

知乎

首发于  
川流不息

## Barra 因子模型中的风险调整



石川

量化交易 话题的优秀回答者

已关注

47 人赞同了该文章

### 摘要

除了 Newey-West 调整，Barra 模型中同时还使用了 Eigenfactor 风险调整和贝叶斯收缩来进一步提高协方差矩阵的估计。本文介绍这两种技巧。

### 1 引言

上一期文章讲了 Barra 对协方差矩阵的 Newey-West 调整（[链接](#)），本文趁热打铁，再简单介绍一下 Barra 对风险的另外两处调整，它们都是 Barra 模型中的核心组成部分。

在《[正确理解 Barra 的纯因子模型](#)》中曾经提到，Barra 的纯因子模型没有什么可投资性，但是仍然有两个重要的作用：

1. 由于每个因子组合都仅仅专注于单一因子，因此它有利于评估每个因子自身的表现；

知乎

首发于  
川流不息

通过多因子模型，股票的超额收益被分解为被因子解释的部分以及各自的特异性收益率：

$$\mathbf{r} = \mathbf{X}\mathbf{f} + \mathbf{u}$$

在上式中， $\mathbf{r}$  为  $N \times 1$  维个股收益率向量（省略了时间下标，假设有  $N$  支股票）、 $\mathbf{X}$  为当期因子暴露矩阵（ $N \times K$  矩阵， $K$  为因子个数）， $\mathbf{f}$  为  $K \times 1$  维因子收益率向量， $\mathbf{u}$  为  $N \times 1$  维个股特异性收益率向量。通过因子模型可知，个股收益率的协方差满足如下关系：

$$\mathbf{V} = \mathbf{X}\mathbf{V}_f\mathbf{X}' + \mathbf{\Delta}$$

其中， $\mathbf{V}$ （ $N \times N$ ）是股票收益率的协方差矩阵， $\mathbf{V}_f$ （ $K \times K$ ）是因子收益率的协方差矩阵，而  $\mathbf{\Delta}$  为  $N \times N$  对角阵，其对角线上的元素对应个股的特异性收益率的方差。为了得到  $\mathbf{V}$  的准确估计，对于  $\mathbf{V}_f$  和  $\mathbf{\Delta}$  的求解至关重要。在上一期《协方差矩阵的 Newey-West 调整》中，我们指出为了得到更准确的估计，Barra 对  $\mathbf{V}_f$  和  $\mathbf{\Delta}$  都进行了 Newey-West 调整。

然而，Barra 对风险调整的脚步并没有止步于此。在 Newey-West 调整之后，它们对因子收益率的协方差矩阵  $\mathbf{V}_f$  进行了 **Eigenfactor risk adjustment**，并对个股的特异性收益率方差矩阵  $\mathbf{\Delta}$  进行了 **Bayesian shrinkage**（贝叶斯收缩）。本文就来对它们分别做简要介绍。

在此之前，先让我们了解另外一个概念：**Bias statistic（偏差统计量）**，这是因为本文关注的两种调整都是以降低 Bias statistic 为目标的。

## 2 偏差统计量

Bias statistic 是评估风险模型准确性的一个常用指标（Menchero et al. 2011），它用来衡量风险预测值和风险实际值（用已实现波动率来评估）之间的误差。如果这二者之间有显著误差，则我们说这个风险预测是有偏的，故这个统计量称为偏差统计量。

在数学上，偏差统计量的定义如下。令  $R_{nt}$  为某投资组合  $n$  在  $t$  期的收益率， $\sigma_{nt}$  是  $t$  期期初 (beginning-of-period) 的预测风险。用  $\sigma_{nt}$  对  $R_{nt}$  进行标准化得到：

$$b_{nt} = \frac{R_{nt}}{\sigma_{nt}}$$

上面这个标准化在数学上的含义是将  $R_{nt}$  的波动率——实际的已实现波动率——使用期初预测的波动率  $\sigma_{nt}$  进行标准化。如果期初风险预测是准确的，那么  $b_{nt}$  的标准差应该为 1。利用总共  $T$  期的  $b_{nt}$  计算样本标准差就得到 Barra 定义的偏差统计量  $B_n$ ：



知乎

首发于  
川流不息

$$\sqrt{\frac{1}{T} \sum_{t=1}^T \dots}$$

在收益率符合正态分布的假设下， $B_n$  的 95% 的置信区间为：

$$B_n \in [1 - \sqrt{2/T}, 1 + \sqrt{2/T}]$$

可见，当样本个数足够大（即 T 足够大）的时候，偏差统计量  $B_n$  应该离 1 不远。如果它离 1 不远则说明我们在期初的风险预测值比较准确，反之则意味着期初的风险预测值是有偏的，这时就需要修正。

在上面的使用中，一个比较强的假设是收益率满足正态分布。当背离这条假设时， $B_n$  很有可能远离 1 而这却并不意味着风险的预测是有偏的。在这方面，Barra 的 USE4 research note (Menchero et al. 2011) 里面有更多的讨论。在本文的最后，我们简单评论一下这个正态分布假设。

好了，现在我们清楚了偏差统计量，马上来看对协方差矩阵的 eigenfactor 风险调整。

### 3 协方差矩阵的 Eigenfactor 风险调整

Barra 的 USE4 模型中，在对因子收益率协方差矩阵进行 Newey-West 调整之后，又进行了 Eigenfactor 调整。关于这个调整介绍的最详细的文献是 Menchero, Wang, and Orr (2011)，这篇文章比 USE4 模型文档中的介绍要更完整，我们就从它讲起。本节的介绍偏重于解释 Eigenfactor 调整的业务意义，会涉及少量必要的数学推导，但是不会详述所有的技术细节。

这里请暂时忘记因子，因为上面这篇文章的研究的对象是个股超额收益的的协方差矩阵。超额收益是相对市场而言的，定义如下：

$$f_{nt} = r_{nt} - R_t^M$$

上式中， $r_{nt}$  是股票 n 在交易日 t 的收益， $R_t^M$  是该日的市场收益， $f_{nt}$  是股票在 t 的超额收益。使用不同股票的超额收益序列，就能计算出样本协方差矩阵：

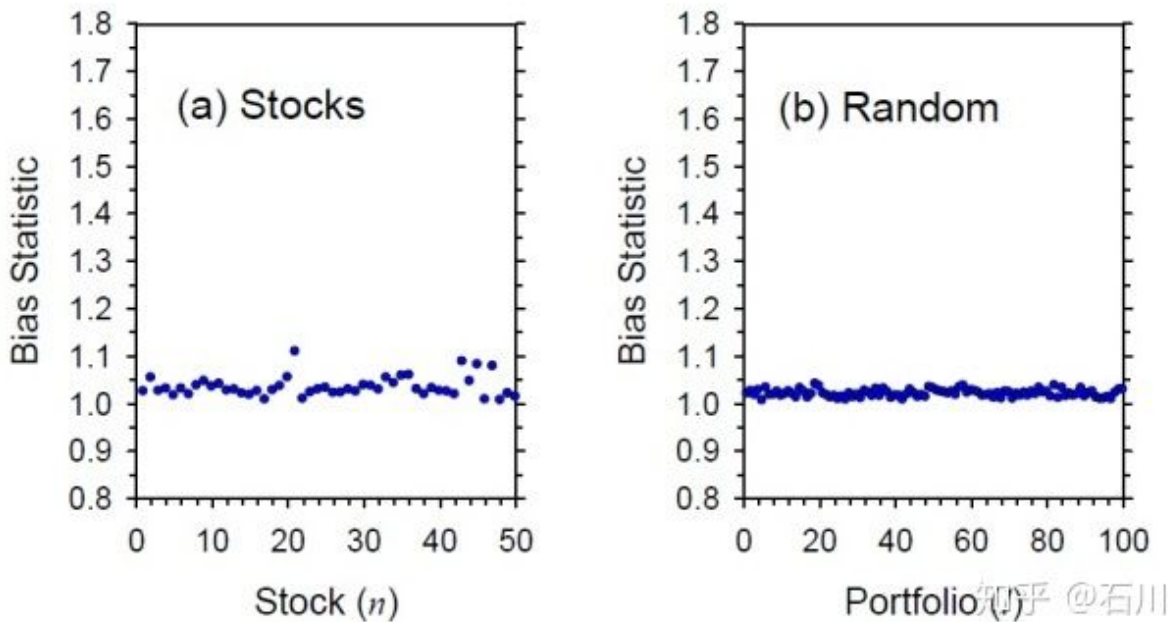
$$\mathbf{V}_0(mn) = \frac{1}{T-1} \sum_{t=1}^T (f_{mt} - \bar{f}_{mt})(f_{nt} - \bar{f}_{nt})$$

这里  $\mathbf{V}_0$  是协方差矩阵，而  $\mathbf{V}_0(mn)$  代表这个矩阵中对应的股票 m 和 n 之间的协方差。

$\mathbf{V}_0$  是样本协方差矩阵，它只是未知总体的估计，它是无偏估计吗？是否会在某些特定使用方式下有很大的偏差？这就是 Barra 关心的。为此采用第二节提到的 Bias statistic 来评估它是否是偏的。



知乎

首发于  
川流不息

然而在现实中，我们并不是随机的投资个股，所以上面的结论虽然不错但可惜是不够的。**现实中我们通常以某种目标来优化投资组合、确定股票在组合中的权重**；比如使用马科维茨 mean-variance optimization 框架得到的最优组合。对于这些最优化得到的投资组合，它们的 Bias statistic 怎样呢？

为了回答这个问题，Eigenfactor 闪亮登场。

假设样本协方差矩阵  $\mathbf{V}_0$  是满秩的。使用它的所有特征向量构建一个矩阵  $\mathbf{U}_0$  —— 这个矩阵的每一列就是  $\mathbf{V}_0$  的一个特征向量。利用线性代数的性质，我们可以用  $\mathbf{U}_0$  把  $\mathbf{V}_0$  “旋转”一下，变成一个对角阵：

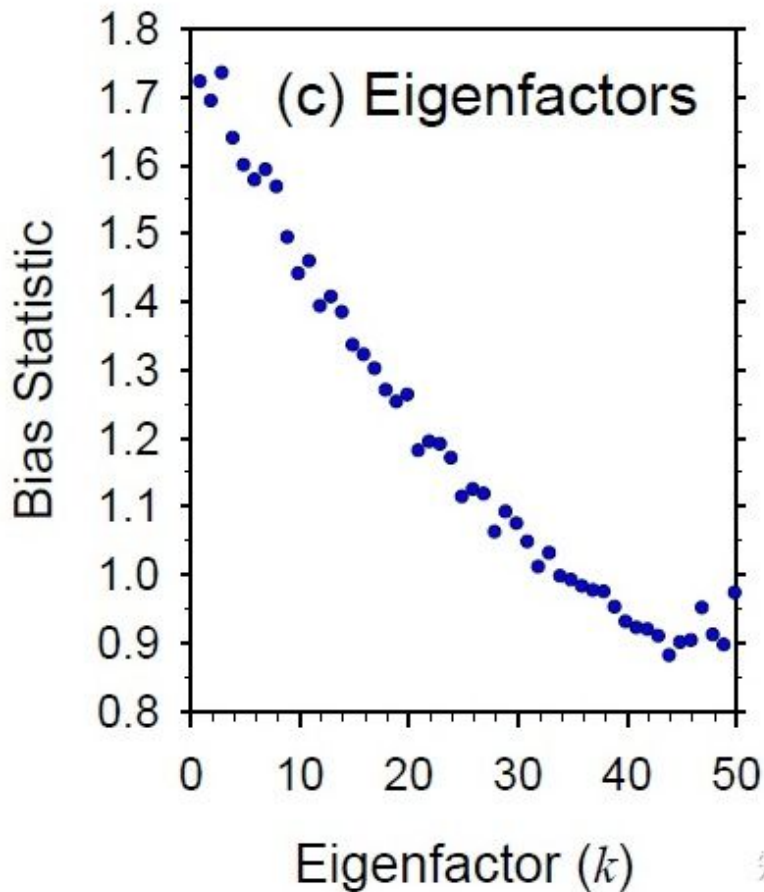
$$\mathbf{D}_0 = \mathbf{U}_0^{-1} \mathbf{V}_0 \mathbf{U}_0$$

这个  $\mathbf{D}_0$  是个对角阵，它对角线上的元素为和那些特征向量对应的特征值。这其实是对  $\mathbf{V}_0$  进行特征分解。这个变换的业务含义是什么呢？假设一共有  $N$  个股票，则每一个特征向量都是一个  $N \times 1$  的向量，它里面第  $n$  个数就代表这第  $n$  支股票的权重。以该特征向量中的数值作为权重就得到了一个 eigenfactor portfolio (eigenfactor 一词是 Barra 发明的，因为来自特征向量 eigenvector)。

由于一共有  $N$  个特征向量，所以有  $N$  个 eigenfactor portfolios。更重要的是，这些投资组合之间相互独立，协方差为 0，这在数学上体现在  $\mathbf{D}_0$  是个对角阵。前面刚说过，它对角线上的元素是  $\mathbf{V}_0$  的特征值，而它们也是这些 eigenfactor portfolios 的方差。

**那么这些 eigenfactor portfolios 的 Bias statistics 如何呢？是否接近 1 呢？很不幸，答案否定的。**下图是按照 eigenfactor portfolios 的样本方差从小到大顺序将这些组合排列（横坐

知乎

首发于  
川流不息

即便如此我们仍然会问——这 eigenfactor portfolios 在实际中有含义吗？如果我们不按照特征向量中的权重来构建投资组合，那么即便它们的 Bias statistics 偏离 1 对我们也没有影响。不幸的是，它们有含义！这就是 Barra 考虑 eigenfactor 调整的原因。以下是 Barra 的原文：

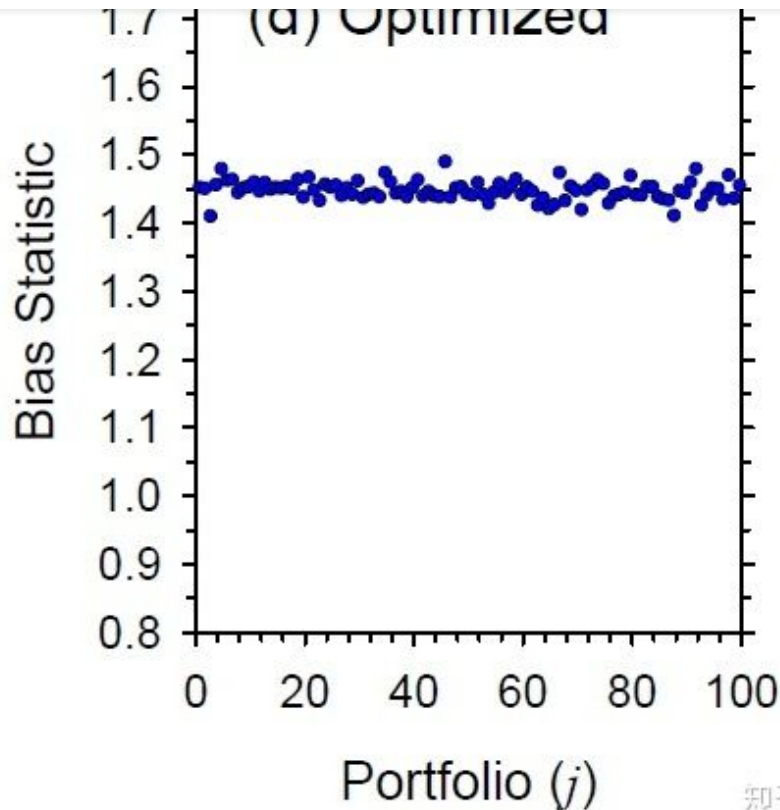
*Eigenfactors are not economically intuitive. However, **they do play an important role in portfolio optimization**. For instance, the first eigenfactor solves for the minimum variance portfolio subject to the constraint that the sum of squared weights adds up to 1. Similarly, the last eigenfactor solves the corresponding maximum variance problem.*

它的意思是**这些 eigenfactor portfolios 在投资组合的最优化构建中有很大的意义**。比如，方差最小的 eigenfactor 组合就等价于我们以最小化组合方差为优化目标构建的组合（这句话逻辑真完美.....）；方差最大的 eigenfactor 组合就等价于我们以最大化方差为优化目标（不确定是否有人这么干.....）而构建的投资组合。

在 Barra 看来，这些 eigenfactor portfolios 和我们最终使用这些个股、以某种最优化目标来构建的最优投资组合有着千丝万缕的联系。因此，如果这些 eigenfactor portfolios 的 Bias statistics 很差，那么我们可以预期，以某种目标最优化得到的投资组合的 Bias statistics 也好不到哪去（图是 Barra 给的模拟产生的最优化组合的 Bias statistics 例子，都高达 1.4 以上）。



知乎

首发于  
川流不息

这就是为什么要对协方差矩阵进行 eigenfactor 调整。读到这里，细心的小伙伴可能会说“你上面这都是个股协方差矩阵啊，但是 USE4 里面是因子协方差矩阵”。

下面说 USE4。

对因子协方差矩阵的 eigenfactor 调整的数学方法和上面完全一致。在 Menchero, Wang, and Orr (2011) 这篇文章的数学推导中，Barra 就说了计算样本协方差矩阵的“assets”可以是 factors，大类资产以及个股：

## Appendix A: Eigenvariance De-biasing

Let  $\mathbf{V}_0$  denote the  $N \times N$  sample covariance matrix, computed as

$$\mathbf{V}_0 = \frac{\mathbf{f} \cdot \mathbf{f}'}{T-1}, \quad (\text{A1})$$

where  $\mathbf{f}$  is the  $N \times T$  matrix of realized asset returns and  $T$  is the number of periods. The “assets” could represent factor portfolios, asset classes, or individual securities. We assume that the sample covariance matrix is full rank, which requires that the number of periods  $T$  exceeds the number of assets  $N$ .

知乎 @石川

因此从数学上来说，不管上面是个股协方差矩阵还是因子协方差矩阵，都没问题。从业务上来说呢，同样对因子协方差矩阵做“旋转”。假设有  $K$  个因子（别忘了，这每一个因子代表着一个个股构建出来的纯因子投资组合），因此得到  $K$  个特征向量。

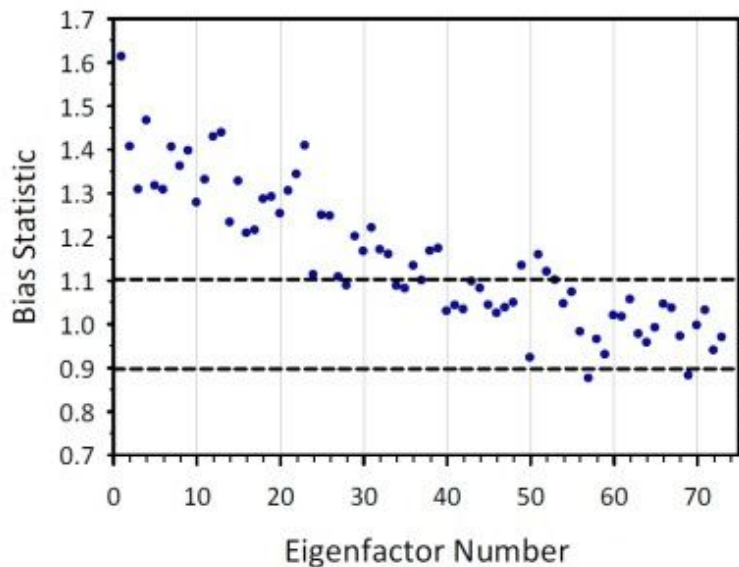
知乎

首发于  
川流不息

析因子的目的是为了针对这些纯因子组合的风险收益特性进行进一步优化，从而把这些因子组合放在一起，得到一个多因子组合（即让我们最终的投资组合暴露于多个优异的  $\alpha$  因子中）。然而，上面的分析指出，**以纯因子组合为输入经过优化后得到的投资组合，它的风险估计是有偏的**（Bias statistic 显著不为 1，见下图）。

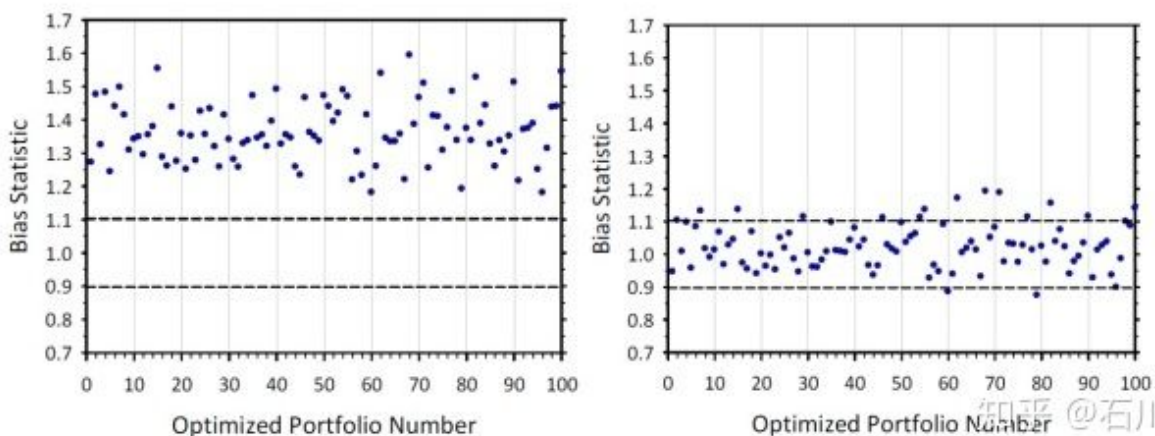
Figure 4.1

Bias statistics of eigenfactors using the unadjusted covariance matrix of Equation 4.1. Results were computed using the USE4S covariance matrix parameters reported in Table 4.1. The sample period comprised 191 months, from July 1995 through May 2011. The 95-percent confidence interval is indicated by the two dashed horizontal lines. While the large eigenfactors lie mostly within the confidence interval, the small eigenfactors fall well outside it.



知乎 @石川

因此，对因子的协方差矩阵做 **eigenfactor risk 调整也是十分必要的**。这就是 Barra 采取这个调整的原因。下图比较了调整前（左图）后（右图）的 Bias statistics，效果显著：



知乎 @石川

上面从业务逻辑的角度解释了为什么 Barra 要对因子协方差矩阵进行 eigenfactor risk 调整。在具体的数学细节上面，其思想是 Bootstrap，请阅读 Menchero, Wang, and Orr (2011) 者是 USE4 模型的文档，这里就不赘述了。

知乎

首发于  
川流不息

本节来解释 Barra 对个股特异性收益率的风险调整 —— 贝叶斯收缩。

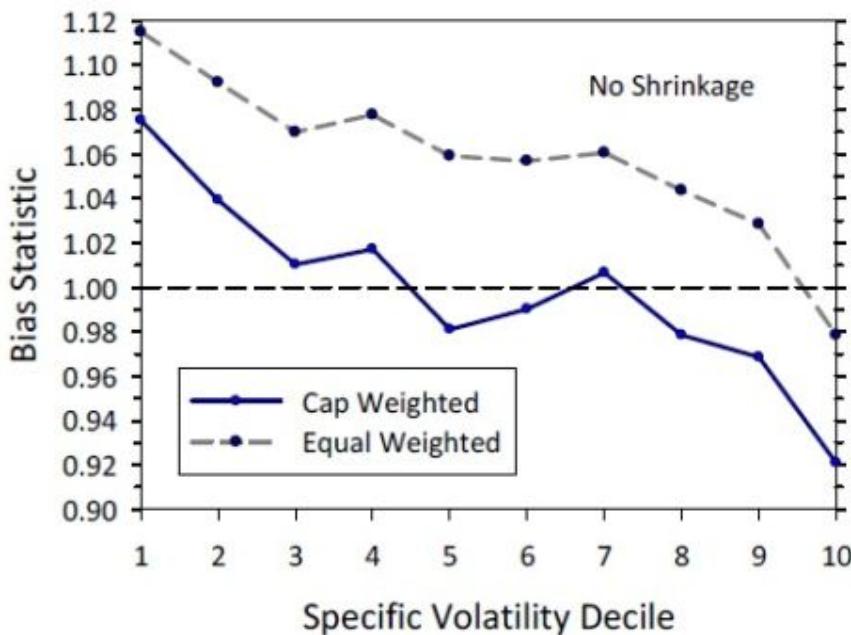
**贝叶斯收缩是一个常见的将先验和样本估计值结合起来的手段；它是先验和样本估计值的线性组合**（见《收益率预测的贝叶斯收缩》）。对（协）方差矩阵进行贝叶斯收缩并不是 Barra 的独创，事实上 Ledoit and Wolf (2003) 就提出了这个思想，并取得了不错的效果。他们的方法也比下面要介绍的 Barra 的收缩方法（主要是在收缩强度系数的选取上）复杂的多，这个咱们以后再说。

来看 Barra 的问题。首先计算出个股的特异性波动率。但是，Barra 指出使用**样本内数据计算出的特异性波动率在样本外的持续性很差**。

下图中，所有股票按照特异性波动率大小分成 10 档（图中第 1 档代表波动率最小；第 10 档代表波动率最大），计算每档的平均 Bias statistic。可以看到，对于波动率小的档，Bias statistic 显著大于 1，说明它低估了样本外这些股票的特异性波动率；而对于波动率大的档，Bias statistic 显著小于 1，说明它高估了这些股票在样本外的特异性波动率。

Figure 5.1

Bias statistics of specific returns grouped by specific risk deciles (sorted from low to high), without Bayesian shrinkage adjustments. Specific risk was computed per Equation 5.5. Both cap-weighted and equal-weighted results are reported. The sample period comprised 191 months, from July 1995 through May 2011. Without Bayesian shrinkage, note the tendency to overpredict high-volatility stocks and underpredict low-volatility stocks.



知乎 @石川

既然使用样本数据得到的这个估计值不准，那就需要使用先验来矫正一下。先验就是我们认为正确的特异性波动率，所以我们把样本数据计算出来的特异性波动率向着先验来靠拢，这就是“收缩”一词的意思，这就是为什么这个技术较贝叶斯收缩。

如何计算先验呢？对于任意给定的个股，Barra 采用**一大堆**个股特异性收益的波动率的**均值**作为先验。这个“一大堆”是什么呢？Barra 把所有个股按照市值分成十档，然后找到我们目标个股



知乎

首发于  
川流不息

计算这一大堆中所有股票的特异性波动率，取它们的平均。怎么取呢？不是简单的等权，而是按照市值加权的。**这个使用和目标股票处在同一市值这一档所有股票（一大堆）按照市值权重计算出来的特异性波动率就是先验。**以  $s_n$  表示市值档位  $\hat{\sigma}_n$  表示  $s_n$  中股票  $n$  的特异性收益率， $w_n$  表示  $s_n$  中股票  $n$  的按照其市值计算出来的权重。则这个先验的表达式为：

$$\bar{\sigma}(s_n) = \sum_{n \in s_n} w_n \hat{\sigma}_n$$

可见，先验就是把属于  $s_n$  内的所有股票的特异性波动率按照它们的市值为权重平均起来。

现在先验、样本估计都有了，最后一步就是把这二者线性组合在一起：

$$\sigma_n^{SH} = \nu_n \bar{\sigma}(s_n) + (1 - \nu_n) \hat{\sigma}_n$$

上式中等式左侧就是收缩后股票  $n$  的最终特异性波动率，等式右侧的第一项中的  $\nu_n$  是在收缩时赋予先验的权重（称为收缩强度系数）。如何确定权重呢？**它和样本估计值与先验的偏离程度有关。**具体的， $\nu_n$  的表达式为：

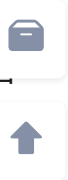
$$\nu_n = \frac{q |\hat{\sigma}_n - \bar{\sigma}(s_n)|}{\sqrt{\frac{1}{N(s_n)} \sum_{j \in s_n} (\hat{\sigma}_j - \bar{\sigma}(s_n))^2 + q |\hat{\sigma}_n - \bar{\sigma}(s_n)|}}$$

上式中， $q$  是一个经验系数， $N(s_n)$  是市值档位  $s_n$  中股票的个数。这个表达式中的分子以及分母中的第二项的  $|\hat{\sigma}_n - \bar{\sigma}(s_n)|$  表示了我们股票  $n$  的样本特异性波动率和其先验之间的偏离程度；而上式分母中的第一项是市值档位  $s_n$  中所有股票的特异性波动率和其先验偏离程度的标准差，它是这一大堆股票的**平均偏离程度**的一个度量。

最终的压缩权重  $\nu_n$  就由这两个偏离程度（以及经验系数  $q$ ）决定：

- $|\hat{\sigma}_n - \bar{\sigma}(s_n)|$  越大， $\nu_n$  就越大，而不要忘记  $\nu_n$  是先验的权重。这就是说，对于目标个股，样本估计值越不靠谱（它的偏离程度和所有小伙伴的平均偏离程度相比更高），我们越不能相信样本估计，而是越要相信先验，所以  $\nu_n$  越大。
- $|\hat{\sigma}_n - \bar{\sigma}(s_n)|$  越小，说明这个目标股票特异性波动率的偏离程度低，我们愿意相信它，所以这时赋予先验的权重  $\nu_n$  就要小点。

在上面  $\nu_n$  的表达式中，唯一剩下的就是要确定经验系数  $q$  了。Barra 没有具体说，但是不难想它一定和 Bias Statistic 有关。贝叶斯收缩的目的就是为了降低个股特异性波动率的 Bias Statistic，所以可以通过综合考虑所有个股特异性波动率收缩前后 Bias statistic 的改进来找到合适的  $q$  值。

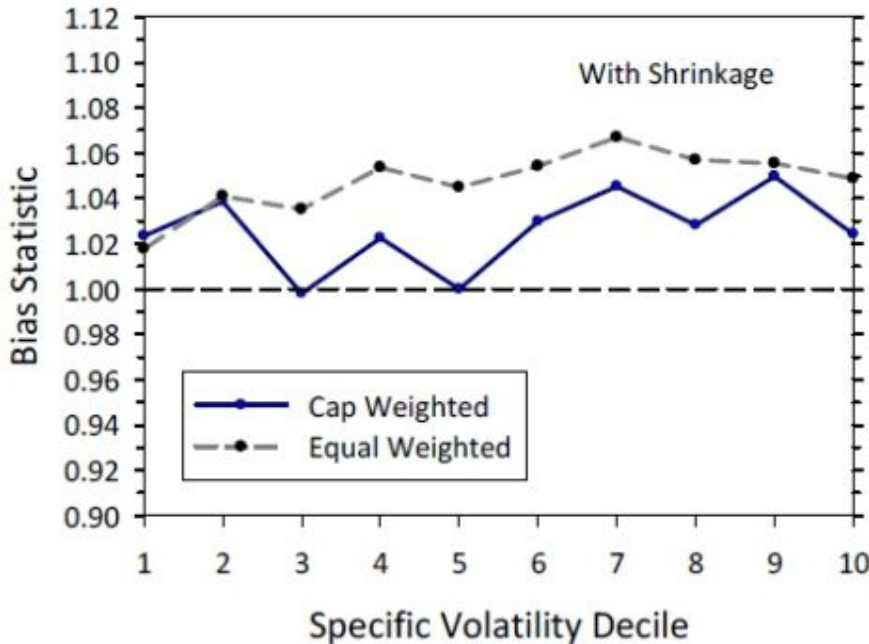


知乎

首发于  
川流不息

Figure 5.2

Bias statistics of specific returns grouped by specific risk deciles (sorted from low to high), with Bayesian shrinkage adjustments. Specific risk was computed using Equation 5.6. Both cap-weighted and equal-weighted results are reported. The sample period comprised 191 months, from July 1995 through May 2011. The Bayesian shrinkage technique largely removes the biases associated with either extreme.



知乎 @石川

以上就是对特异性波动率做的贝叶斯收缩。

## 5 结语

本文介绍了 Barra 对波动率的两种调整方法，它们都是以改善 Bias statistic 为目标。而第二节曾指出，如果收益率不满足正态分布，那么 Bias statistics 可能也是不准的。

既然存在收益率不满足正态分布这个风险，那么 Barra 仍将上述偏差统计量用于因子收益率协方差矩阵和个股特异性收益率方差的调整中，是否合理呢？先来说说因子收益率。对于每个因子，它的收益率是一揽子股票的加权收益率（权重是从模型中根据截面回归来的），因此它是一个投资组合（纯因子组合）的收益率。比起个股，投资组合的收益率应该更加满足正态分布的假设。再来看看个股的特异性收益率。在市场上流行的因子模型中，对因子解释不了的残差（即特异性收益率）通常做的假设大多是正态分布。所以，从这个意义上说，似乎能理解 Barra 坚持使用上述偏差统计量的原因。

在投资实务中，任何模型都需要假设、模型本身并无好坏。所以我们也不用把 Barra 的处理方法当作唯一的、正确的答案。这仅仅是来自 Barra 的选择——我相信这背后自有它的道理和考量。的小伙伴给我们留言，告诉我们国内一些券商报告中有很多其他不错的评价协方差矩阵准确性的方法。那些无疑也是值得我们学习和尝试的。

知乎

首发于  
川流不息

- Ledoit, O. and M. Wolf (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10, 603 – 621
- Menchero, J., D. J. Orr, and J. Wang (2011). *The Barra US Equity Model (USE4)*. MSCI Barra Research Notes.
- Menchero, J., J. Wang, and D. J. Orr (2011). *Eigen-adjusted Covariance Matrices*. MSCI Research Insight.

**免责声明：**文章内容不可视为投资意见。市场有风险，入市需谨慎。

原创不易，请保护版权。如需转载，请联系获得授权，并注明出处，谢谢。已委托“维权骑士”  
([维权骑士\\_免费版权监测/版权保护/版权分发](#)) 为进行维权行动。

编辑于 2019-07-02

[量化交易](#)[BARRA模型](#)[多因子模型](#)

▲ 赞同 47 ▼    3 条评论    分享    ★ 收藏    ...

## 文章被以下专栏收录

**川流不息**

北京量信投资管理有限公司是一家在中国基金业协会备案登记的专业私募基金管理人...

[关注专栏](#)

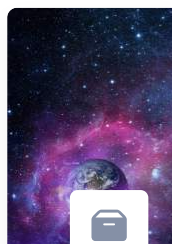
## 推荐阅读



**Barra系列（四）：风险模型之  
异质风险**



**基于数库行业的特质波动率策略**





**量信基金**

3 条评论

⇌ 切换为时间排序

写下你的评论...



 shiqiang 1 年前


写得很清晰了，赞

 赞  回复  踩  举报


 石川 (作者) 回复 shiqiang 1 年前

感谢鼓励!

 赞

 杨宁 8 个月前

请问一下截面回归的R-squared大概是多少? 我自己做了一遍, 只有16%

 1