



我国留学生人数的预测及其对居民消费的影响

摘要

采用 1985—2012 年数据,利用时间序列分析方法,建立了刻画我国出国留学人数变动的模型 ARIMA(2,2,2),得到较好的预测效果.同时研究了出国留学人数与我国城镇居民人均消费性支出之间的关系,建立了相应的模型,并进行了 Granger 因果关系检验.结果表明:出国留学人数数的对数每增加 1%,城镇居民人均消费性支出的对数将增加近 0.06%.

关键词

出国留学人数;时间序列分析;城镇居民人均消费性支出;Granger 因果检验

中图分类号 O213

文献标志码 A

0 引言

随着国民生活水平的提高,出国留学热的现象引起了广泛关注.近年来出国留学人数屡屡攀升,各国之间的人才交流愈加频繁.对未来留学人数走向的准确预测,无疑将对把握人才流动趋势、推动教育改革起到重要的作用;同时,越来越多的学生出国求学,必然会对我国城镇居民的消费结构产生影响.因此对出国留学这一现象的研究有助于政府因地制宜地制定政策,有效地推动我国经济增长.

近些年来,我国学者对有关出国留学开展了多方面的研究,程希等^[1]对 1949—2009 年我国与出国留学有关史实与数据进行了解读和诠释,逢丹^[2]通过对我国自费出国留学诸多现象的综合分析与研究,提出了相关对策与政策性建议,张早玲^[3]通过研究范式的转换,给我国正在或即将进行自费出国留学教育投资决策的家庭以风险警示、决策参考和投资指导,柯普等^[4]利用 GM(1,1)模型建立了我国出国留学人数的预测模型,但是其拟合效果不好且预测结果误差偏大.

本文通过时间序列分析方法,依据 1985—2012 年相关数据,建立我国出国留学人数模型并进行预测;同时研究出国留学对城镇居民消费带来的影响,建立反映相互关系的模型,并给出相关结论.

1 出国留学人数的时间序列模型

时间序列模型是建立在平稳数据的基础上的,对于像留学生人数这种非平稳序列还需要进行处理才能建立一个精确的时间序列模型,才能够较好预测数据走向.模型选择的误差也会导致预测数据的误差,因此在选择比较不同的模型时需要进行多次对比研究,将误差减到最小.

1.1 ADF 平稳性检验^[5-6]

本文首先提取 1985—2012 年我国出国留学人数(记为时间序列 T_t),所有数据取自《中国统计年鉴》,绘制出出国留学人数散点图(图 1),由图 1 可以明显看出数据不平稳且具有指数趋势.

首先对时间序列的平稳性作检验.现实生活中,多数时间序列都是非平稳的,直接应用回归分析往往会产生伪回归现象,从而导致分析的结果无效.本文采用 ADF(Augmented Dickey Fuller)单位根检验序列的平稳性.ADF 检验的一般回归可以表示为

收稿日期 2013-09-26

资助项目 国家自然科学基金(11201356);冶金工业过程系统科学湖北省重点实验室基金(Y201319)

作者简介

冯志平,男,硕士生.287651387@qq.com

余东(通信作者),女,教授,主要研究方向为应用概率统计.yudong8502@163.com

1 武汉科技大学 理学院,武汉,430065

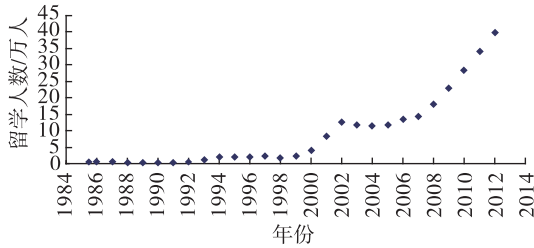


图1 1985—2012年出国留学人数散点
Fig.1 The scatter diagram of number of students studying abroad from 1985 to 2012

$$\Delta x_t = \mu + \beta t + \gamma x_{t-1} + \sum_{i=1}^p \lambda_i \Delta x_{t-i} + \varepsilon_t, \quad (1)$$

其中 Δ 是一阶差分符号, x_t 是所研究的时间序列, μ 、 β 为常数(视时间序列的特点取为零或者非零), ε_t 是随机误差项.对参数 $\gamma=0$ 进行假设检验,如果否定 $\gamma=0$ 的原假设,就得出原序列 x_t 是平稳过程.

借助于Eviews软件的检验结果,如果ADF检验统计量的值比报告的临界值大,即可得出序列非平稳的结论.

为了消除数据序列可能存在的长期趋势及异方差现象,对 T_s 取自然对数,记为 $\ln T_s$,进行单位根检验, $\ln T_s$ 通过二阶差分最终得到平稳序列.

1.2 模型的选择^[7]

要选择模型,先要识别一个模型即判断其是遵循自回归(AR)过程,还是遵循移动平均(MA)过程或是自回归移动平均(ARMA)过程.所使用的工具主要是取对数后数据的二阶差分序列的自相关(ACF)和偏自相关(PACF)图.若序列的偏自相关函数在 p 以后截尾,即 $k > p$ 时 $\rho^* = 0$,而它的自相关函数 ρ_k 是拖尾的,则此序列是自相关AR(p)序列;若随机时间序列的自相关函数截尾,即 q 以后 $\rho_k = 0$ ($k > q$),而它的偏自相关函数是拖尾的,则此序列是 q 阶滑动平均MA(q)序列.但是在观察自相关和偏自相关图时会有主观性,所以还需要理论的筛选,在选择模型的过程中使用AIC准则来作为判断模型优越的依据.其自相关和偏自相关如图2所示.

从图2可以看出该序列比较复杂,需要通过AIC准则来判断.统计出不同阶数的AIC值及 t 统计量来选择最合适的模型.通过多次的试验,本文最终选取ARIMA(2,2,2)模型,其模型的相关性统计指标如表1所示.

以 y_t 记当前的 $\ln T_s$, ε_t 表示单期的误差项,则得到的模型为

Date: 03/30/13 Time: 18: 19
Sample: 1985 2012
Included observations: 26

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.111	-0.111	0.3592	0.549
		2	0.036	0.024	0.3980	0.820
		3	-0.310	-0.307	3.4324	0.330
		4	-0.356	-0.472	7.6371	0.106
		5	0.187	0.080	8.8517	0.115
		6	-0.294	-0.484	11.998	0.062
		7	0.339	-0.095	16.402	0.022
		8	0.163	0.191	17.478	0.026
		9	0.143	0.055	18.351	0.031
		10	-0.083	-0.316	18.661	0.045
		11	-0.060	0.623	18.834	0.064
		12	-0.010	0.157	18.840	0.092

图2 二阶差分后的ACF和PACF值
Fig.2 ACF and PACF after second order difference

表1 ARIMA(2,2,2)模型的参数估计及检验结果

Table 1 Parameter estimation and test results of ARIMA(2,2,2)

变量	系数	标准误差	t 统计量	概率
常数	0.003 423	0.031 558	-0.108 461	0.914 8
AR(1)	-1.092 286	0.192 173	-5.683 854	0.000 0
AR(2)	-0.628 306	0.262 382	-2.394 621	0.027 1
MA(1)	1.838 936	0.495 975	3.707 721	0.001 5
MA(2)	2.562 346	0.712 653	3.595 505	0.001 9
R^2	0.834 677	0.799 872	23.981 61	0.000 000
调整 R^2				1.375 194

$$y_t = 0.003 423 - 1.092 286y_{t-1} - 0.628 306y_{t-2} + 1.838 936\varepsilon_{t-1} + 2.562 346\varepsilon_{t-2}. \quad (2)$$

1.3 模型的预测

同样取2009—2012年的预测数据并将其初始化再令其和实际的数据进行比较分析,通过计算得到表2,可以发现后期的预测效果很好,并给出了2013年出国留学人数的预测.

表2 2009—2012年预测值及相应残差
Table 2 Predictive value and corresponding residual from 2009 to 2012

年份	真实值/人	预测值/人	残差/人	相对误差/%
2009	229 300	229 728	428	0.19
2010	284 700	289 312	4 612	1.60
2011	339 700	343 832	4 132	1.20
2012	399 600	400 311	711	0.18
2013		468 640		

将上述预测结果与文献[4]的结果相比较,可以看出误差远小于文献[4]的误差,本文的模型较好地

描述了出国留学人数的变化规律.

2 出国留学人数与城镇居民消费性支出的关系

首先采用 1985—2012 年间我国出国留学人数与城镇居民人均消费性支出数据,城镇居民人均消费性支出(元)简记为 T_c ,对其取自然对数,记为 $\ln T_c$,作出 $\ln T_c$ 与 $\ln T_s$ 的散点图(图 3),观察其变化趋势.

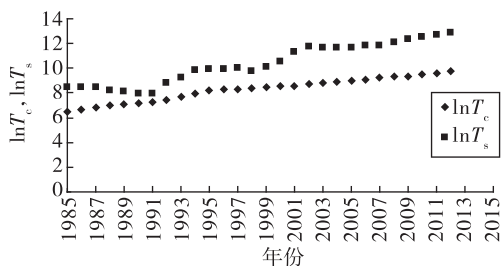


图 3 1985—2012 年 $\ln T_c$ 与 $\ln T_s$ 散点

Fig. 3 The scatter diagram of $\ln T_c$ and $\ln T_s$ from 1985 to 2012

从图 3 中可以发现, $\ln T_c$ 序列与 $\ln T_s$ 序列在波动趋势上基本一致,初步判断 2 组数据可能存在一定的联系.

2.1 平稳性检验(单位根检验)

在此考察 $\ln T_c$ 的平稳性,用 Eviews 软件对 $\ln T_c$ 做 ADF 检验,其结果如表 3 所示.

表 3 单位根检验结果

Table 3 The test results of unit root

变量	ADF 检验统计量	1%临界值	5%临界值	10%临界值	是否平稳
$\ln T_c$	-1.907 144	-4.356 068	-3.595 026	-3.233 456	否
$\Delta \ln T_c$	-1.553 190	-2.656 915	-1.954 414	-1.609 329	是
$\Delta^2 \ln T_c$	-6.477 766	-4.374 307	-3.603 202	-3.238 054	是

注: $\Delta \ln T_c$ 表示对 $\ln T_c$ 进行一阶差分, $\Delta^2 \ln T_c$ 表示对 $\ln T_c$ 进行二阶差分.

从表 3 可以看出, $\ln T_c$ 通过二阶差分最终得到平稳序列.因此, $\ln T_c$ 与 $\ln T_s$ 序列都是二阶单整的, 2 个变量可能存在协整关系.

2.2 协整检验

为了描述 $\ln T_c$ 与 $\ln T_s$ 这两者的关系,需要进行协整检验.进行协整分析时用的是 2 步法,即第一步进行计算非均衡误差,第二步对残差项时间序列进行单位根检验,若残差项时间序列平稳,则 2 组序

列之间存在协整关系.

对 $\ln T_c$ 和 $\ln T_s$ 进行协整分析,设回归模型为

$$\ln T_c = \alpha + \beta \cdot \ln T_s + \mu. \quad (3)$$

协整分析结果如表 4 所示.

表 4 协整分析结果

Table 4 Co-integration analysis

变量	系数	标准误差	t 统计量	概率
常数	2.406 733	0.327 249	7.354 435	0.000 0
$\ln T_s$	0.564 902	0.031 184	18.114 91	0.000 0
R^2	调整 R^2	F 统计量	F 统计量 P	D.W. 值
0.926 585	0.923 761	328.149 8	0.000 000	0.362 811

取模型中残差 μ 的序列,对该序列进行单位根检验,检验结果如表 5 所示.

表 5 残差 μ 单位根检验结果

Table 5 The test results of residual μ unit root

变量	ADF 检验统计量	1%临界值	5%临界值	10%临界值
μ	-3.332 590	-2.656 915	-1.954 414	-1.609 329

由此可知,残差的 ADF 检验值均分别小于 1%、5%、10% 3 个水平的临界值,序列平稳,所以 $\ln T_c$ 与 $\ln T_s$ 确实存在协整关系.回归公式为

$$\ln T_c = 2.406 733 + 0.564 902 \cdot \ln T_s + \mu. \quad (4)$$

但是回归结果显示, D.W. 值为 0.362 811, $\ln T_c$ 存在自相关,模型拟合得并不是很理想.因此,在模型中再加入 $\ln T_c$ 的自回归,回归因变量设定为 $\ln T_c$ 的滞后一阶序列、二阶序列和 $\ln T_s$,通过比较、选择,满意的回归结果如表 6 所示.

表 6 回归模型的参数估计及检验结果

Table 6 Parameter estimation and test results of regression model

变量	系数	标准误差	t 统计量	概率
常数	0.504 639	0.137 214	3.677 737	0.001 2
$\ln T_s$	0.059 693	0.027 051	2.206 701	0.037 1
$(\ln T_c)_{t-1}$	0.877 216	0.047 113	18.619 20	0.000 0
R^2	调整 R^2	F 统计量	F 统计量 P	D.W. 值
0.996 109	0.995 785	3 072.220	0.000 000	1.363 951

重新拟合后,常数项、 $\ln T_c$ 、 $\ln T_s$ 、 $(\ln T_c)_{t-1}$ 都是显著的 ($|t| \geq t_{0.025}(28-3-1) = 2.059 5$), 回归公式为

$$\ln T_c = 0.504 639 + 0.059 693 \ln T_s + 0.877 216 (\ln T_c)_{t-1}, \quad (5)$$

D.W. 值为 1.363 951, 调整 R^2 为 0.995 785, 拟合程

度较好,式(5)能更好地反映出国留学生数与城镇居民消费性支出的关系,因此将式(5)作为出国留学人数与城镇居民消费性支出之间关系的最终模型。

对式(5)残差进行单位根检验,结果如表7所示。

表7 残差单位根检验

Table 7 The unit root test of residual

ADF 检验统计量	1%临界值	5%临界值	10%临界值	是否平稳
-3.654 767	-2.656 915	-1.954 414	-1.609 329	是

根据残差单位根检验分析, $\ln T_c$ 与 $\ln T_s$ 确实存在协整关系,且出国留学生数的对数每增加1%,城镇居民人均消费性支出的对数将增加近0.06%,从量化的角度说明出国留学对城镇居民消费带来了一定的影响。

2.3 Granger 因果检验

虽然 $\ln T_c$ 与 $\ln T_s$ 都是非平稳的,但由于两者存在协整关系,因此可以进行 Granger 因果检验^[8]判断 $\ln T_c$ 与 $\ln T_s$ 是否存在因果关系,检验结果如表8所示。

表8 Granger 因果检验结果

Table 8 Results of Granger causality test

原假设	样本数	F 统计量	P 值
出国留学不是引起城镇居民消费增加的原因	27	1.434 24	0.242 8
城镇居民消费增加不是出国留学人数增加的原因		4.618 93	0.041 9

从检验结果可以看出,在10%水平下,接受原假设“出国留学不是引起城镇居民消费增加的原因”,拒绝原假设“城镇居民消费增加不是出国留学人数增加的原因”,即认为城镇居民消费支出增加是出国留学人数增加的原因。

随着城镇居民人均消费性支出的增加,我国出国留学人数不断上升.特别是进入21世纪之后,中国经济快速发展,居民收入大幅增加,有能力支付在国外的留学费用,依靠奖学金留学已不再是唯一的选择,已经出现了大量的自费出国留学,导致出国留学人数激增,因此城镇居民消费支出增加是出国留学人数增加的原因.另一方面,出国留学只涉及到很少一部分家庭的消费支出,因此出国留学不是引起城镇居民消费增加的原因,说明实际情况与

Granger 因果检验结果完全吻合。

3 结论

1) 从建立的时序模型的预测结果可以看出,整个预测效果较优,2010、2011及2012年出国留学人数的增长率分别为25.93%、18.84%和16.42%,而最后预测出2013年留学生人数将会比2012年增长17.07%,通过这个预测结果,可以对一些相关行业提前进行规划,制定相关政策。

2) 越来越多的求知人士出国深造,并且具有相对稳定的增长率,这对我国较为落后的建筑、法律、经济等高技术含量行业的进步和发展都是有帮助的,也将对我国的经济造成深远影响,而且就我国目前的出国留学市场来看,无论是开办留学中介事务所,还是出国前的各类培训,都使我国形成了一个新的经济增长点,帮助我国解决了部分当下的就业问题.从对出国留学人数和城镇居民消费性支出的关系可以看出,现在人们的消费正在步入多元化,因此国家在调研消费政策时不能局限于常用的几个指标,要了解各种新兴产业对消费的影响,从而制定符合时代特征的经济政策。

参考文献

References

- [1] 程希,苗丹国.出国留学六十年若干问题的回顾与思考(1949—2009年)[J].东南亚研究,2010(1):90-92
CHENG Xi, MIAO Danguo. Reviews on the sixty-year history of Chinese studying abroad (1949—2009) [J]. Southeast Asian Studies, 2010(1):90-92
- [2] 逢丹.当前我国自费出国留学的基本状况、成因分析及对策研究[J].清华大学教育研究,2001,22(4):17-25
PANG Dan. Analysis of the conditions, reasons and countermeasures of studying abroad at one's own expense [J]. Research on Education Tsinghua University, 2001, 22(4):17-25
- [3] 张早玲.我国家庭自费出国留学教育投资决策研究[D].昆明:云南大学公共管理学院,2011
ZHANG Zaoling. Research on decision-making of Chinese family investment on studying abroad at their own expense [D]. Kunming: School of Public Administration, Yunnan University, 2011
- [4] 柯普,吴广.基于GM(1.1)模型的出国留学人数预测研究[J].价值工程,2012(25):318-319
KE Pu, WU Guang. Study on prediction of the number of study abroad based on GM(1.1) model [J]. Value Engineering, 2012(25):318-319
- [5] 李子奈,叶阿忠.高级应用计量经济学[M].北京:清华大学出版社,2012.
LI Zinai, YE Azhong. Advanced applied econometrics

- [M]. Beijing: Tsinghua University Press, 2012
- [6] 王燕.应用时间序列分析[M].北京:中国人民大学出版社,2008
- WANG Yan. Applied time series analysis [M]. Beijing: China Renmin University Press, 2008
- [7] 易丹辉.数据分析与 Eviews 应用[M].北京:中国人民大学出版社,2008.
- YI Danhui. Data analysis and Eviews applications [M]. Beijing: China Renmin University Press, 2008
- [8] 贺红波.非平稳变量的 Granger 因果检验[J].长春师范学院学报,2004,23(3):11-13
- HE Hongbo. The test of Granger causality of nonstationary variables [J]. Journal of Changchun Teachers College, 2004, 23(3): 11-13

Prediction of abroad Chinese students and its impact on household consumption

FENG Zhiping¹ YU Dong¹

1 College of Science, Wuhan University of Science and Technology, Wuhan 430065

Abstract This article uses the data of Chinese students studying abroad during period of 1985 to 2012, and time series analysis method, to construct the model of ARIMA(2,2,2) to describe the change of abroad Chinese student number. The prediction results are in good agreement with the actual annual numbers of Chinese students studying abroad. Meanwhile, the relationship between abroad Chinese students and annual per capita consumption expenditure in China's urban area is studied, and a regression model is established hence. The co-integration relationship is analyzed and Granger causality test is done. The results show that the logarithmic value of annual per capita consumption expenditure will increase nearly 0.06% due to the increase of logarithmic value of abroad Chinese students by 1%.

Key words number of students studying abroad; time series analysis; annual per capita consumption expenditure of urban; Granger causality test