

多因子模型 Beta 估计方法

报告要点

■ 估计 Beta，多因子模型应用的必要条件

前一篇报告中我们展示了多因子模型的巨大潜力，但是实际要运用这个模型，首先要解决的就是 Beta 估计的问题。

■ 常用的几种 Beta 估计方法

我们探讨了多因子模型中的 Beta 估计的几种常用方法，即窗口式的滚动最小二乘、全部历史数据滚动最小二乘、均值回复的状态空间方法和 GARCH 方法。

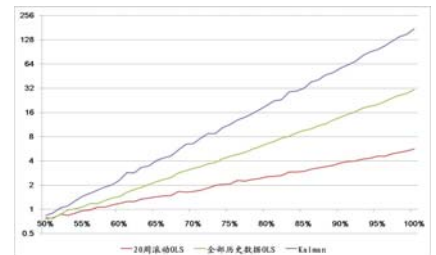
由于周数据收益率序列本身或回归残差没有表现出序列相关性或波动相关性，GARCH 模型可能不适用于周数据的 Beta 估计

■ 使用估计的 Beta 后，多因子选股模型表现如何？

比较了窗口式的滚动最小二乘、全部历史数据滚动最小二乘、均值回复的状态空间方法估计的 Beta 在假定的因子准确度下的多因子选股策略收益率。值回复的状态空间方法的估计结果较为理想。

表明使用估计的 Beta 后，多因子选股模型仍能获得较高的收益，这一收益随着因子预测准确性的提高呈指数形式上升。

多因子选股策略收益——不同因子准确性



数据来源：Wind 资讯，长江证券研究部

相关研究

《不要迷恋 Alpha——多因子选股系列报告(一)》

分析师

俞文冰

(8621) 021-63219519

ywub@cjsc.com.cn

执业证书编号：S0490209100276

联系人：

俞文冰

(8621) 021-63219519

ywub@cjsc.com.cn

研究报告平台：

正文目录

一、多因子选股策略回顾	3
1、多因子选股策略	3
2、理论上的最佳收益	3
二、关于 Beta 系数的理论研究概述	4
三、我们采用的预测 Beta 系数的方法	4
1、全历史数据最小二乘 Beta	4
2、滚动最小二乘计算的 Beta	4
3、均值回复型的状态空间方程	4
4、关于 GARCH 方法	6
四、Beta 序列波动性的比较	7
五、多因子选股效果	9
1、不同因子准确性下的收益	9
2、预测的 Beta 和实际的因子	10

图表目录

图 1：多因子选股收益----实际的因子和全样本回归的 Beta	3
图 2：Kalman 滤波方法图示	5
图 3：状态空间方法预测 Beta 的流程	6
图 4：沪深 300 周收益率序列及其平方的 ACF/PACF	7
图 5：个股回归残差及其平方的 ACF/PACF	7
图 6：市场因子 Beta 标准差直方图	8
图 7：大小盘因子 Beta 标准差直方图	8
图 8：市净率因子 Beta 标准差直方图	9
图 9：多因子选股策略收益----不同因子准确性	10
图 10：多因子选股收益-----实际因子和 20 周滚动 OLS 的 Beta	11
图 11：多因子选股累积收益倍数-----实际因子和 20 周滚动 OLS 的 Beta	11
图 12：多因子选股收益-----实际因子和全部历史数据 OLS 的 Beta	12
图 13：多因子选股累积收益倍数-----实际因子和全部历史数据 OLS 的 Beta	12
图 14：多因子选股收益-----实际因子和状态空间 Beta	13
图 15：多因子选股累积收益倍数-----实际因子和状态空间 Beta	13

一、多因子选股策略回顾

1、多因子选股策略

在《不要迷恋 Alpha——多因子选股系列报告(一)》中我们指出多因子模型能更好的解释收益率，并由此提出了多因子选股策略。我们选取的因子是市场收益率因子、大小盘因子和市净率因子。

我们稍微回顾一下这个方法：首先对于 A 股市场的每只股票预测其对每个因子的 Beta 值，进一步再预测三个因子在下一周的走势，最后依照下面的公式对每只股票的投资价值进行评分：

$$Score_i(t) = 0.55 \cdot X_1(t) \cdot \beta_1(i,t) + 0.27 \cdot X_2(t) \cdot \beta_2(i,t) + 0.18 \cdot X_3(t) \cdot \beta_3(i,t)$$

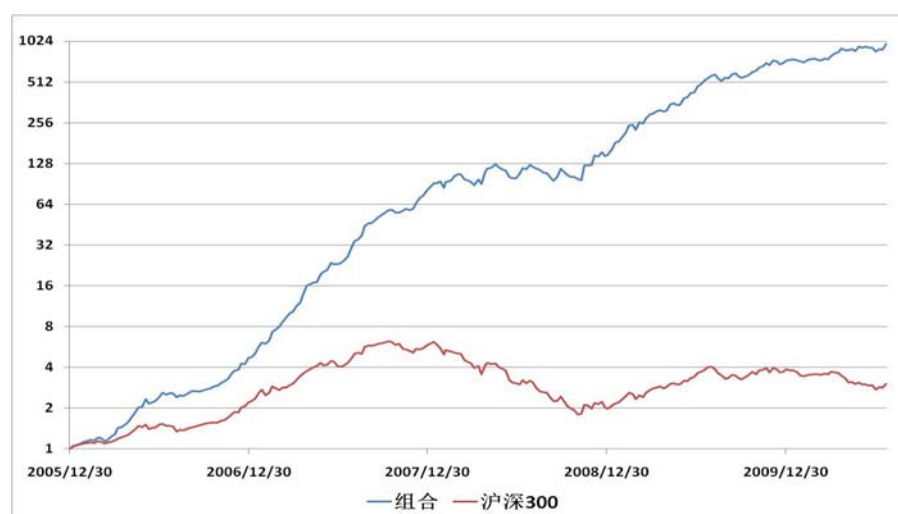
其中 $X_1(t), X_2(t), X_3(t)$ 表示对于 t 时刻的因子走势方向的预测，预测上涨，其值为 1，预测下跌，其值为 -1。

多因子选股策略就是选择得分最高的 n 只股票构成组合。

2、理论上的最佳收益

我们使用“完美预测”的 Beta 和“完美预测”的因子计算了三因子模型可以获得的理论上的最佳收益。可以看出，如果使用事后的 Beta，多因子模型可以获得相当惊人的收益。实际能获得多少收益，取决于对 Beta 和因子的预测。本篇报告主要关注 Beta 的预测。我们的目标就是寻找一种合适的 Beta 预测方法，使得实际收益率相对理论收益率下降得尽量少一些。

图 1：多因子选股收益----实际的因子和全样本回归的 Beta



资料来源：Wind 资讯，长江证券研究部

二、关于 Beta 系数的理论研究概述

Beta 系数是通过模型计算出来的参数，本身是无法直接观测的。我们只能根据过去的数据对其进行估计。

现有研究大部分认同以下观点：

- (1) 组合规模的大小与 Beta 系数预测的稳定程度有关，组合规模越大，未来 Beta 系数的预测准确性越高
- (2) 高 Beta 系数的股票在下一期 Beta 值倾向于变小，低 Beta 值的股票下一期 Beta 值倾向于变大，即 Beta 的变化过程有均值回复特性
- (3) 数据频率的增大会使得 Beta 的估计值变小，即使用日收益率计算的 Beta 值比使用周收益率或月收益率计算的 Beta 值小

目前关于 Beta 系数时变估计的方法主要有：最小二乘（OLS）方法，Garch 方法和状态空间方法。

本文中主要使用状态空间方法，同时考虑两种不同的 OLS 方法与之对比。

三、我们采用的预测 Beta 系数的方法

我们主要使用状态空间方法对 Beta 系数进行预测，同时考虑全历史数据最小二乘方法和滚动最小二乘方法作为参照。

1、全历史数据最小二乘 Beta

$$r_{i,t} = \alpha_i + \beta_i \cdot X_t + \varepsilon_i, t = 1, \dots, T$$

这种方法实际上是认为存在一个真实的、固定不变的 Beta 值，使用所有的历史数据进行回归得到的值就是对这个真实值的最好的估计，也是对 Beta 未来值的最佳预测

2、滚动最小二乘计算的 Beta

$$r_{i,t} = \alpha_i + \beta_i \cdot X_t + \varepsilon_i, t = T - k, \dots, T$$

这种方法放弃了 Beta 一直固定不变的假设，而是认为 Beta 会不断发生变化。但是这个模型对与 Beta 时变的方式没有做进一步的假设，只是简单地使用最近一段数据回归得到的 Beta 值作为 Beta 的估计

3、均值回复型的状态空间方程

$$\begin{aligned} r_{i,t} &= \alpha_{i,t} + \beta_{i,t} \cdot X_t + \varepsilon_{i,t} \\ \alpha_{i,t} - \bar{\alpha}_i &= \rho_{\alpha,i} (\alpha_{i,t-1} - \bar{\alpha}_i) + \mu_{i,t} \\ \beta_{i,t} - \bar{\beta}_i &= \rho_{\beta,i} (\beta_{i,t-1} - \bar{\beta}_i) + \eta_{i,t} \\ \varepsilon_{i,t} &\sim N(0, \sigma^2_{\varepsilon,i}), \mu_{i,t} \sim N(0, \sigma^2_{\mu,i}), \eta_{i,t} \sim N(0, \Sigma_i) \end{aligned}$$

其中：

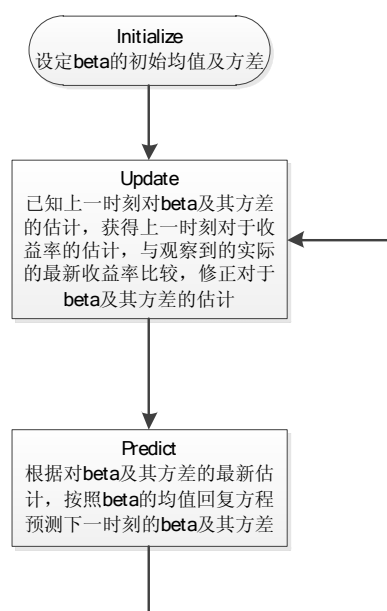
$$\beta_{i,t} = \begin{pmatrix} \beta_{1,i,t} \\ \beta_{2,i,t} \\ \beta_{3,i,t} \end{pmatrix}, \quad \bar{\beta}_i = \begin{pmatrix} \bar{\beta}_{1,i} \\ \bar{\beta}_{2,i} \\ \bar{\beta}_{3,i} \end{pmatrix}, \quad X_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{pmatrix}$$

$$\rho_i = \begin{pmatrix} \rho_{1,i} & 0 & 0 \\ 0 & \rho_{2,i} & 0 \\ 0 & 0 & \rho_{3,i} \end{pmatrix}, \quad \Sigma_i = \begin{pmatrix} \sigma_{1,i}^2 & 0 & 0 \\ 0 & \sigma_{2,i}^2 & 0 \\ 0 & 0 & \sigma_{3,i}^2 \end{pmatrix}$$

这种方法既承认每只股票有一个“内在的”Beta 值，同时又认为每个时刻的 Beta 值并不是简单的与这个内在的 Beta 值相等，而是围绕这个内在 Beta 值波动的。当偏离内在 Beta 值较多的时候，下一期的 Beta 就有向“内在 Beta 值”运动的趋势。

求解上述状态空间模型的方法是 Kalman 滤波。在所有参数都已知的情况下，Kalman 滤波的方法如下图所示：

图 2: Kalman 滤波方法图示



资料来源：Wind 资讯，长江证券研究部

参数估计的目标函数：

$$\max L(\theta) = \max \sum_{t=0}^T \log p(r_t | F_{t-1}, \theta)$$

在正态假设下，上述最优化问题等价于：

$$\max L(r, \theta) = \min \left(\sum_{t=0}^T \log P_t + \frac{(r_t - r_{t|t-1})^2}{P_t} \right)$$

$$P_t = X_t \cdot \Sigma_t \cdot X_t' + \sigma_{r,i}^2$$

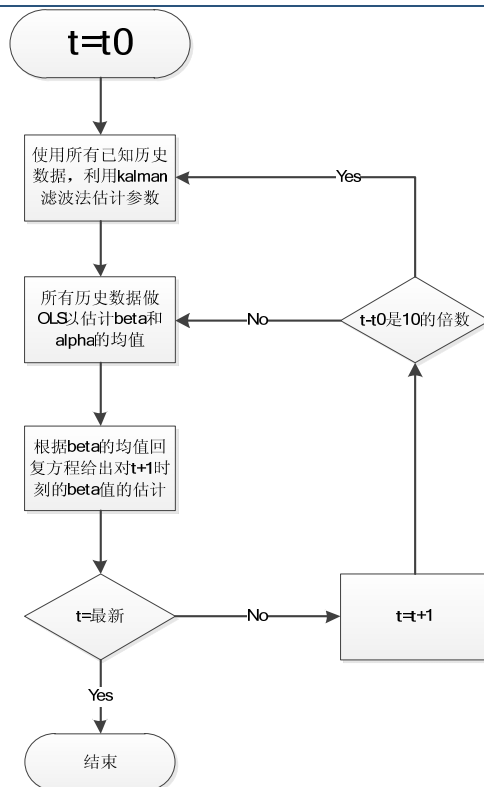
θ 表示模型中的所有除 $\bar{\alpha}_i$ 和 $\bar{\beta}_i$ 之外的参数。如前面所述，全部历史数据的回归得到的

Beta 值就是对“内在 Beta 值”的最好的估计。因此 $\bar{\alpha}_i$ 和 $\bar{\beta}_i$ 是不需要进行极大似然估计的。

由于每次参数优化所需的计算时间非常大，我们没有每周对参数进行更新，而是每 10 周更新一次参数。

使用 Kalman 滤波方法预测 Beta 的流程总结如下：

图 3：状态空间方法预测 Beta 的流程

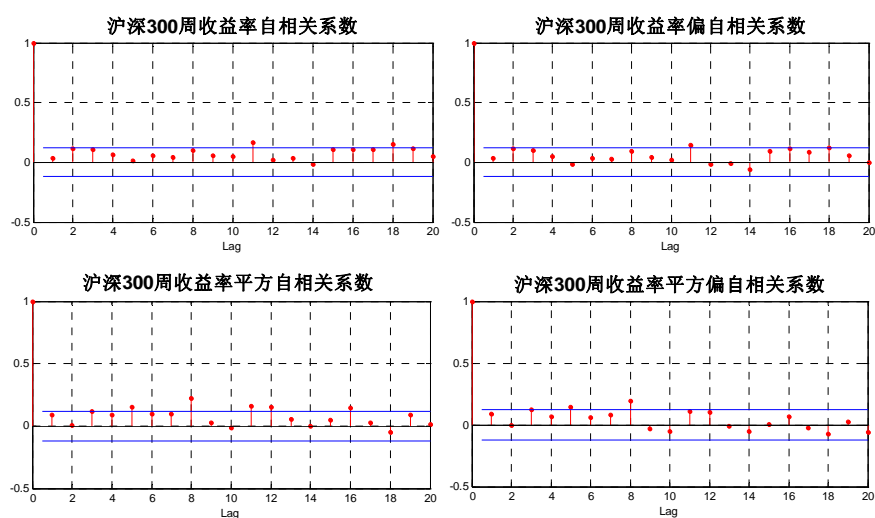


资料来源：Wind 资讯，长江证券研究部

4、关于 GARCH 方法

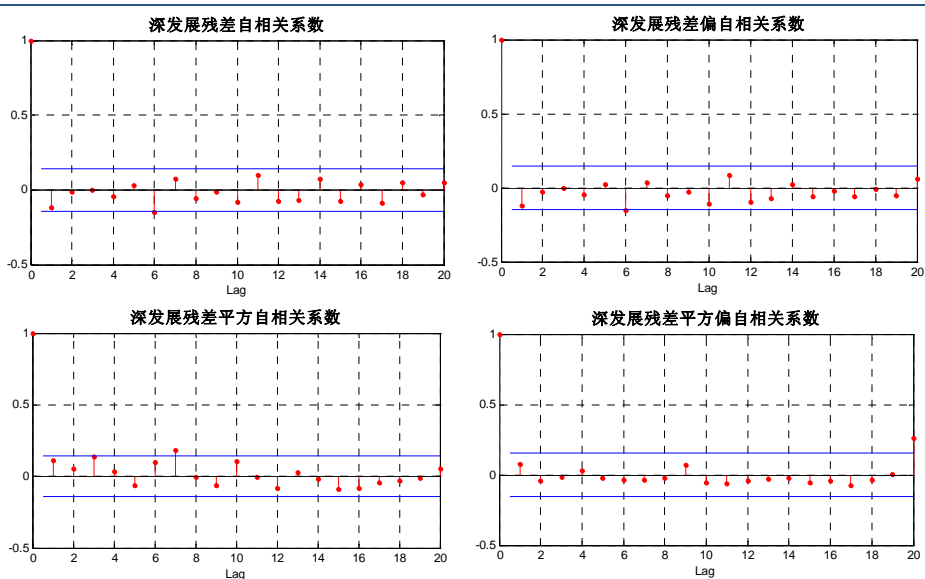
Garch方法的一般假设是序列存在波动相关性。下面给出了沪深300收益率及收益率平方的自相关系数和偏自相关系数、个股（以深发展为例）的回归残差及残差平方的自相关系数和偏自相关系数。可以看出在周数据下，无论是收益率本身还是收益率对三因子做回归之后的残差都没有明显的序列相关性，也没有明显的波动相关性（以序列的平方的相关性表示）。因此我们的Beta估计方法中没有使用Garch方法。

图 4: 沪深 300 周收益率序列及其平方的 ACF/PACF



资料来源: Wind 资讯, 长江证券研究部

图 5: 个股回归残差及其平方的 ACF/PACF



资料来源: Wind 资讯, 长江证券研究部

四、Beta 序列波动性的比较

我们使用上述三种方法计算了 2006 年 1 月 1 日以前上市的全部 A 股 1153 只股票 (剔除了 ST 股、已经退市的股票和停牌时间过长的万方地产) 从 2006 年 1 月到 2010 年 8 月的三因子模型的动态 Beta 值 (每周计算一次)。使用的是周收益率数据。剔除了停牌前后 5

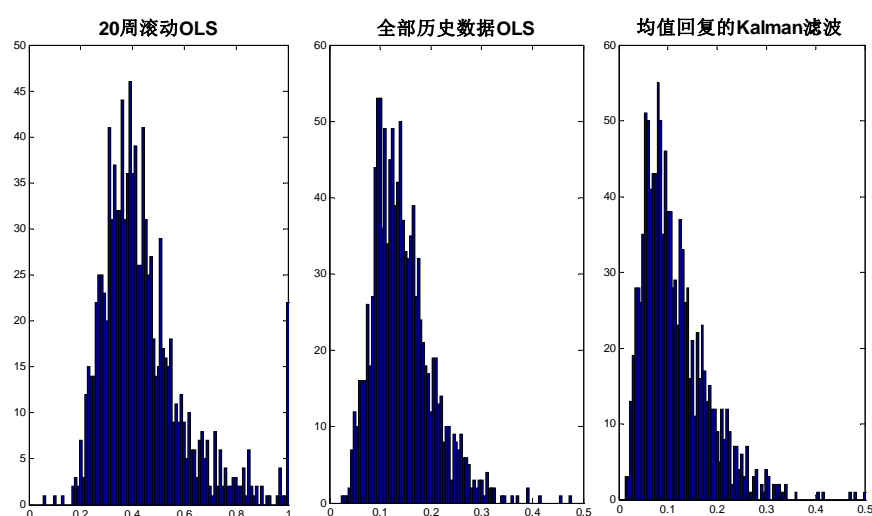
周的数据。

由于计算出来的实际上有 1153×3 个长度为 228 周的 Beta 序列，要将其一一在本报告中展示出来时不切实际的。一个很自然的降维方法就是观察每个 Beta 预测方法的最终投资效果。最终的投资效果也正是我们采用各种方法对 Beta 进行预测的目的。所以我们通过比较最终的策略收益来比较不同预测方法的优劣。

在比较收益之前，我们先从一个侧面来观察一下不同方法预测出的 Beta 的部分性质。

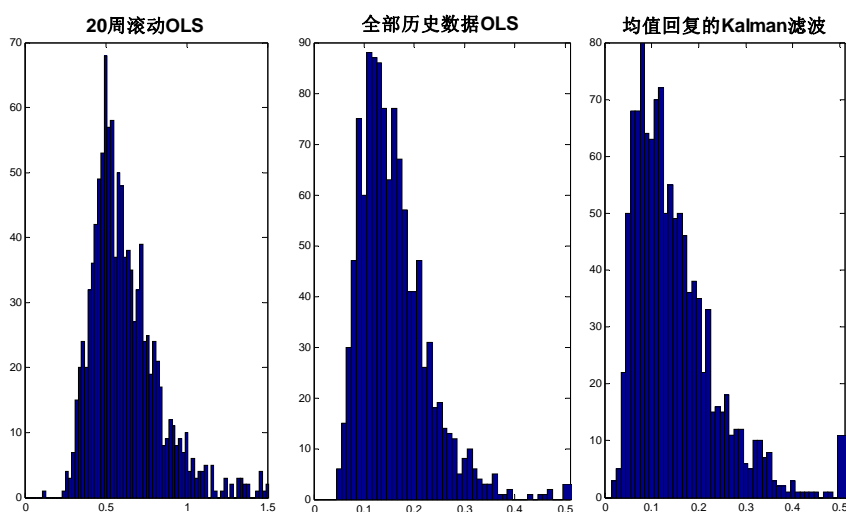
我们对每一个 Beta 序列求标准差，观察标准差对股票的分布。

图 6：市场因子 Beta 标准差直方图



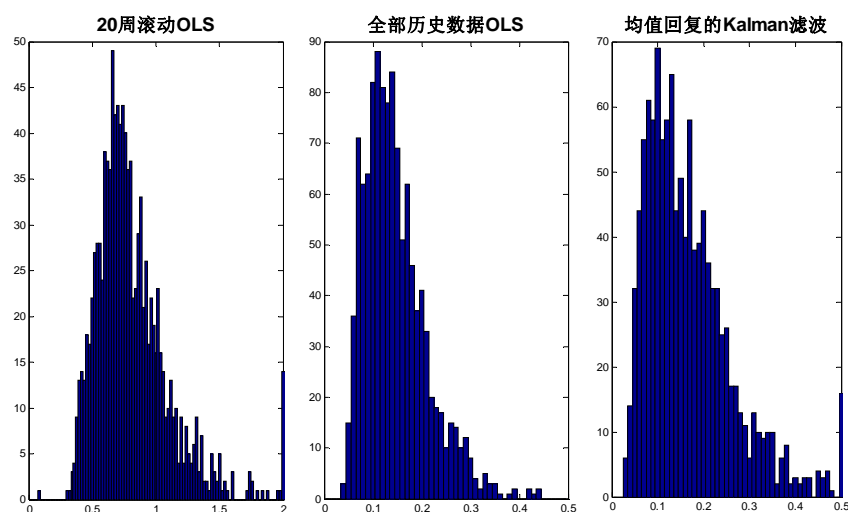
资料来源：Wind 资讯，长江证券研究部

图 7：大小盘因子 Beta 标准差直方图



资料来源：Wind 资讯，长江证券研究部

图 8：市净率因子 Beta 标准差直方图



资料来源：Wind 资讯，长江证券研究部

可以看出，使用 20 周滚动 OLS 方法计算的 Beta 波动性明显比全部历史数据 OLS 和均值回复的状态空间方法大。全部历史数据 OLS 方法和均值回复的状态空间方法获得的 Beta 序列的估计波动性基本一致。不同的是 Kalman 滤波方法获得的 Beta 序列标准差分布要“宽”一些，而全部历史数据 OLS 获得的 Beta 序列标准差分布相对集中。表现在均值回复的状态空间方法获得的 Beta 序列的标准差小于 0.1 的或者大于 0.3 的股票数量都比全部历史数据 OLS 方法多。

滚动 Beta 代表着短期的 Beta 值，它的高波动性预示着 Beta 值是不稳定的，因此全部历史数据 OLS 方法计算出来的“内在 Beta”可能与当前的 Beta 值有较大差异；另一方面，使用当前的 Beta 作为未来 Beta 的估计也不一定能有好效果。均值回复的状态空间方法可以说是对这两种方法的一个折中。我们将以实际的选股效果来进一步评价这几种 Beta 预测方法。

五、多因子选股效果

1、不同因子准确性下的收益

我们用蒙特卡洛模拟的方法构建虚拟的因子预测序列，结合三种方法预测的 Beta，计算了在不同的因子准确性下我们的多因子选股策略能够获得的收益。

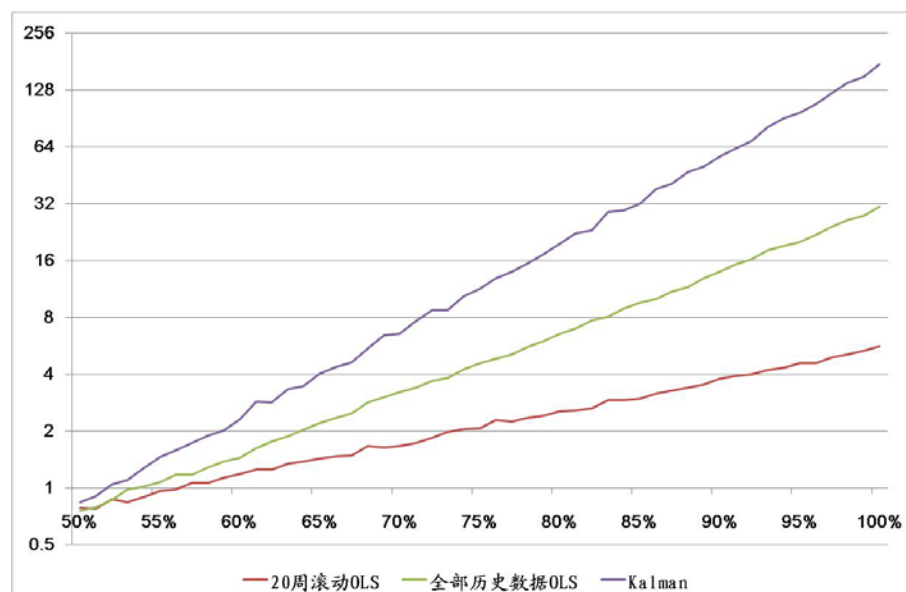
模拟方法的细节如下：

- (1) 2006 年 1 月-2010 年 8 月
- (2) 周调仓
- (3) 因子准确率：由于是回测，因子本来是可以“完美预测”的，为了获得非 100%准确的因子“预测”，我们随机地选取一部分因子，将它改成与实际情况正好相反的“错误值”（即将 1 改成-1，-1 改成 1）。并且进行多次重复实验，最后的收益取所有次实验的收

益的平均值。三个因子的准确性假定是相同的（但选取“错误值”的时候，不同因子之间是独立的）。

- (4) 动态的构建一个包含 20 只股票的组合
- (5) 如《多因子选股系列报告（一）》中所述，构建了 40 只股票的缓冲池以降低换仓成本。
- (6) 交易成本按 1% 计算
- (7) 在每一个准确率下的模拟次数为 100

图 9：多因子选股策略收益-----不同因子准确性



资料来源：Wind 资讯，长江证券研究部

从图中可以看出，随着因子预测准确性的提高，我们的三因子策略可以获得的收益呈指数形式上升。Kalman 滤波方法计算的 Beta 获取的收益明显的高于其他两种方法。

虽然与《多因子选股系列报告之一》中所展示的约 1000 倍的理论收益相比，Kalman 滤波方法的 Beta 预测使得我们可以获取的收益约下降到了理论收益的 20%，但 2006 年 1 月 1 日到 2010 年 8 月总的收益仍有约 200 倍，跑赢大盘约 60 倍。当然这个收益率还是太惊人了。

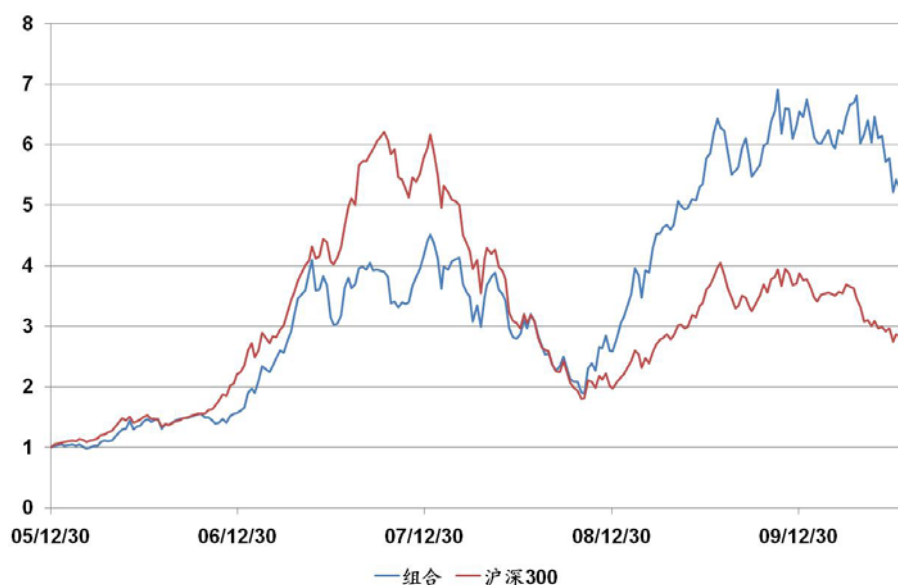
2、预测的 Beta 和实际的因子

我们具体观察预测的 Beta 和实际的因子所构建的组的表现。

使用 20 周滚动最小二乘方法预测的 Beta, 动态的构建包含 20 只股票的组合, 每周调仓。交易成本为 1%, 设置了 40 只股票的缓冲池。周平均换仓为 65.42%。

结果如下图所示:

图 10: 多因子选股收益-----实际因子和 20 周滚动 OLS 的 Beta



资料来源: Wind 资讯, 长江证券研究部

图 11: 多因子选股累积收益倍数-----实际因子和 20 周滚动 OLS 的 Beta



资料来源: Wind 资讯, 长江证券研究部

全部历史数据最小二乘方法预测的 Beta, 动态的构建包含 20 只股票的组合, 每周调仓。交易成本为 1%, 设置了 40 只股票的缓冲池。周平均换仓为 68.90%。

结果如下图所示:

图 12: 多因子选股收益-----实际因子和全部历史数据 OLS 的 Beta



资料来源: Wind 资讯, 长江证券研究部

图 13: 多因子选股累积收益倍数-----实际因子和全部历史数据 OLS 的 Beta

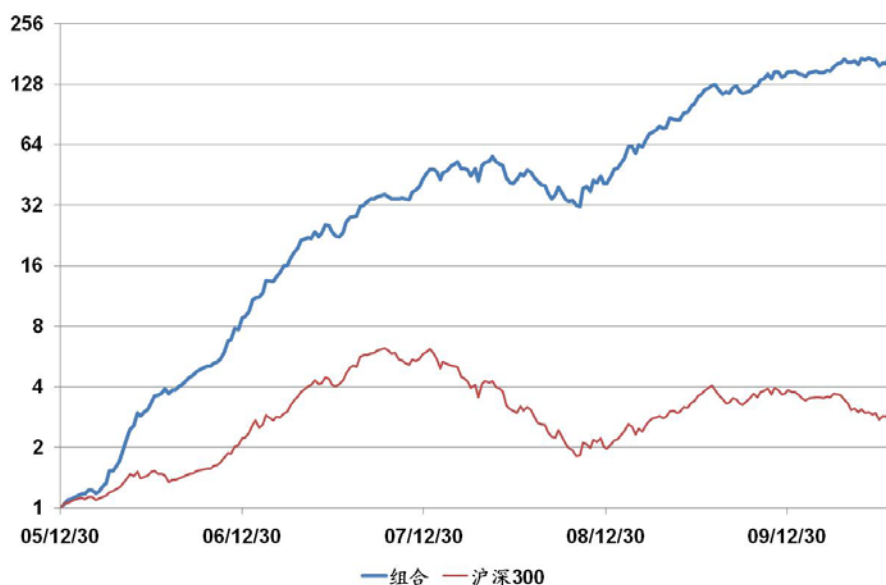


资料来源: Wind 资讯, 长江证券研究部

均值回复的状态空间方法计算的 Beta，动态的构建包含 20 只股票的组合，每周调仓。交易成本为 1%，设置了 40 只股票的缓冲池。周平均换仓为 72.96%。

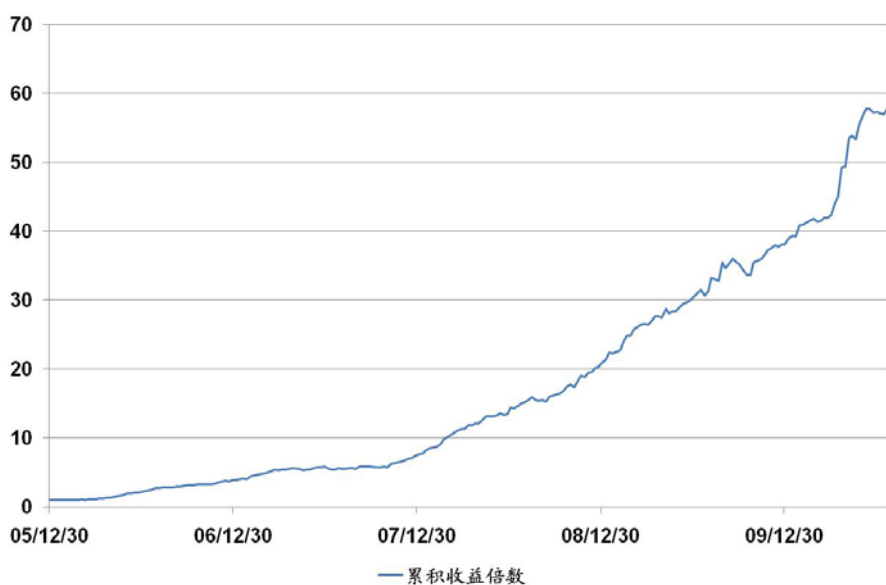
结果如下图所示：

图 14：多因子选股收益-----实际因子和状态空间 Beta



资料来源：Wind 资讯，长江证券研究部

图 15：多因子选股累积收益倍数-----实际因子和状态空间 Beta



资料来源：Wind 资讯，长江证券研究部

2006 年 1 月 1 日至 2010 年 7 月 23 日，在实际的因子下，基于 20 周滚动 OLS 方法、全部历史数据 OLS 方法和均值回复的状态空间方法的三因素选股模型分别获得了 5.61、30.93 和 175.18 倍的绝对收益，并且后两种方法几乎一直是有超额收益的。全部历史数据

请阅读最后一页评级说明和重要声明

OLS 的效果明显优于滚动 OLS 的效果，这也许暗示着“内在 Beta”确实是存在的，至少其变化是缓慢的；均值回复的状态空间方法效果又明显优于全部历史数据 OLS 的效果，表明未来 Beta 值虽然长期来说应等于“内在 Beta”值，但其向“内在 Beta 值”的回复是一个缓慢的过程。

对于多因子选股中的 Beta 预测，我们的探索暂时告一段落。使用预测的 Beta 之后，多因子模型的收益有一定的降低，但仍非常可观。在随后的报告中，我们将继续对多因子选股的其他细节展开讨论，并最终给出一个可操作的多因子选股策略。

写到这里，我们基本上解决了模型构架和 Beta 估计的问题，那么下一篇报告，我们要尝试最难的，也是最重要的因子预测的问题。

分析师介绍

俞文冰，同济大学数学系统计学专业本科，上海财经大学统计学研究生，曾在中海基金负责产品设计，风险控制和金融工程研究工作，目前长江证券研究所担任金融工程研究员，通过 CFA 三级考试

对本报告的评价请反馈至长江证券机构客户部

姓名	分工	电话	Email
伍朝晖	副主管	(8621) 68752398	13564079561 wuzh@cjsc.com.cn
甘露	华东区总经理	(8621) 68751916	13701696936 ganlu@cjsc.com.cn
王磊	华东区总经理助理	(8621) 68751003	13917628525 wanglei3@cjsc.com.cn
鞠雷	华南区副总经理	(8621) 68751863	13817196202 julei@cjsc.com.cn
程杨	华北区总经理助理	(8621) 68753198	13564638080 chengyang1@cjsc.com.cn
张晖	深圳私募总经理	(0755) 82766999	13502836130 zhanghui1@cjsc.com.cn
沈方伟	深圳私募总经理助理	(0755) 82750396	15818552093 shenfw@cjsc.com.cn

投资评级说明

行业评级	报告发布日后的 12 个月内行业股票指数的涨跌幅度相对同期沪深 300 指数的涨跌幅为基准，投资建议的评级标准为： 看好： 相对表现优于市场 中性： 相对表现与市场持平 看淡： 相对表现弱于市场
公司评级	报告发布日后的 12 个月内公司的涨跌幅度相对同期沪深 300 指数的涨跌幅为基准，投资建议的评级标准为： 推荐： 相对大盘涨幅大于 10% 谨慎推荐： 相对大盘涨幅在 5% ~ 10% 之间 中性： 相对大盘涨幅在 -5% ~ 5% 之间 减持： 相对大盘涨幅小于 -5% 无投资评级： 由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级。

重要声明

长江证券系列报告的信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

我公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。

本报告版权仅为我公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处为长江证券研究部，且不得对本报告进行有悖原意的引用、删节和修改。