



金融极值数据波动率建模

◎ 刘威仪 著 ◎

清华大学出版社

明理文丛

金融极值数据波动率建模

◎ 刘威仪 著 ◎



清华大学出版社
北京

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

金融极值数据波动率建模/刘威仪著. —北京:清华大学出版社,2017
(明理文丛)

ISBN 978-7-302-48734-0

I. ①金… II. ①刘… III. ①金融-极值(数学)-研究 IV. ①F830.41

中国版本图书馆CIP数据核字(2017)第266277号

责任编辑:朱玉霞

封面设计:阿东

责任校对:王凤芝

责任印制:杨艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印刷者:北京富博印刷有限公司

装订者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:142mm×210mm 印 张:5.5 字 数:150千字

版 次:2017年9月第1版 印 次:2017年9月第1次印刷

定 价:39.00元

产品编号:075668-01

读书 || 明理



前 言

本书是笔者近几年来在金融计量学领域探索基于金融极值数据对波动率建模的系统性研究成果。这项研究工作得到了国家自然科学基金青年项目——“基于高频极值数据的金融资产跳跃行为建模研究”(71601132)和北京市优秀人才青年骨干项目——“金融高频极值数据建模”(2016000020124G083)的资助。结合这两项基金的研究和笔者前期的工作,在分别利用低频极值数据和高频极值数据两个方面,对金融时间序列的波动率估计、预测和跳跃检验等方面都获得了一系列创造性成果。

金融资产价格过程的波动率在构造投资组合、进行资产定价和风险管理等各类金融决策中都起着至关重要的作用。目前已有的研究主要是基于收盘数据的波动率建模,而基于极值数据的研究则相对匮乏,还有诸多未完善和待改进之处,特别是如何将其与近年来流行的高频数据理论相结合,是亟待探索的重要方向之一。金融极值数据,从广义上来讲同时包括开盘价、最高价、最低价和收盘价,充分利用极值数据能够显著改善波动率的估计效率、提高价格跳跃的收敛速度以及更精确地刻画市场微观结构。另一方面,以高开低收为一组的金融极值数据,无论是低频的日度形式还是高频的分时形式,四个价格在国内外金融数据库中一般都是同时报出的,因此基于极值数据对波动率建模具有较强的技术可行性。已有的研究已充分表明基于高频数据的建模能够得到更加精确的结果,但以目前的条件来看高频数据还时常存在获取和数据质量等方面的问题,因此针对低频数据的扩展研究仍具有一定程度的理论和实践意义。为此,本书将分别针对低频极

值数据和高频极值数据两个角度展开探讨。

基于低频数据的波动率建模是传统的经典方法,例如条件异方差(ARCH、GARCH类)模型和随机波动率(SV)模型是基于日度收盘价格最常用的模型。自从Engle(1982)首创ARCH模型以来,国际上迄今已有三百多种各类扩展的ARCH或GARCH类模型用于刻画不同特点的波动率变化特征。然而,Hansen和Lunde(2005)针对上述各类扩展的波动率模型进行了大量的实证研究,结果表明它们均不能显著超越GARCH(1,1)模型的预测效果,从而进一步奠定了GARCH(1,1)模型的实证地位。无法改进预测效果的原因之一是上述模型都仅使用到了收盘价数据,在高频数据无法获取的情况下,本书充分利用日度的开盘、最高、最低和收盘等低频数据信息,建立了基于日度极值数据的GARCH-X模型框架。通过最新的评价方法实证表明,放弃使用复杂的模型形式而充分利用低频的极值信息,却可以显著地改善波动率的预测。在此基础之上,进一步利用异质市场假说能够进一步解释波动率的“长记忆性”,笔者的研究发现价格极差的主成分能够更加清晰地揭示日、周、月、季所对应的异质投资者类型,使得最终的波动率预测效果能够达到和低频模型相仿的水平。

基于高频数据的波动率建模是近年的热点方向,其中利用高频收盘数据来构造积分波动率估计是当前最为普遍的方式,例如Andersen et al.(2003)提出的已实现波动率(RV)以及Barndorff-Nielsen和Shephard(2006)提出的基于双幂次变差(BV)的跳跃检验方法被广泛地运用于各类经济金融问题的实证研究,但Christensen和Podolskij(2007)的研究仍然表明基于高频极值数据能够显著改善上述研究的结果,不过其统计量的构造只是简单的形式上用高频极差替代了收益率平方,并没有充分利用到高频价格极差的特性。本书中笔者提出了高频双格极差的概念,并基于此构造了已实现信息波动率(RIV)和已实现信息幂次变差(RIPV),

前者是当前渐近方差最小的跳跃稳健一致估计,后者是当前跳跃收敛速度最快的积分波动率估计。以上结果扩展了跳跃稳健波动率估计的理论基础和应用范围。鉴于高频双格极差在波动率估计和价格跳跃度量上的优势,笔者进一步提出基于高频双格极差的跳跃检验方法。另外,考虑人们对风险态度的不对称性,结合高频极值数据的特点,定义了已实现上变差(RUV)和已实现下变差(RDV),并给出了可行的单向跳跃检验方法。模拟和实证分析的结果表明相比已有的跳跃检验,本书给出的跳跃检验方法更具稳健性,特别是在离散化报价机制下可以较好地避免“伪跳跃”的检验结果。

本书可供从事金融计量、金融工程和市场微观结构等方面研究的科研人员,从事实际投资决策和风险管理的专业人员及相关专业的高等院校师生阅读参考。由于作者知识水平有限,选题也限于作者的兴趣,本书难免存在疏漏,欢迎广大读者不吝赐教。

最后,特别感谢我的恩师北京大学王明进教授对本书从框架内容到技术细节的悉心指导和帮助!同时感谢出版社相关工作人员的大力支持!

刘威仪

2017年3月

目 录

第 1 章 绪论	1
1.1 波动率建模的历史背景	1
1.2 金融极值数据波动率建模	2
1.2.1 基于低频数据的研究	2
1.2.2 基于高频数据的研究	4
1.3 本书的结构安排	6
第 2 章 极值数据的理论与方法	8
2.1 随机变量的极值理论	8
2.1.1 随机变量的收敛性	8
2.1.2 次序统计量与极差	15
2.2 随机游动的极值理论	20
2.2.1 带吸收壁的随机游动	20
2.2.2 泛函中心极限定理	26
2.3 布朗运动的极值理论	31
2.3.1 布朗运动的反射原理	31
2.3.2 扩散方程与极差分布	36
第 3 章 基于低频极值数据的动态波动率模型	42
3.1 引言	42
3.2 波动率的静态估计	44
3.3 GARCH-X 模型框架	48
3.4 模型预测能力的评价	51
3.4.1 对波动率的预测评价	51
3.4.2 对风险价值的预测评价	53

3.5	实证分析	54
3.6	小结	60
第 4 章	基于低频价格极差的异质自回归波动率模型	61
4.1	引言	61
4.2	异质市场假说	63
4.3	基于价格极差的异质自回归模型	67
4.3.1	异质自相关分析	67
4.3.2	HAR-P 模型	70
4.4	基于价格极差的异质主成分模型	71
4.4.1	异质主成分分析	71
4.4.2	HPC-P 模型	73
4.5	实证分析	74
4.5.1	与低频波动率模型比较	75
4.5.2	与高频波动率模型比较	79
4.6	小结	80
第 5 章	基于高频双格极差的波动率估计	82
5.1	引言	82
5.2	基本理论框架	84
5.2.1	已实现双格极差波动率	85
5.2.2	已实现双格极差波动率的渐近性质	87
5.3	基于高频双格极差的波动率估计	90
5.3.1	已实现信息波动率	90
5.3.2	已实现信息波动率的渐近性质	92
5.4	微观结构噪声的影响	96
5.5	随机模拟	98
5.6	实证分析	103
5.7	小结	106
5.8	附录	107

第 6 章 基于高频双格极差的价格跳跃检验	123
6.1 引言	123
6.2 基本理论框架	124
6.3 基于高频双格极差的价格跳跃检验	126
6.4 基于高频价格极值的单向跳跃检验	129
6.5 随机模拟	132
6.5.1 双向跳跃检验	133
6.5.2 单向跳跃检验	138
6.6 实证分析	141
6.7 小结	145
6.8 附录	146
参考文献	150

表 格

3.1	几种不同的低频波动率估计	48
3.2	波动率预测的损失函数平均值比较	56
3.3	模型预测能力的 Diebold-Mariano 检验结果	57
3.4	Hansen 的超预测能力 (SPA) 检验结果	58
3.5	对 VaR 预测结果的评估	59
4.1	普通自相关系数与异质自相关系数	68
4.2	异质效应与主成分的对应关系	73
4.3	波动率预测的损失函数平均值比较	76
4.4	模型预测能力的 Diebold-Mariano 检验结果	77
4.5	Hansen 的超预测能力 (SPA) 检验结果	78
4.6	与高频波动率模型的比较结果	79
5.1	IV 估计量对比	95
5.2	IQ 估计量对比	96
5.3	IV 和 IQ 估计的 MSE	100
5.4	IV 估计的统计特征	102
5.5	股指期货波动率估计的描述统计	105
5.6	股指期货波动率估计的动态预测评价	106
6.1	跳跃检验统计量的有限样本性质	134
6.2	单向跳跃检验统计量的有限样本性质	139
6.3	实证数据的跳跃检验结果	143

插 图

2.1	布朗运动路径及其反射路径	33
3.1	无交易时段	47
4.1	普通自相关图 vs. 异质自相关图	69
4.2	异质主成分的系数值	72
5.1	股指期货不同采样频率的波动率估计	103
6.1	跳跃检验的水平与功效	135
6.2	离散化报价下跳跃检验的水平与功效	137
6.3	单向跳跃检验的水平与功效	140
6.4	离散化报价下单向跳跃检验的水平与功效	141
6.5	零收益率比例与成交比例	142
6.6	实证数据的跳跃检验示例	144

第1章 绪 论

1.1 波动率建模的历史背景

金融资产价格的波动率 (Volatility) 作为收益率变异程度的一种统计度量^①, 在资产定价、投资决策和风险管理等问题上都扮演着极其重要的角色。因此, 从 20 世纪 80 年代初期开始, 对波动率的估计和预测一直是金融计量学研究中备受关注的问题。从波动率的研究历程来看, 大致可以分为如下三个阶段:

第一阶段是静态假设, 波动率定义为资产收益率的方差。在经典的 Black-Scholes 模型框架下, 资产的价格被设定为服从几何布朗运动, 此时漂移系数和扩散系数都是常数, 波动率即扩散系数的平方也是一个常数。因此在波动率研究的初期, 人们关注的是一个参数估计问题, 即对资产价格扩散系数的估计。

第二阶段是局部动态假设, 波动率定义为资产收益率关于历史信息的条件方差。传统的静态波动率假设不能解释资产价格变化中的“波动簇集”(volatility clustering) 和收益率分布的厚尾特征等现象, 因此人们意识到资产的波动率是动态变化的。如何描述这种动态特征并由此给出波动率的预测成为这类研究的核心问题, 各种 GARCH 类以及随机波动率的局部动态模型被提出来, 参见 Zivot(2009) 以及 Broto 和 Ruiz(2005) 的综述。

第三阶段是高频动态假设, 波动率定义为资产收益率的二次变差, 即积分波动率 (Integrated Volatility)。20 世纪 90 年代以前, 对于金融时间序列的研究通常是以日度或更低频率的数据为主要

^① 为统一起见, 本书中所提到的波动率均指其度量的二次形式。

对象进行的。近年来,随着电子化以及数据存储技术的快速发展,日内高频数据的获取与使用变得广泛普及,基于高频数据的波动率研究成为当前的一个热点。在这样的背景下,资产价格被设定为更一般的布朗运动半鞅过程,利用高频数据可以给出波动率的更为准确的非参数估计方法,参见 McAleer 和 Medeiros(2008) 的综述。

从以上几个阶段对波动率的定义来看,由资产收益率的二阶矩来给出波动率的估计或预测方法是直观的,这也是上述几个阶段中被广泛采用的一种主流思路。但是,在波动率研究的早期文献里, Parkinson(1980) 和 Garman-Klass(1980) 就曾经指出,充分利用开盘价、最高价、最低价和收盘价 (OHLC) 可以构造出比收益率平方更精确的静态波动率估计,这类研究波动率的方法称为极值方法 (extreme value method)。对应于上述波动率研究的三个阶段,这类极值方法在第一阶段中的研究已经较为全面,而在后面两个阶段的研究则相对缺乏,需要进一步的探索与完善。

1.2 金融极值数据波动率建模

金融极值数据是对一段时间内的开盘价、最高价、最低价和收盘价数据的总称,这里的“一段时间”既可以是日、周、月、季等低频时间段,也可以是小时、分钟或逐笔的高频时间段。本书的研究以价格极值为主线进行展开,一方面考虑基于低频价格极值数据的静态波动率估计和动态波动率预测,另一方面探索基于高频价格极值数据的波动率非参数估计和价格跳跃检验方法。

1.2.1 基于低频数据的研究

已有的基于低频价格极值数据的研究,主要集中在静态波动率估计的问题上,其中比较典型的有 Parkinson(1980) 利用价格的极差 (Range) 给出的估计;以及 Garman 和 Klass(1980) 综合

利用 OHLC 数据给出的最小方差无偏估计二次型, 它们的估计效率分别是收益率平方的 4.9 倍和 7.4 倍。当漂移系数不为零时, 上述波动率估计有偏, 于是 Kunimoto(1992) 以及 Rogers 和 Satchell(1994) 进一步利用价格极值构造了相应的无偏估计。除此之外, 还有 Meilijson(2008), Floros(2009) 以及 Fiszeder(2012) 等最新研究也考虑了各种形式的基于价格极值的静态波动率估计。

针对局部动态波动率的预测问题, 比较常见的是 GARCH 类和随机波动率模型, 但这类模型仅仅使用到了低频的收盘价数据。在高频数据不可获取的情况下, 比较直接的想法是如何充分利用低频的价格极值信息构建波动率的动态预测模型。相比波动率的静态估计而言, 价格极值在波动率的动态预测问题上的研究相对单一, 基本上都是针对价格极差这一种估计形式进行的, 比如 Chou(2005) 的条件自回归极差模型 (CARR)、Brandt 和 Jones(2006) 基于价格极差的 EGARCH 模型, 还有 Fuertes 和 Olmo(2013) 也提出了类似的价格极差动态模型。在使用相同数据的情况下, 只是简单模型形式的区别可能并不能显著改善波动率的预测效果 (Hansen 和 Lunde, 2005a)。为此, 本书考虑了如下两个方面的扩展:

1. 构建基于 OHLC 数据的动态波动率模型

GARCH 模型实际上是将波动率表达成为过去收益率平方的加权, 那么考虑将收益率平方替换为其他静态估计形式嵌入 GARCH 模型中是一个自然的推广, 这样得到的模型称为 GARCH-X 模型。在前期的工作当中, 我们曾将 Parkinson 估计嵌入到 GARCH 模型中提出 GARCH-P 模型, 该模型与传统的 GARCH 模型相比对波动率的预测有显著改善。本书的研究在此基础上, 一方面进一步考虑其他形式的估计对应的 GARCH-X 模型, 另一方面考虑隔夜效应即加入开盘价后对波动率预测的影响。在同一模型形式的条件下, 考察不同数据信息的使用是否能

够显著改善波动率的预测效果。

2. 引入异质市场假说改进基于价格极值的动态波动率

Corsi (2009) 利用异质市场假说 (Heterogenous Market Hypothesis) 解释了波动率“长记忆性”，并进一步提出了基于高频数据的 HAR-RV 模型，改进了波动率的预测效果。异质市场假说具体到波动率的预测问题上，实际上是强调了波动率时间尺度的重要性，比如 Corsi(2009) 同时利用了日、周、月频度的波动率来刻画“长记忆性”并改善波动率的预测效果。针对低频数据的情况，长期的价格极差并不能由短期的价格极差简单相加得到，即这两种极差并不能相互包含对方的信息，因此异质市场假说认为考虑了波动率度量的时间尺度就等同于信息的扩展。我们沿用异质市场假说的基本思想，结合价格极差的特性，构建基于价格极差的异质自回归模型与异质主成分模型，并考察由此得到的波动率预测能否显著超越其他低频波动率模型的预测效果。

1.2.2 基于高频数据的研究

已有的基于高频价格极值数据的研究，基本上是针对高频价格极差并按照已实现波动率 (Realized Volatility) 的研究框架进行的。Martens 和 Van Dijk(2007) 以及 Christensen 和 Podolskij(2007) 首先提出了已实现极差波动率 (Realized Range-based Volaitlity)，作为积分波动率(Integrated Volaliltlity) 的一致估计，其效率是已实现波动率的 4.9 倍。在考虑价格跳跃的情况下，Christensen 和 Podolskij(2006) 又提出了双幂次极差变差 (Range-based Bipower Variation)，其效率是仅基于高频收盘数据的双幂次变差 (Bipower Variation) 的 7.2 倍。除此之外，Christensen et al.(2009) 还考虑了微观结构噪声存在时对已实现极差波动率以及相应跳跃检验方法的修正。

上述一系列研究的基本思路基本上都是将高频区间上的收益率平方直接替换为对应的 Parkinson 估计, 因此相应的推导模式和理论结果都与已实现波动率的框架类似, 总体而言, 没有充分利用价格极差的高频特性。事实上, 两种不同采样频率下的价格极差 (称为双格价格极差) 包含了不同的价格波动信息, 但是却包含了几乎相同的价格跳跃信息。本书的研究从这一思路出发, 从如下两个角度探索高频价格极值的理论和应用:

1. 利用高频双格极差构造波动率的跳跃稳健估计

对于跳跃稳健估计量的构造, Barndorff-Nielsen 和 Shephard (2004) 以及 Christensen 和 Podolskij(2006) 的基本思想是两个相邻的高频时间区间几乎不可能同时出现跳跃, 而本书的基本思想是同一高频区间不同的划分对应的价格极差之间可以近似地抵消跳跃的影响。因此, 这是两种完全不同的思路, 由此得到的波动率估计具有不同的理论性质和实证表现, 我们可以从估计的效率、跳跃的收敛速度以及动态预测效果等方面来进行比较。

2. 利用高频双格极差构造价格跳跃检验方法

在得到波动率的跳跃稳健估计以后, 给出相应的价格跳跃检验统计量是直接的, 因此跳跃检验的理论性质由相应波动率估计的理论性质所决定。由于传统的跳跃稳健波动率估计具有“连乘”的形式, 容易受到“零收益率”的影响导致低估, 因此其对应的跳跃检验倾向于更大的第一类错误概率, 而由本书提出的双格价格极差给出的跳跃检验则倾向于更稳健的表现。另外, 考虑高频区间上最高价和开盘价的差值只反映正向跳跃, 同理最低价和开盘价的差值只反映负向跳跃, 因此本书进一步利用高频价格极值的这一特性给出单向跳跃的检验方法。