
ARIMAX 模型方法及其应用

——重庆城市居民可支配收入与消费支出

程燕¹

(重庆工商大学数学与统计学院, 重庆 400067)

【摘要】:可支配收入是影响消费支出的重要因素,主要讨论 ARIMAX 模型及其在人均可支配收入与消费支出关系上的应用. 首先将重庆人均可支配收入和消费支出取对数并且作一阶和二阶差分序列画图,结合单位根检验来判断其平稳性,并且进行协整检验;其次,用最小二乘法估计参数建立了反应重庆城市居民可支配收入与消费支出的 ARIMAX 模型,并用重庆城市居民人均消费支出和拟合值的拟合图与其相对误差百分比不超过 9.3%来表明 ARIMAX 模型用于研究居民消费效果比较理想.

【关键词】:ARIMAX 模型;协整检验;相对误差百分比

【中图分类号】:F063.2 **【文献标志码】**:A **【文章编号】**:1672-058X(2015)11-0080-06

DOI:10.16055/j.issn.1672-058X.2015.0011.019

1 ARIMAX 模型的理论知识

ARIMAX 模型被称为传递函数模型^[1,2]、动态回归模型,其本质是指带回归项的 ARIMA 模型,又称扩展的 ARIMA 模型.回归项的引入有助于提高模型的预测效果,引入的回归项通常是被解释变量相关程度较高的变量.而 ARIMA 模型的实质是差分运算和 ARMA 模型的组合,其结构为 $\Phi(B) d x_t = \Theta(B) \varepsilon_t$,其中 $d = (1 - B)^d$, $\Phi(B)$ 是自回归系数多项式, $\Theta(B)$ 是移动平均系数多项式.

1976 年,Cox 和 Jenkins 采用带自变量的 ARIMAX 模型^[3],为平稳多元时间序列建模.假设响应序列 $\{y_t\}$ 和自变量序列 $\{x_{1t}\}, \{x_{2t}\}, \dots, \{x_{kt}\}$ 均平稳,首先构建响应序列和自变量序列的回归模型:

$$y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^i x_{it} + \varepsilon_t$$

¹收稿日期:2015-04-02;修回日期:2015-05-26.

作者简介:程燕(1990-),女,重庆丰都人,硕士研究生,从事经济统计研究

其中, $\Phi_i(B)$ 是第 i 个自变量的自回归系数多项式; $\Theta_i(B)$ 是第 i 个自变量的移动平均系数多项式; l_i 是第 i 个自变量的延迟阶数; $\{\varepsilon_t\}$ 是回归残差序列.

由于 $\{y_t\}$ 和 $\{x_{1t}\}, \{x_{2t}\}, \dots, \{x_{kt}\}$ 均平稳, 由于平稳序列的线性组合是平稳的, 则有 $\{\varepsilon_t\}$ 是平稳的残差序列

$$\varepsilon_t = y_t - \left(\mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} \right)$$

使用 ARIMAX 模型继续提取残差序列 $\{\varepsilon_t\}$ 中的相关信息, 得出 ARIMAX 模型:

$$\begin{cases} y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} + \varepsilon_t \\ \varepsilon_t = \frac{\Theta(B)}{\Phi(B)} a_t \end{cases}$$

其中, $\Phi(B)$ 是残差序列的自回归系数多项式; $\Theta(B)$ 是残差序列的移动平均系数多项式; a_t 是零均值白噪声序列:

$$a_t \sim WN(0, \sigma_a^2).$$

2 ARIMAX 模型的应用

2.1 数据来源与处理

重庆 1985 年到 2012 年城市居民人均可支配收入和人均消费支出的数据来源于重庆市统计年鉴. 重庆城市居民人均可支配收入和消费支出分别用 x_t 和 y_t 来表示; $\ln x_t$ 和 $\ln y_t$ 表示对人均可支配收入和消费支出取对数后的数据; $\ln x_t$ 和 $\ln y_t$, $2\ln x_t$ 和 $2\ln y_t$ 分别表示对 $\ln x_t$ 和 $\ln y_t$ 取一阶差分、二阶差分. 首先, 利用原始数据与取对数的数据作图(图 1, 图 2)直观反映出重庆城市居民人均可支配收入^[4]和消费序列呈现出整体不断上涨的趋势关系。

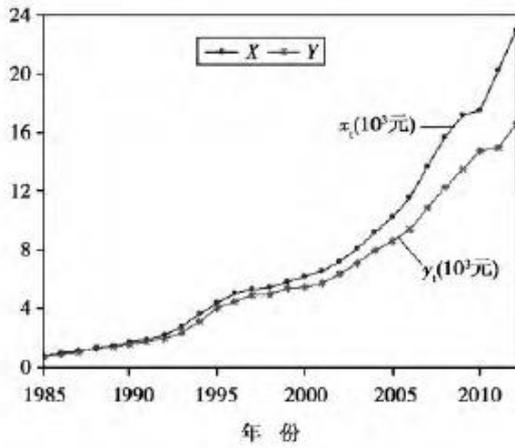


图1 重庆城市居民人均可支配收入 x_t 和人均消费 y_t 序列图

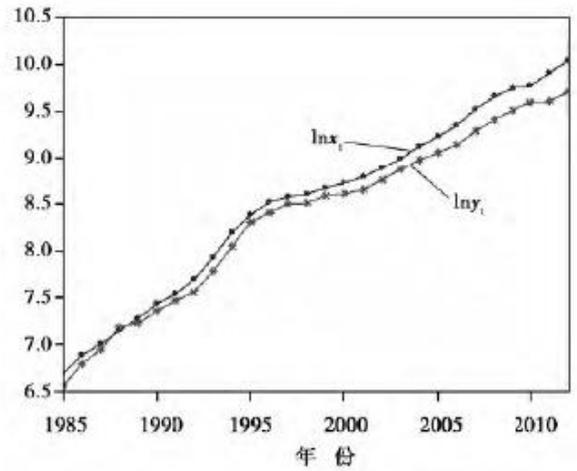


图2 重庆城市居民人均可支配收入对数 $\ln x_t$ 和人均消费支出 $\ln y_t$ 图

从图1和图2中可以看出,原始数据 t 取对数后的数据均是不平稳的,因此,将对数的一阶差分和二阶差分作图(图3,图4).

图3和图4呈现出一阶差分和二阶差分均有可能平稳,因此,对一阶差分和二阶差分序列进行平稳性检验

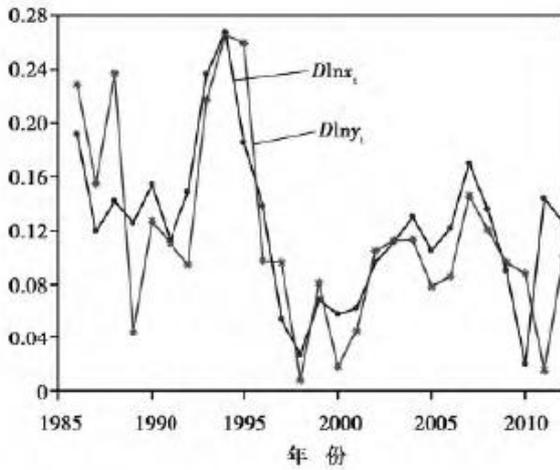


图3 $\nabla \ln x_t$ 和 $\nabla \ln y_t$

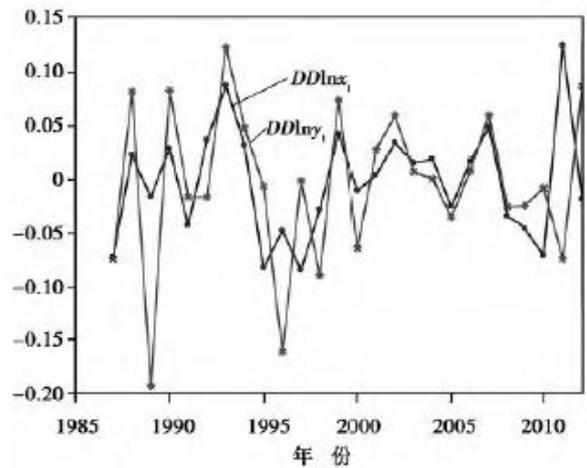


图4 $\nabla^2 \ln x_t$ 和 $\nabla^2 \ln y_t$ 图

2.2 ADF 检验平稳性

目前,最常用的平稳性^[6]统计检验法是单位根检验[6].文章采用 ADF 检验^[7],在 EViews6.0 软件中选视图/单位根检验 (View/ Unit Root Test) 得到结果,结合 AIC 准则、SC 准则^[8]等选出最优的滞后阶数,经整理得到单位根检验结果(表1).

表 1 重庆城市人均可支配收入与消费支出的 ADF 单位根检验结果

变量	(c,t,k)	ADF 统计量	D.W.统计量	5%水平下临界值	结论
$\nabla \ln x_t$	(c,t,1)	-2.777 904	2.006 313	-3.603 202	非平稳
	(c,0,1)	-2.547 632	1.930 740	-2.986 225	非平稳
	(0,0,1)	-0.884 085	1.910 619	-1.955 020	非平稳
$\nabla \ln y_t$	(c,t,0)	-3.517 086	1.999 256	-3.595 026	平稳
$\nabla^2 \ln x_t$	(c,0,1)	-3.617 928	2.028 747	-2.991 878	平稳

注:c,t 分别表示 ADF 检验中常数项和趋势项,k 表示检验所采用的滞后阶数.

表 1 单位根检验结果表明,重庆城市人均可支配收入的一阶差分序列 $\ln x_t$ 在 ADF 检验时无论采用有无常数项、有无趋势项得出的 ADF 统计量值均大于 ADF 中 5%水平下的临界值,不能拒绝有单位根的原假设,所以一阶差分序列 $\ln x_t$ 不平稳;其二阶差分序列 $2\ln x_t$ 的 ADF 统计量值-3.617 928 小于 ADF 中 5%水平下的临界值-2.991 878,则拒绝有单位根的原假设,所以 $2\ln x_t$ 是平稳的,即重庆城市居民人均可支配收入是二阶单整序列,含有常数项的单位根过程.同理可知,重庆城市居民人均消费支出一阶差分序列 $\ln y_t$ 是平稳的,是一阶单整序列,含有常数项和趋势项的单位根过程.这里由于 ADF 检验得出重庆城市居民人均消费支出一阶差分序列是平稳的,就没有对其二阶差分序列 $2\ln y_t$ 进行平稳检验,所以表 1 中没有 $2\ln y_t$ 的 ADF 检验结果.

2.3

$2\ln x_t$ 和 $\ln y_t$ 的 Johansen 协整性检验为了判断用重庆城市人均可支配收入和消费支出来建立动态回归模型是否合适,避免建立的 ARIMAX 模型是个伪回归,需要检验 $2\ln x_t$ 和 $\ln y_t$ 的协整性.选视图/ 协整检验(View/ Co-integration Test)进行协整检验,结果如表 2.

表 2 中的原假设是没有协整关系,伴随概率 Prob.**值<0.05,表明拒绝原假设,说明 $2\ln x_t$ 和 $\ln y_t$ 间存在协整关系.

表 2 $\nabla^2 \ln x_t$ 和 $\nabla \ln y_t$ 的 Johansen 协整检验 (Johansen Co-integration Test)

假 设		Trace	0.05	
原假设	特征根	Statistic	临界值	Prob. * *
None *	0.654 593	36.480 00	20.261 84	0.000 1
At most 1 *	0.366 799	10.967 32	9.164 546	0.022 5

2.4 ARIMAX 模型的建立

前面平稳性检验和协整检验表明,重庆城市居民人均消费支出对数的一阶差分序列 $\{\ln y_t\}$ 和人均可支配收入对数的二阶差分序列 $\{2\ln x_t\}$ 是平稳的,具有协整性,可以建立动态回归模型.首先,构建 $\{\ln y_t\}$ 和 $\{2\ln x_t\}$ 的回归模型:

$$\nabla \ln y_t = \mu + \frac{\Theta(B)}{\Phi(B)} B^k \nabla^2 \ln x_t + \varepsilon_t \quad (1)$$

下面通过 EViews6.0 中选视图/ 协整检验 (View/ Co-integration Test) 画出 $\{\ln y_t\}$ 和 $\{2\ln x_t\}$ 的相互关系图 (图 5), 来考察回归模型的结构.

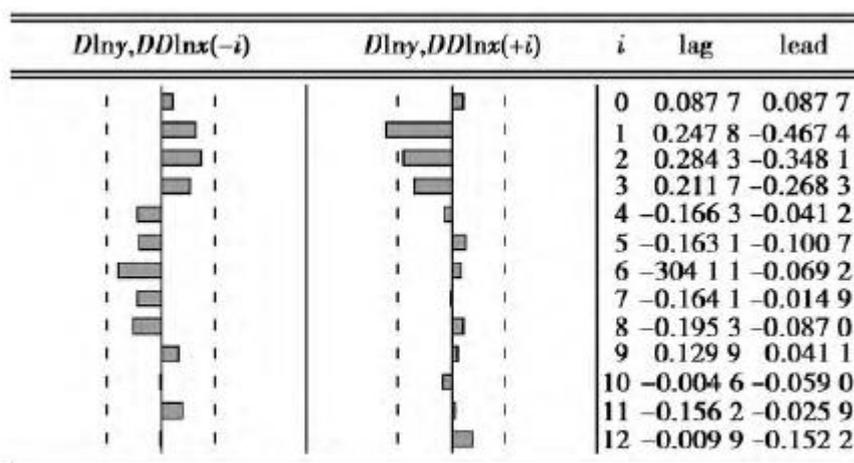


图 5 $\{\nabla \ln y_t\}$ 和 $\{\nabla^2 \ln x_t\}$ 的相互关系图

从协相关图 5 中可知, 延迟一阶的协相关系数显著非 0. 回归模型 (1) 可以进一步表示为

$$\nabla \ln y_t = \mu + \frac{\Theta(B)}{\Phi(B)} B \nabla^2 \ln x_t + \varepsilon_t \quad (2)$$

对回归模型 (2) 中的参数用最小二乘法估计, 回归模型如下:

$$\nabla \ln y_t = 0.096\ 687 + 0.204\ 47B \nabla^2 \ln x_t + \frac{1 + 0.884\ 139B^3}{1 - 0.540\ 564B} \varepsilon_t \quad (3)$$

$$AIC_2 = -3.112\ 665, SC_2 = -2.916\ 323, R_2^2 = 0.550\ 376, S.E._2 = 0.047\ 322$$

这里对上述回归模型 (3) 中的残差序列 ε_t 进行 ADF 检验, 形式采用无截距、时间趋势类型 1, 检验结果如表 3.

表 3 残差 ε_t 的 ADF 检验

被检验变量	ADF 检验值	1%临界值	5%临界值	10%临界值
残差	-6.151 750	-2.679 735	-1.958 088	-1.607 830

从检验结果来看, 残差序列平稳性显著. 考虑残差序列的性质, 下面作出残差序列的自相关和偏自相关图(图 6).

图 6 残差的自相关和偏自相关图显示, 相关和偏自相关均在 2 倍标准差^[9] 范围内, 表明残差序列是白噪声序列^[10], 即回归模型(3)已经提取了大部分信息, 回归模型(3)就是 ARIMAX 模型. 通过建立的 ARIMAX 模型来看, 前期可支配收入新增 1 单位会使消费支出以 0.204 47 的新增幅度增加, 表明居民消费支出受未来可支配收入的增加幅度影响. 对图 6 进行拟合, 得到拟合图(图 7).

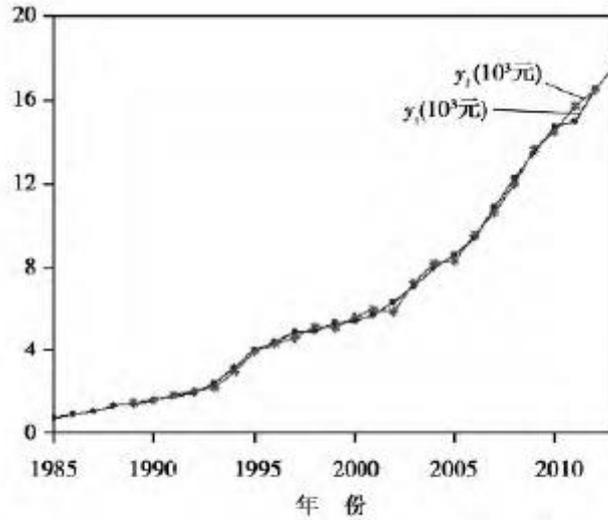


图 7 重庆城市居民人均消费支出实际值 y_t 和拟合值 y_f 图

2.5 ARIMAX 模型的事后检验

利用重庆 1985 年到 2012 年城市居民人均可支配收入和消费支出数据建立的动态回归方程(3)对 1989 年到 2013 年重庆市城市居民人均消费支出进行拟合预测(表 4), 为了考察此动态回归方程(3)的预测精度, 利用各期相对误差^[11] 来测量, 其计算公式:

$$\text{相对误差百分比} = \frac{y_f - y_t}{y_t} \times 100\%$$

表 4 重庆城市居民人均消费支出预测结果

年 份	实际值/元	预测值/元	误差百分比/%	年 份	实际值/元	预测值/元	误差百分比/%
1989	1 382.66	1 462.913 724	5.804 299	2002	6 360.2	5 899.219 717	-7.247 890
1990	1 569.97	1 592.404 856	1.428 999	2003	7 118.06	7 292.644 937	2.452 704
1991	1 754.2	1 806.553 943	2.984 491	2004	7 973.05	8 222.435 492	3.127 856
1992	1 928.63	2 022.722 746	4.878 735	2005	8 623.29	8 309.640 417	-3.637 238
1993	2 397.08	2 175.324 398	-9.251 072	2006	9 398.69	9 539.857 515	1.501 991
1994	3 126.56	2 933.166 855	-6.185 493	2007	10 876.12	10 643.212 61	-2.141 457
1995	4 051.53	3 922.892 781	-3.175 028	2008	12 269.32	12 003.900 02	-2.163 282
1996	4 467.12	4 382.018 463	-1.905 065	2009	13 507.3	13 701.061 29	1.434 493
1997	4 919.63	4 649.941 104	-5.481 894	2010	14 754.74	14 511.892 68	-1.645 894
1998	4 956.8	5 203.754 426	4.982 134	2011	14 974.49	15 721.267 41	4.986 997
1999	5 376.69	5 133.689 062	-4.519 527	2012	16 573.14	16 520.288 25	-0.318 900
2000	5 471.7	5 653.749 35	3.327 108	2013		17 723.757 51	
2001	5 724.9	5 987.390 762	4.585 072				

通过模型,可以预测年度误差绝对百分比都小于 9.3%,另外 24 个年度预测中平均绝对百分误差(MAPE)为 3.325 295 7%。一般认为 MAPE 的值低于 10%时预测精度较高,所以模型拟合效果非常好。

3 结 论

重庆城市居民人均可支配收入和消费支出是两个不平稳的序列,经过取对数、一阶和二阶差分后,得知消费支出对数的一阶差分序列是平稳序列,人均可支配收入的二阶差分序列是平稳的。建立了 ARIMAX 模型,表明重庆城市居民人均消费支出的增加值和前期的人均可支配收入有关。前期可支配收入新增 1 单位会使消费支出以 0.204 47 的新增幅度增加,即政府在促进增加当期人均可支配收入的同时,还需让居民看到未来可支配收入增加的幅度,这样就能在很大程度上促进居民消费支出,促进重庆经济健康发展。

用建立的 ARIMAX 模型对 1989-2012 年消费支出进行拟合,拟合值和实际值的拟合图表明,ARIMAX 模型用于研究居民消费支出的效果比较理想,并且算出 2013 年的消费支出 17 723.757 51 元。模型的事后检验相对误差百分比均不超过 8%,24 个年度预测中平均绝对百分误差 MAPE 为 3.325 295 7%也表明 ARIMAX 模型拟合效果理想。

参考文献:

- [1] 刘建平,岑倩青.线性回归模型与传递函数模型在时间序列应用上的比较研究[J].统计与决策,2005(2):19-21
- [2] 李子奈.计量经济学[M].北京:高等教育出版社,1996
- [3] 王燕.时间序列分析[M].北京:中国人民大学出版社,2008
- [4] 平狄克,鲁宾费尔德.微观经济学[M].北京:中国人民大学出版社,1996

-
- [5] 高启杰. 计量经济学原理与实验[M]. 北京:经济科学出版社, 2013
- [6] 刘贤赵, 邵金花. 烟台地区降水量的 ARIMA 随机模型研究[J]. 数学的实践与认识, 2006, 17(2):58-60
- [7] 王成璋, 尤梅芳. 时间序列分析[M]. 北京:机械工业出版社, 2011
- [8] 胡学锋. B-J 法在储蓄预测中的应用研究[J]. 数理统计与管理, 2001, 20(3):49-53
- [9] 易丹辉. 数据分析与 Eviews 应用[M]. 北京:中国统计出版社, 2002
- [10] 刘巍, 陈昭著. Eviews 操作简明教程[M]. 广州:暨南大学出版社, 2013
- [11] 张玉强, 彭在萍. 城镇居民消费结构及其变化[J]. 重庆教育学院学报, 2000, 13(3)14-18