SEMIPARAMETRIC MODELS FOR NONLINEAR SPATIO-TEMPORAL DATA WITH APPLICATION TO THE UNITED STATES HOUSING PRICES INDEXES

Dawlah Alsulami

Thesis submitted for the degree of Doctor of Philosophy

in

Statistics

at

The University of Adelaide

(Faculty of Mathematical and Computer Sciences)

Department of Applied Mathematics

December 14, 2015

SIGNED STATEMENT

I consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED: .... DATE: ....

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968. The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

ACKNOWLEDGMENTS

First and foremost, tremendous gratitude is due to my supervisor Professor Zudi Lu for his excellent guidance, invaluable support and encouragement. Since the start of my thesis, his patience, caring and solicitude are ineffable. I could not imagine having better mentor for my thesis and he will remain an excellent role model for me, thanks to you.

I would also like to prolong my sincerest thanks and gratefulness to Associate Professor Gary Glonek for encouraging my research and for his brilliant comments and indispensable advice. Thanks must also go to Dr. Nicholas Sim for his support.

During my PhD, I have developed relations with supportive colleagues Dr. Michelle Picard from the University of Adelaide and Professor Jun Zhu from the University of Wisconsin-Madison thanks for both of you for being valuable sources of support and knowledge.

My greatest thanks go to my parents because without their encourage, support and prayer, it would not be possible to complete this thesis. Heartfelt thanks go to my beloved Abdullah, who spent sleepless nights for me. Thanks for being more than a husband. My brothers and sisters, thanks for your love, wishes and support. Finally, I would like to thank my sponsor King Abdulaziz University for providing me this chance to complete the PhD.

ABSTRACT

Modelling spatio-temporal data has received significant attention, recently and is widely applied in many disciplines such as economics, environmental and social sciences. In economics, housing price is a real example that indicates the importance of modeling such data. Estimating the movement in housing prices is an important but challenging problem due to the difficulties associated with spatio-temporal interactions. One main challenge is that there is no natural spatial ordering and thus it is not as straightforward as in time series analysis to transform data to be stationary across space. In addition, it is a challenge to model spatio-temporal data collected at irregularly spaced sampling locations due to the potentially large number of parameters. Moreover, the complexity of the dependence structure requires new effective statistical methods for modeling and analysis.

While nonparametric and semiparametric methods have been popular for nonlinear modeling of time series data in econometrics and statistics, they become increasingly challenging when extended to irregularly located spatio-temporal data with complex nonlinear structures. The literature on nonlinear spatio-temporal modelling is still rather rare except a few contributions recently done, for example, by (Lu et al., 2009), (Wikle and Hooten, 2010) and (Wikle and Holan, 2011). Therefore, the main aim of this thesis is to propose a class of semiparametric spatio-temporal autoregressive partially nonlinear regression models as a practical way to overcome these challenges. The main contributions of this thesis are summarised as follows:

(1) In Chapter 2, a class of semiparametric spatio-temporal autoregressive partially nonlinear regression models is proposed. The proposed models not only permit location-varying nonlinear relationships between the response variable and the covariates but also allow for the dependence structure to be nonstationary over space while alleviating the "curse of dimensionality" by using the popular idea of spatial weight matrix measuring the spatial interaction, which is assumed to be well defined as in spatial econometrics. Both the estimation and its finite- and large-sample properties for the proposed models are established.

(2) In order to more objectively let data decide the spatial interaction in the models, in Chapter 3, we propose a scheme of general data-driven models to estimate the spatial weights in the semiparametric spatio-temporal autoregressive partially nonlinear regression models by applying the adaptive lasso. Both estimation and its finite- and large-sample properties for this class of the general data-driven models are developed.

(3) An improved scheme for such data-driven models with spatial weight matrix is presented in Chapter 4. For this class of the improved data-driven models, we develop a computationally feasible method for estimation and thus enable our methodology to be applicable in practice. Asymptotic properties of our proposed estimates are established and comparisons are made, in theory and via simulations, between estimates before and after spatial smoothing.

(4) In empirical case studies, the proposed methodologies are applied to investigate the housing prices in relation to the mortgage rates in Chapter 2 and to the consumer price index in Chapter 3 and 4 for the 50 states and District of Columbia (DC) in the United States (U.S.). It is empirically found that such relationships could be of nonlinear features that help to improve the predictions and the third class of the improved data-driven models appears to work promisingly better.

CONTENTS

LIST OF TABLES

LIST OF FIGURES