



基于广义谱和 MCS 检验的 VaR 模型预测绩效评估

张玉鹏¹, 洪永森²

1 华东师范大学 金融与统计学院, 上海 200241

2 厦门大学 王亚南经济研究院, 厦门 361005

摘要: 条件 VaR 模型的正确设定检验等价于检验均值化的“撞击序列”是否服从鞅差分序列, 然而通常的反馈检验方法只检验了该序列的部分性质。采用对该鞅差分性质进行直接检验的广义谱检验方法, 全面考察中国股票市场(香港恒生指数、上证综合指数和台湾加权指数)上各参数、非参数和半参数共 22 个 VaR 模型在采用滚动窗口预测机制时的样本外预测绩效。鉴于条件 VaR 模型正确设定检验无法反映超过某 VaR 水平的尾部风险信息, 为避免极端损失的发生以及增加结果的稳健性, 同时采用模型置信集检验方法。研究表明, 采用通常的反馈检验方法常会得出错误的结论; 在 1% 和 5% 置信水平, 与历史模拟法、极值理论模型、CAViaR 模型和 CARE 模型相比, 误差项为 t 分布的 GARCH 模型族在金融危机期间具有较好的样本外预测绩效; 涨跌停板制度对于选取预测绩效最优的 VaR 模型具有重要影响。

关键词: VaR 模型; 预测绩效; 涨跌停板制度; 广义谱检验; MCS 检验

中图分类号: F830.91

文献标识码: A

doi: 10.3969/j.issn.1672-0334.2015.04.010

文章编号: 1672-0334(2015)04-0108-12

1 引言

始于 2007 年的国际金融危机再次引发人们对于金融风险管理问题的关注, 其中一个关键问题是政府监管部门和金融机构能否准确有效地测度金融风险。风险价值(value-at-risk, VaR)模型自被引入风险管理领域以来, 因其直观、易计算的特点, 已成为金融市场风险测量的主流方法。根据是否假设资产收益率的概率分布函数, VaR 模型可分成参数、半参数和非参数 3 类方法, 3 类方法从不同的侧面对 VaR 进行建模, 各类模型相互补充。在实际应用时, 若采用单一的 VaR 模型测度金融风险会产生模型风险, 并可能导致严重的经济损失。因此, 选取科学严谨的 VaR 模型预测评估方法, 准确评价各 VaR 模型在不同资产组合中的样本外预测绩效, 对于政府监管部门和金融机构有效地测度金融风险具有重要的实践意义。

2 相关研究评述

对各 VaR 模型的预测绩效进行评估, 国内外学者

进行了深入的研究。一些学者研究发现基于 GARCH 模型的参数化 VaR 模型具有最优的样本外预测绩效^[1-3]。还有学者认为半参数 VaR 模型具有最优的样本外预测绩效, 如基于极值理论的半参数 VaR 模型^[4-5]、半参数 CAViaR 模型^[6-7]、半参数 CARE 模型^[8-9]、基于滤波历史模拟法的半参数 VaR 模型^[10-11]。对比这些研究可发现, 在评估 VaR 模型的样本外预测绩效时, 关于何种模型具有最优的样本外预测绩效这一问题, 学术界尚未形成一致性的结论, 导致其结论差异性的主要原因在于研究者们选取了不同的样本数据、不同的 VaR 模型估计方法以及不同的 VaR 模型预测评估方法。实际应用时, 研究者常基于其研究的具体目标而选择不同的样本数据, 因此由数据问题导致的实证结论不同是其研究目标差异性的必然结果。但重要的是, 当根据研究目标确定了样本数据后, 研究者能否采用科学严谨的 VaR 模型预测评估方法从广泛的 VaR 模型估计方法集合中选取最优的 VaR 预测模型。

由于已有研究采用 VaR 模型预测评估方法存在

收稿日期: 2015-01-05 修返日期: 2015-04-19

基金项目: 国家自然科学基金(71301053); 教育部人文社会科学研究项目(13YJC790211)

作者简介: 张玉鹏(1980-), 男, 山东烟台人, 经济学博士, 华东师范大学金融与统计学院讲师, 研究方向: 金融计量经济学、国际金融与宏观经济政策等。E-mail: ypzhang@sfs.ecnu.edu.cn

如下3个理论缺陷,会使研究者从VaR备选模型集合中选取错误的最优VaR预测模型。①已有研究采用的反馈检验方法,如Kupiec^[12]的似然比(likelihood ratio, LR)检验、Christoffersen^[13]的LR检验和Engle等^[14]的动态分位数(dynamic quantile, DQ)检验,都只检验条件VaR模型正确设定内涵的部分性质。根据Berkowitz等^[15]的结论,条件VaR模型的正确设定检验等价于检验均值化的“撞击序列”是否服从鞅差分序列。Christoffersen^[13]的LR检验事先假定均值化的“撞击序列”服从一阶马尔科夫过程,因而无法检验均值化“撞击序列”的高阶相依性质。Engle等^[14]的DQ检验将均值化“撞击序列”对常数项、因变量的滞后项以及当期VaR预测值进行线性回归,然后检验回归系数的联合显著性,其不足之处是仅检验了均值化“撞击序列”滞后阶数有限情形以及因变量与解释变量之间的一阶序列相关性。由于上述方法只检验了均值化“撞击序列”的部分性质,这将导致相应检验只有较差的势绩效,即不能正确检测出错误的VaR预测模型。②上述反馈检验方法都没有考虑参数估计效应对检验统计量的影响,由于均值化“撞击序列”中存在未知参数,为构建检验统计量,需要用参数估计值代替未知参数,此时参数估计效应将导致均值化“撞击序列”不再服从鞅差分序列,忽略参数估计效应将导致相应的检验统计量只有较差的水平绩效,即可能会拒绝真实的VaR预测模型。③Abad等^[2]采用的损失函数法和Bao等^[1]采用的Hansen^[16]超预测能力(superior predictive ability, SPA)检验属于VaR模型多元预测评估方法,对模型进行两两比较的损失函数法并不适合VaR模型集较大的情形,通过该方法选择最优模型会存在“数据挖掘”的问题。Bao等^[1]采用的SPA检验需要事先确定基准模型,实际应用时,如果将模型集合中的所有模型依次选为基准模型进行模型评价会违背SPA检验解决“数据挖掘”问题的初衷。

为克服国内外研究中VaR模型预测评估方法的不足,本研究进行如下两方面扩展。①本研究采用Hong等^[17]的广义谱检验考察条件VaR模型的正确设定内涵。区别于Berkowitz等^[15]的谱检验方法,Hong等^[17]的广义谱检验是对均值化“撞击序列”鞅差分序列性质的直接检验,该检验充分利用样本数据中的条件信息,而且明确考虑参数估计效应对检验统计量的影响。因此,与其他的反馈检验方法相比,广义谱方法具有较好的水平绩效和势绩效,本研究的实证结果进一步验证了广义谱检验对VaR模型选择的优越性。②由于所有的条件VaR模型正确设定检验仅关注收益率分布的某个分位点(如1%或5%),无法反映超过分位点的尾部风险信息,在实际应用时可能存在多个具有不同尾部风险的VaR模型通过该检验。为避免极端损失的发生以及增加结果的稳健性,本研究采用Hansen等^[18]的模型置信集(model confidence set, MCS)预测检验方法。为将MCS检验应用于VaR模型的预测绩效评估,研究者须事先设定

损失函数,本研究同时考虑Koenker等^[19]的非对称线性损失函数和Lopez^[20]的幅度损失函数。与Hansen^[16]的SPA检验相比,MCS检验不需要预先设定一个基准模型,并且MCS检验给出了VaR模型集合中未被其他模型显著超越的全部集合特征。

为测度中国大陆股票市场的成熟度以及涨跌停板制度对股票市场风险特征的影响,基于本研究构建的VaR模型预测评估方法,选取包含2007年至2010年国际金融危机时期的上证综指、香港恒生指数和台湾加权指数数据,全面考察金融危机期间各参数法、半参数法和非参数法VaR模型(具体包括历史模拟法、滤波历史模拟法、GARCH模型、极值理论模型、CAViaR和CARE模型,共22个模型)在采用滚动窗口预测机制时的样本外预测绩效。本研究选取国际金融危机时期为预测区间,主要考虑到金融危机为考察大陆股市的成熟度提供了绝佳的试验场(对大陆以及成熟的香港和台湾股市而言,国际金融危机冲击都来自外部且其影响较大),且涨跌停板制度(样本期间大陆和台湾的涨跌幅限制分别为10%和7%,而香港不实行涨跌停板制度)对股市风险特征的影响在金融危机期间可能会表现的更加明显。

3 VaR模型的广义谱检验和MCS预测检验

选择合适的VaR模型对于风险管理者而言是十分重要的工作,只有当VaR模型能准确地预测风险,该模型才是有用的。对VaR模型的预测绩效评估主要包括样本内和样本外方法,从VaR方法的最终目标看,样本外预测绩效的优劣是评价一个VaR模型是否合理的标准。为了评估各VaR模型的样本外预测绩效,本研究将 T 个样本观测值分为 R 个样本内数据和 P 个样本外数据,并且使用滚动窗口预测机制估计模型参数,VaR模型的预测评估主要基于其样本外向前一步预测能力。下面具体介绍Hong等^[17]的广义谱检验和Hansen等^[18]的MCS预测检验。

3.1 VaR模型正确设定的广义谱检验

假设收益率序列 $\{Y_t\}$ 为严格平稳的时间序列过程, t 为时间,其真实条件分位数 $VaR_t(\alpha)$ 可表示为

$$Prob(Y_t < VaR_t(\alpha) | \Omega_{t-1}) = \alpha \quad (1)$$

其中, $Prob(\cdot)$ 为经验累积分布函数, α 为分位数, Ω_{t-1} 为 $(t-1)$ 时刻的信息集。如果存在参数 θ_0 使模型 $VaR_t(\alpha, \theta_0)$ 与真实条件分位数 $VaR_t(\alpha)$ 相吻合,则检验VaR模型正确设定的原假设 $H_0: Prob[Y_t < VaR_t(\alpha, \theta_0) | \Omega_{t-1}] = \alpha$ 等价于检验 $H_0: E[Z_t(\alpha, \theta_0) | \Omega_{t-1}] = 0$, $E[\cdot]$ 为期望, $Z_t(\alpha, \theta_0)$ 为满足鞅差分序列性质的均值化“撞击序列”, $Z_t(\alpha, \theta_0) = I[Y_t < VaR_t(\alpha, \theta_0)] - \alpha$, $I[\cdot]$ 为指示函数。由于 θ_0 是未知参数,根据具体VaR模型设定,本研究采用滚动窗口估计值 $\hat{\theta}_t$ 代替。需强调的是,由于参数估计效应, $E[Z_t(\alpha, \hat{\theta}_t) | \Omega_{t-1}] \neq 0$ 。

为了检验 H_0 是否成立,Hong等^[17]提出了条件VaR模型正确设定的广义谱检验。该方法充分利用样本中的数据信息,并且考虑了参数估计效应对检

验统计量的影响。因而,与其他VaR模型正确设定检验方法相比,广义谱检验更为有效。具体地,其检验统计量可表示为

$$\hat{M}(h) = \frac{\left[\sum_{j=1}^{P-1} k^2\left(\frac{j}{h}\right) (P-j) \int |\hat{\sigma}_{ZT,j}(0,v)|^2 dW(v) - \hat{C}_{ZT}(h) \right]}{\sqrt{\hat{D}_{ZT}(h)}} \quad (2)$$

其中, h 为频宽, $h = h(P)$; j 为滞后阶数; P 为样本外数据个数; $k(\cdot)$ 为对称的核函数, 通常可采用 Bartlett 核函数 $k(z) = (1 - |z|)I(z \leq 1)$ 和 Parzen 核函数 $k(z) = (1 - 6z^2 + 6|z|^3)I(|z| \leq 0.5) + 2(1 - |z|)^3I(0.500 \leq |z| \leq 1)$, z 为任意实数; Z 为均值化“撞击序列”的简化表达式; Y 为样本数据的简化表达式; $\sigma_{ZT,j}(0,v)$ 为总体协方差, $\sigma_{ZT,j}(0,v) = \text{cov}[iZ_t(\alpha, \theta_0), \exp(inY_{t-j})]$, $\hat{\sigma}_{ZT,j}(0,v)$ 为 $\sigma_{ZT,j}(0,v)$ 的样本对应值, i 为虚数单位, \exp 为指数表达式; v 为任意实数; $W(v)$ 为正的、非递减且右连续的函数, 实际应用时可采用标准正态累积分布函数; $\hat{C}_{ZT}(h)$ 和 $\hat{D}_{ZT}(h)$ 分别表示为

$$\hat{C}_{ZT}(h) = \alpha(1 - \alpha) \int |\hat{\sigma}_{YT,0}(v, -v)| dW(v) \sum_{j=1}^{P-1} k^2\left(\frac{j}{h}\right) \quad (3)$$

$$\hat{D}_{ZT}(h) = 2\alpha^2(1 - \alpha)^2 \sum_{j=1}^{P-2} \sum_{\tau=1}^{P-2} k^2\left(\frac{j}{h}\right) k^2\left(\frac{\tau}{h}\right) \cdot \iint |\hat{\sigma}_{YT,j-\tau}(v, v')|^2 dW(v) dW(v') \quad (4)$$

其中, $\hat{\sigma}_{YT,j}(v, v') = \hat{\varphi}_{YT,j}(v, v') - \hat{\varphi}_{YT,j}(v, 0)\hat{\varphi}_{YT,j}(0, v')$, $\hat{\varphi}_{YT,j}(v, v') = \frac{1}{P-1} \sum_{l=R+|j|+1}^T \exp[i(vY_l + v'Y_{l-j})]$, v, v' 为任意实数; τ 为区间 $[1, P-2]$ 内的正整数。Hong 等^[17] 采用非参数插值方法选择数据驱动频宽 $\hat{h}, \hat{h} = h(P) = \max(\ln P, \hat{h}_0)$, 这里 $\hat{h}_0 = \hat{c}_0^{(2q+1)} \sqrt{P}$, q 的取值依赖于核函数 $k(\cdot)$ 的选取, 对于 Bartlett 核函数, $q = 1$, 对于 Parzen 核函数, $q = 2$ 。假设对于核函数 $k(\cdot)$, 存在 $k(q) = \lim_{z \rightarrow 0} \frac{1 - k(z)}{|z|^q} \in (0, \infty)$, 则 \hat{c}_0 可表示为

$$\hat{c}_0 = \frac{2q(k^q)^2}{\int_{-\infty}^{\infty} k^2(z) dz} \cdot \frac{\sum_{j=1}^{P-1} (P-1-j) \bar{k}^2\left(\frac{j}{h}\right) |j|^{2q} \int |\hat{\sigma}_{ZT,j}(0,v)|^2 dW(v)}{\sum_{j=1}^{P-1} (P-1-j) \bar{k}^2\left(\frac{j}{h}\right) \hat{R}_{ZZ}(j) \int \hat{\sigma}_{ZT,j}(v, -v) dW(v)} \Big]^{1/2q+1} \quad (5)$$

其中, \bar{h} 为初始频宽; $\hat{R}_{ZZ}(j) = (P-1-j)^{-1} \sum_{l=R+|j|+1}^T (\hat{Z}_l - \alpha)(\hat{Z}_{l-j} - \alpha)$, T 为样本数据总个数, R 为样本内数据个数, \hat{Z}_t 为 $Z_t(\alpha, \hat{\theta}_t)$ 的简化表达式。Hong 等^[17] 证明, 在适当的正则条件下, 如果 VaR 模型设定正确, $\hat{M}(h) \xrightarrow{d} N(0,1)$, $N(0,1)$ 为标准正态分布函数; 如果

VaR 模型设定错误, $\hat{M}(h) \xrightarrow{d} \infty$ 。在实际应用时, 本研究采用标准正态分布函数的上尾临界值, 如果显著性水平为 5%, 临界值为 1.645。

3.2 MCS 预测检验

Hansen 等^[18] 的 MCS 预测检验依赖于损失函数的准确选择。本研究将同时考虑 Koenker 等^[19] 的非对称线性损失函数 (Lloss) 和 Lopez^[20] 的幅度损失函数 (QLoss)。Koenker 等^[19] 的非对称线性损失函数为

$$L[Var_{t+1}(\alpha, \hat{\theta}_t)] = |\alpha - I[Y_{t+1} < Var_{t+1}(\alpha, \hat{\theta}_t)]| \cdot [Y_{t+1} - Var_{t+1}(\alpha, \hat{\theta}_t)] \quad (6)$$

其中, $L(\cdot)$ 为损失函数, 是 $Var_{t+1}(\alpha, \hat{\theta}_t)$ 的函数, 非对称线性损失函数也被称作“check”函数, 它只考虑了发生极端损失的概率。

Lopez^[20] 的幅度损失函数为

$$L[Var_{t+1}(\alpha, \hat{\theta}_t)] = I[Y_{t+1} < Var_{t+1}(\alpha, \hat{\theta}_t)] \cdot |1 + [Y_{t+1} - Var_{t+1}(\alpha, \hat{\theta}_t)]^2| \quad (7)$$

区别于 Koenker 等^[19] 的非对称线性损失函数, 幅度损失函数既考虑了损失发生的次数, 又考虑了极端尾部事件的幅度, 因而该损失函数更符合经济学意义。Lopez^[20] 的模拟结果表明幅度损失函数能够较好地识别最适合的 VaR 模型。

MCS 预测检验的主要思想是: 考虑一个包含了 m_0 个 VaR 模型的集合 M_0 , 根据损失函数的选择, 模型 m 在 t 时刻的损失定义为 $L_{m,t}$ 。对于所有的 $m \in M_0, n \in M_0$, 定义相对预测绩效变量为 $d_{mn,t}, d_{mn,t} = L_{m,t} - L_{n,t}$ 。对集合 M_0 中的所有模型进行等预测能力检验, 即

$$H_0: E(d_{mn,t}) = 0 \quad m \in M_0, n \in M_0 \quad (8)$$

如果在给定的显著性水平 ψ 下原假设被拒绝, 具有最差预测绩效的模型将从 M_0 中被删去。不断重复上述缩减过程, 直到剩下的模型中不能拒绝原假设为止, 保留下来的模型集合就是模型的置信集 $\hat{M}_{1-\psi}^*$ 。

4 VaR 预测模型

在金融计量中, 常常假设收益率序列 $\{Y_t\}$ 满足如下的波动形式: $Y_t = \mu_t + \sigma_t \varepsilon_t$, μ_t 为 Y_t 的条件均值, σ_t 为 Y_t 的方差函数, ε_t 为满足独立同分布性质的误差项, 服从均值为 0、方差为 1 的标准化概率密度函数 $f_\varepsilon(\cdot; \lambda)$, λ 为密度函数中的参数向量。基于 t 时刻的信息, 向前一步的 VaR 预测值可表示为

$$\widehat{VaR}_{t+1}^\alpha = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} Q_\alpha(\varepsilon) \quad (9)$$

其中, $\hat{\mu}_{t+1}$ 为基于 t 时刻信息的 μ_t 向前一步预测值, $\hat{\sigma}_{t+1}$ 为基于 t 时刻信息的 σ_t 向前一步预测值, $Q_\alpha(\varepsilon)$ 为与 f_ε 相对应的概率分布函数 α 分位数。为得到完全参数化的 VaR 预测值, 需要设定感兴趣的分位数 (即置信水平) α 、条件均值函数 μ_t 、条件方差函数 σ_t 和概率密度函数 f_ε 。本研究主要关注收益率序列的左尾极端损失风险, 假定 α 分别取值为 1% 和 5%。根据是否假设资产收益率的概率密度函数, Engle 等^[14] 将 VaR 估计方法分为参数、非参数和半参数 3 类方法。

下面对这3类VaR估计方法分别进行介绍。

4.1 GARCH模型族

GARCH模型族是最具代表性的VaR模型参数估计方法。为了刻画收益率序列 $\{Y_t\}$ 的序列自相关性,假设 $\mu_t = \rho Y_{t-1}$, ρ 为自回归参数。本研究采用滚动窗口预测机制,对于每个滚动窗口,首先对收益率序列进行零均值化处理,因此没有对条件均值部分设定常数项。为刻画收益率序列的波动聚类效应以及波动的非对称性,本研究采用GARCH(1,1)模型、Glosten等^[21]的GJR-GARCH(1,1)模型和Sentana^[22]的Q-GARCH(1,1)模型。Chen等^[6]和Sener等^[23]研究发现,与对称GARCH模型相比,基于非对称GARCH模型得到的VaR估计值具有更好的样本外预测绩效。考虑到收益率序列的非正态分布特点,本研究考虑3种概率密度函数 f_α ,即标准正态分布、t分布和广义非对称t分布。Abad等^[2]和Polanski等^[24]研究发现,与正态分布相比,基于尖峰厚尾性质的t分布函数所构建的VaR估计值具有更好的预测绩效。Cheng等^[25]和Polanski等^[24]的研究进一步发现,刻画收益率序列的偏态性质对于VaR的准确估计也十分重要。具体地,本研究考虑GARCH-N、GARCH-T、GJR-GARCH-N、GJR-GARCH-T、QGARCH-N、QGARCH-T和GARCH-GAT模型,这里GARCH、GJR-GARCH和QGARCH分别表示收益率条件均值和方差函数服从AR(1)-GARCH(1,1)模型、AR(1)-GJR-GARCH(1,1)模型和AR(1)-Q-GARCH(1,1)模型,N、T和GAT表示各模型的误差项分别服从标准正态分布、t分布和广义非对称t分布(此分布函数可刻画收益率序列的偏态性质)。

为简化计算,JP摩根集团提出的RiskMetrics模型采用简单实用的GARCH(1,1)过程,其向前一步的VaR预测值可表示为 $\widehat{VaR}_{t+1}^\alpha = \hat{\sigma}_{t+1} \Phi^{-1}(\alpha)$, $\Phi(\cdot)$ 为标准正态累积分布函数, $\hat{\sigma}_{t+1}^2 = 0.940\hat{\sigma}_t^2 + 0.060Y_{t-1}^2$ 。

4.2 历史模拟法

历史模拟法(HS)用资产组合收益率数据的样本分位数估计VaR值,该方法计算较为简单。后续研究中,学者们对HS进行了各种改进,以使其更好地估计VaR值,其中表现最优异的是滤波历时模拟法(FHS)。Angelidis等^[10]、Kuester等^[4]和Marimoutou等^[11]研究发现,与其他VaR估计方法相比,基于FHS得到的VaR估计值具有较好的样本外预测绩效。

对于HS, $(t+1)$ 时刻VaR的预测值 $\widehat{VaR}_{t+1}^\alpha = Q_\alpha(Y_t, Y_{t-1}, \dots, Y_{t-w+1})$, $Q_\alpha(\cdot)$ 为 t 时刻之前的 w 个样本观测值的经验 α 分位数,本研究考虑 w 分别取值为500和1000的两种HS模型,即模型HS500和模型HS1000。对于FHS,我们考虑下列AR(1)-GARCH(1,1)模型对原始数据进行滤波,即

$$\begin{aligned} Y_t &= \rho Y_{t-1} + \sigma_t \varepsilon_t \\ \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \gamma \varepsilon_{t-1}^2 \\ t &= 1, 2, \dots \end{aligned} \quad (10)$$

其中, $(\beta_0, \beta_1, \gamma)$ 为GARCH(1,1)模型中的参数向量。

基于该模型,向前一步VaR预测值可表示为 $\widehat{VaR}_{t+1}^\alpha = \hat{\rho} Y_t + \hat{\sigma}_{t+1} Q_\alpha(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-w+1})$, $\hat{\rho}$ 为 ρ 的参数估计值, $\{\hat{\varepsilon}_t\}$ 为标准化残差序列。与HS模型相同,本研究也考虑 w 分别取值为500和1000的两种FHS模型,即模型FHS500和模型FHS1000。

4.3 极值理论模型

极值理论主要关注收益率尾部值的分布,而不必通过直接用于估计整个收益率的分布来推断尾部值,因此通过极值理论对VaR进行估计可以很好地揭示出收益率数据在极端状况下风险的大小。本研究采用对数据量要求较少且能有效使用有限极端观察值的POT模型估计VaR,具体模型介绍参考Bao等^[1]的研究。

由于极值理论的独立同分布假设对于多数资产收益率数据是不合适的,Diebold等^[26]建议首先采用时变波动模型对数据进行拟合,然后利用极值理论模型估计标准化残差的尾部分布。此时,向前一步的VaR预测值可以表示为 $\widehat{VaR}_{t+1}^\alpha = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} Q_\alpha(\hat{\varepsilon}_t)$, $Q_\alpha(\hat{\varepsilon}_t)$ 为用极值理论得到的残差序列 $\{\hat{\varepsilon}_t\}$ 的 α 分位数。如果模型的条件均值和条件方差部分是正确设定的,滤波后的模型残差将几乎是独立同分布的。考虑到收益率序列可能存在的厚尾、偏态以及波动的非对称特征,本研究使用误差项服从t分布或广义非对称t分布的AR(1)-GARCH(1,1)模型消除多数的波动聚类。Marimoutou等^[11]研究发现,基于滤波极值理论模型得到的VaR估计值具有较好的预测绩效。具体地,本研究考虑GARCH-T-EVT和GARCH-GAT-EVT两个滤波极值理论模型,其表示经过AR(1)-GARCH(1,1)滤波且误差项分别服从t分布(T)和广义非对称t分布(GAT)的极值理论模型。

4.4 CAViaR模型

Engle等^[14]把分位数回归模型思想应用于求解VaR,提出条件分位数自回归VaR,即CAViaR模型。其主要思想是将研究的问题由对收益率分布的研究转为直接对分位数的研究,与GARCH模型的建模方法类似,他们用自回归的方法确定分位数随时间的变化规律。具体地,考虑如下4个模型。

(1) 对称绝对值模型(CAViaRSymAbs)

$$Q_t(\alpha) = \delta_0 + \delta_1 Q_{t-1}(\alpha) + \delta_2 |Y_{t-1}| \quad (11)$$

(2) 非对称模型(CAViaRASymSlope)

$$Q_t(\alpha) = \delta_0 + \delta_1 Q_{t-1}(\alpha) + \delta_2 \max[Y_{t-1}, 0] + \delta_3 \max[-Y_{t-1}, 0] \quad (12)$$

(3) 间接GARCH(1,1)模型(CAViaRIndGARCH)

$$Q_t(\alpha) = [1 - 2I(\alpha < 0.500)] [\delta_0 + \delta_1 Q_{t-1}(\alpha)^2 + \delta_2 Y_{t-1}^2]^{1/2} \quad (13)$$

(4) 间接AR(1)-GARCH(1,1)模型(CAViaRIndAR-GARCH)

$$\begin{aligned} Q_t(\alpha) &= aY_{t-1} + [1 - 2I(\alpha < 0.500)] \cdot \\ &\quad \{\delta_0 + \delta_1 [Q_{t-1}(\alpha) + aY_{t-2}]^2 + \\ &\quad \delta_2 (Y_{t-1} - aY_{t-2})^2\}^{1/2} \\ &\quad \{\delta_t\}_{t=0}^2 > 0 \end{aligned} \quad (14)$$

其中, $(\delta_0, \delta_1, \delta_2, \delta_3, a)$ 为未知参数向量。

为估计CAViaR模型中的未知参数,可采用Kocnker等^[19]提出的分位数回归方法,即

$$\min_{\beta} \sum \{ \alpha - I[Y_i < Q_i(\alpha)] \} [Y_i < Q_i(\alpha)] \quad (15)$$

4.5 CARE 模型

基于分位数VaR方法(如CAViaR)的一个不合意性质就是它对极端损失的大小不敏感,无法反映超过分位点的尾部风险,因而基于分位数的VaR方法会使人们忽略了小概率的巨额损失情形甚至是金融危机事件。Newey等^[27]定义的expectile能够很好地解决这一问题,与分位数相比,expectile能够更好地反映不同收益率整体分布函数的差异。基于expectile概念,Taylor^[8]和Kuan等^[28]提出了形式类似的CARE模型。

与CAViaR模型相对应,考虑如下4个CARE模型。

(1) 对称绝对值模型(CARESymAbs)

$$v_i(\tau) = \varphi_0 + \varphi_1 v_{i-1}(\tau) + \varphi_2 |Y_{i-1}| \quad (16)$$

(2) 非对称模型(CAREAsymSlope)

$$v_i(\tau) = \varphi_0 + \varphi_1 v_{i-1}(\tau) + \varphi_2 \max[Y_{i-1}, 0] + \varphi_3 \max[-Y_{i-1}, 0] \quad (17)$$

(3) 间接GARCH(1,1)模型(CAREIndGARCH)

$$v_i(\tau) = [1 - 2I(\tau < 0.500)] \cdot [\varphi_0 + \varphi_1 v_{i-1}(\tau)^2 + \varphi_2 Y_{i-1}^2]^{\frac{1}{2}} \quad (18)$$

(4) 间接AR(1)-GARCH(1,1)模型(CAREIndAR-GARCH)

$$v_i(\tau) = bY_{i-1} + [1 - 2I(\tau < 0.500)] \{ \varphi_0 + \varphi_1 [v_{i-1}(\tau) + bY_{i-2}]^2 + \varphi_2 (Y_{i-1} - bY_{i-2})^2 \}^{\frac{1}{2}} \quad (19)$$

$\{ \varphi_1 \}_{i=0}^{\infty} > 0$

其中, $(\varphi_0, \varphi_1, \varphi_2, \varphi_3, b)$ 为未知参数向量。

为估计CARE模型中的未知参数,本研究采用Newey等^[27]提出的非对称最小二乘法,即

$$\min_{\beta} \sum_i | \alpha - I[Y_i < v_i(\tau)] | [Y_i - v_i(\tau)]^2$$

5 实证分析

5.1 数据描述

本研究选取上证综指(SHCI)、香港恒生指数(HSI)和台湾加权指数(TWII)的日收盘价为数据样本,时间窗口从1999年1月4日至2010年12月31日,所有的股指数据均取自Wind数据库。实证研究建模对象为股指收益率 $r_t, r_t = (\ln \frac{P_t}{P_{t-1}}) \times 100, P_t$ 为 t 时刻的

股指收盘价。对于每一个股指数据,本研究将整个收益率样本分为两部分,用样本数据中最后的1000个收益率数据进行样本外VaR模型的预测绩效评价,用前面的收益率数据进行样本内参数估计。使用滚动窗口预测机制获得向前一步VaR预测值,对于每个滚动窗口,首先对收益率序列零均值化,即 $Y_t = r_t - u, u$ 为滚动窗口内股指收益率的样本均值。

本研究的数据样本区间选取及其区间划分方法主要基于两点考虑。首先,国际金融危机为考察大陆股市成熟度提供了绝佳的试验场。基于本研究的样本内和样本外区间划分标准,上证综指、香港恒生指数和台湾加权指数的预测区间分别始于2006年11月22日、2006年12月11日和2006年12月26日,在这些时间点始于2007年的国际金融危机尚未发生。通过综合比较2007年至2010年国际金融危机期间大陆股票市场与香港、台湾成熟股票市场的风险特征,可在一定程度上说明大陆股票市场的发展程度。其次,中国大陆、香港和台湾实行的是不同的涨跌停板制度,大陆和台湾的涨跌幅限制分别为10%和7%,而香港不实行涨跌停板制度。因而,本研究的样本区间划分有利于分析国际金融危机期间不同涨跌停板制度对中国各股票市场风险特征的差异性影响,即在金融危机期间,相对于大陆和台湾股票市场,不实行涨跌停板制度的香港股票市场是否更有可能发生极端损失。

表1给出了所有股指日收益率在样本内和样本外区间的基本统计特征。①上证综指、香港恒生指

表1 收益率序列的描述性统计

Table 1 The Descriptive Statistics of the Return Series

收益率	样本量	均值	最小值	标准差	偏度	峰度	JB统计量	LB(5)	LB(15)	LB ² (5)	LB ² (15)	
SHCI	样本内	1 895	0.032	-7.920	1.986	0.566	8.088	2 146.000***	10.310*	29.320**	116.000***	220.500***
	样本外	1 000	0.031	-9.256	4.755	-0.375	4.818	161.100***	7.594	26.340**	36.270***	88.730***
HSI	样本内	1 963	0.033	-9.285	1.859	-0.267	6.272	899.100***	12.570**	20.030	68.390***	290.800***
	样本外	1 000	0.020	-13.581	4.641	0.101	8.783	1 395.000***	6.113	25.470**	491.400***	937.000***
TWII	样本内	2 019	0.011	-6.912	2.503	-0.039	4.850	288.400***	17.850***	35.010***	262.000***	535.300***
	样本外	1 000	0.015	-6.735	2.439	-0.356	5.204	223.500***	8.913	31.460***	137.100***	332.000***

注:JB统计量为Jarque-Bera正态性检验;LB(m_1)和LB²(m_1)为收益率序列和其平方序列滞后 m_1 期的Ljung-Box检验统计量;*为相应检验统计量在10%的显著性水平下显著,**为相应检验统计量在5%的显著性水平下显著,***为相应检验统计量在1%的显著性水平下显著。

数和台湾加权指数都有正的平均收益,上证综指和恒生指数的平均收益明显高于台湾加权指数。②无论是在样本内区间还是在样本外区间,相对于上证综指和台湾加权指数,香港恒生指数都具有更小的最小值,说明不实行涨跌停板制度的香港股市更有可能发生极端损失风险。③上证综指和香港恒生指数具有大小和趋势基本相同的样本内和样本外标准差,而台湾加权指数的标准差则相对稳定。这在一定程度上说明,相对于大陆与台湾股票市场的联动性,大陆与香港股票市场的联动性更强。④样本外区间各股指收益率的偏度都不为0,上证综指和台湾加权指数的偏度为负,而香港恒生指数的偏度为正。说明金融危机期间香港恒生指数的主体集中在分布函数的左侧,而上证综指和台湾加权指数的主体集中在分布函数的右侧,因而不实行涨跌停板制度的香港股市在金融危机期间会有更大的概率发生极端损失风险。⑤样本外区间,香港恒生指数比上证综指和台湾加权指数具有更高的峰度,同样说明无涨跌停板制度的香港股市在金融危机期间会产生更为极端的尾部风险(峰度更高,两端的尾部更厚,也就是极值更多)。⑥所有的收益率序列都有显著的一阶和二阶序列自相关,表明所有收益率序列存在明显的序列自相关和波动聚类效应。

5.2 实证结果

表2和表3分别给出在1%和5%置信水平下,Ilong等^[17]的广义谱(HIWZ)检验和Ilansen等^[18]的MCS多元预测检验的实证结果。为了说明基于广义谱和MCS检验方法对VaR模型选择的优劣,表2和表3也同时给出了Christoffersen^[13]的LR检验、Engle等^[14]的DQ检验和Ilansen^[16]的SPA检验的经验 p 值。对HWZ检验,本研究给出基于Bartlett核函数检验统计量的经验 p 值,初始频宽选择为 $\bar{h} = 20$ 。本研究也考察了采用Bartlett核函数和Parzen核函数的广义谱检验统计量的经验 p 值,初始频宽选择为 $\bar{h} = 10$ 和 $\bar{h} = 20$,所有实证检验结果类似,感兴趣的读者可向作者索取相关结果。对于Engle等^[14]的样本外检验统计量 $DQ(m_2)$, m_2 为DQ检验中解释变量所包含的因变量滞后阶数,本研究给出 $m_2 = 20$ 的经验 p 值。也同时考察了 $m_2 = 10$ 和 $m_2 = 30$ 的DQ检验结果,发现 $m_2 = 30$ 时DQ检验结果不太理想, $m_2 = 10$ 和 $m_2 = 20$ 的检验结果类似。与Bao等^[11]类似,Ilansen^[16]的SPA检验的基准模型选择为RiskMetrics模型(原假设为备择模型集中没有模型比RiskMetrics模型的预测效果更好)。需要强调的是,在对表2和表3的实证结果进行分析时,所有统计检验的显著性水平都设定为10%,这主要是为了将VaR模型正确设定检验(即LR检验、DQ检验和HWZ检验)与VaR模型多元预测检验(即SPA检验和MCS检验)的显著性水平选择统一。Hansen等^[16,18]的实证研究均考虑了10%的显著性水平,尽管Ilansen等^[18]的MCS检验也考虑了30%的显著性水平,但实证研究中很少考虑此标准,故本研究将显著性

水平设定为10%。

5.2.1 在1%置信水平下VaR模型的预测绩效评估

由表2可知,对于上证综指,当考虑HWZ检验时,根据其经验 p 值结果,除GARCH-T、GJR-GARCH-T和QGARCH-T模型外,该检验在10%的显著性水平下拒绝了所有其他模型;DQ检验结果与HWZ检验一致;LR检验在10%的显著性水平下只拒绝了10个模型,该结果佐证了LR检验通常具有较差的绩效。上述条件VaR模型正确设定检验结果表明,为了预测上证综指1%置信水平的VaR值,刻画收益率序列的波动集聚效应、波动非对称性和厚尾性质是十分重要的。当损失函数为TLoss时,在10%的显著性水平下MCS包含所有VaR模型中的21个模型,MCS包含更多模型的主要原因是,TLoss损失函数并没有弥补条件VaR正确设定检验未能考虑极端尾部风险的不足,因此基于TLoss的MCS预测检验不能在HWZ的检验结果基础之上提供额外有价值的信息;而当考虑QLoss时,在10%的显著性水平下MCS包含3个模型,这3个模型都包含在损失函数为TLoss时的MCS中,说明两个损失函数的选择具有一定的合理性,QLoss导致MCS中模型较少的原因是这一损失函数对极端风险给予了更大的惩罚因子,一定程度上弥补了条件VaR正确设定检验未能考虑极端尾部风险的问题。与QLoss所对应的MCS中的3个模型就是HWZ检验在10%的显著性水平下没有拒绝的GARCH-T、GJR-GARCH-T和QGARCH-T模型,该结果证实了Lopez^[20]的结论,即QLoss损失函数能够很好地识别正确设定的VaR模型。根据SPA检验结果,在10%的显著性水平下,当考虑TLoss时,SPA检验没有拒绝RiskMetrics模型具有最优预测效果的原假设(经验 p 值为0.520),且该模型包含在此时的MCS置信集中;当考虑QLoss时,SPA检验拒绝了RiskMetrics模型具有最优预测效果的原假设(经验 p 值为0.002),并且该模型也没有包含在MCS集中。SPA检验结果与MCS检验结果相一致,这在一定程度上佐证了MCS检验的合理性。需要强调的是,在实际应用时,不能采用SPA检验将备择模型依次选为基准模型来进行模型评价,这会违背SPA检验解决“数据挖掘”问题的初衷。

对于香港恒生指数,当考虑HWZ检验时,根据其经验 p 值结果,除GARCH-T、GJR-GARCH-T、QGARCH-T、GARCH-T-EVT、GARCH-GAT-EVT和CARESymAbs模型外,该检验在10%的显著性水平下拒绝了所有其他模型;在10%的显著性水平下,DQ检验拒绝了广义谱检验中统计上不显著的QGARCH-T模型,导致该结果的原因是DQ检验没有考虑参数估计效应对检验统计量的影响(即参数估计效应使均值化的“撞击序列”不再服从鞅差分序列,可能存在序列相关性),因而对无序列相关原假设的DQ检验会错误拒绝最适合的模型。此外,在10%的显著性水平下,DQ检验并没有拒绝广义谱检验中统计上显著的CAViaRASymSlope模型、CAREAsymSlope模型和CAREIndGARCH模型,导致该结果的原因是DQ检验仅检验了均值化“撞击序

表2 在1%置信水平下各股指VaR模型的预测绩效评估
Table 2 The Predictive Performance Evaluation of Each Stock Index at 1% Level by VaR Model

模型	SHCI					HSI					TWII				
	LR	DQ	HWZ	QLoss	TLoss	LR	DQ	HWZ	QLoss	TLoss	LR	DQ	HWZ	QLoss	TLoss
GARCHII 模型															
RiskMetrics	0.061	0.000	0.000	0.105	0.084 *	0.385	0.002	0.000	0.043	0.060	0.004	0.000	0.000	0.056	0.018
GARCHII-N	0.002	0.000	0.000	0.119	0.085 *	0.385	0.002	0.000	0.043	0.060	0.006	0.000	0.000	0.051	0.021
GARCH-T	0.524	0.335	0.215	0.034 *	0.078 *	0.886	0.264	0.986	0.014 *	0.058	0.000	0.000	0.000	0.104	0.027
GJRGARCH-N	0.003	0.000	0.000	0.116	0.083 *	0.385	0.001	0.000	0.034	0.056 *	0.015	0.000	0.000	0.047	0.021 *
GJRGARCHII-T	0.886	1.000	0.971	0.035 *	0.078 *	0.886	0.256	0.967	0.012 *	0.055 *	0.944	0.266	0.612	0.013 *	0.027 *
QGARCH-N	0.002	0.000	0.000	0.123	0.085 *	0.385	0.002	0.000	0.040	0.057	0.020	0.000	0.000	0.049	0.021
QGARCHII-T	0.524	0.326	0.215	0.039 *	0.078 *	0.786	0.094	0.997	0.012 *	0.056 *	0.786	0.779	0.993	0.011 *	0.026 *
GARCH-GAT	0.000	0.000	0.000	0.235	0.116	0.500	0.003	0.004	0.019 *	0.053 *	0.000	0.000	0.000	0.325	0.025
历史模拟法															
IIS500	0.205	0.000	0.000	0.091	0.082 *	0.001	0.000	0.000	0.164	0.091	0.446	0.000	0.000	0.033	0.037
HS1000	0.008	0.000	0.000	0.143	0.093 *	0.000	0.000	0.000	0.218	0.097	0.085	0.000	0.000	0.044	0.043
FHS500	0.000	0.000	0.000	0.206	0.105 *	0.000	0.000	0.000	0.082	0.068	0.052	0.000	0.000	0.041	0.026
FHS1000	0.000	0.000	0.000	0.190	0.102 *	0.562	0.059	0.004	0.026	0.057	0.457	0.037	0.000	0.029	0.022 *
极值理论模型															
GARCH-T-EVT	0.002	0.000	0.000	0.135	0.096 *	0.928	0.226	0.351	0.054	0.064	0.278	0.000	0.000	0.035	0.024
GARCH-GAT-EVT	0.076	0.000	0.000	0.088	0.082 *	0.928	0.223	0.351	0.070	0.067	0.000	0.000	0.000	0.527	0.023
CAViaR 模型															
CAViaRSmAbs	0.205	0.000	0.000	0.089	0.086 *	0.446	0.002	0.000	0.029	0.057	0.670	0.768	0.001	0.023	0.029 *
CAViaRA-symSlope	0.500	0.000	0.000	0.068	0.086 *	0.562	0.377	0.003	0.025	0.053 *	0.457	0.206	0.000	0.033	0.028 *
CAViaR-IndGARCH	0.118	0.000	0.000	0.090	0.083 *	0.500	0.002	0.000	0.028	0.057	0.562	0.501	0.000	0.024	0.031 *
CAViaRIndAR-GARCHII	0.205	0.000	0.000	0.099	0.084 *	0.540	0.000	0.015	0.027	0.056 *	0.562	0.289	0.000	0.023	0.028 *
CARE 模型															
CARESymAbs	0.260	0.000	0.000	0.072	0.083 *	0.944	1.000	0.957	0.014 *	0.055 *	0.562	0.224	0.001	0.025	0.030 *
CARE-FA-symSlope	0.385	0.000	0.000	0.063	0.081 *	0.457	0.731	0.000	0.025	0.051 *	0.152	0.007	0.000	0.037	0.029 *
CARE-IndGARCH	0.118	0.000	0.000	0.096	0.084 *	0.446	0.319	0.000	0.035	0.059	0.035	0.005	0.000	0.040	0.028 *
CAREIndAR-GARCHII	0.205	0.000	0.000	0.098	0.084 *	0.148	0.000	0.000	0.039	0.058	0.361	0.201	0.001	0.028	0.030 *
SPA				0.002	0.520				0.026	0.009				0.000	0.091

注: LR为Christoffersen^[13] LR条件覆盖率检验的经验 p 值; DQ为Engle等^[14]样本外检验统计量 $DQ(m_2)$ 的经验 p 值, $m_2 = 20$; HWZ为Hong等^[17]采用Bartlett核函数的检验统计量 $\hat{M}_0(\hat{h})$ 的经验 p 值, \hat{h} 为基于初始频宽 $\bar{h} = 20$ 的数据驱动频宽; TLoss为Koenker等^[19]的非对称线性损失函数, QLloss为Lopez^[20]的幅度损失函数, 表中的数值表示相应VaR模型的平均损失; *为10%显著性水平下模型置信集 $\hat{M}_{90\%}$ 中所包含的VaR模型, 采用10 000个Bootstrap样本计算MCS置信集, Block Bootstrap的区块长度为2; SPA为Hansen^[16]的SPA检验的经验 p 值, SPA检验的基准模型为RiskMetrics模型, Stational Bootstrap的平滑参数为0.500。下同。

列”滞后阶数有限情形以及因变量与解释变量之间的一阶序列相关性,因而具有较差的势绩效;LR检验在10%的显著性水平下只拒绝了3个模型,该结果佐证了LR检验通常具有较差的势绩效。鉴于LR检验和DQ检验具有较差的有限样本绩效,本研究将采用Hong等^[17]的广义谱检验结果。上述条件VaR模型正确设定检验结果表明,为了预测香港恒生指数1%置信水平下的VaR值,刻画收益率序列的波动集聚效应、波动非对称性和厚尾性质是重要的,此外极值理论模型和同时考虑损失概率与损失幅度的CARE-SymAbs模型也可用来对该指数1%置信水平下的VaR进行建模。当损失函数为TLoss时,在10%的显著性水平下MCS包含了所有VaR模型中8个模型,这8个模型中包含了HWZ检验中在10%的显著性水平下没有拒绝的6个模型中的GJRARCH-T、QGARCH-T和CARESymAbs模型;当考虑QLoss时,在10%的显著性水平下MCS包含5个模型,除GARCH-T模型,其余4个模型都包含在损失函数为TLoss时的MCS中。与QLoss对应的MCS中的5个模型中包含HWZ检验中在10%的显著性水平下没有拒绝的6个模型中的GARCH-T、GJRARCH-T、QGARCH-T和CARESymAbs模型。根据SPA检验结果,当分别考虑QLoss和TLoss时,SPA检验统计量的经验 p 值分别为0.026和0.009,在10%的显著性水平下SPA检验都拒绝了RiskMetrics模型具有最优预测效果的原假设,且MCS置信集在10%的显著性水平下都没有包含RiskMetrics模型,说明SPA检验结果与MCS检验结果一致,在一定程度上佐证了MCS检验的合理性。需强调的是,SPA检验只能考察单一基准模型是否比其他备择模型具有更好的预测绩效。

对于台湾加权指数,当考虑HWZ检验时,根据其经验 p 值结果,除GJRARCH-T和QGARCH-T外,该检验在10%的显著性水平下拒绝了所有其他模型;在10%的显著性水平下,DQ检验并没有拒绝广义谱检验中统计上显著的所有CAViaR模型以及CARESymAbs模型和CAREIndARGARCH模型;LR检验也只拒绝了10个模型,该实证结果佐证了DQ检验和LR检验通常具有较差的势绩效。上述条件VaR正确设定检验结果表明,为了预测台湾加权指数1%置信水平下的VaR值,刻画收益率序列的波动集聚效应、波动的非对称性和厚尾性质是十分重要的。当损失函数为TLoss时,在10%的显著性水平下MCS包含所有VaR模型中的12个模型;而当考虑QLoss时,在10%的显著性水平下MCS包含2个模型,且这2个模型都包含在损失函数为TLoss时的MCS中。与QLoss所对应的MCS中的2个模型就是HWZ检验在10%的显著性水平下没有拒绝的GJRARCH-T和QGARCH-T模型。根据SPA检验结果,当分别考虑QLoss和TLoss时,SPA检验统计量的经验 p 值分别为0.000和0.091,SPA检验在10%的显著性水平下都拒绝了RiskMetrics模型具有最优预测效果的原假设,并且MCS置信集在10%的显著性水平下都没有包含RiskMetrics模型,说明SPA检验结果

与MCS检验结果一致,在一定程度上佐证了MCS检验的合理性。需强调的是,SPA检验只能考察单一基准模型是否比其他备择模型具有更好的预测绩效。

5.2.2 在5%置信水平下VaR模型的预测绩效评估

由表3可知,对于上证综指,当考虑HWZ检验时,根据其经验 p 值结果,除GARCH-T、GJRARCH-T、QGARCH-T、CAREAsymSlope和CAREIndGARCH模型外,该检验在10%的显著性水平下拒绝了所有其他模型;在10%的显著性水平下,DQ检验并没有拒绝广义谱检验结果中统计上显著的RiskMetrics模型、服从正态分布的GARCH模型族、所有的CAViaR模型和CARE模型;LR检验也只拒绝了6个VaR预测模型,该结果佐证了DQ检验和LR检验通常具有较差的势绩效。上述条件VaR正确设定检验结果表明,为了预测上证综指5%置信水平的VaR值,刻画收益率序列的波动集聚效应、波动非对称性和厚尾性质是十分重要的,此外同时考虑损失概率和损失幅度的半参数模型CAREAsymSlope和CAREIndGARCH模型也可以很好地刻画该指数5%置信水平的VaR值。当损失函数为TLoss时,在10%的显著性水平下MCS包含所有VaR模型中的16个模型,这16个模型中包含HWZ检验中在10%的显著性水平下没有拒绝的全部5个模型。与HWZ检验结果相比,基于TLoss的MCS预测检验未能提供各VaR模型预测绩效的更多有价值信息。而当考虑QLoss时,在10%的显著性水平下MCS只包含了2个模型,且这2个模型都包含在损失函数为TLoss时的MCS中;与QLoss所对应的MCS中的2个模型是HWZ检验在10%的显著性水平下没有拒绝的5个模型中的GARCH-T和GJRARCH-T模型。根据SPA检验结果,在10%的显著性水平下,当考虑TLoss时,SPA检验没有拒绝RiskMetrics模型具有最优预测效果的原假设(经验 p 值为0.903),且该模型包含在MCS置信集中;当考虑QLoss时,SPA检验拒绝了RiskMetrics模型具有最优预测效果的原假设(经验 p 值为0.000),并且该模型也没有包含在MCS集中。SPA检验结果与MCS检验结果一致,这在一定程度上佐证了MCS检验的合理性。

对于香港恒生指数,当考虑HWZ检验时,根据其经验 p 值结果,除GARCH-T、GJRARCH-T和QGARCH-T模型外,该检验在10%的显著性水平下拒绝了所有其他模型;在10%的显著性水平下,DQ检验并没有拒绝广义谱检验结果中统计上显著的GARCH-N模型、QGARCH-N模型、除CAViaRIndARGARCH以外所有其他CAViaR模型、CARESymAbs模型和CAREIndGARCH模型;LR检验也只拒绝了5个VaR预测模型。上述条件VaR正确设定检验结果表明,为了预测香港恒生指数5%置信水平的VaR值,刻画收益率序列的波动集聚效应、波动的非对称性和厚尾性质是重要的。当损失函数为TLoss时,在10%的显著性水平下MCS包含了所有VaR模型中14个模型。当考虑QLoss时,在10%的显著性水平下MCS包含了3个模型,且其中的GJRARCH-T和QGARCH-T两个模型包含在损失函

表3 在5%置信水平下各股指VaR模型的预测绩效评估
Table 3 The Predictive Performance Evaluation of Each Stock Index at 5% Level by VaR Model

模型	SHCI					HSI					TWII				
	LR	DQ	HWZ	QLoss	TLoss	LR	DQ	HWZ	QLoss	TLoss	LR	DQ	HWZ	QLoss	TLoss
GARCH 模型															
RiskMetrics	0.689	0.486	0.031	0.298	0.264 *	0.441	0.049	0.020	0.185	0.211	0.263	0.278	0.010	0.161	0.181 *
GARCH-N	0.538	0.247	0.003	0.314	0.268 *	0.358	0.125	0.005	0.189	0.211 *	0.463	0.111	0.038	0.155	0.181 *
GARCH-T	0.979	0.624	0.348	0.207 *	0.266 *	0.488	0.173	0.927	0.127 *	0.213	0.000	0.000	0.000	0.419	0.264
GJRGARCH-N	0.538	0.120	0.001	0.306	0.266 *	0.266	0.029	0.010	0.176	0.205 *	0.441	0.167	0.001	0.149	0.179 *
GJRGARCH-T	0.973	0.672	0.479	0.198 *	0.265 *	0.722	0.420	0.778	0.119 *	0.205 *	0.668	0.579	0.640	0.090 *	0.180 *
QGARCH-N	0.538	0.252	0.003	0.317	0.268 *	0.358	0.103	0.045	0.186	0.207 *	0.266	0.311	0.001	0.161	0.179 *
QGARCH-T	0.979	0.594	0.393	0.211	0.266 *	0.766	0.699	0.633	0.130 *	0.207 *	0.732	0.528	0.683	0.096	0.178 *
GARCH-GAT	0.000	0.000	0.000	0.581	0.321	0.220	0.001	0.042	0.156	0.207 *	0.000	0.000	0.000	0.683	0.856
历史模拟法															
HS500	0.182	0.000	0.000	0.360	0.285	0.009	0.000	0.000	0.425	0.272	0.640	0.000	0.000	0.203	0.207
HS1000	0.000	0.000	0.000	0.535	0.308	0.000	0.000	0.000	0.630	0.291	0.023	0.000	0.000	0.277	0.221
FHS500	0.000	0.000	0.000	0.785	0.333	0.000	0.000	0.000	0.352	0.230	0.033	0.000	0.000	0.222	0.187
FHS1000	0.000	0.000	0.000	0.648	0.308	0.014	0.000	0.000	0.223	0.213 *	0.055	0.009	0.000	0.203	0.184 *
极值理论模型															
GARCH-T-EVT	0.000	0.000	0.000	0.543	0.316	0.328	0.002	0.001	0.271	0.222	0.000	0.000	0.000	0.289	0.225
GARCH-GAT-EVT	0.002	0.000	0.000	0.391	0.282 *	0.182	0.003	0.000	0.305	0.227	0.000	0.000	0.000	0.539	0.793
CAViaR 模型															
CAViaRSmAbs	0.689	0.401	0.015	0.302	0.266 *	0.658	0.187	0.010	0.168	0.209 *	0.607	0.186	0.030	0.151	0.181 *
CAViaRA-symSlope	0.689	0.538	0.024	0.302	0.266 *	0.295	0.114	0.004	0.152	0.204 *	0.733	0.813	0.204	0.148	0.179 *
CAViaR-IndGARCH	0.538	0.193	0.012	0.321	0.268 *	0.574	0.301	0.001	0.178	0.210 *	0.607	0.184	0.030	0.150	0.181 *
CAViaRIndAR-GARCH	0.538	0.111	0.006	0.322	0.269 *	0.328	0.083	0.008	0.162	0.207 *	0.800	0.196	0.018	0.145	0.181 *
CARE 模型															
CARESymAbs	0.665	0.636	0.077	0.272	0.263 *	0.684	0.395	0.017	0.158	0.209 *	0.689	0.225	0.001	0.153	0.183 *
CAREAsymSlope	0.767	0.211	0.467	0.252	0.276 *	0.097	0.002	0.000	0.167	0.205 *	0.689	0.166	0.004	0.151	0.180 *
CARE-IndGARCH	0.901	0.592	0.170	0.278	0.264 *	0.457	0.118	0.001	0.185	0.213	0.851	0.145	0.030	0.141	0.182 *
CAREIndAR-GARCH	0.860	0.501	0.068	0.292	0.265 *	0.209	0.000	0.016	0.182	0.209 *	0.640	0.007	0.000	0.150	0.181 *
SPA				0.000	0.903				0.001	0.016				0.000	0.641

数为TLoss的MCS中;与QLoss所对应的MCS中的3个模型就是HWZ检验中在10%的显著性水平下没有拒绝的GARCH-T、GJRGARCH-T和QGARCH-T模型。根据

SPA检验结果,当分别考虑QLoss和TLoss时,SPA检验统计量的经验p值分别为0.001和0.016,SPA检验在10%的显著性水平下都拒绝了RiskMetrics模型具有最

优预测效果的原假设,并且MCS置信集在10%的显著性水平下都没有包含RiskMetrics模型。说明SPA检验结果与MCS检验结果一致,在一定程度上佐证了MCS检验的合理性。

对于台湾加权指数,当考虑HWZ检验时,根据其经验 p 值结果,除了GJRGARCH-T、QGARCH-T和CAViaRAsymSlope模型外,该检验在10%的显著性水平下拒绝了所有其他模型;在10%的显著性水平下,DQ检验并没有拒绝广义谱检验结果中统计上显著的RiskMetrics模型、服从正态分布的GARCH模型族、所有的CAViaR模型以及除了CAREIndARGARCH模型以外其他CARE模型;LR检验也只拒绝了7个模型。上述条件VaR正确设定检验结果表明,为了预测台湾加权指数5%置信水平的VaR值,刻画收益率序列的波动集聚效应、波动的非对称性和厚尾性质是重要的。当损失函数为TLoss时,在10%的显著性水平下MCS包含所有VaR模型中的15个模型;而当考虑QLoss时,在10%的显著性水平下MCS只包含了1个模型,且这1个模型包含在损失函数为TLoss时的MCS中。与QLoss所对应的MCS中的这个模型是HWZ检验在10%的显著性水平下没有拒绝的3个模型中的GJRGARCH-T模型。根据SPA检验结果,在10%的显著性水平下,当考虑TLoss时,SPA检验没有拒绝RiskMetrics模型只有最优预测效果的原假设(经验 p 值为0.641),且该模型包含在MCS置信集中;当考虑QLoss时,SPA检验拒绝了RiskMetrics模型具有最优预测效果的原假设(经验 p 值为0.000),并且该模型也没有包含在MCS集中。SPA检验结果与MCS检验结果一致,这在一定程度上佐证了MCS检验的合理性。

综上所述,结合Hong等^[17]的广义谱检验和基于QLoss损失函数的Hansen等^[18]的MCS预测检验的实证结果,在1%置信水平下,上证综指的最优VaR预测模型为GARCH-T、GJRGARCH-T和QGARCH-T,香港恒生指数的最优预测模型为GARCH-T、GJRGARCH-T、QGARCH-T和CARESymAbs,台湾加权指数的最优VaR预测模型为GJRGARCH-T和QGARCH-T;在5%置信水平下,上证综指的最优VaR预测模型为GARCH-T和GJRGARCH-T,香港恒生指数的最优预测模型为GARCH-T、GJRGARCH-T和QGARCH-T,台湾加权指数的最优预测模型为GJRGARCH-T。需要说明的是,在1%置信水平下,香港恒生指数的最优VaR预测模型包含了CARESymAbs模型,而上证综指和台湾加权指数的最优VaR预测模型并没有包含CARESymAbs模型。该结果表明,对于不实行涨跌停板制度的香港股票市场而言,采用能够测度小概率巨额损失情形甚至是金融危机事件的CARE模型是十分重要的。

6 结论

本研究采用更加严谨和稳健的Hong等^[17]的VaR模型正确设定的广义谱检验和Hansen等^[18]的MCS多元预测检验方法,全面考察中国股票市场(上证综指、恒生指数和台湾加权指数)上各参数、半参数和

非参数VaR模型(共22个模型)在采用滚动窗口预测机制时的向前一步预测绩效。实证结果表明,在1%和5%置信水平下,同时刻而收益率的波动集聚效应、波动非对称性和厚尾性质的GARCH模型族可以很好地预测上证综指、香港恒生指数和台湾加权指数在国际金融危机期间的风险特征,这在某种程度上表明经过20多年的发展,随着监管体系的逐步完善、机构投资者的日益壮大、证券品种结构的日趋合理和发行规模的不断扩大,大陆股票市场与香港、台湾成熟股票市场的风险特性越来越接近;此外,在1%置信水平下,香港恒生指数的最优VaR预测模型包含CARESymAbs模型,说明是否实行涨跌停板制度对于选取最优的VaR预测模型具有重要的影响。

由于通常的反馈检验方法只检验条件VaR模型正确设定内涵(即均值化的“撞击序列”应服从鞅差分序列)的部分性质,在实际应用时采用这些方法常常会得出错误的结论。为避免极端损失的发生以及增加结果的稳健性,政府监管部门和金融机构在测度或预测市场风险时,可结合Hong等^[17]的条件VaR模型正确设定的广义谱检验和Hansen等^[18]的MCS多元预测检验方法。

与GARCH模型族相比,对各股票指数的VaR值进行建模的复杂模型(如经过滤波的极值理论模型)和简单模型(如各种历史模拟法模型)具有相对较差的预测绩效。基于此实证结论,建议监管部门和金融机构在金融危机期间,为了准确地预测未来的市场风险,不应该采用过度参数化和较简单的VaR模型,通常GARCH模型族会具有较为稳健的样本外预测能力。对于政府监管部门和金融机构而言,准确有效地测度金融风险是风险管理中非常重要的工作。本研究为VaR模型预测绩效评价工作提供有益的理论借鉴和具有操作性的实证方法,可采用本研究的方法考察不同VaR水平下各种股指收益率或资产组合在不同市场表现情况(如牛市或熊市)下的风险特征。

本研究尚有不足之处。①本研究只考虑了各VaR模型的样本外向前一步预测能力,实际应用时,政府监管部门和金融研究机构常需了解未来一段时间内各股指收益率或资产组合的风险特征,因而采用本研究的预测评估方法考察各VaR模型的样本外多步预测能力将是未来的一项重要研究内容;②本研究虽然考察了各参数、半参数和非参数VaR模型(共22个模型)的样本外预测绩效,但是仍有一些模型尚未纳入其中,如Ergün等^[29]的高阶矩模型、Yu等^[30]的CAViaR扩展模型等(更详细的VaR模型估计方法见Abad等^[31]的文献综述文章),未来将进一步考察这些VaR模型在中国股票市场中的样本外预测绩效。

参考文献:

- [1] Bao Y, Lee T H, Saltoglu B. Evaluating predictive performance of value-at-risk models in emerging

- markets: A reality check [J]. *Journal of Forecasting*, 2006, 25(2): 101-128.
- [2] Abad P, Benito S. A detailed comparison of value at risk estimates [J]. *Mathematics and Computers in Simulation*, 2013, 94: 258-276.
- [3] 淳伟德, 陈王, 潘攀. 典型事实约束下的上海原油期货期货市场动态 VaR 测度研究 [J]. *中国管理科学*, 2013, 21(2): 24-31.
Chun Weide, Chen Wang, Pan Pan. A study on dynamic VaR predicting models for oil futures market of Shanghai [J]. *Chinese Journal of Management Science*, 2013, 21(2): 24-31. (in Chinese)
- [4] Kuester K, Mittnik S, Paolella M S. Value-at-risk prediction: A comparison of alternative strategies [J]. *Journal of Financial Econometrics*, 2006, 4(1): 53-89.
- [5] 林宇, 黄登仕, 魏宇. 胖尾分布及长记忆下的动态 EVT-VaR 测度研究 [J]. *管理科学学报*, 2011, 14(7): 71-82.
Lin Yu, Huang Dengshi, Wei Yu. Study on financial markets dynamic EVT-VaR measuring based on fat-tail distribution and long memory volatility [J]. *Journal of Management Sciences in China*, 2011, 14(7): 71-82. (in Chinese)
- [6] Chen C W S, Gerlach R, Lin E M H, Lee W C W. Bayesian forecasting for financial risk management, pre and post the global financial crisis [J]. *Journal of Forecasting*, 2012, 31(8): 661-687.
- [7] 张颖, 张富祥. 分位数回归的金融风险度量理论及实证 [J]. *数量经济技术经济研究*, 2012, 29(4): 95-109.
Zhang Ying, Zhang Fuxiang. The theory and empirical research of quantile regression in financial risk measurement [J]. *The Journal of Quantitative & Technical Economics*, 2012, 29(4): 95-109. (in Chinese)
- [8] Taylor J W. Estimating value at risk and expected shortfall using expectiles [J]. *Journal of Financial Econometrics*, 2008, 6(2): 231-252.
- [9] 钟山, 傅强. 基于 CARE 模型的金融市场 VaR 和 ES 度量 [J]. *预测*, 2014, 33(3): 40-44.
Zhong Shan, Fu Qiang. The estimation of value at risk and expected shortfall in financial market based on CARE models [J]. *Forecasting*, 2014, 33(3): 40-44. (in Chinese)
- [10] Angelidis T, Benos A, Degiannakis S. A robust VaR model under different time periods and weighting schemes [J]. *Review of Quantitative Finance and Accounting*, 2007, 28(2): 187-201.
- [11] Marimoutou V, Raggad B, Trabelsi A. Extreme value theory and value at risk: Application to oil market [J]. *Energy Economics*, 2009, 31(4): 519-530.
- [12] Kupiec P H. Techniques for verifying the accuracy of risk measurement models [J]. *The Journal of Derivatives*, 1995, 3(2): 73-84.
- [13] Christoffersen P F. Evaluating interval forecasts [J]. *International Economic Review*, 1998, 39(4): 841-862.
- [14] Engle R F, Manganelli S. CAViaR: Conditional autoregressive value at risk by regression quantiles [J]. *Journal of Business & Economic Statistics*, 2004, 22(4): 367-381.
- [15] Berkowitz J, Christoffersen P, Pelletier D. Evaluating value-at-risk models with desk-level data [J]. *Management Science*, 2011, 57(12): 2213-2227.
- [16] Hansen P R. A test for superior predictive ability [J]. *Journal of Business & Economic Statistics*, 2005, 23(4): 365-380.
- [17] Hong Y, Wang S, Zhang W. Testing for conditional quantile models in time series with application to value at risk [R]. Ithaca, NY: Cornell University, 2011.
- [18] Hansen P R, Lundc A, Nason J M. The model confidence set [J]. *Econometrica*, 2011, 79(2): 453-497.
- [19] Koenker R, Bassett G, Jr. Regression quantiles [J]. *Econometrica*, 1978, 46(1): 33-50.
- [20] Lopez J A. Methods for evaluating value-at-risk estimates [J]. *Federal Reserve Bank of San Francisco Economic Review*, 1999(2): 3-17.
- [21] Glosten L R, Jagannathan R, Runkle D E. On the relation between the expected value and the volatility of the nominal excess return on stocks [J]. *The Journal of Finance*, 1993, 48(5): 1779-1801.
- [22] Sentana E. Quadratic ARCH models [J]. *The Review of Economic Studies*, 1995, 62(4): 639-661.
- [23] Sener E, Baronyan S, Mengütürk I. A. Ranking the predictive performances of value-at-risk estimation methods [J]. *International Journal of Forecasting*, 2012, 28(4): 849-873.
- [24] Polanski A, Stoja E. Incorporating higher moments into value-at-risk forecasting [J]. *Journal of Forecasting*, 2010, 29(6): 523-535.
- [25] Cheng W H, Hung J C. Skewness and leptokurtosis in GARCH-typed VaR estimation of petroleum and metal asset returns [J]. *Journal of Empirical Finance*, 2011, 18(1): 160-173.
- [26] Diebold F X, Schuermann T, Stroughair J D. Pitfalls and opportunities in the use of extreme value theory in risk management [J]. *The Journal of Risk Finance*, 2000, 1(2): 30-35.
- [27] Newey W K, Powell J L. Asymmetric least squares estimation and testing [J]. *Econometrica*, 1987, 55(4): 819-847.

- [28] Kuan C M, Yeh J H, Hsu Y C. Assessing value at risk with CARE, the conditional autoregressive expectile models [J]. *Journal of Econometrics*, 2009, 150(2):261-270.
- [29] Ergün A T, Jun J. Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall [J]. *The Quarterly Review of Economics and Finance*, 2010, 50(3):264-272.
- [30] Yu P L H, Li W K, Jin S. On some models for value-at-risk [J]. *Econometric Reviews*, 2010, 29(5/6):622-641.
- [31] Abad P, Benito S, López C. A comprehensive review of value at risk methodologies [J]. *The Spanish Review of Financial Economics*, 2014, 12(1):15-32.

Evaluating Predictive Performance of Value-at-Risk Models Based on Generalized Spectrum and MCS Tests

Zhang Yupeng¹, Hong Yongmiao²

1 School of Finance and Statistics, East China Normal University, Shanghai 200241, China

2 The Wang Ya'nan Institute for Studies in Economics, Xiamen University, Xiamen 361005, China

Abstract: Conditional VaR models' correct specification test is equivalent to testing the de-meaned hit sequence following a martingale difference sequence (m. d. s), however the commonly used backtesting techniques only test some properties of the sequence. Using generalized spectral test which directly tests the m. d. s property of the de-meaned hit sequence, we evaluate the out-of-sample predictive performance of various parametric, nonparametric and semi-parametric VaR models with a total of 22 models calculated by using rolling predictive method for China's stock markets including Shanghai Composite Index, Hang Seng Index and Taiwan Weighted Index. Because conditional VaR models' correct specification test can not reflect the tail risk information exceeding one specific VaR level, in order to avoid the occurrence of extreme losses as well as to increase the robustness of the results, we adopt MCS (model confidence set) test simultaneously by selecting the asymmetric loss functions proposed by Koenker and Bassett and the magnitude loss function proposed by Lopez. Comparing with SPA (Superior Predictive Ability) test, the main advantage of MCS test is that it does not require a benchmark model to be specified as is the case for SPA tests. It characterizes the entire set of models that are not significantly outperformed by other models, while a test for SPA only provides evidence about the relative performance of a single model (the benchmark).

The empirical results imply the following three conclusions: ① it would cause wrong results using the commonly applied backtesting techniques such as Kupiec likelihood ratio test, Christoffersen likelihood ratio test and Engle and Manganelli dynamic quantile test. However adopting generalized spectral test and MCS test with Lopez loss function simultaneously would give us more accurate and robust results. ② Comparing with historical simulation models, extreme value theory models, CAViaR and CARE models, the out-of-sample predictive performance of the GARCH family with student-t distribution is the best at 1% and 5% significant level during the financial crisis for these three stock indexes. This implies that the risk characteristics of mainland stock market is getting closer and closer to the mature stock markets of Hong Kong and Taiwan after more than 20 years development. ③ At 1% significant level, the optimal VaR predictive models of Hang Seng Index include one of the CARE models which can be used to measure extreme loss situation with small probability. This implies that price limit system implemented by Hong Kong yet not by mainland and Taiwan will make Hong Kong's stock market face more risk during the financial crisis.

Keywords: VaR model; predictive performance; price limit system; generalized spectral test; MCS test

Received Date: January 5th, 2015 **Accepted Date:** April 19th, 2015

Funded Project: Supported by the National Natural Science Foundation of China(71301053) and the Humanity and Social Sciences Research of Ministry of Education(13YJC790211)

Biography: Zhang Yupeng (1980 - , Native of Yantai, Shandong), Doctor in Economics and is a Lecturer in the School of Finance and Statistics at East China Normal University. His research interests include financial econometrics, international finance and macroeconomic policy, etc.

E-mail: ypzhang@sfs.ecnu.edu.cn

□