

# 中国月度 GDP 同比增长率 估算与经济周期分析<sup>\*</sup>

高华川 白仲林

**内容提要:** 为了充分利用我国同比数据的优势、准确反映和及时调控宏观经济,本文提出一种估计同比数据混频近似因子模型的 EM 算法,并得到了我国月度 GDP 同比增长率的估计,然后利用此指标分析了我国经济波动的周期性。研究发现,月度数据的局部极差大于季度数据,尤其在宏观经济经历严重外部冲击时期,月度和季度 GDP 增长率数据相差较大,即季度 GDP 平滑了经济的波动性,低估了外部冲击效应。另外,月度 GDP 增长率数据能够更加精确、及时确定经济周期的转折点。

**关键词:** 月度 GDP 同比增长率; 混频近似因子模型; 经济周期

**DOI:** 10.19343/j.cnki.11-1302/c.2016.11.004

中图分类号: C812

文献标识码: A

文章编号: 1002-4565(2016)11-0023-09

## Estimation of China's Monthly GDP Year on Year Growth Rate and Business Cycle Analysis

Gao Huachuan & Bai Zhonglin

**Abstract:** In order to take full use of the advantages of China's year on year data, and to regulate China's macro-economy timely, this paper proposes an EM algorithm to estimate the mixed frequency approximate factor model, and gets the estimation of China's monthly GDP year on year growth rate. We use the estimated year on year monthly GDP growth to analyze China's business cycle. We found that the local range of monthly GDP is greater than that of quarterly GDP, especially during the period in which the economy undergoes large external shocks. This indicates that quarterly GDP smoothed the volatility of the economy, thus underestimate the impact of external shocks. In determining the turning points of the economy cycle, the monthly GDP is more precise and timely than the quarterly GDP.

**Key words:** Monthly GDP Year on Year Growth Rate; Mixed-frequency Approximate Factor Model; Business Cycle

### 一、引言

把握月度宏观经济状态对经济主体的行为选择和经济调控政策的选择至关重要,GDP无疑是衡量一国宏观经济状态的最佳综合指标。然而,多数国家仅仅公布季度GDP及其增长率。我国季度GDP及其增长率数据通常是在季后20日公布,于是宏观经济调控部门只能在未知新近状态的情况下选择远期目标的(季度性的)经济政策,各经济主体在不明确即期经济形势的状况下参与经济活动。所以,GDP及其增长率数据的披露频率偏低,使得经济政策的选择缺乏时效性和灵活性,导致宏观调控政策

滞后,甚至选择失误。由于缺乏反映月度GDP及其增长率的数据,政府机构与学者利用其他月度指标(如工业增加值或一致指数)进行实时分析。但是,工业增加值并不能全面刻画整体经济运行态势,并且波动剧烈,利用其增长率作为政策制定依据会导致政策的过激反应。另外,由于受春节日期不固定因素的影响,从2012年开始,我国工业增加值1~2月份数据同时调查和发布,不再分别披露1月和2

<sup>\*</sup> 本文得到国家自然科学基金项目“具有Markov体制转换动态因子模型建模方法及其应用研究”(71271142)、天津市哲学社会科学规划项目“大数据背景下的宏观经济实时预测”(TJ16-001Q)资助。

月的当月数据,这也为工业增加值指标的应用带来了困难。而一致指数也因缺乏经济含义而受到批评。例如,Mariano 和 Murasawa (2003)<sup>[1]</sup>指出,由于测度经济周期的一致指数不具有经济意义,不同指标构建的一致指数振幅不同,从而熨平经济波动的政策力度也存在差异。事实上,美国国家经济研究局(NBER)的经济周期定期委员会指出“本委员会将实际 GDP 作为衡量总体经济活动的最佳指标。然而,经济分析局(Bureau of Economic Analysis)的实际 GDP 为季度数据。由于这个原因,本委员会采用大量月度指标来确定经济周期的波峰与波谷。<sup>①</sup>”因此,估算月度 GDP 或其增长率对刻画整体经济运行状况、确定经济周期转折点以及及时准确地选择宏观调控政策具有重要意义。

对于月度实际 GDP 的估计,国外学者普遍采用状态空间方法,即假定季度 GDP 为当季不可观测的月度 GDP 之和,将不可观测的月度 GDP 作为状态向量建立状态空间模型,并使用 Kalman 滤波方法进行估计。例如,Guay 等(1990)<sup>[2]</sup>、Bernanke 等(1997)<sup>[3]</sup>、Cuche 和 Hess (2000)<sup>[4]</sup>利用状态空间模型分别估算了加拿大、美国和瑞士的月度 GDP。然而,状态空间方法只能利用少量变量进行建模,包含的信息量有限。另外,无论从政策分析还是从学术研究的角度,GDP 增长率比 GDP 绝对值更有意义。因此,近年来一些学者将季度 GDP 与其他月度一致指标的增长率建立混频动态因子模型(混频 DFM),利用宏观经济变量之间的协同变动关系来估计月度 GDP 的增长率。例如,Schumacher 和 Breitung(2008)<sup>[5]</sup>、Mariano 和 Murasawa(2010)<sup>[6]</sup>、Mitchell 等(2012)<sup>[7]</sup>利用混频 DFM 分别估算了德国、美国和英国的月度 GDP 环比增长率。

然而,我国环比经济数据发布始于 2011 年,原始环比增长率数据匮乏;另外,从历史同比数据换算环比数据又存在严重的舍弃误差。鉴于我国宏观经济的同比增长率数据较为完备,已基本上剔除了季节效应,也规避了我国特殊的移动假日效应,因此,本文通过建立一种混频因子模型估算了我国月度 GDP 同比增长率,为更实时地反映我国宏观经济运行状况、确定经济周期转折点和及时准确地选择宏观调控政策提供了实证依据和技术支持。

事实上,现有文献中 DFM 的估计方法可划分为三代。基于状态空间模型的卡尔曼滤波估计方法被

称为第一代,第二代是主成分估计方法,而将上述两代估计方法相结合的估计方法以及其他极大似然估计算法归为第三代方法。其中,主成分方法是近年来估计高维因子模型最常用的方法,但是它不能依据混频数据以及有缺失值的数据估计 DFM。虽然第一代估计方法能够估计混频因子模型,然而,在估计参数过程中需使用非线性数值优化算法,尤其,对于超高维 DFM,因待估参数倍增制约了该估计方法的计算能力,因此,第一代估计方法只能用于处理低维精确 DFM,在因子模型发展的早期较常用。例如,Engle 和 Watson (1981,1983)<sup>[8]</sup>、Stock 和 Watson (1989)<sup>[9]</sup>、Sargent (1989)<sup>[10]</sup>等文献。对于第三代估计方法,特别是极大似然估计的 EM 算法不仅能够依据混频数据和有缺失值的数据估计 DFM,而且估计过程无需非线性优化算法、迭代过程能够迅速收敛。即第三代估计方法更适用于高维混频因子模型的估计。因此,鉴于我国宏观经济数据的特殊性,本文将提出基于同比数据估计混频近似因子模型的 EM 算法。

本文的结构如下,第二部分讨论了同比数据混频因子模型及 EM 估计方法;第三部分描述了所采用的数据及数据处理方法;第四部分估算了我国月度 GDP 同比增长率,对该指标进行比较和评价,并根据该指标建立 Markov 转换自回归模型识别我国经济周期的衰退与扩张阶段,与相关文献的结果进行了对比分析;最后是本文的结论。

## 二、混频因子模型及其估计

### (一) 混频因子模型

混频因子模型是指由不同抽样频率的变量所组成的因子模型。因宏观经济变量的季度数据与月度数据较为常见,本文将讨论由季度变量与月度变量构建的混频因子模型。为了将月度变量与季度变量纳入同一模型,首先讨论流量变量的月度同比增长率和季度同比增长率之间的关系。

借鉴 Mariano 和 Murasawa(2003)关于环比增长率及郑挺国和王霞(2013)<sup>[11]</sup>关于同比增长率的构造方法,对于流量变量,可将季度观测值赋予各季度的第 3 个月,即将季度序列视为含缺失值的月度序列。于是,若  $Y_t^Q$  是季度流量变量,则月份  $t = 3, 6, 9, \dots$  时,

① <http://www.nber.org/cycles/resessions.html>.

$Y_t^Q$  有观测值,而在其他月份  $Y_t^Q$  为缺失值。并且,如果  $Y_t^M$  是对应  $Y_t^Q$  的不可观测的月度变量,那么:

$$Y_t^Q = Y_t^M + Y_{t-1}^M + Y_{t-2}^M = 3 \left[ \frac{1}{3} (Y_t^M + Y_{t-1}^M + Y_{t-2}^M) \right]$$

即季度变量可以表示为当季月度变量的算术平均值的3倍。于是,如果以几何平均值近似算数平均值,可得:

$$Y_t^Q = 3 (Y_t^M Y_{t-1}^M Y_{t-2}^M)^{1/3} \quad (1)$$

从而,对式(1)两边取对数后求12阶差分可得:

$$\ln Y_t^Q - \ln Y_{t-12}^Q = \frac{1}{3} [\ln Y_t^M - \ln Y_{t-12}^M + (\ln Y_{t-1}^M - \ln Y_{t-13}^M) + (\ln Y_{t-2}^M - \ln Y_{t-14}^M)]$$

设  $y_t^Q = \ln Y_t^Q - \ln Y_{t-12}^Q$  和  $y_t^M = \ln Y_t^M - \ln Y_{t-12}^M$  分别为季度变量和相应月度变量在  $t$  月份的同比增长率,于是:

$$y_t^Q = \frac{1}{3} (y_t^M + y_{t-1}^M + y_{t-2}^M) \quad (2)$$

所以,季度变量同比增长率为相应月度变量同比增长率的算数平均值。

假设  $x_t$  是  $N_1$  个可观测月度同比增长率组成的向量,  $x_t^Q$  是  $N_2$  个季度同比增长率组成的向量,如果  $x_t^M$  是对应于  $x_t^Q$  的不可观测的月度同比增长率向量,则可构建月度频率的动态因子模型:

$$\begin{pmatrix} x_t \\ x_t^M \end{pmatrix} = \begin{pmatrix} \Lambda \\ \Lambda_m \end{pmatrix} F_t + \begin{pmatrix} e_t \\ e_t^M \end{pmatrix} \quad (3)$$

$$B(L) F_t = \varepsilon_t \quad (4)$$

其中,  $F_t$  为  $r \times 1$  维共同因子,用于捕捉变量间的协同变动信息;  $\Lambda$  和  $\Lambda_m$  分别为  $N_1 \times r$  维、 $N_2 \times r$  维载荷系数矩阵;  $e_t$  和  $e_t^M$  称为异质性部分;  $B(L)$  为由  $p$  阶滞后算子多项式组成的  $r \times r$  维矩阵。假定  $e_t \sim i.i.d. N(0, \sum_1)$ ,  $e_t^M \sim i.i.d. N(0, \sum)$ ,  $\varepsilon_t \sim i.i.d. N(0, \sum)$ 。

由于  $x_t^M$  不可观测,上述模型不能直接估计,考虑将式(3)中的  $x_t^M$  根据式(2)用  $x_t^Q$  替换可得:

$$\begin{pmatrix} x_t \\ x_t^Q \end{pmatrix} = \begin{pmatrix} \Lambda & 0 & 0 \\ (1/3)\Lambda_m & (1/3)\Lambda_m & (1/3)\Lambda_m \end{pmatrix} \begin{pmatrix} F_t \\ F_{t-1} \\ F_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} e_t \\ e_t^M \\ e_{t-1}^M \\ e_{t-2}^M \end{pmatrix} \quad (5)$$

由于式(5)的残差项存在截面相关,称由式(5)和式(4)组成的模型为混频近似动态因子模型。

另外,如果对于共同因子  $F_t$  不施加式(4)的  $\text{VAR}(p)$  过程的约束,则式(5)即为混频近似因子模型。实际上,混频近似因子模型更具灵活性。例如,当设定共同因子  $F_t$  服从 Markov 转换自回归过程时,可以运用这种混频 MS-DFM 模型分析我国宏观经济波动的周期性。

## (二) 混频近似因子模型极大似然估计的 EM 算法

由于混频近似因子模型式(5)不仅存在不可观测的月度共同因子,而且未对共同因子设定状态方程式(4),所以,难以采用状态空间与 Kalman 滤波算法进行估计。为此,本文将给出混频近似因子模型式(5)极大似然估计的 EM 算法。

Dempster 等(1977)<sup>[12]</sup> 率先提出了极大似然估计的 EM 算法, Watson 和 Engle(1983)首次将 EM 算法用于 DFM 的估计, Stock 和 Watson(2002)<sup>[13]</sup> 应用 EM 算法估计了含有缺失值的非混频环比数据 DFM。EM 算法的基本思想是分别计算期望似然函数和最大化期望似然函数的一种迭代算法。其中, E 步(Expectation)是根据前一次迭代得到的参数估计值计算完全信息条件下的期望似然函数; M 步(Maximization)是最大化期望似然函数更新参数的极大似然估计值。EM 算法已被广泛应用于数据缺失值插补和含隐变量计量模型的参数估计。

对于因子模型:

$$X_t = \Lambda F_t + e_t$$

Bai 和 Ng(2002)<sup>[14]</sup> 证明  $F$  和  $\Lambda$  的主成分估计量使得残差平方和

$$V(F, \Lambda) = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_i F_t)^2$$

最小化,其中  $F = (F_1, \dots, F_T)'$ ,  $\lambda_i$  为因子载荷矩阵  $\Lambda$  的第  $i$  行向量。也就是在  $X_{it}$  不存在缺失值的情况下,主成分估计量等价于最小二乘估计量。而当  $X_{it}$  含有缺失值时,最小化目标函数:

$$V^*(F, \Lambda) = \sum_{i=1}^N \sum_{t=1}^T I_{it} (X_{it} - \lambda_i F_t)^2$$

即可得到  $F_t$  和  $\Lambda$  的最小二乘估计量,其中,若  $X_{it}$  为缺失值,则  $I_{it} = 0$ , 否则  $I_{it} = 1$ 。

显然,若  $X_{it} \sim N(\lambda_i F_t, 1)$ , 则  $V(F, \Lambda)$  与样本  $\{X_{it} | i = 1, \dots, N; t = 1, \dots, T\}$  的对数似然成比例,于是,因子模型的最小二乘估计等价于极大似然估

计。另外,注意到  $V^*$  与  $V$  的区别仅在于  $V^*$  是具有缺失值情况下的残差平方和,而  $V$  是完全数据下的残差平方和。因此,可利用 EM 算法最小化  $V^*$ 。

假定 EM 算法已经迭代了  $j$  次,得到估计量  $\hat{A}$  和  $\hat{F}_t$ , 则第  $j+1$  次迭代的步骤如下:

设

$$Q^*(X^*, \hat{F}, \hat{A} | F, A) = E_{\hat{F}, \hat{A}}[V(F, A) | X^*] \quad (6)$$

其中,  $X^*$  表示已经填补了缺失值的完全数据集,  $E_{\hat{F}, \hat{A}}[V(F, A) | X^*]$  表示似然函数  $V(F, A)$  在完全数据下的条件期望在  $(\hat{F}, \hat{A})$  处的取值。

于是,在第  $j+1$  次迭代中,最小化式(6)可得到  $F$  和  $A$  的估计值。另外:

$$Q^*(X^*, \hat{F}, \hat{A} | F, A) = \sum_{i=1}^N \sum_{t=1}^T [E_{\hat{F}, \hat{A}}(X_{it}^2 | X^*) + (\lambda_t F_t)^2 - 2\hat{X}_{it}(\lambda_t F_t)] \quad (7)$$

其中,  $\hat{X}_{it} = E_{\hat{F}, \hat{A}}(X_{it} | X^*)$  说明在第  $j+1$  次迭代中  $X_{it}$  的缺失值可以由观测数据和第  $j$  次迭代中估计值  $\hat{F}$  和  $\hat{A}$  的条件均值进行插补。

并且,因式(7)右边第 1 项  $E_{\hat{F}, \hat{A}}(X_{it}^2 | X^*)$  与  $F$  和  $A$  无关,它可以用  $\sum_{i=1}^N \sum_{t=1}^T \hat{X}_{it}^2$  代替。从而,使式(6)最小化的  $F$  和  $A$  可以由最小化  $\hat{V}(F, A) = \sum_{i=1}^N \sum_{t=1}^T (\hat{X}_{it} - \lambda_t F_t)^2$  得到,即运用完全数据的主成分估计方法可得到第  $j+1$  次迭代中  $F$  和  $A$  的估计量。另外,可根据  $X_{it}$  的某个完全数据子集的主成分估计量  $\hat{F}$  和  $\hat{A}$  作为 EM 算法的初始值。

因此,对于本文讨论的混频近似因子模型式(5),EM 算法的迭代过程如下:

(1) 初始化。

设  $X$  是由月度变量和季度变量所组成有缺失值的原始样本数据。如果原始样本数据中包含某个完全数据子集  $\check{X}$ , 可由  $\check{X}$  的主成分估计量作为迭代初始值。或者使用其他插补缺失值的方法得到完全数据  $\hat{X}^{(0)}$ , 并将基于此完全数据的主成分估计量  $\hat{F}_T^{(0)}$  和  $\hat{A}^{(0)}$  作为算法初始值。

(2) 第  $j+1$  次迭代的 EM 算法。

设  $\hat{F}_t^{(j)}$  和  $\hat{A}^{(j)}$  为第  $j$  次迭代中得到的估计量,  $\hat{\lambda}_i^{(j)}$  为  $\hat{A}^{(j)}$  的第  $i$  行向量。那么第  $j+1$  次迭代为:

E 步: 构造不存在缺失值的完全数据  $\hat{X}^{(j+1)}$ , 即

处理混频与缺失值问题。

缺失值: 若观测到  $X_{it}$ , 则令  $\hat{X}_{it}^{(j+1)} = X_{it}$ , 否则令  $\hat{X}_{it}^{(j+1)} = \hat{\lambda}_i^{(j)} \hat{F}_t^{(j)}$ ;

混频: 由式(5)可令  $\hat{e}_{it}^{(j)} = X_{it} - (1/3) \hat{\lambda}_i^{(j)} (\hat{F}_t^{(j)} + \hat{F}_{t-1}^{(j)} + \hat{F}_{t-2}^{(j)})$ , 其中当  $t=1, 2, 3$  时, 则  $\tau=3$ , 当  $t=4, 5, 6$  时, 则  $\tau=6$ , 等等。然后, 令  $\hat{X}_{it}^{(j+1)} = \hat{\lambda}_i^{(j)} \hat{F}_t^{(j)} + \hat{e}_{it}^{(j)}$ ,  $t=1, 2, 4, 5, 7, 8, \dots$ 。

记  $\hat{X}^{(j+1)}$  为第  $j+1$  迭代得到的完全数据。

M 步: 对  $\hat{X}^{(j+1)}$  进行主成分估计得到因子及其载荷矩阵的新估计量  $\hat{F}_t^{(j+1)}$  和  $\hat{A}^{(j+1)}$ 。

(3) 迭代终止条件。

当前后两次迭代的残差平方和之差小于某一很小的绝对收敛率  $\zeta$  时(例如  $\zeta = 10^{-6}$ ), 终止迭代。

显然,混频近似因子模型的 EM 迭代估计方法充分利用了变量间的协同变动信息实现缺失值插补,具有较确凿的经济学理论依据。另外,因 EM 算法未涉及非线性优化算法,使得迭代收敛迅速。所以,本文提出的 EM 估计方法也适用于高维混频因子模型的估计。

### 三、数据及数据处理

借鉴郑挺国和王霞(2013)、叶光(2015)<sup>[15]</sup> 在利用混频 DFM 构建监测我国经济周期状况一致指标时所采用的变量,本文选取了 GDP、工业增加值、固定资产投资完成额、社会消费品零售总额、进出口总额以及发电量作为原始变量<sup>①</sup>,并且依据可比价格计算变量的同比增长率。另外,由于受春节效应的影响,月度数据在每年的 1~2 月份存在明显的离群值现象,为此本文将各变量每年 1~2 月份的数据删除,作为缺失数据处理。最后,对各变量进行去均值的中心化处理。各变量的观测频率和样本起止日期如表 1 所示。

### 四、实证结果

为了及时反映我国宏观经济运行状况和准确选择宏观调控政策,本部分将利用迭代的 EM 算法估

<sup>①</sup> 由于税收总额不仅与 GDP 数据相关性较弱,而且异常值较多,本文用电量替换了郑挺国和王霞(2013)中的税收总额变量。

表1 原始变量的观测频率和样本起止日期

变量	频率	起始日期	截止日期
GDP	季度	1992年1季度	2015年1季度
工业增加值	月度	1992年1月	2015年3月
发电量	月度	1992年1月	2015年3月
固定资产投资完成额	月度	1992年6月	2015年3月
社会消费品零售总额	月度	1992年1月	2015年3月
进出口总额	月度	1995年1月	2015年3月

数据来源: Wind 数据库。

计我国同比数据的混频近似因子模型,构建月度GDP的同比增长率。具体估计步骤如下:

### (1) 初始化。

将实际季度GDP同比增长率作为每季第3个月的观测值,即将季度GDP同比增长率视为含有缺失值的月度序列,记为 $X_{1t}$ ,当 $t = 3, 6, 9, \dots$ 时 $X_{1t}$ 有观测值, $t = 1, 2, 4, 5, 7, \dots$ 时 $X_{1t}$ 为缺失值。将其他5个月度变量记为 $X_{it}$  ( $i = 2, 3, \dots, 6$ )。令 $X$ 表示由此6个变量组成的含有缺失值的原始数据矩阵。由于EM算法的收敛结果与初始值的选取无关,本文将所有缺失值用标准正态分布产生的随机数填补,得到不含缺失值的完全数据 $\hat{X}^{(0)}$ ,设定因子个数 $r = 1$ 并利用主成分法得到因子与载荷的初始估计 $\hat{F}_t^{(0)}$ 和 $\hat{\Lambda}$ 。

### (2) 第 $j+1$ 次迭代的EM算法( $j=0, 1, 2, \dots$ )。

设 $\hat{F}_t^{(j)}$ 和 $\hat{\Lambda}^{(j)}$ 为第 $j$ 次迭代中得到的估计量, $\hat{\lambda}_i^{(j)}$ 为 $\hat{\Lambda}^{(j)}$ 的第 $i$ 行向量。那么第 $j+1$ 次迭代为:

E步:构造不存在缺失值的完全数据 $\hat{X}^{(j+1)}$ ,即处理混频与缺失值问题。

对于 $X_{it}$  ( $i = 2, 3, \dots, 6$ ) 中的缺失值:若观测到 $X_{it}$ ,则令 $\hat{X}_{it}^{(j+1)} = X_{it}$ ,否则令 $\hat{X}_{it}^{(j+1)} = \hat{\lambda}_i^{(j)} \hat{F}_t^{(j)}$ ;

对于 $X_{1t}$ 的混频:由式(5)可令 $\hat{e}_{1t}^{(j)} = X_{1t} - (1/3) \hat{\lambda}_1^{(j)} (\hat{F}_t^{(j)} + \hat{F}_{t-1}^{(j)} + \hat{F}_{t-2}^{(j)})$ ,其中当 $t = 1, 2, 3$ 时,则 $\tau = 3$ ,当 $t = 4, 5, 6$ 时,则 $\tau = 6$ ,等等。然后,令 $\hat{X}_{1t}^{(j+1)} = \hat{\lambda}_1^{(j)} \hat{F}_t^{(j)} + \hat{e}_{1t}^{(j)}$ ,  $t = 1, 2, 4, 5, 7, 8, \dots$ 。

记 $\hat{X}^{(j+1)}$ 为第 $j+1$ 迭代得到的完全数据。

M步:对 $\hat{X}^{(j+1)}$ 进行主成分估计得到因子及其载荷矩阵的新估计量 $\hat{F}_t^{(j+1)}$ 和 $\hat{\Lambda}^{(j+1)}$ 。

### (3) 迭代终止条件。

重复第ii步,直到前后两次迭代的残差平方和之差小于绝对收敛率 $\xi = 10^{-6}$ ,终止迭代。记最后得

到的不含缺失值的完全数据为 $\hat{X}^*$ ,其第1列 $X_{1t}^*$ 即为月度GDP同比增长率的估计。

当迭代算法终止时,即满足收敛准则时,使用每次迭代的残差平方和绘制的折线图如图1所示。从图1可见,本文所给出的迭代EM算法的收敛速度很快,所以,该方法具有较好的适用性。

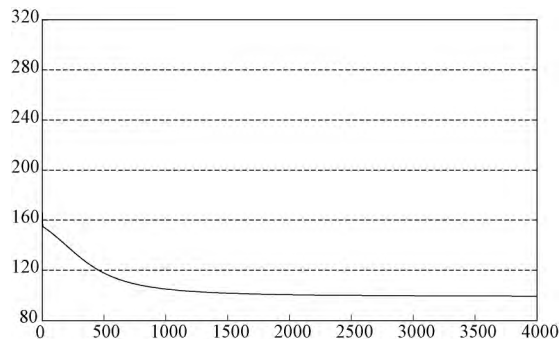


图1 迭代中残差平方和的变化过程

### (一) 月度GDP同比增长率

由于在建立混频近似因子模型时,所有变量均进行了样本中心化,所以,还需对已插补的月度GDP同比增长率数据再进行均值恢复,本文使用原季度GDP同比增长率的样本均值估计月度GDP同比增长率的均值。月度GDP同比增长率与原季度GDP同比增长率的描述性统计分析结果见表2。由表2可见,月度GDP同比增长率与原季度GDP同比增长率的分布具有基本一致的统计特征。因此,本文所建立的混频近似因子模型不仅保持了GDP同比增长率的基本统计性质,而且得到了月度GDP同比增长率的较好估计。

表2 GDP同比增长率的描述性分析

统计量	月度GDP增长率	季度GDP增长率
均值	10.0526	10.0785
中位数	9.6762	9.7000
标准差	2.3866	2.3651
偏度	0.6594	0.6483
峰度	2.8896	2.7429

另外,从月度GDP增长率与季度GDP增长率的折线图①(图2)可见:①两种同比增长率的走势基本一致;②季度GDP增长率不仅比月度GDP增长率平滑,而且在数据的每个“S型”期间,月度数据的局部极差大于季度数据;③在月度GDP增长率与

① 由于季度GDP增长率为当季月度GDP增长率的均值,因此每季度内的月度值就定义为季度GDP增长率。

原季度 GDP 增长率差异较大的时期,中国经济均受到某种较严重的外生冲击。例如,1992 年 10 月份十四大确立了中国经济体制改革的目标为建立社会主义市场经济体制,受此影响,1992 年 11 月的月度 GDP 同比增长率的估值为 18.8%,高于 1992 年第 4 季度的 GDP 同比增长率 17.3%。在 1997 年香港回归期间,月度 GDP 增长率在 1997 年 7 月、8 月和 9 月的估计值分别为 7.6%、9.8% 和 8.5%,而 1997 年第 3 季度的季度 GDP 增长率为 8.5%,显然平滑了香港回归引致的经济波动性。事实上,在香港回归期间,工业增加值和进出口总额等变量的月度同比增长率数据的波动也较大,例如 1997 年 7 月、8 月和 9 月的工业增加值同比增长率分别为 8.4%、10.9% 和 11.1%,进出口总额同比增长率分别为 19.7%、7.2% 和 18.3%。并且,受 1997 年亚洲金融危机的影响,我国披露的季度 GDP 增长率于 1998 年第 2 季度降低为 6.7% 的最低值,但是,根据本文的月度 GDP 同比增长率估计值,亚洲金融危机使我国 4 月的 GDP 同比增长率降低为 5.9%,比季度数据反映的经济状况更严峻。另外,受“非典”的影响,2003 年 5 月月度 GDP 增长率的估计值 7.4% 也低于 2003 年第 2 季度 8.7% 的季度 GDP 增长率。由此可见,与估算的月度 GDP 同比增长率比较,季度 GDP 同比增长率数据不仅光滑了经济的实际波动性,同时“低估了经济上升、高估了经济下降”,而且也低估了经济增长的外生冲击效应。



图 2 季度 GDP 与月度 GDP 增长率走势图

## (二) 经济周期分析

Burns 和 Mitchell (1946) [16] 曾指出“经济周期是一国总体经济活动的一种波动,它由许多经济活动中几乎同时发生的扩展、衰退、紧缩和复苏四个阶段组成。”于是,宏观经济变量中的共变性是经济周期的基本特征之一。所以,为了揭示和反映经济增长率的整体波动性,本文以混频近似因子模型

构建的月度 GDP 同比增长率进行周期性分析。并且,为了反映经济增长率周期的非对称性,以概率分布形式科学和更合理地描述经济增长率周期的状态,以及识别和预测周期状态,本文假定月度 GDP 同比增长率  $z_t$  服从一个两状态 Markov 区制转换自回归过程:

$$z_t - \mu_{s_t} = \sum_{h=1}^p \beta_h (z_{t-h} - \mu_{s_{t-h}}) + \varepsilon_t, \varepsilon_t \sim i.i.d. N(0, \sigma^2) \quad (8)$$

其中,假设状态变量  $s_t$  服从遍历 Markov 链过程

$$P(s_t = j | s_{t-1} = i, s_{t-2} = k, \dots) = P(s_t = j | s_{t-1} = i) = p_{ij}$$

即  $s_t$  根据转移概率矩阵  $P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix}$  进行状态转移。

显然,若  $\mu_1 < \mu_2$ , 则  $s_t = 1$  和  $s_t = 2$  分别表示在  $t$  时期经济处于衰退状态和扩张状态,即扩张状态表现为高的平均月度 GDP 同比增长率,衰退状态表现为低的平均月度 GDP 同比增长率。另外,根据 AIC 准则确定式(8)的滞后阶数,从而得滞后阶数  $p = 1$ ; 最后,采用 Hamilton (1989) [17] 极大似然方法的 Kalman 滤波算法估计式(8),同时,也分别以季度 GDP 同比增长率和一致指数①估计式(8),估计与检验结果如表 3 所示。

表 3 Markov 转换自回归估计结果

参 数	月度 GDP 增长率		季度 GDP 增长率		一致指数 $F_t$	
	衰退区制	扩张区制	衰退区制	扩张区制	衰退区制	扩张区制
$\alpha$	8.6209 *** (0.1039)	12.6976 *** (0.2045)	8.7599 *** (0.3182)	12.8087 *** (0.6234)	-0.9222 *** (0.0441)	0.7568 *** (0.0567)
$\beta$	0.7833 *** (0.0376)	-	0.5811 *** (0.0866)	-	0.8710 *** (0.0303)	-
$p_{ii}$	0.9813	0.9757	0.9538	0.9365	0.9761	0.9869

注: \*\*\*, \*\*, \* 分别表示在 1%、5% 和 10% 的显著性水平下拒绝原假设,括号内为系数估计值的样本标准差。

由表 3 可见,各系数估计值均在 5% 显著性水平下显著,表明我国月度 GDP 增长率存在明显的周期性特征。首先,在衰退和扩张期月度 GDP 同比增长率的均值分别为 8.62% 和 12.70%,相差 4.08 个百分点;衰退区制和扩张区制的平均持续期  $(1 - p_{ii})$  分别为 53.5 个月和 41.2 个月。其次,与根据季度 GDP 同比增长率估计的模型比较, GDP 同比增长

① 当 EM 算法结束后,将所得到的因子估计  $F_t$  作为刻画我国总体经济运行状况的一致指数。

率在各区制长期水平(均值)仅仅相差约0.1,而自回归系数差异显著,月度数据模型较高的自回归系数(0.7833)也充分说明了月度GDP同比增长率存在更强的记忆性;同时,各区制的状态保持概率( $p_{ii}$ )也高于季度数据模型的相应值,这也再次印证了月度增长率数据具有较强的自相关性。而且,从衰退区制与扩张区制平均持续期的差可见,月度数据反映的经济的非对称性(7.7个月)也低于季度数据(5.9个季度)。并且,与一致指数估计的模型比较,两模型的自回归系数和各区制的状态保持概率比较接近(相对于季度数据模型),即除数据量纲区别外,一致指数与月度GDP同比增长率具有较相近的动态行为。但是,一致指数与GDP同比增长率所反映的经济非对称性相反,扩张区制的平均持续期长于衰退区制。显然这与我国近年来的实际经济表现相左。另外,如果将 $s_t = 1$ 的平滑转移概率大于0.5的时期 $t$ 界定为该期经济处于衰退区制,则我国经历了三个经济衰退期,分别是1996年3月至2005年1月、2008年7月至2009年10月和2010年6月至2015年3月,如图3为用月度GDP同比增长率识别的我国经济处于衰退区制的平滑转移概率;而用季度GDP同比增长率和一致指数识别的我国处于衰退区制的平滑转移概率如图4和图5所示。

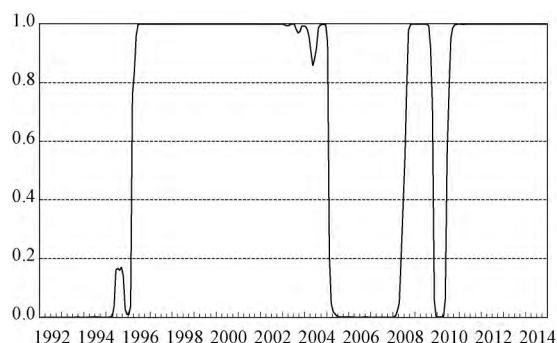


图3 我国经济处于衰退区制的时期及平滑转移概率(月度GDP增长率)

事实上,在1992年年初邓小平发表南方谈话之后,我国全面推进和深化经济体制改革,宏观经济出现“过热”现象。在1992年1月至1993年12月工业增加值平均增速为19.8%,固定资产投资平均增速为50.9%。在有效供给不足的条件下,经济过热导致通货膨胀压力明显上升,CPI高达18.8%。随后我国于1993年6月开始实行紧缩性宏观调控政策,严格控制信贷规模,大幅提高了存贷款利率,并要

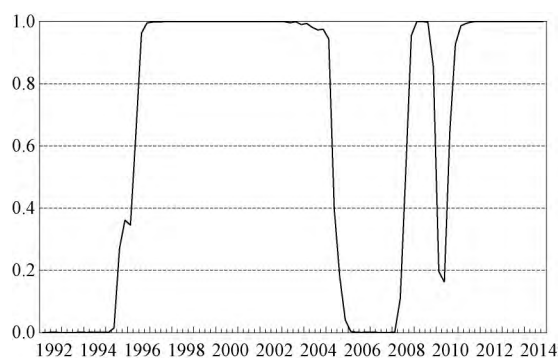


图4 我国经济处于衰退区制的时期及平滑转移概率(季度GDP增长率)

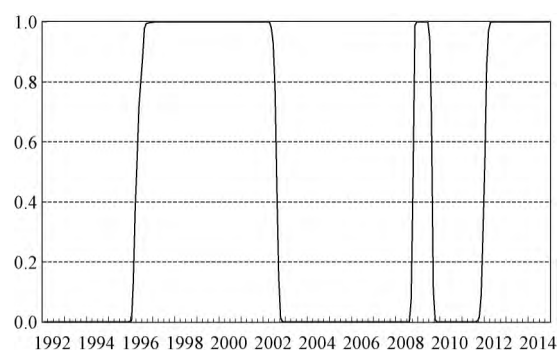


图5 我国经济处于衰退区制的时期及平滑转移概率(一致指数 $F_t$ )

求银行定期回收违规拆借资金。在财政政策方面,实行分税制改革,并适度压缩财政支出。“双紧”的货币与财政政策组合实施,以牺牲GDP增长率赢得了通货膨胀率的大幅回落,有效地抑制了经济过热。各界普遍认为至1996年我国实现了宏观经济的“软着陆”。显然,本文的估计结果也显示我国宏观经济自1996年3月从扩张区制转为衰退区制。之后,在“软着陆”经济政策的滞后效应、1997年的亚洲金融危机、2001年加入WTO和2003年“非典”冲击等诸多因素的影响下,中国经济持续处于衰退区制。2003年下半年投资过度增长,中国经济出现过热现象,2004年1季度的固定资产投资增长高达43%。并且,因2004年政府换届也引致了“政绩工程”建设的投资热潮。中国经济终于在2005年1月走出了长达近9年之久的衰退区制。然而,在这期间大幅增加基础设施建设、积极引进外资和大力推进出口贸易发展等一系列经济战略使中国GDP实现了平均8%以上的增长率,也为“中国增长”神话奠定了基础。

其次,随着2007年2月爆发的金融危机的蔓延和加剧,加之2008年年初“雨雪冰冻”和5月“汶川

地震”等自然灾害造成的直接经济损失,以及受货币政策(始于2007年2月的从紧货币政策直到2008年7月底才调整为“审慎灵活”的货币政策,再到11月调整为适度宽松的货币政策。)错配的影响,2008年7月中国经济再次掉入衰退区制。随着世界经济复苏引致的净出口恢复,以及“4万亿”经济刺激计划和适度宽松货币政策的调控,2009年10月中国经济迅速摆脱了衰退区制。

另外,由于金融危机时期经济刺激计划所产生的“矫枉过正”效应和国际宽松的货币环境对资本流动的影响,2009年下半年固定资产投资持续增长趋势丧失,出现震荡下降,使得GDP增长率连续下滑,并于2010年6月结束了不足一年的经济扩张期。另外,经过30年经济高速增长,由于产业空间趋于饱和、资源环境压力增大、劳动力成本上升和技术创新不足等因素的影响,我国经济增速放缓,进入了一个稳健的发展期——经济发展新常态。根据本文估计的结果,自2010年6月至2015年3月中国经济第三次进入衰退区制。

从上述分析可知,本文估计出的我国月度GDP同比增长率较好地刻画了我国总体经济的运行状况,以月度GDP同比增长率定义的经济周期与国内外经济环境、供需条件以及宏观调控政策选择均具有较高的相关性。

最后,将根据月度GDP同比增长率、季度GDP同比增长率、一致指数 $F_t$ 以及郑挺国和王霞(2013)的一致指数(记为 $F_t^Z$ )所识别出的经济衰退区间列示于表4。

表4 各种一致指标识别出的衰退时期

采用指标	第一次衰退	第二次衰退	第三次衰退
月度GDP	1996m03—2005m01	2008m07—2009m10	2010m06—2015m03
季度GDP	1996q2—2005q1	2008q3—2009q4	2010q3—2015q1
$F_t$	1996m04—2002m07	2008m11—2009m08	2012m01—2015m03
$F_t^Z$	1995m05—2002m12	2008m09—2009m09	2011m11—2011m12

注:m表示月,q表示季度;郑挺国、王霞(2013)采用的样本期为1992年1月至2011年12月。

由表4可见,首先,依据这四种指标均能识别出中国经济自1992年至今的三次衰退期,并且确定的经济周期转折点也比较接近。尤其是季度GDP与月度GDP同比增长率识别出的经济周期转折点非常相似,但月度数据比季度数据更精确和及时。例如,对于第二个衰退区制,从季度GDP同比增长率确定的起始时点为2008年第3季度,而月度GDP

同比增长率识别出的起始日期为2008年7月,比前者提前了2个月。但是,对于第一次衰退区制的截止期,两种共同因子识别出的结果分别是2002年7月和2002年12月,而GDP同比增长率识别的截止期分别为2005年1月和2005年第1季度。造成这种差异的主要原因不仅在于两类指标包含的变量信息不同,除含有GDP变量的信息之外,共同因子还包含固定资产投资增长率和工业总产值增长率等变量的信息。而且,中国经济也于2003—2004年间曾经出现过经济局部过热的现象(陈佳贵,2004)<sup>①</sup>。另外,因采用混频DFM估计共同因子,本文构建的一致指数 $F_t$ 也与郑挺国和王霞(2013)的一致指数所识别的结果非常接近。

## 五、结论

众所周知,国家统计局已公布了完备的GDP季度同比数据,然而,季度数据的时滞性制约了对经济运行状况的及时反映和宏观调控政策的准确选择。并且,同比数据的特殊性也妨碍了环比数据模型的直接应用。为此,本文构建了一种估算月度GDP同比增长率的同比数据混频近似因子模型,并提出了一种估计该模型的EM算法。实证研究发现,本文估算的月度GDP增长率与原季度GDP增长率具有基本一致的趋势循环特征;月度数据的局部极差大于季度数据,尤其在宏观经济经历严重外部冲击时期,月度和季度GDP增长率数据相差较大,充分说明了季度GDP平滑了经济的波动性,低估了外部冲击效应。另外,本文根据月度GDP增长率数据建立Markov转换自回归模型确定了我国经济周期的转折点、识别出了我国自1992年1月起经历的3个衰退期;经过与其他3个一致指标以及实际经济活动的对比分析发现,月度GDP增长率数据能够更加精确、及时地确定经济周期的转折点。

## 参考文献

- [1] Mariano B S, Murasawa Y. A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series [J]. Journal of Applied Econometrics, 2003(18): 427—443.
- [2] Guay R, et al. Estimation du PIB mensuel canadien: 1962—1985, L'Actualité économique [J]. Revue d'analyse économique, 1990, 66(1): 14—30.

① 陈佳贵,抑制局部经济过热的苗头,人民日报,2004年2月24日第九版。



- [3] Bernanke B S, Gertler M, Watson M. Systematic monetary policy and the effects of oil price shocks [J]. Brookings Papers on Economic Activity, 1997(1): 91-157.
- [4] Cuhe N A, Hess M K. Estimating monthly GDP in a general Kalman filter framework: evidence from Switzerland [J]. Economic & Financial Modelling, 2000(7): 153-194.
- [5] Schumacher C, Breitung J. Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data [J]. International Journal of Forecasting, 2008, 24(3): 386-398.
- [6] Mariano B S, Murasawa Y. A coincident index, common factors, and monthly real GDP [J]. Oxford bulletin of economics and statistics, 2010, 71(1): 27-45.
- [7] Mitchell J, Solomou S, Weale S. Monthly GDP estimation for inter-war Britain [J]. Explorations in Economic History, 2012(11): 543-556.
- [8] Engle R F, and Watson M W. Alternative Algorithms for Estimation of Dynamic MIMIC, Factor, and Time Varying Coefficient Regression Models [J]. Journal of Econometrics, 1983(23): 385-400.
- [9] Stock J H, Watson M W. New Indexes of Coincident and Leading Economic Indicators [C]. NBER Macroeconomics Annual, 1989: 351-393.
- [10] Sargent T J. Two Models of Measurements and the Investment Accelerator [J]. Journal of Political Economy, 1989(97): 251-287.
- [11] 郑挺国, 王霞. 中国经济周期的混频数据测度及实时分析, 经济研究, 2013(6): 58-70.
- [12] Dempster A, Laird M, Rubin D. Maximum likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society, Series B, 1977, 39(1), 1-38.
- [13] Stock J H, Watson M W. Macroeconomic Forecasting Using Diffusion Indexes [J]. Journal of Business and Economic Statistics, 2002(20): 147-162.
- [14] Bai J, Ng S. Determining the Number of Factors in Approximate Factor Models [J]. Econometrica, 2002(70): 191-221.
- [15] 叶光. 基于混合数据的一致指数构建与经济波动分析 [J]. 统计研究 2015(8): 17-26.
- [16] Burns A F, Mitchell W C. Measuring business cycles [J]. NBER Books, 1946.
- [17] Hamilton J D. A new approach to the economic analysis of nonstationary time series and the business cycle [J]. Econometrica, 1989(57): 357-384.

#### 作者简介

高华川 男 2016年毕业于南开大学经济学院数量经济研究所 获经济学博士学位 现为天津财经大学统计系讲师。研究方向为计量经济学理论及应用。

白仲林 男 2005年毕业于南开大学经济学院国际经济研究所 获经济学博士学位 2007年吉林大学数量经济研究中心博士后出站 现为天津财经大学统计系教授 博士生导师 中国数量经济学会常务理事。研究方向为数量经济学理论方法及应用研究。

(责任编辑: 方原)

## 《统计研究》“统计图版式”要求

### 基本要求

1. 图的宽度大致以16开刊物的分栏宽度为准,长宽比例以美观适度为宜。
2. 图不取外边框。
3. 图内外文字或数字取6-8号为宜,宋体。
4. 图底色为白色。
5. 坐标刻度单位如为年份,请用“1997”,勿用“1997年”或“97”或“97年”;不得竖排或斜排。如年份较多,可以按间隔一年或数年的方式排出。
6. 坐标刻度单位如为变量值,一般不宜带有小数位,如“10000.0”;最多为5位数字,如需用6位数字标示,如“100000”,请将其变换为“10”,并改变相应的计量单位。图中标志值一般最多带两位小数,在特殊情况下,如整数位为0时,最多可带4位小数。
7. 图标题设在图外的图下方,小五号黑体居中,不得设在图内。
8. 图的标注放在图和图标题之间,以图外框左边界为准前空两格,小五号宋体。
9. 图的序号用“图1”表示,勿用“图1.”、“图一”、“图1-1”等,文中只有1张图时,可用“图1”,亦可用“图”。