

分析师：周建华
登记编码：S0730518120001
zhoujh-yjs@ccnew.com 021-50586758
研究助理：钮若洋
niury@ccnew.com 021-50586356

量化宏观基本面框架：动态因子模型

——宏观专题

证券研究报告-宏观专题

发布日期：2022年10月13日

相关报告

《宏观专题：增量政策有限，企稳势头待确认》
2022-07-27

《宏观专题：内外冲击下的有序复苏》
2022-06-27

《宏观专题：流动性影响有限，宽信用力度或增强》 2022-03-11

联系人：马巍琦

电话：021-50586973

地址：上海浦东新区世纪大道1788号16楼

邮编：200122

投资要点：

● 动态因子模型思想

近年来计量经济学领域通常采取动态因子模型（Dynamic Factor Model）解决宏观数据的追踪预测问题，并在美联储以及欧央行等世界主要地区货币当局获得广泛实际应用。这一专题中，我们使用动态因子的逻辑框架，通过纳入一系列高频宏观数据，对通胀、PMI、GDP进行追踪预测，为分析宏观经济运行状态提供指导。

动态因子模型（Dynamic Factor Model）本质上是一种状态空间模型，即包含观测方程以及状态转移方程两部分。观测方程核心思想是假设大量宏观数据由少量隐含的共同因子驱动；状态转移方程则描述了隐含共同因子所遵循的时间序列过程，即共同因子本身存在的动态演化，一般认为其服从一个向量自回归过程（VAR）。

● 动态因子模型优势

相较于传统的回归模型，动态因子模型在数据处理上具有以下显著优势：

（1）对宏观大数据进行降维处理。（2）解决数据的“混频与缺失问题”（3）反映了状态转移的时变特征（4）解决指数合成的权重问题

● 动态因子模型效果

（1）我们构建高频系统分别对CPI和PPI的环比增速进行追踪预测。从拟合效果来看，动态因子模型对于通胀追踪的整体解释度较好；其中对于环比增速的解释度强于对于同比增速的解释度，对于PPI增速的解释度强于对于CPI增速的解释度。根据截至2022年9月的月内高频数据，我们给出了CPI与PPI增速外推一期的追踪预测值。预计10月PPI同比增速1%，CPI同比增速3%。

（2）我们构建高频系统分别对PMI与GDP进行追踪预测，从拟合效果来看，PMI与GDP增速的整体拟合效果较好，都较好反映了2020年初的疫情冲击影响。样本外预测效果来看，2022年9月29日PMI预测更新值为50.14，次日公布值为50.1，预测接近程度较高。当前模型对于三季度GDP增速的预测值为4.1%。

（3）我们按照全局因子、增长因子、金融因子、通胀因子四个部分对所有观测变量进行约束；按照月度频率在不同因子项下筛选了近100个变量，均进行同比处理消除季节性。模型效果来看，我们从宏观大数据中提取得到增长、货币、通胀的隐含因子，均与历史经验较为符合，并可以进一步用于全局追踪和外推预测。

风险提示：模型效果仅供参考，经济环境的变化可能影响模型效果

内容目录

1. 量化基本面研究导论	3
2. 动态因子模型介绍	3
2.1. 模型主要思想	3
2.2. 模型的数据处理优势	3
2.3. 标准动态因子模型	4
2.4. 数据处理与因子估计	4
3. 动态因子模型实证	5
3.1. 模型逻辑	5
3.2. 数据筛选与处理	5
3.3. 模型实证一：通胀的追踪预测	5
3.4. 模型实证二：GDP 与 PMI 的追踪预测	7
3.5. 模型实证三：提取经济指数	8

图表目录

图 1：从众多宏观变量中提取得到全局通胀因子	3
图 2：通胀高频追踪预测指标筛选分类	6
图 3：CPI 环比追踪预测效果	6
图 4：PPI 环比追踪预测效果	6
图 5：CPI 同比追踪预测效果	7
图 6：PPI 同比追踪预测效果	7
图 7：经济类高频追踪预测指标筛选分类	7
图 8：PMI 追踪预测效果	8
图 9：GDP 同比追踪预测效果	8
图 10：模型提取的全局增长因子	8
图 11：模型提取的全局金融因子	8
图 12：模型提取的全局通胀因子	8
图 13：模型提取的简化增长因子	8

1. 量化基本面研究导论

近年来计量经济学领域通常采取动态因子模型（Dynamic Factor Model）解决宏观数据的追踪预测问题，并在美联储以及欧央行等世界主要地区货币当局获得广泛实际应用。这一专题中，我们使用动态因子的逻辑框架，通过纳入一系列高频宏观数据，对通胀、PMI、GDP 进行追踪预测，为分析宏观经济运行状态提供指导。

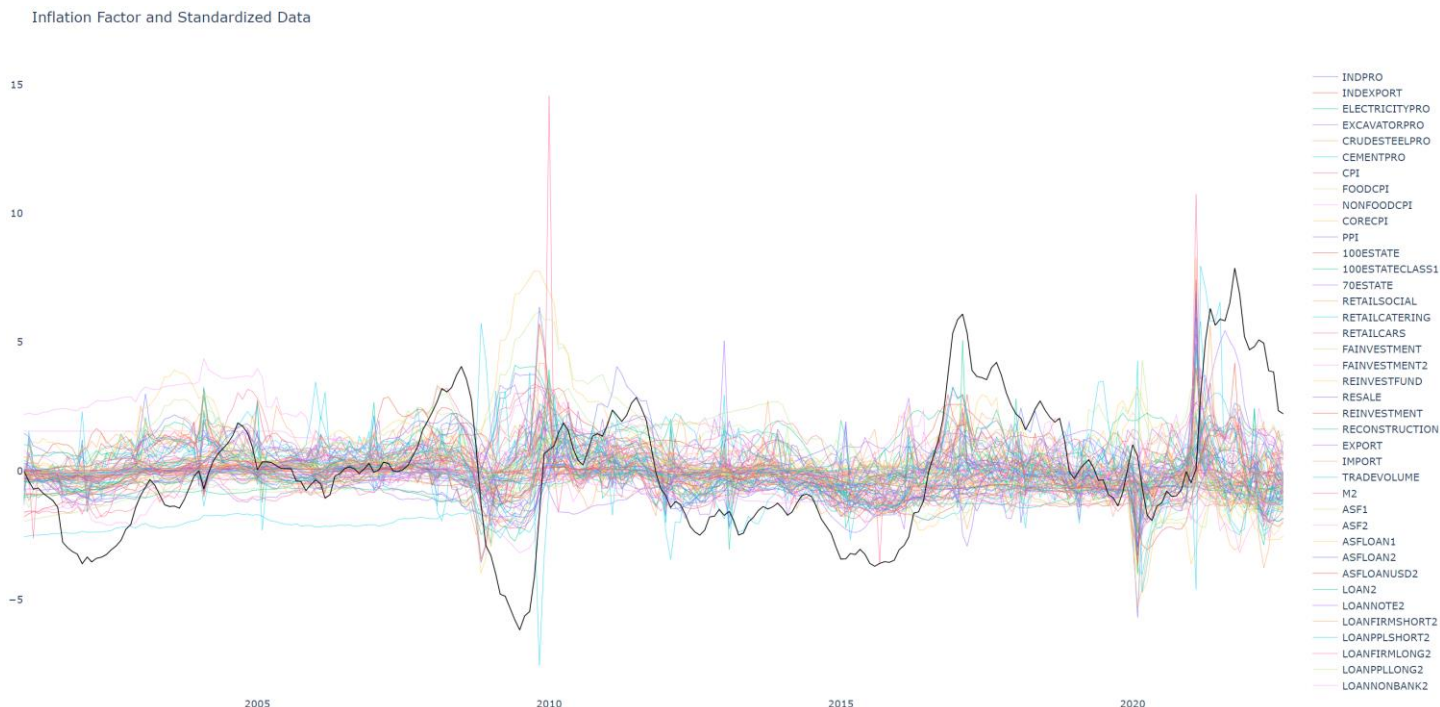
2. 动态因子模型介绍

2.1. 模型主要思想

主要宏观数据的公布频率通常较低，诸如 PMI、CPI、PPI 等数据在月度频率公布，而 GDP 等关键数据更是在季度频率公布。因此有必要通过分析高频宏观数据对当前的经济运行状态进行追踪预测（Nowcasting）。

动态因子模型（Dynamic Factor Model）本质上是一种状态空间模型，即包含观测方程以及状态转移方程两部分。观测方程核心思想是假设大量宏观数据由少量隐含的共同因子驱动；状态转移方程则描述了隐含共同因子所遵循的时间序列过程，即共同因子本身存在的动态演化，一般认为其服从一个向量自回归过程（VAR）。

图 1：从众多宏观变量中提取得到全局通胀因子



资料来源：Wind，中原证券

2.2. 模型的数据处理优势

相较于传统的回归模型，动态因子模型在数据信息的处理上具有以下显著优势：

(1) 对宏观大数据进行降维处理：动态因子模型可以同时纳入大量混频变量，通过提取隐含共同因子，实现数据的降维处理。

(2) 解决数据的“混频与缺失问题”：由于高频数据的公布频率不完全一致，且公布时点较为参差，因此具有锯齿状的数据特性，不可避免的含有缺失值。动态因子模型通过卡尔曼滤波以及 EM 算法对状态变量进行最大似然估计，从而填补缺失数据带来的影响。

(3) 反映了状态转移的时变特征：动态因子模型的状态转移方程中，状态转移矩阵是时变的，即模型有效反映了信息结构变化对隐含经济因子动态变化过程的影响。

(4) 解决指数合成的权重问题：由于系统纳入的宏观变量通常较多，采取主观方法分配权重而合成的指数可能存在较大误差，而动态因子模型采取主成分法识别各变量的因子载荷。

2.3. 标准动态因子模型

令 $x = [x_{1,t}, x_{2,t}, x_{3,t} \dots x_{n,t}]^T$ 表示 n 个标准化处理的观测变量使任意 x_t 的均值为 0 且方差为 1，设定有 r 个不可观测的隐含因子，我们认为 x_t 与 r 个隐含因子之间满足如下关系：

$$x_t = \mu + \Lambda f_t + \varepsilon_t \quad (1)$$

其中 μ 是截距， f_t 是维度为 $r \times 1$ 的隐含因子向量， Λ 是维度为 $n \times r$ 的因子载荷矩阵，它连接了隐含因子和可观测变量， $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t} \dots \varepsilon_{n,t}]^T$ 是不能被隐含因子解释的误差向量，它服从协方差矩阵为 R 的正态分布。我们进一步假设各隐含因子服从一个 VAR 过程如下：

$$f_t = \sum_{i=1}^p A_i f_{t-i} + B u_t, \quad u_t \sim i.i.d.N(0, I_q) \quad (2)$$

其中 A_i 为维度 $r \times r$ 的自回归系数矩阵； u_t 为误差向量，服从正态分布。

以上的等式 (1) 和 (2) 定义了一个标准的动态因子模型，它是一种状态空间模型等式 (1) 被称为测量方程，它连接了可观测变量和隐含变量；等式 (2) 为状态转移方程，它描述了隐含因子的动态演化过程。

2.4. 数据处理与因子估计

动态因子模型的数据处理以及应用大致包括以下三个步骤：

(1) 筛选底层经济指标纳入模型，在指标的筛选上考虑经济指标与目标变量在逻辑上的相关性，经济指标的公布频率，经济指标的可得性；然后对筛选指标进行清洗、标准化处理。

(2) 将动态因子模型写作状态空间模型，设定其状态转移方程以及观测方程。采取相应估计方法对隐含因子和参数进行估计。

(3) 根据得到的因子载荷和系数矩阵，对目标变量进行即时预测和样本外预测。

其中更关键步骤是第二步对于动态因子的估计，目前对动态因子估计的主流方法均借鉴了信号论中的时域分析。具体而言对于一个 n 维观测变量的系统，当 n 数值不大时（例如小于10），通常采取最大似然估计以及卡尔曼滤波进行参数估计；而当模型维度上升，非线性优化会极大提高模型的复杂度，因此对于高维系统，因子估计通常采取 PCA+EM 算法进行估计：

（1）首先通过主成分分析法得到隐含因子与各参数的初始值，随后通过 EM 算法以及卡尔曼滤波迭代计算隐含因子与模型参数的最大似然估计直至模型收敛，得到最优估计量；

（2）EM 算法包括两个步骤：E-step 基于初始值和上次迭代值，使用卡尔曼滤波得到隐含因子序列和模型的似然值，M-step 基于 E 步骤得到迭代值，最大似然估计得到新的参数，重复步骤直至似然估计变动小于阈值时模型收敛。

最终根据得到的因子以及因子载荷，我们可以及时追踪预测目标变量，同时由于假设因子服从一个向量自回归过程（VAR），我们可以进一步对目标变量进行样本外预测。

3. 动态因子模型实证

3.1. 模型逻辑

纽约联储曾使用动态因子模型以月频指标作为底层数据对于 GDP 进行追踪预测，随后又以周频指标作为底层数据推出更加直观的周度追踪预测 WEI 指数。模型逻辑上，前者为了控制模型因子估计的特异性误差限制了因子数量以及各主成分下可参与的观测变量；后者则更加宽松。本文对比两种方法，分别参照 WEI 指数的估计逻辑对通胀等指标在周度频率上进行实时监测，同时参照限制因子相关性的逻辑提取各项经济指数。

3.2. 数据筛选与处理

在指标选择上，考虑经济逻辑较好、公布频率较高、可得性较强、时间序列较长的指标。

数据特征上，我们筛选得到的数据面板包括月度、旬度、周度、日度等不同频次，由于数据存在一定缺失序列以及公布时点不完全一致，导致整体呈现混频锯齿状结构；同时原始价格序列为绝对值，具有一定季节性因素。因此数据处理主要包括周度变频与季节性处理两部分。

对于数据面板的混频锯齿问题，频率上我们将频率统一为周度数据（周内公布值统一至当周最后一日），缺失部分由模型的卡尔曼滤波自行过滤调整。数据的季节性因素处理方面，针对目标最终合成的目标变量，我们采取不同的处理手段：如果跟踪环比变动对于季节性我们不做处理；如果跟踪同比变动，我们则同样使用取同比的方法消除季节性，即将本周数据与上年同期（一年取 52 周）进行同比。

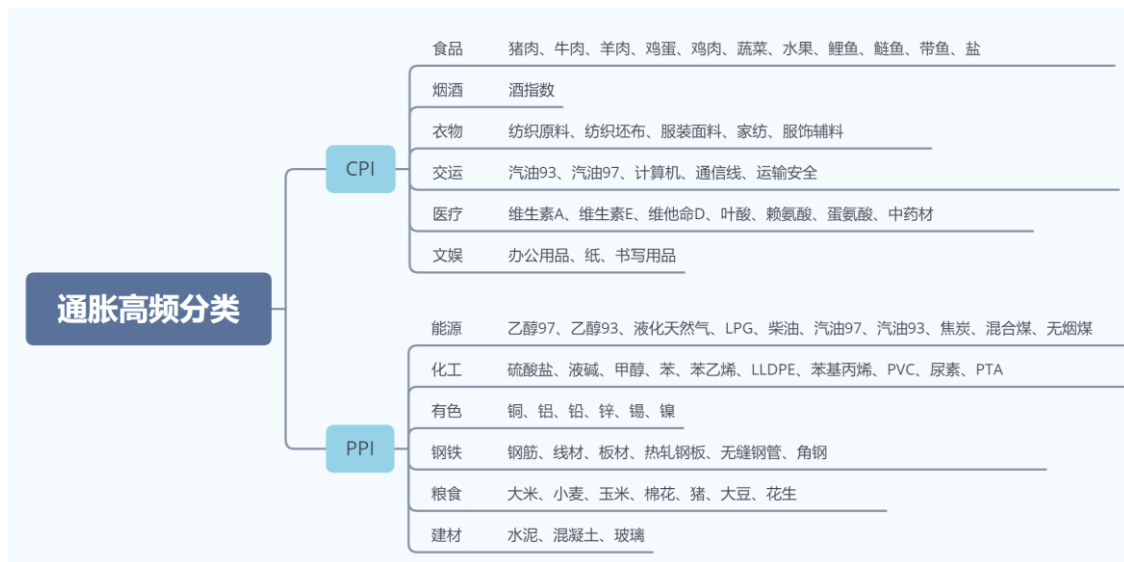
3.3. 模型实证一：通胀的追踪预测

1、底层数据筛选与处理

我们按照居民消费价格指数目录和工业生产者出厂价格目录，筛选出 CPI 与 PPI 的主要大

类科目，并筛选相应大类科目下的主要高频指标；具体筛选如下：CPI 分为食品、烟酒、衣着、交运、医疗、文娱六类；PPI 分为能源、化工、有色、钢铁、粮食、建材六类；各子类下筛选出相关高频指标，利用动态因子模型计算各项因子载荷，而无需通过主观手段赋权合成通胀指数。这里我们根据上述分类，对 CPI 和 PPI 项下共筛选出 70 个高频指标，时间区间上根据各类数据的可得性选取了 2015 年至今的数据，对 CPI 和 PPI 进行追踪拟合。

图 2：通胀高频追踪预测指标筛选分类



资料来源：中原证券

2、模型效果

我们根据构建的高频系统分别对 CPI 和 PPI 的环比与同比增速进行追踪预测，并绘制了当期追踪预测值与实际公布值的对比图。从拟合效果来看，动态因子模型对于通胀追踪的整体解释度较好；其中对于环比增速的解释度强于对于同比增速的解释度，对于 PPI 增速的解释度强于对于 CPI 增速的解释度。根据截至 2022 年 9 月的月内高频数据，我们给出了 CPI 与 PPI 增速外推一期的追踪预测值。预计 10 月 PPI 同比增速 1%，CPI 同比增速 3%。

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_47273

