

DOCTORAL THESIS

Different-based methods in nonparametric regression models

Dai, Wenlin

Date of Award:
2014

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

This thesis develops some new difference-based methods for nonparametric regression models.

The first part of this thesis focuses on the variance estimation for nonparametric models with various settings. In Chapter 2, a unified framework of variance estimator is proposed for a model with smooth mean function. This framework combines the higher order difference sequence with least squares method and greatly extends the literature, including most of existing methods as special cases. We derive the asymptotic mean squared errors and make both theoretical and numerical comparison for various estimators within the system. Based on the dramatic interaction of ordinary difference sequences and least squares method, we eventually find a uniformly satisfactory estimator for all the settings, solving the challenging problem of sequence selection. In Chapter 3, three methods are developed for the variance estimation in the repeated measurement setting. Both their asymptotic properties and finite sample performance are explored. The sequencing method is shown to be the most adaptive while the sample variance method and the partitioning method are shown to outperform in certain cases. In Chapter 4, we propose a pairwise regression method for estimating the residual variance. Specifically, we regress the squared difference between observations on the squared distance between design points, and then estimate the residual variance as the intercept. Unlike most existing difference-based estimators that require a smooth regression function, our method applies to regression models with jump discontinuities. And it also applies to the situations where the design points are unequally spaced.

The smoothness assumption of the nonparametric regression function is quite critical for the curve fitting and the residual variance estimation. The second part (Chapter 5) concentrates on the discontinuities detection for the mean function. In particular, we revisit the difference-based method in Müller and Stadtmüller (1999) and propose to improve it. To achieve the goal, we first reveal that their method is less efficient due to the inappropriate choice of the response variable in their linear

regression model. We then propose a new regression model for estimating the residual variance and the total amount of discontinuities simultaneously. In both theory and simulations, we show that the proposed variance estimator has a smaller MSE compared to their estimator, whereas the efficiency of the estimators for the total amount of discontinuities remain unchanged. Finally, we construct a new test procedure for detection using the newly proposed estimations; and via simulation studies, we demonstrate that our new test procedure outperforms the existing one in most settings.

At the beginning of Chapter 6, a series of new difference sequences is defined to complete the span between the optimal sequence and the ordinary sequence. The variance estimators using proposed sequences are shown to be quite robust and achieve smallest mean square errors for most of general settings. Then, the difference-based methods for variance function estimation are generally discussed.

Keywords: Asymptotic normality, Difference-based estimator, Difference sequence, Jump point, Least square, Nonparametric regression, Pairwise regression, Repeated measurement, Residual variance

Acknowledgements

I would like to express my greatest gratitude to my supervisor, Dr. Tong Tiejun, for his patient guidance, encouragement and advice throughout my Ph.D study. I am so lucky to have a supervisor who cares so much about my work, and who responds to my questions and queries so promptly. This thesis would not have been possible without his constant support.

I would also like to thank all the faculty and staff in the Department of Mathematics for providing me such a supportive environment. In particular, my sincere thanks go to Professor Zhu Lixing and Dr. Peng Heng, for their invaluable advices and stimulating comments. Thanks also go to my friends who have encouraged me to pursuit my dream when I was lost, comforted me when I got dispirited, and accompanied me when I felt lonely. In addition, I would like to thank the following graduate students in the Department of Mathematics: Wang Tao, Chen Chi, Dong Kai, Guo Xu, Zhu Xuehu, Wu Qin, and also my football teammates for making my Ph.D life full of fun and cherished memories.

Last but not least, I would like to thank: my parents, Dai Zhiqiang and Wu Yufeng, for giving birth to me at the first place and for supporting me spiritually throughout my life; my sister, Dai Wenjing, for accompanying our parents when I was absent; and my wife, Zhang Xinying, for her constant love, understanding and accompany. And most importantly, thank her for promoting me to be a daddy. I owe her everything!

Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	x
Chapter 1 Introduction	1
1.1 Nonparametric Regression Model	1
1.2 Residual-based Method	2
1.3 Difference-based Method	3
1.4 Overall Structure	4
Chapter 2 A Unified Framework for Variance Estimation in Nonpara- metric Regression	6
2.1 Introduction	6
2.2 A unified framework	10
2.2.1 Methodology	10
2.2.2 Optimal Estimator	12
2.3 Theoretical results	13
2.3.1 Asymptotic Variance and Bias	13

2.3.2	Comparison	15
2.4	Simulation studies	18
2.4.1	Selection of Tuning Parameters	18
2.4.2	Simulation Study	20
2.4.3	Robustness of the method	23
2.5	Real applications	26
2.6	Proofs	29
2.6.1	Proof of Theorem 1	29
2.6.2	Proof of Theorem 2	30
2.6.3	Proof of Theorem 3	31

Chapter 3 Difference-based Variance Estimation in Nonparametric Regression with Repeated Measurements 37

3.1	Introduction	37
3.2	Main Results	39
3.2.1	Sample Variance Method	40
3.2.2	Partitioning Method	41
3.2.3	Sequencing Method	41
3.3	Simulation Studies	49
3.4	Nonparametric Regression with Unbalanced Repeated Measurements 51	
3.4.1	Methodology	54
3.4.2	A Simulation Study	55
3.5	Real Application	56
3.6	Conclusion	59
3.7	Proofs	60
3.7.1	Proof of Theorem 4	60
3.7.2	Proof of Theorem 5	61
3.7.3	Proof of Theorem 6	66
3.7.4	Proof of Theorem 7	68

Chapter 4	Pairwise Method for Variance Estimation	71
4.1	Introduction	71
4.2	Main Results	74
4.2.1	Difference-based Estimators	74
4.2.2	Pairwise Regression	75
4.2.3	Adaptive Pairwise Regression	77
4.3	Simulation Studies	81
4.3.1	Equidistant Design	81
4.3.2	Non-equidistant Design	82
4.3.3	Simulation Results	82
4.4	Real Application	83
4.5	Discussion	84
Chapter 5	Testing Discontinuities in Nonparametric Regression	89
5.1	Introduction	89
5.2	Main Results	91
5.3	Asymptotic Properties	93
5.4	Simulation Studies	95
5.5	Proofs	97
5.5.1	Proof of Theorem 8	99
5.5.2	Proof of Theorem 9	103
Chapter 6	Future Work	110
6.1	Optimal- p Difference Sequence	110
6.1.1	Definition of New Sequences	111
6.1.2	Properties of Optimal- p Sequence	114
6.2	Difference-based Variance Function Estimation	116
Bibliography		117
Curriculum Vitae		127