

股债 CTA 策略-跨周期与持仓量模型

国投安信期货研究院

王锴 期货投资咨询号 Z0016176

在本篇报告中，我们以量化基本面为核心，结合持仓量模型，形成了多周期结构的策略。其中短周期模型聚焦于市场风格，外部因素，资金面三大高频金融数据板块。在单一类别因子的基本面回测中，我们发现不同类型因子不同时间段表现存在差异。如果采用单因子模型则难以提供稳定收益，这也符合指标在不同行情周期下的有效性规律。将所有因子放入模型中后，我们发现其效果基本要优于单一类别，收益和风险控制上表现都有所提升。长周期模型关注市场预期，聚焦于宏观经济数据等低频指标。主要逻辑是从高维数据中选出有效且相对独立的因子，如果公布的实际金融数据与市场预期存在一定偏差，则部分投资者可能会根据自身投资风格在数据的公布的时间段内进行短期博弈。持仓量模型主要考虑机构多空单持仓量并进行合成，使用机构会员持仓可以为价格提供有效信息，这或是机构研究更专业和资金推动的结果。在 2016-2021 期间，最终策略胜率为 56%。接下来，我们将在以下方向丰富策略：

- 1.在投资方式和投资标的上拓宽。品种池中纳入商品期货合约，做多相对较强的合约，做空相对较弱的合约，灵活调整多空持仓配比。
- 2.嵌套相对独立的策略模型，例如选取稳定的技术面指标组合，协助判断日度和日内的进场时机点，同时起到过滤信号的作用。

一、长周期模型

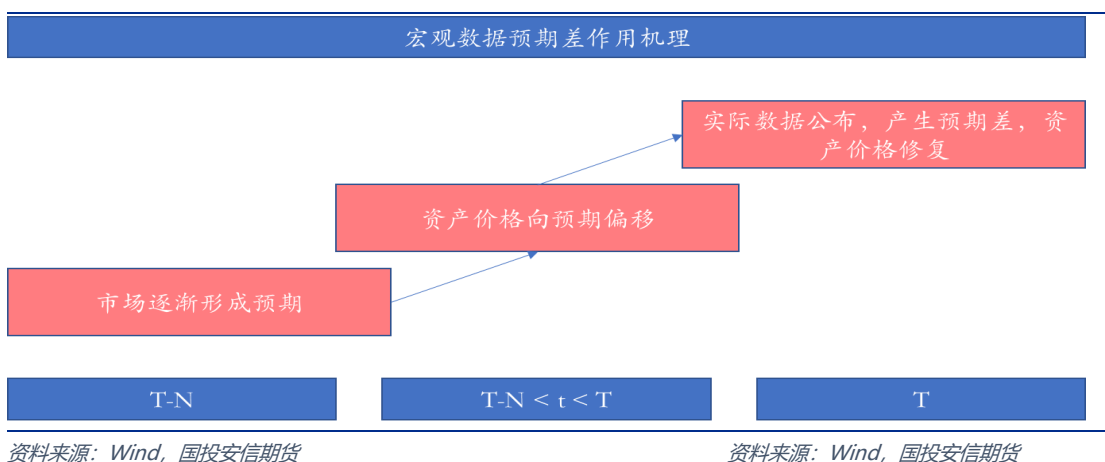
一般情况下，宏观经济金融数据反映的是市场预期，如果公布的实际金融数据与市场预期存在一定偏差，则部分投资者可能会根据自身投资风格在数据的公布的时间段内进行短期博弈。

回顾市场上关于长周期因子的研究偏少，不难发现存在以下问题：1.数据发布时间不固定，且相对滞后，如 CPI、PPI 和 M2 大概在次月 10 号左右公布，而工业增加值、

消费品零售总额和城镇固定资产投资等数据要到次月中旬才会公布。低频宏观数据对于市场的影响取决于数据公布时间，从市场反应的角度来看是不连续的且不对称的。2.如果将低频数据分布到日度的高频数据，可能存在时间点错配的问题，有效性显著降低。

因此，我们对于长周期模型的数据提取机制做了改进，除了定期检测并更新预测变量，当某一预测变量数据更新时，会重新整个模型进行信号预测，产生的信号对应的时间跨度为当日距离下一个标准期限的时间。

图：各合约盘口价差比较(diff(Ask1,Bid1)/

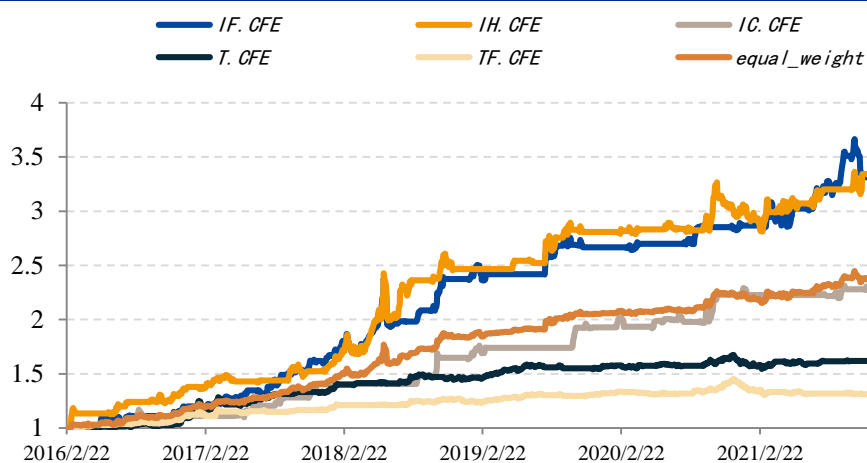


图：长周期模型单因子效果

	预测信号的拟合程度	实际涨跌的胜率
PPI 预测减实际	65.36%	55.31%
社融预测减实际	63.13%	53.07%
固定资产投资预测减实际	53.07%	53.07%
M1 减 M2	60.89%	53.07%
净出口预测减实际	42.46%	52.51%
PMI 预测减实际	53.63%	51.40%
M2 预测减实际	77.09%	50.28%
CPI 预测减实际	48.60%	48.60%
社会消费品零售总额预测减实际	48.04%	46.93%
人民币贷款预测减实际	45.25%	45.25%
工业增加值预测减实际	46.37%	45.25%

资料来源：Wind，国投安信期货

图：组合对冲效果(IF 合约)，横轴为到期前天数



数据来源: wind, 国投安信期货

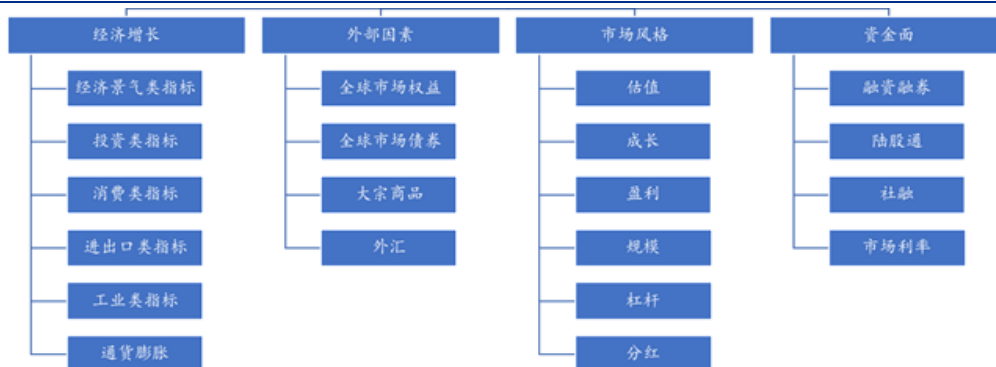


国投安信期货
SDIC ESSENCE FUTURE

资料来源: Wind, 国投安信期货

二、短周期模型

图：基本面因子筛选框架

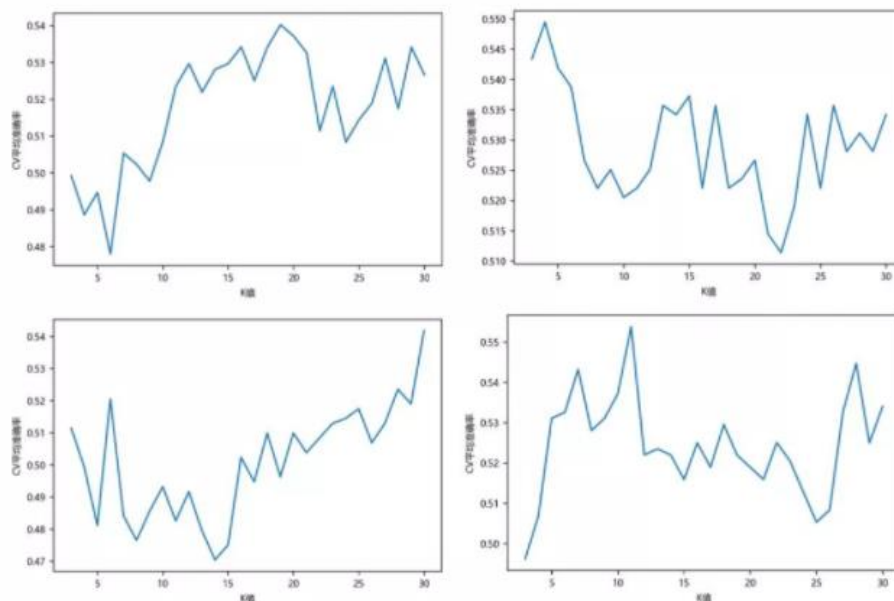


资料来源: Wind, 国投安信期货

我们从经济增长、市场风格，外部因素，资金面等高频金融数据板块中选取了指标，在比较各类机器学习算法后采用 KNN 邻近算法来 KNN 算法属于惰性算法，其特点是不

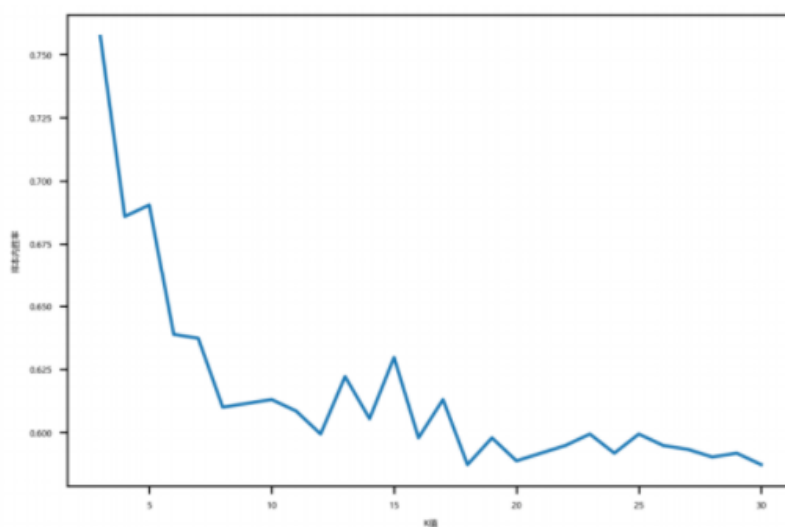
必事先建立全局的判别公式或规则，当新样本需要分类时，根据每个新样本和原样本之间的距离，取最近的 K 个样本点的众数或均值作为新样本的预测值。由于 KNN 算法对于解释变量的类型没有限制，最主要的超参数就是 K，我们运用交叉验证的方法，在样本内确定最优参数 K，力求在拟合和泛化能力上取得平衡。

图：四类因子交叉验证寻找最优参数 K



资料来源：Wind，国投安信期货

图：所有因子在不同 K 值下的样本内胜率



资料来源：Wind，国投安信期货

通过遍历参数，我们使用 $k=20$ 的 KNN 模型，X 使用连续变量，Y 采用二分类法。从结果来看，我们发现使用更多因子的效果虽然可以部分提升最终组合净值，但是仍然不够稳定，我们通过使用混合信号（在数据准备中提到了构建方法，即根据训练集内的胜率，加权平均各类因子给出的信号），达到 50% 以上，同时我们也考虑到了主观逻辑的一致性。

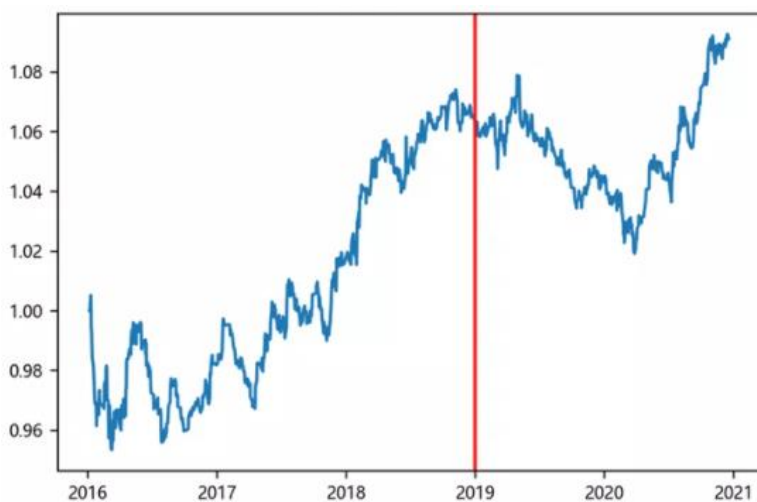
另外值得注意的是，在单一类别因子的基本面回测中，我们发现不同类型因子不同时间段表现存在差异。如果采用单因子模型难以提供稳定收益，这也符合指标在不同行情周期下的有效性规律。若将所有因子放入模型中后，我们发现其效果基本要优于单一类别，收益和风险控制上表现都有所提升。

图：大类因子独立测试情况

	样本内胜率	样本外胜率	最终净值	最大回撤
经济增长	0.50	0.52	0.94	0.22
市场风格	0.54	0.51	0.96	0.16
资金面	0.51	0.51	0.75	0.51
外部因素	0.50	0.50	0.81	0.24

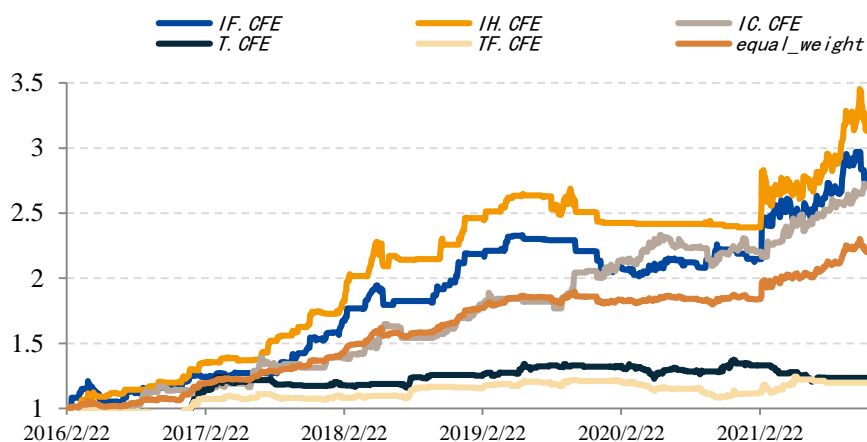
资料来源：Wind，国投安信期货

图：加入训练集的净值曲线



资料来源：Wind，国投安信期货

图：短周期模型品种合约测试



数据来源：wind，国投安信期货

国投安信期货
SDIC ESSENCE FUTURES

资料来源：Wind，国投安信期货

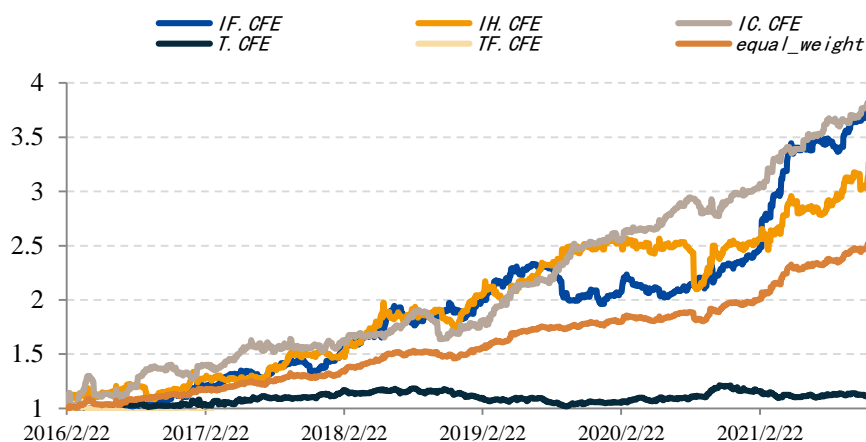
三、持仓量因子策略

持仓因子的构建方法是根据会员持仓，选择会员多头势力强于会员空头势力的合约做多，选择会员空头势力强于多头势力的合约做空，其隐含逻辑是会员投资者与普通投资者之间存在信息不对称，其掌握着比普通投资者更多的信息。

持仓量因子策略关键假设：

1. 前二十大会员多空单持仓比大于过去 3 日均值则记为 1，否则记为-1；
2. 由于持仓量数据受交割日前跨期价差波动影响较大，屏蔽交割日前 7 日的信号。

图：持仓量策略品种合约测试



数据来源：wind，国投安信期货

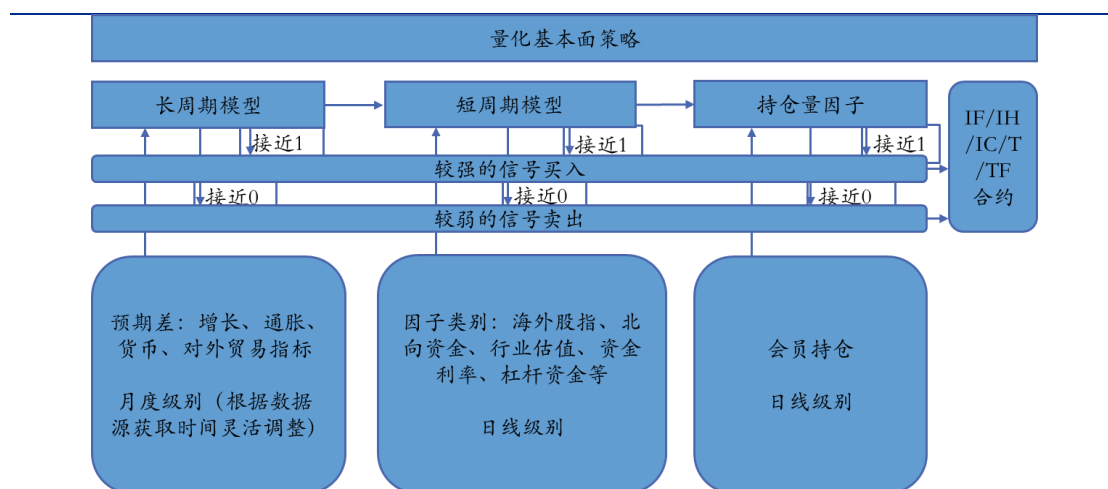
国投安信期货
SDIC ESSENCE FUTURES

资料来源：Wind，国投安信期货

四、合成信号

综合信号强度由3个独立模型的信号加权合成(0-1)。原则上取综合信号强度前2位，且数值大于等于0.6的合约做多。取综合信号强度后2位，且数值小于等于0.4的合约做空。由于持仓量数据受交割日前展期影响较大，屏蔽交割日前7日的信号。日内跌超1%视为止损点。等权重分配资金仓位。若策略未达预期，也可根据自身情况止损或止盈。

图：信号策略生成结构



资料来源：Wind，国投安信期货

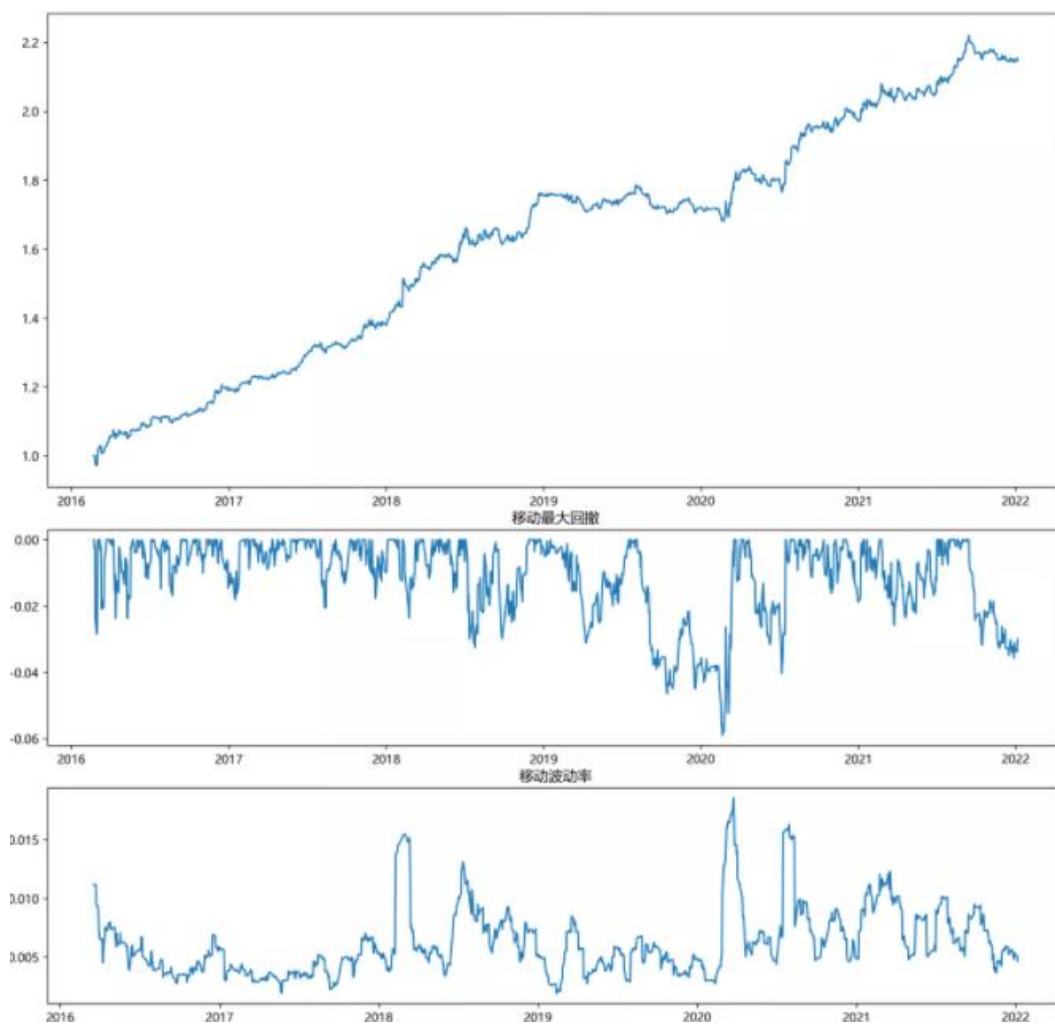
资料来源：Wind，国投安信期货

五、结论与展望

在本篇报告中，我们以量化基本面为核心，结合持仓量，技术面指标，形成了多策略多周期结构的策略。其中短周期模型聚焦于市场风格，外部因素，资金面三大高频金融数据板块；长周期模型关注市场预期，聚焦于宏观经济数据等低频指标。这一部分我们在之前的系列报告《股指量化基本面一：基本面特征筛选》中做过详细的介绍，主要逻辑为从高维数据中选出有效且相对独立的因子，依照主观分析框架，建立具有样本外泛化能力的模型。持仓量主要是考虑机构多空单持仓量进行合成。在 2016-2021 期间，最终策略胜率为 56%。接下来，我们将在以下三个方向上丰富策略：

- 1.在投资方式和投资标的上拓宽。品种池中纳入商品期货合约，做多相对较强的合约，做空相对较弱的合约，灵活调整多空持仓配比。
- 2.对日内具体的进场时间点进行择时和优化，同时测试分钟级别数据滑点所带来的冲击成本影响。
- 3.嵌套相对独立的策略模型，例如选取稳定的技术面指标组合，协助判断日度和日内的进场时机点，同时起到过滤信号的作用。

图：综合信号策略净值曲线



资料来源: Wind, 国投安信期货

免责声明

本研究报告由国投安信期货有限公司撰写,研究报告中所提供的信息仅供参考。报告根据国际和行业通行的准则,以合法渠道获得这些信息,尽可能保证可靠、准确和完整,但并不保证报告所述信息的准确性和完整性。本报告不能作为投资研究决策的依据,不能作为道义的、责任的和法律的依据或者凭证,无论是否已经明示或者暗示。国投安信期货有限公司将随时补充、更正和修订有关信息,但不保证及时发布。对于本报告所提供信息所导致的任何直接的或者间接的投资盈亏后果不承担任何责任。

本报告版权仅为国投安信期货有限公司所有,未经书面许可,任何机构和个人不得以任何形式翻版、复制和发布。如引用发布,需注明出处为国投安信期货有限公司,且不得对本报告进行有悖原意的引用、删节和修改。国投安信期货有限公司对于本免责声明条款具有修改权和最终解释权。