simplicity without loss of efficiency (Koenker, 2005). Nonparametric boundary problems are fascinating and challenging. However, boundary problems at a boundary point have been rarely studied, and the methods used at a boundary point are much more diverse than those used at an interior point.

1.2 Problem Statement

In the sifting process, cubic spline interpolation of maxima and minima is crucial to obtain the lower and upper envelopes. According to Huang et al. (1998), adopting

cubic spline interpolation may not be an ideal choice because the resulting intrinsic mode function (IMF) cannot guarantee symmetric envelops. This situation happens when the endpoints are not extrema.

1.3 Objectives

The main objective of this work is to mitigate boundary problems with EMD; novel two-stage techniques are proposed to reduce bias at the boundaries. The proposed methods are based on a combination of EMD and low-order polynomial quantile regression [i.e., local linear quantile regression (LLQ)] and a combination of EMD and low-order polynomial regression [i.e., local linear regression (LLR)].

1.4 Scope of the Study

The focal point of this work is addressing the boundary problem of independent and correlated noise in EMD. Normal, heavy-tailed, and autoregressive noises will be considered. Two automatic boundary adjustment estimators are examined, namely, Empirical Mode Decomposition Combined with Local Linear Quantile Regression (EMD-LLQ) and Empirical Mode Decomposition Combined with Local Linear Regression (EMD-LL) for one dimension. To provide a convenient framework for the discussion, the two novel boundary adjustment techniques will be compared with four different boundary extension methods. In this comparison, only the equally spaced design points, in which the designer chooses the predictor variable values by using a predefined scheme, will be the focus.

1.5 Significance of the study

EMD is a new nonparametric technique with the ability to reveal unusual aspects that might be observed in noisy data. Considering the boundary problem and its effects on the final estimates, the estimation in the boundary region is remarkably important. We believe that, in real-life situations, much attention should be paid to the estimation in the boundary region. For instance, if the data are related to poverty analysis, obtaining reliable estimates of the income distribution on the left side, close to "0" (left boundary point), is necessary. Similarly, when EMD is employed to describe econometric data (e.g., illustrating the performance of specific groups of people, such as the elderly or the youth and comparing large and small companies), we always focus on the estimates of the boundary. In such cases, data may be generated from the populations where the normality assumption is not met or the noise is correlated. Therefore, boundary problem must be taken seriously. In this regard, robust and effective solutions are required.

1.6 Limitation of Study

In this thesis, we focus on dealing with boundary issues of EMD by proposing two novel methods. Another fundamental problem in EMD-LLQ and EMD-LLR technique is mode mixing. Mode mixing refers to the phenomenon where portions of signal components get distributed over multiple IMFs. This can occur due to noise or intersecting instantaneous frequencies of signal components. Methods to prevent mode mixing, or, if they occur, procedures to optimally combine signal contributions spread over multiple IMFs present yet another area of research problems.

1.7 Outlines

Chapter 2 gives a general review of EMD and other nonparametric regression techniques. Chapter 3 presents the proposed automatic boundary correction methods called EMD-LLQ and EMD-LL. A performance evaluation of the proposed methods is described in chapter 4. Chapter 5 discusses an application of the newly proposed methods, with the classical EMD as benchmark. Finally, conclusions and suggestions for further research are drawn in chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter provides a brief explanation of the various tools available to analyse a non-stationary signal. Their relative merits are compared and a brief description of the EMD algorithm for signal analysis is offered. Finally, an overview of the developments in the field of EMD and their applications is examined.

2.2 Signal Analysis

Time series can be studied using various processes. Practitioners use methods such as short-time Fourier transform (STFT), Fourier transform (FT), wavelet transform, Wigner-Ville and empirical mode decomposition (EMD) for ultimate task of processing data, which often comprises several consecutive steps of solving a statistical decision problem (detection, estimation, classification, recognition, etc.). The advantages and disadvantages of each method will be explored throughout this study.

2.2.1 Fourier Analysis

Fourier analysis is the most frequently used method. It decomposes the signal into sinusoids of differing frequencies in order to determine the frequency content. Fourier Transform (FT) is mostly used for non-periodic signals, whereas periodic signals tend

to utilise Fourier series.

In **Fourier series**, the periodic signal $f(t)$ of period T is depicted as follows:

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k cos(k\omega_0 t) + \sum_{k=1}^{\infty} b_k sin(k\omega_0 t), \tag{2.1}$$

where $\omega_0 = \frac{2\pi}{T}$ represents the fundamental angular frequency in radians per second. The sine and cosine term coefficients are determined using the following:

$$a_0 = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)dt,$$

$$a_k = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)cos(k\omega_0 t)dt,$$

$$b_k = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)sin(k\omega_0 t)dt,$$

$k = 1, 2, ..., \infty$.

2.2.1(a) Fourier Transform

The FT of the square integral function $f(f \in L^2(R))$ can be depicted as follows:

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t} dt, \tag{2.2}$$

and the inverse of the square integral function $f(f \in L^2(R))$ can be depicted as follows:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} dt, \tag{2.3}$$

Signals are analysed by FT using spectral analysis. This is the standard method of obtaining information regarding period signals. FT can be extended using discrete Fourier transform (DFT). The DFT can be computed faster by determining the coefficients using the butterfly algorithm (Cooley and Tukey, 1965). FT does not reveal the temporal location of frequencies and can therefore only be used for stationary signals. Short-time Fourier transform (STFT) can be utilised to solve this problem.

2.2.1(b) Short-Time Fourier Transform

The original signal is broken down into smaller duration particles using STFT. FT is then used to determine the frequency components. FT does not recognise the boundaries and discontinuities and will therefore create higher order harmonics to complement the waveform. Windowing aims to mitigate against this. The smooth window function is utilised, with a unified origin and decay to zero towards the end. Specific criterion must be used so that optimal performance can be achieved. An example of an optimal window has been used by Hamming et. al. (see Carmona et al., 1998), where the FT is concentrated near $\omega = 0$ and possesses functional forms. The FT is known as the 'spectral window' whilst the window in a time-domain is referred to as the 'time window.' A signal is multiplied with window functions $g(t-b)$, where $g(t)$ depicts the functional form and is non-zero in the finite region at time b. FT of $f(t)g(t-b)$ should be determined followed by the movement of the window to a new position and the repetition of the process. The process is depicted below:

$$Sf(\omega, b) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)g(t-b)e^{-i\omega t}dt, \tag{2.4}$$

The signal reconstruction occurs by utilising the following formula:

$$Sf(\omega,b) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Sf(\omega,b) g(t-b) e^{i\omega t}. \tag{2.5}$$

There are some major problems with STFT. Chui (2014) argues that because frequency is proportional to the number of cycles in a specific time interval, locating high-frequency phenomena requires a narrow window, whereas in order to investigate low frequency phenomena, a wide time window is necessary. Thus, for signals with both high and low frequencies, STFT is not an ideal solution.

2.2.2 Wavelet Transform

Wavelet analysis is an emerging method of studying non-stationary signals. The finite energy of a wavelet is concentrated in time or frequency. Wavelet analysis differs from Fourier method as it decomposes signals into individual functions known as wavelets. The locations of wavelets do not match on the T axis and deteriorate to zero. Therefore, they are considered 'local'. Grossmann and Morlet (1984) first used the expression 'wavelet' and defined it to be a two-parameter expansion of a signal within a basic wavelet function. High frequency wavelets are subject to temporal analysis, whereas a low frequency version of the same wavelet will be subjected to frequency analysis.

2.2.3 The Wigner-Ville Distribution (WVD)

WVD is used because of the limited nature of the spectrogram. Ville (1958) utilised the method for spectral analysis after it was introduced for quantum mechan-

ics (Wigner, 1932). Non-stationary signals can be depicted in high-resolution time and frequency using WVD. The satisfaction of frequency and time margin due to the instantaneous power in the time and energy spectrum of frequency is also solved using WVD. However, confusion can arise when the energy distribution is positive, leading to severe cross-terms between components in different t-f regions.

2.2.4 Hilbert-Huang Transform

Huang et al. (1998) introduced a concept known as empirical mode decomposition (EMD) to make a signal ready for Hilbert transform analysis. The combination of EMD and Hilbert transform analysis is known as Hilbert-Huang Tranform (HHT).
Some restrictions must be applied to the data (Boashash, 1992). Firstly, the signal must be mono-component, with no riding waves. The signals that can be studied using the Hilbert transform can only be simple free vibrations. The transform is not applicable to multi-component signals because of the limited data. But band-pass filtering can separate the components and extend the application value to multi-component signals, hence the development of HHT.
HHT has vast uses in time-frequency analysis of dispersive, nonlinear and non-stationary signals and systems. Some uses include decomposing the signal into a series of constituents. Once the method is applied, a set of analytical signals depicting the input signal can be obtained. HTT is able to calculate the instantaneous frequency of each constituent and present the results of time-frequency analysis in a Hilbert spectrum plot. The signal analysis step of HHT, and the EMD, is the subject of this study and the following sections and chapters focus on this technique and its variants.

2.3 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) refers to an adaptive signal-dependent decomposition where any complicated signal can be decomposed into a series of constituents. Amalgamating the constituents can reconstruct the original signal without distortion. Previous sections discusses some of the methods used to simultaneously analyse signals in the time and frequency domains, which are based on the expansion of the signal into a set of basic functions. EMD expands the signal into a set of basic functions defined by the Intrinsic mode functions (IMF). Intrinsic mode functions (IMF) refer to the decomposed constituents. Where Fourier methods have failed, EMD can use time-frequency analysis of transient signals. The uncertainly principle limits the use of the Fourier transform to represent frequencies. Exact information about the frequencies of a signal can be obtained for infinite signal length, but there are restrictions for analysing signals of finite length as there is a limit on the precision of detectable frequencies. The signal frequency at a moment in time, minus the information of signal frequency at alternative times, is known as the instantaneous frequency. EMD can be utilized when a multicomponent signal consists of several intrinsic frequencies. EMD decomposes the signal into its IMF with instantaneous frequency to allow several instantaneous signal frequency components to be calculated. EMD is able to directly extract signal components overlapping in time and frequency by creating an adaptive signal-dependant time-variant filtering procedure as suggested by Rilling et al. (2003). Also, the physical meaning of the intrinsic process underlying the complex signal can be preserved in the decomposed signals, as the results are not influenced by predetermined bases or sub band filtering processes.

EMD offers a new approach to signal analysis. It is an adaptive decomposition with

which complicated signals can be broken down into a finite number of series. EMD has the ability to describe short time changes in frequencies that cannot be resolved by Fourier spectral analysis and can therefore be used for non-linear and non-stationary time series analysis (Huang et al., 2003b). The intended purpose of EMD was to discover a decomposition that made it possible to perform time frequency analysis using the instantaneous frequency of non-stationary signals. This technique is explored in more detail in the following sections.

2.3.1 Intrinsic Mode Function (IMF)

The pioneering researchers defined two criteria that must be satisfied for a component to be classified as an IMF (Huang et al., 1998). Firstly, there cannot be more than a value of one difference between the quantity of extremes and the quantity of zero crossings. Secondly, the average of the envelopes should be equal to zero at any given point, meaning that the functions must be symmetrical.

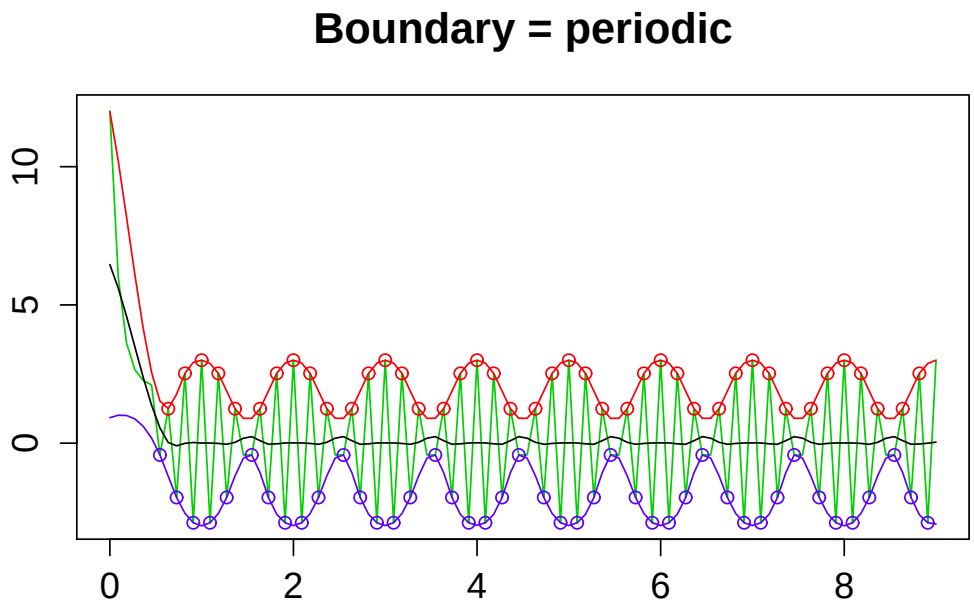

Figure 2.1: An example of extracts of IMF from sifting process

From Figure 2.1, the green line represents the actual data. Both red and blue lines represent the upper and lower envelopes defined by the local maxima and minima, respectively. The mean value of the upper and lower envelopes is given in black.

2.3.2 Sifting Process

Sifting is used to extract each of the IMF's. The goal of sifting is to remove the higher frequency components until only the low frequency components remain. For example, for a signal $x(t)$, the method will divide it into high frequency detail $d(t)$ and low frequency residual (or trend) $m(t)$, so that:

$$x(t) = m(t) + d(t).$$

this becomes the first IMF, and the sifting process is repeated on the residual:

$$x(t) = m(t) + d(t).$$

After K iterations of sifting, the input signal is represented as follows:

$$x(t) = \sum_{k=1}^{K} y_k(t) + m_k(t), \tag{2.6}$$

where $y_k(t)$, $k = 1, ..., K$ refer to IMFs and $m_k(t)$ is the residual, or mean trend.

The effective algorithm of EMD can be summarized as follows (Rilling et al., 2003):

1- Identify all extrema of $x(t)$.

2- Interpolate between minima (respectively maxima), resulting in the envelope $e_{min}(t)$ respectively ($e_{max}(t)$).

3- Compute the mean $m_t = \frac{e_{min}(t)+e_{max}(t)}{2}$.

4- Extract the detail $d(t) = x(t) - m(t)$.

5- If $d(t)$ satisfies all IMF conditions, then set $y_1(t) = d(t)$, the first IMF, else repeat above steps with $d(t)$.

6- Evaluate the residual $m_1(t) = x_t - y_1(t)$.

7- Iterate on the residual $m_1(t)$.

Steps 1-4 may be repeated until the detail $d(t)$ satisfies the IMF conditions. The following sections will discuss practical ways of determining if d(t) satisfies the IMF condition. Figure 2.2 on page 15 shows a practical example which summarizes a sifting process algorithm graphically. EMD decomposed signal into five IMFs and also produced a global trend (residue). IMF_1 represents the high frequency of the signal while IMF_2 and IMF_3 extract mid-range frequency signals. In particular, long-term

behaviors are well described by behavior of IMF_4 through IMF_5. Note that the residue signal is the signal remaining after all oscillatory components have been removed from the orignal signal. Thus the residue might be physically interpreted as a trend.

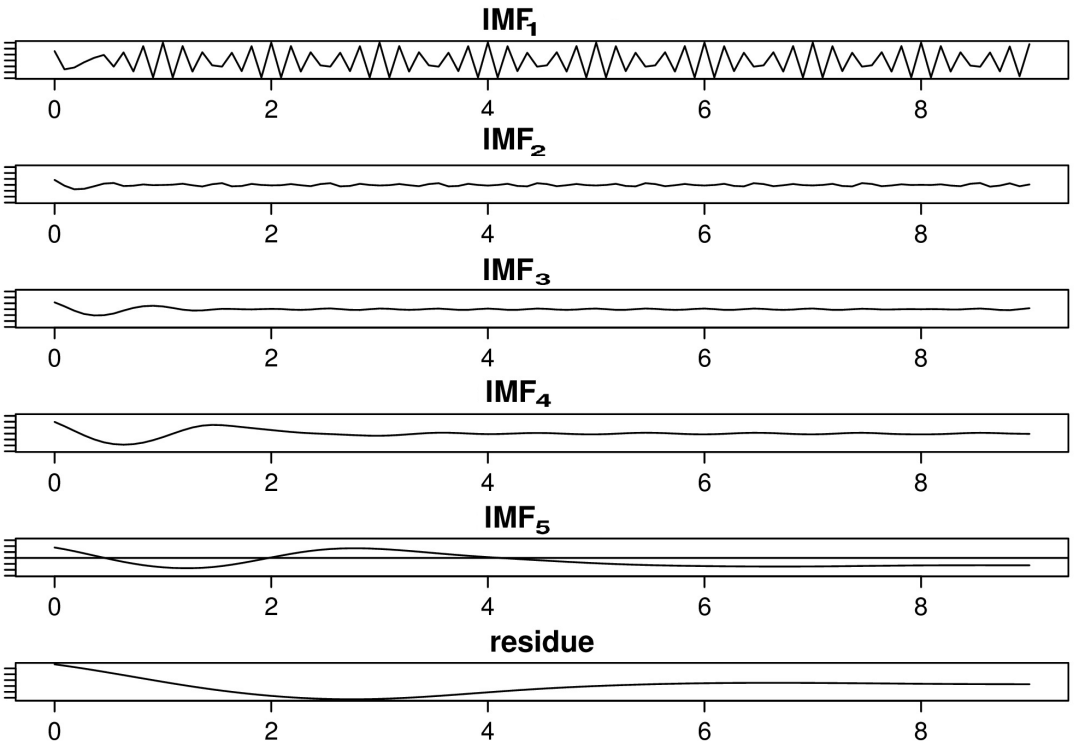

Figure 2.2: An explain on how EMD is working

.

2.3.3 Theoretical Developments

Researchers have attempted to examine some theoretical approaches to EMD, despite it remaining largely algorithmic in nature. Wu and Huang (2004) were the first authors to explore a theoretical approach of EMD. White noise was applied and it was discovered that EMD is a dyadic filter, which can provide octave band frequency decomposition to the input. A filter bank interpretation was then applied to the algorithm

(Flandrin et al., 2004). The dyadic filter nature of EMD has since been recognized by several studies. Although the dyadic filter nature of the EMD algorithm has been quoted by several authors subsequently, it must be remembered that the algorithm behaves so only when presented with white noise-like broadband signal. Studies such as Feldman (2009) and Rilling and Flandrin (2008) developed an analysis in the signal decomposition performance for combined tones.

Sampling rate has consistently been viewed as a research problem. The effect of sampling on decomposition quality and improved performance of the algorithm when the sampling rate is decreased has been extensively explored. Both an empirical and theoretical sampling limit for the algorithm was derived in Stevenson et al. (2005). It was found that the algorithm performance decreases at low sampling rates (nearer to the Nyquist rate). Studies such as Rilling and Flandrin (2006) examined the effect of sampling size on the decomposition quality whereas Xu et al. (2009) developed new methods based on Fourier interpolation which increase the performance of low sampling rates.

2.3.4 Issues of Hilbert Huang Transform

There are three main identified problems with the Hilbert-Huang Transform: side-effect, stopping criterion and mode mixing problem. These will be examined below:

(a) Boundary-effect Boundary effect affects all signal processing and analysis tools. The issue becomes very important in EMD, especially in short sequence. The shape of the spline will be affected by the spline calculation, and different splines will create different IMFs, which changes the overall output of the system. Several methods are used to rectify this issue: (1) characteristic wave extending method, (2) mirror ex-

tending method, (3) data extending method, (4) data reconstruction method and (5) similarity searching method.

(b) Stopping Criterion A stopping criterion must be set so that the iteration number will not be too large as to break the physical meaning of the signal as well as to ensure the EMD can separate each IMF with its instantaneous frequency and sustain the instance magnitude. In previous studies by Huang et al. (1998), the method for a specific IMF stops at the point where the normalised difference in the extracted signal between two consecutive iterations becomes less than a pre-determined threshold ε. A novel stopping criterion was invented in (Linderhed, 2004a). When the envelope mean signal is close to zero the iteration is ceased. Enforcing the envelope mean towards zero will ensure that the envelope remains symmetrical and there is a consistent relationship between the number of zero crossings and number of extremes. An alternative version of this stopping criterion was introduced in Rilling et al. (2003) with an examination of typical threshold values. Another criterion is outlined in Huang et al. (2003a). Sifting is ceased when the number of extremes and zero crossings remains steady after a pre-determined quantity of iterations.

(c) Mode Mixing Problem The specific signal may not be separated to the same EMD each time. This is known as the mode mixing problem. It has the potential to make the feature extraction, model training and pattern recognition difficult as the feature cannot be fixed in one labelling index.

2.4 Statistical Empirical Mode Decomposition(SEMD)

Due to interpolation process in the construction of envelopes, IMFs achieved by the traditional EMD process are sensitive to noise, and outliers. The non-informative

fluctuation effect misleads the subsequent breakdown results. Moreover, this technique emphaszes on a narrow scope that is unable to screen irregularly sampled data. These limitations of its scope steadily reduce the applicability of EMD to numerous signals. A SEMD technique is a statistical EMD technique designed to be an alternative to traditional EMD. SEMD depends on a smoothing technique. In comparison, SEMD has a numerous advantage points over the traditional EMD because SEMD process is robust to noise and outliers, and hence, SEMD is able to breakdown noisy signals into suitable IMFs without bias, and SEMD offers proper boundary condition of an IMF without any boundary solution (Kim et al., 2012). Formally, the SEMD algorithm can be stated as follows:

A. (Modified sifting) Take a signal x to be decomposed, and extract the first mode $h_{1,\lambda}$ by using a smoothing technique.

(A-1) Identify the local maxima (minima) z of the signal $h^0_{1,\lambda}$ where $h^0_{1,\lambda}$ is the original x.

(A-2) Construct an upper envelope $\hat{u}_\lambda$ (lower envelope $\hat{\ell_\lambda}$) by applying smoothing technique with a smoothing parameter λ to maxima (minima) z.

(A-3) compute the local mean $m_\lambda = \frac{1}{2}\hat{u} + \hat{\ell}$ by the average of both the envelopes, and then obtain a candidate intrinsic mode $h^1_{1,\lambda} = h^0_{1,\lambda} - m_\lambda$.

(A-4) Repeat steps (A-1)-(A-3) for the signal $h^1_{1,\lambda}$ until the jth iteration satisfies the IMF conditions.

(A-5) Decompose the signal $x = h_{1,\lambda} + r_\lambda$ defined as limit of $h^j_{1,\lambda}$ and r_λ is the remaining signal.

B. (Conventional sifting) If the remaining signal $r_\lambda = x - h_{1,\lambda}$ has an intrinsic oscillation mode, then r_λ can be further decomposed by conventional sifting.

The only difference between SEMD algorithm and the traditional EMD is step A, where the first mode is breakdown by smoothing instead of interpolation. In particular, step (A-2) in construction of $\hat{u}_\lambda$ and $\hat{\ell}_\lambda$ by smoothing plays the most important role in determining the quality of the decomposition when the signal is corrupted by non-informative random fluctuations. For more explanation, see Kim et al. (2012).

2.5 Non-parametric Regression

To overcome the limitation of classical parametric regression where the restrictive assumption of the regression function form is needed, one may use a nonparametric regression. Unlike traditional regression methodologies, there is no restrictive assumption on regression function form. Earlier variety of nonparametric regression and smoothing methods have been studied extensively. Kernel and local polynomial regressions are popular nonparametric methods. This section provides a brief introduction for both of these methods.

2.6 Kernel density estimation (KDE)

In KDE it is assumed that we are given a sample of n identically and independently distributed (iid) observations $X_1, X_2, ..., Xn$ from which a density will be estimated. Let $f(x)$ be the true probability density function (PDF) and $\hat{f}(x;h)$ be the estimated PDF Hoover et al. (1998). Showed that the kernel density estimate of $f(x)$ is

$$\hat{f}(x;h) = n^{-1}h^{-1}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), \tag{2.7}$$

where K is the kernel function that satisfies $\int K(y)\,dy = 1$, $\int yK(y)dy = 0$, and $0 < \int y^2K(y)dy < \infty$, all odd moments are zero ,and h is the bandwidth or smoothing parameter Wand and Jones (1995). The $\int k\left(y^2\right)dy$ function is denoted by $\mu_2(K)$ which indicates the second moment of a PDF. When $\mu_2 > 0$ then the kernel is said to be of order two. The unsigned integral symbol $\int$ is taken to be over the real line unless otherwise noted.A more compact way to express Equation (2.7) is by letting $u = x - X_i$ and $K_h(u_i) = h^{-1}K(u/h)$:

$$\hat{f}(x;h) = n^{-1}\sum_{i=1}^{n} K_h(u_i). \tag{2.8}$$

To ensure that $\hat{f}(x;h)$ is a proper (probability density function) PDF, the kernel K should be chosen to be a unimodal PDF that is symmetric about zero (Chen and Yue, 2014; Mugdadi and Jeter, 2010). There are two issues, however, that practitioners must be aware of when utilizing KDE. First, it is necessary to specify a kernel function to estimate the density. There are several common types of kernel estimators used in practice such as Normal, Epanechnikov, biweight, triweight, triangular, and uniform as disscussed in Shimazaki and Shinomoto (2010) and Terrell and Scott (1992). Figure 2.3 shows some of the common KDE estimators. There is no consensus about what type of kernel is best, but several authors note that the choice of the kernel is not particularly important because there is a trivial loss of efficiency in picking one kernel over the other (Terrell and Scott, 1992) and (Wand and Ripley, 2006).

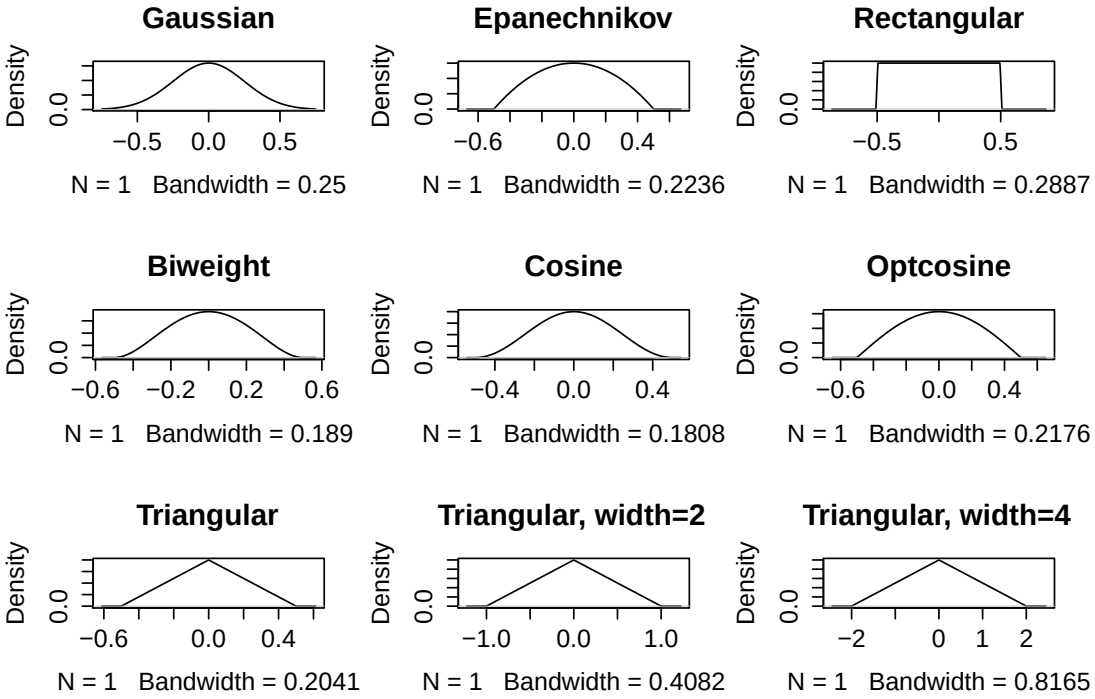

Figure 2.3: Different types of kernel density estimation

2.6.1 Kernel Density Error Criteria

Suitable error criteria must be evaluated to examine the performance of various bandwidth selection methods. There are multiple error criteria that could be used such as mean integrated square error (MISE), mean integrated absolute error $(MIAE)$, mean uniform absolute error $(MUAE)$, and mean Hellinger distance (MHD) (Mugdadi and Jeter, 2010). Most studies, however, have analyzed MISE because it is substantially easier to work with (Jones et al., 1996; Wand and Ripley, 2006). Additionally, Marron and Wand (1992) identified closed form expressions for the exact MISE of Normal mixture densities with a Normal kernel. For these reasons, the present study used the MISE for the error criterion in evaluating bandwidth selection performance.

A discrepancy measure between $f(x)$ and $\hat{f}(x;h)$ at a specific point is the mean squared

error (MSE). The MSE can be expressed as

$$MSE_x\hat{f}(x;h) = \mathbb{E}[\hat{f}(x;h) - f(x)]^2, \tag{2.9}$$

which can be stated in terms of the squared bias and variance as

$$\begin{aligned}\mathbb{E}[\hat{f}(x;h) - f(x)]^2 &= \mathbb{E}[(\hat{f}(x;h) - u_1 + u_1 - f(x))^2] \\ &= \mathbb{E}[(\hat{f}(x;h) - u_1)^2] + 2\mathbb{E}[(\hat{f}(x;h) - u_1)(u_1 - f(x))] + \mathbb{E}[(u_1 - f(x))^2] \\ &= [u_1 - f(x)]^2 + Var[\hat{f}(x;h)] = Bias[\hat{f}(x;h)]^2 + Var[\hat{f}(x;h)],\end{aligned} \tag{2.10}$$

where $u_1 = \mathbb{E}[(\hat{f}(x;h)]$ (here u_1 refers to the mean and not the second moment of the distribution as mentioned previously) (Wand and Jones, 1994). Since Equation (2.10) only calculates the discrepancy at a single point, a loss function is needed to assess the error over the real line. Taking the integral of Equation (2.10) over the real line gives the mean integrated square error $MISE$, which is an average global discrepancy criterion. The MISE is given by

$$MISE_{\hat{f}(x;h)} = \mathbb{E}[\int (\hat{f}(x;h) - f(x))^2 dx] \tag{2.11}$$

According to Fubinis Theorem (see Chu and Marron, 1991) , Equation (2.11) is equivalent to

$$MISE_{\hat{f}(x;h)} = \int \{\mathbb{E}[\hat{f}(x;h) - f(x)]\}^2 dx + \int Var[\hat{f}(x;h)]dx \tag{2.12}$$

Taking the expectation of Equation (2.8) gives

$$\mathbb{E}[\hat{f}(x;h)] = \mathbb{E}[K_h(x-X)], \tag{2.13}$$

using the fact that $\mathbb{E} = \int g(x)f(x)dx$, and the definition of convolution $f*g = \int f(x-y)g(y)dy$,

$$[K_h(x-X)] = K_h * g = \int K_h(x-y)g(y)dy, \tag{2.14}$$

The bias can be written as $(K_h * f)(x) - f(x)$ and the variance $(K_h * f)^2(x) + (K_h * f)^2(x)$ giving the MISE as

$$MISE_{\hat{f}(x;h)} = n^{-1} \int (K_h^2 * f)(x) - (K_h * f)^2(x)dx + \int ((K_h * f)(x) - f(x))^2 dx \tag{2.15}$$

After some manipulation (Rodchuen, 2010; Jeter, 2005), $MISE_{\hat{f}(x;h)}$ can be expressed as

$$MISE_{\hat{f}(x;h)} = (nh)^{-1} \int K_h^2 dx + (1-n^{-1}) \int (K_h * f)^2(x)dx - 2\int (K_h * f)(x)f(x)dx + \int f^2(x)dx. \tag{2.16}$$

Using Equation (2.16) to measure the performance of a bandwidth selection method is straightforward; however, its dependence on the bandwidth h is complex.

To understand the relationship between bandwidth and MISE, an asymptotic approximation to the MISE is used, called the asymptotic mean integrated squared error (AMISE). Using Equation (2.14) and a change of variables by letting $t = (x-y)/h$ the Jacobian of t is h, then $\mathbb{E}[\hat{f}(x;h)] = \int K(t)f(x-ht)dt$. Using a second order Tay-

lor expansion for $f(x-ht)$ gives

$$f(x-ht) = f(x) - htf'(x) + \frac{1}{2}h^2t^2f''(x) + O(h^2)$$

which leads to

$$\begin{aligned}
\mathbb{E}[\hat{f}(x;h)] &= \int \{f(x) - htf'(x) + \frac{1}{2}h^2t^2f''(x) + O(h^2)\}K(t)dt \\
\mathbb{E}[\hat{f}(x;h)] &= \int f(x)dt + \int \frac{1}{2}h^2f''(x)t^2K(t)dt + \int O(h^2)dt \\
\mathbb{E}[\hat{f}(x;h)] &= f(x) + \frac{1}{2}h^2f''(x)\int t^2K(t)dt + O(h^2) \\
\mathbb{E}[\hat{f}(x;h)] - f(x) &= \frac{1}{2}h^2f''(x)\int t^2K(t)dt + O(h^2) \\
Bias_{\hat{f}(x;h)} &= \frac{1}{2}h^2f''(x)\int t^2K(t)dt + O(h^2) \\
Bias_{\hat{f}(x;h)} &\approx \frac{1}{2}h^2u_2(K)f''(x) + O(h^2). \qquad (2.17)
\end{aligned}$$

Note that the symbol $O(h^2)$ indicates that there exists a constant $c > 0$ such that as h approaches zero then the higher order terms in the Taylor expansion remain bounded by ch^4. Using Equation (2.15):

$n^{-1}(K_h^2 * f)(x) - (K_h * f)^2(x) = (nh)^{-1}\int K^2(t)f(x-th)dt - n^{-1}\int K(t)f(x-ht)dt$,

the variance can be approximated via a first order Taylor expansion (see Wolter, 2007, pg. 231 for reasoning) as

$$\begin{aligned}
n^{-1}(K_h^2 * f)(x) - (K_h * f)^2(x) &= (nh)^{-1}\int K^2(t)f(x-ht)dt - (n)^{-1}\int K(t)f(x-ht)dt \\
&= (nh)^{-1}\int K^2(t)\{f(x) + o(1)\}dt - (n)^{-1}\int K(t)\{f(x) + o(1)\}dt \\
&= \frac{1}{nh}R(K)f(x) + o(\frac{1}{nh}) \qquad (2.18)
\end{aligned}$$