

2018 年 06 月 05 日

## 机器学习与量化投资：机器学习结合基本面

## ■商品价格同时受技术面和基本面的驱动

技术类指标对商品有直接的作用，甚至比股票更加显著。

基本面指标是商品价格变化更深刻的逻辑。

二者结合能更能反映出商品价格变化的规律

## ■机器学习结合基本面信息期货策略比纯技术面指标神经网络模型有显著优势

机器学习结合基本面信息期货策略

夏普：1.8

最大回撤：12.43%

纯技术面指标神经网络模型

夏普：1.22

最大回撤：24.36%

## ■风险提示：

某些期货流动性不足。市场结构发生变化导致机器学习模型失效

## 金融工程主题报告

证券研究报告

杨勇

分析师

SAC 执业证书编号：S1450518010002

yangyong1@essence.com.cn

周袤

分析师

SAC 执业证书编号：S1450517120007

zhoumao@essence.com.cn

## 相关报告

美股 Smart Beta 产品巡礼之一：红利因子	2018-06-03
FOF 和资产配置周报：中金 MSCI 中国 A 股国际质量指数基金获批，6 月维持谨慎判断	2018-06-03
机器学习与 CTA：上交所的尾盘集合竞价制度	2018-06-03
华安 CES 港股通精选 100ETF 上市，博时黄金 ETF 份额快速增长	2018-05-27
机器学习与 CTA：钓鱼单策略	2018-05-26

## 内容目录

<b>1. 机器学习结合基本面信息期货策略</b>	<b>3</b>
1.1. 设想和目标	3
1.2. 理论、方法及数据源	3
1.3. 交易成本与策略执行	3
1.4. 算法和模型	3
1.5. 结论	3
1.6. 风险点及未来的改进方向	3
<b>2. 理论分析</b>	<b>3</b>
2.1. 神经网络简介	4
2.2. 商品期货价格形成机制	5
<b>3. 建立数据</b>	<b>5</b>
3.1. 决定数据源	5
3.1.1. 库存指标	5
3.1.2. 商品品种选择	5
3.1.3. 特征因子	5
3.2. 分割数据源为训练集和测试集	5
<b>4. 结论</b>	<b>6</b>
<b>5. 算法选择的一些逻辑</b>	<b>8</b>
5.1. 模型外的异或关系或内在并列关系	8
5.2. 回归还是分类	9

## 图表目录

图 1: 神经网络--单层	4
图 2: 神经网络—多层	4
图 3: 商品品种—库存指标	5
图 4: 机器学习结合基本面信息期货策略—净值	6
图 5: 单纯神经网络—净值	6
图 6: 机器学习结合基本面信息期货策略--收益分布	7
图 7: 单纯神经网络—收益分布	7
图 8: 机器学习结合基本面信息期货策略—回撤	7
图 9: 单纯神经网络—回撤	7
图 10: 机器学习结合基本面信息期货策略—年化夏普	8
图 11: 单纯神经网络—年化夏普	8
表 1: 策略参数比较	8
表 2: 所有数据时期的收益率对比	8

## 1. 机器学习结合基本面信息期货策略

### 1.1. 设想和目标

库存变动作为期货重要的基本面信息，对于价格波动有直观的影响。另一方面，移动平均线等技术指标也是判断趋势的重要依据，而不同于库存值，技术指标因子与商品收益之间更多的是非线性关系。针对于此，采用机器学习中的全连接神经网络对技术指标进行拟合并将结果与库存信息结合成异或关系来预测收益能更好的捕捉获利空间。也即，用当前多种类别的期货商品数据包括价格和库存信息与过去所有数据进行模式匹配，做多或做空根据过去模式显示会大概率上涨下跌的商品。根据库存变动信息的更新周期，设置按周调仓策略。

### 1.2. 理论、方法及数据源

假设一：在流动性充足的市场，现货和期货之间不存在套利机会，即两者具有接近 1 的相关性，可以通过现货数据研究期货价格。

假设二：太阳之下无新鲜事。假设过去发生的事情未来会重复发生。

方法为用全连接神经网络预测对期货价格进行回归预测，并结合库存变动信息选择投资商品品种。数据主要来源于天软、choice，主要为期货主力合约价格信息以及每周库存量。

### 1.3. 交易成本与策略执行

在此策略的历史数据回测中，成交成本假设为隔日单边万分之 3。也即在成交中假设 1.5 个指数点的冲击成本。这样的假设充分包含了目前期货低流动性的现实

### 1.4. 算法和模型

模型本身是监督式学习，在第一阶段用价量指标来预测收益，若基于机器学习的预测值位于高于低位时，且观测点前一段时间的库存仓位增多则做空。判断高位、低位没有固定的阈值，而是通过观测全部商品品种的预测收益表现，取高者、低者。位时，进入第二阶段，若观测点前一段时间的库存仓位减少则做多。

### 1.5. 结论

将 2014 年至 2016 年的数据作为训练集，用 2017 年以后的数据做测试集。下图是测试集结果。

在隔日单边万分之 3 的成交假设下，策略表现如下：

夏普：1.8

最大回撤：12.43%

胜率：61%

盈亏比：1.44

年化：48.80%

### 1.6. 风险点及未来的改进方向

风险点主要有

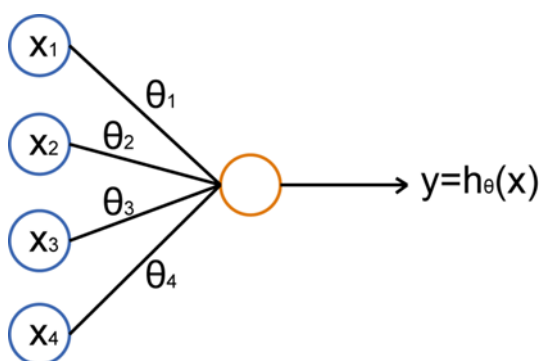
- (1) 期货和现货突然性的暂时偏离(在当前负基差的情况下和低成交量下，尤其可能发生)
- (2) 市场结构发生了深刻的变化(投资者类型，投资者风险偏好等等)，导致过去的数据不再能预测未来。
- (3) 市场流动性不足，导致成交需要付出巨大成本或者无法成交。

## 2. 理论分析

## 2.1. 神经网络简介

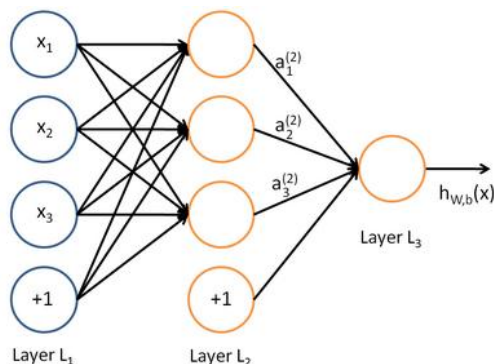
神经网络算法试图模拟现实世界中大脑的架构，神经网络中的每个节点相当于大脑中的神经元。当某个神经元的树突收到上一级感受器或神经元传来的刺激，使其电位超过一定的阈值，那么该神经元会被激活，处于兴奋状态，信号将经胞体沿轴突和末端突触，继续传递至下一级神经元的树突。如此逐级传递形成一个巨型网络。与之类比的是，如下左图输入向量  $x(x_1, x_2, x_3, x_4)$  扮演神经元的树突部分，代表输入的特征信息； $\theta_1$  到  $\theta_4$  称为连接权重，代表不同信息的重要性，需要通过训练调节。 $x_4=1$  时，又称为偏置项。输入层  $x_1$  到  $x_4$  的信息按权重加和，随后传递给“胞体”，进入一个非线性的激活函数  $h\theta(x)$ ，当前神经元将数据融合后的结果作为新的树突传递给下一神经元直到网络终结。常用的激活函数包括 sigmoid 函数、tanh、relu 函数等。

图 1：神经网络--单层



资料来源：安信证券研究中心

图 2：神经网络—多层



资料来源：安信证券研究中心

理论上，多层的神经网络可以近似地拟合出任何一个函数。如上面右图展示了含有一个输入层 (layer L\_1) 隐藏层 (layer L\_2) 和隐藏层节点数均为 4 的神经网络。每次神经网络学习时会初始化权重，再根据反向传播不断修改，最后的结果也因此也不完全相同。为了获得一个比较稳定的模型，我们会令 Bagging 次数为 10，即构建 10 个神经网络模型，对收益的最后预测结果由 10 个模型同权重平均的数值产生。

## 2.2. 商品期货价格形成机制

商品期货价格形成机制 (Commodity Futures Pricing) 的研究主要由库存理论 (The Theory of Storage) 和现货升水理论 (The Theory of Normal Backwardation) 两大山脉构成。而以上两种理论均表明, 商品的库存水平对商品未来走势有很大的影响。

## 3. 建立数据

### 3.1. 决定数据源

#### 3.1.1. 库存指标

真实的社会库存并不局限于交易所仓库中存放的部分 (即交易所库存), 还包括由港口库存、主要城市库存等构成的社会库存。交易所库存包括注册仓单部分和未注册仓单部分。尽管从理论上讲, 在没有准确社会库存数和交易所库存时, 注册仓单数又称期货库存在一定程度上能反映商品库存情况, 但实际操作中, 由于不同商品仓储成本和有效期差异, 部分品种只有真正需要交割时才会被注册成仓单。因此我们选取库存指标的优先级为社会库存>>交易所库存>>期货库存。

#### 3.1.2. 商品品种选择

前述, 库存信息主要有三个来源, 针对具体品种, 库存指标的可靠性还有待商榷。例如黑色系品种属于需要交割才注册成仓单, 而农产品由于其季节性, 可能存在长期无仓单数据的情况, 注册仓单数信号作用较弱。因此, 无社会库存或交易所库存指标的黑色系品种和农产品不在我们的投资品种范围内。最后选定 22 个商品品种如下图所示。

图 3: 商品品种—库存指标

品种代码	品种名称	库存指标	品种代码	品种名称	库存指标
al	沪铝	交易所库存	pb	沪铅	社会库存
au	沪黄金	交易所库存	j	连焦炭	社会库存
cu	沪铜	交易所库存	ME	郑甲醇	期货库存
fu	沪燃油	交易所库存	Ag	沪银	社会库存
ru	沪天胶	交易所库存	FG	郑玻璃	期货库存
zn	沪锌	交易所库存	bu	沪沥青	社会库存
TA	郑甲酸	期货库存	TC	郑动力煤	社会库存
l	连聚乙烯	期货库存	HC	热轧卷板	交易所库存
wr	线材	交易所库存	PP	聚丙烯	期货库存
rb	螺纹钢	社会库存	NI	镍	交易所库存
v	聚氯乙烯	期货库存	SN	锡	交易所库存

资料来源: Choice, 安信证券研究中心

#### 3.1.3. 特征因子

神经网络作为有监督学习, 需要给与“输入” (即特征因子向量) 和相应的“结果”。这里选择的特征因子分别为过去五日的收盘价平均价/观测点收盘价、过去八日的收盘价平均价/观测点收盘价、LLT (低延迟趋势线) 的变化趋势、以及相应的归一化数据等。

### 3.2. 分割数据源为训练集和测试集

天软的期货信息从 2014 年 7 月 25 日开始起较为完善, 故选取从该日的数据至 2016 年 12 月 31 止为训练数据。从 2017 年 1 月 1 日的数据至 2017 年 12 月 31 日的数据为测试集数据。策略主要考察其在测试集的表现。

## 4. 结论

在日内单边千分之一，隔日单边万分之3的成交假设下，策略表现如下：

机器学习结合基本面信息期货策略：

夏普：1.8

最大回撤：12.43%

胜率：61%

盈亏比：1.44

年化：48.80%

单纯神经网络：

夏普：1.22

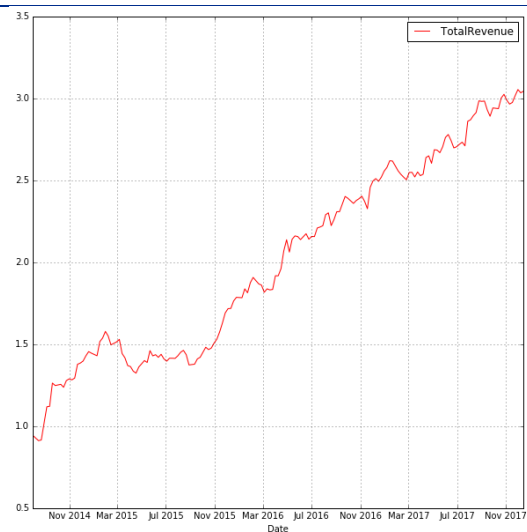
最大回撤：24.36%

胜率：73%

盈亏比：1.48

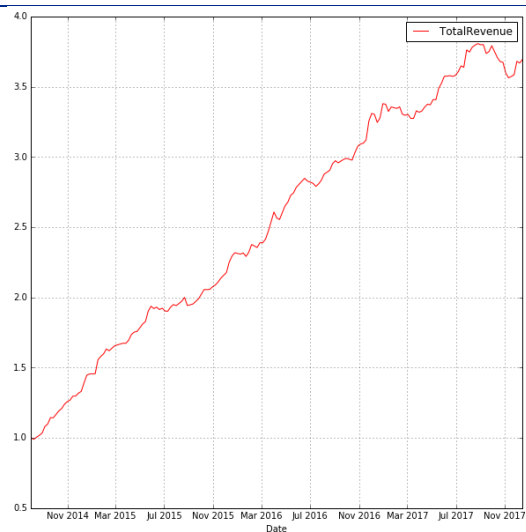
年化：31.50%

图 4：机器学习结合基本面信息期货策略—净值



资料来源：安信证券研究中心

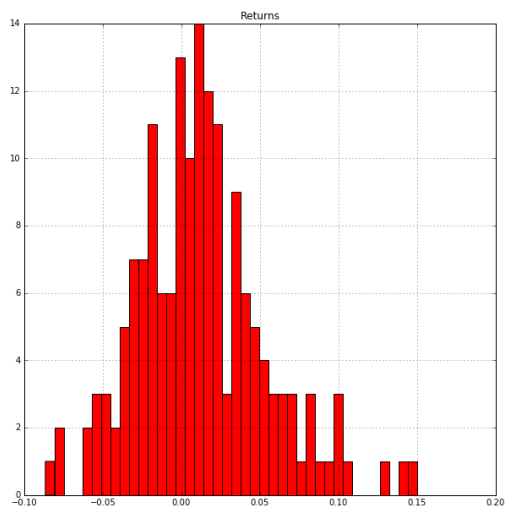
图 5：单纯神经网络—净值



资料来源：安信证券研究中心

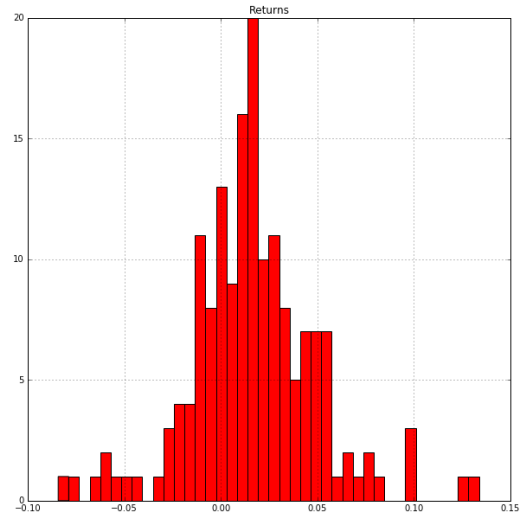


图 6：机器学习结合基本面信息期货策略—收益分布



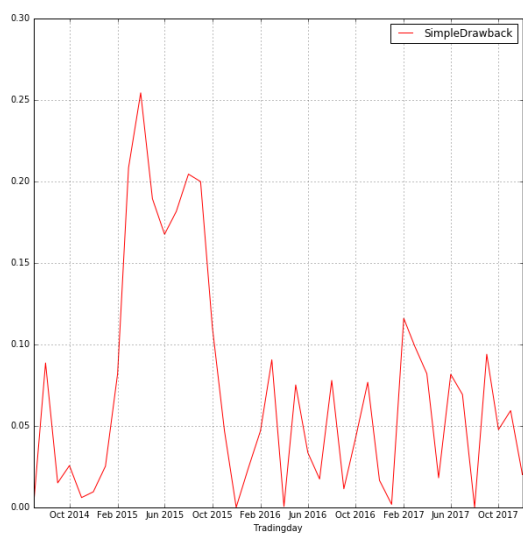
资料来源：安信证券研究中心

图 7：单纯神经网络—收益分布



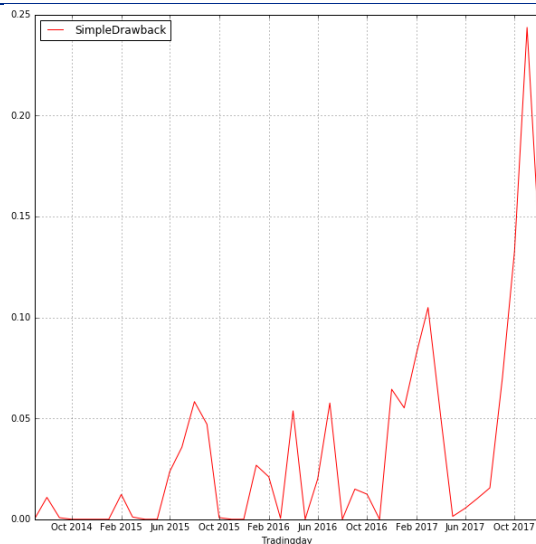
资料来源：安信证券研究中心

图 8：机器学习结合基本面信息期货策略—回撤



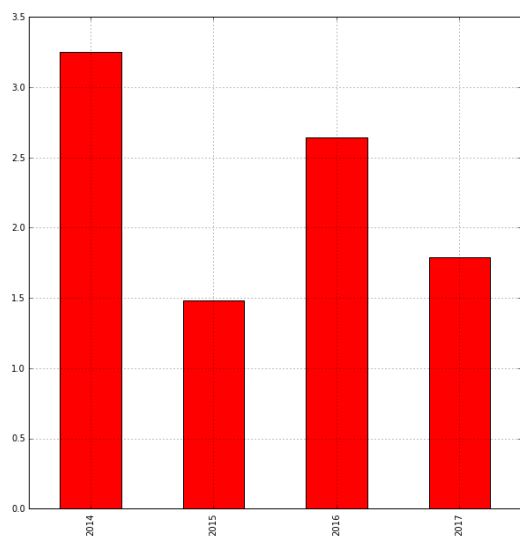
资料来源：安信证券研究中心

图 9：单纯神经网络—回撤



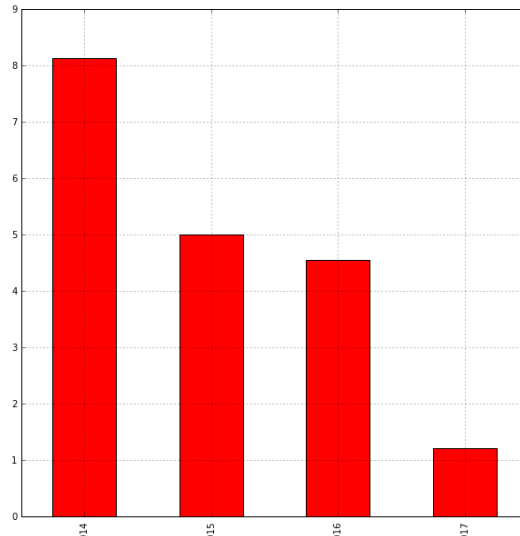
资料来源：安信证券研究中心

图 10：机器学习结合基本面信息期货策略—年化夏普



资料来源：安信证券研究中心

图 11：单纯神经网络—年化夏普



资料来源：安信证券研究中心

表 1：策略参数比较

参数/策略名	机器学习结合基本面信息期货策略	单纯神经网络
夏普	1.8	1.22
最大回撤	12.43%	24.36%
年化收益	48.80%	31.50%
周胜率	61%	73%
盈亏比	1.44	1.40

资料来源：安信证券研究中心

表 2：所有数据时期的收益率对比

年份/策略名	机器学习结合基本面信息期货策略	单纯神经网络
2014	0.445197	0.454479
2015	0.340747	0.863690
2016	0.769929	1.060457
2017	0.487973	0.315122

资料来源：安信证券研究中心

## 5. 算法选择的一些逻辑

### 5.1. 模型外的异或关系或内在并列关系

本策略中，我们首先用神经网络预测收益，然后再观测库存信息变动，如果商品的库存指标显示了预示了相同的变动方向，则买入或者卖出该商品，这属于神经网络模型外的异或关系。之所以选择这种方式在于，神经网络的完成需要大量的数据进行训练，理论上愈多的数据支撑能帮助挖掘到更多的内在逻辑。我们选择的交易期间（2014 年 7 月 25 日至 2017 年 12 月 31 日）22 个期货品种的每日交易数据共约 18000 条，测试数据小于 6000 条；而能获得的可靠库存信息基本上以周更新为主，即  $18000/7$ ，对应的测试数据小于 1000 条。如果将库存变动量作为一个特征因子去训练神经网络，将有“欠拟合”的风险。另一方面，假使库存数



据充足，是否将其放入特征因子中仍需商榷。从上述各年度收益可以得出，训练集的收益大幅度领先于测试集（2014 年数据只从 7 月起）。模型存在“过拟合”的危险，而单纯神经网络的训练集结果远超机器学习结合基本面信息期货策略也说明了这一问题。最终的测试集结果，机器学习结合基本面信息期货策略无论是收益率、夏普、最大回撤状况都好于单纯神经网络，也说明库存信息单独作为判断依据，其 50% 的权重能很好地控制模型的“过拟合”的程度。

## 5.2. 回归还是分类

神经网络根据其输出变量的类型不同，可分为神经网络回归和神经网络分类。具体来说，定量输出称为回归，或者说是连续变量预测；定性输出称为分类，或者说是离散变量预测。我们的策略采用的都是回归模式，用多个特征因子去预测未来五天的收益，而如果采用分类方法，这里预测值选择有两种方式。1) 涨或是跌，可以对应数值数据 1 或者 -1 进行分类。这样的策略选择可能使每个调仓周期交易多个商品品种，尽管有分摊风险的考虑，但算入交易费，以及在测试集上相对较低的准确率，会存在买入或涨幅在 0 附近波动的商品品种，这无疑会降低收益率。2) 涨、平或是跌，对应 (1、0、-1) 通过观测数据，很多商品在一周内的变动量并不大，即在 0 左右波动，这种交易如在 1) 中所述由于有手续费的存在以及可能的预测错误会降低收益，故可以放弃交易，而选择买入预测为 1 的数据并卖出预测为 -1 的数据。理论上该方法有可行性，但是属于 0 类的数据划分阈值难以界定。是根据所有交易时期的涨跌幅程度选择，抑或是通过周内各个商品的涨跌程度来判断？如果该周所有商品都波动不大，那是否还需交易？

实习生胡梦媛对此报告有贡献

## ■ 分析师声明

杨勇、周袁声明，本人具有中国证券业协会授予的证券投资咨询执业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。

## ■ 本公司具备证券投资咨询业务资格的说明

安信证券股份有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得证券投资咨询业务许可。本公司及其投资咨询人员可以为证券投资人或客户提供证券投资分析、预测或者建议等直接或间接的有偿咨询服务。发布证券研究报告，是证券投资咨询业务的一种基本形式，本公司可以对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向本公司的客户发布。

## ■ 免责声明

本报告仅供安信证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因为任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告基于已公开的资料或信息撰写，但本公司不保证该等信息及资料的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映本公司于本报告发布当日的判断，本报告中的证券或投资标的价格、价值及投资带来的收入可能会波动。在不同时期，本公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，本公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。同时，本公司有权对本报告所含信息在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准，如有需要，客户可以向本公司投资顾问进一步咨询。

在法律许可的情况下，本公司及所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务，提请客户充分注意。客户不应将本报告为作出其投资决策的惟一参考因素，亦不应认为本报告可以取代客户自身的投资判断与决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，无论是否已经明示或暗示，本报告不能作为道义的、责任的和法律的依据或者凭证。在任何情况下，本公司亦不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告版权仅为本公司所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复制、发表、转发或引用本报告的任何部分。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“安信证券股份有限公司研究中心”，且不得对本报告进行任何有悖原意的引用、删节和修改。

安信证券股份有限公司对本声明条款具有惟一修改权和最终解释权。

## ■ 销售联系人

上海联系人	葛娇娇	021-35082701	gejy@essence.com.cn
	朱贤	021-35082852	zhuxian@essence.com.cn
	许敏	021-35082953	xumin@essence.com.cn
	孟硕丰	021-35082788	mengsf@essence.com.cn
	李栋	021-35082821	lidong1@essence.com.cn
	侯海霞	021-35082870	houhx@essence.com.cn
	林立	021-68766209	linli1@essence.com.cn
	潘艳	021-35082957	panyan@essence.com.cn
	刘恭懿	021-35082961	liugy@essence.com.cn
	孟昊琳	021-35082963	menghl@essence.com.cn
北京联系人	温鹏	010-83321350	wenpeng@essence.com.cn
	田星汉	010-83321362	tianxh@essence.com.cn
	王秋实	010-83321351	wangqs@essence.com.cn
	张莹	010-83321366	zhangying1@essence.com.cn
	李倩	010-83321355	liqian1@essence.com.cn
	高思雨	021-35082350	gaosy@essence.com.cn
	姜雪	010-59113596	jiangxue1@essence.com.cn
	周蓉	010-83321367	zhourong@essence.com.cn
	胡珍	0755-82558073	huzhen@essence.com.cn
	范洪群	0755-82558044	fanhq@essence.com.cn
深圳联系人	巢莫雯	0755-82558183	chaomw@essence.com.cn
	黎欢	0755-82558045	lihuan@essence.com.cn

## 安信证券研究中心

### 深圳市

地址：深圳市福田区深南大道 2008 号中国凤凰大厦 1 栋 7 层

邮编：518026

### 上海市

地址：上海市虹口区东大名路 638 号国投大厦 3 层

邮编：200080

### 北京市

地址：北京市西城区阜成门北大街 2 号楼国投金融大厦 15 层

邮编：100034