

深度报告

金融工程

量化投资

量化选股专题报告

2010年9月7日

本报告的独到之处

■从基本原理、样本选择、样本处理、适用范围等多个角度全面介绍了 Logistic 模型的体系;

■使用沪深 300 的样本数据,对 Logistic 模型所选股票的绩效进行了详细的分析;

■使用最新的财报数据,通过 Logistic 模型建模,给出了未来一个月建议重点配置的股票。

联系人: 戴军
电话: 0755-82133129
E-mail: daijun@guosen.com.cn

分析师: 黄志文
电话: 0755-82130833-6210
E-mail: huangzw@guosen.com.cn
SAC 执业证书编号: S0980206110185

分析师: 葛新元
电话: 0755-82130833-1870
E-mail: gexy@guosen.com.cn
SAC 执业证书编号: S0980200010107

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。

专题报告

Logistic 选股模型及其在沪深 300 中的实证

从多因素模型产生的思想来看,其目的主要是想解决哪些风险因素以及有多少因子对股票或债券报酬产生影响,以及影响有多大。多因素模型通常有 3 种模型, BARRA 模型、Northfield 模型和对无法识别的风险进行管理的统计模型。其中 BARRA 模型对证券基本面的风险分析做的比较出色, Northfield 模型对宏观经济因素的风险分析十分出色,这两种模型的原理基本相同。BARRA 模型是利用多因素模型的基本原理,识别出风险因子,然后利用这个模型对风险因子导致的风险进行识别并加以控制。

一般主要考虑部分宏观因子,基本面因子和部分技术因子,从中选取可能对投资组合收益率和风险有影响的因子进行降维处理,再进行回归分析,用加权最小二乘法估计回归方程的系数,在下一期的回归方程中保持系数不变,求出因变量(即收益)。但是,预期收益率的确切估计是件比较困难的事情,同时,在投资中的很多策略也许并不要求获取较为确切的预期收益率数据,比如公募基金跑赢跟踪基准的要求以及结合股指期货的 alpha 策略,它们要求的是所选择的股票未来跑赢基准的概率有多大,这一要求比预测股票未来的预期收益率大为降低。而 logistic 模型正是解决这一问题的利器。

样本内研究表明: Logistic 选股模型是一种能够带来超越基准收益,且较为稳定的模型,样本内年化收益率 1.69%,沪深 300 指数年化收益率为 -13.64%,超额收益率年化 17.53%。Logistic 选股模型可用于追求相对排名的公募基金以及追求绝对收益的私募基金(结合股指期货)。

按照最新财报数据,采用等权复合因子的方法进行降维,生成未来一个月的股票配置建议,四十只股票的行业分布中,所占比重最大的行业是机械设备、医药保健、汽车汽配、信息技术、食品饮料,其次是有色金属。

内容目录

背景介绍	4
Logistic 模型原理	4
基于 Logistic 模型的选股策略	5
A 股市场 Logistic 模型实证	6
样本资料描述	6
数据采集	7
极值与标准化	8
报表数据的使用	8
等权复合因子的构造	8
数据处理特别说明	10
实证结果	11
最新建议	12
可能存在的问题	13

图表目录

表 1: 所采用的描述性变量及其分类	6
图 1: Logistic 选股结果走势图	11
图 2: Logistic 选股结果与基准比较图	11
表 2: Logistic 选股模型实证结果比较(2009.11~2010.8).....	12
表 3: Logistic 模型选股最新建议(2010.9.1~2010.9.30).....	12
图 3: 最新选股结果的行业分类	13

背景介绍

从多因素模型产生的思想来看，其目的主要是想解决哪些风险因素以及有多少因子对股票或债券报酬产生影响，以及影响有多大。多因素模型通常有 3 种模型，BARRA 模型、Northfield 模型和对无法识别的风险进行管理的统计模型。其中 BARRA 模型对证券基本面的风险分析做的比较出色，Northfield 模型对宏观经济因素的风险分析十分出色，这两种模型的原理基本相同。BARRA 模型是利用多因素模型的基本原理，识别出风险因子，然后利用这个模型对风险因子导致的风险进行识别并加以控制。

一般主要考虑部分宏观因子，基本面因子和部分技术因子，从中选取可能对投资组合收益率和风险有影响的因子进行降维处理，再进行回归分析，用加权最小二乘法估计回归方程的系数，在下一期的回归方程中保持系数不变，求出因变量(即收益)。但是，预期收益率的确切估计是件比较困难的事情，同时，在投资中的很多策略也许并不要求获取较为确切的预期收益率数据，比如公募基金跑赢跟踪基准的要求以及结合股指期货的 alpha 策略，它们要求的是所选择的股票未来跑赢基准的概率有多大，这一要求比预测股票未来的预期收益率大为降低。而 logistic 模型正是解决这一问题的利器。

Logistic 模型原理

在多元回归分析中，因变量 y 可能有两种情形：

- 1、 y 是一个定量的变量，这时就用通常的多元线性模型对 y 进行回归。
- 2、 y 是一个定性的离散变量，例如， $y = 0$ 或 1 的二分量问题。这时就不能用通常的多元线性模型对 y 进行回归，而要使用所谓的 Logistic 回归。

Logistic 回归的基本思想是，不是直接对 y 进行回归，而是先定义一个概率函数 π ，将 $y = 1$ 的概率记为

$$\pi = P(y = 1 | x_1, x_2, \dots, x_n)$$

此时，如果直接对 π 进行回归，得到的回归方程可能不满足 $0 \leq \pi \leq 1$ 这个条件。

直接求 π 的表达式，是比较困难的一件事，于是考虑

$$\frac{\pi}{1 - \pi} = \frac{P(y = 1)}{P(y = 0)}$$

一般的， $0 \leq \frac{\pi}{1 - \pi} < +\infty$ 。可以令

$$\pi = P(y = 1 | x_1, x_2, \dots, x_n) = \frac{\exp\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}$$

其中 β_j ($j = 0, 1, \dots, n$) 为常数。将 π 变形得到:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{i=1}^n \beta_i x_i$$

于是,即可对 $\log\left(\frac{\pi}{1-\pi}\right)$ 进行通常的多元线性回归。假设 $y = 1$ 代表事件发生,

从而通过 Logistic 回归模型即可算出事件发生的概率 π 。

该模型中的参数通常可以通过最大似然估计,不过在 MATLAB 中已经有现成的非线性回归函数,可以通过 `glmfit` 和 `glmval` 两个函数进行 Logistic 回归。

基于 Logistic 模型的选股策略

我们采用沪深 300 指数作为基准,研究其中个股收益率超越该指数收益率的概率。观察周期为一个月,指数与股价的初末状态为上月最后一个交易日的收盘价和本月最后一个交易日的收盘价。假设在该段时间 t 内,沪深 300 指数的收益率为 R_{it} , 沪深 300 成分中某支股票 k 的收益率为 R_{stk} 。若有 $R_{stk} > R_{it}$, 即股票 k 的收益率在某个周期 t 超越了沪深 300 指数的收益率,则记因变量 $y_k = 1$, 否则记 $y_k = 0$ 。于是这便成为一个二分量问题,可以用 Logistic 回归求解个股收益率在某个周期内超越基准的概率。我们的投资策略为,等权重选取超越基准概率排名前 40 的股票进行组合,在每期(每月)的前一个月最后一个交易日以收盘价买入,在当月最后一个交易日以收盘价卖出。在不计手续费的情况下,记该组合在第 t 期超越基准的收益率 $r_t = R_{stk} - R_{it}$ 。

A 股市场 Logistic 模型实证

首先考虑 logistic 模型中的自变量，即影响股票收益率超越基准收益率概率的因素。我们将这些因素总体分为两类：公司基本面数据、个股市场数据。在这两大类因素中，我们选择了 29 个具有典型性的因素作为自变量，如表 1 所示。

考虑到自变量较多，进行数据处理较慢，我们通过观察其相关性直接将这 29 个因子又分为八组，分别是 BETA、企业规模、相对低估性、流动性、投资成果、波动性、杠杆性和成长性。通过算数平均计算每组内的标准化变量得到新分组的变量值，将这 8 个量作为新的自变量。根据比较，利用简单算术平均法计算合成因子优于其他因子筛选方法的效果¹。结合之前的二分量因变量，即可进行 Logistic 回归。

样本资料描述

由于一些“描述性变量”必须经由过去三年的资料计算而得。另外，本研究采用复权后的股票价格计算股票的收益率，其计算公式如下：

$$r_{it} = \ln \frac{P_{it}}{P_{t-1,i}}$$

其中 r_{it} 为股票 i 在 t 期的收益率， P_{it} 为股票 i 在 t 期的复权后的价格， $P_{t-1,i}$ 为股票 i 在 $t-1$ 期的复权后的价格。本研究所采用的 29 个指标^[3]及其 8 个分类如下表所示：

表 1：所采用的描述性变量及其分类

序号	分类	因素	计算方法
1	Beta	HISTORICAL BETA(每月)	CAMP 的 BETA(每月，期间:36 个月)
2		HISTORICAL BETA(每周)	CAMP 的 BETA(每月，期间:52 个月)
3	企业规模	现价总额(对数值)	总市值(总股数*股价)的对数值
4		流通现价总额(对数值)	流通总市值(流通股数*股价)对数值
5		总资产(对数值)	总资产的对数值
6	相对低估性	营业利润回报	前期营业利润/总市值
7		销售额(业务收入)回报	前期营业收入/总市值
8		B/P	前期所有者权益/总市值
9	流动性	买卖循环率(一个月)	过去一个月内的日平均交易额/流通市值

¹ 请参阅《传统多因素模型及其在沪深 300 中的实证》

10		每日交易额的变动性	过去 120 天的日交易额的标准差
11		买卖资金的变化(25 日/120 日)	过去 25 天的日平均买卖资金/过去 120 天的日平均买卖资金
12		买卖资金的变化(75 日/250 日)	过去 75 天的日平均买卖资金过去 250 天的日平均买卖资金
13	投资 成果	股价变动的平均偏离(25 日)	(25 日变动平均股价-前一天股价)/25 天变动平均股价
14		股价变动的平均偏离(75 日)	(75 日变动平均股价-前一天股价)/75 天变动平均股价
15		Historical Alpha (月)	每支股票的 return 和股价指数的回归切片(周, 期间: 36 个月)
16		Historical Alpha (周)	每支股票的 return 和股价指数的回归切片(周, 期间: 52 周)
17		Specific RETURN(1 个月)	crosssection 回归分析(Beta、size、value)的残差 return(1 个月)
18	波动性	Total risk(月)	每支股票的 RETURN 的标准差(其间:36 个月)
19		Total Risk(周)	每支股票的 RETURN 的标准差(其间:52 周)
20		Residual Risk(月)	每支股票的 RETURN 和股价指数的回归分析的残差项的标准偏差(月、期间:36 个月)
21		Residual Risk(周)	每支股票的 RETURN 和股价指数的回归分析的残差项的标准偏差(周、期间:52 周)
22	杠杆性	负债比率(账本价)	总负债/总资产
23		负债比率(时价)	总负债/总市值
24	成长性	销售额(业务收入)增长度	对与过去 12 个季度的营业收入的时间的回归分析的回归系数/过去 12 个季度的平均营业收入
25		总资产的增长度	对与过去 12 个季度的营业收入的时间的回归分析的回归系数/过去 12 个季度的平均总资产
26		销售额(业务收入)营业利润率	营业利润/营业收入
27		销售额(业务收入)营业利润 Trend	对[营业利润/业务收入]的过去 12 个季度的回归分析的回归系数
28		总资产营业利润率	营业利润/总资产
29		总资产营业利润率 Trend	对[营业利润/总资产]的过去 12 个季度的回归分析的回归系数

数据采集

行情数据:

总股本(日数据)、流通股本(日数据)、日成交额、日收盘价、周收盘价、月收盘价、上证综指(日数据、周数据、月数据)

时间周期: 2000 年 1 月至 2010 年 8 月

财务数据:

总资产、总负债、所有者权益、营业收入、营业利润, 以上皆为季度数据

时间周期: 1999 年 12 月至 2010 年 8 月

沪深 300 成分

考虑沪深 300 样本调整, 取每个月沪深 300 组成。

极值与标准化

1) 变量极值的处理。极值很容易严重扭曲分析结果, 因此去极值的处理步骤是必要的。使用“中位数去极值法(the Skipped Huber Method)”, 以数学式描述该方法如下:

$$D_{i,upper} = D_m + 5.2D_{MAD}, \quad \text{if } D_i \geq D_m + 5.2D_{MAD}$$

$$D_{i,lower} = D_m - 5.2D_{MAD}, \quad \text{if } D_i \leq D_m - 5.2D_{MAD}$$

其中 D_i 记做描述性变量的第 i 个观察值; D_m 记做每一个描述性变量的中位数,

$D_{i,AD}$ 记作描述性变量中每一个观察值与中位数的绝对偏离, 即

$D_{i,AD} = |D_i - D_m|$, D_{MAD} 记作所有绝对偏离 $D_{i,AD}$ 的中位数, $D_{i,upper}$ 和

$D_{i,lower}$ 分别记作经中位数去极值后描述性变量的上下限。

2) 描述性变量的标准化。在将描述性变量组合成风险指数前, 必须标准化每一个描述性变量, 否则所有组合将无意义, 因为每一描述性变量所衡量的单位均为不同。通常以均值取零, 标准差取 1 的正态标准化处理之, 其公式为:

标准化后变量=(原变量-均值)/标准差

取固定时间窗口 (36 个月或 12 个季度), 用时间窗口长度的数据向后滚动作标准化。

报表数据的使用

采用年报或一季报数据 (4 月底出) 代表 5, 6, 7, 8 月数据

采用中报数据 (8 月底出) 代表 9, 10 月数据

采用三季报数据 (10 月底出) 代表 11, 12 月, 下年 1, 2, 3, 4 月数据

等权复合因子的构造²

1、BETA 因子

HISTORICAL BETA (每周): 依据 CAPM 模型, 使用 52 周收盘价收益率序列对相应上证综指收益率序列作回归, 计算得出周 BETA。并将所得出周 BETA 根据其对应时间取平均, 合并成月数据。

HISTORICAL BETA (每月): 依据 CAPM 模型, 使用 36 个月收盘价对相应上证综指收益率序列作回归, 计算得出月 BETA。

² 本部分内容请参阅《传统多因素模型及其在沪深 300 中的实证》。

合成 BETA 因子:

对 HISTORICAL BETA (每周) 月数据与 HISTORICAL BETA (每月) 数据用 36 个月数据长度分别进行标准化, 等权重加权合成 BETA 因子。

2、企业规模因子

总市值 (对数值): 用对应日收盘价*日总股本相乘得到日总市值, 并将所得日数据依据其对应时间合取平均并成月数据, 取对数值。

流通市值 (对数值): 用对应日收盘价*日流通股本相乘得到日流通市值, 并将所得日数据依据其对应时间取平均合并成月数据, 取对数值。

总资产 (对数值): 将总资产数据由季度数据扩充成月数据, 取对数值。

合并成企业规模因子:

对总市值 (对数值)、流通现价市值 (对数值) 和总资产 (对数值) 月数据用 36 个月数据长度分别进行标准化, 等权重合成企业规模因子。

3、相对低估性因子

营业利润回报: 将营业利润扩充成月数据, 前期营业利润/总市值。

营业收入回报: 将营业收入扩充成月数据, 前期营业收入/总市值。

B/P: 将所有者权益扩充成月数据, 前期所有者权益/总市值。

合并成相对低估性因子:

对营业利润回报、营业收入回报和 B/P 月数据用 36 个月数据长度分别进行标准化, 等权重合成相对低估性因子。

4、流动性因子

买卖循环率 (一个月): 日交易额/日流通市值, 根据对应时间取平均并成月数据。

每日交易额的变动性: 过去 120 天日交易额的标准偏差, 将日数据根据相应时间取平均合并成月数据。

买卖资金的变化 (25 日/120 日): 过去 25 天的日平均交易额/过去 120 天的日平均交易额, 将日数据根据相应时间取平均合并成月数据。

买卖资金的变化 (75 日/250 日): 过去 75 天的日平均交易额/过去 250 天的日平均交易额, 将日数据根据相应时间取平均合并成月数据。

合并成流动性因子:

对买卖循环率 (一个月)、每日交易额的变动性、买卖资金的变化 (25 日/120 日)、买卖资金的变化 (75 日/250 日) 月数据用 36 个月数据长度分别进行标准化, 等权重合成流动性因子。

5、投资成果因子

股价变动的平均偏离 (25 日): (25 日变动平均股价-前一天股价)/25 日变动平均股价, 将日数据根据相应时间取平均合并成月数据。

股价变动的平均偏离 (75 日): (75 日变动平均股价-前一天股价)/75 日变动平均股价, 将日数据根据相应时间取平均合并成月数据。

Historical Alpha (周): 用每支股票月收益率序列与上证综指收益率序列回归分析切片 (周, 期间: 52 周), 将周数根据相应时间取平均合并成月数据。

Historical Alpha (月): 用每支股票月收益率序列与上证综指收益率序列回归分析切片 (月, 期间: 36 个月)。

合并成投资成果因子:

对股价变动的平均偏离（25日）、股价变动的平均偏离（75日）、Historical Alpha（周）和 Historical Alpha（月）月数据用36个月数据长度分别进行标准化，等权重合成投资成果因子。

6、波动性因子

Total Risk（周）：每支股票收益率的标准差（期间：52周），将周数据根据相应时间取平均合并成月数据。

Total Risk（月）：每支股票收益率的标准差（期间：36个月）。

Residual Risk（周）：每支股票收益率与上证综指收益率序列作回归分析的残差项的标准差（期间：52周），将周数据根据相应时间取平均合并成月数据。

Residual Risk（月）：每支股票收益率与上证综指收益率序列作回归分析的残差项的标准差（期间：36个月）。

合并成波动性因子：

对 Total Risk（周）、Total Risk（月）、Residual Risk（周）和 Residual Risk（月）月数据用36个月数据长度分别进行标准化，等权重合成波动性因子。

7、杠杆性因子

负债比率（账面价值）：将总负债扩充成月数据，总负债/所有者权益。

负债比率（盯市价值）：将总负债扩充成月数据，总负债/总市值。

合成杠杆性因子：

对负债比率（账面价值）和负债比率（盯市价值）月数据用36个月数据长度分别进行标准化，等权重合成杠杆性因子。

8、成长性因子

营业收入增长速度：对与过去12个季度的营业收入的时间的回归分析的系数/过去12个季度的平均营业收入。

总资产增长速度：对与过去12个季度的总资产的时间的回归分析的系数/过去12个季度的平均总资产。

营业收入营业利润率：营业收入/营业利润。

营业收入营业利润率 TREND：对[营业收入/营业利润]的过去12个季度的回归分析的回归系数。

总资产营业利润率：营业利润/总资产。

总资产营业利润率 TREND：对[营业利润/总资产]的过去12个季度的回归分析的回归系数。

合并成成长性因子：

对营业收入增长速度、总资产增长速度、营业收入营业利润率、营业收入营业利润率 TREND、总资产营业利润率和总资产营业利润率 TREND 季度数据用12个季度数据长度分别进行标准化，根据对应时间将季度数据扩充成月数据，等权重合成成长性因子。

数据处理特别说明

以月为周期进行投资，我们对沪深300的所有股票选取了从2005年11月开始至2010年8月的58期的相关数据，包含58组上述8个自变量的数据以及57组因变量数据（因为是要通过第*i*期的自变量预测当期该股票是否能够超越

基准，而这一因变量数据只有在第 $i+1$ 期才能通过比较上期的收益率得到，即第 i 期的自变量对应的因变量是第 $i+1$ 期的因变量，所以因变量 y 只从第二期开始取)。

回归过程中，一些数据的处理方式如下：

1、由于沪深 300 指数的成分股在不断变化调整，因此利用第 i 期自变量预测当期超越基准的概率时，只取第 i 期存在于沪深 300 成分股中的股票，或者说每期选取的股票样本有可能不同。

2、由于存在部分数据缺失的状况，而太少的样本数据可能会造成回归结果不准确。因此，对于某个自变量，数据量的数目少于 24 期者，该支股票不计入计算范围。

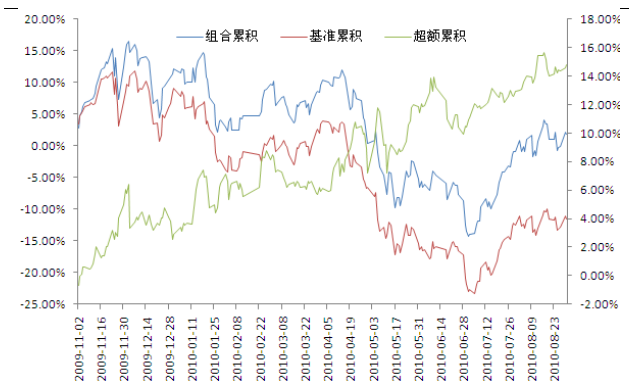
3、同样由于数据缺失，在利用第 i 期自变量预测当期超越基准的概率时，如果第 i 期的某个自变量存在缺失情况，则该支股票不计入计算范围。

4、处理第 i 期预测的回归数据为第 i 期之前各期的数据，因此越靠后的期数所能用到的数据量越多，回归效果越好，预测的概率越准确。

实证结果

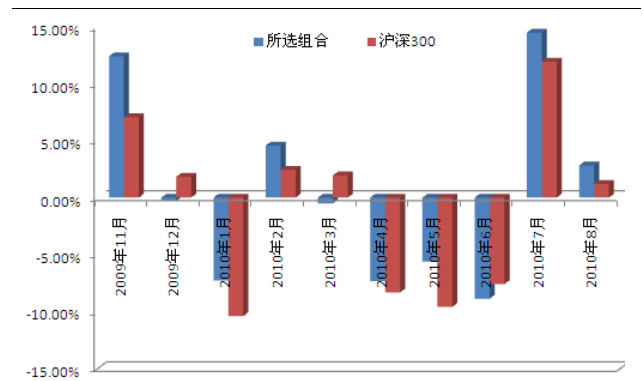
由于所要求的数据量比传统多因素模型要多很多，Logistic 选股模型的实证时间从 2009 年 11 月开始，具体实证结果如图 1 和图 2 所示：

图 1: Logistic 选股结果走势图



资料来源：国信证券经济研究所

图 2: Logistic 选股结果与基准比较图



资料来源：国信证券经济研究所

表 2 详细比较了 Logistic 选股组合与基准的风险收益特征，同时对超额收益率也进行了相应的分析。从表 2 可以看出，Logistic 模型所选股票在样本内有效的超越了基准，这对于公募基金、以及股指期货环境下追求绝对收益的私募

基金有很大的帮助。

表 2: Logistic 选股模型实证结果比较(2009.11~2010.8)

时间	所选组合	沪深 300	超额收益率	组合累积	基准累积	超额累积
2009 年 11 月	12.40%	7.05%	5.35%	12.40%	7.05%	5.35%
2009 年 12 月	-0.31%	1.82%	-2.13%	12.05%	9.00%	3.11%
2010 年 1 月	-7.27%	-10.39%	3.12%	3.91%	-2.33%	6.32%
2010 年 2 月	4.57%	2.42%	2.15%	8.65%	0.04%	8.61%
2010 年 3 月	-0.51%	1.95%	-2.46%	8.10%	1.99%	5.94%
2010 年 4 月	-7.33%	-8.32%	0.99%	0.18%	-6.50%	6.99%
2010 年 5 月	-5.62%	-9.59%	3.97%	-5.45%	-15.46%	11.23%
2010 年 6 月	-8.88%	-7.58%	-1.30%	-13.85%	-21.87%	9.79%
2010 年 7 月	14.48%	11.93%	2.55%	-1.37%	-12.55%	12.59%
2010 年 8 月	2.82%	1.20%	1.62%	1.41%	-11.50%	14.41%
平均值	0.44%	-0.95%	1.39%	1.69%*	-13.64%*	17.53%*
标准差	28.47%	26.43%	9.09%			
Sharpe	0.059	-0.516	1.930			

资料来源: 国信证券经济研究所, *代表年化收益率

由此可见, Logistic 选股模型是一种能够带来超越基准收益的模型, 样本内年化收益率 1.69%, 沪深 300 指数年化收益率为-13.64%, 超额收益率年化 17.53%。Logistic 选股模型可用于追求相对排名的公募基金以及追求绝对收益的私募基金(结合股指期货)。

最新建议

我们根据 2010 年 8 月的市场数据和公司基本面数据预测 2010 年 9 月沪深 300 成分股超越基准的概率, 并选取概率排名前 40 的股票进行等权重配置。所选取股票列表如表 3 所示:

表 3: Logistic 模型选股最新建议(2010.9.1~2010.9.30)

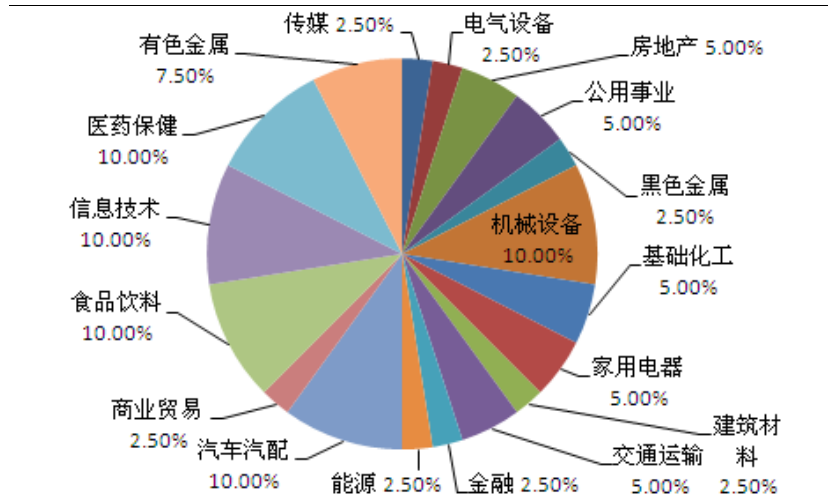
代码	名称	所属行业	代码	名称	所属行业
000039.SZ	中集集团	机械设备	600376.SH	首开股份	房地产
600066.SH	宇通客车	汽车汽配	600516.SH	方大炭素	有色金属
600456.SH	宝钛股份	有色金属	600518.SH	康美药业	医药保健
600597.SH	光明乳业	食品饮料	600779.SH	水井坊	食品饮料
600118.SH	中国卫星	信息技术	600316.SH	洪都航空	机械设备
600216.SH	浙江医药	医药保健	600839.SH	四川长虹	家用电器
600643.SH	爱建股份	金融	000538.SZ	云南白药	医药保健
000900.SZ	现代投资	交通运输	600085.SH	同仁堂	医药保健
600006.SH	东风汽车	汽车汽配	600741.SH	华域汽车	交通运输
600236.SH	桂冠电力	公用事业	000858.SZ	五粮液	食品饮料
000061.SZ	农产品	房地产	600500.SH	中化国际	商业贸易
000528.SZ	柳工	机械设备	600688.SH	S 上石化	能源

600596.SH	新安股份	基础化工	000568.SZ	泸州老窖	食品饮料
600832.SH	东方明珠	传媒	600718.SH	东软集团	信息技术
600660.SH	福耀玻璃	汽车汽配	600089.SH	特变电工	电气设备
600879.SH	航天电子	信息技术	000060.SZ	中金岭南	有色金属
000959.SZ	首钢股份	黑色金属	000951.SZ	中国重汽	汽车汽配
600863.SH	内蒙华电	公用事业	600820.SH	隧道股份	建筑材料
600183.SH	生益科技	信息技术	000059.SZ	辽通化工	基础化工
000768.SZ	西飞国际	机械设备	000527.SZ	美的电器	家用电器

资料来源：国信证券经济研究所，Wind 资讯，国信一级行业分类

四十只股票在行业中的分类如图 3 所示（采用国信一级行业分类），其中所占比重最大的行业是机械设备、医药保健、汽车汽配、信息技术、食品饮料，其次是有色金属。

图 3：最新选股结果的行业分类



资料来源：国信证券经济研究所，Wind 资讯

可能存在的问题

和传统的多因素模型一样，因为涉及到众多的基本面数据，使用 Logistic 模型选择股票，可能存在以下问题：

- 1、众多的实证表明，在财报数据公布之前，信息存在一定程度的泄漏，同时，各个股票财报公布的时间为一个区间，而非某一个固定时点，这对从基本面角度出发的选股模型是一个很大的不利因素；
- 2、对于行业、主题概念等属性很难加入到模型中，但国内市场很多时候的行情确实是一些行业性或者概念类的行情，后续研究中将行业、概念等属性纳入考察的范畴是必不可少的；
- 3、影响股票的因子众多，文中所列的 29 个因子也许不能反映实际情况的全貌，

而所考察的因子是否全面对选股结果至关重要，后续研究中将进一步考察需要纳入的因子范围；

- 4、Logistic 模型只是给出了因子与战胜基准概率之间的定量关系，至于他们之间的逻辑、因子与因子之间的关联则研究的很少，没有合理的逻辑，那么就无法保证未来选股结果的可靠性和稳定性；
- 5、国内股市的时间过短，数据的可靠性、完整性得不到保证，同时，选股绩效是否稳定、可靠还需要大样本验证。

国信证券投资评级

类别	级别	定义
股票 投资评级	推荐	预计 6 个月内，股价表现优于市场指数 20%以上
	谨慎推荐	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数 $\pm 10\%$ 之间
	回避	预计 6 个月内，股价表现弱于市场指数 10%以上
行业 投资评级	推荐	预计 6 个月内，行业指数表现优于市场指数 10%以上
	谨慎推荐	预计 6 个月内，行业指数表现优于市场指数 5%-10%之间
	中性	预计 6 个月内，行业指数表现介于市场指数 $\pm 5\%$ 之间
	回避	预计 6 个月内，行业指数表现弱于市场指数 5%以上

免责声明

本报告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。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或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我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。本报告版权归国信证券所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登。

国信证券经济研究所团队成员

宏观		策略		交通运输	
周炳林	0755-82130638	黄学军	021-60933142	郑武	0755-82130422
林松立	010-66026312	崔嵘	021-60933159	陈建生	0755-82133766
				岳鑫	0755-82130432
				高健	0755-82130678
银行		房地产		机械	
邱志承	021-60875167	方焱	0755-82130648	余爱斌	0755-82133400
黄飙	0755-82133476	区瑞明	0755-82130678	黄海培	021-60933150
谈煜	010-66025229	黄道立	0755-82133397	陈玲	0755-82130646
				杨森	0755-82133343
				李筱筠	010-66026326
汽车及零配件		钢铁		商业贸易	
李君	021-60933156	郑东	010-66026308	孙菲菲	0755-82130722
左涛	021-60933164	秦波	010-66026317	吴美玉	010-66026319
				祝彬	0755-82131528
基础化工		医药		石油与石化	
张栋梁	0755-82130532	贺平鹤	0755-82133396	李晨	021-60875160
陈爱华	0755-82133397	丁丹	0755-82139908	严蓓娜	021-60933165
邱斌	0755-82130532	陈栋	021-60933147		
电力设备与新能源		传媒		有色金属	
皮家银	021-60933160	陈财茂	021-60933163	彭波	0755-82133909
				谢鸿鹤	0755-82130646
电力与公用事业		非银行金融		通信	
徐颖真	021-60875162	邵子钦	0755-82130468	严平	021-60875165
谢达成	021-60933161	田良	0755-82130513	程峰	021-60933167
		童成敦	0755-82130513		
造纸		家电		计算机	
李世新	0755-82130565	王念春	0755-82130407	段迎晟	0755-82130761
邵达	0755-82130706				
电子元器件		纺织服装		农业	
段迎晟	0755-82130761	方军平	021-60933158	张如	021-60933151
高耀华	0755-82130771				
旅游		食品饮料		建材	
廖绪发	021-60875168	黄茂	0755-82138922	杨昕	021-60933168
刘智景	021-60933148				
煤炭		建筑		固定收益	
李然	010-66026322	邱波	0755-82133390	李怀定	021-60933152
陈健	010-66215566	李遵庆	0755-82133055	高宇	0755-82133538
苏绍许	021-60933144			侯慧娣	021-60875161
				张旭	010-66026340
				蔺晓熠	021-60933146
				刘子宁	021-60933145
指数与产品设计		投资基金		量化投资	
焦健	0755-82133928	杨涛	0755-82133339	葛新元	0755-82133332
王军清	0755-82133297	刘舒宇	0755-82133568	董艺婷	021-60933155
彭甘霖	0755-82133259	康亢	010-66026337	林晓明	0755-25472656
阳瑾	0755-82133538	刘洋		赵斯尘	021-60875174
周琦	0755-82133568			程景佳	021-60933166
赵学昂	0755-66025232			郑云	021-60875163
				毛甜	021-60933154
交易策略					
戴军	0755-82133129				
秦国文	0755-82133528				
徐左乾	0755-82133090				
黄志文	0755-82133928				

国信证券机构销售团队

华北区 (机构销售一部)	华东区 (机构销售二部)	华南区 (机构销售三部)
王立法 010-66026352 13910524551 wanglf@guosen.com.cn	盛建平 021-60875169 15821778133 shengjp@guosen.com.cn	万成水 0755-82133147 13923406013 wancs@guosen.com.cn
王晓建 010-66026342 13701099132 wangxj@guosen.com.cn	马小丹 021-60875172 13801832154 maxd@guosen.com.cn	魏宁 0755-82133492 13823515980 weining@guosen.com.cn
焦戡 010-66026343 13601094018 jiaojian@guosen.com.cn	郑毅 021-60875171 13795229060 zhengyi@guosen.com.cn	邵燕芳 0755-82133148 13480668226 shaoyf@guosen.com.cn
李锐 010-66025249 13691229417 lirui2@guosen.com.cn	黄胜蓝 021-60875166 13761873797 huangsl@guosen.com.cn	林莉 0755-82133197 13824397011 linli2@guosen.com.cn
徐文琪 010-66026341 13811271758 xuwxq@guosen.com.cn	刘塑 021-60875177 13817906789 liusu@guosen.com.cn	王昊文 0755-82130818 18925287888 wanghaow@guosen.com.cn
	叶琳菲 021-60875178 13817758288 yelf@guosen.com.cn	甘墨 0755-82133456 15013851021 ganmo@guosen.com.cn
	孔华强 021-60875170 13681669123 konghq@guosen.com.cn	段莉娟 0755-82130509 18675575010 duanlj@guosen.com.cn
		黎敏 0755-82130681 13902482885 limin1@guosen.com.cn
		徐冉 13632580795 xuran1@guosen.com.cn
		颜小燕 13590436977 yanxy@guosen.com.cn