

# Empirical Study on the Impact of Urbanization Rate on Economic Growth in China Based on Smooth Spline Estimation

Xirui Liu

School of Science, Northern University of Technology, Beijing  
Email: 527047039@qq.com

Received: Jun. 7<sup>th</sup>, 2019; accepted: Jun. 21<sup>st</sup>, 2019; published: Jun. 28<sup>th</sup>, 2019

---

## Abstract

In the real world, the relationship between most variables is not linear. When using time series model to explore the relationship between urbanization rate and economic growth in China, the cointegration test fails, showing that there is no linear cointegration relationship between them. Therefore, this paper uses smooth spline estimation to study the relationship between urbanization rate and economic growth in China, and compares it with the classical linear regression method on the actual data set. Through empirical analysis, it is found that the fitting effect of smooth spline is better than that of linear regression model.

## Keywords

Urbanization Rate, Economic Growth, Smooth Spline, Linear Regression

---

# 基于光滑样条估计的我国城镇化率对经济增长影响的实证研究

刘曦蕊

北方工业大学理学院, 北京  
Email: 527047039@qq.com

收稿日期: 2019年6月7日; 录用日期: 2019年6月21日; 发布日期: 2019年6月28日

---

## 摘要

现实世界中, 大多数变量之间的关系都不是线性的。本文在使用时间序列模型对我国城镇化率与经济增长

长之间的关系进行探究时,协整检验失效,说明二者之间不存在线性协整的关系。因此,本文采用光滑样条估计法来研究我国城镇化率与经济增长之间的关系,并在实际数据集上与经典线性回归方法比较。通过实证分析,得知光滑样条拟合效果优于线性回归模型。

## 关键词

城镇化率, 经济增长, 光滑样条, 线性回归

Copyright © 2019 by author(s) 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

## 1. 引言

我们人类社会发展趋势的一个必然表现就是城镇化,城镇化是一个国家或地区现代化重要性的标志,也是工业化的重要性标志。我国城镇化正以惊人的速度快速推进,与此同时,我国的经济也以世界瞩目的姿态蓬勃发展。城镇化、经济增长是两个世界各国争相关注的话题,亦是学者研究的热门话题。Henderson [1]曾尝试使用生产函数和不同国家不同时段的面板数据考察城市化与经济增长的关系,所得出的结论是,推动经济增长的并不是城市化本身,而是城市化的结构。沈坤荣、蒋锐[2]利用 Cobb-Douglas 生产函数,建立计量模型,并利用我国的相关数据进行了实证检验,结果表明:我国城镇化的不断发展能够通过影响物质资本、人力资本、知识资本和产业结构四种方式,进而促进和影响我国经济的不断增长。闫晓红[3]利用城镇化水平内生增长模型对我国 27 个省份 20 年的面板数据进行分析,表明除几个经济较发达的省区外,城镇化水平的提高对经济增长均有很大的推动作用。

因此,本文选取了这二者作为研究对象,试图挖掘出它们之间的关系。然而,本文最初在使用时间序列模型分析时,这二者之间的协整检验无法通过,即认为这二者不存在线性协整关系。因此促使我考虑使用非线性模型来对这两者进行建模、分析。

非参数模型相较于经典的线性模型来说更为灵活,不需要对模型施加过多的假设。光滑样条是一种非常流行的非参数回归方法,它不需要明确指定变量之间的函数关系类型,充分利用数据本身去选择合适的回归函数来拟合数据的变化趋势,能很好的解决变量之间的非线性建模问题。陈生长[4]比较了光滑样条回归、Monotonic 回归、最小二乘回归三种方法,在一个记录了糖尿病患者血糖和胰岛素测定值的数据集上进行拟合,从拟合误差角度,说明了光滑样条的拟合效果优于其他两种回归方法。本文选择光滑样条估计来研究我国城镇化率与经济增长之间的关系。

## 2. 自然三次样条模型

考虑在区间 $[a, b]$ 上选择  $n$  个结点  $t_i (i=1, \dots, n)$ , 满足  $a < t_1 < t_2 < \dots < t_n < b$ , 如果函数  $f(t)$  满足: 1) 在每个区间  $(t_i, t_{i+1}), i=1, \dots, n-1$ , 均是三次多项式; 2) 在每一个结点上,  $f(t)$  的一阶与二阶导数存在且连续。则这样的函数  $f(x)$  被称为三次光滑样条。显然, 三次样条函数可表示为:

$$f(t) = g(t) = d_i(t-t_i)^3 + c_i(t-t_i)^2 + b_i(t-t_i) + a_i$$

$$t_i \leq t \leq t_{i+1}, i = 0, 1, \dots, n \quad (1)$$

如果三次样条函数  $g(t)$  在  $a, b$  两点的二阶、三阶导数为 0, 则函数  $g(t)$  被称做自然三次样条。

设有  $n$  对数据  $(t_i, y_i), i=1, \dots, n$ , 其中  $a < t_1 < t_2 < \dots < t_n < b$ , 则可对数据  $(t_i, y_i)$  拟合模型:

$$y_i = g(t_i) + \varepsilon_i \tag{2}$$

假设  $g(t)$  是满足以上条件的三次自然光滑样条, 则令  $g = (g_1, \dots, g_n)'$ ,  $\gamma = (\gamma_1, \dots, \gamma_n)'$ , 其中  $g_i = g(t_i)$ ,  $\gamma_i = g''(t_i)$ ,  $i=1, 2, \dots, n$ , 由自然边界条件可知  $q, \gamma_1 = \gamma_n = 0$ . 令  $Q$  为  $n \times (n-2)$  维矩阵, 其元素  $q_{ij}$ ,  $i=1, 2, \dots, n; j=2, \dots, n-1$ .  $h_i = t_{i+1} - t_i, i=1, \dots, n-1$ , 则  $q_{j-1, j} = h_{j-1}^{-1}; q_{j, j} = -h_{j-1}^{-1} - h_j^{-1}; q_{i, j} = 0$ , 若  $|i-j| \geq 2$ . 又令  $R$  维  $(n-2) \times (n-2)$  维对称矩阵, 由元素  $r_{ij}$  构成,  $i=j=1, \dots, n-1$ .  $r_{i, i} = \frac{1}{3}(h_{i+1} - h_i)$ ,  $i=1, \dots, n-1; r_{i, i+1} = r_{i+1, i} = \frac{1}{6}h_i, i=2, \dots, n-2; r_{i, j} = 0$ , 若  $|i-j| \geq 2$ . 因此  $R$  是严格的正定矩阵, 定义矩阵  $K: K = QR^{-1}Q'$ . 当且仅当  $Q^T g = R\gamma$  时,  $\int_a^b \{g''(x)\}^2 dx = \gamma' R \gamma = g' K g$ . 满足上述条件时  $g$  和  $\gamma$  可以完全确定一个自然三次样条函数.

可以通过最小化函数:

$$S(g) = \sum_{i=1}^n (Y_i - g(t_i))^2 + \int_a^b \{g''(x)\}^2 dx \tag{3}$$

来拟合得到样条模型. 其中  $g''(x)$  是函数  $g(t)$  的二阶导数,  $\int_a^b \{g''(x)\}^2 dx$  被用来刻画函数的光滑程度;  $\lambda > 0$  是光滑参数,  $\lambda$  的值越大, 函数  $g$  越光滑. 以下用向量来表示求解过程:

$$\sum \{Y_i - g(t_i)\}^2 = (Y - g)'(Y - g)$$

$$\begin{aligned} S(g) &= (Y - g)'(Y - g) + \lambda g' K g \\ &= g'(I + \lambda K)g - 2Y'g + Y'Y \end{aligned}$$

求  $S(g)$  对  $g$  的偏导数,

$$\frac{\partial S(g)}{\partial g} = 2(I + \lambda k)g - 2Y$$

令  $\frac{\partial S(g)}{\partial g} = 0$ , 可得  $g = (I + \lambda k)^{-1}Y$ .  $\gamma$  的值可用  $\gamma = R^{-1}K g$  求得. 在给定光滑参数  $\lambda$  的值后, 三次光滑样条可用以上向量形式表示.

如果  $\lambda$  选的过大, 对模型的光滑度的惩罚力度较大, 会使得曲线过于光滑, 导致过拟合; 如果  $\lambda$  选的太小, 则对曲线的弯曲程度惩罚较小, 会导致曲线出现过拟合.  $\lambda$  可通过交叉验证法 CV、广义交叉验证法 GCV 等方式选出一个最优的  $\lambda$ . 本文选用广义交叉验证(GCV)选择光滑参数  $\lambda$ ,

$$GCV(\alpha) = n^{-1} \frac{\sum_{i=1}^n \{y_i - \hat{g}(t_i)\}^2}{\{1 - n^{-1}trA(\alpha)\}^2}$$

其中  $A(\alpha) = I - \alpha Q(R + \alpha Q'Q)^{-1}Q'$ .

### 3. 实证分析

本节在对我国城镇化水平与经济增长的关系进行研究时, 选取了中国统计年鉴 1978~2017 共 40 年的数据. 选取“城镇人口在总人口所占比重”这一指标分析我国城镇化水平, 记为城镇化率  $X$ . 因为人均国内生产总值将人口规模的影响剔除掉了, 相对来说比较准确所以选取人均国内生产总值作为衡量经济增长的指标, 记做  $Y$ . 为了消除价格因素的影响, 本节所采用的数据是以 1978 年为基期, 经过计算得到

的可比价人均国内生产总值。

### 3.1. 协整检验

对非平稳序列进行建模时，往往会出现“伪回归”现象，为了满足数据的平稳性，我们通常利用差分变换，使序列变平稳，但是这样建模的效果就受到了影响。我们通常可以通过协整来解决此问题，变量的协整检验最常用的方法是 EG 检验。

两变量检验的第二步，再对残差序列  $e_t$  的平稳性进行检验。由 EVIEWS 软件进行 ADF 检验，结果如表 1。

Table 1. The results of ADF test

表 1. ADF 检验结果

变量	ADF 统计量	临界值(1%)	临界值(5%)	临界值(10%)	P 值	平稳性
$e_t$	-2.2368	-3.6268	-2.9458	-2.6115	0.1974	不平稳

通过检验结果发现， $x$  与  $\ln y$  的线性组合所得到的残差序列  $e_t$  并不是平稳的，见表 1，说明  $x$  与  $\ln y$  之间没有线性协整关系。造成此种现象的原因可能是，随着时间的推移，城镇化率对人均 GDP 的影响关系并不是线性的，而是每年都在变化的，因此，本文将使用光滑样条估计对数据进行分析。

### 3.2. 光滑样条估计

根据第二节中介绍的三次光滑样条的理论知识，利用 R 软件对模型进行估计。由于已有的大部分文献都是采用线性回归模型对其进行分析及预测，因此为了和线性回归模型进行比较，本文也对响应变量与预测变量进行了线性回归。图 1 是光滑样条与线性回归模型拟合的图像，较粗的直线是线性回归模型拟合的直线，较细的曲线是光滑样条拟合的图形，从图中可以看出光滑样条模型能够更好的拟合数据。

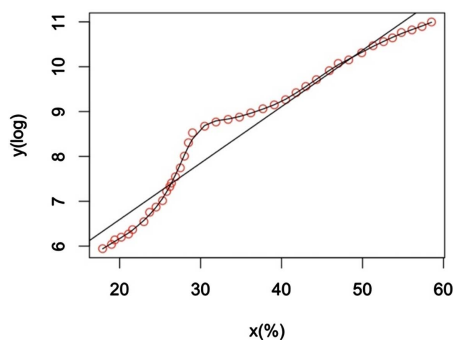


Figure 1. Fit of smooth Spline and linear regression

图 1. 光滑样条与线性回归拟合

根据图 2 光滑样条估计的残差散点图可以发现，所有点都落在了  $-0.2 \sim 0.2$  的水平带状区间中，并且未表现出明显的趋势性，因此我们可以认为该模型具有良好的拟合效果。

因此，可以得出我国城镇化率与人均 GDP 的光滑样条模型：

$$\hat{y} = \exp\left\{\widehat{g(x)}\right\}$$

在构建的线性回归模型中，将做过对数变换后的人均 GDP 作为响应变量城镇化率作为解释变量。利用 R 软件进行计算，并把结果汇总在表 2 中。

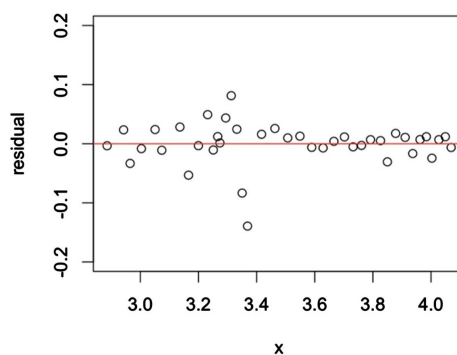


Figure 2. Residual plot of smooth spline

图 2. 光滑样条估计的残差散点图

Table 2. Results of linear regression

表 2. 线性回归的结果

	Estimate	Std. Error	t value	Pr (> t )	显著性
截距	4.075836	0.175769	23.19	<2e-16	***
x	0.125853	0.004669	26.96	<2e-16	***

我们可以得到预测方程：

$$\hat{y} = \exp(4.075836 + 0.125853x)$$

为了分析模型的估计结果，本节把预测的均方误差作为评价标准：

$$MSE = \frac{1}{n} \sum_{i=1}^n \left( y_i - \widehat{g}(t_i) \right)^2$$

通过表 3 可以看出，光滑样条估计的均方误差要比线性回归模型估计的均方误差小，说明光滑样条更好的拟合了数据。因此我们可以知道光滑样条模型的预测效果要优于线性模型。

Table 3. Root MSE predicted by two types of methods

表 3. 两类方法预测的均方误差

	光滑样条估计	最小二乘估计
均方误差	0.00121	0.12619

#### 4. 主要结论

从上一节的分析中，我们得到了光滑样条估计，并通过对比线性模型与光滑样条模型预测的均方误差得出光滑样条模型比线性模型好。同时，分析的结果告诉我们，城镇化率并不是以一种线性的方式影响着人均 GDP 的变化，而是以一种非线性的方式影响着经济增长的步伐。

光滑样条非参数回归分析不必假定反应变量与解释变量之间的函数形式，适用于任意分布的数据，因此光滑样条回归模型可广泛应用于经济学、医学等领域。

#### 基金项目

北京市属高校基本科研业务费 NO.110052971921/103 资助项目。

---

## 参考文献

- [1] Herderson, J.V. (2003) The Urbanization Process and Economic Growth: The So-What Question. *Journal of Economical Growth*, **8**, 47-71.
- [2] 沈坤荣, 蒋锐. 中国城市化对经济增长影响机制的实证研究[J]. 统计研究, 2007(6): 9-15.
- [3] 闫晓红. 城市化水平与长期经济增长关系的实证分析——基于 27 个省份面板数据的考察[J]. 经济经纬, 2011(3): 25-28.
- [4] 陈生长, 徐勇勇, 夏结来. 光滑样条非参数回归方法及医学应用[J]. 中国卫生统计, 1999(6): 342-345.

### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

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

期刊邮箱: [sa@hanspub.org](mailto:sa@hanspub.org)