

分析师:

于明明

yumingming@xyzq.com.cn

S0190514100003

系统化资产配置系列之六：实时预测中国 GDP 增速

2020 年 04 月 24 日

报告关键点

实时预测 (Nowcasting) 是识别当前经济状态的流行方法。它的思想就是利用所有可得的信息来获得当前经济状态的估计。但是,对宏观经济数据建模面临多个难点,包括数据公布时间不同步、有缺失值、截面维度高和数据频率不一致等。为了解决这些问题,本文将介绍一种基于动态因子模型的 Nowcasting 方法。

相关报告

《系统化资产配置系列之五:基于择时的目标风险和风险预算配置模型》2019-12-27

《系统化资产配置系列之四:基于长期、中期、短期择时模型相结合的 A 股市场择时研究》2019-11-03

《系统化资产配置系列之三:基于 AdaBoost 机器学习算法的市场短期择时策略》2019-10-17

《系统化资产配置系列之二:行业的重新分类以及行业轮动策略》2019-09-19

团队成员:

投资要点

- 实时预测 (Nowcasting) 是识别当前经济状态的流行方法。它的思想就是利用所有可得的信息来获得当前经济状态的估计。举例来说, GDP 增速是反映经济状态的最直接指标,但由于它只在每季度发布一次,且发布时间滞后,因此我们总是无法得到当前季度的 GDP 增速。这样一来,我们就必须使用其他更高频指标如工业增加值增速等来推测当前季度的 GDP 增速。
- 宏观经济数据建模面临多个难点,包括数据公布时间不同步、有缺失值、截面维度高和数据频率不一致等。为了解决这些问题,本文介绍了一种基于动态因子模型的 Nowcasting 方法。在这个模型里,所有可观测的宏观经济数据能够被若干不可观测的隐含因子所解释。
- 我们从工业、价格、国内外贸易、固定投资、财政、景气调查、银行与货币等方面选取了 39 个指标建立了中国经济 Nowcasting 模型。我们在模型中设定了 5 个隐含因子,分别是全局因子 (Global), 实际产出因子 (Real), 金融因子 (Finance), 情绪因子 (Sentiment) 和价格因子 (Price)。
- 本文还利用动态因子模型进行了实际测算。通过在样本内构建“伪数据集”的方法,我们展示了对于 2019 年二、三、四季度 GDP 同比的 Nowcasting 过程。结果发现,随着数据集的扩大,对当季 GDP 的预测值与实际发布值会越来越接近,这初步验证了基于动态因子模型的 Nowcasting 方法有效性。
- 最后,基于 2020 年 4 月 13 日的可得数据,我们对 2020 年一季度 GDP 同比做了样本外预测,最终预测值为-2.849%。2020 年 4 月 17 日国家统计局正式公布了我国 2020 年一季度 GDP 同比增速为-6.8%。

风险提示: 结论基于历史数据, 在市场环境转变时模型存在失效的风险。

目 录

1、Nowcasting	- 3 -
1.1、什么是 Nowcasting	- 3 -
1.2、宏观经济数据建模的难点	- 3 -
2、动态因子模型	- 4 -
2.1、标准动态因子模型	- 4 -
2.2、考虑误差项序列自相关	- 5 -
2.3、模型估计算法	- 5 -
2.4、Nowcasting 过程	- 6 -
3、中国经济 Nowcasting 模型	- 7 -
3.1、中国宏观数据处理	- 7 -
3.2、模型设定	- 9 -
4、实证结果	- 11 -
4.1、DFM 模型结果示例	- 11 -
4.2、Nowcasting 样本内预测示例	- 13 -
4.3、Nowcast2020 年一季度 GDP 同比	- 14 -
5、结论	- 16 -
图表 1、中国经济 Nowcasting 模型使用的数据及其因子载荷约束	- 10 -
图表 2、5 个隐含因子历史序列	- 11 -
图表 3、全局因子 (Global) 与标准化宏观数据	- 12 -
图表 4、对全局因子载荷前十大宏观指标	- 12 -
图表 5、实际产出因子 (Real) 与 GDP 和工业增加值数据	- 13 -
图表 6、Nowcasting 过程示例 (2019 年二季度 GDP 同比)	- 14 -
图表 7、Nowcasting 过程示例 (2019 年三季度 GDP 同比)	- 14 -
图表 8、Nowcasting 过程示例 (2019 年四季度 GDP 同比)	- 14 -
图表 9、Nowcasting 过程示例 (2020 年一季度 GDP 同比)	- 15 -
图表 10、2020 年 4 月 13 日 GDP 对数同比预测值变动分解	- 16 -

报告正文

实时预测 (Nowcasting) 是识别当前经济状态的流行方法。它的思想就是利用所有可得的信息来获得当前经济状态的估计。但是, 对宏观经济数据建模面临多个难点, 包括数据公布时间不同步、有缺失值、截面维度高和数据频率不一致等。为了解决这些问题, 本文将介绍一种基于动态因子模型的 Nowcasting 方法, 并利用此方法对我国 2019 年二、三、四季度季度做模拟预测, 结果发现预测准确度较高。最后, 基于 2020 年 4 月 13 日的可得数据, 我们对 2020 年一季度 GDP 同比做了样本外预测, 最终预测值为-2.849%。

1、Nowcasting

1.1、什么是 Nowcasting

对于政府机构来说, 当前经济状态直接影响到宏观政策的制定。例如, 央行需要判断当前经济是否过热, 从而决定是否需要通过加息等措施来给经济降温。对于投资者来说, 经济状态也直接影响投资策略。例如在经济上行时, 投资者应该配置更多的权益资产。因此准确且及时的识别当前经济环境具有重要的意义。

实时预测 (Nowcasting) 是识别当前经济状态的流行方法。它的思想就是利用所有可得的信息来获得当前经济状态的估计。举例来说, GDP 增速是反映经济状态的最直接指标, 但由于它只在每季度发布一次, 且发布时间滞后, 因此我们总是无法得到当前季度的 GDP 增速。这样一来, 我们就必须使用其他更高频指标如工业增加值增速等来推测当前季度的 GDP 增速。

那么如何对数量庞大的各种宏观指标建模, 从而实现实时的经济状态预测呢? 本文将介绍一种基于动态因子模型的 Nowcasting 方法, 它可以解决宏观经济数据建模面临的多种困难。

1.2、宏观经济数据建模的难点

宏观经济数据的特点决定了我们难以用简单模型如线性回归来对其建模和预测, 宏观数据建模的难点主要有以下几点:

1. **宏观数据公布时间不同步, 且可能有缺失值。**一般来说, 各类宏观数据是随着时间推进而逐渐发布的。PMI 这种调查数据可以在当月底获得; 工业、价格等数据在次月上旬发布; 而 M2、社融等数据要在次月中下旬才能获得。也就是说, 在任意时刻我们可得的最新宏观数据呈现出“锯齿状”结构。另外, 我国宏观数据在每年春节前后有“官方缺失”的特点。根据统计报表制度, 2 月份不发布 1 月份规模以上工业和能源生产、固定资产投资、房地产开发投资和销售、社会消费品零售总额和工业经济效益等数据。
2. **宏观数据频率不一致。**为了进行宏观数据研究, 我们需要将数据统一在一

个频率上。对宏观数据来说，月度频率是最常见的，但如 GDP 增速等重要数据却是季度发布的。如何科学的处理这种频率的不一致是一个问题。

3. **宏观数据截面维度高，但时间序列长度较短。**现代市场经济运行极为复杂，各种各样的统计指标都能在一定程度上反映经济运行情况，如工业生产、固定资产投资、国内贸易、对外贸易、财政、银行与货币、价格以及各类景气度调查等等。而由于宏观指标发布频率通常在月频以下，经常会出现宏观数据截面维度高于时间维度的情况，也就是有 $N>T$ 的问题，这给建立计量模型带来一定困难。另外，截面维度过高也给挑选和剔除指标带来了困难。
4. **宏观数据统计口径可能发生变化。**理论上说，统计口径变动是为了让测量值更好的逼近真实值。但统计口径大幅变动带来的不一致性可能会严重影响线性回归等简单模型的预测和推论。因此我们需要某种更加稳健的方法来对宏观数据建模。

2、动态因子模型

2.1、标准动态因子模型

本文使用的动态因子模型 (Dynamic Factor Model) 参考了美国纽约联储的经济实时预测模型，具体请参考论文 Bok et. al.(2017)。在这个模型里，可观测宏观经济数据能够被若干不可观测的隐含变量所解释，下面我们首先介绍动态因子模型的标准形式。

令 $y_t = [y_{1,t}, y_{2,t}, \dots, y_{n,t}]^T$ 表示 n 个标准化可观测变量，其中任意 $y_{i,t}$ 的均值为 0 且方差为 1。假设隐含因子数量为 r ，我们认为 y_t 与隐含因子之间满足如下关系：

$$y_t = \Lambda f_t + \epsilon_t, \quad \epsilon_t \sim i.i.d.N(0, R) \quad (1)$$

其中 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，我们还可以对 Λ 添加系数约束，例如规定某些数据只在特定因子上有载荷； $\epsilon_t = [\epsilon_{1,t}, \epsilon_{2,t}, \dots, \epsilon_{n,t}]^T$ 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：

$$f_t = A f_{t-1} + \mu_t, \quad \mu_t \sim i.i.d.N(0, Q) \quad (2)$$

其中 A 为维度 $r \times r$ 的自回归系数矩阵； μ_t 为误差向量，服从均值为 0 方差为 Q 的正态分布。

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 (State Space Model)。等式 (1) 被称为测量方程 (Measurement Equation)，它连接了可观测变量和隐含变量；等式 (2) 为转移方程 (Transition Equation)，它描述了隐含因子的动态过程。

2.2、考虑误差项序列自相关

标准动态因子模型中假设误差项不存在序列自相关性，而实际数据可能不满足这个条件。因此，本文使用的模型在标准动态因子模型基础上做了改进，我们假设误差项向量 ϵ_t 本身也服从 AR 过程。为了方便模型的估计，我们假设任意误差项 $\epsilon_{i,t} \in \{\epsilon_{1,t}, \epsilon_{2,t}, \dots, \epsilon_{n,t}\}$ 都可被分解为截面上互不相关的 $\tilde{\epsilon}_{i,t}$ 和 $\xi_{i,t}$ ：

$$\begin{aligned} \epsilon_{i,t} &= \tilde{\epsilon}_{i,t} + \xi_{i,t}, & \xi_{i,t} &\sim \text{i.i.d.} N(0, \kappa) \\ \tilde{\epsilon}_{i,t} &= \alpha_i \tilde{\epsilon}_{i,t-1} + e_{i,t}, & e_{i,t} &\sim \text{i.i.d.} N(0, \sigma_i^2) \end{aligned} \quad (3)$$

其中 $\xi_{i,t}$ 服从方差为 κ 的正态分布， κ 是一个极小的常数； $\tilde{\epsilon}_{i,t}$ 服从一阶 AR 过程， α_i 是其自回归系数。 $e_{i,t}$ 是误差项，服从方差为 σ_i^2 的正态分布。

结合 (1) (2) (3) 式，我们便定义了误差序列自相关动态因子模型。通过拓展隐含因子向量，我们可以将上述模型写成如下状态空间模型的形式：

$$\begin{aligned} y_t &= \tilde{\Lambda} \tilde{f}_t + \xi_t, & \xi_t &\sim \text{i.i.d.} N(0, \tilde{R}) \\ \tilde{f}_t &= \tilde{A} \tilde{f}_{t-1} + \tilde{\mu}_t, & \tilde{\mu}_t &\sim \text{i.i.d.} N(0, \tilde{Q}) \end{aligned} \quad (4)$$

其中 $\tilde{f}_t = \begin{bmatrix} f_t \\ \tilde{\epsilon}_t \end{bmatrix}$ 为拓展后的隐含因子向量； $\tilde{\mu}_t = \begin{bmatrix} \mu_t \\ e_t \end{bmatrix}$ 为误差向量；

$\tilde{\Lambda} = [\Lambda \quad I]$ 为因子载荷矩阵； $\tilde{A} = \begin{bmatrix} A & 0 \\ 0 & \text{diag}(\alpha_1, \dots, \alpha_n) \end{bmatrix}$ 为因子的自相关系数

矩阵； $\tilde{Q} = \begin{bmatrix} Q & 0 \\ 0 & \text{diag}(\sigma_1^2, \dots, \sigma_n^2) \end{bmatrix}$ 为转移方程误差向量的协方差矩阵； \tilde{R} 是测

量方程误差向量的协方差矩阵，它是对角阵，且对角线元素均为常数 κ 。

2.3、模型估计算法

动态因子模型中需要估计的变量包括各个隐含因子和模型参数，它们的取值互相依赖且未知变量数量庞大，因此普通的最大似然法难以有效的估计模型。幸运的是，学界已经提出了可以稳健估计动态因子模型的 PCA+EM 算法。

具体来说，PCA+EM 算法流程如下：

1. 用 PCA 方法初始化隐含因子和模型参数:
 - a. 对任意第 i 个隐含因子, 我们找到所有对其有载荷的数据。然后使用 PCA 方法得到其协方差矩阵最大特征值对应的特征向量, 并将数据投影到特征向量方向上得到主成分 $f_{i,t}$ 。主成分序列便是初始化隐含因子值, 特征向量即 Λ 中隐含因子对应的列系数初始值。然后我们从数据序列中减去已得到的主成分序列, 使用剩下的数据重复上述过程, 最终得到所有隐含因子和残差 ϵ_t , 以及拓展后的隐含因子序列 \tilde{f}_t 。
 - b. 对初始化隐含因子做自回归即得到转移矩阵 A 和残差 μ_t 。
 - c. 对残差 ϵ_t 做自回归便得到自回归系数 $\alpha_1, \dots, \alpha_n$ 和残差 e_t 。
 - d. 根据残差 μ_t 和 e_t 序列可得协方差矩阵 \tilde{Q} 。
 - e. 令 \tilde{R} 为对角阵, 且对角线元素均为常数 κ 。
 - f. 至此我们便得到了所有模型参数 \tilde{f}_t , $\tilde{\mu}_t$, $\tilde{\Lambda}$, \tilde{A} , \tilde{Q} 和 \tilde{R} 的初始值。
2. EM 算法迭代计算隐含因子和模型参数:
 - a. E-step: 基于上次迭代得到的模型参数和可观测数据集, 使用 Kalman 滤波得到隐含因子序列和当前模型的似然值。在这一步骤中, 若观测值有缺失, Kalman 滤波会用模型预测值将其自动填充, 这也是为什么动态因子模型可以处理带有缺失值和锯齿状的时间序列数据。
 - b. M-step: 基于 a 步骤得到的隐含因子序列和可观测数据集, 使用最大似然法得到新的模型参数。
 - c. 比较当前模型似然值和上次迭代的模型似然值, 若当前似然值相比上一次迭代似然值的减少幅度大于阈值, 则模型尚未收敛, 重新进入步骤 a; 否则停止 EM 算法迭代, 输出结果。

以上的 PCA+EM 算法将宏观数据映射为若干个隐含因子, 可以解决数据截面维度高的问题。其中的 EM 算法还使用了 Kalman 滤波, 它可以自动处理缺失值, 同时对于统计口径的变动具有较强稳健性。

2.4、Nowcasting 过程

这里我们简要说明基于动态因子模型的 Nowcasting 方法是如何纳入新信息来更新预测值的。我们用 Ω_v 表示在时刻 v 时的可得信息集, 基于 v 时刻的数据集 Ω_v , 我们可以使用 Kalman 滤波来获得 t 时刻可观测变量向量 $y_t = [y_{1,t}, y_{2,t}, \dots, y_{n,t}]^T$ 的预测值:

$$E[y_t | \Omega_v]$$

随着时间 v 的推进，在 $v+1$ 时刻， Ω_{v+1} 中可能因新数据公布或旧数据修正而包含更多信息。基于新数据集，我们可以得到 t 时刻 y_t 的新预测值：

$$E[y_t | \Omega_{v+1}] = E[y_t | \Omega_v] + E[y_t | I_{v+1}]$$

其中 I_{v+1} 表示 Ω_{v+1} 中与 Ω_v 正交的新信息，即新信息带来的预期差。也就是说，每次有新数据发布时，Nowcasting 方法会将实际值与其预测值比较，它们之间的差距 I_{v+1} 会被用来修正其他未公布数据的预测值。可以看出，在每次有新信息纳入时，DFM 模型会同时更新所有可观测变量 $y_t = [y_{1,t}, y_{2,t}, \dots, y_{n,t}]^T$ 的预测值。

在实际操作中，Nowcasting 过程分为两个部分。一方面我们将可以新数据集放入已估计的 DFM 模型来获取新预测值。另一方面，我们也需要用新数据集来重新估计 DFM 模型参数。

3、中国经济 Nowcasting 模型

前文介绍了动态因子模型理论及其估计方法，下面我们开始正式建立中国经济 Nowcasting 模型并介绍宏观数据的具体处理方法。

3.1、中国宏观数据处理

从实际操作便利性角度考虑，我们直接从 Wind 终端提供的宏观经济数据库获取数据集。

与发达经济体通常用环比数据进行分析不同，我国宏观数据类型以同比为核心，且部分指标只有累计同比数据（如固定资产投资完成额）。因此，与海外常见 Nowcasting 模型使用月度环比数据不同，我们将所有数据转换为月度对数同比的形式后再进入模型。

令 $Y_{i,t}^M = \ln(X_{i,t}^M)$ ，其中 $X_{i,t}^M$ 为第 t 年变量 i 的水平值，则我们有如下近似公式：

1. 月度对数同比：

$$y_{i,t}^{MYoY} = Y_{i,t}^M - Y_{i,t-12}^M$$

2. 累计对数同比：

$$\begin{aligned}
 y_{i,t}^{YTD-YoY} &\approx \frac{1}{t} \sum_{j=1}^t Y_{i,j}^M - \frac{1}{t} \sum_{j=1}^t Y_{i,j-12}^M \\
 &\approx \frac{1}{t} \sum_{j=1}^t y_{i,j}^{MYoY}
 \end{aligned}$$

根据上述公式，我们可以将所有月度同比和月度累计同比数据很方便的转换为月度对数同比数据。

而对于季度加总数据如 GDP 同比，由于其按季度发布，我们无法将其直接转换为月度对数同比数据，这就需要我们调整模型形式，将季度加总数据放入模型。一般来说，季度 GDP 对数同比 $y_{GDP,t}^{QYoY}$ 与不可观测的月度 GDP 同比 $y_{GDP,t}^{MYoY}$ 之间有如下近似对应关系：

$$\begin{aligned}
 y_{GDP,t}^{QYoY} &= Y_{GDP,t}^Q - Y_{GDP,t-12}^Q \\
 &\approx \frac{1}{3} (y_{GDP,t}^{MYoY} + y_{GDP,t-1}^{MYoY} + y_{GDP,t-2}^{MYoY})
 \end{aligned}$$

假设月度 GDP 对数同比与其他月度宏观变量服从同一个月度动态因子模型，即：

$$y_{GDP,t}^{MYoY} = \Lambda_{GDP,Q} f_t + \epsilon_{GDP,t},$$

则季度 GDP 对数同比满足：

$$\begin{aligned}
 y_{GDP,t}^{QYoY} &= \frac{1}{3} (1+L+L^2) (\Lambda_{GDP,Q} f_t + \epsilon_{GDP,t}) \\
 &= \bar{\Lambda}_{GDP,Q} [f_t^T, f_{t-1}^T, f_{t-2}^T]^T + \bar{\epsilon}_{GDP,t}
 \end{aligned}$$

其中 L 表示一阶滞后算子； $\bar{\Lambda}_{GDP,Q} = [\frac{\Lambda_{GDP,Q}}{3}, \frac{\Lambda_{GDP,Q}}{3}, \frac{\Lambda_{GDP,Q}}{3}]$ ，这个等式说明季度 GDP 对数同比由隐含因子过去 3 期的取值决定，且对各期载荷相同，这是对载荷系数矩阵的一种约束；同时 $\bar{\epsilon}_{GDP,t} = \frac{(\epsilon_{GDP,t} + \epsilon_{GDP,t-1} + \epsilon_{GDP,t-2})}{3}$ ；这样一来我们就将季度 GDP 对数同比与月度的隐含因子联系起来。

通过将月度对数同比数据 y_t^{MYoY} 与季度对数同比数据 y_t^{QYoY} 分开表示，考虑误差项序列自相关的动态因子模型可写为如下状态空间模型形式：

1. 测量方程 (Measurement Equation):

$$\begin{bmatrix} y_t^{MYoY} \\ y_t^{QYoY} \end{bmatrix} = \begin{bmatrix} \Lambda_M & 0 & 0 & I_{nM} & 0 & 0 & 0 \\ \Lambda_Q & \Lambda_Q & \Lambda_Q & 0 & I_{nQ} & I_{nQ} & I_{nQ} \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ \epsilon_t^M \\ \epsilon_t^Q \\ \epsilon_{t-1}^Q \\ \epsilon_{t-2}^Q \end{bmatrix} + \begin{bmatrix} \xi_t^M \\ \xi_t^Q \end{bmatrix}$$

其中 y_t^{MYoY} 表示月度对数同比数据向量, y_t^{QYoY} 表示季度对数同比数据向量, ξ_t^M 和 ξ_t^Q 为误差项, 且具有固定方差 κ 。 Λ_M 和 Λ_Q 分别表示月度数据和季度数据的载荷系数矩阵。 I_{nM} 和 I_{nQ} 分别是维度为可观测月度数据数量 nM 和可观测季度数据数量 nQ 的单位矩阵。 注意到, 这里我们将隐含因子向量拓展为 $[f_t^T, f_{t-1}^T, f_{t-2}^T, (\epsilon_t^M)^T, (\epsilon_t^Q)^T, (\epsilon_{t-1}^Q)^T, (\epsilon_{t-2}^Q)^T]^T$ 。

2. 转移方程 (Transition Equation):

$$\begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ \epsilon_t^M \\ \epsilon_t^Q \\ \epsilon_{t-1}^Q \\ \epsilon_{t-2}^Q \end{bmatrix} = \begin{bmatrix} A & 0 & 0 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_M & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_Q & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{nQ} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{nQ} & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ \epsilon_{t-1}^M \\ \epsilon_{t-1}^Q \\ \epsilon_{t-2}^Q \\ \epsilon_{t-3}^Q \end{bmatrix} + \begin{bmatrix} \mu_t \\ 0 \\ 0 \\ e_t^M \\ e_t^Q \\ 0 \\ 0 \end{bmatrix}$$

其中 A 表示 f_t 的自回归系数矩阵, 我们假设 A 为对角阵, 即每个隐含因子都服从一个 AR (1) 过程。 $\alpha_M = \text{diag}(\alpha_1, \dots, \alpha_{nM})$, $\alpha_Q = \text{diag}(\alpha_1, \dots, \alpha_{nQ})$ 分别表示月度和季度可观测变量误差项的自回归系数对角矩阵。

3.2. 模型设定

我们从工业、价格、国内外贸易、固定投资、财政、景气调查、银行与货币等度量宏观经济的不同角度选取了 39 个指标来建立中国经济 Nowcasting 模型。参考纽约联储 Nowcasting 模型的隐含因子设定, 我们在中国经济 Nowcasting 模型中设定了 5 个隐含因子, 分别是全局因子 (Global), 实际产出因子 (Real), 金融因子 (Finance), 情绪因子 (Sentiment) 和价格因子 (Price)。

不同的宏观数据根据其类型对 5 个隐含因子的载荷不同。举例来说, GDP 同比增速对全局因子和实际产出因子有载荷, 而对金融、情绪和价格因子没有载荷。

CPI 同比对全局因子和价格因子有载荷，而对实际产出、金融和情绪因子没有载荷。图表 1 详细给出了每个宏观数据的类型及其对 5 个隐含因子的载荷设定。

图表 1、中国经济 Nowcasting 模型使用的数据及其因子载荷约束

指标	频率	Global	Real	Finance	Sentiment	Price	类型
工业增加值:当月同比	M	1	1	0	0	0	工业
工业企业:产销率:当月同比	M	1	1	0	0	0	工业
工业企业:出口交货值:当月同比	M	1	1	0	0	0	工业
工业企业:利润总额:当月同比	M	1	1	0	0	0	工业
产量:钢材:当月同比	M	1	1	0	0	0	工业
货物周转量总计:当月同比	M	1	1	0	0	0	工业
货运量总计:当月同比	M	1	1	0	0	0	工业
产量:发电量:当月同比	M	1	1	0	0	0	工业
固定资产投资完成额:累计同比	M	1	1	0	0	0	固定资产投资
房地产开发投资完成额:累计同比	M	1	1	0	0	0	固定资产投资
民间固定资产投资完成额:累计同比	M	1	1	0	0	0	固定资产投资
公共财政收入:当月同比	M	1	0	0	0	0	财政
公共财政支出:当月同比	M	1	0	0	0	0	财政
进出口金额:当月同比	M	1	1	0	0	0	对外贸易及投资
贸易差额:当月同比	M	1	1	0	0	0	对外贸易及投资
社会消费品零售总额:实际当月同比	M	1	1	0	0	0	国内贸易
销量:零售:狭义乘用车:当月同比	M	1	1	0	0	0	国内贸易
CPI:当月同比	M	1	0	0	0	1	价格
CPI:非食品:当月同比	M	1	0	0	0	1	价格
RPI:当月同比	M	1	0	0	0	1	价格
PPI:全部工业品:当月同比	M	1	0	0	0	1	价格
70 个大中城市二手住宅价格指数:当月同比	M	1	0	0	0	1	价格
70 个大中城市新建商品住宅价格指数:当月同比	M	1	0	0	0	1	价格
PMI	M	1	0	0	1	0	景气调查
PMI:生产	M	1	0	0	1	0	景气调查
PMI:新订单	M	1	0	0	1	0	景气调查
PMI:新出口订单	M	1	0	0	1	0	景气调查
PMI:在手订单	M	1	0	0	1	0	景气调查
PMI:产成品库存	M	1	0	0	1	0	景气调查
PMI:采购量	M	1	0	0	1	0	景气调查
PMI:进口	M	1	0	0	1	0	景气调查
PMI:主要原材料购进价格	M	1	0	0	1	0	景气调查
PMI:原材料库存	M	1	0	0	1	0	景气调查
PMI:从业人员	M	1	0	0	1	0	景气调查
PMI:供货商配送时间	M	1	0	0	1	0	景气调查
M2:同比	M	1	0	1	0	0	银行与货币
金融机构:新增人民币贷款:当月值	M	1	0	1	0	0	银行与货币
社会融资规模:当月值	M	1	0	1	0	0	银行与货币
GDP:不变价:当季同比	Q	1	1	0	0	0	国民经济

资料来源: Wind, 兴业证券经济与金融研究院整理

注: 其中 M 表示月度, Q 表示季度。数字 1 表示此行数据对因子有载荷, 0 表示无载荷。

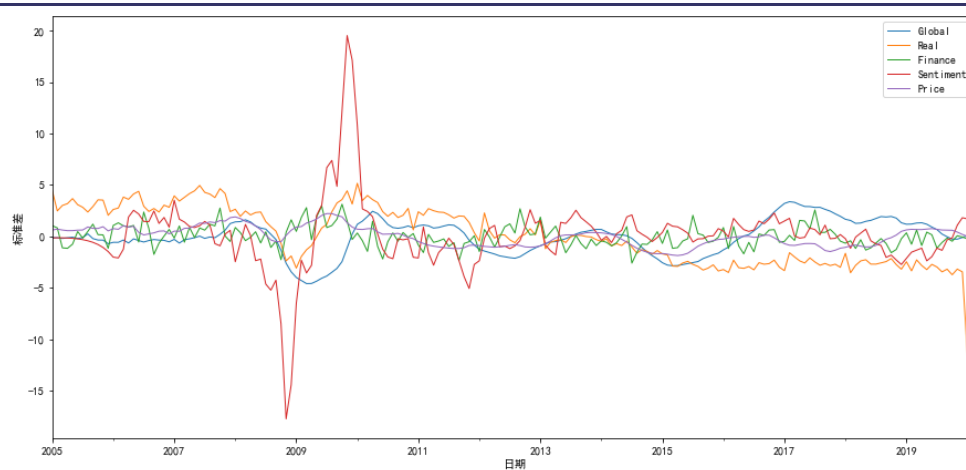
4、实证结果

4.1、DFM 模型结果示例

我们首先给出一个 DFM 模型的估计结果示例。假设当前时点为 2020 年 4 月 13 日，我们从 Wind 终端获取从 2005 年 1 月以来的所有可得宏观数据，然后通过 PCA+EM 算法估计出模型参数和 5 个隐含因子序列。

图表一给出了模型估计的 5 个隐含因子序列，他们分别代表全局因子(Global)、实际产出(Real)、金融(Finance)、情绪(Sentiment)和价格(Price)。我们可以看到在 2008 年金融危机前后各个因子水平出现了剧烈波动，且其中情绪因子变动较其他因子有明显领先性。另外，在 2020 年 2、3 月份，由于疫情冲击，可以看到实际产出和情绪因子出现大幅下跌。

图表 2、5 个隐含因子历史序列

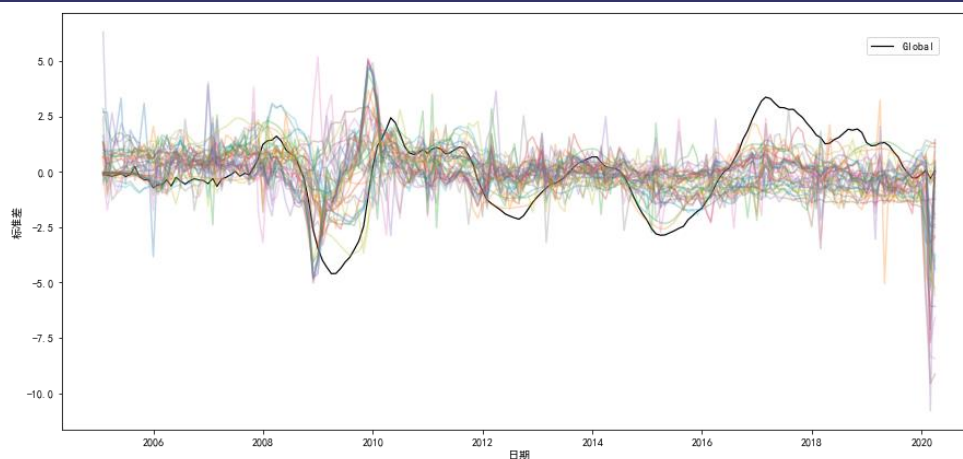


资料来源：Wind，兴业证券经济与金融研究院整理

我们之所以在模型中纳入全局因子，目的是为了捕捉所有宏观变量的共同推动因素，从而在后续做某一宏观指标的预测时能够纳入所有新信息对目标指标的冲击。另外，全局因子捕捉了经济运行中的共同推动因素，某种程度上可以作为经济整体活跃水平的一种度量。图表 3 给出了全局因子与本文使用的所有标准化宏观数据的历史序列，直观地看全局因子确实能够捕捉到变量间的总体变动趋势。

各宏观指标在因子上的载荷大小能在一定程度上表示其蕴含的信息量。图表 4 给出了对全局因子载荷最大的 10 个宏观指标，其中房地产价格载荷最大，这可能表明我国经济在最近 15 年里确实受房地产行业影响较大。

图表 3、全局因子（Global）与标准化宏观数据



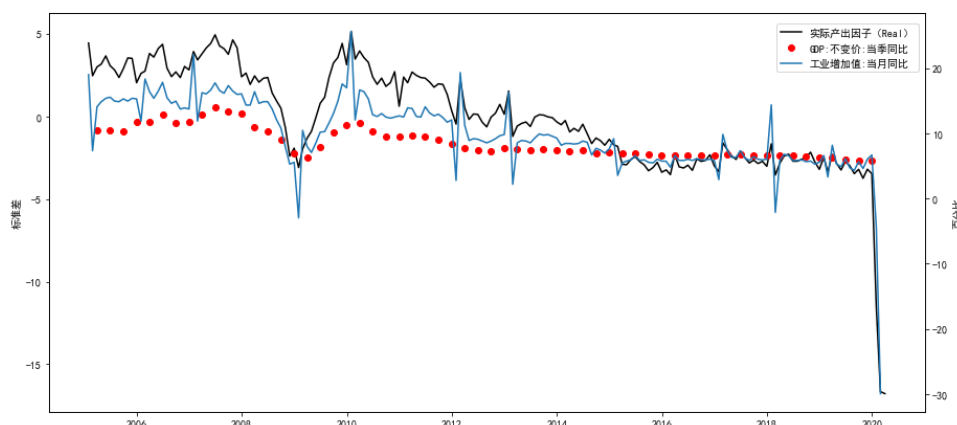
资料来源：Wind，兴业证券经济与金融研究院整理

图表 4、对全局因子载荷前十大宏观指标

排序	宏观指标	因子载荷
1	70 个大中城市二手住宅价格指数:当月同比	0.465
2	70 个大中城市新建商品住宅价格指数:当月同比	0.462
3	PPI:全部工业品:当月同比	0.357
4	工业企业:利润总额:当月同比	0.330
5	CPI:非食品:当月同比	0.195
6	产量:发电量:当月同比	0.186
7	进出口金额:当月同比	0.150
8	PMI:主要原材料购进价格	0.150
9	PMI:在手订单	0.147
10	货运量总计:当月同比	0.132

资料来源：Wind，兴业证券经济与金融研究院整理

图表 5 给出了实际产出因子与 GDP 同比和工业增加值同比的变动关系。从图表看，隐含因子走势与 GDP 同比和工业增加值同比走势非常一致，这说明实际产出因子能够捕捉到经济体中的实际产出变动。

图表 5、实际产出因子 (Real) 与 GDP 和工业增加值数据


资料来源：Wind, 兴业证券经济与金融研究院整理

在以上的讨论中我们发现，各隐含因子具有衡量经济运行状态的潜在能力。不同的隐含因子代表经济运行的不同方面，在未来，我们会进一步探究利用隐含因子序列来对经济周期进行划分的可行性。不过，本文的重点是探究动态因子模型对宏观指标的预测能力，下面我们以 GDP 同比的预测为例，给出 Nowcasting 模型的结果。

4.2、Nowcasting 样本内预测示例

如本文 2.4 部分所述，标准的 Nowcasting 过程是在每个时点获取所有可得数据，然后据此拟合 DFM 模型、更新预测结果。不过由于难以获取各宏观指标的历史发布时间数据，我们在这里通过构建历史月度“伪数据集”的形式来展示 Nowcasting 过程。

举例来说，假设我们处在 2019 年二季度，此时二季度 GDP 同比数据尚未发布，我们要利用其他已有数据来对其进行预测。假设在某时刻（如 2019 年 4 月下旬）2019 年 3 月份及之前的所有数据已经发布完毕，我们便利用 2005 年 1 月至 2019 年 3 月的所有月度数据拟合 DFM 模型，从而获得 2019 年二季度的 GDP 同比增速预测值。图表 6 中的 2019-03 行就表示基于 2019 年 3 月及之前数据对二季度 GDP 同比增速的预测结果。可以看到，此时二季度 GDP 预测值为 6.873%，与二季度实际发布增速 6.2% 差距较大。同样的，基于 2019 年 4 月及以前的数据，我们再次拟合 DFM 模型，获得二季度 GDP 预测值为 6.491%，预测值较上月预测有明显下降。基于 2019 年 5 月数据的二季度 GDP 预测值为 6.238%，较上月预测值继续下降，且与实际值的差距明显缩小。图表 6 中最后一行假设了除二季度 GDP 以外的其他 2019 年 6 月数据均已发布，基于此伪数据集的二季度 GDP 预测值为 6.336%。从以上预测过程可以看出，随着时间推移和数据集的扩大，对于 GDP 增速的预测会越来越准确，逐渐贴近实际发布值。

图表 6、Nowcasting 过程示例（2019 年二季度 GDP 同比）

数据集日期	预测值	实际值
2019-03	6.873%	6.200%
2019-04	6.491%	6.200%
2019-05	6.238%	6.200%
2019-06	6.336%	6.200%

资料来源：Wind，兴业证券经济与金融研究院整理

使用相同的方法，我们再给出对于 2019 年三、四季度 GDP 同比增速的预测过程示例。

从图表 7 看，基于 2019 年 6 月及之前数据的三季度 GDP 同比预测为 6.697%，与 6.0% 的实际值差距较大。而基于 2019 年 8 月和 2019 年 9 月的数据集对三季度 GDP 预测分别为 5.957% 和 5.988%，与实际值已经非常接近。

从图表 8 看，基于 2019 年 9 月及之前数据的四季度 GDP 同比预测为 6.227%。而基于 2019 年 11 月的数据集，四季度 GDP 预测值修正为 6.045%，非常接近实际值。

总体来看，随着数据的不断发布，对于当季 GDP 的预测值与实际发布值会越来越接近，这初步验证了本文基于动态因子模型的 Nowcasting 方法有效性。

图表 7、Nowcasting 过程示例（2019 年三季度 GDP 同比）

数据集日期	预测值	实际值
2019-06	6.697%	6.000%
2019-07	6.290%	6.000%
2019-08	5.957%	6.000%
2019-09	5.988%	6.000%

资料来源：Wind，兴业证券经济与金融研究院整理

图表 8、Nowcasting 过程示例（2019 年四季度 GDP 同比）

数据集日期	预测值	实际值
2019-09	6.227%	6.000%
2019-10	5.737%	6.000%
2019-11	6.045%	6.000%
2019-12	6.111%	6.000%

资料来源：Wind，兴业证券经济与金融研究院整理

4.3、Nowcast2020 年一季度 GDP 同比

4.2 部分的 Nowcasting 结果示例是基于伪数据集的样本内预测，还不能完全展示 Nowcasting 实际过程和证明 DFM 模型的预测有效性。在撰写本报告时，2020 年一季度 GDP 同比数据尚未发布，我们在此利用 DFM 模型对其进行样本外预测。

我们以 2020 年 4 月 6 日和 2020 年 4 月 13 日获取的两组数据集来展示实际

Nowcasting 过程。为了更好的展示一季度 GDP 同比预测值的变动过程，我们首先依然利用本文 4.2 部分的方法构建截止 2020 年 2 月的月度历史伪数据集，并用每个数据集拟合 DFM 模型从而获得一季度 GDP 同比预测值序列。

从图表 9 看，基于 2019 年 12 月及之前数据的一季度 GDP 同比预测值为 6.470%。而随着一月份数据发布，GDP 同比预测降为 4.847%。同时二月份经济数据出现较大下滑，这直接导致一季度 GDP 预测转为负增长 3.398%。

图表 9、Nowcasting 过程示例（2020 年一季度 GDP 同比）

数据集日期	预测值	实际值
2019-12	6.470%	-
2020-1	4.847%	-
2020-2	-3.398%	-

资料来源：Wind，兴业证券经济与金融研究院整理

下面我们不再构建伪数据集，而以在 2020 年 4 月 6 日实际可得的数据集 Ω_1 为基础估计 DFM 模型，模型直接给出的预测值是对数同比形式，数值为 -4.032%，转化为简单同比后为 $\exp(-4.032\%) - 1 = -3.952\%$ 。

到了 2020 年 4 月 13 日，我们再次获得了新的数据集 Ω_2 。这里我们不再重新估计模型，而是将基于数据集 Ω_1 得到的模型用在新数据集 Ω_2 上。由于有新数据发布，此时模型给出的一季度 GDP 对数同比调整为 -2.849%，转化为简单同比后为 -2.809%。注意到此时一季度 GDP 预测值相比 2020 年 4 月 6 日的预测出现了明显回升，那么这种变化由什么新信息导致的呢？幸运的是，我们使用的 DFM 模型可以将每个新发布数据带来的信息冲击清晰的分解出来。

图表 10 给出了 2020 年 4 月 6 日至 2020 年 4 月 13 日间发布的每个新数据对一季度 GDP 对数同比预测值的冲击。其中公布值表示新发布数据的对数同比；预测值表示基于 2020 年 4 月 6 日数据对这些新发布数据的对数同比预测值；预期差是公布值与预测值之差；权重表示新发布数据对一季度 GDP 对数同比的重要程度；信息冲击表示每个新数据对于一季度 GDP 对数同比预测的影响幅度，可以看到所有信息冲击之和为 1.185%，这正是两次 GDP 对数同比预测 -4.032% 和 -2.849% 的差值。换句话说，每个新数据对 GDP 对数同比预测值的影响程度满足如下等式：

$$\text{信息冲击} = \text{权重} * \text{预期差} = \text{权重} * (\text{公布值} - \text{预测值})$$

从图表 10 看，3 月份乘用车销量、货物周转量以及进出口情况超出预期，公布值均高于上次预测值，这使得模型上调了 GDP 对数同比预测值。因此，截止 2020 年 4 月 13 日我们对一季度 GDP 同比的最新预测值为 -2.849%。

值得注意的是，本文使用的动态因子模型是一种纯粹基于数据的计量模型，它的准确与否取决于数据的质量和更新的速度。另外，相比模型给出的预测绝对值，新信息带来的预测值边际变化可能对于指导投资有更重要的意义。

图表 10、2020 年 4 月 13 日 GDP 对数同比预测值变动分解

指标	数据日期	公布值	预测值	预期差	权重	信息冲击
销量:零售:狭义乘用车:当月同比	2020-3	-0.51751	-1.32072	0.80320	0.00985	0.00791
货物周转量总计:当月同比	2020-2	-0.19237	-0.33445	0.14208	0.02041	0.00290
进出口金额:当月同比	2020-3	-0.04082	-0.20445	0.16363	0.01697	0.00278
贸易差额:当月同比	2020-3	-0.45800	-1.55224	1.09424	0.00070	0.00077
社会融资规模:当月值	2020-3	0.55619	0.09545	0.46074	0.00001	0.00001
CPI:当月同比	2020-3	0.04210	0.04967	-0.00757	0.00000	0.00000
M2:同比	2020-3	0.09622	0.08480	0.01142	-0.00013	0.00000
金融机构:新增人民币贷款:当月值	2020-3	0.52259	0.10112	0.42147	0.00000	0.00000
RPI:当月同比	2020-3	0.02762	0.04087	-0.01325	0.00020	0.00000
CPI:非食品:当月同比	2020-3	0.00698	0.00978	-0.00281	0.00204	-0.00001
PPI:全部工业品:当月同比	2020-3	-0.01511	0.00156	-0.01667	0.00073	-0.00001
货物周转量总计:当月同比	2019-12	0.03537	-0.00410	0.03947	-0.00139	-0.00005
货运量总计:当月同比	2019-12	0.04974	0.03647	0.01327	-0.00782	-0.00010

资料来源: Wind, 兴业证券经济与金融研究院整理

5、结论

本文首先介绍了实时预测 (Nowcasting) 的概念和思想, 列举了宏观经济数据建模的难点。然后介绍了动态因子模型及其估计方法, 动态因子模型本身的特性使其可以方便的处理中国宏观经济数据具有的发布时间不一致、缺失值、截面维度高和数据频率不一致等问题。

随后我们从工业、价格、国内外贸易、固定投资、财政、景气调查、银行与货币等度量宏观经济的不同角度选取了 39 个指标建立了中国经济 Nowcasting 模型。参考纽约联储 Nowcasting 模型的隐含因子设定, 我们在中国经济 Nowcasting 模型中设定了 5 个隐含因子, 分别是全局因子 (Global), 实际产出因子 (Real), 金融因子 (Finance), 情绪因子 (Sentiment) 和价格因子 (Price)。

我们还利用动态因子模型进行了实际测算。通过在样本内构建“伪数据集”的方法, 我们展示了对于 2019 年二、三、四季度 GDP 同比的 Nowcasting 过程。结果发现, 随着数据集的扩大, 对当季 GDP 的预测值与实际发布值会越来越接近, 这初步验证了本文基于动态因子模型的 Nowcasting 方法有效性。

最后, 基于 2020 年 4 月 13 日的可得数据, 我们对 2020 年一季度 GDP 同比做了样本外预测, 最终预测值为-2.849%。2020 年 4 月 17 日国家统计局正式公布了我国 2020 年一季度 GDP 同比增速为-6.8%。

风险提示: 结论基于历史数据, 在市场环境转变时模型存在失效的风险

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

投资评级说明

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后的12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅。其中：A股市场以上证综指或深圳成指为基准，香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于15%
		审慎增持	相对同期相关证券市场代表性指数涨幅在5%~15%之间
		中性	相对同期相关证券市场代表性指数涨幅在-5%~5%之间
		减持	相对同期相关证券市场代表性指数涨幅小于-5%
		无评级	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

信息披露

本公司在知晓的范围内履行信息披露义务。客户可登录 www.xyzq.com.cn 内幕交易防控栏内查询静默期安排和关联公司持股情况。

使用本研究报告的风险提示及法律声明

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告仅供兴业证券股份有限公司（以下简称“本公司”）的客户使用，本公司不会因接收人收到本报告而视其为客户。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本报告所包含的材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告并非针对或意图发送予或为任何就发送、发布、可得到或使用此报告而使兴业证券股份有限公司及其关联子公司等违反当地的法律或法规或可致使兴业证券股份有限公司受制于相关法律或法规的任何地区、国家或其他管辖区域的公民或居民，包括但不限于美国及美国公民（1934年美国《证券交易所》第15a-6条例定义为本「主要美国机构投资者」除外）。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。

特别声明

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

兴业证券研究

上海	北京	深圳
地址：上海浦东新区长柳路36号兴业证券大厦15层	地址：北京西城区锦什坊街35号北楼601-605	地址：深圳市福田区皇岗路5001号深业上城T2座52楼
邮编：200135	邮编：100033	邮编：518035
邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn