



Barra 模型专题报告（五）：质量、增长因子

摘要

2018 年 8 月 MSCI 明晟公司发布了针对中国市场的模型 Barra China Equity Model (CNE6)，包括 9 个一级因子、21 个二级因子、46 个三级因子。本文介绍了因子计算方法，以沪深 300 样本股为范围，回溯了 2015 年至 2020 年季报和年报中财务因子表现，并对回测结果给出了现实意义的解释。

本期分析的五类因子中，资产负债率因子表现优异，追平指数增长，说明投资者对于高负债运营企业的看空和较强的避险意识。盈利波动率因子近乎水平直线，说明投资者对财报数据的波动并不敏感。4 个盈利能力因子在 2018 年前后出现分化，与净利润相关的 2 个因子持续有效，与毛利润相关的另 2 个因子急转直下，揭示出市场对盈利结构的关注点发生了转移。增长率因子中，市场认为总资产和资本支出的增长偏向利好，增发股票偏向利空，对每股收益的增长已充分挖掘。

值得注意的是，部分因子的收益曲线在 2018 年年报前后不约而同的出现了反转，也许是受宏观经济预期的影响，投资风格和关注重点发生改变。

格林大华量化研究

联系电话：0371-56518942

独立性声明：

作者保证报告所采用的数据均来自合规渠道，分析逻辑基于本人的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

一. Leverage, 杠杆

1. Market Leverage, 市场杠杆

Style: Leverage

Descriptors: MLEV

Market Leverage

Computed as,

$$MLEV = \frac{ME + PE + LD}{ME}$$

where ME is the market value of common equity on the last trading day, and PE and LD are the preferred equity and long-term debt, respectively, from the last fiscal year.

由于 A 股市场没有优先股，当日的普通股市场价值加上最近财年的长期债务，再除以当日的普通股市场价值，即可得到因子数值，该因子的周期为日度。

由公式可见，因子的核心部分为长期债务除以市场价值，在此基础上加一即因子数值，因子以市场价值为基准来度量企业的债务水平，称为市场杠杆。在生活中的杠杆有跷跷板、启瓶器、扳手等，人们都是利用这些工具，通过省力的方式，完成原本费力的操作。阿基米德也有一句名言，“给我一个支点，我能撬动地球”，可见善用杠杆是可以有所助力的。

在日常生活中，我们会认为不欠别人钱是最好不过的，在金融市场中，零债务并非最优解，适度负债可以帮助企业扩大规模，把握发展的窗口期，等到企业的收入和利润获得增长后，一部分用于偿还借贷的本金和利息，剩余部分可以再投资新业务，形成良性循环。一旦企业负债过多、或者增长速度放缓，利润增长的部分不足以偿还借贷利息时，负债就会变成企业的负担，增加企业的经营风险。负债是一把双刃剑，一方面是希望借力增长，更重要的一方面是莫要徒增风险，因此该因子倾向选择债务率更低的企业。

当然，有些企业在成立之初，需要搭建生产线、物流仓库，投入营销费用抢占市场，这些投资在短期内是无法收回成本的，企业在财务报表上可能出现连续亏损，债务率上升。而单因子的投资体系，是寻找大多数股票的共性，那么就难免忽略掉有些合理存在的特例。市场杠杆因子考虑了市场价值，用意是在上述特殊但却正确的经营行为能够获得投资者的认可时，能够在因子设计中有所体现，即便债务率有所上升，只要市场价值能以更快的速度增长，那么因子数值也会相对较小。

2. Book Leverage, 账面杠杆

Style: Leverage

Descriptors: BLEV

Book Leverage

Computed as,

$$BLEV = \frac{BE + PE + LD}{BE}$$

where BE, PE, and LD are the book value of common equity, preferred equity, and long-term debt, respectively, from the last fiscal year.

由于 A 股市场没有优先股，最近财年的净资产加上长期债务，再除以净资产，即可得到因子数值。

由公式可见，该因子的核心部分为长期债务除以净资产，净资产是总资产减去总负债，总负债可以分为流动负债和非流动负债两部分，流动负债又叫短期负债，是指将在一年以内偿还的债务，包括短期借款、应付票据、应付账款、预收账款、应付工资、应交税金等。非流动负债又叫长期负债，是指期限超过一年的债务，包括银行贷款、金融借款、公司债券等。

短期负债与企业的经营方式相关性较强，短期偿债能力可以通过流动比率，即流动负债除以流动资产来衡量，流动比率一般在 1.5 到 2 之间比较合适，体现了适度负债的含义。

长期负债与企业的真实债务更为接近，且债务利率一般更高，很难提前偿还，长期偿债能力可以通过利息保障倍数，即息税前利润除以利息费用来衡量，利息保障倍数越高越好，说明企业有充足的盈利能力来支付利息费用。

3. Debt-to-Assets, 资产负债率

Style: Leverage

Descriptors: DTOA

Debt-to-Assets

Computed as,

$$DTOA = \frac{TL}{TA}$$

where TL and TA are the total liabilities and total assets, respectively, from the last fiscal year.

最近财年的总负债除以总资产，即可得到因子数值。

资产负债率是最简单最直接的杠杆指标，直观的体现了负债是否适度。超过了 100% 即负债大于资产，资不抵债，是十分危险的信号，一般将资产负债率控制在 60% 以下较为安全。由于总负债中包含了短期负债，因此资产负债率过低，说明企业的应付账款和预收账款较少，在产业链中的处于非主导地位。

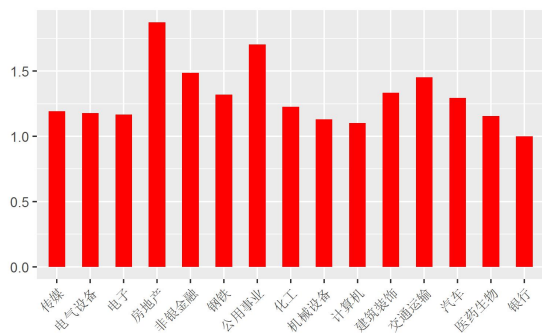


图 1 各行业账面杠杆

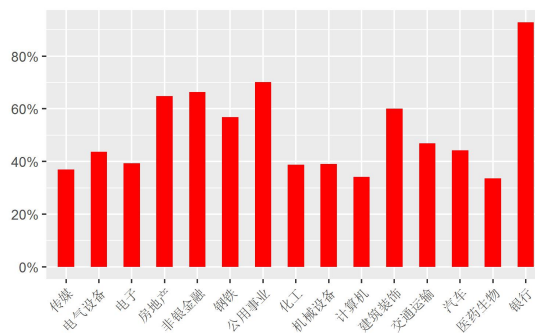


图 2 各行业资产负债率

当然，资产负债率还与行业的经营特征相关，如图 2 所示，银行和房地产行业的资产负债率较高。银行的负债主要来自银行存款、同业拆借及同业存款等，银行正是靠吸收存款和发放贷款之间利差来赚取利润，因此资产负债率接近于 100%。房地产行业需要在施工的各个环节垫付及缴纳资金，属于重资产垫资模式，而且工程回款往往较慢，因此资产负债率也相对较高。

4. 因子回测结果

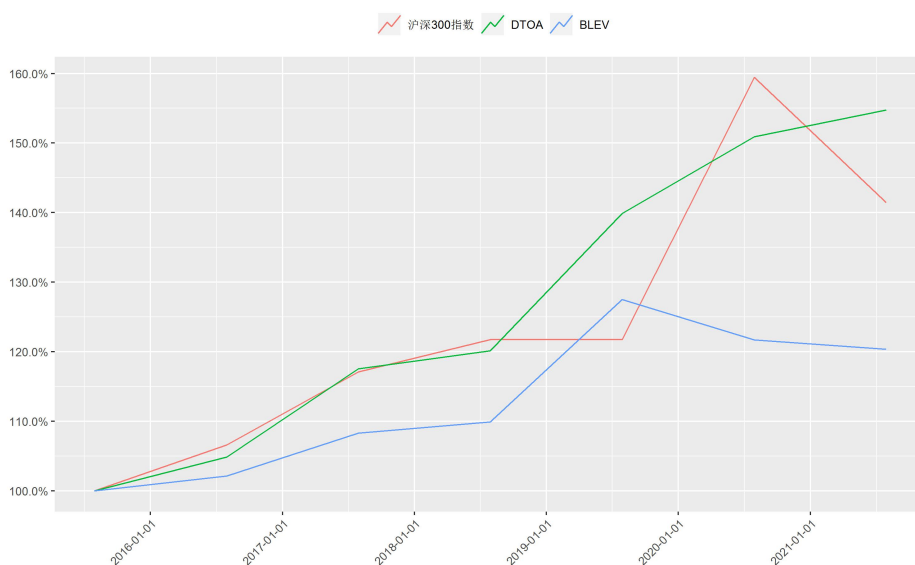


图 3: Leverage 因子累计收益率

请阅读页末免责声明

因子名称	IC	IR	RankIC	RankIC_IR	因子收益率
MLEC	0.97%	7.76%	-1.32%	-9.99%	0.02%
DTOA	18.73%	179.49%	-19.31%	-125.72%	7.67%
BLEV	9.07%	71.91%	-6.04%	-32.67%	3.34%

表 1: Leverage 因子回测结果

如图 3 所示，从因子收益曲线来看，低杠杆率的股票有着稳定而出色的收益，也说明投资者在有意规避高杠杆率的股票。一般来说，单因子投资体系中，只有少部分股票符合因子预设的逻辑，大部分股票对因子不敏感，平均下来少部分股票的高收益会被总体样本所稀释，使得因子表现弱于指数，考察因子的有效性也是以稳定增长且少有反转为佳。资产负债率因子走出了与指数同步增长的表现，说明该因子的有效性是明显强于其他因子的。账面杠杆因子在 2019 年年报之后出现一定下滑，说明投资者对长期债务的包容度有所提高。

二. Earnings Variability, 盈利波动率

1. Variability in Sales, 营业收入波动率

Style: Earnings Variability
 Descriptors: VSAL Variability in Sales
 Computed by dividing the standard deviation of the annual sales of the last five fiscal years by the average annual sales.

取最近 5 年年报的营业收入，其标准差除以均值，即可得到因子数值。

标准差已经是一种波动率的度量，将其再除以均值，主要是剔除企业营收规模的影响。计算中也可以先将营业收入除以 5 年的均值，得到每年营业收入的相对于均值的倍数，再取标准差，可以得到相同的波动率。

营业收入更偏重企业经营规模的度量，虽然会随经营行为波动，但是总体来说波动率较低。相比之下，净利润更偏重企业经营结果的度量，在企业经营不善导致营业收入有所降低的情况下，由于存在与业务量无关的固定成本，例如设备折旧、房屋租金、人员工资等，支付完固定成本后净利润的降低会更加明显，因此总体来说波动率较高。

2. Variability in Earnings, 净利润波动率

Style: Earnings Variability

Descriptors: VERN

Variability in Earnings

Computed by dividing the standard deviation of the annual earnings of the last five fiscal years by the average annual earnings.

取最近 5 年年报的净利润，其标准差除以均值，即可得到因子数值。

有时企业业绩的快速增长也会伴随着高标准差，但是这并不代表我们通常理解的高波动率，因此还需要观察企业业绩是否在增长和衰退之间出现摇摆。单因子的投资体系，是从某一个角度给出股票的度量，我们需要了解单因子所忽略的合理且特殊的情况，结合其他指标，得出更为完整的结论。

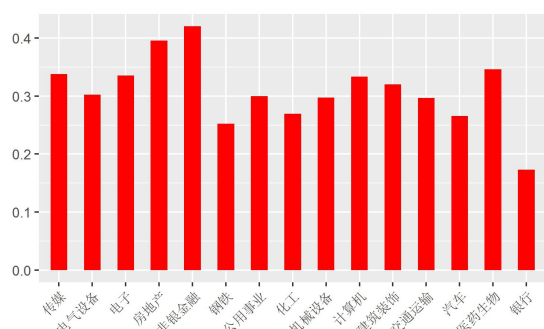


图 4 各行业营业收入波动率

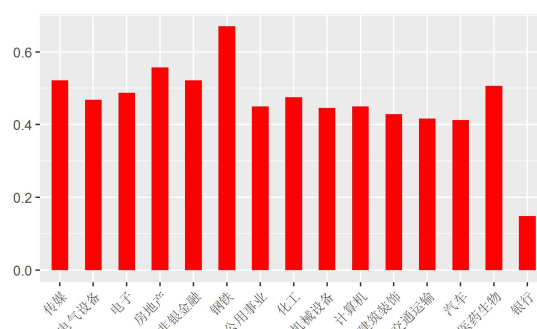


图 5 各行业净利润波动率

企业净利润的波动，除了与企业自身的经营相关，还与国际形势、宏观政策、黑天鹅事件等相关，需要对比行业内其他企业综合分析。如图 4 和图 5 所示，取每只股票最近 8 年报的营业收入，计算其波动率，再将行业内的所有股票取平均值，可以发现除了银行的波动率较低，钢铁行业的净利润波动率较高，其他行业的波动率差别不太明显。

3. Variability in Cash flows, 现金流波动率

Style: Earnings Variability

Descriptors: VFLO

Variability in Cash-flows

Computed by dividing the standard deviation of the annual cash flows of the last five fiscal years by the average annual cash flow.

取最近 5 年年报的现金流，其标准差除以均值，即可得到因子数值。

波动率的三个因子，营业收入、净利润、现金流体现了企业经营中从账面到现实的

过程。营业收入扣除营业成本、各项费用得到净利润，净利润扣除预收预付款项、应收应付款项、融资投资损益等得到现金流。如果一家企业的预收账款和应付账款比较充裕，说明其在产业链中的主导地位 and 议价能力。反之，如果有较多的预付账款和应收账款，那么净利润也会有一部分停留在账面之上，无法真正入袋为安，转化为实际收益。可以说，现金流作为企业的血液，带动起了企业的所有经营活动。波动率的三个因子，通过层层递进的方式，考察了企业经营中三个重要环节的稳定性。

4. 因子回测结果

因子名称	IC	IR	RankIC	RankIC_IR	因子收益率
VSAL	2.29%	11.30%	4.33%	21.42%	1.13%
VERN	1.79%	17.87%	-1.32%	-8.93%	0.78%
VFLO	-0.22%	-8.49%	-0.15%	-1.53%	1.32%

表 2: Earnings Variability 因子回测结果

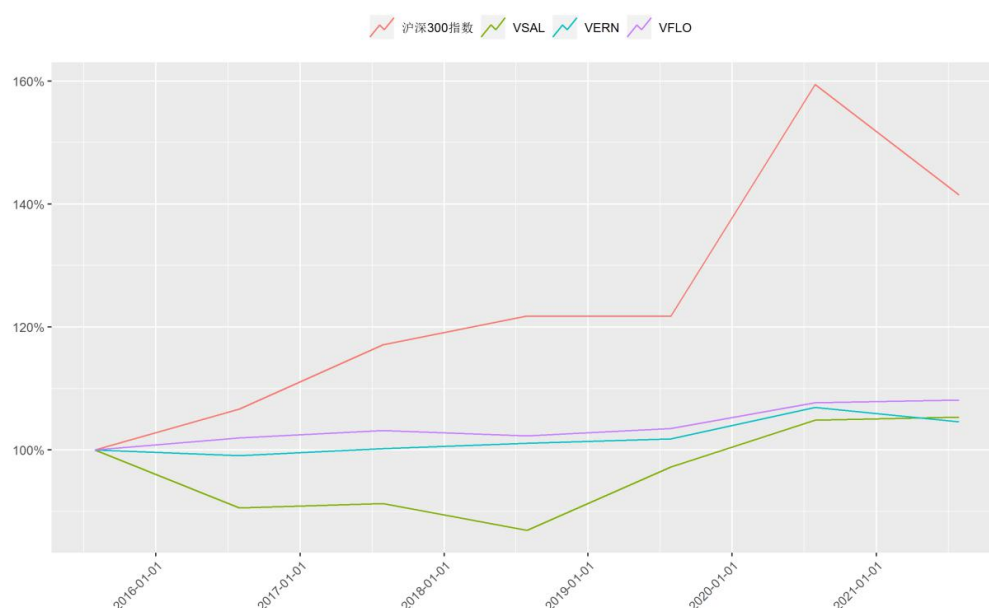


图 6: Earnings Variability 因子累计收益率

如图 6 所示，从因子收益曲线来看，投资者对于营业收入和净利润的波动并不敏感，波动率的大小在收益率上差别很小。对于现金流，投资者起初是轻微厌恶其波动的，波动率大的股票收益率较低，最近几年，这种厌恶有所缓解，因子收益率逐渐回升，总体来说没有十分明显的倾向。

三. Profitability, 盈利能力

1. Asset Turnover, 资产周转率

Style: Profitability

Descriptors: ATO

Asset Turnover

Computed as,

$$ATO = \frac{Sales}{TA}$$

where Sales is the trailing 12-month sales, and TA is the most recently reported total assets.

滚动 12 个月的营业收入除以最新公布的总资产，即可得到因子数值，该因子的周期为季度。

从因子的字面含义理解，就是每 1 元的总资产，创造了多少营业收入。实际上，营业收入除了与企业经营能力有关，还很大程度上与企业规模有关，除以总资产是为了剔除企业规模的影响，使得不同规模的企业之间具有可比性。从另一个角度理解，所谓周转率，是指总资产不管大小，要是躺平在企业账面上，就像躺在过去的功劳簿上，还需要走出去攻城略地，只有将总资产转化为企业收入，那样才算物尽其用，有源之水。

2. Gross Profitability, 总盈利能力

Style: Profitability

Descriptors: GP

Gross Profitability

Computed as,

$$GP = \frac{Sales - COGS}{TA}$$

where Sales, COGS, and TA are the sales, cost of goods sold, and total assets, respectively, from the last fiscal year.

最近财年的营业收入减去营业成本，再除以总资产，即可得到因子数值。

资产周转率衡量了企业的创收能力，该因子进一步衡量了企业的盈利能力，而创造收入和获得盈利之间恰恰相隔的是控制成本。

上述两个因子都是以总资产为基准，而不同行业的经营方式对资产的依赖程度又有所不同，如图 7 所示，轻资产行业主要有医药生物、计算机、电子等，可以依靠较少的总资产获得较多的营业利润，因此总盈利能力表现较高。重资产行业主要有钢铁、非银金融、银行等，除了银行的盈利能力较为突出，钢铁行业的营业利润被重资产稀释，在

总盈利能力显得较低。

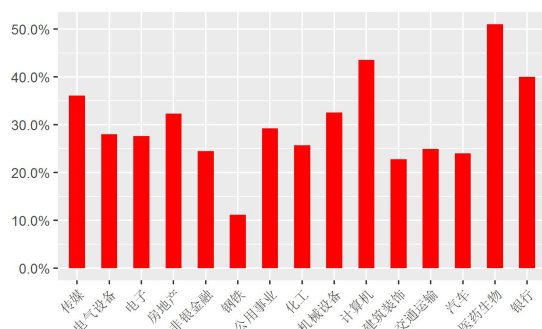


图 7 各行业总盈利能力

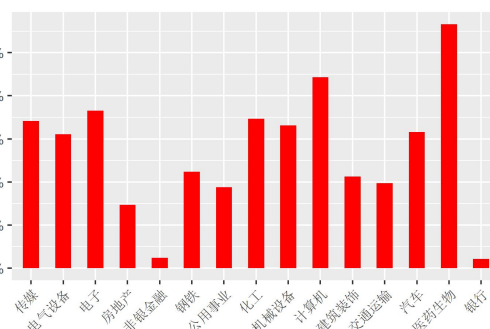


图 8 各行业毛利率

3. Gross Profit Margin, 毛利率

Style: Profitability

Descriptors: GPM

Gross Profit Margin

Computed as,

$$GPM = \frac{Sales - COGS}{Sales}$$

where Sales and COGS are the sales and cost of goods sold, respectively, from the last fiscal year.

最近财年的营业收入减去营业成本，再除以营业收入，即可得到因子数值。

三大财务报表中的利润表，体现的是营业收入到净利润之间抽丝剥茧的过程，营业收入减去营业成本得到毛利，再减去销售费用、管理费用得到息税前利润，再减去财务费用、所得税得到净利润。如图 9 所示，在各行业净利润相差不大的前提下，研发投入较高的行业，如医药生物、计算机，或者营销投入较高的行业，如传媒、房地产，需要的毛利率也相对较高。总体来说，毛利率虽然与所处行业有一定关联，但是更与企业自身的经营能力密切相关。

4. Return on Assets, 总资产收益率

Style: Profitability

Descriptors: ROA

Return on Assets

Computed as,

$$ROA = \frac{Earnings}{TA}$$

where Earnings is the trailing 12-month earnings, and TA is the most recently reported total assets.

请阅读页末免责声明

滚动 12 个月的净利润除以最新公布的总资产，即可得到因子数值，该因子的周期为季度。

考察完毛利，我们来到了净利润，最常用的两个指标是 ROA 和 ROE，用净利润分别除以总资产和净资产。ROA 可以分解为净利率乘以总资产周转率，ROE 可以分解为净利率乘以总资产周转率再乘以杠杆水平，其中净利率是指净利润占营业收入的比重，杠杆水平是指净资产占总资产的比重。由此可见，ROA 将前述两个盈利因子以乘积的形式结合，考察两个因子共同作用的结果，两者之一数值越大，ROA 的数值也越大，两者的数值都很大时，ROA 的数值也会显著增大，是最为常用和有效的盈利因子。

4. 因子回测结果

因子名称	IC	IR	RankIC	RankIC_IR	因子收益率
GP	1.21%	4.68%	6.60%	22.74%	-0.38%
GPM	3.55%	22.19%	12.49%	88.29%	1.29%
ATO	4.36%	42.56%	5.52%	42.52%	0.44%
ROA	5.68%	33.01%	10.41%	54.63%	0.95%

表 3: Profitability 因子回测结果

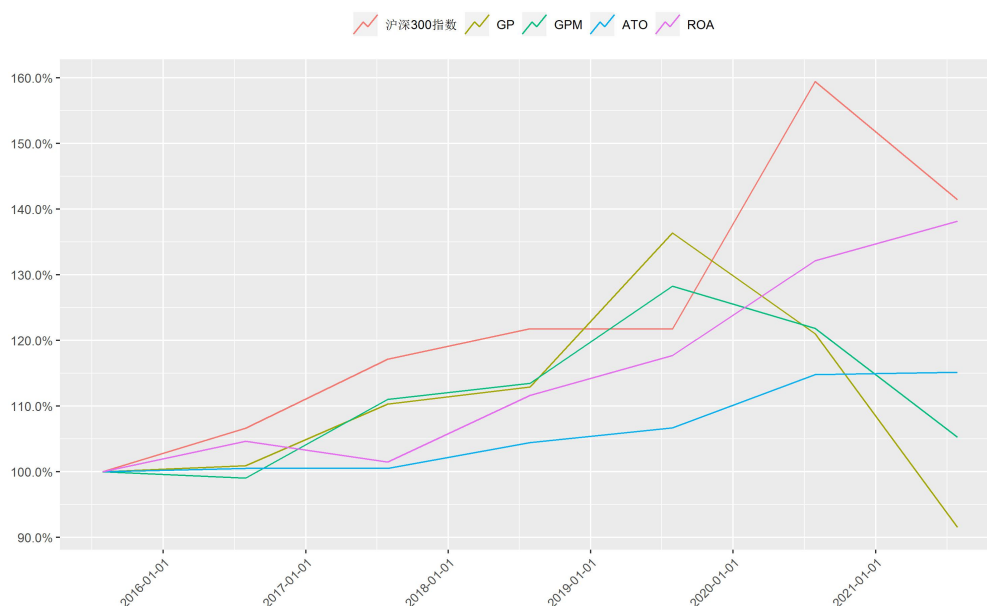


图 9: Profitability 因子累计收益率

请阅读页末免责声明

如图 9 所示，从因子收益曲线来看，总资产周转率和总资产收益率两个因子表现稳定，ROA 又可以分解为总资产周转率和净利率，因此 ROA 收益高于总资产周转率的部分，可以看做净利率的影响，可以投资者对于两者同样看好，共同推高了 ROA 的收益。其余两个因子在经历了一段涨势后急转直下，最终收于零线附近，这两个因子的共同点是以毛利润为主体，不同的是比值分别相对于总资产和营业收入，可见以 2018 年年报为界，投资者对于净利润和毛利润的青睐发生了明显的反转。

四. Investment Quality 投资质量

1. Total Assets Growth Rate，总资产增长率

Style:	Investment Quality
Descriptors	AGRO
	<u>Total Assets Growth Rate</u>
	Computed by first dividing the slope coefficient from the regression of the total assets from the last five fiscal years against time by the average total assets, and then multiplying by -1 to reverse the sign.

取最近 5 年年报的总资产为自变量，以时间为因变量，回归后得到斜率，将其除以 5 年总资产的平均值，最后乘以-1 反转符号，即可得到因子数值。

在 Barra 因子体系之外，我们常见计算增长率的方式有两种，一种是取开始和结束时的数值，得到整个区间的增长率，然后根据区间的时间跨度以复利或者单利的方式计算年化增长率；另一种是计算每年的增长率，然后取平均值。两种方式都有缺陷，前者丢弃了中间年份数据，没有反映增长的过程；后者容易受到某一年增长率突然激增或断崖下跌的影响，使结果过大或过小。

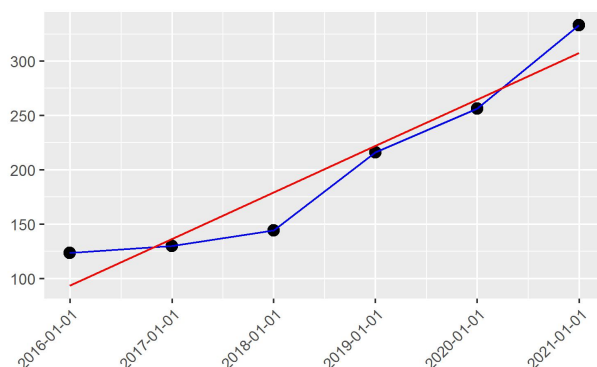


图 10 回归斜率

Barra 因子采用回归斜率计算增长率，如图所示，蓝色折线是每年的增长率，平缓

表示增长较慢，陡峭表示增长较快，红色直线是回归函数拟合的结果，使得真实值与拟合值的累计误差最小，既使用到了每个点的数据，又不会被异常值影响过大。回归斜率可以理解为总资产每年以固定的数量增长，然后除以 5 年间的平均总资产，就可以得到整个区间的平均增长率。在计算过程中，也可以先用每年的总资产除以平均总资产，再求回归斜率。

3. Issuance Growth，股票发行数量增长率

Style: Investment Quality

Descriptors

IGRO

Issuance Growth

Computed by first dividing the slope coefficient from the regression of the number of shares outstanding from the last five fiscal years against time by the average number of shares outstanding, and then multiplying by -1 to reverse the sign.

取最近 5 年年报的发行股票筹资额为自变量，以时间为因变量，回归后得到斜率，将其除以 5 年发行股票筹资额的平均值，最后乘以-1 反转符号，即可得到因子数值。

企业发行股票是一种筹资行为，填补企业的资金缺口，而募集资金用途是评价增发行为是否利好的重要依据。

其一，如果用于产能的扩张，一般主要指新建厂房，购买设备，这样的增发一般为良性的，但是前提是企业的经营本来就比较稳定，业绩持续增长，具备扩张的基础。

其二，如果用于收购别的企业，需要了解被收购的企业，是不是与公司的主营业务相关，能否对企业未来的经营提供帮助。如果与主营业务并不相关，可能存在资本运作，应当属于利空。资本市场对于增发并购，一般是不看好的，需要小心谨慎。

最后，如果用于补充企业的流动性，说明企业经营过程中出现了重大风险，而且无法用其他方式化解，那就属于典型的利空了。

3. Capital Expenditure Growth，资本支出增长率

Style: Investment Quality

Descriptors

CXGRO

Capital Expenditure Growth

Computed by first dividing the slope coefficient from the regression of the capital expenditures from the last five fiscal years against time by the average capital expenditures, and then multiplying by -1 to reverse the sign.

请阅读页末免责声明

取最近 5 年年报的资本支出为自变量，以时间为因变量，回归后得到斜率，将其除以 5 年资本支出的平均值，最后乘以-1 反转符号，即可得到因子数值。

资本支出又称收益性支出，是指为固定资产增值的支出，如房屋、机器设备的购置费，为延长设备使用年限或提高设备性能的支出，以及购建无形资产和其他长期资产所支出的费用。可见资本支出将企业流动的现金流转换为了固定成本，损失了经营的灵活性，必要的资本支出是企业发展的保障，但过多的资本支出显然会阻碍企业的增长。

4. 因子回测结果

因子名称	IC	IR	RankIC	RankIC_IR	因子收益率
AGRO	-7.30%	-32.77%	-5.29%	-25.83%	-3.92%
IGRO	10.74%	73.32%	6.88%	47.54%	3.35%
CXGRO	-3.42%	-20.24%	-2.06%	-14.15%	-3.98%

表 4: Investment Quality 因子回测结果

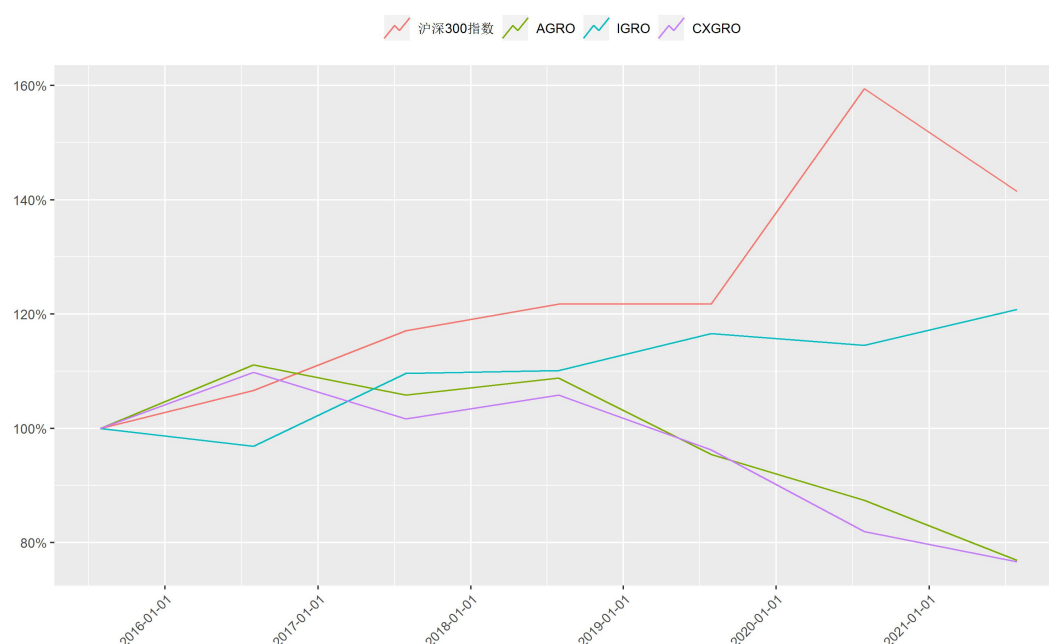


图 11: Investment Quality 因子累计收益率

如图 11 所示，从因子收益曲线来看，2018 年年报之前，三个因子走势较为平稳，之后出现分化，股票发行数量增长率维持稳中有升的格局，说明投资者对增发股票的行

为相对看空。总资产和资本支出在 2019 年年报之后一路缓慢下挫，收至 80%，也许是宏观经济预期的影响，仍有剩余财力投向总资产和资本支出，是一种利好行为，因子设计因为看空出现了亏损。

五. Growth 增长

1. Earnings Per Share Growth Rate, 每股收益增长率

Style: Growth

Descriptors: SGRO

Sales per Share Growth Rate

Computed by dividing the slope coefficient from the regression of the annual sales per share from the last five fiscal years against time, by the average annual sales per share.

取最近 5 年年报的每股收益为自变量，以时间为因变量，回归后得到斜率，将其除以 5 年每股收益的平均值，即可得到因子数值。

每股收益是上市公司披露财报时的主要财务指标，是指归属于母公司净利润除以总股本。其中基本每股收益的总股本指现有发行的股本，稀释每股收益的总股本还包括潜在普通股，例如可转换公司债券、认股权证和股份期权等。每股收益反映了公司的盈利能力，并用于市盈率的计算中。

3. Sales Per Share Growth Rate, 每股营业收入增长率

Style: Growth

Descriptors: EGRO

Earnings per Share Growth Rate

Computed by dividing the slope coefficient from the regression of the annual earnings per share from the last five fiscal years against time, by the average annual earnings per share.

取最近 5 年年报的每股营业收入为自变量，以时间为因变量，回归后得到斜率，将其除以 5 年每股营业收入的平均值，即可得到因子数值。

每股营业收入是指营业收入除以总股本，可以分解为每股收益乘以净利率，净利率即净利润占营业收入的比重，在 Barra 财务类因子中，有些因子在字面含义上平平无奇，再将其分解为两三个指标相乘后，发现因子设计时纳入了更多因子综合评价。

5. 因子回测结果

因子名称	IC	IR	RankIC	RankIC_IR	因子收益率
EGRO	2.44%	21.56%	-0.19%	-1.08%	0.48%
SGRO	-1.49%	-9.67%	4.39%	29.22%	1.37%

表 5: Growth 因子回测结果

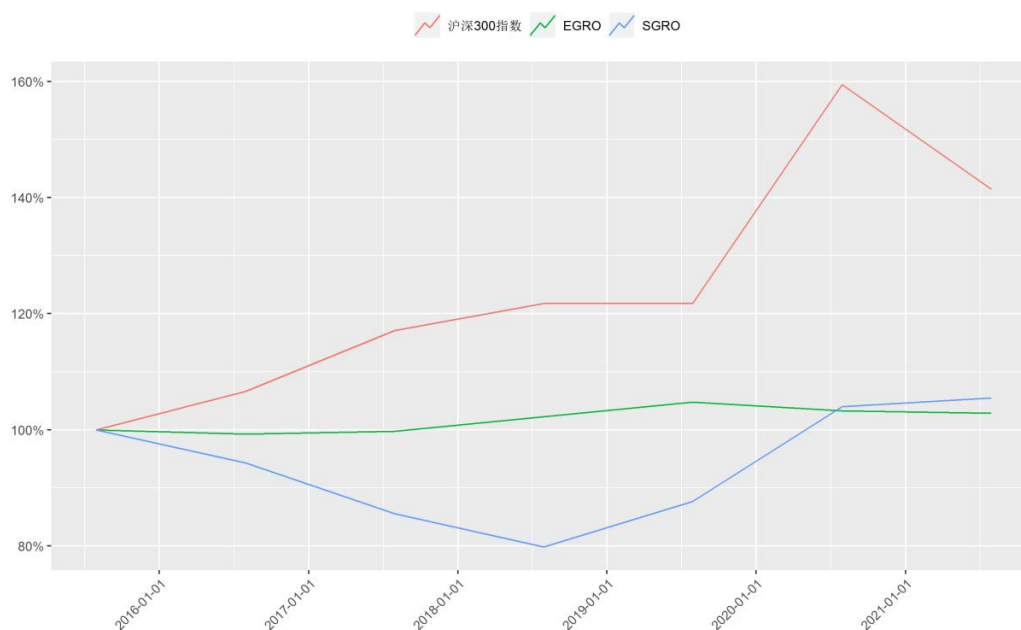


图 12: Growth 因子累计收益率

如图 12 所示，从因子收益曲线来看，每股收益增长率因子几乎是毫无波澜，走出一条水平线，说明每股收益增长快速或者缓慢，在股票收益率上没有呈现出聚集现象，也说明每股收益作为最主要的几个财务指标之一，已被投资者充分挖掘。每股营业收入增长率先跌后涨，2018 年年报之前，每股营业收入增长较快的股票，收益率反而更低，说明投资者有高估增长潜力的倾向。2019 年年报之后，因子收益开始回升，也许是受到宏观经济预期的影响，投资者对增长的预期开始趋向理性和谨慎。

附录 1 Barra 因子结构

Barra 模型是 MSCI 明晟公司研发的金融风控模型，把个股收益分解为市场（国家）因素、行业因素、风格因素，其中风格因素即描述个股特征的风险因子，主要应用于多因子选股分析和结构化风险因子分析。第一个为美国市场设计的多因子模型于 1975 年发布，称为 Barra US Equity Model1（USE1），USE2 于 1985 年发布，USE3 于 1997 年发布，USE4 于 2011 年发布。Barra 模型经受了众多金融机构的实证检验，经久不衰，生机蓬勃。

2012 年 7 月明晟公司终于发布了针对中国市场的模型 Barra China Equity Model（CNE5），包括 10 个大类因子、21 个小类因子。2018 年 8 月发布的 CNE6，将因子进一步细分和扩充，包括 9 个一级因子、21 个二级因子、46 个三级因子，9 个一级因子分别是波动率（Volatility）、动量（Momentum）、价值（Value）、规模（Size）、流动性（Liquidity）、质量（Quality）、成长（Growth）、分红（Dividend Yield）、情绪（Sentiment）。

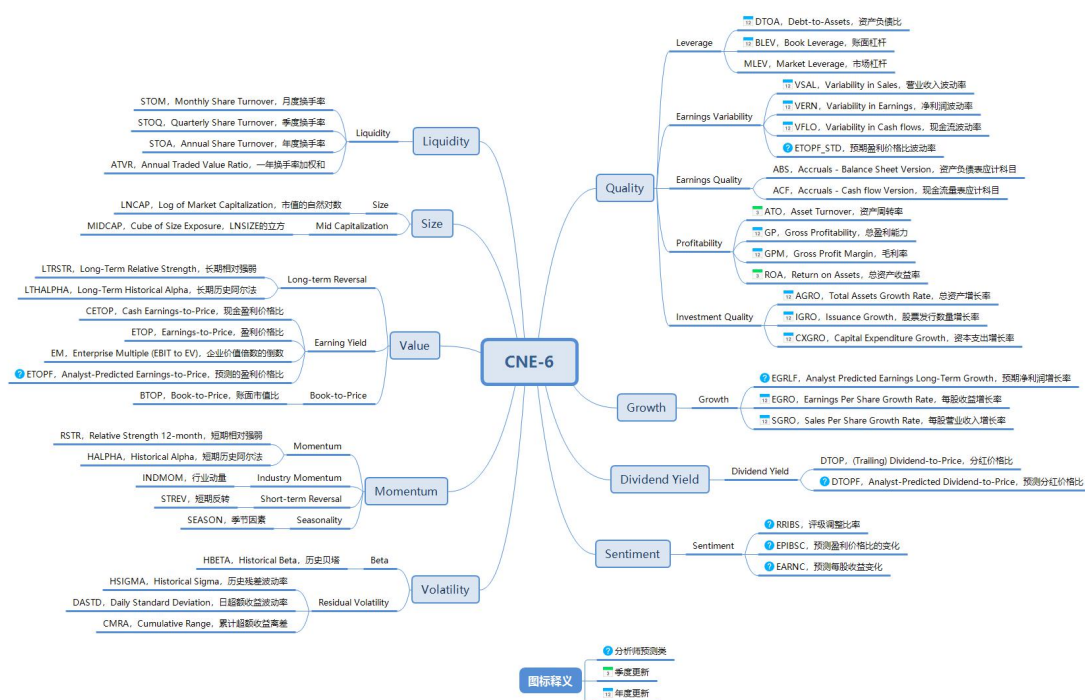


图 13: CNE6 因子结构

附录 2 因子评价指标

一. IC (Information Coefficient, 信息系数)

IC 值是指对于一揽子股票和某一个因子, 当期的因子值序列与下期的收益率序列之间的 Pearson 线性相关系数, 正相关性越高, 说明该因子的数值对于股票收益率的预测能力越好, 因子的有效性越强。当然, 如果呈负相关性, 绝对值越大越好, 即该因子是反向因子, 对股票收益率同样具有良好的预测能力。

日度更新因子在历史上的每一天都有一个 IC 值, 在回测区间的若干天内对若干个 IC 值取平均数, 可以得到因子的 IC 均值, 用来描述该因子在回测区间内的有效性, 通常认为 IC 均值大于 3% 或者小于 -3% 时因子比较有效。

由于 IC 值直接使用数值计算相关系数, 就难免受到异常值和量纲的影响, 因此需要对原始的因子值进行缩尾化和标准化。缩尾化是指以均值加减三倍标准差为边界, 截断边界以外的部分, 保留边界以内的部分, 即将异常值用边界值替代。标准化通常将原序列减去均值再除以标准差, 使新序列的均值为 0 且标准差为 1。这样不同因子间就排除了异常、统一了量纲, 可以在同一水平下进行比较。

另外, 对季度和年度更新的财务因子进行缩尾化和标准化, 是以 $t-1$ 期的因子均值和标准差为基准。首先, 财务因子计算收益率的起始时间点为财报公布日, 这样可以获得市场对于财务数据最及时的反应。其次, 每只股票的财报公布日各有不同, 虽然我们在当前已知所有股票的历史财务数据, 但是回测时是无法知晓回测时点之后公布的数据, 这种现在已知、当时未知的信息在回测中称为未来函数, 无意识的使用未来函数会对回测的真实性有所损害。因此, 如果在对某只股票 t 期的因子数值缩尾化和标准化时, 使用 t 期的全部因子数值计算均值和标准差, 显然属于使用了未来函数, 而 $t-1$ 期的因子数值在任何股票的财报公布日都是已知的。这样的计算方式得到的标准化数值, 其均值并不严格的等于 0, 标准差也不严格的等于 1, 会存在一定的风险敞口, 但总体来说误差很小, 可以控制在 3% 以内。

由于 Pearson 相关系数, 假设序列服从正态分布, 而很多情况下的数据难以满足这个假设。而且 Pearson 相关系数只能描述线性相关性, 对于高度非线性相关的序列也只能得到较低的线性相关性。基于以上两点局限性, Spearman 相关系数先将数值排序, 再计算排名之间的相关性, 可以得到很好的结果。

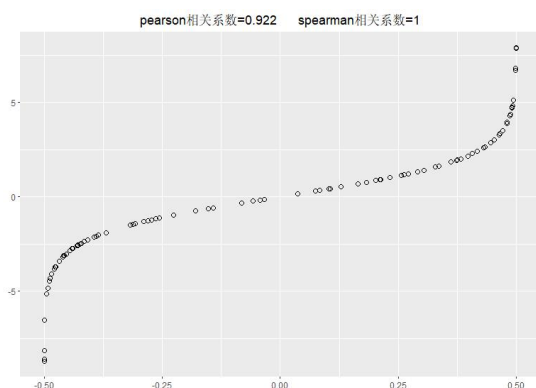


图 14：完全相关时的相关系数

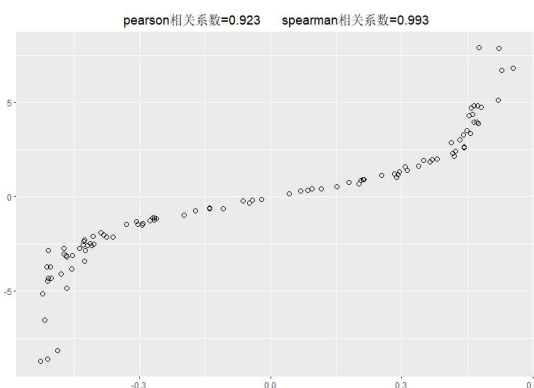


图 15：微小扰动时的相关系数

如图 14 所示，逻辑曲线在中间缓慢增长，在两边迅速增长，符合很多事物分阶段变化的规律。图中曲线序列 Y 完全由序列 X 决定，是完全相关的关系，但是 Pearson 相关系数只有 0.922，而 Spearman 相关系数为 1，准确描述了这种完全相关关系。

如图 15 所示，在图 14 完全决定的关系上施加微小的随机扰动后，Spearman 相关系数仍然高达 0.993，说明在更加真实的环境中，Spearman 相关系数也能得到符合客观规律的结果。

Rank_IC 值是指在 IC 值的计算中用排序代替数值，即因子值的排序和收益率的排序之间的相关关系，更准确的说是因子值和收益率的 Spearman 秩相关系数，是对 IC 值的一种改进。

二. IR（Information Ratio，信息比率）

IR 值是指回测区间内的 IC 均值除以 IC 标准差，即在原有的 IC 均值评价基础上，对于 IC 值波动较大的因子增加惩罚，提高 IC 值较为稳定因子的评价。因此可以看做考虑了风险因素后，对 IC 均值的优化。

然而由于 IC 值大于 3% 或者小于 -3% 都说明因子比较有效，IC 值的绝对值较大直观上会导致波动较大，而 IR 值又会惩罚波动较大的因子，似乎 IR 值会将有效的因子变的平庸。其实不然，IR 值只是会惩罚 IC 值在大于 3% 和小于 -3% 之间多次反转的因子，而这种反转代表着因子的失效以及买卖方向的错误。对于 IC 值稳定的大于 3% 的因子，或者稳定的小于 -3% 的因子，IR 值是不会加以惩罚的。通常认为 IR 大于 5% 因子比较有效。

因此，评价一个因子的有效性，除了 IC 值，还需要观察 IR 值，只有两者都给出有

请阅读页末免责声明

效性较高的结论时，才能说明因子是真正有效的。

RankIC_IR 值是指回测区间内的 Rank_IC 均值除以 Rank_IC 标准差，由于 RankIC 值是对 IC 值的一种改进，RankIC_IR 值也是对 IR 值的一种改进。

三. 因子收益率

因子收益率是指对于一揽子股票和某一个因子，首先对原始的因子值进行缩尾化和标准化。然后将当期的因子值序列与下期的收益率序列做内积，即对应元素相乘再求和。换一个角度理解，以因子值作为权重买入股票，因子值为负的卖空股票，买入权重乘以下期的收益率即该股的盈亏，因子值和收益率都为正则盈利，因子值为正而收益率为负则亏损，再将所有股票的盈亏求和，即整个交易的总盈亏，以此来衡量当期因子值对下期收益率的预测能力，因子收益率越高则因子越有效。

年度更新的因子，其收益率以每只股票年报公布日为起点，计算一年后的个股收益率，在年报公布的截止日，也就是每年的 4 月 30 日可以得到所有股票的收益率，然后取平均值，得到因子收益率。季度更新的因子同理，其中半年报的截止日为 8 月 31 日，季报的截止日为 10 月 30 日。

因子收益率只是一个因子评价指标，并未考虑实际交易中的买入卖出手续费，而模拟的交易中又是当日买入次日卖出的较为频繁交易。在实际交易中可以买入前 20% 的股票，并减去手续费，得到较为真实的收益率曲线。在样本空间的股票数较多时，还可以将股票按因子数值从高到低分为五档，分别计算每天每档的因子收益率，再合成每档的累计收益率曲线，看收益是否仍然呈现从高到低的排列，且有比较明显的差异。

重要声明

本报告中的信息均来源于公开资料，我公司对这些信息的准确性及完整性不作任何保证，不保证报告信息已做最新变更，也不保证分析师作出的任何建议不会发生任何变更。在任何情况下，报告中的信息或所表达的意见并不构成所述期货品种买卖的出价或询价。在任何情况下，我公司不就本报告中的任何内容对任何投资作出任何形式的担保，投资者据此投资，投资风险自我承担。我公司可能发出与本报告意见不一致的其它报告，本报告反映公司分析师本人的意见与结论，并不代表我公司的立场。未经我公司同意，任何人不得对本报告进行任何形式的发布、复制或对本报告进行有悖原意的删节和修改。