

风险模型提速组合优化的另一种方案

——《因子选股系列研究之三十七》



研究结论

- 风险模型主要实现三个功能：估算协方差矩阵、控制风险暴露和组合绩效归因分析。后两者需要用到结构化的因子风险模型（例如 BARRA、Axioma）；估算协方差矩阵可以用结构化因子模型，也可以采用纯统计方法。
- 结构化因子模型的最大好处在于降维，既可以降低参数估计误差，也可以降低协方差相关计算的复杂度，大幅提升组合优化速度；但缺点是模型会存在设定偏误，需要维护更新风险因子库。统计模型没有设定偏误，只需要用到股票收益率数据，计算效率很高，但输入到组合优化时，无法通过因子模型降维的方式实现优化提速。
- 本报告提高了一种方法可以兼顾统计模型的高效便捷和因子模型的计算提速。首先用压缩估计方法（报告用的是线性压缩）给出协方差矩阵估计量 $\hat{\Sigma}$ ，再对其谱分解，保留前 K 个最大特征值对应的部分，残余部分用对角阵近似，这样可以近似拆解出一个结构化因子模型，后续输入到组合优化中提速。
- K 取值越大，近似过程带来的误差越小，但组合优化速度会下降，两者需要权衡。我们分别测试了成分股内和全市场增强沪深 300 和中证 500 指数的策略，发现 K 取 40 以上时，谱分解近似方法得到的结果和采用原始压缩估计量的结果基本一致，但用 `cvxpy+ecos` 优化时，速度可以实现两个数量级的提升。矩阵谱分解的拆解过程我们写成了一个 python 函数，感兴趣的投资者可以与我们联系。

风险提示

- 量化模型失效风险
- 市场极端环境的冲击

报告发布日期

2018 年 03 月 28 日

证券分析师 朱剑涛

021-63325888*6077

zhujiantao@orientsec.com.cn

执业证书编号：S0860515060001

相关报告

A 股小市值溢价的来源	2018-03-04
组合优化的若干问题	2018-03-01
基于风险监控的动态调仓策略	2018-02-22
反转因子择时研究	2018-02-21

东方证券股份有限公司经相关主管机关核准具备证券投资咨询业务资格，据此开展发布证券研究报告业务。

东方证券股份有限公司及其关联机构在法律许可的范围内正在或将要与本研究报告所分析的企业发展业务关系。因此，投资者应当考虑到本公司可能存在对报告的客观性产生影响的利益冲突，不应视本证券研究报告为作出投资决策的唯一因素。

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

风险模型主要实现三个功能：估算协方差矩阵、控制风险暴露和组合绩效归因分析。后两者需要用到结构化的因子风险模型（例如 BARRA、Axioma）；估算协方差矩阵可以用结构化因子模型，也可以采用纯统计方法（例如：压缩估计量，参考前期报告《A股市场风险分析》和《风险模型在时间序列上的改进》）。

用结构化因子模型的好处是可以大幅缩减要估计的参数数量，降低估计误差，同时也降低了组合优化过程中协方差相关计算的计算复杂度，在股票数量较多时可以将组合优化速度提升一至两个数量级，节约策略回溯调试时间，投资者可以利用开源优化器高效实现这个优化过程（例如 python 的 cvxpy+ECOS，参考上篇报告《组合优化的若干问题》）。结构化因子模型不便之处在于投资者需要自己定期维护更新风险因子库，或购买商业版风险模型；而且结构化因子模型存在模型设定偏误风险，理论上不同的股票池会有不同的风险因子，但实际投资中用的都是同一套风险因子。统计模型可以避免这种模型设定偏误，也不需要维护一个风险因子库，只需要用到股票收益率数据，计算效率很高，但在组合优化时无法像结构化因子模型那样享受到降维带来的求解速度大幅提升。

这篇报告里，我们先用线性压缩估计量方法得到协方差矩阵估计 Σ ，再对 Σ 谱分解可以近似的拆出一个“因子结构”，然后输入到组合优化中实现提速。这种处理方案可以同时兼顾结构化因子模型和统计模型的优点。

一、模型原理

假设我们有一个月频调仓的多因子股票组合，每个月底的时候，我们基于 N 个股票过去一年的收益率数据，用 Ledoit(2003)的线性压缩估计量方法（目标阵取为对角阵）得到协方差矩阵估计值 Σ ，它是一个正定矩阵，可以做如下谱分解（Spectral Decomposition）（或称为特征分解）

$$\Sigma = \sum_{i=1}^N \lambda_i * u_i * u_i^T$$

其中 $\lambda_i, i = 1, 2 \dots N$ 是矩阵 Σ 的特征值，并按从大到小的顺序进行排列， u_i 是期对应的特征向量，我们可以参照主成份分析里面方差解释度的概念，统计前 K 个最大的特征值之和占所有特征值之和的比例 $p = \sum_{i=1}^K \lambda_i / \sum_{i=1}^N \lambda_i$ ，结果如图 1 所示，可以看到，不论是在沪深 300 成份股、中证 500 成份股还是全市场范围内，前 20 个最大特征值之和占比就超过了 50%；当 K 取 100 时，沪深 300 成分内这个比例超过 90%，中证 500 成分内超过 80%，全市场范围内超过 70%，可见绝大部分特征值数值都比较小。

图 1：前 K 个最大特征值之和占比

基准指数	K=100	K=80	K=60	K=40	K=20	K=10
沪深300成份股	91.0%	87.4%	82.5%	75.6%	65.0%	56.6%
中证500成份股	85.2%	80.7%	75.1%	67.7%	57.5%	50.4%
全市场	76.6%	72.0%	66.7%	60.3%	52.0%	46.4%

数据来源：东方证券研究所 & Wind 资讯

需要注意的是，这里的前 K 个最大特征值之和占比与主成份分析里的方差解释度的计算式一样，但主成份里用的是样本协方差矩阵的特征值，上文用的是压缩估计量的特征值。之所以采用这个指标是用来估量下一步近似过程中可能的误差。

股票数量较多时， Σ 的大部分特征值都很小，因此可以把谱分解拆成两部分：

$$\Sigma = \sum_{i=1}^K \lambda_i * u_i * u_i^T + \sum_{i=K+1}^N \lambda_i * u_i * u_i^T$$

前一部分可以写成矩阵形式 $B * F * B^T$ 。其中 B 是 $N \times K$ 矩阵，第 j 列即为第 j 个特征值对应的特征向量 $u_j, j = 1, 2 \dots K$ ， F 是一个对角阵，对角线元素为前 K 个特征值 $\lambda_1, \dots, \lambda_K$ 。后面一部分可以直接取矩阵和的对角线元素来做对角阵 S 用以近似， K 取得越大，后面一部分近似带来的整体误差越小。这样 Σ 可以近似表示为：

$$\Sigma \approx B * F * B^T + S$$

我们通过统计方法近似拆解出了一个结构化因子模型，再利用上篇报告《组合优化的若干问题》的方法输入到 `cvxpy+ecos` 中即可实现提速。

二、实证效果

在上述方法中， K 的取值较为关键。 K 取值越大，省略的项越少，造成的误差损失越小，组合优化结果更接近与直接用压缩估计量，但会增加运算复杂度，组合优化速度提升幅度会降低； K 取值变小的话，则会反之，因此两者间需要权衡。

下面我们用四个策略：沪深 300 增强（成分内）、沪深 300 增强（全市场）、中证 500 增强（成分内）、中证 500 增强（全市场），分别测试比较了标准压缩估计量、结构化因子模型和不同 K 取值下的谱分解近似方法的优化速度和效果。组合优化问题设置如下：

$$\max: f'w - \lambda w' \Sigma w$$

$$\text{st: } w'I = 0$$

$$w'Indus = 0$$

$$w'MV = 0$$

$$wmin < w < wmax$$

目标函数里风险厌恶系数取 10，个股权重上限分段设置，行业和市值完全中性。测试结果如图 2 所示

首先需要注意的是选择不同的风险模型，实际构建组合时用到的股票池可能会不一样。例如，为了避免过多噪音数据的干扰，用压缩估计量时会要求过去一年里该只股票至少有一半时间正常交易，这样会剔除备选股票池里的部分股票，而用结构化因子模型时，模型设计人员可能会用不同的筛选标准，造成构建组合时，两个模型实际用到的股票池不一样，会有细微差别，这种差别对组合收益、跟踪误差这些指标影响不大，但最大回撤指标可能会收到较大影响。

其次，策略回溯测试用了不同的风险模型，但风险厌恶系数都一样 ($\lambda = 10$)，但风险模型不同，最优的 λ 取值可能不一样，需要调试。

图 2：不同风险模型的效果对比

沪深300增强(成分内)	压缩估计量	因子模型	压缩估计量谱分解近似					
			k=100	k=80	k=60	k=40	k=20	k=10
IR	2.51	2.42	2.48	2.49	2.48	2.46	2.43	2.44
年化对冲收益	9.49%	8.98%	9.37%	9.41%	9.34%	9.26%	9.14%	9.24%
最大回撤	-2.27%	-2.95%	-2.26%	-2.26%	-2.19%	-2.30%	-2.31%	-2.34%
跟踪误差	3.64%	3.59%	3.64%	3.64%	3.64%	3.63%	3.63%	3.66%
单期优化用时(s)	0.44	0.05	0.12	0.13	0.09	0.06	0.04	0.03
前K个特征值之和占比			90.95%	87.44%	82.55%	75.59%	64.99%	56.64%

中证500增强(成分内)	压缩估计量	因子模型	压缩估计量谱分解近似					
			k=100	k=80	k=60	k=40	k=20	k=10
IR	2.20	2.42	2.21	2.22	2.18	2.20	2.22	2.21
年化对冲收益	10.34%	10.88%	10.33%	10.41%	10.22%	10.14%	10.28%	10.27%
最大回撤	-5.22%	-4.48%	-5.23%	-5.27%	-5.20%	-5.26%	-4.83%	-4.71%
跟踪误差	4.52%	4.31%	4.50%	4.51%	4.51%	4.45%	4.45%	4.47%
单期优化用时(s)	2.07	0.08	0.23	0.15	0.12	0.08	0.05	0.04
前K个特征值之和占比			85.15%	80.75%	75.09%	67.69%	57.50%	50.44%

沪深300增强(全市场)	压缩估计量	因子模型	压缩估计量谱分解近似					
			k=100	k=80	k=60	k=40	k=20	k=10
IR	2.80	2.84	2.81	2.83	2.83	2.86	2.88	2.87
年化对冲收益	10.54%	10.44%	10.58%	10.70%	10.67%	10.75%	10.78%	10.72%
最大回撤	-5.35%	-4.49%	-5.37%	-5.29%	-5.25%	-5.23%	-4.97%	-4.92%
跟踪误差	3.60%	3.53%	3.61%	3.62%	3.61%	3.59%	3.58%	3.58%
单期优化用时(s)	468.60	0.62	2.29	1.42	1.09	0.68	0.36	0.35
前K个特征值之和占比			76.56%	72.00%	66.67%	60.26%	52.04%	46.43%

中证500增强(全市场)	压缩估计量	因子模型	压缩估计量谱分解近似					
			k=100	k=80	k=60	k=40	k=20	k=10
IR	2.90	2.83	2.91	2.89	2.89	2.87	2.82	2.76
年化对冲收益	15.88%	14.69%	15.93%	15.84%	15.84%	15.65%	15.57%	15.24%
最大回撤	-4.89%	-4.29%	-5.00%	-4.77%	-4.74%	-5.08%	-5.15%	-6.07%
跟踪误差	5.13%	4.89%	5.12%	5.14%	5.14%	5.16%	5.18%	5.19%
单期优化用时(s)	384.95	0.61	1.80	1.25	1.00	0.54	0.38	0.25
前K个特征值之和占比			76.56%	72.00%	66.67%	60.26%	52.04%	46.43%

数据来源：东方证券研究所 & Wind 资讯

从图 2 可以看到，由于大部分特征值取值都比较小，所以当 K 取 40 以上的数值时，K 的取值对结果总体影响不大，和直接用压缩估计量的结果非常接近。参照主成分分析的原则，我们应选取特征值累积占比达到 80% 的 K 值，意味着 K 要取 100，不过从检测结果来看，K=100 和 K=40 策略结果相差不大，但优化速度下降明显，实际运用中，投资可以考虑取 K=40。这篇报告里使用线性压缩方法估计协方差矩阵，投资者也可以考虑用其它方法（例如非线性压缩）给出一个估计值，然后再仿照上面步骤进行近似拆解。

风险提示

1. 量化模型基于历史数据分析，未来存在失效风险，建议投资者紧密跟踪模型表现。
2. 极端市场环境可能对模型效果造成剧烈冲击，导致收益亏损。

分析师申明

每位负责撰写本研究报告全部或部分内容的研究分析师在此作以下声明：

分析师在本报告中对所提及的证券或发行人发表的任何建议和观点均准确地反映了其个人对该证券或发行人的看法和判断；分析师薪酬的任何组成部分无论是在过去、现在及将来，均与其在本研究报告中所表述的具体建议或观点无任何直接或间接的关系。

投资评级和相关定义

报告发布日后的 12 个月内的公司的涨跌幅相对同期的上证指数/深证成指的涨跌幅为基准；

公司投资评级的量化标准

买入：相对强于市场基准指数收益率 15%以上；

增持：相对强于市场基准指数收益率 5%～15%；

中性：相对于市场基准指数收益率在-5%～+5%之间波动；

减持：相对弱于市场基准指数收益率在-5%以下。

未评级——由于在报告发出之时该股票不在本公司研究覆盖范围内，分析师基于当时对该股票的研究状况，未给予投资评级相关信息。

暂停评级——根据监管制度及本公司相关规定，研究报告发布之时该投资对象可能与本公司存在潜在的利益冲突情形；亦或是研究报告发布当时该股票的价值和价格分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确投资评级；分析师在上述情况下暂停对该股票给予投资评级等信息，投资者需要注意在此报告发布之前曾给予该股票的投资评级、盈利预测及目标价格等信息不再有效。

行业投资评级的量化标准：

看好：相对强于市场基准指数收益率 5%以上；

中性：相对于市场基准指数收益率在-5%～+5%之间波动；

看淡：相对于市场基准指数收益率在-5%以下。

未评级：由于在报告发出之时该行业不在本公司研究覆盖范围内，分析师基于当时对该行业的研究状况，未给予投资评级等相关信息。

暂停评级：由于研究报告发布当时该行业的投资价值分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确行业投资评级；分析师在上述情况下暂停对该行业给予投资评级信息，投资者需要注意在此报告发布之前曾给予该行业的投资评级信息不再有效。

免责声明

本证券研究报告（以下简称“本报告”）由东方证券股份有限公司（以下简称“本公司”）制作及发布。

本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。本报告的全体接收人应当采取必要措施防止本报告被转发给他人。

本报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的证券研究报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的证券研究报告之外，绝大多数证券研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人作出邀请。

本报告中提及的投资价格和价值以及这些投资带来的收入可能会波动。过去的表现并不代表未来的表现，未来的回报也无法保证，投资者可能会损失本金。外汇汇率波动有可能对某些投资的价值或价格或来自这一投资的收入产生不良影响。那些涉及期货、期权及其它衍生工具的交易，因其包括重大的市场风险，因此并不适合所有投资者。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面协议授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容。不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

经本公司事先书面协议授权刊载或转发的，被授权机构承担相关刊载或者转发责任。不得对本报告进行任何有悖原意的引用、删节和修改。

提示客户及公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告，慎重使用公众媒体刊载的证券研究报告。

东方证券研究所

地址：上海市中山南路 318 号东方国际金融广场 26 楼

联系人：王骏飞

电话：021-63325888*1131

传真：021-63326786

网址：www.dfzq.com.cn

Email：wangjunfei@orientsec.com.cn

