

基于函数性数据分析的中国社会消费品零售总额数据的预测

李小星 徐永利*

(北京化工大学 理学院, 北京 100029)

摘要: 基于数据的函数性特征,采用函数性数据分析方法对中国社会消费品零售总额进行预测。将总额数据分解为长期趋势成分和季节性波动成分分别进行预测。对长期趋势成分引入居民可支配收入数据来辅助预测,对于季节性成分采用自适应选择权重的加权预测法,将两项预测值的和作为总额预测结果。真实数据上的实验结果表明,本文提出的预测方法预测误差小,且预测结果具有很好的可解释性。

关键词: 函数数据分析; 社会消费品零售总额; 居民可支配收入; 长期趋势

中图分类号: F22 **DOI:** 10.13543/j.bhxbzr.2018.03.017

引言

社会消费品零售总额是国民经济中的一项重要经济指标,直接反映居民生活水平、社会消费品购买能力、社会生产力、货币流通力及变化趋势,因此研究和改进社会消费品零售总额预测方法具有重要意义。

目前对于经济数据的预测方法主要有时间序列分析法和函数性数据的预测方法。时间序列分析中常用的模型有自回归移动平均模型(ARMA)和季节时间序列模型(SARIMA)^[1-2],前者可以较好处理时间序列的依存性和数据随机波动的干扰性,对短期数据预测精度较高;后者利用对数差分和季节差分减小序列的波动性建立平稳可逆且拟合效果最优的预测模型。但是该方法的缺点是只针对离散数据而忽略了数据的连续性。

函数性数据分析法从函数的角度出发,重点关注数据的凹凸性、拐点等连续性信息。文献[3-4]运用函数性数据分析法将社会消费品零售总额数据分解为长期趋势成分和季节性波动成分,利用三次函数模型预测长期趋势成分,采用固定权重加权方法预测季节性成分,取得了较满意的预测效果,但是

对于数据波动剧烈的年份预测误差较大。

居民可支配收入是反映居民生活水平的一个重要指标,与社会消费品零售总额存在密切联系^[5];而且家庭人均收入的变化不影响同期人均社会商品零售总额的变化,而是对下一年的人均销售总额有着重要的影响^[6]。因此本文在文献[3]的基础上,引入居民可支配收入指标来修正数据波动较大年份的预测结果,设计了将已知消费数据与收入数据融合的自适应预测方法,利用上一年的居民可支配收入指标辅助预测下一年的社会消费品零售总额长期趋势成分,并基于长期趋势成分的预测结果,采用变权重加权方法预测季节性成分。

1 社会消费品零售总额预测算法

1.1 算法介绍

本文基于函数性数据分析方法^[7],将社会消费品零售总额分解为长期趋势成分和季节性成分分别预测。长期趋势成分的预测采用线性预测模型,该模型的关键环节是预测函数斜率的确定,本文设计“延伸上一阶段的线性趋势”(算法1)和“收入的增长量代替消费的增长量”(算法2)两种算法来确定线性模型斜率,再融合二者的优势设计出算法3,将算法1、2的斜率加权组合用于下一年长期趋势的预测。对于季节性成分设计算法4,采用加权组合前一阶段各年份的季节性成分来进行预测,并且自适应地设置加权组合中的权重。最后将算法3和算法4的结果相加作为社会消费品零售总额对数预测

收稿日期: 2017-05-25

基金项目: 国家自然科学基金(11571031); 国家留学基金

第一作者: 女,1992年生,硕士生

* 通信联系人

E-mail: xuyongli2312@sina.com

值。预测方法的实现过程如图 1 所示。

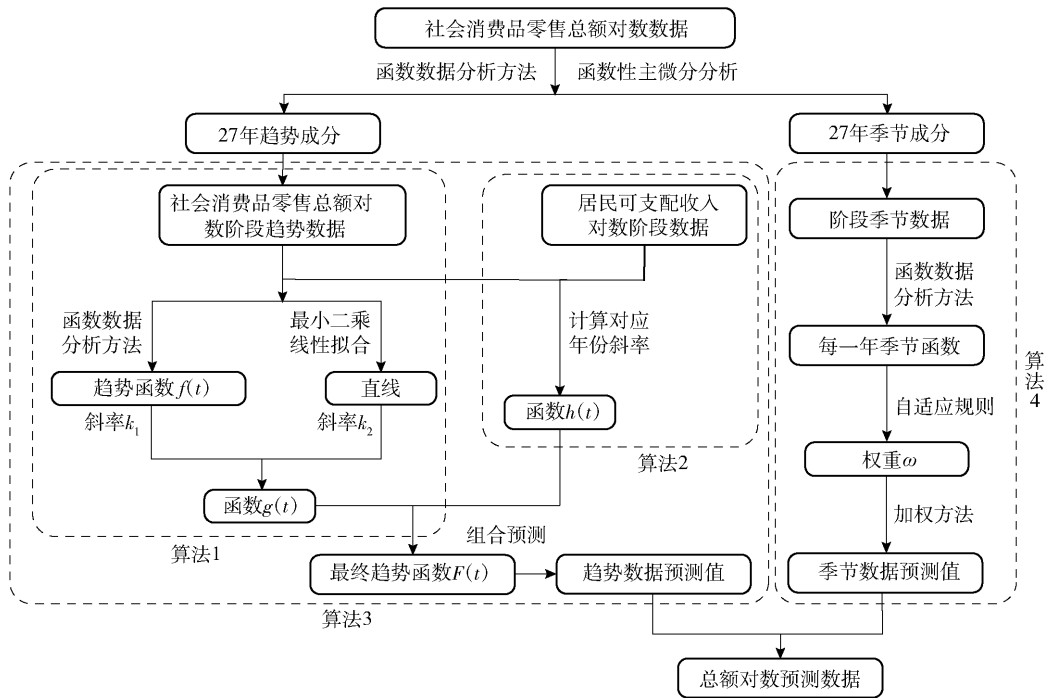


图 1 预测算法流程图

Fig. 1 Prediction algorithm flow chart

1.2 趋势数据预测

消费数据长期趋势成分的预测对于最终总额数据预测的精确度尤为重要。趋势函数即使在较长的时间范围内大致呈现出线性增长的规律,也会在局部时间范围存在增长率变化较快的情况^[3],因此趋势成分的预测需要兼顾线性增长和局部变化。

本文将 1984 ~ 2010 年我国的居民可支配收入年度总额数据以 10 为底数取对数,然后将其按照时间顺序直接相连,图 2 展示了 1984 ~ 2010 年居民可支配收入对数数据与社会消费品零售总额对数数据的比较。可以看到,消费数据与收入数据的长期增长趋势非常相近,而且收入数据的变化大致比消费数据的变化提前一年发生,因此本文引入居民可支配收入来辅助消费数据的预测。

为了预测消费数据的长期趋势成分,首先要对已有数据进行拟合。本文采用如公式(1)的正则化拟合方法

$$\min Pe(c|y) = \sum_i \left[y_i - \sum_{k=1}^K c_k \varphi_k(t_i) \right]^2 + \lambda Pn \left(\sum_{k=1}^K c_k \varphi_k \right) \quad (1)$$

其中, $Pe(\cdot)$ 是带正则项的残差平方和,基函数 φ_k 采用 6 阶 B 样条基, $\varphi_k(t_i)$ 是 φ_k 在时间 t_i 处的取

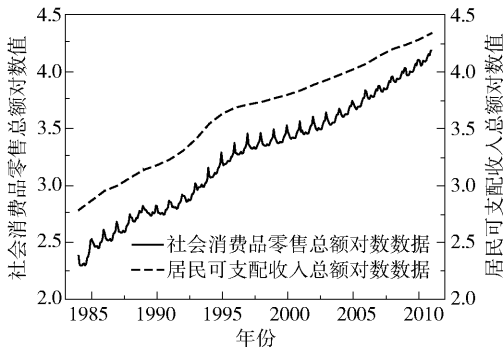


图 2 居民可支配收入对数数据

Fig. 2 Logarithmic data of residents' disposable income
值,基函数的个数 k 与拟合数据的节点数 i 的关系为 $k = i + 6 - 2, Pn \left(\sum_{k=1}^K c_k \varphi_k \right) = \int [D^3 x(t)]^2 dt$ 是标准正则项, λ 是正则化参数。基于公式(1)的拟合方法设计出消费数据的长期趋势成分预测算法 1。

算法 1:利用前期消费数据线性预测下一年消费趋势成分。

- 1) 利用正则化最小二乘算法设置正则化参数为 10^{-2} ,对当前月份之前 6 年对数消费数据进行拟合,获得第一轮长期趋势成分 $f_1(t)$;
- 2) 计算前 6 年对数消费数据与 $f_1(t)$ 在相应时

间点函数值之差, 获得前6年第一轮季节性离散数据;

3) 用初始季节性离散数据的第5年数据替换第6年数据, 获得第一轮季节性离散修正数据;

4) 用对数消费数据减去初始季节性离散修正数据, 获得第一轮长期趋势离散修正数据;

5) 用步骤4)得到的长期趋势离散修正数据替换步骤1)~4)中的对数消费数据, 重复步骤1)~4)2次, 获得季节性离散修正数据和长期趋势离散修正数据;

6) 利用正则化最小二乘算法设置正则化参数为 10^{-2} , 对步骤5)中长期趋势离散修正数据进行非线性拟合, 获得最终长期趋势成分 $f(t)$, $f(t)$ 在最后一个月的函数值记为 b , 线斜率记为 k_1 ;

7) 对步骤5)中长期趋势离散修正数据做线性拟合, 获得直线的斜率, 记为 k_2 ;

8) 计算 $k_3 = \frac{k_1 + k_2}{2}$, 构造 $g(t) = k_3 t + b$, 作为下一年的长期趋势预测函数。

算法1是在拟合已有数据长期趋势的基础上预测下一年长期趋势成分, 但是这种方法是基于上一年长期趋势和前一阶段整体线性趋势的顺延, 在消费数据长期趋势剧烈变化的年份结果不够准确, 所以引入居民可支配收入指标来修正下一年消费数据长期趋势成分, 具体方法如算法2。

算法2: 基于前一年收入数据线性预测下一年消费趋势成分。

1) 将算法1步骤6)中获得的 b 值作为下一年长期趋势成分基准值;

2) 计算上一年收入数据增长量 k_4 ;

3) 构造预测函数 $h(t) = k_4 t + b$, 作为下一年的长期趋势预测函数。

算法2在一定程度上克服了算法1的问题, 但是在收入数据的增量和下一年消费数据长期趋势成分增量上存在明显差异的年份, 算法1的预测又比算法2具有一定优势, 因此将二者优势融合设计了算法3来确定最终趋势成分的预测值。

算法3: 融合前期消费数据与前一年收入数据进行下一年消费趋势成分的线性预测。

1) 将当前月份作为当前时刻点, 利用算法1的步骤1)~6)获得前6年的长期趋势成分 $f(t)$, 并作为这6年的真实趋势函数, 将 $f(t)$ 在最后一个月份处的函数值 b 作为下一年趋势预测的基准值;

2) 将当前月份作为当前时刻点, 利用算法1获得下一年的趋势预测函数 $g(t)$, 斜率记为 k_1 ; 利用算法2获得下一年的趋势预测函数 $h(t)$, 斜率记为 k_2 ;

3) 将去年同月份作为当前时刻点, 利用算法1获得当前月份之前一年的趋势预测函数 $g^1(t)$; 利用算法2获得当前月份之前一年的趋势预测函数 $h^1(t)$;

4) 计算 $f(t)$ 和 $g^1(t)$ 在当前月份之前12个月的平均差异 e_1 , 并作为算法1对去年同月份时刻点到当前月份时刻点趋势成分的预测误差;

5) 计算 $f(t)$ 和 $h^1(t)$ 在当前月份之前12个月的平均差异 e_2 , 并作为算法2对去年同月份时刻点到当前月份时刻点趋势成分的预测误差;

6) 构造线性函数 $F(t) = \frac{k_1 e_2 + k_2 e_1}{e_1 + e_2}$, 作为下一年的长期趋势预测函数。

1.3 季节数据预测

相邻两年的季节数据通常会呈现出相似的变化规律, 用前一年的季节性拟合函数作为下一年季节性成分的预测函数是一种简单的方法, 但是某些相邻年份的季节性成分之间也会出现不规则的波动, 所以对于季节性成分采用短期季节数据变权重加权的方法, 并通过分析已知年份季节成分相似度自适应地选择权重。对于最近两年季节成分相似度高的年份, 预测时主要考虑这两年的季节成分; 相反对于最近两年季节成分相似度低的年份, 则会均衡考虑此前6年中每一年的季节变化规律。

为了预测消费数据的季节性成分, 对于从算法1获得的前6年每一年的季节性离散修正数据 $\{(t_j^i, y_j^i)\}_{j=1}^{12}, i=1, 2, \dots, 6$, 采用如下正则化拟合方法

$$\min Pe(x_i | y_i) = \sum_{j=1}^{12} \left[y_{ij} - \sum_k c_k^i \varphi_k(t_j) \right]^2 + \lambda \int \left[L \left(\sum_k c_k^i \varphi_k(t) \right) \right]^2 dt \quad (2)$$

其中, 基函数 φ_k 采用6阶B样条基; $\varphi_k(t_j)$ 是 $\varphi_k(t)$ 在时间 t_j 处的取值; 基函数的个数 k 与拟合数据的节点数 i 的关系为 $k = i + 6 - 2$; 正则项 $L \left(\sum_k c_k^i \varphi_k(t) \right)$ 是 $\sum_k c_k^i \varphi_k(t)$ 的0, 1, 2, 3阶导数的线性组合, 即 $L = \beta_0 I + \beta_1 D^1 + \beta_2 D^2 + D^3$ 。微分算子的系数 $\beta_0, \beta_1, \beta_2$ 需要基于数据来选取, 具体地, 当已知6年中每一年的季节性成分拟合函数 $x_i(t)$ 时, 可以通过求解以下最优化问题获得

$$\min \left\{ Se(L) = \sum_{i=1}^6 \int (Lx_i(t))^2 dt = \sum_{i=1}^6 \|Lx_i(t)\|^2 \right\} \quad (3)$$

式中 Se 是主微分残差平方和。

由于正则项中的系数 $\beta_0, \beta_1, \beta_2$ 和每一年的季节性成分拟合函数 $x_i(t)$ 在初始状态下均未知,故采用迭代的方法求解,进而通过对迭代求解获得的季节性成分拟合函数进行自适应的加权组合,获得下一年的季节性成分预测。具体方法如算法 4。

算法 4: 基于季节性离散修正数据预测下一年消费季节性成分。

1) 在式(2)中设置正则化参数 $\lambda = 10^{-9}$, 第一轮微分算子参数 $\beta_0 = 0, \beta_1 = 0, \beta_2 = 0$, 对算法 1 步骤 6) 中获得的当前月份前 6 年的每一年季节性离散修正数据进行拟合, 获得第一轮每年季节性成分拟合函数;

2) 对于第一轮每年季节性成分拟合函数, 通过求解式(3)获得第二轮微分算子参数 $\beta_0, \beta_1, \beta_2$;

3) 基于第二轮微分算子参数 $\beta_0, \beta_1, \beta_2$, 重复迭代步骤 1) ~ 2) 3 次, 获得每年季节性成分拟合函数;

4) 计算第三轮第 5 年和第 6 年季节性成分拟合函数之差绝对值的积分, 记为 ε_1 ; 计算第三轮全部任意两年季节性成分拟合函数之差绝对值的积分并求平均值, 记为 ε_2 ;

5) 如果 $\varepsilon_1 < 0.6\varepsilon_2$, 令

$$\omega = \left(\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0 \right)$$

否则, 令

$$\omega = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8} \right)$$

6) 以 ω 为权重(按照此前 1 年至此前 6 年的顺序), 对此前 6 年的第三轮季节性成分拟合函数进行加权组合, 获得下一年季节性成分预测函数。

2 实验部分

2.1 实验设置

现以我国 1984 ~ 2010 年的社会消费品零售总额月度数据作为样本, 数据来源于国家统计局网站。将每个时间点的数据取对数, 作为训练模型和预测检验的数据, 将算法 3 和算法 4 获得的预测之和作为最终预测。算法 3 需要基于前 7 年的数据对下一年的数据进行预测, 算法 4 需要基于前 6 年的数据

对下一年的数据进行预测。将数据按 8 年为阶段逐段进行分组, 得到 20 组数据对象 $\{(t_i^k, y_i^k), i = 1, 2, \dots, 96\}_{k=1}^{20}$, 其中 t_i^k 为第 k 组数据的第 i 个月份, y_i^k 为这个月份的消费数据的对数值。在每一组数据中, 基于前 84 个月的数据利用算法 3 和算法 4 分别获得长期趋势成分和季节性成分, 进而将两项预测结果之和作为预测函数, 对之后 12 个月的数值进行预测。

为了检验本文所提方法的预测准确率, 采用均方根 (X_{RMSE}) 和绝对值均值 (X_{MEAN}) 两种误差来衡量最终的预测结果

$$X_{\text{RMSE}} = \sqrt{\frac{\sum_{t=1}^N (f(t) - x_t(t))^2}{N}} \quad (3)$$

$$X_{\text{MEAN}} = \frac{\sum_{t=1}^N |f(t) - x_t(t)|}{N} \quad (4)$$

式中, $x_t(t)$ 表示真实值, $f(t)$ 表示预测值, N 表示样本数目。

2.2 结果与讨论

采用本文方法对 1991 ~ 2010 年社会消费品零售总额月度数据和趋势成分预测, 并与文献[3]结果进行比较。总额预测结果如图 3, 趋势成分预测

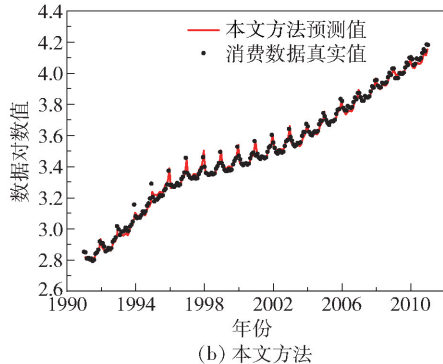
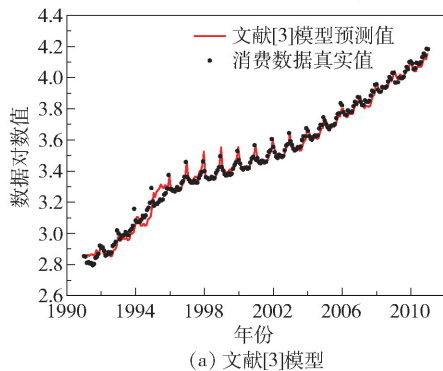


图 3 两种方法社会消费品零售总额的预测结果

Fig. 3 Two methods of forecasting the total retail sales of consumer goods

结果如图 4,预测误差如表 1。

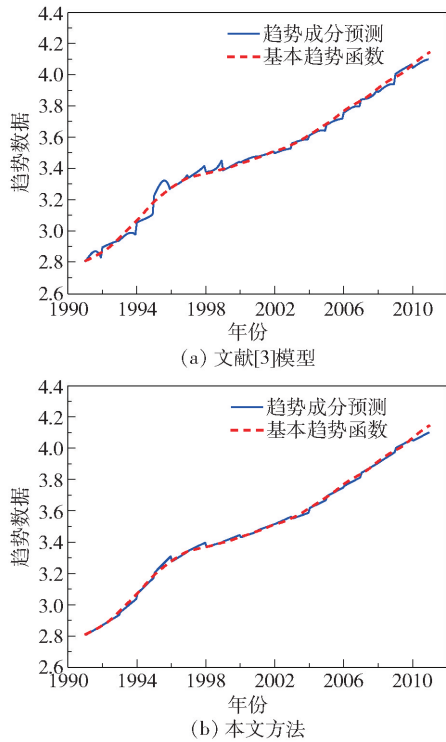


图 4 两种方法趋势成分预测结果

Fig.4 Prediction of trend component using the two methods

表 1 两种方法的预测误差

Table 1 The errors in total data for the two methods		
方法	均方根误差	绝对值均值
文献[3]	0.029 4	0.020 8
本文方法	0.018 9	0.014 9

从图 3 中可以看出,文献[3]方法在多数年份的预测较准确,而在 1994~1997 这 4 年间的预测误差较大,而本文方法的预测结果在整个预测区间都比较接近真实值。一个直观的解释是,这 4 年间数据的增长趋势与前一阶段相比发生了较为明显的变化,文献[3]只基于前期数据所以很难预测这种趋势的改变,而本文方法利用前一年收入数据增长趋势的改变就可以很好的预测出这种变化,因而取得了良好的预测效果。

由图 4 本文方法和文献[3]中基本分解方法的预测结果比较可以看到,本文方法整体呈现出比文献[3]较优的趋势拟合效果,特别地,在波动较剧烈的年份也能很好地捕捉其变化规律。

从表 1 中可以看到,在两种误差度量下本文方法较文献[3]的结果都有很大的改进。

3 结束语

本文在前人研究的基础上,提出了改进的社会消费品零售总额数据的预测方法,将总额数据分解为长期趋势成分和季节性成分分别加以预测。对于长期趋势成分的预测,引入了居民可支配收入指标的信息来辅助预测;对于季节成分的预测,通过分析季节数据相似性设计了自适应加权预测方法,并与文献[3]进行比较。通过在真实数据上的实验验证,发现本文方法能够准确地预测社会消费品零售总额数据,具有较好的实际应用价值。

参考文献:

[1] 翟志荣,白艳萍. 基于 MATLAB 的自回归移动平均模型(ARMA)在股票预测中的应用[J]. 山西大同大学学报(自然科学版), 2010, 26(6):5-7.
ZHAI Z R, BAI Y P. Application of autoregressive moving average model (ARMA) based on MATLAB in stock forecasting[J]. Journal of Shanxi Datong University (Natural Science), 2010, 26(6):5-7. (in Chinese)

[2] 全景月. 我国社会消费品零售总额的预测与分析[J]. 金融经济:理论版, 2013(11):26-29.
QUAN J Y. Forecast and analysis of China's total retail sales of social consumer goods[J]. Finance Economy: A Theoretical Edition, 2013(11):26-29. (in Chinese)

[3] 贾鑫. 面向函数型数据分析的机器学习算法研究及应用[D]. 北京:北京化工大学, 2016.
JIA X. Research and application of machine learning algorithms for functional data analysis [D]. Beijing:Beijing University of Chemical Technology, 2016. (in Chinese)

[4] 贾鑫,徐永利. 基于函数性数据分析的中国社会消费品零售总额数据研究[J]. 北京化工大学学报(自然科学版), 2016, 43(3):115-120.
JIA X, XU Y L. Study on the data of China's total retail sales of consumer goods based on functional data analysis [J]. Journal of Beijing University of Chemical Technology (Natural Science), 2016, 43(3):115-120. (in Chinese)

[5] 付长江. 居民可支配收入与消费需求的关系及对策探究[J]. 企业管理, 2012(增刊1): 274-279.
FU C J. Study on the relationship between residents' disposable income and consumption demand and its countermeasures [J]. Enterprise Management, 2012 (s1): 274-279. (in Chinese)

[6] 丁伊丽. 我国居民收入水平对社会消费品零售量的影响分析[J]. 云南农业大学学报(社会科学版), 2015,

9(2):48-51.

DING Y L. Analysis on the impact of residents' income level on the retail sales of consumer goods in China [J]. Journal of Yunnan Agricultural University (Social Science), 2015, 9(2): 48-51. (in Chinese)

[7] 剡亮亮. 基于函数性视角的经济数据分析——以主微

分分析为例[J]. 统计与信息论坛, 2013, 28(1): 40-46.

YAN L L. Economic data analysis based on functional viewpoint by the principal differential analysis[J]. Statistics & Information Forum, 2013, 28(1): 40-46. (in Chinese)

Prediction of Chinese total retail sales of consumer goods based on functional data analysis

LI XiaoXing XU YongLi*

(Faculty of Science, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: The functional data analysis method is used to predict the data of Chinese total retail sales of consumer goods. The total data is decomposed into a long-term trend component and a seasonal fluctuations component, and then the predictions for these two components are carried out respectively. For the long-term trend component, the data of disposable personal income is introduced to support the forecast. For the seasonal component, a weighted selection method with adaptive selection weights is used. The sum of the two predictions is taken as the total forecast. The experimental results using the real data show that the proposed method can give smaller prediction errors and the predictions clearly explain the observed result.

Key words: functional data analysis; Chinese total retail sales of consumer goods; disposable personal income; long-term trends

(责任编辑:汪 琴)