"订单簿的温度"系列研究(一)

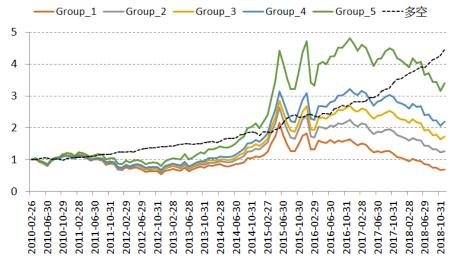
ケータ 新聞 SOOCHOW SECURITIES

反转因子的精细结构

2018年12月13日

观点

- A股市场是订单驱动型市场。从动力学的角度讲,股票行情的所有演化过程,都能由订单簿(order book)自下而上精确决定。逐笔成交与逐笔委托数据的信息量非常丰富。本系列研究取名为"订单簿的温度",旨在分享我们在逐笔数据层面的研究心得。
- 作为"订单簿的温度"系列研究的第1篇,本篇报告我们从最简单的数据 入手,考察了"成交笔数"这个指标。所谓成交笔数,即撮合交易的 次数,是从逐笔成交数据中汇总出来的统计量。我们将看到,这个简 单朴实的信息量,被用于传统反转因子的改进,有着令人震撼的效果。
- 在本报告中,我们借助成交笔数的信息,对传统反转因子进行切割,首次提出一个理想反转因子,其IC均值为-0.057, rankIC均值为-0.070, 五分组净值曲线排序良好,且多头组合与其他4组区分显著。多空对冲的年化收益为19.3%,年化波动为7.68%,月度胜率为74.3%,信息比率高达2.51。在剔除Barra风格因子和行业因子的影响之后,信息比率提升至2.97。



		年化	比收益率		多空对冲的统计指标					
年份	Group_5	Group_1	多空对冲	市场等权组合	月度胜率	年化波动	信息比率	最大回撤		
2010年	24.1%	13.8%	7.1%	15. 5%	60.0%	10.0%	0.71	7.4%		
2011年	-27.3%	-38. 7%	17.1%	-33.0%	83. 3%	4.6%	3. 70	0.7%		
2012年	9.0%	-5.3%	13. 9%	1.6%	83. 3%	4.7%	2. 93	1.0%		
2013年	41.1%	25. 1%	12.6%	30. 4%	66. 7%	4.7%	2. 65	1.8%		
2014年	52.5%	35. 1%	13. 1%	46. 9%	66. 7%	9. 5%	1. 37	5. 3%		
2015年	140.9%	70.4%	37.6%	96. 4%	66. 7%	11. 9%	3. 16	3.0%		
2016年	-2.1%	-17.4%	17.2%	-7. 5%	66. 7%	9. 0%	1. 91	3.4%		
2017年	-10. 9%	-30. 3%	26. 4%	-16. 5%	83. 3%	6. 0%	4. 41	0.1%		
2018年	-28. 1%	-43.3%	25. 6%	-34. 2%	90. 9%	4.4%	5. 81	0. 7%		
2010-2018	15. 5%	-4. 2%	19. 3%	6. 0%	74. 3%	7. 7%	2. 51	7.4%		

■ 风险提示:模型的测算基于历史数据,市场未来可能发生较大变化。

证券分析师 魏建榕 执业证号: \$0600517120003 021-60199793 weijr@dwzq.com.cn 证券分析师 高子剑 执业证号: \$0600518010001 021-60199793 gaozj@dwzq.com.cn 研究助理 傅开波 fukb@dwzq.com.cn

相关研究

金工专题报告



内容目录

1.	引言	. 4
2.	反转因子的切割问题	. 4
3.	反转因子的 W 式切割	. 6
	理想反转因子	
	若干重要的讨论	
	风险提示	
•	× 417 45.4	



图表目录

图	1:	以往模型的交易行为逻辑	. 4
		传统反转因子的稳定性不佳(Ret20因子,分5组,多空对冲)	
图	3:	反转因子的切割问题	. 5
图	4:	因子回测(信息比率与月度胜率的计算是按反转因子使用)	. 6
图	5:	理想反转因子 M 的五分组与多空对冲	. 7
图	6:	理想反转因子 M 的分年度表现	. 7
图	7:	因子间的相关系数矩阵	. 8
图	8:	理想反转因子 M 剔除 Barra 因子与行业因子后的选股能力	. 8
图	9:	N=20,40,60 三种情况下的切割效果(月度 IC 的均值)	. 9
图	10:	沪深 300 成分股上的选股能力(多空对冲净值)	. 9
图	11:	多空对冲收益的累积过程(T+0 为月初建仓日)	10
图	12:	分组比例的影响(纵坐标为 M 因子的信息比率)	10
图	13:	单笔成交金额的日内模式	11



1. 引言

A股市场是一个订单驱动型(order-driven)的市场。从动力学的角度讲,股票行情的所有演化过程,都能由订单簿(order book)自下而上、精确完备地决定。目前,上海证券交易所对外提供了逐笔成交数据,深圳证券交易所对外提供了逐笔成交和逐笔委托数据,这些逐笔数据的信息含量极其丰富。本系列研究取名为"**订单簿的温度**",旨在分享我们在逐笔数据层面的研究心得。

本系列将延续笔者长期以来的报告风格: 天马行空地想, 字斟句酌地写。近几年来, 我们在"高频数据、低频信号"方面的大量研究, 受到了量化同行的普遍认可, 也成功引领了"高频数据"和"价量行为"的研究潮流。事实上, 在笔者心目中, "高频数据"和"价量行为"都不是重要的标签。"从交易行为中寻找alpha", 才是我们这几个系列报告最有价值的内核, 也是我们一贯重视的逻辑根基。作为简单回顾, 图1罗列了我们以往部分模型的交易行为逻辑。

图1: 以往模型的交易行为逻辑

模型名称	基于交易行为的逻辑
蜘蛛网CTA模型	前20大期货会员的持仓动向,对期货次日的涨跌有部分预测能力。
异动罗盘模型	股价在盘中的逆势异动,通常是资金异动的征兆。
十字星模型	十字星形态的逻辑内涵,是交易情绪博弈,应该用于超额收益的K线。
跟踪聪明钱	从分钟行情数据的价量特征中,可以识别出机构参与交易的多寡。
凤鸣朝阳模型	日内的不同时段,交易者行为不同,反转强度也可能有所不同。
枯树生花模型	日内的不同时段,交易者行为不同,反转强度也可能有所不同。
抢跑者的脚步声	大涨大跌前放量,是信息泄露的迹象,可从量价互动中提取选股因子。

数据来源:东吴证券研究所金融工程团队

作为"订单簿的温度"系列研究的第1篇,本篇报告我们从最简单的数据入手,考察了"成交笔数"这个指标。所谓成交笔数,即撮合交易的次数,是从逐笔成交数据中汇总出来的统计量。我们很快将看到,这个简单朴实的信息量,被用于对反转因子进行切割和改进,有着令人震撼的效果。这种简单和深刻的强烈对比,让笔者在落笔时仍旧心潮澎湃。在看完报告全文之后,读者会理解这种心情的由来。

2. 反转因子的切割问题

众所周知,A股市场呈现较为显著的中长期反转效应。以20日收益率因子(Ret20)为例,从2010年至2018年期间,月度IC的均值为-0.056,rankIC的均值为-0.061;将其作



为反转因子对股票进行排序分组,五分组多空对冲的信息比率为1.20,月度胜率为63.8%。但是,令人遗憾的是,反转效应的稳定性很不理想。从图2我们可以看到,至少对于2013年上半年、2014年下半年和2017年这些时段,反转因子基本失效,市场甚至呈现为动量效应。对40日收益率因子(Ret60)和60日收益率因子(Ret60)的考察,也存在类似的结论。

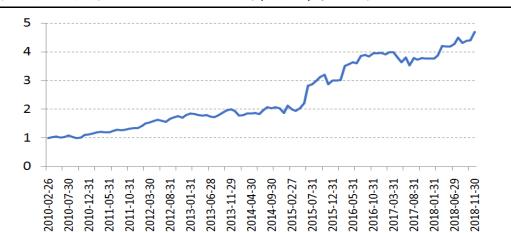
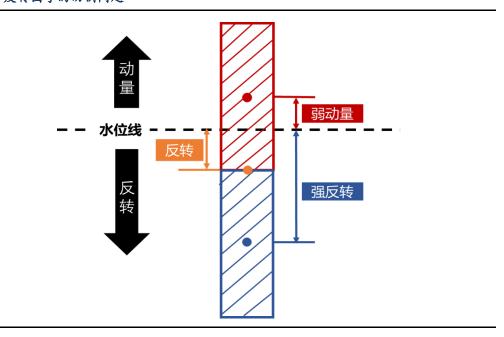


图2: 传统反转因子的稳定性不佳(Ret20因子,分5组,多空对冲)

数据来源: Wind资讯, 东吴证券研究所

传统反转因子在稳定性上的困难,时常将笔者的思路引向"**反转因子的切割问题**"。 思维的过程是这样展开的。首先,我们注意到,传统反转因子本质上是一段区间的涨跌幅,因此可以很自然地拆分为许多更小的时段。那么,我们可以发问,会不会存在这样的情况:组成传统反转因子的各个时段中,某些时段贡献了很强的反转,而剩余时段只是贡献了很弱的反转,甚至有的是贡献了动量效果?

图3: 反转因子的切割问题



数据来源: 东吴证券研究所金融工程团队



图3的示意图,可以提供切割问题的一个直观图景。我们将传统收益率因子想象为一个柱状体,它的重心(橘色圆点)在水位线下方,其寓意是"传统收益率因子呈现反转效应"。所谓切割问题是说,我们能否找到了一个好的切割方案,使得柱状体被分割为蓝色因子和红色因子两个部分呢?在这个理想的切割方案下,蓝色因子的重心(蓝色圆点)在水位线下方更深处,也即呈现为更强的反转因子;红色因子的重心(红色圆点)则略高于水位线,呈现为弱的动量因子。

3. 反转因子的W式切割

经过长期反复的摸索,我们找到了一个反转因子的有效切割方案,简称**W式切割**。 具体操作步骤如下:

- (1) 在每个月底,对于股票s,回溯其过去N个交易日的数据(为方便处理,N取偶数):
- (2)对于股票s,逐日计算**平均单笔成交金额**D(D=当日成交金额/当日成交笔数),将N个交易日按D值从大到小排序,前N/2个交易日称为高D组,后N/2个交易日称为低D组:
- (3) 对于股票s,将高D组交易日的涨跌幅加总[1],得到因子M_high;将低D组交易日的涨跌幅加总,得到因子M_low;
 - (4) 对于所有股票,分别按照上述流程计算因子值。

W式切割的核心步骤是,按照"单笔成交金额"对交易日进行排序分组[2]。我们以20日收益率因子为例(即N取20),来说明W式切割方案的出色效果。样本空间为全部A股(剔除ST和上市未满60日的股票),回测时段仍为2010年至2018年。统计结果如图表4所示,结论是:M_high因子是非常强的反转因子(rankIC均值为-0.082),而M_low因子是较弱的动量因子(rankIC均值为0.018)[3]。

图4: 因子回测(信息比率与月度胜率的计算是按反转因子使用)

因子	IC均值	rankIC均值	信息比率	月度胜率	回归剔除Ret20后的rankIC均值
M_high	-0.069	-0. 082	1. 98	78. 1%	-0.058
M_low	0. 011	0. 018	-0. 31	44.8%	0. 058
Ret20	-0. 056	-0. 061	1. 2	63.8%	

数据来源: Wind资讯, 东吴证券研究所

4. 理想反转因子

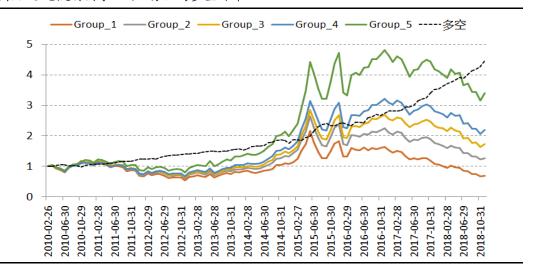
综合前文的讨论, 我们提出一个理想的反转因子M, 其定义式如下:

 $M = M_high-M_low$



从定义式我们很容易预判,由于M_high是强反转、M_low是弱动量,M因子大概率会是更强的反转因子。历史回测显示,对于全部A股(剔除ST和上市未满60日的股票),在2010年至2018年期间,M因子的IC月度均值为-0.057,rankIC月度均值为-0.070。如图5所示,五分组的净值曲线的排序良好,且多头组合Group_5与其他4组有更大的区分度。多空对冲的年化收益为19.3%,年化波动为7.68%,月度胜率为74.3%,信息比率高达2.51。图6我们进一步给出了分年度的统计情况。

图5: 理想反转因子M的五分组与多空对冲



数据来源: Wind资讯, 东吴证券研究所

图6: 理想反转因子M的分年度表现

		年化	七收益率		多空对冲的统计指标				
年份	Group_5	Group_1	多空对冲	市场等权组合	月度胜率	年化波动	信息比率	最大回撤	
2010年	24.1%	13.8%	7.1%	15. 5%	60.0%	10.0%	0.71	7.4%	
2011年	-27.3%	-38. 7%	17.1%	-33.0%	83. 3%	4.6%	3. 70	0.7%	
2012年	9.0%	-5. 3%	13.9%	1. 6%	83. 3%	4.7%	2. 93	1.0%	
2013年	41.1%	25. 1%	12.6%	30. 4%	66. 7%	4.7%	2.65	1.8%	
2014年	52.5%	35. 1%	13. 1%	46. 9%	66. 7%	9. 5%	1. 37	5.3%	
2015年	140. 9%	70. 4%	37.6%	96. 4%	66. 7%	11. 9%	3. 16	3.0%	
2016年	-2.1%	-17.4%	17.2%	-7. 5%	66. 7%	9.0%	1. 91	3.4%	
2017年	-10.9%	-30.3%	26.4%	-16. 5%	83. 3%	6.0%	4.41	0.1%	
2018年	-28. 1%	-43.3%	25.6%	-34.2%	90. 9%	4.4%	5. 81	0.7%	
2010-2018	15. 5%	-4.2%	19. 3%	6. 0%	74. 3%	7. 7%	2. 51	7.4%	

数据来源: Wind资讯, 东吴证券研究所

5. 若干重要的讨论

关于理想反转因子M,有许多值得深入讨论的方面,以下我们做逐一的阐述:

第一,与风格因子的关联。由于M因子是由两个涨跌幅相减得到,我们预判它与传统反转因子的关联会较低,与Beta、波动率等因子的关联可能比较明显。图7给出了M因



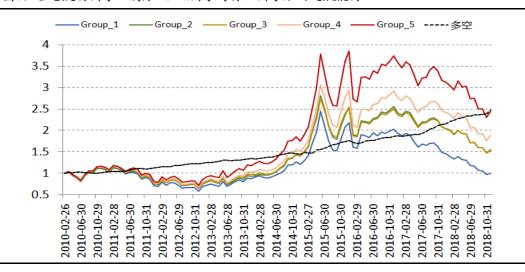
子与10个Barra因子之间的相关系数矩阵。为剔除风格的干扰,我们把M因子在横截面上对10个Barra因子与29个中信一级行业指数的哑变量进行回归,将残差作为新的选股因子。 五分组与多空对冲的净值如图8所示。纯净化后的因子,多空对冲的年化收益为11.3%,年化波动为3.80%,月度胜率为76.2%,信息比率提升至2.97。

图7: 因子间的相关系数矩阵

	M因子	Size	Beta	Momentum	ResVol	NSize	BtoP	Liquidity	EarningsY	Growth	Leverage
M因子	1.00	-0.01	0.16	0.00	0.20	-0.01	-0.08	0.23	-0.07	-0.01	0.00
Size	-0.01	1.00	-0.13	0.11	0.01	0.56	0.15	-0.01	0.26	0.09	0.29
Beta	0.16	-0.13	1.00	-0.06	0.04	-0.01	-0.03	0.32	-0.07	0.01	-0.11
Momentum	0.00	0.11	-0.06	1.00	0.35	0.04	-0.31	0.20	-0.01	0.06	-0.03
ResidualVolatility	0.20	0.01	0.04	0.35	1.00	0.01	-0.41	0.46	-0.22	0.05	0.00
NonlinearSize	-0.01	0.56	-0.01	0.04	0.01	1.00	0.05	-0.05	0.05	0.02	0.15
BooktoPrice	-0.08	0.15	-0.03	-0.31	-0.41	0.05	1.00	-0.16	0.34	-0.12	0.15
Liquidity	0.23	-0.01	0.32	0.20	0.46	-0.05	-0.16	1.00	-0.02	0.12	-0.08
EarningsYield	-0.07	0.26	-0.07	-0.01	-0.22	0.05	0.34	-0.02	1.00	0.24	0.03
Growth	-0.01	0.09	0.01	0.06	0.05	0.02	-0.12	0.12	0.24	1.00	-0.04
Leverage	0.00	0.29	-0.11	-0.03	0.00	0.15	0.15	-0.08	0.03	-0.04	1.00

数据来源: Wind资讯, 东吴证券研究所

图8: 理想反转因子M剔除Barra因子与行业因子后的选股能力



数据来源: Wind资讯, 东吴证券研究所

第二,参数N的敏感度。本报告提供的切割方案,用于改进40日收益率因子(Ret40)与60日收益率因子(Ret60),同样效果出色。图9是参数N在三种取值下,原始因子与切割后因子的IC值。



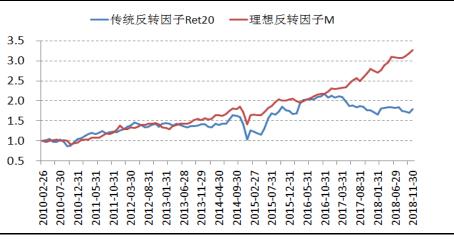
图9: N=20,40,60三种情况下的切割效果(月度IC的均值)

	Ret20	Ret40	Ret60
M_high	-0. 069	-0. 063	-0.064
M_low	0. 011	0. 016	0.018
原始因子	-0. 056	-0. 052	-0. 053

数据来源: Wind资讯, 东吴证券研究所

第三,其他样本空间的情况。理想反转因子在不同样本空间均表现优异。在沪深300成分股中:原始反转因子Ret20的五分组多空对冲年化收益7.2%,年化波动20.4%,信息比率0.35;理想反转因子的多空对冲年化收益15.0%,年化波动12.8%,信息比率1.17。在中证500成分股中:原始反转因子Ret20多空对冲年化收益13.9%,年化波动15.8%,信息比率0.88;理想反转因子的多空对冲年化收益16.3%,年化波动8.2%,信息比率2.00。为方便直观对比,图10给出了在沪深300成分股中,原始反转因子Ret20与理想反转因子M的多空对冲曲线。

图10: 沪深300成分股上的选股能力 (多空对冲净值)



数据来源: Wind资讯, 东吴证券研究所

第四,因子收益的累积过程。在本报告中,因子回测均采用月频调仓。读者可能关心更高频率的交易效果,图11我们展示了N=20时理想反转因子在月初建仓后(全市场股票、分五组),多空对冲收益的累积过程。由于收益累积过程比较均匀,我们定性地判断,可以尝试做周频调仓或半月调仓。

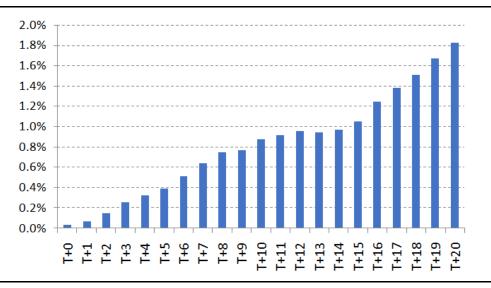
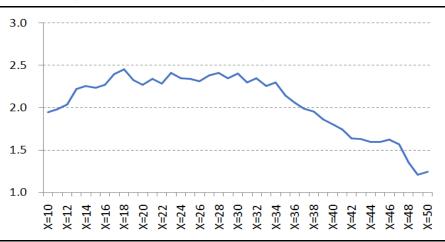


图11: 多空对冲收益的累积过程(T+0为月初建仓日)

数据来源: Wind资讯, 东吴证券研究所

第五,高D组与低D组的分组比例。在W式切割方案中,高D组与低D组的交易日,各占回溯交易日的一半,也即N/2个。如果调整分组的比例,效果会有多大的区别呢?我们以N=60为例,将单笔成交金额大的X个交易日作为高D组,将剩余60-X个交易日作为低D组,遍历X的值,分别计算M因子的信息比率,结果如图12所示。不难发现,X在30附近取值,都有很好的选股效果,这个结论支持了"对半分组"的简易做法。

图12: 分组比例的影响(纵坐标为M因子的信息比率)



数据来源: Wind资讯, 东吴证券研究所

第六, 交易行为逻辑。

我们始终强烈关心量化模型背后的逻辑。在W式切割中,"按单笔成交金额对交易日进行分组",似乎在暗示:对于大单交易活跃(**单笔成交金额高**)的交易日,涨跌幅因子有**更强**的反转特性;相反,对于大单交易不活跃(**单笔成交金额低**)的交易日,涨跌幅因子有**更弱**的反转特性。



熟悉我们研究的读者,可能会联想到我们的另一项工作——反转因子的"日内切割"。 具体做法是,将每日涨跌幅分为5段:隔夜、第1小时、第2小时、第3小时、第4小时。 过去20日的隔夜涨跌幅"加总"成为M0因子,第1小时的涨跌幅"加总"成为M1因子, M2、M3、M4以此类推。我们的实证结论是:M0、M1有弱的动量特性,M2、M3、M4为反 转因子,其中M3的反转特性最强,对应每日下午1点-2点的时段。笔者曾经猜想,反转 强度日内差异的根源,可能与图13中展示的"单笔成交金额的日内模式"有关。对于全 市场平均而言,单笔成交金额在日内呈现"上午高、下午低、1点-2点为全天最低"的 特征。按照我们当时的这个猜想,单笔成交金额低的时段,反转特性反而是更强的。

图13: 单笔成交金额的日内模式



数据来源: Wind资讯, 东吴证券研究所

这么一来,对于"W式切割"与"日内切割"的解释,似乎产生了令人困惑的矛盾。 经过谨慎考虑,我们认为消除矛盾的方法是:承认"单笔成交金额越高,反转特性越强" 的基本假设。这个基本假设直接构成了W式切割的交易行为逻辑。对于日内切割的情形, 需要特别注意的是,图13给出的只是全市场的统计结果,也就是说,"下午1点-2点单笔 成交金额最低"是就全市场平均而言,而不是对"每只股票、每个交易日"都能够成立。 从这一点看,"日内切割"与基本假设,并没有产生直接的冲突。笔者对于日内切割的 解释,目前倾向于认为是由其他市场交易行为的"日内模式"导致,而非与大单分布直 接相关。

最后,我们想讲的是:从实用主义的角度讲,W式切割与日内切割对反转因子的改进效果都相当出色,两者单独使用或联合使用,都是好的选择。此时此刻,如果读者的因子库里,还放着传统反转因子(如Ret20),可以把它扔进垃圾桶里了。

6. 风险提示

模型的测算基于历史数据,市场未来可能发生较大变化。



附注:

[1]这里所说的"加总", 实际上是通过累乘实现, 即:

M_high=(1+R1)*(1+R2)*...*(1+RN/2)-1 (在高D组交易日上进行累乘)

M_low=(1+R1)*(1+R2)*...*(1+RN/2)-1 (在低D组交易日上进行累乘)

[2]分组指标的寻找,并非一步到位,我们也尝试过其他分组方式,比如按"成交金额"或"成交笔数"分组。下表给出了不同分组方式的效果比较(rankIC均值):

	按成交笔数分组	按成交金额分组	按单笔成交金额分组				
M_high	-0. 059	-0. 066	-0. 082				
M_low	-0. 007	0. 007	0. 018				
Ret20	-0. 061						

[3]从图4中,很容易注意到,M_high和M_low在回归剔除Ret20之后,一个是强反转,一个是强动量,选股能力大致是对称的。这个结果几乎是必然的。在这里,图3的图景可以为我们提供判断的直觉:所谓"回归剔除Ret20因子"的操作,实际上是将水位线从原来位置调整到橘色圆点所在的高度;显然,红色圆点与蓝色圆点关于调整后的水位线是上下对称的。



免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准,已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司(以下简称"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客户。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,本公司不对任何人因使用本报告 中的内容所导致的损失负任何责任。在法律许可的情况下,东吴证券及其所属关 联机构可能会持有报告中提到的公司所发行的证券并进行交易,还可能为这些公 司提供投资银行服务或其他服务。

市场有风险,投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息,本公司力求但不保证这些信息的准确性和完整性,也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报告。

本报告的版权归本公司所有, 未经书面许可, 任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发、转载, 需征得东吴证券研究所同意, 并注明出处为东吴证券研究所, 且不得对本报告进行有悖原意的引用、删节和修改。

东吴证券投资评级标准:

公司投资评级:

买入: 预期未来6个月个股涨跌幅相对大盘在15%以上;

增持: 预期未来6个月个股涨跌幅相对大盘介于5%与15%之间:

中性: 预期未来 6个月个股涨跌幅相对大盘介于-5%与5%之间:

减持: 预期未来 6个月个股涨跌幅相对大盘介于-15%与-5%之间;

卖出: 预期未来 6个月个股涨跌幅相对大盘在-15%以下。

行业投资评级:

增持: 预期未来6个月内, 行业指数相对强于大盘5%以上;

中性: 预期未来6个月内, 行业指数相对大盘-5%与5%;

减持:预期未来6个月内,行业指数相对弱于大盘5%以上。

东吴证券研究所

苏州工业园区星阳街5号

