Barra 模型专题报告（一）：波动率因子

格林大华量化研究
联系电话：0371-56518942

独立性声明:

作者保证报告所采用的数据均来自合规渠道，分析逻辑基于本人的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

摘 要

Barra 模型是 MSCI 明晟公司研发的金融风控模型，把个股收益分解为市场（国家）因素、行业因素、风格因素，其中风格因素即描述个股特征的风险因子，主要应用于多因子选股分析和结构化风险因子分析。2012 年 7 月明晟公司发布了针对中国市场的模型 Barra China Equity Model（CNE5），包括 10 个大类因子、21 个小类因子。2018 年 8 月发布的 CNE6，将因子进一步细分和扩充，包括 9 个一级因子、21 个二级因子、46 个三级因子。

本文主要参考 CNE6 的理论说明书《Barra China A Total Market Equity Model for Long-Term Investors》，因子定义及计算细节摘自附录 E。本文介绍了因子计算方法，用简明的语言解释了计算中的数理逻辑，以沪深 300 指数样本股为范围，回溯了 2018-01-01 至 2022-02-28 的因子数值，通过 IC、IR、因子收益率等指标考量了因子有效性，绘制了因子累计收益率曲线，并对因子回测结果给出了现实意义的解释。

本期分析的因子中，Beta 因子表现平平，可能由于投资者风险偏好较为分散。3 个 Residual Volatility 因子均超越沪深 300 指数，表现良好，值得继续关注。

一. beta，贝塔

1. Historical Beta，历史贝塔

Style:	**Beta**	
Descriptors:	HBETA	Historical Beta Computed as the slope coefficient from a time-series regression of stock excess returns against the cap-weighted excess returns of the estimation universe over a trailing window of 504 trading days, with a 252-day half-life. The returns are aggregated over four-day windows to reduce the effect of non-synchronicity and auto-correlation.

以每只股票的收益率为因变量，以样本空间的流通市值加权收益率为自变量，进行指数加权回归，得到的回归系数即该只股票的贝塔值，其中流通市值加权或流通股本加权计算指数常用的加权方式。可见，贝塔值描述的是的个股涨跌中由于指数涨跌所引起的那部分市场风险因素。

具体来说，收益率为日度超额收益率的移动平均值，即在个股每日收益率的基础上，首先减去无风险收益率，得到超额收益率，其次做间隔为 4 天的移动平均，平滑收益率曲线。计算回归的收益率序列长度为 504 天，采用指数加权，对临近日期的收益率赋予更高的权重，对遥远日期的收益率赋予更低的权重。相较于等权重，指数加权更符合人类记忆中近期清晰远期模糊的特性，而原文中的半衰期（half life）正是调节权重高低的参数。

我们在生活中也许听过半衰期，测定生物死亡年代时，放射性元素碳-14 的原子核逐渐发生衰变，放射强度按指数曲线下降，当原子核有半数发生衰变时所需要的时间即半衰期。而指数加权的权重亦是如此，例如序列长度 504 天，半衰期 252 天，即代表第 t - 251 日的权重恰好衰减为第 t 日的一半。相同的序列长度下，半衰期设置的越短，对临近日期赋予的权重越高，衰减速率也越快，详见附录 3。

如图 1 所示，红柱为指数加权的权重，绿柱为线性加权的权重，前 300 天红柱略小于绿柱；后 200 天红柱又大于绿柱，这是因为指数加权在 252 天的半衰期下衰减速率较低。线性加权的权重只与序列长度有关，可以作为比较不同半衰期下指数加权权重变化的参照对象。

如图 2 所示，半衰期缩短至 126 天时，衰减速率明显加快，前 150 天指数加权的权重较大，随后 300 天指数加权的权重较小，在最后 50 天时由于线性加权权重加快衰减，

指数加权又重新大于线性加权。

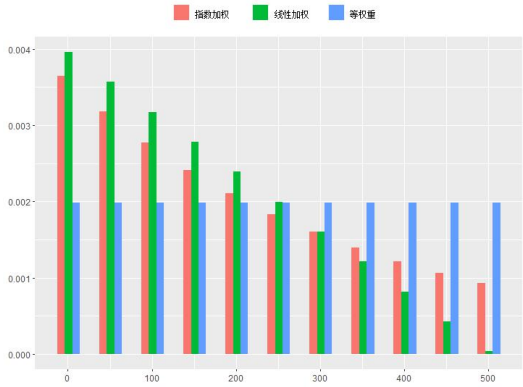

图 1：半衰期 252 天

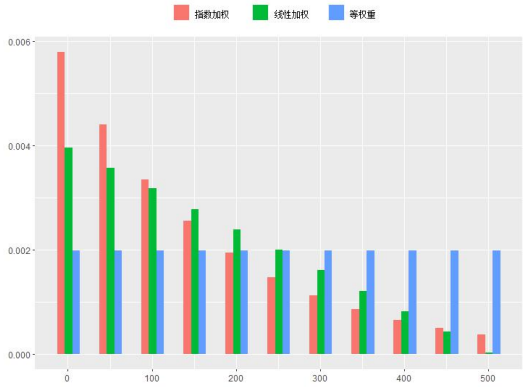

图 2：半衰期 126 天

指数加权回归和指数加权移动平均在 Barra 二级因子残差波动率、动量、长期反转中有广泛的应用，并选取了不同的序列长度和半衰期。通常，更短的半衰期对短期变动更敏感，容易挖掘到转瞬即逝的机会；更长的半衰期得到的结果更稳定，更反映长期趋势。历史贝塔因子选取了 2 年的窗口期和 1 年的半衰期，而且在计算前对收益率序列进行了 4 天平滑，体现了“历史”一词中长期和稳定的含义。

2. Historical Alpha，历史阿尔法

Style:	**Momentum**	
Descriptors:	HALPHA	Historical Alpha The non-lagged Historical Alpha is first computed as the intercept term from the same time-series regression that is used to compute HBETA. The final HALPHA descriptor is then computed as the equal-weighted average of the non-lagged values over an 11-day window lagged by 11 days.

以每只股票的收益率为因变量，以样本空间的流通市值加权收益率为自变量，进行指数加权回归，得到的回归截距即该只股票的阿尔法值。可见，阿尔法值描述的是的个股涨跌中剔除指数相关因素和随机扰动因素后，由于个股自身特性所引起的那部分因素。

具体来说，采用与历史贝塔因子相同的计算方法和参数，首先得到回归的截距项 $alpha_t$；其次将截距项滞后 11 天，即以 $alpha_{t-11}$ 替代 $alpha_t$，以 $alpha_{t-12}$ 替代 $alpha_{t-1}$，以此类推；最后计算新序列间隔 11 天的算术移动平均值。

由于计算历史阿尔法的回归方法中采用了最近 504 天的数据，因此相邻两天的回归函数会涵盖 503 天相同的数据，尽管每天对应的权重有所不同，但是回归的各项结果还

是会产生较高的自相关性。

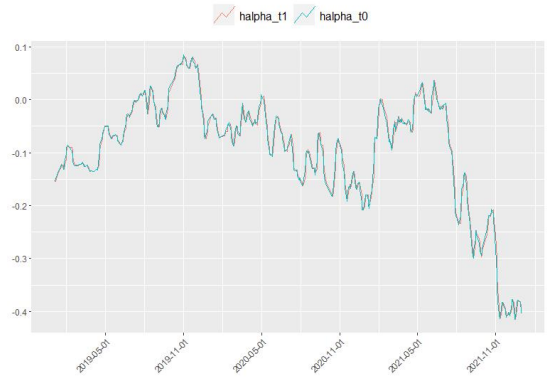

图 3：无滞后的 halpha 序列自相关性

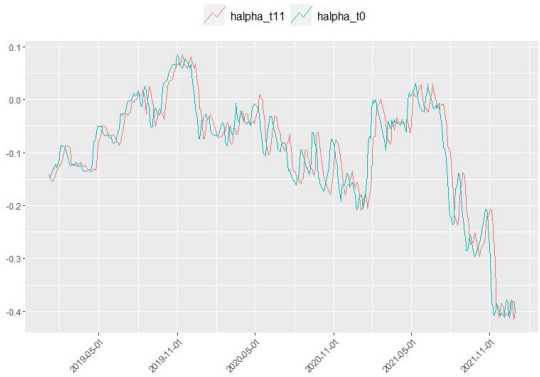

图 4：滞后 11 天的 halpha 序列自相关性

如图 3 所示，红线是原本的 halpha 曲线，绿线是滞后 1 天的 halpha 曲线，相邻两天的红线和蓝线相差无几，有些日期近乎重合。

如图 4 所示，红线是原本的 halpha 曲线，绿线是滞后 11 天的 halpha 曲线，滞后 11 天的红线和蓝线拉开距离，因此同一天的 halpha 数值已经有所区别，数值差距扩大，自相关性降低。

如图 5 所示，红柱是图 3 中两条曲线的同一天的差距，绿柱是图 4 中两条曲线的同一天的差距，滞后 11 天后两条曲线的差距明显增大，同样可以说明自相关性的降低。

如图 6 所示，自相关性随着滞后的天数的增加而逐渐降低。对于沪深 300 指数样本股的历史阿尔法序列，滞后 11 天，平均自相关性可以由 0.9918 降低至 0.8826。

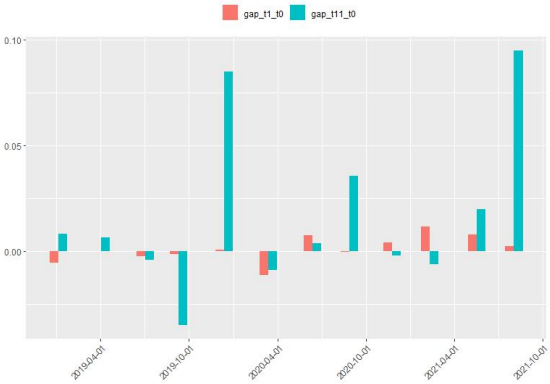

图 5：数值差距扩大

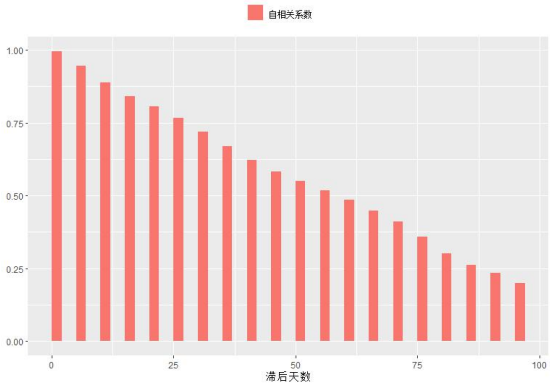

图 6：自相关性降低

在对原始截距项滞后 11 天后，历史阿尔法因子又对新序列进行了 11 天的移动平均，进一步平滑了序列，以期得到更加稳定的因子数值。再加上回归时选取的 2 年的窗口期和 1 年的半衰期，以及在计算前对收益率序列进行的 4 天平滑，可见历史阿尔法因子对

于长期和稳定的重视。

3. Historical Sigma，历史西格玛

Style:	**Residual Volatility**	
Descriptors:	HSIGMA	Historical Sigma Computed as the volatility of the residual returns from the same time-series regression that is used to compute HBETA.

以每只股票的收益率为因变量，以样本空间的流通市值加权收益率为自变量，进行指数加权回归，得到的回归残差再计算标准差即该只股票的西格玛值。可见，西格玛值描述的是随机因素波动的剧烈程度。

样本空间是指考察的所有股票的集合，以样本空间的收益率为自变量，得到的回归结果就不仅仅与个股有关，还与样本空间的选取有关。例如同样一只股票，在所有 A 股、主板或创业板、沪深 300 样本股等不同范围内，得到的回归结果都会不同。

4. 因子回测结果

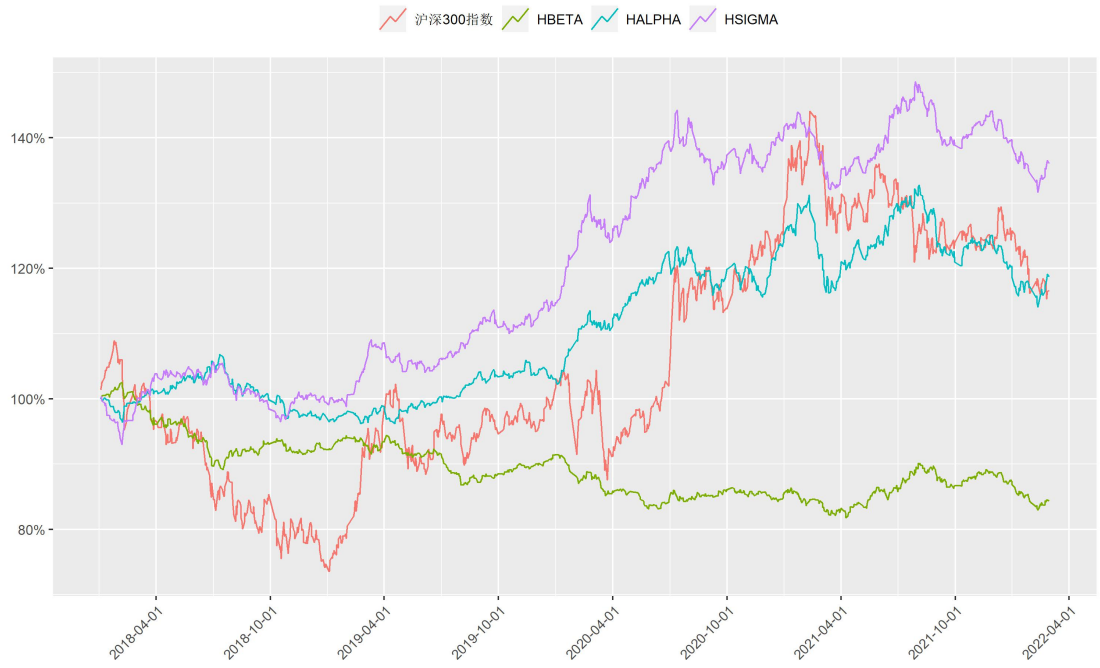

图 7：Beta 因子累计收益率

因子名称	IC	IR	RankIC	RankIC_IR	因子收益率
HBETA	-0.67%	-4.47%	-0.17%	-0.99%	-0.016%
HALPHA	1.03%	5.31%	0.68%	3.24%	0.018%
HSIGMA	1.41%	7.28%	-0.54%	-2.41%	0.032%

表 1：Beta 因子回测结果

如图 7 所示，从因子收益曲线来看，贝塔因子在长期来看比较平稳，波动很小。在 2018 年上半年，随指数同步下跌，收益率在 90%左右企稳。2018 年下半年指数继续下探，2019 年震荡上升，2020 年迅速拉升，2021 年回撤调整。而贝塔因子对于指数这个完整的牛熊周期都缺乏反应，一直平稳的窄幅震荡，将近四年的时间几乎躺平。

贝塔因子的收益率计算中，是买多高贝塔股票，卖空低贝塔的股票，因为高贝塔的股票意味着承担了更多的风险，相应的应该有更高的收益率作为风险补偿。但是在回测中并未展现出这种规律，有两种可能性。第一种是投资者的风险偏好相对分散，没有明显倾向于风险厌恶或风险爱好，因此对于高贝塔的股票不给予风险补偿。第二种是沪深 300 指数的样本股均是 A 股中较优质的公司，他们的风险相对趋同，并没有孰高孰低的区别，因此不能表现出收益率的区别。

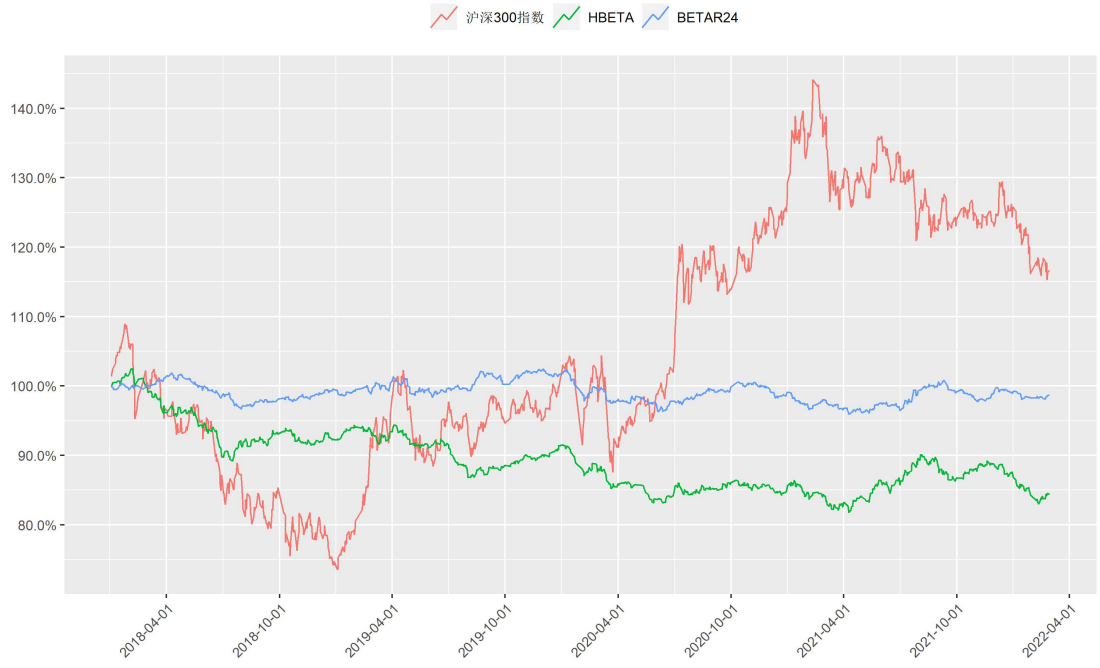

图 8：其他 Beta 因子累计收益率

因此，下载东方财富 Choice 财经数据库 API 中的指标——最近 24 个月 beta 值（BETAR24），然后用同样的方法计算因子收益率和累计收益率。如图 8 所示，BETAR24 因子与 HBETA 因子的走势大体一致，主要区别在于 2018 年上半年指数的快速下跌中，HBETA 因子跌幅较大。HBETA 因子的样本空间为沪深 300 指数样本股，猜测 BETAR24 因子的样本空间为全体 A 股，这可能是两个因子收益差异的来源之一。

二. Residual Volatility 残差波动率

1. Daily Standard Deviation，日超额收益波动率

Style:	**Residual Volatility**	
Descriptors:	DASTD	Daily Standard Deviation
		Computed as the volatility of daily excess returns over the past 252 trading days with a 42-day half-life.

首先将股票的每日收益率，减去无风险利率，得到超额收益率。其次在计算超额收益率方差时，采用指数加权对更近的日期赋予更高的权重。最后得到超额收益率标准差，即超额收益波动率。

具体来说，选择一年期整存整取的定期存款利率作为无风险利率，将其除以 360 天即每日无风险利率。之所以选择一年期利率，是与因子计算中 252 天的窗口期相吻合。而在前述 HBETA 等三个因子的计算中，无风险利率选择两年期利率，也是与 504 天的窗口期相吻合。

回测中随着日期的移动，应当取对应日期的定期存款利率。一是回测中需要尽量还原历史上的各种真实交易环境，不应使用现在已知但在历史那个时点未知的数据；二是如果无风险利率如果是固定值，那么是否减去无风险利率，都不会影响波动率的大小。

加入无风险利率，是将资金的机会成本纳入模型因素，尤其在利率发生变化时，是否考虑无风险利率将对波动率因子产生不同的结果。Barra 模型使用到股票收益率的因子，无一例外的使用了超额收益率，是较为严谨的设计。

另外值得注意的是，HBETA 因子的半衰期是窗口期的 1/2，而波动率因子半衰期设置为 42 天，是窗口期的 1/6，还记得半衰期和权重的关系吗？越短的半衰期将对近期赋予更高的权重，权重的衰减更加快速。这比较符合波动率自身的特性，同时也体现了指数加权参数可调的灵活性。

2. Cumulative Range，累计超额收益离差

Style:	**Residual Volatility**	
Descriptors:	CMRA	Cumulative Range Computed as the gap between the highest and lowest points of the cumulative log excess return in the past 12 months.

首先将股票的每日收益率，减去无风险利率，得到超额收益率。其次以当前日期的一年前为起点，计算最近一年每一天的累计收益率。最后用累计收益率的最大值减去最小值，即得到离差值。

之前的因子使用简单收益率，即收盘价除以昨收价再减 1，对于股票之外的其他资产昨收价可更换为前一期的收盘价。而离差因子使用对数收益率，即收盘价除以昨收价再取对数。

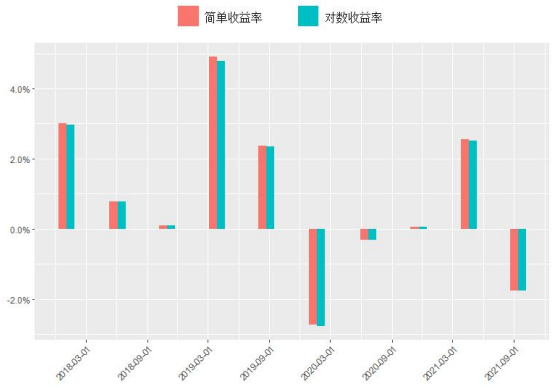

图 9：两种收益率数值差距很小

图 10：简单收益率相加失真

如图 9 所示，两种收益率在数值上差距很小。主要在于计算累计收益率时，如图 10 所示，简单收益率相加会失真，对数收益率直接相加即可。简单收益率需要先加 1，还原为收盘价除以昨收价，然后再相乘，最后再减 1，详见附录 3。Barra 模型使用到累计收益率的因子，均使用了对数收益率。

3. 因子回测结果

如图 11 所示，从因子收益曲线来看，三个残差波动率因子趋势比较一致。2018 年指数下跌时，因子保持在零位小幅震荡，基本未受损失，表现出了良好的抗跌性。2019 年指数收复失地时，因子随之实现 20%左右的盈利。2020 年指数大幅拉升时，因子继续盈利，但盈利幅度不及指数涨幅，被指数逐渐缩小差距。2021 年指数出现调整时，因子

先跌后涨，基本与年初持平，又一次表现出了抗跌性。总体来说，残差波动率因子抗跌性良好，回撤较小，收益不俗。

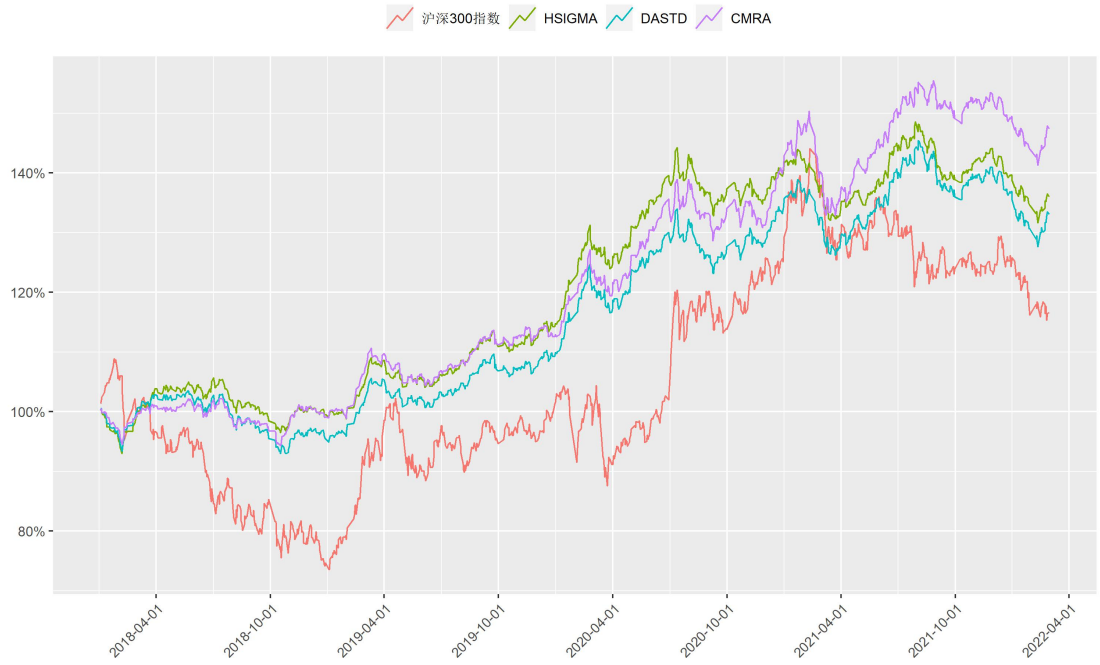

图 11：Residual Volatility 因子累计收益率

因子名称	IC	IR	RankIC	RankIC_IR	因子收益率
DASTD	1.38%	6.55%	-0.34%	-1.40%	0.030%
CMRA	1.93%	9.27%	-0.06%	-0.26%	0.040%

表 2：Residual Volatility 因子回测结果

残差波动率因子的收益率计算中，是买多残差波动率高的股票，卖空残差波动率低的股票。因为残差意味着排除了股票的市场风险和个股特性后，对股票的定价偏差，这种偏差除了非理性和随机性因素外，还包含了投资者对于个股认知的分歧程度。因此残差波动率高可以理解为高估和低估的两方力量相互角力、反复争夺的过程，最终胜利的一方取得了不错的收益，从这个角度也许可以解释残差波动率因子的优秀收益。

附录 1　Barra 因子结构

Barra 模型是 MSCI 明晟公司研发的金融风控模型，把个股收益分解为市场（国家）因素、行业因素、风格因素，其中风格因素即描述个股特征的风险因子，主要应用于多因子选股分析和结构化风险因子分析。第一个为美国市场设计的多因子模型于 1975 年发布，称为 Barra US Equity Model1（USE1），USE2 于 1985 年发布，USE3 于 1997 年发布，USE4 于 2011 年发布。Barra 模型经受了众多金融机构的实证检验，经久不衰，生机蓬勃。

2012 年 7 月明晟公司终于发布了针对中国市场的模型 Barra China Equity Model（CNE5），包括 10 个大类因子、21 个小类因子。2018 年 8 月发布的 CNE6，将因子进一步细分和扩充，包括 9 个一级因子、21 个二级因子、46 个三级因子。9 个一级因子分别是波动率（Volatility）、动量（Momentum）、价值（Value）、规模（Size）、流动性（Liquidity）、质量（Quality）、成长（Growth）、分红（Dividend Yield）、情绪（Sentiment）。

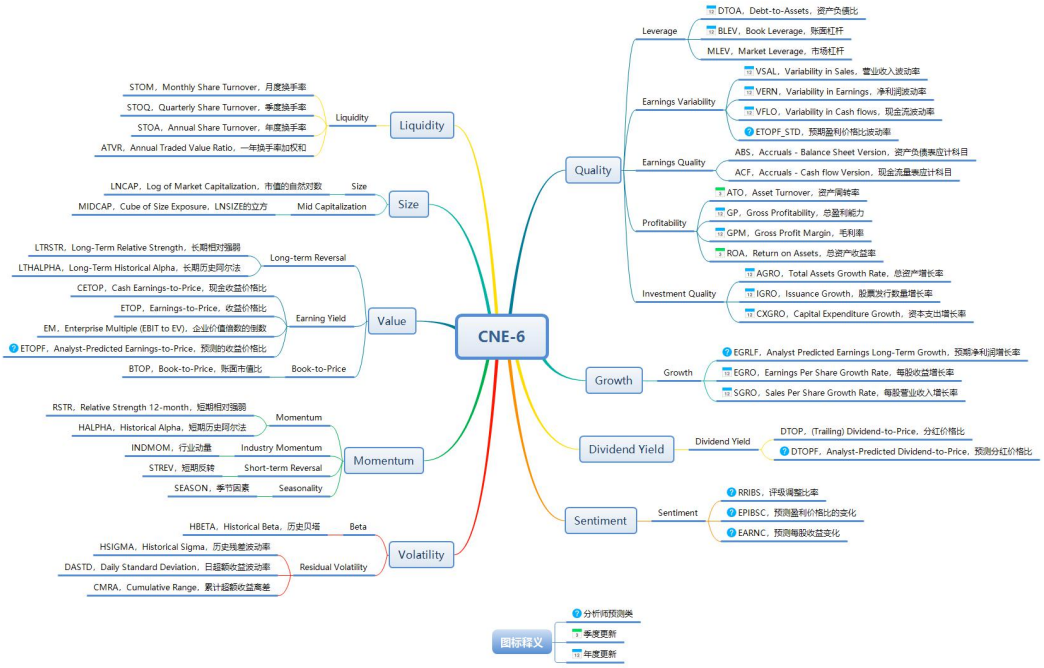

图 12：CNE6 因子结构

附录 2　因子评价指标

一．IC（Information Coefficient，信息系数）

IC 值是指对于一揽子股票和某一个因子，当期的因子值序列与下期的收益率序列之间的 Pearson 线性相关系数，正相关性越高，说明该因子的数值对于股票收益率的预测能力越好，因子的有效性越强。当然，如果呈负相关性，绝对值越大越好，即该因子是反向因子，对股票收益率同样具有良好的预测能力。

日度更新因子在历史上的每一天都有一个 IC 值，在回测区间的若干天内对若干个 IC 值取平均数，可以得到因子的 IC 均值，用来描述该因子在回测区间内的有效性，通常认为 IC 均值大于 3%或者小于-3%时因子比较有效。

由于 IC 值直接使用数值计算相关系数，就难免受到异常值和量纲的影响，因此需要对原始的因子值进行缩尾化和标准化。缩尾化是指以均值加减三倍标准差为边界，截断边界以外的部分，保留边界以内的部分，即将异常值用边界值替代。标准化通常将原序列减去均值再除以标准差，使新序列的均值为 0 且标准差为 1。这样不同因子间就排除了异常、统一了量纲，可以在同一水平下进行比较。 另外，因子原始数值的缩尾化和标准化，是以日期为截面，即对同一日期不同股票的因子值进行数值变换，这样可以保证当期的因子风险敞口为 0。

由于 Pearson 相关系数，假设序列服从正态分布，而很多情况下的数据难以满足这个假设。而且 Pearson 相关系数只能描述线性相关性，对于高度非线性相关的序列也只能得到较低的线性相关性。基于以上两点局限性，Spearman 相关系数先将数值排序，再计算排名之间的相关性，可以得到很好的结果。

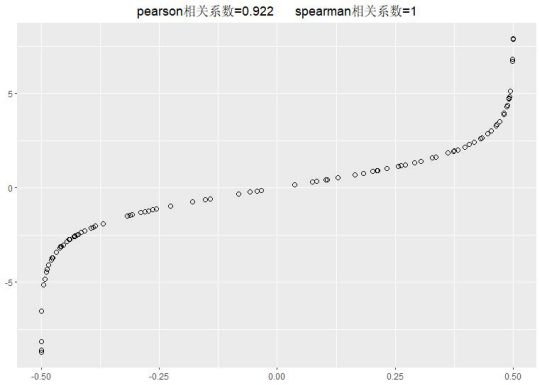

图 12：完全相关时的相关系数

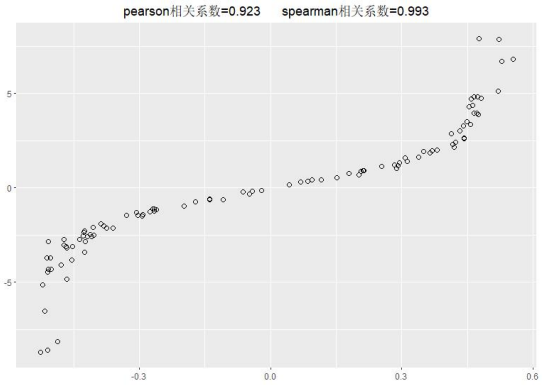

图 13：微小扰动时的相关系数

如图 12 所示，逻辑曲线在中间缓慢增长，在两边迅速增长，符合很多事物分阶段变化的规律。图中曲线序列 Y 完全由序列 X 决定，是完全相关的关系，但是 Pearson 相关系数只有 0.922，而 Spearman 相关系数为 1，准确描述了这种完全相关关系。

如图 13 所示，在图 12 完全决定的关系上施加微小的随机扰动后，Spearman 相关系数仍然高达 0.993，说明在更加真实的环境中，Spearman 相关系数也能得到符合客观规律的结果。

Rank_IC 值是指在 IC 值的计算中用排序代替数值，即因子值的排序和收益率的排序之间的相关关系，更准确的说是因子值和收益率的 Spearman 秩相关系数，是对 IC 值的一种改进。

二. IR（Information Ratio，信息比率）

IR 值是指回测区间内的 IC 均值除以 IC 标准差，即在原有的 IC 均值评价基础上，对于 IC 值波动较大的因子增加惩罚，提高 IC 值较为稳定因子的评价。因此可以看做考虑了风险因素后，对 IC 均值的优化。

然而由于 IC 值大于 3%或者小于-3%都说明因子比较有效，IC 值的绝对值较大直观上会导致波动较大，而 IR 值又会惩罚波动较大的因子，似乎 IR 值会将有效的因子变的平庸。其实不然，IR 值只是会惩罚 IC 值在大于 3%和小于-3%之间多次反转的因子，而这种反转代表着因子的失效以及买卖方向的错误。对于 IC 值稳定的大于 3%的因子，或者稳定的小于-3%的因子，IR 值是不会加以惩罚的。通常认为 IR 大于 5%因子比较有效。

因此，评价一个因子的有效性，除了 IC 值，还需要观察 IR 值，只有两者都给出有效性较高的结论时，才能说明因子是真正有效的。

RankIC_IR 值是指回测区间内的 Rank_IC 均值除以 Rank_IC 标准差，由于 RankIC 值是对 IC 值的一种改进，RankIC_IR 值也是对 IR 值的一种改进。

三. 因子收益率

因子收益率是指对于一揽子股票和某一个因子，首先对原始的因子值进行缩尾化和标准化。然后将当期的因子值序列与下期的收益率序列做内积，即对应元素相乘再求和。换一个角度理解，以因子值作为权重买入股票，因子值为负的卖空股票，买入权重乘以

下期的收益率即该股的盈亏，因子值和收益率都为正则为盈利，因子值为正而收益率为负则亏损，再将所有股票的盈亏求和，即整个交易的总盈亏，以此来衡量当期因子值对下期收益率的预测能力，因子收益率越高则因子越有效。

日度更新因子在历史上的每一天都有一个因子收益率，在回测区间的若干天内对若干个因子收益率取平均数，可以得到因子收益率均值。还可以从回测开始之日起累计每天的因子收益率，得到因子收益率曲线，观察因子有效性随时间变化的走势。

因子收益率只是一个因子评价指标，并未考虑实际交易中的买入卖出手续费，而模拟的交易中又是当日买入次日卖出的较为频繁交易。在实际交易中可以买入前 20%的股票，并减去手续费，得到较为真实的收益率曲线。在样本空间的股票数较多时，还可以将股票按因子数值从高到低分为五档，分别计算每天每档的因子收益率，再合成每档的累计收益率曲线，看收益是否仍然呈现从高到低的排列，且有比较明显的差异。

附录 3 因子计算中的数理推导

一. 半衰期和指数加权

半衰期可以用数学语言表述为公式 1，即以 1 - α 的速率经过 h 天后权重按指数曲线衰减为初始值的一半。因此半衰期 h 和衰减速率α之间存在一一对应关系，即公式 2，通常是在说明时给定半衰期 h，在计算中使用衰减速率α。

$$(1-\alpha)^h = 0.5 \qquad \text{（公式 1）}$$

$$h = \frac{\ln(0.5)}{\ln(1-\alpha)}$$

$$\alpha = 1 - e^{\frac{\ln(0.5)}{h}} \qquad \text{（公式 2）}$$

指数加权移动平均还可以使用迭代式表示为公式 3，即当日的 EMA 可以通过当日的收益率和前一日的 EMA 得到。迭代式在计算中只需要两个数据，不需要整个收益率序列，因此可以节省存储空间。

$$\begin{aligned} EMA_t &= \alpha \cdot r_t + (1-\alpha)\cdot EMA_{t-1} \\ &= \alpha \cdot r_t + (1-\alpha)\cdot\alpha\cdot r_{t-1} + (1-\alpha)^2 \cdot EMA_{t-2} \\ &= \alpha \cdot r_t + (1-\alpha)\cdot\alpha\cdot r_{t-1} + (1-\alpha)^2 \cdot \alpha \cdot r_{t-2} \ldots\ldots + (1-\alpha)^N r_N \end{aligned} \qquad \text{（公式 3）}$$

二. 对数收益率和简单收益率

相隔 1 期的对数收益率定义为公式 4，计算相隔 N 期的累计收益率时，直接相加即可，运算方便，如公式 5 所示。

$$r_t = \log\left(\frac{price_t}{price_{t-1}}\right) \qquad \text{（公式 4）}$$

$$\begin{aligned} \sum_{t=1}^{N} r_t &= \log\left(\frac{price_t}{price_{t-1}}\right) + \log\left(\frac{price_{t-1}}{price_{t-2}}\right) + \ldots\ldots + \log\left(\frac{price_{t-N+1}}{price_{t-N}}\right) \\ &= \log\left(\frac{price_t}{price_{t-1}} \times \frac{price_{t-1}}{price_{t-2}} \times \ldots\ldots \times \frac{price_{t-N+1}}{price_{t-N}}\right) \\ &= \log\left(\frac{price_t}{price_{t-N}}\right) \end{aligned} \qquad \text{（公式 5）}$$

相隔 1 期的简单收益率定义为公式 6，计算相隔 N 期的累计收益率时，直接相加会失真，需要先加 1，还原为收盘价除以昨收价，然后再相乘，最后再减 1，运算复杂，如公式 7 所示。

$$R_t = \frac{price_t - price_{t-1}}{price_{t-1}} = \frac{price_t}{price_{t-1}} - 1 \qquad \text{（公式 6）}$$

$$\prod_{t=1}^{N}(R_t+1) = \frac{price_t}{price_{t-1}} \times \frac{price_{t-1}}{price_{t-2}} \times \ldots\ldots \times \frac{price_{t-N+1}}{price_{t-N}} = \frac{price_t}{price_{t-N}} \qquad \text{（公式 7）}$$

重要声明

本报告中的信息均来源于公开资料，我公司对这些信息的准确性及完整性不作任何保证，不保证报告信息已做最新变更，也不保证分析师作出的任何建议不会发生任何变更。在任何情况下，报告中的信息或所表达的意见并不构成所述期货品种买卖的出价或询价。在任何情况下，我公司不就本报告中的任何内容对任何投资作出任何形式的担保，投资者据此投资，投资风险自我承担。我公司可能发出与本报告意见不一致的其它报告，本报告反映公司分析师本人的意见与结论，并不代表我公司的立场。未经我公司同意，任何人不得对本报告进行任何形式的发布、复制或对本报告进行有悖原意的删节和修改。