

JIYU TUMOXING FANGFA DE
GUSHI XIANGGUANXING DINGLIANG YANJIU

基于图模型方法的 股市相关性定量研究

蔡风景 著



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

本书获温州大学资助

F 830.81
322

基于图模型方法的 股市相关性定量研究

JIYU TUMOXING FANGFA DE
GUSHI XIANGGUANXING DINGLIANG YANJIU

蔡风景 著



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

图书在版编目(CIP)数据

基于图模型方法的股市相关性定量研究 / 蔡风景著.
—杭州: 浙江工商大学出版社, 2015. 4

ISBN 978-7-5178-1055-1

I. ①基… II. ①蔡… III. ①股票市场—研究 IV.
①F830.91

中国版本图书馆 CIP 数据核字(2015)第 088319 号

基于图模型方法的股市相关性定量研究

蔡风景 著

责任编辑 白小平

封面设计 王妤驰

责任印制 包建辉

出版发行 浙江工商大学出版社

(杭州市教工路 198 号 邮政编码 310012)

(E-mail: zjgsupress@163.com)

(网址: <http://www.zjgsupress.com>)

电话: 0571-88904980, 88831806(传真)

排 版 杭州朝曦图文设计有限公司

印 刷 虎彩印艺股份有限公司

开 本 710mm×1000mm 1/16

印 张 18.25

字 数 300 千

版 印 次 2015 年 4 月第 1 版 2015 年 4 月第 1 次印刷

书 号 ISBN 978-7-5178-1055-1

定 价 45.00 元

版权所有 翻印必究 印装差错 负责调换

浙江工商大学出版社营销部邮购电话 0571-88904970

前 言

受经济全球化、金融一体化、金融创新、信息技术和竞争等因素的影响,全球金融市场正发生着根本性的结构变化,金融市场间的相互依赖、相互影响与日俱增,这促使了对金融市场间的相关性如相关程度、协同运动和传导等问题的研究。

股票市场作为金融市场的重要组成部分,其一体化正越来越受到人们的关注。关于各国股票市场间的联系与相关性,学者已进行了大量的研究,各国股市间相关程度的增大已是不争的事实,而且越是股市处于大幅振荡期,各国股市间的相关程度也越大。同时,股票市场风险不仅能演变为资本市场的整体风险,而且还可能引发全面的金融危机和经济危机并危及社会生活的安定,甚至引发国际性危机,其波及范围和后果极其严重,近年来频发的金融危机就是最好的例证。

为了分散投资行为带来的风险,以期达到最佳效用,我们需要对许多不同的资产进行组合,从而产生了投资组合理论。而相关性分析在投资组合理论中处在一个基础性地位,事实上,不仅仅是在投资组合理论中,在涉及多变量金融序列分析的许多场合,比如资产定价、资产选择以及风险管理等问题中,我们都离不开相关性分析。因此,对于金融资产间相关性,尤其是股市的相关性研究,涉及了现代金融理论的方方面面,是一个非常有意义的话题。

目前,股市相关性研究最流行的方法是多元 GARCH 模型,但该模型比较合适处理低维数据,当变量维数较高时,参数过多的弱点使其应用受到很大的限制。图模型为多维(时间序列)随机变量提供了强有力的统计分析工具。图模型方法通过马尔可夫性质利用图上各点之间直观的分

观的工具来推断变量间的条件独立性,可以把一些复杂的高维系统分解成若干简单的部分,对各个子系统进行分别处理,从而将问题简化。在国际股票市场系统中,涉及的变量众多,通过图模型方法从统计方法上不仅可减少待估的参数,节约计算成本,而且可降低问题的复杂性,减少因变量过多给我们带来的一系列问题,提高工作的成效。该方法不仅能直观地描述变量间复杂的关系,而且刻画多维时间序列变量的相关结构时具有相当大的灵活性,能反映多元时间序列的动态依赖关系。通过图模型理论,可揭示国内和国际主要股市之间存在的动态时序关系、相关关系甚至因果关系,有助于我们深入理解和推动国内和世界金融市场的发展,对于分析与研究股市的结构和判断股市的走势及风险传递无疑具有重要的作用。

本书共分为八章。第一章绪论,主要介绍了图模型和股市相关性研究的主要方法,对已有文献进行综述。第二章基于图模型理论的股市信息传导分析,在介绍图模型基本概念和时间序列图模型理论基础上,利用多维时间序列的链图、偏相关图和 Granger 因果图分析股市间的信息传导。第三章基于图模型方法的股市时序相关性研究,将图模型方法应用于单变量时间序列的时序相关性分析,提出经典的时间序列模型,如 MA、ARMA 和 ARCH 等系数显著性检验的新方法,且应用于我国上证综合指数,分析股指收益率的异方差现象。第四章基于有向非循环图方法的我国股指收益率的传导研究,介绍有向非循环图(DAG)理论及 PC 算法,且将该方法应用于我国股票市场。建立我国主要股票指数的结构 VAR 模型,通过检验残差项的偏相关系数,给出主要股票指数的同期相关和因果关系,且利用脉冲响应函数和方差分解给出股票收益率间的动态因果关系和强度,分析我国内地和香港股指的动态传导和反馈机制,系统给出我国股票市场的联动效应。第五章基于体制转换模型的我国行业指数收益率的动态相关关系研究,提出基于体制转换的图模型理论,将我国证券市场行业板块的图结构划分为不同的状态,描述不同状态的条件相关关系,利用蒙特卡洛模拟方法给出图结构参数、状态和转移概率的后验分布,且以概率的方式测定时间可变的行业板块图结构,计算行业指数时变的 beta 系数。第六章基于图结构和 Copula 方法的股市收益率尾部相关性分析,提出基于图结构的 Copula 理论,通过合并条件独立性减少

Copula 方法估计的参数。重点介绍特殊的藤和非循环有向图结构,且将该模型应用于国际股市,描述股市不对称的尾部相关性和局部相关特征。第七章基于状态空间图模型方法的投资组合优化决策,利用 Lasso 图模型理论分析欧洲主要股指的同期条件相关关系,在结合状态空间模型给出残差协方差阵估计的基础上,对原有的 EM 算法进行改进,且将该方法应用于欧洲主要股票指数,进行投资组合优化决策分析。第八章基于稀疏 SUR 模型的系统风险度量,通过对 SUR 模型残差协方差阵的图模型结构分析,提出稀疏的 SUR 模型,并设计 MCMC 算法,最后将该方法应用于我国深圳股票市场度量行业指数的系统风险。

值此书稿完成之际,笔者向南京河海大学王慧敏教授表示由衷的感谢和敬意!王慧敏老师不仅给我植入了管理科学的思维,而且在我感觉迷茫时给了我莫大的信心和勇气。感谢广州大学的李元教授,他引导我进入了统计学新兴的领域:图模型,从而开启了我的科研之路。此外,我还要感谢温州大学数学与信息科学学院的领导高利新院长、全力书记、洪振杰和应裕林副院长给予我生活和工作上的关照和帮助,让我有足够的时间和精力完成书稿。此外,还要感谢我的妻子南赛月女士,她默默地奉献,给予我支持和理解。

囿于知识水平的有限和时间的不足,笔者对于图模型方法在股市相关性应用研究中的理论、模型、算法与实证分析的把握和理解都非常有限,因此,本书难免会存在着错误和遗漏之处,恳请读者给予批评指正。随着大数据时代的来临,金融互联网大数据挖掘已成为热点问题。通过图模型方法能较好地处理高维不确定性和混杂因素,发挥其可视化优势,有效地提取大数据的价值信息,推动大数据挖掘技术的发展。我相信,基于图模型方法的金融大数据挖掘应该大有可为,也希望本书的出版能为大数据分析方法的的发展起到一定的推动作用。

蔡风景

2015 年 1 月

目 录

第一章 绪 论	001
1.1 背景	001
1.2 相关理论的国内外研究进展	003
第二章 基于图模型理论的股市信息传导分析	017
2.1 股市相关性研究概念界定	017
2.2 图模型基本理论	019
2.3 基于链图和偏相关图方法的股市信息传递分析	029
2.4 基于 Granger 因果图方法的股市信息传递分析	036
2.5 本章小结	046
第三章 基于图模型方法的股市时序相关性研究	048
3.1 股市时序相关性简析	048
3.2 基于图方法的滑动平均和自回归滑动平均模型时序相关性 分析	050
3.3 基于图方法的股市收益率的条件异方差研究	056
3.4 本章小结	061
第四章 基于有向非循环图方法的我国股指收益率的传导研究	063
4.1 我国股票指数收益率联动分析的理论探讨	064
4.2 有向非循环图方法理论	066
4.3 我国股票指数收益率的动态传导分析	072
4.4 本章小结	103

第五章 基于体制转换模型的我国行业指数收益率动态相关关系

研究 104

5.1 股市相关性结构变化问题描述 104

5.2 动态相关系数的体制转换模型 106

5.3 基于体制转换图模型方法的动态相关关系分析 124

5.4 本章小结 138

第六章 基于图结构和 Copula 方法的股市收益率尾部相关性分析

..... 140

6.1 Copula 理论及尾部相关性 141

6.2 基于 Copula 函数的我国股市收益率和成交量的尾部相关性分析 147

6.3 Copula 图结构理论及分解 154

6.4 基于图结构和 Copula 方法的亚洲主要股指收益率尾部相关性分析 158

6.5 基于 DAG 的 Pair-Copula 分解方法及其股市相关性应用 163

6.6 本章小结 171

第七章 基于状态空间图模型方法的投资组合优化决策 173

7.1 证券投资组合简析 173

7.2 Lasso 图模型理论 174

7.3 基于 Lasso 图理论和状态空间模型的投资组合优化决策 178

7.4 本章小结 191

第八章 基于 SUR 图模型的我国行业指数系统风险度量 192

8.1 似无关回归模型 192

8.2 算法设计 194

8.3 数值模拟 195

8.4 我国深证行业股票指数的系统风险度量 197

8.5 本章小结 201

附 录 部分 Matlab 程序	203
附录 1 第二程序	203
附录 2 第四章 DAG 识别及方差分解图(R 语言程序)	210
附录 3 第五程序代码	213
附录 4 第六程序代码	224
附录 5 第七章状态空间模型程序	247
附录 6 第八章残差的 SSUR 图	257
参考文献	262

第一章 绪 论

1.1 背 景

人类社会存在于一个错综复杂的关系网中。这个无形的网是那么的神秘又令人难以琢磨,往往表面上看似毫无关联的两个事物之间可能存在着密不可分的联系,而表面看似关系密切的两个事物之间可能事实上并没有本质联系。许多事物之间的相互关系还直接影响到人类的生存和发展,如癌症的病源是什么,全球变暖的原因有哪些,经济危机的起因在哪里,等等。因此,事物间相互关系的研究吸引了无数学者。难怪早在两千多年前,古希腊哲学家 Democritus 就说过:“I would rather discover a single causal relationship than be king of Persia.”

从统计学上讲,研究两个事物之间是否有关系,一个直观的想法就是求出这两个事物的数据之间的相关系数,如果相关系数为零,则说明两事物之间无关系,反之则说明两事物之间存在某种关系。但是,历史上一个关于研究肺癌病因的例子彻底否定了这种思想的正确性。在统计学发展初期,曾经有人对肺癌患者的数据进行分析,发现喝咖啡和患肺癌这两者之间相关系数显著不为零,由此得到了喝咖啡会导致肺癌的结论。这个在今天看来非常荒谬的结论,直到很多年后才被推翻。事实上,喜欢喝咖

啡的人大多喜欢吸烟,即喝咖啡和吸烟之间正相关;而吸烟是导致肺癌的一大病因,因此研究肺癌患者的数据会发现,喝咖啡和患肺癌之间正相关。如果在给定吸烟这个中间因素后,喝咖啡和患肺癌之间是没有本质联系的,即喝咖啡和患肺癌是(在给定吸烟条件下)条件独立的。可见,研究事物之间是否有关系并不是那么简单的问题,而图模型(Graphical Models)方法正是研究事物之间条件相关性的一项有力的工具。

图模型是统计学中一个崭新的研究领域,是概率统计与人工智能理论的交叉学科,它将传统的多元统计中描述多个变量之间关系的统计模型图形化,是推断随机变量间不确定性和复杂性的一种有力工具,利用图的可分解性,可将许多高维问题分解成低维问题,并借助图的直观结构对随机变量间复杂的独立性关系、时序关系或因果关系进行分析。多元统计中的许多技术,如协方差选择模型、对数线性模型、回归分析、因子分析等,都可纳入图模型这一统一框架。目前,图模型方法已广泛应用于人类基因、医学诊断、工程可靠性、生物统计等领域。

20世纪90年代以来,世界上一些国家发生了一系列金融危机,如1997年亚洲金融危机、1998年俄罗斯金融危机、1999年巴西金融危机、2000年美国NASDAQ股票市场危机、2007年美国次贷危机等。随着全球经济一体化进程的深入,金融市场的开放程度不断提高,各个市场之间的相互影响也日益扩大,一国股票市场的走势往往直接受其他市场的影响,这促使了人们考虑金融市场的相关性问题,从随机变量角度来看,这需要研究变量间的相关关系。

美国金融危机和欧洲债务危机的爆发,不仅对美国和欧洲各国的经济产生了重大影响,而且迅速传染到全球,对世界金融体系造成了很大的破坏,其风险传递呈“多米诺骨牌效应”,进一步证实了各国金融体系存在显著的相关性和传导机制。金融危机的传染效应告诉人们一个事实,一国金融市场恶化时,其他国家和地区金融市场往往也会跟着恶化。因此,股票市场之间的相关性分析对国际金融市场的信息传导和风险管理有着重要意义,在金融分析和实践中有着举足轻重的作用。

事物的相关关系和因果关系一直是人们重点研究的问题。在实际活动中,人们经常要描述的对象可能涉及的变量众多,如股票市场的投资组合等,涉及的变量有几十甚至上百个,纯粹的数学或统计方法很难对如此

多的变量理出头绪,也很难直观地表示这些变量间的关系。况且在推断多维随机变量之间的关系时,参数估计的“维数灾难”问题也会是个不小的障碍。在推断多维时间序列变量之间的关系时,变量间的关系可能更加复杂。因为时间序列变量既需考虑变量间的同期相关关系,又要研究变量间的时序关系,甚至因果关系。因此,多维时间序列变量之间关系的推断更是扑朔迷离。

现代股票市场中,不同市场、不同行业指数或不同股票之间,往往存在着相关关系,随着世界金融市场的飞速发展,互动联系越来越紧密。传统的线性相关系数已很难适应越来越一体化的国际股票市场的相关性研究。因此,基于高维数据的统计方法成为研究股市相关性的重要工具。

1.2 相关理论的国内外研究进展

1.2.1 图模型理论和应用的国内外研究进展

图模型是统计学中一个崭新的研究领域,它将传统的多元统计中描述多个变量之间关系的统计模型图形化。

(1) 国外研究综述

国际上对图模型研究,最早起源可追溯到 20 世纪 30 年代,Wright 在有关基因的研究中,用到了所谓路径分析方法^[1-2]。但是,对该领域的大量研究工作,始于 20 世纪 70 年代,Goldberger(1972)^[3]利用路径分析方法讨论线性结构关系模型。Lauritzen(1990)^[4]、Cox 和 Wermuth(1993)^[5]指出图模型方法优于线性结构关系式模型。关于图模型的第一部专著 *Graphical model at applied multivariate statistics* 于 1990 年才出版^[6]。近几年,图模型方法无论在理论和应用研究上都取得了突破性的进展。

对于图模型,图中的点表示随机变量,若点与点间有边相连接,表示两变量间存在相依结构。若两点间无边相连,说明在给定其余变量条件下,此两变量是条件独立的。在用图来建模的过程中,最核心的概念就是图的马尔可夫性,即图中的分离性蕴含概率分布中的条件独立性。

由观测数据构建图模型就是要寻找观测数据的分布 P 的忠实图,但实际中并非每个分布 P 都存在忠实图,而且由 P 观测数据推断的过程中也难免存在误差,因此实际中往往只需在与 P 相适应的图中找一个尽可能接近忠实图且复杂度较低的图。根据观测数据构建图模型一般有两类方法:“条件独立性检验”的方法和“比较得分”的方法。

第一类方法根据图与分布之间忠实性的假设,直接检验数据中变量间的条件独立性。对离散分布,通过检验对数线性模型中相应的交互项是否为 0 来判断条件独立性。对连续分布,一般考虑多元正态分布通过令精度矩阵(precision matrix)的某些非对角线元素为 0,从而确定变量的条件相关性。Whittaker(1990)^[2]提出 Graphical Gaussian model,该模型在变量集服从多元正态分布的假设下,由全模型出发,通过似然比方法检验图模型结构。对于一般分布,可通过变量变换,比如 Box-Cox 幂变换,化成正态分布来处理,或者直接利用相互信息量来判别条件独立性。

第二类方法是基于得分原则的模型选择,对变量集上的每个可能的图 G ,计算它的 AIC 得分, $AIC(G,D)=l(G,D)-Dim(G)$,其中 $l(G,D)$ 是在图 G 所蕴含的独立性假设下数据 D 的最大对数似然函数值, $Dim(G)$ 是对图的复杂度的度量。最后选取得分最高的图。

Lauritzen(1996)^[7]从数学的角度对图模型的条件独立性、马尔可夫性及多维高斯正态模型的协方差选择、可分解图模型的性质等给出严格的理论推导和证明。Edwards(2001)^[8]系统地介绍了离散、连续和混合随机变量的图模型方法,将列联表、对数线性模型、因子分析方法等纳入图模型结构中,且将图模型方法应用于医学领域和专业课程相关性的研究。

图模型方法最初主要应用于基因和医学领域。Beinlich(1989)^[9]对医学外科手术中麻醉问题涉及的 37 个变量进行分析,用一个有向非循环图(DAG)模型直观清楚地表达了这些变量间的复杂关系。Tritchler(1999)^[10]将有向图方法应用于医学领域进行因果关系的统计推断。Friedman(2000)^[11]研究了细胞循环中基因表达水平的数据,该数据集有 800 个基因(变量),但却只有 76 个观测,对这种数据一般统计方法很难处理,但借助图模型和其中的算法,作者对如此多的变量间复杂的独立性结构进行推断和直观表示,而且得到的 DAG 图中的有向边也预测出基

因表达中的一些很有意义的时序关系和因果关系。Pearl(2000)^[12]对抽烟与患肺癌之间的关系进行研究,利用有向非循环图揭示基因型、抽烟量、焦油在肺内的沉积量与是否得肺癌之间的因果联系,研究结论表明:在给定基因型和焦油在肺内的沉积量条件下,抽烟量和是否得肺癌是独立的。2013年,国际著名统计学家范剑青教授提出大数据的图模型建模技术,并应用于基因数据挖掘分析^[13]。在图模型结构推断上,许多学者使用贝叶斯方法,取得了一定的成效^[14-16]。

图模型方法是研究多维时间序列强有力的工具,通常描述时间序列图模型的方法主要有偏相关图、时间序列链图和因果图三种。在偏相关图中,图中的点表示变量的随机点过程,点与点之间通常通过无向边表示变量间的偏相关关系。在时间序列链图中,图中的点表示某一离散时刻的随机变量,即随机点过程用图中一系列点表示,图中的无向边表示变量间的同期相关关系,有向边表示时序关系。第三种图称为混合图,图中的点表示变量的随机点过程,图中的有向边表示变量间存在 Granger 因果关系,无向边表示变量的同期相关关系,混合图又称为 Granger 因果图。

Brillinger(1996)^[17]首先将图模型方法应用于平稳时间序列,直观描述多维过程变量的相关关系。由随机变量的无向图延伸出时间序列变量的偏相关图的概念,即变量的点过程通过图中的一个点来表示,点与点之间的无向边表示变量间的偏相关关系。Dahlhaus^[18-19]利用偏谱一致性分析方法研究多维时间序列变量的交互作用,分别将该方法应用于神经系统突触联结的识别和环境污染变量的多维自回归模型。Timmer(2000)^[20]提出利用交互谱分析方法确定时间序列变量的偏相关图,且将该方法应用于医学领域,对引起颤动症相关因素之间的关系进行探索。Gather(2002)^[21]提出图模型方法在处理高维时间序列数据上有特别的优势,通过偏相关图模型方法可分析多维时间序列变量的偏相关关系,同时将该方法应用于重症特别护理的监控研究,描述重症病人病理指标的相关关系。

Reale(1998)^[22]系统地讨论了多元时间序列的图模型理论,以多维的 AR 模型和 ARMA 模型为例说明时间序列模型的图方法的构建和检验,且将该方法应用于宏观经济领域进行实证分析。Eichler(1999)^[23]提出时间序列的偏相关图、时间序列链图和混合图的定义及其马尔可夫性质,且从

数学角度给出严格的理论推导和证明,并通过非参数的谱域方法分析图中变量的相关和因果关系。Thiesson(2004)^[24]应用图模型方法对一维 ARMA 进行建模,将 ARMA 时间序列表示为一个有向图,且应用 EM 算法估计模型参数。最后,将 ARMA 模型推广为随机 ARMA 模型。

Eichler(2001)^[25]主要介绍描述多维时间序列变量因果关系的混合图模型的有关性质,在 Granger 因果图中,有向边表示变量间的 Granger 因果关系,而无向边表示变量间的同期相关结构。将该图模型方法应用于 VAR 模型,给出 Granger 非因果关系的定义,同时应用因果图模型描述经济变量间的因果关系。Reale(2001)^[26]利用条件独立图识别有递推关系误差项的 AR 结构模型。Eichler(2003)^[27]基于谱域分析方法,通过偏相关分析识别神经系统突触的直接和间接联系。Richardson(2003)^[28]将时间序列的图模型进行拓展,提出时间序列的有向非循环混合图和给出该图的马尔可夫性质。Drton(2006)^[29]提出高斯正态随机变量的时间序列链图方法及马尔可夫性质。Eichler(2007)^[30]引入多维时间序列的路径依赖方法描述变量间的动态关系,分析时间序列的相依结构,同时给出 Granger 非因果关系和在某一时期 Granger 非因果关系的图模型充分条件。由于图模型方法具有全局马尔可夫性质,路径依赖分析方法为多维时间序列的图模型方法提供了较为理想的分析工具。

Eichler(2005)^[31]首次提出含有潜变量的时间序列因果图模型,且将该方法应用于大脑的映射活动,指出变量间的因果关系的图模型描述会受潜变量影响,潜变量可能会导致变量间出现虚假的因果关系。Eichler(2006)^[32]描述了受潜变量影响的多维平稳时间序列相关结构的混合图模型,在混合图中实线有向边表示变量间的直接因果影响,而虚线(有向或无向)显示由潜变量导致的间接联系。同时,将混合图模型方法应用于满足 Granger 因果关系限制的 ARMA 模型图模型,且给出参数估计的方法和数值模拟。

(2) 国内研究综述

在图模型基本理论和应用研究方面,我国最早开始由北京大学耿直教授^[33]于 2000 年提出利用 Simpson 悖论给出相关和因果关系的区分,体现图模型应用的思想。郑忠国教授(2000)^[34]证明了在只包含离散变量的反馈系统产生的因果图中的条件独立关系可以由 d-分离识别。郑忠

国教授(2001)^[35]对有向非循环图中的因果关系进行探讨,并提出虚拟事实模型,在可替换性假设之下,因果效应具有可识别性,即有可能从观察数据直接计算因果效应。梁宇和郑忠国(2003)^[36]针对控制变量作用于协变量的虚拟事实模型进行可识别性研究,指出控制变量是否作用于协变量并不影响因果效应的可识别性和可替换性假设,并给出此类模型因果效应可唯一确定的充要条件。许静和郑忠国(2004)^[37]给出了一个虚拟事实模型中因果效应的贝叶斯估计和经验贝叶斯估计,且提供了可替换性假设的先验分布的选择方法,研究结论表明经验贝叶斯估计要优于其他的估计。耿直(2004)^[38]探讨因果链上总因果效应与局部因果效应的关系及因果效应的可识别性,给出了在因果链上因果效应的传递关系,根据这个关系,还给出了通过控制因果链上中间因素消除混杂偏倚的方法。刘佰军(2005)^[39]给出了一个根据已知链图找出最大链图的算法,给出了直观地判断一个链图是否是与之等价的极大链图的方法。赵慧(2005)^[40]研究了最大祖先图的马尔可夫等价性,得到判断一个最大祖先图与另一个最大祖先图,以及与一个有向非循环图马尔可夫行等价的充要条件,并且提出了一个把最大祖先图变换成与其等价的有向非循环图的快速实现算法。

李开灿(2005)^[41]从可分图协方差矩阵的密度函数出发确定图精度矩阵分布密度函数的一般方法,得到了可分的 Gaussian 图模型中精度矩阵极大似然估计的密度函数表达式。当图协方差矩阵的分布密度分别服从超逆 Wishart 分布、超逆 Γ 分布时,也得到了图精度矩阵密度函数的解析表达式。赵慧(2006)^[42]构建含隐变量和选择变量的图模型方法,利用在观测变量上构造的最大祖先图模型刻画观测变量间的独立性关系和因果结构,并提出了具体的实现算法,从而可由观测数据来推断这类不完全观测下的部分因果关系。赵慧(2007)^[43]主要借助拓扑图的直观形式对多维概率分布进行统计推断和因果分析,对图模型的基本概念、原理、建模进行了简单介绍。张凌洁和张海(2014)^[44]提出一种分层惩罚连接单个图模型估计的多图模型,分析新模型的高维统计性质,给出模型的参数估计,并验证其相合性和稀疏性。

图模型方法在国内的应用则是近几年的事情。李万里(2003)^[45]将图模型方法应用于医学研究领域,利用有向非循环图研究变量之间的因

果关系,分析老年人膳食营养素对 BMD 的影响。赵慧和郑忠国(2007)^[46]在对一种新型中药降脂灵片的药效进行因果分析时,指出传统的统计方法难以处理实验数据样本量小且是混合变量类型问题,提议采用图模型的方法建立一个链图模型,直观地刻画了该药对反应机体抗氧化能力和血脂水平的 4 个指标的因果影响。刘桂梅和李胜宏(2006)^[47]将图模型理论应用于遗传基因领域,利用 DNA 数据,研究了一个在法庭上颇有争议的亲子鉴定问题。通过案例中的家谱图,建立 Bayesian 网络,根据遗传学的孟德尔定律,计算出网络中各结点的概率,给出了一个可供法庭参考的合理推断。赵斌(2013)^[48]将二分图模型方法应用于微博大数据分析和管理,取得了一定的成效。

高伟和田铮(2008)^[49]提出了基于条件互信息的时间序列图模型,且将该模型应用于美国债券市场,研究其相依结构。李元教授(2008)^[50]将时间序列链图应用于自回归和双线性时间序列模型,证明自回归和双线性模型的系数为其他时间序列分量给定的条件下的条件相关系数。最后提出基于图的检验方法来检验自回归和双线性模型系数的显著性,结果表明该检验方法水平和功效表现很好。魏岳嵩和田铮(2010)^[51]对结构向量自回归时间序列的链图模型识别方法进行研究,利用局部密度估计法以及 Bootstrap 方法,给出了时间序列链图模型的概念以及模型结构识别方法,模拟结果显示该方法能有效地识别结构向量自回归模型变量间的相依关系。魏岳嵩和田铮(2011)^[52]提出利用局部密度估计构造相应的检验统计量,并采用 bootstrap 来确定检验统计量的原分布,从而识别 Granger 因果图模型结构。

魏岳嵩和杜翠真(2014)^[53]利用图模型方法辨识非线性结构向量自回归模型变量间的因果关系,给出了一种基于互信息和条件互信息的非线性结构向量自回归因果图模型结构的非参数辨识方法,数值模拟结果验证了方法的有效性。

1.2.2 股票市场相关和传导研究的国内外研究进展

(1) 国外研究综述

许多学者致力于股市相关性研究,在协整方法出现前,学者主要利用