

研究通讯

宏观量化的周期拐点和PPI领先指数设计

广发期货发展研究中心

电话: 020-88830760

E-Mail: zhaoliang@gf.com.cn

摘要:

本文基于周期拐点的视角提出了宏观量化的研究思路,主要包括以下几步:变量选取、数据处理、合成方法、拐点识别、拐点匹配、效果评价,并通过PPI领先指数的设计对这一流程进行实践。

宏观变量的选取和处理部分,对常用的数据归纳为价格、金融条件、工业、进出口、消费、投资、财政7个大类及其细分的28个小类,后续可以继续扩展补充。根据数据的不同特点设置不同的处理步骤,举例分析其中较为重要的步骤:季节调整(X13)、去除噪音(HP滤波、EMD分解、LLT滤波)。周期拐点的识别与匹配部分,介绍了B-B算法步骤和通用的匹配规则,以及效果评价指标。合成方法部分介绍了指数合成法(OECD)、主成分分析法(PCA)、动态因子模型(DFM)三种方法的原理和优劣势。

PPI领先指数设计部分,使用PPI全部工业品当月同比作为基准指标,处理之后进行拐点识别发现:PPI周期的数值规律明显,极值点的持续时间很短,周期长度并不稳定最短为5个月最长为39个月其余大部分在15个月附近。宏观变量主要选用两类:一类是与PPI相关性较高的价格类指标;另一类为工业企业盈利、货运量、进出口等反映经济景气的指标。

对比三种合成方法,从领先效果来看,OECD平均值为4中位数为2.5,PCA平均值为4.7中位数为3,DFM平均值为3.6中位数为3.5;时差相关性方面OECD和PCA分别为0.86和0.92,DFM略差为0.66。整体而言,使用这一宏观量化流程合成得到的领先指数对PPI的预测能取得较好效果存在明显的相关性,基本保持在一个季度左右的领先期数,对于提前预判PPI走势具有重要意义。

宏观量化中的数据、模型、方法种类繁多没有固定的范式可循,这套基于周期拐点的流程通用性强可应用领域广,我们将继续围绕周期分析、经济领先指数、行业领先指数等角度进行研究实践。

投资咨询业务资格:
证监许可【2011】1292号

联系信息



叶倩宁

期货从业资格: F03088416

投资咨询资格: Z0016628

电话: 020-88818051

邮箱: yeqianning@gf.com.cn



陈俊州

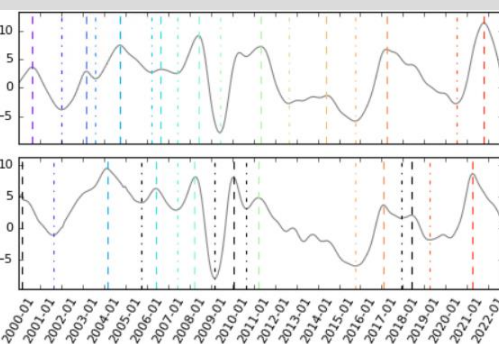
期货从业资格: F3039003

投资咨询资格: Z0014862

电话: 020-88818017

邮箱: chenjunzhou@gf.com.cn

PPI 拐点匹配



拐点匹配的效果评价

评价指标	CI_OECD	CI_PCA	CI_DFM
拐点总数	17	17	17
拐点匹配数	10	7	8
拐点未匹配数	7	5	9
未匹配点前期	0	5	0
未匹配比例	0.411765	0.294118	0.529412
x 冗余	7	4	9
x 冗余比例	0.411765	0.363636	0.529412
时差平均数	4	4.714286	3.625
时差中位数	2.5	3	3.5
时差标准差	4.521553	4.88925	5.289815
时差变异系数	1.130388	1.037114	1.459259
相关性最大值	0.859043	0.923903	0.656528
相关性最大阶数	3	4	2

目录

一、宏观量化的研究思路	1
(一) 研究背景	1
(二) 研究思路	1
二、变量选取与数据处理	2
(一) 变量选取原则	2
(二) 数据处理流程	3
三、周期拐点的识别与匹配	6
(一) 拐点识别	6
(二) 拐点匹配	7
(三) 评价指标	7
四、合成方法	8
(一) 合成指数法 OECD	8
(二) 主成分分析法 PCA	8
(三) 动态因子模型 DFM	9
五、PPI 领先指数设计	9
(一) 变量的选取处理	9
(二) 基准指标的周期拐点	10
(三) 合成方法的效果对比	12
六、结论	14
免责声明	15

一、宏观量化的研究思路

（一）研究背景

资产价格的变化通常会受到宏观经济因素的驱动，宏观经济指标按照影响资产价格的方式可以分为两类：政策变量和统计数据，政策变量更多的通过直接作用于定价机制对资产价格起作用，相对快速高效；而统计数据则更多通过影响人们的预期，然后通过交易影响资产的价格。当前市场中宏观预期的形成以及宏观预期到价格的传导需要时间，这也是研究宏观经济指标对于资产价格预测研究的意义。

由于宏观经济指标体系庞大，涉及的逻辑传导错综复杂，某个经济指标和资产价格的对应关系可能有好几条逻辑链条串联，导致了每种宏观状态事后总有可解释的逻辑，而可解释的逻辑也因为叠加了太多额外的数据和观点仅在某个时刻有用，使得对于宏观经济指标的研究经常困于事后解释和逻辑补充之中。近年来宏观量化研究逐渐兴起，这种方案减少了对宏观经济逻辑传导机制的依赖，而是通过对数据的建模验证获得显著的统计规律，认为在一定的置信水平下内在逻辑将会在概率角度提供给我们较为准确的预测辅助，有效的宏观逻辑是能够被数据量化验证的。

目前宏观量化主要集中在几个方向：一是应用于周期分析，包括宏观周期例如经济周期、信用周期、库存周期的量化分析，也包括中观行业周期比如行业景气度的判别等；二是应用于领先滞后关系的判别，包括合成领先指数、资产收益归因配置等；三是应用于选股因子挖掘，探寻宏观经济指标对上市公司股价影响的有效性，从而提供新的选股因子角度。

宏观量化能够批量处理各类宏观指标和资产类别，通过遍历搜索帮助发现被忽略的内在关系，但是宏观量化研究不像传统量化那样有较为成熟的范式可循，尤其是当研究的对象为低频的经济数据时，合理的研究流程显得更为重要。我们从量化视角针对如何有效处理经济数据，从而定量研究指标的周期规律及领先滞后关系，提出了一整套的量化研究思路，用以解决以下几个核心问题：

- 经济指标的周期、阶段、拐点如何确认？
- 指标间的领先滞后关系如何确认？
- 是否有科学的指数编制方法？
- 合成指数样本外能否有效预测峰与谷？

（二）研究思路

解决上述问题，宏观量化研究的主要步骤如下：

变量选取是研究的目标所在，指标主要分为两类：基准指标与宏观变量。基准指标的选取原则是能够作为当前所研究目标的指代变量，本身具有一定的重要性；宏观变量是对基准指标具有关联关系和解释性的一组数据，宏观变量种类繁多具有体系化、多样化的特点，常见有宏观的如投资类、消费类、进出口类、货币类、信用类等，中观的如行业产量类、库存类等。宏观变量的选取要考虑经济逻辑性、领域重要性、发布及时性等原则。

数据处理是研究效果的前提保障，实际中的经济数据可能出现不完整（存在缺失）、不连续（数据频率变化）、波动异常（季节性因素、统计误差或极端值）等情况，因此需对其进行预处理才能更好的进行后续的分析

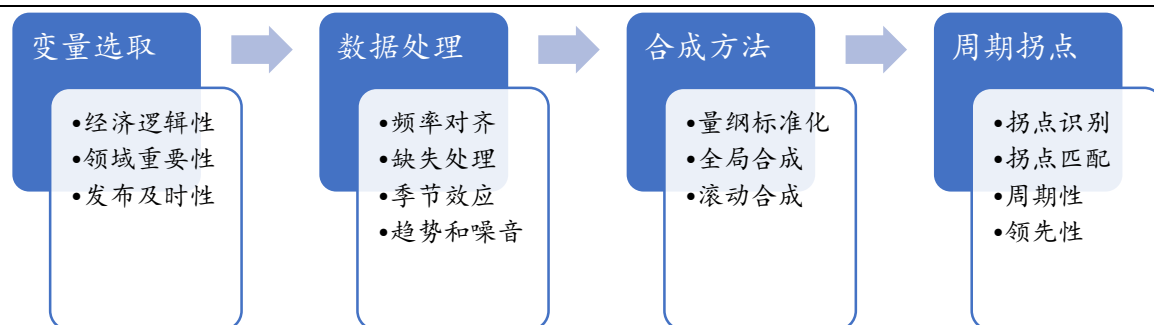
和研究。

合成方法是研究数据整合的重要步骤，选取的宏观变量经过处理之后如何将其复合为一个可跟踪的综合指数，需要考虑数据量纲大小、共线性、数据降维等问题，不同方法之间有其优势或劣势，需要根据场景选择使用。

拐点识别是研究流程中的关键步骤，拐点是时间序列中一类特殊的极值点，它确定了周期、阶段的划分，刻画了周期演绎的过程，同时也蕴含了指标间相位差的信息，确定了拐点也就抓住了该指标周期性的基本特征。

拐点匹配是流程中分析关联关系的重要手段，对序列的拐点进行一一对应，明确彼此的相位对照关系，从而确认两者之间的领先滞后性。传统的关联分析一般使用互相关函数，通过遍历不同滞后期数，使用相关性分析、回归分析等方法测试序列的线性关系，这种做法在各期回归结果区别较大时能得出直观结论，而当某一序列有较强的自相关性时，该方法就存在局限性。而拐点匹配方法的优势在于，不仅能计算平均拐点间隔时间，还能对经济指标的周期和阶段进行定期，并结合历史情形分析拐点的合理性与发生逻辑，从而使量化结果更具有解释力。

图：宏观量化的研究思路



来源：广发期货发展研究中心

二、变量选取与数据处理

（一）变量选取原则

宏观变量的选取要考虑几个选取原则：

经济逻辑性：宏观变量需要对研究目标有着宏观上的传导逻辑。

领域重要性：宏观变量是各个领域较为重要的宏观变量，能够较为全面的反映出当前领域的情况。

发布及时性：领先指标的发布频率不宜较低，且不能有太多滞后。

通过整理筛选宏观研究中常为投资者所关注的指标，我们将指标归纳为价格、金融条件、工业、进出口、消费、投资、财政7个大类及其细分的28个小类，后续可以根据研究目标从中选取关注的宏观变量进行处理使用。

表：宏观变量的划分

大类	小类	举例
价格	总价格指数	CPI:当月同比
	消费品价格	22个省市:猪粮比价:同比
金融条件	货币量	社会融资规模存量:同比
	货币价格与利差	SHIBOR:1年
	外汇汇率	官方储备资产:外汇储备:同比

	资产价格	百城住宅价格指数:同比
工业	工业增加值	工业增加值:当月同比
	PMI	PMI:采购量
	货运量	货运量总计:当月同比
	企业	工业企业:利润总额:当月同比
	生产	产量:发电量:当月同比
	工业品价格	南华综合指数:同比
	生产资料销售	销量:汽车起重机:主要企业:当月同比
进出口	进出口总额	贸易差额:当月同比
	进口	进口金额:当月同比
	出口	出口金额:当月同比
	吞吐量	沿海主要港口货物吞吐量:当月同比
消费	总额	社会消费品零售总额:当月同比
	消费者指数	消费者信心指数
	商品销售量	限额以上企业商品零售总额:当月同比
	消费品产量	零售额:粮油、食品类:当月同比
投资	总投资	固定资产投资完成额:累计同比
	房地产	房屋施工面积:住宅:累计同比
	其他部门投资	固定资产投资完成额:制造业:累计同比
财政	税收	税收收入:当月同比
	收入	中央财政收入:当月同比
	支出	公共财政支出:当月同比
	政府基金	全国政府性基金收入:累计同比

来源：广发期货发展研究中心

（二）数据处理流程

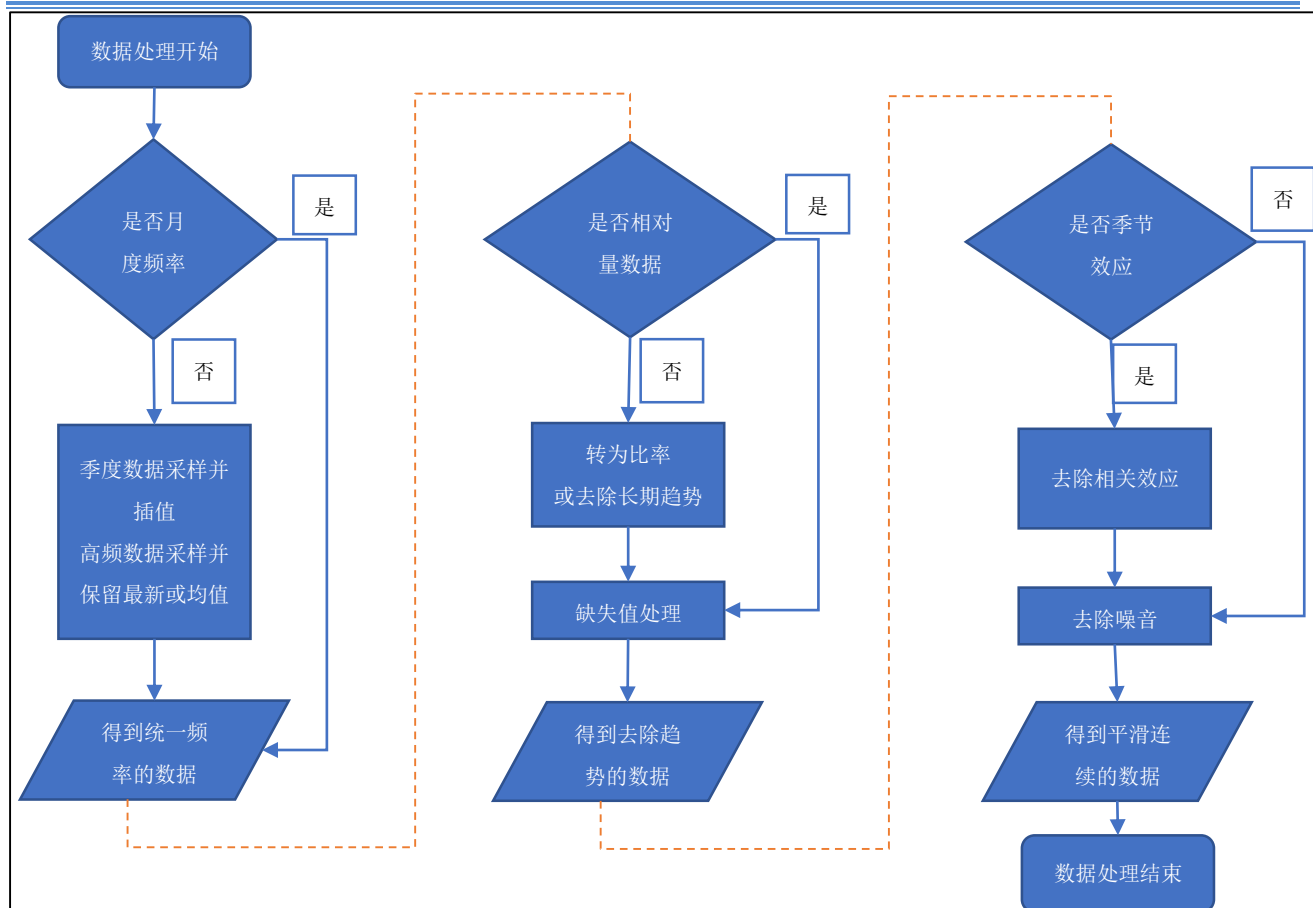
宏观变量的时序数据往往包含四种要素:趋势要素(T)、循环要素(C)、季节要素(S)与不规则要素(I)。不同的数据所需要的处理步骤并不相同,需要结合数据特点进行步骤选择,通常有以下步骤:

首先要得到**同一频率的数据**,若指标频率不一我们无法直接对比,高频数据转低频通常需要采样,保留采样窗口的末期值或窗口均值。低频数据转高频通常需要进行线性插值,补充中间时间点的数值大小。

其次是得到**去除趋势的数据**,宏观数据主要分为两类,一类为比率或利率数据,例如国债十年期到期收益率、PMI,该数据本身并不含有明显的长期趋势性不需要进行趋势去除。第二类为绝对量的数据,例如MO存在着明显的趋势性,可以通过长期滤波或者计算同比的方式转化为增长率序列。

最后是得到**平滑连续的数据**,为了能将循环要素剥离出来,需要尽可能的去除季节要素与不规则要素的影响。季节要素是指每年不同时间段非循环因素对序列的影响,主要由于气候、节日、工作日数等因素组成,其中影响程度最大的是月份效应和节日效应,对此可以使用X-13模型;不规则要素是由偶然因素所引起的噪声,会干扰对指标周期性与拐点的判断,平滑方式主要分为全局平滑与滚动平滑。

图：数据处理流程示意图

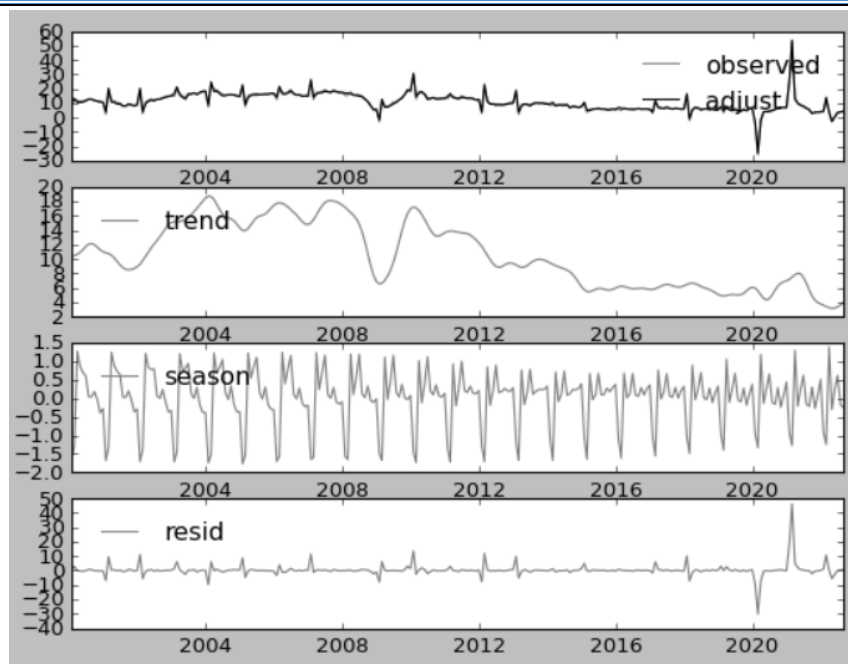


来源：广发期货发展研究中心

需要进行季节效应调整的指标主要有以下几大类：消费与商品价格类，春节往往是购物旺季，会显著增加商品的销量，抬升其价格；货币类，在春节前往往提现的需求较高，因此M0在春节前的增速比较显著，M1与M2的春节效应并不明显；工业生产类，春节的位置会影响1、2月份的实际工作日，进而影响工业产量，如工业增加值、货运量、发电量、汽车产量等；进出口类，企业往往会在春节前减少进口，降低库存成本，因此进出口类指标也需要春节效应调整。

以工业增加值同比增速为例的季节分析如下，可以看到数据的季节要素较为明显，而且每年的1、2月份通常会存在较大幅度的变化。

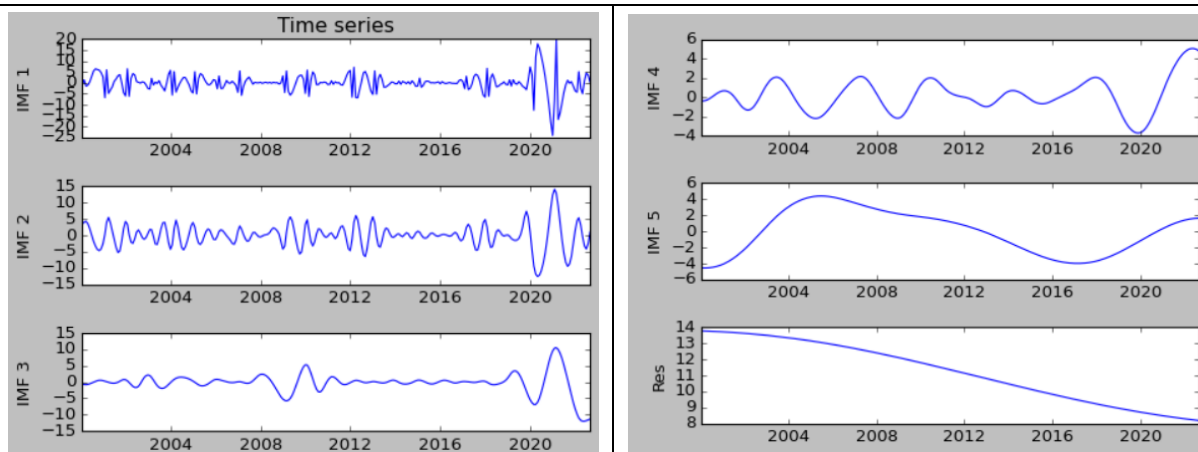
图：工业增加值同比的季节调整



来源：广发期货发展研究中心

去除噪音的平滑方式分为全局或滚动两大类：全局平滑的方法主要是各种滤波（例如BK滤波、HP滤波、EMD分解），滚动平滑的方法主要是各种移动平均（例如MA、EMA、LLT）。HP滤波是宏观经济学中常用的工具，例如OECD在做去除趋势与平滑时都使用此方法，第一步用高参数 λ 剔除趋势，第二步用低参数 λ 去除波动项中的噪声；EMD分解假设任何复杂的信号都是由若干瞬时频率有意义且唯一、但不同时刻瞬时频率可不同的本征模函数（IMF）叠加组成，分解过程不断从信号中提取出本征模态函数，直到信号仅保留趋势项为止。

图：工业增加值同比增速的 EMD 分解

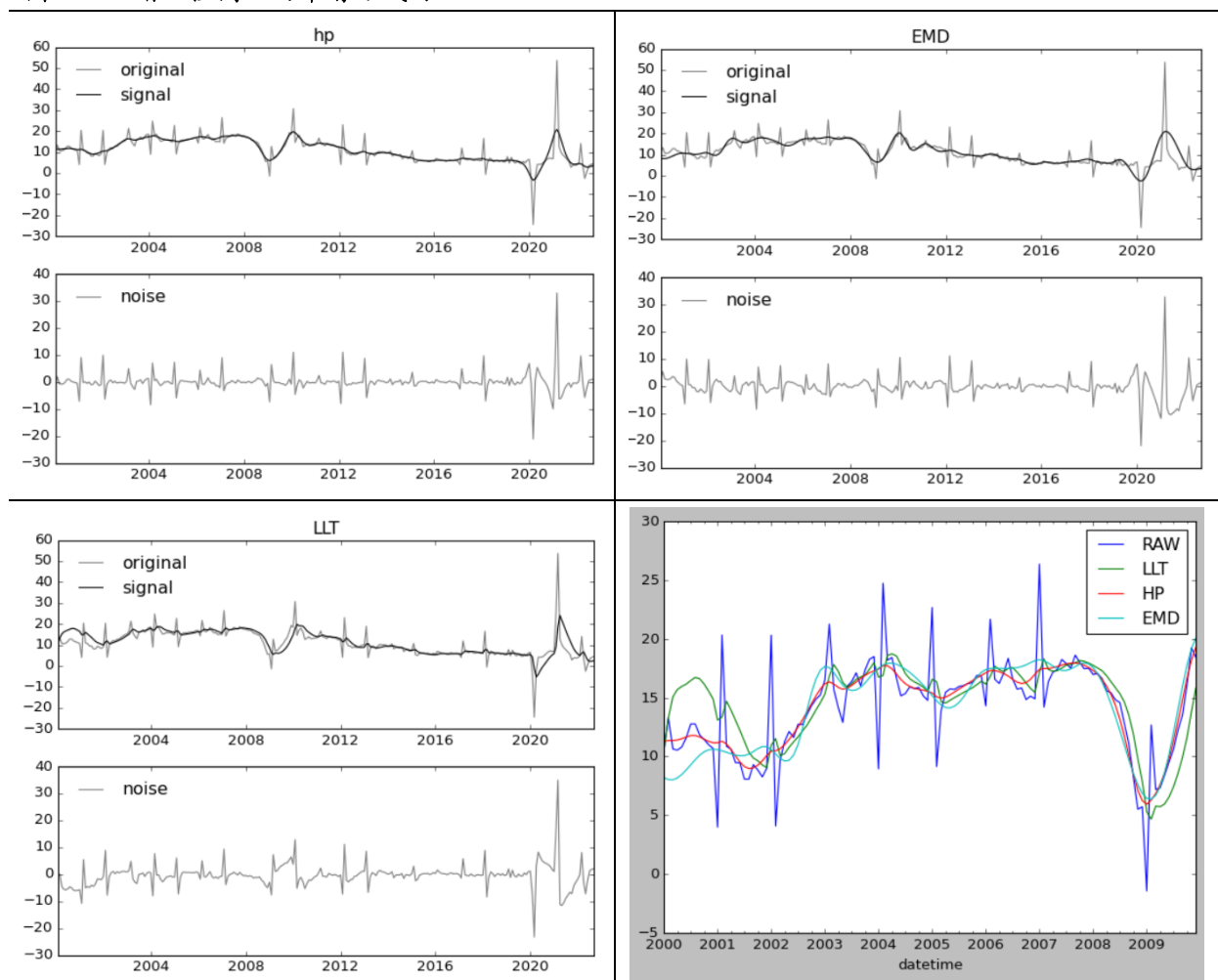


来源：广发期货发展研究中心

以工业增加值同比增速为例进行去除噪音，月度数据的HP滤波参数通常使用14，EMD分解的选用标准可以是层数（通常使用2）或者信号噪音比的阈值（通常使用1.5），LLT类似移动平均的参数（通常使用12）。从结果来看，全局平滑效果较好，但比较依赖经验参数，加入新数据时会改变之

前的平滑结果；滚动平滑则不会产生结果前修，但是受近期数据的影响明显，存在极值点相比原始数据后移的情况。

图：工业增加值同比的平滑方式对比



来源：广发期货发展研究中心

三、周期拐点的识别与匹配

前面介绍了宏观变量必要的数据处理过程，最终得到平滑连续的数据在周期分析中才具有实用价值。周期的演绎模式可以由拐点来定义，拐点划分了指标序列的阶段，确定了拐点也就抓住了该指标周期性的基本特征。

（一）拐点识别

识别周期拐点的最常用模型可以分为参数及非参数两类，参数类主要有Hamilton(1989)的马尔科夫转换模型，非参数类则主要是Bry-Boschan(1971)的B-B算法，以及Bruno and Otranto(2004)提出的拐点全面分类定期方法。B-B算法本质其实是一种极值筛选法，将对拐点的判断方式归纳为一系列简单决策原则，用于层层筛选过滤，最后剩下的极值被认定为拐点。具体算法步骤如下：

- (a) 识别全部极值。寻找前后范围内的最大/最小值（参数通常5个月）。
- (b) 检查相邻极值。首先判断每个极值点在相邻两个极值点之间是否为最大/小值，不满足则删除；其次判断极值点是否交替出现，若有重复峰/谷，

只保留时间最早的较大/小点。

(c) 检查周期长度。峰与峰、谷与谷间的距离是否满足最短时限（通常剔除小于15个月），不满足则删除两个峰值中较低的那个或是两个波谷中最高的那个。

(d) 第二次检查相邻极值。在短周期被删除后序列可能再次包含多个峰值或波谷序列，因此需要重复第二步。

(e) 检查半周期长度。峰与谷的间隔是否满足最短时限（通常剔除小于6个月），不满足则删除偏离均值较小的极值点。

(f) 第三次检查相邻极值。重复第二步，最终剩下的极值点认定为拐点。

（二）拐点匹配

确定拐点后，我们一方面也确定了序列的周期和阶段，另一方面也借此得以进一步研究两个经济指标的关联，最终目的是确定两条时间序列之间的领先或滞后的期数。在此之前就需要对两个序列的拐点进行一一对应，明确彼此的相位对照关系。

拐点匹配在实际操作中并没有被广泛使用或主要机构公布的算法，这一过程可以基于时间相隔的远近来设计匹配逻辑，将拐点分为对应、缺失、多余三种可能的状态。假设基准序列B与变量序列X，主要步骤如下：

(a) 定位匹配点。判断序列X在 $(-M, N)$ （负数表示滞后，一般使用-6至18）的时间范围内，是否存在与序列A中各拐点对应的相同性质（峰/谷）拐点。

(b) 区分缺失的原因。对基准序列B中的每个拐点 $b(j)$ ，在X的上述范围内寻找类型相同的、最近的拐点进行匹配，若存在匹配则进入下一步。否则判断该拐点的日期是否小于两序列时间交集的最早日期，若否则标记为对应拐点缺失，若是则标记为早期缺失，意味着是由于数据序列过短造成的。

(c) 检查重复的情况。判断上一步中寻得的X中的拐点 $x(i)$ 是否此前已有对应拐点。若否，标记 $b(j)$ 和 $x(i)$ 相对应，若是，则比较此前与之对应的拐点与 $x(i)$ 之间的时间间隔是否小于 $b(j)$ 与 $x(i)$ 之间的时间间隔。小于则将 $b(j)$ 标记为拐点缺失，若否则修正 $x(i)$ 的对应关系为 $b(j)$ ，保持最近相邻的原则。

(d) 检查交叉的情况。对所有 $b(j)$ 执行上述两步之后，检查每个有对应点的 $x(i)$ 其对应的 $b(j)$ 是否满足时间上的由远至近。若 $x(i)$ 对应 $b(j)$ 与 $x(i+1)$ 对应 $b(j+1)$ 出现时间戳 $j > j+1$ 的情况，则判断 $x(i)$ 与 $b(j)$ 之间的时间间隔是否小于 $x(i+1)$ 与 $b(j+1)$ 之间的时间间隔，若是则将 $b(j+1)$ 标记为对应拐点缺失，若否则将 $b(j)$ 标记为对应拐点缺失。

(e) 检查X中剩余拐点。将无对应的 $x(i)$ 标记为多余拐点，但如果 $x(i)$ 在B序列最后N月内的除外，因为其对应的 $b(j)$ 仍有可能在未来出现。

（三）评价指标

通过上述方法计算宏观变量的拐点以及匹配结果之后，对结果进行客观评价可以结合历史情景进行分析，也可以借助相关指标进行客观评价，基于匹配结果的评价指标主要有：

未匹配比例：反映基准指标中未能匹配的比例，数值越高说明所选变量的解释能力越弱。

冗余比例：反映宏观变量中无效的拐点比例，数值越高说明所选变量本身的无效信号次数越高。

时差平均值:拐点的平均领先阶数,数值越大说明领先性越强,但而过大的领先性有周期错配的可能性,结果可信度不足。

时差标准差:拐点领先阶数的标准差越小,拐点的领先期稳定性越强。

此外,时差相关性分析也是用来判断经济指标的领先与滞后阶数的常用方法,通过计算领先与滞后N期的候选指标与基准指标间的相关系数,选择相关系数最大的阶数作为领先或滞后阶数,数值结果可以与拐点匹配分析进行相互印证。

四、合成方法

选取的宏观变量经过上述流程处理之后,如何将其复合为一个可跟踪的综合指数从而方便与基准变量进行比较,主要合成方法包括扩散指数法、合成指数法、主成分分析PCA、动态因子模型DFM。其中合成指数法在不同经济研究机构中采用的方法略有差异,经合组织OECD历史最为悠久、应用最为广泛。本章我们简要介绍方法的原理及优缺点,后续再通过实际数据的合成结果与基准变量进行拐点匹配分析不同方法间的差异。

(一) 合成指数法OECD

OECD自1978年始基于增长循环理念编制了各成员国的线性合成领先指数,主要步骤如下:

各个宏观变量为 $C_j(t)$, ($j=1, 2, \dots, m$), m 为宏观变量个数,基准指标为 $X(t)$

(1) 求各个指标的标准化偏差 SD_j

$$SD_j = \left(\sum_{t=1}^T (|C_j(t) - \bar{C}_j|) \right) / T$$

(2) 计算宏观变量的平均标准化序列 $S(t)$

$$S(t) = \left(\sum_{j=1}^m (C_j(t) - \bar{C}_j) / SD_j \right) / m$$

(3) 计算调整因子 k 和 d ,使得合成指数与基准指数具有可比性

$$k = \left(\sum_{t=1}^T |X(t) - \bar{X}| \right) / \left(\sum_{t=1}^T |S(t) - \bar{S}| \right)$$

$$d = \bar{X} - \bar{S}$$

$$CI(t) = k * S(t) + d$$

OECD法计算简便,不存在任何信息损失。对于波动较大的指标,标准化偏差较大时将给予其较低的权重,来降低噪声对合成指数的干扰。不过当指标池中存在经济含义高度重叠的备选指标时,OECD法会重复计算其权重,因此在筛选指标时应尽量避免相关性过高的指标。

(二) 主成分分析法PCA

主成分分析法通过降维的方式将多个指标复合为综合指标,通过正交变换将指标转换为线性不相关,从而提取指标组的主要信息。步骤如下:

(1) 设宏观变量 X 中的指标序列 X_i ($i=1, 2, \dots, p$),将 X 标准化后计算协方差矩阵的特征值 λ_i (从大到小排列)及其对应的特征向量 $U_i =$

$$(U_{i1}, U_{i2}, \dots, U_{ip})^T$$

(2) 计算主成分

$$F_i = U_{i1}X_1 + U_{i2}X_2 + \dots + U_{ip}X_p$$

(3) 对主成分按特征值从大到小排列并计算主成分的方差贡献率，在 p 个主成分中选取方差贡献率之和达80%的前 m 个主成分，以主成分的方差贡献率作为权重对 m 个主成分进行加权求和

$$w_i = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$$

$$F = \sum_{i=1}^m w_i * F_i$$

(4) 计算调整因子 k 和 d ，使得合成指数与基准指数具有可比性

PCA法考虑了多变量的相关性，因而允许备选指标的经济含义高度重叠。但由于丢弃了其他主成分，PCA法存在一定程度的信息丢失。最关键的是，由于不同时间区间第一主成分蕴含的信息量不同，会造成不同时期合成指数的含义不完全一致。

(三) 动态因子模型DFM

DFM的基本思想是经济的周期波动是通过一系列经济变量的活动来传递和扩散的，任何一个经济变量本身的波动都不足以代表宏观经济的整体波动，因此需要提取各个变量波动的共同动态因子。设宏观变量为 $X_t = (X_{1t}, X_{2t}, \dots, X_{Nt})^T$ ，步骤如下：

(1) 隐含状态方程中 $\lambda(L)f_t$ 称为 X_t 的共同部分， $\lambda(L)$ 为由 s 阶滞后算子多项式组成的 $N \times q$ 维矩阵

$$X_t = \lambda(L)f_t + e_t$$

(2) 假定动态因子 f_t 服从某一向量随机过程 $VAR(P)$ ，其中 $A(L)$ 为 p 阶滞后算子多项式组成的 $q \times q$ 维矩阵， u_t 为驱动动态因子的 $q \times 1$ 维结构冲击活原始冲击

$$A(L)f_t = u_t$$

(3) 假定异质部分 e_t 和原始冲击 u_t 在所有时期 k 都不相关

$$E(e_t u_{t-k}) = 0$$

DFM模型不仅允许观测变量受因子滞后项的影响，而且也允许因子本身具有独立的动态演化过程，具有双重动态特征。

五、PPI领先指数设计

(一) 变量的选取处理

生产价格指数(PPI)是衡量工业企业用品出厂价格变动趋势和变动程度，反映一时期生产领域价格变动情况的重要经济指标，对股票、大宗商品等资产价格都有重要影响。如果能通过一系列经济指标设计领先指数，对于提前预判PPI走势具有重要意义。

PPI从两分法的构成上来看，主要受到生产资料中的原材料工业与加工工业的影响，从39个不同行业的细分构成来看，影响较大的为黑色金属、石油、煤炭、化工、有色金属相关行业。因此我们选用的基准指标为PPI全部工业品当月同比，选用的宏观变量主要是两类，一类是与PPI相关性较

高的价格类指标；另一类为工业企业盈利、货运量、进出口等反映经济景气的指标，当货运量提高，工业企业盈利改善，进出口额上涨，工业企业的对原材料需求上升，推动PPI上行。具体指标及所需要的处理步骤如下：

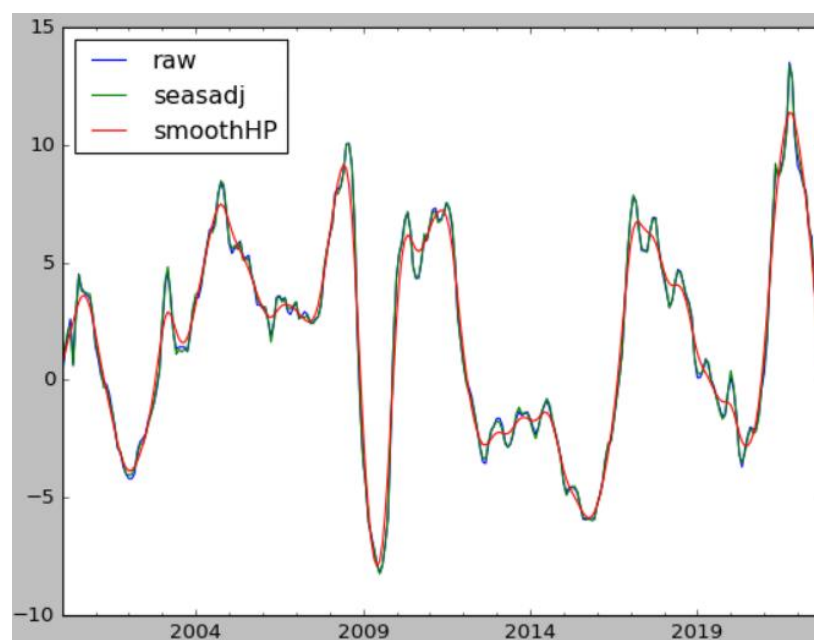
表：不同数据所需的处理步骤

大类	指标名称	调整频率	转为同比	去除季节效应	去除噪音
基准指标	PPI:全部工业品:当月同比			✓	✓
价格	RJ/CRB 商品价格指数	✓	✓		✓
	南华综合指数	✓	✓		✓
工业	波罗的海干散货指数 BDI	✓	✓		✓
	期货收盘价电子盘:LME3 个月铜	✓	✓		✓
	铁路货运量:当月同比				✓
	工业企业:亏损企业单位数:同比				✓
	工业企业:亏损企业亏损额:累计同比				✓
进出口	进出口金额:当月同比			✓	✓
	进口金额:当月同比			✓	✓
	出口金额:当月同比			✓	✓
	工业企业:出口交货值:当月同比			✓	✓

来源：广发期货发展研究中心

基准指标（PPI:全部工业品:当月同比）数据处理之后的结果如下，可以看到季节调整后数据变化不大，原因在于其本身就是同比数据，采用HP滤波去除噪音之后仍能较好的保留主要的周期规律和拐点情况。

图：基准指标的数据处理结果



来源：广发期货发展研究中心

（二）基准指标的周期拐点

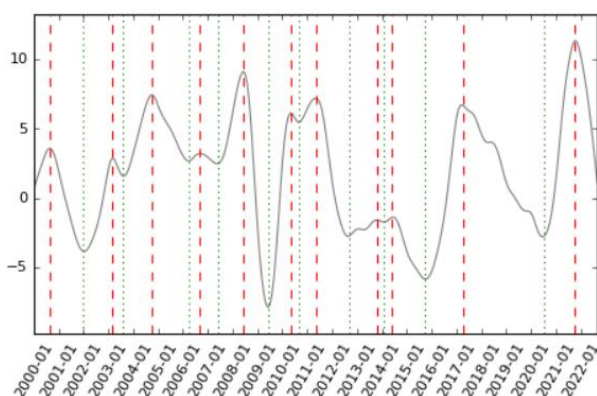
使用上述B-B算法对处理后的基准指标（PPI:全部工业品:当月同比）进行拐点识别，在第一步识别全部极值点时采用的参数为5个月，经过第二步的相邻检查和交叉检查之后得到的拐点数21个，可以看到存在明显的

间隔周期较小的点需要继续剔除；第三步检查周期长度采用的参数为15个月，经过第四步检查之后剔除了4个拐点（2010年5月，2010年9月，2013年11月，2013年2月）；在第五步半周期检查采用参数为6个月，经过第六步检查之后，半周期间隔都满足要求无需剔除。最终得到拐点数量17个，间隔月份的长度并不稳定，最短为5个月最长为39个月其余大部分在15个月附近。

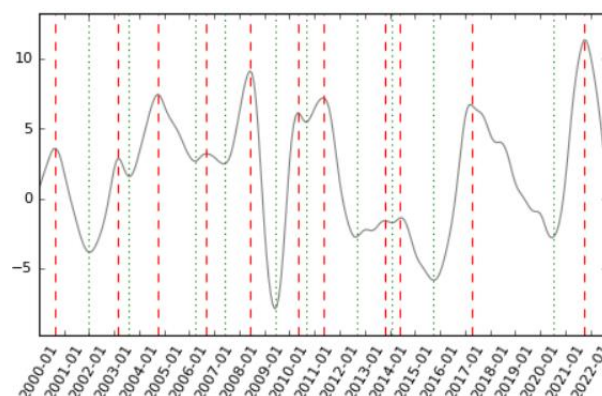
表：拐点识别结果和间隔月份

	拐点	间隔月份		拐点	间隔月份
2000/9/30	1		2009/6/30	-1	12
2002/1/31	-1	16	2011/5/31	1	23
2003/3/31	1	14	2012/9/30	-1	16
2003/8/31	-1	5	2014/6/30	1	21
2004/10/31	1	14	2015/10/31	-1	16
2006/4/30	-1	18	2017/4/30	1	18
2006/9/30	1	5	2020/7/31	-1	39
2007/6/30	-1	9	2021/10/31	1	15
2008/6/30	1	12			

(a) 识别全部极值

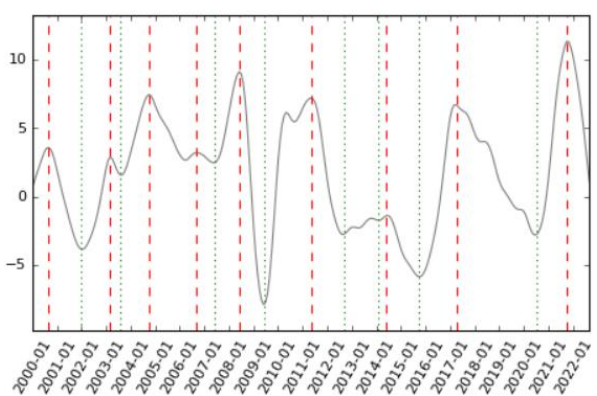


(b) 检查相邻极值

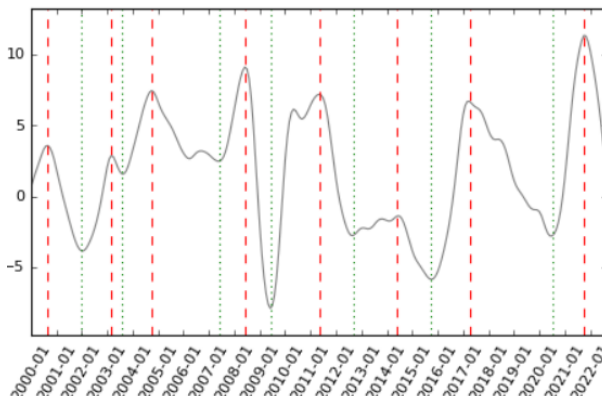


来源：广发期货发展研究中心

(c) 检查周期长度

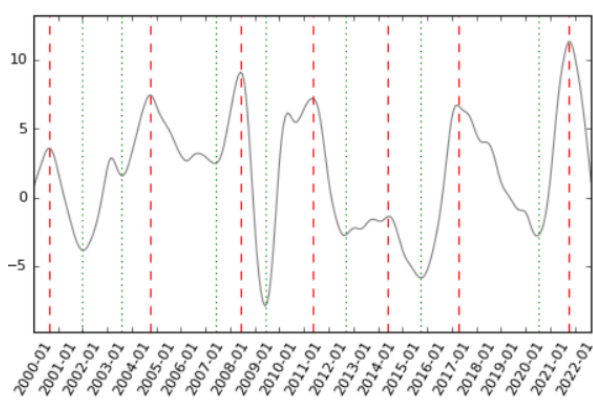


(d) 第二次检查相邻极值

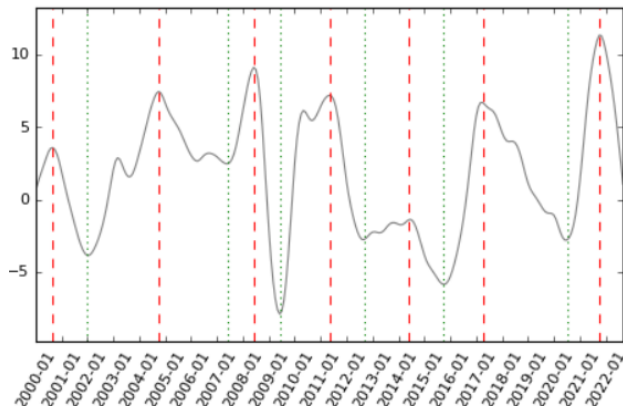


来源：广发期货发展研究中心

(e) 检查半周期长度



(f) 第三次检查相邻极值



来源：广发期货发展研究中心

(三) 合成方法的效果对比

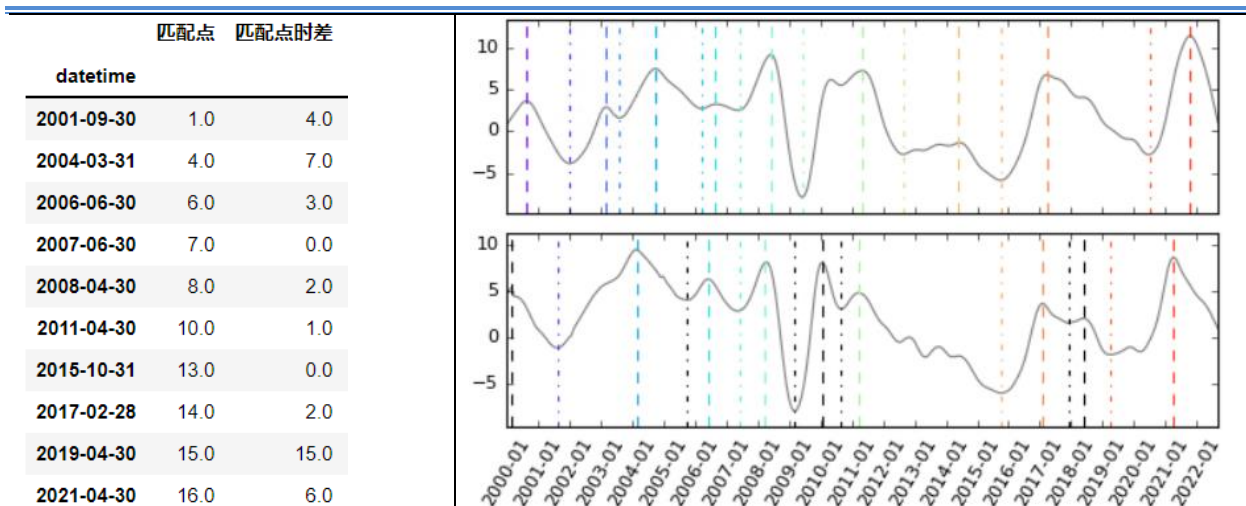
宏观变量经过数据处理之后，采用前述三种不同的合成方法得到领先指数并进行拐点识别，再与基准指标（PPI:全部工业品:当月同比）进行拐点匹配。

从未匹配比例、冗余比例上看，PCA最低OECD其次DFM最高，说明所选取的变量之间还是存在较高的共线性，这种情况下对数据进行降维能取得较好的效果；但是PCA方法依赖数据的完整性，早期数据的缺失会导致合成指数的时间跨度晚于基准指标；从领先效果来看，OECD平均值为4中位数为2.5，PCA平均值为4.7中位数为3，DFM平均值为3.6中位数为3.5；时差相关性方面OECD和PCA分别为0.86和0.92，DFM略差为0.66。整体而言通过一系列宏观变量合成得到的领先指数对PPI的预测能取得较好效果，并且相关性较高走势接近，基本保持在一个季度左右的领先期数。

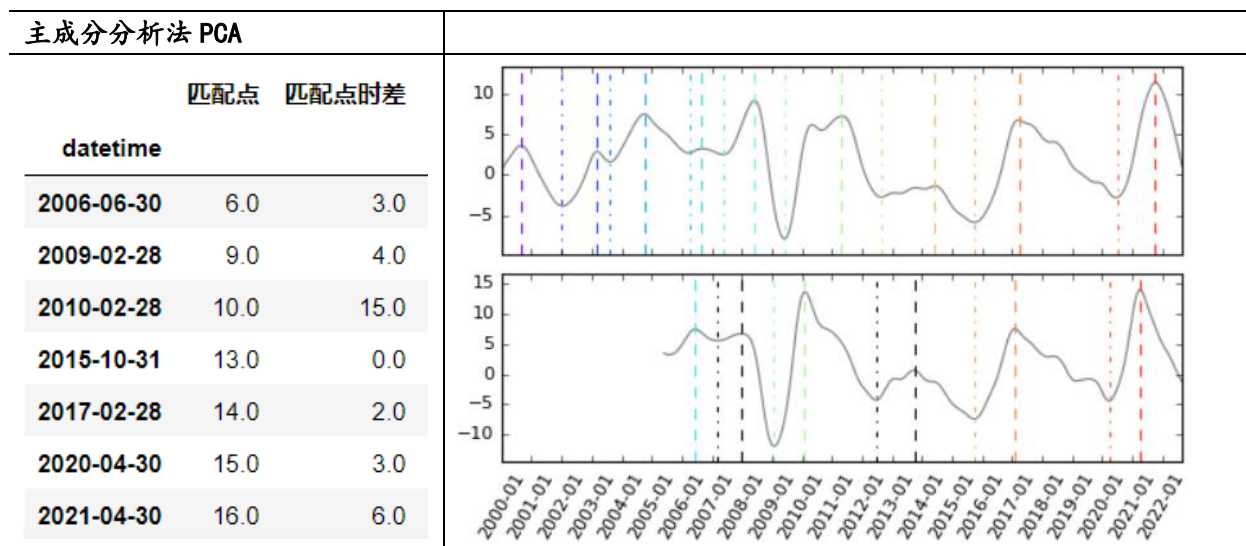
表：拐点匹配的效果评价

评价指标	CI_OECD	CI_PCA	CI_DFM
拐点总数	17	17	17
拐点匹配数	10	7	8
拐点未匹配数	7	5	9
未匹配点前期	0	5	0
未匹配比例	0.411765	0.294118	0.529412
x 冗余	7	4	9
x 冗余比例	0.411765	0.363636	0.529412
时差平均数	4	4.714286	3.625
时差中位数	2.5	3	3.5
时差标准差	4.521553	4.88925	5.289815
时差变异系数	1.130388	1.037114	1.459259
相关性最大值	0.859043	0.923903	0.656528
相关性最大阶数	3	4	2

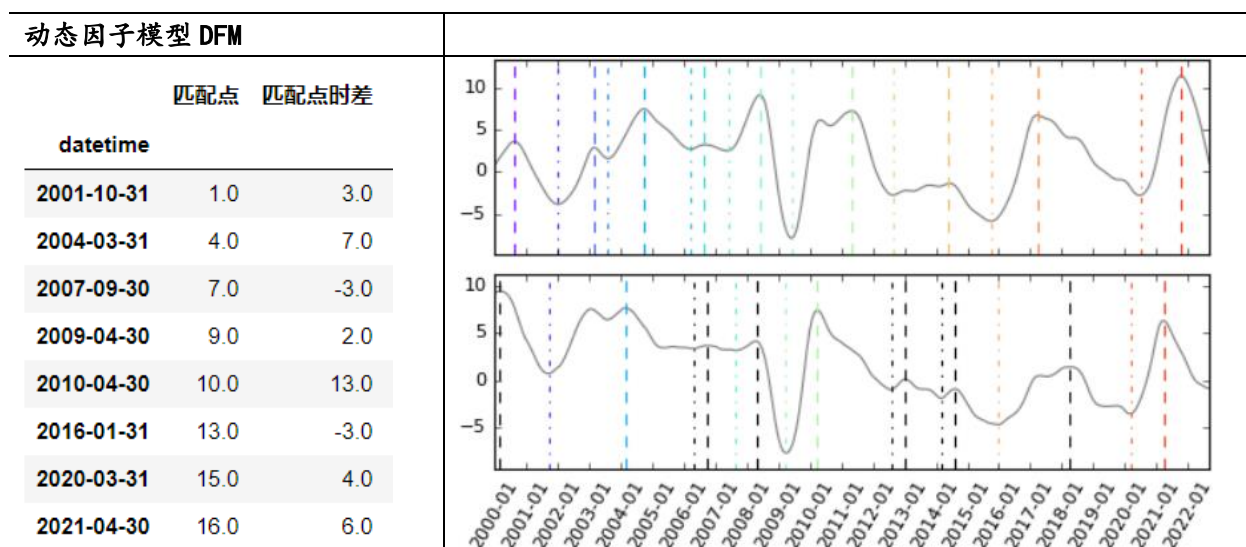
合成指数法 OECD



来源：广发期货发展研究中心



来源：广发期货发展研究中心



来源：广发期货发展研究中心

六、结论

本文基于周期拐点的视角提出了宏观量化的研究思路，主要包括以下步骤：变量选取、数据处理、合成方法、拐点识别、拐点匹配、效果评价。其中变量选取原则要考虑数据的经济逻辑性、领域重要性、发布及时性，通过整理筛选宏观研究中常为投资者所关注的指标，我们将指标归纳为价格、金融条件、工业、进出口、消费、投资、财政7个大类及其细分的28个小类，后续可以继续扩展补充。

数据处理流程在于分离时间序列的四个要素：趋势要素（T）、循环要素（C）、季节要素（S）、不规则要素（I），针对数据的不同特点设置不同的处理步骤，在较为关键的去除噪音环节，对比了不同处理方式的优势和劣势：全局平滑效果较好，但比较依赖经验参数，加入新数据时会改变之前的平滑结果；滚动平滑则不会产生结果前修，但是受近期数据的影响明显，存在极值点相比原始数据后移的情况。实际使用中可以根据数据特点进行精细化的设置。

周期拐点的识别与匹配部分介绍了B-B算法步骤和通用的匹配规则，所使用的相邻点参数、周期参数、半周期参数需要结合研究问题本身的周期大小进行设置，才能达到较好的效果。评价指标方面主要是基于匹配结果的几个指标（未匹配比例、冗余比例、时差平均值、时差标准差）以及时差相关性的最大相关系数和领先滞后阶数。

合成方法部分主要对比了三种常用方法的原理，合成指数法采用OECD的编制方式计算简便但存在共线性问题；主成分分析法PCA考虑了多变量的相关性，因而允许备选指标的经济含义高度重叠，但由于丢弃了其他主成分，存在一定程度的信息丢失，且依赖数据的完整性；DFM模型不仅允许观测变量受因子滞后项的影响，而且也允许因子本身具有独立的动态演化过程，具有双重动态特征。

最后我们通过PPI领先指数的设计对这一宏观量化流程进行实践，根据PPI的定义选取了两类宏观变量：一类是与PPI相关性较高的价格类指标；另一类为工业企业盈利、货运量、进出口等反映经济景气的指标，基准指标使用PPI全部工业品当月同比。对基准指标处理之后进行拐点识别发现：PPI周期的数值规律明显，极值点的持续时间很短，周期长度并不稳定最短为5个月最长为39个月其余大部分在15个月附近，比较符合这一数据的直观印象。

对比三种合成方法，从领先效果来看，OECD平均值为4中位数为2.5，PCA平均值为4.7中位数为3，DFM平均值为3.6中位数为3.5；时差相关性方面OECD和PCA分别为0.86和0.92，DFM较差为0.66。整体而言，使用这一宏观量化流程合成得到的领先指数对PPI的预测能取得较好效果，并且相关性较高走势接近，基本保持在一个季度左右的领先期数，对于提前预判PPI走势具有重要意义。

免责声明

本报告中的信息均来源于被广发期货有限公司认为可靠的已公开资料，但广发期货对这些信息的准确性及完整性不作任何保证。

本报告反映研究人员的不同观点、见解及分析方法，并不代表广发期货或其附属机构的立场。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

在任何情况下，报告内容仅供参考，报告中的信息或所表达的意见并不构成所述品种买卖的出价或询价，投资者据此投资，风险自担。

本报告旨在发送给广发期货特定客户及其他专业人士，版权归广发期货所有，未经广发期货书面授权，任何人不得对本报告进行任何形式的发布、复制。如引用、刊发，需注明出处为“广发期货”，且不得对本报告进行有悖原意的删节和修改。

广发期货有限公司提醒广大投资者：期市有风险 入市需谨慎！

相关信息

广发期货发展研究中心

<http://www.gfqh.cn>

电话：020-88800000

地址：广州市天河区天河北路 183-187 号大都会广场 41 楼

邮政编码：510620