

深度学习神经网络能改进 GDP 的预测能力吗?

肖争艳 刘玲君 赵廷蓉 陈彦斌

【摘要】目前国内外复杂的经济形势加大了预测 GDP 的难度,因此,如何有效地预测 GDP 是值得研究的重要理论与现实问题。有鉴于此,本文构建了既具有宏观经济理论基础又符合中国宏观经济特征的指标体系,并构造了一个用于 GDP 预测分析的 LSTM 模型,将之与 BVAR 模型进行对比研究,以科学地判断 LSTM 模型是否能够提升 GDP 预测的精确度。研究结果表明:第一,本文选择的扩展指标能够提升 BVAR 模型与 LSTM 模型的 GDP 预测能力;第二,相比于 BVAR 模型,LSTM 模型能够更好地挖掘扩展指标对 GDP 的非线性影响,从而提升短期 GDP 预测能力。鉴于 LSTM 模型强大的自我学习能力、良好的泛化能力以及较好的模型可调节性,LSTM 模型在 GDP 预测领域具有广阔前景。

【关键词】神经网络;LSTM 模型;BVAR 模型;GDP 预测

【作者简介】肖争艳,中国人民大学统计学学院教授;刘玲君,中国人民大学经济学院硕士研究生;赵廷蓉,中国人民大学统计学学院硕士研究生;陈彦斌(通讯作者),中国人民大学经济学院教授,博士生导师(北京 100872)。

【原文出处】《经济与管理研究》(京),2020.7.3~17

【基金项目】教育部人文社会科学重点研究基地重大项目“‘十三五’时期中国宏观调控体系的改革与转型问题”(18JD790015);国家自然科学基金应急管理项目“国内经济政策环境与金融风险防范”(71850003)。

一、问题提出

GDP 增速是国家进行宏观调控和企业进行投融资决策的关键参考变量之一,预测 GDP 增速一直是宏观经济学的重要研究领域。然而,目前国内外复杂的经济形势加大了预测 GDP 的难度,给经济增长目标和宏观政策的制定提出了新的挑战。一方面,随着供给侧结构性改革的不断推进,许多新产业的出现使得中国产业结构发生了较大的变化,经济的结构性特征对经济增长的影响越来越重要。另一方面,近年来贸易保护主义抬头,国际贸易摩擦加剧,为中国经济增长带来了极大的不确定性。在此背景下,如何有效地预测 GDP 是值得研究的重要问题。更重要的是,国内经济结构与国际经济环

境的变化要求选择合适的预测方法,以确保预测精确度。

预测 GDP 的已有研究主要使用两类模型。一类是线性模型,该类模型以贝叶斯向量自回归(BVAR)模型为代表。王飞(2011)以青海省为例建立了一个 BVAR 模型,样本内和样本外的预测误差都表明 BVAR 模型的 GDP 预测效果优于向量自回归(vector autoregression, VAR)、差分整合移动平均自回归(autoregressive integrated moving average, ARIMA)等线性模型^[1]。周建和况明(2015)构建 BVAR 模型对中国 GDP 预测误差进行分析,发现 BVAR 模型优于现有文献中传统的 VAR 方法^[2]。

另一类 GDP 预测模型是非线性模型,该类模型

以人工神经网络 (artificial neural network, ANN) 为代表。ANN 作为人工智能领域的非线性方法,能够处理非线性、不连续和高频多维的数据,因而该模型在 GDP 预测领域得到越来越多的关注。特卡茨 (Tkacz, 2001) 构建神经网络对加拿大年度 GDP 增速进行预测分析,发现用该模型训练得到的预测误差比线性模型降低约 25%^[3]。欧邦才 (2004) 采用 BP 神经网络对中国年度 GDP 进行预测,发现预测效果较好^[4]。郭庆春和何振芳 (2014) 使用改进的 BP 神经网络建立 GDP 预测模型,研究表明该模型具有较高的预测精度,并且还具有一泛化能力强、训练速度快、稳定性好等优点^[5]。传统的 ANN 模型虽然能保证网络学习过程的最终收敛,但不能反映样本之间的时序关系,而在经济领域普遍存在的这种时序关系及其背后的逻辑关系与理论关系对于预测分析是大有裨益的。相比之下,深度学习模型中的长短期记忆神经网络模型 (long short term memory, LSTM) 由于其出色的时序数据处理能力,在经济与金融预测领域得到了较为广泛的应用与推广。崔和李 (Choi & Lee, 2018) 使用 LSTM 模型来预测实时时间序列数据,结果表明 LSTM 模型取得了已知各类模型中的最好预测效果^[6]。赵军豪等 (2018) 等使用融合微博情感分析和 LSTM 模型进行预测,发现该方法对数据抖动有良好的适应性,并具有很好的稳健性^[7]。

使用神经网络方法预测 GDP 的已有文献较丰富,也取得了显著的学术进展,但仍有可进一步改进的地方。首先,虽然以往研究对短期 GDP 预测算法做了改进,但鲜有文献对线性与非线性两类模型的预测效果进行比较。宏观经济指标体系是一组时间序列数据,在这组数据中部分指标会对核心变量 GDP 产生非线性的影响。但是,非线性模型是否能够抓住这种非线性影响?非线性模型与线性模型相比预测效果是否会有显著提高?对于这两个问题已有文献尚未给出明确答案。其次,已有研究的宏观指标选择往往没有以经济学理论为依据,可能会增加模型预测误差。已有 GDP 预测方面的文

献在进行预测指标选择时,大多存在预测指标与宏观经济理论分离的现象,这可能会导致遗漏重要指标的问题出现,从而增大了 GDP 预测的误差。最后,对于如何选择影响模型短期预测能力的指标,现有文献往往依据国际相关经验,而对于是否符合国情考虑较少。由于中国经济增长具有明确的国情特征尤其是结构性特征,与欧美等其他国家差异较大,所以有必要依据中国经济特点来选择预测指标。以房地产市场为例,陈小亮 (2017) 测算发现,2016 年中国房地产对 GDP 增长的总贡献率高达 18.3%^[8]。因此,在选择中国 GDP 预测影响因素时,不能忽略房地产等切实反映中国国情的宏观指标。

为了有效解决以上问题,本文在已有研究文献的基础上进行改进,主要包括两个方面。第一,本文选择线性模型的代表 BVAR 模型和非线性模型的代表 LSTM 模型进行预测,并且采用常用的经济预测误差指标对两种模型的预测能力进行度量。BVAR 模型在预测效果和稳健性方面均有较好表现^[1],是在经济预测领域得到广泛应用的线性模型。LSTM 模型基于当前热门的深度学习神经网络技术,并且具有较为优异的时序数据处理能力,是 GDP 预测领域较为值得推广的非线性模型。因此,本文选择 BVAR 模型作为参照,比较分析 LSTM 模型的预测能力。此外,为对两模型的预测能力进行充分的对比,本文分别在动态滚动预测与固定估计窗口预测两种方式下,采用 BVAR 模型与 LSTM 模型预测 GDP,从而充分比较两个模型的预测误差以判断二者的差异及优劣。第二,本文构建了既具有宏观经济理论基础又符合中国国情的宏观经济指标体系,本文首先以主流宏观理论中的 IS-LM 模型为理论基础构建了基准指标体系,然后在此基础上纳入商品房本年新开工面积同比增速、宏观经济景气先行指数、沪市成交额同比增速等多个可以反映中国宏观经济特征的扩展指标。

本文的发现主要有两点:第一,本文选择的扩展指标能够提升 BVAR 模型与 LSTM 模型的 GDP

预测能力。一方面, BVAR 模型和 LSTM 模型的样本外预测结果在扩展指标体系下对 GDP 实际值的拟合度更高。另一方面, 加入扩展指标后, BVAR 模型和 LSTM 模型的预测误差值呈现下降的趋势。因此, 本文选择的扩展指标能够提升 BVAR 模型与 LSTM 模型的 GDP 预测能力。第二, 相比于 BVAR 模型, LSTM 模型能够更好地挖掘扩展指标对 GDP 的非线性影响, 从而提升短期 GDP 预测能力。在动态滚动预测与固定估计窗口预测情形下, LSTM 模型的预测效果均优于 BVAR 模型。并且, 当预测步距增大时, LSTM 模型预测效果要优于 BVAR 模型。具体而言, 在各个预测步距下, LSTM 模型的预测效果优于 BVAR 模型, 并且随着向前步距的增加, 预测优势显著提升。

二、BVAR 模型和 LSTM 模型的设定与求解

为了分析 LSTM 模型是否能够提升 GDP 的预测能力, 本文首先选取了在经济预测中得到广泛应用的 BVAR 模型, 对 GDP 进行预测。同时, 本文选取了深度学习神经网络中的 LSTM 模型进行了相同的预测, 旨在对两模型的 GDP 预测能力进行对比分析。LSTM 模型由于其网络结构能够反映样本之间的时序关系, 是众多神经网络算法中较为适合于经济学研究的模型。

(一) BVAR 模型

1. BVAR 模型介绍

比较常见的 GDP 预测方法主要有单变量 ARIMA 模型和多变量 VAR 类模型等线性模型。ARIMA 模型作为一种经典的非平稳时间序列分析方法, 擅长发现时间序列自身的影响规律, 在 GDP 预测领域有较为广泛的应用。然而, 由于 ARIMA 模型仅包含了 GDP 单个变量, 并没有考虑与其他经济变量之间的联系, 显然会损失大量有价值的信息。为此, 西姆斯 (Sims, 1980) 将之进一步扩展到多变量模型, 提出了 VAR 模型^[9]。VAR 模型目前已成为各国中央银行广泛使用的宏观预测模型。然而, 高维 VAR 模型面临参数过多的问题, 对模型中参数进行准确估计存在着明显困难。为了有效解决

这一问题, 越来越多的文献采用贝叶斯方法来估计 VAR 模型。BVAR 由多恩等 (Doan et al., 1984)^[10] 提出, 模型的优势在于利用具有信息含量的先验分布来弥补样本中信息缺乏的问题, 从而降低参数估计的不确定性。

BVAR 模型用贝叶斯方法对动态面板的 VAR 系统进行分析。建立一个由 m 个变量组成、滞后阶数为 p 、含有常数项的 VAR(p) 模型, 即:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + A_c + u_t \quad (1)$$

其中, A_c 是常数矩阵; $y_t = (y_{1,t}, y_{2,t}, \dots, y_{m,t})^T$, 包含 GDP、CPI、M2 等宏观变量; 残差 $u_t \sim i. i. d. N_m(0, \Sigma)$, Σ 为 $m \times m$ 正定矩阵; A_j 为 $m \times m$ 阶系数矩阵, $\alpha_{i,j,\tau}$ 为方程 i 中变量 j 滞后 τ 期的系数。

2. 先验分布与后验分布

BVAR 模型利用了来源于经验和历史资料的先验信息, 减少了需要赋值的参数的数量, 因此, 在估计 BVAR 模型的过程中, 需要对式 (1) 中 Σ 与 A_j 的参数分布做出先验假设, 本文采用实际应用中最为广泛使用的明尼苏达 (Minnesota) 先验分布。明尼苏达先验分布假定协方差阵 Σ 的先验分布取为扩散先验分布。同时, 假设 $\alpha_{i,j,\tau}$ 和 Σ 相互独立, 且满足:

$$\alpha_{i,j,\tau} \sim i. i. d. N(\delta_{i,j,\tau}, V_{i,j,\tau}) \quad (2)$$

由于宏观经济变量具有较强的持续性, 因此模型设定: 当 $i=j, \tau=1$ 时, $\delta_{i,j,\tau} = 0.9^{[11]}$, 其他情况下等于 0, 即方程中变量自身的一阶滞后变量系数的先验期望为 0.9, 其他系数的先验期望为 0。明尼苏达先验分布假设方程 i 中各系数的方差矩阵 $V_{i,j,\tau}$ 为对角矩阵, $V_{i,j,j}$ 表示 V_i 中的对角线元素, 则:

$$V_{i,j,j} = \begin{cases} \frac{a_1}{r^2}, & i=j \\ \frac{a_2 \sigma_{ii}}{r^2} & i \neq j \end{cases} \quad (3)$$

由此得到 α 的后验分布服从正态分布:

$$\alpha_{i,j,\tau} | y \sim N(\bar{\alpha}_{i,j,\tau}, \bar{V}_{i,j,\tau}) \quad (4)$$

3. 超参数的选择与设定

BVAR 模型将众多参数的估计问题转化为对

a_1 和 a_2 这两个超参数的估计。超参数 (a_1, a_2) 有两种取值方法。第一, 参照多恩等 (1984) 的做法, 给定超参数的可能性取值, 通过比较所有超参数组合的均方根误差 (RMSE) 来选择最优的参数组合^[10]。第二, 使用贝叶斯方法, 在最大似然函数的基础上计算出超参数的取值。由于第二种方法在使用软件计算时较为困难, 因此并没有得到广泛的实际应用。而且, 由于本文 BVAR 模型采用的明尼苏达先验分布设定和多恩等 (1984) 的方法十分接近, 因此本文采用第一种处理方法。给定超参数的可能性取值 (具体取值参照多恩等 (1984) 的做法), 其中, $a_1 = \{0.0005, 0.0008, 0.001, 0.05, 0.1, 0.2, 0.3, 0.4\}$, $a_2 = \{0.00025, 0.0004, 0.0005, 0.025, 0.05, 0.1, 0.15, 0.2\}$ 。

在给定可能性取值的基础上, 本文参照肖争艳等 (2009)^[12] 的做法, 最终所使用的最优参数的具体选择过程如下。首先, 确定 BVAR 模型的最优滞后阶数。本文采用与 VAR 模型一致的标准, 即根据 LR、FPE、AIC、HQ 和 SC 等准则, 选取最优滞后阶数为 2 阶。然后, 在超参数的可能性取值中选取 a_1 与 a_2 的组合, 并使用泰尔混合估计方法对 BVAR 模型参数进行估计。最后, 比较所有超参数组合的均方根误差 (RMSE) 选择最优的参数组合为 $\{0.05, 0.025\}$ 。

(二) LSTM 模型

1. LSTM 模型介绍

目前, ANN 各类模型是非线性模型的代表模型。ANN 作为人工智能领域的非线性方法, 能够处理非线性、不连续和高频多维的数据, 因而该模型

在 GDP 预测领域得到越来越多的关注。传统的 ANN 模型虽然能保证网络学习过程的最终收敛, 但是也存在两点显著的局限性: 其一, 容易陷入局部最小值; 其二, 不能反映样本之间的时序关系, 而在经济领域普遍存在的这种时序关系及其背后的逻辑关系与理论关系, 对于预测分析是大有裨益的。因此, 本文选用深度学习模型中的 LSTM 模型进行 GDP 预测。

LSTM 模型最早由霍克赖特和施米德胡贝 (Hochreiter & Schmidhuber, 1996)^[13] 提出, 由于其能更好地发现长期依赖关系而被广泛用于处理序列信息, 如语音识别、机器翻译等。后来由于其出色的时序数据处理能力, 在经济与金融预测领域得到了较为广泛的应用与推广。LSTM 模型的具体预测过程通过其基本单元模型的激活函数及神经元之间数据结果的传递来实现。换言之, 可以将 LSTM 模型视为若干的线性和非线性函数相互嵌套的模型。

对于 GDP 预测而言, 由于近期和远期历史数据都会对 GDP 产生影响, 但近期历史数据影响会更强, 故而本文选择设有三个门控 (gates) 的 LSTM 模型来实现对信息传递的控制, 筛选出对预测有用的部分。每个神经元结构如图 1 所示。

其中, x_t 是指 t 时刻的输入向量, 包含 GDP、CPI、M2 等宏观变量的历史数据。 h_{t-1} 表示上一时刻的输出, c_{t-1} 表示上一时刻的记忆。遗忘门 (forget gate) 决定了要遗忘多少旧记忆, 更新门 (update gate) 决定要更新多少新记忆, 并由新记忆和旧信息相加形成最终的记忆, 最后由输出门 (output gate) 从最终记忆中筛选出对预测有用的部分。

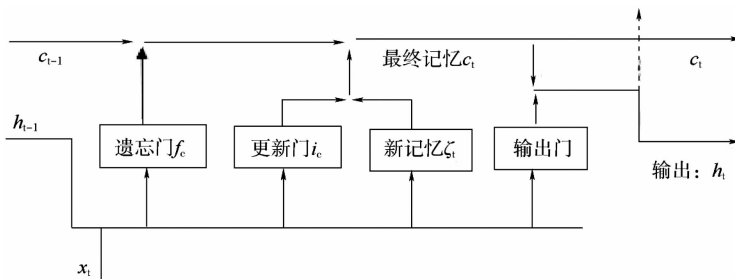


图 1 LSTM 神经元结构

LSTM 模型相比 BVAR 模型,主要具备如下两个优点:第一,BVAR 模型只能捕捉数据之间的线性关系,而 LSTM 模型通过其非线性的激活函数,能够识别与挖掘变量之间、历史数据与新数据之间的非线性关系;第二,BVAR 模型利用的历史信息受到模型设定的滞后阶数的限制,然而,对于 LSTM 模型而言,由于遗忘门、更新门的存在,其对数据具有长记忆性,从而能够利用更长时间序列的数据。

2. 训练方法及优化器选择

LSTM 模型通过数据训练帮助模型更新权重,在训练 LSTM 模型的过程中,选择合适的训练方法和优化器能够提升模型的训练效率。第一,选择小批量(Mini - Batch)方法作为本文的训练方法。具体而言,对 LSTM 模型训练时,根据深度学习神经网络的常用做法,每一次训练均基于随机抽取的 30 个样本点,避免了用全部样本点训练导致的训练过程慢以及过拟合问题。第二,为使 LSTM 模型快速收敛,本文采用作为自适应学习率优化算法的适应性矩估计(adaptive moment estimation, Adam)优化器进行优化训练。与传统的优化算法将学习率设为一个常数不同,Adam 优化算法会采取一些措施来更新学习率,并且较为容易实现,有较高的计算效率和较低的内存需求。

3. 超参数的设定和网络结构的选择

由于训练 LSTM 模型需要对数据进行训练集与验证集的划分,本文参照已有研究的常用划分方式,在初始训练中采用 1997 - 2004 年共 32 个季度的数据作为训练集,2005 - 2006 年共 8 个季度数据

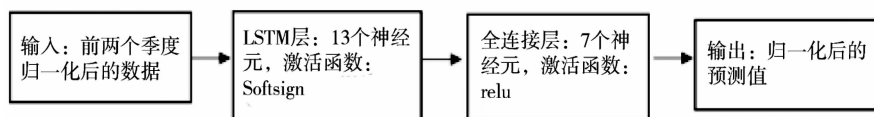
作为验证集。训练集数据用来帮助模型更新权重,而验证集的数据则主要用来评估模型。并且,在后续预测过程中,模型会随样本量增加而不断扩充训练集。模型需要调试的目标包括 LSTM 层的神经元个数、训练周期的期数等。训练时,采用 Mini - Batch Adam 算法进行更新,再根据模型在验证集上的均方误差(MSE)来衡量模型的精度,选择精度最高的模型结果。本文对于基准指标体系和扩展指标体系选用的网络结构如图 2 所示。

(三) 动态预测过程

令 $y_{\tau+h}$ 表示在时刻 τ 对 $\tau+h$ 时的宏观经济变量 y 的预测值, $y_{\tau+h}^0$ 表示 $y_{\tau+h}$ 的实际观测值。为了全面分析模型的预测效果,本文不仅进行相邻一步预测,而且进行远期预测。当进行相邻一步预测时, h 设置为 1; 当进行远期预测时, h 分别设置为 2, 3, 4。由于相邻一步预测与远期预测的差别主要在于预测值的滞后期数不同,所以下面以相邻一步预测为例,介绍模型具体的预测过程。

参照张劲帆等(2018)^[14]的做法,本文的预测方式包括动态滚动预测与固定估计窗口预测两种。当进行动态滚动预测时,具体的预测步骤为:(1)把 1997Q1 - 2006Q4 作为样本来预测滞后一期的数值,即 2007Q1 时 y 的数值。(2)扩充样本区间为 1997Q1 - 2007Q1,将真实观测值作为样本来预测 2007Q2 时 y 的数值,预测步骤同(1)。(3)进一步扩充样本区间为 1997Q1 - 2007Q2,将真实观测值作为样本来预测 2007Q3 时 y 的数值。重复以上扩充样本滚动预测过程,直至全部样本结束。当进行

基准模型:



扩展模型:

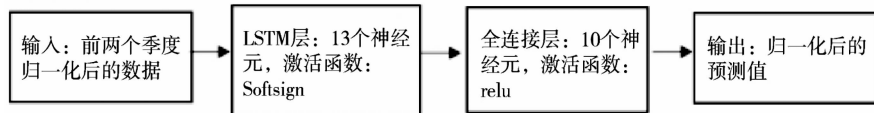


图 2 模型网络结构

固定窗口预测时,具体的预测步骤为:(1)把1997Q1-2006Q4作为样本来预测滞后一期的数值,即2007Q1时 y 的数值。(2)扩充样本区间为1997Q2-2007Q1,将真实观测值作为样本来预测2007Q2时 y 的数值,预测步骤同(1)。(3)进一步扩充样本区间为1997Q3-2007Q2,将真实观测值作为样本来预测2007Q3时 y 的数值。重复以上扩充样本滚动预测过程,直至全部样本结束,即可得到从2007Q1开始的GDP的预测值序列。

(四) 预测误差指标

为对模型预测能力进行更为严谨的分析,本文参照已有文献使用均方误差(MSE)、平均绝对百分比误差(MAPE)、对称平均绝对百分比误差(SMAPE)作为模型样本外预测能力的测度指标。以上三个指标具体计算方法如下:

$$MSE = \frac{1}{N} \sum_{\tau=t_0}^{t_{N-1}} (y_{\tau} - y_{\tau}^0)^2 \quad (5)$$

$$MAPE = \frac{100\%}{N} \sum_{\tau=t_0}^{t_{N-1}} \left| \frac{y_{\tau} - y_{\tau}^0}{y_{\tau}^0} \right| \quad (6)$$

$$SMAPE = \frac{100\%}{N} \sum_{\tau=t_0}^{t_{N-1}} \frac{|y_{\tau} - y_{\tau}^0|}{(|y_{\tau}| + |y_{\tau}^0|)/2} \quad (7)$$

这里, y_{τ} 表示第 τ 期的预测值, y_{τ}^0 表示第 τ 期的实际观测值, t_0 表示2007Q1。

除此之外,MSE的提高虽然在绝对数量上能够体现不同模型的预测效果的改善程度,但并不能在统计分布上检验改善的效果是否显著。因此本文在强调MSE的基础上引入D-M检验来检验预测效果。D-M检验由戴博和马里亚诺(Diebold & Mariano, 1995)提出^[15],该统计量服从标准正态分布,具体表达式如下:

$$DM = \frac{S - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \sim N(0,1) \quad (8)$$

其中, $S = \sum_{\tau=1}^{\tau} I + (d_{\tau} \text{rank}(|d_{\tau}|))$, $I + (d_{\tau}) = \begin{cases} 1, & d_{\tau} > 0 \\ 0, & d_{\tau} < 0 \end{cases}$ 是指示函数, d_{τ} 表示两种预测方法对应损

失函数的差分序列, $\text{rank}(|d_{\tau}|)$ 表示 d_{τ} 取绝对值,按从小到大的顺序进行排序后的对应序数值。

三、宏观经济指标选择

在模型的预测指标选择方面,已有文献大多存在指标选择与宏观经济理论分离的现象,这可能导致遗漏重要指标的问题出现。同时,由于中国经济增长具有明确的国情特征尤其是结构性特征,与欧美等其他国家差异较大,如果在指标选择过程中忽略中国经济特点,极易导致GDP预测误差增加。有鉴于此,本文在已有研究基础上做出两方面改进:一,以IS-LM模型为基础构建了基准指标体系;二,纳入反映中国国情的扩展指标,以建立科学合理的宏观经济指标体系,从而更加准确地对GDP进行预测。

本文首先基于凯恩斯的理论,构建短期GDP预测的基准指标体系。具体来说,凯恩斯四部门IS-LM模型的表达式为:

$$\text{IS 曲线: } Y = C + I(r) + G + NE$$

$$\text{LM 曲线: } M/P = L(r, Y)$$

其中, Y 代表总产出, C 代表消费, I 代表投资, G 代表政府购买, NE 代表净出口, M 代表货币供给, P 代表价格水平。模型中经济体的均衡是IS曲线和LM曲线的交点,该均衡点给出了同时满足产品市场与货币市场均衡条件的利率 r 与总产出 Y 的组合^[16]。

本文参照陈彦斌(2019)^[17]的做法,选择如下具体指标:(1)季度实际GDP增长率(GDP),该指标用来衡量总产出(Y)的增速,是宏观经济预测的核心指标;(2)季度工业增加值同比增速(IVA),该指标体现了工业企业在一定时期内工业生产活动创造的价值,反映了生产单位或部门对国内生产总值的贡献,是国内生产总值常用的补充指标;(3)季度实际社会消费品零售总额增长率(C),该指标能够反映国内消费情况,对判断当前宏观经济发展状况和未来经济走势具有重要的指导作用;(4)季度实际固定资产投资增长率(I),本文选择该指标作为投资的指标,原因在于固定资产投资是投资中最

重要的组成部分,是监测宏观经济走势、进行宏观调控的重要依据;(5) 贷款实际利率(r),实际利率反映了企业实际面临的间接融资成本,同时,基于数据可得性,本文以该指标为模型的利率指标;(6) 季度一般公共预算支出当期同比增速(G),该指标是较为常用的政府购买指标;(7) 季度非金融部门杠杆率($DEBT$),该指标体现了一定时期内的债务负担,会对经济产出水平产生重要影响^[17-18],本文以该指标作为季度一般公共预算支出当期同比增速(G)的补充指标;(8) 季度出口额同比增长率($EXPORT$),该指标能够较好地衡量市场开放程度,当该指标提高时,出口增速加快,表明经济运行情况良好,反之若该指标下降,出口增速下降,表明宏观经济不景气,增长速度放缓;(9) 消费者物价指数(CPI),本文以该指标衡量通货膨胀水平^[2,19];(10) 季度 $M2$ 同比增长率($M2$)。该指标是常用的货币供给变化指标。

为完整刻画中国宏观经济特征,除基准指标体系中的指标外,本文在扩展指标体系中加入了房地产市场、第三产业增加值、宏观经济景气先行指数、发电量、金融市场等影响因素。扩展指标体系中包括了反映中国宏观经济结构性特征的结构指标和常用的经济先行指标:(1) 商品房本年新开工面积同比增速($HSTARTS$),该指标能够反映建筑业对国民经济的影响,而建筑业的发展可以带动近 30 多个上下游相关产业发展,是第二产业的重要组成部分;(2) 第三产业增加值当期同比实际增速($THIRDI$),第三产业增加值在国内生产总值中占据着较大的比例,与国内生产总值具有较一致的变化趋势^[14];(3) 季度的宏观经济景气先行指数(LI),该指标由国家统计局发布,是当前较为权威的经济先行指标,在 GDP 预测领域得到了广泛应用;(4) 发电量同比增速($ELEC$),该指标为克强指数^①的重要组成部分,是反映地区工业生产活跃度以及工厂开工率的重要指标;(5) 沪市成交额同比增速($STOCK$),该指标是衡量股票市场流动性的重要

指标,同时金融指标对宏观经济变量有显著的预测能力^[12]。

以上指标的数据均来自于中经网统计数据库,其中 5 个为季度频次指标,余下 10 个为月度频次指标。由于核心指标 GDP 为季度数据,为保持数据频率的一致性,本文参考周建和况明(2015)^[2]的做法,将指标均处理为季度数据。各指标的统计频率及本文对指标的具体处理方法如表 1 的第 4 列、第 5 列所示。基于数据的可获得性,所有指标时间跨度为 1997 年第一季度(1997Q1)至 2018 年第四季度(2018Q4),各指标定义及描述性统计详见表 1 和表 2。

四、预测结果分析

基于上一部分所构建的模型,本文试图回答如下问题:第一,在预测 GDP 的过程中,是否有必要在指标体系中纳入更适宜中国国情的扩展指标? 第二,作为线性模型代表的 BVAR 模型与作为非线性模型代表的 LSTM 模型,预测 GDP 的能力究竟孰强孰弱? 第三,各个扩展指标的加入对于模型预测效果的影响程度如何? 为了回答这些问题,本文从多个角度对比各模型的预测误差:通过比较 BVAR 模型、LSTM 模型在纳入扩展指标前后预测误差的变化,即可分析在指标体系中纳入扩展指标的必要性;通过比较 BVAR 模型与 LSTM 模型的预测精确度,即可判断 BVAR 模型与 LSTM 模型预测能力强弱;通过比较剔除各个扩展指标后模型的预测误差,即可定量分析 LSTM 模型下各个扩展指标的预测效果。

(一) 在指标体系中纳入扩展指标的必要性分析

结合图 3、图 4 与表 3 可以看出,加入反映中国国情的扩展指标能够提升 BVAR 模型与 LSTM 模型的 GDP 预测能力。本部分分别将 BVAR 模型的基准指标体系与扩展指标体系、LSTM 模型的基准指标体系与扩展指标体系的预测效果进行了对比,发现本文选取的扩展指标能够提升模型的短期 GDP 预测能力。

表 1 指标的含义与处理方式

指标类型	指标	指标名称	统计频率	指标的处理
基准指标	<i>GDP</i>	实际 <i>GDP</i> 同比增长率	季度	
	<i>CPI</i>	<i>CPI</i> 同比增长率	月度	月度数据减去 100, 以三个月的平均值为季度值
	<i>M2</i>	货币与准货币期末同比增长率	季度	
	<i>r</i>	贷款实际利率	月度	月度贷款基准利率数据减去 <i>CPI</i> 增长率, 以三个月的平均值为季度值
	<i>C</i>	社会消费品零售总额同比增长率	月度	以三个月的平均值为季度值
	<i>I</i>	实际固定资产投资增长率	季度	
	<i>EXPORT</i>	出口额(美元)同比增长率	月度	以三个月的平均值为季度值
	<i>IVA</i>	工业增加值同比增速	月度	以三个月的平均值为季度值
	<i>DEBT</i>	非金融部门杠杆率	季度	季度数据减去 100
	<i>G</i>	一般公共预算支出当期同比增速	月度	以三个月的平均值为季度值
扩展指标	<i>HSTARTS</i>	商品房本年新开工面积同比增速	月度	以三个月的平均值为季度值
	<i>LI</i>	宏观经济景气先行指数	月度	月度数据减去 100, 以三个月的平均值为季度值
	<i>ELEC</i>	发电量同比增速	月度	以三个月的平均值为季度值
	<i>STOCK</i>	沪市成交额同比增速	月度	以三个月的平均值为季度值
	<i>THIRDI</i>	第三产业增加值当期同比实际增速	季度	

表 2 主要指标的描述性统计

指标	样本数	平均值	标准差	最小值	最大值
<i>GDP</i>	88	0.091	0.021	0.064	0.150
<i>CPI</i>	88	0.019	0.021	-0.022	0.080
<i>M2</i>	88	0.156	0.042	0.080	0.293
<i>r</i>	88	0.040	0.022	-0.006	0.088
<i>C</i>	88	0.110	0.032	0.018	0.177
<i>I</i>	88	0.177	0.085	-0.002	0.403
<i>EXPORT</i>	88	0.150	0.152	-0.235	0.410
<i>IVA</i>	88	0.115	0.040	0.056	0.190
<i>DEBT</i>	88	0.727	0.458	0.074	1.559
<i>G</i>	88	0.173	0.084	-0.086	0.378
<i>HSTARTS</i>	88	0.167	0.183	-0.220	0.674
<i>LI</i>	88	0.013	0.019	-0.028	0.055
<i>ELEC</i>	88	0.083	0.058	-0.056	0.265
<i>STOCK</i>	88	0.010	0.024	-0.008	0.119
<i>THIRDI</i>	88	0.099	0.021	0.072	0.173

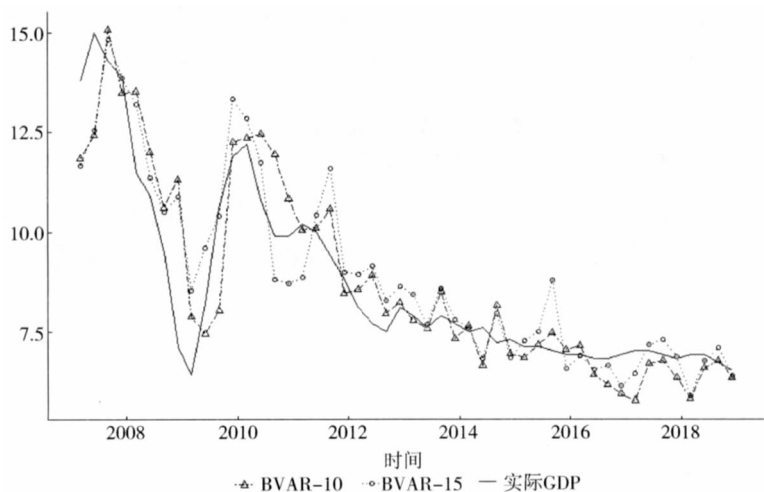


图3 BVAR模型的GDP样本外预测结果

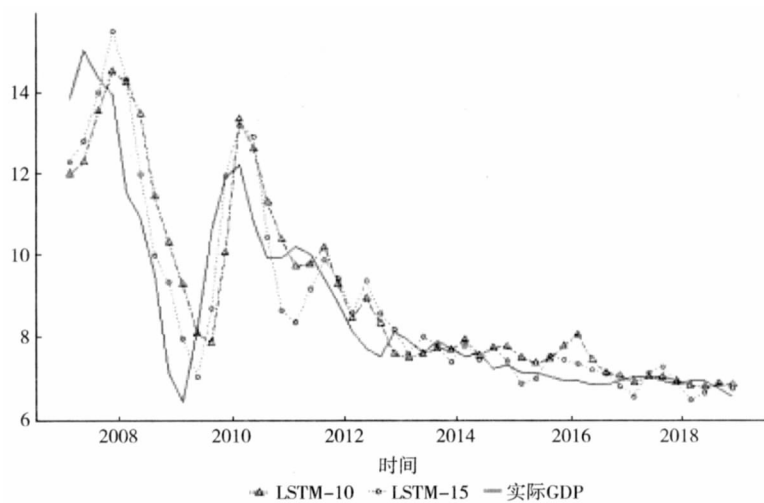


图4 LSTM模型的GDP样本外预测结果

首先,对BVAR模型和LSTM模型的样本外预测图进行定性分析。一方面,从图3和图4可以较为直观地对比分析不同指标体系下两类模型的预测结果。从图3可以看出,纳入扩展指标后,BVAR模型在2007Q1-2009Q1时间区间内的预测值与纳入扩展指标前没有明显差异,在2009Q2-2013Q4时间区间内预测稳定性显著提高,在2014Q1-2018Q4的时间区间内预测值对GDP实际值的拟合度更强。因此,从总体上看,纳入扩展指标能够提升BVAR模型的预测能力。同时,由图4可以看出,纳入扩展指标后,LSTM模型拟合的时效性有了显著提升,具体体现在扩展指标体系下LSTM模型对2007Q1-2012Q1时间区间的预测滞后性问题得

到了较大程度的解决,同时模型对于2012Q2-2018Q4的预测结果与基准指标体系下的预测结果没有明显差异,均与实际值拟合程度较高。

其次,采用预测误差更加严谨地对不同指标体系下两类模型的预测效果进行对比分析。分别对比表3列(1)、列(3)和列(2)、列(4),可以看出,加入扩展指标之后,BVAR模型的MSE、SMAPE下降,MAPE略有上升,由于更多的指标表示BVAR模型的预测误差下降,所以从总体上看BVAR模型的预测能力得到了提升。对于LSTM模型,MSE、MAPE、SMAPE均下降,因此可以认为加入扩展指标提升了LSTM模型的预测能力。同时,可以初步判断,加入扩展指标对于BVAR模型和LSTM模型预测精确度

的提升程度不同,本文将在后面对 BVAR 模型和 LSTM 模型的预测能力进行更为细致的比较分析。

表 3 模型预测效果对比(动态滚动预测)

指标体系	(1)BVAR	(2)LSTM	(3)BVAR	(4)LSTM
	基准指标体系		扩展指标体系	
MSE	1.332	1.520	1.258	1.033
MAPE	9.19%	9.32%	9.42%	8.13%
SMAPE	8.92%	8.83%	8.89%	7.95%
D - M Test	0.8628(0.3883)		1.754(0.0795)	

注:括号内为 P 值,后表同。

综合上述两方面的分析,可以发现:如果预测指标体系仅基于凯恩斯 IS - LM 理论来设置,将难以与中国的现实情况相拟合,从而导致 GDP 预测的误差较大。这也进一步说明,本文选取的商品房本年新开工面积同比增速、宏观经济景气先行指数、发电量同比增速等扩展指标能够使模型本身更符合中国宏观经济的运行规律,从而提高模型预测的精确度。

(二)BVAR 模型与 LSTM 模型的预测能力比较

为了更充分地对 BVAR 模型与 LSTM 模型的预测能力进行比较,本部分基于扩展指标体系展开分析。参照张劲帆等(2018)^[14]的常用做法,本文首先在动态滚动情形下进行预测。然后在固定估计窗口情形下进行预测,从而比较 BVAR 模型与 LSTM 模型的预测精确度。需要说明的是,以上操作是在向前预测步距设置为 1(即 $h = 1$)时进行的 GDP 预测,为了确保研究结论稳健可靠,本文增加预测步距作为稳健性检验,从而更加全面地比较 BVAR 模型与 LSTM 模型的远期预测能力。

第一,动态滚动预测时,与 BVAR 模型相比,LSTM 模型能够进行更精确的拟合和预测。通过分析图 5 可知,相比 BVAR 模型,LSTM 模型预测曲线的预测表现更为稳定。具体来说,在 2007Q1 - 2012Q2 预测中,LSTM 模型总体上看与 GDP 实际值的拟合度更高,与之相比,BVAR 模型预测曲线对局部极值点的预测较为粗糙,较大程度上存在放大短

期趋势的现象。同时,在 2012Q3 - 2018Q4 预测中,LSTM 模型的预测结果表现出更加平稳的特征,并且预测值与 GDP 的实际值非常接近,没有出现显著异常的预测值。将 BVAR 模型与 LSTM 模型的预测误差相比,可以得到相同的结论。GDP 动态滚动预测中,LSTM 模型的 MSE、MAPE、SMAPE 均小于 BVAR 模型的相关指标。这意味着,用 LSTM 模型预测 GDP 的效果优于 BVAR 模型,并且 D - M 检验表明该差异在 10% 的显著性水平上显著。

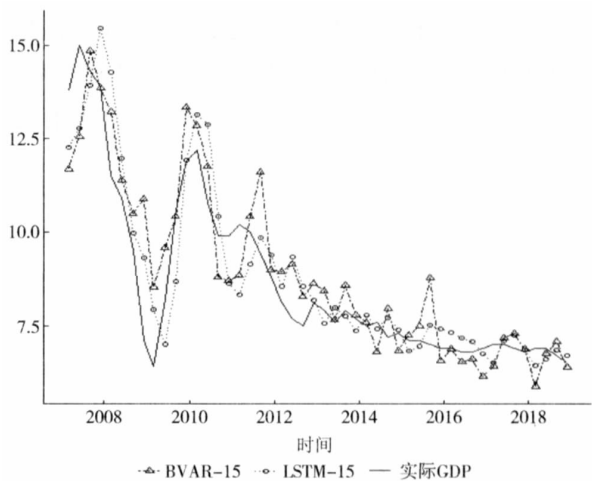


图 5 扩展指标体系下 GDP 样本外预测结果

这表明,相对于 BVAR 模型,LSTM 模型能够更大程度地改善预测结果。究其原因,LSTM 模型能够识别与挖掘非线性因素对短期 GDP 预测的影响,因此,选取 LSTM 模型构建方式能够更加准确地刻画中国宏观经济特征,从而提升 GDP 预测模型的预测能力。

第二,在固定估计窗口情形下,LSTM 模型的预测效果依然优于 BVAR 模型。为了更充分地比较 BVAR 模型与 LSTM 模型的预测效果,本文参考张劲帆等(2018)^[14]的做法,首先选取 1997Q1 至 2006Q4 这个区间的时间序列,并将预测窗口固定为 40 个季度,进行固定估计窗口的短期 GDP 预测。通过对比表 4 的列(1)与列(2),可以发现,固定预测时间窗口后,LSTM 模型样本外预测的误差同样小于 BVAR 模型,与上文所得结论保持一致。

从模型自身特性角度看,BVAR 模型利用的历

史信息会受到滞后阶数的限制,与之相比,LSTM 模型具有更为优越的学习能力,会按照需要自行决定保留更久远的记忆或者遗忘过去的信息,从而能够利用更长的时间序列的相关信息预测。本文通过固定估计时间窗口,控制了 LSTM 模型能够利用的样本量,由此得到的结论再度证明了,LSTM 模型样本外预测能力的提升是因为 LSTM 模型能够更精准地刻画宏观指标对于 GDP 的非线性影响。

表 4 模型预测效果对比(固定估计窗口)

	(1) BVAR	(2) LSTM
ME	1.509	1.318
MAPE	10.40%	9.37%
SMAPE	10.16%	8.90%

第三,稳健性检验结果显示,增加预测步距情形下,LSTM 模型的远期预测能力依然优于 BVAR 模型。以上两种情形下进行的是相邻一步预测,即 h 设置为 1(滞后一季)。作为稳健性检验,当进行远期预测时,本文将 h 设置为 2、3、4(滞后二、三、四季度)。参数设定参照前文,预测过程分别采用动态滚动预测方式、固定估计窗口方式进行。BVAR 模型与 LSTM 模型各步距预测的实证结果对比见图 6、图 7。图 6 显示,在动态滚动预测 GDP 时,LSTM 模型不同步距的 MSE 均小于 BVAR 模型,且平均预测效果改善 39.4%。从图 7 中可以看出,在固定估计窗口预测 GDP 时,当 h 为 1、2 时,BVAR 模型与 LSTM 模型预测能力较为接近;然而,随着预测步距的增加,当 h 增加为 3、4 时,BVAR 模型的 MSE 大幅度增加,预测能力出现较大程度的折损,而 LSTM 模型的 MSE 变化较为平稳,预测能力更为稳定,平均预测效果改善 26.42%。这一结果表明,LSTM 模型具有较强的泛化能力,对 GDP 不同期限的预测效果均较为稳定。

此外,BVAR 模型与 LSTM 模型在正常(非危机)状态下的预测效果更优。考虑到 2008 - 2011 年正好跨越了一个经济周期,为了考察经济周期波动对模型预测能力的影响,首先计算 2007 - 2018 年全样本的预测误差,然后以 2011 年为起点^②计算

2011 - 2018 部分样本的预测误差,最后对二者进行比较分析。通过对比表 5 的相邻两列(如表 5 的列(1)与列(2))可以看出,固定时间窗口后,2011 - 2018 年部分样本的预测效果较 2007 - 2018 年全样本的预测效果有大幅度提升。由此可以得出结论,排除 2007 - 2011 年周期性影响能够显著提升模型样本外预测能力,即模型在正常(非危机)状态下的预测效果更优。

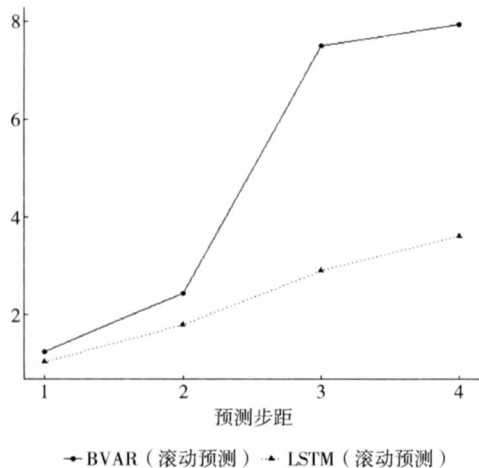


图 6 GDP 远期预测结果对比(动态滚动预测)

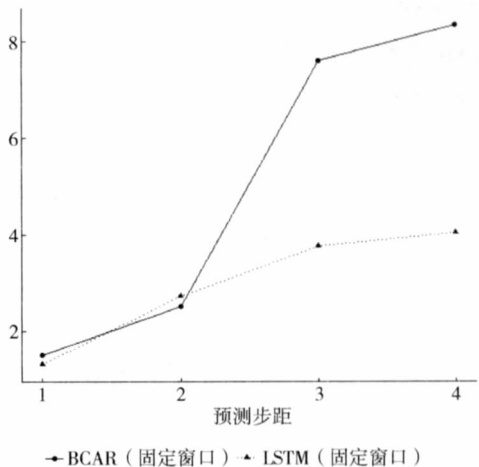


图 7 GDP 远期预测结果对比(固定估计窗口)

(三) 扩展指标的预测效果比较

基于本文构造的扩展指标体系,可进一步探讨房地产部门、第三产业、宏观经济景气先行指数、发电量、金融市场对模型 GDP 预测效果的影响。本部分采用预测效果更优越的 LSTM 模型,逐个剔除模型中的扩展指标,通过分析剔除各个扩展指标后的

模型与原模型的 MSE 变化,比较各扩展指标的预测能力。

表 5 模型预测效果对比(全样本与部分样本)

时间区间	BVAR 模型		LSTM 模型	
	(1)	(2)	(3)	(4)
	2007-2018	2011-2018	2007-2018	2011-2018
MSE	1.518	0.560	1.520	0.239
MAPE	10.33%	7.64%	9.32%	5.04%
SMAPE	9.88%	7.64%	8.83%	4.87%

为保持与上文中原模型预测效果的可比性,仍选取相同区间的时间序列。具体预测步骤如下:首先,从扩展指标体系中剔除商品房本年新开工面积同比增速;随后,通过对比 MSE 的大小,综合比较剔除该指标后的模型相较于原模型预测精度的变化,由此即可得到剔除该指标对于 GDP 预测精确度的影响。重复上述两个步骤,依次在模型中剔除第三产业增加值当期同比实际增速、宏观经济景气先行指数、发电量同比增速、沪市成交额同比增速,即可得到各扩展指标的预测效果。

表 6 给出的是剔除各扩展指标后, LSTM 模型对实际 GDP 增长率进行样本外预测的 MSE。从表中可以看出,在相同的参数设定和网络结构下,样本外预测精确度下降最大的是沪市成交额同比增速,其次是第三产业增加值当期同比实际增速,排名第三位的是宏观经济景气先行指数。这说明,不同指标对模型预测准确度的影响是明显不同的。其中,沪市成交额同比增速对近十余年 GDP 增速的非线性变化具有最强的解释能力。需要强调的是,2008 年国际金融危机前,资本市场流动性显著下滑,显示了股市流动性对宏观经济的先导性,引发了有关股市流动性和宏观经济相互关系的研究。

表 6 剔除各个扩展指标后模型的均方误差

扩展指标体系	HSTARTS	THIRD	LI	ELEC	STOCK	
MSE	1.033	1.1998	1.2773	1.2476	1.1854	1.341
百分比变化	—	12.81%	20.09%	17.30%	11.45%	26.08%

奈斯等(Naes et al., 2011)运用美国数据发现,股票市场流动性显著领先于实体经济指标^[20]。王东旋等(2014)基于回归分析和格兰杰因果检验的样本内拟合,发现中国股市流动性领先于宏观经济^[21]。本文的这一实证结果从样本外预测的角度提供了关于股市流动性与宏观经济联动的证据。

考虑中国宏观现实对于提升 GDP 预测的精确度具有重要意义。在宏观经济预测中,如果忽略中国实际问题,则极易导致预测值相比实际值的结构性偏离,从而增大 GDP 预测的误差。有鉴于此,本文认为要想切实提升宏观经济预测的准确性,除了探索、改进预测方法外,还需要科学合理地确定宏观经济预测指标,构建既基于经济学理论又符合中国国情的宏观经济预测指标体系。

五、结论

本文旨在探讨深度学习神经网络能否改进 GDP 的预测能力。为此,本文首先基于宏观经济学经典理论与中国宏观经济现实,构建了宏观经济预测指标体系。并结合目前最前沿的深度学习神经网络优化技术构造了一个 LSTM 模型,将其应用于中国 GDP 的预测分析,并与 BVAR 模型预测结果进行了基准指标体系和扩展指标体系的对比。

本文研究结果表明,第一,本文选择的扩展指标能够提升 BVAR 模型与 LSTM 模型的 GDP 预测能力。一方面, BVAR 模型和 LSTM 模型的样本外预测图在扩展指标体系下对 GDP 实际值的拟合度更强。另一方面,加入扩展指标后, BVAR 模型和 LSTM 模型的预测误差值呈现下降的趋势。第二,相比于 BVAR 模型, LSTM 模型能够更好地挖掘扩展指标对 GDP 的非线性影响,从而提升短期 GDP 预测能力。在动态滚动预测与固定估计窗口预测

注:每一个模型只剔除一个扩展指标。例如, HSTARTS 表示剔除商品房本年新开工面积同比增速的扩展指标体系。

情形下,LSTM模型的预测效果均优于BVAR模型。此外,当预测步距增大时,LSTM模型预测效果要优于BVAR模型,具体而言,在各种步距条件的设定下,LSTM模型的预测效果均优于BVAR模型,并且随着向前步距的增加,预测优势发生显著提升,在动态滚动预测与固定估计窗口方式下预测效果分别平均改善39.4%、26.42%。因此,本文认为鉴于LSTM模型具有深度学习能力、良好的泛化能力及高度可调节性,其在GDP预测领域具有广阔的应用前景。

本文在宏观经济预测领域利用深度学习前沿技术进行积极探索,并以中国GDP预测为实例,验证了LSTM模型的预测优越性,为将深度学习神经网络广泛应用于宏观经济预测、形成新的研究范式提供了实践经验。鉴于深度学习神经网络的高度可调节性,后续还可以在两大主题上继续探索:一是将深度学习神经网络应用于其他宏观经济数据的预测,利用大数据的优势更有效地进行风险预警与经济分析;二是改进深度学习神经网络技术,例如添加多种非同质信息作为神经网络输入,附加小波分解或主成分分析等数据预处理技术进行模型优化,或从神经网络本身进行结构优化等。

注释:

①克强指数是以李克强总理命名的衡量经济发展状况的指标,包括耗电量、铁路货运量和银行贷款发放量三个指标组成,分别占比重40%、25%及35%。

②2010年10月中国宣布退出全面扩张、全面刺激经济的反危机政策,因此,以2011年为起点可以衡量正常(非危机)状态下GDP预测的准确度。

参考文献:

[1]王飞.基于贝叶斯向量自回归的区域经济预测模型:以青海为例[J].经济数学,2011(2):95-100.
[2]周建,况明.中国宏观经济动态传导、可靠性及货

币政策机制[J].经济研究,2015(2):31-46.

[3]TKACZ G. Neural network forecasting of Canadian GDP growth[J]. International Journal of Forecasting, 2001, 17(1):57-69.

[4]欧邦才.基于BP神经网络的经济预测方法[J].南京工程学院学报(自然科学版),2004(2):11-14.

[5]郭庆春,何振芳.基于人工神经网络的经济预测模型[J].计算技术与自动化,2014(1):132-136.

[6]CHOI J Y, LEE B. Combining LSTM network ensemble via adaptive weighting for improved time series forecasting[J]. Mathematical Problems in Engineering, 2018:1-8.

[7]赵军豪,李玉华,霍林,等.融合微博情感分析和深度学习的宏观经济预测方法[J].计算机应用,2018(11):3057-3062.

[8]陈小亮.房地产对经济增长的短期与长期影响[J].中国高校社会科学,2017(6):52-60,154-155.

[9]SIMS C A. Macroeconomics and reality[J]. Econometrica, 1980, 48(1):1-48.

[10]DOAN T, LITTERMAN R B, SIMS C A. Forecasting and conditional projection using realistic prior distributions[J]. Econometric Reviews, 1984, 3(1):1-100.

[11]KOOP G. Bayesian methods for empirical macroeconomics with big data[J]. Review of Economic Analysis, 2017, 9(1):33-56.

[12]肖争艳,安德燕,易娅莉.国际大宗商品价格会影响我国CPI吗——基于BVAR模型的分析[J].经济理论与经济管理,2009(8):17-23.

[13]HOCHREITER S, SCHMIDHUBER J. LSTM can solve hard long time lag problems[C]//NIPS 96: Proceedings of the 9th International Conference on Neural Information Processing Systems. Cambridge, MA:MIT Press, 1996:473-479.

[14]张劲帆,刚健华,钱宗鑫,等.基于频规向量自回归模型的宏观经济预测[J].金融研究,2018(7):34-48.

[15]DIEBOLD F X, MARIANO R S. Comparing predictive accuracy[J]. Journal of Business & Economic Statistics, 1995, 13(3):253-263.

[16]MANKIW N G. Macroeconomics[M]. 8th ed. New

York: Worth Publishers, 2012.

[17] 陈彦斌. 宏观政策评价报告 2019[J]. 经济研究参考, 2019(4): 13 - 20.

[18] REINHART C M, ROGOFF K S. Growth in a time of debt[J]. American Economic Review, 2010, 100(2): 573 - 578.

[19] 刘哲希, 李子昂. 结构性去杠杆进程中居民部门

可以加杠杆吗[J]. 中国工业经济, 2018(10): 42 - 60.

[20] NAES R, SKJELTORP J A, DEGAARD B A. Stock market liquidity and the business cycle[J]. The Journal of Finance, 2011, 66(1): 139 - 176.

[21] 王东旋, 张峥, 殷先军. 股市流动性与宏观经济——有关中国股市的新证据[J]. 经济科学, 2014(3): 61 - 71.

Can Deep - learning Neural Networks Improve GDP Forecast?

Xiao Zhengyan Liu Lingjun Zhao Tingrong Chen Yanbin

Abstract: The prediction of GDP growth rate is an important research field of macroeconomics. However, the complex economic situation at home and abroad makes it more difficult to forecast GDP, so how to forecast GDP effectively is an important theoretical and practical problem worth studying. This paper constructed with the macroeconomic theory basis and in line with the characteristics of China's macroeconomic index system, and constructed a LSTM model for GDP forecast analysis, and compared with the BVAR model, in order to scientifically judge whether the LSTM model can enhance the accuracy of the GDP forecast. The results are shown as follows. Firstly, the expansion index selected in this paper can improve the GDP forecasting ability of the BVAR model and the LSTM model, so it is necessary to take into account the expansion index reflecting China's national conditions when forecasting China's GDP. Secondly, compared with the BVAR model, the LSTM model can better explore the non-linear influence of extended indicators on GDP, so as to improve the short-term GDP forecast ability. On this basis, because of strong self-learning ability, good generalization ability and good model adjustability, the deep-learning neural network has a broad prospect in the field of GDP forecasting.

Key words: neural network; LSTM model; BVAR model; GDP forecast