

Faculty of Engineering and Information Technology
University of Technology, Sydney

Copula-based High Dimensional Dependence Modelling

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

by

Wei Wei

November 2014

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisors Prof. Longbing Cao for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

I also would like to appreciate my co-supervisor Associate Prof. Jinyan Li for providing me with continuous support throughout my PhD study and research. Without his professional guidance and persistent help, this thesis would not have been possible.

I thank my fellow labmates in Advanced Analytics Institute: Xuhui Fan, Yin Song, Junfu Yin for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three years.

I place on record, my sense of gratitude to the team leader, Zhigang Zheng in Australian Taxation Office for his expert and sincere help in the project. I am also grateful to my team members, Allen Lin, Mu Li, Chunming Liu and Wei Cao, for their hard working in project teams.

Last but not the least, I would like to thank my family: my wife and my parents, for their unconditional support, both financially and emotionally throughout the whole PhD studying.

Wei Wei

June 2014 @ UTS

Contents

Certificate	i
Acknowledgment	iii
List of Figures	xi
List of Tables	xiii
List of Publications	xv
Abstract	xvii
Chapter 1 Introduction	1
1.1 Background	1
1.2 Research Issues	7
1.2.1 RI 1: High Dimensional Data	7
1.2.2 RI 2: Dependence Structure Assumptions	7
1.2.3 RI 3: Dependence Structure Truncation and Optimi- sation	8
1.2.4 RI 4: High Dimensional Dependence Evaluation	8
1.3 Research Contributions	8
1.4 Thesis Structure	10
Chapter 2 Literature Review and Foundation	13
2.1 Copula	13
2.1.1 Elliptical Copulas	15
2.1.2 Archimedean Copulas	16
2.1.3 Bivariate Copula Families	17
2.1.4 The Tail Dependence of Copulas	22

CONTENTS

2.2	Copula-based Models	23
2.2.1	Time-varying Copula Models	24
2.2.2	Estimate and Inference	28
2.2.3	Goodness-of-Fit Testing and Model Selection	34
2.3	Multivariate Copula Models	36
2.3.1	Multivariate Elliptical Copulas	36
2.3.2	Multivariate Archimedean Copulas	37
2.4	Vine Copula Models	41
2.4.1	Pair Copula Construction	41
2.4.2	Graph Theory	44
2.4.3	Regular Vine	47
2.4.4	Vine Tree Structure Selection Strategies	50
2.4.5	Parameter Estimate	52
2.4.6	Copula Families Selection	53
2.4.7	Vine Optimisation	54
2.4.8	Probability Graphical Models and Copula	54
2.5	Model Comparison Test – Vuong Test	57
2.6	Partial Correlation	59
2.6.1	Partial Correlation Definition	59
2.6.2	Partial Correlation and Conditional Correlation	60
2.7	Time Series Analysis	63
2.7.1	Basic Definition	63
2.7.2	The ARMA Models	65
2.7.3	The GARCH Models	66
2.7.4	The ARMA(1,1)-GARCH(1,1) Model	67
2.7.5	The Ljung Box Test	68
2.8	Applications in Financial Market	69
2.8.1	Risk Managements	69
2.8.2	Derivative Contracts	70
2.8.3	Portfolio Decision Problems	71
2.8.4	Financial Contagion	71

2.8.5	Regular Vine Copula Applications	72
2.8.6	Other Applications	72
2.9	Summary	73

**Chapter 3 Modelling the Complex Dependence Structures
by Using Truncated Partial Canonical Vine 75**

3.1	Introduction	75
3.2	Canonical Vine Autoregressive Model	80
3.3	Partial Canonical Vine Model	82
3.3.1	Canonical Vine	83
3.3.2	Partial Canonical Vine Specification	87
3.3.3	Dependence Structure Construction	88
3.3.4	Copula Families Selection	91
3.3.5	Vine Tree Structure Truncation and Simplification	91
3.3.6	Parameter Estimate	94
3.4	Marginal Models Specification	96
3.4.1	ARMA-GARCH Model	96
3.4.2	Error Distribution	97
3.5	Portfolio of Value at Risk	98
3.6	Dependence Analysis on European Stock Returns	101
3.6.1	Data and Marginal Model Specification	101
3.6.2	The Truncation and Simplification Analysis of Partial Canonical Vine	104
3.6.3	Experiments on Value at Risk	105
3.7	Summary	108

**Chapter 4 Modeling Asymmetry and Tail Dependence among
Multiple Variables by Using Partial Regular Vine 116**

4.1	Introduction	116
4.2	Regular Vine Copula Model with Asymmetric Dependence	119
4.2.1	Regular Vine Copula Specification	119
4.2.2	Regular Vine Tree Structure Construction	120

4.2.3	Bivariate Copula Family Selection	123
4.3	Marginal Distribution Specification and Parameter Estimate	124
4.4	Value at Risk– A widely Used Evaluation in Financial Market	126
4.5	Case Study	128
4.5.1	Non-parametric Dependence Analysis	128
4.5.2	Regular Vine Copula Specification and Tail Dependence Analysis	129
4.5.3	Value at Risk Forecasting	133
4.6	Summary	134
Chapter 5	High-dimensional Dependence Modelling through Truncated Partial Regular Vine Copula	145
5.1	Introduction	145
5.2	Truncated Partial Regular Vine	150
5.2.1	Partial Regular Vine Specification	150
5.2.2	Partial Regular Vine Tree Structure Construction	151
5.2.3	Partial Regular Vine Tree Structure Truncation	155
5.2.4	Bivariate Copula Selection	156
5.2.5	Marginal Distribution Specification and Parameter Estimation	158
5.3	Evaluation Methods	159
5.3.1	In-Sample Evaluation Methods	160
5.3.2	Out-of-Sample Evaluation Method	160
5.4	Experimental Results	162
5.4.1	Data and Marginal Distributions Specification	162
5.4.2	Partial Regular Vine Truncation Analysis	164
5.4.3	In-Sample Performance Analysis	166
5.4.4	Out-of-Sample Performance Analysis	167
5.5	Summary	169
Chapter 6	Optimal Allocation of High Dimensional Assets through Truncated Partial Canonical Vines	174

6.1	Introduction	174
6.2	The Portfolio Optimisation	177
6.2.1	CRRA Optimisation Function	177
6.2.2	Truncated Partial Canonical Vine	178
6.2.3	Marginal Distribution and Estimate	180
6.3	Evaluation Methods	181
6.4	Case Study	182
6.4.1	Data and Model Specification	182
6.4.2	Experiment Results and Analysis	184
6.5	Summary	186
Chapter 7 Conclusions and Future Work		193
7.1	Conclusions	193
7.2	Future Work	195
Appendix A Dependence Measurements		197
A.1	Pearson Correlation	197
A.2	Kendall's Tau	198
Appendix B The Determinant of Partial Correlation-based Canonical Vine and Regular Vine		200
Appendix C Model Selection Criteria		203
Appendix D List of Symbols		204
Bibliography		206

List of Figures

1.1	The scatter plot of copulas	3
1.2	The profile of work in this thesis	12
2.1	Fully nested Archimedean construction	38
2.2	Partially nested Archimedean construction	39
2.3	Hierarchically nested Archimedean construction	40
2.4	An example graph with 5 nodes and 5 edges	46
2.5	An example graph with a cycle	47
2.6	An example tree with 5 nodes and 4 edges	47
3.1	An example of dependence structure	76
3.2	CAVA dependence structure	81
3.3	PCV model work flow	82
3.4	PCV dependence structure	84
3.5	The canonical vine dependence structure with 7 variables	86
3.6	The results of truncated and simplified partial canonical vine against ρ_{sign}	114
3.7	The VaR forecasting with portfolio returns	115
4.1	The partial regular vine tree structure	131
4.2	The tail dependence of v_4 and v_6	137
4.3	The tail dependence of v_4 and v_5	138
4.4	The tail dependence of v_4 and v_8	139
4.5	The tail dependence of v_4 and v_7	140

LIST OF FIGURES

4.6	The tail dependence of v_2 and v_7	141
4.7	The tail dependence of v_2 and v_3	142
4.8	The tail dependence of v_1 and v_2	143
4.9	The VaR forecasting of portfolio returns	144
5.1	Performance analysis of truncated partial regular vines	172
5.2	The VaR forecast of portfolio returns by using $TPRV_{0.1}$	173
6.1	The truncated and simplified vine for $S\&P500$	190
6.2	The truncated and simplified vine for $Stoxx50E$	191
6.3	The portfolio value results analysis	192

List of Tables

2.1	The Tail Dependence of the Bivariate Copula Family	24
3.1	Indices and Stocks	102
3.2	Descriptive Statistics for the Indices	103
3.3	Results of Ljung Box Test	104
3.4	Determinants and Numbers of Pair Copulas	106
3.5	Truncated and Simplified Partial Canonical Vine Analysis . . .	109
3.6	Number of Exceedances of VaR Forecast	112
3.7	The Results of VaR Backtesting	113
4.1	The Tail Dependence of Copula Family	123
4.2	Non-parametric Tail Dependence Analysis	129
4.3	Results of BP and LP Tests	130
4.4	Tail Dependence Analysis by Using Various Copula	135
4.5	The Backtesting Results of Value at Risk Forecasting	136
5.1	The Tail Dependence of Bivariate Copulas	157
5.2	Ljung-Box Test Results	163
5.3	The Total Number of Edges in Truncated Partial Regular Vine against Truncation Value	165
5.4	The Comparison between $TPRV_{0.1}$ and PRV	166
5.5	Results of Bivariate Copula Selection and Model Selection . .	167
5.6	Vuong Test Results	168
5.7	Backtesting Results of Value at Risk for Currencies	170

LIST OF TABLES

5.8	Backtesting Results of Value at Risk for Currencies	171
6.1	The Partial Correlation Analysis	183
6.2	The Comparison between Truncated, Simplified and Non-truncated Vines	184
6.3	The Results of the Opportunity Costs, Final Amounts and Utilities for <i>S&P500</i>	188
6.4	The Results of the Opportunity Costs, Final Amounts and Utilities for <i>Stoxx50E</i>	189

List of Publications

Papers Published

- **Wei Wei**, Junfu Yin, Jinyan Li, Longbing Cao (2014), Modelling Asymmetry and Tail Dependence among Multiple Variables by Using Partial Regular Vine. *in* 'Proceedings of the SIAM International Conference on Data Mining Conference (**SDM14**)', pp.776-784.
- Junfu Yin, Zhigang Zheng, Longbing Cao, Yin Song, **Wei Wei** (2013), Efficiently Mining Top-K High Utility Sequential Patterns. *in* 'Proceedings of the 13th IEEE International Conference on Data Mining (**ICDM13**)', pp. 1259-1264.
- **Wei Wei**, Jinyan Li, Longbing Cao, Jingguang Sun, Chunming Liu, Mu Li (2013), Optimal Allocation of High Dimensional Assets through Canonical Vines. *in* 'Proceedings of the 17th pacific-asia conference on knowledge discovery and data mining (**PAKDD13**)', pp. 366-377.
- **Wei Wei**, Xuhui Fan, Jinyan Li, Longbing Cao (2012), Model the complex dependence structures of financial variables by using canonical vine. *in* 'Proceedings of the 21st ACM international conference on Information and knowledge management (**CIKM12**)', ACM, pp, 1382-1391.

Papers to be Submitted/Under Review

- Zhigang Zheng, **Wei Wei**, Chunming Liu, Wei Cao, Longbing Cao, and Maninder Bhatia (2014) An Effective Contrast Sequential Pattern Mining Approach on Taxpayer Behavior Analysis, *World Wide Web Journal*.
- **Wei Wei**, Longbing Cao, Jinyan Li, and Philip S. Yu (2014) High-dimensional Dependence Modelling through Truncated Partial Regular Vine Copula, to be submit the 14 th IEEE International Conference on Data Mining (**ICDM14**).

Research Reports of Industry Projects

- **Wei Wei**, Zhigang Zheng, Maninder Bhatia. Self-finaliser analysis (debt cases) of activity statement – data mining modelling and evaluation report , Debt Collection and Optimisation Project, Australian Taxation Office, Dec 2012.
- Zhigang Zheng, **Wei Wei**, Maninder Bhatia. Self-finaliser analysis (due date cases) of activity statement – data mining modelling and evaluation report , Debt Collection and Optimisation Project, Australian Taxation Office, May 2013.
- Zhigang Zheng, **Wei Wei**, Wei Cao, Maninder Bhatia. Action response model of activity statement – data mining modelling and evaluation report, Debt Collection and Optimisation Project, Australian Taxation Office, Nov 2013.
- **Wei Wei**, Zhigang Zheng and Wei Cao. Self-finalising Lodgment Analysis Project – data mining modelling and evaluation report, Lodge-ments Collection and Optimisation Project, Australian Taxation Office, May 2014.

Abstract

Big data applications increasingly involve high-dimensional and sophisticated dependence structures in complex data. Modelling high-dimensional dependence, that is, the dependence between a set of high-dimensional variables, is a critical but challenging issue in many applications including social media analysis and financial markets. A typical example concerns the interplay of financial variables involved in driving complex market movements. A particular problem is understanding the dependence between high-dimensional variables with tail dependence and asymmetric characteristics which appear widely in financial markets. Typically, existing methods, such as the Bayesian logic program, relational dependency networks and relational Markov networks, build a graph to represent the conditional dependence structure between random variables. These models aim at high-dimensional domains, and have the advantage of learning latent relationships from data. However, they tend to force the local quantitative part of the model to take a simple form such as the discretized form of the data when multivariate Gaussian or its mixtures cannot capture the data in the real world. The complex dependencies between high-dimensional variables are difficult to capture.

In statistics and finance, the copula has been shown to be a powerful tool for modelling high-dimensional dependencies. The copula splits the multivariate marginal distributions from dependence structures, so that the specification of dependence structures can be investigated independently of the marginal distributions. It can provide a flexible mechanism for modelling real world distributions that cannot be handled well by graphical models. Thus,

researchers have tried to combine copula and probability graphical models, such as the tree-structured copula model and copula Bayesian networks. These copula-based models aim to resolve the limitations of discretizing data, but they impose assumptions and restrictions on the dependence structure. These assumptions and restrictions are not appropriate for dependence modelling among financial variables.

In order to address these research limitations and challenges, this thesis proposes the use of the truncated partial correlation-based canonical vine copula, partial correlation-based regular vine copula and truncated partial correlation-based regular vine copula to model the dependence of high-dimensional variables. Chapter 3 introduces a new partial correlation-based canonical vine to identify the asymmetric and non-linear dependence structures of asset returns without any prior dependence assumptions. To simplify the model while maintaining its merit, a partial correlation-based truncation method is proposed to truncate the canonical vine. The truncated partial correlation-based canonical vine copula is then applied to construct and analyse the dependence structures of European stocks as a case study.

Chapter 4 introduces the truncated partial correlation-based regular vine copula to explore the relations in multiple variables. Very often, strong restrictions are applied on a dependence structure by existing high-dimensional dependence models. These restrictions disabled the detection of sophisticated structures such as the upper and lower tail dependence between multiple variables. A partial correlation-based regular vine copula model may relax these restrictions. The partial correlation-based regular vine copula model employs a partial correlation to construct the regular vine structure, which is algebraically independent. This model is able to capture the asymmetric characteristics among multiple variables by using a two-parametric copula with flexible lower and upper tail dependence. The method is tested on a cross-country stock market data set to analyse the asymmetry and tail dependence in the dynamic period.

Chapter 5 proposes a novel truncated partial correlation-based regular

vine copula model which can capture more flexible dependence structures without making pre-assumptions about the data. Specifically, the model employs a new partial correlation to build the dependence structures via a bottom-up strategy. It can identify important dependencies and information among high-dimensional variables, truncating the irrelevant information to significantly reduce the parameter estimate time. The in-sample and out-of-sample performance of the model are examined by using the data in currency markets over a period of 17 years.

Chapter 6 discusses how to resolve the high-dimensional asset allocation problem through a partial correlation-based canonical vine. Typically, the mean-variance criteria which is widely used in asset allocation, is actually not the optimal solution for asset allocation as the joint distribution of asset returns are distributed in asymmetric ways rather than in the assumed normal distribution. The partial correlation-based canonical vine can resolve the issue by producing the asymmetric joint distribution of asset returns in the utility function. Then, the utility function is then used for determining the optimal allocation of the assets. The performance of the model is examined by using data in both European and United State stock markets.

In summary, this thesis proposes three dependence models, including one canonical vine and two regular vines. The three dependence models, which do not impose any dependence assumption on the dependence structure, can be used for modelling different high-dimensional dependencies, such as asymmetry or tail dependencies. All of these models are examined by the datasets in the real world, such as stock or currency markets. In addition, the partial correlation-based canonical vine is used to resolve optimisation allocation of assets in stock markets. This thesis works to show that there is great potential in applying copula to model complex dependence, particular in modelling time-varying parameters, or in developing efficient vine copula simplification methods.

