

A 股财务数据的整理和使用

刀疤连，方信光

2019-03-08

摘要：基本面量化是应用量化研究领域的重头戏，财务数据的整理和加工是基本面量化的第一步。本文梳理了财务数据的基本知识，包括报表类型、数据来源、调整更正和使用原则等，并给出了单季度和 TTM 数据的计算流程。在本报告的基础上，可以搭建财务因子计算框架，为后续基本面量化研究铺好路。

1. 财务数据

经典的有效市场理论认为，公司的一切信息都反映在了股价上面，基本面研究可能并无用武之处。然而近年来各种异象的发现，表明市场并没有那么有效。目前大量研究把精力放在交易数据的挖掘上，像工厂一样制造出了形形色色的因子；反而更加符合直觉的基本面研究，时常遭到莫名其妙的攻击。事实上，基本面量化是应用量化研究领域的重头戏，也是我们后续努力的方向之一。千里之行，始于足下，财务数据的整理和加工是基本面量化的第一步，也是本文要讨论的主要内容。

财务数据就是与公司财务有关的数据，主要是反映企业一定期间的经营成果和财务状况变动。交易所为了保护投资者的权益，要求上市公司定期公布财务报告，以便投资者了解其经营以及财务状况。

1.1 报告期

在内地，《会计法》规定会计年度从1月1日开始，12月31日结束。一般要求上市企业一年定期披露四次财报，分别是一季报、中报、三季报、年报¹。财务报告在披露日期上没有严格要求，只要在规定的截止日之前披露就行。每个报告期都有一个标识，一般用年份+季度末表示，如2018年三季报用“20180930”标识，2018年年报用“20181231”标识，具体定义见表1。

表1 财务报告期

报告期	起始点	结束点	常用标签	最晚披露日
一季报	1月1日	3月31日	xxxx0331	4月30日
半年报	1月1日	6月30日	xxxx0630	8月31日
三季报	1月1日	9月30日	xxxx0930	10月31日
年报	1月1日	12月31日	xxxx1231	4月30日

数据来源：CQR

¹ A股上市公司从2002年开始披露季度资产负债表和利润表，从2003年开始披露季度现金流量表

1.2 财务报表

财务报告主要由财务报表构成²，财务报表包括五表一附注，即资产负债表、利润表、现金流量表、财务状况变动表、附表和附注，具体定义和特点见表 2。

表 2 五表一附注

表名	含义	特点
资产负债表	反映企业在一定日期（通常为各会计期末）的资产、负债和所有者权益情况	时点数据
利润表	反映企业在一定时期经营成果（盈利或亏损）和利润分配情况	区间数据
现金流量表	反映企业资产负债表中各个项目对现金流的影响，分为经营、投资和融资三个类别	区间数据
财务变动表	反映企业在一定时期内，资金取得的来源渠道和资金的用途去向等整个过程，分析流动资金净额增减的原因及由此引起的财务状况变动情况的报表	区间数据
报表附注	财务报表附注旨在帮助财务报表使用者深入了解基本财务报表的内容，财务报表制作者对资产负债表、利润表和现金流量表的有关内容和项目所作的说明和解释。财务报表附注中的内容非常重要，主要包括：企业所采用的主要会计处理方法；会计处理方法的变更情况、变更的原因及对财务状况和经营业绩的影响；发生的非经常性项目；一些重要报表项目的明显情况；或有事项；期后事项；以及其他对理解和分析财务报表重要的信息	—

数据来源：CQR

² 财务报表是财务报告中的标准数据，除了五表一附注外，财务报告还包括董事报告、管理分析和财务情况说明书等非标准内容

1.3 业绩预告和快报

业绩预告和业绩快报是财报预披露的两种形式，旨在提前释放重要的会计信息，减少股价波动。业绩预告主要是对公司当期净利润情况的预计，包括预告类型、净利润上下限、净利润变动幅度上下限和业绩变动原因等内容；业绩快报更加全面，一般披露公司主要的财务数据，包括营业收入、净利润和总资产等关键科目，可以提前用来计算各类财务因子。

表 3 整理了业绩预告的基本规则，包括披露条件和披露时间；表 4 为业绩快报的披露规则。

表 3 业绩预告披露规定

	上交所	深交所		
	主板	中小板	创业板	
披露要求	对于年度报告，如果上市公司预计全年可能出现亏损、扭亏为盈、净利润较前一年度增长或下降 50% 以上（基数过小的除外）等三类情况； 如果不存在上述三类情况，可以不披露年度业绩预告； 对于半年报和季度报告，也没有作强制要求。	预计报告期内（第一季度、半年度、第三季度和年度）出现以下情况的，应进行业绩预告：净利润为负、扭亏为盈、实现盈利且净利润与上年同期相比上升或者下降 50% 以上（基数过小的除外）、期末净资产为负、年度营业收入低于 1 千万元； 如果不存在上述情况，可以不披露业绩预告	上市公司应披露半年度、三季度、年度业绩预告； 上市公司预计第一季度业绩将出现归母净利润为负值、净利润与上年同期相比上升或者下降 50% 以上（基数过小的除外）、扭亏为盈，需要披露一季度业绩预告； 中小板业绩预告是强制性披露的，全部上市公司均要进行业绩预告	上市公司应披露一季度、半年度、三季度、年度业绩预告； 创业板业绩预告是强制性披露的，全部上市公司均要进行业绩预告
披露时间	报告次年 1 月 31 日前	一季度业绩预告：报告期当年的 4 月 15 日前； 半年度业绩预告：报告期当年的 7 月 15 日前； 三季度业绩预告：报告期当年的 10 月 15 日前； 年度业绩预告：报告期次年的 1 月 31 日前	上市公司应在第一季度报告、半年度报告和第三季度报告中披露对年初至下一报告期末的业绩预告； 如果满足第一季度业绩预告条件，应不晚于 3 月 31 日（在年报摘要或临时公告）进行业绩预告	一季度业绩预告：上年度年度报告预约披露时间在 3 月 31 日之前的，应当最晚在披露年度报告的同时，披露本年度第一季度业绩预告；年度报告预约披露时间在 4 月份的，应当在 4 月 10 日之前披露第一季度业绩预告； 半年度业绩预告：报告期当年的 7 月 15 日前； 三季度业绩预告：报告期当年的 10 月 15 日前； 年度业绩预告：报告期次年的 1 月 31 日前

数据来源：CQR，上交所，深交所

表 4 业绩快报披露规定

上交所		深交所	
主板		中小板	创业板
披露要求	上市公司可以在年度报告和中期报告披露前发布业绩快报；	鼓励上市公司在定期报告披露前，主动披露定期报告业绩快报；	年度报告预约披露时间在3-4月份的公司，强制披露业绩快报；
	上交所业绩快报不是强制性披露的	拟发布第一季度报告业绩预告但其上年年报尚未披露的上市公司，应当在发布业绩预告的同时披露其上年度的业绩快报	年度报告预约披露时间在3月份之前的公司可不披露年度业绩快报； 鼓励半年度报告预约披露时间在8月份的公司披露业绩快报； 半年报和季报业绩快报不是强制性披露的
披露时间	正式报告披露前	正式报告披露前	年度报告预约在3~4月份披露的公司，在2月底前披露快报； 鼓励半年度报告预约披露时间在8月份的公司，在7月底前披露半年度业绩快报

数据来源：CQR，上交所，深交所

1.4 数据源

财务报告一般可通过上市公司官网或交易所网站进行查询，但要做量化研究则需要整理好的结构性数据。目前可以从多个数据源获取财务数据，表 5 整理了我们使用过的一些数据源及它们的优势劣势³。

表 5 常见数据源

数据源	优势	劣势
Wind底层	数据齐全，质量高	收费不低
Wind API	相对便宜，接口获取	流量限制，未来数据
Tushare	免费，数据齐全，质量高	财务指标表有未来数据
新浪和雪球	免费，无门槛	需要爬虫采集，没有调整和修改

数据来源：CQR

2.调整 and 更正

2.1 基准报告期

³ 从 wind 接口提取会计科目或者财务指标，如果报表有过更正，只能提取到更正后的数据，例如柳工（000528.SZ），提取的每股收益和净利润/营业总收入，均是用更正后的数值计算。另外，从 wind 接口提取财务指标，财务指标均是用调整前的数据计算，例如中铁工业（600528.SH）的流动资产/总资产，虽然后面进行了追溯调整，但并没有用更新该指标的值。和 wind 一样，Tushare 财务指标表中，如果有过更正，只能提取到更正后的数据；如果有过实质性调整，也不对指标进行更新。

在最新报告期财务报表披露时，会同时披露对应的某个历史基准报表。不同报表对应的基准报表不同，资产负债表对应上年年报，如图 1 白云机场披露 2018 年三季报的时候，会同时披露年初余额（2017 年年报）；利润表和现金流量表对应为上年同期值，如图 2 白云机场 2018 年三季报现金流量表时，会同时披露上年同比报告期（2017 年三季报）。

图 1 白云机场 2018 年三季报资产负债表

合并资产负债表
2018 年 9 月 30 日
编制单位:广州白云国际机场股份有限公司

单位:元 币种:人民币 审计类型:未经审计

项目	期末余额	年初余额
流动资产:		
货币资金	3,481,537,664.27	2,836,957,414.70
结算备付金		
拆出资金		
交易性金融资产		
衍生金融资产		

数据来源：白云机场三季报

图 2 白云机场 2018 年三季报现金流量表

合并现金流量表
2018 年 1—9 月
编制单位:广州白云国际机场股份有限公司

单位:元 币种:人民币 审计类型:未经审计

项目	年初至本报告期 期末金额 (1-9月)	上年年初至本报告期 期末金额 (1-9月)
一、经营活动产生的现金流量:		
销售商品、提供劳务收到的现金	5,877,994,678.15	4,715,182,016.24
客户存款和同业存放款项净增加额		
向中央银行借款净增加额		
向其他金融机构拆入资金净增加额		
收到原保险合同保费取得的现金		
收到再保险业务现金净额		

数据来源：白云机场三季报

无论一季报、半年报、三季报还是年报，在每期资产负债表披露时，其对应的基准报表都为上年年报；而现金流量表和利润表，其对应的基准报表总为对应的同比报告期，对应关系见表 6，表 7 给出了一个示例。

表 6 基准报告期

最新报告期	资产负债表	利润表	现金流量表
一季报	上年年报	上年一季报	上年一季报
半年报	上年年报	上年半年报	上年半年报
三季报	上年年报	上年三季报	上年三季报
年报	上年年报	上年年报	上年年报

数据来源：CQR

表 7 基准报告期示例

最新报告期	资产负债表	利润表	现金流量表
20180331	20171231	20170331	20170331
20180630	20171231	20170630	20170630
20180930	20171231	20170930	20170930
20181231	20171231	20171231	20171231

数据来源：CQR

2.2 调整和更正

财务报表披露后，由于某些原因可能会进行一些修正，更改原始的科目数值。修正按照时间频率分为两种，一种是调整，一种是不定期更正。

财务报表第一次披露时的值，称为原始报表；在后续定期报告中再次披露基准报表时，如果基准报表和原始报表科目数值出现了不同，则发生了调整。以中铁工业（600528.SH）为例，在 2017/3/31 首次公布了其 2016 年年报（原始报表），如图 3 所示；2018/3/28 公布 2017 年年报时同时披露了 2016 年年报（基准报表），如图 4 所示。图 3 期末余额对应 2016 年原始报表，图 4 期初余额对应 2016 年基准报表。可以看到，货币资金、应收票据、应收账款和预付款项均进行了调整。

图 3 2016 年中铁工业合并资产负债表（部分）

合并资产负债表
2016 年 12 月 31 日

编制单位：中铁高新工业股份有限公司

单位：元 币种：人民币

项目	附注	期末余额	期初余额
流动资产：			
货币资金	(七)1	5,029,065,933.80	5,804,069,205.13
结算备付金			
拆出资金			
以公允价值计量且其变动计入当期损益的金融资产			
衍生金融资产			
应收票据	(七)4	1,446,912,807.36	457,243,020.17
应收账款	(七)5	15,372,891,499.20	15,703,344,098.78
预付款项	(七)6	12,583,410,842.95	11,330,011,031.13

数据来源：中铁高新年报

图 4 2017 年中铁工业合并资产负债表（部分）

合并资产负债表
2017 年 12 月 31 日

编制单位：中铁高新工业股份有限公司

单位：元 币种：人民币

项目	附注	期末余额	期初余额（已重述）
流动资产：			
货币资金	(七)1	6,847,797,345.18	8,370,064,573.52
结算备付金			
拆出资金			
以公允价值计量且其变动计入当期损益的金融资产			
衍生金融资产			
应收票据	(七)4	752,356,525.30	2,066,512,377.56
应收账款	(七)5	6,770,694,211.15	21,101,007,729.71
预付款项	(七)6	1,370,385,544.24	13,611,701,647.39

数据来源：中铁高新年报

不定期更正指的是，由于会计政策变更或者会计统计差错等原因，对之前某期的报告进行更正。更正需要发更正公告，没有明确的时间规律。以柳工

（000528.SZ）为例，其 2018/8/30 披露了 2018 年半年报，在 2018/9/29 日对半年报进行了更正，表 8 展示了利润表部分科目更正前和更正后的数值。

表 8 柳工 2018 年半年报利润表更正

会计科目	更正前	更正后
营业总成本	9,062,061,659.26	9,065,019,260.18
投资收益	53,176,657.27	54,502,547.68
营业利润	766,448,249.39	774,665,801.19
净利润	627,967,425.36	636,184,977.16
基本每股收益	0.5207	0.5295

数据来源：CQR

2.3 基本原则

在进行财务数据的量化研究时，调整和更正的存在会增加数据处理的复杂度。一个处理方式是，不考虑调整和修正，只使用第一次披露的数值。这种方式简单易实现，不会引入未来数据，逻辑上也讲得通；缺点是无法反映调整或更正等新信息，信息利用不充分。另外一种方法是始终采用调整后和修正后的最新数据，这种方式也简化了计算过程，缺点是历史回溯时会利用未来数据，导致结果偏差较大。

上面两种方式为了追求简单而牺牲了精确性，在实际应用中并不推荐。量化研究是一个精细的过程，事实上应该遵循以下原则：在当前时点利用可得的最新信息（point in time）。按照这个原则，既能保证充分利用最新的数据，例如考虑调整和更正，或者将业绩预告和业绩快报考虑进来，也可以完全避免使用到未来的信息。

2.4 数据存储

无论是 Wind 还是 Tushare，三大表在处理调整和更正时采用了相同的步骤。以大智慧（601519.SH）2014 年年报合并资产负债表为例，表 9 记录了其首次披露、调整和更正过程。

表 9 大智慧 2014 年合并报表的调整和更正

报告期	披露日期	说明
20141231	20150129	2014年年报首次披露
20141231	20160429	披露2015年年报的同时，还披露2014年年报，部分科目做了调整
20141231	20150129	在2015/1/29日对2014年年报进行了更正

数据来源：CQR

因此，底层表中同一个报告期可能会有多条记录，一般包括 4 条：类型 1-初始报表（最新值）、类型 2-基准报表（最新值）、类型 3-初始报表（原始值）和类型 4-基准报表（原始值）。这四种类型与 wind 和 Tushare 的对应关系如表 10 所示。

表 10 合并报表在底层表中的四种记录类型

类型	简称	Wind	Tushare	说明
1	初始报表（最新值）	合并报表	合并报表	首次披露的财务报表，如果有更正，则为更正后值
2	基准报表（最新值）	合并报表（调整）	调整合并报表	最新报告期披露的基准报告期报表，基准报告期为上年同期或上年年报；如果有更正，则为更正后值
3	初始报表（原始值）	合并报表（更正前）	调整前合并报表	如果有更正公告，则在更正公告发出后，修改初始报表相应的会计科目，修改前的记录标识为初始报表（原始值），修改后的记录标识为初始报表（最新值）
4	基准报表（原始值）	合并调整（更正前）	调整前合并报表	如果有更正公告，则在更正公告发出后，修改基准报表相应的会计科目，修改前的记录标识为基准报表（原始值），修改后的记录标识为基准报表（最新值）

数据来源：CQR

需要注意的是：1) 如果没有任何一个会计科目发生调整，则初始报表和基准报表记录相同；2) 由于对上市公司财务报表披露规范的指定，1998 之前的数据没有调整和更正，每个报告期只有一条记录；3) 如果更正发生在基准报告之前，则只有三条记录：初始报表（最新值）、初始报表（原始值）和基准报表（最新值）；4) 如果更正发生在基准报告之后，则上面四条记录都有。

2.5 如何使用

为了说明如何正确使用数据，图 5 展示了几种常见的情况。

第一种情况，如图 5a 所示，x 年年报在 t1 日披露后，如果没有发生调整和更正，此时只有 1 条数据即类型 1。t1 日后如果要使用该财报数据，则提取类型 1 即可。

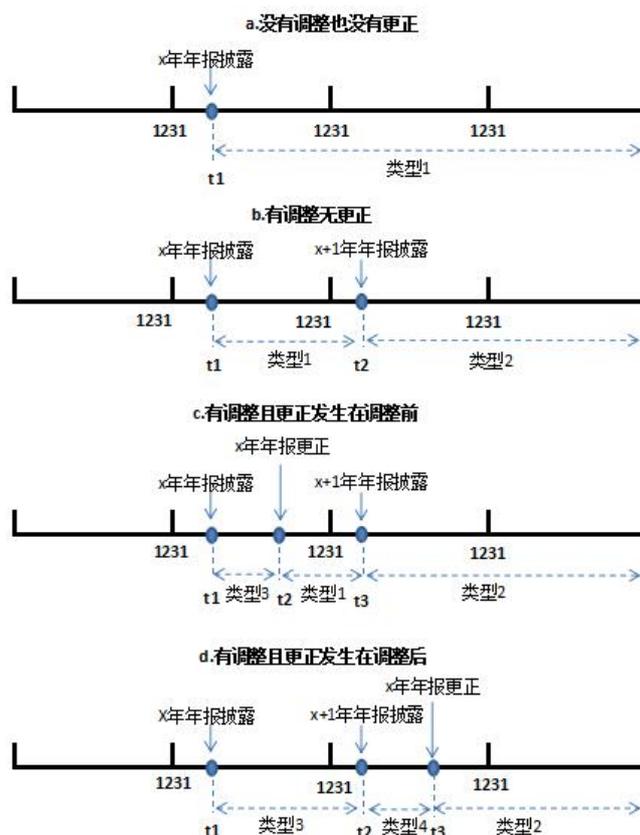
第二种情况如图 5b 所示，x 年年报在 t1 日首次披露，在 t2 日 x+1 年年报披露时顺便再一次披露了 x 年年报（基准报表）。则在 t1~t2 日如果要使用该财报数据，提取类型 1；在 t2 日后如果要使用该数据，则提取类型 2。

图 5c 展示了第三种情况，即除了常规的初始报表和基准报表，还发生了数据更正，且更正发生在基准报表之前⁴。在 t1 日 x 年年报披露，此时记录为类型 1；在 t2 时刻对年报进行了更正，此时更正后的数据记为类型 1，更正前的数据记为类型 3；在 t3 日披露调整数据，此时记录为类型 2。因此，如果在 t1~t2 日要使用 x 年年报，应提取类型 3；在 t2~t3 如果要使用该数据，则提取类型 1；在 t3 之后，提取类型 2 即可。

最后一种情况见图 5d，即更正发生在基准报表之后。t1 日披露 x 年年报，此时记为类型 1；t2 日披露 x 年年报基准报表，记为类型 2；t3 日发出 x 年年更正公告，此时修改类型 1 和类型 2 为更正后的最新值，原来的数据分别记为类型 3 和类型 4。因此，如果在 t1~t2 要使用该年报，则应提取类型 3；在 t2~t3 要使用该数据，则提取类型类型 4；t3 之后提取类型 2 即可。

⁴ 还有一种极其特殊的情形，即更正公告和基准报表在同一日披露。这种情况比较少见，小差错往往在基准报告中就可以做修正调整，除非影响比较大。如股票 600087.SH 在 2010 年年报，初次披露日期为 20110304，在 20120412 日披露了 2011 年年报及 2010 年年报基准报表，同一天也披露了 2010 年会计差错更正。这种情形在提取数据时，可以忽略掉会计更正记录，因为在新的基准报告中也会体现这种更正。

图 5 调整和更正示例



数据来源：CQR

3. 财务计算

3.1 单季度数据

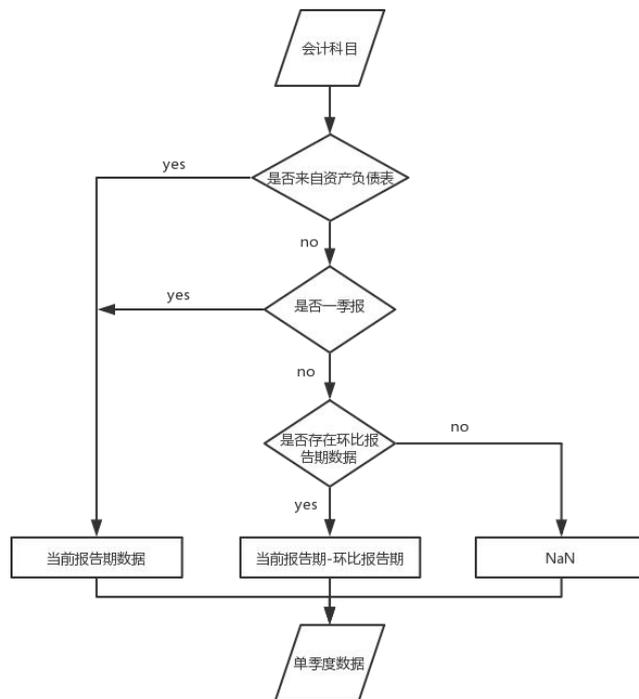
每年 12 个月可以分为四个季度，每个季度包含三月份，单季度数据对每个季度作单独衡量。由于资产负债表是一个时点数据，因此单季度数据和对应的报告期数据相同；对于利润表和现金流量表，一季度数据和一季报相同，其余三个季度数据等于对应报告期减去环比报告期数据；如果环比报告期缺失，则无法计算单季度数据，单季度数据为空。具体对应关系见表 11，计算流程图见图 6。

表 11 单季度指标计算

季度	资产负债表	利润表	现金流量表
一季度	一季报	一季报	一季报
二季度	半年报	半年报减一季报	半年报减一季报
三季度	三季报	三季报减半年报	三季报减半年报
四季度	年报	年报减三季报	年报减三季报

数据来源：CQR

图 6 单季度指标计算流程图



数据来源：CQR

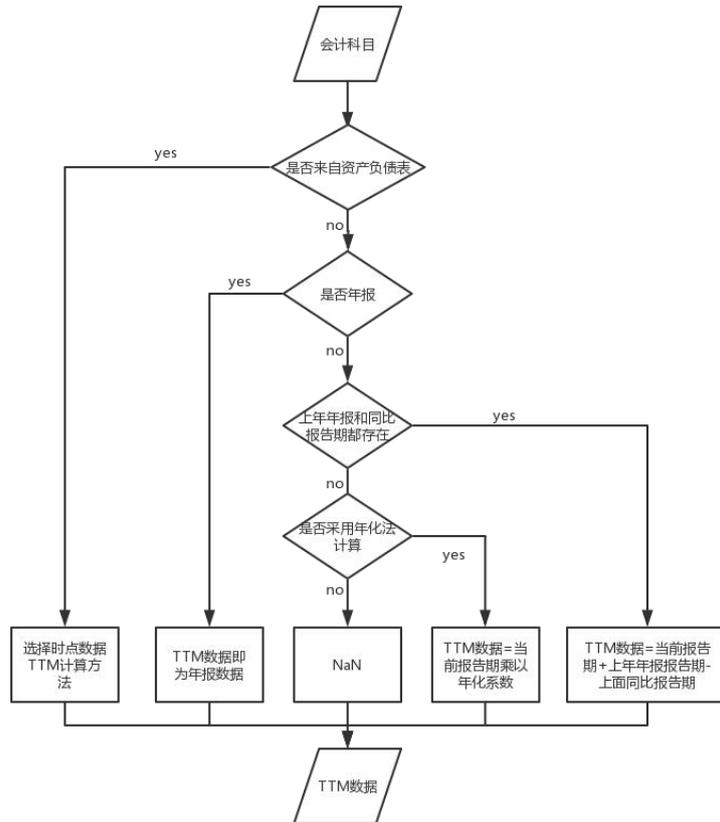
3.2 TTM 数据

TTM 是 Trailing Twelve Months 首字母的缩写，即滚动 12 个月数据，以消除季节性差异使得横向可比，具体计算流程图见图 7。

对于资产负债表来说，由于其为时点数据，如果要计算其科目 TTM 数据，常见有三种算法：1) 当前最新数据；2) 最近四个报告期平均值；3) 当前最新报告期和其同比报告期的平均值。

对于利润表和现金流量表来说，如果最新报告期为年报，则 TTM 数据即为年报数据；如果是其他报告期，则 TTM 数据等于当前最新报告期数据，加上上年年报数据，减去上年同比报告期数据；如果上年年报或同比报告期不存在，可以使用年化算法，即当前最新报告期数据乘以年化系数⁵。

图 7 单季度指标计算流程图



数据来源：CQR

⁵ 一季报、中报和三季度报的年化系数分别为 4、2、4/3

欢迎提供建议和沟通交流

公众号：

Chihiro Quantitative Research



刀疤连个人微信：



知乎专栏 1：跨资产的因子配置

https://zhuanlan.zhihu.com/c_152990241

知乎专栏 2：因子投资那些事儿

<https://zhuanlan.zhihu.com/factor-investing>