CSSCI来源集刊

中国金融学

China Journal of Finance

四川大学金融研究所
复旦大学财务金融学系

2008 总第十四辑

中国金融出版社

CSSCI 来源集刊

中国金融学

2008　总第十四辑

四川大学金融研究所
复旦大学财务金融学系

中国金融出版社

责任编辑：吕　楠
责任校对：孙　蕊
责任印制：丁淮宾

图书在版编目（CIP）数据

中国金融学（Zhongguo Jinrongxue）2008. 总第十四辑/四川大学金融研究所，复旦大学财务金融学系. —北京：中国金融出版社，2009.2
ISBN 978-7-5049-4912-7

Ⅰ. 中…　Ⅱ. ①四…　②复…　Ⅲ. 金融—中国—文集　Ⅳ. F832-53

中国版本图书馆 CIP 数据核字（2009）第 007038 号

出版
发行　中国金融出版社
社址　北京市广安门外小红庙南里 3 号
市场开发部　（010）63272190，66070804（传真）
网上书店　http://www.chinafph.com　（010）63286832，63365686（传真）
读者服务部　（010）66070833，82672183
邮编　100055
经销　新华书店
印刷　利兴印刷有限公司
装订　东兴装订厂
尺寸　180 毫米×255 毫米
印张　9.75
字数　183 千
版次　2009 年 2 月第 1 版
印次　2009 年 2 月第 1 次印刷
印数　1—3290
定价　30.00 元
ISBN 978-7-5049-4912-7/F.4472

《中国金融学》编委会

目　录

论　文

Contents

Aticles

各类 VaR 方法的回测比较：基于中国股市的实证研究

戴晓凤　史敬

摘　要　本文通过建立一个较全面的 VaR 模型回测比较评价体系，从模型的准确性、保守性和有效性三个方面，对所建立的基于中国股市的 15 类 60 个 VaR 模型在 95% 和 99% 置信水平下的样本外预测值的特征进行回测比较。回测比较的结果显示，虽然没有任何一个 VaR 模型能在所有的考察指标中都表现出具有绝对的优势，但仍得出了一些较有普遍意义的结论。

关键词　VaR，股市风险，回测检验

1. 引　　言

1995 年，Beder（1995）在其研究中，对设定的三个证券组合分别用八种不同的方法计算其 VaR 值，发现计算结果相差最大的有 14 倍之多，由此引发了众多对各类 VaR 模型的评价与比较研究。其中，以 Hendricks（1996）、Kupiec（1995）以及 Lopez（1998，1999）三人的工作为代表分别产生了三类不同的比较工具。Hendricks（1996）选取十二种 VaR 计算方法，对每种方法计算电脑随机组合的 1 000 个资产组合的 VaR 值；并通过构建十个指标，来比较考察各种 VaR 模型的性质和差异。之后的学者将这种思路引入到具体的实证比较中来，Engel（1999）、Bredin 等（2002）在比较各 VaR 模型时分别都采用了 Hendricks（1996）提出的均值相对偏差（The Mean Relative Bias，MRB）和均方根相对偏差（The Root Mean Squared Relative Bias，RMSRB）两个指标；Sinha 等（2000）提出了一种滚动的绝对平均百分比误差（The Rolling Meaning Absolute Percentage Error，RMAPE）指标用来衡量模型的准确性。另外，Cassidy 等（1997）、Barone-Adesi（2000）、Caporin 等（2003）也分别根据比较需要建立自己的评价指标。由此构成了第一类指标评价工具。

第二类工具被称为假设检验工具，它主要是通过假设检验的方式来接受或拒绝一个 VaR 模型。自从 Kupiec（1995）提出了其经典的 Kupiec 检验之后，研究以这种方式来评价 VaR 模型的文献最多。Christoffersen（1998）在 Kupiec 的基础上提出了考察时间易变性的 Christoffersen 检验；出于同样的考虑，Marcus Haas（2001）提出了一种混合 Kupiec 检验。Christoffersen 等（2003）针对其 1998 年提出的检验方法的缺陷又重新开发了一种基于久期的检验方式。而

密度预测技术的引入也使得对 VaR 模型的检验设计更加丰富。Crnkovic 和 Drachman（1997）提出了著名的 C&D 检验方法，Diebold、Gunther 和 Tay（1998）则采用 CUSUM 和 CUSUM 的平方统计量来检验模型。Berkowitz（1999，2001a，2001b）更是在前人工作的基础上在密度预测领域内开发了一套完整的基于似然比检验的框架来评价 VaR 模型。另外，Gabriela 等（2002a，2002b），Joe 等（2003）都提出了自己的检验方式。

第三类工具是以 Lopez（1998，1999）为代表提出的比较评价工具，它根据管理者的自身偏好来构建损失函数，并依损失函数的大小来排序，借此评价 VaR 模型。之后，Imad 等（2002）通过建立一个“realistic volatility”的概念来求取 realistic volatility 以建立一个基准，从而得到风险值，然后用其他的度量方法与之相比，进行准确性排序。Christoffersen 等（2001）以 Kitamura 和 Stutzer 提出的理论并结合 Hansen 提出的广义矩过分识别检验，开发了一种检验方法用来比较两个 VaR 模型的优劣，如果将所有模型都两两比较的话，那么就能将各种 VaR 方法进行准确性排序了。

借助以上所述的众多比较工具，许多学者对各种 VaR 模型进行了大量的实证比较分析。Angelidis 等（2003）采用 Kupiec 检验和 Christoffersen 检验对各种 ARCH 类模型用于 VaR 估计进行了比较研究，他发现总体来说 ARCH 类模型产生的 VaR 估计都能有较高的精度，但比较来看，采用厚尾分布如 t 分布或 GED 分布能更好地改进模型。Sinha 等（2000）和 Lee 等（2001）在使用 Kupiec 和 Christoffersen 检验后，也采用自己构建指标的方式来进行比较分析。Sinha 等建模的对象是墨西哥和拉丁美洲。他认为，对于这些发展中的新兴市场，由于波动率较大，使用历史模拟法和 EWMA 方法将使 VaR 的估计产生严重偏差。而 Lee 等在对日本市场总共建立了 27 个 VaR 模型进行比较后认为：极值理论用于 VaR 建模并不如许多文献中说的那么好，反而一些传统的模型，如 TGARCH 和 Mento-Carlo 模拟法表现较好；但是，没有任何一个模型能在各种评价指标下都做到最好。Angelidis 等（2004）针对希腊股票市场采用 Kupiec 和 Christoffersen 检验结合 Lopez 的损失函数的评价方式，对建立的 14 个 VaR 模型进行排序，最终得出在 99% 置信水平下，Filter History Simulation 方法最好，而在较低的置信水平下时，各模型都差不多。

从以上的回顾可以看出，各种比较分析的结果差异很大。而建模方式、比较方法以及各个不同市场的具体情况，可能都是导致这些差异的原因。Diebold 和 Lopez（1996）指出：仅仅依赖某一个特定的属性是否被满足并据此来评价一个 VaR 估计值的精确性仅仅只捕捉到了关于 VaR 值精确性中十分有限的信息。因此，到目前为止，哪一种 VaR 模型是最优模型，甚至构建怎样的比较框架才能选择出最优模型，仍没有一个统一的标准。

本文以 VaR 方法用于我国股市风险度量为基本前提，针对我国股市风险进行建模分析的结果，通过尝试建立一个较全面的 VaR 模型回测评价体系，来对

各类 VaR 模型在用于我国股市风险度量时的特点进行全面、客观的比较和评价。

2. 研究对象与样本选择

2.1. 研究对象的选择

本文选择在中国股市上最具代表性的两种股价指数：上证综合指数（代码：0000001，以下简称上证指数）和深证成分指数（代码：399001，以下简称深成指数）作为实证研究对象。而 VaR 的实际计算我们则采用对数收益率的概念，即：$r_t = \ln(P_t) - \ln(P_{t-1})$。

由于股价指数的市场因子就是其本身，因此在计算 VaR 时，我们对资产价值的映射模块不作考虑，而主要考虑股价指数的波动性预测和股价指数的价值变化预测。

由于本文主要对我国股市风险进行分析，所以对于 VaR 持有期的选择为一个交易日，而对置信水平的选择由于存在一定的主观因素，在之后的实证分析中，本文选择 95% 和 99% 这两种比较普遍的置信水平来计算 VaR 值。

2.2. 样本区间的选择

考虑到能将各种不同的 VaR 模型放在同一起跑线上进行客观公正的评价和比较，同时也考虑到比较评价的最终结论能对目前我国的风险管理者选择构建自己的 VaR 模型提供参考，本文先将上证指数和深成指数的所有历史样本数据选取以下五个区间：

样本段 1：股价指数公布日至 1993 年 1 月 4 日（上证指数 518 个样本，深成指数 472 个样本）

样本段 2：1993 年 1 月 5 日至 1996 年 12 月 15 日（995 个样本）

样本段 3：1996 年 12 月 16 日至 2001 年 2 月 21 日（1 006 个样本）

样本段 4：2001 年 2 月 22 日至 2005 年 4 月 25 日（1 006 个样本）

样本段 5：1996 年 12 月 16 日至 2005 年 4 月 25 日（2 012 个样本）

如此划分的理由在于：鉴于 1993 年前我国股票还处于初期发展阶段，这些不规范的数据用于分析整个股市特征可能会造成扭曲，同时也可能对模型的比较和评价产生误导，而在 1996 年 12 月 16 日我国正式实行了涨跌停盘限价交易制度后，股市并没有出现大的制度性的变化。在这两个历史时点上比较特殊，应加以区别。同时，为了考察 1996 年 12 月 16 日后我国股市的基本特征是

否存在较大的变化，我们将其平分后再比较分析。

通过比较分析两个时段的基本信息，我们看到在1996年12月16日前的两个样本时段，不论从标准差还是从峰度来看，两市的收益率分布基本上都不满足正态性假设，沪市较深市而言，尖峰厚尾特点稍强，而深市较沪市而言波动性稍大，股市的特征明显不同于这之后的样本时段。为了保证比较实证的结果能有指导意义，建模时的样本区间选择不应落在这之内。另外，从1996年12月16日后的两段子样本和总体样本比较来看，几乎没有什么明显的差别，因此可以认为在1996年12月16日后的历史样本数据特征比较平稳，没有大的变动。在之后的VaR建模中我们将在1996年12月16日至2005年4月25日总共2 012个历史数据中，选择具体建模时的样本数据。

2.3. 样本外预测设计

参数VaR方法中最核心的部分在于如何构建较好的波动率模型来进行预测。在金融时间序列分析中，一般对波动率的预测分为两种：样本内预测和样本外预测。因此，对于一个VaR模型的样本内预测能力和样本外预测能力的评价结论我们也应当区别对待。从VaR方法提出的初衷和使用者的最终目标来看，样本外预测能力的好坏应该才是评价一个VaR模型是否合理的标准。但是，目前关于VaR模型的比较研究大都采用样本内预测评价的方式（Engle等，1999；Bredin等，2002；Sinha等，2000；Angelidis等，2003），采用样本外预测的则较少（Angelidis等，2004；Sama等，2000）。为此，本文设计了一种样本外预测的方式。

考虑到波动率模型一般都是短期预测模型，在进行样本外预测时通常只有向前一步预测的结果才是最精确的，如果要再往下预测的话，从理论上来说最合理的做法应该是移动样本窗口，并重新估计模型的参数后再进行预测。但是，在风险管理实务中，这种理论上最好的做法基本上不可能付诸实现，风险管理者不可能每天重新估计自己的模型参数再来测算VaR值。基于上面两点考虑，本文设计的样本外预测方式如下所述：本文拟采用各种VaR模型的500天样本外预测结果作为评价和比较各VaR模型的基础。假设风险管理者每半年（取125个交易日）调整或重新评估一次自己的模型。那么在对波动率模型进行估计时，每隔125天，移动估计样本的窗口，重新估计模型的参数并用来预测后125天的实际损益。如此循环估计4次后将其组合，那么就能得到一组500天样本外预测的VaR值了。由于考虑到了风险管理者的实际操作情况，所以对各VaR模型500天样本外预测结果进行评价和比较所得出的结论，对于风险管理者而言应该更具有参考价值。

本文选取8种参数VaR方法、4种非参数VaR方法以及3种半参数VaR方法总共15种方法，针对我国股票市场上的上证指数和深成指数分别建立了在

95%和99%置信水平下的60个VaR模型并计算出其500天的样本外VaR值。

3. VaR模型回顾测试体系构建

3.1. 回测体系的基本框架设计

“回顾测试”（backtesting）一词在VaR体系中专指：用于比较和评价各个建立的VaR模型好坏优劣的一系列方法。本文立足于作为我国金融机构引入VaR体系管理市场风险前的一个基础性研究，因此回测的目的是考察各类VaR模型的通用性和基础性，着重在于比较各类VaR方法在各个不同方面的特性，以期为各类机构选择具体的VaR模型提供参考。因此，本文设计的回测比较框架为：首先通过VaR模型的准确性检验，初步评价VaR模型的基本特性，然后，在VaR模型基本的准确性得到保证的前提下，选择构建一系列的评价指标，全面比较VaR模型在各个方面的性质。

在本节构建的VaR回测框架中主要考察以下三个方面的内容，并据此来选择假设检验和统计量。（1）准确性。即检验实际损益超出VaR值的比例是否在统计上与VaR建模中设定的置信水平保持一致。（2）保守性。即看被检验模型其产生的风险估计值总是相对于其他VaR模型来说要高。（3）有效性。在模型准确的条件下，再比较各个VaR模型的资本准备金机会成本，期望超出损失以及模型与实际波动的联动效果等方面的差异，以了解各个模型的特点和效率。

3.2. 回测体系中的假设检验设计

3.2.1. Kupiec 检验

关于VaR模型准确性最直接的检验方法就是考察实际损失超过VaR的概率。假定实际考察天数为 T，设定超出值序列 $I_{t+1}=\begin{cases}1, & \text{如果 } y_{t+1}<VaR_{t+1|t}\\ 0, & \text{如果 } y_{t+1}\geqslant VaR_{t+1|t}\end{cases}$，其中，$y_{t+1}$代表 $t+1$ 时刻的实际损益。那么失败的天数 $N=\sum_{t=1}^{T} I_t$，这时失效率应为 N/T，同时每一个 I_t 都是一个经典的贝努里试验。如果假定VaR估计具有时间独立性，则失败观察的二项式结果代表了一系列独立的贝努里试验。而失效率 N/T 的期望概率就应等于VaR的置信水平 p。因此，对VaR模型准确性的评估就转化为检验失效率 N/T 是否显著不同于 p。Kupiec（1995）提出了针对

原假设 $N/T=p$ 的似然比检验：

$$LR_{uc} = -2\ln[(1-p)^{T-N}p^{N}] + 2\ln\{[1-(N/T)]^{T-N}(N/T)^{T}\} \quad (1)$$

在原假设的条件下，统计量 LR_{uc} 服从自由度为 1 的 χ^2 分布。在一定的显著性水平下，我们可以构造相应的非拒绝区。不同置信水平下的非拒绝区如表 1 所示。

表 1　Kupiec 检验 95% 置信水平下的非拒绝置信区

VaR 置信水平	$T=255$	$T=510$	$T=1\ 000$
99%	$N<7$	$1<N<11$	$4<N<17$
97.50%	$2<N<12$	$6<N<21$	$15<N<36$
95%	$6<N<21$	$16<N<36$	$37<N<65$
92.50%	$11<N<28$	$27<N<51$	$59<N<92$
90%	$16<N<36$	$38<N<65$	$81<N<120$

注：N 是样本规模 T 中可观测到的失效次数。

资料来源：Kupiec（1995），参考文献［8］。

3.2.2.　Christoffersen 的区间预测评价法

1998 年 Christoffersen（1998）在 Kupiec 检验的基础上，进一步考察了 VaR 超出值的时间易变性。他认为，一个好的区间预测应该在市场波动平稳的时候较窄而在市场波动较大的时候自动变宽，这样落在区间预测之外的观察值应该均匀地分布在整个样本区内，而不应该存在聚类性（clustering）。

Christoffersen 的检验具体设计如下：在一天中，如果 VaR 没被超出，我们将偏差指标 I_t 标定为 0，否则为 1。定义 n_{ij} 为发生状态 i 的后一天发生状态 j 的天数。那么：

（1）对于无条件覆盖的似然比检验与 Kupiec 的 LR_{uc} 检验一样。

（2）对于例外值独立性的似然比检验为：

$$LR_{ind} = -2\cdot\ln\frac{(1-\hat{\pi}_2)^{(n_{00}+n_{10})}\hat{\pi}_2^{(n_{01}+n_{11})}}{(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}} \sim \chi^2(1) \quad (2)$$

其中：n_{ij} 为发生状态 i 的后一天发生状态 j 的天数。$(i,j=0,1)$

$$\hat{\pi}_{01}=\frac{n_{01}}{n_{01}+n_{00}},\ \hat{\pi}_{11}=\frac{n_{11}}{n_{10}+n_{11}},\ \hat{\pi}_2=\frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}$$

如果满足独立性假设的话，原假设即为：$\hat{\pi}_{01}=\hat{\pi}_{11}=\hat{\pi}_2$

（3）对于有条件覆盖的似然比检验则为：

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2(2) \quad (3)$$

经整理可写成：

$$LR_{cc} = -2 \cdot \ln \frac{(1-p)^{n_0} p^{n_1}}{(1-\hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1-\hat{\pi}_{11})^{n_{10}} \hat{\pi}_{11}^{n_{11}}} \sim \chi^2(2) \tag{4}$$

3.2.3. Mixed Kupiec 检验

与 Christoffersen 基于相同的考虑，Marcus Haas（2001）也提出了一种改进的 Kupiec 检验。他首先定义 v_i 为例外值 i 与例外值 $i-1$ 之间的天数。如果 VaR 模型是正确的话，那么在给定的置信水平下，每个 v_i 都应显著相同于 $1/p$。因此，对于 n 个超出值，就可以建立 n 个相互独立的检验统计量，并且都服从 $\chi^2(1)$ 分布。这样，当我们定义原假设为：超出值都是相互独立的，那么其联合检验的统计量就为

$$LR_{ind} = \sum_{i=2}^{n} \left[-2\ln\left(\frac{p(1-p)^{v_i-1}}{\hat{p}_i(1-\hat{p}_i)^{v_i-1}}\right)\right] - 2\ln\left[\frac{p(1-p)^{v-1}}{\hat{p}(1-\hat{p})^{v-1}}\right] \sim \chi^2(n) \tag{5}$$

其中，$\hat{p}_i$ 和 $\hat{p}$ 为每一个超出值所对应的失效率，v 表示第一个例外值出现前所经历的天数。将其与 Kupiec 的无条件似然比检验联立，即得

$$LR_{mix} = LR_{ind} + LR_{uc} \sim \chi^2(n+1) \tag{6}$$

3.2.4. 超出值序列的自相关检验

正如之前所述，像 Christoffersen 检验和 Mixed Kupiec 检验这样的似然比检验方法在小样本下的检验能力有限（特别是在 99% 置信水平）。而从考察超出值序列独立性的角度来看，最简单最直观的方法莫过于对超出值序列 I_t 的自相关分析。因此，在回测体系中将其加入作为辅助分析的工具以便更深入地分析超出值序列的性质也不失为一种很好的选择。序列自相关的原假设如下：

$$H_0: corr\{I_t, I_{t-s}\} = 0, \forall s \; vs, \; H_1: \exists s \; corr\{I_t, I_{t-s}\} \neq 0$$

对于该原假设我们采用 Ljung-Box 的 Q 统计量来进行检验。至于自相关分析时的最优滞后阶数选择，Boudoukh 等（1998）建议使用 1 周（即 5 天）的滞后来作为判断标准。因此，我们主要分析超出值序列滞后 5 阶的自相关性质，但同时，如果在其他滞后阶数上出现自相关异常的情况，我们也会加以考察。

3.2.5. Spearman 秩相关检验

在实际应用中，不仅要求 VaR 估计对实际损失的充分涵盖，还要求 VaR 估计值与实际的风险暴露高度相关，大的 VaR 估计值应伴随较大的损益发生，小的 VaR 估计值应伴随小的损益发生，因此，一个有效的 VaR 模型对实际损益分布的方差变动应具有较强的跟踪能力。

由于 VaR 估计序列和实际损益序列都具有明显的非正态性。而对于非正态分布的数据，采用 Spearman 秩相关检验比较适合（王星，2005）。所以，设计 Spearman 检验如下所示：

H_0：序列 X 与序列 Y 不相关 *vs*，H_1：系列 X 与序列 Y 正相关对序列 X 中的每一个元素 X_i 赋秩 R_i，对序列 Y 中的每一个元素 Y_i 赋秩 Q_i，则秩相关系数化简后可写成：$r_S = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2$ (7)

在原假设下，定义 T 统计量为：$T = r_S\sqrt{\frac{n-2}{1-r_S^2}}$ (8)

该统计量服从自由度为 $n-2$ 的 t 分布，并取右单侧检验。如果要进行双侧检验，那么备择假设应改为：系列 X 与序列 Y 存在相关性。

3.3. 回测体系中描述性统计量的选择

3.3.1. VaR 的均值和标准差

作为最基本的描述性统计量，均值和方差的内涵是相当丰富的。很明显，由于 VaR 值是以历史收益率计算得出，其均值和方差也一定随历史样本的不同而有不同，所以，单独考察每一个 VaR 模型预测的均值和标准差的绝对数量似乎不能说明模型的任何性质。但是，如果在相同的样本区间内，考察不同 VaR 模型的均值和标准差的相对大小对于模型的保守性和有效性就具有一定的反映能力了。在 VaR 模型的准确性得到满足的前提下，一个相对更大的 VaR 均值说明模型相对来说会产生更高的风险估计，这时要求的平均资本准备金也会更高。同样，标准差过大的 VaR 模型则说明 VaR 估计值的波动性很强，模型或许过于敏感。由于风险管理者调整风险准备金或改变交易头寸等都需要付出成本，对实际损益太过于敏感的模型无疑会大大提高管理者调整头寸和准备金的频率，这时将不利于管理者提高管理效率。

3.3.2. MRB 和 RMSRB 统计量

Hendricks（1996）开发了两个评价 VaR 模型的指标：MRB 和 RMSRB。

MRB（The Mean Relative Bias）指均值相对偏差。如果给定 T 为样本大小，N 为 VaR 模型的个数，那么第 i 个 VaR 模型的 MRB_i 定义如下：

$$MRB_i = \frac{1}{T}\sum_{t=1}^{T}\frac{VaR_{it} - \overline{VaR_t}}{\overline{VaR_t}}，其中\ \overline{VaR_t} = \frac{1}{N}\sum_{i=1}^{T}VaR_{it} \quad (9)$$

MRB 的作用是为了衡量每一个 VaR 模型的风险度量相对于所有 VaR 模型风险度量的平均值的偏差程度，如果偏差越小，说明模型越接近风险度量的平均水平。它是一种百分位数的表示方法，例如，如果 $MRB = \pm 0.01$，则说明这个 VaR 模型产生的风险估计比所有模型的平均度量水平高（或低）1%。

RMSRB（The Root Mean Squared Relative Bias）则指均值平方根相对偏差，

表达式如下：

$$RMSRB_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{VaR_{it} - \overline{VaR_t}}{\overline{VaR_t}}\right)^2}，其中\ \overline{VaR_t} = \frac{1}{N}\sum_{i=1}^{T} VaR_{it} \quad (10)$$

可以看出，对于所有 VaR 模型产生的平均风险估计值而言，RMSRB 度量了每一个 VaR 模型围绕这一平均值的变异性大小。它能捕捉到两个效应：(1) 一个 VaR 模型产生的风险估计是否显著不同于风险估计的平均水平。(2) 一个 VaR 模型产生的风险估计是否也显著不同于模型自身的变异性。为了区分这两种效应，我们再定义一个修正的 RMSRB 统计量：$RS_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(VaR_{it} - 0)^2}$，相比传统 RMSRB 而言，由于在任意时点上，这里假设其“均值”都为零，所以，它能消除由于各时点上 VaR 均值的不同引起的度量偏差而将所有模型放在同一水平下进行比较。

3.3.3. RMAPE 统计量

RMAPE（The Rolling Meaning Absolute Percentage Error）被称为滚动的绝对平均百分比误差，由 Sinha 等（2000）提出。假设在整个检验样本 T 中，给定一个 K 天的窗口，并计算在此窗口中的超出值个数。如果 VaR 模型准确的话，那么超出值的期望值应为 pK。

将这个窗口从第一个样本数据开始滚动直至最后，那么总共就能产生 $m = T - K + 1$ 个大小为 K 的窗口，记为 i。令 O_i 为在窗口 i 上观察到的超出值个数，那么其期望都应等于 pK。此时，RMAPE 被定义为

$$RMAPE_k = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{O_i - pK}{K}\right| \quad (11)$$

这个值越小，说明估计的效果越好。但是，值得注意的是，窗口大小的选取存在一定的问题。因为在较高的置信水平（如 99%）下，VaR 超出值的个数较少，这时如果窗口选取过小，必然会导致统计量反映的信息失真。因此，我们选取 4 种不同的 K 值，分别为 100、200、300、400，并计算其 *RMAPE* 值。只有 VaR 模型的 *RMAPE* 值在各个不同 K 值下与其他 VaR 模型相比都处于一个相对稳定的位置时，才能利用 *RMAPE* 进行解释，由此得出来的结果才能更具有说服力。

3.3.4. 损失发生时的期望超出损失

VaR 只给出在某一概率下可能发生损失的值，而对损失真正发生时损失的数量却不能考察，但这恰恰又是风险监管者和公司经营者都不能忽视的一个问题。一个有效的 VaR 模型不仅要能正确地涵盖损益，而且要能在损失发生时尽量地降低损失的数量。这个问题在 Hendrick（1996）和 Lopez（1998）的论文中都有所体现。Lopez 也正是出于这种考虑，将这个指标纳入了他的损失函数

中。本文不采用损失函数来评价，但由于该因素不能忽视，故在此我们定义：

$$SSEL_{it} = \begin{cases} (VaR_{it} - y_t) & \text{如果 } y_t - VaR_{it} < 0 \\ 0 & \text{如果 } y_t - VaR_{it} \geqslant 0 \end{cases} \tag{12}$$

则

$$SEL_i = \sum_{t=1}^{T} SSEL_{it} \tag{13}$$

$$AEL_i = \frac{SEL_i}{M} \tag{14}$$

其中，M 为观察到的超出值次数。

另外定义：

$$AUL_i = \frac{1}{M}\sum_{m=1}^{M} X_m ; MUL_i = \max\{X_m\} \tag{15}$$

其中，$X_m = \frac{y_t}{VaR_{it}}$，如果 $y_t - VaR_{it} < 0$；M 为超出值次数。

SEL_i 表示在整个样本中当实际损益超出 VaR 值的总和，AEL_i 为超出值的平均值，即期望超出损失。AUL_i 反映了超出值相对于 VaR 值的平均倍数。MUL_i 则为超出值相对于 VaR 值的最大倍数。对于 VaR 模型而言，这些指标越小模型就越有效。另外，MUL_i 还为损益分布的厚尾性提供了另一种度量的方式，同时也对 VaR 模型捕捉极端事件的能力进行了考察。

3.3.5. 金融机构的资金机会成本

对于金融监管当局而言，考虑到期望超出损失也许已经足够了，因为他们的目的是为了维护金融体系的稳定。但是对于金融机构的管理者而言，过高的资本准备金虽然能够避免巨额损失的发生，可同时也占用了过多的资金，使得公司存在较多的资源浪费，从而产生高额的机会成本。因此对于金融机构而言，一个有效的模型应该能在保证风险损失的大小在金融机构承受范围之内的同时将资金的机会成本减至最低。

对于此，我们设计的评价指标基本思路是：考察在实际损益未超出时，VaR 模型相对于实际损失的超出水平，这个值越小，说明 VaR 值跟实际损失的差异越小，机会成本也应越小。值得注意的是，真实损益有正也有负，但是，即使正的收益出现，风险损失准备也不可能为正。为此，在计算这个指标时，如果收益为正，我们就将其截平，令它等于零，然后再求其与 VaR 值的差。具体的数学表达式如下所示：

$$令\ TY_t = \begin{cases} y_t & \text{如果 } y_t \leqslant 0 \\ 0 & \text{如果 } y_t > 0 \end{cases}$$

$$则：TSCC_{it} = \begin{cases} (TY_t - VaR_{it}) & \text{如果} \quad TY_t - VaR_{it} > 0 \\ 0 & \text{如果} \quad TY_t - VaR_{it} \leqslant 0 \end{cases}$$

$$SCC_i = \sum_{t=1}^{T} TSCC_{it}, \quad ACC_i = \frac{1}{N} SCC_i, 其中, N 为未超出的次数。$$

SCC_i 为整个样本区间内的机会成本总和，ACC_i 为平均机会成本。

4. 各类 VaR 方法在我国股市上的回测实证分析

通过各模型的 VaR 回测，初步结果显示：（1）除简单移动平均方法外，其他参数 VaR 模型和半参数 VaR 模型相较于非参数 VaR 模型波动都更加剧烈。（2）基于 t 分布和 GED 分布的 GARCH 模型以及各类非参数模型在 95% 和 99% 置信水平下的 VaR 值差距更加明显。（3）Bootstrap 方法在两种不同的历史样本（分别为 1 512 天和 250 天）下，产生的 VaR 估计图形具有明显的差别，这也从实证的角度证明了在 VaR 估计中，样本长度和时期的选择对最终结果具有重要的影响。下面，笔者根据比较的三个原则分别进行分析。

4.1. 准确性分析

4.1.1. 基本的准确性分析

表 2 给出了所有 60 个模型的 VaR 超出值个数。对照表 1 中给出的 95% 置信水平下 Kupiec 检验的非拒绝置信区，我们可以对其进行 Kupiec 检验。

表 2 各 VaR 模型超出值个数表

VaR 模型	上证指数		深成指数	
	95%	99%	95%	99%
SMA（K = 125）	23	2	23	4
EWMA	25	1 *	22	3
GARCH	24	8	23	4
GARCH – M	40 *	13 *	24	5
EGARCH	14 *	4	19	2
EGARCH – M	19	4	22	3
GARCH – T	27	2	27	4
GARCH – GED	20	2	23	2
Hybrid – HS	30	9	29	14 *
Filter – HS	23	3	21	3
Bootstrap – G	21	4	24	3
Bootstrap1	13 *	1 *	16 *	1 *