

全方位融合：价值投资与量化交易的深度基本面量化（Quantamental）研究报告

摘要

金融市场的演进历程始终伴随着两种主流投资范式的博弈与融合：一是以本杰明·格雷厄姆（Benjamin Graham）和沃伦·巴菲特（Warren Buffett）为代表的价值投资（Value Investing），强调对企业内在价值的定性挖掘与安全边际的坚守；二是以詹姆斯·西蒙斯（James Simons）和 AQR 资本为代表的量化交易（Quantitative Trading），主张利用数学模型与统计规律捕捉市场定价错误。随着大数据技术、算力爆发以及非结构化数据处理能力的跃升，这两种范式正在经历一场前所未有的融合，即“基本面量化”（Quantamental）投资体系的兴起。

本报告旨在提供一份详尽的研究框架，系统性解构如何将深度的价值投资逻辑转化为严谨的量化模型。报告内容涵盖从底层的数据治理基础设施（Point-in-Time 数据架构）到核心因子的构建（价值因子去冗余、质量因子 QMJ 复现），再到经典选股模型（神奇公式、Z-Score）的量化实现与优化。进一步，本报告将深入探讨高阶交易策略，包括价值与动量的因子择时机制、基于基本面价差的统计套利（配对交易）体系，以及行业中性化多空策略的构建。最后，报告将视野拓展至前沿领域，论述自然语言处理（NLP）和另类数据（卫星图像、供应链数据）如何在量化模型中识别企业“护城河”与管理层信号。

第一章 范式融合：Quantamental 投资框架的理论与演进

1.1 从对立到统一：人机结合的必然性

长期以来，基本面投资与量化投资被视为硬币的两面。传统的基本面投资者（Discretionary Investors）倾向于深度分析，他们通过研读财报、调研管理层、分析商业模式来构建投资组合，其优势在于对因果逻辑的深刻理解和对非量化信息（如企业文化、品牌护城河）的感知，但劣势在于覆盖广度有限、易受认知偏差（如确认偏误、过度自信）影响，且难以在全市场范围内进行标准化的横向对比。相比之下，量化投资者（Systematic Investors）利用计算机算法处理海量数据，寻找具有统计显著性的市场规律，其优势在于纪律性强、覆盖面广、执行速度快，但往往面临模型缺乏经济学解释力（Black Box）、过度拟合（Overfitting）以及无法应对结构性市场突变（Regime Shift）的风险¹。

“基本面量化”（Quantamental）并非两者的简单叠加，而是一种深层次的辩证统一。它试图保留价值投资的经济学内核（Economic Intuition），即确保每一个因子的构建都有坚实的商业逻辑支撑，同时利用量化手段赋予其广度与纪律性。Morgan Stanley 的研究指出，全球股票管理人的超额收益（Alpha）中，约 65%可由广泛的市场因子（如价值、成长、质量、动量）解释，而剩余

35%则来自特质性的选股能力³。Quantamental 的核心目标正是通过因子化手段系统性地获取那 65%的 Beta 收益，并利用算法辅助的深度研究去捕捉那 35%的 Alpha，实现“人机结合”（Man + Machine）的最优解。

1.2 QVGS 多维因子框架体系

在构建 Quantamental 体系时，必须建立一个能够涵盖资产全生命周期的分类框架。业界广泛采用 QVGS 框架（Quantamental Value Growth Sentiment）作为核心支柱，该框架强调因子之间的正交性（Orthogonality）与互补性，以构建全天候的投资组合⁴。

维度	核心逻辑	典型指标与量化实现	经济学解释
Q (Quality/Quantamental)	寻找高质量的商业模式与管理层	ROIC (投入资本回报率), Gross Margin Stability (毛利率稳定性), Accruals (应计项目), F-Score	高质量企业能穿越周期，低应计项目代表盈余质量高，非财务造假。
V (Value)	寻找价格低于内在价值的标的	EBIT/EV (企业价值倍数), Shareholder Yield (股东收益率), Adjusted P/B	均值回归理论；投资者对暂时性负面消息的过度反应导致定价过低。
G (Growth)	捕捉盈利与收入的内生增长	SUE (标准化预期外盈利), Revenue Surprise, Forward EPS Growth	市场往往低估复利效应，高增长企业能通过业绩消化高估值。
S (Sentiment)	利用市场情绪与资金流向辅助择时	Price Momentum (价格动量), Analyst Revision (分析师修正), News Sentiment (新闻情绪)	行为金融学中的“羊群效应”与“反应不足/过度反应”。

QVGS 框架不仅是一个静态的分类，更是一个动态的生态系统。例如，在经济衰退初期，**质量因子（Q）**往往表现优异，因为投资者寻求避险；在复苏期，**价值因子（V）**接棒，因为受损资产估值修复；在繁荣期，**成长因子（G）与动量因子（S）**成为主导。Quantamental 策略通过量化识别宏观环境（Regime Identification），动态调整这四个维度的权重。

1.3 护城河的量化转译：从艺术到科学

巴菲特所推崇的“经济护城河”（Economic Moat）通常被视为一种定性的艺术判断，但在 Quantamental 框架下，这些定性特征必须被“转译”为可计算的代码。

- 品牌优势：** 可通过**毛利率（Gross Margin）**相对于行业平均水平的溢价及其稳定性来量化。拥有定价权的公司通常能维持高于同行的毛利率。
- 网络效应：** 对于平台型企业，可通过**资产周转率（Asset Turnover）**的非线性增长或用户获取成本（CAC）的边际下降来量化。
- 管理层诚信：** 可通过 **M-Score** 或**应计项目（Accruals）**分析来量化。如果一家公司的净利润持续高于经营性现金流，且应收账款周转天数异常增加，模型将判定其管理层可能存在激进会计处理，从而给予低评分⁵。

第二章 数据基础设施：Python 生态与 Point-in-Time 数据治理

2.1 数据的“时间机器”：Point-in-Time (PIT) 架构设计

在将价值投资理念回测于历史数据时，最致命且最隐蔽的错误是“前视偏差”（Look-ahead Bias）。在传统的基本面分析中，分析师清楚地知道某份财报是在哪一天发布的，但在量化回测中，如果数据库结构设计不当，模型极易“穿越”未来。例如，一家公司在 2024 年 4 月发布了 2023 年年报，如果在 2024 年 1 月的回测时间点上，模型读取到了这份年报的数据，那么它实际上是在利用未来的信息进行交易，这将导致回测业绩的严重虚高⁷。

Point-in-Time (PIT) 数据库的核心特征：

- 双重时间戳（Dual Timestamping）：** 每一个数据点必须包含两个时间属性——“生效时间”（For Period）和“发布时间”（As-Reported Date）。例如，2023 Q4 的营收数据，生效时间是 2023-12-31，但发布时间可能是 2024-03-15。模型只能在 2024-03-15 之后访问该数据。
- 版本控制与重述管理（Restatement Handling）：** 企业经常会因为会计准则变更、审计修正或错误更正而追溯调整历史报表。例如，中兴通讯（ZTE）在 2018 年 7 月重述了其 Q1 财报，导致 ROE 数据发生剧变。如果回测系统直接使用最终修正后的“正确”数据，就会产生偏差。PIT 架构必须存储数据的每一个历史版本（Vintage），确保回测时使用的是当时那个时刻市场所能看到的“错误”或“初步”数据，而非事后上帝视角的“正确”数据⁸。

3. **滞后处理（Lagging Assumption）**：如果缺乏精确的 PIT 发布日期，保守的量化模型通常会引入“最大发布滞后”假设（例如，假设财报在季度结束后的 45 天或 90 天后才发布），以物理隔离未来信息。

2.2 Python 量化技术栈的选择与应用

Python 以其强大的数据处理生态成为了 Quantamental 领域的通用语言。构建一个完整的价值量化系统，通常涉及以下库的深度应用：

库名称	功能定位	在价值量化中的具体应用场景
Pandas / NumPy	核心数据处理	处理时间序列财务数据，执行数据对齐（Alignment）、重采样（Resampling）和滚动窗口计算（Rolling Metrics）。例如，计算 TTM（Trailing Twelve Months）市盈率时，利用 Pandas 的 rolling 功能聚合过去 4 个季度的 EPS ⁹ 。
Zipline / Backtrader	回测引擎	Zipline （曾由 Quantopian 维护）是 PIT 数据处理的标杆，其 Pipeline API 允许用户定义动态股票池和因子计算逻辑，并自动处理拆股、分红和数据对齐，特别适合防止未来函数泄露。 Backtrader 则以其灵活性著称，支持多资产类别和复杂的事件驱动策略，适合模拟实盘交易逻辑 ¹⁰ 。
QuantLib	金融数学计算	虽然主要用于衍生品定价，但在计算企业信用风险时至关重要。例如，利用

		QuantLib 计算 Merton 模型中的期权定价公式，从而推导出企业的“距离违约点”（Distance to Default），作为深度价值策略的安全边际指标 ¹² 。
Statsmodels / Scikit-learn	统计与机器学习	用于因子有效性检验（Factor Validity Check）。利用 OLS 回归分析价值因子对收益率的解释力度（t-stat），或使用随机森林（Random Forest）等非线性模型预测下一季度的盈余惊喜（Earnings Surprise） ¹³ 。
BeautifulSoup / NLTK	非结构化数据处理	用于爬取 EDGAR 数据库中的 10-K/10-Q 文件，并利用 NLTK 进行文本清洗和词频统计，构建“文本情绪因子” ¹³ 。

2.3 财务报表自动化清洗与异常值检测

在将财报数据输入模型前，必须经过严格的清洗流程，这通常通过编写 Python 自动化脚本完成¹⁴：

- **会计恒等式校验：** 自动检查 $资产 = 负债 + 所有者权益$ 。如果源数据不满足此恒等式，说明数据源可能存在录入错误，需自动标记并剔除。
- **缺失值插补（Imputation）：** 对于缺失的财务数据，不能简单填充 0。应根据行业分类（GICS），利用同行业、同市值规模公司的中位数进行插补，或者基于该公司的历史增长率进行线性外推。
- **异常值（Outlier）处理：** 利用 MAD（Median Absolute Deviation）或 Z-Score 方法识别极端值。例如，如果某公司的 P/E 高达 5000 倍，可能是因为净利润接近于 0，这类数据点会扭曲线性回归模型，需进行 Winsorization（缩尾处理），将其压缩至分布的 1%或 99%分位点。

第三章 价值因子的深度解构与重构

3.1 传统价值因子的失效与“冗余”之谜

经典的价值投资依赖于账面市值比（Book-to-Market, B/M）、市盈率（P/E）或现金流收益率（CF/P）进行筛选。然而，近年来 Fama-French 五因子模型的研究提出了一个挑战性的观点：当模型中引入**盈利能力（Profitability, RMW）和投资模式（Investment, CMA）**因子后，传统的 HML（High Minus Low）价值因子在统计上变得“冗余”（Redundant），即其解释超额收益的能力被其他因子吸收了¹⁶。

这一现象的深层原因在于驱动机制的混淆：

1. **无形资产的缺失：**传统的账面价值（Book Value）主要计量厂房、设备等有形资产，而忽略了品牌、专利、研发能力等无形资产。这导致轻资产的高科技公司（如 Google、Microsoft）长期看起来 P/B 过高，被价值因子剔除，而筛选出的往往是重资产、低增长的夕阳产业。
2. **现金流冲击 vs. 贴现率冲击：**价值股的“便宜”可能源于两种截然不同的原因。
 - **现金流冲击（Cash Flow Shock）：**市场预期公司未来的现金流将永久性恶化（基本面变差）。这类股票是真正的“价值陷阱”，买入它们无法获得溢价。
 - **贴现率冲击（Discount Rate Shock）：**公司的现金流预期稳定，但由于市场风险厌恶情绪上升或流动性紧缩，导致投资者要求更高的回报率，从而压低了股价。只有买入这类由贴现率冲击导致的低估值股票，投资者才能获得真正的风险溢价（Risk Premium）¹⁶。

3.2 修正策略：基于贴现率冲击的价值因子构建

为了恢复价值因子的有效性，Quantamental 策略必须对股票池进行提纯，剥离掉那些因基本面恶化而变便宜的股票。

构建与清洗步骤：

1. **横截面盈利预测：**利用回归模型预测全市场股票的未来盈利（Expected Profitability）。
2. **残差提取：**股票的实际回报减去由盈利变化解释的部分，即为由贴现率变化驱动的部分。
3. **因子重构：**仅在那些估值变动主要由贴现率冲击驱动的股票池中构建价值因子。研究显示，经过这种调整后的价值因子，其年化超额收益比传统价值因子高出约 50%，且不再被五因子模型中的投资因子所吸收¹⁶。

3.3 “纯”价值因子（Pure Value）的工程化实现

在实际投资中，简单的“买入低 P/E”策略往往会暴露于巨大的行业风险（如长期超配银行和能源）或宏观风险（如利率敏感性）。为了获取纯粹的价值溢价，需构建“纯”价值因子（Pure Value）³。方法论：

- **行业中性化（Sector Neutrality）：**不在全市场范围内比较 P/E，而是在每个二级行业（如

半导体、生物医药、区域银行）内部进行标准化排名。例如，买入半导体行业中最便宜的股票，而不是买入全市场中最便宜的银行股。

- **正交化处理（Orthogonalization）**：通过多元回归，剔除价值因子对动量、规模、波动率等其他风格因子的暴露。

$$R_{stock} = \beta_1 \cdot Value + \beta_2 \cdot Momentum + \beta_3 \cdot Size + \epsilon$$

提取残差或构建对冲组合，确保所获收益纯粹来自于价值回归，而非动量崩盘或小盘股效应。

- **目标敞口法（Target Exposure）**：FTSE Russell 等机构采用优化算法，在最大化价值因子载荷（Factor Loading）的同时，将其他因子的载荷约束为零。历史回测表明，这种“纯”价值因子在金融危机后的表现显著优于传统价值因子，且回撤控制更佳。

第四章 质量因子（Quality）：价值的安全边际与风控

4.1 AQR 的“质量减去垃圾”（QMJ）模型详解

格雷厄姆强调的“安全边际”不仅指价格便宜，更指企业质地优良。在 Quantamental 框架下，这一思想被 AQR 资本系统化为 **QMJ (Quality Minus Junk)** 因子。AQR 的研究表明，高质量公司通常具有更高的风险调整后收益，且在市场动荡期具有极强的防御性¹⁸。

QMJ 因子的四维评分体系与公式⁶：QMJ 得分为以下四个维度的 Z-Score 均值：

$$Quality = z(Profitability + Growth + Safety + Payout)$$

1. **盈利能力（Profitability）**：衡量企业利用资产创造利润的效率。
 - **GPRO (Gross Profits / Assets)**: 毛利润率。这是最纯粹的盈利指标，因为它剔除了管理费用、广告费和利息支出的干扰，且难以通过会计手段操纵。
 - **ROE (Net Income / Book Equity)**: 净资产收益率。
 - **ROA (Net Income / Total Assets)**: 总资产收益率。
 - **ACC (Accruals)**: 应计项目。
 $ACC = (NetIncome - CFO) / TotalAssets$ 。低应计项目（现金流好于净利润）代表高质量，反之则暗示盈余质量低或存在虚增利润的嫌疑。
2. **成长性（Growth）**：过去五年盈利能力指标（GPRO, ROE, ROA 等）的变动率。高质量企业不仅当前盈利强，且盈利能力在持续提升。
3. **安全性（Safety）**：衡量企业陷入财务困境的风险。
 - **BAB (Betting Against Beta)**: 低 Beta 值。高质量股票通常对市场波动不敏感。
 - **IVOL (Idiosyncratic Volatility)**: 低特质波动率。
 - **LEV (Leverage)**: 低财务杠杆（Total Debt / Assets）。
 - **O-Score / Z-Score**: 破产风险评分。

4. 支付能力 (Payout)：衡量管理层对股东的友好程度。
- **EISS (Equity Issuance)**: 净股票发行。回购股票（负发行）为正分，增发股票稀释股东权益为负分。
 - **DISS (Debt Issuance)**: 净债务发行。

4.2 量化风控：Altman Z-Score 与 Merton 模型的深度应用

在价值投资中，识别并避开即将违约的企业是生存的关键。量化模型通过计算“违约距离”来实现这一目标。

1. Altman Z-Score 模型：财务报表的体检表 Altman Z-Score 是基于会计数据的经典破产预测模型，特别适用于剔除制造业和周期性行业的潜在违约者²¹。核心公式：

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5$$

- X_1 (Working Capital / Total Assets): 衡量短期流动性。
- X_2 (Retained Earnings / Total Assets): 衡量累积盈利能力和企业年龄（新公司此项通常较低）。
- X_3 (EBIT / Total Assets): 核心资产盈利能力，权重最大，是衡量企业生命力的关键。
- X_4 (Market Value of Equity / Total Liabilities): 衡量偿债杠杆，引入了市场价格信息，反映投资者对企业前景的信心。
- X_5 (Sales / Total Assets): 资产周转率，衡量管理层利用资产产生收入的能力。

应用阈值：一般认为 $Z < 1.8$ 为破产高危区 (Distress Zone)， $1.8 < Z < 3.0$ 为灰色区域， $Z > 3.0$ 为安全区域。在构建价值组合时，强制剔除所有 $Z < 1.8$ 的股票，可以显著降低“踩雷”概率。

2. Merton 距离违约模型 (Distance to Default, DD)：市场视角的风控 相比于滞后的会计数据，基于期权定价理论的 Merton 模型利用股票市场的实时波动率来评估信用风险。它将股权视为看涨期权 (Call Option)，标的资产是公司总资产，执行价格是公司债务面值。利用 QuantLib 库，可以计算出公司资产价值跌破债务面值的概率。研究表明，在中国市场和美股市场，Altman Z-Score 与 Merton 模型结合使用，能提供比单一模型更准确的违约预警²³。

第五章 经典选股模型的量化实现与优化

5.1 格雷厄姆“捡烟蒂”策略的现代量化复刻

格雷厄姆在《证券分析》中提出的 NCAV (Net Current Asset Value) 策略, 即以低于“净流动资产价值”(流动资产 - 总负债) 三分之二的价格买入股票, 是深度价值投资的鼻祖。

Python 量化实现逻辑与现代改良:

1. **基础筛选:** $MarketCap < (CurrentAssets - TotalLiabilities) \times 0.67$ 。
2. **现代市场的困境:** 在现代高效率市场, 严格符合此条件的股票极少, 且多为流动性枯竭的微盘股。
3. **改良版双重过滤器:**
 - **估值放宽:** 改为筛选 $P/B < 0.8$ 且 $P/E < 10$ 。
 - **负债约束:** 增加 $Debt/Equity < 0.5$, 确保资产负债表健康。
 - **流动性过滤器:** 必须加入日均成交额 (ADV) > 500 万美元的条件, 否则回测出的高收益在实盘中无法成交 (冲击成本过大) ²⁴。

5.2 乔尔·格林布拉特的神奇公式 (Magic Formula)

神奇公式是 Quantamental 的早期代表, 它通过两个简单的因子排名来选股, 试图以低廉的价格买入优质的公司 ²⁵。

核心公式与逻辑:

1. **资本回报率 (ROC) = $EBIT / (Net\ Working\ Capital + Net\ Fixed\ Assets)$**
 - **衡量好公司:** 剔除财务杠杆和现金储备的影响, 纯粹衡量企业利用有形资本 (运营资本 + 固定资产) 创造息税前利润的效率。
2. **盈利率 (Earnings Yield) = $EBIT / Enterprise\ Value$**
 - **衡量好价格:** 相比 P/E, 使用 EBIT/EV 剔除了资本结构 (债务水平) 和税率的影响。它本质上是计算如果买下整家公司 (股权+债权), 能获得多少息税前收益率。

量化执行步骤 (Python 伪代码逻辑):

1. **剔除行业:** 剔除金融股 (资产负债表结构特殊) 和公用事业股 (受管制)。
2. **数据计算:** 计算全市场股票的 ROC 和 Earnings Yield。
3. **双重排名:** 分别对 ROC 和 Earnings Yield 进行降序排名。例如, ROC 最高的股票得 1 分, 第 100 名得 100 分。
4. **综合评分:** $Score = Rank(ROC) + Rank(EarningsYield)$ 。
5. **组合构建:** 按总分从小到大排序, 买入排名前 30 的股票, 等权持有。
6. **优化:** 研究表明, 在神奇公式的基础上叠加**动量因子** (即在排名前 30 的股票中, 只买入过去 6 个月价格涨幅为正的股票), 可以显著降低策略的最大回撤, 避免买入正在暴跌的“价值陷阱” ²⁸。

5.3 Piotroski F-Score: 基本面趋势扫描

斯坦福教授 Joseph Piotroski 提出的 F-Score 主要用于在低市净率（High B/M）的股票池中筛选出基本面正在改善的公司。它是一个 0-9 分的二元计分系统，不仅关注当前状态，更关注**边际变化**。

9 分制评分模型细则：

- **盈利性（4 分）：**
 1. $ROA > 0$ （当前盈利）
 2. $CFO > 0$ （经营性现金流为正）
 3. $ROA(t) > ROA(t-1)$ （盈利能力同比提升）
 4. $CFO > Net\ Income$ （盈余质量高，非纸面富贵）
- **杠杆与流动性（3 分）：**
 5. $Long\ Term\ Debt(t) < Long\ Term\ Debt(t-1)$ （去杠杆）
 6. $Current\ Ratio(t) > Current\ Ratio(t-1)$ （流动性改善）
 7. 未增发新股（无股权稀释）
- **运营效率（2 分）：**
 8. $Gross\ Margin(t) > Gross\ Margin(t-1)$ （毛利率提升，暗示产品竞争力增强）
 9. $Asset\ Turnover(t) > Asset\ Turnover(t-1)$ （资产周转率提升，运营效率提高）

策略融合： 在价值因子打底的基础上，叠加 $F - Score \geq 7$ 的筛选条件，可以显著提升

Alpha 收益。尤其是在空头端，做空高估值且 $F - Score \leq 3$ 的股票，是量化对冲基金常用的 Alpha 策略。

第六章 高阶策略一：多因子融合与择时机制

6.1 价值与动量的互补性之谜：负相关的力量

学术界和业界公认的一个现象是：****价值（Value）和动量（Momentum）****在长期均能产生正向超额收益，但两者的相关性却为负（尤其是在短期，相关系数常在 -0.4 到 -0.6 之间）。

- **行为学解释：** 价值投资者倾向于“左侧交易”，在价格下跌、市场悲观时买入被低估的资产；动量投资者倾向于“右侧交易”，在价格上涨、趋势形成时追入。
- **互补效应：** 当市场发生反转时（如 2009 年复苏或 2020 年 3 月反弹），价值因子往往表现强劲，而动量因子（持有前期的赢家，即抗跌的防御股）则会遭受重创（Momentum Crash）；反之，在趋势延续的市场中，动量表现优异，价值则可能表现平平。将两者结合，利用其负相关性，可以极大地平滑组合波动，提升夏普比率（Sharpe Ratio）²⁹。

6.2 融合策略的构建路径

1. 组合层面融合（Portfolio Level Combination）：

- 分别构建纯价值组合（Top 10% Value）和纯动量组合（Top 10% Momentum）。
- 将资金 50/50 分配给两个组合，并定期（如季度或年度）再平衡。
- **优势：** 操作简单，且保留了各自因子的纯度。研究表明，这种简单的等权组合通常能获得优异的风险调整后收益³⁰。

2. 信号层面融合（Signal Level Combination）：

- 对每只股票分别计算价值 Z-Score 和动量 Z-Score。
- 计算综合得分： $Score_{Total} = w_1 \times Z_{Value} + w_2 \times Z_{Momentum}$ 。
- 买入综合得分最高的股票。
- **劣势：** 同时具备深度价值（极便宜）和强劲动量（涨得好）的股票非常稀缺，可能导致选股池过于狭窄，或者选出的股票在两个维度上都只是平庸（既不够便宜，涨势也不够强）。因此，在 A 股等市场，组合层面融合往往优于信号层面融合³⁰。

3. 动态风险加权（Dynamic Risk Weighting）：

- 不采用固定的 50/50 权重，而是根据两个因子的近期波动率进行反向加权（Risk Parity 思想）。
- **机制：** 当动量因子的波动率飙升时（通常发生在市场崩盘前夕或剧烈反转期），模型自动降低动量因子的权重，增加价值因子的权重。S&P Global 的研究显示，这种动态加权策略在历史上能有效规避动量崩溃（Momentum Crashes），显著降低最大回撤³³。

第七章 高阶策略二：统计套利与行业中性化多空

7.1 统计套利的进化：从价格协整到基本面收敛

传统的配对交易（Pairs Trading）主要基于价格序列的协整关系（Cointegration）或最小距离法（Distance Method），寻找价格走势高度相关的两只股票（如可口可乐与百事可乐）。当价差（Spread）偏离历史均值超过 2 个标准差时，做空赢家、做多输家，赌其均值回归³⁴。

痛点： 纯价格驱动的配对交易面临**“结构性断裂”（Structural Break）**风险。如果两只股票价差扩大是因为其中一家公司基本面彻底恶化（如爆出财务丑闻或丢失大客户），价差可能永远不会回归，导致交易者在“死扛”中遭受巨大亏损。

7.2 基于估值价差（Valuation Spread）的配对策略

为了解决上述问题，Quantamental 策略将基本面引入配对逻辑，构建基于估值的统计套利体系。

策略逻辑与步骤：

1. **配对选择：** 依然在同行业内寻找业务模式相似、市值相近的公司。

2. **信号触发：** 不仅观测价格比率（Price Ratio），更观测估值比率（P/E Ratio Spread）。
 - 如果 Stock A 的 P/E 从历史均值 15x 跌到 10x，而 Stock B 维持 15x，且两者基本面（如预期增长率）未发生重大偏离，则做多 A / 做空 B。
3. **基本面过滤器（Fundamental Filter）：** 这是关键一步。在开仓前，必须确认 P/E 的偏离不是由盈利（E）的永久性下滑引起的。
 - **检查手段：** 结合 **Piotroski F-Score** 或 **分析师盈利预测修正（Analyst Revision）**。如果 Stock A 的 P/E 下降是因为分析师大幅下调了明年的盈利预测，说明其“便宜”是有原因的，模型应禁止开仓。
 - **Ornstein-Uhlenbeck (OU) 过程：** 利用 OU 随机过程对估值价差进行建模，计算均值回归的半衰期（Half-life）。只交易那些半衰期较短（回归速度快）的配对，避免资金长期被占用³⁶。

7.3 行业中性化多空策略（Sector Neutral Long/Short）

这是一种更宏观的配对概念，不局限于两只股票，而是在整个行业板块内进行对冲，旨在剥离市场风险（Beta）和行业风险（Industry Beta），纯粹博取选股能力的 Alpha。

构建方法：

1. **行业划分：** 将全市场股票按 GICS 行业分类。
2. **行业内打分：** 在每个行业内部，利用多因子模型（如 Value + Quality + Momentum）对所有股票进行相对排名。
3. **多空构建：**
 - **多头（Long）：** 买入每个行业内得分前 20% 的股票。
 - **空头（Short）：** 卖空每个行业内得分后 20% 的股票（或卖空该行业的 ETF）。
4. **优势：** 这种策略完全对冲了行业波动。例如，在油价暴跌导致能源板块整体下跌 30% 时，高质量、低成本的石油公司可能只跌 20%，而低质量、高杠杆的公司可能跌 40%。行业中性策略通过做多前者、做空后者，依然可以获得 10% 的净收益（ $-20\% - (-40\%) = +20\%$ ），实现“市场下跌中赚钱”³⁸。

第八章 前沿探索：另类数据与 NLP 在价值发现中的应用

8.1 自然语言处理（NLP）：量化“软信息”

传统量化只能处理结构化数据（如财报数字），而 NLP 技术使得计算机能够阅读新闻、研报和财报电话会议（Earnings Calls），从而量化管理层信心、企业文化等“软信息”，填补了基本面量化的最后一块拼图。

1. 财报电话会议的深度解析：

- **技术工具：** 使用 **BERT** 或专门针对金融领域的 **FinBERT** 预训练模型处理电话会议的文本记

录 (Transcripts) ⁴¹。

- **核心信号提取：**
 - **净情绪得分 (Net Sentiment)：** 基于 Loughran and McDonald (2011) 金融情感词典，计算正面词汇与负面词汇的频率差。研究发现，高管在**问答环节 (Q&A)** 的情绪比演讲环节 (Presentation) 更具预测性，因为 Q&A 更难照本宣科，更能暴露真实情况 ⁴³。
 - **模糊度 (Vagueness)：** 当管理层使用大量模糊词汇 (如 "might", "could", "approximately", "hope") 时，往往预示着未来业绩波动加大或存在隐瞒。
 - **回答回避度 (Non-answer metrics)：** 利用向量空间模型计算分析师提问与高管回答之间的语义距离 (Cosine Similarity)。如果距离过大，说明答非所问，通常是负面信号 ⁴⁵。
- **Quantamental 应用：** 将 NLP 情绪得分作为一个因子，叠加在价值模型上。对于低 P/E 股票，如果管理层在电话会议中表现出极高的信心且措辞精准 (低模糊度)，则大概率是市场错杀的黄金坑。

2. 经济护城河的文本量化：

- 通过分析企业年报 (10-K) 中关于“竞争优势”、“品牌忠诚度”、“转换成本”等关键词的上下文，可以构建“护城河指数”。研究表明，该指数与高质量因子 (Quality) 高度相关，可用于筛选那些定性优势明显的长牛股 ⁴⁶。

8.2 卫星图像与供应链数据：营收预测的“上帝视角”

对于零售、大宗商品和制造业，另类数据 (Alternative Data) 提供了实时的基本面数据，往往比财报提前数月，为价值投资者提供了巨大的信息优势。

- **零售业营收预测：** 通过卫星图像分析沃尔玛 (Walmart) 或家得宝 (Home Depot) 停车场在季度内的车辆密度变化，结合机器学习模型，可以高精度预测其季度营收。如果模型预测营收大增，而股价仍因宏观悲观情绪处于低位 (低 P/E)，这是一个极高置信度的买入信号 ⁴⁷。
- **原油库存监测：** 通过卫星监测储油罐的浮顶高度 (利用阴影长度计算)，可以实时计算全球原油库存水平。这对于能源股的价值判断至关重要，帮助投资者在库存下降 (需求回暖) 初期布局被低估的能源公司 ⁴⁸。

第九章 执行、风险管理与回测陷阱

9.1 换手率与交易成本管理

价值投资通常是低换手的，而引入量化因子 (尤其是动量) 容易导致高换手。在结合两者时，必须严格控制交易成本，否则微薄的 Alpha 会被手续费和滑点 (Slippage) 吞噬。

优化策略：交易缓冲区（No-trade Buffer）

- **原理：** 设定一个缓冲区（如目标权重的 $\pm 10\%$ ）。假设某股票的目标权重是 5%。
 - 如果当前权重在 4.5% - 5.5% 之间，即使模型给出的新目标是 5.2%，也不进行交易。
 - 只有当权重偏离超过缓冲区（例如跌至 4.4% 或涨至 5.6%）时，才触发再平衡交易，将其调回目标权重。
- **效果：** 模拟回测显示，这种简单的启发式规则可以将换手率降低 30%-50%，而对组合收益率的影响微乎其微，显著提升了扣费后的净回报⁴⁹。

9.2 回测中的常见陷阱与防御

在开发 Quantamental 策略时，需警惕以下陷阱，确保回测结果的真实性：

1. **幸存者偏差（Survivorship Bias）：** 数据库必须包含已经退市或破产的股票。如果只测试当前存在的股票（S&P 500 成分股），会人为高估价值策略的回报，因为那些因为“便宜”买入但最终破产的公司被剔除了。必须使用包含“死股票”的全历史数据库（Dead-listing Database）²⁸。
2. **前视偏差（Look-ahead Bias）：** 再次强调，必须使用 PIT 数据。特别是计算 Z-Score 或 F-Score 时，这些指标对财务数据的时效性非常敏感。
3. **流动性陷阱：** 许多高 Alpha 的价值策略（如 NCAV）往往集中在微盘股。回测时必须加入严格的流动性约束（如剔除市值后 20% 的股票，或限制单笔交易量不超过日均成交量的 2%），并设置合理的冲击成本模型（Transaction Cost Model）²⁴。

9.3 风险模型：VaR 与回测检验

除了追求收益，量化系统必须包含严格的风险度量。

- **Value at Risk (VaR):** 计算在 95% 置信度下，组合在未来一天的最大可能损失。
- **VaR 回测（Backtesting VaR）：** 使用 Kupiec's POF test 或 Christoffersen's independence test 来检验风险模型的准确性。如果实际损失超过 VaR 的频率显著高于 5%，说明风险模型低估了尾部风险（Fat Tail），需引入极值理论（EVT）进行修正⁵⁰。

结论

价值投资与量化交易并非非此即彼的对立关系，而是现代投资管理中不可或缺的两个维度。价值投资提供了深刻的商业洞察和安全边际思想，是策略的灵魂；量化交易提供了处理大数据的能力、严格的纪律和风险管理框架，是策略的骨架。

未来的 Quantamental 投资将呈现以下趋势：

1. **因子更纯粹：** 通过精细的数学处理（正交化、去噪），剥离干扰项，还原价值和质量因子的本质。

2. **数据更多维：** 从枯燥的财务报表扩展到管理层语调（NLP）、卫星图像等另类数据，全息重构企业基本面。
3. **模型更智能：** 利用机器学习捕捉因子之间非线性的相互作用（例如，在不同市场环境下动态调整价值与动量的权重）。

对于投资者而言，掌握 Python 编程和数据科学技能，并将其植根于深厚的价值投资哲学之中，将是在未来高效率市场中获取超额收益的关键路径。

主要参考文献引用索引

- **Quantamental 框架与定义：** ¹
- **价值因子与贴现率冲击：** ¹⁶
- **AQR 质量因子（QMJ）：** ⁶
- **价值与动量结合：** ²⁹
- **配对交易与统计套利：** ³⁴
- **Python 与回测技术：** ⁹
- **NLP 与另类数据：** ⁵
- **选股模型（神奇公式、Z-Score）：** ²¹
- **执行与风控：** ⁷

Works cited

1. Quantamental Investing - quant Mutual Fund, accessed January 27, 2026, https://quantmutual.com/Admin/Pdf/quant_quantamental_Fund_Presentation.pdf
2. Quantamental Investing: Merging Quantitative and Fundamental Analysis | by mids.capital, accessed January 27, 2026, <https://midscapital.medium.com/quantamental-investing-merging-quantitative-and-fundamental-analysis-c7656eeecf41>
3. Quantamental investing: The future is now - Morgan Stanley, accessed January 27, 2026, https://www.morganstanley.com/im/publication/insights/investment-insights/ii_quantamentalinvesting_us.pdf
4. The QVGS Framework – Implementing Quantamental Investing - Modulor Capital, accessed January 27, 2026, <https://modulorcapital.com/the-qvgs-framework-implementing-quantamental-investing/>
5. LEVERAGING TEXT MINING TO EXTRACT INSIGHTS FROM EARNINGS CALL TRANSCRIPTS - AllianceBernstein, accessed January 27, 2026, <https://www.alliancebernstein.com/content/dam/global/insights/insights-whitepapers/leveragingtextminingtoextractinsights-chinfan.pdf>
6. An Implementation of Quality Minus Junk - CRAN, accessed January 27, 2026,

- <https://cran.r-project.org/web/packages/qmj/vignettes/qmj.Rnw>
7. a Standardized Benchmark of Look-ahead Bias in Point-in-Time LLMs for Finance - arXiv, accessed January 27, 2026, <https://arxiv.org/pdf/2601.13770>
 8. Importance of Point-in-Time for Alternative Data - Eagle Alpha, accessed January 27, 2026, <https://www.eaglealpha.com/2024/05/06/point-in-time-alternative-data/>
 9. Python for Algorithmic Trading: Essential Libraries - LuxAlgo, accessed January 27, 2026, <https://www.luxalgo.com/blog/python-for-algorithmic-trading-essential-libraries/>
 10. 10 Best Python Backtesting Libraries for Trading Strategies - QuantVPS, accessed January 27, 2026, <https://www.quantvps.com/blog/best-python-backtesting-libraries-for-trading>
 11. Backtesting.py – An Introductory Guide to Backtesting with Python - Interactive Brokers, accessed January 27, 2026, <https://www.interactivebrokers.com/campus/ibkr-quant-news/backtesting-py-an-introductory-guide-to-backtesting-with-python/>
 12. 7 Essential Python Packages for Finance | by Silva.f.francis - Medium, accessed January 27, 2026, <https://medium.com/@silva.f.francis/7-essential-python-packages-for-finance-9161dbdb5926>
 13. Automating Data Analysis with Python Dashboards - The CPA Journal, accessed January 27, 2026, <https://www.cpajournal.com/2025/09/17/automating-data-analysis-with-python-dashboards/>
 14. Financial Analytics Using Python and AI Tools - Dan Gode, accessed January 27, 2026, <https://www.dangode.com/financialanalytics/>
 15. Quantitative analysis model with Python. Extrapolates variance, standard deviation, correlation and beta from a list of securities in a few clicks. - GitHub, accessed January 27, 2026, <https://github.com/AI0010/quantitative-model-analysis-python>
 16. Resurrecting the Value Factor from its Redundancy - Financial ..., accessed January 27, 2026, <https://www.fma.org/assets/docs/Consortium2023/Paper.pdf>
 17. Comparing Value Factor Performance in Global Equity Markets | LSEG, accessed January 27, 2026, https://www.lseg.com/content/dam/ftse-russell/en_us/documents/research/comparing-value-factor-performance-gobal-equity-markets.pdf
 18. Quality Minus Junk - NHH, accessed January 27, 2026, <https://www.nhh.no/globalassets/departments/business-and-management-science/seminars/2015-fall/020915.pdf>
 19. Quality Minus Junk - AQR Capital Management, accessed January 27, 2026, <https://www.aqr.com/Insights/Research/Working-Paper/Quality-Minus-Junk>
 20. Quality Minus Junk - Research@CBS, accessed January 27, 2026, [https://research-api.cbs.dk/ws/portalfiles/portal/60211462/lasse heje pedersen et al quality minus junk publishersversion.pdf](https://research-api.cbs.dk/ws/portalfiles/portal/60211462/lasse%20heje%20pedersen%20et%20al%20quality%20minus%20junk%20publishersversion.pdf)

21. Applications of Distress Prediction Models: What Have We Learned After 50 Years from the Z-Score Models? - MDPI, accessed January 27, 2026, <https://www.mdpi.com/2227-7072/6/3/70>
22. Financial distress prediction using integrated Z-score and multilayer perceptron neural networks - PMC - PubMed Central, accessed January 27, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9675979/>
23. Application of the Merton Model and the Altman Z-score Model in Credit Risk Assessment - Lund University Publications, accessed January 27, 2026, <https://lup.lub.lu.se/student-papers/record/9118955/file/9118972.pdf>
24. Understanding The Benjamin Graham Formula Correctly | GrahamValue, accessed January 27, 2026, <https://www.grahamvalue.com/article/understanding-benjamin-graham-formula-correctly>
25. How to implement the Magic Formula investment strategy - Quant-Investing, accessed January 27, 2026, <https://www.quant-investing.com/blog/how-to-implement-a-magic-formula-investment-strategy>
26. What Is Magic Formula Investing? The Joel Greenblatt Systematic Strategy, accessed January 27, 2026, <https://www.investing.com/academy/analysis/what-is-magic-formula-investing/>
27. Magic Formula has its magic and Momentum has its moments. - Diva-portal.org, accessed January 27, 2026, <https://www.diva-portal.org/smash/get/diva2:1326926/FULLTEXT01.pdf>
28. Backtesting Greenblatt's Magic Formula : r/SecurityAnalysis - Reddit, accessed January 27, 2026, https://www.reddit.com/r/SecurityAnalysis/comments/gyyx4a/backtesting_greenblatts_magic_formula/
29. accessed January 27, 2026, <https://alphaarchitect.com/using-value-and-momentum-for-stock-selection-and-market-timing/#:~:text=Both%20the%20Valuation%20and%20Momentum,highest%20Sharpe%20and%20Sortino%20ratios.>
30. Should You Combine Value and Momentum Investing?, accessed January 27, 2026, <https://www.quant-investing.com/blog/should-you-mix-value-and-momentum-strategies>
31. Value and Momentum Everywhere: Factors, Monthly - AQR Capital Management, accessed January 27, 2026, <https://www.aqr.com/Insights/Datasets/Value-and-Momentum-Everywhere-Factors-Monthly>
32. Multi-factor: Why it takes value, quality, and momentum - Aberdeen Investments, accessed January 27, 2026, <https://www.aberdeeninvestments.com/en-us/institutional/insights-and-research/io-2024-multi-factor-why-it-takes-value-quality-momentum>
33. Value and Momentum: Everywhere, But Not All the Time - S&P Global, accessed January 27, 2026, <https://www.spglobal.com/content/dam/spglobal/mi/en/documents/general/S-P->

[Global-Market-Intelligence-Value-Momentum-January-2019.pdf](#)

34. Pairs Trading: Performance of a Relative-Value Arbitrage Rule - Wharton Statistics, accessed January 27, 2026, <http://stat.wharton.upenn.edu/~steele/Courses/434/434Context/PairsTrading/PairsTradingGGR.pdf>
35. Statistical Arbitrage Pairs Trading Strategies - Institute for Economics, accessed January 27, 2026, <https://www.iwf.rw.fau.de/files/2016/03/09-2015.pdf>
36. Statistical Arbitrage: Strategies, Risks, and How It Works - QuantInsti Blog, accessed January 27, 2026, <https://blog.quantinsti.com/statistical-arbitrage/>
37. A Diversification Framework for Multiple Pairs Trading Strategies - MDPI, accessed January 27, 2026, <https://www.mdpi.com/2227-9091/11/5/93>
38. Long-Short Equity Strategy using Ranking: : Simple Trading Strategies Part 4 | by Auquan, accessed January 27, 2026, <https://medium.com/auquan/long-short-equity-trading-strategy-daa41d00a036>
39. Tutorials/Long-Short Strategies using Ranking.ipynb at master - GitHub, accessed January 27, 2026, <https://github.com/Auquan/Tutorials/blob/master/Long-Short%20Strategies%20using%20Ranking.ipynb>
40. Long-Short Equity (L/S) | Fund Investing Strategy - Wall Street Prep, accessed January 27, 2026, <https://www.wallstreetprep.com/knowledge/long-short-equity-investing/>
41. Evaluating Local Language Models: An Application to Financial Earnings Calls - Federal Reserve Bank of Kansas City, accessed January 27, 2026, <https://www.kansascityfed.org/documents/9862/rwp23-12cookkazinnikhansenmcadam.pdf>
42. News to Numbers: NLP Stock Return Predictions - Stanford University, accessed January 27, 2026, <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1254/final-reports/256968470.pdf>
43. Natural language processing (nlp) for financial text analysis - ResearchGate, accessed January 27, 2026, https://www.researchgate.net/publication/385860012_Natural_language_processing_nlp_for_financial_text_analysis
44. Natural Language Processing - Part I: Primer - S&P Global, accessed January 27, 2026, <https://www.spglobal.com/content/dam/spglobal/mi/en/documents/general/SP-Global-Market-Intelligence-NLP-Primer-September-2018.pdf>
45. Assessing the Predictive Power of Earnings Call Transcripts on Next-Day Stock Price Movement: A Semantic Analysis Using Large Language Models, accessed January 27, 2026, <https://www.ischool.berkeley.edu/projects/2024/assessing-predictive-power-earnings-call-transcripts-next-day-stock-price-movement>
46. Text as Data - Stanford University, accessed January 27, 2026, <https://web.stanford.edu/~gentzkow/research/text-as-data.pdf>

47. Deep learning with satellite images enables high-resolution income estimation: A case study of Buenos Aires | PLOS One - Research journals, accessed January 27, 2026, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0338110>
48. How Satellite Imagery is Revolutionizing the Way we Invest | by Kolemnn Lutz | Medium, accessed January 27, 2026, <https://kolemannlutz.medium.com/how-satellite-imagery-is-revolutionizing-the-way-we-invest-da44bfb95d89>
49. Reducing Quant Trading Costs: Simple Turnover Solution - Robot Wealth, accessed January 27, 2026, <https://robotwealth.com/a-simple-effective-way-to-manage-turnover-and-not-get-killed-by-costs/>
50. A review of backtesting for value at risk - The University of Manchester, accessed January 27, 2026, <https://pure.manchester.ac.uk/ws/files/60673220/back4.pdf>
51. A Review of Backtesting and Backtesting Procedures - Federal Reserve Board, accessed January 27, 2026, <https://www.federalreserve.gov/pubs/feds/2005/200521/200521pap.pdf>
52. What is Quantamental Investing? - International Banker, accessed January 27, 2026, <https://internationalbanker.com/brokerage/what-is-quantamental-investing/>
53. Quality Minus Junk: Factors, Daily - AQR Capital Management, accessed January 27, 2026, <https://www.aqr.com/Insights/Datasets/Quality-Minus-Junk-Factors-Daily>
54. FACTOR MOMENTUM WITHIN FACTORS | Northern Trust, accessed January 27, 2026, <https://www.northerntrust.com/content/dam/northerntrust/pws/nt/documents/investment-management/factor-momentum-within-factors.pdf>
55. Mean Reversion II: Pairs Trading Strategies - RavenPack, accessed January 27, 2026, <https://www.ravenpack.com/research/mean-reversion-pairs-trading-strategy/>
56. How I Built an Event-Driven Backtesting Engine in Python | by Timothy Kimutai | Medium, accessed January 27, 2026, <https://timkimutai.medium.com/how-i-built-an-event-driven-backtesting-engine-in-python-25179a80cde0>
57. Natural Language Processing, Part II: Stock Selection | S&P Global, accessed January 27, 2026, <https://www.spglobal.com/market-intelligence/en/news-insights/research/natural-language-processing-part-ii-stock-selection>