

基于高维混频信息甄选的宏观经济 总量实时预报与短期预测

刘 汉 刘 营 卓杏轩*

摘 要：本文采用阈值组惩罚无约束混频数据抽样 (T-GP-U-MIDAS) 模型对中国季度实际 GDP 增长率进行实时预报和短期预测，并研究其精度提升机理。实证结果显示：与 AR 等传统模型相比，T-GP-U-MIDAS 模型在预测精度方面具有显著的比较优势；T-GP-U-MIDAS 模型通过引入具有先行特征的阈值变量识别经济运行中可能存在的转变点，并依据不同的经济状况筛选出对预测实际 GDP 增长率存在重要影响的核心变量，在一定程度上提高了模型的预测和解释能力，是现有混频数据模型的有效改进和补充。

关键词：高维混频数据；T-GP-U-MIDAS 模型；实际 GDP 增长率

DOI: 10.13821/j.cnki.ceq.2023.03.17

一、引 言

自 2008 年金融危机爆发以来，全球经济面临流动性收缩和金融动荡等复杂多变的外部经济环境，尤其是本次新冠肺炎疫情的突发使得阐释和防范外部经济冲击对我国经济增长带来的不确定性影响变得十分重要。在此背景下宏观经济调控需要及时、准确地预判经济态势中出现的新变化。这一需求也大大激发了政策制定者与学术界对宏观经济实时预报和短期预测的关注。随着信息技术的并行发展，越来越多的官方统计数据可被用于宏观经济预报和预测研究，如何构建恰当的计量模型以纳入更多可用数据往往面临两方面的难题：

第一，如何处理“混频数据”问题？例如，本文的目标变量——国内生产总值 (GDP) 按照季度频率发布，而与其相关的其他经济变量，如进出口总额却是以月度频率发布 (王维国和于扬, 2016)。利用不同频率数据进行建模时，传统的做法是采用依时性加总、跳跃取样或插值等方法对数据进行同频化处理。但根据同频化处理得到的数据往往会存在信息漏损或虚增 (高华川和白仲林, 2016; 刘汉等, 2017)。由 Ghysels et al. (2004) 提出的混频数据抽样 (Mixed Data Sampling, MIDAS) 模型，因其能无

* 刘汉，吉林大学数量经济研究中心、吉林大学商学与管理学院；刘营，吉林大学商学与管理学院；卓杏轩，福州大学经济与管理学院。通信作者及地址：卓杏轩，福建省福州市福州大学城龙江北大道 2 号，350108；电话：15060069957；E-mail: zhuoxx1991@fzu.edu.cn。本文得到国家自然科学基金青年科学基金项目 (72004077、72103044)、教育部人文社会科学重点研究基地重大项目 (22JJD790066)、教育部人文社会科学研究项目青年基金项目 (20YJC790076)、福建省自然科学基金青年项目 (2020J05124)、吉林大学研究生创新基金资助项目 (101832020CX066)、吉林大学“学科交叉融合创新”培育项目 (JLUXKJC2020312) 和中央高校哲学社会科学创新团队项目 (2022CXTD25) 的资助。作者感谢匿名审稿人的建设性意见，文责自负。

损地攫取原始数据信息，被广泛应用于宏观经济预测中（郑挺国和尚玉皇，2013）。

第二，如何处理引入高维数据所导致的“维度灾难”问题？即模型中的待估参数数量可能远大于样本量（Nakajima and Sueishi, 2022）。近年来随着统计数据的收集日益系统化，研究者们倾向于使用更为丰富的高维数据进行宏观经济预测。然而，高维数据的使用可能会导致“维度灾难”，从而出现模型过拟合和参数估计不稳定等问题。针对这一难题，目前在统计学和计量经济学领域常用“损失函数+惩罚函数”形式构建模型的高维稀疏化建模方法（Uematsu and Tanaka, 2017），通过将不（弱）相关的解释变量系数约束为零从而实现降维目的（曾津和周建军，2017）。高维稀疏化建模方法的优点在于它能够同时兼顾“变量选择”与“参数估计”。例如，Tibshirani（1996）提出的具有开创性意义的最小绝对收缩与选择算子（LASSO）惩罚函数，因其能够同时实现模型的稀疏化和变量选择功能且具备高效的求解算法，被广泛应用于机器学习、统计分析与预测等领域。Fan and Li（2001）提出的平滑削边绝对偏离（SCAD）惩罚函数和Zhang（2010）提出的最小最大凹惩罚（MCP）函数，因能同时满足稀疏性、无偏性和连续性性质，从而提高了参数估计和变量选择的准确性。

目前已有一些研究开始在 MIDAS 类模型中引入各类惩罚函数进行建模，并应用于宏观经济分析与预测。例如，Marsilli（2014）首次将 LASSO 惩罚函数与 MIDAS 模型结合构建了 LASSO-MIDAS 模型，并将其应用于预测美国季度 GDP 增长率。实证结果表明该模型有效地利用混频数据信息，在同时实现降维、参数估计和变量选择的情况下，提高了预测精度。随后，Siliverstovs（2017）将弹性网（Elastic Net）与无约束混频数据抽样（U-MIDAS）模型结合，构建了 MIDASSO 模型，并用于预测瑞士季度实际 GDP 增长率，结果表明 MIDASSO 模型的预测精度较基准模型更为准确。Uematsu and Tanaka（2017）将 LASSO、SCAD 和 MCP 惩罚函数与 U-MIDAS 模型相结合，构建了惩罚无约束混频数据抽样（P-U-MIDAS）模型，并从理论上证明了惩罚函数应用于高维混频时间序列数据的合理性。然而，U-MIDAS 模型在处理混频数据时需要将高频解释变量通过多阶滞后操作进行频率对齐转换，以实现与低频目标变量相匹配。此时，若由同一解释变量转换生成的任意频率对齐项对目标变量存在重要影响，均代表该解释变量对目标变量存在重要影响。因此，应将同一高频变量生成的低频频率对齐向量组视为不可分割的整体。但基于单变量选择的惩罚函数所构建的 P-U-MIDAS 模型倾向于随机选择同一变量的某一频率对齐项，而忽略变量选择的完整性。为此，Xu et al.（2018）将基于组变量选择的组 LASSO（Yuan and Lin, 2006）、组 SCAD（Wang et al., 2007）和组 MCP（Huang et al., 2012）惩罚函数与 U-MIDAS 模型相结合，构建了组惩罚无约束混频数据抽样（GP-U-MIDAS）模型，并用于预测美国季度 GDP 增长率。该方法通过组惩罚函数将同一变量的所有滞后项视为不可分割的整体进行同时保留或删除，改善了 P-U-MIDAS 模型变量选择结果分散且随机的缺陷。

上述高维混频数据模型能够无损地利用高维混频数据信息对季度实际 GDP 增长率进行实时预报和短期预测。然而，诸如 2020 年年初爆发的重大公共卫生事件——新冠肺炎疫情给我国经济带来了极大的外部冲击，该类突发事件会导致宏观经济变量对 GDP 呈现不同的非线性影响。因此，在应用高维混频数据建模的同时，如何识别经济运行过程中由于外部环境不稳定导致可能存在的转变点所带来的非线性影响呢？近年来，

Ciuperca (2014) 研究了可以识别多个转变点的阈值回归模型, 并应用 LASSO 等惩罚函数来估计模型, 以同时实现参数估计和变量选择。然而, 该文章仅在解释变量个数较少的情形下验证了所提出模型的优势。为此, Lee et al. (2016) 提出了可以同时进行参数估计、变量选择和转变点分析的高维阈值回归模型, 蒙特卡洛模拟结果表明无论是在低维还是高维数据环境中, 该模型的预测效果均为最优。然而, 该模型不能直接用于混频数据分析, 从而无法满足本文需求。

基于此, 本文将 Xu et al. (2018) 提出的 GP-U-MIDAS 模型与 Lee et al. (2016) 提出的高维阈值回归模型相结合, 构建了阈值组惩罚无约束混频数据抽样 (T-GP-U-MIDAS) 模型, 该模型的优势在于能够同时处理混频数据 (U-MIDAS 模型)、高维数据 (组惩罚函数) 以及识别经济运行中可能存在的转变点 (阈值函数) 问题。具体来说, 与 Xu et al. (2018) 相比, T-GP-U-MIDAS 模型通过引入阈值变量以捕捉在经济运行中可能存在的转变点, 并基于最优阈值参数的结果判断当下经济状况的好坏, 从而在不同的经济增长阶段下选择对 GDP 具有重要影响的解释变量。特别是在面对重大危机事件导致预测失灵时, 以具有先行特征的阈值变量作为经济状况阶段性转变的指南针, 能够为非平稳时期的预测提供更多有效的信息。与 Lee et al. (2016) 相比, T-GP-U-MIDAS 模型一方面能直接利用原始混频数据, 减少了数据信息的漏损; 另一方面本文对原有模型的惩罚方法进行了扩充, 在原有仅包含 LASSO 惩罚函数的基础上添加了 SCAD 和 MCP 惩罚函数, 以及能够进行组变量选择的组 LASSO、组 SCAD 和组 MCP 惩罚函数, 以更全面地分析各类惩罚方法所具有的优势。在本文实际应用中, 第一, 使用 24 个月度宏观和金融变量检验了 T-GP-U-MIDAS 模型对我国实际 GDP 增长率进行实时预报与短期预测的适用性与有效性; 第二, 重点分析了实际 GDP 增长率实时预报与短期预测的精度提升机理。

本文剩余内容安排如下: 第二部分阐述了 T-GP-U-MIDAS 模型的构建、估计和评价方法; 第三部分阐述了数据选取及处理方法; 第四部分展示了 T-GP-U-MIDAS 模型对实际 GDP 增长率进行实时预报和短期预测的实证结果, 并分析了模型的精度提升机理; 第五部分为结论与政策建议。

二、T-GP-U-MIDAS 模型

下文将从 T-GP-U-MIDAS 模型的构建、参数估计方法和模型评价等方面进行详细阐述。

(一) 模型构建

1. 基础 U-MIDAS 模型

本文首先给出基础 U-MIDAS 模型的一般形式 (Foroni et al., 2015), 以更清晰地理解 T-GP-U-MIDAS 模型的构建过程:

$$y_t = \sum_{i=1}^N \sum_{j=0}^{l_i} \beta_{i,j} x_{i,t-h-j/m}^{(m)} + \epsilon_t, \quad t=1, \dots, T, \quad (1)$$

其中, y_t 为第 t ($t=1, \dots, T$) 期的低频目标变量; N 表示解释变量个数; l_i 为第 i 个解释变量的最大滞后阶数; $\{x_{i,t-h-j/m}^{(m)}\}_{j=0}^{l_i}$ 为第 i 个高频解释变量在 $t-h$ 时期通过频

率对齐转换所形成的低频向量组 $(x_{i,t-h-1/m}^{(m)}, x_{i,t-h-2/m}^{(m)}, \dots, x_{i,t-h-l_i/m}^{(m)})'$; $\{\beta_{i,j}\}_{j=0}^{l_i}$ 为第 i 个高频解释变量对应的 (l_i+1) 个待估系数; m 表示高频解释变量与低频目标变量之间的频率倍差; h 表示向前预测步长; ϵ_i 为误差项, 以下出现的相同符号均与上述符号含义一致。

U-MIDAS 模型中的参数估计结果, 可通过最小化下列方程求得:

$$\hat{\beta} = \arg \min_{\beta} \sum_t (y_t - \sum_{i=1}^N \sum_{j=0}^{l_i} \beta_{i,j} x_{i,t-h-j/m}^{(m)})^2 = \arg \min_{\beta} \|Y - X\beta\|_2^2. \quad (2)$$

在上述模型中, 若 N 个高频解释变量的滞后阶数均为 l 时, 则有 $N(l+1)$ 个待估参数, 当待估参数数量大于样本量时, 则会导致“维度灾难”问题。为此, Uematsu and Tanaka (2017) 以 U-MIDAS 模型为基础, 通过在式 (2) 中添加惩罚项构建了 P-U-MIDAS 模型。P-U-MIDAS 模型可以剔除对低频目标变量影响较小的高频解释变量从而实现降维的目的。P-U-MIDAS 模型的具体形式可参考 Uematsu and Tanaka (2017)。

2. GP-U-MIDAS 模型

GP-U-MIDAS 模型是 Xu et al. (2018) 将具有组变量选择功能的组 LASSO、组 SCAD 和组 MCP 惩罚函数分别与 U-MIDAS 进行结合构建而成。具体模型形式分别如下:

(1) G-LASSO-U-MIDAS 模型

G-LASSO-U-MIDAS 模型是由 Yuan and Lin (2006) 提出的组 LASSO 惩罚函数与 U-MIDAS 模型结合构建而成, 待估参数可表示为:

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_{2,1}, \quad (3)$$

其中, $\|\beta\|_{2,1}$ 表示变量选择只在组与组之间进行, 即在不同的解释变量之间进行, 而不在同一解释变量的不同滞后阶数内进行选择。

(2) G-SCAD-U-MIDAS 模型

G-SCAD-U-MIDAS 模型是由 Wang et al. (2007) 提出的组 SCAD 惩罚函数与 U-MIDAS 模型结合构建而成, 待估参数可表示为:

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \|Y - X\beta\|_2^2 + \sum_{i=1}^N p_{\lambda,\gamma}(\|\beta^{(i)}\|_{2,1}), \quad (4)$$

其中, $\beta^{(i)}$ 为第 i 个解释变量及其滞后项构成的组变量所对应的回归系数; $p_{\lambda,\gamma}(\|\beta^{(i)}\|_{2,1})$ 代表惩罚函数, 其具体形式为:

$$p_{\lambda,\gamma}(\|\beta^{(i)}\|_{2,1}) = \begin{cases} \lambda \|\beta^{(i)}\|_{2,1}, & \text{当 } \|\beta^{(i)}\|_{2,1} \leq \lambda \text{ 时} \\ -\left(\frac{\|\beta^{(i)}\|_{2,1}^2 - 2\gamma\lambda \|\beta^{(i)}\|_{2,1} + \lambda^2}{2(\gamma-1)}\right), & \text{当 } \lambda < \|\beta^{(i)}\|_{2,1} \leq \gamma\lambda \text{ 时,} \\ \frac{(\gamma+1)\lambda^2}{2}, & \text{当 } \|\beta^{(i)}\|_{2,1} > \gamma\lambda \text{ 时} \end{cases} \quad (5)$$

γ 表示调整参数, 且 $\gamma > 2$ 。

(3) G-MCP-U-MIDAS 模型

G-MCP-U-MIDAS 模型是由 Huang et al. (2012) 提出的组 MCP 惩罚函数与 U-MIDAS 模型结合构建而成, 待估参数可表示为:

$$\hat{\beta} = \arg \min_{\beta \in \mathbf{R}^p} \frac{1}{2} \| \mathbf{Y} - \mathbf{X}\beta \|_2^2 + \sum_{i=1}^N p_{\lambda, \gamma}(\| \beta^{(i)} \|_{2,1}), \quad (6)$$

其中 $p_{\lambda, \gamma}(\| \beta^{(i)} \|_{2,1})$ 的具体形式为:

$$p_{\lambda, \gamma}(\| \beta^{(i)} \|_{2,1}) = \begin{cases} \lambda \| \beta^{(i)} \|_{2,1} - \frac{\| \beta^{(i)} \|_{2,1}^2}{2\gamma}, & \text{当 } \| \beta^{(i)} \|_{2,1} \leq \gamma\lambda \text{ 时} \\ \frac{1}{2} \gamma \lambda^2, & \text{当 } \| \beta^{(i)} \|_{2,1} > \gamma\lambda \text{ 时} \end{cases}. \quad (7)$$

3. T-GP-U-MIDAS 模型

T-GP-U-MIDAS 模型是在 GP-U-MIDAS 模型基础上引入阈值函数构建而成。为了更清晰地展示 T-GP-U-MIDAS 模型的构建过程, 本文首先给出不带有惩罚函数的 T-U-MIDAS 模型的一般形式:

$$y_t = \beta_0 + \sum_{i=1}^N \sum_{j=0}^{l_i} \beta_{i,j} x_{i,t-h-j/m}^{(m)} + \mathbf{1}\{Q_t < \theta_t\} \sum_{i=1}^N \sum_{j=0}^{l_i} \delta_{i,j} x_{i,t-h-j/m}^{(m)} + \varepsilon_{t+h}, \quad (8)$$

其中, $\mathbf{1}\{\cdot\}$ 是示性函数, $Q_t (t=1, 2, \dots, T)$ 是阈值变量且频率与 y_t 一致, $\theta_t (t=1, 2, \dots, T)$ 是阈值参数, 通过 Q_t 与 θ_t 的大小可判断解释变量在第 t 期是否存在阈值效应, $\delta_{i,j}$ 为存在阈值效应的解释变量所对应的阈值系数。其中, 阈值变量的选择需要根据实际研究问题来确定。本文的目的是对季度实际 GDP 增长率进行实时预报和短期预测, 并识别在不同的经济增长阶段下对季度实际 GDP 增长率存在重要影响的变量。因此, 本文参考现有文献 (刘汉和刘金全, 2011; 郑挺国和尚玉皇, 2013; 郑挺国和王霞, 2013), 选取了工业增加值 (VAI), 社会消费品零售总额 (TRSCG), 进口 (IMPORT) 及出口 (EXPORT) 贸易额, 税收收入 (TRGR) 等一系列反映中国经济运行态势的指标与实际 GDP 增长率进行相关性检验, 并经过对时间序列数据进行分段检验等多次尝试, 最终选取相关性最高的月度工业增加值 (VAI) 作为阈值变量¹, 用以识别经济增长阶段中可能存在的转变点。

T-GP-U-MIDAS 模型是在式 (8) 的基础上加入组惩罚函数构建的, 其参数估计形式可表示为:

$$E(\alpha) = \arg \min_{\alpha} \frac{1}{2} \| \mathbf{Y} - \mathbf{X}(\theta)\alpha \|_2^2 + p_{\lambda, \gamma}(\| \mathbf{D}(\theta)\alpha \|_{2,1}), \quad (9)$$

其中, $\alpha = (\beta, \delta)' = (\beta_{1,0}, \dots, \beta_{1,l_1}, \dots, \beta_{N,0}, \dots, \beta_{N,l_N}, \delta_{1,0}, \dots, \delta_{1,l_1}, \dots,$

$\delta_{N,0}, \dots, \delta_{N,l_N})'$ 是 $2 \sum_{i=1}^N (l_i + 1)$ 维的系数向量; $\mathbf{X}(\theta) = (\mathbf{X}, \mathbf{1}\{Q < \theta\}\mathbf{X})'$ 是由原始解

释变量矩阵 $\mathbf{X}$ 与存在阈值效应的解释变量矩阵 $\mathbf{1}\{Q < \theta\}\mathbf{X}$ 连接形成的 $T \times 2 \sum_{i=1}^N (l_i + 1)$ 维

的新解释变量矩阵; $\mathbf{D}(\theta) = \text{diag}\{\mathbf{X}(\theta)\}$ 是 $2 \sum_{i=1}^N (l_i + 1) \times 2 \sum_{i=1}^N (l_i + 1)$ 维的对角矩阵;

¹ U-MIDAS 模型是通过频率对齐将高频解释变量转换为低频的频率对齐向量。因而, 本文在使用月度工业增加值 (VAI) 作为阈值变量时, 通过将当下最新可获取的月度数据取平均的方式转换为季度频率, 以放入 U-MIDAS 模型中构建阈值函数。

$p_{\lambda,\gamma}(\cdot)$ 为组惩罚函数, 在 T-GP-U-MIDAS 模型中可分别表示为组 LASSO、组 SCAD 以及组 MCP 的形式, 由于与 GP-U-MIDAS 模型中惩罚函数形式一致, 故不再赘述。此外, 本文还构建了由单变量选择方法与 U-MIDAS 模型结合构建而成的阈值惩罚无约束混频数据抽样 (T-P-U-MIDAS) 模型, 以检验对变量及其滞后项进行分组能否提高预测精度。由于 T-P-U-MIDAS 模型与 T-GP-U-MIDAS 模型仅在惩罚函数上存在区别, 限于篇幅故不再赘述。

此外, 为了更清晰地阐述 T-GP-U-MIDAS 模型中阈值效应的作用机制, 本文以全样本估计为例进行详细说明, 如图 1 所示。首先, 本文假设根据每一期阈值变量 Q_t 与阈值参数 θ_t 的比较结果, 得出解释变量在 $t=t_0$ 时期 (以 $t_0=\{1, 3\}$ 为例) 存在转变点, 即表示解释变量在 $t=1, 3$ 期存在阈值效应。其次, 将原始解释变量矩阵 X 与存在阈值效应的解释变量矩阵 $1\{Q<\theta\}X$ 连接构成新的解释变量矩阵 $X(\theta)=(X, 1\{Q<\theta\}X)$ 。最后, 将 $X(\theta)=(X, 1\{Q<\theta\}X)$ 带入式 (9) 中同时进行参数估计和变量选择, 最终选择出影响实际 GDP 增长率的核心变量。

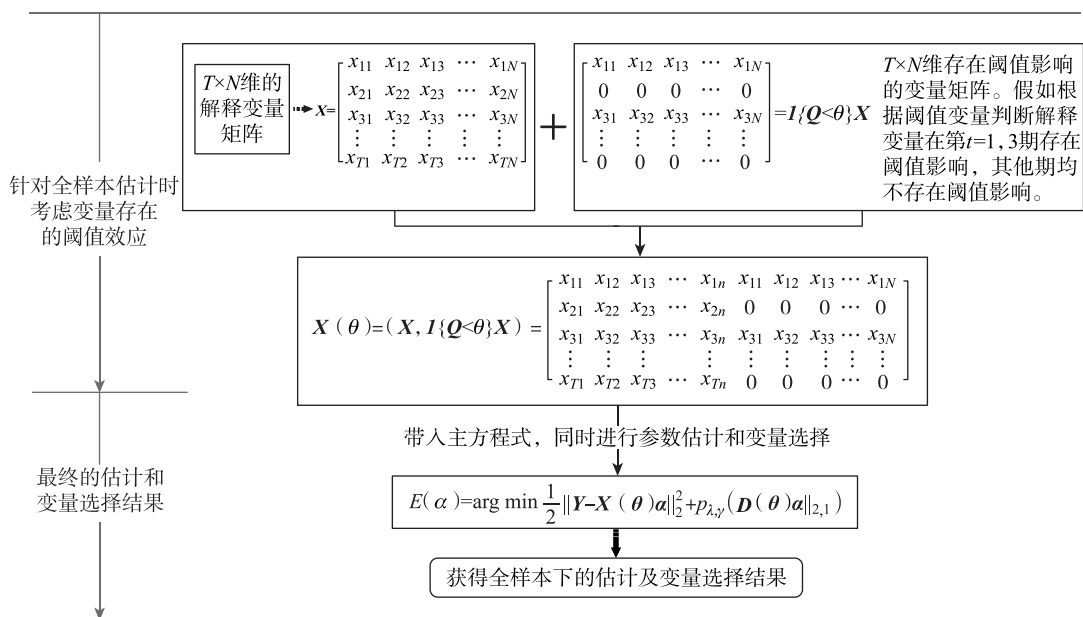


图 1 T-GP-U-MIDAS 模型中阈值效应作用机制

(二) 模型参数估计方法

T-GP-U-MIDAS 模型涉及两个模型估计问题: 其一是如何估计模型的待估参数。由 Breheny and Huang (2015) 提出的组下降算法 (Block Coordinate Descent) 可被用来求解组 LASSO、组 SCAD 和组 MCP 等惩罚函数的参数估计问题, 该方法针对单个组优化目标函数, 循环遍历组直到收敛, 并不需要复杂的数值优化步骤。其二是如何确定每一期的最优阈值参数 θ_t 和惩罚参数 λ 。本文参考 Lee et al. (2016) 和 Xu et al. (2018) 的做法, 首先采用网格搜索方法确定每一期的最优阈值参数 θ_t , 然后采用时间

序列交叉验证方法确定最优惩罚参数 λ 。²

(三) 模型评价

本文选取均方根误差 (Root Mean Square Error, RMSE)、平均绝对误差 (Mean Absolute Error, MAE)、平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 三种误差损失函数来评价实际 GDP 增长率的实时预报与短期预测结果, 具体公式分别表示为:

$$RMSE = \sqrt{\frac{1}{J} \sum_{j=1}^J (\hat{y}_j - y_j)^2}, \quad (10)$$

$$MAE = \frac{1}{J} \sum_{j=1}^J |\hat{y}_j - y_j|, \quad (11)$$

$$MAPE = \frac{100\%}{J} \sum_{j=1}^J \left| \frac{\hat{y}_j - y_j}{y_j} \right|. \quad (12)$$

上式中 J 为样本外预测集长度, y_j 表示样本外第 j 期目标变量的实际值, $\hat{y}_j$ 表示第 j 期目标变量的预测值, 三种评价指标的值越小, 表明预测值 $\hat{y}_j$ 与真实值 y_j 之间差距越小, 预测结果越准确。

通过对比上述三种预测评价指标的大小可以初步判断模型的优劣, 但由于评价指标的大小非常容易受到异常值的影响, 使得仅使用损失函数作为评价指标存在一定的局限性。因此, 本文引入 Diebold and Mariano (1995) 提出的 DM (Diebold-Mariano) 检验, 从统计意义上检验不同模型间的预测能力是否存在显著的差异。具体的 DM 检验过程请参见 Diebold and Mariano (1995)。在对比分析任意两个模型预测能力的优劣时, DM 检验的原假设 H_0 假设模型 1 和模型 2 的预测能力无差异, 本文选择的备择假设 H_1 为模型 2 的预测能力优于模型 1。若 DM 统计量显著, 则拒绝原假设; 反之, 则接受原假设。

三、变量选择与数据检验

(一) 变量及样本数据选取

本文以中国宏观经济总量——季度实际 GDP 增长率作为目标变量; 以国家统计局、中国经济景气监测中心, 以及现有文献 (刘汉和刘金全, 2011; 尚玉皇和郑挺国, 2018; 郑挺国和尚玉皇, 2013; 周建和况明, 2015; 费兆奇和刘康, 2019; 周德才等, 2019) 所选取的相关指标作为参考依据, 最终选取了九个月度先行指标、七个月度一致指标和八个月度金融变量作为解释变量。鉴于数据的可获得性, 本文选取的高频月度数

² 受篇幅限制, 关于网格搜索方法和时间序列交叉验证的具体步骤, 如有需要可见上述参考文献, 也可来信向作者索要。

据样本区间为2002年1月（2002M1）至2020年12月（2020M12），对应的低频季度实际GDP数据的样本区间为2002年第1季度（2002Q1）至2020年第4季度（2020Q4）。所选变量的名称、英文缩写、数据形式和数据来源如表1所示。

表1 选取变量信息表

类型	指标名称	英文缩写	数据形式	数据来源
先行指标	房地产土地购置面积	IAAREDE	同比增长	国家统计局
	商品房新开工面积指标	HSTARTS	同比增长	国家统计局
	社会货运量	TFT	同比增长	国家统计局
	沿海主要港口货物吞吐量	VFH	同比增长	国家统计局
	货币供应量	M2	同比增长	中国人民银行
	国债利率差	TBIRD	原序列	中债估值中心
	固定资产投资新开工项目	NNP	同比增长	国家统计局
	消费者预期指数	CEI	1阶差分	国家统计局
	工业企业产品产销率	RSO	同比增长	国家统计局
一致指标	工业增加值	VAI	同比增长	国家统计局
	工业企业利润总额	TPIE	同比增长	国家统计局
	国家财政收入中各项税收收入	TRGR	同比增长	财政部
	进口额	IMPORT	同比增长	国家统计局
	出口额	EXPORT	同比增长	国家统计局
	固定资产投资完成额	CIFA	同比增长	国家统计局
	社会消费品零售总额	TRSCG	同比增长	国家统计局
金融指标	社会融资规模	AFRE	同比增长	中国人民银行
	银行间7天同业拆借利率	IWLR7	原序列	中国外汇交易中心
	人民币实际有效汇率	REER	同比增长	国际清算银行（BIS）
	上证综合指数收益率	SSEC	同比增长	上海证券交易所
	深证成份指数收益率	SZI	同比增长	深圳证券交易所
	房地产销售均价	HPI	同比增长	国家统计局
	长期国债收益率	LBOND	原序列	中债估值中心
	短期国债收益率	SBOND	原序列	中债估值中心

（二）单位根检验

为了保证建模的稳健性，避免数据因季节性趋势导致难以识别变量之间的真实影响关系。本文首先采用 Hylleberg et al.（1990）提出的季节性单位根（HEGY）检验方法，分别以每6个、12个月为循环周期，检验数据是否存在季节性单位根。表2的检验结果显示解释变量在所有情况下均不存在季节性单位根；其次，采用 ADF 检验（Dickey

and Fuller, 1981) 变量是否存在单位根, 检验结果表明所选变量均为平稳序列; 最后, 对所有变量进行标准化处理以消除量纲影响。

表 2 变量的平稳性检验结果

变量	HEGY 检验 <i>P</i> 值		ADF 检验	
	6 个月	12 个月	检验类型 (<i>C, T, K</i>)	<i>P</i> 值
IAAREDE	0.0000	0.0002	(<i>C, T, 1</i>)	0.0000
HSTARTS	0.0000	0.0000	(<i>C, T, 1</i>)	0.0000
TFT	0.0000	0.0000	(<i>C, T, 1</i>)	0.0007
VFH	0.0000	0.0000	(<i>C, T, 1</i>)	0.0001
M2	0.0000	0.0000	(<i>C, T, 3</i>)	0.0433
TBIRD	0.0000	0.0105	(<i>C, T, 1</i>)	0.0024
NNP	0.0000	0.0000	(<i>C, T, 4</i>)	0.0005
D(CEI)	0.0156	0.0080	(0, 0, 4)	0.0000
RSO	0.0000	0.0000	(<i>C, T, 1</i>)	0.0007
VAI	0.0000	0.0000	(<i>C, T, 1</i>)	0.0000
TPIE	0.0000	0.0001	(<i>C, T, 4</i>)	0.0000
TRGR	0.0157	0.0000	(<i>C, T, 4</i>)	0.0001
IMPORT	0.0000	0.0000	(<i>C, T, 1</i>)	0.0042
EXPORT	0.0000	0.0000	(<i>C, T, 2</i>)	0.0304
CIFA	0.0008	0.0008	(<i>C, T, 4</i>)	0.0040
TRSCG	0.0000	0.0000	(<i>C, T, 6</i>)	0.0255
AFRE	0.0000	0.0000	(<i>C, T, 5</i>)	0.0013
IWLR7	0.0000	0.0002	(<i>C, T, 1</i>)	0.0041
REER	0.0000	0.0001	(<i>C, T, 6</i>)	0.0307
SSEC	0.0000	0.0000	(<i>C, T, 5</i>)	0.0049
SZI	0.0000	0.0000	(<i>C, T, 5</i>)	0.0033
HPI	0.0000	0.0001	(<i>C, T, 4</i>)	0.0001
LBOND	0.0000	0.0001	(<i>C, T, 4</i>)	0.0158
SBOND	0.0000	0.0000	(<i>C, 0, 1</i>)	0.0041
GDP	0.0000	0.0038	(<i>C, T, 1</i>)	0.0131

注: (1) 滞后阶数 *K* 以 AIC、BIC 值最小为准则进行选择。(2) ADF 单位根检验采纳费兆奇和刘康 (2019) 的检验流程。

四、中国季度实际 GDP 增长率实时预报与短期预测的实证研究

下文采用 T-GP-U-MIDAS 模型对中国季度实际 GDP 增长率进行实时预报和短期预测, 并依据不同的经济状况, 筛选出对实际 GDP 增长率预报和预测最为有效的核心解释变量。

（一）基准模型和对比模型的选取

为了检验 T-GP-U-MIDAS 模型对我国季度实际 GDP 进行实时预报和短期预测的适用性和有效性,本文选取了一系列对比模型:(1)选取 AR 模型作为基准模型。需要说明的是,ARDL、U-MIDAS 作为 T-GP-U-MIDAS 的基础模型本应被选入作为基准模型,但由于数据长度有限以及变量个数过多,使得 ARDL 和 U-MIDAS 模型在一些情况下失效,故本文未选取 ARDL、U-MIDAS 模型作为基准模型。(2)选取由 LASSO、SCAD 和 MCP 惩罚函数与 U-MIDAS 模型相结合构建的 P-U-MIDAS 类模型 (Uematsu and Tanaka, 2017),以检验对解释变量及其滞后项进行分组能否提高预测精度。(3)选取由 G-LASSO、G-SCAD 和 G-MCP 三种惩罚函数与 U-MIDAS 模型结合构建的 GP-U-MIDAS 类模型 (Xu et al., 2018),以检验在对变量及其滞后项进行分组的基础上,阈值函数的加入能否提高预测精度。(4)选取由 T-LASSO、T-SCAD 和 T-MCP 三种惩罚函数与 U-MIDAS 模型相结合构建的 T-P-U-MIDAS 模型,以检验在阈值函数的基础上,对变量及其滞后项进行分组能否提高预测精度。

此外,关于 AR 模型中自回归项的最优滞后阶数,P-U-MIDAS、GP-U-MIDAS、T-P-U-MIDAS 和 T-GP-U-MIDAS 模型中解释变量的最优滞后阶数均由 AIC 和 BIC 准则进行确定。

（二）实时预报和短期预测策略说明

为了评估模型预测能力的好坏,本文首先采用滚动窗口策略,在固定估计窗口长度后将总样本数据划分为训练集和预测集,通过滚动窗口生成直接一步向前预测结果,以避免全样本数据下一次实验的偶然性和随机性。具体来说,本文将数据集划分为:(1)训练集:基于 2002Q1—2015Q4 区间的子样本进行样本内估计;(2)预测集:基于滚动窗口生成从 2016Q1—2020Q4 期间的样本外预测结果。其次,Marsilli (2014) 表明 MIDAS 类模型适用于短期预测,且当混频变量的频率倍差较小时,U-MIDAS 类模型可以提高实时预测精度 (Foroni et al., 2015)。因此,本文参考 Marcellino and Schumacher (2010) 和 Uematsu and Tanaka (2017) 的方法,考虑了实时预报 ($h=0, 1/3, 2/3$) 和短期预测 ($h=1$) 下的四种预测步长情况,其中实时预报是指在当季实际 GDP 增长率正式发布之前,使用截至目前可获取的当季内最新月度信息来预测当季实际 GDP 增长率。

（三）实时预报与短期预测结果

表 3 报告了在不同预测步长下对中国季度实际 GDP 增长率进行实时预报和短期预测时,以 AR 模型为基准,P-U-MIDAS 类模型、GP-U-MIDAS 类模型、T-P-U-MIDAS 类模型和 T-GP-U-MIDAS 类模型的相对 RMSE、MAE 和 MAPE 结果。由表 3 可得:(1)表中 RMSE、MAE 和 MAPE 三种评价指标的所有比值均小于 1,说明 P-U-MIDAS 类模型、GP-U-MIDAS 类模型、T-P-U-MIDAS 类模型以及 T-GP-U-MIDAS 类模型,在不同步长下的预测精度均明显优于基准 AR 模型;(2)T-GP-U-MIDAS 类模型的预测精度明显高于基准模型和其他对比模型,这表明了 T-GP-U-MIDAS 模型应用于宏观经济

表 3 各种模型的实时预报和短期预测评估结果

模型	P-U-MIDAS				GP-U-MIDAS				T-P-U-MIDAS				T-GP-U-MIDAS									
	LASSO-U-MIDAS		SCAD-U-MIDAS		MCP-U-MIDAS		GLASSO-U-MIDAS		GSCAD-U-MIDAS		GMCP-U-MIDAS		T-LASSO-U-MIDAS		T-SCAD-U-MIDAS		T-MCP-U-MIDAS		T-GLASSO-U-MIDAS		T-GMCP-U-MIDAS	
rRMSE	$h=0$																					
	0.6375		0.6120		0.5867		0.5841		0.5796		0.5726		0.5582		0.5468		0.5442		0.4819		0.4514	
	$h=1/3$		0.6704		0.6689		0.6040		0.5724		0.5782		0.5633		0.5629		0.5488		0.5103		0.4708	
	$h=2/3$		0.7365		0.7451		0.7458		0.6692		0.6585		0.6509		0.6137		0.6375		0.5952		0.4793	
rMAE	$h=1$																					
	0.7501		0.7501		0.7501		0.7344		0.7107		0.7038		0.6480		0.6429		0.6381		0.6319		0.6165	
	$h=0$		0.5156		0.5211		0.4934		0.4607		0.4529		0.4578		0.4597		0.4483		0.4109		0.4036	
	$h=1/3$		0.6479		0.6016		0.5859		0.5361		0.5001		0.5055		0.4738		0.4977		0.4671		0.4287	
rMAPE	$h=2/3$																					
	0.6803		0.6560		0.6483		0.6418		0.5599		0.5545		0.5564		0.5493		0.5139		0.5093		0.4835	
	$h=1$		0.7322		0.6585		0.6570		0.6663		0.5712		0.5763		0.6073		0.5494		0.5535		0.5337	
	$h=0$		0.4234		0.4317		0.4294		0.3635		0.3467		0.3469		0.3485		0.3513		0.3407		0.3162	
rMAPE	$h=1/3$																					
	0.5021		0.5012		0.4913		0.4066		0.4157		0.4096		0.3996		0.3771		0.4121		0.3364		0.3278	
	$h=2/3$		0.5355		0.5150		0.5136		0.5628		0.4295		0.4263		0.4108		0.4138		0.4176		0.3889	
	$h=1$		0.5979		0.5302		0.5299		0.5677		0.4625		0.4651		0.4950		0.4340		0.4289		0.3897	

注：表格中报告是以 AR 模型为基准，各类惩罚模型的相对 RMSE、MAE 和 MAPE 结果。

总量实时预报和短期预测的有效性；(3)模型的预测精度整体呈现出 T-GP-U-MIDAS 类模型 > T-P-U-MIDAS 类模型 > 或 $\approx$ GP-U-MIDAS 类模型 > P-U-MIDAS 类模型 > AR(1) 模型的规律，这表明对变量及其滞后项进行分组和引入具有先行特征的阈值变量以捕捉在经济运行过程中可能存在的转变点，有助于对我国季度实际 GDP 增长率进行实时预报和短期预测；(4)随着预测步长 h 的增大，所有模型的预测效果均随可用信息的减少而逐步变差，该结果说明对实际 GDP 增长率进行实时预报和短期预测时，要求高频数据信息的实时更新是必要且合理的。

基于上述 RMSE、MAE 和 MAPE 的比值结果，可以得出 T-GP-U-MIDAS 模型表现最优，但正如上文所说，评价指标的大小易受到异常值的干扰，导致预测结果具有一定的偶然性和局限性。因而，本文进一步采用 DM 检验方法，从统计意义上检验模型间预测能力的优劣。图 2(a)至图 2(c) 分别以热图形式展示了基于 RMSE、MAE 及 MAPE 三种评价指标的 DM 检验结果，若 DM 检验结果的值小于 0.1（在 10% 的显著性水平下显著），则表明在横轴上的模型预测结果在统计意义上显著优于纵轴上的模型。由图 2 可得，基于 RMSE、MAE 和 MAPE 的 DM 检验结果虽略微存在差异，但均显示 T-GP-U-MIDAS 类模型的预报及预测能力在 10% 的显著性水平下优于 T-P-U-MIDAS 类模型、GP-U-MIDAS 类模型、P-U-MIDAS 类模型和 AR(1) 模型，从统计意义上证实了 T-GP-U-MIDAS 模型的有效性。

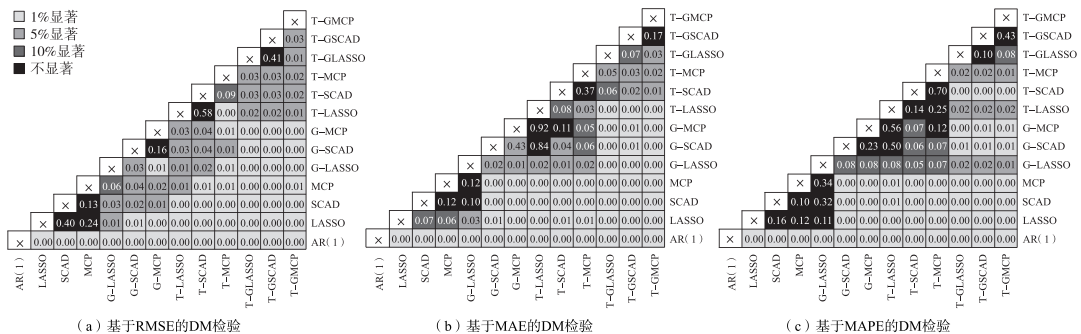


图 2 基于评价指标的 DM 检验结果

(四) 变量选择机理分析

下文将从 T-GP-U-MIDAS 模型中包含的阈值函数作用机制和惩罚函数的变量选择功能具体分析模型预测效果改进的内在机理。本文以预测结果最优的 T-GMCP-U-MIDAS 模型的变量选择结果为例进行分析。

1. 预测步长 $h=0$ 时的变量选择结果

图 3(d) 报告了 T-GMCP-U-MIDAS 模型在 $h=0$ 时的变量选择结果，为了对比说明不同模型变量选择结果的差异性，本文将 MCP-U-MIDAS 模型、GMCP-U-MIDAS 模型和 T-MCP-U-MIDAS 模型在 $h=0$ 预测步长下的变量选择结果分别列入图 3(a)至图 3(c) 中以作对比，图形中阴影部分表示在每一期预报和预测中被选中的变量。在图 3(c)至图 3(d)中，分割线以上阴影部分选择的变量代表在每一期预报和预测中正常发挥作用的解释变量，分割线以下代表的是在 $Q < \hat{\theta}$ 情况下，即经济增速较低时产生阈值效应的解释变量。具体分析如下：

第一,在模型中将高频解释变量与其滞后项进行分组,通过组惩罚函数对组变量进行同时保留或删除是否发挥了重要作用?对比有无组惩罚变量选择机制的 MCP-U-MIDAS 模型(图 3(a))和 GMCP-U-MIDAS 模型(图 3(b))的结果可得,MCP-U-MIDAS 模型选择的变量较为分散且不稳定,由于高频解释变量与其滞后项构成的变量组之间存在高度相关,使得 MCP 惩罚方法会倾向于随机选择任意一个滞后项,导致变量选择结果具有随机性,而 G-MCP-U-MIDAS 模型将高频解释变量与其滞后项进行同时保留或删除,在一定程度上提高了变量选择结果的一致性和完整性,从而增强了模型的预测能力和实证结果的解释能力。类似地,对比 T-MCP-U-MIDAS 模型(图 3(c))和 T-GMCP-U-MIDAS 模型(图 3(d))的变量选择结果,同样可以证实在解释变量个数增多时,组惩罚变量选择机制的存在使得 T-GMCP-U-MIDAS 模型的变量选择结果更为精简。

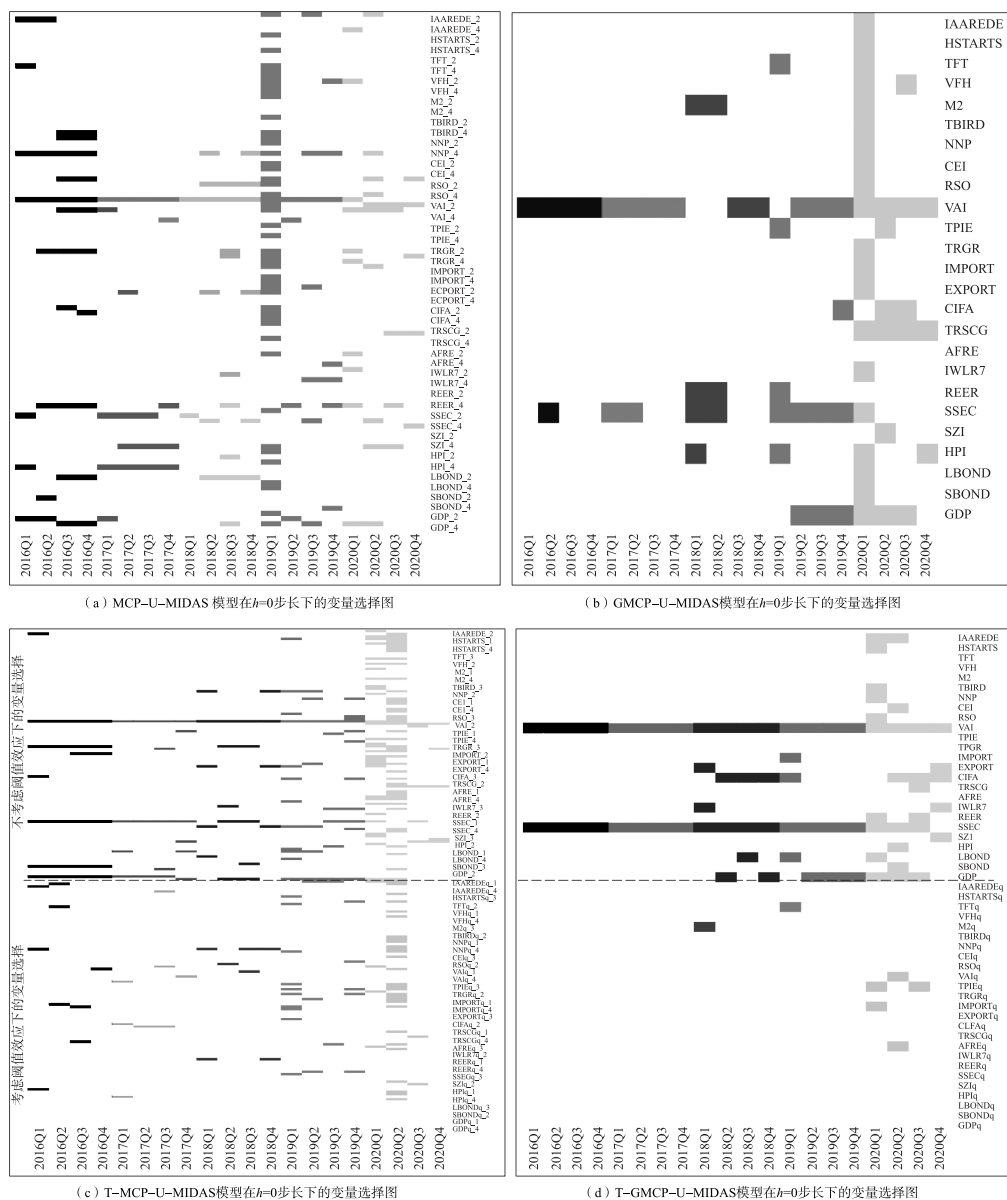


图 3 MCP 惩罚类模型在 $h=0$ 步长下变量选择对比

第二，在组惩罚混频数据模型的基础上加入阈值函数是否发挥了作用？对比有无阈值函数的 GMCP-U-MIDAS 模型（图 3(b)）与 T-GMCP-U-MIDAS 模型（图 3(d)）的变量选择结果可得，T-GMCP-U-MIDAS 模型的变量选择结果更为一致，且除正常发挥作用的解释变量以外，还在某些预测期依据经济状况的变化选择了存在阈值效应（ $Q < \theta$ ）的解释变量，那这些存在阈值效应的解释变量具体是在什么时期发挥了作用呢？受篇幅所致，本文以 $h=0$ 步长下 T-MCP-U-MIDAS 模型的结果为例说明预测过程中阈值效应的具体体现。

图 4(a) 展示了 T-MCP-U-MIDAS 模型在 $h=0$ 预测步长下进行每一期样本外预测时阈值变量和最优阈值参数的变化。首先，由样本外每一期阈值变量 Q_t 与最优阈值参数 $\hat{\theta}_t$ 的比较结果可得，2016Q1—2020Q3 期间中国经济一直处于经济增速较低的阶段。实际上，自 2008 年全球金融危机以来，中国经济增速就开始逐渐下滑，并一直处于经济下行阶段，其次，由于样本外预测期数过多，本文以 2016Q1、2017Q1、2018Q1、2019Q1 以及 2020Q1 为例，分别展示了在上述样本外预测期时解释变量在经济运行中存在阈值效应的时期，见图 4(b)至图 4(f) 所示，存在阈值效应的时期主要集中于 2008 年全球金融危机时期，以及自 2011 年以来的经济下行阶段。结合 T-GMCP-U-MIDAS 模型的变量选择结果（图 3(d)）可得，工业企业利润总额（ $TPIE$ ）、进口（ $IMPORT$ ）和社会融资规模（ $AFRE$ ）等指标主要在新冠疫情暴发时期存在额外的阈值效应，这表明在经济受到重大突发公共事件的冲击而处于衰退期（阈值变量 Q_t 低）时，进出口的停滞、工业企业利润总额的下滑以及受疫情影响导致流动性资金供给短缺均对 GDP 产生了重要冲击，可以从进出口、投资以及货币政策等方面进行调控来促进经济回温。

第三，由 T-GMCP-U-MIDAS 模型的变量选择结果可得：（1）在 2016Q1—2017Q4 期间工业增加值（ VAI ）始终被选入作为预测当季 GDP 增长率最为核心的解释变量；众所周知，第二产业是构成我国 GDP 的重要组成部分，虽然近年来第二产业对我国 GDP 的贡献率占比逐渐下降，但及时发布的月度指标工业增加值增速仍然是预测实际 GDP 增长率的重要参考依据之一。除此之外，高频金融变量上证综指收益率（ $SSEC$ ）也为我国 GDP 增长率的预测提供了有用的短期领先信息。（2）进入 2018 年后，受中美贸易摩擦对经济带来的不确定性影响，影响实际 GDP 增长率的变量个数逐渐增加，除了上述指标外，综合反映中国对外贸易情况的进出口（ $IMPORT$ 和 $EXPORT$ ）指标，在较大程度上反映了市场的真实供求关系的长期国债收益率（ $LBOND$ ），实时发布的银行间 7 天同业拆借利率（ $IWLR7$ ）以及代表我国固定资产投资主要构成部分的固定资产完成额（ $CIFA$ ）指标也在不同预测期内对季度实际 GDP 增长率的预测发挥了重要作用。（3）进入 2020 年以来，重大突发公共卫生事件新冠疫情的暴发给经济带来了严重的外在冲击，T-GMCP-U-MIDAS 模型选取的变量变得错综复杂，除了上述始终发挥重要作用的工业增加值（ VAI ）和上证综指收益率（ $SSEC$ ）变量以外，剩余被选入变量主要集中于投资、消费和利率等方面，这表明受新冠疫情的影响，线下消费的流失、投资需求的下降以及进而导致的利率下降都对季度实际 GDP 增长率的预测产生了或多或少的影响。

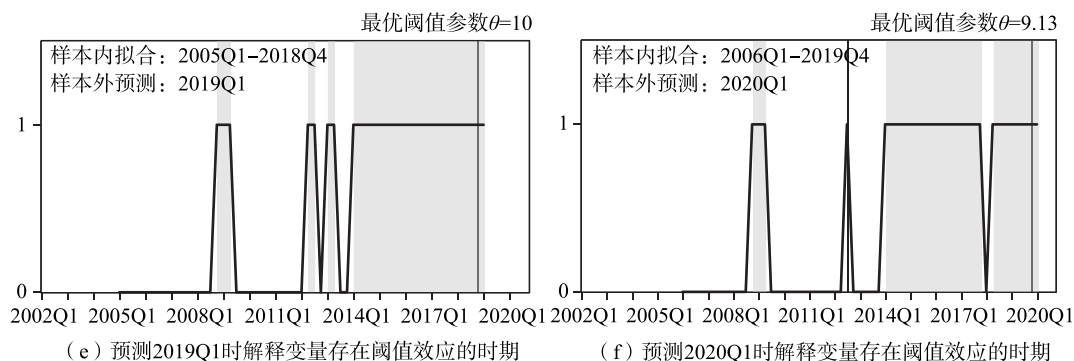
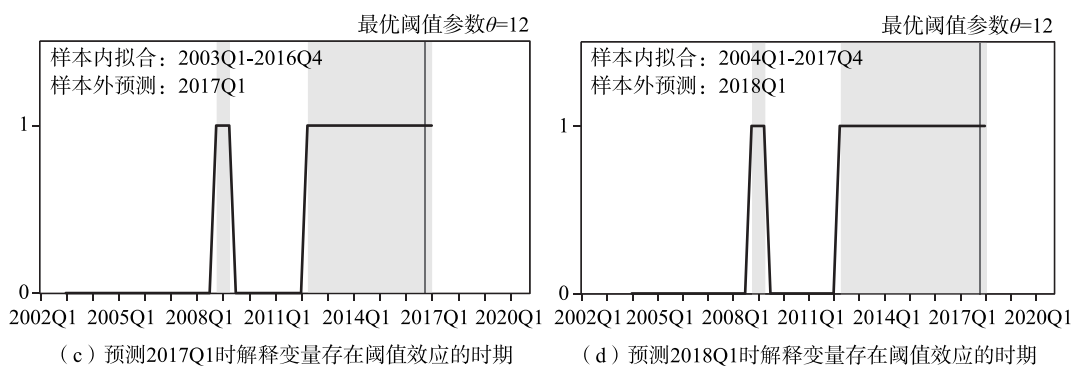
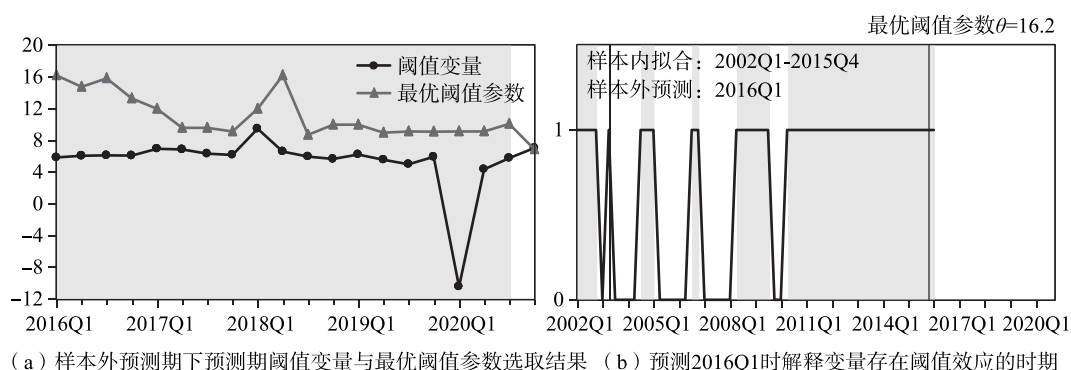


图 4 T-MCP-U-MIDAS 模型在 $h=0$ 步长下的阈值效应结果

2. 预测步长 $h=1/3, 2/3, 1$ 时的变量选择结果

除了上述 $h=0$ 预测步长下的变量选择结果外, 本文还展示了 T-GMCP-U-MIDAS 模型在 $h=1/3, 2/3$ 和 1 预测步长下的变量选择结果说明不同区制内所选取变量的差异性, 如图 5 所示。首先, 从整体上与 $h=0$ 结果对比可得, 随着预测步长 h 的增加, T-GMCP-U-MIDAS 模型倾向于选择更多的变量以保持准确的预测。其次, 从具体的变量选择结果分析可得: (1) 与 $h=0$ 结果一致, 工业增加值 (VAI) 在绝大多数预测期内对季度实际 GDP 增长率的预测产生了持续且重要的影响。(2) 在不同预测步长下其他变量产生的作用有所差异, 例如在预测步长 $h=1/3$ 时, 实际 GDP 增长率滞后项在绝大多数预测内都被选入作为影响实际 GDP 增长率的核心变量; 在预测步长 $h=2/3$ 和 1

时,高频金融变量上证综合指数(SSEC)始终对实际GDP增长率的预测发挥了重要作用。(3)受新冠疫情暴发的影响,自2020Q1以来变量选择的情况变得多而杂,除了发挥重要影响的工业增加值(VAI)和上证综合指数(SSEC)以外,在 $h=1/3, 2/3$ 和1步长下,被选择的变量主要集中于投资、消费、利率和货币供应量(M2)等方面。(4)与 $h=0$ 预测步长下的结果相比,随着预测步长的增加,存在阈值效应的变量逐渐增多,在新冠疫情暴发时期该现象更为明显,且主要集中于运输业、投资、消费、工业企业指标以及债券收益率等方面,这表明在受到重大突发公共卫生事件的冲击而导致经济处于衰退甚至停滞期时,通过恢复中断的交通运输、开放线下消费市场、恢复工厂企业的运行等方面可能会更快且更容易地拉动经济回升。特别地,在 $h=1$ 预测步长下,短期国债收益率(SBOND)存在持续的阈值影响,这表明在信息短缺的背景下该变量通过提供有用的短期领先信息为实际GDP增长率的预测起到了先导作用,该结论与费兆奇和刘康(2019)得出的通过上证综指和国债利率差等变量构建的日度先行指数对GDP的运行具有先导作用的结论类似。

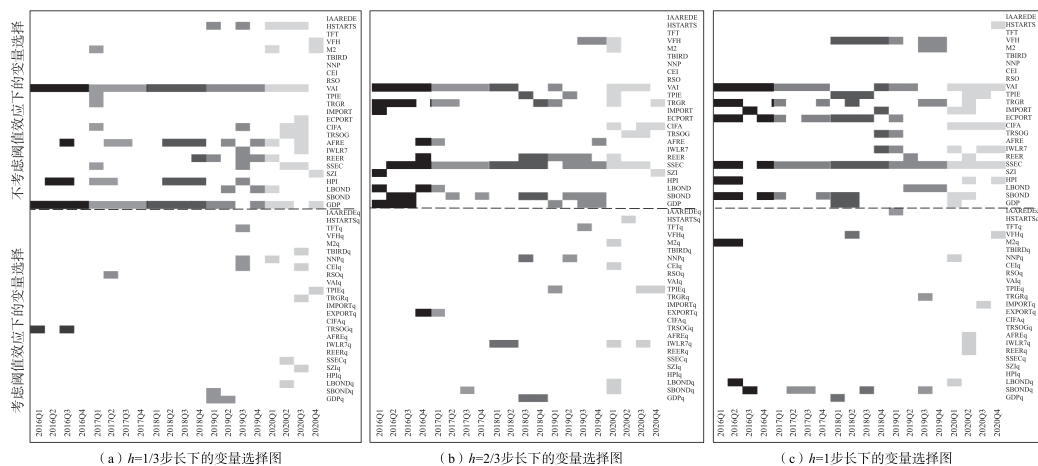


图5 T-GMCP-U-MIDAS模型在不同步长下变量选择对比

五、结论与政策建议

混频数据建模因其能直接利用不同频率数据进行分析的优势,由此获取的预测结果更为精准和及时等优点,近年来备受经济学界的关注(王维国和于扬,2016)。本文基于高维混频数据检验了T-GP-U-MIDAS模型对中国季度实际GDP增长率进行实时预报和短期预测的适用性和有效性,并得出以下几点重要结论与启示:

第一,实时预报与短期预测的结果表明,T-GP-U-MIDAS类模型通过同时实现混频数据分析、变量选择、参数估计,以及引入阈值函数以识别经济运行中可能存在的转变点,在一定程度上改善了我国季度实际GDP增长率实时预报和短期预测的精度,且随着预测步长的增加,预测效果随着可用信息的减少而呈现出递减的趋势。此外,DM检验的结果进一步从统计意义上证实了T-GP-U-MIDAS类模型在实时预报与短期预测上的比较优势。

第二,模型的变量选择结果显示,首先,工业增加值始终与实际 GDP 增长率的预测密切相关,建议对重点地区工业增加值与 GDP 的占比指标建立动态监测和评估机制,进而改善宏观经济管理的绩效。其次,进出口、投资和消费这三辆拉动经济增长的马车,对预测实际 GDP 增长率起到了非常重要的作用,这表明适度的合理投资、以自主创新为主驱动经济发展,促进消费和出口以拉动内需和融入国际国内双循环是促进经济稳定增长的重要掌控方向。最后,少数高频金融变量对实际 GDP 增长率的实时预报和短期预测具有持续且重要的影响,在我国大多数宏观经济变量的发布均存在时滞的背景下,实时更新的高频金融数据为预测我国实际 GDP 增长率提供了及时且有效的信息,对其进行有效监管能够减少股票市场及债券市场的意外冲击对我国投资、物价等方面带来的负向影响,进而降低对我国经济带来的影响。

第三,不同预测步长下的变量选择结果显示,随着预测步长的增加,T-GP-U-MIDAS 模型倾向于选择越来越多的变量以保持准确的预测。存在阈值效应的变量主要集中于新冠疫情暴发时期,且主要集中于投资、消费、进出口、运输业以及工业企业指标等方面,这表明在经济受重大突发公共卫生事件冲击而导致经济处于衰退期时,通过加快企业复工复产、促进消费拉动内需以及恢复中断的交通运输可能会对拉动经济回升发挥重要的作用。

综上所述,当前我国经济容易受到经济转型、重大危机和政策法规调整等非平稳复杂环境的冲击和影响,此时经济预测往往遭遇困境,甚至是失灵。本文的研究以具有先行特征的指示变量为阈值变量,并将其作为阶段性转变的指南针,采用惩罚机制从高维混频变量中筛选出能提高经济预报、监测和预警的核心解释变量,这将有助于非平稳时期对宏观经济总量进行实时预报和短期预测,对把握宏观调控的力度、节奏和方向,实现宏观政策的针对性、前瞻性和有效性,有着非常重要的理论和现实意义。此外,在未来的研究中,可以考虑将 T-GP-U-MIDAS 模型向以下两个方向进行改进:其一是如何在目前模型中包含信息更多的高频阈值变量;其二是目前的模型仅考虑了单变点问题,如何扩展至多变点问题也值得探究。

参 考 文 献

- [1] Breheny, P., and J. Huang, "Group Descent Algorithms for Nonconvex Penalized Linear and Logistic Regression Models with Grouped Predictors", *Statistics and Computing*, 2015, 25 (2), 173-187.
- [2] Ciuperca, G., "Model Selection by LASSO Methods in a Change-Point Model", *Statistical Papers*, 2014, 55 (2), 349-374.
- [3] Dickey, D. A., and W. A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 1981, 49 (4), 1057-1072.
- [4] Diebold, F. X., and R. S. Mariano, "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, 1995, 13 (3), 253-263.
- [5] Fan, J., and R. Li, "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties", *Journal of the American Statistical Association*, 2001, 96 (456), 1348-1360.
- [6] Foroni, C., M. Marcellino, and C. Schumacher, "Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2015, 178 (1), 57-82.

- [7] 费兆奇、刘康, “中国宏观经济波动的高频监测研究——基于混频模型对日度经济先行指数的构建和分析”, 《管理世界》, 2019年第35卷第6期, 第27—38页。
- [8] Ghysels, E., P. Santa-Clara, and R. Valkanov, “The MIDAS Touch: Mixed Data Sampling Regressions”, *Mimeo, Chapel Hill, N C*, 2004.
- [9] 高华川、白仲林, “中国月度GDP同比增长率估算与经济周期分析”, 《统计研究》, 2016年第33卷第11期, 第23—31页。
- [10] Huang, J., P. Breheny, and S. Ma, “A Selective Review of Group Selection in High-Dimensional Models”, *Statist Sci*, 2012, 27 (4), 481-499.
- [11] Hylleberg, S., R. F. Engle, C. W. J. Granger, and B. S. Yoo, “Seasonal Integration and Cointegration”, *Journal of Econometrics*, 1990, 44 (1), 215-238.
- [12] Lee, S., M. H. Seo, and Y. Shin, “The LASSO for High Dimensional Regression with a Possible Change Point”, *Journal of the Royal Statistical Society Series B, Statistical Methodology*, 2016, 78 (1), 193-210.
- [13] 刘汉、刘金全, “中国宏观经济总量的实时预报与短期预测——基于混频数据预测模型的实证研究”, 《经济研究》, 2011年第46卷第3期, 第4—17页。
- [14] 刘汉、王永莲、陈德凯, “混频Granger因果关系检验的功效和稳健性分析”, 《数量经济技术经济研究》, 2017年第34卷第10期, 第144—161页。
- [15] Marcellino, M., and C. Schumacher, “Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP”, *Oxford Bulletin of Economics and Statistics*, 2010, 72 (4), 518-550.
- [16] Marsilli, C., “Variable Selection in Predictive MIDAS Models”, *Banque de France Working Paper* 520, 2014.
- [17] Nakajima, Y., and N. Sueishi, “Forecasting the Japanese Macroeconomy Using High-Dimensional Data.”, *The Japanese Economic Review*, 2022, 73 (2), 299-324.
- [18] Siliverstovs, B., “Short-Term Forecasting with Mixed-Frequency Data: A MIDASSO Approach”, *Applied Economics*, 2017, 49 (13), 1326-1343.
- [19] 尚玉皇、郑挺国, “中国金融形势指数混频测度及其预警行为研究”, 《金融研究》, 2018年第3期, 第21—35页。
- [20] Tibshirani, R., “Regression Shrinkage and Selection via the LASSO”, *Journal of the Royal Statistical Society Series B (Methodological)*, 1996, 58 (1), 267-288.
- [21] Uematsu, Y., and S. Tanaka, “High-Dimensional Macroeconomic Forecasting and Variable Selection via Penalized Regression”, *The Econometrics Journal*, 2017, 22 (1), 34-56.
- [22] Wang, L., G. Chen, and H. Li, “Group SCAD Regression Analysis for Microarray Time Course Gene Expression Data”, *Bioinformatics*, 2007, 23 (12), 1486-1494.
- [23] 王维国、于扬, “基于混频回归类模型对中国季度GDP的预报方法研究”, 《数量经济技术经济研究》, 2016年第33卷第4期, 第108—125页。
- [24] Xu, Q., X. Zhuo, C. Jiang, X. Liu, and Y. Liu, “Group Penalized Unrestricted Mixed Data Sampling Model with Application to Forecasting US GDP Growth”, *Economic Modelling*, 2018, 75, 221-236.
- [25] Yuan, M., and Y. Lin, “Model Selection and Estimation in Regression with Grouped Variables”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, 68 (1), 49-67.
- [26] Zhang, C.-H., “Nearly Unbiased Variable Selection under Minimax Concave Penalty”, *The Annals of statistics*, 2010, 38 (2), 894-942.
- [27] 张劲帆、刚健华、钱宗鑫、张龄琰, “基于混频向量自回归模型的宏观经济预测”, 《金融研究》, 2018年第7期, 第34—48页。
- [28] 周建、况明, “中国宏观经济动态传导、可靠性及货币政策机制”, 《经济研究》, 2015年第50卷第2期, 第31—46页。
- [29] 郑挺国、尚玉皇, “基于金融指标对中国GDP的混频预测分析”, 《金融研究》, 2013年第9期, 第16—29页。
- [30] 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年第48卷第6期, 第58—

70 页。

[31] 曾津、周建军, “高维数据变量选择方法综述”, 《数理统计与管理》, 2017 年第 36 卷第 4 期, 第 678—692 页。

[32] 周德才、朱志亮、纪应心、余永奎, “混频非对称金融景气指数编制及应用研究——基于 MF-MS-DFM 模型的经验分析”, 《管理评论》, 2019 年第 31 卷第 8 期, 第 71—83 页。

Nowcasting and Short-Term Forecasting Macroeconomic Aggregates Based on the High-Dimensional Mixed-Frequency Information Identification and Selection

LIU Han LIU Ying

(Jilin University)

ZHUO Xingxuan*

(Fuzhou University)

Abstract: The Threshold Group Penalized Unrestricted Mixed Data Sampling (T-GP-U-MIDAS) model is proposed to nowcast and forecast China's quarterly real GDP growth rate, and the accuracy improvement mechanism of the model is studied. The empirical results show that the forecast performance of the T-GP-U-MIDAS model is superior to benchmarks. The T-GP-U-MIDAS model could identify possible change points in economic operations by introducing a leading threshold indicator, and then select a series of core variables that affect the real GDP growth rate under different economic conditions, resulting in an improvement in predictive and explanatory capabilities of the model.

Keywords: high-dimensional mixed-frequency data; T-GP-U-MIDAS model; real GDP growth rate

JEL Classification: C12, C22, C53

* Corresponding Author: Zhuo Xingxuan, School of Economics and Management, Fuzhou University, No.2 Xue Yuan Road, University Town, Fuzhou, Fujian 350108, China; Tel: 86-15060069957; E-mail: zhuoxx1991@fzu.edu.cn.