银行业存款预测——基于starma模型

**摘要：**本文基于2015-2020年27个省份银行业境内本外币存款余额的月度数据，使用starma模型进行预测，最终发现我国银行业存款无论是在地理空间维度上还是时间维度上存在较强关系。经过MSE和MAE的比较下，发现在中国整体上和大部分省份中，使用动态样本外预测测试集，starma模型预测效果要优于arma模型预测。

**关键词：**starma模型 arma模型 空间计量

引言

近年来，人民生活水平日益提高，经济快速发展，消费水平也在快速上升。人民手中的钱拥有越来越多的决策方案，更多用于消费或者储蓄。市场经济体系越来越开放，银行业通过各种金融产品大量吸引资金流入，而不是仅仅依靠存款利率吸引更多客户进行存款。银行业能够吸纳存款越多不仅说明其规模有扩大的优势，而且往往说明了国民对于储蓄投资欲望的强烈，对于我国资本市场开放和进一步扩大起着巨大作用。不仅如此，在金融行业的发展过程中，我们可以明显看出不同地区省市之间的发展存在差异，在东南沿海省份依靠其市场开发程度高和贸易水平发达，往往有更加充裕的资金去进行储蓄投资，从而可以更加快速的进一步发展金融行业和银行业。因此考虑到银行业在中国整体虽然发展巨大，但存在地区差异化发展，因此引入空间差异性对各省市银行业存款进行预测。

文献综述

存款预测的方法在国内外文献内现使用过的方法十分广泛，其中根据切入视角的不同进行预测的方法使用也不同，主要分为两大类，一类是通过过去的存款数（宏观数据or银行数据）进行一些机器学习或者计量模型的方法进行未来的估计。早期的文献大部分采用ARMA模型基础上进行改进进行估计，胡学锋（2001）使用ARIMAX模型预测居民储蓄存款；李明明王旭皓等（2015）使用ARIMA与季节指数组合模型进行居民储蓄存款预测。近年来将机器学习和深度学习的方法引入到计量模型中进行预测，王艳（2019）由于数据不平滑的特点，通过HP滤波-ARMA的方法预测湖北省居民个人储蓄存款余额；宋良美和胡金辉（2018）由于数据呈非齐次指数性的特点，采用GM(1,1,k)模型对上海市各项存款余额预测；舒服华（2019）使用LSTM神经网络进行武汉市居民存款余额预测; Lin Canrong & Kim Hee Chel则使用ARIMA进行储蓄银行住房抵押贷款预测。该类方法的优势在于当未存在外部冲击时，模型预测效果较好，但由于缺少外部信息因此由于突发性事件导致的存款余额大幅变动会产生较大影响难以预测。

另一种方法则是通过引入部分外部数据，其中包括利率、国内生产总值、货币供应量等数据进行分析，张凯和李旭（2007）考虑了医疗、住房等重大体制变革前后通货膨胀对居民储蓄的影响后进行预测;S. Chiraphadhanakul & P. Dangprasert等使用遗传算法进行预测，捕捉到了国内生产总值（gdp）、货币供应量（公众持有）、利率、商业银行分行数和贷款与银行存款之间的关系。

本文综合考虑了地区之间银行业存款预测存在空间相关性，同时arma模型能够较好预测存款，选取了starma模型进行分地区银行业存款余额预测。

模型和空间权重矩阵设定

starma模型是由Phillip E. Pfeifer & Stuart. Jay Deutsch（1980）所提出，其实质为时空模型，往往模型被多用于区域交通流量预测，而在Serkan Kurt & K. Batu Tunay（2015）应用于土耳其区域银行业存款预测。

starma（p,q）模型：



其中为的t时刻截面数据，为的k阶ar项的截面数据，为的k阶ma项截面数据，为l阶空间权重矩阵，和为待估参数。

当，即不存在ar项时，模型则退化为stma(q)模型:



当，即不存在ma项时，模型则退化为star(p)模型:



空间权重矩阵：

为0阶空间权重矩阵，为单位阵，在starma模型中引入用来给与每个个体自身的ar项和ma项对其影响，当时，即不存在更高阶空间权重矩阵时，starma(p,q)将退化为arma(p,q)模型：



为1阶空间权重矩阵，该权重矩阵即反映了部分空间相关性，通常采用1阶邻接权重矩阵即可，或者通过一定准则进行判断，但是要满足Giacomini & Granger（2004）所提出的3个基本条件：

。

为高阶权重矩矩阵，通常认为是存在空间滞后性，该空间滞后性的含义是变量在一个时点或者多个时点存在多个空间方向进行移动，因此导致了空间有多种变化。基于这种考虑，通常引入高阶邻接权重进行刻画。在2阶邻接权重矩阵条件下，元素则满足：



由于所选用的数据为分省的地区数据，3阶及以上的空间权重矩阵实际经济效应过小，因此不纳入考虑。

数据选取及预处理

本文所选数据为2015-2020年27个省份银行业境内本外币存款余额（亿元）的月度数据，数据来源同花顺。

表格1罗列了27个省份的存款余额描述性统计，图1展示了浙江、云南、新疆和西藏2015-2020年的存款余额变化。从图1可以明显看出该月度数据存在明显的上升趋势，并且无法通过平稳性检验，因此本文选择取对数后十二步差分剔除季节效应，再进行一阶差分剔除增长趋势，即，。

表格3分别对psd进行的无截距项、有截距项和有趋势项的Levin-Lin-Chu面板单位根检验结果，均拒绝原假设，说明处理后的数据平稳。

表格 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| y | 均值 | 标准差 | 最小值 | 最大值 | 偏度 | 峰度 |
| 浙江 | 106451.3 | 15345.62 | 82758.96 | 140364.9 | 0.3 | -0.98 |
| 云南 | 28308.33 | 3319.53 | 22244.55 | 33689.07 | -0.34 | -1.13 |
| 新疆 | 19544.32 | 2391.85 | 15074.87 | 23159.8 | -0.39 | -1.13 |
| 西藏 | 4427.99 | 638.74 | 3198.58 | 5131.42 | -0.71 | -0.99 |
| 四川 | 67332.78 | 9367.5 | 51726.89 | 84112.81 | -0.07 | -1.24 |
| 陕西 | 36552.59 | 4732.1 | 28441.44 | 44482.3 | -0.09 | -1.2 |
| 山西 | 32597.05 | 3771.09 | 27130.83 | 40028.65 | 0.19 | -1.21 |
| 山东 | 85245.36 | 11012.16 | 67166.55 | 107612.7 | 0.08 | -1.08 |
| 青海 | 5558.56 | 422.7 | 4468.71 | 6163.61 | -1.19 | 0.67 |
| 宁夏 | 5551.99 | 648.93 | 4266.28 | 6670.14 | -0.32 | -0.97 |
| 内蒙 | 20774.06 | 2449.99 | 16177.43 | 23975.73 | -0.59 | -1.07 |
| 辽宁 | 49656.62 | 5534.11 | 41186 | 60223 | 0.34 | -1.07 |
| 江西 | 30788.23 | 5470.53 | 21943.39 | 41012.85 | -0.02 | -1.26 |
| 湖南 | 42960.79 | 6369.22 | 30630.35 | 53526.05 | -0.44 | -1.05 |
| 湖北 | 50912.44 | 7106.3 | 38447.79 | 63249.6 | -0.13 | -1.2 |
| 北京 | 118154.4 | 15639.68 | 91080.46 | 142611.4 | -0.03 | -1.19 |
| 海南 | 8679.64 | 1159.5 | 6257.98 | 9859.87 | -0.8 | -0.96 |
| 贵州 | 22856.8 | 3275.03 | 15392.61 | 25809.34 | -1.07 | -0.34 |
| 广西 | 26447.74 | 3536.36 | 20100.6 | 32366.72 | -0.22 | -1.24 |
| 甘肃 | 17504.39 | 1435.07 | 14395.44 | 20432.36 | -0.22 | -0.54 |
| 福建 | 38916.63 | 4600.16 | 30536.83 | 47622.35 | -0.18 | -1.1 |
| 安徽 | 43592.62 | 7277.13 | 30452.01 | 56093.79 | -0.23 | -1.23 |
| 上海 | 87737.73 | 11497.99 | 68300.95 | 108868 | 0.05 | -1.16 |
| 重庆 | 31773.43 | 3680.18 | 25072.65 | 38330.76 | -0.11 | -1.11 |
| 江苏 | 126415.6 | 16845.11 | 97876.16 | 162141.9 | 0.03 | -0.96 |
| 广东 | 161207.1 | 26034.75 | 120179.8 | 210519.8 | 0.08 | -1.18 |
| 河北 | 58951.73 | 8966.98 | 43930.28 | 76272.16 | 0.07 | -1.13 |

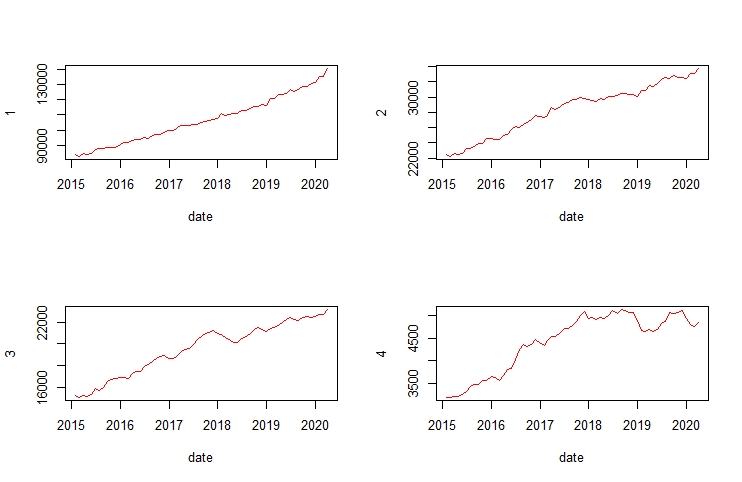


图 1

表格 2

|  |  |  |
| --- | --- | --- |
|  | 统计量值 | P值 |
| 无截距项 | -32.70 | 0.000 |
| 有截距项 | -31.03 | 0.000 |
| 有趋势项 | -30.10 | 0.000 |

模型设定检验

starma模型提供两种检验，用以判断数据是否存在时空效应；一、stacf（空间自相关）和stpacf（空间偏自相关）；二、时空相关性检验。

图2为stacf，图3为stpacf；stacf分别在，，下1阶截尾，显著性弱，显著性弱，说明在，下时空相关性较弱较弱；stpacf在，，下均1阶截尾，因此选择1阶模型，，阶数则要具体根据BIC信息准则最小原则进行选择。表格3为时空相关性检验，原假设为数据不存在时空相关性，拒绝原假设说明在时空双维度存在相关性。

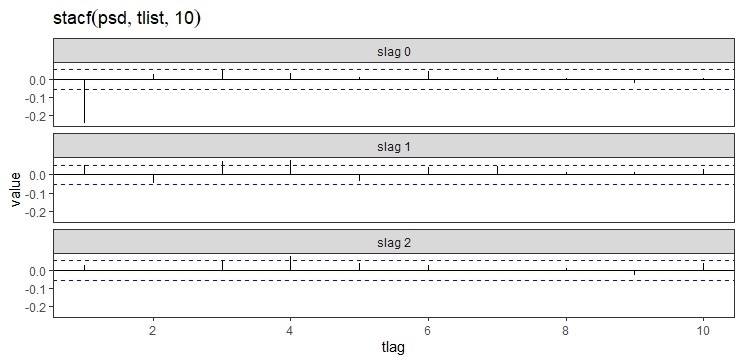


图 2

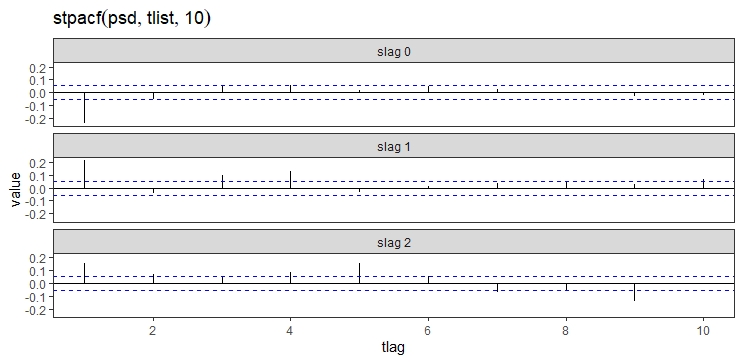


图 3

表格 3

|  |  |  |
| --- | --- | --- |
|  | 统计量 | P值 |
| 时空相关性检验 | 254.36 | 0.000 |

模型估计结果

为了检验模型预测及拟合程度，提取前45个时间点的数据作为训练集，后5个时间点的数据作为测试集。最终模型选择在BIC信息准则最小以及系数显著性检验通过情况下选择了star(3)模型，模型估计结果见表格4。

表格 4

|  |  |  |  |
| --- | --- | --- | --- |
| 参数 | 估计值 | 标准误 | P值 |
|  | -0.35 | 0.034 | 0.000\*\*\* |
|  | 0.20 | 0.065 | 0.002\*\*\* |
|  | 0.19 | 0.075 | 0.01\*\* |
| BIC | | -6656.93 | |

模型结果为：

转换为原数据模型为：

模型预测结果：

将star(3)模型在训练集中与arma模型（通过AIC信息准则判断阶数）的MSE进行比较，最终发现，训练集中四川、青海、宁夏、内蒙古、江西、湖南、北京、重庆、江苏。广东和河北star(3)模型拟合情况优于arma模型。

表格5为在测试集进行比较，该测试集的预测值通过样本外动态预测进行预测并与真实值进行比较的MSE、MAE，整体上MSE和MAE，star(3)模型拟合情况更优，在总体上应该选择star(3)。

图4展示了分别在训练集和测试集的估计结果和预测结果以及原值，分别展示了云南、新疆、山西。山东、青海、辽宁、江西、湖南、湖北、北京、海南、贵州、广西、安徽和重庆，其中原值为红色，star(3)模型为绿色，arma模型为蓝色，2015年1月-2016年2月为差分时期，因此该部分时期无模型估计结果；2016年3月-2019年10月为模型的估计时间段，即训练集阶段；2019年11月-2020年3月为模型的预测时间段，即测试集阶段。罗列的这些省份由star(3)模型的预测结果精度要高于arma模型。

表格 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | MSE(star) | MSE(arma) | 结果 | MAE(star) | MAE(arma) | 结果 |
| 总体 | 0.00014 | 0.00065 | star | 0.00804 | 0.0117 | star |

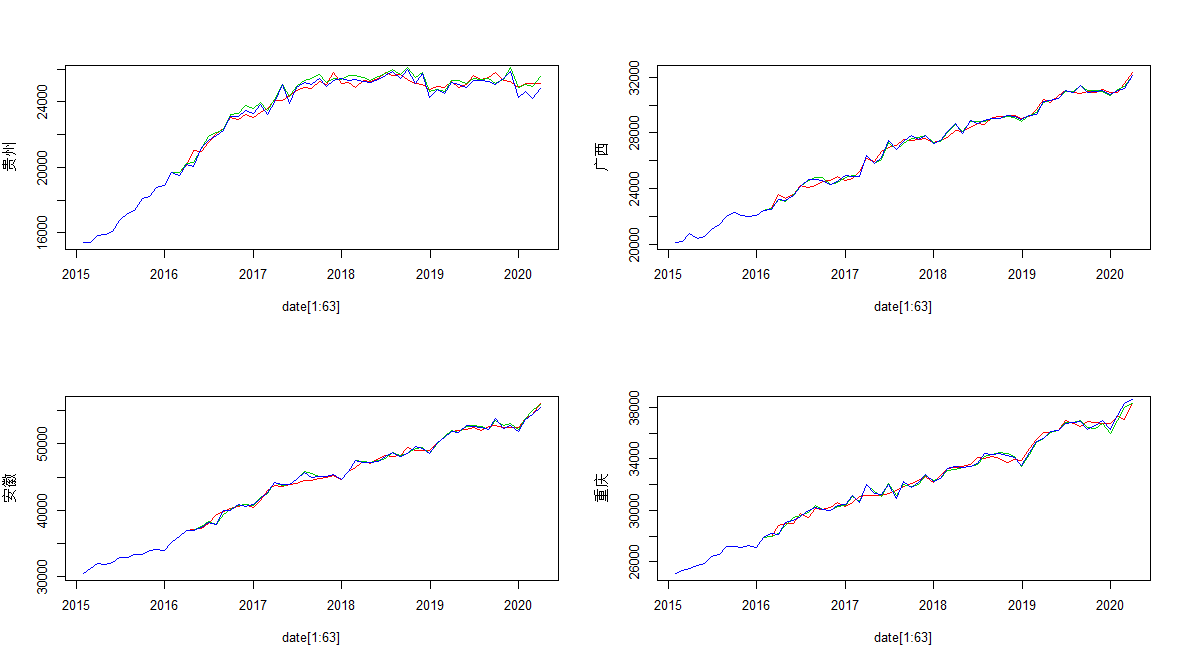
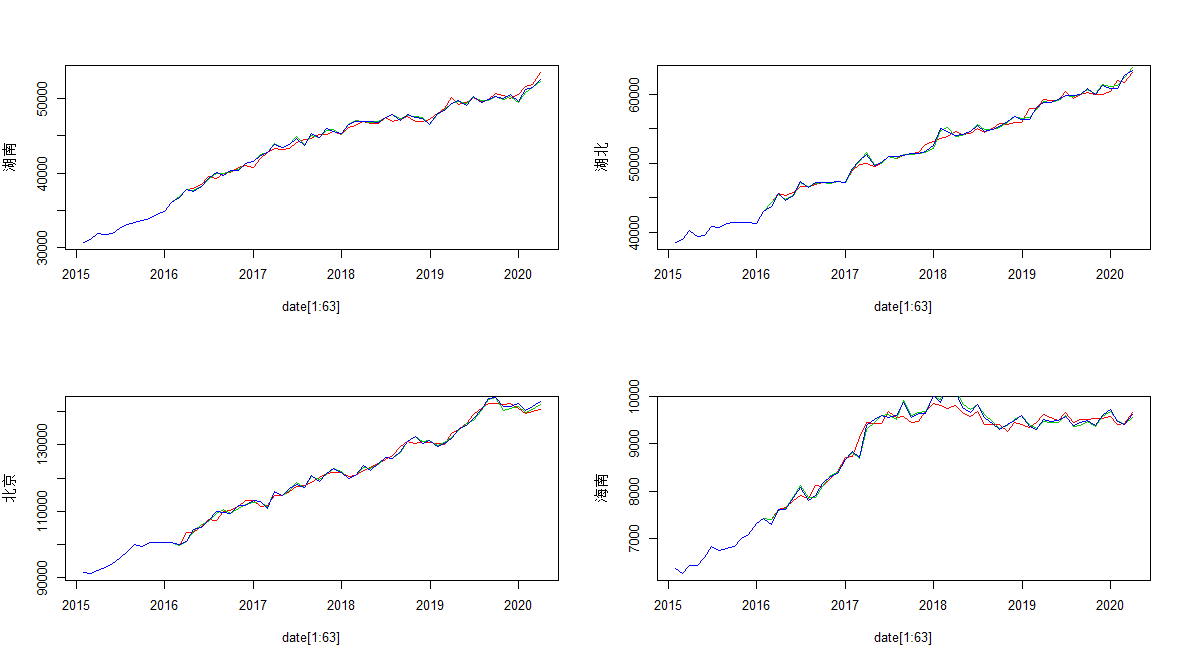
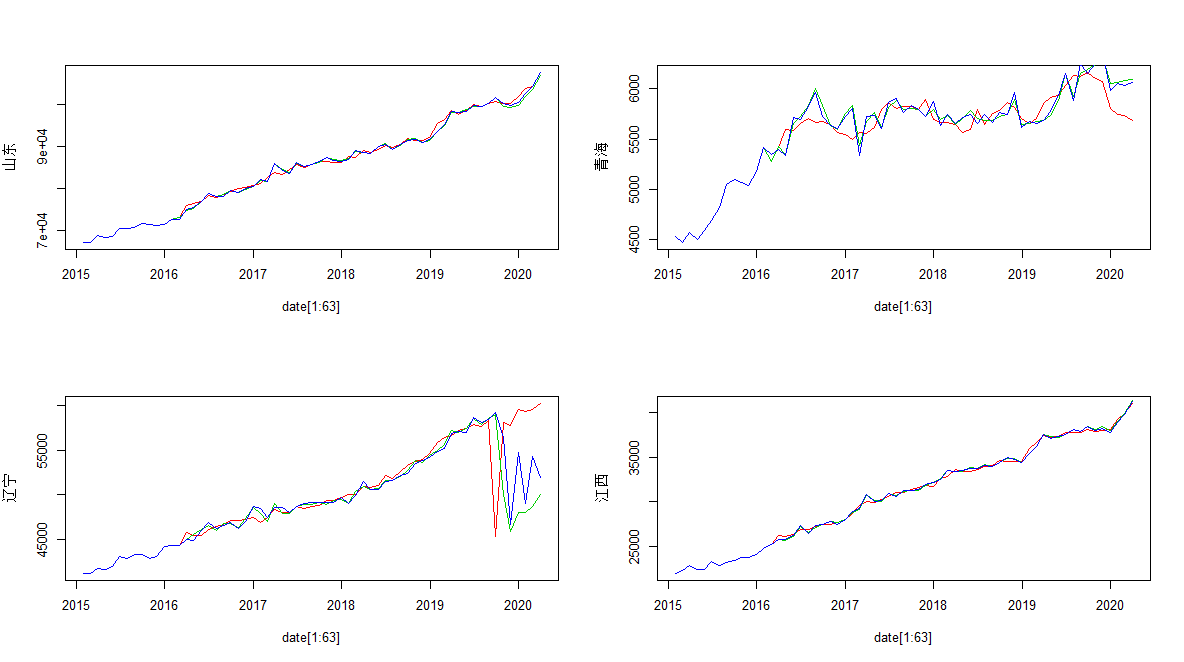
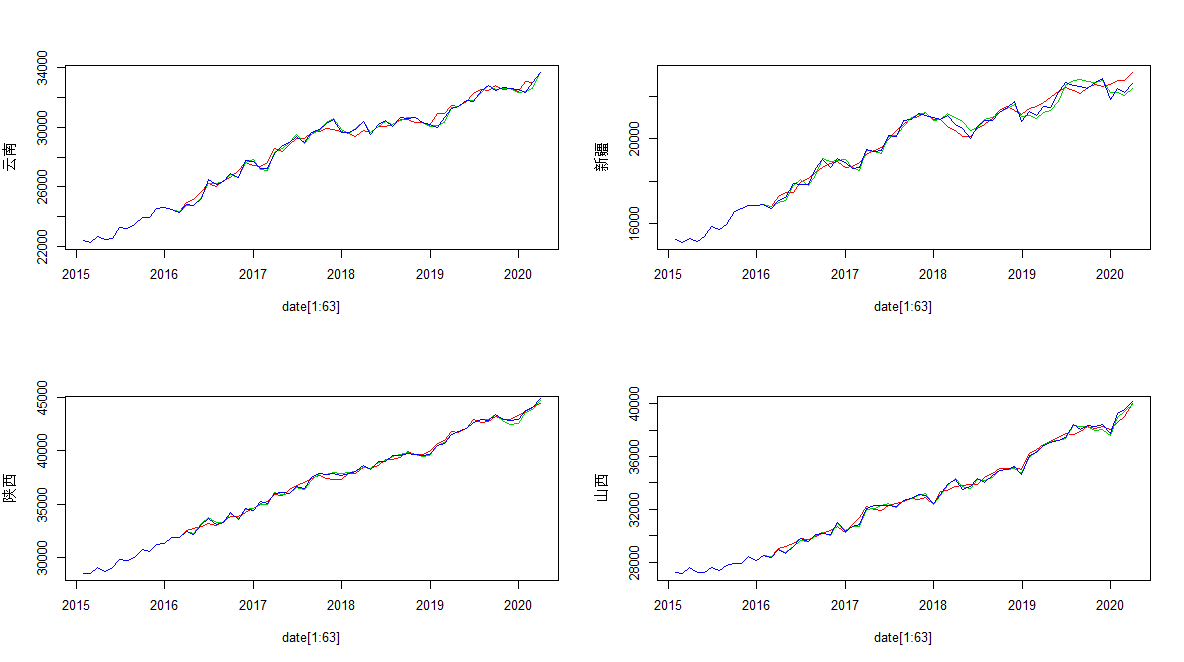


图 4

结论

银行存款与各个省份的经济发展和人民生活水平有着关联，因此银行存款预测有着更多可以外部数据可以进行考虑，本文主要从存款上的空间异质性出发，在arma模型的基础上引入空间效应，采用starma模型进行预测，在时空双维度上进行合理地预测。

在进行starma模型和arma模型的MAE和MSE结果比对中可以发现，在引入空间效应后中国绝大部分省份和整体上，模型预测结果的精度更好，因此在银行业存款预测上可以考虑空间效应。

参考文献

[1]胡学锋.ARIMAX模型在居民储蓄存款预测中的应用[J].财经问题研究,2001(01):70-72.

[2]李明明,王旭皓,孙舒曼,李智明.基于ARIMA与季节指数组合模型的居民储蓄存款预测[J].经济研究导刊,2015(05):192-193.

[3]张凯,李旭.我国居民储蓄存款总额预测模型[J].时代经贸(中旬刊),2007(S2):75-76.

[4]王艳.基于HP滤波-ARMA的湖北省居民个人存款预测[J].金融理论与教学,2019(05):11-1

[5]宋良美,胡金辉.基于改进GM(1,1,k)模型的上海市各项存款余额预测[J].上海立信会计金融学院学报,2018(06):70-78.

[6] S. Chiraphadhanakul, P. Dangprasert and V. Avatchanakorn, "Genetic algorithms in forecasting commercial banks deposit," 1997 IEEE International Conference on Intelligent Processing Systems (Cat. No.97TH8335), Beijing, China, 1997, pp. 116-121 vol.1.

[7] Lin Canrong and Kim Hee Chel, A Study on the Analysis and Prediction of Housing Mortgage in Deposit Bank Using ARIMA Model.[J][Journal of Korea Institute of Information, Electronics, and Communication Technology,](http://scholar.cnki.net/new/journal/index/SJNV2005081X1020)Volume 12, Issue 3. 2019. PP 265-272

[8] Serkan Kurt,K. Batu Tunay. STARMA Models Estimation with Kalman Filter: The Case of Regional Bank Deposits[J]. Procedia - Social and Behavioral Sciences,2015,195.

[8] Pgillip E.P., Stuart. J. D.. Stationarity and invertibility regions for low order starma models[J]. Communications in Statistics - Theory and Methods, Volume 9, Issue 5. 1980. PP 551-562.

附录

表格 6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dsp | MSE(star) | MSE(arma) | 结果 | MAE(star) | MAE(arma) | 结果 |
| 浙江 | 7.04E-05 | 7.03E-05 | arma | 0.006893 | 0.006799 | arma |
| 云南 | 5.96E-06 | 6.34E-06 | star | 0.002069 | 0.002245 | star |
| 新疆 | 6.02E-05 | 6.44E-05 | star | 0.006844 | 0.007346 | star |
| 西藏 | 1.28E-04 | 1.00E-04 | arma | 0.009624 | 0.008059 | arma |
| 四川 | 6.88E-05 | 6.68E-05 | arma | 0.007606 | 0.007248 | arma |
| 陕西 | 5.18E-05 | 5.20E-05 | star | 0.005351 | 0.005314 | arma |
| 山西 | 1.20E-04 | 1.20E-04 | star | 0.00799 | 0.008335 | star |
| 山东 | 9.82E-05 | 9.94E-05 | star | 0.008702 | 0.008826 | star |
| 青海 | 3.79E-04 | 4.06E-04 | star | 0.015187 | 0.015652 | star |
| 宁夏 | 3.60E-05 | 3.60E-05 | arma | 0.004913 | 0.004909 | arma |
| 内蒙 | 6.63E-05 | 5.34E-05 | arma | 0.007588 | 0.006782 | arma |
| 辽宁 | 1.18E-03 | 1.51E-02 | star | 0.023659 | 0.118025 | star |
| 江西 | 5.84E-05 | 5.88E-05 | star | 0.006227 | 0.00629 | star |
| 湖南 | 3.09E-05 | 3.64E-05 | star | 0.004816 | 0.005465 | star |
| 湖北 | 2.75E-05 | 2.79E-05 | star | 0.004943 | 0.004989 | star |
| 北京 | 1.24E-04 | 1.25E-04 | star | 0.008933 | 0.009275 | star |
| 海南 | 8.47E-05 | 1.27E-04 | star | 0.008853 | 0.010422 | star |
| 贵州 | 4.21E-05 | 5.22E-05 | star | 0.004593 | 0.006665 | star |
| 广西 | 8.29E-06 | 9.47E-06 | star | 0.002471 | 0.002594 | star |
| 甘肃 | 4.02E-05 | 3.35E-05 | arma | 0.005121 | 0.004756 | arma |
| 福建 | 1.57E-04 | 1.57E-04 | arma | 0.011426 | 0.01141 | arma |
| 安徽 | 3.02E-05 | 3.04E-05 | star | 0.003255 | 0.0034 | star |
| 上海 | 2.14E-04 | 2.12E-04 | arma | 0.011787 | 0.011432 | arma |
| 重庆 | 2.27E-04 | 2.34E-04 | star | 0.012565 | 0.013186 | star |
| 江苏 | 1.42E-04 | 1.41E-04 | arma | 0.008911 | 0.008902 | arma |
| 广东 | 1.03E-04 | 1.02E-04 | arma | 0.008287 | 0.008112 | arma |
| 河北 | 9.37E-05 | 9.22E-05 | arma | 0.008707 | 0.008629 | arma |

该表格展示了在MSE和MAE计算准则下，每个省份计算的结果，虽然starma模型在解释变量上要多余arma模型，但在MSE和MAE计算结果上，starma模型并非全都优于arma模型。