

数据纵横：拓展差分操作域

——机器学习系列报告之三

金融工程深度

本篇报告将探讨在运用序列数据的过程中，不同差分操作对数据在模型有效性上带来的影响；利用后移算子将差分概念的操作域从正整数拓展到正实数，并给出一些非整差分的应用。

- ◆ **差分操作域可以从正整数拓展到正实数域。**在直观逻辑上，序列差分的阶数仅能是正整数，而通用利用后移算子与级数展开，可以将操作域扩大到正实数。
- ◆ **大部分指数与股票价格序列无须一阶差分即可满足平稳性要求。**很多序列差分的目的是为了满足不同平稳性条件，但过度差分会使得数据丧失序列记忆性。通过对各个指数与股票的测试，可以发现指数大多仅需 0.7 阶差分即可满足平稳性要求，此时差分序列与原序列仍有 0.9 左右相关性。而股票的要达到平稳性所需的最低差分阶数差异较大，但大多数也都不需要至 1 阶差分。
- ◆ **非整差分下的信息熵特征效果更好。**运用不同差分阶数下的价格数据计算信息熵效果不同，差分阶数较小时保留信息更多。用 0.7 阶差分价格对数序列计算的信息熵数据配合 RSRS 择时策略能小幅提升信号效果。
- ◆ **利用非整差分能用以优化传统量价因子。**用收益率标准差计算的低波因子是常用有效 α 因子。我们通过测试不同差分下的价格对数序列标准差来尝试优化低波因子效果。经测试，最优差分阶数并非默认的 1 阶（收益率）而在 1.4 至 1.5 之间。在 1.5 阶差分下计算的低波因子，在全市场、中证 500、沪深 300 股票池内选过效果均有明显提升。
- ◆ **风险提示：**测试结果均基于模型和历史数据，模型存在失效的风险。

分析师

刘均伟(执业证书编号：S0930517040001)

021-52523679

liujunwei@ebcn.com

联系人

胡骥聪

021-52523683

hujicong@ebcn.com

相关研报

《机器学习：开拓金融量化新前沿——机器学习系列报告之一》2019.01

《数据纵横：探秘 K 线结构新维度——机器学习系列报告之二》2019.02

目 录

1、 时间序列上预测信息与可推论性的悖论	5
1.1、 被牺牲的预测信息.....	5
1.2、 再看作为输入特征的价格与收益率	5
2、 探寻记忆性与平稳性的平衡	6
2.1、 差分与非整差分	6
2.2、 非整差分下价格序列的记忆性与平稳性.....	6
3、 非整差分的一些简单应用	9
3.1、 不同差分程度下价格序列的信息熵	9
3.2、 波动因子构造的延伸	12
4、 风险提示.....	17
5、 附录	17

图目录

图 1：上证综指差分测试.....	7
图 2：上证 50 差分测试	7
图 3：沪深 300 差分测试.....	7
图 4：中证 500 差分测试.....	7
图 5：上证综指价格对数 0.3 阶差分下单年测试 ADF 统计值.....	8
图 6：上证综指价格对数 0.7 阶差分序列与上证综指价格对数比较.....	8
图 7：不同股票在各自最小差分阶数下与原始序列相关性箱型图.....	9
图 8：收益率序列在分位数编码下的信息熵	11
图 9：0.7 阶差分序列在分位数编码下的信息熵.....	11
图 10：收益率序列在等距编码下的信息熵	12
图 11：0.7 阶差分序列在等距编码下的信息熵.....	12
图 12：上证综指 RSRS 策略配合信息熵指标效果	12
图 13：沪深 300 指数 RSRS 策略配合信息熵指标效果	12
图 14：主流低波因子在不同样本长度参数下的 IC 与 IR 表现.....	13
图 15：主流低波因子 (STD(10)) 在不同差分阶数参数下的 IC 与 IR 表现.....	14
图 16：全市场股票池下不同波动因子多空净值.....	16
图 17：中证 500 股票池下不同波动因子多空净值	16
图 18：沪深 300 股票池下不同波动因子多空净值	16
图 19：全市场股票池下不同波动因子多头组合	17
图 20：中证 500 股票池下不同波动因子多头组合	17
图 21：沪深 300 股票池下不同波动因子多头组合	17

表目录

表 1：价格序列差分测试设定	7
表 2：不同宽基指数达到平稳条件所需差分阶数大致相同	9
表 3：价格序列差分测试设定	11
表 4：因子测试设置	13
表 5：各股票池内不同波动因子构造的效果差异	15
表 6：不同低波因子构造下多空组合统计表现	15
表 7：不同低波因子构造下多头组合统计表现	16
表 8：不同股票整体最小差分阶数与年度最小差分阶数	18
表 9：波动因子在不同样本长度与差分阶数参数下的 IC 均值	19
表 10：波动因子在不同样本长度与差分阶数参数下的 IR 值	19

1、时间序列上预测信息与可推论性的悖论

“历史会重复”。无论是传统量化，亦或机器学习，都是基于这样一个假设在探索金融领域下每一个可能获利的角落。因此要构建一个成功的模型，算法的有效性与历史数据信息，缺一不可。然而在实际研究中容易发现，要提升其中一方面往往要以削弱另一方面为代价。在该篇报告我们研究讨论上述现象和对这种现象的处理手段。

1.1、被牺牲的预测信息

数据本身是否含有预测信息是我们在做预测时最关心的的一个特质。所谓“垃圾进、垃圾出”，如果数据本身完全不具备预测信息，那么无论用什么模型或方法，都不能获得有价值的预测模型。而在满足一定预测信息的条件下，另一个会影响到最终预测效果的因素在于这些预测信息能多大程度上被模型利用。如果要处理的是金融中常见的时序数据，从模型推论性的角度出发，我们希望模型中的数据为平稳序列。更进一步，倘若模型是机器学习算法，那么这个需求就更为迫切。

然而众所周知，绝大部分的金融时序数据都是不平稳的，无论是股票的价格、期权的隐含波动率、亦或上市公司的季度收入等财报数据。因此在运用以上各种金融时序数据前，往往需要进行各种预处理操作，比如去周期化、作差分等等，使其最终达到一定程度上的平稳性。然而这种获取平稳性的方式并非完全没有缺点，它会使得每个数据点丢失之前序列所蕴含的信息或记忆。如果丢失的内容是噪声，那么处理后的数据不仅获得了平稳性还提高了信噪比。可并非在所有场景下都会有这么理想的情形，绝大多数情况下丢失的恰恰是更为重要有用的预测信息。

1.2、再看作为输入特征的价格与收益率

作为最直接体现市场参与者观点的交易数据，同时也是大部分预测模型的预测目标，股票价格序列一直是金融业者的重点研究对象。每一个新的价格都是在上一个价格的基础上融入了新的市场信息后的交易结果，因此当前价格除了体现新入信息以外，另一部分还保留了上一个价格的信息或者说是记忆。但正如上一节所论述，价格序列不平稳，例如一个正常运作不断产生价值的公司，它的股票价格大概率会时不时产生新高。在大部分统计或机器学习模型里，如果直接把价格数据作为输入特征或预测变量，最终得到的模型往往在股票创新高或创新低时就完全失去了推论（预测）能力。

将价格对数序列做一阶差分，得到了另一个我们关心的时间序列：收益率。它比价格序列有更有利的平稳性条件，但同时它仅有最新交易的信息，一阶差分操作把之前价格序列的历史信息抹去了。有时我们正需要这些历史信息来对未来做预测，这时虽然模型对数据有较好的推论能力，但一阶差分后的数据已经丧失了大部分的预测信息。

2、探寻记忆性与平稳性的平衡

如果数据序列记忆性与平稳性不可兼得，那无论是哪个极端——为了信息而放弃平稳性或为了平稳性而放弃全部信息——都不太可能是最理想的数据状态。那是否可以在两个极端之间，找到一个点，在这一点上时间序列在保持平稳性特征的同时，每个数据点都尽可能地保留足够多的信息呢？在这一章我们借鉴 Macros Lopez 在《Advances in Financial Machine Learning》一书中提出的非整差分概念，探索我国 A 股指数与股票价格（或者说价格对数）序列记忆性与平稳性的平衡。

2.1、差分与非整差分

无论是经典的计量经济学模型或主流统计模型，对于最终运用算法的数据都有平稳性要求。金融数据大多都是不平稳的，因此一些预处理操作是运用算法前的必要手段。最为主流常见的方式有两种：一种是协整（cointegration），将多个不平稳时间序列整合成一个平稳序列，配对交易（pair-trading）正是基于此构建而生，但金融市场中能存在长期稳定协整关系的序列少之又少，因而协整方式能适用的场景其实很小；而另一种就是更为简单通用的差分，“没有什么平稳条件是做一次差分解决不了的，如果有，那就再差分一次”。

差分的逻辑与运算操作简单直白，但差分的阶数只能是正整数。这就意味着最小的差分程度就是 1 阶差分，而从第一章的探讨我们知道即使是 1 阶差分也可能损失太多的信息或记忆。为了处理这个问题，引入非整差分的概念，将差分的操作域从整数域扩充到正实数域就很重要。

非整差分的逻辑原理从直观想象上难以理解，而从后移算子 B (backshift operator) 的角度则能形成一个自洽的解释。后移算子 B 是针对时间序列的运算符，其规则为对于序列 X_t ， $B^k X_t = X_{t-k}$ 。那么相应的一阶差分就是 $(1-B)X_t = X_t - X_{t-1}$ ，二阶差分就是 $(1-B)^2 X_t = X_t - 2X_{t-1} + X_{t-2}$ 。

沿着上述规则，对于任何正实数 d ， d 阶差分就是 $(1-B)^d X_t$ ，通过级数展开 $(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$ 即可算出序列的 d 阶差分。当 d 非整数时，级数展开后是一个无穷级数，很容易验证这些无穷级数的收敛性，因此在实际计算中，只需取级数足够多的前一部分进行求和计算即可：
$$(1-B)^d \approx \sum_{k=0}^N \binom{d}{k} (-B)^k。$$

2.2、非整差分下价格序列的记忆性与平稳性

探究价格序列差分程度的平衡点，我们测试了不同宽基指数与沪深 300 成分股中于 2005 年之前发行且尚未退市的股票在经过不同程度差分后的统计特征：记忆性的定义方法为：差分后序列与原始序列的相关性，相关性越大说明保留的记忆与信息越多；平稳性以 ADF 检验（单位根检验）检查其统计值是否能在 95% 置信程度下拒绝平稳假设。

通过这样的测试，我们试图发现在基本满足平稳性条件下，不同价格序列实际需要的差分程度到底有多高。在尽可能不过分差分的情况下，原序列的记忆能保留多少。

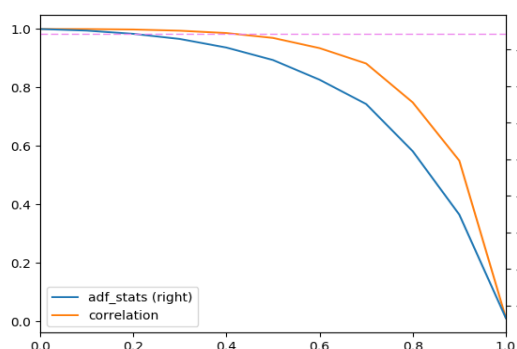
表 1：价格序列差分测试设定

设定项	设定值
原始序列	价格序列对数
数据区间	2005/01/01 – 2018/12/31
差分程度最小间隔	0.1
记忆性检验	相关系数
平稳性检验	ADF 检验

资料来源：光大证券研究所

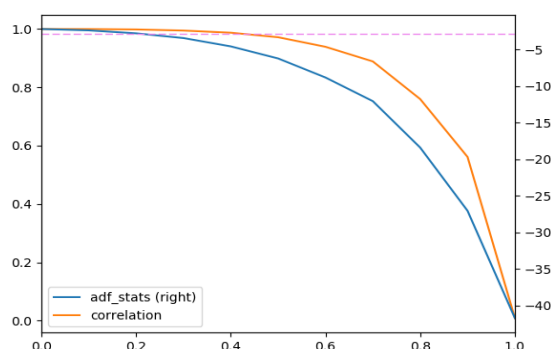
首先在不同宽基指数的测试表明：**A 股大部分宽基指数所需要最小差分程度较为类似，皆在 0.3；此时差分序列与原始序列的相关系数都在 0.99 以上，每个数据点几乎保留了所有前序序列的记忆，而不用像收益率序列那样仅保留前一个点的记忆。差分序列与原始序列之间的相关性并非随着差分程度线性变化，而是更偏向于指数变化，在 1 阶差分下相关系数基本为 0，而在 0.9 阶差分下仍有 0.5 左右的相关性。**

图 1：上证综指差分测试



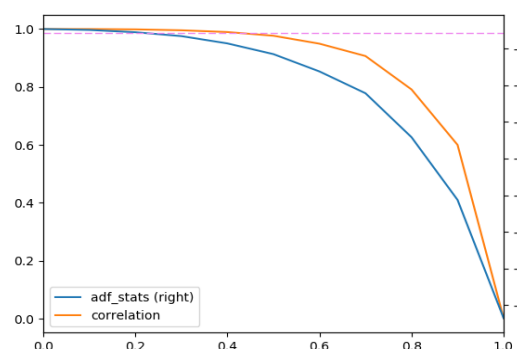
资料来源：光大证券研究所，Wind

图 2：上证 50 差分测试



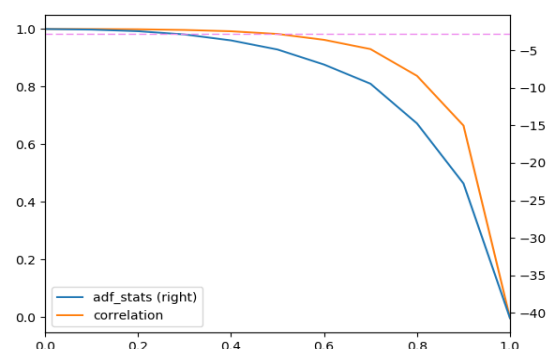
资料来源：光大证券研究所，Wind

图 3：沪深 300 差分测试



资料来源：光大证券研究所，Wind

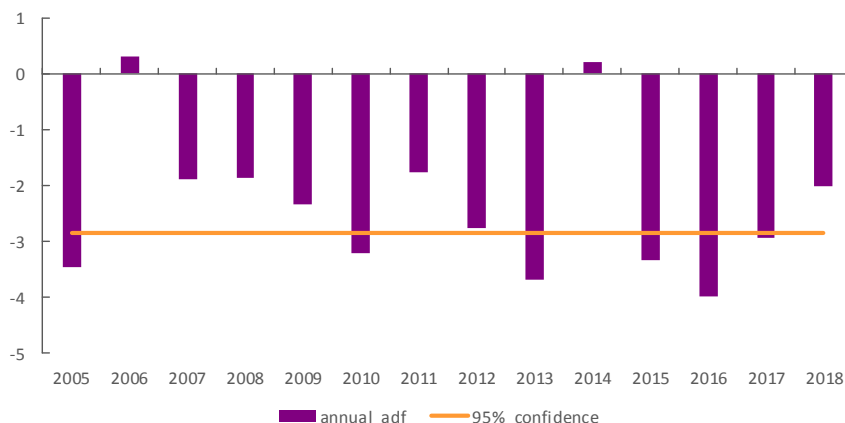
图 4：中证 500 差分测试



资料来源：光大证券研究所，Wind

虽然只需要 0.3 阶差分就能使得整体序列基本满足平稳要求，然而指数价格序列在不同时期的特征不同。有的区间下是单边上涨牛市，有的是震荡下跌的慢熊。过低的差分阶数并不能保证在末段市场环境下依然较好满足平稳性假设。我们计算宽基指数在 0.3 阶差分下的价格对数序列在不同年份及时长下的 ADF 统计值，发现大多以年为单位的单边趋势行情下，0.3 阶差分远远不能满足平稳性要求。下图仅展示上证综指单年的平稳性测试结果，可以看出在 0.3 阶差分下仅不到一半的年份通过平稳性检验。

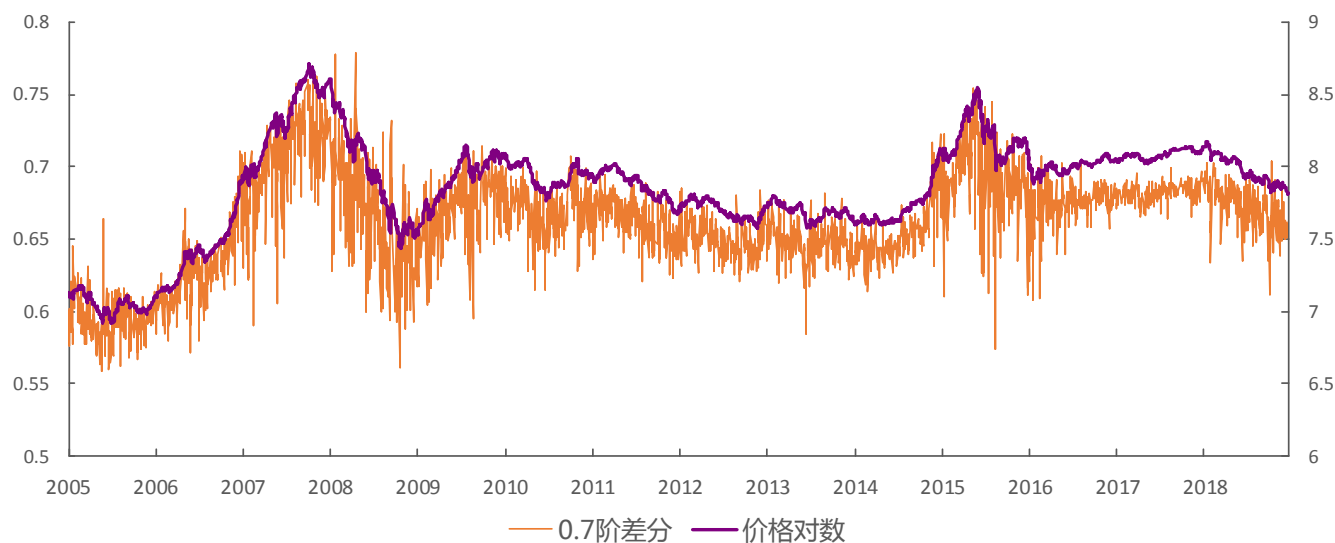
图 5：上证综指价格对数 0.3 阶差分下单年测试 ADF 统计值



资料来源：光大证券研究所，Wind

而若要上证综指价格对数在所有年份及时长下都满足平稳性检验，则需要最少 0.7 阶的差分，此时差分序列与原始序列的相关系数仍有 0.88。同样上证 50、沪深 300 与中证 500 也都在差分阶数为 0.7 时满足年度平稳性条件。而与原始序列的相关性也都在 0.9 左右。

图 6：上证综指价格对数 0.7 阶差分序列与上证综指价格对数比较



资料来源：光大证券研究所，Wind； 注：纵坐标左轴为 0.7 阶差分坐标，纵坐标右轴为价格对数坐标

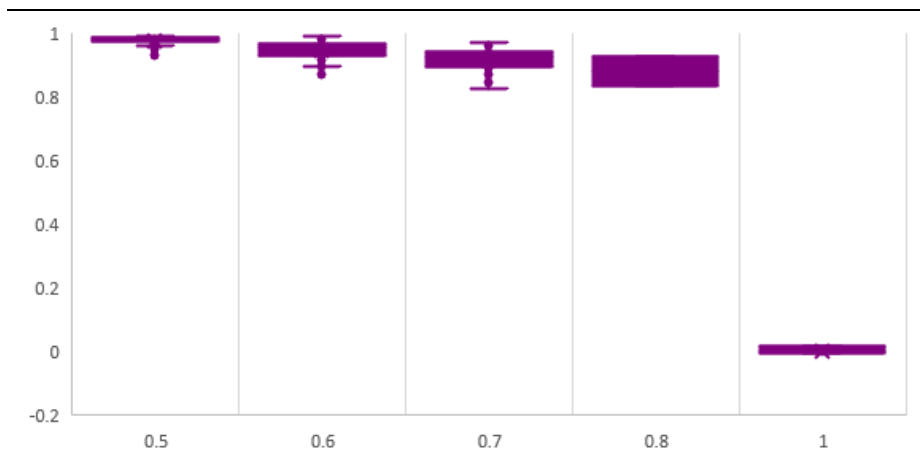
表 2：不同宽基指数达到平稳条件所需差分阶数大致相同

	上证综指	上证 50	沪深 300	中证 500
整体最小差分阶数	0.3	0.3	0.3	0.3
整体最小差分阶数下与原始序列相关系数	0.994	0.995	0.996	0.997
年度最小差分阶数	0.7	0.7	0.7	0.7
年度最小差分阶数下与原始序列相关系数	0.882	0.889	0.907	0.930

资料来源：光大证券研究所，Wind

相比于不同指数之间所需差分阶数的相似性，不同股票要达到平稳性所需的最低差分阶数差异较大。测试的股票数量为 142 只（沪深 300 成分股中于 2005 年之前发行且尚未退市的股票共 142 只），以这 142 只股票为样本，整体最小差分阶数的均值为 0.31，标准差为 0.12；而年度最小差分阶数的均值为 0.61，标准差为 0.09。股票样本中，仅 3 只股票的年度最小差分阶数为 1。这说明对于绝大部分股票与指数而言，如果仅是要求达到平稳性条件，并不一定需要做一阶差分，0.7 阶差分已经可以令大部分股票或指数满足平稳性要求。股票非整差分的详细数据可见附录部分。

图 7：不同股票在各自最小差分阶数下与原始序列相关性箱型图



资料来源：光大证券研究所，Wind；注：纵坐标为差分序列与原始序列相关系数，横坐标为差分阶数。

3、非整差分的一些简单应用

上一章我们探讨非整差分的方法与 A 股指数与股票中要达到平稳条件，这一章我们探索非整差分在量化研究中的一些应用方式。

3.1、不同差分程度下价格序列的信息熵

在运用机器学习工具时，平稳处理后的价格序列除了本身可以作为特征或预测对象以外，一些对平稳性有一定要求的衍生指标也可以基于不同差分程度的价格序列来构造。其中交易序列的信息熵正是这样一种对交易决策有信息价值的指标。

3.1.1、信息熵的概念

信息熵作为信息理论的重要概念由 Shannon 于 1948 年提出,最开始用于描述要传达一段信息序列最少要用多少字节,如果信息越规律所需要的字节就越少。这个概念随后被延申到所有平稳序列,用来描述一段信息序列蕴含的平均信息量。其数学定义为:

$$H := - \sum_{i=1}^n p_i \log p_i$$

其中: p_i 为第 i 个事件(或字段)发生的概率; n 为所有可能发生的事件(字段)个数。

而在金融领域中,很多市场参与者的投资策略表现与市场环境有关。例如趋势交易者的策略往往在单边趋势行情下表现更好、套利交易者的策略往往在市场有效性弱的时候表现更好。如果能更深入了解当前市场的环境特征,便能更灵活地实施交易策略;而连续交易形成的价格可以理解为市场向所有人展示的信息,将信息熵的概念应用到交易序列中来表达当前市场内部环境便成为一个自然而然的需求。

由于信息熵的估计与计算对于序列有平稳性要求,传统上收益率序列往往成为了计算信息熵的底层数据。我们现在已经知道达到平稳的要求并不需要差分到收益率的程度,价格对数序列在足够大的非整差分下也可以成为计算信息熵的底层数据,而保留更多记忆的特征也使得它从逻辑上会是更好的底层数据。

3.1.2、量价序列数据编码

在上述信息熵的数学定义中,涉及的概念是事件、字段这种极为离散的数据类型,这与量价数据这种较为连续的数据并不一致。因此在构建量价数据的信息熵指标之前,需要先对其进行编码成为离散的字段。

主流的编码方式一般有二元编码、分位数编码、等距编码等:

- 二元编码是天然契合于收益率数据的直观编码方式,只需把收益率大于 0 与收益率小于 0 分别编码为 1 与 0 即可;
- 分位数编码是将数据按其所在分位数组别来编码,例如如果按 10% 分位数编码,那么处于 0-10% 分位数的数据编码为 A, 10%-20% 分位数的数据编码为 B, 以此类推,直到 90% 到 100% 分位数的数据编码为 J;
- 等距编码的规则是相邻编码之间的边界之间的差距是固定值,一般情况下,这个差距由数据样本的标准差决定,例如 0 到 1 倍标准差内的数据编码为 A, 0 到 -1 倍标准差内的数据编码为 B, 1 到 2 倍标准差内的数据编码为 C, -1 到 -2 倍标准差内的数据编码为 D, 以此类推,直到把所有数据都编码完毕。这样的编码方式不像前两种编码方式有固定个数的编码符号,它的编码符号个数会根据样本特征有所差异。

3.1.3、较少差分阶数下价格信息熵指标有更多信息

从信息熵的构建逻辑上，我们预期它能反映出市场的有效性程度。信息熵越高，表明价格序列越没有规律，市场有效性越强；反过来说信息熵降低的时候，市场有效性较弱。因此在一些策略中，或机器学习模型里，信息熵可以作为一个额外特征用以辅助判断一些交易决策是否适合于当前市场环境。例如避免在市场有效性强时的趋势跟随信号或市场有效性弱时的反转信号等。

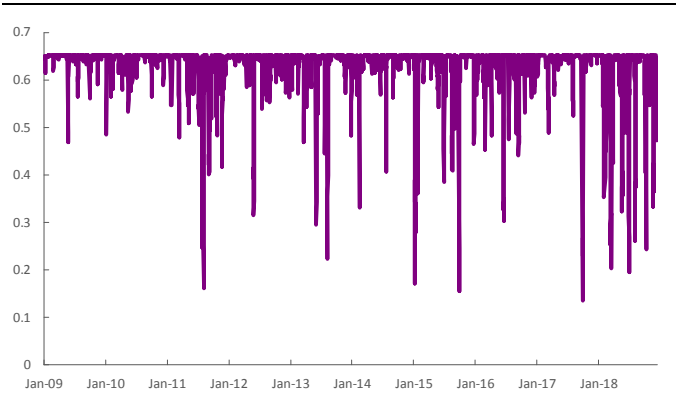
以上证综指指数为例，我们比较收益率序列计算的信息熵指标与以 0.7 阶差分下价格对数序列计算的信息熵。若采用分位数编码，在收益率序列上与在 0.7 阶差分序列上计算的信息熵在不同时间段皆有明显变化；而若采用 1 倍标准差等距编码，基于收益率序列结算的信息熵在整个时间段内基本没有太大变化，而基于 0.7 阶差分序列计算的信息熵仍有一定信息。

表 3：价格序列差分测试设定

设定项	设定值
测试序列	上证综指收益率序列 & 价格对数 0.7 阶差分序列
数据区间	2009/01/01 – 2018/12/31
价格序列频率	5 分钟
编码方式	分位数编码 & 等距编码 (1 倍标准差)
词段长度	12 (1 小时)
句段长度	240 (1 周)

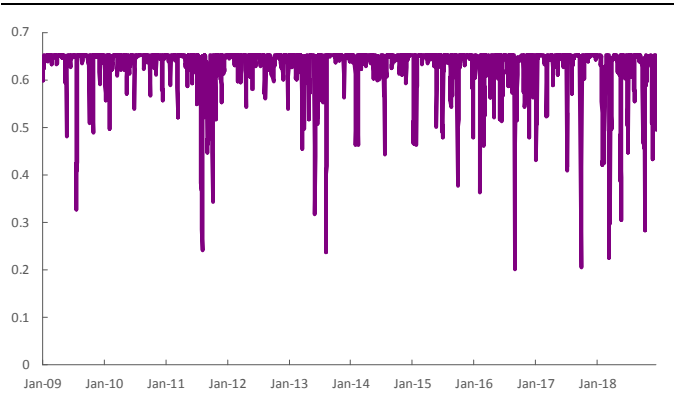
资料来源：光大证券研究所

图 8：收益率序列在分位数编码下的信息熵



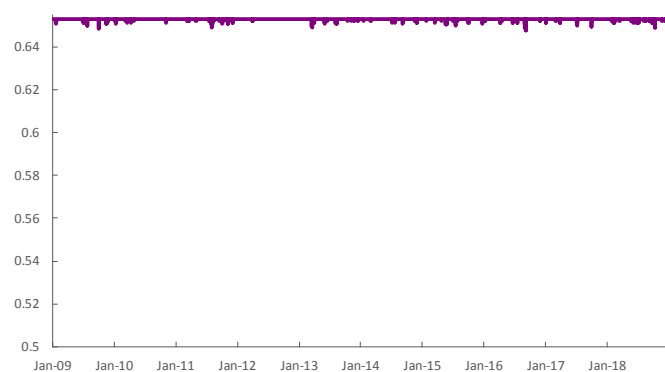
资料来源：光大证券研究所

图 9：0.7 阶差分序列在分位数编码下的信息熵



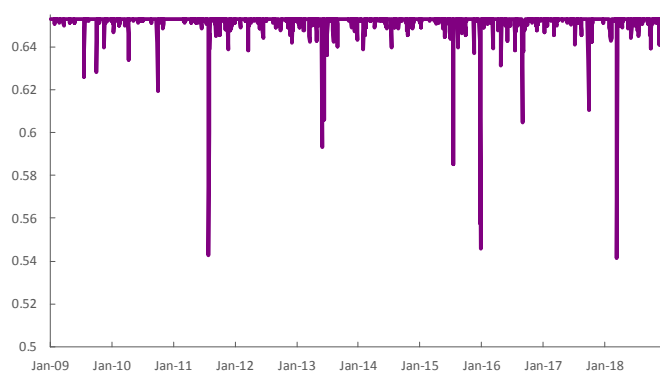
资料来源：光大证券研究所

图 10: 收益率序列在等距编码下的信息熵



资料来源: 光大证券研究所

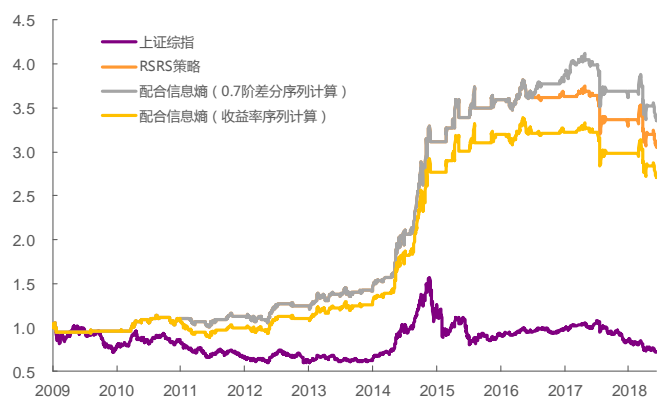
图 11: 0.7 阶差分序列在等距编码下的信息熵



资料来源: 光大证券研究所

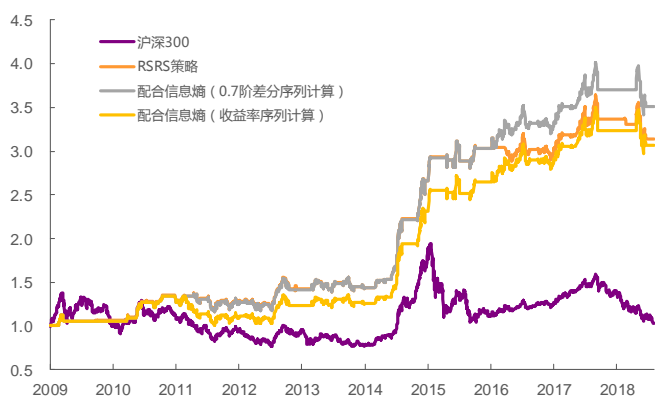
我们尝试将信息熵简单配合在之前报告中构建的 RSRs 择时策略, 整体策略不变, 仅在发出信号时根据信息熵的大小判断是否要执行该信号, 若信息熵很小则拒绝新的交易信号。从结果上看, 配合信息熵后 RSRs 策略的确避免了一些错误信号, 但实际边际改善较低且多集中在 2017 与 2018 年信息熵变化较大的时期, 因此在此不多做展开。从下图可以看出基于 0.7 阶差分计算的信息熵指标对原策略改善效果好于收益率计算的信息熵。

图 12: 上证综指 RSRs 策略配合信息熵指标效果



资料来源: 光大证券研究所; 注: 测试样本区间为 2009/1/1 至 2018/12/31

图 13: 沪深 300 指数 RSRs 策略配合信息熵指标效果



资料来源: 光大证券研究所; 注: 测试样本区间为 2009/1/1 至 2018/12/31

对信息熵的实际运用并不会像上述示例那样简单, 在机器学习算法里只有配合其它在特定市场环境下有明显预测效果的特征时才能最大化运用它带来的信息。接下来我们尝试非整差分另一个更直观有用的应用场景。

3.2、波动因子构造的延伸

虽然我们通过机器学习特征细节的角度引入了非整差分概念与操作的研究与讨论, 但其应用并非仅限定在机器学习领域。在传统量化里, 很多基于价格序列数据构造的模型或策略也能基于非整差分做更多的思考。比如波

动因子在不同差分阶数下的表现是否有明显的特征，在原有的构造下还有没有进一步改善的空间。

表 4：因子测试设置

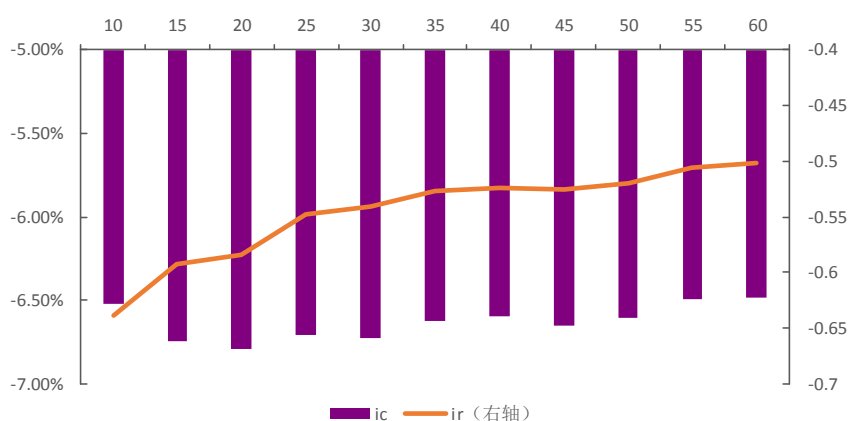
设定项	设定值
数据区间	2005/01/01 – 2018/12/31
股票池	全部 A 股（剔除选股日 ST/PT 股票；剔除上市不满一年的股票；剔除选股日由于停牌等因素无法买入的股票）
因子预处理	截面标准化处理
因子中性化	市值、一级行业
因子检验项	IC 均值、IR
ICIR 统计频率	月频

资料来源：光大证券研究所

3.2.1、主流低波因子的构建与效果

低波因子是 A 股较为稳定的有效因子，其因子主流构建方式是以一定区间内的日收益率序列（或者价格对数的一阶差分）为样本计算其样本标准差。样本大小有的以自然月为基准，有的以最近固定 N 个交易日为基准。我们先测试不同样本大小参数下低波因子的效果。可以看出在 N 从 10 个交易日到 60 个交易日的区间里，主流低波因子的 IC 均值不太受样本长度参数影响，基本都稳定在 -6.5% 到 -6.8% 之间。而因子 IR 的绝对值明显随着样本长度参数的增加而单调下降，以最近 10 个交易日的收益率计算的标准差作为当月波动因子值的效果最佳，此时 IR 绝对值 0.63。

图 14：主流低波因子在不同样本长度参数下的 IC 与 IR 表现



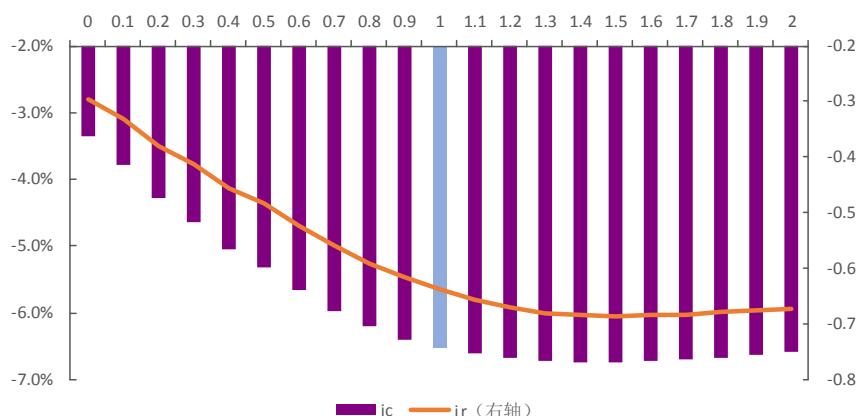
资料来源：光大证券研究所，Wind；注：数据统计区间 2005/01/01 – 2018/12/31

3.2.2、非整差分拓宽波动因子优化空间

除了通过对样本长度参数的优化，寻找主流波动因子最佳的构建方式之外，从非整差分的角度也可以进一步拓宽因子优化空间，我们验证了从因子效果来说一阶差分并非是波动因子的最佳差分阶数，通过选择合适的差分阶数，可以继续提升波动因子的选股效果。

在固定样本长度为 10 个交易日的情形下，测试差分阶数从 0 阶到 2 阶所有间隔 0.1 阶的差分阶数。以价格对数差分数据的标准差作为因子值的因子 IC 绝对值与 IR 绝对值在差分阶数小于 1 时单调上升。在差分阶数大于 1 时，因子 IC 与 IR 的变化变缓，IC 绝对值在 1.4 阶处达到最大后开始缓缓下降，IR 绝对值在 1.5 阶处达到最大后缓缓下降。因此从波动因子效果角度出发，价格对数差分阶数的最优参在 1.4 与 1.5 左右，此时相比于原来波动因子，IR 大概有 0.05 左右的提升。

图 15：主流低波因子（STD(10)）在不同差分阶数参数下的 IC 与 IR 表现



资料来源：光大证券研究所，Wind；注：数据统计区间 2005/01/01 – 2018/12/31

在对各不同样本大小参数下的情况做了综合测试后（详细统计数据可见文后附录），我们得到以下结论：

1. 在所有样本大小参数下，1 阶差分都不是最优差分阶数参数（实际都在 1.4 或 1.5 以上）。
2. 在所有样本大小参数下，IC 均值的绝对值都在差分阶数为 1.4、1.5 之间达到最大。
3. 从 IR 绝对值的角度看，最优差分阶数参数会随着样本大小参数增大而变大；但最大 IR 绝对值依然会随着样本大小变大而减小。

综合上述测试，我们较为推荐的月频波动因子构建方式为：在对股票价格对数序列作 1.5 阶差分后，以月末最近的 10 个交易日数据为样本计算标准差负数，作为次月的低波动因子值。为了之后表述方便我们用 $STD(n, m)$ 表示 n 阶差分、 m 个交易日数据样本计算标准差的低波因子。那么原来的主流波动因子就是 $STD(1, 10)$ ，新构建的波动因子就是 $STD(1.5, 10)$ 。我们继续测试如上述构造的低波因子在选股能力与组合表现上与主流波动因子的差异。

◆ 各股票池内差分优化后的低波因子 IC、IR 更高。

从 IC、IR 的统计可以看出 $STD(1.5, 10)$ 在全市场、中证 500 成分股内、沪深 300 成分股内的表现皆优于 $STD(1, 10)$ 。在中证 500 成分股内的优势最为明显。而两种波动因子之间的相关性很高，在全市场、中证 500、沪深

300 内的相关性分别为 0.95、0.94、0.95，说明该改进方式是在没有改变原有因子逻辑下使得效果得到一定提升。

表 5：各股票池内不同波动因子构造的效果差异

股票池	因子名称	IC 均值	IC 标准差	IC>0 比例	IC>0.02 比例	IR
全市场	STD(1,10)	6.5%	10.2%	72.9%	66.3%	0.63
	STD(1.5,10)	6.7%	9.8%	73.5%	66.9%	0.68
中证 500	STD(1,10)	4.8%	10.4%	68.5%	61.5%	0.46
	STD(1.5,10)	5.2%	9.9%	73.1%	66.2%	0.52
沪深 300	STD(1,10)	3.0%	11.0%	63.6%	56.3%	0.27
	STD(1.5,10)	3.2%	11.1%	61.6%	55.6%	0.29

资料来源：光大证券研究所，Wind；注：沪深 300 股票池内的统计数据为 2005/5/1-2018/12/31，中证 500 股票池内的统计数据为 2007/1/1 - 2018/12/31

◆ 各股票池内差分优化后的低波因子多空组合表现更佳。

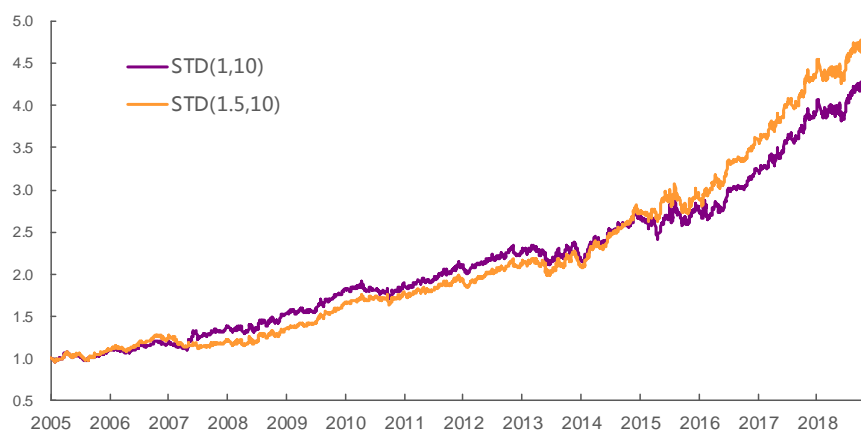
按因子大小将股票池内股票等分 5 组，以第五组与第一组股票构建多空组合。在各股票池内，STD(1.5,10)的多空组合都有更高的年化收益与夏普比率。以相比于 STD(1,10)的提升幅度这个角度，可以发现新构建的低波因子在沪深 300 与中证 500 股票池内的改善幅度更明显：在中证 500 成分股内，年化收益增加 1.5%，夏普比率上升 0.17；在沪深 300 成分股内，年化收益增加 2.3%，夏普比率上升 0.23。

表 6：不同低波因子构造下多空组合统计表现

统计项	全市场		中证 500		沪深 300	
	STD(1,10)	STD(1.5,10)	STD(1,10)	STD(1.5,10)	STD(1,10)	STD(1.5,10)
年化收益	11.0%	11.8%	5.3%	6.8%	3.0%	5.3%
累计收益	329.2%	377.1%	86.3%	120.1%	50.8%	103.6%
年化波动率	9.7%	9.3%	10.0%	9.5%	10.6%	10.2%
夏普比率	1.12	1.25	0.57	0.74	0.33	0.56
最大回撤	12.8%	13.0%	13.8%	13.5%	34.4%	30.3%

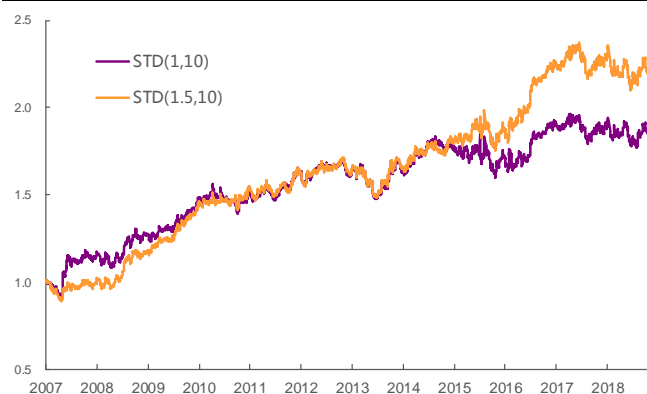
资料来源：光大证券研究所，Wind；注：沪深 300 股票池内的统计数据为 2005/5/1-2018/12/31，中证 500 股票池内的统计数据为 2007/1/1 - 2018/12/31

图 16: 全市场股票池下不同波动因子多空净值



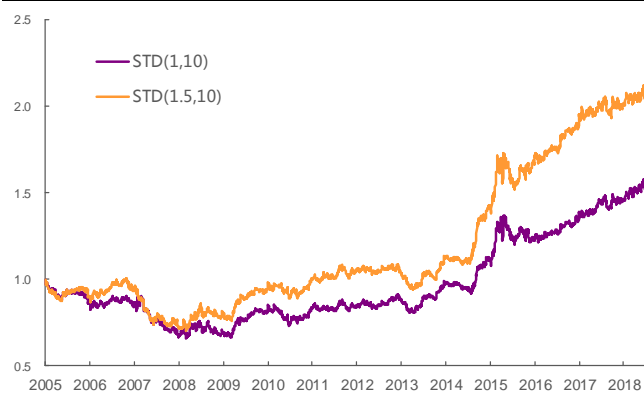
资料来源: 光大证券研究所, Wind

图 17: 中证 500 股票池下不同波动因子多空净值



资料来源: 光大证券研究所, Wind

图 18: 沪深 300 股票池下不同波动因子多空净值



资料来源: 光大证券研究所, Wind

◆ 差分优化后的低波因子多头组合增益有限。

按因子大小将股票池内股票从大到小排序后挑选前 100 只股票构建等权组合。在各股票池内, STD(1.5,10) 的多头组合的年化收益与夏普比率都有上升, 但上升幅度有限: 在全市场年化收益增加 1.2%; 中证 500 成分股内, 年化收益增加 0.8%; 在沪深 300 成分股内, 年化收益仅增加 0.4%。

表 7: 不同低波因子构造下多头组合统计表现

统计项	全市场		中证 500		沪深 300	
	STD(1,10)	STD(1.5,10)	STD(1,10)	STD(1.5,10)	STD(1,10)	STD(1.5,10)
年化收益	10.8%	12.0%	5.4%	6.2%	9.7%	10.1%
累计收益	319.9%	387.9%	88.6%	106.3%	257.0%	278.0%
年化波动率	28.3%	28.4%	31.1%	31.3%	27.8%	27.9%
夏普比率	0.50	0.54	0.33	0.35	0.47	0.49
最大回撤	71.5%	72.8%	71.2%	70.5%	71.2%	70.9%

资料来源: 光大证券研究所, Wind; 注: 沪深 300 股票池内的统计数据为 2005/5/1-2018/12/31, 中证 500 股票池内的统计数据为 2007/1/1-2018/12/31

图 19：全市场股票池下不同波动因子多头组合



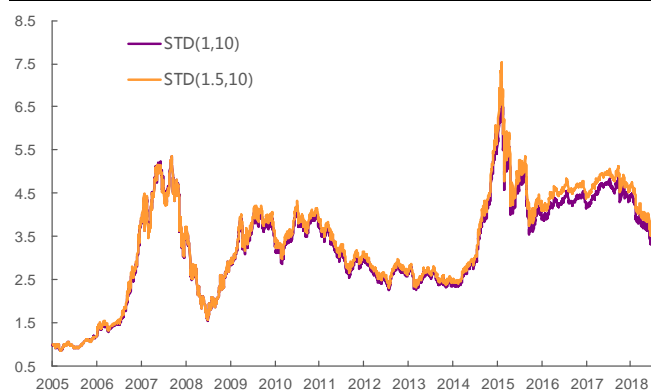
资料来源：光大证券研究所，Wind

图 20：中证 500 股票池下不同波动因子多头组合



资料来源：光大证券研究所，Wind

图 21：沪深 300 股票池下不同波动因子多头组合



资料来源：光大证券研究所，Wind

综上测试可得出结论，在引入非整差分的操作后，原有低波因子的构建有进一步优化的空间。同理，其它基于价格或收益率序列数据构建的量价因子也可以基于非整差分的操作尝试进一步优化。

4、风险提示

本报告中的测试结果均基于模型和历史数据，历史数据存在不被重复验证的可能，模型存在失效的风险。

5、附录

1. 当前沪深 300 成分股中历史长于 14 年的所有股票达到平稳性要求的最小差分阶数

整体差分阶数：结果仅需通过 2005/1/1 到 2018/12/31 整个序列进行的 ADF 检验

年度差分阶数：结果需通过数据样本内所有连续年份组合的数据序列上进行的 ADF 检验。（例如如果整个数据样本是 2010 年初到 2012 年末，那么需要在以下所有序列上分别进行的 ADF 检验：2010 年单年、2011 年单年、2012 年但单年、2010 年初到 2011 年末、2011 年初到 2012 年末、2011 年初到 2012 年末，共 6 次。同理，整个数据样本从 2005 年初到 2018 年末的话，共需满足全部 105 次检验。）

表 8：不同股票整体最小差分阶数与年度最小差分阶数

股票代码	整体差分阶数	年度差分阶数	股票代码	整体差分阶数	年度差分阶数	股票代码	整体差分阶数	年度差分阶数
000001.SZ	0.3	0.6	600000.SH	0.3	0.6	600372.SH	0.4	0.5
000002.SZ	0.4	0.7	600004.SH	0.3	0.6	600383.SH	0.3	0.6
000063.SZ	0.2	0.6	600009.SH	0.3	0.6	600390.SH	0.3	0.6
000069.SZ	0	0.6	600010.SH	0.2	0.6	600398.SH	0.3	0.6
000100.SZ	0.3	0.5	600011.SH	0.2	0.6	600406.SH	0.4	0.6
000157.SZ	0.3	0.7	600015.SH	0.3	0.6	600415.SH	0.3	0.6
000402.SZ	0	0.6	600016.SH	0.3	0.7	600436.SH	0.5	0.5
000408.SZ	0.4	0.6	600019.SH	0.2	0.7	600438.SH	0.1	0.6
000413.SZ	0.4	0.5	600028.SH	0.2	0.6	600482.SH	0.3	0.6
000415.SZ	0.3	0.5	600029.SH	0.3	0.7	600487.SH	0.4	0.5
000423.SZ	0.4	0.5	600030.SH	0.1	0.7	600489.SH	0	0.6
000425.SZ	0.3	0.6	600031.SH	0.3	0.7	600498.SH	0.3	0.5
000503.SZ	0.3	0.6	600036.SH	0.4	0.7	600516.SH	0.4	0.7
000538.SZ	0.4	0.5	600038.SH	0.2	0.5	600518.SH	0.5	0.8
000553.SZ	0.3	0.6	600050.SH	0.2	0.6	600519.SH	0.5	0.7
000568.SZ	0	0.7	600061.SH	0.3	0.7	600522.SH	0.4	0.6
000625.SZ	0.4	0.6	600066.SH	0.4	0.6	600535.SH	0.4	0.5
000627.SZ	0.3	0.6	600068.SH	0.3	0.6	600547.SH	0.3	0.6
000630.SZ	0.2	0.6	600085.SH	0.4	0.5	600549.SH	0.1	0.6
000651.SZ	0.5	0.6	600089.SH	0.3	0.6	600566.SH	0.4	0.6
000661.SZ	0.5	0.6	600100.SH	0.2	0.6	600570.SH	0.5	0.6
000671.SZ	0.4	0.5	600104.SH	0.4	0.6	600583.SH	0	0.6
000703.SZ	0.4	0.6	600109.SH	0.3	0.7	600585.SH	0.3	0.6
000709.SZ	0.3	0.7	600111.SH	0.4	0.6	600588.SH	0.3	0.6
000725.SZ	0.2	0.6	600115.SH	0.2	0.7	600606.SH	0.2	0.6
000728.SZ	0.2	0.6	600118.SH	0.4	0.6	600637.SH	0.4	0.6
000768.SZ	0	0.6	600153.SH	0.4	0.7	600660.SH	0.4	0.6
000776.SZ	0.3	0.6	600157.SH	0.4	0.8	600674.SH	0.4	0.6
000783.SZ	0.2	1	600170.SH	0.3	0.6	600688.SH	0.2	0.6
000786.SZ	0.3	0.5	600176.SH	0.3	0.6	600690.SH	0.4	0.6
000792.SZ	0.3	0.6	600177.SH	0.3	0.6	600703.SH	0.5	0.6
000826.SZ	0.4	0.6	600188.SH	0.2	0.5	600704.SH	0.3	0.6
000839.SZ	0.2	0.6	600196.SH	0.4	0.5	600705.SH	0.5	0.7
000858.SZ	0.4	0.7	600208.SH	0.2	0.6	600739.SH	0	0.7
000876.SZ	0.2	0.6	600219.SH	0.2	0.6	600741.SH	0.4	0.6
000895.SZ	0.3	0.6	600221.SH	0.2	0.6	600760.SH	0.4	0.6
000898.SZ	0.3	0.6	600233.SH	0.3	0.6	600795.SH	0.1	0.7

000938.SZ	0.3	0.5	600271.SH	0.3	0.5	600809.SH	0.3	0.6
000959.SZ	0.2	0.6	600276.SH	0.5	0.6	600816.SH	0.4	0.7
000961.SZ	0.3	0.6	600297.SH	0.3	0.5	600837.SH	0.2	0.7
000963.SZ	0.5	0.6	600309.SH	0.3	0.6	600867.SH	0.5	0.6
000983.SZ	0.3	0.7	600332.SH	0.3	0.5	600886.SH	0.4	0.7
002001.SZ	0.4	0.6	600339.SH	0.1	0.5	600887.SH	0.4	0.6
002007.SZ	0.4	0.7	600340.SH	0.6	0.6	600893.SH	0.4	1
002008.SZ	0.4	0.6	600346.SH	0.4	0.5	600900.SH	0.4	0.7
002024.SZ	0	0.7	600352.SH	0.4	0.6	601607.SH	0.3	0.5
002027.SZ	0.4	0.6	600362.SH	0.2	0.6			
002032.SZ	0.4	0.6	600369.SH	0.4	1			

资料来源：光大证券研究所，Wind

2. 低波因子参数测试

表 9：波动因子在不同样本长度与差分阶数参数下的 IC 均值

	10	15	20	25	30	35	40	45	50	55	60
0	-3.4%	-3.4%	-3.4%	-3.0%	-2.7%	-2.5%	-2.1%	-1.9%	-1.8%	-1.8%	-1.8%
0.1	-3.8%	-4.1%	-3.9%	-3.5%	-3.1%	-3.0%	-2.6%	-2.3%	-2.1%	-2.1%	-2.1%
0.2	-4.3%	-4.6%	-4.3%	-4.0%	-3.7%	-3.6%	-3.3%	-3.0%	-2.8%	-2.7%	-2.7%
0.3	-4.6%	-5.1%	-4.7%	-4.5%	-4.3%	-4.2%	-4.0%	-3.7%	-3.5%	-3.4%	-3.4%
0.4	-5.0%	-5.5%	-5.2%	-5.0%	-4.8%	-4.7%	-4.6%	-4.4%	-4.2%	-4.1%	-4.1%
0.5	-5.3%	-5.9%	-5.6%	-5.4%	-5.3%	-5.3%	-5.1%	-5.1%	-4.9%	-4.8%	-4.8%
0.6	-5.7%	-6.2%	-6.0%	-5.8%	-5.8%	-5.7%	-5.6%	-5.6%	-5.5%	-5.4%	-5.4%
0.7	-6.0%	-6.5%	-6.3%	-6.2%	-6.1%	-6.1%	-6.0%	-6.0%	-5.9%	-5.9%	-5.9%
0.8	-6.2%	-6.6%	-6.6%	-6.4%	-6.4%	-6.3%	-6.3%	-6.3%	-6.3%	-6.2%	-6.2%
0.9	-6.4%	-6.7%	-6.7%	-6.6%	-6.5%	-6.5%	-6.5%	-6.6%	-6.5%	-6.4%	-6.4%
1	-6.5%	-6.7%	-6.8%	-6.7%	-6.7%	-6.6%	-6.6%	-6.6%	-6.6%	-6.5%	-6.5%
1.1	-6.6%	-6.8%	-6.9%	-6.8%	-6.7%	-6.7%	-6.6%	-6.7%	-6.6%	-6.5%	-6.5%
1.2	-6.7%	-6.9%	-6.9%	-6.9%	-6.8%	-6.8%	-6.7%	-6.8%	-6.7%	-6.6%	-6.5%
1.3	-6.7%	-7.1%	-7.0%	-6.9%	-6.8%	-6.8%	-6.8%	-6.8%	-6.7%	-6.6%	-6.6%
1.4	-6.7%	-7.1%	-7.0%	-6.9%	-6.9%	-6.8%	-6.8%	-6.8%	-6.8%	-6.6%	-6.6%
1.5	-6.7%	-7.1%	-7.0%	-7.0%	-6.9%	-6.8%	-6.8%	-6.9%	-6.8%	-6.6%	-6.6%
1.6	-6.7%	-7.1%	-7.0%	-7.0%	-6.9%	-6.8%	-6.8%	-6.8%	-6.7%	-6.6%	-6.6%
1.7	-6.7%	-6.9%	-7.0%	-6.9%	-6.9%	-6.8%	-6.8%	-6.8%	-6.7%	-6.6%	-6.5%
1.8	-6.7%	-6.9%	-7.0%	-6.9%	-6.8%	-6.8%	-6.8%	-6.8%	-6.7%	-6.5%	-6.5%
1.9	-6.6%	-6.9%	-7.0%	-6.9%	-6.8%	-6.8%	-6.7%	-6.8%	-6.7%	-6.5%	-6.5%
2	-6.6%	-6.8%	-6.9%	-6.9%	-6.9%	-6.7%	-6.7%	-6.7%	-6.6%	-6.5%	-6.5%

资料来源：光大证券研究所，Wind

表 10：波动因子在不同样本长度与差分阶数参数下的 IR 值

	10	15	20	25	30	35	40	45	50	55	60
0	-0.297	-0.293	-0.289	-0.248	-0.215	-0.191	-0.161	-0.144	-0.137	-0.143	-0.139
0.1	-0.331	-0.353	-0.319	-0.281	-0.244	-0.224	-0.194	-0.174	-0.159	-0.163	-0.161
0.2	-0.379	-0.396	-0.355	-0.318	-0.288	-0.267	-0.242	-0.222	-0.203	-0.202	-0.199
0.3	-0.414	-0.434	-0.392	-0.358	-0.332	-0.311	-0.290	-0.268	-0.255	-0.249	-0.246
0.4	-0.455	-0.468	-0.431	-0.397	-0.374	-0.354	-0.335	-0.324	-0.306	-0.299	-0.295

0.5	-0.483	-0.499	-0.468	-0.432	-0.414	-0.396	-0.379	-0.373	-0.358	-0.350	-0.346
0.6	-0.523	-0.527	-0.501	-0.462	-0.449	-0.434	-0.420	-0.417	-0.405	-0.398	-0.394
0.7	-0.559	-0.555	-0.531	-0.494	-0.478	-0.467	-0.454	-0.453	-0.443	-0.436	-0.431
0.8	-0.591	-0.576	-0.553	-0.517	-0.502	-0.493	-0.485	-0.485	-0.481	-0.468	-0.464
0.9	-0.617	-0.578	-0.571	-0.535	-0.522	-0.514	-0.508	-0.509	-0.505	-0.491	-0.487
1	-0.639	-0.593	-0.583	-0.548	-0.541	-0.527	-0.524	-0.525	-0.520	-0.506	-0.502
1.1	-0.658	-0.606	-0.595	-0.558	-0.544	-0.536	-0.534	-0.534	-0.528	-0.513	-0.509
1.2	-0.671	-0.618	-0.606	-0.569	-0.555	-0.547	-0.545	-0.546	-0.537	-0.522	-0.517
1.3	-0.680	-0.644	-0.617	-0.580	-0.565	-0.557	-0.555	-0.556	-0.546	-0.530	-0.526
1.4	-0.685	-0.650	-0.623	-0.587	-0.572	-0.564	-0.562	-0.564	-0.552	-0.536	-0.532
1.5	-0.686	-0.653	-0.627	-0.593	-0.578	-0.570	-0.568	-0.570	-0.557	-0.540	-0.536
1.6	-0.685	-0.655	-0.630	-0.596	-0.582	-0.574	-0.572	-0.573	-0.560	-0.543	-0.539
1.7	-0.683	-0.639	-0.631	-0.599	-0.584	-0.576	-0.574	-0.575	-0.562	-0.544	-0.541
1.8	-0.680	-0.640	-0.632	-0.601	-0.586	-0.578	-0.576	-0.578	-0.564	-0.545	-0.542
1.9	-0.677	-0.640	-0.632	-0.602	-0.587	-0.579	-0.577	-0.579	-0.565	-0.546	-0.543
2	-0.674	-0.640	-0.633	-0.603	-0.594	-0.580	-0.577	-0.580	-0.566	-0.546	-0.543

资料来源：光大证券研究所，Wind

行业及公司评级体系

评级	说明
买入	未来 6-12 个月的投资收益率领先市场基准指数 15% 以上；
增持	未来 6-12 个月的投资收益率领先市场基准指数 5% 至 15%；
中性	未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差 -5% 至 5%；
减持	未来 6-12 个月的投资收益率落后市场基准指数 5% 至 15%；
卖出	未来 6-12 个月的投资收益率落后市场基准指数 15% 以上；
无评级	因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，致使无法给出明确的投资评级。

基准指数说明：A 股主板基准为沪深 300 指数；中小盘基准为中小板指；创业板基准为创业板指；新三板基准为新三板指数；港股基准指数为恒生指数。

分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，并对本报告的内容和观点负责。负责准备以及撰写本报告的所有研究人员在此保证，本研究报告中任何关于发行商或证券所发表的观点均如实反映研究人员的个人观点。研究人员获取报酬的评判因素包括研究的质量和准确性、客户反馈、竞争性因素以及光大证券股份有限公司的整体收益。所有研究人员保证他们报酬的任何一部分不会与，不与，也将不会与本报告中的具体的推荐意见或观点有直接或间接的联系。

特别声明

光大证券股份有限公司（以下简称“本公司”）创建于 1996 年，系由中国光大（集团）总公司投资控股的全国性综合类股份制证券公司，是中国证监会批准的首批三家创新试点公司之一。根据中国证监会核发的经营证券期货业务许可，本公司的经营范围包括证券投资咨询业务。

本公司经营范围：证券经纪；证券投资咨询；与证券交易、证券投资活动有关的财务顾问；证券承销与保荐；证券自营；为期货公司提供中间介绍业务；证券投资基金代销；融资融券业务；中国证监会批准的其他业务。此外，本公司还通过全资或控股子公司开展资产管理、直接投资、期货、基金管理以及香港证券业务。

本报告由光大证券股份有限公司研究所（以下简称“光大证券研究所”）编写，以合法获得的我们相信为可靠、准确、完整的信息为基础，但不保证我们所获得的原始信息以及报告所载信息之准确性和完整性。光大证券研究所可能将不时补充、修订或更新有关信息，但不保证及时发布该等更新。

本报告中的资料、意见、预测均反映报告初次发布时光大证券研究所的判断，可能需随时进行调整且不予通知。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。客户应自主作出投资决策并自行承担投资风险。本报告中的信息或所表述的意见并未考虑到个别投资者的具体投资目的、财务状况以及特定需求。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。

不同时期，本公司可能会撰写并发布与本报告所载信息、建议及预测不一致的报告。本公司的销售人员、交易人员和其他专业人员可能会向客户提供与本报告中所载观点不同的口头或书面评论或交易策略。本公司的资产管理子公司、自营部门以及其他投资业务板块可能会独立做出与本报告的意见或建议不相一致的投资决策。本公司提醒投资者注意并理解投资证券及投资产品存在的风险，在做出投资决策前，建议投资者务必向专业人士咨询并谨慎抉择。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务。投资者应当充分考虑本公司及本公司附属机构就报告内容可能存在的利益冲突，勿将本报告作为投资决策的唯一信赖依据。

本报告根据中华人民共和国法律在中华人民共和国境内分发，仅向特定客户传送。本报告的版权仅归本公司所有，未经书面许可，任何机构和个人不得以任何形式、任何目的进行翻版、复制、转载、刊登、发表、篡改或引用。如因侵权行为给本公司造成任何直接或间接的损失，本公司保留追究一切法律责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

光大证券股份有限公司 2019 版权所有。

联系我们

上海	北京	深圳
静安区南京西路 1266 号恒隆广场 1 号 写字楼 48 层	西城区月坛北街 2 号月坛大厦东配楼 2 层 复兴门外大街 6 号光大大厦 17 层	福田区深南大道 6011 号 NEO 绿景纪元大 厦 A 座 17 楼