基于时间序列的山东省人均 GDP 分析与预测

许付常
(山东农业大学 经济管理学院,山东 泰安 271018)

摘要: 人均国内生产总值综合考虑了一个地区经济总量和人口基数,能够较好地反映一个地区经济增长和居民经济生活水平。本文以 1978 - 2013 年的山东省人均 GDP 数据为样本,用时间序列分析方法建立自回归预测模型。根据预测,山东省"十二五"期间人均 GDP 呈现出先慢后快的增长特点,可能原因是前期受金融危机影响较大,后期影响逐渐减弱。政府应保持宏观经济政策连续性稳定性,不断扩大开放,实施创新驱动,加快转方式调结构促升级,深化国民收入分配体制改革,实现经济持续健康发展和社会和谐稳定。

关键词: 人均 GDP; 时间序列; AR 模型; 预测
中图分类号: F201 **文献标识码:** A **文章编号:** 2095 - 0098(2015)01 - 0057 - 07

山东省位于中国东部沿海,黄河下游,陆地总面积 15.71 万平方公里,近海域面积 17 万平方公里。总人口 9579.72 万,辖 17 个市 137 个县(市、区)。[1] 山东省的省情和我国的国情有一定的相似性,即经济总量大、人口多,因此对山东省的人均 GDP 研究具有重要的意义。

一、方法选择与数据来源

时间序列是指将某一现象的某个统计指标在不同时间上的各个数值,按时间先后顺序排列形成的数列。1968 年,美国预测学家乔治·博克斯(George Box)和格威林·詹金斯(Gwilym Jenkins)创造了博克斯 - 詹金斯预测模型(B - J 模型),该模型包括 AR(自回归)模型、MA(滑动平均)模型、ARMA(自回归滑动平均)模型以及 ARIMA(差分自回归滑动平均)模型。[2] 这种模型分析经济问题的特点比较鲜明,优点是变量单一,不考虑其他变量的影响,完全根据变量自身建立模型,可以有效规避多变量引起的误差问题;缺点是它不以经济理论作指导,但是经济问题往往都是由很多因素决定的,因此这种方法在影响经济问题的大环境(比如经济政策、政治、环境等)不发生巨大变化的前提下才比较有效。[3] 本文中所用 AR 模型可以用数学公式近似表示为:$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} \cdots + \alpha_p Y_{t-p} + e_t$,它既可以利用 Y_t 过去和现在客观数据,依靠数据的惯性原理进行分析预测,又加入了白噪声随机误差项。所以,通过 AR 模型我们可以得到与数据类型配合最佳的预测。张先兵、张松(2013)建立了以时间序列为基础的 AR 模型研究我国人均 GDP,模型通过了 EViews 软件中的所有检验,模型的精度较高。[4]

本文选取了山东省自 1978 - 2013 年的历年人均 GDP 数据,共 36 个样本数据,所选数据均来自 2014 年《山东省统计年鉴》。[5]

二、建立与检验模型

(一)数据平稳性检验

1. 数据基本特征分析

收稿日期: 2014 - 10 - 30
作者简介: 许付常(1987 -),男,山东临沂人,在读硕士研究生,研究方向为货币银行理论与政策。

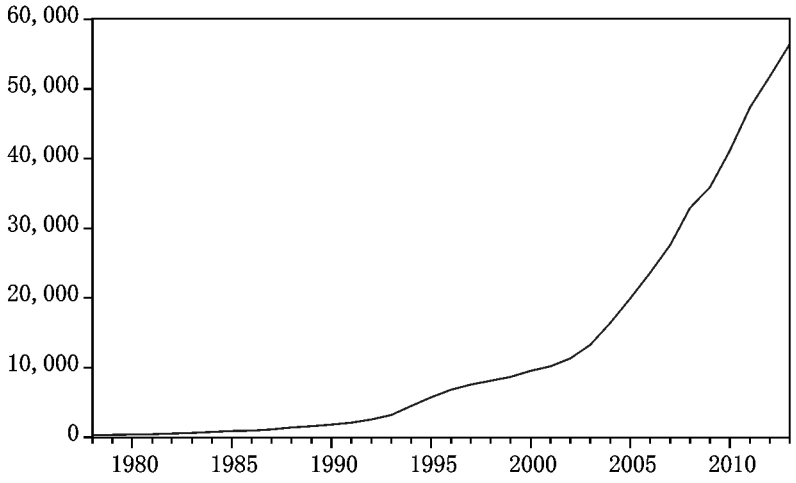


图 1　人均 GDP 的时序图

如图 1 所示，山东省人均 GDP 在 1978 年改革开放以后实现了突飞猛进的指数式增长，因而原始序列是非平稳的。为防止出现因异方差带来的虚假回归问题，我们需要对时间序列 Y_t（代表人均 GDP，下文同）进行平稳化处理。

2. 数据非纯随机性、无季节性检验

Correlogram of GDP

Date: 12/13/14 Time: 15:04
Sample: 1978 2013
Included observations: 36

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.876	0.876	29.978	0.000
		2	0.753	-0.058	52.822	0.000
		3	0.633	-0.060	69.453	0.000
		4	0.526	-0.017	81.293	0.000
		5	0.430	-0.024	89.440	0.000
		6	0.334	-0.064	94.527	0.000
		7	0.253	-0.006	97.546	0.000
		8	0.182	-0.020	99.162	0.000
		9	0.121	-0.020	99.899	0.000
		10	0.070	-0.009	100.16	0.000
		11	0.030	-0.004	100.21	0.000
		12	-0.005	-0.024	100.21	0.000
		13	-0.038	-0.033	100.30	0.000
		14	-0.071	-0.041	100.61	0.000
		15	-0.104	-0.037	101.31	0.000
		16	-0.136	-0.043	102.58	0.000
		17	-0.168	-0.041	104.60	0.000
		18	-0.197	-0.037	107.55	0.000
		19	-0.222	-0.031	111.52	0.000
		20	-0.242	-0.026	116.53	0.000

图 2　人均 GDP 的自相关和偏自相关图

分析图 2 我们清楚的发现，时序 Y_t 的自相关系数值（AC 值）是存在拖尾现象的，这说明山东省的人均 GDP 序列是非平稳、非纯随机序列。另外，我们所用数据都是年度 GDP，因此不存在季节性问题。

(二) 数据平稳化处理

对原始数据做一阶差分,时序图和相关图分别如图 3 和图 4 所示。分析图 3 我们发现一阶差分序列 D(GDP) 仍然存在较为显著的上升趋势,并且图 4 中自相关系数值也是逐渐衰减的,故 D(GDP) 仍然是非平稳序列。

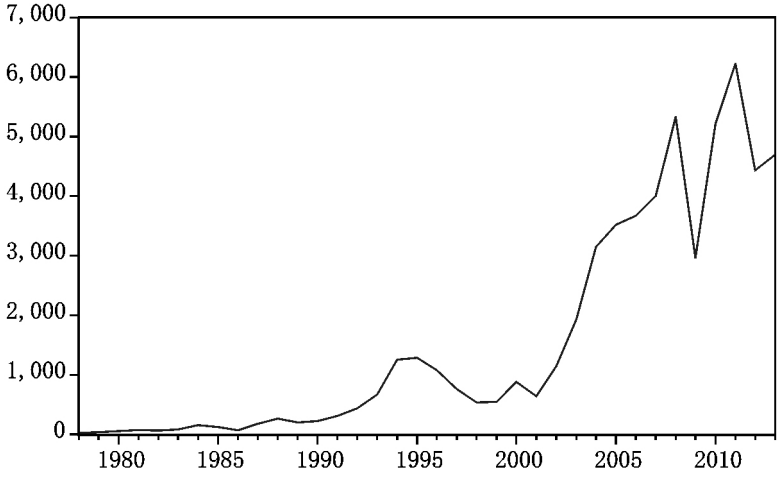
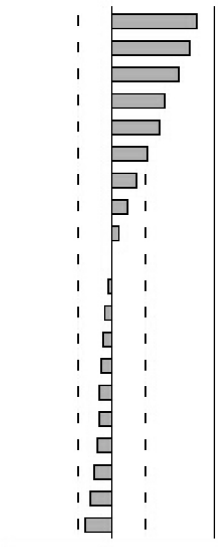


图 3　一阶差分的时序图

Correlogram of D(GDP)

Date: 12/13/14 Time: 15:48
Sample: 1978 2013
Included observations: 35



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.865	0.865	28.476	0.000
		2	0.787	0.156	52.782	0.000
		3	0.685	-0.096	71.798	0.000
		4	0.547	-0.240	84.288	0.000
		5	0.477	0.144	94.122	0.000
		6	0.357	-0.146	99.829	0.000
		7	0.261	-0.058	102.97	0.000
		8	0.169	-0.080	104.35	0.000
		9	0.077	0.005	104.64	0.000
		10	0.008	-0.045	104.64	0.000
		11	-0.033	0.103	104.70	0.000
		12	-0.067	-0.005	104.95	0.000
		13	-0.084	0.008	105.37	0.000
		14	-0.107	-0.081	106.08	0.000
		15	-0.121	0.017	107.03	0.000
		16	-0.123	-0.004	108.05	0.000
		17	-0.141	-0.075	109.49	0.000
		18	-0.179	-0.196	111.93	0.000
		19	-0.220	-0.080	115.83	0.000
		20	-0.262	-0.035	121.78	0.000

图 4　一阶差分的自相关和偏自相关图

为了获取平稳的序列,我们对 D(GDP) 做对数处理后再做差分,公式为: $D(\ln Y) = \ln(Y_t) - \ln(Y_{t-1})$。同样我们用 Eviews 8.0 对序列 D(lnY) 进行处理,得到相应时序图 5、相关图 6,通过观察图 5、图 6 发现 D(lnY) 序列已经不存在明显趋势,初步确定是平稳序列。

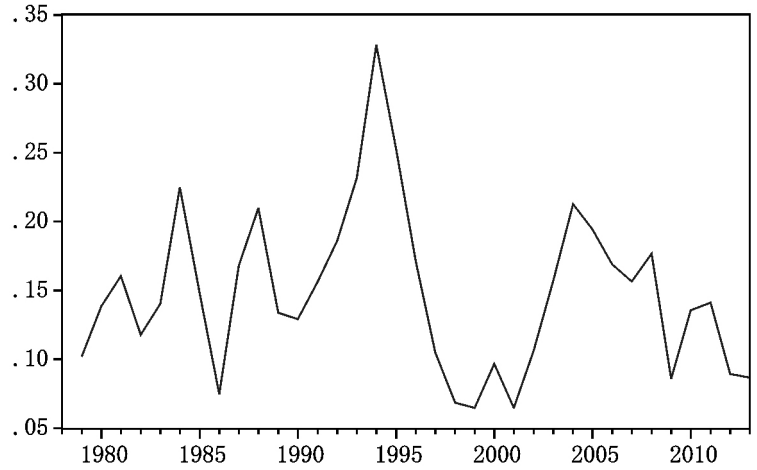


图 5　D(lnY) 的时序图

Correlogram of D(lnY)

Date: 12/13/14 Time: 15:58
Sample: 1978 2013
Included observations: 35

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.573	0.573	12.500	0.000
		2	0.134	-0.289	13.206	0.001
		3	0.020	0.140	13.222	0.004
		4	-0.183	-0.374	14.613	0.006
		5	-0.441	-0.236	23.021	0.000
		6	-0.335	0.136	28.022	0.000
		7	-0.262	-0.347	31.196	0.000
		8	-0.233	0.122	33.802	0.000
		9	-0.050	-0.068	33.925	0.000
		10	0.245	0.234	37.031	0.000
		11	0.279	-0.080	41.244	0.000
		12	0.195	-0.083	43.387	0.000
		13	0.082	-0.119	43.782	0.000
		14	-0.019	-0.035	43.805	0.000
		15	-0.101	0.115	44.460	0.000
		16	-0.051	0.050	44.638	0.000
		17	-0.120	-0.196	45.673	0.000
		18	-0.179	0.031	48.119	0.000
		19	-0.091	0.009	48.785	0.000
		20	0.030	-0.024	48.864	0.000

图 6　D(lnY) 的自相关和偏自相关图

为了确认对 D(lnY) 平稳性的判断，我们用不含趋势项的 ADF 模型进行检验，检验结果如表 1 所示。ADF 检验的原假设是: 原序列存在单位根，原序列为非平稳序列。从表 1 可以看出序列 DY、lnY 的 ADF 检验值均显著大于 5% 的临界值，皆为非平稳序列。而 D(lnY) 的 ADF 检验值(－3.280022) 小于 5% 显著水平的临界值(－2.954021) ，因此 D(lnY) 序列是平稳的，这和之前的预判是吻合的。

表 1　各变量序列的单位根检验结果

序列	检验方式	ADF 检验值	5% 临界值	P 值	结论
DY	(C,T,0)	－2.633582	－3.548490	0.2689	非平稳
lnY	(C,T,1)	－2.563993	－3.548490	0.2979	非平稳
D(lnY)	(C,N,1)	－3.280022	－2.954021	0.0241	平稳

(三) 模型识别与定阶

图 6 给出了序列 D(lnY) 的相关图，结果显示，自相关系数值具有波动性，可近似认为是拖尾的，而偏自相关系数值(PAC 值) 只在延迟 1 阶显著大于 2 倍标准差，可以近似认为是截尾的，同时为更加准确定阶，我们分别建立 AR(1)、AR(2) 模型，通过比较两个模型的 AIC 值，我们认为 AR(2) 模型模拟效果更好，因此采用 AR(2) 模型进行模拟预测。

(四) 模型建立与适应性检验

因为序列 D(lnY) 的均值为非零的，所以在建立模型时加入漂移项，加入漂移项后的 AR(2) 模型输出结果如表 2，具体估计结果如下:

$$D(\ln Y)_t = 0.1487 + 0.7721(D(\ln Y)_{t-1} - 0.1487) - 0.3034(D(\ln Y)_{t-2} - 0.1487) + e_t$$
$$(9.5) \quad (4.4) \qquad\qquad (-1.7)$$

整理上述模型得: $D(\ln Y)_t = 0.0790 + 0.7721D(\ln Y)_{t-1} - 0.3034D(\ln Y)_{t-2} + e_t$，漂移项 $\alpha = 0.0790$，特征根位于单位圆内，满足平稳性要求。根据图 7 所示对残差序列进行白噪声检验，可以看出自相关和偏自相关系数都在零值左右小范围波动，Q 统计量的相伴概率显著大于 5%，可认定残差序列为白噪声序列。另外，由表 2 可以看出 AR(1)、AR(2) 和常数项的相伴概率都很小，参数显著，因此可用该模型作预测。

表 2　AR(2) 模型输出结果

Dependent Variable: D(lnY)
Method: Least Squares
Date: 12/14/14 Time: 10:50
Sample (adjusted): 1981 2013
Included observations: 33 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.148682	0.015641	9.506207	0.0000
AR(1)	0.772101	0.174669	4.420377	0.0001
AR(2)	-0.303423	0.175793	-1.726027	0.0946

R-squared	0.406008	Mean dependent var	0.149845
Adjusted R-squared	0.366408	S.D. dependent var	0.059875
S.E. of regression	0.047659	Akaike info criterion	-3.162965
Sum squared resid	0.068143	Schwarz criterion	-3.026919
Log likelihood	55.18892	Hannan-Quinn criter.	-3.117189
F-statistic	10.25286	Durbin-Watson stat	1.913055
Prob(F-statistic)	0.000404		

Inverted AR Roots	.39+.39i	.39-.39i

Correlogram of RESID

Date: 12/13/14 Time: 20:14
Sample: 1978 2013
Included observations: 34

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.183	0.183	1.2388	0.266
		2	-0.227	-0.269	3.2054	0.201
		3	0.063	0.183	3.3598	0.339
		4	-0.001	-0.142	3.3598	0.500
		5	-0.363	-0.311	8.9293	0.112
		6	-0.095	0.052	9.3236	0.156
		7	-0.071	-0.299	9.5518	0.215
		8	-0.161	-0.044	10.778	0.215
		9	-0.088	-0.182	11.162	0.265
		10	0.243	0.161	14.173	0.165
		11	0.159	0.001	15.523	0.160
		12	0.085	0.070	15.928	0.195
		13	0.056	-0.037	16.111	0.243
		14	0.015	-0.080	16.125	0.306
		15	-0.049	0.094	16.277	0.364
		16	0.121	0.151	17.270	0.368
		17	-0.036	-0.009	17.364	0.430
		18	-0.199	-0.059	20.393	0.311
		19	-0.086	0.014	20.999	0.337
		20	0.069	0.029	21.421	0.373

图 7　AR(2) 模型残差的自相关和偏自相关图

三、模型预测

Eviews8.0 提供的预测(Forecast)功能包含静态预测(Static Forecast)和动态预测(Dynamic Forecast)两种,动态预测是多步预测,这种预测的结果只有第一个预测值是用解释变量的真实值预测的,其后各期预测值都是用递推预测得出。静态预测则是进行一系列的一步预测,用解释变量的真实值来进行预测。[6] 对于一步向前预测,静态预测比动态预测更准确,因此,本文采用静态预测方法,输出结果如图 8 所示。

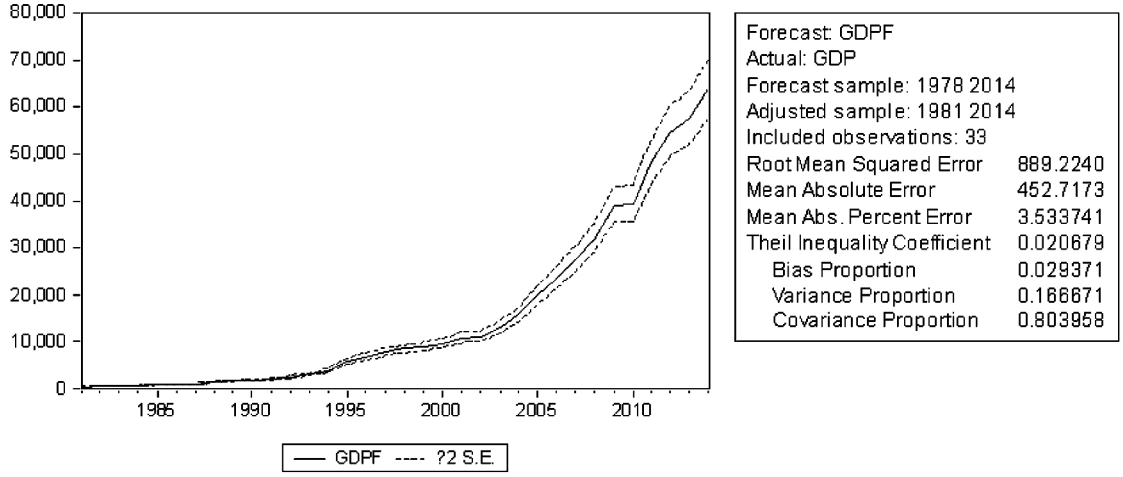


图 8 AR(2)模型预测结果

从图 8 可以看出,预测人均 GDP 的两倍标准差置信带比较窄,平均绝对百分误差(Mean Abs. Percent Error)小于 5%,希尔不等系数(Theil Inequality Coefficient)接近于 0,协变率(Covariance Proportion)接近于 1,以上说明 AR(2)模型预测精度较高。表 3 为 2011 -2013 年实际值与预测值的比较,以及根据 AR(2)模型预测的 2014、2015 年的人均 GDP。需要指出的是,由于静态预测是一步向前预测,因此 2015 年的预测值是根据 1978 -2014 年数据(2014 年数据为预测数据)重新建立模型预测的,所以在预测精度方面略有降低。

表 3 实际值与预测值比较

年份	实际值	预测值	相对误差(%)
2011	47335	48122	1.66
2012	51768	54820	5.90
2013	56463	57517	1.87
2014	- - -	63590	- - -
2015	- - -	73470	- - -

预测结果显示,2014 年山东省人均 GDP 将突破六万大关达到 63590 元,名义增长率为 12.62%。2015 年人均 GDP 则有望增至 73470 元,名义增长率为 15.54%。因此,在没有大的人口变动情况下,山东省人均 GDP 仍将快速稳定增长。

四、结语

本文以 1978 -2013 年的山东省人均 GDP 数据为样本,用时间序列分析方法建立自回归预测模型。根据预测,山东省"十二五"期间人均 GDP 呈现出先慢后快的增长特点,可能原因是前期受金融危机影响较大,后期影响逐渐减弱。另外,山东省"十二五"以来实施的蓝黄两区科技创新、特色产业培育、未利用地开发等重大项目成果逐渐显现,这也为后期经济增长提供了强有力的支持。

2015 年是我国"十二五"规划的最后一年,政府应保持宏观经济政策的连续性、稳定性,不断扩大开放,实施创新驱动,加快转方式、调结构、促升级,深化国民收入分配体制改革,[7] 实现经济持续健康发展和社会和谐稳定。

参考文献:

[1] 山东省情网. 山东概况 [EB/OL]. http://www.sdsqw.cn/,2014 - 08 - 21.

[2] 博克斯等. 时间序列分析:预测与控制 [M]. 北京:中国统计出版社,1997.

[3] 田永强. 谈谈 AR 模型在短期经济预测中的应用 [J]. 数理统计与管理,1988(05):26 - 30.

[4] 张先兵,张松. 我国人均国内生产总值的时间序列分析(1952 - 2011)与静态预测(2012 - 2015) [J]. 现代管理科学,2013(06):64 - 67.

[5] 山东省统计局. 山东省统计年鉴 2014 [M]. 北京:中国统计出版社,2014.

[6] 高铁梅. 计量经济分析方法与建模 [M]. 北京:清华大学出版社,2009.

[7] 李克强. 政府工作报告:2014 年 3 月 5 日在第十二届全国人民代表大会第二次会议上 [R]. 北京:人民出版社,2014.

The Analysis and Forecast of the Per Capita GDP in Shangdong Province Based on Time - Series

XU Fuchang

(School of Economics and Management, Shandong Agricultural University, Taian, Shangdong 271018, China)

Abstract: The per capita GDP considering the total economy and population base of a region, and it is an efficient tool to reflect a region's economic development and living standards of residents. In this paper, we utilize time series analysis method to establish autoregressive model according to the per capita GDP index of Shandong Province from 1978 to 2013. According to the forecast, Shandong Province's Per Capita GDP will grow from slow to fast during "the Twelfth Five - Year Plan", the reason might be that the affects of financial crisis in the early period are more obvious than the affects in the later period. The government should maintain the continuity and stability of our macroeconomic policies; constantly expand opening up; drive development through innovation; accelerate transformation of the growth model, structural adjustments, and industrial upgrading; deepen reform of the income distribution system. Thus we will achieve the sustained and healthy economic development and social harmony and stability.

Key words: per capita GDP; time series; AR model; forecast

(责任编辑:黎　芳)