**基于宏观大数据的GDP 即时预测**

易艳萍、黄德金、王熙

**目录**

[附录I 卡尔曼滤波与卡尔曼平滑 1](#_Toc164774388)

[附录II 估计混频动态多因子模型 4](#_Toc164774389)

[附录III 信息冲击的分解 13](#_Toc164774390)

[附录IV 稳健性检验 15](#_Toc164774391)

# 附录I 卡尔曼滤波与卡尔曼平滑

Ⅰ.1卡尔曼滤波

我们将因子记为, 观测值记为，假设两者符合如下动态因子模型:

|  |  |
| --- | --- |
| , | (I1) |
| , | (I2) |

其中与相互独立。本文假设为对角矩阵，可以为非对角矩阵。记信息集, 以及。当动态因子模型参数已知，我们有：

|  |  |
| --- | --- |
|  | (I3) |
|  | (I4) |
|  | (I5) |
| *.* | (I6) |

因此，给定初值以及，我们可以使用式(I3)-(I4)计算以及，进而通过式(I5)-(I6)获得以及。类似的，对于任意，我们可以获得，，以及。

Ⅰ.2卡尔曼平滑

当我们观测到所有数据，我们可以使用信息集对隐因子进行推断，即我们需要估计。我们可以得到[[1]](#footnote-0)：

|  |  |
| --- | --- |
|  | (I7) |
| . | (I8) |

因此，给定从卡尔曼滤波中获得的以及（上节中已给出计算方法），我们可以通过式(I7)-(I8)迭代计算。

考虑。不失一般性令，由de Jong and Mackinnon (1988)可知：

|  |  |
| --- | --- |
| . | (I9) |

因此，对于任意，给定初值，我们可以通过式（I9）迭代计算出。如果，则通过计算。

给定信息集，我们还可以计算：

|  |  |
| --- | --- |
|  |  |
| 在正态分布假设下此等式成立 |  |
|  | (I10) |

特别的，本文需要计算和。

Ⅰ.3 有缺失观测值的卡尔曼滤波与卡尔曼平滑

数据生成过程仍为式(I1)-(I2)。由于中部分指标可能缺失, 式(I1)不再是观测方程。我们引入维对角矩阵作为掩码矩阵，其中当可观测时，的第个对角元素等于1; 当缺失时, 的第个对角元素等于0。我们将分解为可观测部分和不可观测部分：，其中为时间t时能观测到的数据, 为单位矩阵。我们有：

|  |  |
| --- | --- |
|  | (I11) |
| . |  |

当有缺失观测值的情况下，观测方程为式(I11), 而状态方程保持不变。

当缺失时, 向量、 和 的第个元素为0，的第行为0向量，因此我们可以抽取出式(I11)中非零元素对应的行，形成简化后等价的观测方程：

|  |  |
| --- | --- |
| *.* | (I12) |

根据卡尔曼滤波公式(I3)-(I6)，在有缺失观测值的情况下，卡尔曼滤波的计算公式需做如下调整[[2]](#footnote-1)：

|  |  |
| --- | --- |
|  | (I13) |
|  | (I14) |
|  | (I15) |
| . | (I16) |

简单的说，我们在计算卡尔曼滤波时，只需将中不可观测的元素（以及对应的系数行）去掉，而具体的迭代算法则与第(一)节类似。对于卡尔曼平滑的迭代算法, 由于不直接涉及到，仍可使用式(I7)-(I10)来计算。不失一般性，本文假设。

# 附录II 估计混频动态多因子模型

记, , 。代表在时间T能观察到的所有数据。分别代表联合概率密度函数，边际概率密度函数以及条件概率密度函数。

Ⅱ.1 EM算法的原理

因为隐因子不可观测，我们需要选取使得观测数据的对数似然函数最大。由于不可观测的隐因子的维度较高，导致最大似然估计方法难以实施，因此我们采用EM 算法，以下将简述EM 算法的基本原理。因为，我们有:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  | (II1) |

EM 算法通过不断改进式(II1) 右侧取值而改进, 直至收敛。换言之, 我们会通过选取最优的使得式(II1) 右侧取值最大:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | (II2) |

式(II2) 中最后一个等式是因为。此处，是基于参数估计值和信息集所计算的条件期望值。EM 算法的每一次迭代优化都分两步：在第一步中(期望(E) 步骤)，已知上一次迭代中的估计值，通过卡尔曼平滑(参考上文) 计算；在第二步(最大化(M) 步骤)，通过极大化得到此次迭代的估计值。EM 算法不断重复以上两个步骤，直至参数的估计值收敛。

Ⅱ.2 EM算法估计动态多因子模型

上文动态因子模型(I1)-(I2)关于和的联合对数似然函数如下：

|  |
| --- |
| (II3) |

1. EM 算法估计完整数据的动态多因子模型

此小节中 均可直接观测, 因此。EM 算法通过极大化可得到第次迭代的估计值。

首先，对于系数矩阵求导我们有一阶条件

|  |
| --- |
|  |
|  |
|  |

因此，我们有

|  |  |
| --- | --- |
|  | (II4) |

类似的, 我们可以分别通过一阶条件、以及得到第次迭代中关于、的估计：

|  |  |
| --- | --- |
|  | (II5) |
|  | (II6) |
|  | (II7) |

式(II4)-(II7)的证明也可参考Banbura and Modugno (2014) 等。由和的表达式可知，它们的每一行均可通过单方程的多元回归得到。

2.EM 算法估计具有缺失观测值的动态多因子模型

此小节中[[3]](#footnote-3)，我们考虑中部分指标可能缺失的情况, 因此。由于隐因子和部分不可观测，我们需要选取 使得观测数据的对数似然函数 最大。用类似于第（一）节的推导，我们可以证明EM 算法可通过极大化得到第次迭代的估计值。由 和的联合对数似然函数式(II3)可知，和的表达式仍由式(II4)和式(II6)给出，而 和的表达式则需做调整。

我们将分解为为可观测部分和不可观测部分：

|  |  |
| --- | --- |
|  | (II8) |

此处为时间t时能观测到的数据，为单位矩阵。因此我们有

|  |  |
| --- | --- |
|  | |
|  | (II9) |

由式(II8)的分解，我们有:

|  |  |
| --- | --- |
|  | |
|  | (II10) |

由模型(I1)-(I2)的假设，可知，因此，

|  |
| --- |
|  |
|  |

由期望的迭代法则可知

|  |  |
| --- | --- |
|  | (II11) |

同理可证。同样由模型(I1)-(I2)的假设，我们有。由期望的迭代法则可知，

|  |  |
| --- | --- |
|  | (II12) |

因此由式(II9), (II10), (II11), (II12), 我们有:

|  |
| --- |
|  |
| (II13) |

对于系数矩阵求导我们有一阶条件:

|  |  |
| --- | --- |
|  | |
|  |  |

由式(II13)和(II14)以及和 ，我们可以得到：

|  |
| --- |
|  |

因而我们有

|  |  |
| --- | --- |
| (II15) |  |

此处代表克罗内克积(Kronecker product)。类似于上小节，由的表达式可知的每一行均可通过单方程的多元回归得到，这一观察启发我们后续使用Lasso等压缩算法估计GDP增速的因子载荷向量。同样的，我们可以通过一阶条件得到关于的估计：

|  |  |
| --- | --- |
|  | |
|  |  |

Ⅱ.3 在以上EM 算法分析框架内进行GDP 即时预测

此节将推导正文的式(6) - (9)。采用Mariano and Murasawa (2003) 的方法[[4]](#footnote-4)，我们将季度GDP 建模为三个（不可观测）月频指标的几何平均，。季度GDP同比增长率满足以下关系式：

|  |  |
| --- | --- |
|  | |
|  | (II17) |

此处 为不可观测的月频指标。事实上, 我们可以将所有季度指标罗列为，均满足式(II17)（同正文式（1））。根据正文式(1)-(5)，我们有:

|  |  |
| --- | --- |
|  | |
|  | |
|  | |
|  | (II18) |

定义

|  |  |
| --- | --- |
|  | (II19) |
|  | (II20) |

式(II19)和(II20)中的定义同正文式(8)和(9), 从而式(II18)可写成：

|  |  |
| --- | --- |
|  | (II21) |

这里我们引入主要是为了能使用前文的EM算法分析框架来估计混频动态多因子模型(参考Banbura and Modugno (2014))，因此在实操中我们会假设观测方程,其中既不存在横截面相关性，也不存在时序相关性，且每个分量的方差为。

由正文关于隐因子和特质因子的假设（正文式(3)-(5)), 拓展后的状态变量服从以下状态动态方程：

|  |  |
| --- | --- |
|  | |
|  |  |

其中（见正文式(5)），

|  |
| --- |
|  |
|  |

与Banbura and Modugno (2014)类似，我们也考虑了特质因子、可能存在的时序相关性。

拓展后的模型(II21)和(II22)可以用此部分第（二）节中第2小节的EM算法框架来估计, 但需调整状态变量和待估计参数的定义：, ,。由于拓展后模型对 和进行了参数约束，我们在估计的时候需要考虑到这些约束，以下将简述如何在EM算法框架中做约束回归。

Ⅱ.4 线性约束

由于本文对因子载荷矩阵进行了参数约束，此节将考虑在EM算法框架中如何计算的受约束估计量。为了更清楚阐明受约束估计量的性质，我们在附录中采用了与正文等价但形式不同的线性约束方式：，其中为 维矩阵, 是的维度，为维列向量，为维列向量。我们将目标函数写作拉格朗日形式：

|  |
| --- |
|  |

由式（II13）以及，vec 和 的性质，我们有:

对于 求导我们有一阶条件

|  |  |
| --- | --- |
|  | |
|  | (II23) |

式(II23)中未知的 ，我们用第 次迭代的估计值 来替代，记

求解式（II23），可得 的受约束估计量：

|  |  |
| --- | --- |
|  | |
|  | (II24) |

其中 的表达式与式(II15)是等价的，只是采用了不同的矩阵表达方式。从 的表达式可以看出，的每一行均是通过单方程的多元回归得到。因为 满足线性约束，即，所以 。因此 的受约束估计量可表示为：

|  |
| --- |
| . (II25) |

我们可以类似的考虑对 的约束回归，在此不再赘述。对于某些具体的约束形式，式(II25)可进一步简化。  
 在本文所研究的模型(II21)中，我们分别对每个观测值时间序列的载荷向量 的第 行)施加了约束（见式(II20)）。由式(II20)可知，

其中 是 维矩阵， 是 维列向量， 是对载荷向量 的线性约束 并且 。记 ，由于 是对角矩阵，式(II25)中 的受约束估计量可表示为：

|  |  |
| --- | --- |
|  | (II26) |

其中

|  |
| --- |
| (II27) |

是对 的无约束估计量，与式(II15)和(II24)中无约束估计量的表达式是等价的。由式(II26)可知， 受约束估计量的每一行均可通过单方程的（受约束）多元回归得到。其直觉是，由于 是对角矩阵, 个观测方程不相关，所以可分别进行单方程回归。  
 式(II26)也可以通过将单方程受约束回归改写成无约束回归得到。

Ⅱ.5 拓展状态变量后的压缩估计GDP增速因子载荷向量

我们将GDP增速作为观测向量的最后一个元素记为，其因子载荷向量为的最后一行，我们在此节将详述 的估计。由式(II20)可知，

|  |  |
| --- | --- |
|  | (II28) |

其中 是正文式（2）中 矩阵的最后一行，是需要估计的。（即）可以用上文中的式(II26)估计，而另外一个等价做法是将方程的受约束回归改写成无约束回归。模型（II21）中GDP的观测方程为：

|  |  |
| --- | --- |
|  | (II29) |

式（II29）可等价地表示为一个无约束方程：

|  |  |
| --- | --- |
|  | (II30) |

式(II30)中的 可以用类似于此部分第（二）节中第2小节的方法进行估计（见式(II15)和(II27)）：

|  |  |
| --- | --- |
|  | |
|  | (II31) |

其中 可以直接观测 。式(II31)中的条件期望均可通过卡尔曼平滑来计算。式（II31）对 的估计，也等价于如下 OLS 回归所得到的参数估计：

|  |
| --- |
|  |
|  |
| (II32) |

对式(II32)的回归，我们采用Lasso的方法对因子变量进行选择，并且在完成因子筛选后进行OLS回归估计因子载荷。

类似的，其他观测方程中的因子载荷矩阵和转移方程中的系数矩阵均可通过OLS post-Lasso方法估计。但是由于本文仅聚焦于GDP预测，我们采用了最为保守的因子筛选方式：仅对GDP观测方程使用变量选择。

# 附录III 信息冲击的分解

随着新的数据不断在不同时间点公布，即时预测模型会基于当前最新公布的信息更新对隐因子的推断(如式(I5))，然后进一步对GDP以及其他数据进行即时预测[[5]](#footnote-5)。假设为上一个数据发布日的数据集，在新数据发布日，我们获得了新公布数据，即，。

我们考虑在时间点和之间的GDP即时预测变动，此处是尚未公布的季度GDP同比增速。在本文的模型设定下(正文式(6)和(7))，此处的预测是线性投影 (linear projection)。换言之，如果我们用表示线性投影，则。

由线性投影的性质（参考Hamilton (1994) 式 [4.5.30]），我们有：

|  |  |
| --- | --- |
|  | |
|  | |
| , | (III1) |

此处， 。是中包含的新信息。我们将基于式(III1)进行信息增益分解，具体而言，我们需要计算和 。

文中 的生成机制如下（同式(II21)和正文式(6)）： 的第行记为 , 为的因子载荷向量 。我们有：

由此可得， 为 行向量, 其中第 个元素为：

|  |  |
| --- | --- |
|  | (III2) |

而 为 矩阵，其 行, 列的元素为：

|  |
| --- |
| (III3) |

此处 是 的第 个对角线元素。式(III2)和(III3)中的期望值，均可通过卡尔曼平滑式（I9）来计算。因此，我们可以计算得到 列向量：

由式(III1)可知:

|  |  |
| --- | --- |
|  | (III4) |

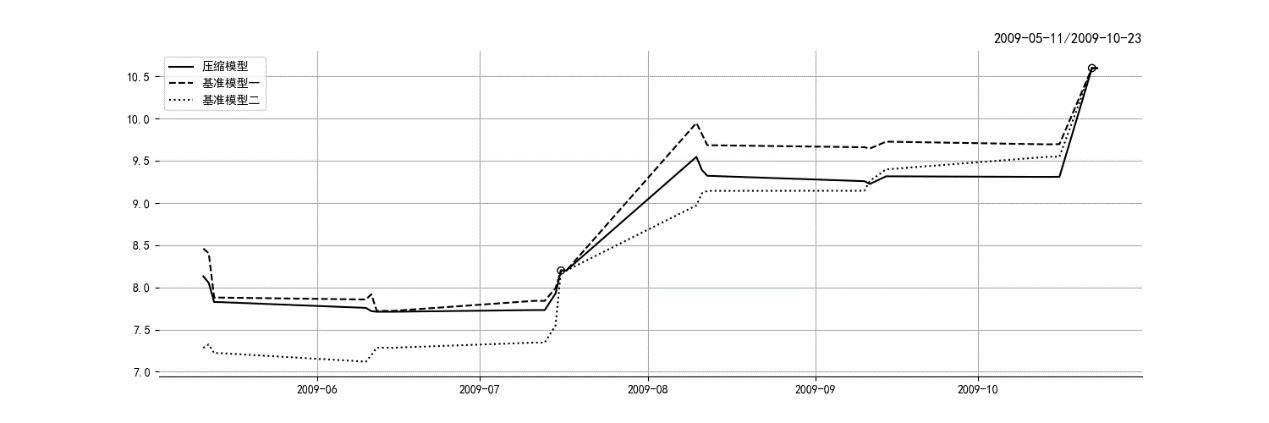
从式(III4)可见，各公布数据对于即时GDP预测值的修正的不同贡献。

# 附录IV 稳健性检验

Ⅳ.1 2009年二季度-2009年三季度的时序预测

在正文中，我们汇报了5因子压缩模型、基准模型一、基准模型二在经济下滑期的时序预测。这里，我们在图IV1中汇报2009年二季度至2009年三季度的GDP即时预测图，这段时间代表典型的宏观经济上行情形。与经济下行时正文图2展示的结果类似，随着数据的不断更新，三个模型的GDP即时预测值都在不断逼近真实的GDP增速。压缩模型与基准模型一、基准模型二表现相近。此外，较2009年第三季度的GDP数据发布日，我们的模型可以提前72日预判2009年三季度经济增速上涨。

**图IV1 5因子模型的GDP即时预测时序图（2009年二季度-2009年三季度，圆圈代表真实**值）



Ⅳ.2 样本外预测评价：固定窗口训练模型

为了验证压缩模型的稳健性，我们进一步使用了固定估计窗口的估计方式，将训练窗口固定为10年，并使用与预测期最近的两个季度作为Lasso超参数调试的验证集。以最近的2个季度或者3个季度作为最优因子（个数）模型的验证集。我们将各类即时最优模型的样本外MSFE表现汇报在表IV1之中。与正文类似，不论是整体预测结果还是新冠疫情前的预测结果，即时最优压缩模型均优于即时最优的基准模型一和基准模型二。但是在新冠疫情后，我们可以发现基准模型一具有更好的预测效力。我们认为这是由于新冠疫情后我国宏观经济受到较大冲击，而固定窗口使用了更少且更近的训练数据，更加贴近结构变迁后的经济结构，如我们可以发现在新冠疫情期间，各类模型的固定窗口MSFE误差低于延展窗口MSFE误差。此外，在新冠疫情期间，我国不同类别的宏观经济数据可能对于GDP预测的有效性并不稳健，亦或对于GDP或多或少的均存在预测性，因此，基于稀疏模型预测效力可能并不如基准模型一。但在经济结构较为稳定的情况下，固定窗口训练却损失了部分数据，如我们发现在新冠疫情前，各类模型的固定窗口MSFE误差大于延展窗口MSFE误差。随着我国进一步优化新冠肺炎疫情防控措施，精准做好疫情防控工作，我国宏观经济逐步走向正轨，随着我们不断积累后疫情时代的宏观经济数据，基于Lasso方法的混频因子模型也会提供更好的GDP即时预测表现。

我们以MAFE指标[[6]](#footnote-8)说明这一预测效果最终转化为预测GDP增速的误差，以2季度长度的验证集为例，在整体预测效果上压缩模型、基准模型二、基准模型一和单因子模型的预测绝对误差分别为0.980、1.055、1.032和1.059个百分点。在疫情前预测效果上压缩模型、基准模型二、基准模型一和单因子模型的预测绝对误差分别为0.672、0.669、0.754和0.687个百分点。在疫情后预测效果上压缩模型、基准模型二、基准模型一和单因子模型的预测绝对误差分别为3.037、3.622、2.886和3.544个百分点。可见MAFE指标与MSFE指标有着些许不同，这一点与我们的模型目标函数设定相关：EM算法中存在MLE步骤，以高斯分布为基础分布的MLE与最小预测误差的平方类似，即基于EM算法的估计值理论上可以达到（样本内）最小MSFE，但这一估计值并不一定能达到最小MAFE。因此，既然基于MAFE误差表现选择了基于Lasso方法的混频因子模型，也从一定程度上让我们对基于Lasso方法的混频因子模型更加乐观，但我们更偏好使用样本外MSFE评价模型的稳健性。

表IV1 GDP即时预测的样本外MSFE表现（即时最优模型、固定窗口）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 最佳模型 | 模型类型 | 2季度验证集 | | 3季度验证集 | |
| 2005-2022 | 整体样本 | 因子压缩模型 | 压缩模型 | 3.322 | | 3.311 | |
| 基准模型一 | 3.555 | | 3.545 | |
| 基准模型二 | 3.779 | | 3.827 | |
|  |  |  | MIDAS | 18.196 | | | |
| 2005-2019 | 新冠疫情前 | 因子压缩模型 | 压缩模型 | 1.005 | | 0.992 | |
| 基准模型一 |  | 1.373 |  | 1.406 |
| 基准模型二 |  | 1.067 |  | 1.406 |
|  |  |  | MIDAS | 25.231 | | | |
| 2020-2022 | 新冠疫情后 | 因子压缩模型 | 压缩模型 |  | 18.768 |  | 18.768 |
| 基准模型一 |  | 18.104 |  | 17.804 |
| 基准模型二 |  | 21.859 |  | 22.338 |
|  |  |  | MIDAS | 17.141 | | | |

我们将固定因子个数的两个基准模型和压缩模型对于GDP即时预测的样本外表现汇报在表IV2中。由表IV2可以看出，不论是整体样本还是新冠疫情前后的两个子样本，压缩模型一直是最佳的GDP即时预测模型。以整体样本预测为例，GDP即时预测效果最好的五个模型依次是：3因子压缩模型、3因子基准模型一、4因子基准模型一、4因子压缩模型以及6因子基准模型一。而次优的四个模型相较于3因子压缩模型，预测误差上升了1.80%、6.81%、9.08%以及11.85%。

其次，在新冠疫情前区间内，最佳的五个GDP即时预测模型为3因子压缩模型、单因子模型、3因子基准模型一、3因子基准模型二以及4因子基准模型一。可以看出整体而言，新冠疫情前，较小因子数目的GDP即时预测模型表现更加优秀。相较于3因子压缩模型，其他四个次优模型，其预测误差分别上升了1.38%、6.56%、9.42%以及11.11%。此外，从预测误差上看，使用固定窗口的预测误差整体而言大于使用延展窗口的预测误差，表明在没有突发疫情的情况下，使用更长时间的估计数据可以提供更稳定准确的预测结果。

最后，在新冠疫情后区间内，最佳的五个GDP即时预测模型为3因子压缩模型、3因子基准模型一、6因子基准模型一、4因子基准模型一以及6因子压缩模型。由于数据原因，3因子压缩模型并未对模型进行压缩，因此，3因子压缩模型与3因子基准模型一表现相同。相对于3因子压缩模型，其他三个次优模型，其预测误差分别上升了3.26%、5.08%以及5.91%，且显著优于单因子模型。从预测误差上看，使用固定窗口的预测误差整体而言小于使用延展窗口的预测误差，表明在疫情期间，宏观经济的内生结构的确发生了改变，需要使用近期的数据以提供更稳定准确的预测结果。由此可见，压缩模型的样本外预测优势十分稳健。

表IV2 各类GDP即时预测的样本外MSFE表现（固定窗口）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 最佳模型 | 模型类型 | 3 因子 | 4 因子 | 5 因子 | 6 因子 |
| 2005-2022 | 整体样本 | 3因子压缩模型 | 压缩模型 | 2.995 | 3.267 | 3.977 | 3.747 |
| 基准模型一 | 3.049 | 3.199 | 3.718 | 3.35 |
| 基准模型二 | 3.545 | 3.459 | 3.586 | 3.814 |
|  |  |  | 单因子模型 | 3.798 | | | |
|  |  |  |  |  |  |  |  |
| 2005-2019 | 新冠疫情前 | 3因子压缩模型 | 压缩模型 | 0.945 | 1.034 | 1.504 | 1.662 |
| 基准模型一 | 1.007 | 1.053 | 1.206 | 1.272 |
| 基准模型二 | 1.050 | 1.066 | 1.131 | 1.135 |
|  |  |  | 单因子模型 | 0.958 | | | |
|  |  |  |  |  |  |  |  |
| 2020-2022 | 新冠疫情后 | 3因子压缩模型/基准模型一 | 压缩模型 | 16.659 | 18.159 | 20.465 | 17.644 |
| 基准模型一 | 16.659 | 17.506 | 20.466 | 17.202 |
| 基准模型二 | 20.178 | 19.408 | 19.958 | 21.675 |
|  |  |  | 单因子模型 | 22.729 | | | |

Ⅳ.3 基于固定窗口因子模型的信息分解

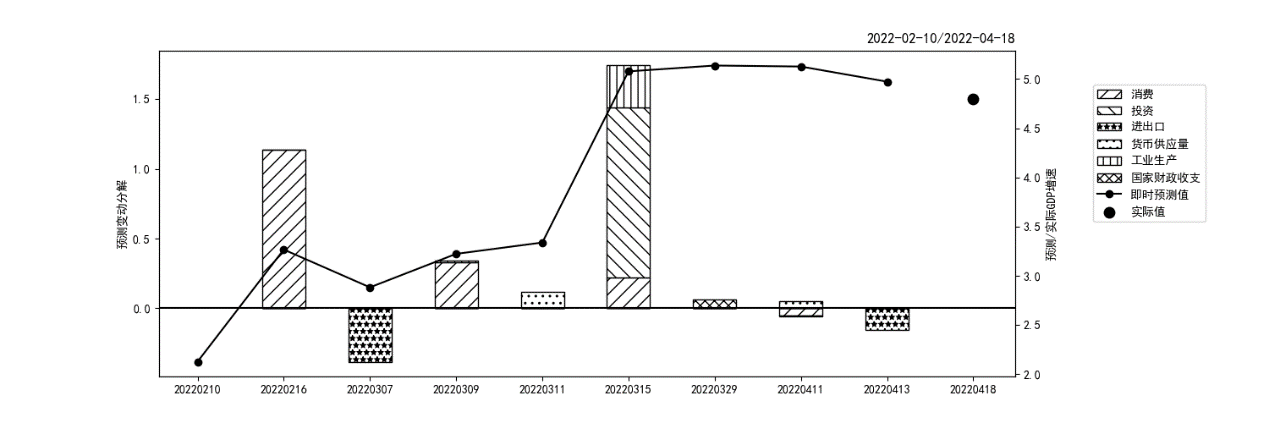
在表IV3中, 我们基于3因子压缩模型（固定窗口），汇报了2005年第一季度至2022第一季度GDP即时预测的变动分解。我们从各个宏观变量引起GDP即时预测的平均变动（以变动绝对值的平均值衡量）和总变动（以变动平方的总计衡量）两个角度来分析我国不同宏观经济变量可以为GDP即时预测所带来的信息增益，下文中我们简称这两者为平均信息增益和总信息增益。表IV3汇报了3因子压缩模型分析框架下的GDP即时预测变动分解。我们可以看到在平均信息增益方面，能给GDP即时预测带来最多信息增益的五个宏观变量分别为发电量同比增长（ELEC）、工业增加值同比增长（IP）、工业生产者出厂价格指数（PPI）、社会消费品零售总额同比增长（TRADE）以及国家财政支出同比增长（FEXP）。从总信息增益方面，能给 GDP 即时预测带来最多信息增益的五个宏观变量分别为社会消费品零售总额同比增长（TRADE）、工业增加值同比增长（IP）、发电量同比增长（ELEC）、出口总值同比增长（EXPORT）以及工业生产者出厂价格指数（PPI）。我们可以发现，在两种衡量信息增益的测算中，带来信息增益最多的前五个宏观变量中有四个是一致的。

表IV3 基于3因子压缩模型的GDP即时预测变动分解（2005年1季度-2022年1季度）

|  |  |  |
| --- | --- | --- |
| 指标名称 | 平均信息增益 | 总信息增益 |
| 发电量同比增长（ELEC） | 0.0428 | 0.9758 |
| 工业增加值同比增长（IP） | 0.0313 | 1.0118 |
| 工业生产者出厂价格指数（PPI） | 0.0309 | 0.7788 |
| 社会消费品零售总额同比增长（TRADE） | 0.0308 | 1.7496 |
| 国家财政支出同比增长（FEXP） | 0.0257 | 0.1579 |
| 进口总值同比增长（IMPORT） | 0.0240 | 0.3876 |
| 出口总值同比增长（EXPORT） | 0.0200 | 0.9191 |
| 居民消费价格指数（CPI） | 0.0171 | 0.1243 |
| 房地产投资同比增长（REINV） | 0.0140 | 0.1158 |
| 固定资产投资额累计增长（INV） | 0.0127 | 0.2262 |
| M0 供应量同比增速（M0） | 0.0127 | 0.1377 |
| M1 供应量同比增速（M1） | 0.0104 | 0.1208 |
| M2 供应量同比增速（M2） | 0.0099 | 0.0969 |
| 国家财政收入同比增长（FREV） | 0.0078 | 0.0366 |
| 商品房销售额同比增长（CMS） | 0.0055 | 0.0200 |

类似的，我们在图IV2中汇报了基于3因子压缩模型（固定窗口）对2022年第一季度GDP即时预测做的贡献分解。2022年2月10日，我们对2022年第一季度GDP增速的即时预测值为2.13%。2月16日CPI、PPI数据公布，为GDP即时预测带来了积极影响，使得预测值上升1.14%。之后3月7日进出口数据公布使得预测值略微下降0.38%，3月9日CPI、PPI 数据公布使得预测值略微上升 0.34%，3 月 11 日货币供应量数据公布使得预测值略微上升0.12%。3月15日，消费、投资、工业生产相关数据公布，均给GDP即时预测带来了明显的正向冲击，使得预测估计值大幅上升，其中固定资产投资额累计增长（INV）数据更新使GDP增长预测值提高了0.85个百分点，房地产投资同比增长（REINV）贡献了0.37个百分点，工业增加值同比增长（IP）贡献了0.28个百分点，总共使得预测值提高了1.74个百分点，上升至5.08%。在这之后预测估计值基本稳定在5%上下，直到4月18日官方发布第一季度 GDP 同比增速为 4.8%。我们可以发现，3因子压缩模型在3月15日已经有效预测到我国2022年第一季度的经济增速上升的情况。

图IV2 基于3因子压缩模型（固定窗口）2022年一季度GDP即时预测的贡献分解



Ⅳ.4 宏观变量分类情况

我们将宏观变量分类情况汇报在表IV4中。

表IV4 宏观变量分组情况

|  |  |
| --- | --- |
| 组别名称 | 指标名称 |
| 消费 | 居民消费价格指数、社会消费品零售总额、商品房销售额同比增长 |
| 投资 | 房地产投资、固定资产投资额累计同比增长 |
| 进出口 | 进口总值、出口总值同比增长 |
| 货币供应量 | M0 供应量、M1 供应量、M2 供应量同比增速 |
| 工业生产 | 工业生产者出厂价格指数、工业增加值、发电量同比增长 |
| 国家财政收支 | 国家财政收入、国家财政支出同比增长 |

参考文献

1. Banbura, M., and M. Modugno, “Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data”, *Journal of Applied Econometrics*, 2014, 29(1), 133–160.
2. de Jong, P., and M. J. Mackinnon, “Covariances for Smoothed Estimates in State Space Models”, *Biometrika*, 1988, 75(3), 601–602.
3. Durbin, J., and S. Koopman, *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press, 2012.
4. Hamilton, J. D., *Time Series Analysis.* Princeton: Princeton University Press, 1994.
5. Mariano, R. S., and Y. Murasawa, “A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series”, *Journal of applied Econometrics*, 2003, 18(4), 427-443.

**注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处**。

1. 式(I7)-(I8)的证明可以参考Hamilton (1994), Durbin and Koopman (2012)。 [↑](#footnote-ref-0)
2. 公式(I13)-(I16)的证明也可以参考Hamilton (1994) 第13.8 节。 [↑](#footnote-ref-1)
3. 本小节内容参考了Banbura and Modugno (2014)，但做了适当调整。 [↑](#footnote-ref-3)
4. 由于本文使用同比增长率, 技术细节与Mariano and Murasawa (2003) 有所不同。 [↑](#footnote-ref-4)
5. 为了读者阅读方便，此部分的内容引自Banbura and Modugno (2014)，并根据我们处理的数据和模型进行了适当调整。 [↑](#footnote-ref-5)
6. 由于篇幅限制，我们未将GDP即时预测的样本外MAFE表现单独制表，仅对实证结果进行了梳理。 [↑](#footnote-ref-8)