

金融高频协方差阵的 估计及应用研究

Estimation and Application Study on Financial
Covariance Matrix of High Frequency Data

刘丽萍/著



科学出版社

金融高频协方差阵的估计 及应用研究

刘丽萍 著

科学出版社

北 京

内 容 简 介

本书基于金融市场的高频数据,考虑市场微观结构噪声和跳跃对协方差阵的影响,提出了修正的门限预平均已实现协方差阵(MTPCOV),并采用分块策略和正则化技术对其进行修正。将高频数据波动理论、计量分析方法及实证研究进行结合,研究了新估计量的理论性质及其在投资组合中的应用。

本书适合统计学、金融学和数学等相关专业的研究人员和高校师生阅读,也可供从事金融高频数据研究的工作者和对金融计量感兴趣的读者阅读。

图书在版编目(CIP)数据

金融高频协方差阵的估计及应用研究/刘丽萍著. —北京:科学出版社, 2016.6

ISBN 978-7-03-048698-1

I. ①金… II. ①刘… III. ①金融—经济数学—研究 IV. ①F830

中国版本图书馆CIP数据核字(2016)第129409号

责任编辑:郭勇斌 邓新平/责任校对:杜子昂

责任印制:张 伟/封面设计:众轩企划

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016年6月第 一 版 开本:720×1000 1/16

2016年6月第一次印刷 印张:11 1/4

字数:151 000

定价:65.00 元

(如有印装质量问题,我社负责调换)

前 言

在金融计量领域里，多维的金融高频协方差阵在投资组合和风险管理中起着非常重要的作用。在资产组合中，资产权数的确定跟它们之间的协方差阵是密切相关的，得到的协方差阵越精确，权数的分配越合理，组合效果将会更好。因此，对基于高频数据的协方差阵估计方法进行研究具有重要的理论和实践意义。

高频数据包含了丰富的市场信息，是金融领域的研究重点之一。当市场上不存在市场微观结构噪声或跳跃时，已实现协方差阵（RCOV）是积分协方差阵的一致估计量。而在现实的金融市场上，噪声和跳跃往往是同时存在的，在二者都存在的情况下，对高频协方差阵进行估计从而得到积分协方差阵的一致估计量，是值得深入研究的问题。但是现有的研究要么只考虑市场微观结构噪声的影响，要么只考虑跳跃的影响，很少有文献同时考虑噪声和跳跃对高频协方差阵估计的影响。而噪声和跳跃有可能同时存在于金融市场上，在二者同时存在的情况下对高频数据的协方差阵进行估计，仍是一个较为困难的问题。本书在前人研究的基础之上，针对目前研究的不足，提出了新的高频协方差阵估计量，同时处理市场微观结构噪声和跳跃的影响，提高高频协方差阵的估计效率。本书将高频数据波动理论、计量分析方法及实证研究紧密地结合起来，采用定性与定量相结合的方法，既有理论的梳理与构建，又有详细的实证分析。

本书的亮点主要表现在三个方面。首先，提出了一个新的高频协方差阵估计量——修正的门限预平均已实现协方差阵（MTPCOV）估计量。该估计量不仅可以处理市场微观结构噪声的影响，还可以剔

除跳跃对高频协方差阵估计的影响。其次，将分块策略和正则化方法应用在 MTPCOV 的估计中，减少数据量的损失，得到了基于流动性调整的修正的门限预平均已实现协方差阵 (RnBMTPCOV) 估计量。该方法是对 MTPCOV 的调整，克服了其由于刷新时间采样而导致大量数据损失的缺点，在不对参数施加任何限制的情况下，提高了估计精度。最后，将新的高频协方差阵估计方法应用在投资组合中，并与其他高频协方差阵估计方法进行了比较分析。

国内对高频数据的研究大都集中在一维的研究，而对多维的高频协方差阵的研究还非常少，个别的文献涉及高频数据的 RCOV，但对于能够消除市场微观结构噪声和跳跃的 MTPCOV 及其在投资组合中的应用，国内还没有相关研究。通过本书的研究，希望能够给统计学、金融学和数学等相关专业的研究人员及高校师生一些启发。

书中难免有不足之处，恳请读者指正。

刘丽萍

2015 年 12 月

缩 写 释 义

KCOV	Multivariate Realized Kernels Covariance Matrix	多元已实现核协方差阵
MPCOV	Modified Pre-averaging Realized Covariance Matrix	修正的预平均已实现协 方差阵
MPCV	Modified Pre-averaging Realized Covariance	修正的预平均已实现协 方差
MRC	Modulated Realised Covariance	调整的已实现协方差
MRCOV	Modulated Realised Covariance Matrix	调整的已实现协方差阵
MTPCOV	Modified Threshold Pre-averaging Realized Covariance Matrix	修正的门限预平均已实 现协方差阵
MTPCV	Modified Threshold Pre-averaging Realized Covariance	修正的门限预平均已实 现协方差
PCOV	Pre-averaging Realized Covariance Matrix	预平均已实现协方差阵
PCV	Pre-averaging Realized Covariance	预平均已实现协方差

RBPCOV	Realized BiPowe Covariance Matrix	已实现双幂次协方差阵
RBPCV	Realized BiPower Covariance	已实现双幂次协方差
RCOV	Realized Covariance Matrix	已实现协方差阵
RCV	Realized Covariance	已实现协方差
RnBMTPCOV	Blocking and Regularization Modified Threshold Pre-averaging Realized Covariance Matrix	基于流动性调整的修正的门限预平均已实现协方差阵
ROWCOV	Realized Outlyingness Weighted Covariance Matrix	已实现离群加权协方差阵
thresholdCOV	Threshold Covariance Matrix	门限协方差阵
TPCOV	Threshold Pre-averaging Realized Covariance Matrix	门限预平均已实现协方差阵
TPCV	Threshold Pre-averaging Realized Covariance	门限预平均已实现协方差
TSCOV	Two time scale Realized Covariance Matrix	双频已实现协方差阵
TSCV	Two time scale Realized Covariance	双频已实现协方差

目 录

前言

缩写释义

1	绪论	1
2	金融高频数据研究现状	4
2.1	高频数据及其特征分析	4
2.1.1	什么是金融高频数据	4
2.1.2	金融高频数据的主要特征	4
2.2	金融高频数据分析的主要动因	5
2.3	金融高频数据分析研究的现状	5
2.3.1	金融高频数据统计特征的研究	6
2.3.2	金融市场微观结构的研究	8
2.3.3	金融高频数据建模的研究	9
2.3.4	基于金融高频数据已实现波动的研究	12
2.3.5	基于金融高频数据协方差阵的研究	21
2.4	我国研究金融高频数据的必要性	22
3	常见的高频协方差阵估计方法及其应用	24
3.1	RCOV 估计方法	24
3.2	基于市场微观结构噪声的 RCOV 估计方法	27
3.2.1	市场微观结构噪声对 RCOV 的影响	28
3.2.2	考虑了市场微观结构噪声影响的 RCOV 估计方法	29
3.3	考虑跳跃影响的高频协方差阵估计方法	37
3.3.1	RBPCOV 估计方法	37

3.3.2	ROWCOV 估计方法	40
3.3.3	thresholdCOV 估计方法	42
3.4	金融高频协方差阵在投资组合中的应用情况	43
3.5	本章小结	46
4	TPCOV 估计方法的提出及其修正	48
4.1	预平均协方差阵估计方法	49
4.1.1	改进的预平均方法	49
4.1.2	基于预平均方法的 MRCOV 估计法	53
4.2	新估计量的提出——TPCOV 及其修正	55
4.2.1	高频数据的基本设定	55
4.2.2	MTPCOV 的构造形式	56
4.2.3	积分方差的一致估计量——MTPRV	57
4.2.4	积分协方差的一致估计量——MTPCV 估计量	60
4.3	基于 MTPCV 的模拟研究	65
4.3.1	窗宽及门限函数的选择	65
4.3.2	基于随机波动模型的数据模拟研究	70
4.4	本章小结	80
5	RnBMTPCOV 的估计	82
5.1	基于刷新时间方案的 MTPCOV 的数据损失分析	84
5.1.1	刷新时间方案	84
5.1.2	基于刷新时间方案的数据损失分析	85
5.2	RnBMTPCOV 估计方法	87
5.2.1	基于分块策略的协方差矩阵	87
5.2.2	协方差阵的正则化处理方法	91
5.3	RnBMTPCOV 的估计及有效性分析	93
5.3.1	RnBMTPCOV 估计结果的描述性统计分析	93

5.3.2 基于 Mincer-Zarnowitz 回归的协方差阵的比较分析	94
5.4 本章小结	97
6 多维协方差阵预测模型的比较分析	98
6.1 基于高频数据的协方差预测模型	100
6.1.1 CF-ARMA (2, 1) 模型	101
6.1.2 FI-VAR 模型	102
6.1.3 多元异质自回归 (MHAR) 模型	104
6.1.4 基于 Wishart 分布的自回归 (WAR) 模型	106
6.2 基于低频数据的协方差阵预测模型	107
6.2.1 DCC 模型	108
6.2.2 BEKK 模型	108
6.3 预测模型的比较方法	109
6.3.1 损失函数	110
6.3.2 MCS 检验	111
6.4 模型预测结果的比较	113
6.4.1 数据的描述	113
6.4.2 多维协方差阵预测模型的比较分析	116
6.5 本章小结	119
7 金融高频协方差阵在投资组合中应用的实证分析	120
7.1 高频数据在投资组合中应用问题的提出	120
7.1.1 引言	120
7.1.2 投资组合优化问题	122
7.2 实证分析方法介绍	124
7.2.1 动态投资组合策略——波动择时策略	124
7.2.2 动态投资组合的比较方法	125
7.3 实证分析	129

7.3.1 样本数据的处理	129
7.3.2 各投资组合的收益和波动分析	130
7.3.3 各投资组合的经济收益分析	131
7.3.4 各投资组合 Sharpe 比率的比较	134
7.4 本章小结	137
参考文献	138
附录 A 书中用到的部分程序代码	149
附录 B 部分模拟数据	159

1 绪 论

金融资产的协方差阵在投资组合中起着非常重要的作用,资产权数的确定与它们之间的协方差阵是密切相关的。如何准确估计金融资产的协方差阵,使组合效果达到最优,是金融领域研究的热点问题之一。不同方法计算的协方差阵存在较大的差异,目前主流的计算方法是基于低频数据对协方差阵进行估计,但是在交易频繁的金融市场上,低频数据损失了很多有用的信息使协方差阵的估计不够理想。随着高频数据可获得性的提高及计算方法的飞速发展,越来越多的学者开始对高频数据进行研究,高频数据提供了关于市场微观结构的大量信息,从而使基于高频数据的协方差阵估计方法更为精确。

不同的协方差阵估计方法估计得到的矩阵存在较大差异,将其应用于投资组合时也会存在显著的差异。因而将高频协方差阵应用到投资组合时,首先应该考虑估计得到的协方差阵是否是有效的、无偏的。当金融市场满足有效市场的假定,即市场上不存在市场微观结构噪声和跳跃时,Andersen 等(2003a)提出的 RCOV 是积分协方差阵的一致估计量,但是当市场微观结构噪声或跳跃存在时,RCOV 估计的往往不是资产的有效价格的协方差阵,而是市场微观结构噪声或跳跃的协方差阵。

为了降低市场微观结构噪声或跳跃对高频协方差阵估计的影响,很多学者提出了新的协方差阵估计方法,如双频已实现协方差阵(TSCOV)估计方法、多元已实现核协方差阵(KCOV)估计方法等,这类协方差阵估计方法考虑了市场微观结构噪声对高频协方差阵估计的影响。又如已实现双幂次协方差阵(RBPCOV)估计方法、已实现离

群加权协方差阵 (ROWCOV) 估计方法、门限协方差阵 (thresholdCOV) 估计方法等, 这类协方差阵估计方法考虑了跳跃对高频协方差阵估计的影响。这些研究, 要么只考虑了市场微观结构噪声的影响, 要么只考虑了跳跃的影响, 而在现实的金融市场上, 噪声和跳跃往往是同时存在的。如何在噪声和跳跃二者都存在的情况下, 对高频协方差阵进行估计从而得到积分协方差阵的一致估计量, 是值得深入研究的问题。

目前, 学者们大致从三个方面对这一问题进行了研究。首先, 如何准确地估计高频协方差阵, 尤其是当市场微观结构噪声或跳跃同时存在时, 这也是本书的研究重点和难点。其次, 如何选择预测模型对高频协方差矩阵进行预测。最后, 在构造投资组合时, 全面考虑目标函数的选择及组合调整频率的变化对投资组合收益的影响。

中国的金融市场正处于转轨阶段, 市场还不够成熟, 仍存在诸多问题。近年来, 已有很多学者对中国金融市场进行了深入的研究, 并提出了许多宝贵的建议。但是国内的研究大多是对一维的高频数据进行研究, 对于多维的金融高频协方差阵的研究少有涉及。而多维的金融高频协方差阵在投资组合和风险管理中起着非常重要的作用, 对高频协方差阵进行研究有助于防范或者降低金融风险。当市场微观结构噪声和跳跃同时存在时, 如何精确地估计和预测高频协方差阵? 高频协方差阵应用在投资组合中真的比低频数据的效果更好吗?

为了在市场微观结构噪声和跳跃都存在的条件下, 对金融高频数据的协方差阵进行估计, 本书提出了新的估计量——MTPCOV 估计量, 该估计量能够通过预平均方法降低市场微观结构噪声的影响, 通过门限剔除跳跃的影响。首先对 MTPCOV 估计量的理论性质进行探讨, 然后采用模拟数据对其进行模拟研究, 最后采用中国金融市场真实的股票高频数据对 MTPCOV 估计量进行实证检验。在估计 MTPCOV

时,为了解决不同的交易问题,往往采用刷新时间方案对高频数据进行同步化处理,这使数据的损失量非常大,尤其是在考虑的资产数目较多时。为了减少数据的损失得到更为精确的高频协方差阵估计量,将分块策略与正则化方法应用到 MTPCOV 的估计中,得到了 RnBMTPCOV 。另外,还对常用的高频协方差阵预测模型进行了比较分析,选择了较好的协方差阵预测模型对高频协方差阵进行预测,并对 MTPCOV 估计量及 RnBMTPCOV 估计量在投资组合中的应用情况进行了深入研究。

2 金融高频数据研究现状

2.1 高频数据及其特征分析

2.1.1 什么是金融高频数据

近年来, 计算工具与计算方法的发展, 极大地降低了数据记录和存储的成本, 使得对大规模数据库的分析成为可能。所以, 许多科学领域的数据都开始以越来越精细的时间刻度被搜集, 这样的数据被称为高频数据。高频金融数据是指日内的金融时间序列, 以小时、分和秒为采集频率的、按时间先后顺序排列的金融类数据。金融市场中, 逐笔交易数据或逐秒记录数据就是高频数据的例子。相比以日、月、年为频率的低频数据, 金融高频数据中除交易价格外, 还包括与交易相连的询价和报价、交易数量、交易之间的时间间隔、相似资产的现价等方面的具有高度持续性的交易信息。

2.1.2 金融高频数据的主要特征

按照更短时间间隔所取得的金融高频数据较传统的低频观测数据而言, 呈现出了其独有的一些特征。它主要有四个典型的特征: 一是高频数据的记录间隔是不等的, 这是因为在金融市场上, 某支股票的交易并不一定按相同的时间间隔发生, 因而所观测到交易价格等变量的时间间隔也就不等; 二是高频数据所记录的交易价格是离散的变量, 某项资产的价格变动往往是以计量单位的若干倍发生的; 三是高频数据存在着 U 型的日内周期模式, 其交易量往往在每一天的开盘和收盘时间较大, 在中午较小; 四是多笔交易有可能

同时发生,这种现象归因于高频数据每天的交易量都很大,从而使秒计量时间都成为一个较大的时间单位。

2.2 金融高频数据分析的主要动因

伴随着社会经济的发展,金融交易的频率不断加快,金融高频数据成为金融研究中备受瞩目的焦点。学术界对金融高频数据关注的原因主要可以归结为两个方面。一是金融高频数据本身的特征值。交易数据本身即存在多个维度,如交易价格、交易数量、同类或类似资产价格、交易时间间隔等。因此,对金融高频数据的分析可以是不同时间间隔,既有离散变量又有连续变量的多维度复杂问题。这样,问题就成为既可以总体分析,又可以局部特殊处理的多层次富有挑战性和乐趣的课题。二是金融高频数据由微观交易产生,分析对于理解微观市场非常重要。对于高频数据的分析也得到了非常重要的成果,转变了原有的一些错误或不精确的观念。如短期价格波动不再被认为是无搜集意义的噪声,可以通过对高频数据的考察进行分析;再如一些古典经济假定短期价格波动服从高斯随机游程(Gaussian Random Walk)、金融市场同类性(Homogeneous)等随着对高频数据的统计认识而受到质疑。总之,对金融市场微观结构的考察,需要对基础经济理论、研究方法和计量模型等不断地进行完善,金融高频数据及其分析为这些转变的实现提供了条件。

2.3 金融高频数据分析研究的现状

尽管对金融高频数据和超高频数据的分析研究历史并不长,但是目前的发展状况却着实令人鼓舞。众多研究者对此都表现出极大的兴趣,分别从不同的角度对金融高频数据和超高频数据进行了探索和研究。

2.3.1 金融高频数据统计特征的研究

在讨论金融高频数据如何应用时, 证券价格会受到金融市场信息的影响, 所获得的金融市场信息的多少取决于金融数据的抽样频率, 数据的抽样频率越高, 所获得的市场信息越多。要对金融高频数据建立模型, 就很有必要对金融高频数据的统计特征进行分析。因为统计特征不仅是认识数据的基本依据, 也是正确使用数据的前提。早期的研究表明, 金融高频数据是不稳定的, 在较短期间内有厚尾 (Heavy-tail) 趋势。相比较而言, 近期对金融高频数据统计特征的研究更为深入和具体。有研究发现高频收益数据具有非正态性, 随着数据频率的增加, 其数据的峰度也随之增加, 以分钟为频率的数据, 其峰度就已经达到 100 以上 (Andersen and Bollersle, 1998)。有学者采用高频数据对美国股票市场和外汇市场的日内 (Intra Day) 波动性和长记忆性进行研究, 证明了在这些市场中存在着波动的长记忆性 (Andersen and Bollerslev, 1997)。在 T.G. Andersen 和 T. Bollerslev 研究中, 他们还进一步利用高频数据对日本股票市场进行了研究, 通过滤波的方法证明了波动长记忆性的存在 (Andersen et al., 2000b)。

研究发现高频数据除了具有低频数据所具有的长记忆性、非正态性, 其自身还具有一些独特的统计特征:

(1) 具有日内“U”型走势

在对日内的高频数据进行实证研究时, 会发现日内波动率在开盘和收盘时段比较高, 而在中间时段, 波动率相对较低, 整体来看, 其表现出典型的“U”型走势。有些学者分别从不同的角度对波动率的“U”型结构进行研究, 对其结构模式给出了合理的理论解释 (Admati and Pfleiderer, 1988; Brock and Kleidon, 1992); 有的学者则采用单个