

## DOCTORAL THESIS

### Variable selection in high dimensional semi-varying coefficient models

Chen, Chi

*Date of Award:*  
2013

[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

# Variable Selection in High Dimensional Semi-Varying Coefficient Models

**CHEN Chi**

A thesis submitted in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

Principal Supervisor: Dr. PENG Heng

Hong Kong Baptist University

Aug 2013

# Abstract

With the development of computing and sampling technologies, high dimensionality has become an important characteristic of commonly used science data, such as some data from bioinformatics, information engineering, and the social sciences. The varying coefficient model is a flexible and powerful statistical model for exploring dynamic patterns in many scientific areas. It is a natural extension of classical parametric models with good interpretability, and is becoming increasingly popular in data analysis. The main objective of this thesis is to apply the varying coefficient model to analyze high dimensional data, and to investigate the properties of regularization methods for high-dimensional varying coefficient models.

We first discuss how to apply local polynomial smoothing and the smoothly clipped absolute deviation (SCAD) penalized methods to estimate varying coefficient models when the dimension of the model is diverging with the sample size. Based on the nonconcave penalized method and local polynomial smoothing, we suggest a regularization method to select significant variables from the model and estimate the corresponding coefficient functions simultaneously. Importantly, our proposed method can also identify constant coefficients at the same time. We investigate the asymptotic properties of our proposed method and show that it has the so called “oracle property.”

We apply the nonparametric independence Screening (NIS) method to varying coefficient models with ultra-high-dimensional data. Based on the marginal varying coefficient model estimation, we establish the sure independent screening property under some regular conditions for our proposed sure screening method. Combined with our proposed regularization method, we can systematically deal with high-dimensional or ultra-high-dimensional data using varying coefficient models.

The nonconcave penalized method is a very effective variable selection method. However, maximizing such a penalized likelihood function is computationally challenging, because the objective functions are nondifferentiable and nonconcave. The local linear approximation (LLA) and local quadratic approximation (LQA) are two popular algorithms for dealing with such optimal problems. In this thesis, we revisit these two algorithms. We investigate the convergence rate of LLA and show that the rate is linear. We also study the statistical properties of the one-step estimate based on LLA under a generalized statistical model with a diverging number of dimensions. We suggest a modified version of LQA to overcome its drawback under high dimensional models. Our proposed method avoids having to calculate the inverse of the Hessian matrix in the modified Newton Raphson algorithm based on LQA.

Our proposed methods are investigated by numerical studies and in a real case study in Chapter 5.

# Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	vi
List of Tables	vii
Chapter 1 Introduction	1
1.1 High Dimension Regression Models . . . . .	1
1.2 Variable Selection via Penalty . . . . .	3
1.3 Numerical Algorithms for Nonconcave Penalized Methods . . . . .	7
1.4 Sure Independence Screening in Sparse Ultra-high-dimensional models	11
1.5 Varying Coefficient Models . . . . .	13
1.6 Organization of the Dissertation . . . . .	19
Chapter 2 Model Selection in Semivarying Coefficient Models with Diverging Number of Variables	21
2.1 Introduction . . . . .	21
2.2 Regularized Estimation for Semivarying Coefficient Models By Local Linear Fitting . . . . .	24
2.3 Asymptotical Properties of the Proposed Estimation . . . . .	26
2.4 Practical Issues . . . . .	31

2.5	Appendix: Proofs . . . . .	33
Chapter 3 Independent Sure Screening in Ultra-High-Dimensional Semi-varying		
	Coefficient Models . . . . .	43
3.1	Introduction . . . . .	43
3.2	NIS in Varying Coefficient Models . . . . .	44
3.3	Asymptotic Sure Screening Properties of NIS . . . . .	46
3.4	Practical Issues . . . . .	50
3.5	Appendix: Proofs . . . . .	52
Chapter 4 Revisit Local Linear and Quadratic Approximation for Nonconcave		
	Penalized Methods . . . . .	61
4.1	Introduction . . . . .	61
4.2	Revisiting the LLA for Nonconcave Penalized High Dimensional Like- lihoods . . . . .	63
4.3	Revisiting the LQA Algorithm . . . . .	69
4.4	Proofs . . . . .	72
Chapter 5 Numerical Study of High-dimensional Varying Coefficient Models 79		
5.1	Implementation of Group-SCAD Method . . . . .	79
	5.1.1 Simulation Examples . . . . .	81
5.2	Numerical Study for Ultra-high-dimensional Varying Coefficient Models 87	
	5.2.1 Implementation of NIS-Group-SCAD Method . . . . .	87
	5.2.2 Simulation Example . . . . .	91
5.3	Real Case study . . . . .	92
5.4	Numerical Study of Penalized Spline . . . . .	96
	5.4.1 Implementation of some algorithms . . . . .	96
	5.4.2 Simulation Examples . . . . .	98
Chapter 6 Conclusion and Discussion 100		