

量化期货 A 组中期报告

数 41 李博扬 仇泽民

一、前言

本组经过讨论后确定本学期量化期货研究的目标是：在平稳过程的 CTA 策略增强框架的基础上，实现单品种仓位控制的效益，进而再进行多品种趋势策略中的仓位管理，实现进一步的效益。其中我们会改进原始两篇报告中的方法，期望找到更有效的仓位管理策略。

前半学期我们着重于仓位管理的目标策略选取以及在平稳过程条件下 CTA 策略增强框架的实现。在这个过程中，我们首先在 TB 上实现了若干经典策略在螺纹钢数据上的回测，选取了 Dualthrust 策略作为我们后期仓位控制的目标策略，之后我们实现了原始报告《基于平稳过程的 CTA 策略增强框架》中量比因子与波动趋势因子的可行性测试，其中遇到了不少难题，将在下文逐一展示。

本篇报告暂不涉及 Dualthrust 策略直接应用的结果（在上次 proposal 中已经展示），本报告着重于基于平稳过程的 CTA 策略增强框架的实现。

二、理论基础

众所周知，大数定律能够保证在具有正的期望收益的情况下，收益的不确定性能够被大量相互独立的类似事件抵消，从而使得期望收益不变的情况下，收益波动大大减小，进而提高收益风险比和最终盈利的概率。但是相互独立这一条件在金融市场中几乎不可能实现，我们拿到的商品期货时间序列数据也是如此。此时平稳时间序列所特有的遍历性定理，类似于时间序列情况下的大数定律，为我们带来了在择时问题中利用大数定律的可能性。

要实现单品种的仓位控制，我们主要是需要解决两个问题：1、策略信号的预期收益的计算。2、在已知信号预期收益情况下，信号的执行交易量的配置。下面我们来分别解决这两个问题。

1、策略信号的预期收益的计算。

我们主要利用多因子模型来对第 i 次信号触发的 t_i 时刻信号预期收益率 f_{t_i} 进行预测，我们暂时采用线性回归的方法，之后我们也会换成 SVM 等机器学习方法进行尝试。本次我们与方正报告选取的因子相同：量比因子与波动趋势因子。

(1) 量比因子：最近 5min 成交均量除以最近 20min 成交均量

$$VRatio_t = MA_t^5(V)/MA_t^{20}(V)$$

可以证明，在成交量的对数变化率为平稳过程的前提下，该量比指标 $VRatio_t$ 为平稳过程。

因子预测能力测试：（按时间顺序，以若干个信号为一组，分为若干组）

(a) 对每一组，用 t_i 时刻的预测收益率 f_{t_i} 对单因子 $VRatio_{t_i}$ 作线性回归，回归数据使用组内 t_i 之前时刻的数据。得到 t_i 时刻的预测值 f_{t_i} 。

(b) 计算每组组内 $\{g_{t_i}\}$ 与 $\{f_{t_i}\}$ 之间的相关系数，计算相关系数均值，并计算相关系数大于 0 的组胜率，绘制直方图。

(2) 波动趋势因子

布林带公式：

$$\begin{cases} C_t = MA_t^n(p) \\ SD_t = std(\{p_i - C_i, i = t - n + 1, \dots, t\}) \\ U_t = C_t + 2SD_t \\ L_t = C_t - 2SD_t \end{cases}$$

其中 p_i 为 i 时刻的对数价格

波动趋势因子定义为：

$$BRatio_t = \frac{U_t^5 - C_t^5}{U_t^{20} - C_t^{20}}$$

即使用 $n=5$ 的布林带带宽除以 $n=20$ 的布林带带宽。可以证明在对数收益率为平稳过程的假设下， $BRatio_t$ 为平稳过程。

因子预测能力测试：同量比因子

(3) 多因子的混合

$$g_{t_i} = f_{t_i} + \varepsilon_{t_i} = a_0 + a_1 VRatio_{t_i} + a_2 BRatio_{t_i} + \varepsilon_{t_i}$$

在实际操作中，系数 a_0, \dots, a_m 的估计采用向前滚动回归的方法。例如，在第 i 个交易信号发生的时刻 t_i ，使用之前所有的开仓信号对应的因子值与交易信号的真实收益率 $g_k, k < i$ 作为训练样本，通过最小二乘法拟合上述方程，得到估计值 $\widehat{a_0}, \dots, \widehat{a_m}$ ，随着训练样本的增长， $\widehat{a_0}, \dots, \widehat{a_m}$ 会逐渐收敛到真实的 a_0, \dots, a_m 。将该估计值代入 t_i 时刻的方程，得到 t_i 时刻的 f_{t_i} ：

$$\hat{f}_{t_i} = \hat{a}_0 + \hat{a}_1 VRatio_{t_i} + \hat{a}_2 BRatio_{t_i}$$

多因子预测能力测试：（按时间顺序，以若干个信号为一组，分为若干组）

计算每组组内 $\{g_{t_i}\}$ 与 $\{f_{t_i}\}$ 之间的相关系数，计算相关系数均值，并计算相关系数大于 0 的组胜率，绘制直方图。

2、已知信号预期收益情况下的交易量配置

我们假设第 i 个信号触发时需要对原始策略执行交易量 w 增强的权重为 w_i （可以为负，表示减弱权重）。当 $\sum w_i = 0$ 时，称为增强中性。若 $\sum w_i \neq 0$ ，我们也可通过调整原始交易信号手数为 $w + \bar{w}$ ，其中 \bar{w} 为 w_i 的均值，于是增强的手数变为 $\{w_i - \bar{w}\}$ ，满足增强中性条件。于是以下我们假设增强中性条件恒成立。

可以证明如下结论：在未来 N 次交易信号的预测值 f_{t_i} 均已知，且增强中性的条件下，使得增强收益 $R = \sum w_i g_{t_i}$ 达到最大信息比率 IR 的增强权重 $\{w_i\}$ 需满足： w_i 正比于标准化后的预期收益率 $\frac{f_{t_i} - \text{mean}(f)}{\text{std}(f)}$ ，其中 $\text{mean}(f)$ 表示未来 N 次 $\{f_{t_i}\}$ 的均值， $\text{std}(f)$ 表示未来 N 次 $\{f_{t_i}\}$ 的标准差。

但是由于 $\text{mean}(f)$ 与 $\text{std}(f)$ 的计算需要一次性用到所有 $\{f_t\}$ ，这在预测过程中是做不到的。但是如果假设 $\{f_t\}$ 为平稳过程，则利用平稳过程性质未来一段时间 $\{f_{t_i}\}$ 的均值和标准差可以通过历史数据的相关统计量逼近，并且最终随着数据量增大会收敛到真实水平。于是我们给出 $\text{mean}(f)$ 与 $\text{std}(f)$ 的估计如下：

$$\widehat{\text{mean}}(f)_{t_i} = \sum_{\{t_k < t_i\}} \frac{f_{t_k}}{N_{t_i}}$$

$$\widehat{\text{std}}(f)_{t_i} = \sqrt{\frac{\sum_{\{t_k < t_i\}} f_{t_k}^2}{N_{t_i}} - \left(\frac{\sum_{\{t_k < t_i\}} f_{t_k}}{N_{t_i}} \right)^2}$$

其中， N_{t_i} 为 t_i 时刻之前交易信号出现次数。

因此，当 w_i 正比于 $\frac{f_{t_i} - \widehat{\text{mean}}(f)}{\widehat{\text{std}}(f)}$ 时，增强策略相对于原始策略可以达到最高的增强信息比率。可以计算超额收益的信息比率近似于 $IR \approx \sqrt{N-1} \text{Corr}(f_{t_i}, g_{t_i})$ ，这也是为什么判断因子预测能力好坏时我们选择计算 f_{t_i}, g_{t_i} 的相关系数。

在期货交易中，由于交易手数总为整数，所以光使得 w_i 正比于 $\frac{f_{t_i} - \widehat{\text{mean}}(f)}{\widehat{\text{std}}(f)}$ 还

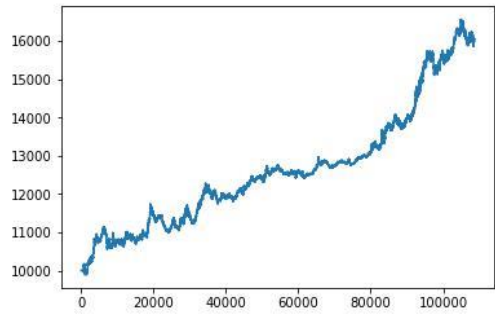
不够,我们需要一种类似四舍五入的近似方法来判断到底是选择增强信号还是减弱信号。我们设置原始策略每次交易过后最大持仓为1手,增强权重上限为1手,下限为-1手,于是经过策略增强后我们的持仓量取值范围为 2,1,0,-1,-2 (负数代表空仓)。由于 $\frac{f_{t_i}-\widehat{mean}(f)}{std(f)}$ 近似标准正态,于是 w_i 也近似均值为0的正态分布,我们选择以 ± 0.5 为界决定是否增强或减弱信号。具体操作见下表:

当前仓位	触发信号	w_i 预测值	增强交易后仓位
1	sellshort	> 0.5	-2
	sellshort	$-0.5\sim 0.5$	-1
	sellshort	< -0.5	0
0	buy	> 0.5	2
	buy	$-0.5\sim 0.5$	1
	buy	< -0.5	0
	sellshort	> 0.5	-2
	sellshort	$-0.5\sim 0.5$	-1
	sellshort	< -0.5	0
-1	buy	> 0.5	2
	buy	$-0.5\sim 0.5$	1
	buy	< -0.5	0

三、 实证探究

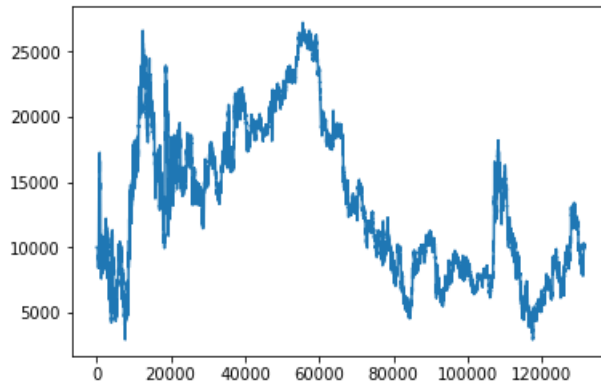
我们希望对三种不同的商品期货:螺纹钢、铜、橡胶实施如上仓位控制方法,以验证上述方法的普适性。以下是 DualThrust 策略在这三种期货数据上回测的资产曲线。(纵坐标为资产总量,横坐标为 index,表示随时间变化)

螺纹钢



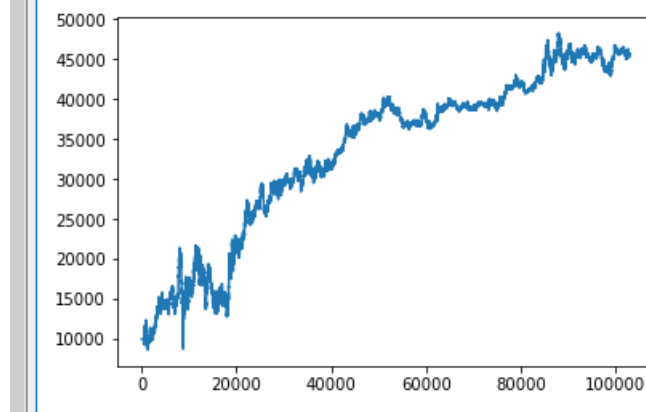
铜:

Out[13]: [<matplotlib.lines.Line2D at 0x1fe5fc8dcf8>]



橡胶:

```
...: plt.plot(x,propertyrecord)
Out[20]: [<matplotlib.lines.Line2D at 0x1fe5ec2da58>]
```



可以看到，Dualthrust 策略在螺纹钢和橡胶上表现较好，在铜上表现很差。

下面我们根据理论基础中的方式来实现仓位控制效果。

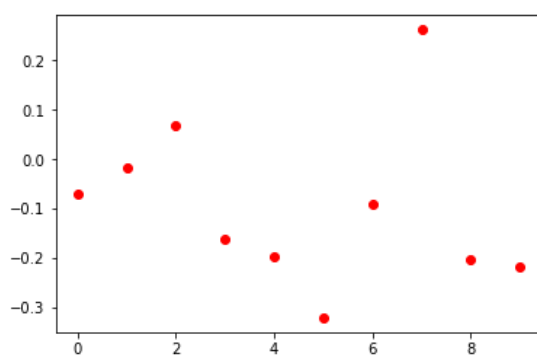
1、因子预测能力测试

理论基础部分提到，衡量因子预测能力好坏的标准是 $\{g_{t_i}\}$ 与 $\{\widehat{f_{t_i}}\}$ 之间相关系数的高低，为了衡量相关系数高低的同时兼顾判断相关系数随时间变化的稳定性，我们将信号按照时间排序后，以每 30 个信号为一组，检验每组组内 $\{g_{t_i}\}$ 与 $\{\widehat{f_{t_i}}\}$ 之间的相关系数，并观察相关系数大于 0 的组胜率。下图分别展示了 Dualthrust 在三个不同品种上的做多信号与做空信号每组中 $\{g_{t_i}\}$ 与 $\{\widehat{f_{t_i}}\}$ 之间的相关系数数值。

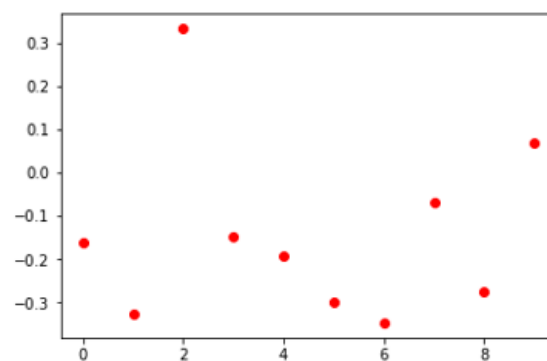
(1) 对于量比因子估计出的 $\{\widehat{f_{t_i}}\}$ 与 $\{g_{t_i}\}$ 的相关系数图

A、螺纹钢:

做多信号:

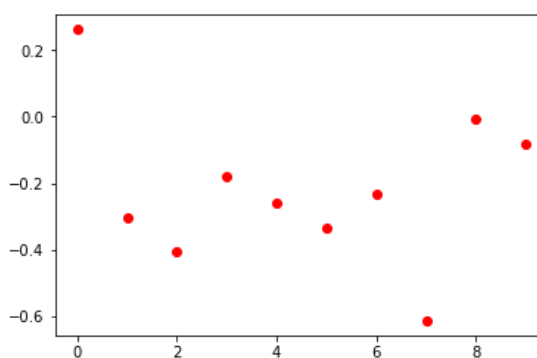


做空信号:

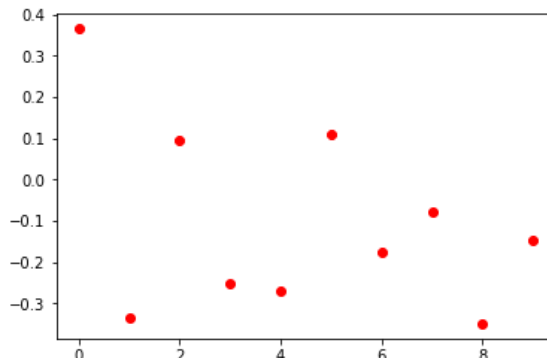


B、铜:

做多信号:

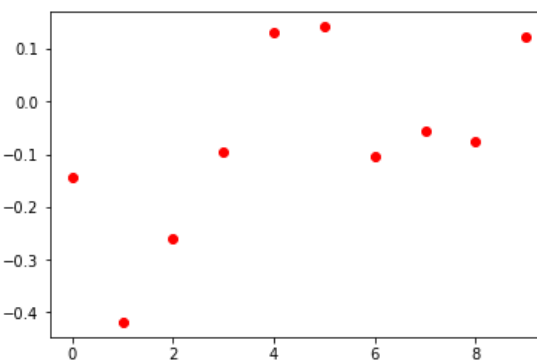


做空信号:

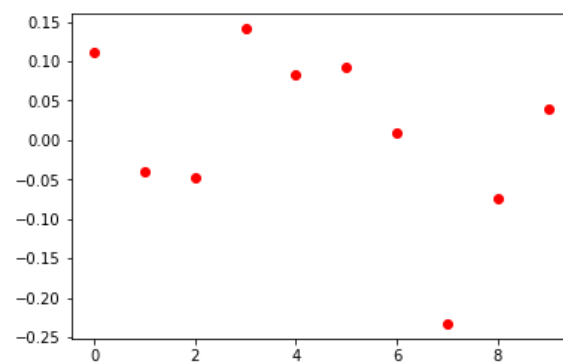


C、橡胶

做多信号:



做空信号:

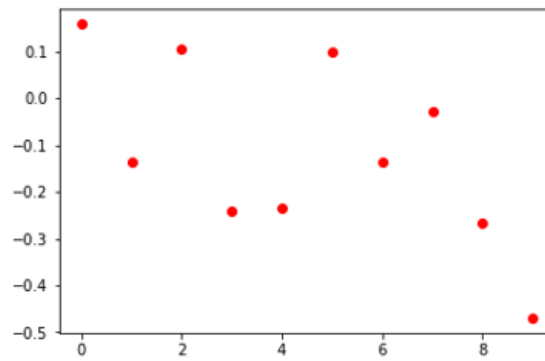
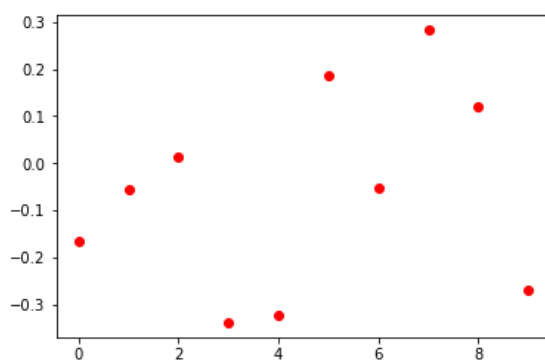


(2) 对于波动趋势因子估计出的 $\{\widehat{f}_{t_i}\}$ 与 $\{g_{t_i}\}$ 的相关系数图

A、螺纹钢:

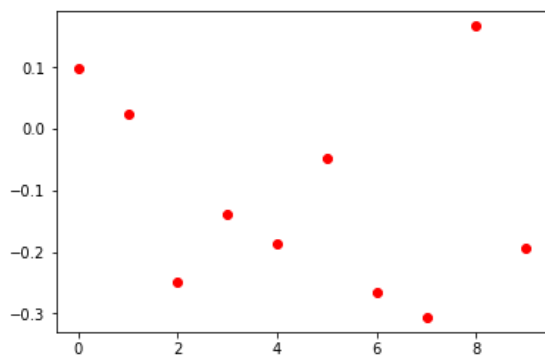
做多信号:

做空信号:

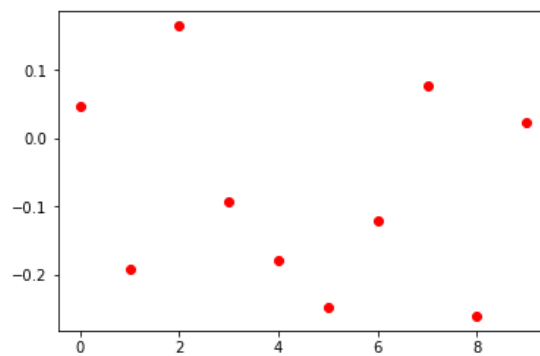


B、铜:

做多信号:

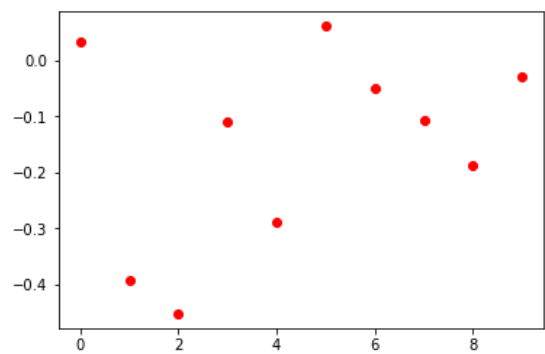


做空信号:

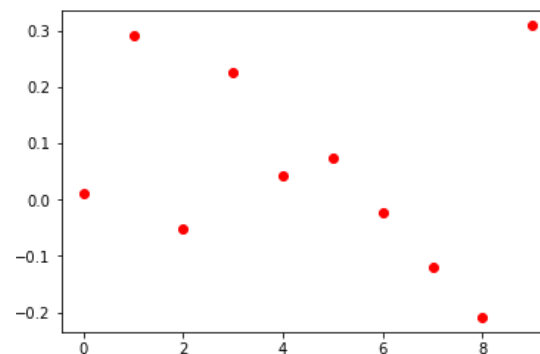


C、橡胶:

做多信号:



做空信号:



因子预测能力如下表:

因子种类	期货品种	信号类型	胜率
量比因子	螺纹钢	多	20%
		空	20%
	铜	多	10%
		空	30%
	橡胶	多	30%

		空	50%
波动趋势因子	螺纹钢	多	40%
		空	30%
	铜	多	30%
		空	40%
	橡胶	多	20%
		空	60%

从预测值 $\{\widehat{f}_t\}$ 与 $\{g_t\}$ 的相关系数大于 0 的胜率来看，预测结果很不乐观，因此量比因子和波动趋势因子并不适合作为预期收益率的预测因子，我们必须另外选取合适的因子完成这个实验。

四、 总结

单品种仓位控制方法大体上已经在理论基础部分进行了阐述，但是在实证研究中，我们的实验卡在因子的预测能力不合格上，所以下一步必须得到更加适合预测策略收益率的因子才能完成实证实验。

在实验过程中我们也发现了若干可以改进的地方。比如我们可以使用 SVM 等机器学习方法代替原有的线性回归预测，或许能得到更准确的预测结果。