

商品基本面量化系列（七）：PTA 基本面因子初探

王 锴 投资咨询证号 Z0016943
王雪忆 从业资格证号 F03125010

国投期货研究院

PTA 作为聚酯产业链的中间产品，主要满足下游原料聚酯的需求，并进一步用于服装纺织和包装材料中。随着国内聚酯产业的需求增强和生产装置的技术更新，我国 PTA 产能、产量和需求均大幅上升，在全球 PTA 产业中具有较大的定价权。

从 PTA 上下游情况来看，其产能和消费集中度都比较高。原油存储和炼化战略布局集中在东部沿海地区，而服装产业集群也多集中东部、南部沿海地区。相比于上游原料，PTA 受宏观政策和国际局势影响较小，上下游产业链价格的相关性较高。因此，PTA 价格走势具有较高的研究价值。

从交易角度看，PTA 作为 2006 年就挂牌上市的期货品种，是我国首个化工期货品种，交易量大、流动性好、市场透明度高。PTA 品种数据充足，将其量化模型分析具有可行性。另一方面，市场参与结构较为完善，意味着基本面对价格的驱动传导机制较为通畅，市场有效性大部分情况下相对稳定。

在构建策略模型前，我们根据 PTA 产业特征充分考虑季节性因素的影响。PTA 生产装置每年需要进行检修，生产企业往往会选择淡季进行装置检修从而减少供给侧的影响。秋冬季节是纺织行业的传统旺季，服装和家纺行业的需求会显著增加，进而带动聚酯纤维需求的上升，可能推动 PTA 价格上涨。

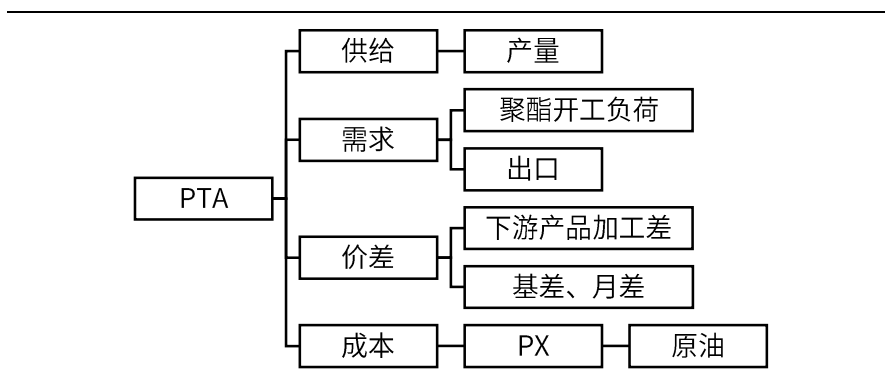
在之前的量化研究中，我们运用多种机器学习模型来进行拟合和预测，在基本面因子探索方面取得了一定的进展。我们仍然会以尊重主观基本逻辑为前提，以合理的分析方法，建立多维度模型加以比较和论证。本文中我们梳理了影响 PTA 价格的产业链逻辑，深度挖掘影响价格的基本面数据。

基于供给-需求-成本-价差四因子模型构建的周频策略样本外年化收益 18.05%，夏普比值 2.83。样本内价差因子表现最佳，年化收益 49.97%；样本外需求因子较好，夏普比 1.81。

1. 基本面相关因子

PTA 既是原油产业链的末端产品，同时又是化纤的前端产品，起到承前启后的作用，影响其价格变化的因素较多。本文将 PTA 的基本面因子分为四个大类，分别为供给因子、需求因子、价差因子和成本因子。

图 1：PTA 指标结构分类



资料来源：同花顺 ifind，隆众，国投期货

供给因子：PTA 装置运行负荷的变化是导致供应变化的重要因素，生产企业的集中检修通常集中在每年的春季或者秋季。多数情况下，装置的开工负荷受制于行业利润水平的高低，当预期下游需求较大或者利润空间变大时，企业生产积极性会提高。因生产周期的滞后性，行情在上涨过程中，产品出厂时利润空间会放大，工厂生产负荷往往会加大，但随着产品价格的进一步提升，下游承受能力逐渐减弱，同时供应量的加大会对价格的继续上涨形成压制。相反，在行情下跌过程中，生产利润下降，工厂或停车检修，市场供应量减少，将会减缓行情的下跌速度，若出现供不应求的情况，价格也有可能出现反弹。

需求因子：PTA 主要用于生产聚酯产品，生产 1 吨聚酯产品需要 0.85 吨左右的 PTA，聚酯的生产运行负荷变化是最能反映下游对于 PTA 需求变化的指标。同时，纺织增长决定终端需求，聚酯产品中涤纶对 PTA 的需求量占比较大。化纤产量占我国纺织工业纤维加工总量的近 85%，而化纤中涤纶占 80%左右，可见涤纶是纺织行业的主要原料。也就是说，纺织行业的景气程度、发展情况直接影响涤纶市场消费，进而对 PTA 需求产生影响，可以通过重点关注纺织企业以及印染厂开机率等指标来进行评估。

成本因子：PTA 的上游原料来源单一，基本都是由 PX 生产而来。PTA 的价格与上游原料（如 PX、原油）有较强的关联性，成本传导机制较为清晰。近年来，PTA 行业产能过剩形势加剧，行业利润受到大幅挤压，成本端驱动效应更加显著。市场上 90%以上的 PX 是用来生产 PTA 的，因此 PX 的价格以及供需情况会对 PTA 的市场价格产生直接的影响。PX 的原料是石油，石油经过馏分获得石脑油，从石脑油中提炼出 MX，再由 MX 生产 PX，最终获得 PTA。原油作为石化行业的源头，其价格的变化会影响 PTA 的市场价格，尤其在期货市场上的影响会被放大。此外，原油价格的涨跌会影响下游市场的心态，指导原料的采购，进而主导整条产业链产品的行情。

价差因子：主要由华东地区期现价差以及聚酯产品加工差构成。随着交割制度的不断完善，产业参与 PTA 期货意愿增加，期现联动性增强，其中华东基差的变化会影响套期保值和套利活动的积极性。聚酯产品加工差包括瓶片、DTY、POY、FDY 以及短纤加工差，可以说是 PTA 下游行业开工负荷的领先指标。如果加工差扩大，意味着下游企业盈利情况改善，企业开工积极性提升，市场对未来需求持乐观态度，预期推动 PTA 价格上涨；若加工差缩小或转负，表明下游企业面临成本压力，则 PTA 价格可能承压。

2. 数据清洗

2.1 数据频率和季节性处理

数据频率方面，我们选择日频和周频数据，更低频次的数据由于样本数量过少不利于拟合，且发布时间往往滞后不足以支持样本外实际跟踪。考虑到 PTA 日内价格波动较大，第一种方法是将周频数据映射为日频数据，但是这里并没有采用插值法来直接估算实际数据未发布前的数据，而是统一采用原数据，因为这样可以匹配更为真实客观的对应关系。第二种方法是统一将所有数据低频化处理，这主要是考虑到 PTA 所获取的基本面数据中供应以及需求大类因子均为周频，对应的信号变化频率也偏向周度层面的变化，因此考虑将周频模型的表现与日频模型进行比较，同时考虑周期同比、环比和均线平滑来处理这些数据。

2.2 数据滞后性处理

日频数据基本都会在交易日当天收盘后更新，而周频数据基本在周五发布。因此我们在做特征有效性检测和建模过程中统一将这些数据延后一阶处理，即当天获得的数据生成信号在下一个交易日进行交易。

2.3 去极值和标准化处理

数据中的极值往往会影响到线性模型假设下的拟合，同时对于映射关系也存在错误推断的可能，因此我们将高于时序数据 3 倍标准差的数值调整为平均值偏离 3 倍标准差。此外，所有的基本面因子采用 W 检验的方式处理其数据的正态关系，通过 P-value 值观测，检验中只有少量因子存在相对正态分布关系，因此集中处理其影响模型预测精度的异常值并按照各个大类因子数据特征进行标准化处理，使得大类数据和 PTA 价格数据处于同一假设关系。

整体数据选取 2019 年 1 月至 2024 年 12 月的基本面数据，进行差值处理和数据填充整合，分别采用相关性及平稳性检验、回归测试单因子有效性，进而设计基本面因子指导的交易策略。

图 2：特征筛选排序-节选

因子名称	相关性	p值	Adj_R2	显著性
印染厂：开机率：江苏（周）	0.0799	0.0024	0.0003	***
印染厂：开机率：华东地区（周）	-0.0731	0.0055	-0.0004	***
PTA:华东:主流价	-0.0596	0.0237	0.0019	**
印染厂：开机率：浙江（周）	-0.0570	0.0306	-0.0004	**
PX(对二甲苯):中国台湾:到岸中间价	0.0556	0.0349	0.0050	**
PTA成本	-0.0475	0.0716	0.0015	*
PTA：原料：中国聚酯工厂：库存可用天数（周）	-0.0439	0.0959	0.0005	*
纺织原料类：纺织企业：库存可用天数：中国（周）	0.0404	0.1249	0.0015	0
期货收盘价(活跃):布伦特原油:ICE	-0.0387	0.1424	0.0002	0
短纤加工差	0.0376	0.1542	0.0001	0
PTA：产能利用率：中国（周）	-0.0367	0.1643	0.0014	0
聚酯纤维短纤：厂内权益库存可用天数：中国（周）	0.0330	0.2110	-0.0007	0
聚酯纤维短纤：产销率：中国（周）	0.0302	0.2513	0.0001	0
DTY加工差	-0.0250	0.3425	0.0009	0

资料来源：同花顺 ifind，隆众，国投期货

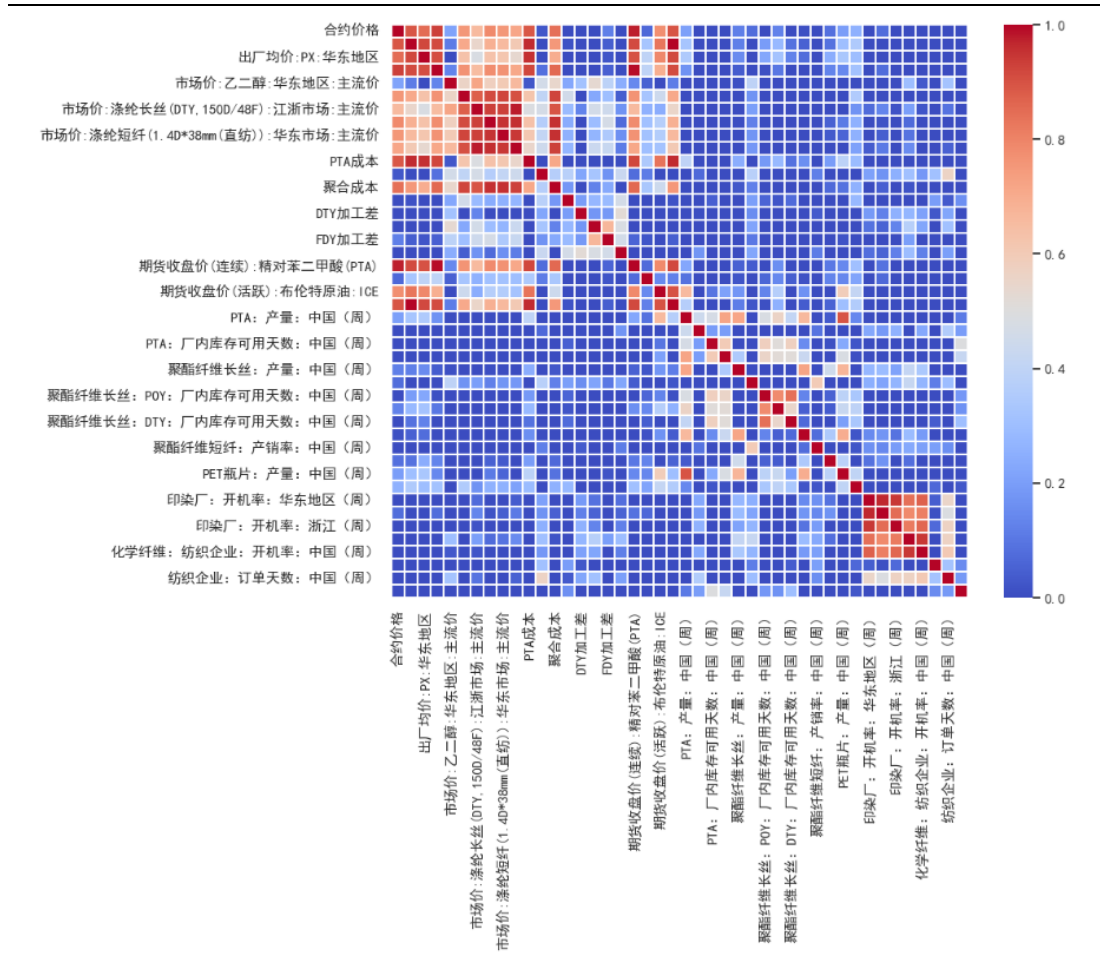
供给因子中，PTA 产能利用率及厂内库存可用天数的胜率较高。高产能利用率意味着市场供应较为充足，可能导致 PTA 价格承压。PTA 库存可用天数的变化可以反映市场对后期供应的预期，如果市场货源充裕，下游需求一般，库存可用天数可能延长，一般长假前后该指标的变化较为显著。成本因子中，布伦特原油期货活跃合约收盘价的胜率较高。该因子代表国际市场对未来油价的预期，会直接影响 PX 的成本，进而传导至 PTA 的价格。并且原油作为商品之王，价格的波动还受到宏观经济、地缘政治等多重因素影响，这些因素通过油价的变化传递到整个化工产业链，对 PTA 价格的变化产生较大影响。

需求因子中，聚酯纤维长丝厂内库存可用天数及印染厂开机率的胜率较高，主要因为它们直接反映了产业链中下游的需求情况，具有较强的参考性。价差因子中，聚酯纤维加工差的胜率较高。该因子代表下游聚酯工厂在生产过程中的利润空间，能够提前反映市场的变化趋势，不仅包含了 PTA 和 EG 的价格信息，还考虑了下游产品的市场价格和其他加工成本，综合体现了多个因素的影响，可以看作下游需求的领先指标。

通过计算信息系数 IC，我们可以了解因子值对于下期 PTA 收益率的预测能力以及因子间的内在相关性。从整个单因子线性回归结果看，成本因子以及价差因子对于价格的影响较大，相关性也更为显著。需求端印染厂以及纺织企业开机率等存在一定同步性，导致多重共线性的可能提升，我们通过特征筛选模型进行识别，并在合成因子时引入相关性罚值。

整体来看，因子构建逻辑主要根据估值+驱动的方法，首先从基本面上找到估值过低或过高的行情，然后再结合技术面以及产业装置变化等情况来寻找行情驱动因子。

图 3: IC 相关性



资料来源: 同花顺 ifind, 隆众, 国投期货

3. 基本面多空策略周期信号的嵌套和比较

3.1 时序多空模型

我们将筛选得到的供给、需求、价差和成本四大类中的每个因子,从时间序列维度出发,根据 IC 正负关系,遍历参数得到单因子开仓阈值,随后我们将这些单因子开仓阈值在每个大类中进行加权合成,产生大类因子的开仓信号。

$$Class_j = c_1 \dots c_i \dots c_n$$

$$c_i = \begin{cases} 1 & c_i > percentile_a \quad a \in [lower, upper] \\ 0 & \text{other} \\ -1 & c_i < percentile_b \quad b \in [lower, upper] \end{cases}$$

$$c_i = \begin{cases} 1 & c_i > Class_len_a \quad a \in [lower, upper] \\ 0 & \text{other} \\ -1 & c_i < Class_len_b \quad b \in [lower, upper] \end{cases}$$

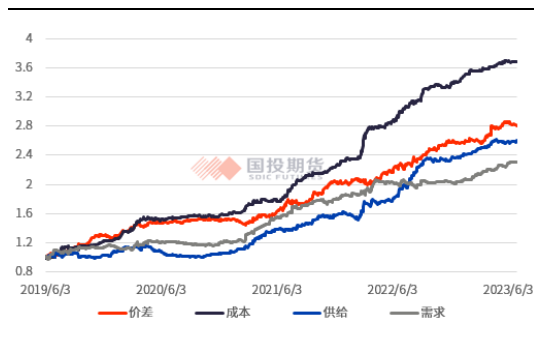
$$j \in \{supply, demand, stock, index\}$$

$$OR = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}$$

3.1.1 日频信号

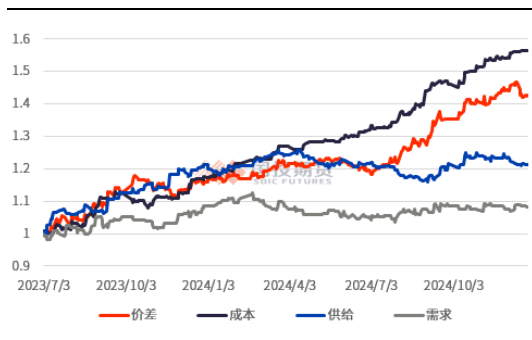
PTA 在化工品种中日内价格波动较大，因此从信号生成和调仓频率出发，我们优先考虑将周度数据高频化处理，与日度数据合成日频信号。在进行分类后选取 IC 值存在有效性的因子做阈值筛选，并按照 IC 的正负相关性方向作为信号的变化方向，当与标的处于正相关关系的时候时同方向变化，反之信号反方向变动。我们使用 Odds Ratio 作为筛选指标对每个大类下的因子进行遍历，通过排序选取 OR，从而增加合成开仓信号的胜率大小。对于不同的因子的最高 OR 筛选的阈值进行合成后折算成不同因子相应的胜率，并按照胜率以及相关性的罚值对不同大类下的合成因子进行加权。通过单因子净值的统计指标分析，价差、成本、供给以及需求大类单因子的平均胜率分别为 0.47、0.70、0.55 以及 0.57 左右，同时各个大类的平均年化收益集中在 45.46%、67.55%、40.40%以及 32.82%左右。从大类单因子入手观察对于收益变化影响，成本因子中，布伦特原油期货活跃合约收盘价的胜率较高。该因子不仅与 PTA 的生产成本密切相关，还能作为市场情绪和预期的领先指标，不论是样本内还是样本外的收益表现都相对较好。

图 4：日频策略训练集因子表现



资料来源：Wind，国投期货

图 5：日频策略测试集因子表现



资料来源：Wind，国投期货

图 6：日频策略训练集收益表现

	价差	成本	供给	需求
累积收益	180.205041	267.794582	160.153499	130.111738
年化收益	45.460404	67.556655	40.401993	32.823345
最大回撤	9.737397	5.300602	16.449210	10.587284
夏普比率	2.320500	3.911300	2.327500	2.165800
卡玛比率	4.673400	12.758000	2.458600	3.103400
年化开仓次数	101.160444	82.744702	101.160444	74.167508

资料来源：Wind，国投期货

图 7：日频策略测试集收益表现

	价差	成本	供给	需求
累积收益	37.738616	61.397634	21.345644	8.144138
年化收益	26.062580	42.401681	14.741467	5.624404
最大回撤	7.272141	5.384009	9.932951	8.914665
夏普比率	2.443000	3.725200	1.506000	0.601000
卡玛比率	3.593800	7.897300	1.488200	0.632700
年化开仓次数	91.850829	84.944751	97.375691	78.729282

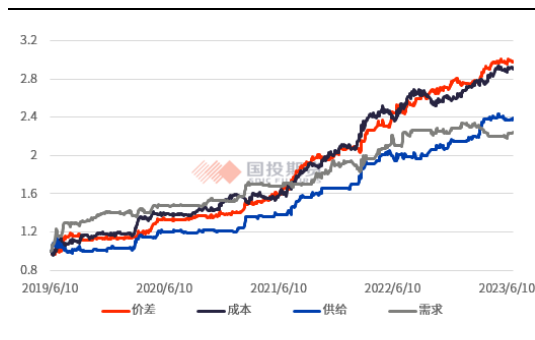
资料来源：Wind，国投期货

3.1.2 周频信号

商品基本面逻辑通常较少在短时间内发生较大改变，将周频因子进行高频化处理也可能造成信息缺失。因此如果在牺牲一定交易机会的条件下将日度数据低频化处理，能够更好地发挥与周度数据的协同性，更长的时间区间也有利于反映基本面的变化。同时，在周度层面上也更便于检验分析数据在同比、环比等时间周期方面的有效性。

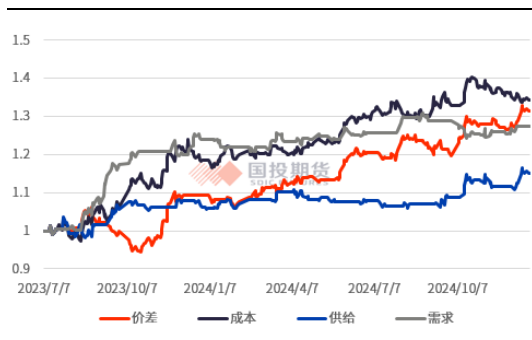
PTA 周度数据整体预测表现从收益以及稳定性来说没有日频策略效果好，我们尝试对数据集进行切片化处理，优化在极端行情下出现的过拟合问题。从整体数据角度出发，PTA 基本面的周度数据量占比超过日度数据，尤其体现在需求大类。从年化收益的角度出发，周度数据的价差、成本、供给以及需求大类单因子的平均年化收益集中在 50.00%、48.3%、35.3%、31.5%左右。相较而言，周度数据占比更高的大类因子在周频策略中表现更佳。需求因子在周频策略中样本外的年化收益、最大回撤以及夏普比率均优于日频策略。价差因子在周频策略中若只保留下游聚酯产品加工差因子的整体收益表现更好，因为该因子能够提前反映市场的变化趋势，不仅包含了原料的价格信息，还考虑了下游产品的市场价格和其他加工成本，综合体现了多个因素的影响，可以看作下游需求变动的领先指标。

图 8：周频策略训练集因子表现



资料来源：Wind，国投期货

图 9：周频策略测试集因子表现



资料来源：Wind，国投期货

图 10：周频策略训练集收益表现

	价差	成本	供给	需求
累积收益	197.281973	190.665275	139.436433	124.320683
年化收益	49.970105	48.294143	35.318245	31.489535
最大回撤	9.885009	17.006452	16.962429	15.497913
夏普比率	2.360800	1.915100	1.997100	1.640300
卡玛比率	5.060300	2.842600	2.084300	2.033900
年化开仓次数	24.569402	24.569402	17.223911	17.477204

资料来源：Wind，国投期货

图 11：周频策略测试集收益表现

	价差	成本	供给	需求
累积收益	31.396146	34.360814	15.088151	27.559243
年化收益	21.863611	23.928143	10.507069	19.191674
最大回撤	11.690578	7.435403	5.313348	6.573876
夏普比率	1.772000	1.610900	1.042100	1.818800
卡玛比率	1.875400	3.227100	1.983000	2.927500
年化开仓次数	23.676880	21.587744	17.409471	20.194986

资料来源：Wind，国投期货

4. 实证分析

从周频策略的表现来看，年化开仓次数大多在 17-25 之间，夏普比均达到 1 以上水平，其中价差因子表现最佳，年化收益 49%以上。日频策略的整体收益和风险控制能力均优于周频策略，成本因子的收益能力显著突出，风险控制能力也最佳，样本内外最大回撤不超 6%，夏普比达到 3 以上水平。表现相对较差的是需求因子，2022 至 2023 年俄乌冲突后原油和汽油调油需求成为影响市场的重要边际变量，过剩周期下产业链利润低位，终端需求较差，单因子有显著回撤，样本外收益表现一般。

近年来由于 PTA 行业过剩格局加剧，价格成本驱动较强，装置运行负荷以及下游开工的变化对行情的影响相对弱化，体现在模型中供给以及需求因子在样本外的有效性不如价差以及成本因子表现好，因子的解释力度相对较为有限。后续我们可以就供给和需求因子季节项以及趋势项的拆解进行进一步探索，优化模型的大类因子表现。

免责声明

国投期货有限公司是经中国证监会批准设立的期货经营机构，已具备期货投资咨询业务资格。

本报告仅供国投期货有限公司（以下简称“本公司”）的机构或个人客户（以下简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。如接收人并非国投期货客户，请及时退回并删除。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、意见及推测只提供给客户作参考之用。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的期货或期权的价格、价值可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。客户不应视本报告为其做出投资决策的唯一因素。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所导致的任何损失负任何责任。

本报告可能附带其它网站的地址或超级链接，本公司不对其内容的真实性、合法性、完整性和准确性负责。本报告提供这些地址或超级链接的目的纯粹是为了客户使用方便，链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。