

文章编号: 1003-207(2009)06-0001-08

基于价格极差的金融波动率建模: 理论与实证分析

李红权^{1,2}, 汪寿阳²

(1. 湖南师范大学商学院, 湖南 长沙 410081; 2. 中国科学院数学与系统科学研究院, 北京 100190)

摘要:经典的波动率模型(GARCH等)是收益率为基础的模型,利用的是收盘价信息,忽略了价格波动的日内信息,这将导致信息与效率的损失。为了弥补这一缺陷并获得满意的波动率预测效果,本文引入并扩展了基于价格极差的自回归波动率模型。实证研究表明新模型能够有效刻画波动率的动态变化规律,其预测效果一致性地优于经典的GARCH模型。同时,我们的研究还证实了在波动率模型中加入收益率的滞后项能够提高模型的解释能力,并且存在明显的“杠杆效应”。

关键词:波动率建模;价格极差;日内信息;预测绩效

中图分类号:F713

文献标识码:A

1 引言

金融市场的典型特征之一就是其复杂多变的波动性。在金融学领域,波动率是资产定价与风险管理理论的核心变量;在现实中,资产波动状况和其收益率是紧密相连的,它直接决定了投资者的投资损益,因而金融市场的波动率是市场参与各方关注的焦点。然而,遗憾的是,资产的真实波动率或内在波动率(Latent Volatility)是未知的,我们只能通过可观测到的波动率代理变量(Volatility Proxy)去评估真实波动率(会导致测量误差),另一方面我们还必须对未来的波动率做出评估和预测,这就需要有良好的波动率预测模型(Volatility Model)。

经典资本市场理论(始于Markowitz(1952))引入方差(Variance)来描述股票市场收益率变化,所采用的理论模型一般都假定收益率方差保持不变。这些模型传承了金融市场中有效市场假说的理念,运用简便,常用来预测和估算股票价格的变动。但对金融数据的大量实证研究表明,有些假设不甚合

理。如对股票收益率建模,其随机扰动项往往在较大幅度波动后面伴随着较大幅度的波动,在较小波动幅度后面紧接着较小幅度的波动,这种性质称为波动率聚丛(Volatility Clustering)。该现象的出现源于外部冲击对股价波动的持续性影响,在收益率的分布上则表现为出尖峰厚尾(Fat Tails)的特征。这类序列随机扰动项的无条件方差是常量,条件方差是变化的量。Engle(1982)、Bollerslev(1986)、Nelson(1991)等人提出的ARCH/GARCH模型组很好地捕捉了波动率的时变性^[1-3]。相对于GARCH模型组,Taylor(1986)提出的随机波动率模型(Stochastic Volatility, SV)^[4]也能反映条件方差的时变性,并且要比前者更有弹性。另外,为了更好地描述波动率的非线性动力学特征如波动率的长期记忆性(Long Memory),研究人员在ARCH模型组和SV模型基础上又提出了相应的分数差分模型及双线性ARCH模型等。

虽然经典的波动率模型GARCH、SV等显著增强了我们对于资产波动率与金融市场的认识,然而它们都是基于收益率的模型(Return-based model)。常用的对数收益率被定义为金融资产收盘价对数的一阶差分,而收盘价仅仅是一个时点上的资产价格取值(观测点),没用充分利用大量的日内价格信息(样本区间),这将导致信息与效率的损失。为了弥补这一缺陷,学术界提出了已实现波动率(Realized Volatility,记为RV, Andersen等, 2001,

收稿日期:2008-03-11;修订日期:2009-11-23

基金项目:国家自然科学基金委员会优秀创新研究群体基金(70221001);教育部人文社会科学研究项目;湖南师范大学社会科学青年学术骨干培养计划(基金)

作者简介:李红权(1976-),男(汉族),河南南阳人,中国科学院数学与系统科学研究院管理学博士,博士后,湖南师范大学商学院副教授,研究方向:金融工程与风险管理。

2003)^[5,6]和基于价格极值(或称极差(Range),即在固定样本区间如1天或1周内的最高价对数与最低价对数之差)的波动率(Garman 和 Klass, 1980; Parkinson, 1980; Rogers 和 Satchell, 1991; Yang 和 Zhang, 2000; Christensen 和 Podolskij, 2007)^[7-11]。虽然基于日内高频收益率平方和计算的已实现波动率RV是非常有前景的波动率代理变量,然而基于日内极值的波动率却被证明在某些方面更有优势。两者的共同点都是利用了日内的价格路径信息,在理论上被证明要比样本方差等基于日收益率的波动率代理变量有效的多,而且都是无偏估计量;但RV在真实金融环境中的表现却逊色于基于日内极值的波动率(Alizadeh 等, 2002)^[12],主要原因是在高频数据环境下,市场微观结构噪声对RV的影响是相当显著的,而且这种影响会随频率的增加而增加。尽管研究人员试图消除微观噪声的影响(例如,Zhang 等, 2005; Barndorff-Nielsen 等, 2006)^[13,14],然而对于二次变差的精确估计量仍是未知的(Christensen 和 Podolskij, 2007)^[11]。而后者(基于极值的波动率测度)对微观结构噪声具有鲁棒性(Robustness)。

理论上,基于日内极值的波动率(Extreme-Value Volatility, 记为EV)可以作为有效的代理变量,极差在实际金融市场中也被广泛运用(如K线图分析)。然而,比较有挑战性的问题集中在如何对基于极值的波动率进行有效建模以获得满意的预测效果(Alizadeh 等, 2002; 周雨田(Chou, 2005); Brandt 和 Jones, 2006)^[12,15,16]。

在基于日内信息的波动率测度方面,国内学术界也展开了相关的研究。典型的研究成果如黄后川和陈浪南(2003)、唐勇和张世英(2006)、王春峰(2006, 2007)等^[17-20]。多数的研究侧重于讨论高频波动率的测度,特别是如何减小噪声对于RV的影响等;关于高频波动率建模与预测的研究亟待加强。

鉴于此,本章拟对基于日内极值的波动率(EV)度量与建模做进一步的深入研究,探索基于极值波动率的有效建模方法,以获得良好的样本内拟合效果与优良的样本外预测性能;并应用于金融市场波动率的有效测度与管理。

2 波动率建模研究

对投机资产收益率行为的建模是金融经济学与计量经济学领域关注的焦点问题之一。最简单的处理方式是假定收益率是一个零均值的正态随机变

量。虽然收益率零均值的假定得到了多数金融实证的支持,然而实证研究却表明收益分布是负偏的、尖峰态的非高斯分布。尖峰态可以由条件异方差建模来解释。在金融实证领域,研究人员常将收益率过程表示为:

$$r_t = \sigma_t \varepsilon_t, \varepsilon_t \sim i.i.d.(0, 1) \quad (1)$$

在公式(1)中, ε_t 为一零均值的白噪声过程且服从正态分布, σ_t 为波动率且是时变的。这个模型考虑了波动率的动态变化即时变性,忽略均值过程可能存在的时变性,这种处理方式被广泛采用(比如J.P.Morgan开发的RiskMetrics系统)。假定 ε_t 是高斯白噪声,那么收益率的条件分布也将服从正态分布。虽然条件分布是正态的,但由于波动率是时变的,收益率的无条件分布将是尖峰态的。对波动率过程的不同定义决定了模型的不同形式。

2.1 ARCH类模型

ARCH模型是金融建模的常用工具,它能够描述金融市场的典型特征特征即条件异方差现象。ARCH模型始于Engle(1982),后经过不断发展成为ARCH/GARCH模型组。其中,最常用的是Bollerslev(1986)提出的GARCH(1,1)模型:

$$(2a) \quad r_t = \sigma_t \varepsilon_t, \varepsilon_t \sim i.i.d. N(0, 1)$$

$$(2b) \quad \sigma_t^2 = \gamma + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (2)$$

ε_t 是正态的,收益率的条件分布也是正态的,条件方差和前期信息有关。参数 α 和 β 决定了模型的平稳性条件($\alpha + \beta < 1$)。虽然收益率是不相关的,收益率的平方却是相关的,服从自回归过程(具体而言是ARMA(1,1)过程)。收益率的无条件分布是关于零均值对称的,无条件方差为 $\gamma/(1 - \alpha - \beta)$,分布的峰态值大于3(尖峰态分布)。模型组中的ARCH(p)与GARCH(p,q)也具有类似的性质。在对标准GARCH模型的改进中,指数GARCH模型或称EGARCH模型是比较有代表性的一类模型。EGARCH模型由Nelson(1991)提出,它能够解释收益率对条件方差的非对称影响。典型的EGARCH(1,1)模型中方差项为:

$$\ln \sigma_t^2 = \gamma + \beta \ln \sigma_{t-1}^2 + \alpha \left| \frac{r_{t-1}}{\sigma_{t-1}} \right| + \omega \frac{r_{t-1}}{\sigma_{t-1}} \quad (3)$$

其中,参数 ω 代表了收益率对于条件方差的非对称冲击。方差的对数形式确保了条件方差为正。

由于正态分布的假定,研究人员可以很方便地用极大似然估计方法来估计GARCH模型的参数。释放正态分布的假定比如假设 ε_t 服从学生t分布,这会增强模型捕捉收益率分布厚尾特征的能力,但

模型参数的估计却更加困难了。GARCH 模型中方差由条件方差方程给出,事实上条件方差的一步预测值是由方程(2b)完全确定的,这一简明的处理是 GARCH 模型区别于其它波动率模型的主要特征。

2.2 基于极差的自回归波动率模型(Range-based Autoregressive Volatility Model, 记为 AV)

传统的波动率度量方式仅仅利用了资产价格的收盘信息,而极差波动率则充分利用了整个价格形成过程的信息量,这使得基于极差的波动率(EV)具有相对优越性。例如,当天的收盘价和昨日收盘价相等时,收益率为 0,然而当日的价格波动却有可能是非常剧烈的,传统波动率度量手段反映不了这一事实,而包含了日内信息量的极差波动率则可以有效地反映其波动状况。理论研究结果均表明基于极差的波动率要比基于日收益率的传统波动率代理变量更有效、更优越。Beckers (1983)^[21]的研究表明:如果包含了日内最高价与最低价信息,波动率估计的有效性将得到显著性提高。最近的研究结果(Alizadeh 等, 2002; Shu 和 Zhang, 2006)^[12, 22]从已实现波动率的角度也给出了支持极差波动率 EV 的有利证据:EV 作为真实波动率的代理变量是无偏、高度有效的,而且对于市场微观结构噪声(比如买卖价差)具有很好的鲁棒性。EV 的有效性与简明性奠定了它作为波动率度量方式的理论基础。经典的极差波动率代理变量是由 Parkinson (1980)提出的,其表达式为:

$$\sigma_{\text{RNG}}^2 = \left[\frac{1}{4 \ln 2} \right] (\ln H_t - \ln L_t)^2 \quad (4)$$

其中, H_t 和 L_t 分别是日内(或周内、月内)的最高价和最低价。Parkinson 假定资产价格服从无漂移的几何布朗运动,并在理论上证明了这个波动率代理变量的有效性是传统度量方式(基于收盘价的)的 5.2 倍。Garman 和 Klass (1980), Beckers (1983), Rogers 和 Satchell (1991) 及 Yang 和 Zhang (2000) 等在此基础上提出了多种修正后的极差波动率,分别考虑了收盘价、开盘效应、漂移率以及其它因素。

虽然基于极差的波动率在理论上及模拟实验研究中被证明是有效的,但是在实际波动率预测中的表现却不尽人意。主要的原因在于常用的波动率模型没能捕捉波动率的动态变化(Chou, 2005)。基于此,本文引入并扩展了基于极差的自回归波动率模型(Autoregressive Volatility Model, 简记为 AV)以对波动率的动态变化进行有效建模。

AV 模型最初是由 Hsieh (1991) 提出的

Hsieh (1991, 1993, 1995)^[23-25]的研究表明 AV 模型能够有效刻画波动率的动态特征。模型的具体形式如下:

$$(5a) \quad r_t = \sigma_{\text{RNG}, t} e_t, e_t \sim i.i.d.(0, 1)$$

$$(5b) \quad \ln \sigma_{\text{RNG}, t}^2 = \alpha + \sum \beta_i \ln \sigma_{\text{RNG}, t-i}^2 + v_t, v_t \sim i.i.d.(0, \sigma_v^2) \quad (5)$$

其中 $\sigma_{\text{RNG}, t}$ 是基于极差的波动率代理变量。Hsieh 提出 AV 模型的主要动因之一是发现波动率是高度自相关的。从公式(5b)可以看出,当期波动率和前期的波动率有关,通过最小二乘方法 OLS 可以估计出方程(5b)中的参数。简化起见,我们假定 v_t 和 e_t 是相互独立的。

我们可以很容易地扩展条件方差方程(5b)以包含其它的解释变量。例如,我们将方程(5b)扩展成方程(6)的形式以解释收益率冲击对条件方差的非对称影响

$$\ln \sigma_{\text{RNG}, t}^2 = \alpha_0 + \sum \beta_i \ln \sigma_{\text{RNG}, t-i}^2 + \alpha_1 \left| \frac{r_{t-1}}{\sigma_{\text{RNG}, t-1}} \right| + \alpha_2 \cdot \frac{r_{t-1}}{\sigma_{\text{RNG}, t-1}} + v_t, v_t \sim i.i.d.(0, \sigma_v^2) \quad (6)$$

这是一个非对称的自回归波动率模型(简记为 AV- α)。参数 α_2 量化了非对称性的影响。

与经典的 GARCH 模型相比较, AV 模型有三个主要的特点:(1)研究表明 AV 模型的波动率持续性要比 GARCH 模型的弱一些,这使得 AV 模型能够更好地捕捉波动率的动态变化趋势;(2)使用极大似然方法估计 GARCH 模型时需要误差项 e_t 的分布作出特定的假设,而 AV 模型的估计不需要任何分布假设;(3)AV 模型的波动率方程含有一个随机项(与随机波动率 SV 模型类似),这使得 AV 模型更一般化也更具弹性。此外与 Chou (2005) 提出的 CARR 模型(借用了 GARCH 模型的形式)相比, AV 模型的参数估计更加便捷、模型形式更具弹性、能够充分捕捉波动率的均值回复行为(CARR 模型侧重于波动率的持续性行为描述)。

3 数据与模型参数估计

3.1 数据及其基本特征

本文的数据是 S&P500 指数的日度最高、最低、开盘、收盘价(数据来源于雅虎财经)。数据集包括 1994 年 5 月 27 日至 2008 年 4 月 22 日近 14 年的日度数据,得到 700 个周度数据点(每周定义为 5 个连续交易日,不是“日历周”)。前 600 个数据点作为估计样本,最后 100 个数据(April 26,

2006——April 22, 2008)作为样本外预测区间。

图 1 显示的是 S&P500 指数周度收益率, 直观的一个判断是收益率在零均值左右变动、且有明显的波动集从显现。表 1 是数据样本的基本统计特征。其中 ADF 检验的结果表明收益率序列是平稳的。偏度为负、峰度大于 3、J-B 统计量拒绝了正态分布的原假设, 这都说明了收益率分布是显著偏离正态分布的, 这和多数金融实证的结果一致。ARCH 检验的结果也证实了收益率过程确实存在波动集从特征。对于收益率平方序列的 LB Q 检验说明收益率过程存在非线性相关结构; 而对于收益率序列的 Q 检验值较小, 这表明收益率序列的相关

性结构主要是非线性相关的。

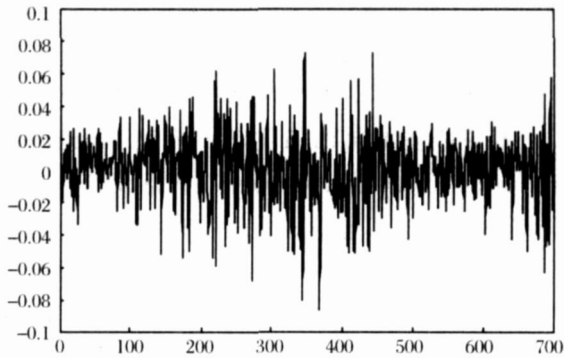


图 1 S&P500 指数周度收益率序列
(May 27, 1994—April 22, 2008)

表 1 S&P500 指数周度收益率的基本统计特征

Mean	Std. Dev.	Skewness	Kurtosis	Jar. Bra	ADF	Q(15)	QS(15)	ARCH(5)
0.0016	0.0218	-0.36*	3.93*	39.80*	-27.63	20.44	146.93*	56.71*
0.0018	0.0221	-0.34*	3.92*	32.83*	-25.47	22.72	125.59*	50.93*

注: * 代表显著性水平为 1%。第二行是全样本区间的统计数据; 第三行是样本内估计区间的统计量。ADF 检验包括截距项与趋势项; ADF 检验的临界值为 -3.43(1%)和 -2.86(5%)。Q(15)和 QS(15) 分别代表的是 r_t 和 r_t^2 序列的 Ljung-Box Q 统计量。Ljung-Box Q 检验的临界值(滞后阶数为 15): 25.00 (5%)、30.58 (1%)。

图 2 显示的是周度极差波动率序列。表 4.2 是不同波动率代理变量的统计特征。由图表可知, 周度波动率的均值约为 2%, 也就是说年化波动率为 14%。波动率变量的标准差为 1%说明 S&P500 波动率本身的波动性是比较剧烈的。对比不同波动率代理变量(收益率均值 $|r_t|$ 、极差波动率 σ_{RNG})的自相关系数(ACF)与 Ljung-Box Q 统计量, 我们会发现 σ_{RNG} 序列的相关系数与检验统计量的值都远远大于 $|r_t|$ 序列的值, 这表明极差波动率序列的相关性程度非常强烈。这也是我们在波动率建模与预测中采用自回归模型 AV 的原因之一。

EGARCH(1,1)是非常常用的两类 ARCH 模型, 也是作为波动率预测的理想比较基准。

表 2 波动率代理变量的统计特征

Statistic	Mean	Max.	Min.	Std. Dev.	ACF (1)	ACF (15)	Q (15)
$ r_t $	1.67	8.67	0.00	1.40	0.185	0.076	215.26
σ_{RNG}	1.95	7.27	0.42	1.06	0.609	0.247	1664.79

3.2 GARCH 模型估计

GARCH(1,1)模型的参数估计可以由极大似然方法得到, 表 3 给出了参数估计的详细结果。

首先参数 α 和 β 的估计值之和小于但非常接近 1, 这表明序列即便是平稳的, 方差的持续性也非常强。另外, 参数估计的值和标准风险管理模型 RiskMetrics 设定的值非常接近, RiskMetrics 本质上使用的是 IGARCH 模型即设定 $\alpha + \beta = 1$ 。估计出参数的值, 再由方程(2b)我们可以直接得到条件方差的一步预测值。GARCH(1,1)模型成功地解释了序列中的二阶矩相依性, 这一点可以从表 7 中的统计结果得到验证。表 7 中, 对于 GARCH 模型标准残差平方的 Q 统计量表明残差序列是不相关的, GARCH 模型已经捕捉了收益率序列的非线性相关性; ARCH 检验的结果也表明残差序列不再具有 ARCH 效应。表 7 中的结果还表明, 即使标准残差序列的峰度或偏度有所降低(这体现在 J-B 统计量的值显著减小), 但仍不符合正态分布的原假设。

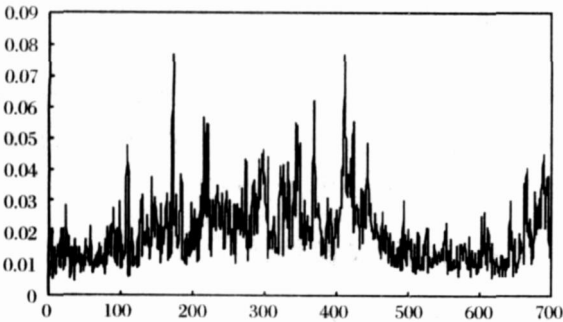


图 2 S&P500 指数周度波动率序列
(May 27, 1994—April 22, 2008)

下面我们将估计并比较本文中的四个波动率模型: GARCH(1,1)、EGARCH(1,1)、基于极差波动率的 AV 与 AV - α 模型。GARCH(1,1)和

这个结果并不令人惊奇,多数实证研究的结果也表明 GARCH 只能部分地解释收益率的厚尾分布特征(Hsieh, 1991)。

EGARCH(1,1)模型参数的估计结果见表 4。模型满足平稳性条件($\beta < 1$);非对称项的系数 ω 显著不等于 0 且为负值,说明收益率冲击对于条件波动率的影响具有非对称性,负收益对波动率的冲击要大些。EGARCH 模型的标准残差也是独立不相关的,不再含有 ARCH 效应,这和 GARCH(1,1)模型的分析类似(详见表 7)。

表 3 GARCH(1,1) 模型参数估计

Parameter	Estimate	Std. Error	p-value
γ	0.0000	0.0000	0.0622
α	0.0687	0.0175	0.0001
β	0.9215	0.0199	0.0000
$\alpha + \beta$	0.99		

表 4 EGARCH(1,1)模型参数估计

Parameter	Estimate	Std. Error	p-value
γ	-0.4109	0.1026	0.0001
α	0.1341	0.0349	0.0001
β	0.9606	0.0113	0.0000
ω	-0.1008	0.0274	0.0002

3.3 AV 模型估计

基于极差波动率的 AV 模型和 AV - α 模型的 OLS 估计结果分别见表 5 和表 6。AV 模型组的最优滞后阶数由 Schwarz 信息准则确定。波动率的持续性由回归参数 β_i 之和来度量,分别是 0.80(AV 模型)、0.89(AV - α 模型)。它们的值均小于 1,说明波动率过程是严格平稳的。和 GARCH 模型的估计结果(0.99)比较,AV 模型的波动率持续性较低。这一结果和 Hsieh (1995)的研究结论一致,Hsieh (1995)的研究表明 GARCH 模型倾向于高估实际波动率的持续性,而 AV 模型能够比较好地描述波动率过程的动态行为包括波动率集从特征与波

表 5 AV 模型参数估计

Parameter	Estimate	Std. Error	p-value
α	-1.5762	0.2517	0.0000
β_1	0.3277	0.0380	0.0000
β_2	0.1923	0.0356	0.0000
β_3	0.1981	0.0374	0.0000
β_4	0.0878	0.0369	0.0178
$\sum \beta_i$	0.80		

动率的均值回复行为。AV 模型的优越性能在表 7 和表 8 以及波动率样本外预测结果中都得到了实证

的支持。表 7 中 Q 统计量和 ARCH 检验证实了 AV 模型能够捕捉序列的非线性相关性、解释异方差现象,AV 模型标准化的收益率序列是独立的。

表 6 AV - α 模型参数估计

Parameter	Estimate	Std. Error	p-value
α_0	-1.4708	0.2030	0.0000
β_1	0.5166	0.0282	0.0000
β_2	0.1844	0.0298	0.0000
β_3	0.1388	0.0275	0.0000
β_4	0.0541	0.0288	0.0609
α_1	0.6405	0.0245	0.0000
α_2	-0.0835	0.0137	0.0000

表 7 样本内基于标准化收益率的统计检验

	GARCH (1, 1)	EGARCH (1, 1)	AV	AV - α
JB	14.56 (0.00)	15.32 (0.00)	18.71 (0.00)	31.10 (0.00)
ARCH (1)	0.68 (0.78)	0.01 (0.94)	0.17 (0.68)	0.07 (0.78)
ARCH (5)	7.37 (0.19)	6.96 (0.22)	8.10 (0.15)	8.97 (0.11)
Q (15)	17.33 (0.30)	17.14 (0.31)	16.64 (0.34)	20.15 (0.17)
QS (15)	13.85 (0.54)	16.62 (0.34)	18.55 (0.24)	20.21 (0.16)

注:括号内为 p 值;JB 是 Jarque - Bera 统计量;ARCH(1) 和 ARCH(5)是 ARCH 检验统计量;Q(15) 和 QS(15)分别代表对标准化收益率及其平方序列的 Ljung-Box Q 统计量,15 为滞后阶数。

为了比较 GARCH 模型和 AV 模型的样本内预测效果,我们采用 R^2 来作为模型建模效果的评价指标。对于方程:

$$\alpha_t = \gamma_0 + \gamma_1 \hat{\alpha}_t + \epsilon_t \tag{7}$$

这个回归方程的 R^2 表明了依据波动率模型得到的波动率预测值($\hat{\alpha}_t$)对真实波动率(α_t)的解释程度。回归方程的 R^2 越高,说明波动率预测值对真实波动率的解释能力越强,也就意味着与之对应的波动率模型的建模效果越佳。表 8 是不同已实现波动率条件下方程(7)的回归结果,包括参数值与回归的 R^2 值。表 4.8 的结果表明:不论以那个指标来度量真实波动率,AV 模型的 R^2 都比 GARCH 模型的高;相对于 GARCH 模型,AV 模型波动率的回归方程参数更接近其无偏性条件即 $\gamma_0 = 0, \gamma_1 = 1$,说明 AV 模型的波动率建模效果要优于 GARCH 模型,其误差更小;此外,包含了不对称性因素的模型要比原始模型更好,比如 EGARCH 模型的建模效果(R^2)要好于 GARCH、AV - α 模型的效果要优于 AV 模型;在所有的模型中,AV - α 模型的表现最

优,这体现在具有对真实波动率最好的解释力(R^2 最高)以及最接近回归方程的无偏性条件($\gamma_0 = 0$ 、 $\gamma_1 = 1$)。

图 8 样本内波动率预测效果评价

	γ_0	γ_1	R^2
Panel A· 已实现波动率测度为 σ_{RNG}			
GARCH(1,1)	-0.00(-0.82)	0.97(16.17)	0.30
EGARCH(1,1)	-0.00(-3.25)	1.13(20.38)	0.41
AV	-0.00(-0.02)	1.07(21.64)	0.44
AV - α	-0.00(-0.21)	1.05(36.37)	0.69
Panel B· 已实现波动率测度为 σ_{SSDR}			
GARCH(1,1)	-0.00(-1.78)	1.10(16.52)	0.31
EGARCH(1,1)	-0.01(-4.56)	1.29(21.18)	0.43
AV	-0.00(-1.06)	1.20(21.72)	0.44
AV - α	-0.00(-1.40)	1.05(27.04)	0.55

注:括号内是 T 统计量(采用 Newey-West 方法计算标准差)。已实现波动率 σ_{RNG} 是基于极差(周内最高价、最低价)的波动率代理变量(见公式 4)。SSDR 代表周内的日收益率平方和, σ_{SSDR} 是 SSDR 的平方根。

4 模型预测效果评价

波动率模型的建模效果最终要看模型的样本外预测能力。为了客观评价模型的预测效果,我们使用多种评价方法与评价指标。在已有的预测评价指标中,比较有代表性的是均方根误差(RMSE)、绝对误差均值(MAE)、Theil-U 统计量及前述的回归方法 R^2 统计量,其定义分别如下:

$$RMSE = \sqrt{\frac{1}{I} \sum_{i=1}^I (\sigma_i - \hat{\sigma}_i)^2} \quad (8)$$

$$MAE = \frac{1}{I} \sum_{i=1}^I |\sigma_i - \hat{\sigma}_i| \quad (9)$$

$$Theil-U = \frac{\sum_{i=1}^I (\hat{\sigma}_i - \sigma_i)^2}{\sum_{i=1}^I (\sigma_{i-1} - \sigma_i)^2} \quad (10)$$

在这里, $\hat{\sigma}_i$ 是由波动率模型得到的预测值; σ_i 是真实波动率或已实现波动率,本文中真实波动率的测度方式有两个:一个是基于周内极差的已实现波动率,记为 σ_{RNG} (见公式 4);另外一个是基于日度收益率的已实现波动率,记为 σ_{SSDR} ,它是周内日收益率平方和的二次方根。

RMSE 和 MAE 是最常用的预测评价指标,它们具有数学上的简洁性,但容易受量纲与比例转换的影响。Theil-U 统计量克服了这一缺陷,它是个不变量,是评价不同预测模型精度的一个比较合适的指标(Armstrong and Fildes, 1995)^[26]。在 Theil-U 统计量中,随机游走模型的预测误差是分

母(作为标准化的基础)。如果某一特定模型的预测误差小于随机游走模型的天真预测误差,那么 Theil-U 统计量的值将小于 1。表 9 给出了评价结果。

表 9 中的结果表明:不论采用那个评价指标,AV 模型对未来波动率的预测能力都比相应的 GARCH 模型要好。具体而言,AV 模型的预测误差(RMSE 和 MAE)均小于 GARCH 模型的预测误差,AV - α 模型的预测误差小于对应的非对称 GARCH 即 EGARCH 模型的误差。仔细对比还可以发现,当以 σ_{RNG} 作为已实现波动率的代理测度时,AV 模型相对于对应的 GARCH 模型的优势更为突出,这体现在两个方面:一方面,以 σ_{RNG} 作为测度时,AV 模型的预测误差最小;更重要的是,AV 模型的误差低于 GARCH 模型预测误差的幅度要比以 σ_{SSDR} 作为已实现波动率测度时的幅度为大。这个现象并不令人奇怪:我们知道极差利用了更多的价格变动信息包括日内信息而 SSDR 只利用了每日的收盘价信息,所以极差波动率 σ_{RNG} 是对真实波动率更准确的测度,以此作为比较基准,更容易反映 AV 和 GARCH 两类模型预测效果的差距。

在 Theil-U 统计量下,两类模型的 Theil-U 值均小于 1,说明无论 AV 模型还是 GARCH 模型都比随机游走模型的预测效果要好。两类模型中,最好的模型依然是 AV - α 模型,其 U 值在不同的测度下分别是 0.54 和 0.68。回归分析 R^2 的分析结论和前述评价指标的评价结论是一致的。AV 模型的 R^2 均大于相对于的 GARCH 模型的值,再一次表现出相对于 GARCH 模型比较优势。

表 9 模型的样本外预测效果评价

	RMSE	MAE	Theil-U	R^2
Panel A· 已实现波动率测度为 σ_{RNG}				
GARCH(1,1)	7.27	5.83	0.93	0.38
EGARCH(1,1)	6.90	5.51	0.83	0.48
AV	6.77	5.00	0.80	0.45
AV - α	5.57	3.68	0.54	0.65
Panel B· 已实现波动率测度为 σ_{SSDR}				
GARCH(1,1)	8.85	6.93	0.86	0.35
EGARCH(1,1)	8.20	6.54	0.74	0.44
AV	8.68	6.46	0.83	0.42
AV - α	7.90	5.83	0.68	0.49

注:表中的 RMSE 和 MAE 都已经乘了 10^3 。

5 结语

基于高频数据的波动率是非常有前景的波动率代理变量,特别是基于日内极差的波动率不仅有效

地利用了日内信息,同时还具有对于金融市场微观噪声的鲁棒性。然而,与此有关的一个核心问题是缺少对于高频波动率有效建模与预测的工具。本文的研究工作正是基于这一现实问题的思考,引入并扩展了基于价格极差的自回归波动率(AV)模型,研究结果表明:

(1) AV 模型能够有效地解释时间序列中存在的异方差性等非线性相关性结构,具有良好的样本内拟合效果;

(2) AV 模型的样本外预测能力一样出众,在多种不同的预测评价指标下 AV 模型的预测绩效一致优于经典的波动率模型包括 GARCH 模型组、随机游走 RW 模型等,这表现为 AV 模型的预测具有更小的误差与更高的预测精度;

(3) AV 模型优良的性能不仅来源于极差波动率作为波动率代理变量的高效性,更在于 AV 模型能够描述波动率的动态变化特征:波动率的持久性与均值回复行为;

(4) 我们的研究还证实:在波动率模型中加入收益率的滞后项能够提高模型的解释能力,而且正负收益率对于波动率的冲击是非对称的。

AV 模型为预测波动率的动态变化提供了一个简明且有效的框架。我们的实证研究是基于日度、周度的,后续的研究还可以验证 AV 模型在月度、季度波动率建模与预测中的有效性,以及与其它 SV 模型、CARR 模型等作对比研究。总体而言,我们的研究工作证实了 AV 模型应用于金融市场的有效性,AV 模型组能够有效地捕捉波动率的时变性、相关性结构与均值回复行为,因而具有良好的波动率预测效果。这无论对于学术研究人员还是实际从业人员都具有重要的意义,特别是对关注全球风险管理的人员而言我们的研究工作更具有直接的参考价值。

参考文献:

- [1] Engle, R. F. . Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation [J]. *Econometrica*, 1982, 50: 987—1008.
- [2] Bollerslev, T. . Generalized autoregressive conditional heteroskedasticity [J]. *Journal of Econometrics*, 1986, 31: 307—327.
- [3] Nelson, D. B. . Conditional heteroscedasticity in asset returns: a new approach [J]. *Econometrica*, 1991, 59: 347—370.
- [4] Taylor, S. J. . *Modelling Financial Time Series* [M]. Wiley, Chichester, UK, 1986.
- [5] Andersen, T. G. , Bollerslev, T. , Diebold, F. X. , Ebens, H. . The distribution of realized stock return volatility [J]. *Journal of Financial Economics*, 2001, 61: 43—76.
- [6] Andersen, T. G. , Bollerslev, T. , Diebold, F. X. , Labys, P. . Modeling and forecasting realized volatility [J]. *Econometrica*, 2003, 71: 529—626.
- [7] Garman, M. B. , Klass, M. J. . On the estimation of price volatility from historical data [J]. *Journal of Business*, 1980, 53: 67—78.
- [8] Parkinson, M. . The extreme value method for estimating the variance of the rate of return [J]. *Journal of Business*, 1980, 53: 61—65.
- [9] Rogers, L. C. G. , Satchell, S. E. . Estimating variances from high, low, and closing prices [J]. *Annals of Applied Probability*, 1991, 1 (4): 504—512.
- [10] Yang, D. , Zhang, Q. . Drift-independent volatility estimation based on high, low, open, and close prices [J]. *Journal of Business*, 2000, 73: 477—491.
- [11] Christensen, K. , Podolskij, M. . Realized range-based estimation of integrated variance [J]. *Journal of Econometrics*, 2007, 141: 323—349.
- [12] Alizadeh, S. , Brandt, M. W. , Diebold, F. X. . Range-based estimation of stochastic volatility models [J]. *Journal of Finance*, 2002, 57: 1047—1091.
- [13] Zhang, L. , Mykland, P. A. , Ait-Sahalia, Y. . A tale of two time scales: Determining integrated volatility with noisy high-frequency data [J]. *Journal of the American Statistical Association*, 2005, 100 (472): 1394—1411.
- [14] Barndorff-Nielsen, O. E. , Hansen, P. R. , Lunde, A. , Shephard, N. . Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise [R]. Working paper, Nuffield College, University of Oxford, 2006.
- [15] Chou, R. Y. . Forecasting financial volatilities with extreme values: the Conditional Autoregressive Range (CARR) model [J]. *Journal of Money, Credit and Banking*, 2005, 37(3): 561—582.
- [16] Brandt, M. W. , Jones, C. S. . Volatility forecasting with range-based EGARCH models [J]. *Journal of Business & Economic Statistics*, 2006, 24 (4): 470—486.
- [17] 黄后川,陈浪南. 中国股票市场波动率的高频估计与特性分析 [J]. *经济研究*, 2003, (2): 75—82.
- [18] 唐勇,张世英. 高频数据的加权已实现极差波动及其实证分析 [J]. *系统工程*, 2006, (8): 52—57.

- [19] 蒋祥林, 吴晓霖, 王春峰. 基于日内价格幅度与回报的随机波动率模型[J]. 系统工程, 2006, (6): 68—73.
- [20] 王春峰, 张蕊等. 高频数据下投资组合风险预测模型比较[J]. 系统工程, 2007, (3): 23—28.
- [21] Beekers, S. . Variances of security price returns based on high, low, and closing prices [J]. Journal of Business, 1983, 56: 97—112.
- [22] Shu, J. , Zhang, J. E. . Testing range estimators of historical volatility [J]. Journal of Futures Markets, 2006, 26: 297—313.
- [23] Hsieh, D. A. . Chaos and nonlinear dynamics; Application to financial markets[J]. Journal of Finance, 1991, 46: 1839—1877.
- [24] Hsieh, D. A. . Implication of nonlinear dynamics for financial risk management [J]. Journal of Financial and Quantitative Analysis, 1993, 28(1): 41—64.
- [25] Hsieh, D. A. . Nonlinear dynamics in financial markets; evidence and implications[J]. Financial Analysts Journal, 1995, 51(4): 55—62.
- [26] Armstrong, J. S. , Fildes, R. . On the selection of error measures for comparisons among forecasting methods[J]. Journal of Forecasting, 1995, 14: 67—71.

Modeling financial volatilities based on price range: theoretical research and empirical study

LI Hong-quan^{1,2}, WANG Shou-yang²

(1. School of Business, Hunan Normal University, Changsha 410081, China;

(2. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: The classical volatility models, such as GARCH, are return-based models, which are constructed with the data of closing prices. It might neglect the important intraday information of the price movement, and will lead to loss of information and efficiency. This study introduces and extends the range-based autoregressive volatility model to make up for these weaknesses and obtain satisfactory volatility predicting performance. The results consistently show that the new model successfully captures the dynamics of the volatility and gains good performance relative to GARCH model. Furthermore, we find that the inclusion of the lagged return can significantly improve the forecasting ability of the volatility model, and the leverage effect does exist in volatility.

Key words: volatility model; price range; intraday information; forecasting performance