中国科学 A 辑: 数学 2009 年 第 39 卷 第 3 期: 373～384
www.scichina.com math.scichina.com
《中国科学》杂志社
SCIENCE IN CHINA PRESS

基于经验似然的 Value-at-Risk 模型的评价方法

魏正红 [①②*], 温松桥 [①②], 朱力行 [②]

① 深圳大学数学与计算科学学院, 深圳 518060
② 香港浸会大学数学系, 香港
* E-mail: weizhenghong2006@yahoo.com.cn

收稿日期: 2007-10-29; 接受日期: 2008-11-20
香港研究局研究基金 (批准号: HKBU 2030/07P) 和广东省自然科学基金 (批准号: 2008276) 资助项目

摘要 Value-at-Risk (VaR) 是度量市场风险的一个基本工具. 自从 VaR 概念提出以来, 涌现出大量方法用于 VaR 估计, 因此在统计意义下, 如何检验这些方法的有效性, 以及如何比较不同 VaR 模型从而选择出最好的方法, 就成为人们非常关注的问题. 本文提出了利用经验似然法来评估和比较不同的 Value-at-Risk 模型. 模拟和实证分析表明经验似然方法比已有的方法有效和稳健.

关键词 Value-at-Risk 波动率 经验似然 设定检验 非嵌套检验

MSC(2000) 主题分类 62G10, 62P20, 91B30

1 引言

风险管理是金融机构、金融监管当局、非金融机构和资产管理公司非常关注的重要问题. Value-at-Risk (VaR) 是一种度量某一个金融资产的市场风险的方法. 它是指当市场正常波动时, 在一定的持有期和置信水平下, 某一金融资产所面临的最大可能的损失. 关于 VaR 的详细介绍, 可参见综述文献 [1] 以及专著 [2–4].

风险管理领域的发展非常迅速, 涌现出了大量的 VaR 的估计方法. 目前已有的方法大致包括六类. 最简单的方法是历史模拟法 (History Simulation), 该方法是利用基于历史数据的样本分位数去估计 VaR, 见文献 [5, 6]. 另一类普遍使用的方法是基于 GARCH 模型的参数估计方法, 见文献 [7, 8]. 极值理论是测量极端市场情况下风险损失的另一种方法, 它是基于最小或最大次序统计量的分布, 焦点在分布的尾部概率, 关于这种方法见文献 [9, 10]. 第四类方法是条件自回归 Value-at-Risk 模型, 简称为 CAViaR 模型, 该方法是由 Engle 等 [11] 提出的, 主要思想是将研究的问题由收益的分布转为直接对分位数的研究. 他们用自回归的方法确定分位数随时间的变化规律. 文中使用了 Koenker 等提出的分位数回归的方法来估计未知的参数. 为了扩展模型的使用范围, Fan 等 [13] 提出了时间相依的半参方法, 这个方法是 Chan 等 [14] 提出的用于动态期限结构的时间齐性半参模型的推广. 最后一类方法是 Chen

引用格式: 魏正红, 温松桥, 朱力行. 基于经验似然的 Value-at-Risk 模型的评价方法. 中国科学 A, 2009, 39(3): 373–384
Wei Z H, Wen S Q, Zhu L X. Empirical likelihood-based evaluations of Value at Risk models. Sci China Ser A, 2009, 52, DOI: 10.1007/s11425-009-0050-6

等 [15] 提出的基于特殊的密度函数的核密度估计法, 该方法可以解决一些相依问题. Chen 等 [16]利用平滑的经验似然构造了弱相依过程的 VaR 的置信区间. 由于该方法依赖于分块样本, 虽然可以处理相依数据, 但是窗宽的选择却是一个挑战性的问题.

对银行和监管当局来说, VaR 的估计方法无疑是非常重要的, 但是评估不同 VaR 模型的准确性同样是必须解决的问题. 首先, 给定一个 VaR 的估计方法, 风险管理者如何检验这个估计模型的准确性? 其次, 给定两个不同的 VaR 估计模型, 比如一个是 GARCH 模型, 另一个是隐含波动率模型, 那么在统计意义上如何比较这两个模型, 从而选择出一个好的模型? 目前基于假设检验的 VaR 模型的评估方法有三类, 每一个评估方法的零假设都是 VaR 模型准确地估计了实际的 VaR. 第一个检验方法是无条件覆盖检验, 第二个是条件覆盖检验, 分别由 Kupiec[17] 和 Christoffersen[18] 提出. 这两个检验方法都是基于例外 (实际损失超过 VaR) 的模型检验, 它们仅仅考虑了整个分布的一个百分点, 所以 Crnkovic 等 [19] 提出第三个检验方法, 称为分布预测检验. 他们认为应该以整体概率分布函数, 而不是仅仅以某一点的概率特性为基础来评估预测的质量. 在这些检验中, 如果零假设被拒绝, 则 VaR 的预测值将不能准确地预测实际的 VaR, 也就是说 VaR 的估计模型是 "不正确的"; 如果零假设不被拒绝, 则 VaR 的估计模型 "可以认为是正确的". 由于这些检验的功效往往很低, 也就是常常将不正确的模型误判为正确的模型, Lopez[20] 提出了一种不基于假设检验的 VaR 模型的评估方法, 该方法使用一种非统计的预测评价标准, 其基本思想是指定一个预测的损失函数, 使用该损失函数对 VaR 模型的预测结果打分, 根据 VaR 模型的得分评估其准确性, 得分越高, 模型的预测效果越差, 如果得分太高则拒绝该 VaR 模型.

在 Kullback-Leibler 信息准则下, Christoffersen 等 [21] 提出了设定检验和非嵌套检验, 并利用实际数据作了实证分析, 但却没有对这些方法进行必要的模拟. 一般来说, 为了全面地评估一种方法, 常规的模拟是不可缺少的. 本文中, 我们对 Christoffersen 的两种检验方法做了模拟, 发现这种渐进方法不是总能得到希望的结果, 因此, 我们在同样的条件下考虑另一种方法, 即经验似然法. 经验似然法由 Owen[22, 23] 首先提出并用于构造置信区间, Hall 等 [24] 总结了该方法与传统的统计方法相比所具有的一些优点: 如, 用经验似然法构造置信区间时, 置信区间的形状由数据自行决定, 域保持性, 变换不变性, 还有 Bartlett 纠偏性等. 由于这些原因, 经验似然方法被应用到各种统计模型和众多领域, 如均值的光滑函数的应用 [25], 非参数密度和回归参数的估计 [26−28], 分位数估计 [29] 等等. 关于经验似然方法及其应用的详细介绍, 请参阅 Owen 的综合性专著 [30]. 模拟和实证分析表明经验似然方法比已有的方法有效和稳健.

本文的结构如下. 在第 2 节, 我们提出了基于经验似然的对 VaR 模型进行检验和比较的新方法, 并详细介绍了这种方法是如何对 VaR 模型进行检验和比较的. 第 3 节中给出这种方法的模拟和实证结果. 最后在附录中我们给出主要定理的证明.

2 基于经验似然方法的 Value-at-Risk 模型的检验和比较

关于 VaR 模型的检验和比较, 包括 Christoffersen 在内的一些学者已经作了一些研究. 但是我们的数值研究表明这些方法不是总能给出希望的结果 (参见下一节的模拟结果). 在这一节里我们将提出另一个方法 — 经验似然方法. 我们将分别讨论设定检验和非嵌套检验.

2.1 基于经验似然方法的设定检验

设 S_t 是 t 时刻的资产价格, 若

$$r_t = \log S_t - \log S_{t-1} = \log \frac{S_t}{S_{t-1}}$$

是 t 时刻资产的对数收益, 根据 VaR 的定义, 它是指在一定的持有期 τ 和置信水平 $1-p$ 下, 某一金融资产所面临的最大可能的损失 (在本篇文章中, 我们取 $\tau=1$). 更准确地说, VaR_t 就是下列方程的解

$$P(r_t \leqslant \mathrm{VaR}_t|\Psi_{t-1}) = p, \tag{2.1}$$

其中, $\Psi_t = \sigma(S_s, s \leqslant t)$ 表示 t 时刻的历史信息.

实际中, 金融资产的收益往往具有被广泛介绍的三个特征: (i) 波动的群聚性, (ii) 高峰厚尾的分布, (iii) 收益带有偏态特性, 而且可能是时变的. 考虑到这些因素, 我们引入下面更一般的模型

$$r_t = \mu_t + \sigma_t \varepsilon_t, \tag{2.2}$$

这里 μ_t 和 σ_t 都是 Ψ_{t-1}-可测的, ε_t 是均值为 0, 方差为 1 的随机变量. 记给定 Ψ_{t-1} 的条件下, r_t 的条件分布函数为 $F_t(w) = P\left(r_t \leqslant w|\Psi_{t-1}\right), \forall\ w$. 则根据 (2.1) 式和上述的位置-刻度模型, 我们得到

$$\mathrm{VaR}_t(\beta) = F_t^{-1}(p) := \mu_t + \beta\sigma_t,$$

这里 β 是 ε_t 的 p^{th} 分位数, 即 $P(\varepsilon_t \leqslant \beta) = p$. 不失一般性, 我们可以假设 $\mu_t = 0$ (例如, 首先对原序列零均值化).

由于 (2.1) 可以重新写成下列形式

$$E(I\{r_t \leqslant \mathrm{VaR}_t(\beta)\} - p|\Psi_{t-1}) = 0, \tag{2.3}$$

如果 $\{z_{t-1}, z_{t-2}, \ldots\}$ 是 Ψ_{t-1}-可测, 这里 Ψ_{t-1} 是 $t-1$ 时刻的信息集, 那么由 (2.3) 可以导出

$$Ef(x_t, \beta) =: E\left\{\left(I\{r_t \leqslant \mathrm{VaR}_t(\beta)\} - p\right) k(z_{t-1}, z_{t-2}, \ldots)\right\} = 0, \tag{2.4}$$

这里 $x_t = (r_t, z_{t-1}, z_{t-2}, \ldots)$, $f(x_t, \beta) = (I\{r_t \leqslant \mathrm{VaR}_t(\beta)\} - p)k(z_{t-1}, z_{t-2}, \ldots)$. 随机变量序列 $\{z_{t-1}, z_{t-2}, \ldots\}$ 指的是在计量经济文献中常用的工具变量. 在后面的模拟分析中, 我们采用的是波动率的一阶滞后作为工具变量. 向量值函数 $k(\cdot)$ 是最简单的截距为 0 斜率为 1 的线性函数.

定义 $V_t = \gamma' f(x_t, \beta)$, 如果 VaR 的估计模型是正确的, 那么对任意的 $\gamma \in R^{d+1}$, 我们都有 $EV_t = 0$. 因此我们构造如下基于 V_t 的经验似然,

$$L(\beta, \gamma) = \max \prod_{i=1}^{T} p_i,$$

满足条件

$$\sum_{i=1}^{T} p_i = 1, \qquad \sum_{i=1}^{T} p_i V_i = 0.$$

根据标准的经验似然的推导过程可以证明 (见文献 [30])

$$p_t = \frac{1}{T(1+\lambda V_t)},$$

这里 λ 满足条件

$$\sum_{t=1}^{T}\frac{V_t}{1+\lambda V_t}=0.$$

那么, 对数经验似然比 (乘以 -2) 是

$$\tilde{\rho}_T^{(1)}:=-2\log\prod_{t=1}^{T}(Tp_t)=2\sum_{t=1}^{T}\log(1+\lambda V_t).$$

上述的经验似然比 $\tilde{\rho}_T^{(1)}$ 包含两个未知的参数 β 和 γ, 这两个参数可以用下面的方法估计 [31],

$$\begin{aligned}(\hat{\beta}_T,\hat{\gamma}_T)&=\arg\max_{\beta}\min_{\gamma}M_T(\beta,\gamma)\\&=\arg\max_{\beta}\min_{\gamma}\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f(x_t,\beta)\right).\end{aligned}\tag{2.5}$$

上述的估计是基于这样的直觉: 如果估计模型是正确的, VaR 估计应该最小化 Kullback Leibler Information Criterion (KLIC). 设

$$\hat{V}_t=\hat{\gamma}_T' f(x_t,\hat{\beta}_T),\qquad 和 \qquad \rho_T^{(1)}=2\sum_{t=1}^{T}\log(1+\lambda\hat{V}_t).$$

这样我们得到下面的定理, 定理的证明见附录.

定理 2.1 在 Christoffersen 等 [21] 的假设 1–9 下, 对数经验似然比 (乘以 -2)

$$\rho_T^{(1)}=2\sum_{t=1}^{T}\log(1+\lambda\hat{V}_t)\longrightarrow a\chi_1^2,\quad 依分布.$$

这里 $a=\sigma_\infty^2{}'/EV_1^2$, $\sigma_\infty^2{}'=\lim_{T\to\infty}\mathrm{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_t)$. 等价地,

$$a^{-1}\rho_T^{(1)}\longrightarrow\chi_1^2,\quad 依分布.$$

上述定理表明, 一个较大的 $a^{-1}\rho_T^{(1)}$ 值表明采用的模型可能是错误的. 在应用定理 2.1 进行设定检验之前, 我们首先需要估计未知的尺度参数 a. 注意到 $\sigma_\infty^2{}'=\lim_{T\to\infty}\sigma_T^2$, 这里

$$\begin{aligned}\sigma_T^2&=\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_t\right)\\&=\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{Cov}(V_s,V_t)\\&=\mathrm{Var}(V_1)+\frac{2}{T}\sum_{j=1}^{T-1}(T-j)\mathrm{Cov}(V_1,V_{1+j}),\end{aligned}\tag{2.6}$$

最后一个等式是基于序列 V_t 的平稳性. 在实际中我们可以用 (2.6) 式的前 r 项的相应的样本版本去近似的估计 σ_T^2, 为此我们定义

$$\hat{\sigma}_{r,T}^2=\frac{1}{T}\sum_{t=1}^{T}(\hat{V}_t-\bar{V})^2+\frac{2}{T}\sum_{j=1}^{r}\sum_{t=1}^{T-j}(\hat{V}_t-\bar{V})(\hat{V}_{t+j}-\bar{V}),$$

这里 $\bar{V}=\sum_{t=1}^{T}\hat{V}_t/T$. 但是为了保证我们有足够的数据去估计相应的协方差, r 不能取得太大. 在后面的模拟研究中, 我们发现上述估计量对 r 的选择是相当不敏感的. 当然我们也可以用别的方法去估计 $\sigma_\infty^2{}'$, 但是相应的估计的相合性需要作进一步的研究. 现在我们定义 $\hat{a}=\hat{\sigma}_{r,T}^2/(T^{-1}\sum_{t=1}^{T}\hat{V}_t^2)$, 那么如果 $|\hat{a}^{-1}\rho_T^{(1)}|>\chi_1^2(\alpha)$, 我们就拒绝零假设.

2.2 基于经验似然方法的非嵌套检验

现在我们将问题转向不同的 VaR 估计模型的比较上. 假定给出两个 VaR 估计模型 $\mathrm{VaR}_t(\beta)$ 和 $\mathrm{VaR}_t(\theta)$, 那么矩的条件可以重新写成: $f(x_t,\beta)=(I\{r_t\leqslant \mathrm{VaR}_t(\beta)\}-p)k(z_{t-1},z_{t-2},\ldots)$, $g(x_t,\theta)=(I\{r_t\leqslant \mathrm{VaR}_t(\theta)\}-p)k(z_{t-1},z_{t-2},\ldots)$. 注意到 VaR 模型评估的特殊性, 传统的非嵌套检验在这里就不能够使用. 这一节里, 我们将利用经验似然方法研究这个问题. 首先我们定义

$$M(\beta^*,\gamma^*)=\max_\beta\min_\gamma M(\beta,\gamma)=\max_\beta\min_\gamma E[\exp(\gamma' f(x_t,\beta))],$$
$$N(\theta^*,\lambda^*)=\max_\theta\min_\lambda N(\theta,\lambda)=\max_\theta\min_\lambda E[\exp(\lambda' g(x_t,\theta))].$$

那么在零假设成立的条件下, 两个估计模型应该具有同样的 KLIC 距离, 也就是 $M(\beta^*,\gamma^*)=N(\theta^*,\lambda^*)$. 相应的样本的 KLIC 距离定义为

$$M_T(\hat\beta_T,\hat\gamma_T)=\max_\beta\min_\gamma M_T(\beta,\gamma)=\max_\beta\min_\gamma \frac{1}{T}\sum_{t=1}^T \exp(\gamma' f(x_t,\beta)), \tag{2.7}$$

$$N_T(\hat\theta_T,\hat\lambda_T)=\max_\theta\min_\lambda N_T(\theta,\lambda)=\max_\theta\min_\lambda \frac{1}{T}\sum_{t=1}^T \exp(\lambda' g(x_t,\theta)). \tag{2.8}$$

关于 KLIC 距离的差, Christoffersen 等 [21] 证明了下列性质,

$$\sqrt{T}(M_T(\hat\beta_T,\hat\gamma_T)-N_T(\hat\theta_T,\hat\lambda_T))\to N(0,\sigma_\infty^2),\quad \text{依分布}. \tag{2.9}$$

这里

$$\sigma_\infty^2=\lim_{T\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T[\exp(\gamma^{*\prime} f(x_t,\beta^*))-\exp(\lambda^{*\prime} g(x_t,\theta^*))]\right). \tag{2.10}$$

下面我们构造经验似然比检验来比较两个 VaR 估计模型的差异. 设

$$W_t=\exp\{\gamma^{*\prime} f(x_t,\beta^*)\}-\exp\{\lambda^{*\prime} g(x_t,\theta^*)\},$$
$$\hat W_t=\exp\{\hat\gamma_T' f(x_t,\hat\beta_T)\}-\exp\{\hat\lambda_T' g(x_t,\hat\theta_T)\},$$

在零假设成立的条件下, 我们有 $EW_t=0$. 类似于前面构造设定检验的方法, 我们有下面的定理, 定理的证明见附录.

定理 2.2 在 Christoffersen 等 [21] 的假设条件 1–9 下. 如果零假设 $M(\beta^*,\gamma^*)=N(\theta^*,\lambda^*)$ 成立, 我们有

$$\rho_T^{(2)}=2\sum_{t=1}^T\log(1+\lambda\hat W_t)\longrightarrow b\chi_1^2,\quad \text{依分布}.$$

这里 $b=\sigma_\infty^2/E{W_1}^2$, λ 由下式确定:

$$\sum_{t=1}^T\frac{\hat W_t}{1+\lambda\hat W_t}=0.$$

等价地有

$$b^{-1}\rho_T^{(2)}\longrightarrow\chi_1^2,\quad \text{依分布}.$$

类似于定理 2.1, 当我们实际使用定理 2.2 时, 我们需要估计未知的刻度参数 b, 这可以用与设定检验里相类似的方法来处理, 具体的细节在这里就省略了.

3 模拟和实证分析

3.1 随机模拟

在这一节, 我们用随机模拟的方法来评估本文提出的经验似然比检验的功效. 模拟数据来源于 GARCH(1,1) 模型,

$$r_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = 0.1 r_{t-1}^2 + 0.85 \sigma_{t-1}^2,$$

这里 ε_t 是独立同分布的 $N(0,1)$ 随机变量. 所有报告的结果都是基于 $B=500$ 次重复试验, 即 500 次模拟. 每次试验产生 5 000 个来源于 GARCH(1,1) 模型的观察值. 为了消除初始值对结果的影响, 我们删掉了初始的 2 000 个观测值. 剩下的 3 000 个观测值中, 2 000 个用于估计模型参数, 1 000 个用于验证 VaR 模型. 我们利用四个波动率的一阶滞后作为工具变量. 在所有的模拟中, 显著性水平都是 10%.

四个波动率模型分别是: 历史模拟法 (History Simulation), RiskMetrics, GARCH$(1,1)$ 和 GJR$(1,1)$. 下面是这四个模型的简单介绍.

1. GARCH$(1,1)$ 模型 [8]:

$$r_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = c_0 + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2,$$

这里 ε_t 是均值为 0 方差为 1 的独立随机变量序列, $c_0 > 0$, $\alpha \geqslant 0$, $\beta \geqslant 0$, $\alpha + \beta < 1$.

2. RiskMetrics[32]:

$$r_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2,$$

对于日交易数据, 这里 λ 取为 0.94.

3. GJR$(1,1)$ 模型:

$$r_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = c_0 + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 + L r_{t-1}^2 I\{r_{t-1} < 0\}.$$

当 $L \neq 0$ 时, GJR 模型可以刻画金融资产收益的非对称性, 当 $L > 0$ 时, 我们就说存在杠杆效应.

4. 历史模拟法:

历史模拟法是最简单最直观的计算波动率的方法, 即用 t 时刻之前的若干观察值的标准差作为 t 时刻的标准差. 我们的模拟用 t 时刻之前的 500 个观察值的标准差作为 t 时刻的波动预测.

3.1.1 设定检验的比较

关于设定检验的模拟结果见表 1 , 表中的数据分别是用渐进方法和经验似然方法得到的检验的功效. 从表 1 中我们可以得到下列的一些结论. 由于样本数据来源于 GARCH$(1,1)$ 模型, 因此在下面的功效比较中我们将不考虑这个模型.

1. 从表 1 中我们可以看到, 经验似然方法比渐进方法更有效. 在某些情形下, 经验似然方法的功效比渐进方法大几倍, 如 GJR(1,1) 模型.

2. 在所有的波动率估计方法中, 历史模拟法是最粗糙的, 因此, 毫不意外的, 这两种方法在识别这种错误的波动估计模型时都显得非常有效. 当然, 相比之下, 经验似然方法还是略好于渐进方法.

3. 另一方面, 从表 1 中可以看出, 这两种方法识别 RiskMetrics 和 GJR(1,1) 模型的功效要小于识别历史模拟法的功效, 这是因为 RiskMetrics 和 GJR(1,1) 模型和 GARCH 属于同类模型, 而我们的数据来源于 GARCH, 因此要成功识别它们就显得更加困难.

3.1.2 非嵌套检验的比较

表 2 是关于非嵌套检验的模拟结果. 表中的数据分别是用渐进方法和经验似然方法检验时的拒绝率, 从表 2 中我们可以得到下面的一些结论.

1. 渐进方法在区别 GARCH(1,1) 模型和历史模拟法时有很高的拒绝率, 然而在进行其它比较时, 其拒绝率都是很低的, 这表明渐进方法不能够有效的区别 RiskMetrics, GJR(1,1) 和历史模拟法. 然而当用经验似然方法分别来比较 GARCH(1,1) 和 RiskMetrics, GARCH(1,1) 和 GJR(1,1), RiskMetrics 和 GJR(1,1) 时, 拒绝率都是非常低的, 这表明 GARCH(1,1), RiskMetrics 和 GJR(1,1) 这三种模型是一个等价类. 表 2 的结果进一步说明了这三个模型明显区别于历史模拟法. 因此用经验似然方法我们可以将这四个模型分成两类 {GARCH(1,1), Risk Metrics, GJR(1,1)}和历史模拟法.

表 1 渐进方法 (Asymp) 和经验似然方法 (E. L.) 在设定检验中的功效比较

		p=0.01	p=0.05	p=0.10	p=0.15	p=0.25
历史模拟法	Asymp	0.956	0.990	0.988	0.974	0.874
	E. L.	0.986	0.994	0.994	0.996	0.982
RiskMetrics	Asymp	0.588	0.636	0.636	0.554	0.314
	E. L.	0.828	0.888	0.878	0.836	0.708
GJR(1,1)	Asymp	0.246	0.112	0.114	0.086	0.078
	E. L.	0.544	0.426	0.426	0.416	0.418

2. 既然 RiskMetrics 和 GJR(1,1) 同属于 GARCH 类模型, 我们自然希望它们彼此的表现相似, 但不同于历史模拟法. 从这个观点来看, 上面的结论表明在对模型进行分类时经验似然法要比渐进方法优越.

3. Nelsen[33] 证明过一个结论, 即如果真实的数据来源于某一个扩散过程, 比如 GARCH (1,1), 则所有的 GARCH 类模型都可以给出相合的波动率估计. 从这个观点来看, 由结论 1 表明经验似然方法要比渐进方法更稳健, 因为前者不会因为模型的错误设定而受太大的影响.

4. 综上所述, 经验似然方法要比通常的渐进方法有效和稳健.

表 2 渐进方法 (Asymp) 和经验似然方法 (E. L.) 在非嵌套检验中的功效比较

VaR		p=0.01	p=0.05	p=0.10	p=0.15	p=0.25
RiskMetrics vs GJR(1,1)	Asymp	0.058	0.000	0.000	0.002	0.000
	E. L.	0.246	0.128	0.156	0.136	0.070
GARCH(1,1) vs RiskMetrics	Asymp	0.072	0.290	0.288	0.190	0.036
	E. L.	0.256	0.132	0.174	0.156	0.068
GARCH(1,1) vs GJR(1,1)	Asymp	0.008	0.000	0.000	0.000	0.000
	E. L.	0.038	0.006	0.004	0.002	0.002
历史模拟法 vs RiskMetrics	Asymp	0.076	0.004	0.008	0.002	0.002
	E. L.	0.536	0.572	0.596	0.560	0.374
历史模拟法 vs GJR(1,1)	Asymp	0.050	0.002	0.000	0.002	0.000
	E. L.	0.672	0.798	0.830	0.792	0.566
GARCH(1,1) vs 历史模拟法	Asymp	0.784	0.916	0.918	0.896	0.708
	E. L.	0.660	0.806	0.842	0.796	0.574

3.2 实证分析

在这一节中, 我们将利用中国股票市场的实际数据, 分别对渐进方法和似然方法的检验结果作比较. 我们选取的是上证指数, 时间从 1990 年 12 月 19 日到 2005 年 4 月 1 日, 数据来源于深圳证券交易所的交易数据库. 由于深圳主版市场自 2000 年以来停止发行新股, 所以深证指数基本不能反映大盘的走势, 因此我们没有考虑. 注意到, 日收益定义为

$$r_t = \ln(S_t) - \ln(S_{t-1}),$$

这里 S_t 是上证指数的 t 日收盘价. 计算结果列在表 3–6 中. 表中的数据是检验统计量的值. 我们同时考虑 5% 和 1% 的检验水平. 不同水平的临界值 (CV) 也在表中给出. 我们用 "$*$" 表示 "在 1% 水平下不拒绝", 用 "$**$" 表示 "在 1% 和 5% 水平下都不拒绝".

对于设定检验, 根据表 3 和表 4 我们得到下列结论:

1. 首先我们看 1% 显著水平下的结果. 当 $p = 0.01$ 时, 除了历史模拟法和 RiskMetrics 模型之外, 其它所有模型都没有被拒绝. 当 $p \neq 0.01$ 时, 两种检验方法都拒绝所有的模型. 这与 Christoffersen 等 (见文献 [21]) 所得到的结论是一致的. 即设定检验不容易拒绝极端 VaR 估计, 如 p=0.01, 易拒绝低分位数 VaR, 如 p=0.25, 0.15 等.

2. 在 5% 的检验水平下, 结果和 1% 的水平类似. 只有一点区别, 那就是 RiskMetrics 模型, 如果采用渐进方法则不被拒绝, 而用经验似然方法则被拒绝.

表 3　基于渐进方法的设定检验

VaR	p=0.01	p=0.05	p=0.10	p=0.15	p=0.25
历史模拟法	19.5506	28.4326	45.1047	59.3348	50.3273
RiskMetrics	9.6797*	22.7542	14.2153	21.4903	19.9675
GARCH(1,1)	3.8747**	28.9094	13.5687	17.3633	23.4247
GJR(1,1)	1.9248**	14.7838	11.6474	14.3736	20.2767

(当显著水平为 5% 时, CV = 7.81; 当显著水平为 1% 时, CV = 11.34)

表 4　基于经验似然方法的设定检验

VaR	p=0.01	p=0.05	p=0.10	p=0.15	p=0.25
历史模拟法	24.5751	25.9190	37.0130	48.6278	44.1488
RiskMetrics	10.6660	23.7472	12.0050	18.0905	18.5827
GARCH(1,1)	4.1666**	25.9527	11.9767	15.3586	21.4619
GJR(1,1)	1.6727**	14.3605	10.1250	12.8433	18.7422

(当显著水平为 5% 时, CV = 3.84; 当显著水平为 1% 时, CV = 6.63)

关于非嵌套检验的结果, 由表 5 和表 6 给出. 从这两个表中的数据, 我们可以得出下面一些结论: 当 $p = 0.05$ 时, 四个波动模型用两种检验方法检验, 在统计意义上都是不显著的; 当 $p = 0.10$, 0.15, 0.25 时, GARCH(1,1), RiskMetrics 和 GJR(1,1) 这三个波动模型在两种检验方法下都是不显著的, 而历史模拟法与这三个模型的区别是显著的; 当 $p = 0.01$ 时, 结论则比较复杂.

4　附录

本节给出定理 2.2 的证明, 定理 2.1 的证明可类似得出. 首先看两个引理.

引理 4.1 在定理 2.2 的条件下, 我们有

(a) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{W}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}W_t + o_p(1) \to N(0,\sigma_\infty^2)$.

(b) $\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t^2 \to E{W_1}^2$, 依概率.

(c) $Z_T := \max_{1\leqslant t\leqslant T}|\hat{W}_t| = o(T^{\frac{1}{2}})$ a.s.

(d) $\lambda = O_p(\frac{1}{\sqrt{T}})$.

证明 (a) 可以由文献 [21] 得出.

表 5 基于渐进方法的非嵌套检验

VaR	p=0.01	p=0.05	p=0.10	p=0.15	p=0.25
历史模拟法 vs RiskMetrics	–1.7266**	–0.6161**	–3.7731	–3.7134	–3.2499
历史模拟法 vs GARCH(1,1)	–4.1781	0.0428**	–3.9885	–4.6739	–2.6360
历史模拟法 vs GJR(1,1)	–5.9080	–1.7774**	–4.5301	–5.5301	–3.1765
RiskMetrics vs GARCH(1,1)	–1.5510**	0.5531**	–0.0824**	–0.4639**	0.3413**
RiskMetrics vs GJR(1,1)	–2.6058	–1.0394**	–0.3504**	–0.8836**	0.0329**
GARCH(1,1) vs GJR(1,1)	–0.6561**	–1.8393**	–0.2622**	–0.3716**	–0.3350**

(当显著水平为 5% 时, CV = 1.96; 当显著水平为 1% 时, CV = 2.58)

表 6 基于经验似然方法的非嵌套检验

VaR	p=0.01	p=0.05	p=0.10	p=0.15	p=0.25
历史模拟法 vs RiskMetrics	1.6567**	0.2486**	3.9777*	5.1641*	6.4872*
历史模拟法 vs GARCH(1,1)	4.6358*	0.0014**	5.1185*	6.8435	6.8506
历史模拟法 vs GJR(1,1)	5.6294*	1.2511**	5.1548*	7.4473	7.8196
RiskMetrics vs GARCH(1,1)	2.2793**	0.4039**	0.0121**	0.6602**	0.3472**
RiskMetrics vs GJR(1,1)	2.1566**	2.3947**	0.2439**	1.8763**	0.0027**
GARCH(1,1) vs GJR(1,1)	0.7857**	3.0060**	1.1641**	3.2636**	2.0415**

(当显著水平为 5%, CV = 3.84; 当显著水平为 1%, CV = 6.63)

(b) 根据文献 [21] 附录中的结论, 我们有 $\sqrt{T}(\hat{\beta}_T-\beta^*) = O_p(1)$, $\sqrt{T}(\hat{\gamma}_T-\gamma^*) = O_p(1)$, $\sqrt{T}(\hat{\theta}_T-\theta^*) = O_p(1)$, 和 $\sqrt{T}(\hat{\lambda}_T-\lambda^*) = O_p(1)$. 最后根据 Taylor 展开式和大数定律就可以得到 (b).

(c) 因为对任意的 $t\geqslant 1$, 有 $E\hat{W}_t^2<\infty$, 由著名的不等式 $\sum_{n=1}^{\infty}P(|\xi|\geqslant n)\leqslant E|\xi|\leqslant 1+\sum_{n=1}^{\infty}P(|\xi|\geqslant n)$, 有 $\sum_{T=1}^{\infty}P(\hat{W}_t^2\geqslant T)<\infty$. 根据 Borel-Cantelli 引理, $P(\limsup_T\{\hat{W}_t^2\geqslant T\}) = P(\hat{W}_t^2\geqslant T, \text{i.o.}) = 0$, 即, 仅有有限多的 T 使得, 对所有的 $1\leqslant t\leqslant T, \hat{W}_t^2\geqslant T$, a.s., 由此可以推得, 仅有有限多的 T 使得

$$\max_{1\leqslant t\leqslant T}\hat{W}_t^2\geqslant T \quad \text{a.s.} \qquad \Longleftrightarrow \qquad \frac{Z_T}{\sqrt{T}}\geqslant 1 \quad \text{a.s.},$$

所以, $\limsup_T Z_T/\sqrt{T}\leqslant 1$ a.s., 现在我们得到, 对任意的整数 $m\geqslant 1$ 及 $t\geqslant 1$, $E(m\hat{W}_t)^2<\infty$ 成立, 类似的推理可以证明,

$$\limsup_T\frac{Z_T}{\sqrt{T}}\leqslant\frac{1}{m} \text{ a.s.,} \quad 对所有的 m\geqslant 1$$
$$\Longleftrightarrow P\bigg(\bigcap_{m=1}^{\infty}\bigg\{\limsup_T\frac{Z_T}{\sqrt{T}}\leqslant\frac{1}{m}\bigg\}\bigg) = 1,$$

$$\Longleftrightarrow P\left(\limsup_T \frac{Z_T}{\sqrt{T}} = 0\right) = 1 \Longleftrightarrow Z_T = o(\sqrt{T}) \text{ a.s.}$$

(d) 注意到

$$0 = \frac{1}{T}\sum_{t=1}^{T} \frac{\hat{W}_t}{1+\lambda\hat{W}_t} = \frac{1}{T}\sum_{t=1}^{T}\hat{W}_t - \frac{\lambda}{T}\sum_{t=1}^{T}\frac{\hat{W}_t^2}{1+\lambda\hat{W}_t}.$$

因此, 我们有

$$\left|\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t\right| = \frac{|\lambda|}{T}\left|\sum_{t=1}^{T}\frac{\hat{W}_t^2}{1+\lambda\hat{W}_t}\right| \geqslant \frac{|\lambda|}{1+|\lambda|\max_t|\hat{W}_t|}\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t^2,$$

所以,

$$\left|\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t\right|\left(1+|\lambda|\max_t|\hat{W}_t|\right) \geqslant |\lambda|\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t^2.$$

那么我们可以导出

$$|\lambda| \leqslant \frac{|\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t|}{\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t^2 - |\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t|\max_t|\hat{W}_t|}.$$

根据引理的 (a)–(c) , 我们有

$$\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t = O_p\left(\frac{1}{\sqrt{T}}\right), \qquad \frac{1}{T}\sum_{t=1}^{T}\hat{W}_t^2 \to \sigma_\infty^2, \qquad \max_{1\leqslant t\leqslant T}|\hat{W}_t| = o_p(T^{\frac{1}{2}}).$$

因此, $\lambda = O_p(1/\sqrt{T})$.

引理 4.2 在定理 4.2 的条件下, 有 $\sum_{t=1}^{T}\lambda\hat{W}_t = \sum_{t=1}^{T}(\lambda\hat{W}_t)^2 + o_p(1)$.

证明 根据 Taylor 展式, 我们可以得到

$$\begin{aligned} 0 &= \lambda\sum_{t=1}^{T}\frac{\hat{W}_t}{1+\lambda\hat{W}_t} = \lambda\sum_{t=1}^{T}\hat{W}_t(1-\lambda\hat{W}_t + O_p((\lambda\hat{W}_t)^2)), \\ &= \sum_{t=1}^{T}(\lambda\hat{W}_t) - \sum_{t=1}^{T}(\lambda\hat{W}_t)^2 + \lambda^3 O_p\left(\sum_{t=1}^{T}\hat{W}_t^3\right). \end{aligned}$$

注意到 $\lambda^3|\sum_{t=1}^{T}\hat{W}_t^3| \leqslant \lambda^3\max_{1\leqslant t\leqslant T}|\hat{W}_t|\sum_{t=1}^{T}\hat{W}_t^2 = O_p(T^{-3/2})o_p(T^{\frac{1}{2}})O_p(T) = o_p(1)$. 因此有 $\sum_{t=1}^{T}\lambda\hat{W}_t = \sum_{t=1}^{T}(\lambda\hat{W}_t)^2 + o_p(1)$.

定理 2.2 的证明 由引理 4.2 , 我们有 $\sum_{t=1}^{T}\hat{W}_t = \lambda\sum_{t=1}^{T}\hat{W}_t^2 + o_p(\lambda^{-1})$. 所以,

$$\begin{aligned} \lambda &= \frac{\sum_{t=1}^{T}\hat{W}_t}{\sum_{t=1}^{T}\hat{W}_t^2} + o_p\left(\lambda^{-1}\left(\sum_{t=1}^{T}\hat{W}_t^2\right)^{-1}\right) = \frac{\sum_{t=1}^{T}\hat{W}_t}{\sum_{t=1}^{T}\hat{W}_t^2} + o_p(O_p(T^{\frac{1}{2}})O_p(T^{-1})) \\ &= \frac{\sum_{t=1}^{T}\hat{W}_t}{\sum_{t=1}^{T}\hat{W}_t^2} + o_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

因此, 我们有

$$\begin{aligned} \rho_T^{(2)} &= 2\sum_{t=1}^{T}\log(1+\lambda\hat{W}_t) \\ &= 2\sum_{t=1}^{T}\lambda\hat{W}_t - \sum_{t=1}^{T}(\lambda\hat{W}_t)^2 + 2\sum_{t=1}^{T}o_p((\lambda\hat{W}_t)^2) = \lambda\sum_{t=1}^{T}\hat{W}_t + o_p(1) \\ &= \frac{(T^{-1/2}\sum_{t=1}^{T}\hat{W}_t)^2}{T^{-1}\sum_{t=1}^{T}\hat{W}_t^2} + o_p(1) \longrightarrow \left(\frac{\sigma_\infty^2}{E{W_t}^2}\right)\chi_1^2, \quad \text{依分布}. \end{aligned}$$

最后的渐进结果是根据引理 4.1 的 (a) 和 (b) 得到的.

致谢 非常感谢编委和审稿人对本文提出的宝贵意见和建议.

参考文献

1 Duffie D, Pan J. An overview of value at risk. *J Derivatives*, **5**: 7–49 (1997)

2 Alexander C. Risk Management and Analysis, Vol. I & II. Chichester: Wiley, 1998

3 Dowd K. Beyond Value at Risk: The New Science of Risk Management. New York: Wiley, 1998

4 Jorion P. Value at Risk: The New Benchmark for Managing Financial Risk, 2nd ed. New York: McGraw-Hill, 2000

5 Hendricks D. Evaluation of value-at-risk models using historical data. *Federal Reserve Bank of New York Economic Policy Review*, **2**: 39–69 (1996)

6 Mahoney J M. Forecast biases in value-at-risk estimations: evidence from foreign exchange and global equity portfolios. Working paper, Federal Reserve Bank of New York, 1996

7 Engle R F. ARCH, Selected Readings. Oxford: Oxford University Press, 1995

8 Bollerslev T. Generalized autoregressive conditional heteroscedasticity. *J Econometrics*, **31**: 307–327 (1986)

9 Embrechts P, Kluppelberg C, Mikosch T. Modeling Extremal Events for Insurance and Finance. Berlin: Springer, 1997

10 Christoffersen P F. Elements of Financial Risk Management. Amsterdam: Academic Press, 2003

11 Engle R F, Manganelli S. CAViaR: conditional autoregressive value at risk by regression quantiles. *J Bus Econom Statist*, **22**: 367–381 (2004)

12 Koenker R, Bassett G. Regression quantiles. *Econometrica*, **46**: 33–50 (1978)

13 Fan J Q, Gu J. Semiparametric estimation of Value-at-Risk. *Econometrics J*, **6**: 261–290 (2003)

14 Chan K C, Karolyi A G, Longstaff F A, et al. An empirical comparison of alternative models of the short-term interest rate. *J Finance*, **47**: 1209–1227 (1992)

15 Chen S X, Tang C Y. Nonparametrical inference of Value-at-Risk for dependent financial returns. *J Financial Econometrics*, **3**: 227–255 (2005)

16 Chen S X, Wong C M. Smoothed block empirical likelihood for quantiles of weakly dependent process. *Statist Sinica*, to appear (2009)

17 Kupiec P H. Techniques for verifying the accuracy of risk measurement models. *J Derivatives*, **3**: 73–84 (1995)

18 Christoffersen P F. Evaluating interval forecasts. *Internat Econom Rev*, **39**: 841–862 (1998)

19 Crnkovic C, Drachman J. Quality control. *Risk*, **9**: 139–143 (1996)

20 Lopez J. Regulatory evaluation of value-at-risk models. *J Risk*, **23**: 470–472 (1997)

21 Christoffersen P, Hahn J, Inoue A. Testing and comparing value-at-risk measures. *J Empirical Finance*, **8**: 325–342 (2001)

22 Owen A B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**: 237–249 (1988)

23 Owen A B. Empirical likelihood ratio confidence regions. *Ann Statist*, **18**: 90–120 (1990)

24 Hall P, LaScala B. Methodology and algorithms of empirical likelihood. *Internat Statist Rev*, **58**: 109–127 (1990)

25 DiCiccio T S, Hall P, Romono J. Empirical likelihood is Bartlett correctable. *Ann Statist*, **19**: 1053–1061 (1991)

26 Owen A B. Empirical likelihood for linear models. *Ann Statist*, **19**: 1725–1747 (1991)

27 Chen S X. Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, **83**: 329–341 (1996)

28 Chen S X, Qin Y S. Empirical likelihood confidence intervals for local linear smoothers. *Biometrika*, **87**: 946–953 (2000)

29 Chen S X, Hall P. Smoothed empirical likelihood confidence intervals for quantiles. *Ann Statist*, **21**: 1166–1181 (1993)

30 Owen A B. Empirical Likelihood. London: Chapman and Hall, 2001

31 Kitamura Y, Stutzer M. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, **65**: 861–874 (1997)

32 Morgan J P. Riskmetrics-Technical Document, 4th ed. New York: Morgan Guaranty Trust Company, 1996

33 Nelson D. Filtering and forecasting with misspecified GARCH models: getting the right variance with the wrong models. *J Econometrics*, **52**: 61–90 (1992)