

目录结构

- [一、引言](#)
- [二、MIDAS模型简介](#)
- (一) 基础MIDAS模型及其设定
- (二) MIDAS模型中的设定问题
- (三) MIDAS模型的扩展
- [三、中国宏观经济中运...](#)

文内图表

- 图1 四种不同形态的指数Almon权重函数
- 图2 中国混频数据图形
- 表1 MIDAS模型有效性的蒙特卡洛模拟
- 图3 混频数据参数估计的有效性对比分析(快速递减权重的情形(1))
- 表2 混频数据模型在中国宏观经济波动模型的有效性估计结果
- 图4 混频数据模型估计的权重
- 图5 混频数据和其他模型预测结果的比较
- [参考文献](#)

中国宏观经济混频数据模型应用——基于MIDAS模型的实证研究

刘金全 刘汉 印重

吉林大学数量经济研究中心

导出/参考文献

关注

分享

收藏

摘要: 传统宏观计量模型需要利用加总或插值等方法将混频数据统一到同频数据再应用于宏观经济模型。混频数据模型是直接利用混频数据构建模型, 避免了因数据加总或插值导致的信息损失和人为信息的虚增, 充分利用现有高频数据的信息, 改进了宏观计量模型估计的有效性和预测的精度。混频数据抽样模型(MIDAS)是混频数据模型的一种, 它使用参数控制的滞后权重多项式函数对高频滞后数据进行有权重的加总并构建模型, 再通过参数估计非线性的方法估计混频数据模型中的最优参数。MIDAS模型是攫取现有高频数据的全样本信息用于宏观经济模型的分析与预测的有效方法, 本文基于该模型的实证研究探寻混频数据在中国宏观经济应用中的有效性。

关键词: 混频数据; MIDAS模型; 宏观经济; 有效性;

作者简介: 刘金全, 吉林大学数量经济研究中心教授, 经济学博士, 博士研究生导师;

作者简介: 刘汉, 吉林大学数量经济学专业博士研究生;

作者简介: 印重, 吉林大学数量经济学专业博士研究生。

基金: 国家自然科学基金项目“非线性随机波动模型估计方法及应用研究”(70971055); 教育部人文社会科学研究2008年度重大项目“我国经济周期波动态势与宏观经济总量内在关联机制的动态计量研究”(08JJD790133); 学“211工程”和“985工程”建设项目资助;

一、引言

宏观经济中有诸多能反映当前宏观经济状态和未来宏观经济走势的经济数据, 如季度GDP数据、月度CPI数据、金融市场收益的日数据、股票市场波动的日内数据等等。这些数据受到经济个体、企业、组织和国家、国际社会的广泛关注, 人们试图使用不同的数据处理方法和构建各种模型从这些纷繁复杂的数据中攫取信息, 得出当期宏观经济的准确预报和未来一段时间内宏观经济走势的精确预测。然而, 在构建宏观计量模型时却经常遇到抽样的频率高低有别的问题, 而大多数宏观计量模型都要求模型等式两边的数据频率一致, 因此要想利用宏观计量模型去估计、预报和预测宏观经济就必须对混频数据进行处理, 有的采用加总或替代的方法将高频数据转换为低频数据(Silvestrini和Veredas, 2008), 有的采用插值法将低频数据处理为高频数据(Chow和Lin, 1971; 1973; 进文和薛艳, 2009), 但是这两种方法经常受到质疑, 人们认为加总或替代法在数据处理过程中忽视了高频数据的波动, 抹杀了高频数据的波动, 在一定程度上人为地减少了样本信息, 但是由于该方法的简便易行, 在实际应用中基本上都采用这种方法; 而插值法在应用上相对较少, 但是对该方法的研究却相当多。人们普遍认为插值法虽然能较好地处理混频数据, 但是这种高频数据一定程度上有人为构造的嫌疑, 因此使用这种高频数据建模时都非常谨慎, 这也是插值法在实际应用中使用较少的重要原因。

混频数据模型不对混频数据做任何处理, 而是充分利用原始数据的信息构建混频数据模型。当前处理混频数据模型主要有混合数据抽样(MIDAS, MIxed DAta Sampling)模型和混频向量自回归移动平均(VARMA)模型。MIDAS模型是吉塞尔斯等人(Ghysels, Santa-Clara和Valkanov, 2004)在分布滞后模型的基础上提出来的; VARMA模型是马塞利诺等人(Marcellino和Schumacher, 2007)提出的。MIDAS模型的提出的初衷是想利用混频数据模型从高频金融数据中攫取信息来预测金融市场的波动(MIDAS模型方面的应用如: Ghysels, Santa-Clara和Valkanov, 2005; 2006; Forsberg和Ghysels, 2007; Chen和Ghysels, 2008等等)。克莱蒙茨和加维奥(Clements和Galvão, 2005)开始将MIDAS模型应用于宏观经济领域, 马赛利诺和马赫(Marcellino和Schumacher, 2007)将因子模型引入到MIDAS模型, 模型的预测结果显示MIDAS模型在很多实际预测中具有最佳的预测效果。克莱蒙茨和加维奥(Clements和Galvão, 2008)使用MIDAS-AR模型并运用1959年到2005年的工业产出、就业率和产能利用率这三个月度指标预测了美国的GDP增长率。

产出增长率,结果显示MIDAS模型在宏观经济总量的短期预测方面具有比较优势。霍格雷夫 (Hogrefe, 2008) 单频,混频、邹和林 (Chow和Lin1971) 的插值法在预测美国GDP的修订数据中的优劣,研究结果认为MIDAS样本外预测中表现最好。吉塞尔斯和赖特 (Ghysels和Wright 2009) 利用MIDAS和卡尔曼滤波,利用高频日度对实际GDP增长率、通货膨胀率、短期国库券和失业率的预测,其结果表明MIDAS比简单的随机基准预测更佳。莱蒙茨和加维奥 (Clements和Galvão, 2009) 使用多种先行指数的多变量MIDAS模型来预测美国产出,结果显示MIDAS具有比较优势,预测结果比AR模型要好。另一种处理混频数据的模型是由扎德罗兹尼 (Zadrozny, 1990) 的用于连续混频时间序列的VARMA模型,随后扩展到离散混频时间序列的VARMA (Zadrozny, 1990;2008) 的基本思想将低频数据看作是有循环缺省值的高频数据(如季度数据看可以看作是该季末月度数据的值,如2010年第一季度数据可以看作是2010年3月份的数据,其它月份的数据视作缺省值),然后运用卡尔曼滤波方法估计具有间断形式的混频VARMA模型和缺省的高频数据。扎德罗兹尼 (Zadrozny, 2008) 的研究结果显示使用状态空间VARMA模型预测美国的GNP比使用自回归模型要有更好的预测效果。

本文首先模拟不同样本长度、滞后阶数和各种高频数据形式的MIDAS模型,并使用蒙特卡洛模拟探寻影响模型有效性的主要因素。然后,通过比较传统同频数据模型(加总法和插值法)和MIDAS混频数据模型在中国经济分析和预测方面的优劣,分析混频数据模型在我国宏观经济中的有效性⁽¹⁾,模型估计结果显示混频数据模型的估计和方差都要比传统的同频数据要好,初步说明MIDAS模型在中国宏观经济中的应用是有效的,值得进一步研究和挖掘。

二、MIDAS模型简介

吉塞尔斯等人 (Ghysels, Santa-Clara和Valkanov, 2004) 提出的MIDAS混频数据模型源于分布滞后模型,主要应用于金融市场波动和预测方面,目前该方法在预报当前宏观经济状态,分析和预测未来宏观经济走势等方面有广泛的应用。

(一) 基础MIDAS模型及其设定

为了更加清晰理解MIDAS的构建过程,本文首先给出分布滞后模型,其表示方法如下:

$$y_t = \beta_0 + B(L)x_t + \varepsilon_t \quad (1)$$

其中, $B(L)$ 是由有限或者无限滞后多项式算子,模型假设数据具有相同的频率。

MIDAS模型不是严格意义上的分布滞后模型,其最显著的特点是能够处理混频数据,并能获得优于分布滞后模型的参数估计结果 (Ghysels, Santa-Clara和Valkanov, 2004)。假定在MIDAS回归模型中,等式左边的低频数据($t=1, \dots, T$),等式右边的高频数据为 x_{τ} ($\tau=1, \dots, mT$),令 $x^{(m)} = x_T$,其中 m 表示混频数据的频率倍差,即 $x^{(m)}$ 在期进行了 m 次抽样。则MIDAS回归模型即可表示为如下的形式:

$$y_t = \beta_0 + \beta_1 B(L^{1/m}; \theta) x_t^{(m)} + \varepsilon_t^{(m)} \quad (2)$$

其中,滞后算子多项式 $B(L^{1/m}; \theta) = \sum_{k=0}^K B_k(\theta) L^{k/m}$ 是参数向量 θ 的一个函数, $L^{1/m}$ 是高频数据的滞后算子, $L^{1/m}x_t^{(m)} = x_{t-1/m}^{(m)}$, K 是高频数据的滞后阶数。

假设宏观经济中, y_t 是季度数据序列, $x_t^{(m)}$ 是与 y_t 在同一样本期间内抽样 m 次的高频数据,若 $m=3$,则 $x_t^{(m)}$ 说就是一个月度数据序列,若 y_t 为2010年第一季度的数据,则 $x_t^{(3)}$ 表示2010年3月的数据, $x_{t-1/3}^{(3)}$ 表示2010年2月的数据, $x_{t-1}^{(3)}$ 表示2009年12月的数据。一个滞后阶数 $K=12$ 的MIDAS宏观经济模型为:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 B(L^{1/3}; \theta) x_t^{(3)} + \varepsilon_t^{(3)} \\ &= \beta_0 + \beta_1 [B(0; \theta) x_t^{(3)} + B(1; \theta) x_{t-1/3}^{(3)} + \dots + B(12; \theta) x_{t-12/3}^{(3)}] + \varepsilon_t^{(3)} \end{aligned} \quad (3)$$

(二) MIDAS模型中的设定问题

MIDAS模型估计中关键问题是权重函数 $B(k; \theta)$ 中的参数向量 θ 和滞后阶数 K 的选取,这涉及权重函数的选取,给出应用于宏观经济中的三种常用的有限和无限权重多项式函数 $B(k; \theta)$ 。

第一种是Almon多项式函数,其基本形式为:

第二种为Almon多项式函数的变形,称为指数Almon多项式函数,它是目前使用最多的一种多项式函数形式构造各种不同的权重函数(如图1所示),它能够保证权重数为正数,且能使方程获得零逼近误差的良好性质(和Valkanov, 2006),具体形式为:

$$B(k; \theta) = \frac{\exp(\theta_0 + \theta_1 k + \theta_2 k^2 + \cdots + \theta_p k^p)}{\sum_{k=1}^K \exp(\theta_0 + \theta_1 k + \theta_2 k^2 + \cdots + \theta_p k^p)} \quad (5)$$

第三种为 β 多项式函数,该多项式函数是仅带有两个参数的 β 多项式函数,它同样也能构造多种形态的权重函数的具体形式可以表示为:

$$B(k; \theta_1, \theta_2) = \frac{f(k/K, \theta_1; \theta_2)}{\sum_{k=1}^K f(k/K, \theta_1; \theta_2)} \quad (6)$$

其中,

$$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \quad (7)$$

$$\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx \quad (8)$$

MIDAS模型中,上述三种权重函数均能保证高频滞后阶数的权重函数为正,且上述权重的函数定义中暗含之和为1的假设。第一个多项式函数在金融市场波动的预测和分析中使用较多,第二和第三个多项式函数在经济的分析与预测中应用比较多,MIDAS模型应用于宏观经济,其多项式函数的形式常常选取两参数的指数Almon函数,吉塞尔斯等人(Ghysels, Saino和Valkanov, 2007)的表明指数Almon多项式函数能够构造丰富的滞后阶数(如图1所示),但在宏观经济模型的构造过程常常约束指数Almon多项式函数的参数 $\theta_1 \leq 300$, $\theta_2 < 0$ (Cle Galvão, 2005),以保证滞后高频数据对应的权重函数在宏观经济应用中是递减的。

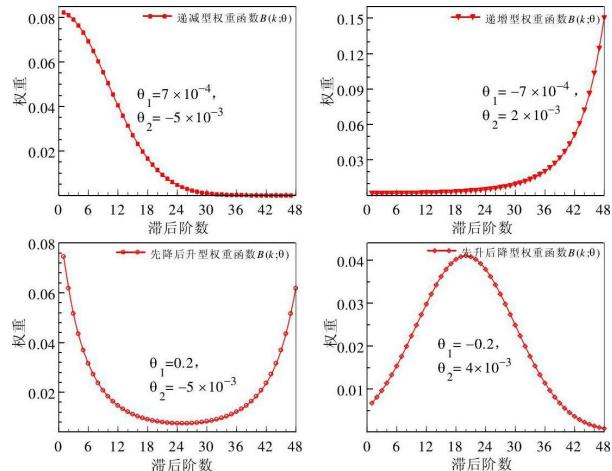


图1 四种不同形态的指数Almon权重函数

[下载原图](#)

MIDAS模型设定的另一个关键问题是高频数据滞后多项式的选择问题,这涉及在模型的估计和预测中使用间的高频数据来预测低频数据⁽²⁾。吉塞尔斯等人(Ghysels, Santa-Clara和Valkanov, 2004)认为通过非线性法优化MIDAS回归方程中权重函数中的参数向量 θ ,得出的参数 $\hat{\theta}$ 所绘出的权重函数图形中可以获得滞后阶数优长度⁽¹⁾,这样确定的滞后阶数完全是数据驱动的,所以是最优的。然而在宏观经济的实际运用中,滞后阶数具有任意性和经验性的特点,也有部分的学者使用AIC和BIC准则作为滞后阶数的选择标准。

(三) MIDAS模型的扩展

MIDAS模型自吉塞尔斯等人 (Ghysels, Santa-Clara和Valkanov, 2004) 提出以来, 发展相当迅速, 具有扩展性。目前该模型已经结合AR模型、非对称模型、因子模型、状态空间模型, 以及半参数和非参数模型, 几种扩展形式做简单的介绍。

1、MIDAS-AR模型

MIDAS法的一个基本拓展是在模型中加入自回归因子, 吉塞尔斯等人 (Ghysels, Santa-Clara和Valkanov) 认为在MIDAS模型中引入滞后被解释变量将导致模型有效性的损失, 只有当解释变量中包含季节因素时才能(莱蒙茨和加维奥 (Clements和Galvão, 2005) 提出将动态自回归作为一般因素引入模型解决了这个问题, 则自回归的h步向前预测模型可以表示为:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/m}; \theta)(1 - \lambda L)x_{t-h}^{(m)} + \varepsilon_t^{(m)} \quad (9)$$

该模型的估计可以先通过估计方程 (2) 得到标准MIDAS方程的残差, 然后估计 λ
 $\hat{\lambda}_0 = \left(\sum \hat{\varepsilon}_{t-h}^2 \right)^{-1} \sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-h}$ 。接着
 $y_t^* = y_t - \hat{\lambda}_0 y_{t-h}$ 和 $x_{t-h}^* = x_{t-h} - \hat{\lambda}_0 x_{t-2h}$, 并使用非线性最小方法的模型, 即可得到模型的参数估计。

$$y_t^* = \beta_0 + \beta_1 B(L^{1/m}; \theta)x_{t-h}^* + \varepsilon_t \quad (10)$$

2、非对称MIDAS模型

吉塞尔斯等人 (Ghysels, Santa-Clara和Valkanov, 2005) 引入了非对称MIDAS模型, 模型形式如下:

$$y_t = \beta_0 + \beta_1 \left(\phi B(L^{1/m}; \theta^+) I_{t-1}^+ x_{t-1}^{(m)} + (2 - \phi) B(L^{1/m}; \theta^-) I_{t-1}^- x_{t-1}^{(m)} \right) + \varepsilon_t^{(m)} \quad (11)$$

其中 I_{t-1}^i ($i=\{+, -\}$) 是一个指标函数, 分别当 $x_{t-1}>0$ 或 $x_{t-1}<0$ 时其值都为1。该模型度量正负x冲击对y的影响。参数 ϕ 的取值区间为[0, 2]保证了权重和为1。通过估计参数 ϕ 即可获得高频数据中正负冲击对低频数据的

3、半参数或非参数MIDAS模型

半参数或非参数MIDAS模型是陈和吉塞尔斯 (Chen和Ghysels, 2008) 基于林顿和马门 (Lin和M 2005) 的半参数ARCH (∞) 提出来的, 主要应用于高频金融时间序列的混频建模, 基本形式如下:

$$y_{t+1} = \beta_0 + \sum_{i=1}^K \sum_{j=1}^L B_{ij} \left(L^{1/M_i, \theta} m(x_t^{(m)}) \right) + \varepsilon_t \quad (12)$$

其中 $m(\cdot)$ 是一个未知函数, 陈和吉塞尔斯 (Chen和Ghysels, 2008) 提供了该模型的估计方法, 并通过实该模型的有效性优于同频数据模型。

除此之外, 该模型在经济金融领域应用还结合了因子模型、状态空间模型等, 这里就不再赘述。

二、MIDAS方法在我国宏观经济中的实证分析

MIDAS模型在宏观经济和金融中运用都无一例外地显示了其在利用高频数据建模方面的比较优势, 然而M型在中国股票市场波动的实证研究中, 徐剑刚、张晓蓉、唐国兴 (2007) 使用MIDAS模型和ABDL模型 (安德森 (Andersen, Bollerslev, Diebold和Labys, 2003) 采用分数自回归 (ARFI) 模型预测未来波动的模型) 并利用和深成指的5分钟数据来预测股票市场的波动率, 其实证结果表明ABDL模型在预测波动方面优于MIDAS模型。吉塞尔斯等人 (Ghysels, Santa-Clara和Valkanov, 2006) 对美国股票市场波动的分析认为MIDAS模型优于AR模型相矛盾。孔等人 (Kong, liu和wang, 2008) 使用MIDAS和GARCH方法研究中国股票市场波动过程, 其结果表明GARCH模型比MIDAS模型更好地预测可实现方差中的极值波动, 而MIDAS模型在预测条件方差方面比GARCH模型要好。鉴于MIDAS模型在中国金融市场波动方面并不完全具有比较优势, 因此下文将通过MIDAS的蒙特卡洛模拟的实证研究初步探讨MIDAS模型在中国宏观经济方面的应用有效性。

(一) 中国宏观经济中的混频数据

在分析MIDAS模型在中国宏观经济应用有效性方面涉及到中国混频数据的使用,本文使用1992年第1季度第4季的GDP增长率的同比变化率 y_t 和1992年1月到2009年12月的月度通货膨胀率的同比变化率 x_t ⁽³⁾ MIDAS模型,混频数据的图形如图2所示,从图中可以看出中国宏观经济数据多是混频数据,在年度数据的区间有大量的季度和月度数据,如果仅使用低频数据建模,就会忽略原始数据的信息,方程的参数估计准确性和模型精确度就会降低,甚至会出现偏误,因此如何充分利用高频数据的信息用于宏观经济分析和预测是值得分析的。

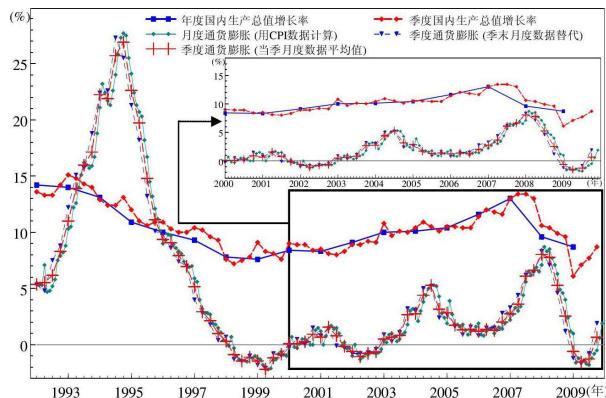


图2 中国混频数据图形

[下载原图](#)

(二) MIDAS模型有效性的蒙特卡洛模拟

本文参考安德烈乌等人 (Andreou, Ghysels和Andreou, 2010) 的蒙特卡洛方法构建服从正态分布 (具有平稳性的序列)、自回归过程 (具有自相关性特征的序列) 和自回归条件异过程 (具有波动聚类特征的序列) 三种不同高频率数据,并设定不同的样本长度、权重函数形式、滞后阶数和信噪比程度来度量MIDAS模型在不同情况下对间序列数据估计的有效性。

1、蒙特卡洛模拟的数据生成

首先,构造三种不同的高频率时间序列数据用于MIDAS模型有效性的模拟。

第一种:构造服从正态分布的平稳高频率时间序列数据,即 $x_t^{(m)} \sim i.i.d.N(0, 1)$;

第二种:构造服从自回归过程的高频率时间序列数据,即 $x_t^{(m)} = \sqrt{(\sigma_t^{(m)})^2} e_t^{(m)}$,
 $e_t^{(m)} \sim i.i.d.N(0, 1)$, $(\sigma_t^{(m)})^2 = c_0 + c_1 (\tilde{x}_{t-1}^{(m)})^2$,
并选取 $c_0=0.25$, $c_1=0.85$ 。

第三种:构造服从自回归条件异方差的高频率时间序列数据,即 $x_t^{(m)} = c + \phi \tilde{x}_{t-1}^{(m)} + e$
其中 $e^{(m)} \sim i.i.d.N(0, 1)$, $\tilde{x}_{t-1}^{(m)} \sim i.i.d.N(0, 1)$, 并选取 $c=0.25$, $\phi=0.85$ 。

其次,使用两参数的指数Almon权重构造低频数据。本文约束 $\theta_0 \leq 300$, $\theta_1 < 0$, 即以构建符合宏观经济实型权重函数,并在此约束的基础上构建快速衰减权重函数 ($\theta_0=7 \times 10^{-4}$, $\theta_1=-5 \times 10^{-2}$) 和缓慢衰减权重函数 ($\theta_0=-5 \times 10^{-4}$),并依据 $y_t = \beta_0 + \beta_1 B(L^{1/m}; \theta) x_t^{(m)} + u_t$ 构建低频数据,其中 $B(L^{1/m}; \theta)$ 表示两参数的指数Almon权 $x_t^{(m)}$ 是上述三种不同形态的高频率数据, $u_t \sim i.i.d.N(0, 0.125)$ 。本文将参数 β_0 和 β_1 初始值分别选取 $\beta_0^0=0.5$, $\beta_1^0=2$ 和 $\beta_1^0=4$,其中上述 β 的初始值分别表示低、中和高信噪比 (SNR, Signal to Noise Ratios)。

2、蒙特卡洛模拟的数据

我们选择样本长度T=200、T=500和T=1000的低频数据,选择频率倍差m=5、m=50和m=100(为了模拟起见,本文选择与频率倍差相等的滞后阶数)分别构建MIDAS模型的数据,再使用NLS(非线性最小二乘估计型中的参数 β 进行估计,同时将MIDAS模型中的高频率数据通过等权重处理为低频数据,再使用OLS(普通最小二乘估计同频数据中的参数 β)。最后,本文对上述估计进行1000次的模拟⁽¹⁾,得出两种方法的参数估计值 $\hat{\beta}_1$ 和 β_1^0 的误差平方和MSE,并使用RMSE=MSE_{LS}/MSE_{NLS}度量MIDAS模型在不同的情形下有效性,若RMSE

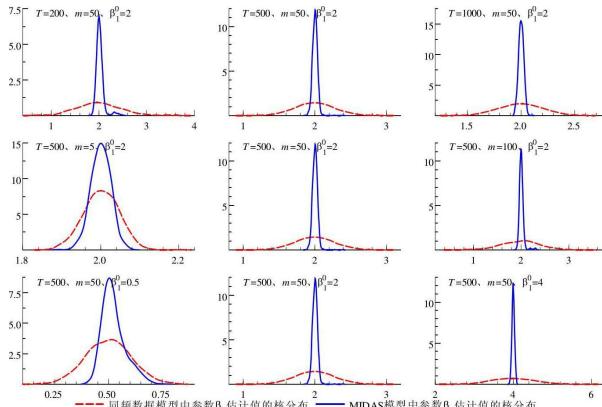
MIDAS模型估计 β 的误差相对于同频数据估计的误差就越小,说明MIDAS模型估计有效。表1给出了在不同条件下蒙特卡洛模拟得出的RMSE,具体情况如表1所示。

从蒙特卡洛模拟的结果可以看出MIDAS模型的有效性受以下几个方面因素的影响,首先,MIDAS模型的结果受到样本长度T、滞后阶数m,一般来说样本长度越长、滞后阶数越长,表中的RMSE就越大⁽¹⁾, MIDAS模型的有效性就更加突出;其次,本文在构建MIDAS模型时设定了三种不同信噪比,从表中可以看出信噪比和RMSE就越大,使用MIDAS模型估计就非常有必要(从图3中第三行的三个图可以直观看出,其他情况一致的信噪比程度越高,估计结果的分布就越集中于参数的真实值);第三,通过比较三种形态的时间序列数据我们可知MIDAS模型在估计服从AR(1)的高频数据中最为有效,随后是ARCH(1),估计服从独立同分布的正态高频率优势不是很明显,尤其是在样本长度短、信噪比低的时候,MIDAS模型的优势体现不出来。最后,从表1上部分的对比可以看出,MIDAS模型在快速递减权重的模型中具有非常大的优越性(从图3也可以看出),而在缓慢递减的模型中,MIDAS模型的优越性并不突出,甚至在滞后阶数较低的情况下,其估计比OLS估计的同频数据模型还要差。

表1 MIDAS模型有效性的蒙特卡洛模拟

[下载原表](#)

频率偏差(m) β_1	T	5			50			100		
		0.5	2	4	0.5	2	4	0.5	2	4
快速递减的权重函数										
I.I.D.	200	1.0794	3.0444	8.1807	2.8848	28.1280	188.8182	9.4533	54.0285	183.0625
	500	1.0973	2.9014	8.4184	2.9110	58.6462	249.1517	8.7971	70.3041	372.3593
	1000	1.1124	2.9676	7.8427	3.1418	69.4206	264.0422	8.8933	101.6384	556.7430
AR(1)	200	1.2062	4.2544	14.4643	4.0553	52.6957	454.0431	12.8193	82.8420	219.9291
	500	1.2653	4.2560	14.7205	3.5456	128.9388	473.2316	11.8205	136.6199	381.1640
	1000	1.1949	3.8674	14.3213	4.9834	118.2634	458.7969	11.0664	173.5934	991.2074
ARCH(1)	200	1.2652	4.6446	14.2465	3.3393	29.1616	291.7100	9.5441	53.2860	130.6855
	500	1.2057	4.1686	15.6477	3.0445	69.3719	306.2574	9.7111	105.2313	421.0620
	1000	1.2048	4.8423	15.7449	2.8806	67.9944	326.6478	9.7448	106.9783	585.3597
缓慢递减的权重函数										
I.I.D.	200	0.9897	0.9869	0.9821	0.9166	1.1397	1.5522	1.1780	1.6016	3.6510
	500	0.9847	0.9886	1.0013	0.9938	1.1280	1.5372	1.1721	1.6285	3.6301
	1000	0.9961	0.9968	0.9981	1.0012	1.1514	1.5839	1.1823	1.8338	3.4213
AR(1)	200	0.9885	0.9954	0.9896	0.9526	1.2579	2.0212	1.1828	2.4023	5.0782
	500	0.9955	1.0045	0.9943	1.0273	1.2536	2.0077	1.2626	2.1703	5.3891
	1000	1.0008	0.9966	1.0026	1.0070	1.2435	1.9452	1.2090	2.4082	5.4884
ARCH(1)	200	0.9686	0.9966	0.9785	0.9279	1.1511	1.6418	1.1200	1.6507	3.7289
	500	0.9920	0.9983	0.9882	0.9873	1.1601	1.7166	1.1429	1.7750	3.4498
	1000	0.9968	0.9974	1.0027	0.9803	1.2113	1.6501	1.2203	1.9093	3.8129

图3 混频数据参数估计的有效性对比分析(快速递减权重的情形⁽¹⁾)[下载原图](#)

(三) MIDAS模型在中国宏观经济应用中的有效性分析

本文首先使用1992年第1季度到2008年第4季的GDP增长率的条件方差 h_{gt} 和1992年1月到2008年12月的货膨胀率的条件方差 h_{mt} ⁽³⁾来构建MIDAS混频数据模型⁽²⁾;其次,将混频数据处理为低频的同频数据,即将上月的月度通货膨胀率处理为季度通货膨胀率⁽³⁾,并构建回归模型;同时,使用邹和林(Chow和Lin, 1971; 1976)将低频的季度GDP数据处理为高频的月度GDP数据,并构建回归模型;再次,本文对上述模型分别进行估计,将估计值和模型的残差平方和来判定各种模型的优劣,进而确定MIDAS模型在中国宏观经济应用中的可行性;最后利用2009年1月到2010年6月的月度通货膨胀率条件方差的数据,以及混频数据模型的估计结果来预测2009度到2010年第2季的GDP增长率的条件方差,并与同频数据模型的预测结果和GARCH模型度量出的原始GD的条件方差进行比较分析,进而确定MIDAS模型在中国宏观经济应用中的有效性。构建中国宏观经济波动的模型⁽⁴⁾如下:

$$h_{g_t} = \beta_0 + \beta_1 B(L^{1/3}; \theta) h_{\pi_t^{(3)}} + \varepsilon_t^{(3)} \quad (12)$$

其中, 权重函数B($k; \theta$) 分别选择两参数的指数Almon权重函数和 β 多项式权重函数, 并分别对约束的和非约束函数进行估计。本文的实证研究中滞后阶数的选取方面, 首先参考AIC和BIC准则, 然后实验了克莱蒙茨(Clements和Galvão, 2008) 的在对美国GDP预测的实证研究选择滞后阶数为24, 最后根据中国实际情况最终滞后阶数为 $k=12$, 具体估计结果如下:

表2 混频数据模型在中国宏观经济波动模型的有效性估计结果 [下载原表](#)

参数	MIDAS 混频数据模型				同频数据的回归模型			同频数据的ADL模型		
	指数 Almon 权重		β 多项式权重		同低频		同高频		同低频	
	约束	无约束	约束	无约束	平均	替代	插值	平均	替代	插值
$\hat{\beta}_0$	1.1061	1.1068	1.1058	1.1061	1.0944	1.0567	1.0976	1.1150	1.0612	1.1033
$\hat{\beta}_1$	0.1226	0.1224	0.1232	0.1226	0.1943	0.2593	0.1887	0.1353	0.2280	0.1718
$\hat{\theta}_1$	[-100 300]	10.0000	[0 + ∞]	1.0295	—	—	—	—	—	—
$\hat{\theta}_2$	[-∞ 0]	-3.1561	[-∞ 300]	8.0352	—	—	—	—	—	—
σ_e	0.2963	0.2968	0.2966	0.2969	0.3584	0.3539	0.3747	0.3369	0.3353	0.3705

上述混频数据模型和同频数据模型的估计结果可以看出:

第一, 月度通货膨胀率的条件方差 $h_{\pi t}^{(3)}$ 与季度GDP增长率的条件方差 g_t 之间存在正向关系, 说明自20世纪以来, 较高的通货膨胀率波动是引起GDP增长率出现波动的重要因素, 这说明要想实现季度GDP增长率的“滑”(刘树成等人, 2005) 就必须对通货膨胀的波动进行控制, 若想经济在金融危机后有所转机(即经济增长启动), 则需要有适度的通货膨胀率来刺激, 但同时也应该警惕较高通货膨胀所带来的GDP增长率的大幅波动。

第二, 从模型的估计结果看, MIDAS混频数据模型中参数 β 的估计值 $\hat{\beta}_1$ 要明显比同频数据模型的估计值 $\hat{\beta}_1$ 稳定在12%左右, 且模型的残差比同频数据模型的残差要小, 这说明MIDAS混频数据模型比同频(无论是同低频还是同高频都是如此) 数据模型更为有效地利用了样本信息, 得出了更加精确的估计结果, 也说明混频数据模型估计数比普通的同频数据模型更好拟合我国宏观经济波动模型;同时, 在MIDAS混频数据模型中, 无论是指数Almon权数, 还是 β 多项式权重的MIDAS模型的估计结果都显示约束权重估计参数 β 的估计值 $\hat{\beta}$ 也比无约束的权重估计结果要大且残差要小, 这说明本文在对参数的约束的提高了估计结果的精度。

第三, 从混频数据模型估计结果中的权重函数可以看出(如图4所示), 多项式权重函数在滞后1~4阶时权重很大, 尤其是滞后1阶和滞后2阶的权重尤其大, 这说明通货膨胀率的波动对GDP增长率波动的影响不具有长期说明模型估计中的滞后阶数可以选取更短一些。从混频数据模型和同频数据模型模型预测我国2009年至今的增长率的结果可以看出(如图5所示), 混频数据比同频数据的估计结果更为接近真实值, 证实混频数据模型在预测比传统的同频数据更为有效。

总之, 从MIDAS混频模型和同频模型估计和预测中国宏观经济波动的结果可以看出, 混频数据模型具有数据模型更为优秀的估计参数, 模型估计的残差波动相对较小, 模型的样本内预测更加接近真实值, 说明MIDAS综合、估计和样本内预测方面都有较大优势, MIDAS模型在中国宏观经济应用中的有效性得到了证实。

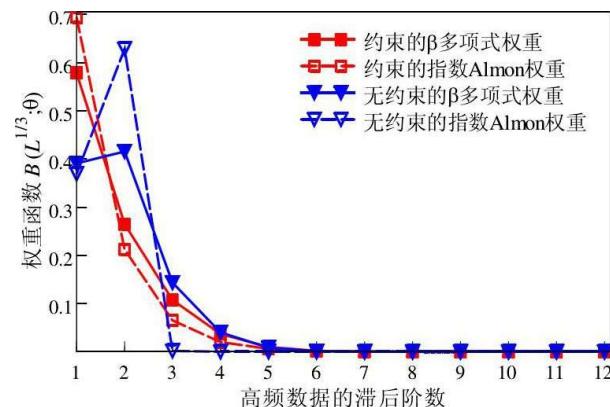


图4 混频数据模型估计的权重 [下载原图](#)

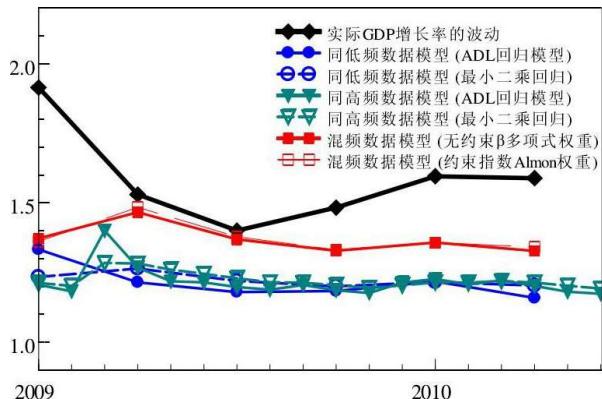


图5 混频数据和其他模型预测结果的比较

[下载原图](#)

三、中国宏观经济中运用混频数据的经验结论与展望

本文简单介绍了混频数据的模型，并从模型的基础理论、模型的扩展和模型的应用重点介绍了MIDAS模型，在此基础上使用蒙特卡洛模拟对MIDAS模型在各种情形下的有效性进行了详细的分析，最后结合我国宏观动模型对MIDAS模型在我国宏观经济中运用的有效性的进行验证，由此本文得出了以下三点基本结论：

第一，MIDAS模型具有能直接使用不同频率数据构建模型，充分挖掘高频数据中有效样本信息的优越性。模型估计结果的有效性无论是在蒙特卡洛模拟中，还是在中国宏观经济的应用中都得到了非常好的体现，说明模型在中国宏观经济运用中的有效性的确存在。

第二，MIDAS模型的有效性的显著程度是有差别的。其有效性的显著程度主要受到样本长度、滞后阶数、以及序列本身所具有的内在特征的影响。一般来说，样本长度越大、滞后阶数越长、信噪比越强，MIDAS模型有效性就越显著。而且MIDAS模型在估计具有持续性和波动聚类性特征的数据序列时，其有效性比估计独立分位数为突出，说明MIDAS模型在非平稳时间序列数据的建模和估计上比平稳时间序列数据的建模和估计更为有效。

第三，MIDAS模型在中国宏观经济中的运用是有效的。无论是从攫取高频数据信息方面，模型估计的总方和、还是从模型的样本内预测上，MIDAS模型都显示出了的有效性都比使用同频数据模型显著，初现模型在我国宏观经济运用中的有效性。但是这种有效性还是比较微弱的，从蒙特卡罗模拟可以看出，这主要和宏观经济数据的样本长度和模型优化后的权重函数形式有关。

总之，本文认为在构建模型中如果遇到混频数据，应该考虑使用混频数据模型，而不应该通过加总或插值处理为同频数据来建模。当然，本文是对MIDAS模型有效性的初步探讨，MIDAS模型是否在我国宏观经济中具有有效性，这还需要我们构建更为复杂的混频数据模型，应用到宏观经济分析和预测中，让实证检验该模型的有效性。

参考文献

- [1].刘树成、张晓晶、张平:《实现经济周期波动在适度高位的平滑化》[J],《经济研究》2005年第11期。
- [2].刘金全、谢卫东:《中国经济增长与通货膨胀的动态相关性》[J],《世界经济》2003年第6期。
- [3].徐剑刚、张晓蓉、唐国兴:《混合数据抽样波动模型》[J],《数量经济技术经济研究》2007年第11期。
- [4].赵进文、薛艳:《我国分季度GDP估算方法的研究》[J],《统计研究》2009年第10期。
- [5].Andersen, T., Bollerslev, T., Diebold, F.X.and P.Labys, 2003, "Modeling and forecasting realized volatility," *Econometrica*, 71, 529-626.
- [6].Andreou, E., Ghysels E., and A.Kourtellos, 2010, "Regression models with mixed sampling frequencies"[J], *J Econometrics*, doi:10.1016/j.jeconom.2010.01.004.
- [7].Chen, X., and E.Ghysels, 2008, "News-good or bad and its impact on volatility predictions over multiple time horizons," *mimeo*, Chapel Hill, N.C.
- [8].Chow, G., and A.Lin, 1971, "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series"[J], *The Review of Economics and Statistics*, 53 (4), 372-375.
- [9].Chow, G., and A.Lin, 1976, "Best linear unbiased estimation of missing observations in an economic time series," *Journal of the American Statistical Association*, 71 (355), 719-721.
- [10].Clements, M.P., and A.B.Galvão, 2005, "Macroeconomic forecasting with mixed-frequency data:Forecasting U.S. growth administration", Warwick working paper.

- [11].Clements, M.P., and A.B.Galvão, 2008, "Macroeconomic forecasting with mixed-frequency data:Forecasting US output growth"[J], Journal of Business and Economic Statistics, 26 (4), "546-554.
- [12].Clements, M.P., and A.B.Galvão, 2009, "Forecasting US output growth using leading indicators:an appraisal of MIDAS models"[J], Journal of Applied Econometrics, 24 (7), 1187-1206.
- [13].Forsberg, L., and E.Ghysels, 2007, "Why do absolute returns predict volatility so well?"[J], Journal of Econometrics, 5 (1), 31-67.
- [14].Ghysels, E., A.Sinko, and R.Valkanov (2007, "MIDAS regressions:Further results and new directions"[J], Economic Reviews, 26 (1), 53-90.
- [15].Ghysels, E., and J.Wright, 2009, "Forecasting professional forecasters"[J], Journal of Business and Economic Statistics, 27 (4), 504-516.
- [16].Ghysels, E., and R.Valkanov, 2006, "Linear time series processes with mixed data sampling and MIDAS regressions models", mimeo, Chapel Hill, N.C.
- [17].Ghysels, E., Santa-Clara, P.and R.Valkanov, 2004, "The MIDAS touch:Mixed data sampling regressions"[J], Chapel Hill, N.C.
- [18].Ghysels, E., Santa-Clara, P.and R.Valkanov, 2005, "There is a risk-return trade-off after all"[J], Journal of Economics, 76 (3), 509-548.
- [19].Ghysels, E., Santa-Clara, P.and R.Valkanov, 2006, "Predicting volatility:How to get most out of returns data sampled at different frequencies"[J], Journal of Econometrics, 131, 59-95.
- [20].Hogrefe, J., 2008, "Forecasting data revisions of GDP:A mixed frequency approach"[J], Advances in Statistical Analysis, 92 (3), 271-296.
- [21].Kong, D., H.Liu, and L.Wang, 2008, "Is there a risk-return trade-off?Evidences from Chinese stock markets"[J], Journal of Economics in China, 3 (1), 1-23.
- [22].Linton, O., and E.Mammen, 2005, "Estimating semiparametric models by kernel smoothing methods"[J], Econometrics, 73 (3), 771-836.
- [23].Marcellino, M., and C.Schumacher, 2007, "Factor nowcasting of German GDP with ragged-edge data:Comparison with comparison using MIDAS projections", Bundesbank Discussion Paper, Series 1 No.34.
- [24].Silvestrini, A., and D.Veredas, 2008, "Temporal aggregation of univariate and multivariate time series survey"[J], Journal of Economic Surveys, 22 (3), 458-497.
- [25].Zadrozny, P.A., 1988, "Gaussian-Likelihood of continuous time AR-MAX models when data are sampled at different frequencies"[J], Econometric Theory, 4 (1), 108-124.
- [26].Zadrozny, P.A., 1990, "Estimating a multivariate ARMA model with mixed-frequency data:An application to forecasting U.S.GNP on monthly intervals", Federal Reserve Bank of Atlanta Working Paper Series No.90-6.
- [27].Zadrozny, P.A., 2008, "Estimating a multivariate ARMA model with mixed frequency data:An application to forecasting U.S.GNP at monthly intervals"[J], CESifo Working Paper. (M)

注释

- ① 目前将MIDAS混频模型应用到我国宏观经济研究的文献还没有,但是对于中国股票市场的研究已经有相关的文献(徐剑刚和唐国兴,2007年;Kong, Liu和Wang, 2008),但MIDAS的有效性并没有得到完全证实,徐剑刚、张晓蓉和唐国兴(2007)认为MIDAS不如安德森等人(Andersen, Bollerslev, Diebold和Labys, 2003)的分数自回归(ARFI)模型。
- ② β 滞后多项式函数也能构造出不同的形态的权重函数,限于篇幅所限,本文不列出 β 滞后多项式函数构造的常见形式,如果感兴趣,可向作者索取。
- ③ 理论上来说,选择的滞后阶数越长就能获得更多的信息,估计模型就越有效,预测就越精准。但是高频数据的滞后会导致失真,长滞后阶数在高频金融数据也许可行,但是宏观经济数据的长度都非常短,滞后阶数越长,能用于MIDAS模型中的样本这样反而会影响模型的估计与预测。
- ④ 如果优化后的参数 θ^* 所绘出的权重函数图形在第j阶滞后趋于零,就可以将滞后阶数选为K=j。
- ⑤ 这里选择进行1000次的蒙特卡洛模拟是根据对RMSE结果再进行100次的模拟发现模拟出的RMSE波动非常小,其偏差不超过5%,因此认为1000次蒙特卡罗模拟出的RMSE是稳定的,当然如果要求更加稳定的RMSE值,蒙特卡罗模拟的次数越多越好,计算机计算速率和本文精度要求,本文选择模拟1000次是有说服力的。
- ⑥ 如图3中第一行(第二行),当其他条件不变的情况下,样本长度越长(滞后阶数越长),MIDAS模型估计出的 β^1 的分布就更接近真实值,说明估计效果越明显。而同频数据估计出的 β^1 的分布都比较发散,说明其估计结果不具有优越性,只有当滞后阶数较短时,混频数据和同频数据模型估计结果才比较接近。
- ⑦ 缓慢递减权重的情形下的参数估计值的各种图形也可以容易得到,由于文章篇幅所限,本文省略。
- ⑧ GDP增长率和通货膨胀率的条件方差是通过构建响应的GARCH模型计算而来的(刘金全和谢卫东,2003),具体地说,GDP的GARCH模型是由AR (1) 的均值过程和GARCH (1, 1) 的方差方程组成;通货膨胀率的GARCH模型是由ARMA (2, 1) 的均值

GARCH (1, 1) 的方差方程组成。

[9] 季度通货膨胀变化率的构建方法有两种, 一种是加总法, 即通过平均季度内月度通货膨胀率获得季度通货膨胀率的数据; 另一种方法, 即使用季末月份的通货膨胀率来代替整个季度的通货膨胀率。

[10] 本文在构建中国宏观经济波动的MIDAS模型时使用了只有一个解释变量和一个被解释变量的方程简单方程, 中国的产出波动当然不仅仅取决于物价水平的波动, 还取决于很多因素的共同作用, 但是为了本文对比MIDAS优越性的需要, 本文没有涉及的情形, 在今后的研究还将进一步深入。