

第五届泰迪杯 数据挖掘挑战赛

优秀
作品

作品名称：基于市场资金流向分析的商品期货量化交易策略

荣获奖项：特等奖

作品单位：暨南大学

作品成员：卢国豪 丘延君 叶诚

指导老师：侯雅文

基于市场资金流向分析的商品期货量化交易策略

摘要

本文旨在开发基于期货市场资金流向分析的商品期货量化交易策略，通过分析商品期货历史数据，结合期货市场相关背景，来刻画了期货市场资金流向规律，并构建策略模型。

针对问题一，本文构建了**动态权重资金流向模型**。在股票市场经典资金流向计算公式的基础上和商品期货市场交易机制背景下，我们运用分钟级别 K 线数据，引进了持仓量、基于价格和持仓量变化的动态权重，建立了动态权重资金流向模型来计算资金流向，有效刻画了商品期货市场资金流向的规律。

针对问题二，首先，本文计算了 38 个期货合约在不同分钟级别下的资金流向与未来商品期货价格的**相关系数**并建立了**线性回归模型**来研究 2011 年至 2013 年间资金流向对未来商品期货价格的影响，得到了存在强正相关关系的 6 个期货合约组合和强负相关的 16 个期货合约组合；其次，本文建立了 **Logistic 回归模型**，分析了短期内不同分钟级别下资金流向与未来商品期货价格涨跌的关系，基本上所有都得到了超过 50% 的预测准确率。最后，本文对资金流向流量数据进行 **ADF 检验**并计算其**自相关系数**，发现各个板块的资金流向在 10 个滞后期内都具有正相关持续性的。

针对问题三，本文基于动态权重资金流向模型统计了各个板块的总资金流向，发现各个板块的资金流向在 2011 年至 2013 年间总体上均是上升的。特别地，发现在 2012 年年末前后，推动整个商品期货市场资金流的的增长的主力板块由化工、农产品、油类变成了化工、金属、黑色系，说明板块间具有**轮动效应**。

针对问题四，本文建立了**基于资金流向的二分类 Logistics 回归策略模型**，从资金流向与期货商品未来价格具有较大相关系数的期货合约组合中挑选部分合约作为回测标的，在 Auto-Trader 平台进行量化交易，回测报告显示：各个期货合约的初始收益率在 5% 到 35% 之间浮动，总体收益率为 15.71%。

关键词：资金流向、相关关系、板块轮动效应、二分类 Logistics 回归策略模型

Quantitative Trading Strategy of Commodity Futures Based on The Analysis of Money Flow

Abstract: This paper aims to develop quantitative trading strategies of commodity futures based on the analysis of money flow. By analyzing the historical data of commodity futures and the context of futures market, this paper describes the law of money flow in futures market and establishes a strategy model as well.

To solve the first problem, the money flow model with the dynamic weight is introduced in this paper. During the course of forming the futures-market money flow model, this paper studies the classical formula of money flow in the context of domestic stock market. The money flow model developed in this paper is a variation of the mentioned stock-market money flow model. The difference between the two models lies in open interest, the dynamic weights of the change in the futures contract prices and the change in open interest. The established model is proved to be effective in characterizing the law of money flow in commodity futures market.

To solve the second problem, this paper develops two models. The first one is a liner regression model which also calculates the correlation coefficient between the money flow and the future commodity futures contract prices. Such calculation is repeated among 38 different futures contracts, with each in different levels of minutes. The data used cover the range from the year of 2011 to 2013. The results show that 6 of the 38 futures contracts display a strong positive correlation and 16 of them display a strong negative correlation. The second one is a logistic regression model, which analyzes the relationship between the money flow and the movement of the commodity futures contract prices in near future in different levels of minutes. The forecasting accuracy of this model is greater than 50%. At the end of this part, the autocorrelation coefficient of the money flow is tested in this paper. According to the ADF test, the money flow of each block possesses positive correlation persistence in 10 lag periods.

To solve the third problem, this paper calculates the total money flow of each block using the money flow model established in the first problem. According to the results, the total money flow of the each block has been increasing from the year of 2011 to

2013. In particular, in the main plate that promoted the growth of money flow in the commodity futures market was changed in the end of 2012 from chemical, agricultural and oil to chemical, metal and black series, which proved the entity of sector rotation.

To solve the fourth problem, a two-classification Logistics regression strategy model based on money flow is established in this paper. Some contracts are selected as part of the futures contract portfolio, which has a large correlation coefficient between the future price of futures and money. We trade in the Auto-Trader platform using this strategy. The retest report shows that the rate of return all of the futures contract fluctuates between 5% to 35%, and the overall return is 15.71%.

Key words: Money Flow, Correlation Coefficient, Sector Rotation, Two-classification Logistic Regression Strategy Model

目 录

1 问题重述	5
1.1 问题背景	5
1.2 要解决的问题	5
2 问题分析	5
2.1 问题一的分析	5
2.2 问题二的分析	6
2.3 问题三的分析	6
2.4 问题四的分析	6
3 符号说明	7
4 模型的建立与求解	7
4.1 问题一：构建基于期货市场的资金流向模型	7
4.1.1 模型思想	7
4.1.2 模型建立	8
4.1.3 模型的解释	9
4.1.4 模型应用	10
4.2 问题二：资金流向对未来商品期货价格的影响及其持续性分析	12
4.2.1 资金流向对未来商品期货价格的影响	13
4.2.2 基于 Logistic 模型的资金流向与未来期货价格涨跌关系实证分析	17
4.2.3 资金流向的持续性分析	21
4.3 问题三：资金流向特征分析	23
4.3.1 板块间资金流向的相似特点分析	23
4.3.2 各个板块的资金流向统计分析	25
4.4 问题四：基于资金流向的二分类 Logistics 回归策略设计	32
4.4.1 策略设计	32
4.4.2 回测结果	33
5 模型的评价与推广	33
5.1 模型的优点	33
5.2 模型的局限性	34
参考文献	34

1 问题重述

1.1 问题背景

“水涨船高”是个很简单的道理，在股票市场也是如此。一般来说，大量资金不断往某只股票流入，这只股票的价格就会上涨；当某只股票的资金大量流出，股价可能下跌，甚至一泻千里。股票市场中常说“量于价先行”，这个“量”包括资金的流向和资金的流量，统称为资金流向(money flow)。资金流向在国际上是一个成熟的技术指标，反映市场当前对某只股票超额的需求或者供给，它能够帮助投资者透过指数（价格）涨跌的迷雾看到其他投资者的投资行为。

资金流向的计算方法多种多样，没有统一的公式。例如可以根据成交价格的涨跌来界定资金的流入或流出，或者根据主动买卖双方的力量强弱来界定资金的流入或流出。

在期货市场中因为有做空机制，所以直接采用股票市场中的资金流向公式是无法刻画出期货标的合约真实的资金流向规律。比如某一期货标的合约价格下跌，可能是资金流入导致的，也可能是资金流出导致的。所以需要考虑持仓量与成交量以及价格的涨跌等综合因素，来刻画期货产品的资金流向。

1.2 要解决的问题

1) 使用 2011 年 1 月 1 日至 2013 年 12 月 31 日的商品期货历史数据，分析资金流向的规律，设计一个商品期货量化交易策略。

2) 使用策略研究系统 Auto-Trader 回测引擎对 2014 年 1 月 1 日至 2015 年 12 月 31 日的数据进行策略回测。

2 问题分析

基于要解决的问题，本文拆分为以下四个子问题进行求解，从如下角度进行分析，并以此构建策略模型：

2.1 问题一的分析

充分运用期货市场分钟级 K 线数据，构造资金流向模型。

问题一旨在根据期货市场的交易分钟级 K 线数据结合期货市场特点来改进经典的资金流向技术指标公式，充分持仓量、成交量以及价格涨跌等综合因素来刻画期货产品的资金流向，使之更加符合期货市场交易机制。

2.2 问题二的分析

资金流向对未来商品期货的价格有什么影响？资金流向是否具有持续性？

问题二旨在分析资金流向对未来商品期货价格的影响，并对资金流向的持续性展开分析。

首先，本文将利用 2011 年到 2013 年的全部样本数据，基于问题构建的期货市场资金流向公式计算得资金流向，计算了资金流向与未来商品期货价格的相关系数并进一步建立线性回归模型来研究资金流向对未来商品期货价格的总体影响；

其次，为了进一步量化资金流向对未来期货价格的影响，建立 Logistic 回归模型，分析近期内期货市场资金流向与未来商品期货价格涨跌的关系；

最后，关于资金流向的持续性问题，本文将建立资金流向自相关模型，我们利用时间序列方法对板块资金流向的流量数据进行分析来研究板块资金流向自相关系数，探讨进而分析资金流向得持续性。

2.3 问题三的分析

分析资金流向是否存在相似的特点，是否集中在某些行业板块，是否有轮动效应？统计各个板块的资金流向，对比整个市场的资金流向、某一板块的资金流向与单个标的的资金流向之间的关系。

问题三旨在对各板块的资金流向进行统计分析对比，我们按常见板块将期货划分为黑色系、化工、金属、农产品以及油类 5 个板块，根据期货合约数据来计算板块的资金流向，并进行对比分析。

2.4 问题四的分析

商品期货量化交易策略设计。

本文在程序中设置了平仓时间，规定在上午 9:30 之前下午 14:50 之后进行平仓，不进行交易。为了控制风险，本文对交易设置严格的止损止盈机制，在模型中添加 `openprice`, `histExtre` 变量记录开盘价和交易期间的极值，用来跟踪止赢和止损。在模型中设置止损率 `stoploss` 为 0.05，止盈率为 0.1，追踪止盈率为 0.03；本文选取 20140101 到 20151231 时期的期货分钟数据用于回测，采样频率为 30 分钟，交易模型的训练集样本数为 2000。从资金流向与期货商品未来价格具有较大相关系数的合约中挑选出一部分合约作为回测标的，利用基于资

金流向的二分类 logistics 回归模型进行策略设计。

3 符号说明

$MoneyFlow$	资金流向	$Volume_i$	第 i 时刻的成交量
P_i	第 i 时刻的成交价	I_i	第 i 时刻的持仓量
α_i	第 i 时刻价格变动的权重	β_i	第 i 时刻持仓量变动的权重

4 模型的建立与求解

4.1 问题一：构建基于期货市场的资金流向模型

4.1.1 模型思想

首先,根据资金流向的定义以及资金流向指标在股票市场及股指期货市场的应用情况,我们在建立商品期货市场的资金流向公式时,并不完全推翻价格的涨跌对于资金流向的影响。

侯丽薇等在 A 股市场资金流向指标应用分析中将股票市场资金流向设为:

$$MoneyFlow = \sum_i^n (Volume_i) \times P_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|}$$

陈鸿在股指期货市场有效性与资金流向分析中将股指期货资金流向设为:

$$MoneyFlow = \sum_i^n \left(\frac{P_i \cdot P_{i-1}}{|P_i - P_{i-1}|} \right) \times (Volume_i) \times \left(\frac{P_i - P_{i-1}}{P_{i-1}} \right)$$

E Olaison 在量化交易研究中也把资金流向的公式设为:

$$MoneyFlow = \sum_i^n \left(\frac{P_i \cdot P_{i-1}}{|P_i - P_{i-1}|} \right) \times (Volume_i) \times \left(\frac{P_i - P_{i-1}}{P_{i-1}} \right)$$

结合已有的研究及商品期货市场资金流向的变动规律,我们认为价格变动对于商品期货市场资金流向的影响是不容忽视的。可能存在这样的相关关系,价格的上涨使得商品期货的资金有流入趋势,价格下跌使得商品期货的资金有流出趋势。

其次,在期货市场中因为有做空机制,所以直接采用股票市场中的资金流向公式是无法刻画出期货标的合约真实的资金流向规律。比如某一期货标的合约价

格下跌，可能是资金流入导致的，也可能是资金流出导致的。所以除了考虑成交量以及价格的涨跌等因素的作用，还要考虑持仓量及其变动带来的影响，才合理刻画出期货产品的资金流向。

在《期货市场教程》中，强调：持仓量是从期货合约开始起计算的未平仓的合约数量。持仓量越大，说明该合约到期前平仓交易量和实物交割量就越大。因此，分析持仓量的变化可以推测资金在期货市场的流向。持仓量增加，表明资金流入期货市场；持仓量减少，则说明资金流出市场。

在《期货及衍生品基础》中，强调：由于持仓量是从期货合约开始交易起计算的未平仓合约数量，持仓量增加，表明资金流入期货市场，反之，则说明资金正流出期货市场。分析持仓量的变化可能推测资金在期货市场的流向和主力资金的交易行为，有助于投资者判断下一步价格波动的方向。

需要强调的是，本文认可了价格变动对商品期货市场资金流向的变化趋势有影响，但并非利用价格的变动直接完全界定商品期货市场资金流入或流出，而是进一步引入持仓量及其变动带来的影响，这是与上述文献中建立的资金流向公式不同的地方。

4.1.2 模型建立

根据股票市场中资金流向指标的定义，结合期货标的合约真实的资金流向规律，我们在原有的资金的流向公式继续引入持仓量因素，使用下式可表示商品期货产品的资金流向：

$$MoneyFlow = \sum_i^n (Volume_i) \times P_i \times \operatorname{sgn} \left(\alpha_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|} + \beta_i \times \frac{I_i - I_{i-1}}{|I_i - I_{i-1}|} \right)$$

$$\alpha_i = \frac{\frac{|P_i - P_{i-1}|}{|P_i - P_{i-1}| + |I_i - I_{i-1}|} + \varepsilon}{\frac{|P_i - P_{i-1}|}{|P_i - P_{i-1}| + |I_i - I_{i-1}|} + 2\varepsilon}$$

$$\beta_i = \frac{\frac{|I_i - I_{i-1}|}{|P_i - P_{i-1}| + |I_i - I_{i-1}|} + \varepsilon}{\frac{|P_i - P_{i-1}|}{|P_i - P_{i-1}| + |I_i - I_{i-1}|} + 2\varepsilon}$$

其中， P_i 是 i 时刻的成交价格， P_{i-1} 是 i 时刻之前 ($i-1$) 时刻的成交价格， I_i 是 i 时刻的持仓量， I_{i-1} 是 i 时刻之前 ($i-1$) 时刻的持仓量， $Volume_i$ 是 i 时刻的成交量， $\frac{|P_i - P_{i-1}|}{|P_i - P_{i-1}| + |I_i - I_{i-1}|}$ 、 $\frac{|I_i - I_{i-1}|}{|P_i - P_{i-1}| + |I_i - I_{i-1}|}$ 分别是对 $|P_i - P_{i-1}|$ 、 $|I_i - I_{i-1}|$ 进行 0-1 标准化处理的标准化变量，

α_i 是 i 时刻成交价格变化的权重值, β_i 是 i 时刻持仓量变化的权重值, ε 是大于零的无穷小量, $\text{sgn}(x)$ 是符号函数:

$$\text{sgn}(x) = \begin{cases} -1: x < 0 \\ 0: x = 0 \\ 1: x > 0 \end{cases}$$

由于考虑到资金流向的计算公式中有 $|P_i - P_{i-1}|$ 和 $|I_i - I_{i-1}|$ 以及实际计算数据时有可能出现分母 $|P_i - P_{i-1}|$ 及 $|I_i - I_{i-1}|$ 为零的事实, 我们对资金流向的公式规定了以下 4 种情形:

1) 当 $|P_j - P_{j-1}| \neq 0$ 且 $|I_j - I_{j-1}| \neq 0$, 对 $j=1, \dots, n$ 恒成立时:

$$\text{MoneyFlow} = \sum_i^n (\text{Volume}_i) \times P_i \times \text{sgn} \left(\alpha_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|} + \beta_i \times \frac{I_i - I_{i-1}}{|I_i - I_{i-1}|} \right)$$

2) 当 $|P_j - P_{j-1}| = 0$ 且 $|I_j - I_{j-1}| = 0$, $1 \leq j \leq n$ 时:

$$\text{MoneyFlow} = \sum_{i \neq j}^n (\text{Volume}_i) \times P_i \times \text{sgn} \left(\alpha_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|} + \beta_i \times \frac{I_i - I_{i-1}}{|I_i - I_{i-1}|} \right) + \sum_{i=j} (\text{Volume}_i) \times P_i \times \text{sgn}(0)$$

3) 当 $|P_j - P_{j-1}| = 0$ 且 $|I_j - I_{j-1}| \neq 0$, $1 \leq j \leq n$ 时:

$$\text{MoneyFlow} = \sum_{i \neq j}^n (\text{Volume}_i) \times P_i \times \text{sgn} \left(\alpha_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|} + \beta_i \times \frac{I_i - I_{i-1}}{|I_i - I_{i-1}|} \right) + \sum_{i=j} (\text{Volume}_i) \times P_i \times \text{sgn} \left(0 + \beta_i \times \frac{I_i - I_{i-1}}{|I_i - I_{i-1}|} \right)$$

4) 当 $|P_j - P_{j-1}| \neq 0$ 且 $|I_j - I_{j-1}| = 0$, $1 \leq j \leq n$ 时:

$$\text{MoneyFlow} = \sum_{i \neq j}^n (\text{Volume}_i) \times P_i \times \text{sgn} \left(\alpha_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|} + \beta_i \times \frac{I_i - I_{i-1}}{|I_i - I_{i-1}|} \right) + \sum_{i=j} (\text{Volume}_i) \times P_i \times \text{sgn} \left(\alpha_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|} + 0 \right)$$

4.1.3 模型的解释

在期货市场中因为有做空机制, 某期货标的合约价格下跌, 可能是资金流入导致的, 也可能是资金流出导致的, 股票市场中常用的资金流向指标公式无法刻画出现货合约真实的资金流向规律。因此我们进一步考虑了持仓量因素, 构建的资金流向模型是可以用来解释期货市场的做空机制。

1. 权重的定义

我们将权重的计算定义为以下的公式:

$$\alpha_i = \frac{|P_i - P_{i-1}|' + \varepsilon}{|P_i - P_{i-1}|' + |I_i - I_{i-1}|' + 2\varepsilon}$$

$$\beta_i = \frac{|I_i - I_{i-1}|' + \varepsilon}{|P_i - P_{i-1}|' + |I_i - I_{i-1}|' + 2\varepsilon}$$

通过赋予期货价格方向变化和持仓量方向变化权重，充分考虑两种变化率的不同对资金流向的影响，能够较好地度量了期货价格变化与持仓量变化对资金流向的影响。

2. 数据的 0-1 标准化

0-1 标准化也称离差标准化，是对原始数据的线性变换，使结果落在[0, 1]区间，转换函数如下：

$$x^* = \frac{x - \min}{\max - \min}$$

其中 max 为样本数据的最大值，min 为样本数据的最小值。注意：当有新数据加入时，需要重新获取 max 和 min。

在计算过程中，我们将利用对 $|P_i - P_{i-1}|$ 、 $|I_i - I_{i-1}|$ 进行了如下标准化处理：

$$|P_i - P_{i-1}|' = \frac{|P_i - P_{i-1}| - \min(|P_i - P_{i-1}|)}{\max(|P_i - P_{i-1}|) - \min(|P_i - P_{i-1}|)}$$

$$|I_i - I_{i-1}|' = \frac{|I_i - I_{i-1}| - \min(|I_i - I_{i-1}|)}{\max(|I_i - I_{i-1}|) - \min(|I_i - I_{i-1}|)}$$

经过标准化后，消除单位差别和数量级差别，利用标准化后的数据 $|P_i - P_{i-1}|'$ 、 $|I_i - I_{i-1}|'$ 来计算权重，经过标准化的数据可以更加有效地反映价格和持仓量各自的变化速度，权重满足 $\alpha_i + \beta_i = 1$ 。

4.1.4 模型应用

根据本文的资金流向的计算公式，界定资金流入或流出的关键部分是函数 $\text{sgn}\left(\alpha_i \times \frac{P_i - P_{i-1}}{|P_i - P_{i-1}|} + \beta_i \times \frac{I_i - I_{i-1}}{|I_i - I_{i-1}|}\right)$ 的值，由于 $\frac{P_i - P_{i-1}}{|P_i - P_{i-1}|}$ 、 $\frac{I_i - I_{i-1}}{|I_i - I_{i-1}|}$ 的值为 ± 1 以及 $\alpha_i + \beta_i = 1$ 、 $\alpha_i > 0$ 、 $\beta_i > 0$ ， α_i 为价格变动的权重， β_i 为持仓量变动的权重；我

们可以得到以下表格。（注：表格中的+、-号表示 $\frac{P_i - P_{i-1}}{|P_i - P_{i-1}|}$ 、 $\frac{I_i - I_{i-1}}{|I_i - I_{i-1}|}$ 的取值为 1

或-1)

1) 当 $\alpha_i = \beta_i$ 时:

表 1 : 当 $\alpha_i = \beta_i$ 时的符号取值情况

α_i	β_i	$\alpha_i + \beta_i$	sgn
+	+	1	1
+	-	0	0
-	-	-1	-1
-	+	0	0

由表 1 可以看出, 当 $\alpha_i = \beta_i$ 时, 价格变动对资金流方向的影响与持仓量变动对资金流方向的影响是相当的。当价格上涨时, 资金可能是流入的, 也可能是不变的; 当价格下跌时, 资金可能是流出的, 也可能是不变的; 同理, 当持仓量增加时, 资金可能是流入的, 也可能是不变的; 当持仓量减少时, 资金可能是流出的, 也可能是不变的。此时, 没有哪一方势力可以单独地决定资金流向。另外, 在样本范围内 $\alpha_i = \beta_i$ 的情形是较少出现的。

2) 当 $\alpha_i > \beta_i$ 时:

表 2 : 当 $\alpha_i > \beta_i$ 时的符号取值情况

α_i	β_i	$\alpha_i + \beta_i$	sgn
+	+	1	1
+	-	> 0	1
-	-	-1	-1
-	+	< 0	-1

由表 2 可以看出, 当 $\alpha_i > \beta_i$ 时, 价格变化对资金流向的影响要大于持仓量变动对资金流向的影响; 当价格上涨时, 无论持仓量增加还是减少, 资金流均被界定为流入的; 当价格下跌时, 无论持仓量增加还是减少, 资金流均被界定为流出

的。此时，价格势力的变动单方面地决定了资金的流向。

3) 当 $\alpha_i < \beta_i$ 时:

表 3 : 当 $\alpha_i < \beta_i$ 时的符号取值情况

α_i	β_i	$\alpha_i + \beta_i$	sgn
+	+	1	1
+	-	< 0	-1
-	-	-1	-1
-	+	> 0	1

由表 3 可以看出，当 $\alpha_i < \beta_i$ 时，持仓量变化对资金流向的影响要大于价格变动对资金流向的影响；当持仓量增加时，无论价格上涨还是下跌，资金流均被界定为流入的；当持仓量减少时，无论价格上涨还是下跌，资金流均被界定为流出的。此时，持仓量势力的变动单方面地决定了资金的流向。同时，由表 3 也可以看出，当持仓量的变动影响比较大时，价格下跌，可能对应的是资金流入，也可能对应的是资金流出。这就涵盖了期货标的合约价格下跌，可能是资金流入导致的，也可能是资金流出导致的情况。

4.2 问题二：资金流向对未来商品期货价格的影响及其持续性分析

资金流向作为一个技术指标，不仅具有反映了当期股票的超额供给与需求，而且具有反映了未来股票的超额供给与需求的实用价值。

前人研究表明：1) 从“羊群效应”理论的角度看，当期资金流向与未来资金流向之间、资金流向与股票收益率之间存在着正相关关系。2) 交易中冲击成本的存在，使得机构投资者持续的建仓或者减仓，使得当期资金流向与未来资金流向之间、资金流向与股票收益率之间存在着正相关关系。3) 从行为金融从行为金融学的角度出发，动量效应（正反馈）可以解释当期资金流向与未来资金流向之间、资金流向与股票收益率之间的正相关性，而反转效应（负反馈）可以解释当期资金流向与未来资金流向之间、资金流向与股票收益率之间的负相关性。

在期货市场中因为有做空机制，所以直接采用股票市场中的资金流向公式是无法刻画出期货标的合约真实的资金流向规律。本文考虑持仓量与成交量以及价

格的涨跌等综合因素，来刻画期货产品的资金流向，接下来我们将量化资金流向对未来商品期货价格的影响，并对持续性进行分析。

4.2.1 资金流向对未来商品期货价格的影响

1. 资金流向与未来期货价格之间的相关关系

为了检验当期资金流向与未来期货价格之间的相关关系，我们使用 2011 年至 2013 年样本数据中的 38 个期货合约分钟级交易数据，交易时间晚于 2011 年 1 月 1 日的，从开始交易时间采集。资金流向由基于期货市场构建的资金流向模型计算所得。

我们计算了所有期货合约的资金流向序列与未来期货价格序列的相关系数。其中资金流向序列 $MoneyFlow_{t-m,j}$ 指 j 期货合约在 t 时刻之前 m 分钟的标准化的资金流向，未来期货价格序列 $P_{t+m,j}$ 指 j 期货合约在 t 时刻之后 m 分钟的收盘价。如果考察过去 m (1-30) 分钟资金流向对未来 m (1-30) 分钟资金流向的相关关系，共 900 组可能组合。在下文中，我们分别列出了 1、5、10、15、20、30 分钟的等长信息序列周期与测试周期的期货合约相关系数（如过去 10 分钟的资金流向对未来 10 分钟的期货合约价格的相关系数），如表 4、表 5、表 6 所示：

表 4 相关系数小于-0.6 的期货合约列表

板块	期货合约	1/1	5/5	10/10	15/15	20/20	30/30
黑色系	j0000	-0.63	-0.63	-0.63	-0.63	-0.63	-0.63
黑色系	rb0000	-0.73	-0.73	-0.73	-0.73	-0.73	-0.73
化工	ru0000	-0.92	-0.92	-0.92	-0.92	-0.92	-0.92
化工	TA000	-0.81	-0.81	-0.81	-0.81	-0.81	-0.81
化工	v0000	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79
金属	ag0000	-0.65	-0.65	-0.65	-0.65	-0.65	-0.65
金属	cu0000	-0.87	-0.87	-0.87	-0.86	-0.86	-0.86
金属	zn0000	-0.91	-0.91	-0.91	-0.91	-0.91	-0.91
农产品	CF000	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80
农产品	p0000	-0.94	-0.94	-0.94	-0.94	-0.94	-0.94
农产品	PM000	-0.80	-0.80	-0.79	-0.78	-0.79	-0.78
农产品	SR000	-0.93	-0.93	-0.93	-0.93	-0.93	-0.93
农产品	WH000	-0.88	-0.88	-0.88	-0.88	-0.88	-0.88
农产品	y0000	-0.92	-0.92	-0.92	-0.92	-0.92	-0.92
油类	p0000	-0.94	-0.94	-0.94	-0.94	-0.94	-0.94
油类	y0000	-0.92	-0.92	-0.92	-0.92	-0.92	-0.92

如表 4, 共有 16 个期货合约的资金流向与未来商品期货的相关系数小于-0.6, 这说明这些的资金流向与未来商品期货价格存在强相关关系, 并且呈负相关关系。

表 5 相关系数介于-0.6 和 0.6 之间的期货合约列表

板块	期货合约	1/1	5/5	10/10	15/15	20/20	30/30
黑色系	i0000	0.06	0.06	0.07	0.07	0.07	0.07
黑色系	jm0000	-0.02	-0.02	-0.01	-0.01	-0.01	-0.01
化工	bb0000	-0.06	-0.07	-0.08	-0.09	-0.10	-0.13
化工	fb0000	-0.32	-0.32	-0.32	-0.31	-0.31	-0.30
化工	l0000	-0.31	-0.31	-0.31	-0.31	-0.31	-0.31
化工	bu0000	-0.61	-0.61	-0.61	-0.60	-0.60	-0.59
金属	al0000	0.15	0.15	0.15	0.15	0.15	0.15
金属	au0000	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50
农产品	a0000	0.45	0.45	0.45	0.45	0.45	0.45
农产品	b0000	-0.46	-0.45	-0.45	-0.44	-0.45	-0.43
农产品	c0000	0.30	0.30	0.30	0.30	0.30	0.30
农产品	jd0000	0.38	0.38	0.38	0.38	0.38	0.38
农产品	JR000	-0.44	-0.44	-0.42	-0.42	-0.39	-0.38
农产品	m0000	0.26	0.26	0.26	0.26	0.26	0.26
农产品	RM000	0.15	0.15	0.15	0.15	0.15	0.15
油类	fu0000	-0.40	-0.40	-0.40	-0.41	-0.41	-0.41

如表 5, 共有 16 个期货合约的资金流向与未来商品期货的相关系数介于-0.6 和 0.6 之间, 这说明这些资金流向与未来商品期货价格存在较弱的相关关系。

表 6 相关系数大于 0.6 的期货合约列表

板块	期货合约	1/1	5/5	10/10	15/15	20/20	30/30
化工	FG000	0.73	0.73	0.73	0.73	0.73	0.73
化工	wr0000	0.77	0.77	0.78	0.78	0.78	0.77
金属	pb0000	0.74	0.74	0.74	0.75	0.75	0.75
农产品	OI000	0.90	0.90	0.90	0.90	0.90	0.90
农产品	RI000	0.81	0.81	0.81	0.81	0.81	0.81
油类	RS000	0.81	0.81	0.81	0.81	0.81	0.82

如表 6, 共有 6 个期货合约的资金流向与未来商品期货的相关系数大于 0.6, 这说明这些的资金流向与未来商品期货价格存在强相关关系, 并且呈正相关关系。

综合表 4、表 5、表 6, 我们可以看出, 对于不同期货合约, 资金流向对未来商品期货价格的影响是不同的。因此在期货市场进行交易时, 可考虑选择分贝将存在强正相关性的期货合约和存在负正相关性的期货合约组成两个组合, 并针对其明显的相关性进行着重考察其收益率情况, 进而开展组合投资交易。

2. 资金流向与未来期货价格关系模型构建

由上文可知, 诸多期货合约存在较强的相关性关系, 为了量化资金流向对未来期货价格的影响, 我们通过线性回归方法, 来获取它们之间关系的定量结果。我们构建如下模型:

$$P_{t+m,j} = \varphi_{11} + \varphi_{12} \times MoneyFlow_{t-m,j} + \varepsilon_j$$

其中, $P_{t+m,j}$ 指 j 期货合约在 t 时刻之后 m 分钟的收盘价; $MoneyFlow_{t-m,j}$ 指 j 期货合约在 t 时刻之前 m 分钟内的标准化资金流向。

同样地, 如果考察过去 m (1-30) 分钟资金流向对未来 m (1-30) 分钟资金流向的线性关系, 共 900 组可能组合。在下文中, 我们分别考察了 1、5、10、15、20、30 分钟的等长信息序列周期与测试周期的期货合约线性相关关系 (如过去 10 分钟的资金流向对未来 10 分钟的期货合约价格的), 通过比较结果, 我们发现 1、5、10、15、20、30 分钟线性回归的结果相似, 鉴于篇幅, 在此我们仅罗列 1 分钟和 10 分钟的线性回归的结果, 其中包括解释度 R^2 、回归系数 φ_{12} 以及相应的 P 值, 如表 7、表 8 所示:

表 7 强负相关关系的期货合约资金流量与未来商品期货价格的关系

板块	期货 合约	1min/1min			10min/10min		
		R2	ϕ_{12}	pvalue	R2	ϕ_{12}	pvalue
黑色系	j0000	0.4002	-1.54715E-08	0.00	0.4001	-1.54688E-08	0.00
黑色系	rb0000	0.5275	-3.23765E-09	0.00	0.5273	-3.23745E-09	0.00
化工	ru0000	0.8524	-9.16337E-09	0.00	0.8522	-9.16232E-09	0.00
化工	TA000	0.6563	-1.51202E-08	0.00	0.6560	-1.51187E-08	0.00
化工	v0000	0.6176	-2.45227E-07	0.00	0.6176	-2.45275E-07	0.00
金属	ag0000	0.4181	-1.26389E-08	0.00	0.4179	-1.26387E-08	0.00
金属	cu0000	0.7484	-1.91336E-08	0.00	0.7482	-1.91301E-08	0.00
金属	zn0000	0.8290	-5.44943E-08	0.00	0.8289	-5.44871E-08	0.00
农产品	CF000	0.6398	-6.15603E-08	0.00	0.6400	-6.15776E-08	0.00
农产品	p0000	0.8821	-6.14507E-08	0.00	0.8821	-6.14482E-08	0.00
农产品	PM000	0.6376	-6.43357E-05	0.00	0.6296	-6.22948E-05	0.00
农产品	SR000	0.8715	-1.55971E-08	0.00	0.8714	-1.55953E-08	0.00
农产品	WH000	0.7730	-4.44379E-07	0.00	0.7727	-4.44355E-07	0.00
农产品	y0000	0.8478	-2.33229E-08	0.00	0.8477	-2.33235E-08	0.00
油类	p0000	0.8821	-6.14507E-08	0.00	0.8821	-6.14482E-08	0.00
油类	y0000	0.8478	-2.33229E-08	0.00	0.8477	-2.33235E-08	0.00

如表 7，具有强负相关关系的 16 个期货合约的线性回归结果显示，解释度 R2 均大于 0.4，回归系数 ϕ_{12} 均为负数，相对应的 P 值均接近 0，由此说明：具有强负相关关系的 16 个期货合约，资金流向为正，其未来的商品期货合约价格上升，资金流向为负，其未来的商品期货合约价格下跌。

表 8 强正相关关系的期货合约资金流量与未来商品期货价格的关系

板块	期货 合约	1min/1min			10min/10min		
		R2	ϕ_{12}	pvalue	R2	ϕ_{12}	pvalue
化工	FG000	0.5343	1.32181E-08	0.00	0.5336	1.32092E-08	0.00
化工	wr0000	0.5989	0.000446519	0.00	0.6007	0.000441462	0.00
金属	pb0000	0.5449	1.15187E-05	0.00	0.5538	1.16318E-05	0.00
农产品	OI000	0.8048	8.41357E-07	0.00	0.8050	8.41973E-07	0.00
农产品	RI000	0.6589	1.2086E-06	0.00	0.6587	1.20601E-06	0.00
油类	RS000	0.6589	5.24833E-07	0.00	0.6605	5.25512E-07	0.00

如表 8，具有强正相关关系的 6 个期货合约的线性回归结果显示，解释度 R^2 均大于 0.5，回归系数 ϕ_{12} 均为正数，相对应的 P 值均接近 0，由此说明：具有强正相关关系的 6 个期货合约，资金流向为负，其未来的商品期货合约价格上升，资金流向为正，其未来的商品期货合约价格下跌。

4.2.2 基于 Logistic 模型的资金流向与未来期货价格涨跌关系实证分析

1. Logistic 模型思想

在一般线性模型中反应变量 y 的值是有实际意义的，并假定 $y \sim N(\mu, \sigma^2)$ ，当 y 是二分类或 0-1 变量时， y 的取值为 0 或 1，仅是名义上的但并没有实际意义，此时 y 是服从 Bernoulli 分布的随机变量，即 $y \sim b(n, p)$ ，针对 0-1 变量，回归模型须做如下改进：

(1) 回归函数应该改用限制在 $[0,1]$ 区间内的连续曲线，而不能再沿用线性回归方程，应用较多的是 Logistic 函数（也称 Logit 变换），其形式为：

$$y = f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$$

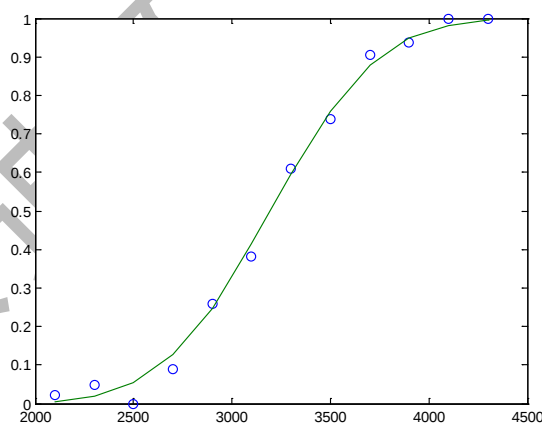


图 1 Logistic 函数 $f(x) = \frac{e^x}{1 + e^x}$ 的“S”型曲线

(2) 因变量 y_i 本身只取 0, 1 值，不适于直接作为回归模型中的因变量，设 P 表示 $y=1$ 的概率，Q 表示 $y=0$ 的概率， $Q=1-P$ 。概率 P 是有实际意义的，它表示 y 取值为 1 的可能性的。假定在观察反应变量的同时，观察了 p 个解

释变量 $x_1, x_2, \dots, x_p$ ，用向量 X 记 $(x_1, x_2, \dots, x_p)'$ ，与线性模型不同的是，我们不是研究反应变量的值与解释变量之间的关系，而是研究反应变量取某值的概率 P 与解释变量之间的关系。实际观察结果表明概率 P 与解释变量之间的关系不是呈线性关系，而是呈“S”型曲线关系。这是因为概率分布函数是一条“S”型曲线。Logistic 函数是呈“S”型的曲线，见图 1，故此一般用 Logistic 曲线来描述 P 与解释变量 x 之间的关系：

$$P = P(y = 1 | X) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} = \frac{\exp(X\beta)}{1 + \exp(X\beta)}$$

对该式做 Logit 变换，得

$$\text{Logit}(y) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = X\beta$$

上式称为 Logistic 回归模型，其中 $\beta_0, \beta_1, \dots, \beta_p$ 为待估参数。

2. Logistic 模型建立

由上文可知，很多期货合约在长期存在较强的相关性关系，为了进一步量化资金流向对未来期货价格的影响，我们通过 Logistic 模型，利用资金流向预测下一期商品期货价格的涨跌。我们构建如下模型：

$$\text{Logit}(y) = \ln\left(\frac{p_{t+m,j}}{1-p_{t+m,j}}\right) = \beta_0 + \beta_1 \times \text{MoneyFlow}_{t-m,j}$$

其中， $p_{t+m,j}$ 指 j 期货合约在 t 时刻之后 m 分钟的期货价格上涨的概率； $\text{MoneyFlow}_{t-m,j}$ 指 j 期货合约在 t 时刻之前 m 分钟内的资金流向； $p_{t+m,j}$ 表示 $y = 1$ （期货价格上涨）的概率， $q = 1 - p_{t+m,j}$ 表示 $y = 0$ （期货价格下跌）的概率。

3. Logistic 模型实证分析

本文利用 2011 年至 2013 年每个期货合约最后的 1000 个数据（数据长度不足 1000 则使用全部数据），将最后 40 个数据作为测试集，前面的数据作为训练集，建立 logistic 回归模型，分析近期内期货市场资金流向与未来商品期货价格涨跌的关系。

同样地，如果考察过去 m （1-30）分钟资金流向对未来 m （1-30）分钟期货价格的 Logistic 回归关系，共 900 组可能组合。在本文中，我们分别考察了 1、10、20、30 分钟的等长信息序列周期与测试周期的期货价格涨跌（如过去 10 分

钟的资金流向对未来 10 分钟的期货合约价格涨跌的预测)，通过比较，我们发现 1、10、20 分钟 Logistic 回归结果略有不同。其中显示结果包括回测准确率 Q、回归系数 β_1 以及相应的 P 值，其中 P 值小于 0.15 的用浅绿色底纹标注，如下列表格所示：

表 9 黑色系板块期货合约 Logistic 回归结果

期货 合约	1 分钟			10 分钟			20 分钟		
	Q	回归系数	p 值	Q	回归系数	p 值	Q	回归系数	p 值
i0000	0.78	2.76E-10	0.98	0.73	-4.1E-09	0.47	0.58	-3.3E-09	0.61
j0000	0.65	-2.7E-11	0.00	0.73	-1.7E-11	0.00	0.65	-1.2E-10	0.57
jm0000	0.70	3.1E-09	0.58	0.73	1.25E-11	0.97	0.68	-1.5E-10	0.63
rb0000	0.68	-1.9E-12	0.00	0.55	-1.1E-12	0.00	0.55	-9.1E-13	0.00

如表 9 所示，黑色系板块中，短期内的回测准确率均在 55%以上，我们经过参数检验发现，仅有 rb0000 期货合约在 1、10、20 分钟级别数据均是通过检验的，i0000、j0000 期货合约在 1、10 分钟级别数据也能通过检验。

表 10 化工板块期货合约 Logistic 回归结果

期货 合约	1 分钟			10 分钟			20 分钟		
	Q	回归系数	p 值	Q	回归系数	p 值	Q	回归系数	p 值
FG000	0.70	2.12E-09	0.13	0.70	1.41E-10	0.04	0.65	-2.6E-12	0.91
TA000	0.73	-4.4E-12	0.00	0.63	-1.9E-12	0.00	0.50	-1.6E-12	0.00
bb0000	0.68	4.37E-10	0.97	0.50	-5E-10	0.95	0.58	-1.2E-09	0.92
bu0000	0.83	1.13E-08	0.11	0.55	-2.9E-10	0.33	0.55	3.91E-10	0.03
fb0000	0.70	5.44E-08	0.01	0.68	8.3E-09	0.49	0.60	6.07E-09	0.75
l0000	0.83	-2.9E-12	0.00	0.55	-1.1E-12	0.00	0.45	-5.5E-13	0.03
ru0000	0.68	-1.7E-13	0.00	0.60	-8.9E-14	0.01	0.65	-9.1E-14	0.01
v0000	0.83	-1.5E-10	0.00	0.70	9.65E-10	0.73	0.63	-1.8E-09	0.38
wr0000	0.60	2.55E-07	0.07	0.60	1.78E-07	0.34	0.58	-8.1E-08	0.79

如表 10 所示，在化工板块中，短期内的回测准确率均在 50%以上，我们经过参数检验发现，仅有 bb0000 期货合约在 1 分钟级别数据是不通过检验的。

表 11 金属板块期货合约 Logistic 回归结果

期货 合约	1 分钟			10 分钟			20 分钟		
	Q	回归系数	p 值	Q	回归系数	p 值	Q	回归系数	p 值
ag0000	0.70	-3.8E-12	0.00	0.70	5.37E-13	0.95	0.55	3.51E-12	0.55
al0000	0.78	-1.8E-09	0.51	0.65	9.99E-10	0.09	0.58	1.13E-09	0.00
au0000	0.78	-3.7E-08	0.17	0.68	4.67E-09	0.53	0.65	2.18E-09	0.29
cu0000	0.60	-7.7E-13	0.00	0.48	-2.1E-13	0.00	0.50	-1.8E-13	0.01
pb0000	0.68	-7.6E-09	0.67	0.63	3.64E-09	0.30	0.65	-6.7E-10	0.83
zn0000	0.70	-1.1E-11	0.00	0.63	-6.1E-12	0.00	0.55	-3.1E-12	0.00

如表 11 所示，金属板块中，短期内的回测准确率均在 55%以上，我们经过参数检验发现，cu0000、zn0000 期货合约在 1、10、20 分钟级别数据均通过检验。

表 12 农产品板块期货合约 Logistic 回归结果

期货 合约	1 分钟			10 分钟			20 分钟		
	Q	回归系数	p 值	Q	回归系数	p 值	Q	回归系数	p 值
CF000	0.63	-5.9E-12	0.00	0.55	-3.4E-12	0.00	0.48	-1.6E-12	0.00
JR000	0.75	-1.3E-07	0.10	0.55	-4.8E-09	0.84	0.50	1.62E-08	0.66
OI000	0.65	-6.8E-11	0.69	0.60	1.14E-11	0.90	0.60	5.87E-11	0.58
PM00	0.48	-1.3E-07	0.34	0.55	-6.1E-09	0.96	0.48	7.67E-08	0.81
RI000	0.75	3.14E-08	0.21	0.70	6.24E-10	0.81	0.73	2.19E-09	0.05
RM000	0.55	-4.4E-11	0.59	0.45	-7.5E-11	0.22	0.48	-6.2E-11	0.17
SR000	0.68	-5.9E-12	0.00	0.68	-2.6E-12	0.00	0.60	-1.9E-12	0.00
WH000	0.85	-6E-09	0.47	0.60	1.22E-09	0.59	0.65	5.86E-11	0.89
a0000	0.58	-2.2E-11	0.00	0.55	-1E-11	0.00	0.50	-6.5E-12	0.00
b0000	0.50	8.08E-08	0.37	0.48	3.33E-08	0.47	0.53	-4.4E-08	0.52
c0000	0.73	-1.1E-10	0.00	0.75	-1.6E-09	0.17	0.53	-4E-10	0.51
jd0000	0.68	-1.1E-09	0.66	0.55	1.61E-10	0.46	0.50	1.34E-11	0.97
m0000	0.70	-2.1E-12	0.00	0.53	-1.2E-12	0.00	0.45	-1.2E-12	0.00
p0000	0.55	-7E-12	0.00	0.60	-2.4E-11	0.46	0.55	-2.9E-11	0.32
y0000	0.70	-3.7E-12	0.00	0.60	-1.5E-12	0.00	0.60	-5E-12	0.48

如表 12 所示，农产品板块中，短期内的回测准确率多数保持在 50%以上，我们经过参数检验发现，CF000、SR000、a0000、m0000 期货合约在 1、10、20 分钟级别数据均通过检验。

表 13 油类板块期货合约 Logistic 回归结果

期货 合约	1 分钟			10 分钟			20 分钟		
	Q	回归系数	p 值	Q	回归系数	p 值	Q	回归系数	p 值
RS000	0.60	-8.9E-09	0.16	0.55	-2.1E-09	0.33	0.425	2.03E-10	0.66
fu0000	0.50	-2.2E-07	0.43	0.425	-4.5E-09	0.01	0.575	-9.6E-10	0.30
p0000	0.55	-7E-12	0.00	0.6	-2.4E-11	0.46	0.55	-2.9E-11	0.32
y0000	0.70	-3.7E-12	0.00	0.6	-1.5E-12	0.00	0.6	-5E-12	0.48

如表 13 所示，油类板块中，短期内的回测准确率多数保持在 50%以上，我们经过参数检验发现，有 y0000 期货合约在 1、10 分钟级别数据均通过检验。

综合表 9、10、11、12、13，我们利用资金流向预测下一期商品期货价格的涨跌，大多数情况下能获得 50%以上的准确率，同时也会随时间级别数据变化产生微小波动。

4.2.3 资金流向的持续性分析

我们利用板块资金流向的流量数据（即将板块的累加资金流向值序列进行一阶差分）来研究板块资金流向的持续性。在以下述图中，M1 表示“黑色系”板块，M2 表示“化工”板块，M3 表示“金属”板块，M4 表示“农产品”板块，M5 板块表示“油类”板块。我们首先对板块资金流向的流量数据进行了 ADF 检验，发现所有的板块的流量数据均是平稳的时间序列，并进一步通过计算自相关系数 ACF 来判断资金流向的持续性。

表 14 ADF 平稳性检验表

板块	Statistics	p.value
黑色系	-51.694	0.01
化工	-53.307	0.00
金属	-51.609	0.00
农产品	-52.082	0.01
油类	-48.004	0.01

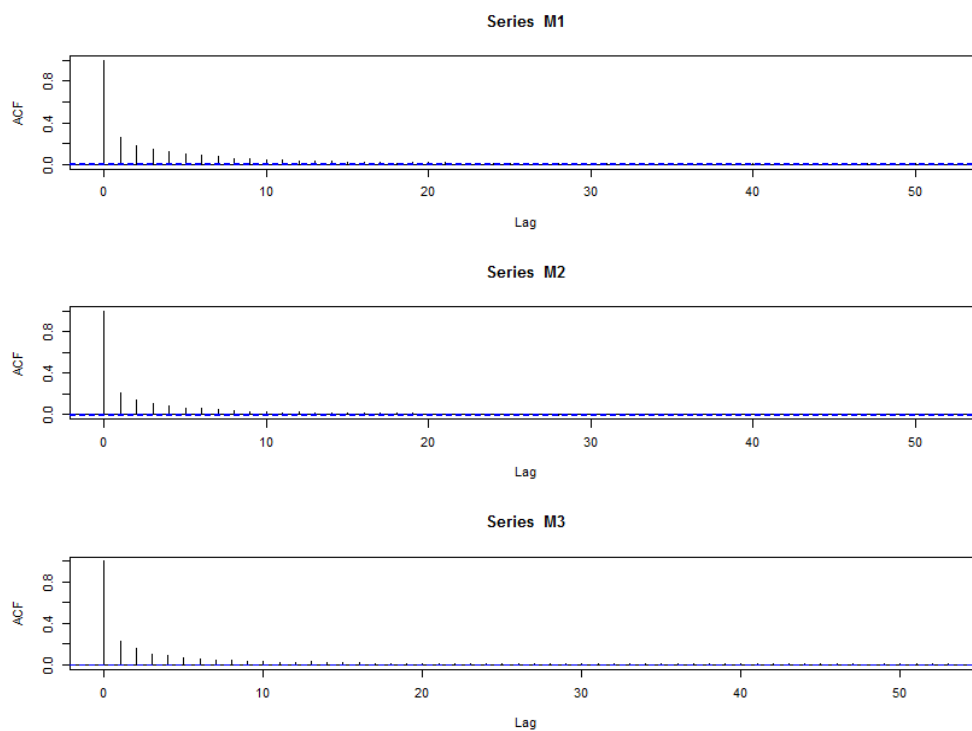


图 2 板块自相关系数图

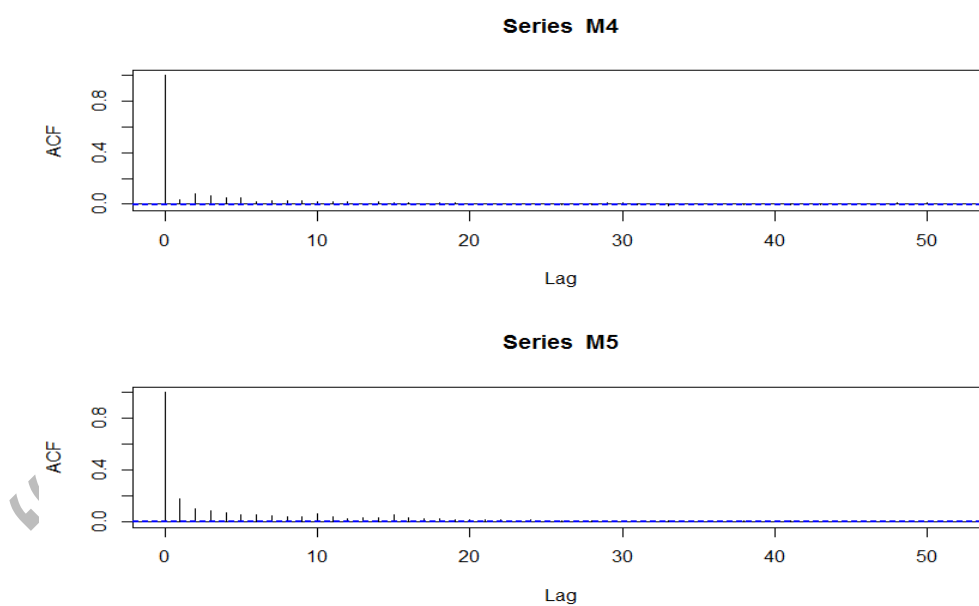


图 3 板块自相关系数图

由图 2 和图 3 可知，各板块资金流向的流量数据呈现出正的自相关现象，在各板块的前 10 个滞后期长度内，各自相关系数是落在均值为 0，置信度为 0.95。

置信区间之外，所以我们可以认为各个板块资金流向的流量变化是具有正的持续性的，也就意味着，在短期内，各个板块的资金流向是具有正相关持续性的。

4.3 问题三：资金流向特征分析

4.3.1 板块间资金流向的相似特点分析

1. 数据处理

首先，我们将利用 1 分钟级的数据计算 MoneyFlow（即 $n=1$ ，每一个分钟的时间点都计算出一个资金流向），并且计算出累加的资金流向序列用于资金流向的分析；另一方面，我们将商品期货市场的 38 个主力合约按照题目附件参考给出的板块分类方法分成 5 大类，分别为“黑色系”、“化工”、“金属”、“农产品”以及“油类”。由于所获取的 2011 到 2013 年的样本数据中，并不是每个合约的数据长度都一样（交易时间晚于 2011 年 1 月 1 日的期货合约，数据从开始交易时间采集），我们在计算板块在 3 年间各个时间段的资金流向时，将合约缺失的时间处对应的资金流向值赋为零，继而将板块内各合约的累加资金流向值序列横向加总，得到整个板块的累加资金流向值序列。

2. 板块间的资金流向分析

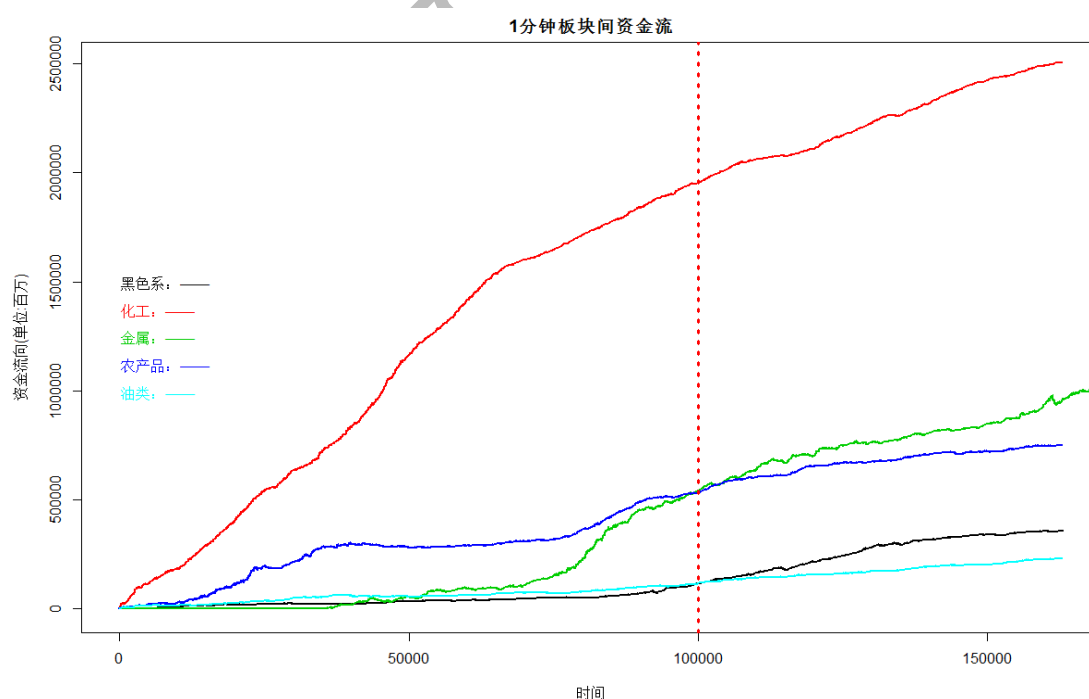


图 4 整个市场的资金流向概况

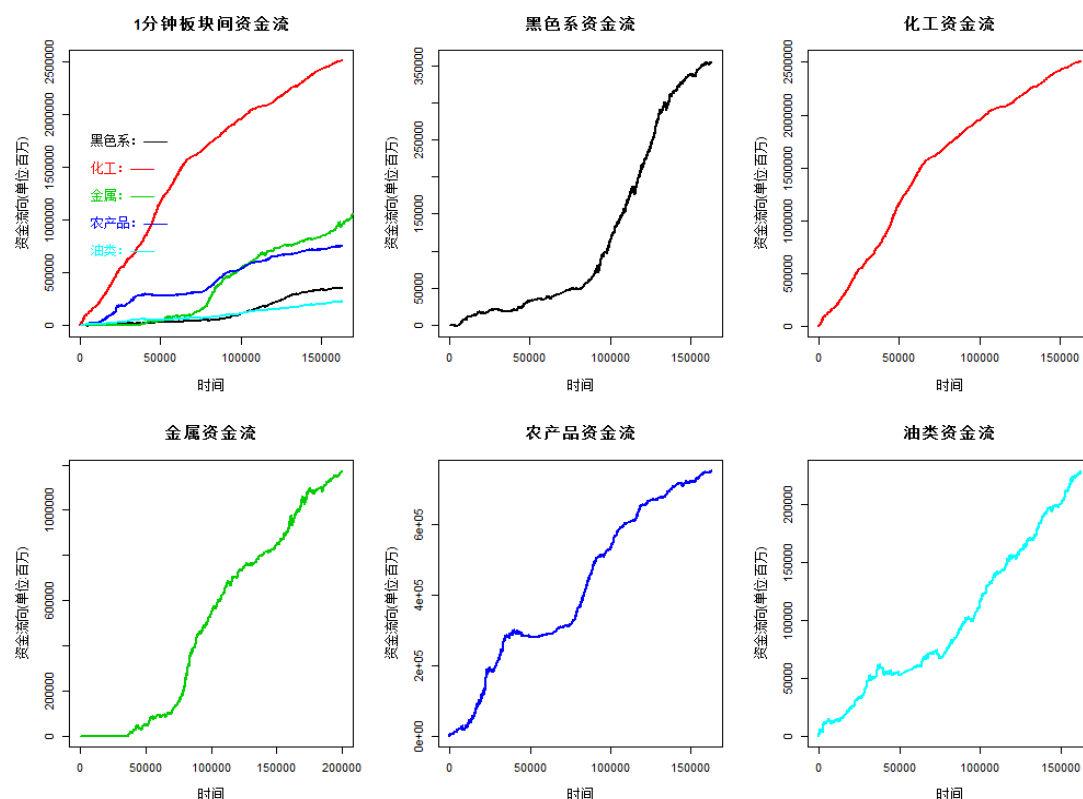


图 5 各个板块资金流向

由图 4 和图 5，可以发现，各个板块的累加资金流向值虽然在较短的时间段内存在波动，但就长期而言是在增长的；这是符合商品期货市场的变化规律的，由于我们此处比较的是累加的资金流向（总量），所以图 4 和图 5 表明，随着商品期货市场的发展以及其他经济因素，其曲线会呈现总体上升的趋势；而资金流方向的变动则是体现在曲线的波动上，曲线在某时间点向上转折表示资金流入，向下转折表示资金流出。

3. 板块间的轮动效应

由图 4 可以清晰地看到，就整个商品期货市场而言，“化工”板块所占资金流的比重最大，而另外几个板块，则随着时间的变化其在整个市场所占有的资金流的比重呈交替领先的情况。

表 15 轮动效应概况

	2012 年 12 月前		2012 年 12 月后	
	板块	增长速度	板块	增长速度
主力	化工	较快	化工	平缓
	农产品	较快	金属	较快
	油类	较快	黑色系	较快
非主力	金属	较慢	农产品	较慢
	黑色系	较慢	油类	较慢

特别地，我们发现 1 分钟级时间点 100000（即大概为 2012 年年末）可以作为一个时间分割点，在此之前，“农产品”板块的累加资金流向值高于“金属”板块，“油类”板块的累加资金流向值高于“黑色系”板块；而在此之后，“农产品”板块的累加资金流向值低于“金属”板块，“油类”板块的累加资金流向值低于“黑色系”板块。这说明，在 2012 年年末之后，“农产品”板块及“油类”板块的资金流增速减缓，而“金属”板块及“黑色系”板块的资金流增速加快，整个市场的资金流缓慢地向“化工”、“金属”及“黑色系”这三个板块转移，推动商品期货市场资金流增长的主力军在这个时间分割点之后变为“化工”、“金属”及“黑色系”三大板块。

4.3.2 各个板块的资金流向统计分析

除了板块间的累加资金流向值序列具有这样的关系外，板块内各合约的累计资金流向也具有类似的关系，通过以下图，我们可以发现各板块中的“主力”合约。

（注：以下图中，线条是为 0 的水平直线表明该段时间上，相应的合约缺少数据）

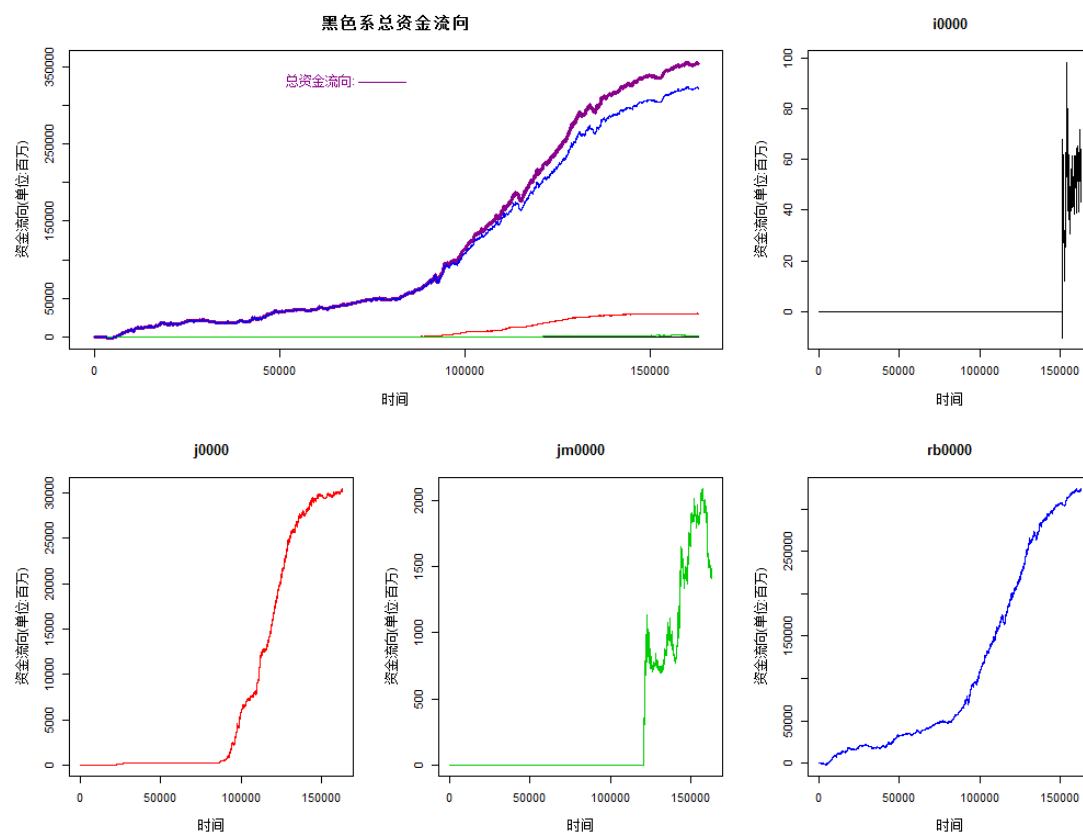


图 6 黑色系板块资金流向概况

从图 6 我们可以看出，“黑色系”板块的 4 个合约中合约“rb0000”的资金流所占的比重最高，涵盖了板块总资金流的绝大部分；其余的合约所占的资金流向则很少，并不会对板块的总资金流造成太大的影响；因此，对于“黑色系”板块，我们可以通过观察合约“rb0000”的资金流向的变动来获得板块总资金流变动的近似变动。

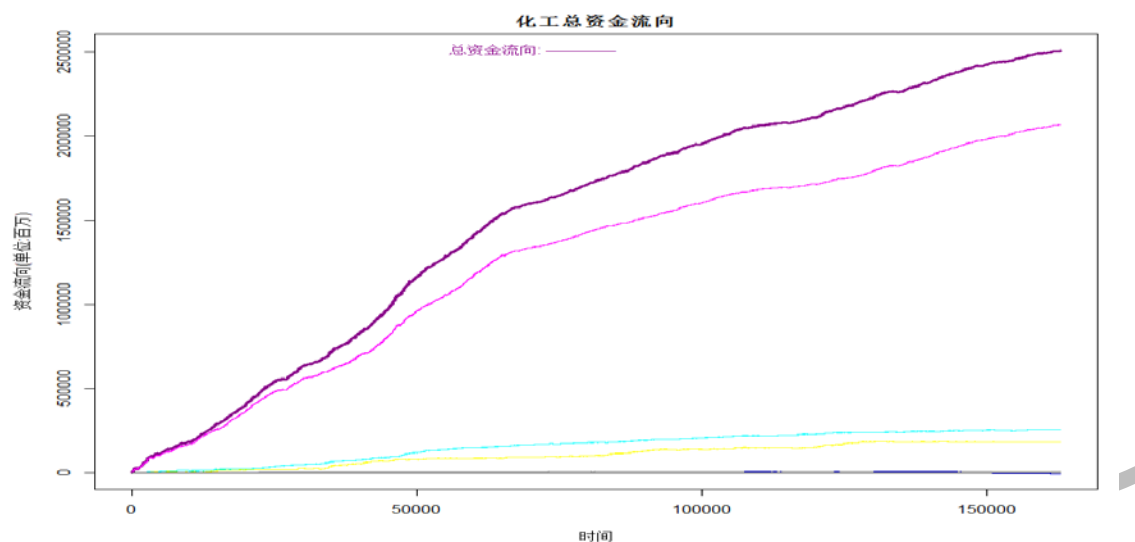


图 7 化工板块总体资金流向概况

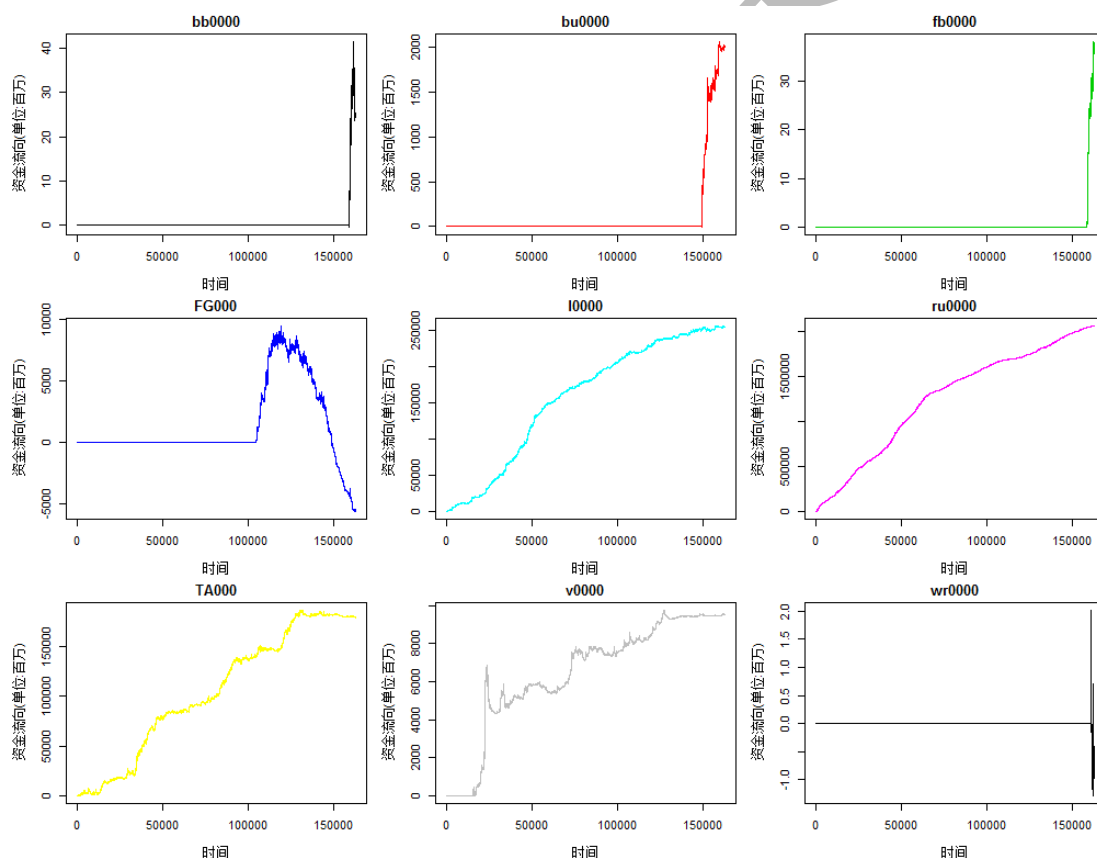


图 8 化工板块期货合约资金流向概况

从图 7 和图 8 我们可以看出，“化工”板块的 9 个合约中“ru0000”的资金流所占的比重最高，涵盖了板块总资金流的绝大部分；其余的合约所占的资金流则很少，不对板块的总资金流造成太大的影响；因此，对于“化工”板块，我们可以通

过观察合约“ru0000”的资金流向的变动来获得板块总资金流变动的近似变动。另外，“FG000”、“wr0000”的资金流向是向上升后下降的。

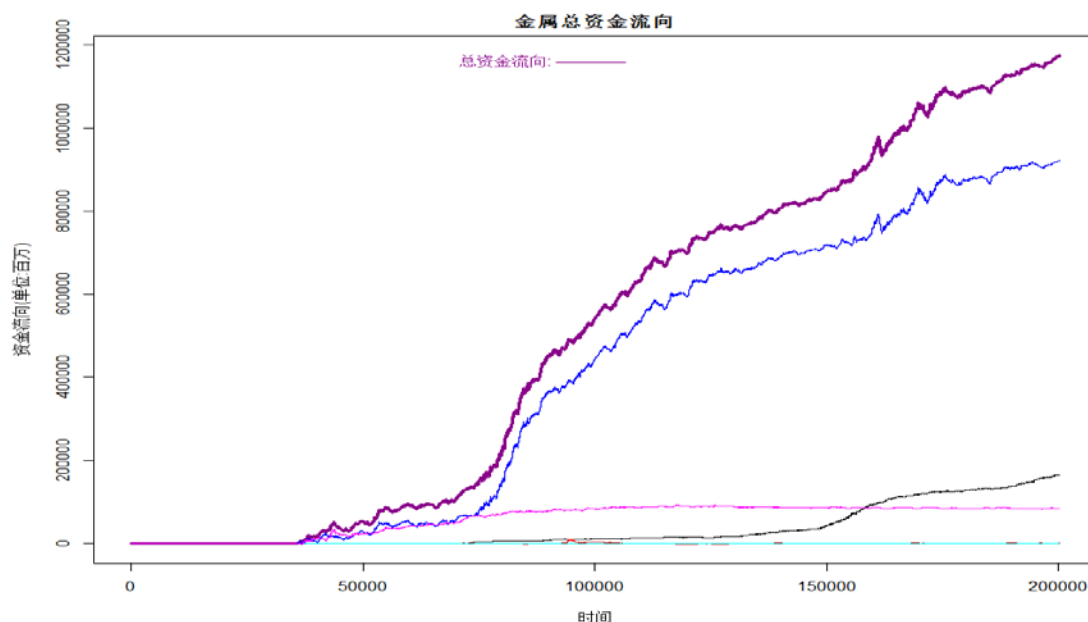


图9 金属板块资金流向概况

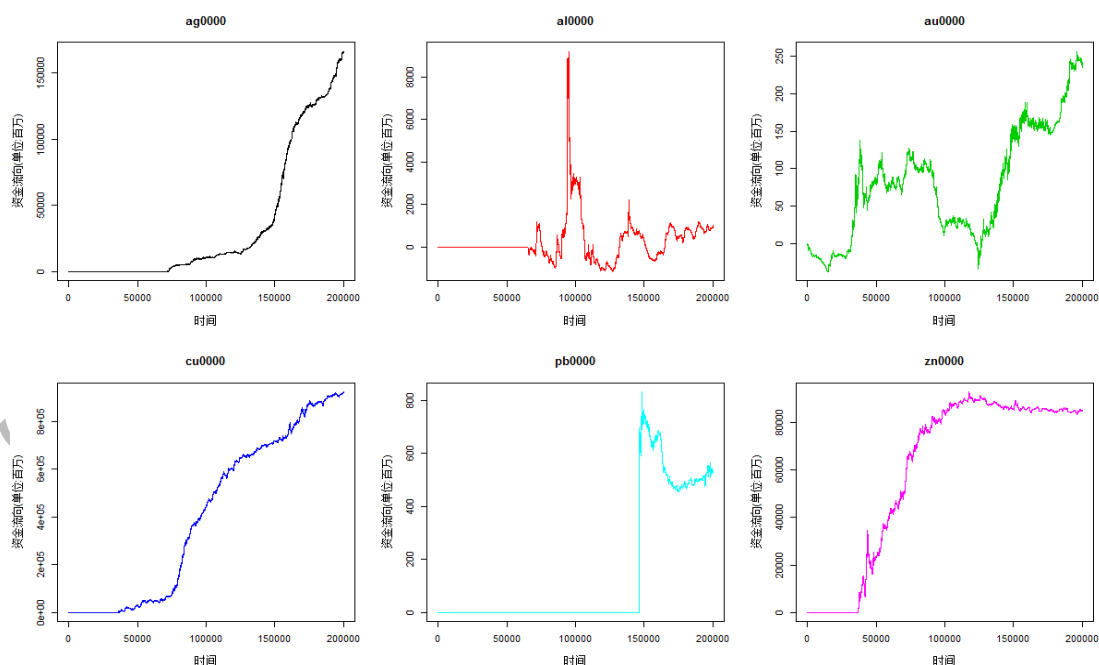


图10 金属板块期货合约资金流向概况

从图9和图10我们可以看出，“金属”板块的6个合约中“cu0000”的资金流所

占的比重最高，涵盖了板块总资金流的绝大部分；其余的合约所占的资金流则很少，不对板块的总资金流造成太大的影响；因此，对于“金属”板块，我们可以通过观察合约“cu0000”的资金流向的变动来获得板块总资金流变动的近似变动。另外，“al0000”、“au0000”的资金流向具有较大的波动趋势。

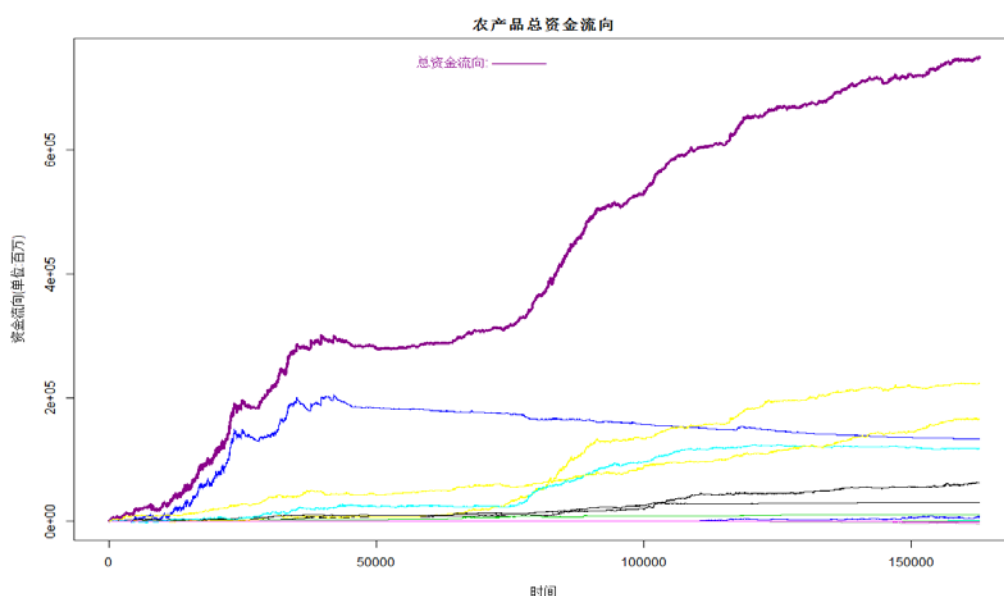


图 11 农产品板块资金流向概况

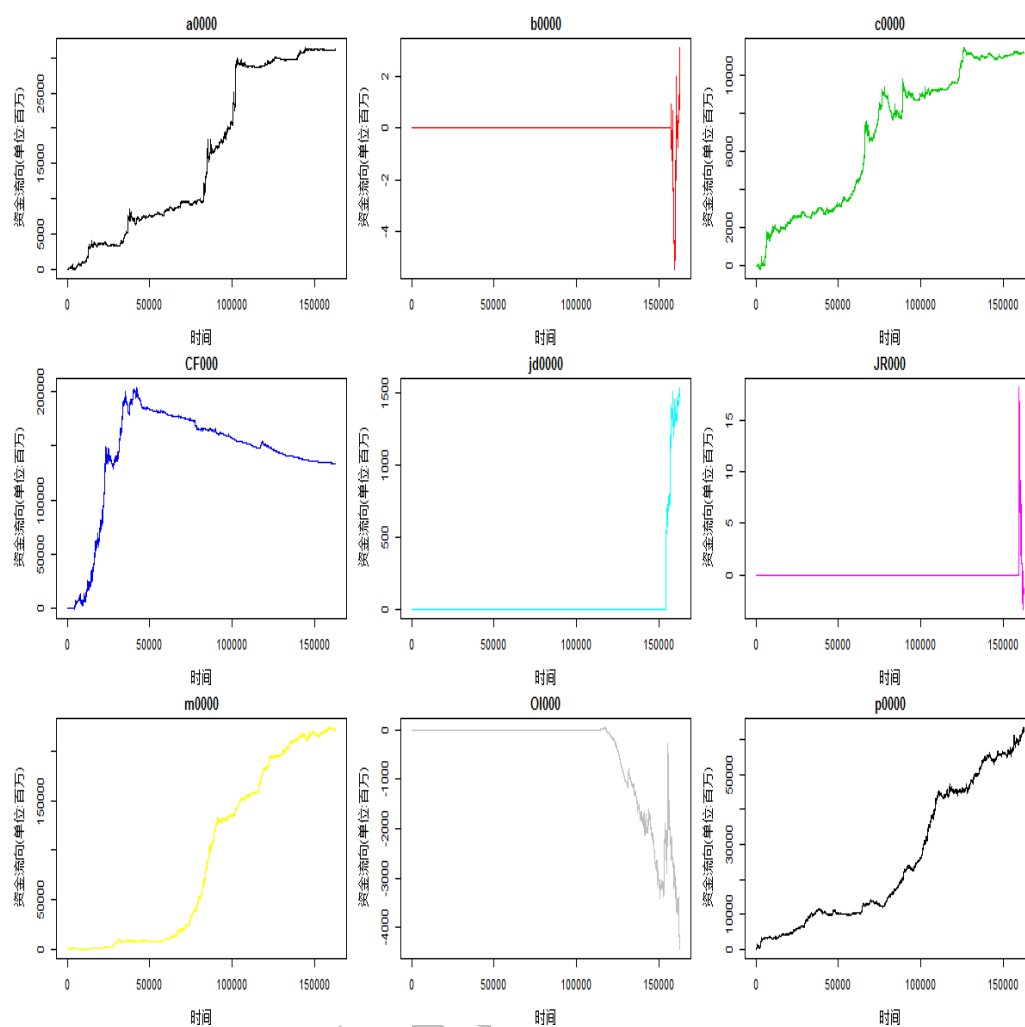


图 12 农产品板块期货合约资金流向

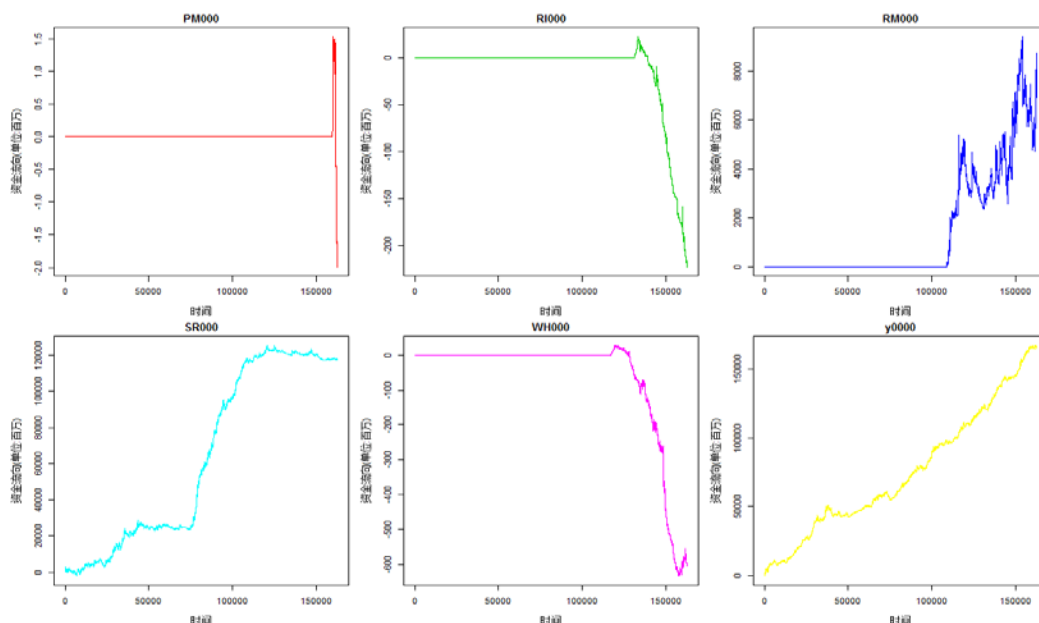


图 13 农产品板块期货合约资金流向

从图 11、图 12 和图 13，我们可以看出，“农产品”板块的 15 个合约所体现出的规律与其他板块有所不同。该板块中，没有一个合约能够单独决定板块的资金流向大小；其中“CF000”、“m0000”、“y0000”及“SR0000”的资金流向值比较大且相接近，其他合约的资金流向值则相对地较小。“CF000”的资金流向变化比较突出，在 11 年快速上升，随后便是逐步下降。

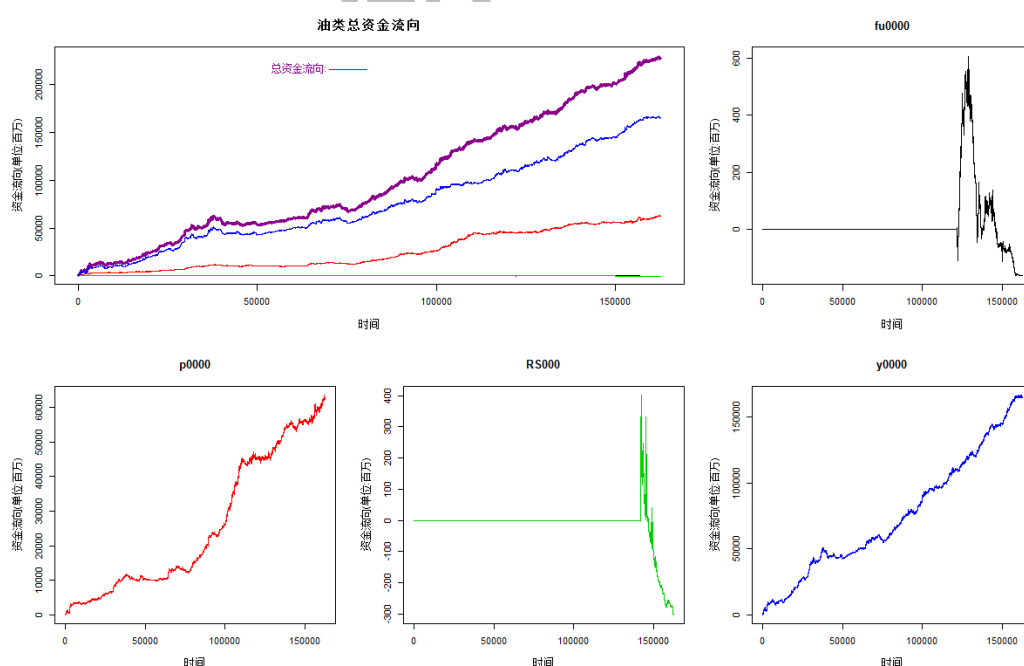


图 14 油类板块期货合约资金流向

从图 14 们可以看出，“油类”板块的 4 个合约中合约“y0000”和“p0000”的资金流所占的比重较大，涵盖了板块总资金流的绝大部分；其余的合约所占的资金流则很少，不对板块的总资金流造成太大的影响；因此，对于“黑色系”板块，我们可以通过观察合约“y0000”和“p0000”的资金流向的变动来获得板块总资金流变动的近似变动。

4.4 问题四：基于资金流向的二分类 Logistics 回归策略设计

4.4.1 策略设计

1. 交易机制

平仓时间：本文在程序中设置了平仓时间，规定在上午 9:30 之前下午 14:50 之后进行平仓，不进行交易。

止赢、止损、追踪止盈止损机制：为了控制风险，本文对交易设置严格的止损止盈机制；在模型中添加 openprice, histExtre 变量记录开盘价和交易期间的极值，用来跟踪止赢和止损。在模型中设置止损率 stoploss 为 0.05，止盈率 stopprofit 为 0.1，追踪止盈率 trailinggap 为 0.03。

采样频率：本文选取 20140101 到 20151231 时期的期货分钟数据用于回测，采样频率为 30 分钟，交易模型的训练集样本数为 2000，交易手数 1000 手。

2. 二分类 Logistic 模型建立

为了进一步预测未来期货价格涨跌的情况，我们建立了二分类 Logistic 回归模型，引入了资金流向、持仓量、成交量、收盘价等指标来预测下一期商品期货价格的涨跌。我们构建如下的二分类 Logistic 模型：

$$\text{Logit}(y) = \ln\left(\frac{p_j}{1-p_j}\right) = \beta_0 + \beta_1 \times d\text{MoneyFlow}_{t,10,j} + \beta_2 I_{t,5,j} + \beta_3 \text{Volume}_{t,20,j} + \beta_4 P_{t,5,j} + \beta_5 P_{t,20,j}$$

其中， p_j 指的是 j 期货合约下一期商品期货价格上涨的概率； p_j 表示 $y=1$ （期货价格上涨）的概率， $q_j=1-p_j$ 表示 $y=0$ （期货价格下跌）的概率； $d\text{MoneyFlow}_{t,10,j}$ 指 j 期货合约差分资金流向在 t 时刻之前 10 期的移动平均值； $I_{t,5,j}$ 指 j 期货合约在 t 时刻之前 5 期的持仓量的移动平均值； $\text{Volume}_{t,20,j}$ 指 j 期货合约在 t 时刻之前 20 期的成交量的移动平均值； $P_{t,5,j}$ 指 j 期货合约在 t 时刻之前

5 期的收盘价的移动平均值； $P_{t,20,j}$ 指 j 期货合约在 t 时刻之前 20 期的收盘价的移动平均值。

3、交易指令（信号）

通过二分类 logisitc 模型的输出结果，得到期货合约下一期涨跌的概率 p ，如果概率值 p 大于 0.5，则认为该合约下一期的价格会上涨，此时程序将会发出一个买入的指令，建立多仓，否则程序将会给出一个卖空的指令，建立一个空仓。

4.4.2 回测结果

表 16 策略在 Auto-Trader 平台的回测结果

	合约	初始收益率	年化收益率	总体收益率
收益	pb0000	19.46	11.49	15.71
	TA000	5.49	3.04	
	y0000	12.36	6.19	
	p0000	34.76	17.48	
	WH000	6.47	3.44	

由表 16 可以发现，初始收益率最高的合约是“p0000”，收益率 34.76；其次，初始收益率较高的合约为“pb0000”、“y0000”，收益率分别为 19.46、12.36；而初始收益率较低的合约为“TA000”、“WH000”，收益率也有 5.49、6.47。另一方面，从总体收益率，我们可以看出，总体收益率为 15.71。

5 模型的评价与推广

5.1 模型的优点

本文的资金流向模型不仅考虑了价格和成交量的影响，也考虑了持仓量变动和期货价格变动对商品期货市场资金流向的影响。我们构建了期货价格变动权重 α_i 以及持仓量变动权重 β_i ，通过赋予期货价格方向变化和持仓量方向变化权重，充分考虑两种变化率的不同对资金流向的影响，能够较好地度量了期货价格变化与持仓量变化对资金流向的影响。本文建立的资金流向模型涵盖了商品期货市场资金流向与价格变动之间的关系，即期货价格下跌，可能对应于资金流入，也可能对应于资金流出，这在一定程度上避免了直接套用股市资金流向公式所带来累加性错误。

5.2 模型的局限性

价格变动权重 α_i 以及持仓量变动权重 β_i 的计算方法对于本文的资金流向模型具有较大的影响。由于市场交易复杂，在本文中计算权重的方法基于资金流向的变化决定于价格和持仓量的变化速度，可能并不能完全体现出商品期货市场中资金流向与价格、成交量及持仓量等因素的真实关系。

参考文献

- [1] 陈鸿, 股指期货市场有效性与资金流向分析, Finance Economy, 2015
- [2] 侯丽薇, 谢赤, 戴军, 葛新元, A 股市场资金流向指标应用分析, 证券市场导报, 2010
- [3] E Olaison, High-Frequency Trading: How Money Flow Affects Stock Returns, 《Business & Economics》, 2013, 27(8):2267-2306
- [4] 俞乔.市场有效、周期异常与股价波动[J].经济研究, 1994 (8)
- [5] 沈艺峰.会计信息披露和我国股票市场半强式有效性的实证分析[J].会计研究, 1996 (1)
- [6] 中国期货业协会, 期货市场教程, 北京: 中国财政经济出版社, 2013
- [7] 弗兰克 J.法博齐, 金融市场与金融机构基础, 北京: 机械工业出版社, 2013
- [8] 欧阳红兵, 量化投资, 北京: 北京大学出版社, 2016
- [9] 卓金武, 周英, 量化投资数据挖掘技术与实践 (MATLAB 版), 北京: 电子工业出版社, 2015.5
- [10] 石榴红, 期货、期权理论与实务, 北京, 科学出版社, 2016