

My Quant Notes

Hans Yang

2016 年 9 月 14 日

目录

第一部分 高管增持事件策略	3
1 高管增持事件概述	3
1.1 高管增持事件公告日	3
1.2 高管增持事件最优投资策略	3
第二部分 Log-Periodic Power Law模型	3
2 概述	3
3 模型拟合	4
4 LPPL Confidence 和Trust 指数	4
4.1 LPPL Confidence	4
4.2 LPPL Trust	5
5 模型应用	6
5.1 LPPL Confidence	6
5.1.1 上证综指 (000001.SH)	6
5.1.2 南华螺纹钢指数 (NH0016.NHF)	10
5.1.3 南华豆粕指数 (NH0015.NHF)	14
5.1.4 南华焦煤指数 (NH0029.NHF)	19
5.2 LPPL Trust	24
6 结论	24
6.1 LPPL Confidence指数的应用分析	24
6.2 后续研究方向	25

目 录	2
7 附录	25
7.1 LPPL模型的拟合	25
7.2 禁忌(Tabu)搜索算法	27
8 参考文献	27

第一部分 高管增持事件策略

[高管增持事件选股：追踪真正的聪明钱，丁鲁明，中信建投，2016.08.03]

1 高管增持事件概述

高管增持事件是上市公司董事、监事和高级管理人员利用自有资金在二级市场增持自己公司的行为。高管大额增持通常意味着其看好公司未来的发展，或者认为公司股票受到低估，也表明公司大概率内不存在很大的潜在利空，改善了市场对于上市公司的预期，通过一段时间的持续发酵，在未来的一段时间内，通常能够显著地跑赢市场。

1.1 高管增持事件公告日

根据规定，上市公司董事、监事和高级管理人员所持本公司股份发生变动的，应当自该事实发生之日起2个交易日内，向上市公司报告并由上市公司在证券交易所网站进行公告。深圳交易所的上市公司一般满足这种情况，Wind将持股变动日后的第1个交易日作为公告日期。而上海证券交易所的上市公司高管发生持股变动时，由上市公司在交易所网站上自行填报，填报日就是官网上显示的填报日期，但是不能明确肯定填报发生在盘前还是盘后，所以我们将填报日期的后1个交易日作为分析的关键时点。

1.2 高管增持事件最优投资策略

根据上面的分析，在当日高管增持金额达到300万元以上时，可以认为是有效增持。同时，高管增持事件的最优持有区间为公告日T+1到公告日后50个交易日。因此，我们可以构建策略：在公告日T+1收盘时买入增持金额大于300万的股票，公告日以后50个交易日卖出。该策略的组合均剔除掉交易日一字板涨跌停和停牌的股票。

- (i) 在当日高管增持金额达到300万元以上时，可以认为是有效增持。
- (ii) 在公告日T+1收盘时买入增持金额大于300万的股票
- (iii) 在公告日T+50日卖出持有股票
- (iv) 剔除掉交易日一字板涨跌停和停牌的股票

第二部分 Log-Periodic Power Law模型

2 概述

LPPL全称为对数周期幂律(Log - Periodic Power Law)，此模型源自于地球物理中地震与材料断裂等临界现象的理论模型，后被物理学家Sornette应用于预测金融市场的泡沫形成与破裂，即研究金融市场的拐点。

LPPL模型描述的现象是基于交易者之间的相互模仿，这些局部相互作用可形成正反馈，从而导致泡沫和反泡沫的产生，因此该模型可用于金融泡沫和反泡沫的建模和预测。金融市场反泡沫表现在价格演化中，即价格演化呈现出对数周期性振荡且振荡周期不断延长。金融泡沫表现为振荡周期不断缩短。

3 模型拟合

对于在 $[t_1, t_2]$ 区间内的时间序列 p_t , LPPL模型认为其服从

$$\ln p_t = A + B(t_c - t)^m + C(t_c - t)^m \cos[\omega \ln(t_c - t) + \phi] \quad (1)$$

t_c 是临界时间(泡沫破裂时间); A 是指若泡沫持续到临界时间 t_c 时, p_t 能达到的价格; $B < 0$ 是指在临界时间前价格是向上的加速过程, $B > 0$ 是指在临界时间前价格是向下的加速过程; C 是围绕指数增长的一个波动幅度量值, 量化对数周期震动; m 是幂指数, 衡量价格上涨的加速程度; ω 是泡沫期波动的角频率; ϕ 是周期波动的初相位。

公式(1)第二项 $B(t_c - t)^m$ 幂律项描述了价格的加速来自正向反馈机制, 第三项 $C(t_c - t)^m \cos[\omega \ln(t_c - t) + \phi]$ 可以看做是对超指数行为的修正

LPPL模型存在两个显著特征

- 对数周期震荡, 在线性尺度下, 越接近临界时间, 振荡频率越快。但是在对数尺度下, 振荡频率为常数
- 幂律增长, 即超指数增长, 即价格的增长率不是常数, 而是单调递增

关于模型参数的有效约束条件在章节(4.1)中有进一步的讨论。

模型的拟合 经过降维变换, 优化问题的参数最终从七个变成了三个, 即 (t_c, m, ω) , 大大减小了拟合的复杂度。降维和拟合的具体方法见章节(7.1)。

拟合过程中有两个值得注意的问题

- 时间序列 t 使用的是整数序列, 即 $t = 1, 2, 3, \dots$, 表示第几个交易日(如果使用日频数据)或者表示第几个分钟(如果使用分钟数据)
- 拟合起始时间点的选取对最终结果 t_c 影响较大, 建议选取明显的低点或者高点, 作为泡沫或者反泡沫估算的最佳起始点

4 LPPL Confidence 和 Trust 指数

4.1 LPPL Confidence

*LPPL Confidence*和*LPPL Trust*是Sornette于2015年提出的判断泡沫破裂点的新方法。

计算方法 *LPPL Confidence*是指不同数据长度的窗口下有多少比例是有效的拟合。步骤为:

- (i) 对于每一个时间点(拟合终点) t_2 , 选取不同长度的时间序列进行模型拟合。不断往前延伸拟合起点 t_1 ($t_1 < t_2$), 得到若干个时间区间。
- (ii) 对于每一组时间序列进行LPPL模型拟合, 拟合出的结果根据Sornette提出的表1中过滤条件Filtering Condition 1进行过滤。

[

表1中约束条件的解读

- 对 m 的约束源自于模型推导中hazard rate小于等于1的要求, 以及拟合经验。也有算法设置成 $0 \leq m \leq 1$
- ω 的约束条件是确保对数周期的波幅不至于过大或者过小

- t_c 的约束条件是指转折点在拟合终点附近, 如果 t_c 离拟合终点太远则警示意义很小, 可以忽略
- Damping比率的约束源自对于hazard rate非负的要求
- 相对误差项则要求每一点的拟合误差不差过一定程度, 比如5%

]

(iii) 统计样本中满足过滤条件的时间窗口占有所有时间窗口的比例, 即为时间点 t_2 的LPPL Confidence指数。

指数含义 指数值越大代表时间窗口内有越多的不同长度的价格序列为有效拟合, 因此该时间点上泡沫生成的判断可信度越高; 指数值越小代表时间窗口内有越少的不同长度的价格序列为有效拟合, 因此该时间点上泡沫生成的判断可信度也越低。

4.2 LPPL Trust

LPPL Trust是为了衡量残差项对于拟合结果的影响。步骤为:

- (i) 每一个不同长度的时间窗口 $[t_1, t_2]$ 先进行一次拟合, 获取该时间窗口下对应的残差序列。然后将残差序列 ε_t 重新随机排列并与模型拟合后估计的对应价格序列 \hat{p}_t 相加合成一组新的时间序列 p_t^* 。

[

定义 $f(t)$ 为 t 到 t 的随机映射函数, 一次resampling可以定义为 $p_{t_i}^* = \hat{p}_{t_i} + \varepsilon_{f(t_i)} = \hat{p}_{t_i} + \varepsilon_{t_j}$

]

- (ii) 对新合成的时间序列 p_t^* 重新使用LPPL模型拟合, 拟合出的结果根据表1中的过滤条件Filtering Condition 2进行判断是否为有效拟合。
- (iii) 对同一个时间窗口重复reampling100次, 统计新的时间序列中有效拟合比例。
- (iv) 统计出时间点 t_2 对应的所有时间窗口计算结果的中位数, 即为该时间点的LPPL Trust 指数。

指数含义 LPPL Trust指数衡量了LPPL理论模型与实际价格序列的相似程度, 0代表完全不符合, 1代表完全符合; 通过经验数据来说, 一般LPPS Trust指数值大于5%, 表明目前的价格序列不稳定且存在泡沫破灭的可能性。

Item	Notation	Search Space	Filtering Condition 1	Filtering Condition 2
3 Nonlinear Parameters	m	$[0, 2]$	$[0.01, 1.2]$	$[0.01, 0.99]$
	ω	$[1, 50]$	$[2, 25]$	$[2, 25]$
	t_c	$[t_2, t_2 + 0.2dt]$	$[t_2, t_2 + 0.1dt]$	$[t_2, t_2 + 0.1dt]$
Damping	$\frac{m B }{\omega C }$	-	$[0.8, +\infty)$	$[1, +\infty)$
Relative Error(ε)	$\frac{p_t - \hat{p}_t}{\hat{p}_t}$	-	$[0, 0.05]$	$[0, 0.2]$

表 1: LPPL Filtering Table

5 模型应用

5.1 LPPL Confidence

本节应用(4.1)所述的基本方法计算不同标的的LPPL Confidence指数

- LPPL模型拟合所使用的参数 m , ω 和 t_c 搜索空间为表1 中搜索空间 (*Search Space*) 列中给出的范围; Confidence指数计算所使用的过滤条件对应表1中过滤条件1(*Filtering Condition 1*)
- 每一个标的指数序列选取三类时间窗口集合($t_{window} = t_2 - t_1$)分别进行Confidence指数的计算, 以进行比较。具体区间分为小区间($t_{window,min} = 50, t_{window,max} = 100$), 中等区间($t_{window,min} = 150, t_{window,max} = 250$)和长区间($t_{window,min} = 250, t_{window,max} = 500$)三类, 时间间隔为5(或者1) 个交易日
- 保持其他过滤条件不变的情况下, 计算了不同相对误差条件下(分为 $\varepsilon < 5\%$, $\varepsilon < 10\%$ 和 $\varepsilon < 20\%$)的Confidence指数

下列各图中, 蜡烛线展示的是标的指数在测试区间内的走势, 而图中红线则代表了区间内各时间点上Confidence指数值。

5.1.1 上证综指 (000001.SH)

图1-图7展示了对标的上证综合指数 (000001.SH) 使用(5.1)描述的分析方法进行Confidence指数计算的结果。测试数据的时间范围为2014年9月1日至2016年3月1日, 数据频率为日频, 共计363 个交易日。

图1、2和3分别展示的是短周期窗口下($t_{window,min} = 50, t_{window,max} = 100$), 时间间隔为1个交易日($\Delta t = 1$), 过滤条件相对误差 (*Relative Error*) 分为三种情况: $\varepsilon < 5\%$ 、 $\varepsilon < 10\%$ 和 $\varepsilon < 20\%$, 其他条件对应表1中过滤条件1(*Filtering Condition 1*)。



图 1: 蜡烛线: 上证综指(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年9月1日至2016年3月1日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1中过滤条件1(*Filtering Condition 1*)



图 2: 蜡烛线: 上证综指(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年9月1日至2016年3月1日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 10\%$, 其余过滤条件对应表1中过滤条件1(*Filtering Condition 1*)



图 3: 蜡烛线: 上证综指(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年9月1日至2016年3月1日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1中过滤条件1(*Filtering Condition 1*)

图4和5分别展示的是中等周期窗口($t_{window,min} = 150, t_{window,max} = 250$) 下, 时间间隔为2个交易日($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表1 中过滤条件1(*Filtering Condition 1*)。



图 4: 蜡烛线: 上证综指(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年9月1日至2016年3月1日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1中过滤条件1(Filtering Condition 1)

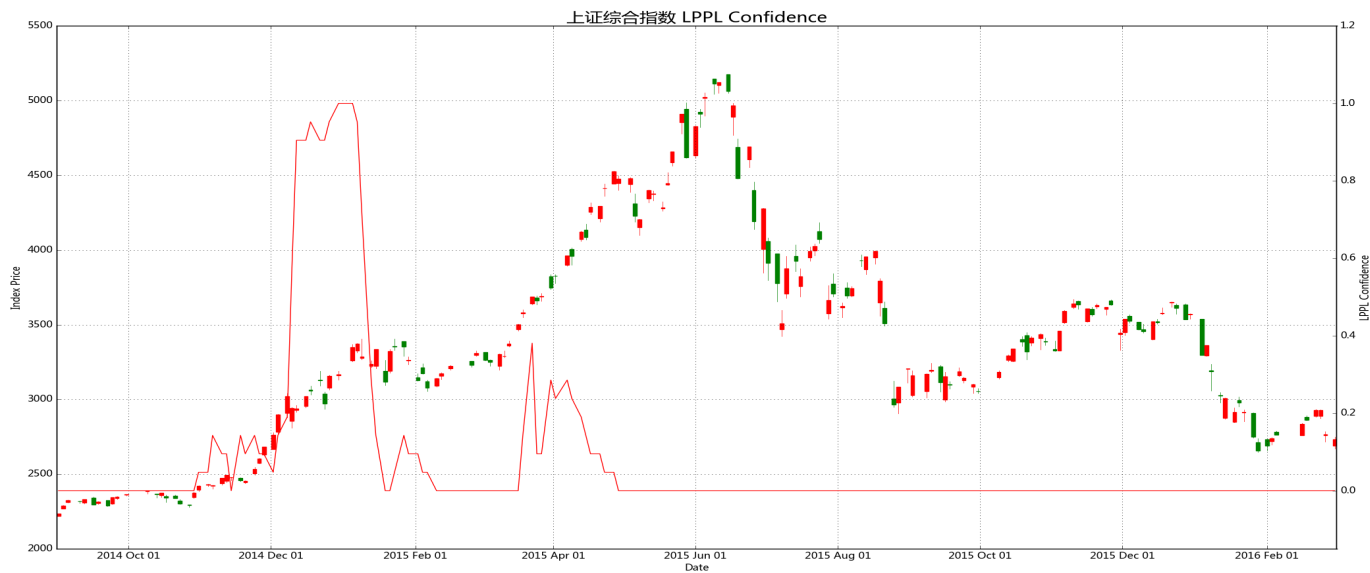


图 5: 蜡烛线: 上证综指(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年9月1日至2016年3月1日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1中过滤条件1(Filtering Condition 1)

图6和7分别展示的是长周期窗口($t_{window,min} = 250, t_{window,max} = 500$) 下, 时间间隔为2个交易日($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)。



图 6: 蜡烛线: 上证综指(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年9月1日至2016年3月1日; 时间窗口: 250-500; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1中过滤条件1(*Filtering Condition 1*)



图 7: 蜡烛线: 上证综指(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年9月1日至2016年3月1日; 时间窗口: 250-500; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1中过滤条件1(*Filtering Condition 1*)

结果分析 从不同情形下Confidence指数计算的结果与上证指数的走势比较, 可以看到

- 根据长周期窗口($t_{window,min} = 250, t_{window,max} = 500$)计算的Confidence指数在2015年6月初有明显异常的正值,能够正确判断指数拐点的形成。其他周期窗口计算的指数除了在2015年6月以外,在2014年12月-2015年1月期间也有明显异常的正值,而且强度比6月份更大,从实际走势来看,2014年12月-2015年1月期间发出的拐点信号是错误的。

另外短周期窗口下在2015年9月还Confidence指数出现了一定程度的负值,预测市场将反转向上(见图(2)和图(3)),后来走上证指数的走势也证明了此类信号的错误。

从信号的准确率上来说,长周期窗口计算的Confidence指数效果更好

- 过滤条件中相对误差项 ε 对Confidence指数的形态影响不大,改变相对误差项的阈值并不能明显提高Confidence指数对拐点预测的准确率
- 长周期窗口计算的Confidence指数目前并没有发出市场见底反转向上的信号,说明从2015年6月以后至今的下跌不符合指数式下跌并于近期可能反转的形态
- 长周期窗口下计算的Confidence指数绝对值低于20%时,信号的准确率较低。

5.1.2 南华螺纹钢指数 (NH0016.NHF)

图8至图11展示了对标的南华螺纹钢指数 (NH0016.NHF) 使用(5.1)描述的分析方法进行Confidence指数计算的结果,测试数据的时间范围为2014年11月25日至2016年5月24日,数据频率为日频,共计366个交易日。

图8和图9分别展示的是短周期窗口($t_{window,min} = 50, t_{window,max} = 100$)下,时间间隔为1个交易日,过滤条件相对误差 (Relative Error) 分为两种情况: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果,其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)。



图 8: 蜡烛线: 南华螺纹钢指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)



图 9: 蜡烛线: 南华螺纹钢指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1中过滤条件1(Filtering Condition 1)

图10和图11分别展示的是中等周期窗口($t_{window,min} = 150, t_{window,max} = 250$)下, 时间间隔为2个交易日($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)。



图 10: 蜡烛线: 南华螺纹钢指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1中过滤条件1(Filtering Condition 1)

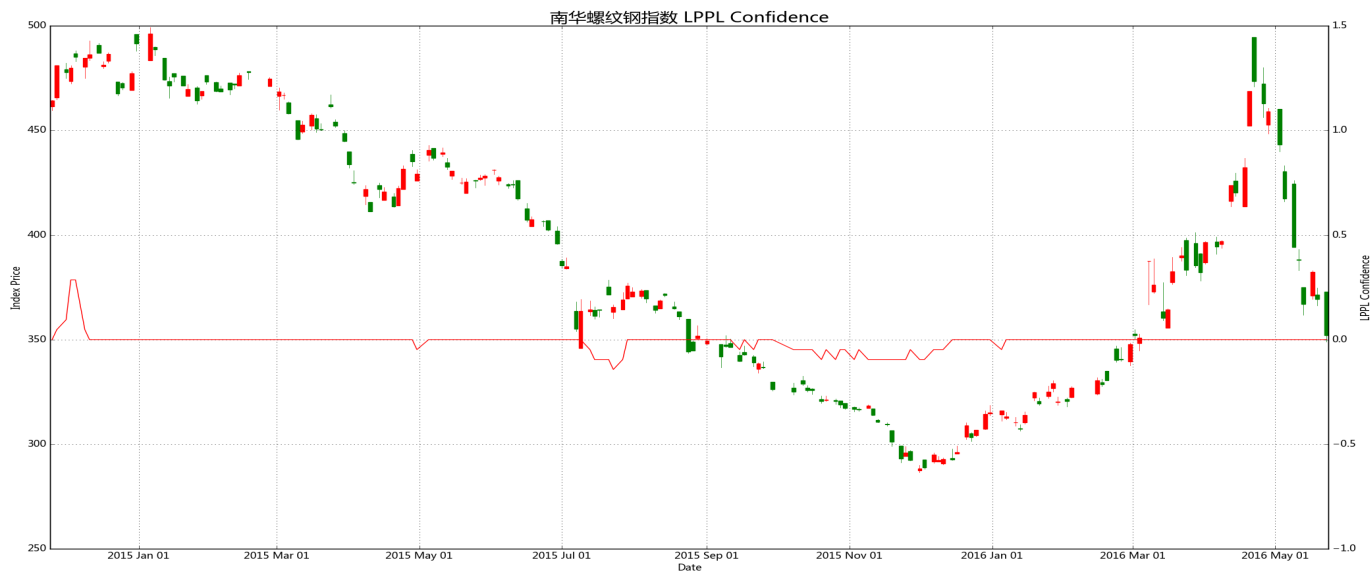


图 11: 蜡烛线: 南华螺纹钢指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

图12和图13分别展示的是长周期窗口($t_{window,min} = 250, t_{window,max} = 500$)下, 时间间隔为2个交易日($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)。



图 12: 蜡烛线: 南华螺纹钢指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 250-500; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1中过滤条件1(Filtering Condition 1)



图 13: 蜡烛线: 南华螺纹钢指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 250-500; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

结果分析 通过比较不同情形下Confidence指数计算的结果与南华螺纹钢指数 (NH0016.NHF)的走势比较, 可以看到:

- 根据短周期窗口($t_{window,min} = 50, t_{window,max} = 100$)所计算的Confidence指数在2015年11月末有明显的负值, 可以正确判断指数拐点。此外, 短周期窗口下在2015年4月和2015年7月初也有明显异常的负值, 强度弱于11月份的信号。但从指数的实际走势来看, 2015年4月至2015年7月之间发出的拐点信号都是错误的。

此外, 短周期窗口且相对误差项 $\varepsilon < 20\%$ 条件下在2016年3月至4月底, Confidence指数出现了连续正值信号, 指示市场出现正泡沫(见图17)。而南华螺纹钢指数在4月21日达到最高点后连续下跌。Confidence指数正确判断了该次拐点的形成。

- 中等周期窗口、长周期窗口计算的Confidence指数分别在在2014年11月出现的较小幅度的正值, 2015年5月至6月出现的小幅负值被螺纹钢指数的走势证明所发出信号为错误的。
- 过滤条件中相对误差项 ε 在短周期窗口下对生成负信号强度的影响不大, 但完全过滤了生成的正泡沫信号。而在中等周期窗口、长周期窗口下, 改变相对误差项阈值并没有明显改善Confidence信号对拐点预测的准确率。

从信号的准确率和指示趋势拐点的成功率来看, 短周期窗口, 相对误差 $\varepsilon < 20\%$ 条件下Confidence指数绝对值超过50%时, 计算的指数准确率较高。

5.1.3 南华豆粕指数 (NH0015.NHF)

图14至图17展示了对标的南华豆粕指数 (NH0015.NHF) 使用LPPL模型拟合所得出的结果, 测试数据的时间范围为2014年11月25日至2016年5月24日, 数据频率为日频, 共计366个交易日。

图14和图15分别展示的是短周期窗口($t_{window,min} = 50, t_{window,max} = 100$) 下, 时间间隔为1个交易日, 过滤条件相对误差 (*Relative Error*) 分为两种情况 $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(*Filtering Condition 1*)。



图 14: 蜡烛线: 南华豆粕指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1 中过滤条件1(*Filtering Condition 1*)



图 15: 蜡烛线: 南华豆粕指数(左轴); 红线: LPPL Confidence 指数(右轴); 时间范围: 2014 年 11 月 25 日至 2016 年 5 月 24 日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表 1 中过滤条件 1 (Filtering Condition 1)

图 16 和图 17 分别展示的是中等周期窗口 ($t_{window,min} = 150, t_{window,max} = 250$) 下, 时间间隔为 2 个交易日 ($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表 1 中过滤条件 1 (Filtering Condition 1)。



图 16: 蜡烛线: 南华豆粕指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

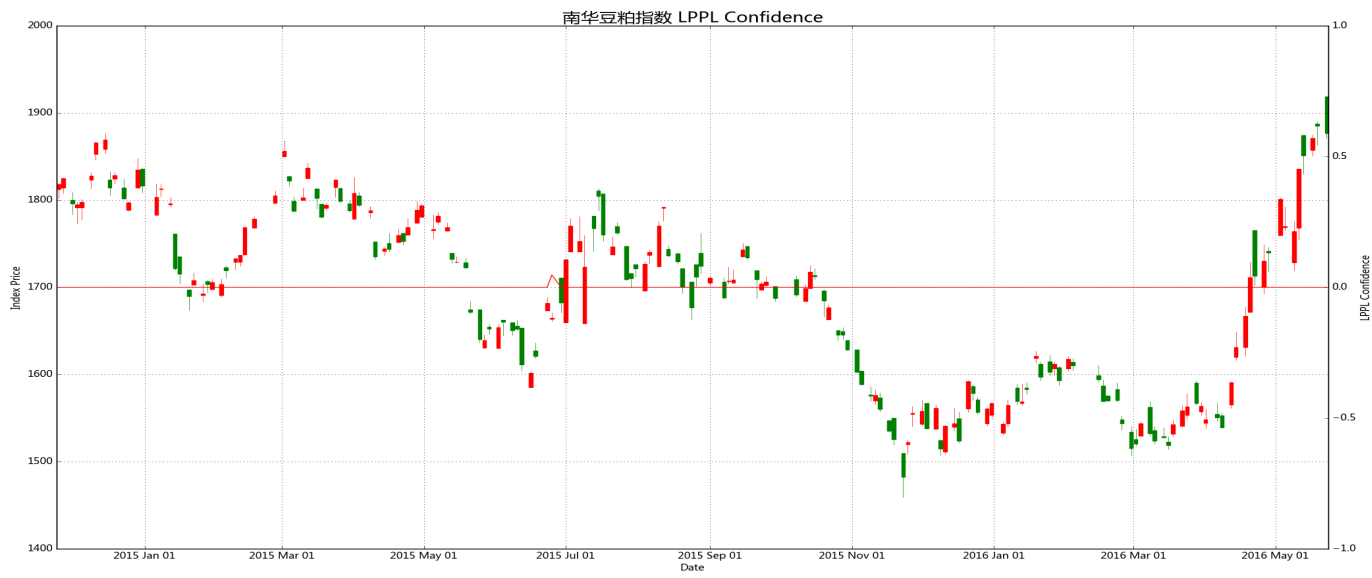


图 17: 蜡烛线: 南华豆粕指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

图18和图19分别展示的是长周期窗口($t_{window,min} = 250, t_{window,max} = 500$)下, 时间间隔为2个交易日($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)。



图 18: 蜡烛线: 南华豆粕指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 250-500; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

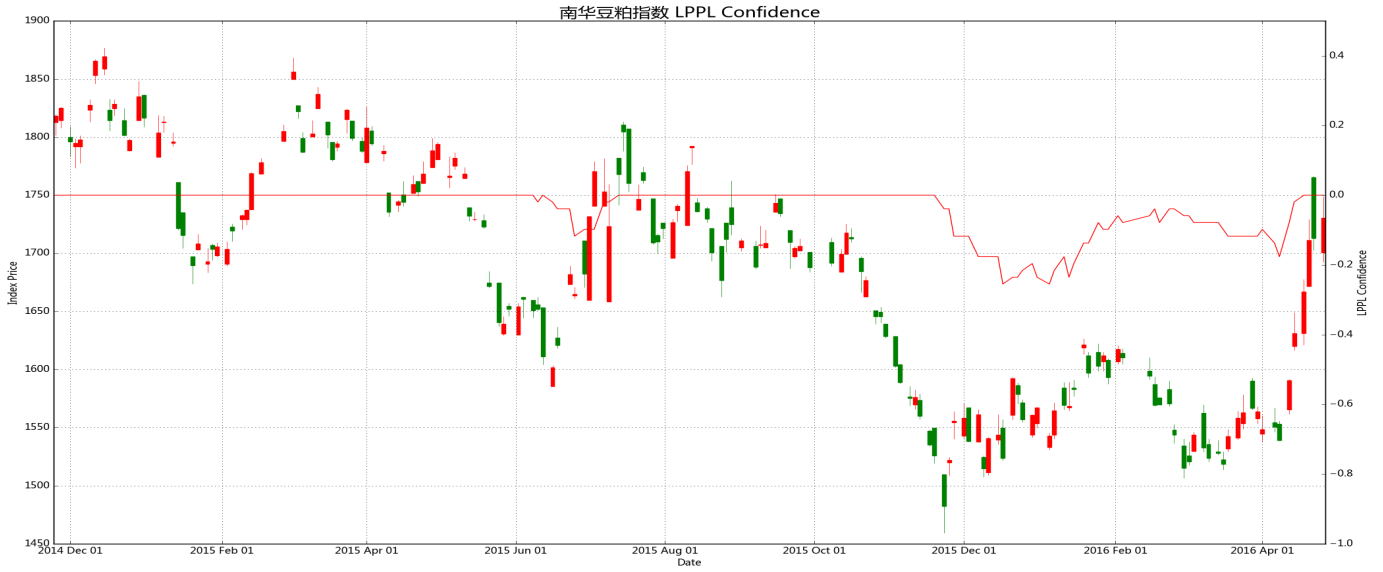


图 19: 蜡烛线: 南华豆粕指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 250-500; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

结果分析 通过比较不同情形下Confidence指数计算的结果与南华豆粕指数 (NH0015.NHF)的走势, 可以看到:

- 根据短周期窗口($t_{window,min} = 50, t_{window,max} = 100$)所计算的Confidence指数分别在2015年1月末、2015年5月末和2015年11月中出现了明显异常的负值信号, 其中2015年11月中旬出现的信号强度最强, 信号绝对值大于50%。而标的指数相应区间和随后期间的实际走势进行说明仅第三次(2015年11月中旬)的信号指示的止跌回升信号正确。

另外, 短周期窗口下在2016年5月中起至近日(2016年5月24日), Confidence指数出现了连续正值信号(图14和图15), 其信号的正确性仍有待后续指数走势进行确认。

- 中等周期窗口在测试区间内没有任何出现异常值, 而长周期窗口计算的Confidence指数在2015年6月末出现的较小幅度的信号(图18), 但该信号指示的反转信号是错误的。

综上, 从信号的准确率来看, 短周期窗口下计算的Confidence指数出现异常且指数绝对值超过50%时准确率较高。

- 此外, 过滤条件中相对误差项 ε 在短周期窗口下对Confidence指数的形态影响较小, 使用不同的相对误差项的阈值并没有显著提高Confidence指数对拐点预测的准确率。

5.1.4 南华焦煤指数 (NH0029.NHF)

图20至图25展示了对标的南华焦煤指数 (NH0015.NHF) 使用LPPL模型拟合所得出的结果,

图20至图21测试数据的时间范围为2014年11月25日至2016年5月24日, 数据频率为日频, 共计366个交易日。图22至图23测试数据的时间范围为2015年9月15日至2016年5月23日, 数据频率为日频, 共计168个交易日。

图20和图21分别展示的是短周期窗口($t_{window,min} = 50, t_{window,max} = 100$) 下, 时间间隔为1个交易日($\Delta t = 1$), 过滤条件相对误差 (*Relative Error*) 分为两种情况: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$, 其他条件对应表1中过滤条件1(*Filtering Condition 1*)。



图 20: 蜡烛线: 南华焦煤指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1 中过滤条件1(*Filtering Condition 1*)



图 21: 蜡烛线: 南华焦煤指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2014年11月25日至2016年5月24日; 时间窗口: 50-100; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

图22和图23分别展示的是中等周期窗口($t_{window,min} = 150, t_{window,max} = 250$)下, 时间间隔为2个交易日($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)。

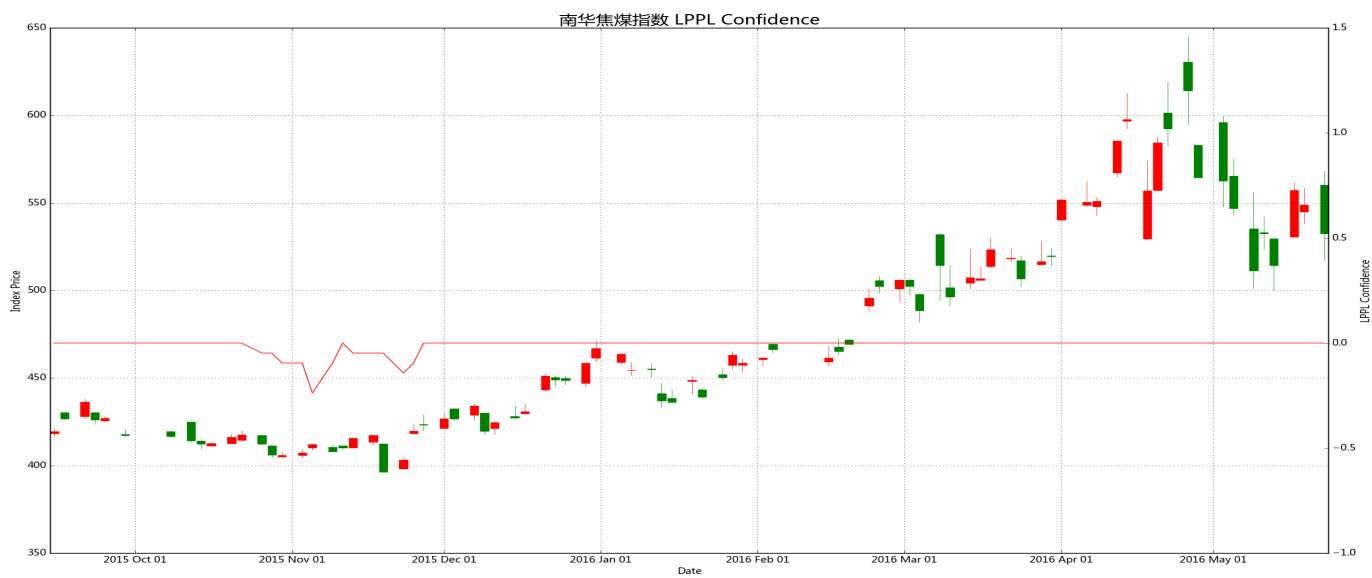


图 22: 蜡烛线: 南华焦煤指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2015年9月15日至2016年5月23日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

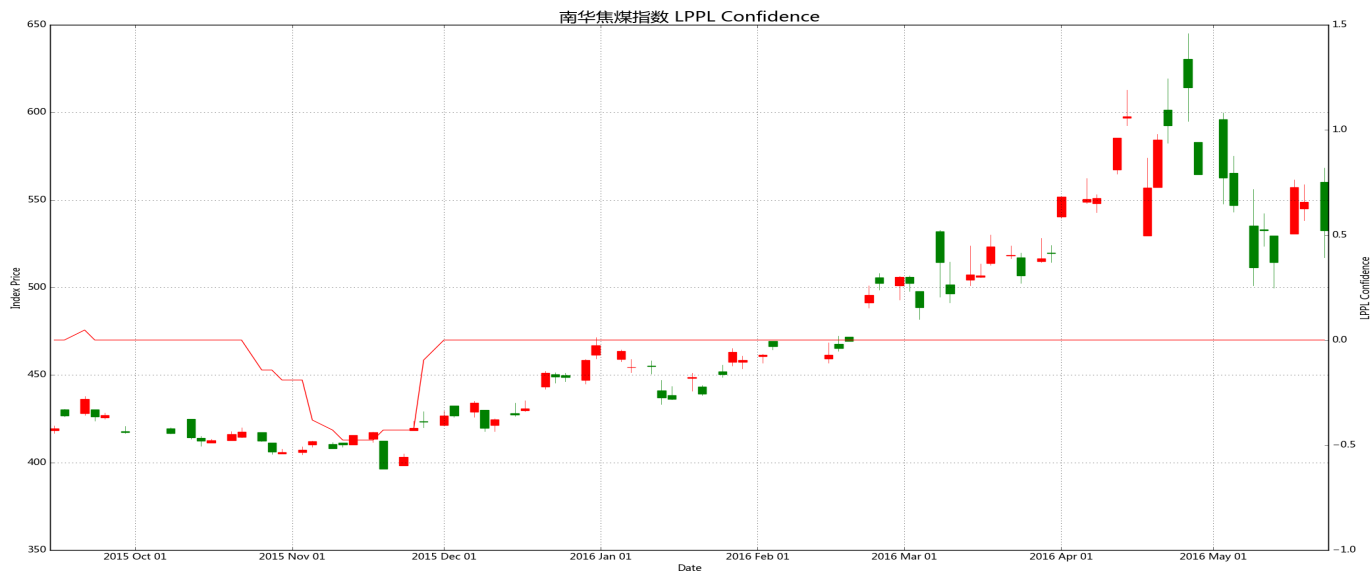


图 23: 蜡烛线: 南华焦煤指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2015年9月15日至2016年5月23日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

图24和图25分别展示的是长周期窗口($t_{window,min} = 250, t_{window,max} = 500$)下, 时间间隔为2个交易日($\Delta t = 2$), 过滤条件相对误差 (Relative Error) 采用标准: $\varepsilon < 5\%$ 和 $\varepsilon < 20\%$ 的结果, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)。

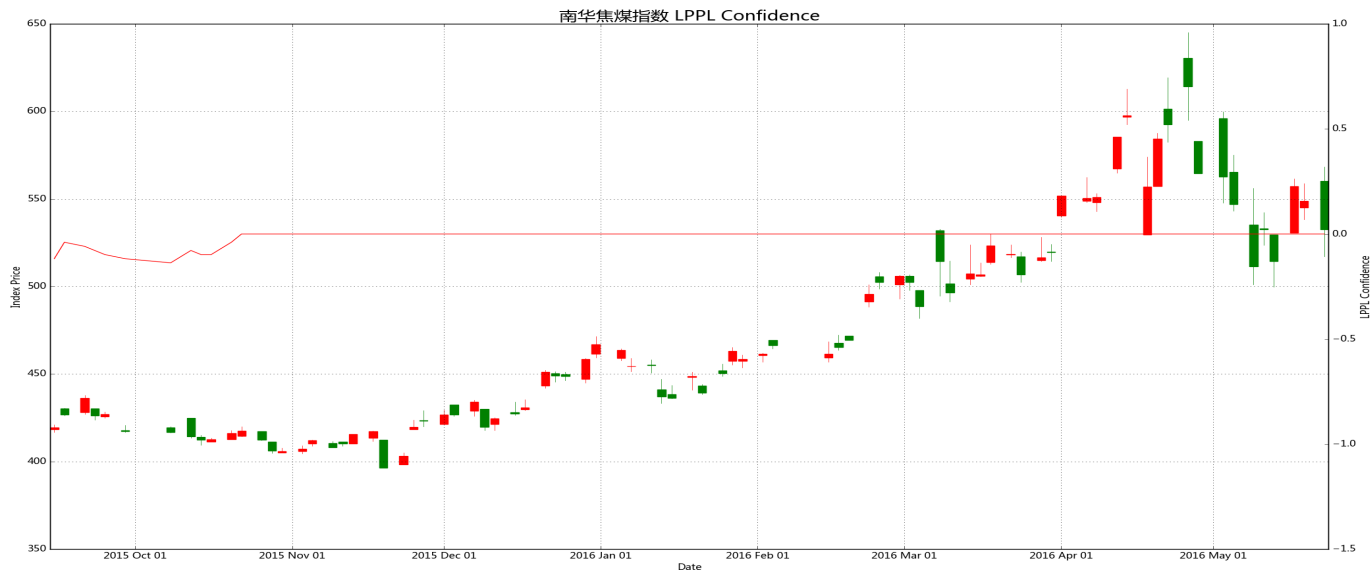


图 24: 蜡烛线: 南华焦煤指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2015年9月15日至2016年5月23日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 5\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

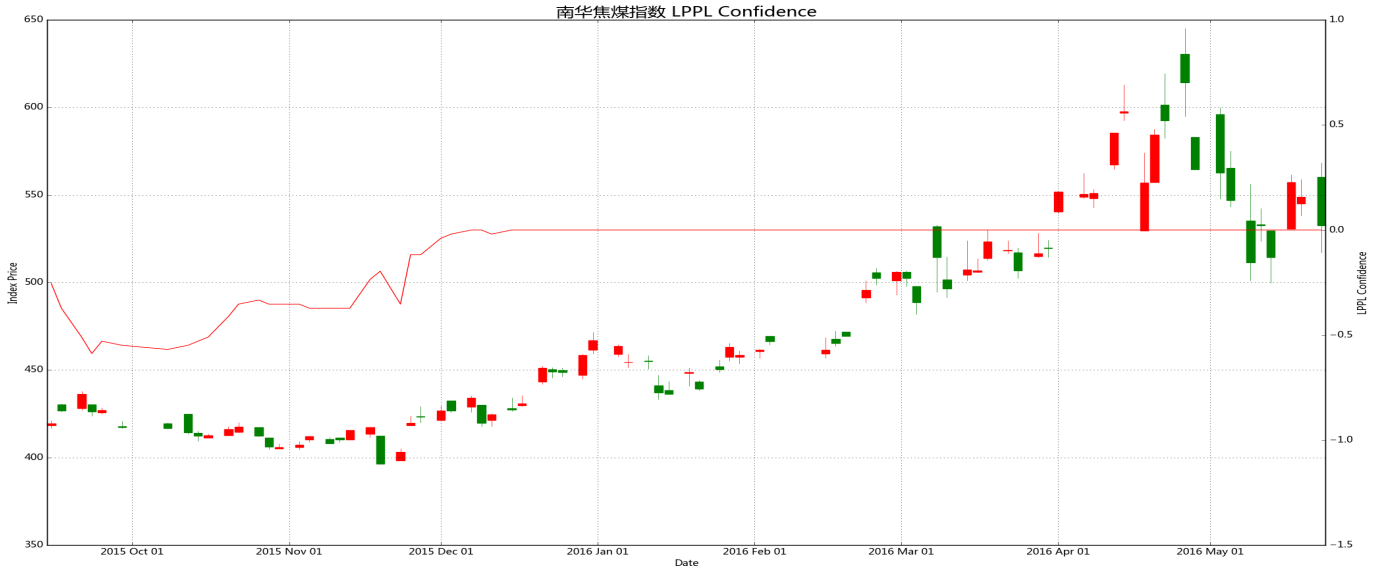


图 25: 蜡烛线: 南华焦煤指数(左轴); 红线: LPPL Confidence指数(右轴); 时间范围: 2015年9月15日至2016年5月23日; 时间窗口: 150-250; 过滤条件: $\varepsilon < 20\%$, 其余过滤条件对应表1 中过滤条件1(Filtering Condition 1)

结果分析 通过比较不同情形下Confidence指数计算的结果与南华焦煤指数 (NH0029.NHF)的走势, 可以看到:

- 中等周期窗口下($t_{window,min} = 150, t_{window,max} = 250$)所计算的Confidence指数在2015年11月初至11月末出现了多次明显的负值信号, 指示焦煤指数即将止跌回升, 而市场验证了该信号, 焦煤指数的拐点在区间内形成。长周期窗口计算的指数在2015年10月底出现了连续正值, 从实际走势来看, 发出的拐点信号略早于实际拐点。
 - 短周期窗口下2015年3月末至4月末Confidence指数出现了多次正值信号(图20和图21), 预测市场将反转向下。目前焦煤指数仅小幅下跌, 信号的正确性仍有待市场验证。
- 从信号的准确率来说, 目前中等周期窗口下计算的Confidence指数效果最好。
- 过滤条件中相对误差项 ε 在各周期窗口下对Confidence指数的形态影响较小, 不同的相对误差项的阈值并没有显著提高Confidence指数对拐点预测的准确率。

5.2 LPPL Trust

因所涉计算量较大, 故暂未应用。

6 结论

6.1 LPPL Confidence指数的应用分析

从(5.1)讨论的针对不同标的以及不同过滤条件的LPPL Confidence指数的比较可以得出如下结论

- 上证综指：使用长周期($t_{window,min} = 250, t_{window,max} = 500$)进行Confidence指数的计算，信号的正确率较高。使用中短周期窗口进行计算Confidence指数效果较差

[Confidence指数较好的预测了2015年6月的拐点。对于随后至今的下跌，未发出反转信号（至少说明下跌过程不符合指数式下跌并反转的形态）]

- 螺纹钢，豆粕和焦炭期货指数：由于此类标的指数的投机气氛较浓，趋势较为明显且持续周期短，经测试短周期($t_{window,min} = 50, t_{window,max} = 100$)为效果最佳的Confidence指数计算周期

[Confidence指数较好预测了螺纹钢指数2015年12月的最低点和2016年5月的最高点，效果最好；对于豆粕指数也有非常类似的预测，但是2016年5月底的高点还有待验证；对于焦煤指数，Confidence指数没有给出最低点反转的信号，但是对于2016年5月的最高点有泡沫反转信号]

- 在放宽的相对误差 $\varepsilon < 20\%$ 条件下进行有效拟合的判断比原始误差条件 $\varepsilon < 5\%$ (其他条件对应表1 中过滤条件1(Filtering Condition 1))，效果更好

6.2 后续研究方向

下一步的研究方向为继续优化LPPL模型算法，将计算过程中现有使用的穷举法改为Levenberg-Marquardt最优化算法，根据含参数的残差平方和函数的梯度及二阶导数近似估计残差平方和函数，从初始解开始不断逼近最优解来减少LPPL 拟合所用的时间，并将Sornette提出的LPPL Trust指数应用市场泡沫的判断中去。

7 附录

7.1 LPPL模型的拟合

根据公式(1),LPPL模型有7个参数需要估计，即 $A, B, C, t_c, m, \omega, \phi$ ；其中 t_c, m, ω, ϕ 为非线性参数， A, B, C 为线性参数。为降低拟合参数的数量，同时为确保参数估计的稳定性，需要首先将线性参数用非线性参数估计值表达出来，从而只需要估计非线性参数即可。有诸多论文提出了拟合方法，此处引用了Sornette于2011年提出的拟合改进方法。¹

首先，令 $y_t = \ln p_t, C_1 = C \cos \phi, C_2 = C \sin \phi, f_t = (t_c - t)^m, g_t = (t_c - t)^m \cos [\omega \ln (t_c - t)],$
 $h_t = (t_c - t)^m \sin [\omega \ln (t_c - t)],$ 公式(1) 可改写成 $y_t = A + B f_t + C_1 g_t + C_2 h_t$

[公式(1)等价于

$$\ln p_t = A + B (t_c - t)^m + C (t_c - t)^m \cos [\omega \ln (t_c - t)] \cos \phi + C (t_c - t)^m \sin [\omega \ln (t_c - t)] \sin \phi$$

令

$$C_1 = C \cos \phi, C_2 = C \sin \phi$$

所以(1)可改写为

$$\ln p_t = A + B (t_c - t)^m + C_1 (t_c - t)^m \cos [\omega \ln (t_c - t)] + C_2 (t_c - t)^m \sin [\omega \ln (t_c - t)] \quad (2)$$

¹见A Stable and Robust Calibration Scheme of LPPL Model, Vladimir Filimonov, Didier Sornette, 2011

]

其次, 假设区间内有 N 个 p_t 取值点, LPPL模型的拟合可以整理为最小化残差和的优化问题

$$\min_{A, B, C_1, C_2, t_c, m, \omega} S(A, B, C, t_c, m, \omega) = \sum_{i=1}^N (y_{t_i} - A - Bf_{t_i} - C_1g_{t_i} - C_2h_{t_i})^2$$

再其次, 将线性参数 A, B, C_1, C_2 用非线性参数 t_c, m, ω 表达出来, 那么优化问题可以转化为

$$\min_{A, B, C_1, C_2, t_c, m, \omega} S(A, B, C, t_c, m, \omega) \equiv \min_{t_c, m, \omega} S_1(t_c, m, \omega)$$

其中

$$S_1(t_c, m, \omega) = \min_{A, B, C_1, C_2} S(A, B, C_1, C_2, t_c, m, \omega) \quad (3)$$

那么优化问题(3)等价于

$$\{\hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2\} = \arg \min_{A, B, \hat{C}_1, \hat{C}_2} S(A, B, C, t_c, m, \omega) = \arg \min_{A, B, \hat{C}_1, \hat{C}_2} \sum_{i=1}^N (y_{t_i} - A - Bf_{t_i} - C_1g_{t_i} - C_2h_{t_i})^2 \quad (4)$$

联立公式(4)的一阶条件可以得到关于 $\{\hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2\}$ 的方程组, 通过LU方法或者Cholesky方法可以得到 $\{\hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2\}$ 的解。

[

$\{\hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2\}$ 应该满足公式(4)的一阶条件, 联立方程之后可得到如下方程组

$$\begin{pmatrix} \sum_{i=1}^N y_{t_i} \\ \sum_{i=1}^N y_{t_i} f_{t_i} \\ \sum_{i=1}^N y_{t_i} g_{t_i} \\ \sum_{i=1}^N y_{t_i} h_{t_i} \end{pmatrix} = \begin{pmatrix} N & \sum_{i=1}^N f_{t_i} & \sum_{i=1}^N g_{t_i} & \sum_{i=1}^N h_{t_i} \\ \sum_{i=1}^N f_{t_i} & \sum_{i=1}^N f_{t_i}^2 & \sum_{i=1}^N f_{t_i} g_{t_i} & \sum_{i=1}^N f_{t_i} h_{t_i} \\ \sum_{i=1}^N g_{t_i} & \sum_{i=1}^N f_{t_i} g_{t_i} & \sum_{i=1}^N g_{t_i}^2 & \sum_{i=1}^N g_{t_i} h_{t_i} \\ \sum_{i=1}^N h_{t_i} & \sum_{i=1}^N f_{t_i} h_{t_i} & \sum_{i=1}^N g_{t_i} h_{t_i} & \sum_{i=1}^N h_{t_i}^2 \end{pmatrix} \begin{pmatrix} A \\ B \\ C_1 \\ C_2 \end{pmatrix} \quad (5)$$

$$\text{令 } b = \begin{pmatrix} A \\ B \\ C_1 \\ C_2 \end{pmatrix}, X = \begin{pmatrix} 1 & f_{t_1} & g_{t_1} & h_{t_1} \\ 1 & f_{t_2} & g_{t_2} & h_{t_2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & f_{t_N} & g_{t_N} & h_{t_N} \end{pmatrix}$$

则公式(2)可以用矩阵形式表示为 $X^T y = (X^T X) b$, 所以可得到 b 的解

$$\hat{b} = (X^T X)^{-1} X^T y \quad (6)$$

]

所以优化问题的参数最终从七个变成了三个, 即 (t_c, m, ω) 。

参数估计²的方法分成两步

²也可以用穷举法, 效率较低

- 使用遍历方法或者禁忌搜索算法(Taboo)³找出10个局部解
- 这10个局部解作为初始值，使用Levenberg-Marquardt非线性最优化算法来拟合曲线，使得 $\sum_{i=1}^N (y_{ti} - \hat{y}_{ti})^2$ 取得最小值

7.2 禁忌(Tabu)搜索算法

算法的伪代码如下

Algorithm 1 禁忌搜索算法

Initialization: $s = s_0$, $sBest = s$, $tabuList = []$

Iteration:

```
1: while stoppingCondition() == false do
2:   candidateList = [];
3:   bestCandidate = null;
4:   for sCandidate ∈ sNeighborhood do
5:     if (not tabuList.contains(sCandidate)) & (fitness(sCandidate) > fitness(bestCandidate)) then
6:       bestCandidate = sCandidate
7:     end if
8:   end for
9:   s = bestCandidate
10:  if fitness(bestCandidate) > fitness(sBest) then
11:    sBest = bestCandidate
12:  end if
13: end while
14: return sBest;
```

8 参考文献

股市泡沫反泡沫研究，国泰君安证券，刘富兵，蒋瑛琨，2012

Everything You Always Wanted to Know about LPPL, *Petr Geraskin*, 2010

A Stable and Robust Calibration Scheme of LPPL Model, *Vladimir Filimonov, Didier Sornette*, 2011

Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash, *Didier Sornette, Guilherme Demos*, 2015.08]

Diagnosing Bubbles in Fixed Income Markets, *Tuncay Michael Irmak, Burgdorf*, 2015.10

³算法见章节(7.2)