

寻找股潮澎湃下的暗流

2011 年 5 月 23 日

—— 资金流特征分析

从市场微观结构理论入手，首先提出了资金流信息含量的概念，并引入资金流的强度、杠杆倍数等信息指标，全面系统地分析了资金流的特点及对未来收益率的影响。

- **资金流原理**，资金流是一种反映股票供给信息的指标。传统的量价无法区分流动性和私有信息对股价的影响，而市场微观结构理论认为根据委托流测算的资金流能够有效反映潜在私有信息：资金流入（流出）导致价格上涨（下跌），净流入（净流出）越大，则对股票价格上涨（或者下跌）影响越大。
- **CMSMF 测算模型**，资金流测算方法很多，其结果也大不相同，招商证券资金流模型（CMSMF）采用高频数据进行测算（leve 1 为每 5 秒一个数据包），同时采取报价（最优买价、卖价）作为比较基准，成交价大于等于上期最优卖价时为流入；成交价小于等于上期最优买价则为卖出，最后流入减流出即为资金流净额。
- **资金流的信息含量**，本文首先引进资金流信息含量的概念，即为资金流净额除以交易金额（其他报告中将此公式定义为标准化资金流是存在误区的），用来度量信息含量的高低。信息含量较低，则受噪音干扰较大，导致测算的资金流存在误差，是无效信息；超过 10% 的信息含量的资金流数据才是有效信息。
- **资金流的标准化**，与其他报告不同，我们采用资金流除以流通市值来标准化资金流（即资金流强度），可以有效地规避不同流通市值大小股票的差异，同时为了表明资金流导致市值变化的效果，引入资金流杠杆倍数的指标。实证表明：通过资金流强度、杠杆倍数与当期收益率的回归拟合优度 R^2 得到了明显提高。
- **资金流指标预测效果**，通过资金流强度、杠杆倍数、信息含量指标以及当期收益率，能够较好地预测对下期收益率。资金流信息含量越高、持续的时间越长，其预测效果越好，说明信息含量和持续时间能够表明信号的强度和持续性，给市场造成深远影响。
- **市场对信息过度反应**，资金流强度指标的回归系数均值在不同信息含量和时间区间内都为负值，即大部分情况下，股票的当期资金流则对其未来收益率产生相反的作用。根据市场微观结构的信息模型解释为：在非有效市场，知情交易者会利用私有信息，引诱其他投资者，导致股价对潜在信息过度反应。

罗业华

0755-25310140
luoyh@cmschina.com.cn
S1090511040010

杨向阳

0755-82943205
yangxy@cmschina.com.cn
S1090511040036

研究助理

陈军华

chenjh2@cmschina.com.cn

正文目录

一、 概述.....	3
1. 资金流原理.....	3
2. 资金流测算方法.....	3
二、 资金流测算模型.....	5
1. 资金流测算模型.....	5
2. 数据说明.....	5
三、 资金流特征分析.....	7
1. 资金流信息假设.....	7
2. 资金流的信息含量.....	7
3. 资金流的信号处理.....	9
四、 资金流信息预测.....	11
4. 市场微观结构理论.....	11
5. 实证分析.....	11
五、 结论.....	14

图表目录

表格 1: 资金流测算模型的种类.....	3
表格 2: 资金流入流出计算方法.....	5
表格 3: 当日资金流净额与股票收益率的拟合优度 R^2	9
表格 4: 当期（多日）资金流净额与股票收益率的拟合优度 R^2	9
表格 5: 当期（多日）资金流净额与股票收益率的拟合优度 R^2	10
表格 6: 当期资金流强度与下期收益率的拟合优度 R^2	12
表格 7: 当期资金流强度、收益率与下期收益率的拟合优度 R^2	13
表格 8: 当期资金流强度、杠杆倍数、信息含量、收益率与下期收益率的 R^2	13
表格 9: 当期资金流强度与股票收益率的 $\beta 1$	14
图表 1: 不同时间频率下测算的资金流量.....	6
图表 2: 日内资金流净额与股票价格变化图.....	6
图表 3: 2010 年资金流净额与股票价格变化图.....	6
图表 4: 资金流正态密度曲线直方图.....	8

一、概述

巴菲特曾说过：“股市从短期看是投票机，但从长期看一定是称重机”。有些投资者对股票基本面情况进行分析和预测，来判断股票的长期价值。而有些人则通过技术指标来进行短期投机交易。不管是长期还是短期，投资者必须通过投票机的方式来交易股票，而这种投票行为反映潜在的信息。股市高频数据行情准确地记录了交易者每天数以亿万计的投票数据，我们将这些涓涓细流的委托信息，按照驱动股票方向的委托流汇总成资金流净额，揭示了潜藏在股潮澎湃下的暗流，这些主导了个股涨跌和大盘此起彼伏，值得我们深入地研究。

1. 资金流原理

资金流 (Money Flow) 是一种反应股票供求信息的指标。传统的量价无法区分市场微观结构中的流动性和私有信息对股价的影响。而根据委托流测算的资金流，能够有效地观察微观市场交易者的真实意图以及对价格造成的影响，因此资金流可能反映潜在私有信息。信息模型理论把投资者分为不知情交易者（噪音交易者）以及知情交易者，不知情交易者总是基于投机或者流动性而交易，没有明显的方向性，在没有新信息传到市场时，其买卖价格也是基本均衡的；而知情交易者总是基于某一个方向的交易，而且在当前价格没有反应私有信息时，倾向于采取持续交易的方式。这样就造成买卖单不均衡（order imbalance），即资金流入流出失衡，资金流入（流出）导致价格上涨（下跌）。因此净流入（净流出）越大，则对股票价格上涨（或者下跌）影响越大。

2. 资金流测算方法

严格意义上讲，每一个买单必须要有一个相应的卖单，因此真实的资金流入无法准确计算，只能通过其他替代方法来区分资金的流入和流出，通过高频数据，将每笔交易按照驱动股价上涨和下跌的差异，确定为资金的流入或流出，最终汇聚成一天的资金流净额数据。测算资金流的方法很多，主要差别：一是在时间频率上，二是在参考价格上（见表格 1），比较流行的方法分类有以下几种：

表格 1：资金流测算模型的分类标准

	类型	说明	优点	缺点
频率	日数据	日数据计算，	计算简单，国内外应用较广	丢失许多微观信息
	高频数据	分时或者分笔数据进行计算	日内信息较多	计算量较大
参考 价	报价	如果成交价大于等于卖方最优价则为流入，如果成交价小于等于买方最优价则为流出	分类较细，适合采取分笔数据测算	损失位于买卖最优价中间价的交易量信息
	成交价	如果大于前成交价则为流入，小于前期成交价则为卖出。	分类较细，适合采取分时数据测算	损失成交价相等的交易量信息
	其他方法	分时收盘价、成交均价等	能够根据市场实际效果进行取舍，信息含量高	比较复杂

资料来源：招商证券

在测算资金流时时间频率的选择非常重要，目前有些是采取日数据计算，即当日价格上涨则全部交易量计为流入，否则计为流出；而有些是采取高频数据计算，比如每笔数据

或者 1 分钟数据等时间间隔，在选择时间频率上的差异会导致测算的资金流净额存在差异（如图表 1），从 5 秒到 1 天的时间间隔，计算出来的交易量。

在参考价格上，有些采取前收盘价格，有些采取最近成交价，或者前一个（或者后一个）报价，参考价的选择不同，也会导致差异较大。通常区分资金流入流出时，按照交易价格与之前成交价比较，如果大于等于前成交价则为流入，小于等于前期成交价则为卖出，对于分时数据测算时，使用该方法较好，但是其损失部分相等价格的交易量。也有采取成交价与最近报价进行比较，如果成交价大于等于卖方最优价则为流入，如果成交价小于等于买方最优价则为流出。然而指令驱动市场，报价是实时匹配的，有可能买卖价格同时到达而配对成功，尤其在 A 股市场报价不是实时广播的，而是间隔 3-5 秒生成成交均价和报单行情，这样导致采取该种策略，许多成交价位于最优买卖报单之间的交易量未被统计进来。

二、 资金流测算模型

1. 资金流测算模型

目前采用高频数据进行资金流测算，主要是从如下几个方面考虑：第一，采用高频数据进行测算（leve 1 为每 5 秒一个数据包），这样能够尽可能多的反应真实的市场信息；第二采取报价（最优买价、卖价）作为比较基准，成交价大于等于上期最优卖价时为流入；成交价小于等于上期最优买价则为卖出；即传统意义上的主买和主卖，这样能反映交易者主导买卖的行为和对股价的影响。其他成交价格不计入资金流。该方法命名为招商证券资金流模型（简称 CMSMF，见表格 2）。

表格 2：资金流入流出计算方法

	类型	条件说明
开盘集合竞价	流入	集合竞价的成交价大于昨收盘价时的交易金额
	流出	集合竞价的成交价小于昨收盘价的交易金额
连续竞价	流入	成交价大于等于最近卖方最优价的交易金额
	流出	成交价小于等于最近买方最优价的交易金额
涨跌停	流入	涨停时成交金额
	流出	跌停时成交金额
尾盘集合竞价	流入	集合竞价的成交价大于最近卖方最优价的交易金额
	流出	集合竞价的成交价小于最近买方最优价的交易金额

资料来源：招商证券

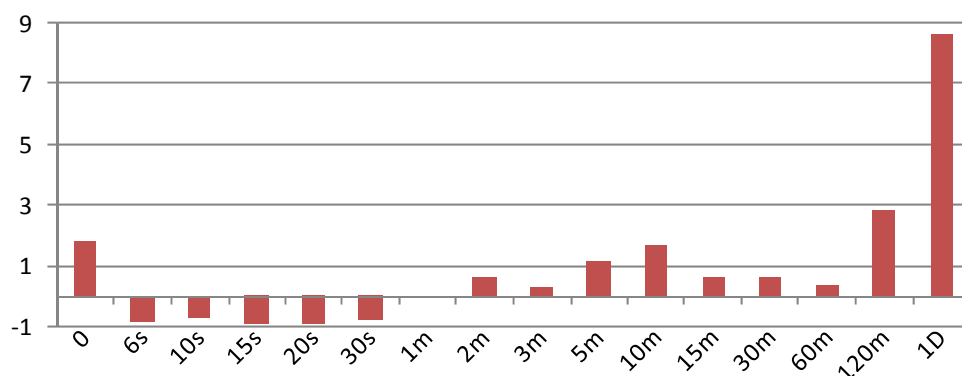
2. 数据说明

目前的高频数据最小间隔时间（5 秒），每只股票每日大约 2900 个数据点，按照 CMSMF 模型对每个数据点进行测算，得出流入、流出量，流入减流出即为当日资金流净额。目前测算了 1600 多只股票在 2005 年到 2010 年的每日资金流净额数据，共计 2800 多个日数据，以下的实证分析都是采用日资金流金额进行测算的。

我们摘取浦发银行（600000）2010 年 12 月 31 日数据（见图表 1）进行分析，按照 CMSMF 模型测算的结果与其他方法（采用前交易价格作为参考价的分时方法计算）计算的结果存在差异，从图中可以看出，采取分时的方法，在秒级别时流入净额为负，而分钟级别及一天时，其流入为正，CMSMF 模型测算的结果也为正，但与其他方法测算存在差别，原因很可能是因为资金流信息含量不高，噪音导致各种算法误差较大。

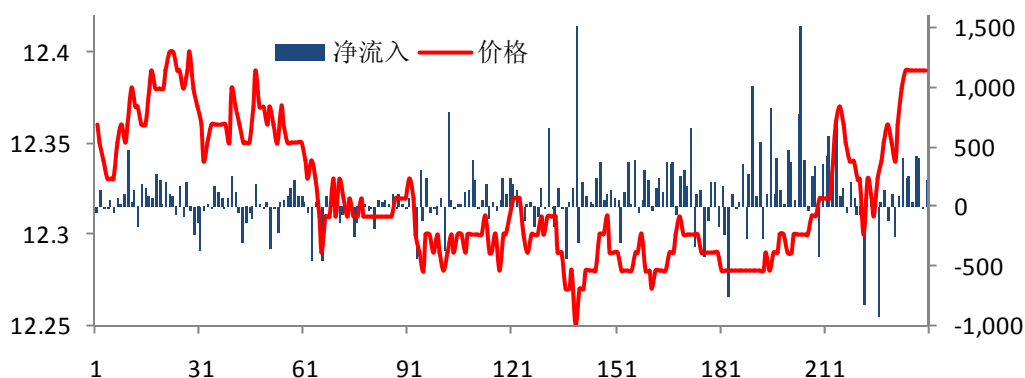
粗略地浏览资金流的反映信息情况，图表 2 是日内每分钟的股价变化与资金流入流出对比，从中可以看出资金流入流出对市场价格造成了影响，流出较多时，股票价格上涨。从日数据上看（见图表 3），2010 年武钢股份的涨跌与资金流有一定关系。

图表 1: 不同时间频率下测算的资金流量



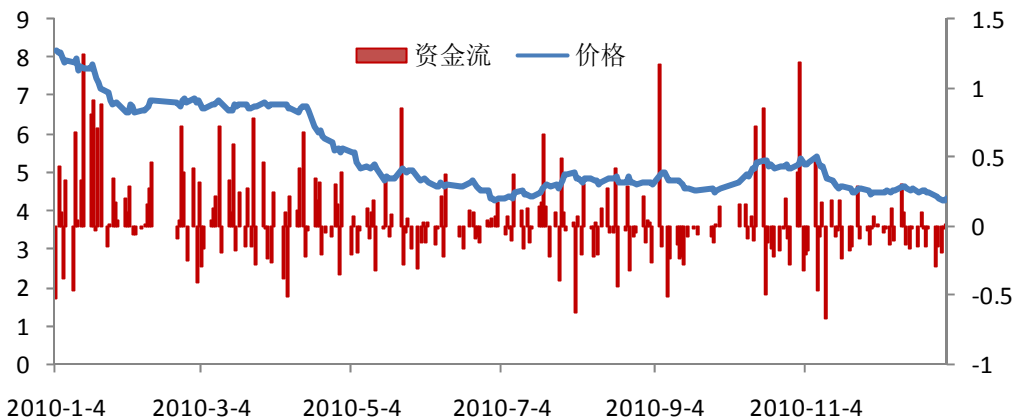
注: 根据浦发银行 (600000) 在 2010 年 12 月 31 日的高频数据测算, 其中 0 表示 CMSMF 模型测算的数据, 其他为按照涨跌幅的分时数据测算的结果, s 表示秒, m 表示分钟, D 表示天, 单位为亿。
资料来源: 招商证券

图表 2: 日内资金流净额与股票价格变化图



注: 根据浦发银行 (600000) 在 2010 年 12 月 31 日的高频数据测算, 价格 (左轴) 为日内每分钟收盘价; 资金流净额 (右轴, 单位万元) 表示根据 CMSMF 模型测算的每分钟流入流出净额。
资料来源: 招商证券

图表 3: 2010 年资金流净额与股票价格变化图



注: 根据武钢股份 (600005) 在 2010 年的日数据测算, 价格 (左轴) 为日收盘价; 资金流净额 (右轴, 单位亿元) 表示根据 CMSMF 模型测算的每日流入流出净额。
资料来源: 招商证券

三、 资金流特征分析

1. 资金流信息假设

通过上述测算方法得到股票的资金流净额的数据，这是从市场直接获取的相当重要的潜在信息。许多其他机构也提供资金流的测算和做一些简单的处理，对投资缺乏实质性的指导。我们认为直接研究绝对金额的资金流缺乏实际意义，所以必须对资金流信息进行分析处理，找出有价值的指标，现提出以下问题：

1. 资金流是否反映潜在信息？也就是说资金流的信息含量应该怎样鉴定，由于资金流信息是通过高频数据测算的，如果信号本身比较微弱，由于噪音的干扰可能导致测算结果有较大误差。因此如何度量信息含量、以及获取有效信息是我们首要研究的问题。
2. 资金流的强度如何度量？一般认为资金流净额越大其对股价影响也就越大，即信号越强则可能作用的效果越好，用绝对金额可能忽略了股票市值的差异，所以必须结合市场实际情况进行标准化。
3. 资金流的效应怎样？由于信息的含量不一样，信息的传播和暴露程度不一样，以及流动性的差异，资金流对市值变化的影响效应也会不一样，有些时候很小的资金流量就能撬动大的市值变动，有些则遇到较大的阻力，我们将研究资金流的效应。
4. 资金流信息对未来收益率的影响如何？如果资金流反映潜在的信息，那么我们就能够根据现有的资金流信息，建立有效的模型对未来收益率进行预测？所以必须研究资金流信息的特点以及对未来收益率的影响。

2. 资金流的信息含量

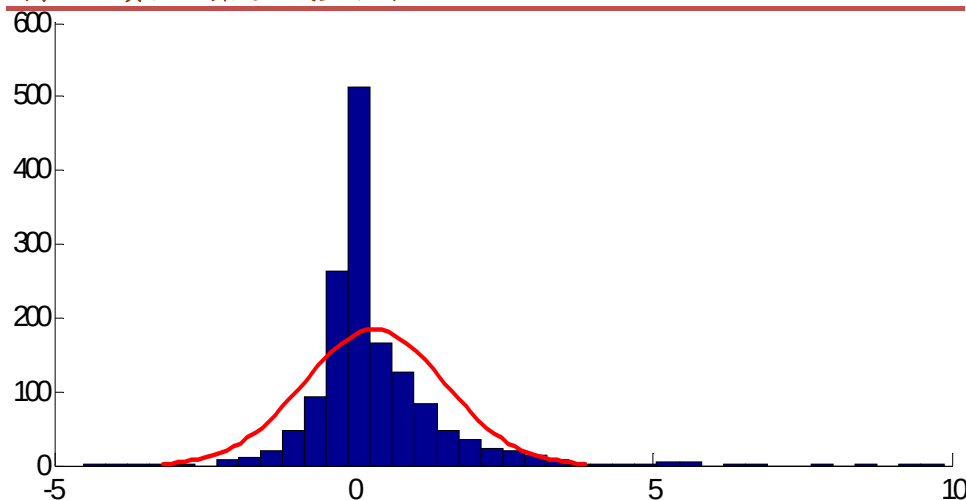
1) 信息含量度量

通过对历史每日资金流净额的密度分布进行分析（见图表4），发现浦发银行（600000）的资金流信息在历史6年中中有三分之二的的数据（1000/1500）位于0附近，即大多数时候资金信息都比较微弱，可能存在噪音的干扰。资金流信息是通过高频数据测算的，如果信号本身比较微弱，噪音的干扰可能导致测算值的误差较大，对其流入流出的方向性判断产生根本性影响。因此必须对信息含量进行分析和度量，信息含量进行定义为：资金流金额与交易金额比值的绝对值（公式1），我们认为比值越高，则有效信息越多，如果比值较低则是无效信息，应该剔除。

$$IC_{t,i} = \text{abs}(MF_{t,i} / \text{Amount}_{t,i}) \quad (1)$$

其中， $IC_{t,i}$ 表示第t期股票i的信息含量； $MF_{t,i}$ 表示第t期股票i的资金流净额； $\text{Amount}_{t,i}$ 表示第t期股票i的交易金额；abs求绝对值函数。

图表 4: 资金流密度曲线直方图



注：根据浦发银行（600000）在2005年1月至2010年12月的资金流净额计算，共计1456个数据，单位为亿。

资料来源：招商证券

2) 当日信息含量分析

对信息含量进行测试，确定怎样的信息是有效的。通常用回归分析方法，日股票收益率作为因变量，资金流净额作为自变量，用因子解释的比重（即统计学中常用的拟合优度 R^2 ）来衡量信息的含量。首先将资金流按照信息含量高低进行分组，再对分组后的资金流与股票收益率做回归分析（见公式 2），得出全部样本股票不同信息含量组别的平均拟合优度 R^2 。

$$R_{t,i} = MF_{t,i} * \beta_i + \alpha_i \quad (2)$$

其中， $R_{t,i}$ 表示第 t 期股票 i 的收益率； $MF_{t,i}$ 表示第 t 期股票 i 的资金流净额； β_i 表示股票 i 的回归系数。 R^2 为回归的拟合优度，用来检验自变量对因变量的解释程度，越高越好。

从测试的结果看（见表 3），信息含量在 0~5% 之间的组别平均 R^2 最差，只有 5% 左右，随着含量的增加， R^2 逐渐变大，信息含量超过 10% 之后， R^2 维持在 30% 以上。另外从市值大小的维度上看， R^2 的大小没有明显的差异，也就是说资金流信息含量相同时，对大小盘股票的解释作用是无差异的。

表格 3: 当日资金流净额与股票收益率的拟合优度 R^2

	0~5%	5~10%	10~15%	15~20%	20~30%	30~50%	50%~100%	0~100%
1	6.58%	22.78%	35.67%	42.63%	44.09%	49.18%	39.90%	30.65%
2	7.02%	22.43%	35.52%	41.82%	43.63%	49.43%	48.34%	30.77%
3	7.91%	21.15%	35.18%	43.11%	43.40%	48.65%	41.68%	30.28%
4	6.46%	21.86%	36.11%	41.89%	41.62%	47.06%	42.75%	29.61%
5	7.07%	21.03%	34.23%	39.50%	41.16%	48.26%	42.66%	28.91%
6	7.80%	26.16%	35.29%	42.01%	43.69%	47.25%	43.78%	30.73%
7	6.67%	24.18%	33.36%	42.68%	43.75%	48.70%	39.98%	28.97%
8	7.00%	24.83%	37.18%	40.49%	43.82%	48.92%	44.08%	30.09%
9	7.33%	22.91%	34.38%	40.74%	44.46%	46.39%	42.57%	28.64%
10	8.07%	20.05%	33.87%	39.97%	44.37%	48.00%	39.82%	29.03%
All	7.16%	22.89%	35.15%	41.75%	43.94%	48.48%	42.97%	30.25%

注: 回归时, 按照信息含量进行分组统计, 得出拟合优度。再将全市场股票市值按照从小到大分为 10 组, 第 10 组表示市值最大, all 表示全市场, 求每组的平均拟合优度。

资料来源: 招商证券

3) 多日信息含量分析

再按照时间间隔分为 8 组期限 (5 天, 10 天~40 天), 计算当期多日资金流净额总和与当期股票收益率的回归的拟合优度 (见表 4), 其信息含量越高的组别的平均 R^2 的值越高, 信息含量在 10% 之后, R^2 的大小维持在 50% 左右, 保持稳定。从时间间隔维度上看, 随着间隔时间的延长, R^2 逐渐变大, 但是总体差异不明显。

表格 4: 当期 (多日) 资金流净额与股票收益率的拟合优度 R^2

	0~5%	5~10%	10~15%	15~20%	20~100%	30~50%	50%~100%	0~100%
5	9.52%	29.09%	38.73%	45.70%	49.76%	51.33%	36.73%	28.33%
10	10.30%	28.82%	38.42%	46.52%	52.01%	50.33%	33.89%	26.55%
15	10.71%	28.55%	39.64%	49.77%	53.83%	48.57%	32.93%	25.88%
20	10.92%	29.18%	42.38%	50.19%	53.04%	52.18%	31.61%	25.66%
25	10.99%	29.53%	43.18%	49.26%	57.08%	39.23%	28.80%	25.22%
30	11.12%	29.78%	43.71%	51.19%	59.73%	34.03%	31.14%	24.72%

注: 按照交易日期分为 8 组, 第 1 组 5 表示间隔为 5 天。5-40 天为取样的时间段, 单位为天。

资料来源: 招商证券

3. 资金流的信号处理

一般认为资金的流入流出导致了股价的涨跌, 且净额越大其对股价影响也就越大。但是同样的资金流量净额为 1 亿元, 对于市值为 1000 亿的股票和 100 亿的股票的作用肯定

不一样,因此用绝对金额可能忽略了股票市值的差异,所以必须除以流通市值做标准化,我们称之为资金流强度指标(公式3)。资金流的流入、流出对股价会产生影响,其影响的效果就是流通市值的变动,但是同样的资金流带来不同效果,也就是说市场本身存在阻力,我们引入资金流杠杆倍数来度量资金流的效应,即流通市值变化除以资金流的比值(公式4),越高说明效应越好。

$$MFP_{t,i} = MF_{t,i} / Value_{t,i} \quad (3)$$

其中, $MFP_{t,i}$ 表示第 t 期股票 i 的资金流强度 (Money Flow Power); $MF_{t,i}$ 表示第 t 期股票 i 的资金流净额; $Value_{t,i}$ 表示第 t 期股票 i 的流通市值。

$$MFL_{t,i} = \Delta Value_{t,i} / MF_{t,i} \quad (4)$$

其中, $MFL_{t,i}$ 表示第 t 期股票 i 的资金流杠杆倍数 (Money Flow Lever); $MF_{t,i}$ 表示第 t 期股票 i 的资金流净额; $\Delta Value_{t,i}$ 表示第 t 期股票 i 的流通市值的变化额。

通过对资金流强弱进行标准化之后,再将资金流强度 (MFP)、杠杆倍数 (MFL) 与当期收益率做回归分析 (见公式5),其 R^2 得到了明显提高,说明标准化后的资金流比直接使用资金流净额数据要好,更符合市场实际情况。

$$R_{t,i} = MFP_{t,i} * \beta_{1,i} + MFL_{t,i} * \beta_{2,i} + \alpha_i \quad (5)$$

其中, $R_{t,i}$ 表示第 t 期股票 i 的收益率; $MFP_{t,i}$ 表示第 t 期股票 i 的标准化资金流; $MFL_{t,i}$ 表示第 t 期股票 i 的资金流杠杆倍数, $\beta_{1,i}$ 、 $\beta_{2,i}$ 表示股票 i 的两个因子的回归系数。 R^2 为回归的拟合优度,用来检验自变量对因变量的解释程度,越高越好。

表格 5: 当期 (多日) 资金流强度、杠杆倍数与股票收益率的拟合优度 R^2

	0~5%	5~10%	10~15%	15~20%	20~30%	30~50%	50%~100%	0~100%
5	13.87%	41.41%	55.84%	65.89%	70.49%	71.40%	61.06%	39.19%
10	14.97%	42.63%	57.49%	66.92%	71.64%	74.64%	56.08%	36.83%
15	15.47%	43.54%	58.86%	69.57%	74.06%	75.48%	48.34%	35.65%
20	15.87%	44.85%	61.13%	72.32%	76.31%	81.51%	48.46%	34.82%
25	15.89%	45.97%	63.58%	72.77%	79.58%	83.27%	52.70%	33.60%
30	15.87%	47.45%	65.97%	74.42%	84.09%	85.12%	61.79%	32.45%
35	15.92%	49.28%	68.93%	75.69%	85.77%	88.80%	63.47%	31.42%
40	15.95%	50.60%	71.84%	79.10%	87.12%	88.04%	70.47%	29.94%

注: 按照交易日间隔分为 8 组, 第 1 组 5 表示间隔为 5 天。5-40 为取样的时间段, 单位为天。
资料来源: 招商证券

从统计的结果看,较之前单一未标准化的资金流指标回归,其拟合优度 R^2 都有明显提高,说明加入两个标准化的因子之后,其回归的解释度得到明显提高。同时,信息含量大于 5% 之后,其各个区间的 R^2 都比较好,因此接下来的分析都将 5% 以内的信息含量样本去掉。

四、 资金流信息预测

4. 市场微观结构理论

资金流能够反映潜在私有信息，那它是如何作用于市场并影响未来收益的呢，即在普遍情况下，前期资金持续流入是否会导致未来市场上涨呢？首先我们必须从市场微观结构理论框架下来研究私有信息是如何注入市场和产生影响的。在非有效的市场，由于信息成本、流动性效应等因素的存在，导致价格对私有信息的反应是逐步缓慢实现的。信息模型理论认为市场存在知情交易者和非知情交易者，根据信息暴露程度和反应效果可分为以下三种情况：

1. **信息反应不足**：如果只存在市场垄断性的知情交易者，那么他会尽量控制交易节奏来降低信息泄露，以及尽量多地在有利价位交易足够的数量，并在两者之间进行平衡以获取利益最大化，待信息披露之后，价格才会迅速反应。
2. **信息正常反应**：实际市场是存在信息竞争者，在价格没有反映潜在信息之前，多个知情交易者必须尽可能多地抢先交易，这就不可避免地造成信息暴露，也就会体现市场持续性的资金流入流出。如果大家都是理性投资者，则价格反映潜在信息之后，资金流持续流便会停止，未来价格便处于均衡状态，资金流的持续性关键因素是潜在信息注入市场的速度和效率。
3. **信息反应过度**：也有研究表明，知情交易者还能够“两次”利用私有信息，如果提前收到利好信息，就在真实信息公告之前买入，当预测市场会在公告之时反应过度，就选择在公告之时卖出。这个实现的前提是市场信息效率较低的非强式有效市场，例如：前期 A 股市场的“做庄”者，故意拉抬或打压股票暴露信息，促使中小投资者追涨杀跌，知情交易者利用自己私有信息中的噪声误导其他投资者，以达到浑水摸鱼的效果。

既然我们分析了三种信息注入市场的情况情况，那么我们对每种情况进行比较分析，通过对资金流信息与未来收益的回归分析，可得出以下假设推论，以待后续验证：

1. 如果是第一种情况，则我们无法从资金流信息中获得必要的信息，也就无法作出准确的预测，表现的形式就是回归预测的拟合优度非常低，预测效果极差。
2. 如果是第二类情况，则可以从资金流中获得必要的信息，信息是逐步注入市场的，也不存在过度反映和诱饵的出现，因此资金流的持续性处决于信息注入市场的速度和市场定价的效率。但是不管怎样，前期资金流的流入对未来有一定的正面效应。回归时的拟合优度较好，回归系数为正值。
3. 如果是第三类情况，则证明市场的定价效率较低，知情交易者可以多次利用潜在信息，引诱或者打压投资者导致股价过度反映，因此如果当期信号强烈，则未来可能会出现反转，表现的形式是拟合优度较好，回归系数为负值。

5. 实证分析

1) 回归公式

为了确定 A 股市场股票到底属于那种情况。我们将资金流指标与下一期收益率做回归分析，建立以下三个回归分析公式，公式 6 主要单独检验 MFP 对未来收益率的影响，公

式 7 主要增加了当期收益率的影响，公式 8 则加入所有的信号指标。最后测算不同回归公式的 R^2 得在不同信号强度下及不同间隔时间内的平均值，以及各个回归系数的平均值。

$$R_{t+1,i} = MFP_{t,i} * \beta_{1,i} + \alpha_i \quad (6)$$

$$R_{t+1,i} = MFP_{t,i} * \beta_{2,i} + R_{t,i} * \beta_{3,i} + \alpha_i \quad (7)$$

$$R_{t+1,i} = MFP_{t,i} * \beta_{4,i} + MFL_{t,i} * \beta_{5,i} + R_{t,i} * \beta_{6,i} + IC_{t,i} * \beta_{7,i} + \alpha_i \quad (8)$$

其中， $R_{t,i}$ 表示第 t 期股票 i 的收益率； $R_{t+1,i}$ 表示第 $t+1$ 期股票 i 的收益率； $MFP_{t,i}$ 表示第 t 期股票 i 的标准化资金流； $\beta_{1,i}$ 、 $\beta_{2,i}$ 、 $\beta_{3,i}$ 、 $\beta_{4,i}$ 、 $\beta_{5,i}$ 、 $\beta_{6,i}$ 表示股票 i 在各个公式因子的回归系数。 R^2 为回归的拟合优度，用来检验自变量对因变量的解释程度，越高越好。

2) 拟合优度

从根据公式 6 的回归结果看（见表格 6），其平均拟合优度 R^2 还是比较低，信息含量越高，其值逐渐变大。在考虑到资金流在作用当期股价时，影响的效果可能不同，公式 6 的回归公式把当期收益率作为自变量加入，实测证明平均 R^2 在不同信号强度下及不同间隔时间内都得到了明显提高（见表格 7）。

表格 6：当期资金流强度与下期收益率的拟合优度 R^2

	0~5%	5~10%	10~20%	20~30%	30~40%	40~50%	0%~100%
5	3.73%	4.58%	14.42%	17.18%	18.40%	29.56%	1.26%
10	2.78%	3.62%	12.73%	16.73%	17.68%	30.76%	0.95%
15	2.52%	3.78%	13.31%	16.09%	17.66%	32.45%	0.88%
20	2.69%	4.03%	13.13%	16.52%	18.37%	28.93%	0.94%
25	3.49%	4.74%	13.39%	17.98%	19.29%	26.56%	1.09%
30	3.98%	4.90%	13.38%	17.89%	19.49%	27.50%	1.07%
35	3.98%	5.38%	13.59%	18.71%	20.60%	26.47%	1.14%
40	4.23%	5.68%	13.76%	18.74%	22.69%	25.90%	1.22%

注：按照交易日期分为 8 组，第 1 组 5 表示间隔为 5 天。5-40 为取样的时间段，单位为天。

资料来源：招商证券

表格 7: 当期资金流强度、收益率与下期收益率的拟合优度 R^2

	0~5%	5~10%	10~20%	20~30%	30~40%	40~50%	0%~100%
5	2.53%	4.06%	7.09%	13.10%	15.64%	18.36%	1.33%
10	2.21%	4.05%	8.04%	15.41%	17.39%	19.66%	1.46%
15	2.47%	4.75%	11.08%	19.97%	19.79%	22.78%	1.90%
20	2.73%	5.97%	14.70%	23.26%	22.46%	18.57%	2.33%
25	3.22%	7.02%	17.58%	24.17%	24.94%	25.22%	2.80%
30	4.08%	8.38%	20.33%	28.94%	30.87%	31.67%	3.45%
35	5.47%	9.75%	24.77%	31.95%	31.03%	38.56%	4.34%
40	6.42%	11.31%	28.45%	34.04%	34.96%	43.57%	4.86%

注: 按照交易日期分为 8 组, 第 1 组 5 表示间隔为 5 天。5 - 40 为取样的时间段, 单位为天。

资料来源: 招商证券

为了获得更多信息, 我们试着将资金流杠杆倍数、信息含量以及当期收益率加入回归公式 (公式 8) 来预测未来收益率, 结果表明平均拟合优度 R^2 得到很大的提高。具体从以下几个方面解读:

1. 在低信息含量 (10%以下) 组别中, 其平均 R^2 较低, 可能是无信息或者信息较微弱, 在噪音的干扰下泄露较少, 该种情况下无法预测未来收益率, 所以应该属于无效信息, 予以剔除。
2. 同一时间间隔内, 信息含量越高其平均 R^2 越高, 某些情况下平均 R^2 甚至超过了 50%, 说明私有信息暴露越多, 对未来收益率预测性越好。符合信息竞争理论, 各种投资者是信息竞争者, 大家都尽可能多的交易, 则信息就越多, 从而对未来预测性越好。
3. 同一信息含量下, 时间越长平均 R^2 越高, 则其对未来预测的效果更好。说明持续时间长的高信息含量, 给市场的信号越有力, 从而对未来预测性就越好。

表格 8: 当期资金流强度、杠杆倍数、信息含量、收益率与下期收益率的 R^2

	0~5%	5~10%	10~20%	20~30%	30~40%	40~50%	0%~100%
5	4.49%	6.56%	11.78%	22.60%	27.01%	35.90%	2.20%
10	3.73%	7.05%	14.47%	28.04%	32.65%	42.21%	2.48%
15	4.10%	8.39%	19.65%	35.22%	35.30%	48.92%	3.11%
20	4.68%	10.20%	25.32%	39.17%	40.28%	51.78%	3.73%
25	5.17%	11.73%	29.65%	41.26%	46.03%	60.37%	4.29%
30	6.17%	13.89%	35.20%	46.29%	51.46%	62.24%	5.19%
35	7.74%	16.25%	41.93%	51.18%	50.64%	67.22%	6.39%
40	9.03%	18.94%	46.44%	51.88%	55.47%	68.44%	7.16%

注: 按照交易日期分为 8 组, 第 1 组 5 表示间隔为 5 天。5 - 40 为取样的时间段, 单位为天。按照时间间隔和信息分组统计平均 R^2 。

资料来源: 招商证券

3) 回归系数

从公式 8 的各个资金流指标的平均回归系数看，其资金流强度的平均系数均为负，表明当期资金流对未来收益成反作用。该现象符合信息注入市场的第三种情况，即知情交易者两次利用了私有信息，首先利用私有信息提前买入（卖出）股票，然后引诱贪婪（恐慌）的散户投资买入（卖出），从而导致股票反应过度，最后反向操作，以达到获利的目的。尽管有部分股票的未来收益率跟其当期资金流是正相关的，但是在不成熟的 A 股市场，大部分股票都会对潜在的私有信息反应过度。

表格 9：当期资金流强度的回归系数 β_4

	10~20%	20~30%	30~40%	40~50%	0%~100%
5	-0.36	-0.46	-0.41	0.00	-0.32
10	-0.20	-0.38	-0.28	-0.28	-0.33
15	-0.06	-0.23	-0.13	-0.48	-0.24
20	-0.10	-0.29	-0.20	-0.50	-0.18
25	-0.25	-0.83	-0.18	-0.20	-0.25
30	-0.53	-0.29	-0.03	-0.25	-0.29
35	-0.64	-0.31	-0.20	-0.42	-0.31
40	-0.46	0.50	-0.20	-0.66	-0.20

注：按照交易日期分为 8 组，第 1 组 5 表示间隔为 5 天。5 - 40 为取样的时间段，单位为天。按照时间间隔分组统计平均回归系数。

资料来源：招商证券

五、 结论

从上述实证分析可以得到如下结论：

1. 通过不同资金流测算方法测算的资金流存在差异，必须谨慎选择资金流测算模型；我们根据中国市场特点，设计了一个独特的招商证券资金流测算模型（CMSMF），通过高频数据测算了的资金流数据。
2. 资金流的信息含量对研究资金流非常重要，如果资金流的信息含量较低，则可能受噪音干扰，导致测算结果存在误差。实测证明，信息含量比率越高，则其与当期收益率的回归拟合优度 R^2 越高，说明有效信息含量高，其中超过 10% 的信息含量才是有效的资金流信息。
3. 绝对资金流含量无法反映市值大小，必须进行标准化，将资金流除以流通市值的比值记作资金流强度。同时为了表明资金流导致市值变化的效果，引入资金流杠杆倍数的指标。实证表明，通过资金流强度、杠杆倍数与当期收益率的回归拟合优度 R^2 得到了明显提高。
4. 利用资金流指标预测未来收益率与信息含量有关，资金流信息含量越高、持续的时间越长，其预测效果越好，说明信息含量和持续时间能够给市场造成深远影响。
5. 资金流对未来收益率是负的效应，在散户较多的非有效市场，知情交易者会利用私有信息，引诱其他投资者，导致股价对潜在信息反应过度，形成收益率的反转效应。从而使股票的当期资金流则对其未来收益率产生相反的作用。

分析师承诺

负责本研究报告的每一位证券分析师，在此申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告中的具体推荐或观点直接或间接相关。

罗业华：江西财经大学计算机工学和金融经济学双学位，MBA，金融工程分析师。从业七年，曾主持开发股指期货套利系统、权证评价系统、金融工程实验室，目前工作是运用数理、编程知识，进行股指期货、量化投资、算法交易等策略研究。

杨向阳：西安交通大学计算数学理学硕士，信息与计算科学理学学士。07 年进入招商证券研发中心，主要研究方向包括数量化投资策略、投资组合构建及风险评估以及期权期货等金融衍生产品。

陈军华：华中科技大学管理科学与工程专业硕士、信息管理与信息系统专业学士，现招为招商证券研究发展中心金融工程分析师。

投资评级定义

公司短期评级

以报告日起 6 个月内，公司股价相对同期市场基准（沪深 300 指数）的表现为标准：

- 强烈推荐：公司股价涨幅超基准指数 20%以上
- 审慎推荐：公司股价涨幅超基准指数 5-20%之间
- 中性：公司股价变动幅度相对基准指数介于±5%之间
- 回避：公司股价表现弱于基准指数 5%以上

公司长期评级

- A：公司长期竞争力高于行业平均水平
- B：公司长期竞争力与行业平均水平一致
- C：公司长期竞争力低于行业平均水平

行业投资评级

以报告日起 6 个月内，行业指数相对于同期市场基准（沪深 300 指数）的表现为标准：

- 推荐：行业基本面向好，行业指数将跑赢基准指数
- 中性：行业基本面稳定，行业指数跟随基准指数
- 回避：行业基本面向淡，行业指数将跑输基准指数

重要声明

本报告由招商证券股份有限公司（以下简称“本公司”）编制。本公司具有中国证监会许可的证券投资咨询业务资格。本报告基于合法取得的信息，但本公司对这些信息的准确性和完整性不作任何保证。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。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价，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。除法律或规则规定必须承担的责任外，本公司及其雇员不对使用本报告及其内容所引发的任何直接或间接损失负任何责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突。

本报告版权归本公司所有。本公司保留所有权利。未经本公司事先书面许可，任何机构和个人不得以任何形式翻版、复制、引用或转载，否则，本公司将保留随时追究其法律责任的权利。