



基于HMM和GARCH模型的中国期货市场波动性研究

景楠¹, 吕闪闪¹, 江涛²

1 上海大学 悉尼工商学院, 上海 201899

2 哈尔滨工业大学(深圳) 经济管理学院, 广东 深圳 518055

摘要:期货市场波动性反映了市场的活跃度和流动性,是政府管控市场的重要决策来源,是投资者测量风险、实现资产保值的有利工具。已有研究表明,因加入了对过去时期的预测方差,GARCH模型比ARCH模型更能反映市场数据信息。然而在实际应用中,GARCH模型经常因数据的离散性而无法适应金融市场的结构突变,进而导致波动性预测效果不够理想。

为解决上述问题,结合HMM和GARCH模型预测中国期货市场收益率的波动性。通过GARCH模型计算期货的波动率序列;利用K均值法对波动率序列聚类得出观察序列;根据HMM划分波动率的状态,将不同状态对应的收益率代入HMM-GARCH模型,以得到不同状态下的波动率;通过VIX公式计算波动指数,以测量市场的波动性。基于以上逻辑,选用沪深300股指期货作为标的,以2015年5月至2016年4月为样本期,验证模型的有效性。

研究表明,一方面,HMM-GARCH模型的MAD和MSE两种损失函数值均比GARCH模型的低,表明拟合损失和错误少,可见与GARCH模型相比,HMM-GARCH模型能更好地拟合样本数据并预测市场信号;另一方面,基于HMM-GARCH模型的波动率指数显示,在样本期内沪深300股指期货由前期的小幅频繁波动转为大幅跳跃性波动,此后波动幅度保持较高水平并呈现增长态势,最终继续转为大幅跳跃性波动,与样本期内沪深300股指期货价格的实际波动态势一致。

因此,所述HMM-GARCH模型能够较好地测量中国期货市场波动状况,反映期货投资者对未来中国期货市场的预期。同时,能够为政府设置金融衍生品定价提供决策依据,为投资者测量市场风险、择取投机策略、合理配置资产提供客观的量化指标,有助于培养投资者的投资理性,促进中国期货市场的繁荣稳定发展。

关键词:期货市场波动率;隐马尔科夫模型;广义自回归条件异方差方法;HMM-GARCH模型;波动率指数

中图分类号: TP391

文献标识码: A

doi: 10.3969/j.issn.1672-0334.2019.05.013

文章编号: 1672-0334(2019)05-0152-11

引言

金融是国家经济发展的血液和重要支撑,期货

市场作为金融市场的重要组成部分,成为支持国家经济社会持续健康发展的重要平台。期货市场存在

收稿日期: 2018-01-20 **修返日期:** 2018-07-04

作者简介: 景楠,计算机科学博士,上海大学悉尼工商学院副教授,研究方向为决策分析、金融工程和信息推荐等,代表性学术成果为“Modeling co-construction processes in a socio-technical framework to support collaborative engineering design”,发表在2011年第3期《IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews》,E-mail:jingnan@shu.edu.cn

吕闪闪,上海大学悉尼工商学院硕士研究生,研究方向为数据建模、金融工程和信息推荐等,E-mail:silc_Cassie@163.com

江涛,哈尔滨工业大学(深圳)经济管理学院博士研究生,研究方向为信息推荐、金融计量、经济建模与预测等,代表性学术成果为“Personalized recommendation based on online reviews mining and sentiment assessment on electronic commerce website”,发表在2017年第5期《Electronic Commerce Research》,E-mail:er7248@126.com

明显的波动性,适度的价格波动将增加期货市场活跃度,提高整体市场流动性,加快资源配置效率;而过度的波动将严重阻碍市场的价格发现,降低资源配置效率,并影响宏观经济的正常运行^[1-3]。此外,金融全球化的进一步开放使中国金融市场受到更多冲击,波动率是金融资产投资和风险管控的重要因素^[4]。对投资者而言,增强波动率预测效果将为投资者投机、套利、避险提供决策依据^[5];对市场而言,增强波动率预测效果有助于减轻市场冲击影响,促进投资收益稳步增长^[6]。因此,在一个不完全成熟的期货市场中,对于市场波动性的分析、估计和预测尤为重要。

根据市场结构的突变性、波动率的持续性等特点,结合中国期货市场发展现状,本研究以沪深300股指期货作为标的,探究预测中国期货市场波动性的方法。首先,通过广义自回归条件异方差(generalized auto regressive conditional heteroskedasticity, GARCH)模型计算期货的波动率序列;其次,利用K均值法对波动率序列聚类得到观察序列;再次,根据隐马尔科夫模型(hidden Markov model, HMM)划分波动率的状态,将不同状态对应的收益率代入HMM-GARCH模型,得到不同状态下的波动率;最后,通过波动率指数公式计算波动指数,用该指数测量市场的波动性。本研究提出的将HMM与GARCH模型相结合的波动性指标计算方法,可用于测量中国期货市场波动状况,揭示投资者对未来风险的预期,为相关学者研究中国期货市场波动性提供新思路;同时,本研究的波动率指标为政府和投资者提供市场波动性的客观量化指标,对于投资者的理性投资和中国期货市场的繁荣稳定发展有重要意义。

1 相关研究评述

20世纪90年代以来,随着美国金融市场的波动加大和风险加剧,投资者迫切需要对市场波动性进行动态测量,波动率指数应运而生。1993年,美国芝加哥期权交易所(Chicago board options exchange, CBOE)编制和发布了基于S&P100指数的波动率指数(volatility index, 简称为VIX指数)^[7]。VIX指数一经推出便很快得到市场的广泛认同,被理论界和实务界作为测量美国股票市场即期波动率和市场风险的重要指标。由此可见,将市场波动率量化为波动指数,可以更直观地反映波动率的变化。

在过去的几十年里期货市场波动率估计模型成为实证金融学和时序计量经济学中最为活跃的研究领域之一。目前,按照计算方法和应用的不同可将波动率的预测方法分为隐含波动率法和历史波动率法两类。第1类,隐含波动率法^[8-10]通过将期权等衍生品的市场价格代入其理论价格模型,反推市场对于未来波动率的预期。期权的市场价格中包含了大量市场前瞻性信息,隐含了市场对标的资产未来波动率的预期^[11-12]。第2类,历史波动率法^[13]假设未来是过去的延伸,基于对资产价格序列过去某一段

时期内波动率的统计分析,试图发现波动率变化的规律,从而预测未来的波动率。

在实践应用中,市场波动率预测方法的选择取决于市场环境的成熟度和预测时间的长短。CORRADO et al.^[14]的研究表明,期权隐含波动率包含的信息量随着期权市场的发展得到显著提高。在成熟市场中,隐含波动率包含历史波动率所能反映的市场交易信息,而在新兴市场中则相反。在预测时间方面,杨小玄等^[15]认为预测时间的长短直接影响隐含波动率表现的好坏。预测时间较短时,隐含波动率反映信息的能力不及历史波动率,导致隐含波动率的预测能力也不及历史波动率。因此,在不同的期货市场中,需要根据市场的成熟度和预测期限确定具体使用哪种方法。国外波动率指数的编制均基于成熟的期权市场交易,因而可根据波动率预测时长选择采用隐含波动率法或历史波动率法预测市场未来波动。而在中国,2015年2月9日上证50ETF期权合约品种刚刚上市交易,还没有正式推出基于沪深300的股指期货,具有期权性质的权证也于2011年8月告别市场,因此与隐含波动率法相比,历史波动率法更适用于计算中国期货市场的波动性。

历史波动率法包括ENGLE^[16]在1982年首次提出的自回归条件异方差(auto regressive conditional heteroskedasticity, ARCH)模型和BOLLERSLEV^[17]提出的广义自回归条件异方差模型。GARCH模型是ARCH模型的推广,其在ARCH模型的基础上加入了对过去时期的预测方差,即误差项条件方差的滞后期,更能反映实际数据中的长期记忆性质,是专门针对金融数据的回归模型。在基于历史波动率法的波动率预测研究中,王佳妮等^[18]以1999年至2004年欧元、日元、英镑、澳元等4种货币兑美元的收市价为数据,用GARCH模型对外汇市场的波动性进行预测评估,发现GARCH模型比ARCH模型预测效果好;黄海南等^[19]以上证指数收益率为数据样本,通过GARCH族模型对其进行全面估计和样本外预测,用已实现波动率作为波动率预测的评价标准,通过M-Z回归和损失函数评价GARCH族模型的波动率预测表现,研究结果表明,GARCH族模型在样本内外均能较好地预测上证指数的收益波动率;魏宇^[20]认为,传统GARCH模型在预测未来的波动率时,基于从样本中总结得到的规律,在实际应用中效果并不好,或出现过度拟合问题,对沪深300股指期货的波动率预测能力较差。但由于其数据来源于中国金融交易所的股指期货仿真交易,因此该结论的普适性有待检验。

在实际应用中,传统GARCH模型大都是基于离散且低频的数据,而在金融市场中信息是连续影响金融市场的价格运动过程的,即收益波动具有持续性,离散时间模型必然导致金融市场信息的缺失,无法适用于金融市场的结构突变。因此,基于低频数据的GARCH模型往往因为损失了大量的日内交易信息^[21-22],造成对收益率波动性的估计和预测存在较大的偏差,无法达到描述金融时间序列频繁变化的目

的。为解决该问题, HAMILTON et al.^[23]最早提出将马尔可夫状态转换与GARCH模型结合起来用于金融时间序列的拟合, 实验结果表明, 马尔科夫模型在一定程度上弥补了GARCH模型的弊端, 两者的结合能较好地拟合实验数据, 但可能存在波动预测值偏高、持续性过强的问题; HENRYÓ^[24]将带有机制转换的马尔科夫链融入GARCH模型, 构建RS-EGARCH (regime switching-GARCH)模型, 该模型解决了早期模型中波动预测值偏高和持续性过强的问题, 解释了压力释放效应和波动的非对称性; 张锐等^[25]在HENRYÓ^[24]的基础上提出了一种新的RS-EGARCH模型, 为了刻画金融收益的厚尾特征可能随时间变化的特性, 该模型允许收益残差序列分布的自由度与所处状态相关, 实证结果表明, 在各种分布下, RS-EGARCH均能得到较为精确的预测值。

然而, 马尔科夫转换模型在对模型参数估计时, 需要人为地划分波动率状态, 以至对模型参数估计造成一定的干扰, 使模型对市场拟合效果不佳, 而隐马尔科夫模型利用可见变量进行客观分类的特性弥补了马尔可夫模型人为划分波动状态的不足。隐马尔科夫模型与马尔科夫模型不同, 它描述一个含有隐含未知参数的马尔科夫过程, 从可观察的参数中确定该过程的隐含未知参数, 用于探测一些引起某事件发生的隐藏事件的状态和转移概率等相关信息^[26]。曲大成等^[27]通过历史数据进行参数估计, 运用参数和当前观测值对未来波动率进行预测, 将预测结果与马尔科夫模型和GARCH模型的预测结果进行比较, 实证结果表明隐马尔科夫模型在沪深300指数波动性预测中具备更高的有效性和准确性。在马尔可夫模型中, 状态对于观察者来说是直接可见的, 研究者需要人为划分各种状态, 这样状态的转换概率便是全部的参数; 而在隐马尔可夫模型中, 状态是不可见的, 由算法根据可见的状态变量进行分类。若将HMM模型引入波动率的计算模型中, 可以在一定程度上规避了马尔科夫模型中需要主观分类的弊端。

综上所述可知, 在隐含波动率法和历史波动率法两种常见波动率估计模型中, 以ARCH模型和GARCH模型为代表的历史波动率法更适合中国期货市场的成熟度和预测期限。其中, GARCH模型对历史波动率的预测能力相对于ARCH模型更优秀, 因而被广泛运用于金融市场波动率的预测。但是, 传统GARCH模型大都基于离散且低频的数据, 而信息是连续影响金融市场的价格运动过程的, 即收益波动具有持续性, 离散时间模型必然导致金融市场信息的缺失, 不能有效反映市场因结构突变呈现多波动状态的信息。马尔科夫状态转换模型能够描述不同阶段、状态和机制下经济行为具有的不同特性, 因此, 部分学者考虑将GARCH模型与马尔科夫模型结合, 以在描述离散、低频的金融序列结构变化的基础上更准确地预测波动率。然而, 马尔科夫转换模型在对模型参数估计时需要人为地划分波动率状态,

对模型参数估计造成一定的干扰。因此, 本研究尝试以波动率研究中的传统GARCH模型为基础, 结合隐马尔科夫模型建立组合模型, 即隐马尔科夫-广义自回归条件异方差模型 (hidden Markov-generalized auto regressive conditional heteroskedasticity, HMM-GARCH), 对中国期货市场的波动率进行研究, 并通过波动指数的估计来反映波动率变化。本研究的波动率指标可以为政府及投资者提供市场波动性的客观量化指标, 对于投资者的理性投资、中国期货市场的繁荣稳定发展有重要意义。

2 基础模型

2.1 GARCH 模型

GARCH模型被广泛用于金融收益和风险的预测。ENGLE^[16]提出ARCH模型分析时间序列的异方差性以后, BOLLERSLEV^[17]又提出GARCH模型。GARCH模型在ARCH模型的基础上对误差的方差进一步建模, 使其更适用于波动性的分析和预测, 其分析对投资者决策的指导作用甚至超过了对数值本身的分析和预测^[28]。

ARCH的具体模型^[29]为, 对于时间序列 $\{a_t\}$, 若有

$$\begin{cases} a_t = \sigma_t u_t & u_t \sim i. i. d. N(0, 1) \\ \sigma_t^2 = \alpha_0 + \sum_{c=1}^Q \alpha_c a_{t-c}^2 \end{cases} \quad (1)$$

则时间序列 $\{a_t\}$ 是一个ARCH(Q)序列, (1)式为ARCH(Q)模型。其中, a_t 为 t 时刻的收益或者收益残差; σ_t 为波动率的标准差; u_t 为服从独立标准正态分布的随机变量序列, $u_t \sim i. i. d. N(0, 1)$; σ_t^2 为波动率; Q 为ARCH(Q)模型的阶数; c 为任意阶数, $c = 1, 2, \dots$; $Q; \alpha_0$ 和 α_c 为参数且大于等于0。

在ARCH(Q)模型的基础上, BOLLERSLEV^[17]提出了GARCH(P, Q)模型, 即

$$\begin{cases} a_t = \sigma_t u_t & u_t \sim i. i. d. N(0, 1) \\ \sigma_t^2 = \alpha_0 + \sum_{c=1}^Q \alpha_c a_{t-c}^2 + \sum_{d=1}^P \beta_d \sigma_{t-d}^2 \end{cases} \quad (2)$$

其中, P 为阶数; d 为任意阶数, $d = 1, 2, \dots, P; \beta_d$ 为参数且大于等于0。

当 $P = Q = 1$ 时, (2)式即可推导为GARCH(1, 1)模型, 即

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (3)$$

近年来, 关于GARCH(1, 1)模型的研究逐步深入。郑振龙等^[30]对香港恒生指数期权市场所含信息的研究发现, 在预测期限较短(1周)时, GARCH(1, 1)模型所含信息较多, 预测能力最强; 邹建军等^[31]选用移动平均法、GARCH(1, 1)模型和RiskMetrics分别预测上海证券交易所日收益率的波动性, 并计算每日VaR, 返回式检验表明, 在所选的预测方法中, GARCH(1, 1)模型最能准确预测中国上海证券交易所的风险, 更适用于中国股权价值波动率的计算。另有洪晶晶等^[32]基于上证指数的实验表明, 以标准差定义的波动率建立的GARCH(1, 1)模型的波动率

预测效果较好。本研究在研究沪深300股指期货的短期中、低频数据时,通过对比多个GARCH模型的赤池信息准则(Akaike information criterion, AIC)结果,发现GARCH(1,1)模型的AIC值最小,模型最优。因此,本研究采用GARCH(1,1)模型计算波动率。

2.2 HMM 模型

马尔科夫模型(Markov model, MM)是由Markov于1913年提出的统计模型,可利用概率建立随机时间序列,并用于市场趋势分析^[33]。作为对马尔科夫模型的一种扩充,隐马尔科夫模型诞生于20世纪60年代末,其在标准马尔科夫模型的基础上,添加了可见状态和可见状态与隐状态之间的概率分布。由于隐马尔科夫模型的隐状态不可观测,而是需要通过计算观测向量序列才能得到,每个观测向量都是通过一些概率密度分布产生各种状态表现形式,而一个概率密度分布的状态序列可相应产生一个观测向量。因此,隐马尔科夫模型是一个双重随机过程^[34],经常用于探测引起某事件发生的隐藏事件的状态和转移概率等相关信息。隐马尔科夫模型通过前向-后向算法不断对样本进行调试和迭代,使其能更准确地刻画波动率的变化状态^[35]。

在处理状态方面,马尔科夫模型需要主观设定状态,而隐马尔科夫模型可以根据可见状态数据运用模型进行隐状态的客观分类。郭存芝^[36]的研究表明,隐马尔科夫模型的状态变量可以用来描述金融市场的正常波动状态和异常波动状态,同时不可观测的状态变量能够对波动的集聚现象给出很好的解释。隐马尔科夫模型的状态链包括两条,一条为可见状态链,一条为隐含状态链,根据转换概率和可见状态,可以推断出隐含状态的分类。基于此特性,结合现有模型能够观测到波动率的表现却不能对波动率的结构突变进行准确判断的现状,本研究将隐马尔科夫模型引入波动率模型,构建HMM-GARCH的组合模型,对中国期货市场的结构突变特征进行捕捉和刻画。

具体而言,一个隐马尔科夫模型可以用一个五元组 (S, V, A, B, δ) 表示^[37]。

(1) 隐藏的状态集合

S 为隐状态集合, $S = \{s_1, s_2, \dots, s_N\}$, 并记 t 时刻的状态为 $q_t, q_t \in \{s_1, s_2, \dots, s_N\}$, s_i 为隐状态, $1 \leq i \leq N, N$ 为隐状态数目。

(2) 观察符号集合

V 为所有观察符号的集合, $V = \{v_1, v_2, \dots, v_M\}$, v_k 为观察符号, $1 \leq k \leq M, M$ 为每一个状态的不同符号数。

(3) 状态转移概率分布

A 为状态转移矩阵, $A = \{a_{i,j}\}$, $a_{i,j}$ 为由一个表现状态转变为另一个表现状态的概率, $a_{i,j} = p(q_{t+1} = s_j | q_t = s_i), 1 \leq i \leq N, 1 \leq j \leq N, p(\cdot)$ 为条件概率。

(4) i 状态中可观察符号的概率分布

B 为观测概率矩阵, $B = \{b_i(k)\}$, $b_i(k)$ 为在隐状态为 s_i 时对应观察符号为 v_k 的概率, $b_i(k) = p(o_t = v_k | q_t = s_i), 1 \leq k \leq M, 1 \leq i \leq N, o_t$ 为在 t 时刻的观察值。

(5) 初始状态概率分布

δ 为初始状态概率向量, $\delta = \{\delta_i\}$, δ_i 为初始状态为 s_i 的概率, $\delta_i = p(q_0 = s_i), 1 \leq i \leq N$ 。

一个隐马尔可夫模型完全由 A, B, δ 确定,为了方便,简记为 $\lambda = (A, B, \delta)$ 。

2.3 Baum-Welch 算法

本研究采用Baum-Welch算法解决HMM模型中的参数估计问题,具体思想是:给定一个观察序列 O ,调整模型参数 λ ,使产生观察序列的概率 $Prob(O | \lambda)$ 最大^[38-39]。根据前向变量和后向变量的定义,推出前向变量和后向变量的混合概率公式为

$$Prob(O | \lambda) = \sum_{i=1}^N \sum_{j=1}^N \alpha_i(i) a_{i,j} \beta_{i+1}(j) b_j(o_{t+1}) \quad (4)$$

其中, t 为当前状态的时刻,也作为样本的序号, $1 \leq t \leq T-1, T$ 为总时间长度,即样本总数; $\alpha_t(i)$ 为前向变量; $\beta_{t+1}(j)$ 为后向变量; $b_j(o_{t+1})$ 为 j 状态观察值为 o_{t+1} 的概率; o_{t+1} 为 $(t+1)$ 时刻的观察值。

实际上,在给定的观察序列有限时,不存在一个最佳的方法来估计,只能找到某些方法,让参数在特定的几个性能上有较强的优势。Baum-Welch算法利用递推的思想,使条件概率局部最大,并得到模型参数的估计值。具体训练步骤如下:

定义 $\varepsilon_t(i, j)$ 为 t 时刻状态为 $s_i, (t+1)$ 时刻状态为 s_j 的概率, $\varepsilon_t(i, j) = Prob(q_t = s_i, q_{t+1} = s_j | O, \lambda)$ 。根据前向-后向算法可推出

$$\varepsilon_t(i, j) = \frac{\alpha_t(i) a_{i,j} \beta_{t+1}(j) b_j(o_{t+1})}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{i,j} \beta_{t+1}(j) b_j(o_{t+1})} \quad (5)$$

定义 $\gamma_t(i)$ 为 t 时刻状态为 s_i 的概率, $\gamma_t(i) = \sum_{j=1}^N \varepsilon_t(i, j)$, $\sum_{i=1}^N \gamma_t(i)$ 为从 s_i 状态转移出去的期望次数, $\sum_{i=1}^N \sum_{j=1}^N \varepsilon_t(i, j)$ 为从 s_i 状态转移到 s_j 状态的期望次数,因此

可以得到 $a_{i,j}$ 和 $b_i(k)$ 的重估值 $\tilde{a}_{i,j}$ 和 $\tilde{b}_i(k)$ 。则得到Baum-Welch算法的重估公式为

$$\tilde{\mu}_{i,j} = \frac{\sum_{i=1}^N \sum_{j=1}^N \varepsilon_t(i, j) o_t}{\sum_{i=1}^N \sum_{j=1}^N \varepsilon_t(i, j)} \quad (6)$$

其中, $\tilde{\mu}_{i,j}$ 为重估后从 s_i 状态转移到 s_j 状态的期望概率。

$$\tilde{\sigma}_{i,j} = \frac{\sum_{i=1}^N \sum_{j=1}^N \delta_i(o_t - \tilde{\mu}_{i,j})(o_t - \tilde{\mu}_{i,j})'}{\sum_{i=1}^N \sum_{j=1}^N \delta_i(i, j)} \quad (7)$$

其中, $\tilde{\sigma}_{i,j}$ 为从 s_i 状态转移到 s_j 状态的概率离散程度, δ_i 为初始概率矩阵中从 s_i 状态转移到 s_j 状态的初始状态转移概率。

$$\tilde{\omega}_{i,j} = \frac{\sum_{i=1}^N \sum_{j=1}^N \delta_i(i, j)}{\sum_{i=1}^N \delta_i(i)} \quad (8)$$

其中, $\tilde{\omega}_{i,j}$ 为初始概率矩阵中从 s_i 状态转移到 s_j 状态的概率。

$$\tilde{\omega}_{i,j} = \frac{\sum_{i=1}^r \alpha_i(i) a_{i,j} \beta_{i+1}(j) b_j(o_{i+1})}{\sum_{i=1}^r \alpha_i(i) \beta_i(i)} \quad (9)$$

其中, $\tilde{\alpha}_{i,j}$ 为修正后的状态转移矩阵。

通过隐马尔科夫模型的训练,可以得到符合预期的波动率状态的划分结果。一方面,证明了本研究将HMM模型引入中国期货市场波动率研究有利于后续工作的展开;另一方面,HMM的训练为对结构突变前后不同状态的波动率建模提供了基础。

3 基于HMM-GARCH模型的金融市场波动性计算模型

本研究基于隐马尔科夫-广义自回归条件异方差模型研究中国期货市场波动性,具体研究思路见图1。以沪深300股指期货作标的,①通过GARCH模型计算期货的波动率序列;②利用K均值聚类法对波动率序列聚类得出观察序列;③根据HMM模型划分波动率的状态,将不同状态对应的收益率代入HMM-GARCH模型中,得到不同状态下的波动率;④通过VIX公式计算波动指数,用该指数测量市场的波动性。

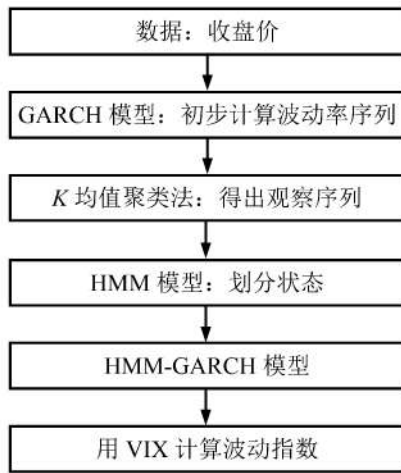


图1 研究思路

Figure 1 Research Method

3.1 HMM-GARCH 模型

假定 r_t 为收益率序列, $t = 1, 2, \dots, T-1$, 对该序列建立一个隐马尔科夫模型。

假设下一个交易日的波动情况只与今天的波动情况有关,则期货市场的收益率时间序列服从马尔科夫过程,将其构造成一个马尔科夫链,对其进行模拟。由于波动只能向某一个时刻的下一个时刻进行转移,因此仅使用一步转移概率作为统计特征,即 $Prob(r_{t+1} | r_t)$ 。

将实际波动情况分为 N 个状态,时间轴上每个时

刻所在的状态 q_t 随机出现,两个状态之间服从马尔科夫转移概率矩阵。转移概率矩阵为 $A, A = (a_{i,j}), a_{i,j} \geq 0$ 并且 $\sum_j a_{i,j} = 1$ 。

$$a_{i,j} = \begin{pmatrix} p(\frac{q_{t=1}}{q_{t=1}}) & p(\frac{q_{t=1}}{q_{t=2}}) \\ p(\frac{q_{t=2}}{q_{t=1}}) & p(\frac{q_{t=2}}{q_{t=2}}) \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \quad (10)$$

当收益率序列处于 j 状态时,对应的观察值概率矩阵为 $B, B = \{b_j(k)\}, b_j(k) = p(r_t = v_k | q_{t-1} = s_i), v_k \in (-\infty, +\infty)$ 。本研究模型中假设观察值概率分布为高斯分布。初始化分布为 $\delta, \delta = \{\delta_1, \delta_2\}$ 。

为叙述方便,隐马尔科夫模型参数记为 $\lambda, \lambda = (A, B, \delta)$ 。本研究将Baum-Welch算法用于可观测收益率序列,从而估计模型的未知参数 λ , 并通过Viterbi算法^[40]估算各收益率序列对应的隐状态序列。

建立HMM-GARCH(1,1)模型。

均值方程为

$$\begin{cases} y_t = C_{s_t} + a_t \\ a_t = h_t v_t \end{cases} \quad (11)$$

条件方差方程为

$$h_t^2 = \alpha_{s_t} + \alpha_{s_t} a_{t-1}^2 + \beta_{s_t} h_{t-1}^2 \quad (12)$$

其中, y_t 为收盘价的条件收益率; h_t^2 为沪深300指数对数收益率的条件波动率; h_t 为沪深300指数对数收益率的条件波动率的标准差; $C_{s_t}, \alpha_{s_t}, \beta_{s_t}$ 为依赖于 t 时刻的变量 s_t 的参数; v_t 为服务标准正态分布的随机变量, $v_t \sim N(0,1)$ 。

本研究运用损失函数对模型的模拟效果进行评估,进而比较HMM-GARCH与GARCH模型所拟合波动率的精确度。损失函数用来测量波动率模拟值与实际值之间的差距,损失函数值越小,拟合波动率越接近实际波动率。由于实际波动率未知,因此本研究使用近似收益率替代实际波动率进行比较^[41]。用到的两个损失函数为

$$MAD = \frac{1}{N} \sum_{i=1}^N |\sigma_i^2 - h_i| \quad (13)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (\sigma_i^2 - h_i)^2 \quad (14)$$

其中, MAD 为平均绝对误差, MSE 为平均平方误差。

3.2 基于HMM-GARCH模型的VIX计算

基于FERNANDES et al.^[42]提到的VIX指数计算方法,本研究将HMM-GARCH模型中得出的预测波动率放大100倍,并且对近月波动率和次近月波动率进行加权平均,计算距到期日时间为30天的加权平均波动率指数。

VIX指数计算公式为

$$VIX = 100 \sqrt{T_1 \sigma_1^2 \frac{U_2 - NT_{30}}{U_2 - U_1} + T_2 \sigma_2^2 \frac{NT_{30} - U_1}{U_2 - U_1} \cdot \frac{NT_{365}}{NT_{30}}} \quad (15)$$

其中, T_1 为数据样本中各月的近月合约交割时间在1年中所占比例, T_2 为数据样本中各月的次近月交割时间在1年中所占比例。根号中第1项为近月波动率,记为 VIX_1 ; 第2项为次近月波动率,记为 VIX_2 。

(1) NT 和 U 的单位都是分钟。 NT_{30} 为30天的总分钟数,具体为30天 \times 24小时 \times 60分钟; NT_{365} 为1年中的总分钟数,具体为365天 \times 24小时 \times 60分钟; U_1 为当前时间距离近月交割时间所剩的分钟数; U_2 为当前时间距离次近月交割时间所剩的分钟数。

(2) R 为指数期权合约剩余到期时间在1年中所占比例。计算公式为

$$R = \frac{M_{\text{当天}} + M_{\text{到期日}} + M_{\text{其他时间}}}{NT_{365}} \quad (16)$$

其中, $M_{\text{当天}}$ 为当前时间到24点剩余的分钟数; $M_{\text{到期日}}$ 为从到期日的0点到到期日上午8点30分的分钟数; $M_{\text{其他时间}}$ 为从今天24点到到期日当天0点的分钟数。另外,为了避免由于距离交割时间过近出现异常波动,影响所得指数的准确性,要求当前时间距离近月交割时间必须满足至少7天,若到期日不足7天,则近月和次近月均往后顺移,即以次近月的交割时间距离作为近月交割时间距离,以此类推。

通过计算每日波动率指数得到代表一段时间内期货市场波动性的指数序列,并形成波动走势图,以波动走势图为一段时间内标的价格走势、波动情况的参考。

4 实证分析

4.1 数据描述

本研究选择中国金融期货交易所推出的沪深300股指期货作为标的指数,选取在2015年5月至2016年4月交易日中的每小时收盘价共1225个数据作为数据样本,按30个交易日为时间窗口,分别计算沪深300股指期货在该时间段内每月的近月波动率指数。沪深300股指期货的编制目标是准确有效地反映中国金融市场价格波动的情况以及整个金融市场的运行情况,是能够较全面地反映中国A股整体走势的指数^[43]。相关研究表明,沪深300股指期货走势与沪深300指数相关程度高,对市场信息的反应更迅速,并且300股指期货领先沪深300指数^[44-46]。因此,以沪深300股指期货合约作为未来市场波动预测指标的计算依据具有一定的理论意义和现实意义。

4.2 统计特性分析

本研究探究期货每日收盘价对应的收益率的分布函数,为波动性计算奠定基础。考虑到正态分布的普遍性和可行性,假设收益率符合正态分布。

计算偏度、峰度和J-B统计量,这3个统计量在很大程度上反映收益率是否呈正态分布,结果见表1。

由表1可知,该组数据收益率的整体均值为 -4.12% ,说明整体收益是小于0的;标准差为0.013;偏度为0.095,说明市场收益率小于均值的时间占多数;峰度为5.445,高于正态分布的峰度值3;J-B统计量为51.390,概率为0。峰度和偏度的数据说明收益率序列具有尖峰和厚尾的特征,J-B统计量取值大于样本计算的J-B值的概率,说明该对数收益率序列服从正态分布的假设是不准确的。这一结论为推出收益率序列适用的分布模型奠定了基础。

表1 数据样本的描述性统计

Table 1 Descriptive Statistics for Data Sample

属性名称	属性值
均值	-4.120E-04
中间值	-1.520E-04
最大值	0.047
最小值	-0.440
标准差	0.013
偏度	0.095
峰度	5.445
J-B 值	51.390
概率	0

本研究绘制沪深300股指期货样本数据的对数收益率走势图,见图2。由图2可知,收益率序列在一段时间内波动较小,在某些时刻波动较大,说明收益率序列具有聚集性,不符合正态分布。

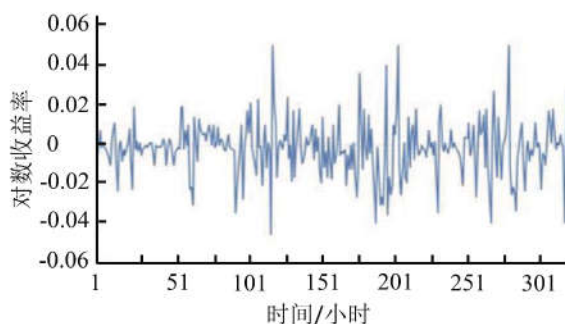


图2 对数收益率走势图

Figure 2 Trend Chart of Logarithmic Return

残差平方值可以对ARCH效应进行检验,如果残差序列不存在ARCH效应,自相关系数(AC)和偏自相关系数(PAC)在所有的滞后阶数均为0,且Q统计量也不显著;否则说明序列中存在ARCH效应。表2给出2015年5月至2016年4月沪深300股指期货交易日每小时收盘价对数收益率的残差平方情况。

由表2可知,AC、PAC值不为0且Q统计量显著,因此收益率序列具有ARCH效应,可以应用GARCH模型。

4.3 GARCH

1987年,为解决时间序列定阶问题,赤池弘次提出赤池信息准则(AIC)^[47],模型的该值越小,表明该模型对数据的拟合情况越好。AIC的诞生有效解决了在多个可供选择的模型中选定最优模型的问题^[48]。本研究通过对比多个GARCH模型的AIC值,

发现GARCH(1,1)模型更适用于本研究数据样本。对比结果见表3。

表2 对数收益率的残差平方值
Table 2 Residual Square of Logarithmic Return

月份	自相关系数	偏自相关系数	Q 统计量	概率
1	0.286	0.286	17.054	0
2	0.061	-0.023	17.823	0
3	0.106	0.104	20.195	0
4	0.113	0.061	22.898	0
5	0.130	0.089	26.499	0
6	0.063	-0.005	27.356	0
7	0.197	0.187	35.651	0
8	0.031	-0.109	35.859	0
9	0.081	0.107	37.297	0
10	0.166	0.084	43.312	0
11	0.126	0.049	46.772	0
12	-0.018	-0.123	46.874	0

表3 不同模型下的AIC 值
Table 3 AIC Value of Different Models

	GARCH(1,1)	GARCH(2,1)	GARCH(2,2)	GARCH(3,3)
常数	9.390E-06 (2.352)	9.310E-06 (2.001)	1.930E-05 (2.775)	1.080E-05 (2.056)
α_1	0.132 (3.490)	0.245 (2.438)	0.144 (3.716)	0.227 (2.811)
α_2		-0.126 (-1.246)	0.141 (3.438)	-0.118 (-0.818)
α_3				0.116
β_1	0.718 (17.950)	0.730 (13.405)	-0.286 (-4.723)	1.290 (6.786)
β_2			0.695 (19.127)	-1.370 (-6.989)
β_3				0.497 (7.061)
L	3 168.790	3 169.800	3 189.460	3 178.490
AIC	-9.321	-9.314	-9.307	-9.292

注：L为对数似然函数值，括号内数据为t检验值，AIC值取自然对数。

由表3可知，AIC 值随着滞后阶数的增大而增大，GARCH(1,1)模型的 AIC 值最小，模型最优。因此，本研究运用GARCH(1,1)模型分别计算2015年5月至2016年4月沪深300股指期货的近月波动率，以2015年5月的近月波动率为例给出计算过程。

取IF1505（交割时间为2015年5月的期货合约）从2015年4月30日15:15的收盘价到5月15日15:15的收盘价和IF1506（交割时间为2015年6月的期货合约）从2015年5月18日10:15的收盘价到5月29日15:15的收盘价作为计算5月近月波动率的样本。其中，取2015年4月29日最后1小时的收盘价计算5月第1个交易日第1个小时的波动率。由于近月合约的交割日为每个月第3个周五，如果离交割日不到7天，近月合约和次近月合约顺移，因此5月的近月合约交易价格从5月4日至5月15日取IF1505的交易价格、5月18日至29日取IF1506的交易价格。随后，对该样本中的收盘价建立对数收益率序列，并对该收益率序列去均值化，用去均值化后的收益率序列建立GARCH(1,1)模型，计算得出5月近月的波动率序列。

将收益率导入EViews 8.0中形成时间序列，在其Method工具包中选择GARCH模型，即可得到GARCH模型的描述性统计量，见表4。

表4 GARCH 模型描述性统计
Table 4 Descriptive Statistics of GARCH Model

变量	系数
α_0	6.680E-06
α_1	0.114
β_1	0.864

由表4可知(3)式差分方程为

$$\sigma_t^2 = 6.680E-06 + 0.114\alpha_{t-1}^2 + 0.864\sigma_{t-1}^2 \quad (17)$$

将序列中的 α_{t-1}^2 和 σ_{t-1}^2 重复迭代，可得 σ_t^2 预测波动率序列，生成波动率指数。

4.4 损失函数的比较

为比较模型与实际样本的拟合程度，结合两种模型估计得出的波动率，通过MAD和MSE两种损失函数观察比较结果，损失函数比较结果见表5。

表5 损失函数比较
Table 5 Comparison of Loss Function

模型	MAD	MSE
GARCH	1.462E-02	3.494E-04
HMM-GARCH	1.176E-02	1.627E-04

由表5可知，HMM-GARCH模型的两种损失函数值均比GARCH模型的低。损失函数的数值越小，代

表损失和错误越少。通过比较HMM-GARCH模型和GARCH模型的MAD和MSE可知,HMM-GARCH模型对中国期货市场的波动率预测更为准确,更能体现实际波动率的变化趋势。

4.5 波动率的实证计算结果

本研究取每小时沪深300股指期货的收盘价作为数据,因此将VIX公式的单位变更为小时,基础公式不变。分别计算数据样本中各月的近月和次近月合约交割时间在一年中所占比例 T_1 和 T_2 ,如当前时间为2015年5月4日上午10点15分,则

$$M_{\text{当天}} = \text{当前时间到24点剩余的小时} \\ = 24:00 - 10:15 = 13.750 \text{ 小时}$$

$$M_{\text{到期日}} = \text{到期日的0点到到期日上午8:30的小时数} \\ = 8:30 - 0:00 = 8.500 \text{ 小时}$$

$$M_{\text{其他时间(近)}} = \text{今天24点到近月交割时间当天0点的小时数} \\ = 2015年5月15日00:00 - 2015年5月4日24:00 = 240 \text{ 小时}$$

$$M_{\text{其他时间(次近)}} = \text{今天24点到次近月交割时间当天0点的小时数} \\ = 2015年6月19日00:00 - 2015年5月4日24:00 = 1080 \text{ 小时}$$

则

$$T_1 = \frac{13.750 + 8.500 + 240}{8760} = 0.030$$

$$T_2 = \frac{13.750 + 8.500 + 1080}{8760} = 0.126$$

将得到的两个波动率加权平均处理,并扩大100倍作为波动率指数。例如,在2015年5月4日上午10点15分的波动率指数满足

$$VIX_1 = 0.030 \times 6.240E-05 \times \frac{1102.25 - 720}{1102.25 - 262.250}$$

$$VIX_2 = 0.126 \times 6.010E-05 \times \frac{720 - 262.250}{1102.25 - 262.250} \times \frac{8760}{720}$$

$$VIX = 100 \sqrt{VIX_1 + VIX_2} = 0.715$$

根据上述流程,分别计算本研究沪深300股指期货数据样本各时段的波动率指数。

根据GARCH模型的特性,本研究预测到的波动率展示的是未来1个小时的波动幅度。图3给出本研究对沪深300股指期货波动率指数的预测情况,图4给出沪深300股指期货实际的价格走势。观察图3可知,本研究预测沪深300股指期货将先呈现一段时间的小幅频繁波动,此后波动幅度加大,而后一段时间内波动幅度保持较高水平并呈增长态势,最终继续转为大幅跳跃性波动。对比两图发现,图3中的波动率指数对图4中的期货价格波动情况有较好的预测效果。图4中的沪深300股指期货价格在一段时间的小幅波动后快速增长,价格变化区间增大,而后价格大幅跳水,这与图3的预测波动情况相吻合。因此,本研究建立的HMM-GARCH组合模型是有效可行的,

能够计算出与中国期货市场走势相关的波动率。通过对中国期货市场波动性的研究,能够反映期货投资者对未来期货市场波动性的预期,并能起到警示大盘涨跌幅度的作用,对于现货的走势预测也有一定的参考意义。

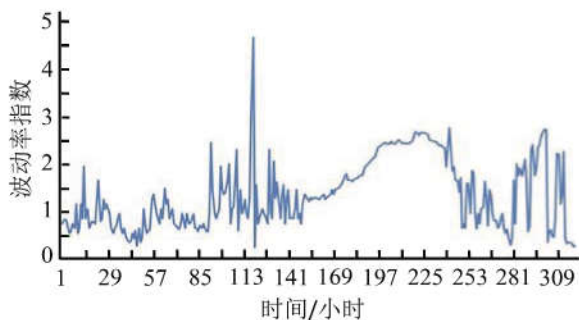


图3 沪深300股指期货的波动率指数

Figure 3 Volatility Index of CSI300 Index Futures

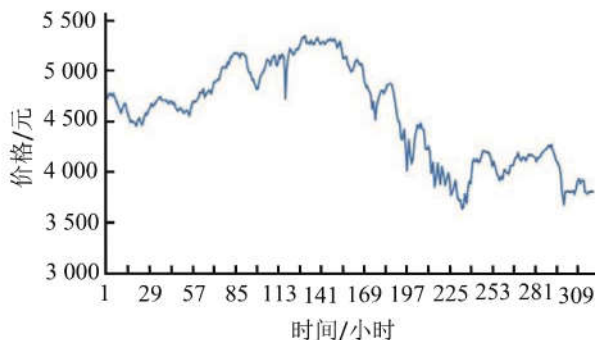


图4 沪深300股指期货价格走势

Figure 4 Price Trend Chart of CSI300 Index Futures

5 结论

本研究以沪深300股指期货作标的,将GARCH模型与隐马尔科夫模型相结合计算波动率,可规避GARCH模型本身的缺陷,并利用以隐马尔科夫模型拟合出的隐状态更准确地计算期货市场波动率。

研究结果表明,本研究所选数据符合GARCH模型的使用条件;通过MAD和MSE两种损失函数观察比较GARCH模型和HMM-GARCH模型估计得出的波动率,从计算结果看HMM-GARCH模型的两种损失函数值均比GARCH模型的低,表明HMM-GARCH模型对中国期货市场的波动率预测更为准确,更能体现实际波动率的变化趋势。此外,基于HMM-GARCH模型的波动率指数计算结果表明,沪深300股指期货由前期的小幅频繁波动转为大幅跳跃性波动,此后波动幅度保持较高水平并呈增长态势,最终继续转为大幅跳跃性波动,该结果能较好地反映期货价格的波动情况。

通过对中国期货市场的研究,本研究创新性地提出以隐马尔科夫模型和GARCH模型相结合的方法研究中国期货市场波动性,为相关学者研究中国期货市场波动性提供了新思路。本研究的波动率指标

可以为政府、投资者提供市场波动性的客观量化指标,为政府设置金融衍生品定价提供决策依据,为投资者测量市场风险、择取投机策略、合理配置资产提供参考,有助于培养投资者的投资理性,促进中国期货市场的繁荣稳定发展。

未来研究可以扩大波动率的计算范围,以获得更加客观的研究成果;还可以检验模型在不同类型金融市场中的有效性,根据检验结果对模型进行适当的拓展提升,以增加模型的普适性。

参考文献:

- [1] 刘志东,姜玲. 基于贝叶斯参数估计的期货市场交易成本、流动性与资产定价研究. *管理科学*, 2017, 30(1):146-159.
LIU Zhidong, JIANG Ling. Trading costs, liquidity and asset pricing for futures markets with Bayesian parameter estimation. *Journal of Management Science*, 2017, 30(1):146-159.
- [2] BARUNIK J, KOCEK E, VACHA L. Asymmetric connectedness on the U.S. stock market: bad and good volatility spillovers. *Journal of Financial Markets*, 2016, 27:55-78.
- [3] 宋逢明,江捷. 中国股票市场波动性特性的实证研究. *金融研究*, 2003(4):13-22.
SONG Fengming, JIANG Jie. Empirical research on the volatility characteristics of Chinese stock market. *Journal of Financial Research*, 2003(4):13-22.
- [4] 瞿慧,刘焯. 沪深300指数收益率及已实现波动联合建模研究. *管理科学*, 2012, 25(6):101-110.
QU Hui, LIU Ye. A joint model for CSI300 index return and realized volatility. *Journal of Management Science*, 2012, 25(6):101-110.
- [5] 郭彦峰,黄登仕,魏宇. 上海期货市场收益和波动的周日历效应研究. *管理科学*, 2008, 21(2):58-68.
GUO Yanfeng, HUANG Dengshi, WEI Yu. A study of the day of the week effect in return and volatility at the Shanghai futures market. *Journal of Management Science*, 2008, 21(2):58-68.
- [6] 周士元. 基于结构转换非参数GARCH模型的股市波动率预测拟合性评估. *统计与决策*, 2015, 31(17):162-165.
ZHOU Shiyuan. Stock market volatility forecast fitting evaluation based on RS-APGARCH. *Statistics & Decision*, 2015, 31(17):162-165.
- [7] FERMANDES M, MEDEIROS M C, SCHARTH M. Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 2014, 40:1-10.
- [8] AN B J, ANG A, BALI T G, et al. The joint cross section of stocks and options. *Journal of Finance*, 2014, 69(5):2279-2337.
- [9] LEUNG T, SIRCAR R. Implied volatility of leveraged ETF options. *Applied Mathematical Finance*, 2015, 22(2):162-188.
- [10] LEUNG T, LORIG M, PASCUCCI A. Leverage ETF implied volatilities from ETF dynamics. *Mathematical Finance*, 2016, 27(4):1035-1068.
- [11] BLACK F, SCHOLES M. The pricing of options and corporate liabilities. *Journal of Political Economy*, 1973, 81(3):637-654.
- [12] 黄慧舟. 隐含波动率的信息含量及其在我国的应用. *商业经济与管理*, 2011(7):70-76.
HUANG Yizhou. The information content with implied volatility and its application in China. *Journal of Business Economics*, 2011(7):70-76.
- [13] SHABARISHA N, MADEGOWDA P J. Forecasting volatility in options trading-nexus between historical volatility and implied volatility. *International Journal of Research in Finance and Marketing*, 2016, 6(9):1-14.
- [14] CORRADO C J, MILLER T W, Jr. The forecast quality of CBOE implied volatility indexes. *Journal of Futures Markets*, 2005, 25(4):339-373.
- [15] 杨小玄,刘立新. 人民币汇率波动率预测模型的研究. *财贸研究*, 2016, 27(3):80-90.
YANG Xiaoxuan, LIU Lixin. Forecast quality comparison: a research of RMB exchange rate volatility. *Finance and Trade Research*, 2016, 27(3):80-90.
- [16] ENGLE R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 1982, 50(4):987-1007.
- [17] BOLLERSLEV T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 1986, 31(3):307-327.
- [18] 王佳妮,李文浩. GARCH模型能否提供好的波动率预测. *数量经济技术经济研究*, 2005, 22(6):74-87.
WANG Jiani, LI Wenhao. Do GARCH models give poor out-of-sample forecasting volatility. *The Journal of Quantitative & Technical Economics*, 2005, 22(6):74-87.
- [19] 黄海南,钟伟. GARCH类模型波动率预测评价. *中国管理科学*, 2007, 15(6):13-19.
HUANG Hainan, ZHONG Wei. Evaluation on volatility forecasting performance of GARCH-type models. *Chinese Journal of Management Science*, 2007, 15(6):13-19.
- [20] 魏宇. 沪深300股指期货的波动率预测模型研究. *管理科学学报*, 2010, 13(2):66-76.
WEI Yu. Volatility forecasting models for CSI300 index futures. *Journal of Management Sciences in China*, 2010, 13(2):66-76.
- [21] 陈杰. 基于已实现极差的高频数据VaR研究. 杭州:浙江财经大学, 2014:13-14.
CHEN Jie. *High frequency data VaR research based on realized range*. Hangzhou: Zhejiang University of Finance and Economics, 2014:13-14.
- [22] 伍习丽. 基于高频数据的中国股市VaR风险研究. 重庆:重庆大学, 2013:15-16.
WU Xili. *Research on VaR of Chinese stock market based on high frequency data*. Chongqing: Chongqing University, 2013:15-16.
- [23] HAMILTON J D, SUSMEL R. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 1994, 64(1/2):307-333.
- [24] HENRYÓ T. Regime switching in the relationship between equity returns and short-term interest rates in the UK. *Jour-*

- nal of Banking & Finance*, 2009, 33(2):405-414.
- [25] 张锐,魏宇,金炜东. 基于MRS-EGARCH模型的沪深300指数波动率预测研究. *系统工程学报*, 2011, 26(5):628-635.
ZHANG Rui, WEI Yu, JIN Weidong. HS300 index volatility forecast based on Markov regime-switching EGARCH model. *Journal of Systems Engineering*, 2011, 26(5):628-635.
- [26] TANG L, DIAO X D. Option pricing based on HMM and GARCH model // WEN C Y, JIA Z P. *Proceedings of the 29th Chinese Control and Decision Conference*. Chongqing: IEEE, 2017:3363-3368.
- [27] 曲大成,房振明. 基于隐马尔科夫模型的波动率预测探究. *电子设计工程*, 2014, 22(18):1-3.
QU Dacheng, FANG Zhenming. Predictive research of volatility on hidden Markov model. *Electronic Design Engineering*, 2014, 22(18):1-3.
- [28] 于志军,杨善林. 基于误差校正的GARCH股票价格预测模型. *中国管理科学*, 2013, 21(S1):341-345.
YU Zhijun, YANG Shanlin. A model for stock price forecasting based on error correction. *Chinese Journal of Management Science*, 2013, 21(S1):341-345.
- [29] BOLLERSLEV T, CHOU R Y, KRONER K F. ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, 1992, 52(1/2):5-59.
- [30] 郑振龙,黄意舟. 波动率预测:GARCH模型与隐含波动率. *数量经济技术经济研究*, 2010, 27(1):140-150.
ZHENG Zhenlong, HUANG Yizhou. Volatility forecast: GARCH model vs implied volatility. *The Journal of Quantitative & Technical Economics*, 2010, 27(1):140-150.
- [31] 邹建军,张宗益,秦拯. GARCH模型在计算我国股市风险价值中的应用研究. *系统工程理论与实践*, 2003, 23(5):20-25.
ZOU Jianjun, ZHANG Zongyi, QIN Zheng. The application of GARCH model in computing the var of Chinese stock market. *System Engineering - Theory & Practice*, 2003, 23(5):20-25.
- [32] 洪晶晶,吴鹏跃. 基于GARCH模型的金融市场波动性分析与预测. *价值工程*, 2016, 35(1):46-47.
HONG Jingjing, WU Pengyue. Analysis and prediction of finance market volatility based on GARCH model. *Value Engineering*, 2016, 35(1):46-47.
- [33] PUTERMAN M L. *Markov decision processes: discrete stochastic dynamic programming*. New York: John Wiley & Sons, 1994.
- [34] EPHRAIM Y, DEMBO A, RABINER L R. A minimum discrimination information approach for hidden Markov modeling. *IEEE Transactions on Information Theory*, 1989, 35(5):1001-1013.
- [35] YU S Z, KOBAYASHI H. An efficient forward-backward algorithm for an explicit-duration hidden Markov model. *IEEE Signal Processing Letters*, 2003, 10(1):11-14.
- [36] 郭存芝. 股票投资收益率预测的随机分析方法研究. *数量经济技术经济研究*, 2002, 19(2):52-55.
GUO Cunzhi. Research on stochastic analysis method of stock investment yield prediction. *The Journal of Quantitative & Technical Economics*, 2002, 19(2):52-55.
- [37] 黄晓彬,王春峰,房振明,等. 基于隐马尔科夫模型的中国股票信息探测. *系统工程理论与实践*, 2012, 32(4):713-720.
HUANG Xiaobin, WANG Chunfeng, FANG Zhenming, et al. Detecting Chinese stock information based on hidden Markov model. *System Engineering - Theory & Practice*, 2012, 32(4):713-720.
- [38] LINDBERG D V, OMRE H. Inference of the transition matrix in convolved hidden Markov models and the generalized Baum-Welch algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(12):6443-6456.
- [39] LINDBERG D V, OMRE H. Inference of the transition matrix in convolved hidden Markov models and the generalized Baum-Welch algorithm. *IEEE Transactions on Geoscience & Remote Sensing*, 2015, 53(12):6443-6456.
- [40] VITERBI A J. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 1967, 13(2):260-269.
- [41] PAGAN A R, SCHWERT G W. Alternative models for conditional stock volatility. *Journal of Econometrics*, 1990, 45(1/2):267-290.
- [42] FERNANDES M, MEDEIROS M C, SCHARTH M. Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 2014, 40:1-10.
- [43] 袁彩萍. 期货市场高频数据的长记忆性研究:以沪深300股指期货为例. 南京:南京理工大学, 2013:14-17.
YUAN Caiping. *Research on long memory of high frequency data in futures market: a case study Shanghai and Shenzhen 300 index futures*. Nanjing: Nanjing University of Science and Technology, 2013:14-17.
- [44] 贾尚晖,江令. 股指期货和股票指数的关联性分析:来自沪深300市场的实证分析. *数学的实践与认识*, 2013, 43(2):90-94.
JIA Shanghui, JIANG Ling. Stock index futures and stock index correlation analysis: an empirical analysis based on the CSI-300 stock market. *Mathematics in Practice and Theory*, 2013, 43(2):90-94.
- [45] 高伟,唐国强,林同智. 基于变结构协整方法的期货市场与证券市场关联性分析:沪深300股指期货与沪深300指数的实证分析. *数学的实践与认识*, 2014, 44(16):65-72.
GAO Wei, TANG Guoqiang, LIN Tongzhi. The relevance analysis between the futures market and stock market based on structure break co-integration: evidence from the CSI300 index future and the spot. *Mathematics in Practice and Theory*, 2014, 44(16):65-72.
- [46] 陶利斌,潘婉彬,黄筠哲. 沪深300股指期货价格发现能力的变化及其决定因素. *金融研究*, 2014(4):128-142.
TAO Libin, PAN Wanbin, HUANG Junzhe. The determinants of price discovery in CSI300 stock index futures. *Journal of Financial Research*, 2014(4):128-142.
- [47] AKAIKE H. Factor analysis and AIC. *Psychometrika*, 1987, 52(3):317-332.
- [48] BUMHAM K P. Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 2004, 33(2):93-101.

Research on Volatility of Chinese Futures Market Based on Hidden Markov Model and GARCH Model

JING Nan¹, LYU Shanshan¹, JIANG Tao²

1 SHU-UTS SILC Business School, Shanghai University, Shanghai 201899, China

2 School of Economics and Management, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China

Abstract: The volatility of a futures market reflects the activity and liquidity of the market. For governments, it is a vital source of decision-making to control the market; while for investors, it is a useful tool to measure risks and achieve asset maintenance. Previous studies have shown that the GARCH (generalized auto regressive conditional heteroskedasticity) model is more responsive to market data than ARCH (auto regressive conditional heteroskedasticity) model because of the inclusion of the prediction variance over the past period. In practice, however, the GARCH model often fails to adapt to the structural changes in the financial market because of the discreteness of the data, which leads to the inefficiency of the volatility prediction effect.

To solve the above problems, this paper attempts to predict the volatility of Chinese futures market returns by combining HMM (hidden Markov model) and GARCH. Firstly, this paper calculates the volatility series of futures by GARCH model. Then it clusters the volatility series to obtain the observation sequence through K-means method. After that, it divides the state of volatility according to the HMM model, and obtains the volatility under different conditions by substituting the corresponding rates of different states into the HMM-GARCH model. Finally, it calculates the volatility index using the VIX (volatility index) formula. Followed by the above logic, this paper selects the CSI 300 stock index futures as the experimental subject, and chooses May 2015 to April 2016 as the sample period to verify the validity of the model.

In the experiment results, the two loss function values (MAD: mean squared difference and MSE: mean squared error) of the HMM-GARCH model are both lower than those of the GARCH model, indicating that the fitting loss and errors of the HMM-GARCH model are less. Compared with the GARCH model, the HMM-GARCH model can better fit sample data and predict market signals. In addition, according to the forecasting of volatility index, the price of CSI 300 stock index futures changed from a small amount of frequent volatility in the previous period to a large jump in volatility. Since then, the volatility remained at a high level and had a growth trend. Eventually, it continued to turn into sharp jumps. Such fluctuation of CSI 300 stock index futures is consistent with its actual fluctuation.

Therefore, the HMM-GARCH model described in this paper can better measure the volatility of Chinese futures market and reflect the expectations of futures investors for future domestic futures markets. At the same time, it can provide the basis for the government to set the pricing of financial derivatives, provide objective quantitative indicators for investors to measure market risk, choose a speculative strategy, and rationally allocate assets, which will help cultivate investors' rational investment and promote the prosperity and stability of Chinese futures market.

Keywords: futures market volatility; hidden Markov model; generalized autoregressive conditional heteroscedasticity method; HMM-GARCH model; volatility index

Received Date: January 20th, 2018 **Accepted Date:** July 4th, 2018

Biography: JING Nan, doctor in computer science, is an associate professor in the SHU-UTS SILC Business School at Shanghai University. His research interests cover decision-making analysis, financial engineering, and information recommendation. His representative paper titled "Modeling co-construction processes in a socio-technical framework to support collaborative engineering design" was published in the *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews* (Issue 3, 2011). E-mail: jingnan@shu.edu.cn

LYU Shanshan is a master degree candidate in the SHU-UTS SILC Business School at Shanghai University. Her research interests cover data modeling, financial engineering, and information recommendation. E-mail: silc_Cassie@163.com

JIANG Tao is a Ph.D candidate in the School of Economics and Management at Harbin Institute of Technology, Shenzhen. His research interests cover information recommendation, financial econometrics, and economic modeling and forecasting. His representative paper titled "Personalized recommendation based on online reviews mining and sentiment assessment on e-commerce website" was published in the *Electronic Commerce Research* (Issue 5, 2017). E-mail: er7248@126.com □