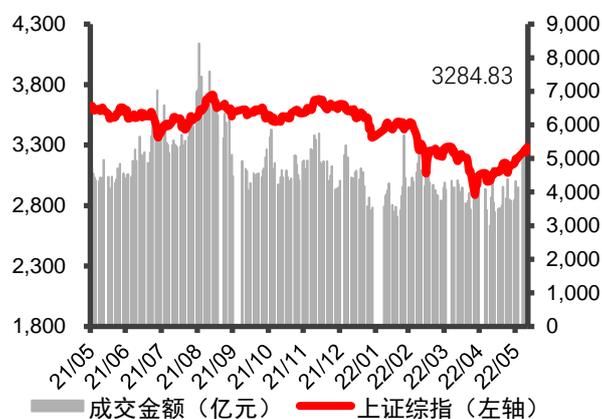
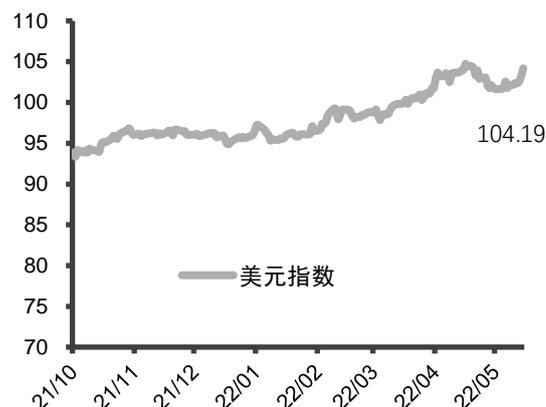


上证综合指数近一年走势图



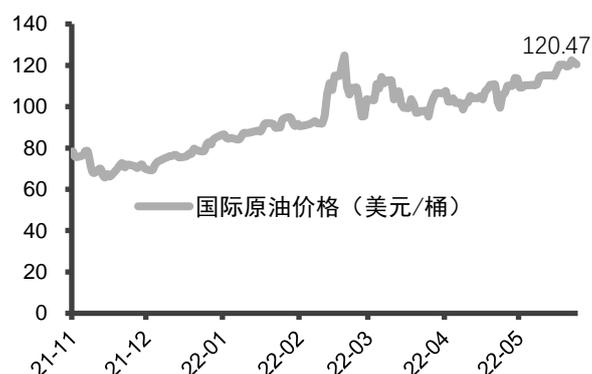
数据来源: 万得, 截止日期: 2022.6.10

美元指数近一年走势图



数据来源: 万得, 截止日期: 2022.6.10

国际原油价格近一年走势图



数据来源: 万得, 截止日期: 2022.6.10

## 基金经理论市: 美国通胀创历史新高, 国内社融超预期

本周 A 股表现继续强于海外市场, 上证综指上涨 2.80%, 深证成指上涨 3.50%, 创业板指数上涨 4.00%。分板块看, 有色金属、化工、非银金融、电力设备和食品饮料等板块表现较好; 轻工、电子、家电、通信和建材等板块表现较差。国内的疫情逐渐好转, A 股开始逐渐脱离美股的影响, 走出独立行情。

海外经济方面, 美国通胀创历史新高。CPI 通胀同比上涨 8.6%, 创 1981 年 12 月以来新高, 高于上月及预期的 8.3% 增幅, 核心 CPI 5 月同比上涨 6%, 较 4 月的 6.2% 回落 0.2%。能源成本主导了整体 CPI 的飙升, 5 月商品价格同比增长 8.5%, 较上月 9.7%, 继续回落 1.2%; 5 月服务价格同比增长 5.2%, 较上月 4.9% 上行 0.2 个百分点。通胀高企, 将在未来一段时间内制约美联储的货币政策。

国内经济方面, 5 月份国内新增社融超预期。5 月新增社融 2.79 万亿元, 较去年同期多增 8399 亿元, 较 4 月多增 1.88 万亿元, 高于市场预期。5 月社融存量同比增长 10.5%, 较 4 月上升 0.3 个百分点。5 月社融数据中, 贷款数据结构仍然较差, 企业中长贷呈现同比少增, 票据融资创历史新高, 金融机构存在以票据冲量的现象。随着疫情之后, 经济内生增长动能的逐渐修复, 国内的社融有望逐渐企稳回升, 结构也将得以改善。

总体上, 随着国内疫情的好转, 市场的风险偏好逐渐修复, A 股的走势受到美股的压制逐渐减小。目前中美十年期国债的利差基本上处于相对稳定的水平, 美联储加息对 A 股流动性的影响正在逐渐消除。北京和上海的新冠疫情已经取得了阶段性胜利, 市场的信心有望显著修复, 中央政策对经济的刺激政策的重要抓手是刺激汽车等居民消费, 相关板块的估值可能显著提升。

## 专栏：文本挖掘在因子开发中的运用

在评估一个量化模型的优势时，我们常常会考量它在因子储备上所能达到的深度和广度。因为只有因子积累得足够多，才有机会具备在任何复杂的市场环境下都能获取阿尔法回报的可能性。但随着量化投资的发展进入到一个相对成熟的阶段，大部分有效因子都已成为各家量化团队的标配，想要在传统数据框架下开发新的阿尔法因子已经愈发困难。不过近年来逐步兴起的人工智能技术却为新因子的开发打开了视野，本文就主要介绍文本挖掘技术在因子开发中的一些应用案例。

首先一个简单的案例就是可以通过文字匹配的方式来判断一篇研究报告的投资倾向。比如我们要从一篇分析师所撰写的研究报告中了解到他对这家上市公司的看法究竟是正面或是负面，乃至推荐力度如何，我们通常会从该分析师在行文中的语义表达去进行感性认知。但显然我们没有足够的精力去阅读所有分析师和所有上市公司的相关研报，那么通过文字匹配的方法来进行模糊判断就不失为一个次优选择。比如我们可以在每一篇研报中去搜索已经被我们定义为“正向”或“负向”的字眼，进而对该分析师在该公司上的投资倾向进行评估。“正向”的字眼可以包括“推荐”、“利好”、“改善”和“强劲”等，“负向”的字眼可以包括“低于预期”、“不及预期”、“下行”和“停滞”等。尽管不能说通过简单的文字匹配就能精确地度量分析师的投资倾向，但在纳入所有分析师和所有上市公司的相关研报后，误差还是能够逐步下降到一个可以接受的范围之内，我们据此也就可以开发出像分析师推荐度等一系列因子。

更复杂的文本挖掘技术则需要运用到像新闻、论坛、微博等舆情数据上。因为这些信息渠道不存在统一、标准化的语言表达方式，所以不能使用简单的文字匹配方法去进行判断（这样只会导致大量的漏判和误判），而需要引入一个叫做“情感分析”的文本分析工具。情感分析也可以被理解为一个文本分析的黑箱子，只要向这个箱子输入一段文字，他就可以输出相应的文本分析结果，比如这段文字的情感色彩如何，是乐观还是悲观等。当然，考虑到中文语言的博大精深，情感分析的精确度难以无限度趋近于 100%，但值得肯定的是，这个黑箱子通常也具备反复迭代和优化的能力，所以随着经验的累积我们终可以获得能尽可能让我们满意的分析结果。通过对语义明确的文本数据进行统计分析我们又可以生产出诸如投资者关注热度、看好程度等一系列舆情因子。

最后，从上市公司公告中我们也能通过文本挖掘的方式获取到一些有意思的另类数据。比如有些上市公司会披露其与其他公司签订的合同或达成的协议等信息，我们就可以通过文本挖掘的方式将这些信息给抓取下来，然后搭建标准化的上市公司关联数据库，在其中记录交易双方信息、合同金额、交易时间等信息，这样就可以便于我们对由某一家上市公司延展开的整个供应链网络进行全貌分析，也能藉此开发相应的选股因子。以上均是文本挖掘技术在因子开发中的一些案例展示，具体研究方向还有很多，就不一一展开了。

（专栏作者：摩根士丹利华鑫基金数量化投资部 王联欣）