**期货品种性质研究进展报告**

**山东大学齐鲁证券金融研究院**

**期货各品种基础性质研究课题组**

**项目负责人：嵇少林 教授**

**项目组成员：刘浩东 时晓敏 陈俊 孔垂柳 王丽 王文达 张洁 曹旭 谌强 邢元杰 韩强 徐翔**

**期货品种基础性质研究**

**1 概述**

成熟稳健的的交易应该包括下列环节：

1. 决定买（卖）什么，确定资产配置，即品种选择。
2. 何时买（卖），仓位控制——交易策略。
3. 极端事件的风险防范——量化投资的风险控制。
4. 同期同类收益率的比对——绩效评估。

基于项目服务对象是万家基金管理有限公司。其主要投资方式为买入平仓，而非实物交割；其偏好运作模式为短期的量化投资，而非完全的基本面分析，长期持有。因此，依据项目要求，研究的重点主要是量化投资下的品种选择和策略研究两个方面，最后再依据风险控制和绩效评估的要求，完善量化投资程序的开发，彻底实现项目的实用性要求。

期货品种可分为两大类：一类是金融期货，一类是商品期货。我国期货市场现状是，金融期货只有5年国债期货和沪深三百指数期货两种；商品期货有农产品、金属和能源化工三大类40个品种。

因此，在选择期货品种时，可以首先根据投资者偏好以及对不同板块的熟悉度选择商品期货或金融期货。在商品期货品种选择中，可以先在不同板块中选取能够发挥量化投资优势的优良品种，然后按照适当的资产配置，对不同的品种以不同的比例持有。对于金融期货品种的选择，优于国内只有5年国债期货和沪深三百指数期货两种，所以可以根据投资者偏好及熟悉度进行选择。

在仓位建立完成后，持有比例并不是一成不变的。一者，长期来看，随着市场行情的波动，新数据的产生，依据原有的品种选择策略就应该有一定幅度地调整持有比例。再者（也更为重要的是），短期来看，量化投资更能够抓住短期的相关变量的变动，更能够捕捉到短期的非理性波动，并能确定买卖时机以期获得短期资本利得。

如此便构成了品种选择与交易策略。综上，有必要对这两个问题进行进一步的界定。

品种选择即按照“从上到下”的顺序，先选择期货类，再选择具体品种，然后分配资金。

交易策略即通过数据分析，先确定大概价格区间，再确定价格走势（等同于确定买卖时机），实现低买高卖，获取资本利得。

**2 交易品种选择**

在选择商品期货种类时，可采用因子分析法分析历史交易数据，量化商品期货各品种的流动性及波动性等情况，最后选取波动性及流动性均表现良好的品种作为优良品种，作为投资组合待选品种。

**2.1 数据选择与处理**

我们选取的原始数据是，2009年1月1日至2013年12月31号（未上市品种从上市之日起算起）各品种期货的日交易数据。变量包括前结算价，开盘价，最高价，最低价，收盘价，结算价，成交量，成交金额，持仓量，涨跌1（收盘价-前结算）以及涨跌2（结算价-前结算）。

分析原始数据样本各变量的实际意义可发现：

（1）原始数据变量中，开盘价、最高价、最低价和收盘价这四个价格表示的是一天价格波动中的4个极端：时间的两个极端(开盘价和收盘价)与价格的两个极端(最高价和最低价)。收盘价的高低往往反映出市场资金对某只个股的关注程度，有预示下一个交易日的演绎方向的功能，所以收盘价的表现是值得关注的。

（2）成交量的高低不仅取决于市场的投资热情，还取决于相应产品的吸引力大小，以及投资者对该产品的熟悉程度。成交额与其具有相同意义。而持仓量的大小则反映了市场交易规模的大小，也反映了多空双方对当前价位的分歧大小。

（3）涨跌1和涨跌2描述了当日收盘价和结算价与前日相比涨跌情况。

考虑到日交易数据的季节性变动差异和不稳定性，我们选用各品种每年日交易数据平均值作为该交易品种该年的样本。同时，选取结算价，开盘价，最高价，最低价，收盘价，结算价，成交量，成交金额，持仓量，涨跌幅1（（收盘价-前结算）/前结算）以及涨跌2（（结算价-前结算）/前结算）作为原始变量值指标。

2013年有交易的40个品种。数据初期处理使用Excel，因子分析使用SPSS实现。

**2.2 实证分析**

**2.2.1 因子分析简介**

因子分析通过研究相关阵或协方差阵的内部依赖关系，在处理具有多个指标的问题中用来寻找出支配多个指标的少数几个公因子或者是共性因子，这些公因子具有一个性质，就是他们是彼此独立或者是不相关的。在研究问题时，以公因子代替原指标作为研究对象，可以达到到在不损失或很少损失原指标所包含信息的效果。

因子分析的理论基础非常成熟，应用也十分广泛，其理论知识附在文末，下面简单介绍一下因子分析的步骤。

（1）根据将要解决的问题，选定要分析的原始变量，搜集相关数据并处理，做因子分析适用性检验。

（2）提取公因子。首先要确定因子求解的方法（这里采用主成分方法）及因子的个数，因子个数根据累积方差贡献率确定，一般认为要达到60%以上才符合标准。

（3）因子旋转。因子分析不仅仅要求出公共因子，更重要的是能够解释公共因子的实际意义。如果每个公共因子的意义含糊不清，则不利于根据实际背景对因子进行解释。由于因子载荷矩阵不唯一，所以应该对因子载荷矩阵实施旋转变换，使得各个因子载荷矩阵的每一列个元素的平方按列向0和1两级转化，以达到简化结构的目的。

（4）计算因子得分。有时候我们需要把公因子表示成指标变量的线性组合，或是对每一个样本计算公因子的估计值，这就是因子得分。因子得分有两方面的用途：一是用来诊断模型；而是用来作为进一步分析的原始数据。因子得分函数可以表示成如下形式：



**2.2.2 结果分析**

**(1).因子分析适用性检验**

因子分析的前提是观测变量之间具有较强的相关性，通常因子分析使用性检验采用KMO and Bartlett 检验。

KMO检验，通过比较各变量间简单相关系数和偏相关系数的大小判断变量间的相关性，相关性强时，偏相关系数远小于简单相关系数，KMO值接近1。kmo值大于0.5可以做因子分析，kmo值越大，因子的贡献率也就越高，效果越好。

Bartlett’s球型检验，用于检验相关阵是否是单位阵，即各变量是否独立。它是以变量的相关系数矩阵为出发点，零假设：相关系数矩阵是一个单位阵。如果巴特利球形检验的统计计量数值较大，且对应的相伴概率值小于用户给定的显著性水平，则应该拒绝零假设，可做因子分析。

检验结果如下表：

| **KMO and Bartlett's Test** | | |
| --- | --- | --- |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .685 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 6215.807 |
| df | 55 |
| Sig. | .000 |

KMO值为0.685，根据KMO度量标准知原变量适合做因子分析；接下来我们再看巴特利特球度检验，Sig值小于0.05，此时应该拒绝变量相互独立的假设，这说明因素分析是合适的。综上所述，因子分析适用于本文样本数据。

**(2).提取公因子**

提取公因子：采用主成分法，使得累积方差贡献率大于85%，即根据，确定因子个数同时考虑特征值大于或接近于1。



原始方差贡献率以及旋转后的方差贡献率如下表所示：

| **Total Variance Explained** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 6.026 | 54.780 | 54.780 | 6.026 | 54.780 | 54.780 | 5.973 | 54.300 | 54.300 |
| 2 | 2.299 | 20.902 | 75.682 | 2.299 | 20.902 | 75.682 | 2.184 | 19.851 | 74.151 |
| 3 | 1.067 | 9.701 | 85.383 | 1.067 | 9.701 | 85.383 | 1.005 | 9.133 | 83.284 |
| 4 | .750 | 6.821 | 92.204 | .750 | 6.821 | 92.204 | .981 | 8.920 | 92.204 |
| 5 | .650 | 5.907 | 98.111 |  |  |  |  |  |  |
| 6 | .182 | 1.656 | 99.767 |  |  |  |  |  |  |
| 7 | .023 | .209 | 99.977 |  |  |  |  |  |  |
| 8 | .003 | .023 | 100.000 |  |  |  |  |  |  |
| 9 | 6.771E-6 | 6.156E-5 | 100.000 |  |  |  |  |  |  |
| 10 | 5.220E-7 | 4.746E-6 | 100.000 |  |  |  |  |  |  |
| 11 | 9.417E-8 | 8.561E-7 | 100.000 |  |  |  |  |  |  |
| Extraction Method: Principal Component Analysis. | | | | | | | | | |

根据上表，考虑到使得累积方差贡献率大于85%，同时考虑特征值大于或接近于1，我们选取4个主因子。此时，累计方差贡献率为92.204%，可见保留了大部分原始信息，效果良好。

**(3).因子解释**

正交旋转后的因子载荷矩阵如下表所示：

| **Rotated Component Matrixa** | | | | |
| --- | --- | --- | --- | --- |
|  | Component | | | |
| f1 | f2 | f3 | f4 |
| VAR00001 | .996 | -.030 | -.026 | .022 |
| VAR00002 | .996 | -.008 | .009 | .070 |
| VAR00003 | .996 | -.008 | .009 | .070 |
| VAR00004 | .995 | -.008 | .009 | .070 |
| VAR00005 | .991 | -.013 | .009 | .109 |
| VAR00006 | .996 | -.028 | .021 | .026 |
| VAR00007 | .043 | .892 | .046 | .073 |
| VAR00008 | -.096 | .702 | -.037 | .074 |
| VAR00009 | .019 | .925 | .047 | .063 |
| VAR00010 | .151 | .191 | .093 | .965 |
| VAR00011 | .004 | .033 | .994 | .085 |
| Extraction Method: Principal Component Analysis.  Rotation Method: Varimax with Kaiser Normalization. | | | | |
| a. Rotation converged in 4 iterations. | | | | |

由上表可以看出：

（1）第一主因子f1描述第一个变量到第六个变量（前结算价，开盘价，最高价，最低价，收盘价，结算价），衡量的是期货品种日交易的基本数据，反映市场资金对该产品的关注度或者说是该产品吸引资金的能力，体现一个期货品种的基本性能和整体特点，因此可以命名为基能因子。

（2）第二主因子f2描述第7、8、9个变量（成交量，成交金额，持仓量），成交量反映了市场的投资热情及投资者对该产品的熟悉程度。成交额与其具有相同意义。而持仓量的大小则反映了市场交易规模的大小。因此第二主因子可以命名为活跃度因子

（3）第三主因子f3描述第10个变量：涨跌幅1，反映该品种期货的波动情况，命名为主波动情况因子。

（4）第四主因子f4描述第11个变量：涨跌幅2，亦反映该品种期货的波动情况，命名为副波动情况因子。

**(4).得分分析**

利用回归法估计因子得分系数，根据各因子得分函数，计算出的因子得分。再以各个旋转后因子的方差贡献率占四个因子总方差贡献率的比重作为权数，就得到了各期货品种的综合得分函数：f=(54.3f1+19.851f2+9.133f3+8.92f4)/92.204

根据得分分析和综合得分函数对2013年在交易的40个商品期货品种进行排序：



由上表可以看出，综合得分排名靠前的期货品种，大都在第一主因子——基能因子上得分排名较为靠前，同时在第二主因子——活跃度因子得分排名居中前方，在主波动因子（f3）和副波动因子(f4)上得分排名居中或者稍靠后。这表明，在期货市场表现良好的期货品种，本身具有优良的基础和资金吸引能力，活跃度良好，具有较好的波动性但波动幅度不会过大以致风险过高。

我们再看综合得分排名靠后的几个品种，焦煤、铁矿石、黄金、纤维板和胶合板。这几个品种的基能因子和活跃度因子得分排名居中，但是主波动因子排名非常靠前，导致风险很高。

综上，我们可以得到如下结论：得分排名靠前的品种在基本性能，活跃度和受欢迎度以及波动情况上综合表现良好，一般说来这样的期货品种比较活跃，波动情况足够良好存在获利机会的同时又不会存在过大风险，我们应该选取排名靠前的品种进行交易。

综合得分按板块排序如下表：



**（5）结论**

得分排名靠前的品种在基本性能，活跃度和受欢迎度以及波动情况上综合表现良好，一般说来这样的期货品种比较活跃，波动情况足够良好存在获利机会的同时又不会存在过大风险，我们应该选取排名靠前的品种进行交易。

总的来说，农产品类和金属类期货品种在日交易基本情况、活跃度和波动情况上综合表现比较突出，排名相对靠前。其中铜、豆粕、天然橡胶、棉花和锌排名前五；同时排在后五位的是早籼稻2、甲醇、动力煤、普麦和燃料油。因此，如若不考虑板块区别，我们推荐优先选择铜、豆粕、天然橡胶和锌作为交易品种。

分板块来看，农产品类期货中排名比较靠前的是豆粕、棉花和豆油；金属类期货中排在前面的是铜、锌和铝；化工类产品中排在前面的是天然橡胶、聚乙烯和精对苯二甲酸（PTA）；能源类期货中焦炭、沥青和焦煤排在前面。因此，在考虑分版块投资组合时，建议优先考虑在豆粕、棉花和豆油，铜、锌和铝，天然橡胶聚乙烯和精对苯二甲酸（PTA），焦炭、沥青和焦煤中进行投资组合品种选择。

**2.3 构建投资组合**

**2.3.1 资产组合策略**

如同公司的目标杠杆比率一样，构建资产组合得到结果只是目标资产组合。目标组合并不是说在动态的交易过程中严格按照这一比例交易，而是在满足动态策略的前提下，尽可能的向这一目标靠拢。按照资产组合的资产配置比例开始建仓在，以后的动态交易过程中可以改变。或者，在交易了一段较长的时间后（比如：一年），根据新的数据生成新的组合策略也是很有必要且合理的。

依据上述因子分析的结果，下一步我们采用马克维茨的投资组合理论构建最初的期货产品投资组合。依据商品期货市场上现存的农产品、金属、能源和化工四大板块。在因子模型已经给出每种期货产品综合得分的情况下，在4个不同板块中，我们选择每一板块中得分最高的期货产品作为投资组合备选品种。根据上述结果，我们选择铜、燃料油、豆粕、天然橡胶。当然，选择的个数并非局限为每一板块只选一个。一种更好的办法是：在交易量较大的市场中（固然因子分析得分亦高）可以多选择几个品种。而且，从经验上来看，能源与化工的产品联系比较紧密，其价格相关性应该较强。比如：石油这种重要的能源产品能够同时用来生产多种化工产品。所以，能源化工甚至可以按照一个品种来选取期货配置。

对于每一个期货品种，我们选取该品种当日的主力合约价格作为该品种的交易价格，由此价格计算每天的收益率。其中主力合约是指该期货品种每一天的交易中成交量最大的合约。每天收益率使用相邻两个主力合约价格比值的对数。

依据我们选出的交易价格数据，计算出铜、焦炭、豆粕和天然橡胶的平均收益率（%）：-0.046420605、-0.09795625、0.015307879和-0.157531152。相应的协方差矩阵（受收益率影响，真实值放大为一万倍）：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 铜 | 焦炭 | 豆粕 | 天然橡胶 |
| 铜 | 1.226525754 | 0.694080957 | 0.364526829 | 1.281596882 |
| 焦炭 | 0.694080957 | 2.264669418 | 0.447730295 | 1.437533761 |
| 豆粕 | 0.364526829 | 0.447730295 | 3.603521497 | 0.734093364 |
| 天然橡胶 | 1.281596882 | 1.437533761 | 0.734093364 | 2.947718993 |

简单从协方差矩阵来看，对角线元素为方差，而两种商品的协方差比方差小。说明可以通过分散化组合降低风险。

我们要求的一种组合方式，使得每天的收益率最大，而总方差最小。也就是说，单位总风险对应的收益率最大。经过二次最优化运算，我们得到如下结果：

组合比例：W=[1.5,0.1,0.5,-1.1].

日平均收益率：R=0.1015.

日平均收益率方差：2.6994680851

所以，在一个没有明显的交易机会出现的市场中，应该持有1.5单位的铜，0.1单位的焦炭，0.5单位的豆粕，卖空1.1单位的天然橡胶，来构成我们的初始仓位。

**2.3.2 结论**

考虑投资组合策略时，建议初始仓位为有1.5单位的铜，0.1单位的焦炭，0.5单位的豆粕，卖空1.1单位的天然橡胶。在后续交易中，根据市场实际情况进行调整，但一般说来各品种持仓量保持在这个水平上下较小幅度波动。

**3 交易策略选择（趋势交易策略）**

**3.1 趋势交易策略简介**

金融价格波动可以分为两大类：趋势行情与振荡行情，相应地交易方式也可以分为趋势交易与振荡交易两类。对于期货投资而言，由于具有保证金杠杆机制与双向交易特性，每年的趋势投资机会非常多，如果能很好地把握一些品种的趋势性行情机会，投资收益是非常可观的。

从实际交易的情况来看，虽然趋势性行情的机会众多，但大多数投资者并不能很好地抓住这些机遇，主要原因在于：对趋势的定义和特性没有清晰的概念；没有简洁明了的判断方法；主观因素影响理性判断，建仓时机把握不好；加仓方式不是很科学；对各品种特性没有仔细分析，选择了一些并不是很适合做趋势交易的期货品种。

针对这些客观存在的问题，接下来这一部分旨在根据历史交易数据，判断固定窗口期内期货价格走势，识别趋势行情，并分析相关宏观经济指标与期货价格走势的关系，以实现利用宏观经济指标预测未来期货行情。在此基础上，调整投资组合中不同期货品种的权重，以达到整体投资收益最大的目的。

**3.2趋势识别方法**

**3.2.1.构建收益率指标（固定窗口期）**

固定一个月为窗口期（大约20个交易日），以日数据计算一个月内的收益率：

以为指标来判断这段时期内的期货价格走势是震荡还是趋势。

判断方法：

1. ,则判断为趋势。若则判断为上升趋势。否则为下跌趋势。

（2）则判断为震荡行情。

其中 >0为上海银行间同业拆放月利率。

**3.2.2斜率k值指标（固定窗口期）**：

固定一个月为窗口期（大约20个交易日），以5分钟数据计算。

步骤：

（1）对一个月内的5分钟数据进行平滑处理（平滑曲线用局部加权回归散点平滑法）；

（2）找出平滑后的数据的极大值点和极小值点；

（3）对极大值和极小值分别进行线性拟合，拟合后的斜率为；

（4）对进行加权平均得到。权重为线性拟合的拟合度。

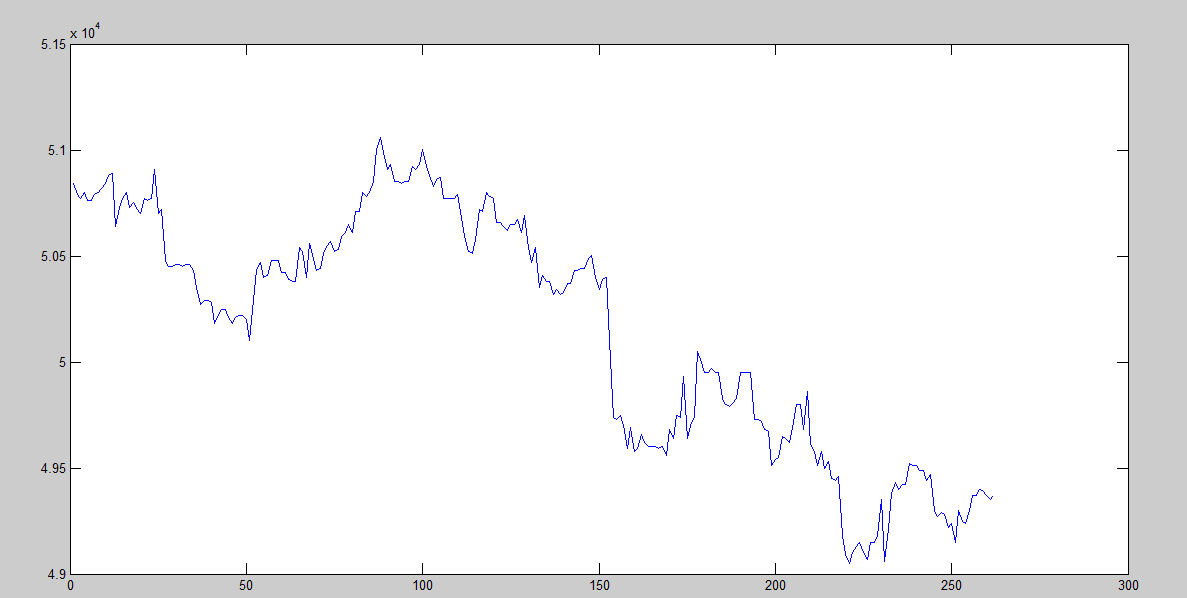
判断方法：

1. 若则判断为趋势。若则为上升趋势，否则为下跌。
2. 若，则判断为震荡行情。

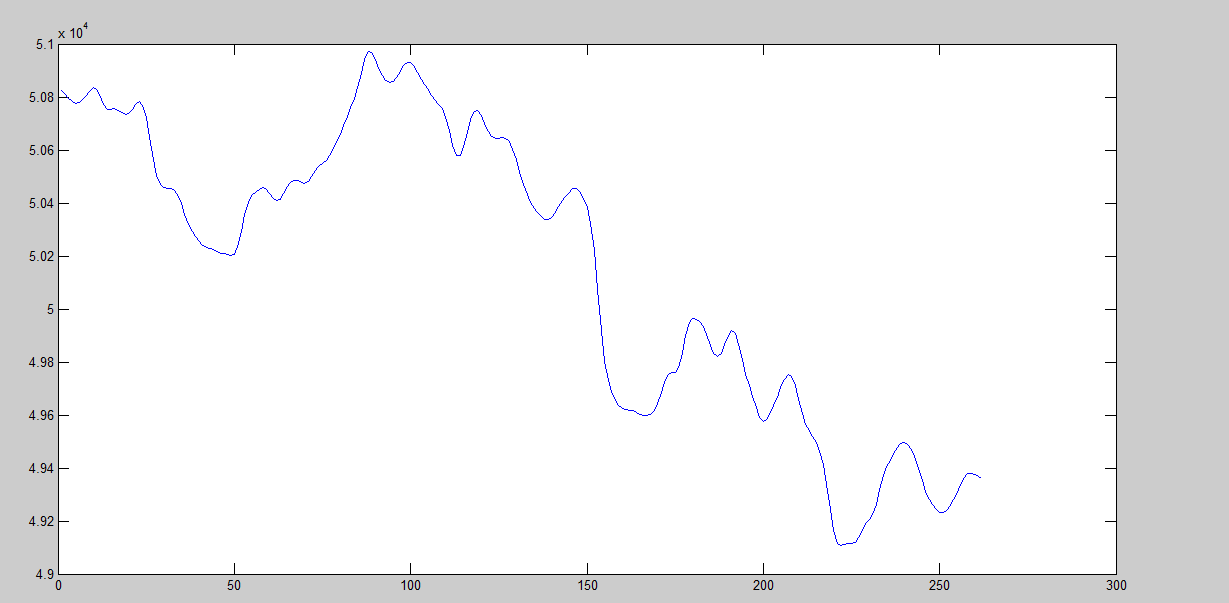
运用MATLAB实现，选取几个运行结果做展示：

2014年2月份5分钟收盘价：(下跌趋势)

平滑前：



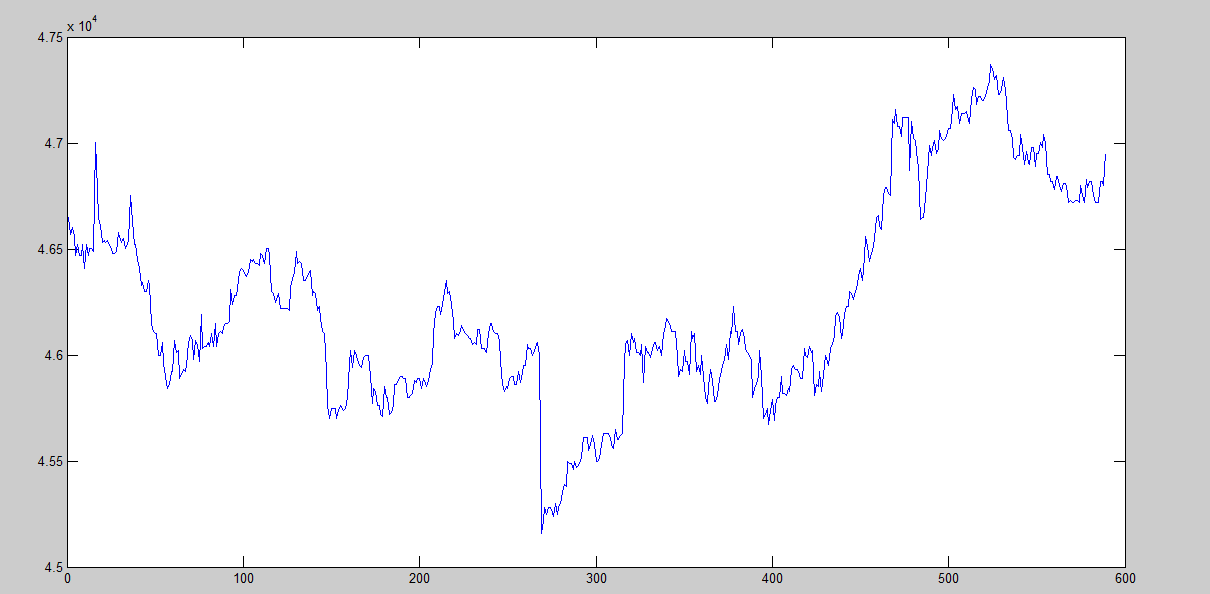
平滑后：



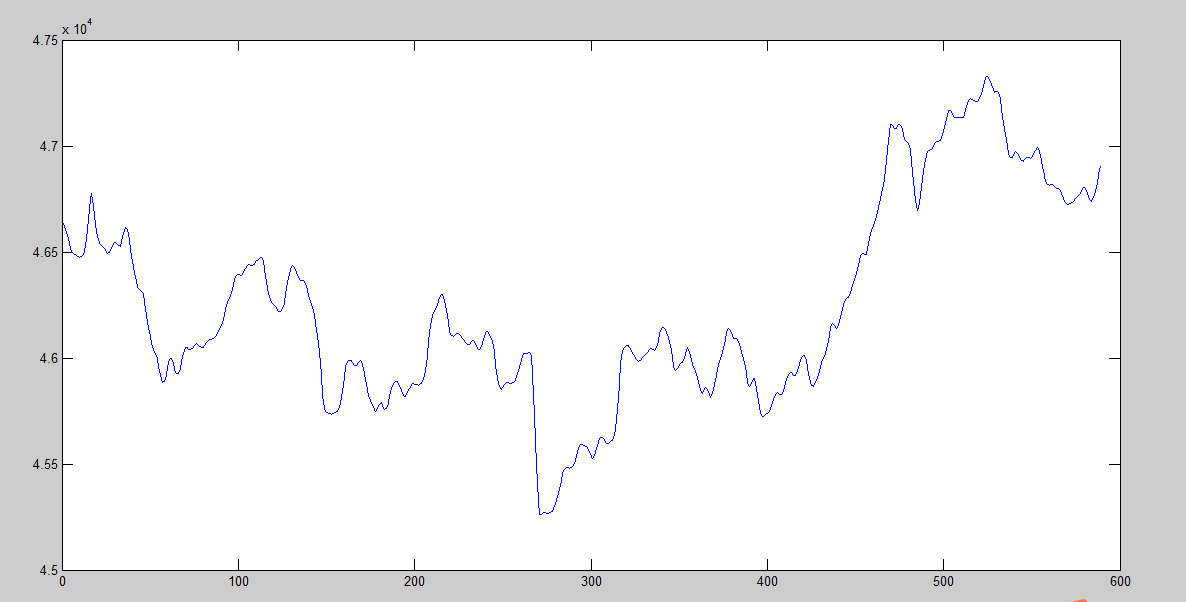
计算得K=-6.2046，由判别方法判定为下降趋势，真实看图也是如此。

2014年4月份5分钟收盘价：（震荡）

平滑前：



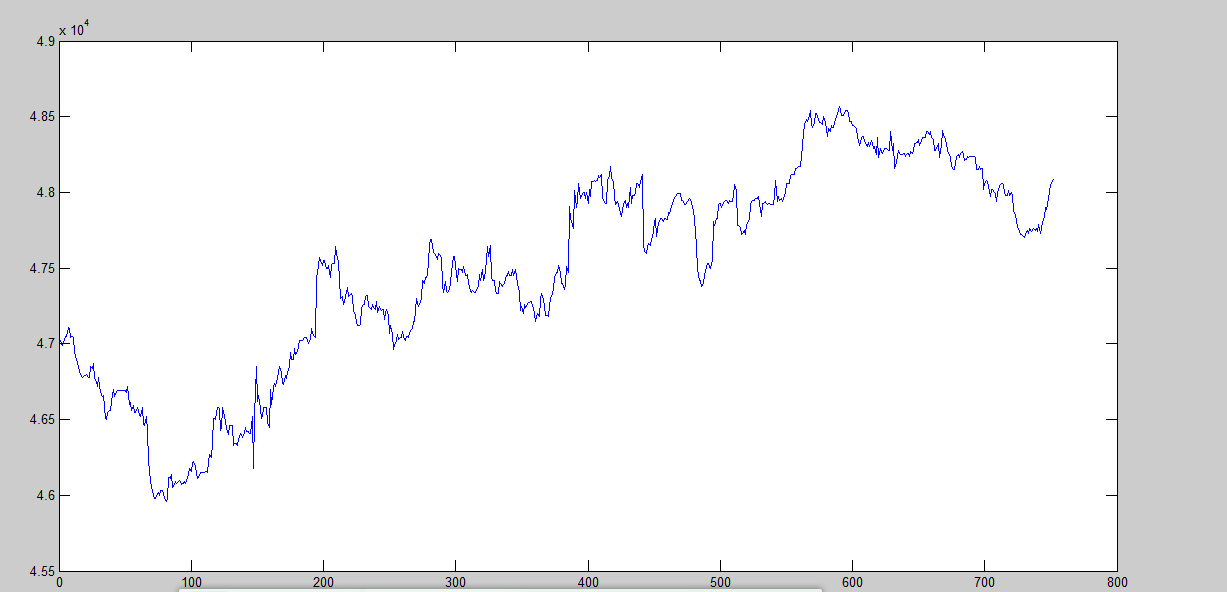
平滑后：



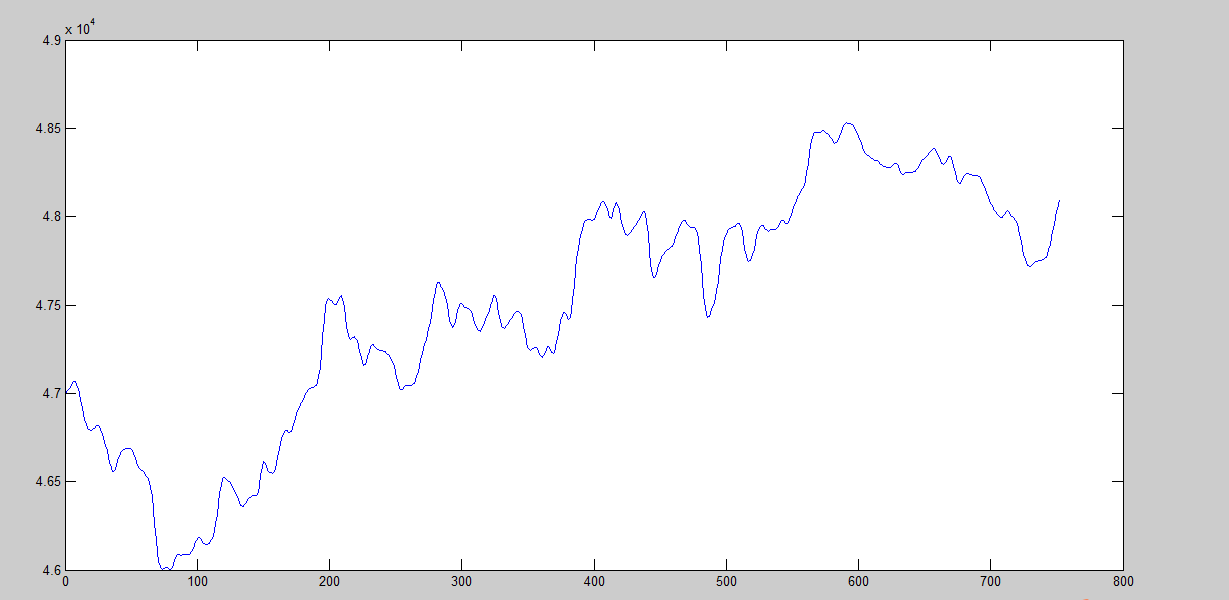
计算得k=1.4772，由判别方式判别为震荡，直观判断也是如此。

2014年5月份5分钟收盘价：（上升趋势）

平滑前：



平滑后：



计算得K=3.1163，按判别方法判为上升趋势，只管也是如此。

考虑到我们能够获取到的数据范围有限，对于第二个方法，五分钟数据只有2013年11月至今，样本量太小，无法实现分析，而如果选用日线数据，误差很大，所以，对于固定窗口期我们目前选用第一种方法来进行趋势识别，如果能解决数据量不足的问题，可以考虑两种联合识别。

**3.3影响期货价格走势的潜在宏观经济指标**

宏观经济运行状况对金融证券期货市场的价格行为有着重要的影响。在多因素资产定价模型中，任何影响未来投资机会或消费水平的经济变量都应予以体现。宏观经济发展状况对投资者判断产品需求格局、未来现金流和调整后的风险贴现率有直接或间接的影响。因此，宏观经济信息对金融市场的价格行为应有直接的作用。国民生产总值、国内生产总值、国民收入、消费品价格指数、生产品价格指数、航运指数、货运指数、采购经理人指数、贸易余额、失业率，以及货币供给、利率和新房开工率等指标，从不同方面反映了一国宏观经济的运行情况，因而对金融市场的价格行为会产生不同的影响。

宏观经济形势对期货价格的影响主要通过以下几个渠道表现出来：首先，当GDP 增速加快时，投资者对后期的经济形势充满信心，投资者就会以多头持仓为主；反之，投资者对后市普遍看淡的时候，则会选择平仓或卖空，从而导致市场抛压加大，使得期货价格走低。其次，当经济增速开始进入回升期时，投资增速也会加快、居民消费额就会不断上升、外贸出口也会增加，在它们带动下，商品需求量会不断扩大、现货供应压力会逐步减轻，期货价格会稳步上扬；反之，期货价格则可能步步走低。第三，随着宏观经济形势变化，国家的宏观调控政策也会随之调整。当经济增速加速初期，央行一般采取宽松的货币政策，充足的流动性会减轻现货压力，从而对期货价格产生利多影响。而当经济增速过快、CPI 失控、经济出现过热苗头时，央行则会采取紧缩的货币政策，随着市场上流动性减少，现货压力会逐步加大，期货价格就会冲高回落。

为寻找影响期货价格走势的潜在宏观指标，这里我们选取反映货币供应、通货膨胀、国内和国际经济景气程度的诸多宏观经济指标（注：一些前期选取的指标中，相关性表现不好的指标没有将结果展示。），用南华商品指数（NH）表示大宗商品价格水平，并以月度收盘价代表指数的变化情况。

**3.2.1相关性分析**

相关性分析是指对两个或多个具备相关性的变量元素进行分析，从而衡量两个变量因素的相关密切程度。相关性的元素之间需要存在一定的联系或者概率才可以进行相关性分析。相关性不等于因果性，也不是简单的个性化。这里采用简单的相关分析，通过计算Pearson相关系数r来衡量宏观经济指标与南华商品价格指数的相关度。

Pearson相关系数：



对于r的解释如下（即r的特点）

1. r的取值正负取决于分子协方差；
2. R的绝对值在0和1之间；
3. R的绝对值大小可以说明现象之间相关关系的紧密度。

一般标准如下：



宏观指标数据选取2005年1月至2014年10月的月度数据，南华商品期货指数选取月度收盘价（数据来源：万德数据库）。利用SPSS软件实现相关分析，结果如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 相关系数 | | | | | | | | | | | | | |
|  | | | cpi | hjj | dollar | m2 | ppi | rpi | ep | tc | c | NH | |
|  | cpi | 相关 | 1.000 | -.001 | -.497 | -.427 | .692 | .979 | .489 | .593 | .291 | .532 | |
| 显著性 | . | .990 | .000 | .000 | .000 | .000 | .000 | .000 | .002 | .000 | |
| df | 0 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | |
| hjj | 相关 | -.001 | 1.000 | .234 | .651 | .231 | -.004 | .451 | .142 | .465 | .339 | |
| 显著性 | .990 | . | .012 | .000 | .013 | .964 | .000 | .132 | .000 | .000 | |
| df | 112 | 0 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | |
| dolla | 相关 | -.497 | .234 | 1.000 | .266 | .019 | -.476 | .019 | -.264 | .491 | -.607 | |
| 显著性 | .000 | .012 | . | .004 | .842 | .000 | .843 | .005 | .000 | .000 | |
| df | 112 | 112 | 0 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | |
| m2 | 相关 | -.427 | .651 | .266 | 1.000 | -.258 | -.374 | -.122 | .170 | -.002 | .040 | |
| 显著性 | .000 | .000 | .004 | . | .006 | .000 | .198 | .071 | .980 | .669 | |
| df | 112 | 112 | 112 | 0 | 112 | 112 | 112 | 112 | 112 | 112 | |
| ppi | 相关 | .692 | .231 | .019 | -.258 | 1.000 | .735 | .803 | .514 | .556 | .304 | |
| 显著性 | .000 | .013 | .842 | .006 | . | .000 | .000 | .000 | .000 | .001 | |
| df | 112 | 112 | 112 | 112 | 0 | 112 | 112 | 112 | 112 | 112 | |
| rpi | 相关 | .979 | -.004 | -.476 | -.374 | .735 | 1.000 | .499 | .680 | .286 | .519 | |
| 显著性 | .000 | .964 | .000 | .000 | .000 | . | .000 | .000 | .002 | .000 | |
| df | 112 | 112 | 112 | 112 | 112 | 0 | 112 | 112 | 112 | 112 | |
| ep | 相关 | .489 | .451 | .019 | -.122 | .803 | .499 | 1.000 | .306 | .527 | .405 | |
| 显著性 | .000 | .000 | .843 | .198 | .000 | .000 | . | .001 | .000 | .000 | |
| df | 112 | 112 | 112 | 112 | 112 | 112 | 0 | 112 | 112 | 112 | |
| tc | 相关 | .593 | .142 | -.264 | .170 | .514 | .680 | .306 | 1.000 | .184 | .319 | |
| 显著性 | .000 | .132 | .005 | .071 | .000 | .000 | .001 | . | .050 | .001 | |
| df | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 0 | 112 | 112 | |
| c | 相关 | .291 | .465 | .491 | -.002 | .556 | .286 | .527 | .184 | 1.000 | -.060 | |
| 显著性 | .002 | .000 | .000 | .980 | .000 | .002 | .000 | .050 | . | .529 | |
| df | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 0 | 112 | |
| NH | 相关 | .532 | .339 | -.607 | .040 | .304 | .519 | .405 | .319 | -.060 | 1.000 | |
| 显著性 | .000 | .000 | .000 | .669 | .001 | .000 | .000 | .001 | .529 | . | |
| df | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 112 | 0 | |

其中：

MY：实际美元指数：广义；

EP：指进出口额；

TC：指社会消费品零售总额；

HG：指宏观经济景气指数：先行指数；

PPI：工业生产者出厂价格指数

CPI：居民消费价格指数

RPI：商品零售价格指数

**结果分析**：由上表知南华期货价格指数与CPI,MY,RPI显著相关，与HJJ，PPI，EP低度相关，与其他指标不相关。

**3.2.2格兰杰因果关系检验**

在[时间序列](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=693050)情形下，两个经济变量X、Y之间的[格兰杰](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=2709819)因果关系定义为：若在包含了变量X、Y的过去信息的条件下，对变量Y的预测效果要优于只单独由Y的过去信息对Y进行的预测效果，即变量X有助于解释变量Y的将来变化，则认为变量X是引致变量Y的[格兰杰](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=2709819)原因。

进行格兰杰因果关系检验的一个前提条件是[时间序列](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=693050)必须具有平稳性，否则可能会出现虚假回归问题。因此在进行格兰杰因果关系检验之前首先应对各指标[时间序列](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=693050)的平稳性进行[单位根检验](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=53386008)(unit root test)。常用增广的迪基—富勒检验(ADF检验)来分别对各指标序列的平稳性进行[单位根检验](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=53386008)。

运用eviews进行实证分析（数据来源：万德数据库）。

1.单位根检验：

从运行结果看：若ADF值小于level(1),level(2),level(3).的三个水平条件下的值，则以该序列为一个平稳序列，反之，如果ADF值大于三个水平给定值，那么要对该序列的差方项作ADF检验，分别赋予不同滞后周期，直到ADF值小于三个水平的值为止。

CPI：一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(CPI) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 11 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -4.674597 | 0.0014 |
| Test critical values: | 1% level |  | -4.051450 |  |
|  | 5% level |  | -3.454919 |  |
|  | 10% level |  | -3.153171 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

NH ：一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(NH) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -7.514583 | 0.0000 |
| Test critical values: | 1% level |  | -4.042042 |  |
|  | 5% level |  | -3.450436 |  |
|  | 10% level |  | -3.150549 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

EP：一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(EP) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -16.25035 | 0.0000 |
| Test critical values: | 1% level |  | -4.042042 |  |
|  | 5% level |  | -3.450436 |  |
|  | 10% level |  | -3.150549 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

HG：一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(HG) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -4.938762 | 0.0005 |
| Test critical values: | 1% level |  | -4.042042 |  |
|  | 5% level |  | -3.450436 |  |
|  | 10% level |  | -3.150549 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

MY :一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(MY) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 1 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -7.246173 | 0.0000 |
| Test critical values: | 1% level |  | -4.042819 |  |
|  | 5% level |  | -3.450807 |  |
|  | 10% level |  | -3.150766 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

PPI：一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(PPI) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 1 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -4.445341 | 0.0028 |
| Test critical values: | 1% level |  | -4.042819 |  |
|  | 5% level |  | -3.450807 |  |
|  | 10% level |  | -3.150766 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

RPI:一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(RPI) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 11 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -5.121226 | 0.0003 |
| Test critical values: | 1% level |  | -4.051450 |  |
|  | 5% level |  | -3.454919 |  |
|  | 10% level |  | -3.153171 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

TC:一阶平稳

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: D(TC) has a unit root | | | |  |
| Exogenous: Constant, Linear Trend | | | |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=12) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -19.01156 | 0.0000 |
| Test critical values: | 1% level |  | -4.042042 |  |
|  | 5% level |  | -3.450436 |  |
|  | 10% level |  | -3.150549 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

由以上结果可以看出，宏观指标变量和南华商品期货指数时间序列均为一阶平稳序列，下面对这些序列做格兰杰因果检验。

2.格兰杰检验

在0.05的置信区间下，P值小于0.05拒绝原假设，即有格兰杰因果关系。

CPI and NH:

|  |  |  |  |
| --- | --- | --- | --- |
| Pairwise Granger Causality Tests | | | |
| Date: 12/17/14 Time: 11:41 | | | |
| Sample: 2005M01 2014M06 | | | |
| Lags: 1 | |  |  |
|  |  |  |  |
|  |  |  |  |
| Null Hypothesis: | Obs | F-Statistic | Prob. |
|  |  |  |  |
|  |  |  |  |
| NH does not Granger Cause CPI | 113 | 15.0829 | 0.0002 |
| CPI does not Granger Cause NH | | 4.34786 | 0.0394 |
|  |  |  |  |
|  |  |  |  |

EP and NH

|  |  |  |  |
| --- | --- | --- | --- |
| Pairwise Granger Causality Tests | | | |
| Date: 12/17/14 Time: 11:42 | | | |
| Sample: 2005M01 2014M06 | | | |
| Lags: 1 | |  |  |
|  |  |  |  |
|  |  |  |  |
| Null Hypothesis: | Obs | F-Statistic | Prob. |
|  |  |  |  |
|  |  |  |  |
| NH does not Granger Cause EP | 113 | 8.72988 | 0.0038 |
| EP does not Granger Cause NH | | 0.90901 | 0.3425 |
|  |  |  |  |
|  |  |  |  |

HG and NH：

|  |  |  |  |
| --- | --- | --- | --- |
| Pairwise Granger Causality Tests | | | |
| Date: 12/17/14 Time: 11:43 | | | |
| Sample: 2005M01 2014M06 | | | |
| Lags: 1 | |  |  |
|  |  |  |  |
|  |  |  |  |
| Null Hypothesis: | Obs | F-Statistic | Prob. |
|  |  |  |  |
|  |  |  |  |
| NH does not Granger Cause HG | 113 | 9.81904 | 0.0022 |
| HG does not Granger Cause NH | | 6.20581 | 0.0142 |
|  |  |  |  |
|  |  |  |  |

MY and NH：

|  |  |  |  |
| --- | --- | --- | --- |
| Pairwise Granger Causality Tests | | | |
| Date: 12/17/14 Time: 11:44 | | | |
| Sample: 2005M01 2014M06 | | | |
| Lags: 1 | |  |  |
|  |  |  |  |
|  |  |  |  |
| Null Hypothesis: | Obs | F-Statistic | Prob. |
|  |  |  |  |
|  |  |  |  |
| NH does not Granger Cause MY | 113 | 15.3661 | 0.0002 |
| MY does not Granger Cause NH | | 3.84170 | 0.0525 |
|  |  |  |  |
|  |  |  |  |

PPI and NH：

|  |  |  |  |
| --- | --- | --- | --- |
| Pairwise Granger Causality Tests | | | |
| Date: 12/17/14 Time: 11:45 | | | |
| Sample: 2005M01 2014M06 | | | |
| Lags: 1 | |  |  |
|  |  |  |  |
|  |  |  |  |
| Null Hypothesis: | Obs | F-Statistic | Prob. |
|  |  |  |  |
|  |  |  |  |
| PPI does not Granger Cause NH | 113 | 5.07380 | 0.0263 |
| NH does not Granger Cause PPI | | 38.7581 | 9.E-09 |
|  |  |  |  |
|  |  |  |  |

RPI and NH：

|  |  |  |  |
| --- | --- | --- | --- |
| Pairwise Granger Causality Tests | | | |
| Date: 12/17/14 Time: 11:45 | | | |
| Sample: 2005M01 2014M06 | | | |
| Lags: 1 | |  |  |
|  |  |  |  |
|  |  |  |  |
| Null Hypothesis: | Obs | F-Statistic | Prob. |
|  |  |  |  |
|  |  |  |  |
| RPI does not Granger Cause NH | 113 | 6.14158 | 0.0147 |
| NH does not Granger Cause RPI | | 23.6976 | 4.E-06 |
|  |  |  |  |
|  |  |  |  |

TC and NH：

|  |  |  |  |
| --- | --- | --- | --- |
| Pairwise Granger Causality Tests | | | |
| Date: 12/17/14 Time: 11:46 | | | |
| Sample: 2005M01 2014M06 | | | |
| Lags: 1 | |  |  |
|  |  |  |  |
|  |  |  |  |
| Null Hypothesis: | Obs | F-Statistic | Prob. |
|  |  |  |  |
|  |  |  |  |
| NH does not Granger Cause TC | 113 | 2.82735 | 0.0955 |
| TC does not Granger Cause NH | | 2.58315 | 0.1109 |
|  |  |  |  |
|  |  |  |  |

**结果分析**：由以上结果，可以得出如下结论，居民消费价格指数（CPI）、宏观经济景气指数：先行指数（HG）、工业生产者出厂价格指数（PPI）、商品零售价格指数（RPI）和南华商品价格指数有显著的格兰杰因果关系。

结合相关性分析和格兰杰因果检验，下面重点研究居民消费价格指数（CPI）、宏观经济景气指数：先行指数（HG）、工业生产者出厂价格指数（PPI）、商品零售价格指数（RPI）和广义美元指数（MY）与期货价格行情的关系，再根据不同期货品种，增加相应的宏观指标，挖掘出是否这些宏观指标能预测走势震荡或者趋势。

**3.3预测期货行情**

**3.3.1 逻辑回归模型**

逻辑模型（Logit model）是[离散选择法](http://baike.baidu.com/view/99578.htm)模型之一，属于[多重变量分析](http://baike.baidu.com/view/793754.htm)范畴，

逻辑分布（Logistic distribution）[公式](http://baike.baidu.com/view/645857.htm)P(Y=1│X=x)=exp(xβ)/(1+exp(xβ))

其中参数β常用[极大似然估计](http://baike.baidu.com/view/185250.htm)。

建立宏观经济指标和期货行情（趋势=1，震荡=0）之间的逻辑回归模型。以铜一月份期货合约为例，初步分析结果如下：

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.97277 11.30959 0.086 0.931

e 0.02179 0.11433 0.191 0.849

m -0.01273 0.03662 -0.348 0.728

ppi 0.06396 0.08329 0.768 0.443

coal -0.02267 0.01705 -1.330 0.184

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 162.73 on 117 degrees of freedom

Residual deviance: 160.26 on 113 degrees of freedom

(1 observation deleted due to missingness)

AIC: 170.26

Number of Fisher Scoring iterations: 4

其中e为宏观经济景气指数，m为美元指数，ppi工业生产者出厂价格指数，coal为原煤价格指数:当月同比。分析结果，可以看出回归效果并不好（p值显著大于置信度0.05，不能拒绝原假设，即回归系数为0）。原因：自变量间存在共线性。

接下来的计划：

1. 加入其他宏观指标，消除自变量之间的共线性，观察回归效果；
2. 考虑到实际应用中的情况，尝试对每个月的主力合约做分析。

**3.3.2 AR模型**

AR模型，即自回归（Auto Regressive, AR）模型，又称为时间序列模型，目标是通过分析要素(变量)随时间变化的历史过程, 揭示其变化发展规律, 并对未来状态进行分析预测。数学表达式为：

其中，e(t)为[均值](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=399288&ss_c=ssc.citiao.link)是0，[方差](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=151934&ss_c=ssc.citiao.link)为某值的白噪声序列。

白噪声序列：[随机变量](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=116078)X（t）t=1，2，3…，如果是由一个不相关的随机变量的序列构成的，即对于所有S不等于T，随机变量Xt和Xs的[协方差](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=392413)均为零，则称其为纯[随机过程](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=180092)。对于一个纯随机过程来说，若其期望和[方差](http://baike.sogou.com/lemma/ShowInnerLink.htm?lemmaId=151934)均为常数，则称之为白噪声过程。白噪声过程的样本实称成为白噪声序列，简称白噪声。之所以称为白噪声，是因为他和白光的特性类似，白光的光谱在各个频率上有相同的强度，白噪声的谱密度在各个频率上的值相同。

模型建立：以收益率为判断指标，建立收益率和宏观因素的模型。

传统的基于收益率AR模型：



在此基础上，加入宏观指标为外生变量的AR模型：



其中为宏观指标变量以及和宏观指标有关的虚拟变量。

下一步展望：用数据进行实证分析。估计出模型的系数。将预测的收益率和同业拆解率对比，判断趋势和震荡，进一步进行交易策略设计。

**4 交易策略选择（期限结构法）**

**4.1 期货价格期限结构（主成分分析方法，动态分析）**

近年来，资产价格信息含量的研究越来越受到学者们的关注，已成为金融市场重要研究领域之一。其目的是从各种金融资产的市场价格中提取出市场隐含的对各种风险和价格走向的预期，从而为投资者制定合理的投资策略及进行风险管理提供依据。

商品期货因其套期保值和价格发现的功能，经过几十年的发展，已成为资本市场的重要角色，也是企业制定生产决策和投资策略不可忽视的因素之一。与金融市场相比，虽然商品已具有一定的金融属性，但商品市场仍有其自身的特性。例如，大部分商品期货价格都表现出明显的季节性：商品期货通常呈现为反向市场，即近期期货合约的价格比远期期货合约的价格要高，而月反向市场的程度和价格的波动率基本成正比，等等。

**4.1.1 期货价格期限结构简介**

期货价格的期限结构，指某一时点期货价格与不同到期期限的期货合约的关系。同时也是不同到期期限的期货合约跨期价格关系的呈现，描述了不同时点期货价格与现货价格的变动关系，同时也反映出不同到期合约期货价格与现货价格的变动关系。随着期货市场不断完善，许多期货合约到期期限越来越长，因此期限结构的概念也变得越来越重要。国内外涉及期货价格期限结构的研究主要集中在以下几个方面：

其一，构建期限结构模型，并用于未来期货价格走势预测。这些模型考虑影响期货价格的诸多因素，如现货价格、便利收益率、利率、长期价格和季节性等。实证结果均表明：通过单因素模型得到的期限结构曲线与实际观察到的期限结构曲线吻合得较好，具有较好的预测效果。认为引进不对称便利收益能改进模型的性能。实证发现该仿射模型对沪铜是适用的，而且模型中因素数目越多，模型拟合与预测能力越强。

其二，将主成分分析应用到商品期货价格期限结构上，考察期限结构的动态变化，进而挖掘期限结构的信息价值，以用于交易策略设计。应用主成分分析法来研究商品期货价格期限结构，发现前三个主成分能够解释期限结构的动态变化。

本文主要应用主成分分析方法来研究期货价格期限结构，挖掘其隐含信息。

由于不同因素可能由相同的驱动因素影响，所以主成分分析可以被用来减少特定数据

集的维数，把包含的信息浓缩到一个正交的因素中，通常少于现存的数据集。在商品期货价格的研究中，主成分分析允许把期限结构中可得的信息浓缩到相关的因素，然后我们通过主成分分析检验商品期货价格期限结构的信息内容，揭示其隐藏的信息含量。

国外研究表明，期限结构的变动总体的方差绝大部分来自于三个因素的贡献。Litterman

andschein腼an(2991)[40]把这三个风险因素称为“水平因素(Level)”、“斜度因素(slope)”和“凸度因素(Curvature)”。

水平因素对应期货价格相关系数矩阵的最大特征根，反映了平行移动的因素在期货价格期限结构曲线变动中发挥了主导作用，该因素的增加会使所有期货合约近似同等程度的增加，主要是影响现货价格的宏观因素，如经济周期，货币政策等；

斜度因素对应第二特征根，它和期限结构曲线的倾斜程度有关，该因素使得短期期货合约和长期期货合约朝着不同的方向变化，从而将改变期限结构曲线的斜率；

凸度因素对应第三特征根，该因素和期限结构曲线的曲率关系密切，该因素对中期产生影响，与对短期和长期的影响相反，从而改变期限结构曲线的曲率。

**4.1.2 主成分分析简介**

主成分分析是将多个变量通过正交线性变换以选出较少个数重要变量的一种多元统计分析方法，又称主分量分析。具体的来说，主成分分析是设法将原来众多具有一定相关性(比如p个)指标，重新组合成一组新的互相无关的综合指标来代替原来的指标。

通常数学上的处理就是将原来p个指标作线性组合，作为新的综合指标。最经典的做法就是用F1 (选取的第一个线性组合，即第一个综合指标)的方差来表达，即Var(F1)越大，表示F1包含的信息越多。因此在所有的线性组合中选取的F1应该是方差最大的，故称F1为第一主成分。如果第一主成分不足以代表原来p个指标的信息，再考虑选取F2，即选第二个线性组合，为了有效地反映原来信息，F1已有的信息就不需要再出现在F2中，用数学语言表达就是要求，则称F2为第二主成分，依此类推可以构造出第三、第四……，第P个主成分。具体步骤如下：



(1)指标的标准化

设有n个样本，每个样本有P个指标的值，为了排除数量级和量纲不同带来的影响，首先对原始数据作标准化处理，使得每个指标的平均值为0，方差为1，标准化公式为:



(2)计算样本的相关矩阵R



(3)计算相关矩阵R的特征值和对应的标准正交特征向量

计算特征值：令



解方程得到p个非负的特征值，这p个特征值实际上分别是主成分的方差。



计算特征值的贡献率和累计贡献率

第K个主成分的方差贡献率



主成份的累积贡献率为



主成分的名义按特征值的大小排列，在实际应用中，一般取m()个主成分，且使累计贡献率全85%。 。计算特征向量



(4)主成分



(5)构造综合评价函数



**4.2 实证分析**

**4.2.1 数据来源及处理**

以铜、豆粕和天然橡胶期货合约为例，取样时间为从2008年1月4日到2014年10月29日，价格均为每个交易周最后交易日的结算价。之所以这样选择样本，是因为日数据和月数据相对于周数据，都有各自的缺陷，日数据因为变化频繁，波动较大，可能会导致期货价格期限结构的信息失真，而月数据由于时间跨度太大，又不能完全反映不同到期时间的期货合约的价格。数据来源为万德数据库。

**4.2.2 期货价格时间序列和收益率时间序列**

通常期货价格时间序列是从实际期货价格按到期日延展滚动得到，本文亦采用此方法来构建。例如，第一个期货合约F1的时间序列使用的价格数据来自当前的合约从某时到一个月到期的合约。当前合约一到期，第一个期货合约F1就采用第二个近期合约，从而使得第二个近期合约就变成了当前合约。同样，第二个通用合约F2的时间序列从实际的第二个近期合约的期货价格得来。这样，完全的期货价格期限结构(到期日T<1年)就可以跨越到任何给定的时间。

本文除了对期货价格的时间序列进行分析，还对收益率时间序列进行主成分分析。这个数据也是直接从期货价格时间序列得来的。我们计算每个到期时间序列T的收益率对数

rt=lnFt,T-lnFT-1,T，Ft,T和FT-1,T同属于相同的期货价格时间序列。连续收益率的时间序列的非平稳假设可以被拒绝，对任何给定的合约的收益率序列都可以用于主成分分析。

**4.2.3 期货价格期限结构**

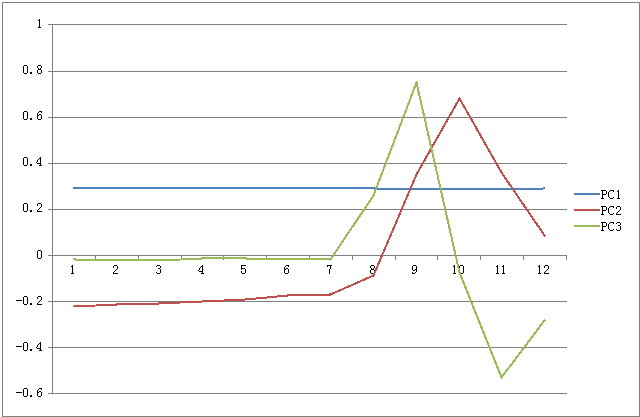
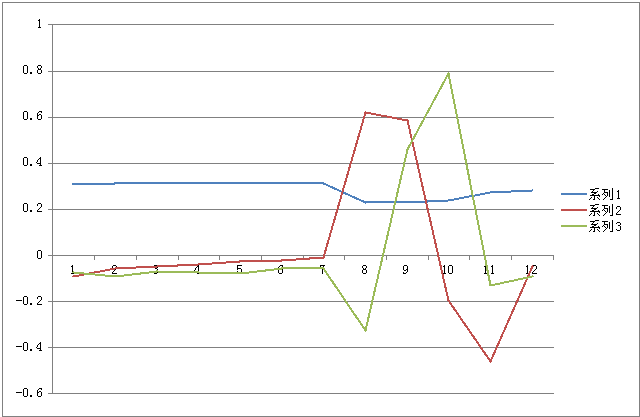
**一、主成分静态分析**

我们分别看一下铜、豆粕、天然橡胶、豆粕和铝的主成分分析结果。

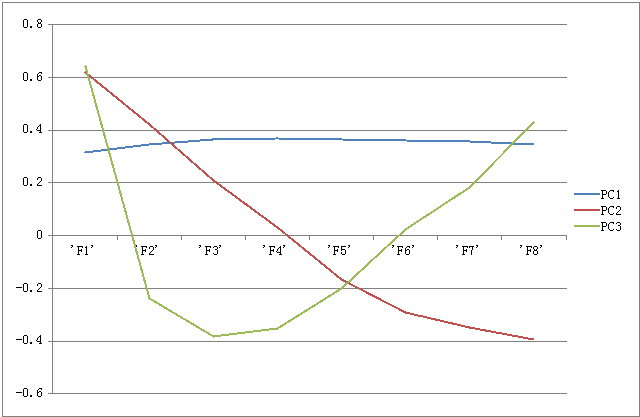
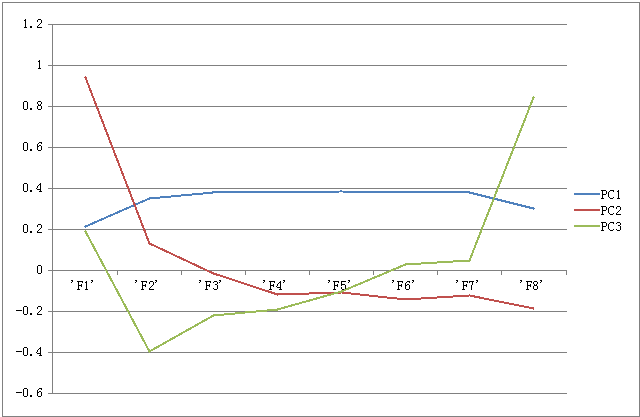
期货价格时间序列的主成分解释能力（即累计方差贡献率）单位：%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **品种** | **PC1** | **PC2** | **PC3** | **累计贡献率** |
| **铜** | 99.0164 | 0.4532 | 0.3689 | 99.8385 |
| **豆粕** | 89.8078 | 7.9122 | 1.2410 | 98.9609 |
| **天然橡胶** | 98.8268 | 0.5203 | 0.2080 | 99.5551 |
| **铝** | 95.3760 | 1.8272 | 1.0940 | 98.2972 |

1.下图分别是铜的价格时间序列和期货收益率时间序列主成分因子载荷图：



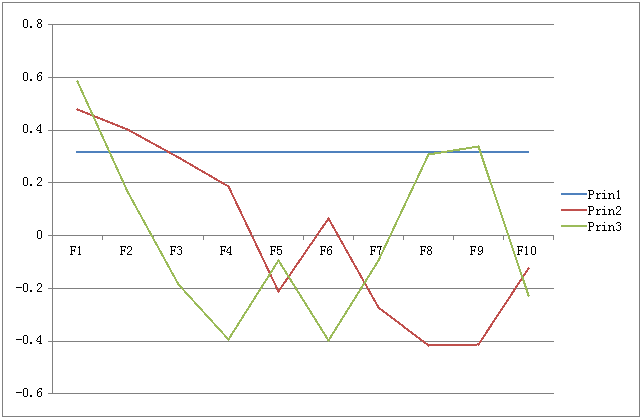
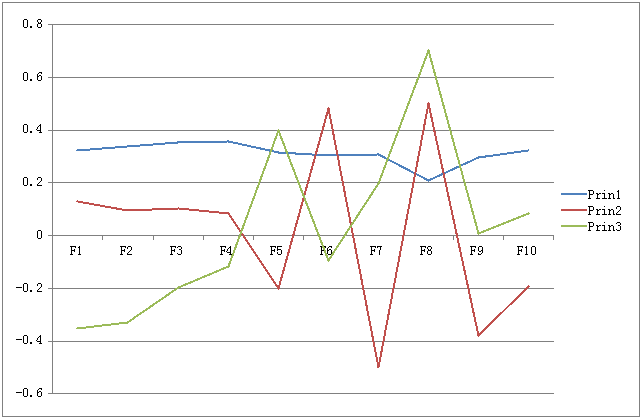
2.下图分别是豆粕价格时间序列和收益率时间序列的主成分因子载荷图：



（2）豆粕期货收益率时间序列主成分因子载荷图

**（3）天然橡胶的收益率时间序列分析**

3.下图分别是天然橡胶价格时间序列和收益率时间序列主成分因子载荷图：



由此可以总结出，商品期货价格和收益率的主成分分析结果有四个共同特征:

1.第一主成分的因子载荷图近似是一条水平的直线，因此第一主成分因子对任何合约的的影响大致相同，这个效应可以被解释为曲线的平行移动。第一主成分因子的解释能力是最重要的，其期货价格时间序列的第一主成分解释能力都超过89%。

2.观察第二主成分因子载荷图，可以看出其斜率要么为正要么为负。说明该主成分因子的变化对短期合约与长期合约的影响方向相反，因此是期限结构曲线发生相对移动的最好写照，因此称为斜率因素。

3.第三主成分的解释能力相对较小，特别是在期货价格的时间序列的主成分分析

结果中，有些品种几乎可以忽略。如果我们绘制它的因子载荷图，发现它是个凸(或倒凸)

的形状。由于它使中间到期日期货的价格和收益率向到期日长和短的反方向移动，它被称

做凸度因素。

4.进一步的主成分基本没有解释能力。特别是期货价格时间序列，前三个主成分

解释了总方差的98%以上，累计贡献率大大超过85%。而收益率序列前三成分的解释能力

要比期货价格时间序列低。

这些发现从经济角度可以解释为，所有商品期货价格主要受期限结构的平行移动影响。主成分1可以解释为宏观经济影响，主成分2可以解释为期货预期价格影响，主成分3由于影响较小，定为其他影响。

**二、滚动主成分分析**

为了检验主成分分析结果的稳定性，我们进一步以前249周作为初始时段进行滚动主成分分析，每次向前滚动一周，共形成80个数据集，进行80次滚动主成分分析。滚动主成分分析考察了我们提取的主成分的解释力在样本期内随着时间的变动是否具有稳定性。结果如下（单位：%）：

|  |  |  |  |
| --- | --- | --- | --- |
| **品种** | **收益率** | | |
| **Std\_PC1** | **Std\_PC2** | **Std\_PC3** |
| 铜 | 0.2513 | 0.109 | 0.0717 |
| 豆粕 | 0.8494 | 0.2698 | 0.3275 |
| 天然橡胶 | 0.390805 | 0.232773 | 0.183812 |
| S\_PCA为主成分成分的标准差 | | | |

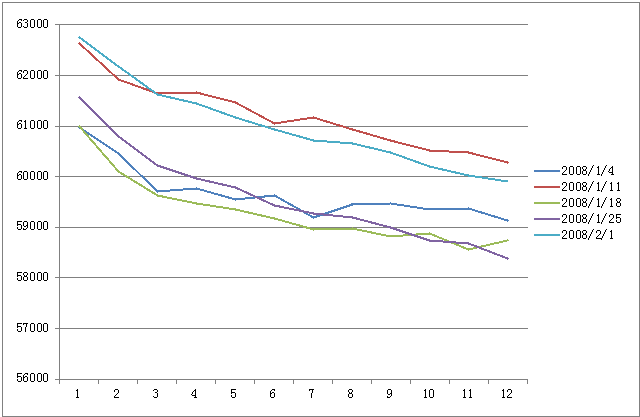
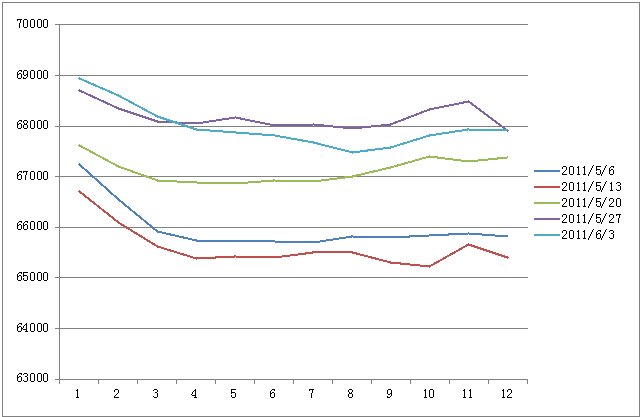
**结果分析：**

各主成分的标准差都在1%以下，解释力较稳定。第二主成分的解释能力则与第一主成分有很大关系，即高的第一主成分解释能力的标准差则伴随着高的第二主成份解释能力的标准差。第三主成分的解释能力较小，其标准差的参考意义相对较小。因此，我们可以断定，提取的前三个主成分不仅能够反映期限结构的动态变动，而且其解释力在样本期内具有较好的稳定性。

在滚动主成分分析的过程中，我们对期限结构的变动与各主成分间的关系有了更深入的了解：在不同期限合约价格变动较大从而使期限结构平行移动的时期或期限结构斜率和曲率变化较小的时期，第一主成分表现了较高的解释能力；在期限结构的斜率和曲率变化较大的时期，第一主成分的解释能力降低，第二和第三主成分的解释能力增加。这些发现支持了我们把前三个主成分理解为水平因素、斜率因素和曲率因素的观点。更为重要的是，通过滚动主成分分析，各品种前三个主成分的解释力在样本期内具有较好的稳定性。

下面附上铜、豆粕和天然橡胶的所有数据前五周、中间五周以及后五周价格期限结构图

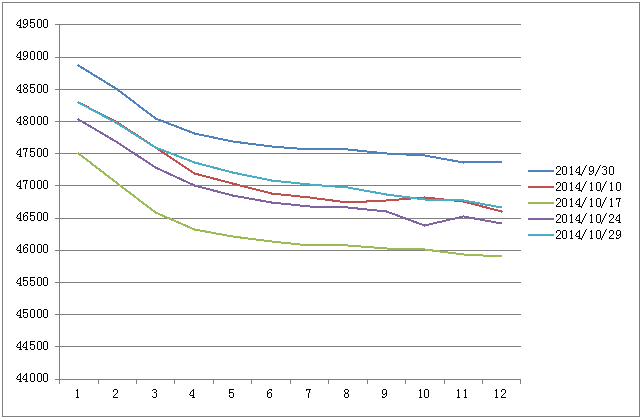
1. **铜的价格期限结构**



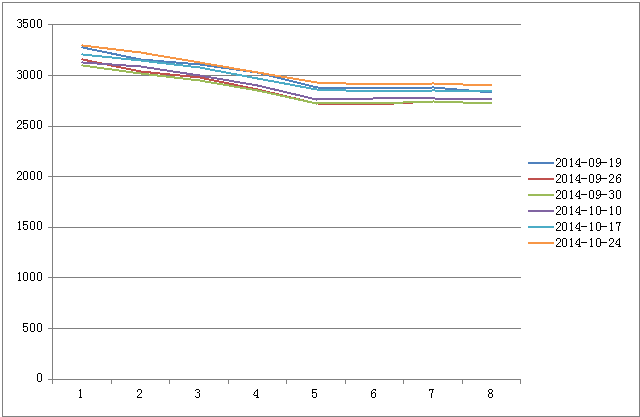
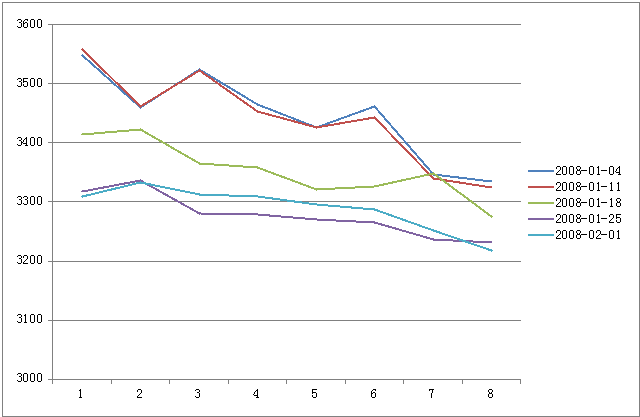
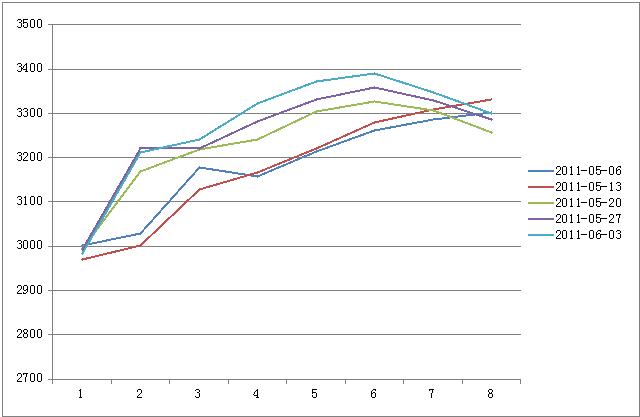
。

（2）中间五周的铜价格期限结构

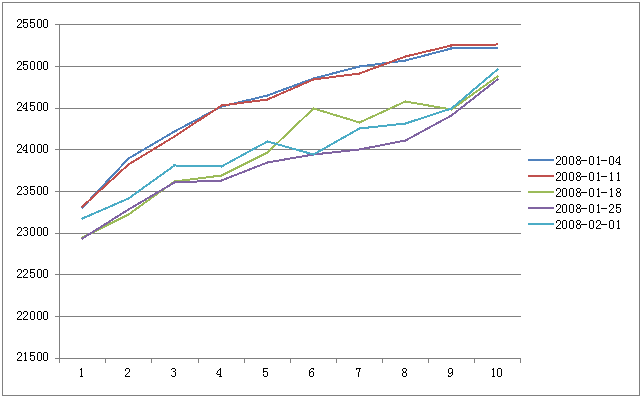
（3）后五周的铜价格期限结构



**（2）豆粕的价格期限结构**



**（3）天然橡胶的价格期限结构**



**（4）期货价格期限结构分析和解释**

反向市场表示近期期货价格或现货价格高于远期价格，相反的情况则表示正向市场。

反向市场状态的出现一般有两个原因:一是近期对某种商品的需求非常迫切，远大于近期产量和库存量，使得现货价格大幅提高，高于期货价格；二是预计将来该商品的供给会大幅提高，导致期货价格大幅下降，低于现货价格。

总之，由于人们对现货的需求很迫切，价格再高也愿意承担，造成现货价格大幅提高，近期合约价格也随之上升，并高出远期月份合约的价格，导致基差为正，出现反向市场状态。对于容易储存的商品，反向市场就违背了持有成本理论。

我们可以发现铜期货基本处于反向市场，豆粕在前五周和后五周处于反向市场、中间五周处于正向市场，天然橡胶在前五周和后五周处于正向向市场、中间五周处于反向市场。

**4.3交易策略**

根据前面的结论，第一主成分(平行移动)高的解释能力源于期限结构的形状不随时间而

显著变化(因此陡度和曲率因素并不重要)，或者，价格水平的整体移动解释了曲线波动的

大部分从而让其他因素变得不重要。基于此，投资商品期货，我们考虑从以下两点切入。

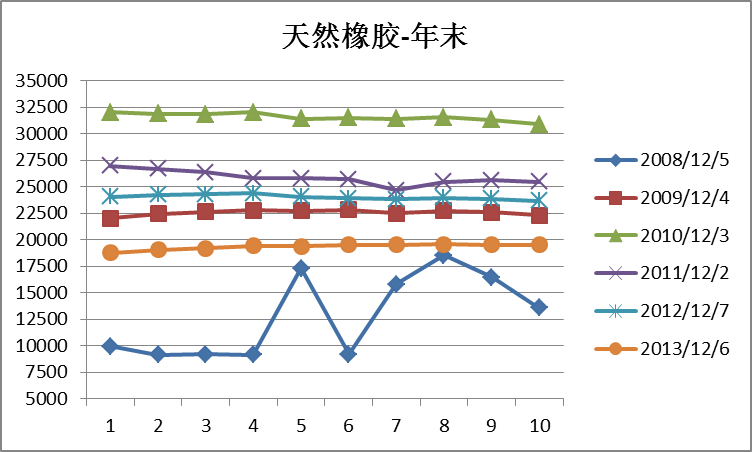
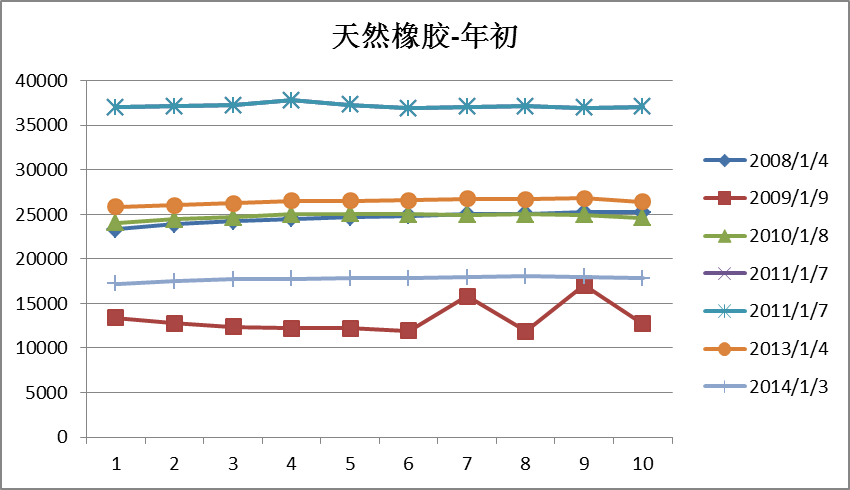
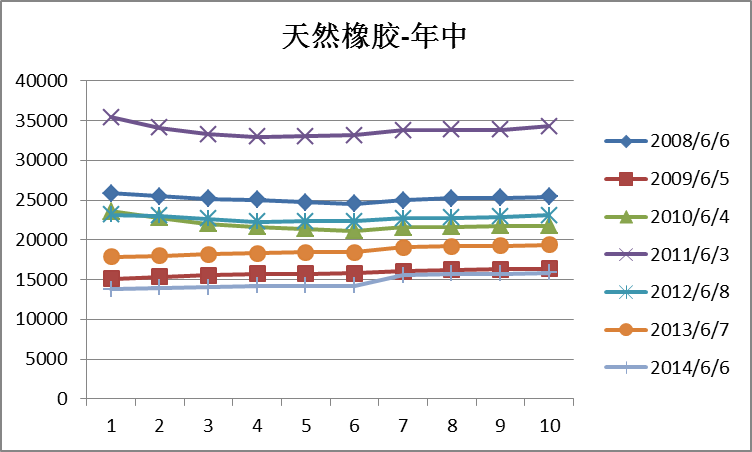
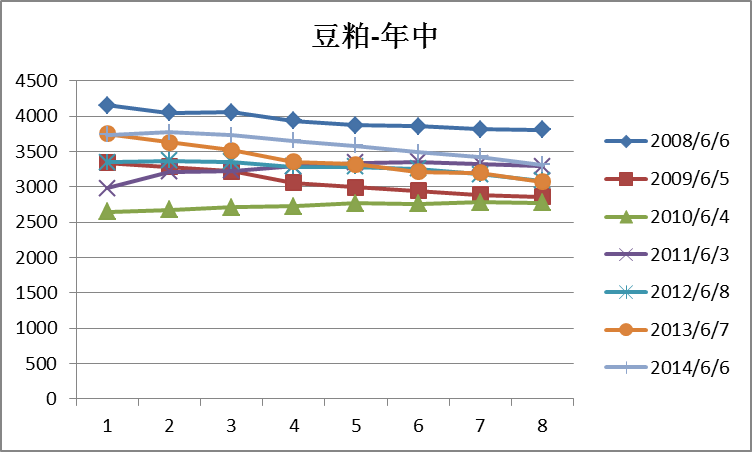
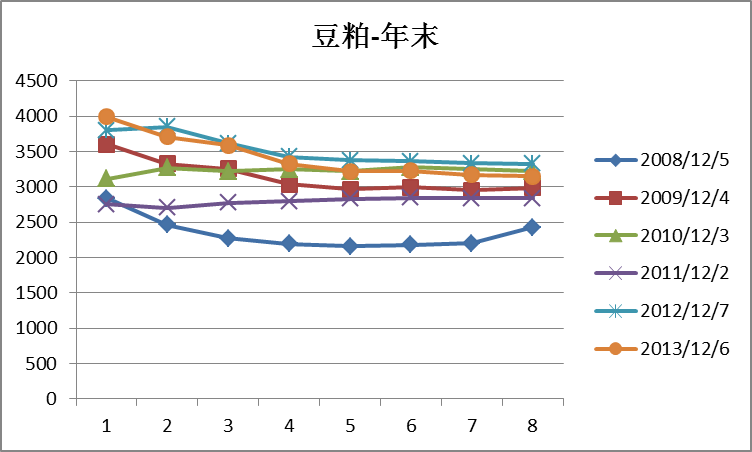
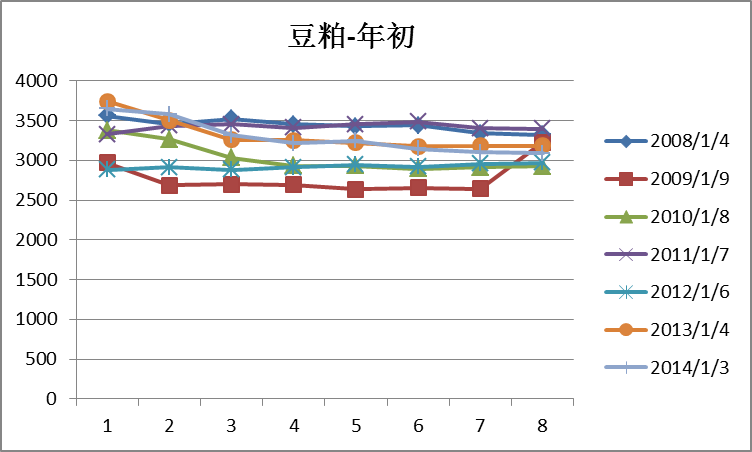
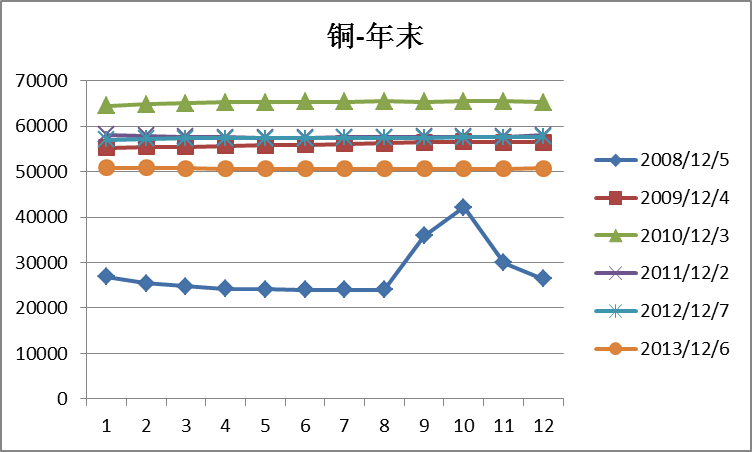
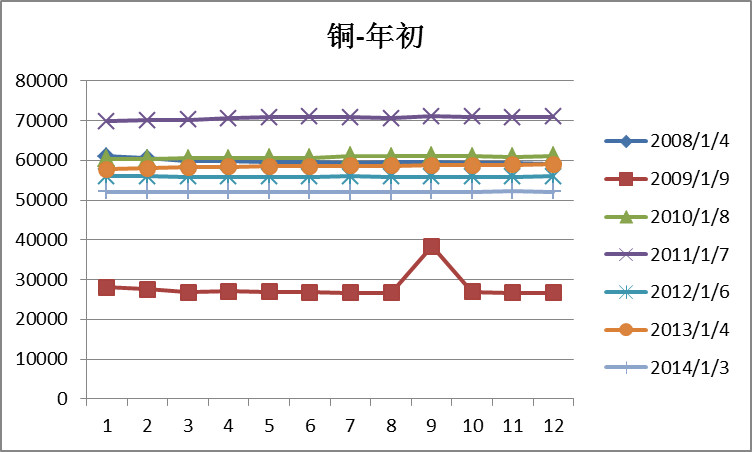
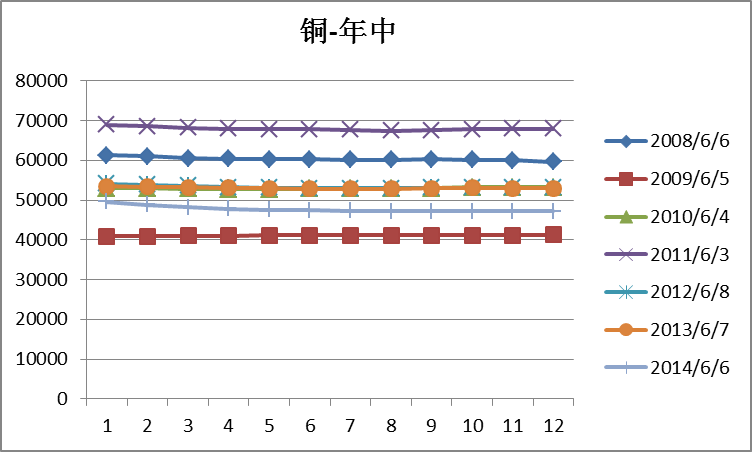
首先，投资那些期限结构的形状是恒定不变的品种。这对于第一主成分有高解释能力的商

品是很重要的。第二，通过选择期限结构曲线上期望滚动收益最高的到期日来最优化持有

的头寸，此点也是到期日T和T+1之间曲线陡度最高的情况。无论第一主成分的解释能力

如何，这都使滚动收益率最优。

为了更直观的说明年与年报之间期货价格期限结构不随时间显著变化，我们取每年年初、年中和年末的铜、豆粕和天然橡胶的期货价格时间序列画出折线图如下：



在期限结构上，曲线的陡度在决定选择哪个合约时起到很重要的作用。通常很难去预

测一笔交易的价格走势，因为期货价格己经反映了那些因素。如果由不同到期时间到一年

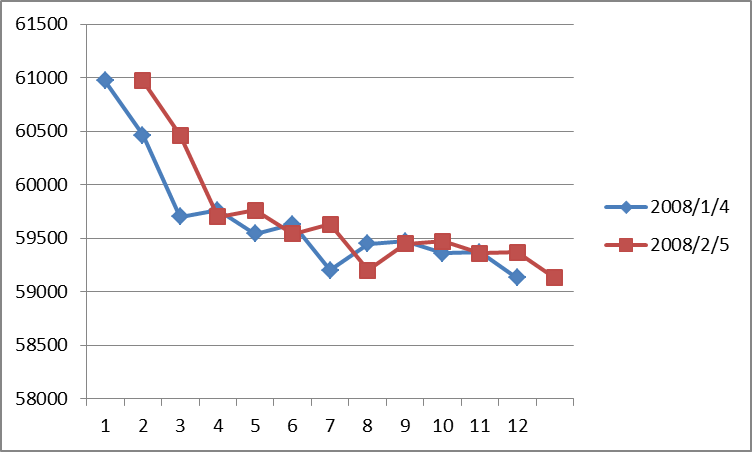
浙江工商大学硕士论文期货价格期限结构隐含信息及其应用研究的期货组成的期货期限结构的正向市场，那么市场期望的现货价格在未来就会上升。为了让期货头寸获得一个正的收益，当前合约价格的增加必须大于期货价格所隐含的增加值。

恒定时间到到期日的总收益率是展期收益和现货价格变化的总和。展期收益是指在期货合约临近到期时，卖出原有合约买入新合约获取的价差，这与该商品期货价格期限结构挂钩。如果期限结构的形状(水平，正向或反向)从一个到期日到下一个到期日不发生变化，那么总收益率就只来自于展期。因此，在这种情况下总收益就取决于选择的合约到期时间和头寸的方向(多头或者空头)。例如，如果今天的燃油期货是正向市场，我们持有近期合约的多头头寸，且持有至该合约到期，当前合约到期后再向下一近月合约展期，结果我们的展期收益为负。类似的，如果是反向市场，我们就可以获得一个正的展期收益。

如果期限结构曲线稳定，我们可以选择使收益最大化或损失最小化的合约来最优化滚动收益率，这就是在反向市场斜率(绝对值)最大点及在正向市场斜率最小点处持有期货合约。基于以上的分析，我们提出下面三种交易策略，通过多头头寸或空头头寸来选择最优合约(假设没有做空限制)。

**交易策略**：根据期货价格期限结构曲线的形状或持有合约多头或持有合约空头。在正向市场，我们持有空头合约，该合约的到期日对应期限结构曲线最大斜率点处，以获得最大滚动收益。在反向市场，我们持有多头合约，该合约的到期日对应期限结构曲线最大斜率(绝对值)点处，以获得期望的最大滚动收益。

下面以铜2008年1月4日数据为例，探讨如何制定交易策略。

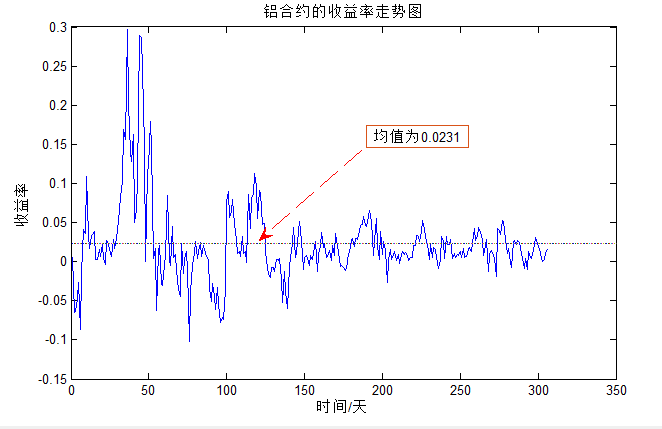
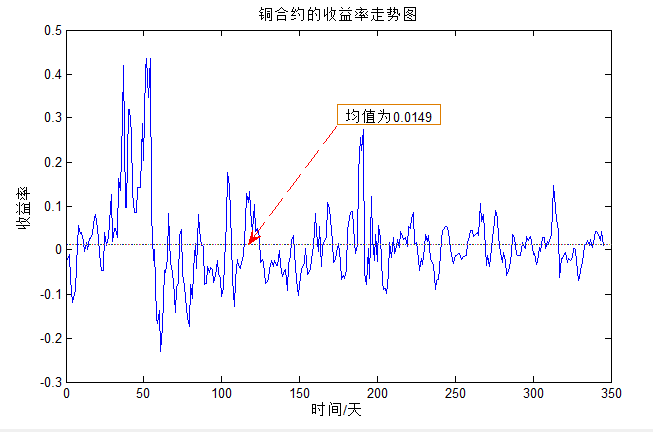


从图中可以看出，2008年1月4日铜期货期限价格结构处于反向市场。蓝色线代表一月合约到十二月合约的价格在2008年1月4日这一天的价格，按照之前的理论证明，期货期限结构没有显著变化，即期限结构可以平移，我们预测至2008年2月5日，代表一月合约到十二月合约的价格如图中红色线所示。

例如我们在1月4日买入2月份合约，那么到下一个月这个合约价格升高，此时卖出该合约，我们可以赚取差价。关于买入或卖出合约选择，我们选取斜率最大或者说价差最大的两个合约中的后一个月份合约。在2008年1月4日这一天，我们就买入3月份合约，在四周后平仓，然后在观察当日期限结构，同样分析，选择期货合约，决定买入（反向市场）或卖出（正向市场）。

一个月中有四个周，每一周如此分析，做决策。对2008年至2014年铜、豆粕、天然橡胶和铝的数据进行验证分析，得到收益率并计算平均收益率，再乘以12得到年收益率。结果如下表所示。

|  |  |  |
| --- | --- | --- |
| **品种** | **月平均收益率** | **年平均收益率** |
| **铜** | 0.0149 | 0.1788 |
| **铝** | 0.0231 | 0.2772 |
| **天然橡胶** | 0.0432 | 0.5164 |
| **豆粕** | -0.0111 | -0.1332 |



再考虑到期货收益是保证金交易，此策略的收益情况比较可观。

**5 交易策略选择（期货价格时间序列重标极差法）**

**5.1期货价格时间序列重标极差法简介**

分形时间序列的重标极差分析，又称R/S分析方法。水文专家Hurst在研究尼罗河泛滥的记录时发现其水流量并不是一个随机过程,它们似乎呈现出循环,长度是非周期的。赫斯特提出了R /S分析法，不限定时间序列的线性特征，对序列本身数值进行结构分析，没有将其与稳步递增的时间标示相关联，将原序列按比例分段，用逐段讨论与综合分析相结合，研究序列的长期关联性和记忆性，最终是得出Hurst指数来衡量序列的长期性。

重标极差的一般形式：设有一时间序列：, t=1,2,。



对于任意的整数n,定义均值序列：



累积离差：

极差为：



相应标准差：

为了比较不同类型的时间序列，Hurst取其比值作为重置标度，有以下关系：



取对数，即为：



其中, c为常数, n为时间序列的长度, 中n为对应长度为n的时间序列的R /S值,H称为Hurst指数。



H的取值范围是[0,1],对于标准的布朗运动其Hurst指数为0.5，以此分界：

1. H =0.5时,说明时间序列是随机游动的，标准布朗运动，马氏链特性。对于金融市场而言,信息被立即反映到价格中,昨天的事件不影响今天,今天的事件不影响明天,事件互不相关,旧新闻已经被市场吸收和计算进去了,肯定有效市场理论。
2. 当H=1时，表示完全预测，时间序列为一条直线，未来完全可以用现在预测。
3. 当0.5<H<1时,持久性分形布朗运动，隐含今天的事件确实影响明天,今天接受到的信息在其被接收到之后继续被市场计算进去,信息可以在很长一段时间内影响未来,并且穿越时间标度。如果H值越大,则说明时间序列的趋势更明显,期货价格增量之间的相关性越强。
4. 当0≤H <0.5时, 反持久性分形布朗运动。此时，时间序列的未来数据倾向于返回历史点，即均值回复特性。

**5.2实证分析**

**5.2.1数据处理及程序实现**

选取沪铜与豆粕1月份到期期货合约的铜期货和豆粕期货的当日收盘价作为样本数据，时间区间为2004年1月2日到2014年11月28日。数据来源：万德数据库。

按照如下定义，定义价格百分比和对数收益率：

=



其中， 为t时的对数收益率, 为t时的价格,数据量为2619个,用对数收益率和价格百分比做出的R/S分析结果及其相近。（这里，我们用对数收益率）

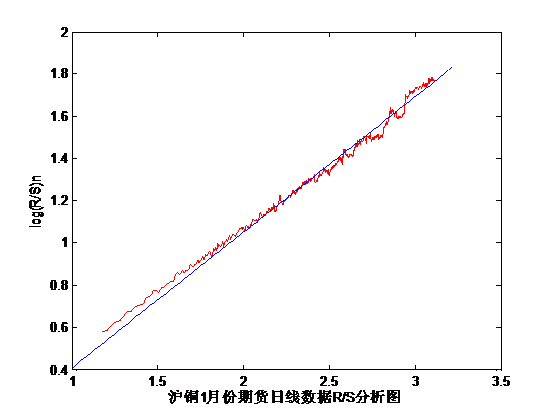


沪铜1月份期货的每日收盘价波动及对数收益率波动如下图所示：



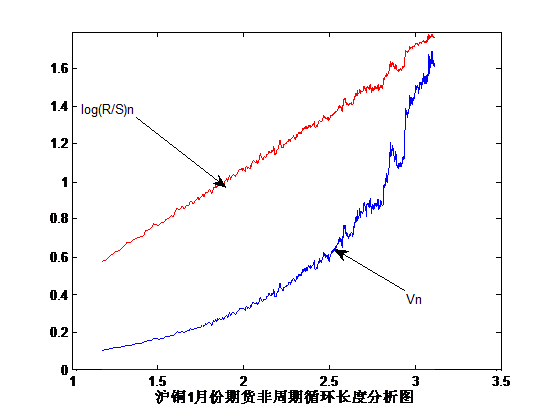
用matlab相关函数编程实现了整个计算过程。

首先均分这个时间长度为T(T =M -1)的时间序列为A个相邻的子区间,使A·n =N,这样每组都有n个观察值。根据彼得斯的经验法则,从n>10开始计算。然后对每个子样本计算其重标极差R和标准差S,可以得到A个R /S值,再求出平均值,计为。接下来,取n +1,n+2,,,循环计算,直到n = (M -1) /2。这样,对于沪铜1月份期货的日数据处理共得到1309组n和(R /S)n值,运用最小二乘法再对上式进行回归,得到的斜率是Hurst指数的估计值,如下图所示：

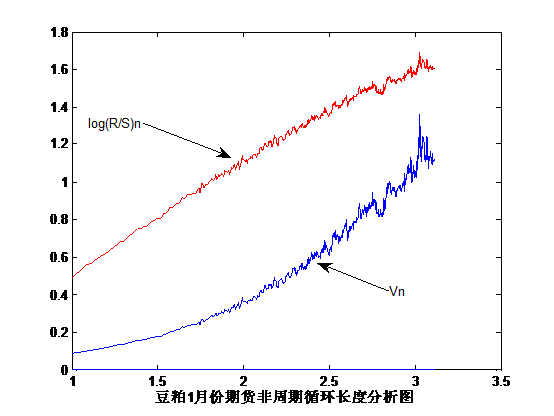
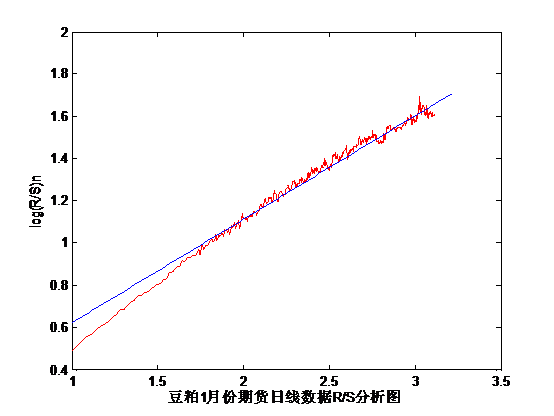
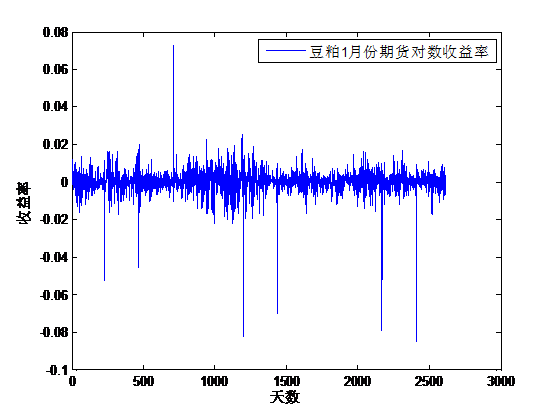
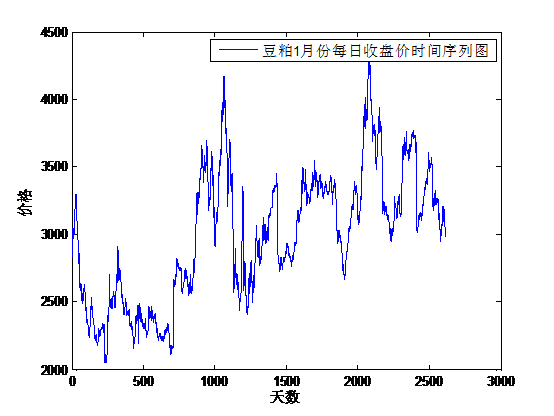


计算得到H =0.6410,大于0.50,说明沪铜1月份期货价格时间序列不是随机游动, 价格波动并不遵循有效市场理论,期货价格时间序列的观测值之间不是相互独立的,期货价格时间序列具有持久性趋势。沪铜期货价格具有记忆性,今天发生的事件将对以后的期货价格产生影响。

独立的随机过程系统,统计量关于是平坦的(Hurst提出的统计量用于估计非周期循环长度)。对于具有状态持续性的过程（）来说, 关于是向上倾斜,会一直上升。如反持续（）或随机游走（）时，统计量就会出现单调下降或保持不变。当值发生突变时,如下图，统计量曲线由上升转为下降的分界点即为序列长期记忆的消失点。通过计算统计量,分析下图中数据发现, 关于的曲线在n =501时发生突变,估计沪铜1月份期货时间序列的非周期循环长度大约是501天。



相应的豆粕合约进行同样处理，得到如下结果：



注意：其上豆粕的Hurst指数为0.4812，但右边的统计量却是递增的，与事实不是很相符，需要进一步对数据进行进一步核对，分析和解释。

**5.2.2 阶段性结论**

根据以上分析,得出以下结论：

沪铜期货价格波动不是随机游动,是典型的有偏随机游动,沪铜的H指大于0.5，,期货市场收益率时间序列存在状态持续性,即如果价格在前一个时期是向上(或向下)走的,那么它在下一个时期延续这一趋势的可能性更大,这与实际情况是一致的,市场经常表现为期货价格的连续上涨或连续下跌。价格具有记忆性,今天发生事件的影响会一直持续下去,直到长记忆消失。沪铜1月份期货合约的非周期循环长度大约是501天,这里的501天的循环不是“价格”的循环,是统计特征的循环。501天是铜期货价格长记忆消失所需要的时间,这个时间长度对于投资者决策有一定的指导作用。

**5.2.3下一步计划**

1.进一步处理：首先需要对每支期货的不同月份按日，周，月时间跨度进行分析，得出比较全面的Hurst指数，进行综合分析。接着以跳跃点为界，分割时间序列，分别再次计算出Hurst指数，进一步验证非周期循环的存在性，并进行分析。最后需要计算不同期货品种和股指期货的Hurst指数。如在投机中可以寻找Hurst指数异乎寻常大的期货，然后研究过去的走向形势，判断其基本形势是看涨还是看跌进行投机。

2.H值稳定性检验

彼得斯认为要检验H的稳定性必须检验在不同的时间频率上H值如何缩放。

继续运用R /S分析方法分别研究**以周和月为间隔（待做,）**的沪铜期货价格时间序列,比较在不同的时间频率上H值如何缩放。

3.有效性检验

为了进一步论证上述结论,采用彼得斯的方法,通过**随机打乱沪铜周收益率（待做）**的时间序列来进行检验。

**6 其他相关研究**

一、趋势分析（非固定窗口期）

1.移动标准差

考虑到移动均值的技术分析经常被人们使用，为了更好的精确性和实用性，我们采用移动标准差（如果较好，可以结合移动均值来分析）

设计了三种显然的趋势：震荡、上升、下降。震荡趋势时，标准差较小，上升及下降趋势时，因为涵盖了趋势的力量，所以标准差大。且变化节点处标准差会有明显的上升或下降的一小段。（绿线为聚类分析）



以分钟铜的价格数据绘图：红线是期限较长的移动标准差，绿线较短：



分成5类（按照时间，价格，标准差聚类）



分成3类（按照时间，价格，标准差聚类）零附近的红线为移动标准差。



分成3类（按照时间，价格，持仓量）

可见：加入持仓量后分类信息居然和含标准差的几乎一致，所以，标准差中确实蕴含了一定的信息，或者说，人们在不自觉地按照标准差制定交易策略。



后来有考虑联合偏度和峰度一起考虑，蓝线是偏度，绿线是峰度。但是还没考虑到是用综合指标（均值、标准差、偏度、峰度的综合函数）还是加上（时间、价格）做多维聚类。

这张图的红线是移动标准差，0附近的蓝线是移动均值的差，可以看出。移动均值的峰点对应移动标准差的峰点，谷点对应移动标准差的峰点，也就是说：移动均值所涵盖的信息量更大。故，放弃了移动标准差。

而且，聚类分析首先要确定分几类，固然可以分3类，但是又要求对数量级的处理，技术上没达到，且，项目主要要求宏观分析，故，后来没做。

2、移动K值

思想：移动均值是求固定天数下的移动

Cu5分钟价格走势



对数收益率走势图：



在求k值的程序中：令n=2,则求出的数据：(p2-p1)/p1约=ln(p2)-ln(p1)



可以想象：K值与间隔收益率其实差不多。

图：n=20 ,(p20-p1)/p1；



图：20天移动K值



蓝线为价格，绿线为移动K值。可见前一张图的绿线（20天的首尾收益率），与上图的绿线（20天数据的移动回归系数K值）有很强的走势相似性。

K可以视为(Pj-Pi)/pi而(Pj-Pi)/pi其实就是收益率。

由于j和i可能间隔时间长，可以忽略掉价格中间的变化，故有一定的模拟价格走势的能力。移动K的图与间隔收益率的图相似。也就是说：直线回归的办法其实就是考虑其收益率的变化。

因此，舍弃移动K值。

3、移动均值，并找到趋势点：

铜的月数据macd线：



依据：大于0买入，小于0卖出，并且认为，连续出现10天买入机会就是一个确切的趋势。此时，可识别趋势：



可见，效果并不理想。原因：

1、因为月结算价格并非交易者特别关注的一个价格，只是每天交易结算价的月末数据，不具有特殊性。

2、时间间隔被月结算数据拉大以后交易价格点数较少，波动较大。不宜用盈动平均线来追踪其趋势。

结论：

找到其中间隔的日数据，补充到大曲线中。再找趋势。

二、宏观分析：

１、持仓量分析

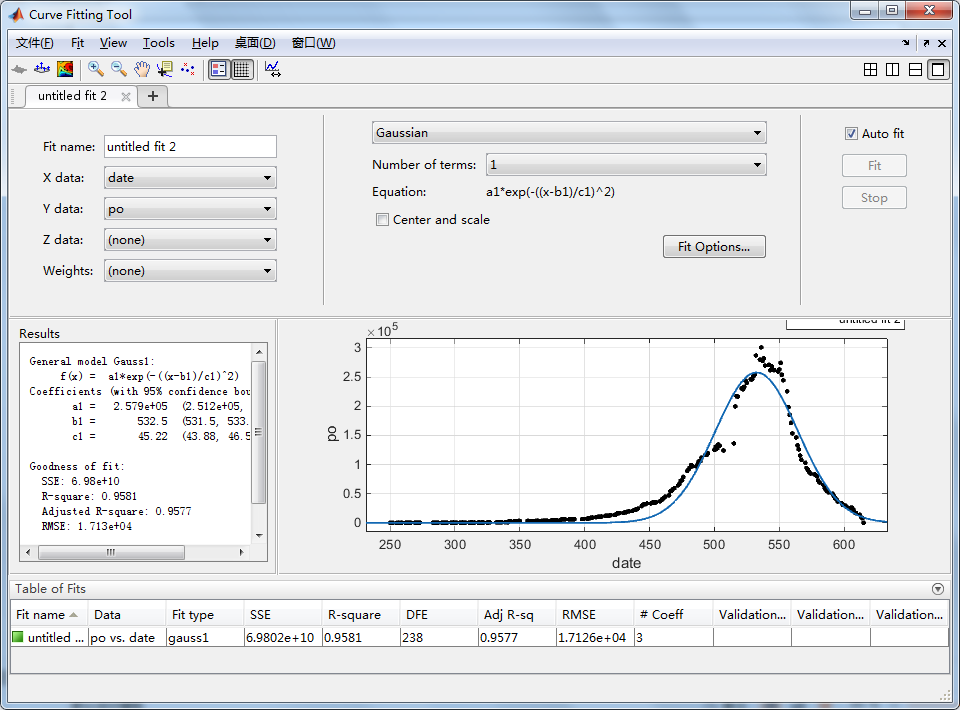
这个并不算宏观分析，但是分析方法与持仓量相同：

铜的每日持仓量（1995-2014）

20年，20个倒U型曲线。如果能分析只有拿出一个倒U型分析



2013-14铜01合约拟合



将价格与离差在一个图里绘制（ｘ，ｙ），观察回归关系

价格走势图及离差（ｄｅｌｔａ）走势：



可见此时，delta与价格无关。

推测：

1、由于价格受主力合约的影响，所以尾部交易量较小，与价格波动没有关系。故：取持仓量90%以上的时间内的交易数据试图证明：价格急剧变化的时刻会有持仓量的变化。使用简单二次函数拟合，较不精确。

90%持仓量下的拟合图，价格（ｐ）～（ｐ－ｐｈａｔ）离差散点图，（９０％持仓的）价格曲线图，（ｐ－ｐｈａｔ）离差曲线图



使用更精确的拟合Fourier（解释力９８％）



可见：精确拟合并没有优势。

改进：1、先找到趋势点。大家一定会在该点买入。故交易量必须增加。同理：找到关键点处有没有回归关系。趋势买入点与持仓量离差（ｐ－ｐｈａｔ）趋势之间有关系：买入点价格走势图和买入点离差走势图如下：



结论：

当买入点出现，且delta值小于零时，即使持有并错过卖出点，在下一个买入点时价格依然比买入价高，也就是说，delta小于0（过低持仓）时是一个长期买入点。而delta大于0（过高持仓）时即使有买入机会也应该短期买入迅速平仓。卖出点，同理，反之。

二次多项式：

3天及5天移动平均。连续3天认为有趋势。

上下：趋势点价格~离差。（b,s,b&s,b&s&be&se）试图找到这些关键点处的回归关系。

买入点，即上升趋势的起点，卖出点，即下降趋势的起点。

买入点价格、买入点离差；卖出点价格、卖出点离差；趋势起点价格，趋势起点离差；趋势始末点价格，始末点离差，如下图所示：



其实，更应该考虑趋势内的delta的变化，这才符合人们连续分析的实际情况。（人们并不知道下一个趋势什么时候开始，这一个什么时候结束）买入、卖出趋势内，P~delat



结论：

p↓趋势明显，po↓-pohat=delta↓

p↑趋势明显，po↑-pohat=delta↑

图中5段趋势中有3段支持，其他2段可能是交易者过于谨慎，不认为趋势明显。也有一种说法："涨的慢，跌得快。投资者对下跌敏感而对上涨不敏感。"刚好在此处得以印证。

推测：n更大时，趋势更明显，该关系成立。更严格得识别趋势（或者说，只识别大趋势）n=7；8



建议

首先做出回归估计pohat,时时计算delta。

卖出机会（移动平均法）出现时，只要delta也在减小应立即卖出，否则可认为时短期波动，忽略不计。

买入机会（移动平均法）出现时，如果delta也在增加可以认为是一个较长的买入时机。而delta并没有增加反而下降。价格也可能增加，但不会持续。且此时delta走势与价格单调性相反。

同时观察delta值。结合delta是否大于0（超额持仓或过低持仓来分析机会点）

还可以分析的地方：

波动趋势相似性。

delta与0的关系。

但意义不大，毕竟：“（与持仓量类比）股票成交量只是价格预测的辅助指标。”

**7 下一步研究计划**

**7.1评价各种交易策略模型，选择最终交易策略。**

完善三种策略选择方法，比较优劣。可以采用基于RSI指标做交易策略的收益率作为标准，衡量上述模型的优劣。

RSI的原理简单来说是以数字计算的方法求出买卖双方的力量对比，譬如有100个人面对一件商品，如果50个人以上要买，竞相抬价，商品价格必涨。相反，如果50个人以上争着卖出，价格自然下跌。

强弱指标理论认为，任何市价的大涨或大跌，均在0-100之间变动，根据常态分配，认为RSI值多在30-70之间变动，通常80甚至90时被认为市场已到达超买状态(Overbought)，至此市场价格自然会回落调整。当价格低跌至30以下即被认为是超卖(Oversold)，市价将出现反弹回升。

RSI计算公式为

RSI（n）=n日内上涨各日涨幅和÷（n日内上涨各日涨幅和+n日内下跌各日跌幅和）×100

一般教科书上所给出的计算公式为：RSI（n）=100－100÷［1+RS（n）]

其中RS（n）=n日内上涨各日涨幅和÷n日内下跌各日跌幅和。以 上 两 个 公 式 的 计 算结 果一致，含义也相同，RSI的值在0~100。各种行情软件中所使用的RSI指标计算公式的具体形式可能会稍有变化，但基本思想是相同的。RSI指标一般会同时给出计算周期一般为7天和14天的两条曲线，不同的使用者可以根据需要进行修改。

一般而言，RSI值为50是市场强与弱的分界线，RSI值介于50~80时为强，表明市场多方力量强于空方，此时市场适宜做多；RSI值介于20~50时为弱，表明市场空方力量强于多方，此时市场适宜做空。当RSI值介于0~20与80～100时，预示着市场处于超买超卖中，表明目前的下跌或上涨行情已经超出了理性的范围，市场行情反转的可能性很大.

如若可能，综合运用三种方法，进行策略选择以减小风险。

**7.2优化策略及控制风险**

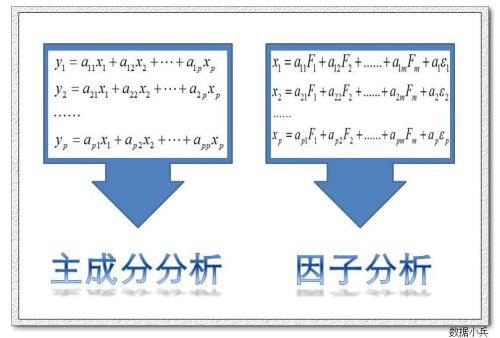
划分期货市场风险，将市场风险，信用风险，流动性风险等各类风险，用方法将风险量化，然后进行风险管理的量化建模，确定各项风险的频度和强度，为选择适当的风险处理方法提供依据。

**7.3程序的初步研发**

根据建立的模型完成程序开发，利用等编写代码实现程序化的计量和监测。并导入模拟数据进行测试并调试程序。导入万家基金的期货管理信息，对模型进行进一步优化和实际应用。根据优化后的传统模型，对已经开发出来的程序进行后期优化、调试，以开发出一套适合万家基金实际情况的程序。

**附录1：**

**因子分析与主成分分析的区别**



主成分分析：研究如何通过少数几个主成分来解释多变量的方差一协方差结构的分析方法，也就是求出少数几个主成分(变量)，使它们尽可能多地保留原始变量的信息，且彼此不相关。它是一种数学变换的方法，即把给定的一组变量通过线性变换，转换为一组不相关的变量(两两相关系数为0，或样本向量彼此相互垂直的随机变量)。在这种变换中，保持变量的总方差(方差之和)不变。同时具有最大方差，称为第一主成分；具有次大方差，称为第二主成分。

 在主成分分析中，最终确定成分是原始变量的线性组合。每个主成分都是由原有p个变量线性组合得到。在诸多主成分Z i中，Z 1在方差中占的比重最大，说明它综合原有变量的能力最强，越往后主成分在方差中的比重也小，综合原信息的能力越弱。

      因子分析: 因子分析是寻找潜在的起支配作用的因子模型的方法。因子分析是根据相关性大小把变量分组，使得同组内的变量之间相关性较高，但不同的组的变量相关性较低。每组变量代表一个基本结构，这个基本结构称为公共因子。对于所研究的问题就可试图用最少个数的不可测的所谓公共因子的线性函数与特殊因子之和来描述原来观测的每一分量。通过因子分析得来的新变量是对每个原始变量进行内部剖析。因子分析不是对原始变量的重新组合，而是对原始变量进行分解，分解为公共因子和特殊因子两部分。具体地说，就是要找出某个问题中可直接测量的具有一定相关性的诸指标，如何受少数几个在专业中有意义、又不可直接测量到、且相对独立的因子支配的规律，从而可用各指标的测定来间接确定各因子的状态。

1、主成分分析在于对原始变量的线性变换，注意是转换、变换；而因子分析在于对原始变量的剖析，注意是剖析，是分解，分解为公共因子和特殊因子。

2、这两种分析法得出的新变量，也就是成分或者因子，并不是原始变量筛选或者提出后剩余的变量。

3、因子分析只能解释部分变异（指公共因子），主成分分析能解释所有变异（如果提取了所有成分）。

4、主成分分析，有几个变量就至少有几个成分，一般只提取能解释80%以上的成分；因子分析，有几个变量不一定有几个公共因子，因为这里的因子是公因子，潜在的存在与每一个变量中，需要从每一个变量中去分解，无法解释的部分是特殊因子。

**附录2：**

**因子分析理论基础介绍**

因子分析起源于心理学领域。20世纪初心理学家斯皮尔曼在研究智力活动时首先提出了因子分析的思想，后来经过汤普森、凯勒、霍特林以及瑟斯通等人的完善改进，直至20世纪50年代，因子分析技术形成，但是仍然存在很多问题。最初的因子提取方法只是近似，决定因子数量的校标对于相关解释也是近似的。

近年来，由于不断深入的理论研究以及引入了计算机技术，一些问题得到了解决或者是有了明确的界定。因子分析从最初应用于心理学扩展到多个领域，目前已经在心理学，社会学，经济学、考古学、医学等领域有着广泛的应用。

因子分析法通过研究相关阵或协方差阵的内部依赖关系，在处理具有多个指标的问题中用来寻找出支配多个指标的少数几个公因子或者是共性因子，这些公因子具有一个性质，就是他们是彼此独立或者是不相关的。在研究问题时，以公因子代替原指标作为研究对象，可以达到到在不损失或很少损失原指标所包含信息的效果。简而言之，因子分析是多元统计中处理降维的一种方法。

根据研究对象的不同，可将因子分析分为两类。一类是R型因子分析，研究指标之间的相关关系,分析相关阵或是协方差阵内部结构，找出控制所有变量的几个公共因子；另一类是Q型因子分析，研究样本之间的关系,通过对样品相似矩阵的内部结构的研究找出控制所有样本的几个主要指标[1]。

总的来说，因子分析应用可概括为两大类：第一类是简化观测系统，在能包含原始变量信息的条件下，将关系错综复杂的对象（变量或是样品）综合为少数几个因子；第二类是用于分类，将个指标或是个样本分类。



**2.1 正交因子模型**

假设要进行分析的原指标有个，记为可观测的随机向量，，且设是不可观测的随机向量，，又设与互不相关，且满足：



。



假设随机向量满足



则称模型为正交因子模型，用矩阵表示如下：，



其中，为的公共因子；，称称为X的特殊因子。而且各个特殊因子之间以及特殊因子与所有公共因子之间均是不相关的。



模型中称为因子载荷矩阵，为待估系数矩阵。



为了更好的解正交因子模型，有必要来探讨一下正交因子模型中各个量的统计意义。

由正交因子模型的假设条件可知：



由上论述可知，



若变量是标准化变量，则有：



，即因子载荷就是第个变量与第个公因子之间的相关系数。



令，称为变量的共同度。



由此可知的方差有两部分构成，第一部分是公因子对变量的总方差的共吸纳称为公因子方差，第二部分是由特殊因子产生的方差，仅与变量有关，亦称剩余方差[1]。



在因子载荷矩阵中，求得的各列平方和，记为，



，表示第个公因子对的所有分量的总影响，越大，对的贡献就越大。



**2.2 参数估计方法**

假设已知个相关变量的次观测值为。因子分析的目标是用少数几个公共因子（设为个）描述个描述变量的协方差结构，即：。



由个相关变量的观测值计算出样本相关阵作为协方差阵的估计，为了建立因子模型，首先要估计出因子载荷矩。常用的估计方法有三种：主成分法、主因子法和极大似然法。下面着重介绍一下主成分法。



设样本相关阵的特征值为，相应的单位正交特征向量为，则的谱分解式为。当最后个特征值较小时，可以近似的分解为：



其中：



上式给出的就是因子模型的一个解，由此可看出因子载荷矩阵中的第列（即第个公共因子在上的载荷）和的第个主成分之间相差一个倍数，所以上式给出的解称为因子模型的主成分解。



记，则有。



证明：



因此如果选择合适的，则可以使得近似式的误差平方和很小。



一般说来公因子个数的选择有两种方式，一是根据实际问题的意义和专业理论知识确定；二是用确定主成分个数的原则，选择使得下式成立的最小整数：



[1]。



**2.3 正交旋转**

因子分析不仅仅要求出公共因子，更重要的是能够解释公共因子的实际意义。如果每个公共因子的意义含糊不清，则不利于根据实际背景对因子进行解释。由于因子载荷矩阵不唯一，所以应该对因子载荷矩阵实施旋转变换，使得各个因子载荷矩阵的每一列个元素的平方按列向0和1两级转化，以达到简化结构的目的。有三种主要的正交旋转方法：方差最大法、四次方最大法和等量最大法。下面介绍方差最大的正交旋转方法。

因子模型为，对作正交变换：



令，则有：，且满足：



（2.1）



式（2.1）表明，如果是正交因子模型的公因子向量，则对于任意正交矩阵，也是公因子向量，相应的有是公因子的因子载荷矩阵[1]。



利用这个性质，在实际应用中，当我们求得初始因子载荷矩阵后，反复右乘正交矩阵使得具有更明显的实际意义。我们称这种变换载荷矩阵的方法为因子轴的正交旋转。



对于因子载荷矩阵，如果的因子载荷向量（即每一列）数值越分散，相应的因子载荷向量的方差就越大。为了消除由于符号不同的影响以及各个变量对于公共因子依赖程度不同的影响，我们做如下变化：



令，并将第列的个数据的方差定义为，



其中



则因子载荷矩阵A的方差为



值越大，的第个因子载荷向量数值越分散，如果载荷值或是趋于1，或是趋于0，则这时相应的公因子即具有简化结构，所以我们希望因子载荷矩阵的方差尽可能大。



当时，因子载荷矩阵为，取正交阵，



则有：



为的因子载荷矩阵，这相当于将有确定的因子平面旋转了一个角度。



因子载荷向量的方差：



令,整理后知欲满足上式则需要有下列关系式成立：



其中



当时，可以逐次对每两个因子进行上述旋转，选择满足上式的旋转角度。个因子全部配对旋转共需旋转次，全部旋转结束第一次循环，经过第一次循环后计算所得的因子载荷矩阵方差。从旋转后的在和矩阵方差出发，再进行第二循环，第三循环，直到不能再增大。



**2.4 因子得分**

经过上述讨论我们了解了如何从样本相关阵R出发，得到公因子和因子载荷矩阵，并能对公因子作出合理解释。但是有时候我们需要把公因子表示成指标变量的线性组合，或是对每一个样本计算公因子的估计值，这就是因子得分。因子得分有两方面的用途：一是用来诊断模型；而是用来作为进一步分析的原始数据。

因子得分函数可以表示成如下形式：



要求得每个因子的得分，则需求得每个函数的系数。但是由于，所以不能求出精确得分，只能进行估计。



接下来介绍两种常用的估计因子得分的方法。

**2．4．1**  **巴特莱特因子得分（加权最小二乘法）**

因子模型为，假定因子载荷矩阵和特殊因子方差已知，而把特殊因子看做误差。一般不相等，所以我们利用加权最小二乘法估计公因子的值。用误差方差的倒数作为权重的误差平方和：



上式中，已知，X可观测亦是已知量，求的估计值，使得关于的函数达到最小值，即，令,得到的估计值：



,即为因子得分的加权最小二乘估计。



在处理实际问题时，是未知量。但是可以用它们的某个估计值代替。



对于样本，它的因子得分可以进行如下表示：



本文用主成分法估计因子载荷矩阵，在估计因子得分时，通常用不加权的最小二乘法。



，令



得到F的估计值：，即为因子得分的最小二乘估计。



同样对于样本，因子得分为：



**2．4．2**  **汤普森因子得分（回归法）**

假定变量为标准化变量，公因子也已标准化。



在最小二乘法下对因子得分函数进行估计，得到公因子对变量满足如下回归方程：



下面估计回归方程中的系数。



上述个回归方程中是未知待估量，已知条件是因子载荷矩阵。



由因子载荷的意义知：



即：



其中为相关阵，



令，则有:



由此可得，利用回归法建立的公因子对变量的回归方程为：



由样本观测值计算相关阵，估计因子载荷矩阵，代入上式，就得到了因子得分函数，即为汤普森因子得分。



目前为止，对于两种得分方法的优劣评价不一。

**2.5 因子分析的计算步骤**

根据以上的讨论可以总结出因子分析的计算步骤如下：

1. 选择要分析的变量并计算所选原始变量指标的相关系数矩阵。因子分析的前提是观测变量之间具有较强的相关性，而相关系数矩阵描述了变量之间的关系，可以辅助判断原始变量之间是否具有相关关系，同时相关系数矩阵也是估计因子结构的基础。



1. 提取公因子。首先要确定因子求解的方法及因子的个数，本文采用主成分方法。因子个数根据累积方差贡献率确定，一般认为要达到60%以上才符合标准。
2. 因子旋转。通过坐标变换使每个原始变量在尽可能少的因子之间具有密切关系，这样因子解才具有更明显的实际意义。
3. 计算因子得分。