

Forecasting Model of Year-on-Year Probability

Yujuan Min¹, Wei Zhao¹, Yun Yang^{2*}

¹Information Department, The First People's Hospital of Lianyungang, Lianyungang Jiangsu

²Computer Science Department, College of Information Engineering, Yangzhou University, Yangzhou Jiangsu

Email: *yyang@yzu.edu.cn

Received: May 18th, 2020; accepted: Jun. 1st, 2020; published: Jun. 8th, 2020

Abstract

The verification system of personal foreign exchange business uses the big data of foreign exchange transactions to directly list individuals who borrow others' quota to handle the settlement and sale of foreign exchange, through the algorithm analysis of big data. It can find out the subject of abnormal transaction in time, predict and analyze the possible clues, which provides the basis for legal authorities to lock in targets and find out abnormalities, screen violations and fast law enforcement. The effective combination of off-site data analysis and on-site inspection can be achieved, which can improve the actual effect of on-site inspection. It is necessary to predict and analyze the amount and total amount of the individual foreign exchange business in order to analyze the change range of split amount and total split amount, as well as master the quantitative and qualitative changes of the individual foreign exchange business data. Each forecasting model has its application scope. Based on the analysis of the existing prediction algorithms, we establish the forecasting of year-on-year probability based on probability. The simulation verification of the big foreign exchange data of the year-on-year probability forecasting model shows that it can not only obtain the data change trend of time series, but also make data fluctuate changes according to the seasonal characteristics. The forecasting model for year-on-year probability is more accurate than the gray forecasting model based on time series when predicting data with large data gaps between years and fluctuations.

Keywords

Forecasting Model, Grey Forecasting, Probability Forecasting, Personal Foreign Exchange Business

基于同比的概率预测模型

闵玉涓¹, 赵巍¹, 杨云^{2*}

¹连云港市第一人民医院, 信息部, 江苏 连云港

²扬州大学信息工程学院, 计算机系, 江苏 扬州

*通讯作者。

Email: *yyang@yzu.edu.cn

收稿日期: 2020年5月18日; 录用日期: 2020年6月1日; 发布日期: 2020年6月8日

摘要

个人外汇业务核查系统利用外汇交易大数据, 通过对大数据的算法分析, 将存在借用他人额度办理结售汇行为的个人, 直接列入“关注名单”, 能够及时地查找异常交易主体, 预测分析可能成案的线索因素, 为执法部门锁定目标、发现异常、甄别违规、快速执法提供依据, 实现非现场数据分析与现场检查的有效结合, 进而提高现场检查的实际效果。为了分析个人外汇业务中的分拆量、分拆金额总量的变化区间, 掌握个人外汇业务数据的量变与质变, 需要对个人外汇业务的分拆量、分拆金额量等进行预测分析。各种预测模型都有其应用范围, 在对现有预测算法分析的基础上, 我们基于同比预测模型, 以概率为基础建立了同比概率预测模型。同比概率预测模型的外汇大数据的仿真验证表明: 同比概率预测算法不仅能得到时间序列的数据变化趋势, 同时可以使数据根据季节性特点呈现波动性变化。同比概率预测模型对年份间数据差距较大, 且存在波动的数据进行预测时, 精度高于以时间序列为主的灰色预测模型。

关键词

预测模型, 灰色预测, 概率预测, 个人外汇业务

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

Open Access

1. 引言

个人外汇业务核查系统是一个非现场核查系统, 是为了加强个人外汇业务管理, 防范跨境资本过度流动, 将在现有个人外汇管理框架下, 进一步便利银行及个人办理外汇业务, 提升个人外汇业务监管效率。对个人外汇业务预测分析的目的是了解分拆量、分拆金额总量等数据的变化区间和发展趋势, 掌握个人外汇业务数据的量变与质变, 评估当前个人购付汇管理政策实施的效果, 在检查实践和技术手段层面提出针对性的解决方案, 在政策储备和监管理念方面提出相应的对策建议, 对于促进个人外汇业务健康发展、改进外汇管理方式和手段、完善外汇管理体制框架具有重要实现意义。

每年的个人外汇支出业务数据量差距较大, 同时受开学季的影响, 以留学为理由的个人外汇业务明显增多。由于灰色预测模型是对小样本数据序列建模, 预测的值以时间序列为主, 反应数据变化的总体趋势, 不能反应数据的波动情况(比如不能反应随着季节变动的数据波动)。若采用同比预测, 虽然可以反应数据的波动情况, 但是若前后两年数据差距很大, 会造成预测值和真实值每个月都存在一个差值。为了能更准确地预测个人外汇业务数据的变化趋势, 我们采用基于同比的概率预测算法, 不仅可以消除年份之间数据差距较大带来的问题, 还可保持以时间序列为主的总体变化趋势, 同时也能够突出季节性数据的变化特点。

2. 相关工作

预测在我们生活中起着至关重要的作用, 基于数学建模的预测方法种类繁多, 根据各学科领域的需要以及数据自身的特点, 各国专家学者致力研究提出了许多预测算法: 时间序列分析法、灰色模型预测算法和马尔科夫预测算法等。

1982年3月,中国学者邓聚龙教授在国际上首先提出了灰色系统理论[1]。灰色系统是既含有已知信息,又含有未知信息的系统。灰色预测模型使用的不是原始数据的序列,而是生成的数据序列。核心体系是 Grey Model,即对原始数据作累加生成(或其他处理生成)得到近似的指数规律再进行建模的方法。适用于在处理较少的特征值数据,不需要数据的样本空间足够大,就能解决历史数据少、序列的完整性以及可靠性低的问题,能将无规律的原始数据进行生成得到规律较强的生成序列。只适用于中短期的预测,适合近似于指数增长的预测。由于灰色预测具有所需原始信息少、计算过程简单、可测试性强、可预测等优点,因此它被广泛使用。

时间序列预测法[2]根据客观事物发展的这种连续规律性,运用过去的历史数据,通过统计分析,进一步推测市场未来的发展趋势。适用于时间序列在时间序列分析预测法处于核心位置,一般用 ARMA 模型拟合时间序列,预测该时间序列未来值, Daniel 检验平稳性。主要有自动回归 AR (Auto Regressive)和移动平均 MA (Moving Average)预测模型。

马尔科夫预测[3] [4]适用于随机现象的数学模型(即在已知现情况的条件下,系统未来时刻的情况只与现在有关,而与过去的历史无直接关系)。适用于研究一个商店的未来某一时刻的销售额,当现在时刻的累计销售额已知。不适宜用于系统中长期预测。

差分方程预测[5]利用差分方程建模研究实际问题,常常需要根据统计数据用最小二乘法来拟合出差分方程的系数。适用于商品销售量的预测、投资保险收益率的预测。数据系统的稳定性还要进一步讨论代数方程的求根。

Facebook 时间序列预测算法[6], Facebook 所提供的 prophet 算法不仅可以处理时间序列存在一些异常值的情况,也可以处理部分缺失值的情形,还能够几乎全自动地预测时间序列未来的走势。prophet 算法是基于时间序列分解和机器学习的拟合来做的,其中在拟合模型的时候使用了 pyStan 这个开源工具,因此能够在较快的时间内得到需要预测的结果。除此之外,为了方便统计学家,机器学习从业者等人群的使用,prophet 同时提供了 R 语言和 Python 语言的接口。适用于一般的商业分析或者数据分析的需求。

同比是一种描述数据变动的方法,是指历史同期数据进行比较,同比预测是指使用今年第 n 月与去年第 n 月进行比较,可以计算出相对发展速度,计算公式为:同比发展速度 = 本期发展水平/去年同期水平 $\times 100\%$ 。同比更倾向于反映一直长期的大趋势,也就规避了季节这些影响因素,消除季节变动的影响能够更好的反映数据的变化趋势。若以同比为基础进行简单的加权均值法对未来数据进行预测,将会受限于数据的特点,若数据虽然保持了以季节性变得的特性,但年份间可能存在数值差距较大,权值难以确定,会对预测结果造成很大误差。

3. 算法设计

分析灰色预测模型、同比预测模型的算法特点,基于同比预测模型,提出同比概率预测模型。

3.1. 经典灰色 GM(1,1)预测模型

灰色预测模型是对系统因素之间发展趋势相异成都进行关联分析,其特点是不使用原始数据序列,对原始数据进行处理生成有较强规律性的数据序列,并建立相应的微分方程模型,从而通过计算来预测数据信息在未来的发展趋势[7]。

使用 GM(1,1)模型,必须保证已知数据的可行性。设原始数据序列为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 计算该数据序列的级比:

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)} \quad (k = 2, 3, \dots, n) \quad (1)$$

如果所有的级比都落在可覆盖区间 $X = \left(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}} \right)$ 内, 则该数据序列可以使用 GM(1,1)模型进行数据预测。否则需要对数据做变换处理, 可以使用平移变换:

$$y^{(0)}(k) = x^{(0)}(k) + c \tag{2}$$

以数据序列 $x^{(0)}$ 建立 GM(1,1)模型, 其相应微分方程为:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \tag{3}$$

通过回归分析求得 a, b 的值, 解方程得到:

$$x^{(1)}(t) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-a(t-1)} + \frac{b}{a} \tag{4}$$

从而相应得到 GM(1,1)模型预测值:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (k = 1, 2, 3, \dots, n-1) \tag{5}$$

灰色预测模型的缺点是: 一般情况下, GM(1,1)模型适用于近似灰色指数率和灰饱和率的单调小样本数据序列进行建模, 而对于波动性较强的序列则没有有效的进行预测[8] [9]。

3.2. 同比预测模型

为了保持每一年的月份或季节的特点, 我们设计了同比预测模型, 如式(6)所示:

$$l_k^1 = \frac{1}{2} \times \left[\frac{1}{3} \times \sum_{i=k-3}^{k-1} l_i^1 + l_k^2 \right] \tag{6}$$

其中 l_k^1 表示要预测数据的第 k 月份的数据, l_k^2 表示前一年的第 k 月份的数据。先取当前年前三个月的数据的均值, 然后再与 l_k^2 取均值。 l_k^2 为历史同期数据, 加入 l_k^2 这个参数能利用历史同期数据的季节性特点来预测本期数据, 但是若各个年份数据变化幅度大, 即 l_k^1 与 l_k^2 数值差距大, 将会导致预测数据的不准确性。

3.3. 同比概率预测模型

基于同比的概率预测模型分为两种情况:

第一, 当月份小于 4 月份时, 将历史同期数据加上权重来预测本期数据。根据往年数据。根据往年数据与本期数据的关系, 相距越远的年份, 相互关系越低, 因此设历史前三年的权重分别为 0.618、0.236、0.146。如式(7)所示:

$$l_k^1 = l_k^2 \times 0.618 + l_k^3 \times 0.236 + l_k^4 \times 0.146 \quad (1 \leq k \leq 3) \tag{7}$$

第二, 当月份大于等于 4 月份时, 需要利用已经预测出的月份来向后续预测。先计算出当前月份历史同期数据在前三个月占的比重。若要分析季节性特点, 当前月份与本期数据的前三个月所占比例将与前一年相同, 基于同比概率预测模型公式如下:

$$\frac{l_k^1}{l_k^1 + \sum_{i=k-3}^{k-1} l_i^1} = \frac{l_k^2}{l_k^2 + \sum_{i=k-3}^{k-1} l_i^2} \tag{8}$$

可以推导出:

$$l_k^1 = \frac{l_k^2 \times \left(l_k^1 + \sum_{i=k-3}^{k-1} l_i^1 \right)}{l_k^2 + \sum_{i=k-3}^{k-1} l_i^2} \tag{9}$$

用 P 表示概率，则概率 P 的值为：

$$P = \frac{l_k^2}{l_k^2 + \sum_{i=k-3}^{k-1} l_i^2} \tag{10}$$

无论历史同期数据与本期数据差距多大，概率 P 是固定的，因此可以解决年份间数据差距大的问题，最后基于同比概率预测模型的公式为：

$$l_k^1 = \frac{P \times \sum_{i=k-3}^{k-1} l_i^1}{1 - P} \tag{11}$$

3.4. 同比概率预测算法基本步骤

- 第一步：用 i 代表月份，设置 i 为 1；
- 第二步：判断 i 是否小于 4，是则转第三步，否则转第四步；
- 第三步：根据式(7)计算 l_i^1 的值，并使 $i = i + 1$ ，转第二步；
- 第四步：根据式(10)计算概率 P ，根据式(11)计算 l_i^1 的值，并使 $i = i + 1$ ，转第五步；
- 第五步：判断 i 是否小于等于 12，是则转第四步，否则转第六步；
- 第六步：输出每月的预测值。

4. 实验与算法分析

实验环境：集成开发环境 IDE 选为 PyCharm，开发语言选用 python3.0；

实验数据：由于数据涉及个人隐私，我们使用某省某个地级市表示数据来源，实际实验数据是该省该地级市 2015 年至 2018 年个人外汇涉外支出业务量的原始数据；

实验模型：简单统计模型、灰色 GM(1,1)预测模型和同比概率预测模型。

分别用简单统计模型、灰色 GM(1,1)预测模型和同比概率预测模型，在该实验环境下建模进行仿真，对比分析三种模型的预测准确性。

4.1. 同比算法实验分析

通过分析 2015 年~2018 年各月份的分拆量，可以得到如图 1 所示结果。

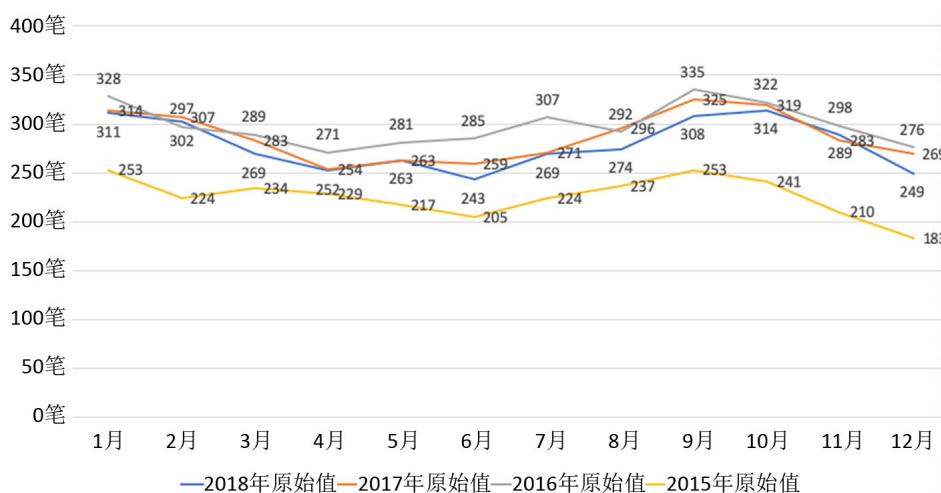


Figure 1. Breakdown of each month in 2015~2018 (simple statistical model)

图 1. 2015~2018 年各月份分拆量(简单统计模型)

由图 1 可以发现, 各年度每月的数据有两个特点, 第一, 每一个年度的数据变化趋势不平稳, 且年度之间分拆量差距较大; 第二, 分拆量在有的月份或者季节有着明显特性, 例如 9 月份, 分别在 2015、2016、2017、2018 年的分拆量是一个凸点, 相比较临近月份数值较高。

4.2. 灰色预测模型与同比概率预测模型实验分析

以 2018 年个人外汇涉外支出业务量作为数据, 建立 GM(1,1)模型, 得到原始数据序列 $x^{(0)} = (311, 302, 269, 252, 263, 243, 269, 274, 308, 314, 289, 249)$, 计算数据序列的级比 $\lambda(k)$, 得到级比数列 $\lambda(k) = (1.02, 1.12, 1.06, 0.95, 1.08, 0.90, 0.98, 0.88, 0.98, 1.08, 1.16)$ 。同时将 $n=12$ 代入可覆盖区间, 得 $X = (0.857, 1.166)$ 。可以发现级比的 $\lambda(k)$ 全部落在可覆盖区间内, 使用 GM(1,1)模型对数据进行预测, 结果如表 1 所示:

Table 1. GM(1,1) model prediction results

表 1. 灰色 GM(1,1)模型预测结果

时间	原始值	灰色 GM(1,1)预测值	相对误差/%
2018 年 1 月	311	311	0
2018 年 2 月	302	269	-10.927
2018 年 3 月	269	270	0.372
2018 年 4 月	252	271	7.540
2018 年 5 月	263	273	3.802
2018 年 6 月	243	274	12.7572
2018 年 7 月	269	275	2.230
2018 年 8 月	274	276	0.730
2018 年 9 月	308	277	-10.065
2018 年 10 月	314	279	-11.147
2018 年 11 月	289	280	-3.114
2018 年 12 月	249	281	12.851

使用同比概率预测模型, 以 2015~2017 年个人外汇数据作为输入, 对 2018 年的数据进行预测, 结果如表 2 所示:

Table 2. Forecasting model of year-on-year probability model prediction results

表 2. 同比概率模型预测结果

时间	原始值	同比概率模型预测值	相对误差/%
2018 年 1 月	311	308	-0.965
2018 年 2 月	302	292	-3.311
2018 年 3 月	269	277	2.974
2018 年 4 月	252	246	-2.381
2018 年 5 月	263	253	-3.802
2018 年 6 月	243	251	3.292
2018 年 7 月	269	261	-2.974

Continued

2018年8月	274	285	4.014
2018年9月	308	313	1.623
2018年10月	314	307	-2.229
2018年11月	289	272	-5.882
2018年12月	249	258	3.614

将 GM(1,1)预测模型与同比概率预测模型的预测值进行对比分析, 结果如图 2 所示:

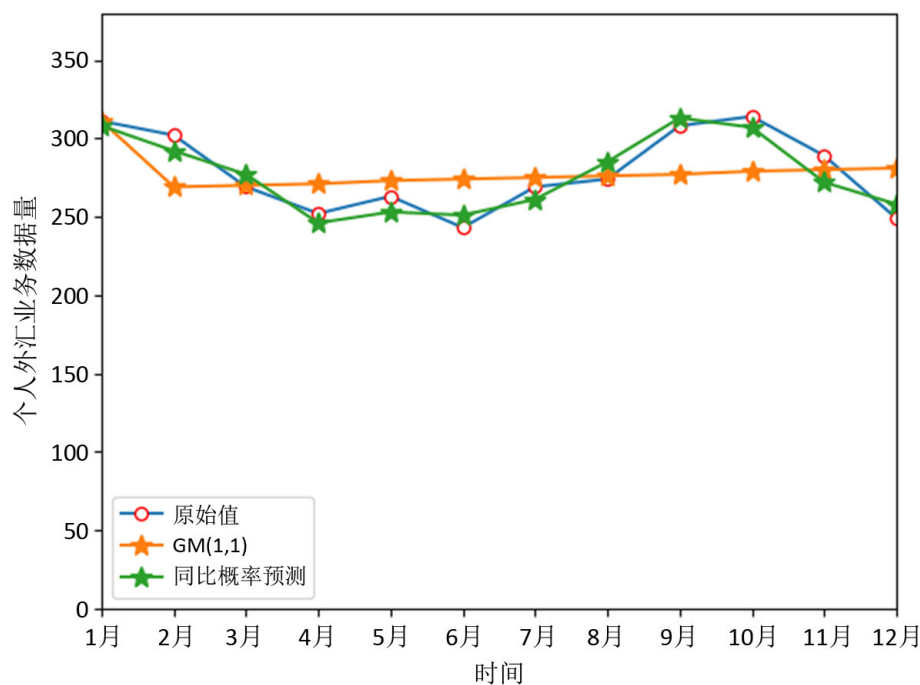


Figure 2. GM(1,1) and forecasting model of year-on-year probability comparison results
图 2. GM(1,1)与同比概率预测对比图

对表 1、表 2 和图 2 进行分析, 灰色 GM(1,1)模型在 1 月, 3 月, 8 月预测值的相对误差较小, 但是整体波动较大, 而采用同比概率预测的相对误差整体偏小, 且无较大波动。根据图 1 对比分析两种算法的准确度, 可以发现 GM(1,1)模型仅仅预测了数据根据时间的整体变化趋势, 在个别月份会造成较大误差, 同比概率预测模型, 采用概率 P 来消除年份之间数据误差大的问题, 能够很好的预测出个人外汇业务数据量的季节性波动变化的特点。

5. 结束语

准确有效地预测未来个人外汇业务数据量, 能够有效地促进个人外汇业务的健康发展, 推动外汇管理体制的完善。对往年个人外汇业务数据进行分析, 发现数据呈季节性变化特点, 且年份间数据差距较大, 在使用灰色 GM(1,1)模型对未来个人外汇业务数据量进行预测时, 虽然数据序列在可覆盖区内, 但是预测结果仅仅反映了数据的整体变化趋势, 未能突出季节性特点。采用基于同比的概率预测模型, 以概率为基础对往年数据进行分析, 同时结合已经预测的结果对未来结果进行递归预测, 能够有效地消除年份间数据差距大的问题, 同时能够很好地预测个别季节数据的波动变化。

为了对数据进行更加准确的预测，日后的工作重点是将灰色预测模型和基于同比的概率预测模型进行结合，借助灰色模型预测出的以时间序列为主的整体变化趋势，加上同比概率预测模型的波动变化，可以完善针对此类数据的有效预测模型。

基金项目

国家自然科学基金(No.61872312)；江苏省产学研前瞻性联合项目(No.BY2016069-16)。

参考文献

- [1] 刘领坡. 我国社会消费品零售总额时间序列模型及预测[J]. 经济论坛, 2011(6): 5-8.
- [2] Begleiter, R., El-Yaniv, R. and Yona, G. (2011) On Prediction Using Variable Order Markov Models. *Journal of Artificial Intelligence Research*, **22**, 385-421. <https://doi.org/10.1613/jair.1491>
- [3] 魏庆征, 杨云, 李凌燕. 优化的灰色马尔科夫外汇支出预测模型[J]. 计算机科学与应用, 2019, 9(11): 2105-2114.
- [4] 舒服华. 基于小波 GM(2, 1)模型的我国外汇储备预测[J]. 嘉兴学院学报, 2018, 30(3): 1-6.
- [5] 姜田利. 基于差分方程的河流污染物含量动态预测算法研究[J]. 环境科学与管理, 2019, 44(4): 145-148.
- [6] <https://otexts.org/fpp2/components.html>
- [7] 王子赞, 纪志成. 基于 GM(2,1)和辨识算法的风电功率短期预测研究[J]. 系统仿真学报, 2015, 27(11): 2762-2769.
- [8] 官兴国, 张博, 吴琪. 基于 GM(2,1)灰色复合模型的财务风险预测[J]. 统计与决策, 2015(17): 179-182.
- [9] 丁松, 李若瑾, 党耀国. 基于初始条件优化的 GM(1, 1)模型及其应用[J]. 中国管理科学, 2020, 28(1): 153-161.