

# 商业银行 VaR 模型预测能力的验证<sup>1</sup>

刘晓曙 郑振龙

(厦门大学金融系 福建厦门 361005)

内容摘要：本文运用 Hong & Li [7]于 2004 年提出的用非参数方法来检验时间序列动态模型设定正确性的方法，来研究商业银行使用的各种 VaR 模型的正确性。本文通过实证研究发现，当前商业银行模型风险管理工具回顾测试中常用的巴塞尔规则、Kupiec 统计量、Christofferson 统计量方法可能具有一定的误导性。这种误导可能使得商业银行的风险预测能力受到影响，从而影响盈利能力，甚至危及整个金融系统安全性。因此，商业银行在利用 VaR 模型时需要仔细的选择合适的方法。

关键词：VaR 回顾测试 模型验证

JEL:G21

市场风险的量化工具，如 VaR，只有对风险测量的计算方法及其风险衡量与实际业绩的关系都进行了充分的考虑时，才具有指导意义。

艾伦·格林斯潘 (Alan Greenspan)

## 一、引言

只有能准确预测风险的 VaR 模型才是有效的。在实践中，模型的运用过程是一个不断检验、证明与发展的过程。模型验证 (model validation) 是指检验一个模型是否正确，它可以通过一系列的工具体，如回测等来完成。

回测是用来检测实际损失与预期损失是否一致的有效的统计方法，它将 VaR 的历史预测与相关的资产 (组合) 收益率进行系统的比较。由于金融市场的复杂性以及监管的压力，商业银行在实施 VaR 管理时必须成本收益取舍中作出选择。VaR 模型常常需要对金融资产模型作出一些假设，比如资产收益的分布、马尔可夫性质等等。因而对于需要检查 VaR 预测能力的 VaR 使用者和风险管理者来说，回测是至关重要的。

对于商业银行系统来说，从决策的角度来看，VaR 风险预测的不精确性可能导致两个重要方面的影响：

- 第一、如果风险预测不准确，商业银行的业务决策包括目标设定、风险定价、风险资产组合构造以及套期策略等将也变得不准确，从而可能影响商业银行的盈利；
- 第二、如果风险模型预测能力低，又假如巴塞尔协议的检验规则具有某种误导性，那么银行系统可能蕴含系统性风险。我们知道，巴塞尔协议或本国银监局关心的不是单个商业银行的盈利能力而是金融系统良好、稳健的运行。

本文将从实证的角度讨论当前回顾测试中常用的基于失效率 (failure rate) 的巴塞尔规则、Kupiec 统计量、Christofferson 统计量方法的不足。文章接下来的部分包括：在第二部分中，我们简单回顾一下基于失效率的巴塞尔规则、Kupiec 统计量、Christofferson 统计量方法，在第三部分中我们给出一个实证研究表明传统方法可能带有一定的误导性。

<sup>1</sup> 本文为教育部人文社科基地重大项目—金融制度设计与经济增长 (项目号 05JJD790026) 的阶段性研究成果。郑振龙为通讯作者。

## 二、回顾测试方法回顾

### 1. 巴塞尔规则

验证模型准确性的最简便的方法是记录失效率。失效率是在给定样本中 VaR 被超越的次数比率。假设银行  $T$  天的 VaR 值中有  $N$  个 VaR 例外数，则  $N/T$  就是失效率。失效率方法的最简单应用就是巴塞尔规则[1]。内部模型回测的巴塞尔规则直接产生于失效率试验。设计这种试验时，首先必须选择第一类错误率，即模型正确却被拒绝的概率。当这类错误发生时，银行只能怪自己运气不好，因为它们不应当受到这个错误引起的不适当的惩罚。因此，一般说来，应该选择一个较低的第一类错误率。巴塞尔规则认为在 250 个观测值中 99% 的置信水平（也就是 1% 的第一类错误率）下 4 个以内的损失例外数是可以接受的，也就是银行的“绿灯”区域。如果例外的个数大于或等于 5，银行便会进入“黄灯”或“红灯”区域，并因此遭受乘数因子从 3 到 4 的累进惩罚。如表一所示[1]。

表一 巴塞尔回顾测试规则（250 个观测值，第一类错误率为 1%）

区域	例外个数	惩罚乘数
绿灯	0~4	3.00
黄灯	5	3.40
	6	3.50
	7	3.65
	8	3.75
	9	3.85
红灯	10 个以上	4.00

### 2. Kupiec 统计检验量方法

Kupiec 统计检验量方法[2]是目前金融机构中应用非常广泛的一种回顾测试检验方法。令  $N$  表示回顾测试中  $T$  个观测值样本的 VaR 例外个数。如果 VaR 模型是正确的，那么，例外个数  $N$  服从于二项分布：

$$f(N) = \binom{T}{N} p^N (1-p)^{T-N} \quad (1)$$

这里， $p = 1 - \text{VaR 置信水平}$ ，简称 VaR 水平。

Kupiec 在零假设即 VaR 模型是正确的情况下，构造了一个统计量，（以下简称 PF 统计量）

$$LR_{PF} = -2 \ln \left[ (1-p)^{T-N} p^N \right] + 2 \ln \left[ \left( 1 - \frac{N}{T} \right)^{T-N} \left( \frac{N}{T} \right)^N \right] \quad (2)$$

在零假设下，PF 统计量渐近服从自由度为 1 的  $\chi^2$  分布。表二给出了不同 VaR 水平和 PF 统计量不同显著水平下拒绝零假设的 VaR 例外数。

表二 250 个观察样本下 PF 统计量的零假设的拒绝区间

VaR 水平	PF 统计量显著水平	拒绝零假设 VaR 例外数区间
1%	1%	$N < 8$
5%	1%	$4 < N < 23$
1%	5%	$N < 7$
5%	5%	$6 < N < 19$

### 3. Christoffersen 统计检验量

我们注意到，Kupiec 检验仅仅关注 VaR 例外个数，而没有考虑 VaR 例外发生的时间变

化。而如果 VaR 模型是正确的话，那么，VaR 例外的发生应该是独立的、同分布的。因此，假若 VaR 例外表现出某些团簇（clustering）现象，VaR 模型则可能在一定条件很难捕捉到资产变化的波动特性。

Christoffersen 统计检验量[3]的主要贡献在于它对 PF 进行了扩充，以保证 VaR 例外在每一时刻都是独立的，也就是联合检验失败率和 VaR 例外的独立性。

Christoffersen 构造了零假设为 VaR 模型是正确条件下的统计量：

$$LR_c = -2 \ln \{p^N (1-p)^{T-N}\} + 2 \ln \{(1-p_{01})^{n_{0,0}} p_{01}^{n_{0,1}} (1-p_{11})^{n_{1,0}} p_{11}^{n_{1,1}}\} \quad (3)$$

其中  $p_{ij} = p(I_t = j | I_{t-1} = i)$ ，表示两天之间 VaR 例外的跃迁几率，这里哑变量  $I_t$  表示若

没有 VaR 例外发生  $I_t = 0$ ，否则  $I_t = 1$ 。 $n_{i,j}$  表示回顾测试样本期间内发生状态  $i$  的后一天

发生状态  $j$  的天数。统计上  $p_{ij}$  可以用  $\hat{p}_{ij} = \frac{n_{i,j}}{(n_{i,j} + n_{i,j-1})}$  来估计。

在零假设下，Christoffersen 统计量渐近服从自由度为 2 的  $\chi^2$  分布。

### 三、实证研究结果与分析

本部分中，我们将首先给出一个通过巴塞尔规则、Kupiec 统计量、Christofferson 统计量检验的实证结果，然而应用 Hong & Li 统计量对实证数据进行检验，发现巴塞尔规则、Kupiec 统计量、Christofferson 统计量方法在回顾测试中可能具有一定的误导性。

#### （一）实证分析

我们采用上海综合指数，时间自 2000 年 1 月 4 日至 2006 年 12 月 29 日。收益率  $r_t$  定义为：

$$r_t = 100 * (\log p_t - \log p_{t-1}) \quad (4)$$

这里  $p_t$  表示第  $t$  天指数值。

我们采用 GARCH、EGARCH 与 APARCH 三类模型：

#### 1. GARCH(1,1)类模型

$$\begin{aligned} r_t &= \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} z_t \\ h_t &= \omega + \alpha h_{t-1} + \beta \varepsilon_{t-1}^2 \end{aligned} \quad (5)$$

这里  $z_t$  是随机扰动项，按照  $z_t$  是否服从 i.i.d. N(0,1) 分布、i.i.d. student t 分布和 i.i.d. GED 分布模型可以分为 GARCH(1,1)、t-GARCH(1,1) 与 GED-GARCH(1,1) 三个模型。其中 student t 分布的密度函数是：

$$t(x, v) = \frac{\Gamma(\frac{v+1}{2})(1+x^2/2)^{-\frac{v+1}{2}}}{\Gamma(\frac{v}{2})\sqrt{v\pi}} \quad (6)$$

GED 分布的密度函数是：

$$f(x, v) = \frac{v \exp\left(-\left|\frac{x}{\lambda}\right|^{\frac{v}{2}}\right)}{\lambda 2^{1+v^{-1}} \Gamma(v^{-1})} \quad (7)$$

这里  $\lambda = \left[ \frac{2^{-2/v} \Gamma(v^{-1})}{\Gamma(3/v)} \right]^{1/2}$ ,  $\Gamma(\cdot)$  是 Gamma 函数。

### 2. EGARCH(1,1)类模型

$$\begin{aligned} r_t &= \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} z_t \\ \log h_t &= \omega + \alpha \log h_{t-1} + \beta \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \end{aligned} \quad (8)$$

这里  $z_t$  是随机扰动项, 按照  $z_t$  是否服从 i.i.d. N(0,1)分布、i.i.d. student t 分布和 i.i.d. GED 分布模型可以分为 EGARCH(1,1)、t-EGARCH(1,1)与 GED-EGARCH(1,1)三个模型。

### 3. APARCH(1,1)类模型

$$\begin{aligned} r_t &= \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} z_t \\ h_t^{\frac{\delta}{2}} &= \omega + \alpha h_{t-1}^{\frac{\delta}{2}} + \beta (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^{\delta} \end{aligned} \quad (9)$$

这里  $z_t$  是随机扰动项, 按照  $z_t$  是否服从 i.i.d. N(0,1)分布、i.i.d. student t 分布和 i.i.d. GED 分布模型可以分为 APARCH(1,1)、t-APARCH(1,1)与 GED-APARCH(1,1)三个模型。我们用前 1432 个数据估计模型的参数, 用后 250 个数据作 VaR 模型回顾测试。模型参数估计结果见表三。

表三 九个模型参数估计结果

模型	$\omega$	$\alpha$	$\beta$	$\gamma$	$v$	$\delta$
GARCH(1,1)	0.066272 (0.010867) 0.0000	0.850277 (0.012108) 0.0000	0.119597 (0.011014) 0.0000	--	--	--
GED-GARCH(1,1)	0.064803 (0.020968) 0.0020	0.867245 (0.011014) 0.0000	0.097678 (0.019607) 0.0000	--	1.192679 (0.051826) 0.0000	--
t-GARCH(1,1)	0.063811 (0.021602) 0.0031	0.877793 (0.024156) 0.024028	0.089972 (0.019519) 0.0000	--	4.793459 (0.640269) 0.0000	--
EGARCH(1,1)	-0.105040 (0.012485) 0.0000	0.977257 (0.003296) 0.0000	0.150642 (0.017060) 0.0000	-0.070373 (0.007598) 0.0000	--	--
GED-EGARCH(1,1)	-0.106129 (0.021899) 0.0000	0.974212 (0.007747) 0.0000	0.152597 (0.030515) 0.0000	-0.073470 (0.015417) 0.0000	1.236543 (0.052045) 0.0000	--
t-EGARCH(1,1)	-0.106988 (0.021295) 0.0000	0.973068 (0.008987) 0.0000	0.155838 (0.030550) 0.0000	-0.078741 (0.016929) 0.0000	5.361577 (0.726270) 0.0000	--
APARCH(1,1)	0.024184 (0.003418) 0.0000	0.928719 (0.006750) 0.0000	0.065216 (0.008399) 0.0000	0.664065 (0.079364) 0.0000	--	0.412488 (0.119020) 0.0000

GED- APARCH(1,1)	0.028966 (0.008476) 0.0006	0.920202 (0.014679) 0.0000	0.074623 (0.015325) 0.0000	0.606700 (0.139743) 0.0000	1.248355 (0.054061) 0.0000	0.634974 (0.269756) 0.0186
t-APARCH(1,1)	0.030907 (0.009762) 0.0015	0.916778 (0.015822) 0.0000	0.078452 (0.015227) 0.0000	0.603976 (0.131538) 0.0000	5.484637 (0.763656) 0.0000	0.7193697 (0.276174) 0.0092

注：每个参数估计栏中第一行为参数估计值，圆括号中是参数估计的标准差，最后一行表示参数估计  $t$  统计量  $p$  值。

在 VaR 框架下，给定日收益率  $r_t$  与条件波动率  $\sqrt{h_t}$ ，则  $t+1$  时刻的 VaR 值是

$$VaR_{t+1}(p) = Z_p \sqrt{h_{t+1}} \quad (10)$$

这里  $\sqrt{h_{t+1}}$  是前向一天预测的条件标准方差，由 (5)、(8) 与 (9) 得到，而  $p$  是 VaR 水平， $Z_p$  是假设分布的  $p$  分位数，对于 GED 分布来说，若  $p < 0.5$ ，则不难得到

$$Z_p = \frac{2\lambda}{\nu} \left[ \log(p) + \frac{1}{\nu} \log 2 + \log\left(\Gamma\left(\frac{1}{\nu}\right)\right) \right].$$

通过计算，我们得到在 PF 统计量、Christoffersen 统计量在各显著水平下的  $p$ -值及 VaR 例外数，见表四。

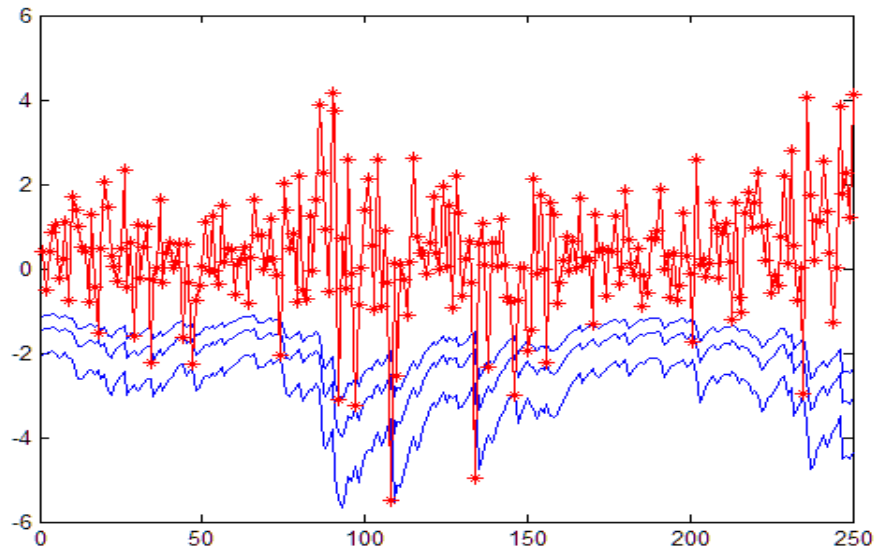
表四 回顾测试中 PF 统计量、Christoffersen 统计量的  $p$ -值及 VaR 例外数

模型	VaR 水平	VaR 例外数 $N$	PF 统计量 $p$ 值	Christoffersen 统计量 $p$ 值
GARCH(1,1)	1%	3	0.7580	0.9083
	5%	9	0.2860	0.3893
	10%	17	0.0751	0.0549
t-GARCH(1,1)	1%	2	0.7419	0.9245
	5%	6	0.0366*	0.0947
	10%	11	0.0010*	0.0026*
GED-GARCH(1,1)	1%	0	0.0000*	0.0000*
	5%	2	1.7086e-004*	8.3520e-004*
	10%	8	3.6943e-005*	1.4933e-004*
EGARCH(1,1)	1%	5**	0.1619	0.3325
	5%	13	0.8853	0.4582
	10%	19	0.1884	0.0808
t-EGARCH(1,1)	1%	2	0.7419	0.9245
	5%	10	0.4529	0.4766
	10%	16	0.0435*	0.0406*
GED-EGARCH(1,1)	1%	1	0.2781	0.5508
	5%	4	0.0042*	0.0154*
	10%	13	0.0058*	0.0102*
APARCH(1,1)	1%	8**	0.0054*	0.0155*
	5%	18	0.1331	0.0738
	10%	19	0.1884	0.0808
t-APARCH(1,1)	1%	2	0.7419	0.9245
	5%	10	0.4529	0.4766
	10%	19	0.1884	0.0808
GED-APARCH(1,1)	1%	1	0.2781	0.5508
	5%	6	0.0366*	0.0947
	10%	14	0.0121*	0.0176*

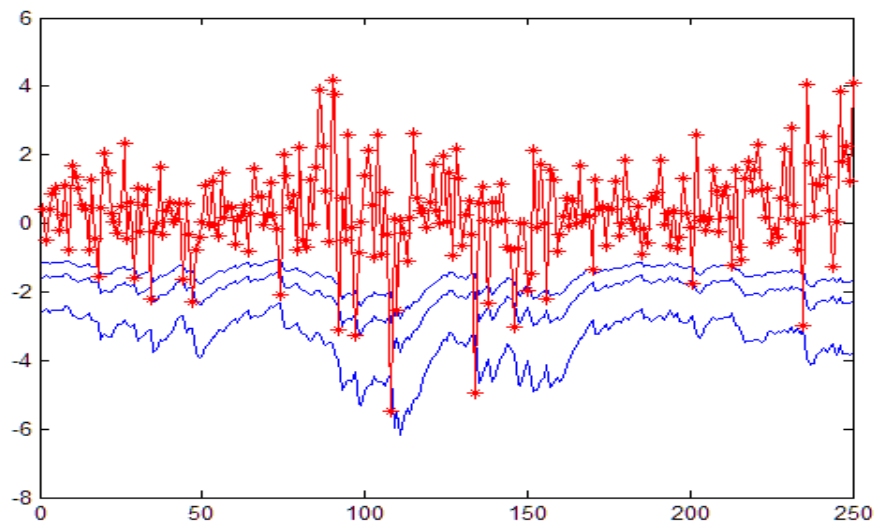
注：\*表示在通常使用的 5% 显著水平下，模型被拒绝；\*\*表明模型处在巴塞尔规则的“绿灯”区外。

与表一相比较，我们发现除 APARCH(1,1)与 EGARCH(1,1)外所有模型回顾测试结果都落在巴塞尔规则的“绿灯”区；但是在“绿灯”区的模型中只有 GARCH(1,1)与 t-APARCH(1,1)模型在 VaR 水平为 1%、5%与 10%同时在通常使用的 5%显著水平下模型被接受。

在接下来的分析中我们将集中于 GARCH(1,1)与 t-APARCH(1,1)这两个模型。从图一和图二我们可以很直观地观察这两个模型的回顾测试结果。



图一：GARCH(1,1)模型在 VaR 水平分别为 1%、5%与 10%时的回顾测试结果。VaR 例外数  $N = 3、9、17$ 。



图二：t-APARCH(1,1)模型在 VaR 水平分别为 1%、5%与 10%时的回顾测试结果。VaR 例外数  $N = 2、10、19$ 。

## (二) Hong & Li 统计量分析

我们这里进一步应用最近发展起来的模型验证的工具来检验上面我们提到的 GARCH(1,1)与 t-APARCH(1,1)模型。

Rosenblatt[4]证明了如果模型是设定正确的,那么广义残差序列  $\{Z_t\}$  服从  $i.i.d. U[0,1]$ 。

这里,  $Z_t \equiv \int_{-\infty}^r p(r | \tilde{I}_{t-1}, t) dy$ .  $\tilde{I}_{t-1}$  表示直到  $(t-1)$  时刻的信息集. Diebold[5]和 Elerian、Chib & Shephard [6]将之广泛应用于计量经济学中.  $p(r | \tilde{I}_{t-1}, t)$  表示  $t$  时刻可观测变量  $r$  在  $t$  时刻的条件密度函数。

Hong & Li [7]于 2004 年提出了用非参数方法来检验时间序列动态模型设定的正确性,也就是用非参数的方法来检验广义残差序列  $\{Z_t\}$  是否服从 *i.i.d.*  $U[0,1]$ 。这一方法适用于各种动态概率分布模型。由于使用非参数方法,这种检验方法对各种各样的模型设定正确与否有很强的检验能力。而且,检验统计量的渐近行为与参数估计的数目无关,因此,如果一个模型包含了一些没有解释能力的变量就不可能会被当作最好的模型。Hong & Li 检验统计量因而可以看作是衡量模型正确性的一个标准。此外, Hong & Li 统计量显示若检验统计量越小,表明模型就越接近正确性设定。

具体的讲, Hong 等使用估计核

$$\hat{g}_j(z_1, z_2) \equiv (n-j)^{-1} \sum_{t=j+1}^n K_h(z_1, Z_t) K_h(z_2, Z_{t-j}) \quad (11)$$

对  $\{Z_t, Z_{t-j}\}$  的联合密度进行估计。这里,  $K_h(x, y)$  是边界修正核函数:

$$K_h(x, y) \equiv \begin{cases} h^{-1} k\left(\frac{x-y}{h}\right) / \int_{-(x/h)}^1 k(u) du & \text{如果 } x \in [0, h) \\ h^{-1} k\left(\frac{x-y}{h}\right) & \text{如果 } x \in [h, 1-h] \\ h^{-1} k\left(\frac{x-y}{h}\right) / \int_{-1}^{(1-x)/h} k(u) du & \text{如果 } x \in (1-h, h] \end{cases} \quad (12)$$

其中  $h = \hat{\sigma}_z n^{-1/6}$ ,  $\hat{\sigma}_z$  是广义残差  $\{Z_t\}_{t=1}^n$  的标准差,  $k(\cdot)$  是预设的对称密度函数。

Hong & Li 统计量  $\hat{Q}(j)$  构造如下:

$$\hat{Q}(j) \equiv \left\{ (n-j)h \int_0^1 \int_0^1 [\hat{g}_j(z_1, z_2) - 1]^2 dz_1 dz_2 - hA_h^0 \right\} / V_0^{1/2} \quad (13)$$

其中

$$A_h^0 \equiv \left[ (h^{-1} - 2) \int_{-1}^1 k^2(u) du + 2 \int_0^1 \int_{-1}^b k_b^2(u) du db \right]^2 - 1$$

$$V_0 \equiv 2 \left[ \int_{-1}^1 \left[ \int_{-1}^1 k(u+v) k(v) dv \right]^2 du \right]^2$$

这里  $k_b(\cdot) \equiv k(\cdot) / \int_{-1}^b k(v) dv$

如果模型设定正确, 则当样本数  $n \rightarrow \infty$  时, Hong & Li 统计量  $\hat{Q}(j) \rightarrow N(0,1)$ , 否则,

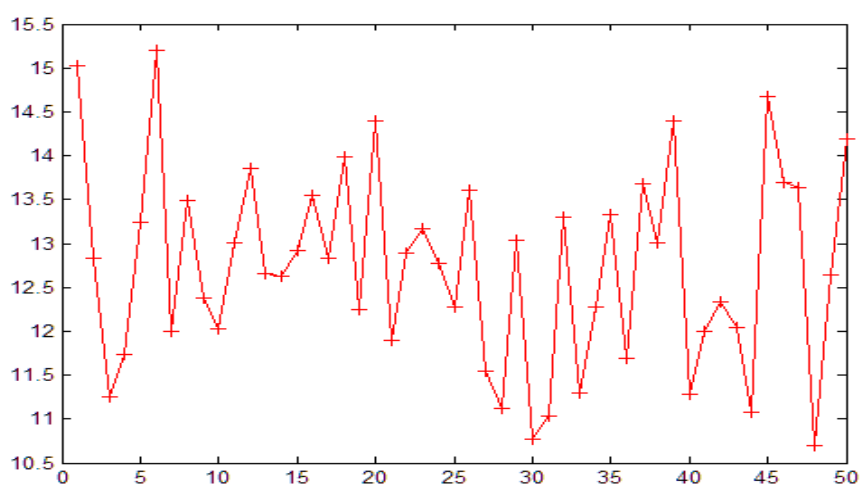
$\hat{Q}(j) \rightarrow \infty$ 。这里  $j$  表示计算统计量时所用到的滞后的阶数。Hong 证明了当  $n \geq 500$  时,

该统计量具有相当的验证模型正确性的效力。当我们使用全部观测值，计算得到的  $Q(j)$  如表五：

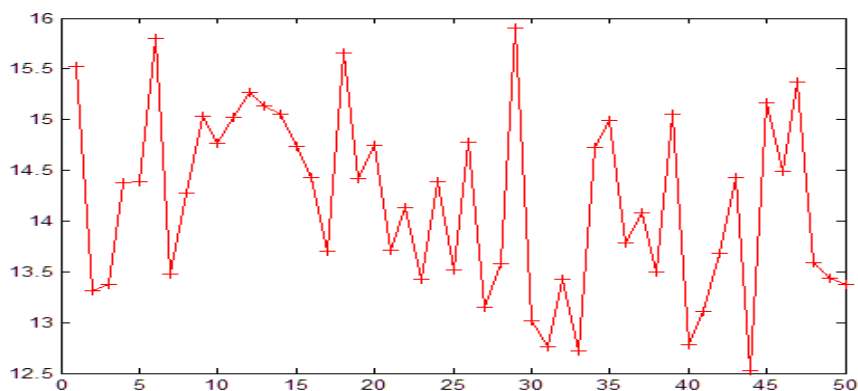
表五 Hong & Li 统计量  $\hat{Q}(j)$  的相应各阶滞后项的统计值

	$j=1$	$j=5$	$j=10$	$j=15$
GARCH(1,1)	<b>15.0304</b>	<b>13.2340</b>	<b>12.0320</b>	<b>12.9196</b>
t-APARCH(1,1)	<b>15.5231</b>	<b>14.3917</b>	<b>14.7672</b>	<b>14.7376</b>

更直观的，我们将计算得到各阶  $\hat{Q}(j)$  表示如图三与图四。这里  $j=1, 2, \dots, 50$ .



图三 GARCH(1,1)模型的 Hong & Li 统计量  $\hat{Q}(j)$  直观图。横轴表示滞后阶数，纵轴表示统计量的值。



图四 t-APARCH(1,1)模型的 Hong & Li 统计量  $\hat{Q}(j)$  直观图。横轴表示滞后阶数，纵轴表示统计量的值。

我们从表五与图三、四看到 GARCH (1,1) 与 t-APARCH(1,1)模型都被显著拒绝。因此，



我们可以得出结论说，巴塞尔规则、Kupiec 统计量、Christofferson 统计量方法在回顾测试中可能具有一定的误导性。

#### 四、结论

我们通过实证研究显示商业银行或银监局在管理模型风险时，常用的回顾测试方法（如巴塞尔规则、Kupiec 统计量、Christofferson 统计量方法）可能带有一定的误导性。这种误导可能使得商业银行的风险预测能力受到影响，从而影响盈利能力，甚至危及整个金融系统安全性。

仔细分析，我们可以看到这些常用的回顾测试是基于点预测或区间预测（即条件分位数预测）。这是导致它们可能具有误导性的原因。可能的选择方法是选用基于整个密度函数的预测与模型验证。但是，基于整个密度函数的预测与模型验证在新兴的金融市场中会遇到因为政策性等原因导致市场结构变化而引起合理的回顾检验样本数不够的问题（比如，Hong & Li  $Q(j)$  需要至少 500 个数据）。此外，由于一切计量经济学模型都是“错误”模型，如何在“错误”模型与设定正确的苛刻要求之间作出选择，商业银行需要在模型假设的复杂性与成本间实现权衡，因此，未来商业银行的模型风险管理还需在前行中寻找新的、合适的方法。

#### 参考文献：

- [1] 菲利普·乔瑞，风险价值-金融风险管理新标准[M].中信出版社,2000.
- [2] Kupiec, P., Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*[J] 3. 73-84
- [3] Christoffersen, P. (1998) Evaluating Interval Forecasts, *International Economic Review*[J], 39, 841-862.
- [4] Rosenblatt M. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*[J], 1952, 23,470-472.
- [5] Diebold, F X, T A Gunther and A S Tay, Evaluating density forecasts with applications to financial risk management. *International Economic Review*[J], 1998, 39,863-883.
- [6] Elerian O, Chib S and Shephard N, Likelihood inference for discretely observed non-linear diffusions, *Econometrica*[J] 2001 69 959-93
- [7] Hong, Y. and H. Li, Nonparametric Specification Testing for Continuous-Time Models With Applications to Interest Rate Term Structures. *Review of Financial Studies*[J], 2004, forthcoming.

## Test the forecast ability of VaR models

Xiaoshu Liu , Zhenlong Zheng

(Department of Finance, Xiamen University, Xiamen, 361005, China)

**Abstract:** In this paper we empirically find that the current back-test tools including Basle test, Kupiec test and Christofferson test used in the business banks' model risk management can be some misleading. So, business banks should choose appropriate back-test method carefully.

**Key words:** VaR, backtest, model validation