

Per Capita GDP Prediction Research of Shanghai Bases on Time Series Analysis

Wei Shan, Xiang Yuan

Shanghai Maritime University, Shanghai
Email: yuanx2109@163.com

Received: Jun. 24th, 2016; accepted: Jul. 8th, 2016; published: Jul. 15th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Shanghai is an important financial center city in our country, researching the development of its per capita GDP has important significance. According to the Shanghai per capita GDP from 1978 to 2014 and with the aid of Eviews software, this article uses the time series analysis method, and chooses the optimal data model. Then the data from 2015 to 2018 is predicted by the established model and some relevant suggestions are given by the analysis of the results.

Keywords

Per Capita GDP, Times Series, Predict, Advice

基于时间序列分析的上海人均GDP预测研究

单伟, 袁象

上海海事大学, 上海
Email: yuanx2109@163.com

收稿日期: 2016年6月24日; 录用日期: 2016年7月8日; 发布日期: 2016年7月15日

摘要

上海作为我国的重要的金融中心城市, 研究其人均GDP的发展状况具有重要的现实意义。文章采用时间

序列分析方法, 根据1978~2014年的上海人均GDP数据, 借助Eviews软件选择出最优的数据模型。通过所建立的模型预测2015~2018年的数据, 并结合分析的结果, 给出相关的建议。

关键词

人均GDP, 时间序列, 预测, 建议

1. 引言

人均 GDP 是一个国家或地区的生产总值与人数的比值, 通常作为发展经济学中衡量经济发展状况的指标, 是最重要的宏观经济指标之一, 也是反应人民生活水平的重要标准。2010 年上海人均生产总值为 76,074 元, 2011 年上涨至 82,560 元。2014 年人均 GDP 更是达到了 97,370 元。可见上海人均 GDP 一直保持着从低速到高速的发展势头。目前我国已经进入经济转型期, 未来上海人均 GDP 发展趋势是否有较大的波动或者仍以较快的速度发展是国内外关注的焦点。本文将通过时间序列分析手段, 来探讨和分析未来几年上海人均 GDP 的发展情况。

现代计量经济学研究表明, 经济时间序列绝大多数是非平稳的。而在时间序列分析中特别是单整自回归移动平均模型在处理非平稳经济时间序列时有着良好的效果, 因而运用较为广泛。国外文献中, Andrei 和 Bugudui [1] (2011)两位学者在研究美国 GDP 中采用了 ARIMA(1,1,1)模型, 结果表明美国经济呈持续上升的发展趋势。国内文献中, 何新易[2] (2012), 张丙见[3] (2013)均采用了时间序列的方法对我国的 GDP 进行了分析和预测。郭景威、李宏斌[4] (2012)采用时间序列分析对我国的人均 GDP 进行分析和预测。华鹏和赵学民[5] (2010)利用 ARIMA 模型预测了广东省 GDP。熊志斌[6] (2011), 龙会典、严广乐[7] (2013), 汪丽艳、叶明海[8] (2013)均引入了神经网络的集成模型对 GDP 进行分析和预测。梅沁、刘宴先、景小南[9] (2011)建立了多因素动态系统 VAR 模型分析上海 GDP 的发展趋势。

在研究已有文献中可以发现, 学者们采用不同的方法对国内主要城市 GDP 进行了分析研究, 但尚未发现有学者采用时间序列分析的方法对上海市人均 GDP 的未来发展趋势进行分析和预测。本文以上海市 1978~2014 年 GDP 时间序列为例, 借助 Eviews 软件处理数据, 并对未来几年内的变化趋势做出判断。

2. 时间序列模型

时间序列是研究事物动态变化最有效的方法, 几乎涉及经济中的各个领域。基本思路是根据有限长度的数据, 建立能够比较精确的反映序列中所包含动态依存关系的数学模型, 通过对模型的检验最终确定模型形式, 再利用模型进行预测, 时间序列含有自回归模型 AR, 移动平均模型 MA, 自回归移动平均模型 ARMA, 单整自回归移动平均模型 ARIMA。

2.1. AR(p)模型

定义: 为非零均值平稳序列, 满足方程

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

其中 ϕ_p 为 X_t 对 X_{t-p} 的依赖程度, $\{\varepsilon_t\}$ 为白噪声序列, 满足 $E\varepsilon_t = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$, 且 $E(X_s \varepsilon_t) = 0$, 则称 $\{X_t\}$ 满足 p 阶自回归模型, 记作 AR(p)。

2.2. MA(q)模型

定义: 如果一个随机过程可用下式表示

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

其中 $\theta_1, \theta_2, \dots, \theta_q$ 是回归参数, $\{\varepsilon_t\}$ 为白噪声过程, 满足 $E\varepsilon_t = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$, 则上式称 q 阶移动平均模型, 记作 $MA(q)$ 。

2.3. ARMA(p,q)模型

定义: 如果一个随机过程满足

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

其中 $E\varepsilon_t = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$ 且 $E(X_t \varepsilon_t) = 0$, 则上式称为满足自回归滑动平均模型, 记作 $ARMA(p,q)$ 。

2.4. ARIMA 模型

定义: 如果时间序列的 d 阶差分 $Y_t = (1 - B^d)X_t$ 是一个平稳的 $ARMA(p,q)$ 模型, d 为大于 0 的整数, 则称 $\{X_t\}$ 是具有阶 p , d 和 q 的自回归求和移动平均模型, 即 $ARIMA(p,d,q)$, 它是一种非平稳序列, 在实际研究当中要先转化为平稳序列在进行分析。

3. ARIMA(p,d,q)模型的预测步骤

3.1. 平稳化处理

生活中常见的数据大多数是非平稳的, 不能直接进行拟合。首先进行平稳化处理。方法是先作出原始序列的时序图判断序列是否具有明显的趋势和季节性变化。这是最直观的方法, 但是带有主观性。可以通过单位根检验(ADF), 通过 t 统计量和 p 值做出判断; 若原始序列是非平稳可对其进行对数处理或直接逐次差分, 再分别用同样的方法检验, 直到结果检验具有平稳性。差分的次数即为 d 的值。

3.2. 模型识别

用样本自相关图和样本偏相关图识别模型形式。在实际中自相关系数(ACF)和偏自相关系数(PACF)是未知的, 对于给定的序列 x_1, x_2, \dots, x_t , 我们可以使用样本的自相关系数和偏相关系数对其进行估计, 由于两者均为随机变量, 对应的模型不可能具有严格的“截尾性”, 而是在几步之后围绕着零值上下波动, 因此我们可以借助截尾性来判断 p 和 q 的值。

3.3. 模型定阶和参数估计

由于模型具有多样性, 利用 B-J 建模方法定出相应阶数, 并得到多个模型, 在对不同的模型进行参数估计。在对各个模型的 AIC, SC 和 DW 等统计量的大小做出比较, 以确定最优的模型。

3.4. 模型检验和预测

通过残差序列图对模型进行适应性检验和参数的显著性检验。最后利用所建立的模型进行预测。

4. 上海人均 GDP 预测的实证分析

4.1. 数据的来源与说明

本文所研究的上海市人均 GDP 来自中国统计年鉴[10], 历史数据时间是从 1978~2014 年, 记原始序列, 对数序列, 对数一阶差分及二阶差分序列分别为 $\{X_t\}, \{Y_t\}, \{Z_t\}, \{W_t\}$ 。首先利用 Eviews 做出的时序图, 如图 1。

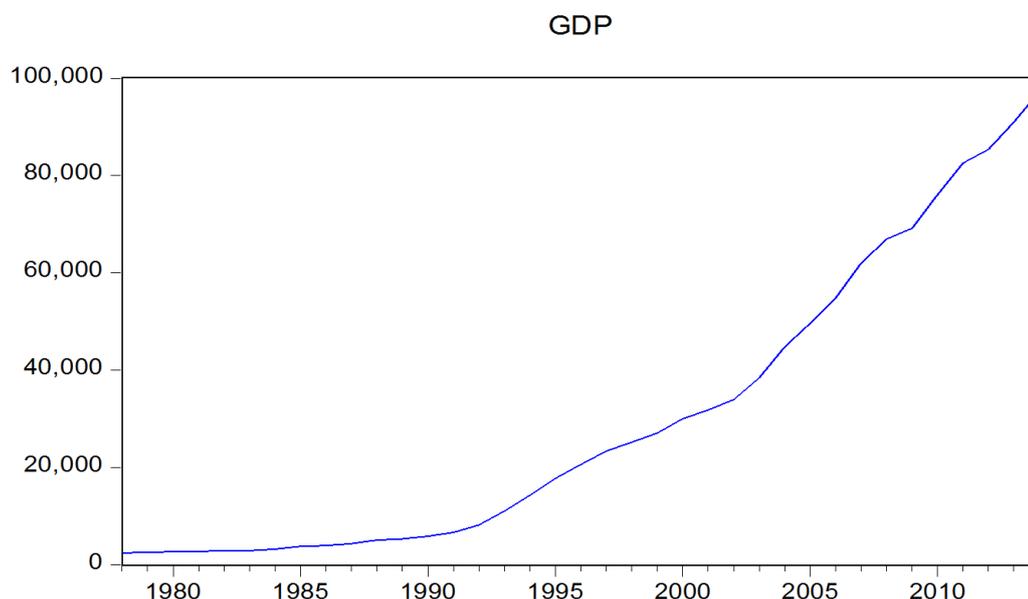


Figure 1. Change trend chart of per capita GDP of Shanghai in 1978-2014

图 1. 1978~2014 年上海人均 GDP 变化态势图

4.2. 确定 ARIMA 模型中的差分系数的确定

通过 Eviews 作出 X_t 的时序图, 如图 1 呈明显上升的趋势。由于观察时序图具有主观性。因此可对 X_t 进行平稳性检验。结果如表 1 所示 ADF 统计量大于临界值且 p 值大于 0.05 接受原假设, 都证明 X_t 是非平稳序列。

观察序列有明显的指数上升趋势, 可以取 X_t 的对数 Y_t 进行检验, 但检验结果显示仍然是非平稳的。同理 Y_t 的一阶差分 Z_t 也是非平稳的, 而二阶差分 W_t 的 ADF 统计量和 P 值分别为 -6.38618 和 0.0000, 均通过相关临界值。因此 W_t 为平稳时间序列, ARIMA 模型的差分系数为 2。

4.3. 模型的参数估计与检验

通过前面所述可知 d 为 2, 所以在 ARIMA($p,2,q$) 还需要确定自回归和移动平均系数。需要利用软件做出 W_t 的相关图。

由图 2 所示, 序列 W_t 的自相关系数和偏相关系数在 2 处接近临界水平。因此可取相关系数 $p=2$, $q=2$, 所以根据 B-J 方法, 可以分别尝试建立 ARIMA(2,2,0), ARIMA(0,2,2) 和 ARIMA(2,2,2) 模型。在检验的过程中对 AR(i), MA(i) 不显著的进行剔除, 再得到不同模型的统计量值如表 2 所示。根据最优原则, 当 AIC, SC 越小时, 所建立的模型越好。DW 表示模型的残差序列是否有自相关, DW 等于 2 是表示无自相关, 越接近 0 表示越有强的正相关, 越接近 4 表示越有强的负相关。

结合表的数据, ARIMA(0,2,2) 的 DW 值相对偏离 2, 代表残差序列具有较强的相关性, 可以先舍弃。而 ARIMA(2,2,2) 的 DW 值更接近 2, 且 AIC 与 SC 值比 ARIMA(2,2,0) 小, 因此最终可以确定选择 ARIMA(2,2,2) 模型, 结果如表 3 所示。

参数估计后, 应对拟合的模型进行适应性检验, 实质是对模型的残差序列进行白噪声检验。通常是采用残差相关图检验。对上述检验结果做出残差相关图, 由图 3 可知, P 值都大于 0.05, 说明所有的 Q 值都小于显著水平为 0.05 的卡方分布临界值。因此, ARIMA(2,2,2) 模型的残差序列是白噪声序列, 表明拟合结果有效。

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	-0.099	-0.099	0.3733	0.541
** .	** .	2	-0.278	-0.291	3.4059	0.182
. * .	. * .	3	0.205	0.155	5.1062	0.164
. .	** .	4	-0.150	-0.218	6.0420	0.196
. .	. .	5	-0.068	0.011	6.2410	0.283
. .	*** .	6	-0.200	-0.404	8.0222	0.236
. .	. * .	7	0.071	0.133	8.2575	0.310
. * .	. .	8	0.097	-0.194	8.7079	0.368
. .	. * .	9	-0.075	0.162	8.9845	0.439
. * .	. .	10	0.134	-0.120	9.9150	0.448
. .	. * .	11	-0.021	0.088	9.9382	0.536
. .	. .	12	0.041	-0.050	10.031	0.613
. .	. .	13	0.001	0.062	10.031	0.691
. .	. .	14	-0.112	-0.090	10.805	0.701
. .	. .	15	0.059	0.073	11.034	0.750
. .	. .	16	-0.006	-0.012	11.036	0.807

Figure 2. Correlation diagram of sequence W

图 2. 序列 W 的相关图

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	-0.045	-0.045	0.0728	0.787
. .	. .	2	-0.136	-0.138	0.7616	0.683
. ** .	. ** .	3	0.242	0.234	3.0126	0.390
. .	. .	4	-0.130	-0.142	3.6805	0.451
. .	. .	5	-0.125	-0.070	4.3268	0.503
. .	** .	6	-0.138	-0.255	5.1420	0.526
. .	. * .	7	0.043	0.093	5.2253	0.632
. .	. .	8	-0.007	-0.043	5.2276	0.733
. .	. .	9	-0.048	0.050	5.3397	0.804
. * .	. .	10	0.088	-0.026	5.7274	0.838
. .	. .	11	0.012	0.006	5.7350	0.890
. * .	. .	12	0.076	0.071	6.0526	0.913
. .	. .	13	-0.011	-0.017	6.0596	0.944
. .	. .	14	-0.098	-0.088	6.6462	0.947
. .	. .	15	0.041	-0.002	6.7542	0.964
. .	. .	16	-0.016	0.016	6.7710	0.977

Figure 3. Residual sequence correlation diagram

图 3. 残差序列相关图

Table 1. Sequence unit root test

表 1. 序列单位根检验

	X	Y	Z	W
检验形式(C,T,P)	(C,T,0)	(C,T,0)	(C,T,0)	(0,0,0)
ADF 值	-0.704938	-1.98169	-2.6432	-6.38618
P 值	0.9641	0.5607	0.2649	0
1%的临界值	-4.273277	-4.24364	-4.24364	-2.63473
5%的临界值	-3.557759	-3.54428	-3.54428	-1.951
10%的临界值	-3.212361	-3.2047	-3.2047	-1.61091

注: (C,T,P)分别代表单位根方程含有常数项, 趋势和滞后阶数。

Table 2. Different p and q corresponding statistical quantity
表 2. 不同 p 与 q 对应的统计量值

	AIC	SC	DW
ARIMA(2,2,0)	-3.033413	-2.988064	2.133777
ARIMA(0,2,2)	-3.334093	-3.289654	1.711952
ARIMA(2,2,2)	-3.340168	-3.249471	0.053145

Table 3. Model fitting estimation results
表 3. 模型拟合的估计结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.421881	0.154933	2.722988	0.0105
MA(2)	-0.999832	0.064801	-15.42919	0.0000
R-squared	0.362551	Mean dependent var		0.001230
Adjusted R-squared	0.341989	S.D. dependent var		0.054525
S.E. of regression	0.044229	Akaike info criterion		-3.340168
Sum squared resid	0.060643	Schwarz criterion		-3.249471
Log likelihood	57.11277	Hannan-Quinn criter.		-3.309651
Durbin-Watson stat	2.053145			
Inverted AR Roots	0.65	-0.65		
Inverted MA Roots	1.00	-1.00		

4.4. 模型的预测

根据表 3 得到序列 W_t 的模型为:

$$W_t = 0.421881W_{t-1} + \varepsilon_t - 0.999832\varepsilon_{t-2}, \varepsilon_t \sim WN(0, \sigma^2)$$

由 W_t 可推导 Z_t 的预测公式

$$Z_t = Z_{t-1} + W_t$$

同理 $Y_t = 2Y_{t-1} - Y_{t-2} + W_t$

再由 Y_t 推导 X_t 的预测公式

$$X_t = \exp(2Y_{t-1} - Y_{t-2} + W_t)$$

由模型推导出的公式对进行预测。考虑到文章的篇幅, 表 4 只做出 2001~2014 年的真实值与预测值大小的比较。数据结果显示真实值与预测值之间的相对误差均仅在 5% 以下。从侧面看出 ARIMA(2,2,2) 模型的预测效果较好。由于 ARIMA 模型具有较强的短期预测能力, 因此在所建立模型的基础上预测出 2015~2018 年上海市人均 GDP 预测值分别为 103,255 元, 109,494 元, 116,408 元, 123,720 元。整体拟合的结果如图 4 所示, 可以发现预测值基本上与真实值在一条线上。图像表明在未来几年内上海人均 GDP 仍以较快的速度发展并且没有较大的波动。

5. 上海提高人均 GDP 的相关措施分析

如前所述上海人均 GDP 依然保持着稳中求进的发展速度, 这基本符合我国进入经济转型期的发展特

Table 4. ARIMA (2,2,2) predicted values and the relative error table of per capita GDP in Shanghai
表 4. 上海人均 GDP 的 ARIMA(2,2,2)预测值及相对误差表

年份	真实值	预测值	误差(%)	年份	真实值	预测值	误差(%)
2001	31,799	33,244	4.54	2010	76,074	72,817	-4.28
2002	33,958	33,537	-1.24	2011	82,560	84,051	1.81
2003	38,486	36,714	-4.6	2012	85,373	88,175	3.28
2004	44,839	43,188	-3.68	2013	90,993	89,146	-2.02
2005	49,648	51,194	3.11	2014	97,370	97,952	0.59
2006	54,858	54,887	0.05	2015	未知	103,255	未知
2007	62,040	61,510	-0.85	2016	未知	109,494	未知
2008	66,932	70,003	4.59	2017	未知	116,408	未知
2009	69,165	72,153	4.32	2018	未知	123,720	未知

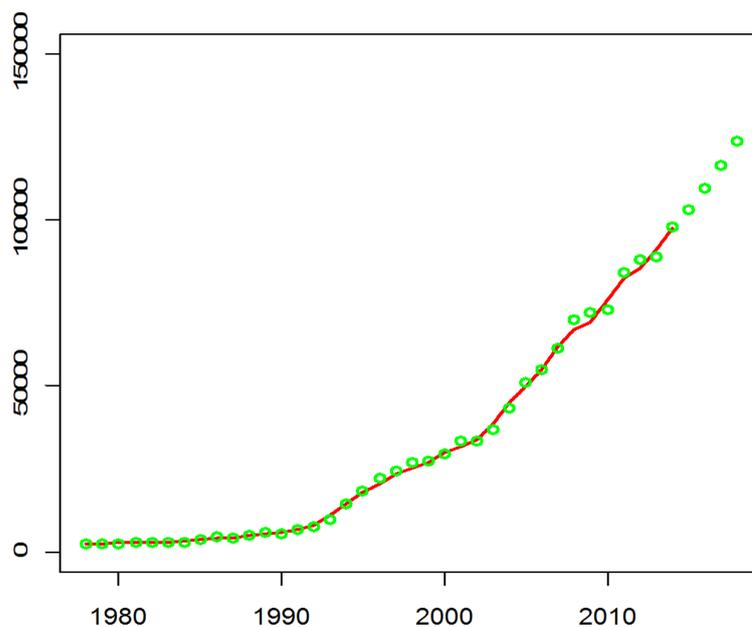


Figure 4. Prediction effect diagram
图 4. 预测效果图

点。但是上海以其优越的地理位置和竞争优势，仍然可以有较大的发展空间，因此找出更有利于区域经济发展的方法至关重要。首先上海的土地开发还有很大的发展空间特别是浦东地区南汇部分还有较大的开发潜力，可以由内到外，有限的扩充繁华经济圈；其次在第三产业快速发展的同时，也要加快带动第一、二产业的发展，形成协调性发展；最后上海的自贸区改革红利以及上海建设国际科技中心等也为上海的未来发展提供了强大的发展动力。

6. 结束语

本文根据 1978~2014 年上海市人均 GDP 数据建立了 ARIMA(2,2,2)模型，经模型参数估计与诊断检验，证明模型拟合结果有效。通过真实值与预测值对比的检验结果发现，预测相对误差较小，预测精度较高，可用于对上海市人均 GDP 做短期预测，为上海市制定经济计划提供参考和依据。

参考文献 (References)

- [1] Andrei, E.-A. and Bugudui, E. (2011) Econometric Modeling of GDP Time Series. *Theoretical & Applied Economics*, 2011, 91-98.
- [2] 何新易. 基于时间序列模型的中国 GDP 增长预测分析[J]. 财经理论与实践, 2012, 33(4): 96-99.
- [3] 张丙见. 我国 GDP 时间序列的模型建立与实证分析[J]. 商, 2013(23): 93-96.
- [4] 郭景威, 李宏斌. 中国人均 GDP 时间序列的实证分析与预测[J]. 经济论坛, 2012(3): 9-12.
- [5] 华鹏, 赵学民. ARIMA 模型在广东省 GDP 预测中的应用[J]. 统计与决策, 2010(12): 166-167.
- [6] 熊志斌. 基于 ARIMA 与神经网络集成的 GDP 时间序列预测研究[J]. 数理统计与管理, 2011, 30(2): 306-314.
- [7] 龙会典, 严广乐. 基于 SARIMA, GM(1,1)和 BP 神经网络集成模型的 GDP 时间序列预测研究[J]. 数理统计与管理, 2013, 32(5): 814-822.
- [8] 汪丽艳, 叶明海. 基于 RBF 神经网络的 GDP 时间序列预测模型研究[J]. 经济论坛, 2013(3):10-13.
- [9] 梅沁, 刘宴先, 景小南. 基于多因素 VAR 分析的上海 GDP 预测[J]. 湖北工业大学学报, 2011, 26(3): 28-31.
- [10] 中华人民共和国国家统计局网[EB/OL]. <http://www.stats.gov.cn/>

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>