

文章编号: 1004-4353(2021)04-0329-05

基于非参数 Panel data 模型的居民边际消费 区域差异性分析—以山西省为例

张慧芳¹, 张德生²

(1. 忻州师范学院 数学系, 山西 忻州 034000; 2. 西安理工大学 理学院, 西安 710054)

摘要: 基于非参数局部线性估计方法, 建立了一种固定效应非参数 Panel data 模型, 并利用 2001—2019 年山西省 11 个地市的农村居民消费与收入数据进行了实证检验. 结果显示: 非参数 Panel data 模型的拟合误差显著小于参数 Panel data 模型的拟合误差, 同时非参数模型变化的边际消费比固定的参数模型的边际消费更为合理. 该研究结果可为政府相关部门制定不同区域的消费政策提供参考.

关键词: Panel data 模型; 边际消费; 可支配收入; 核函数

中图分类号: F318; O212.7

文献标识码: A

Analysis on regional differences of residents' marginal to consume based on nonparametric Panel data models — Taking Shanxi Province for example

ZHANG Huifang¹, ZHANG Desheng²

(1. Department of Mathematics, Xinzhou Teachers University, Xinzhou 034000, China;
2. School of Science, Xi'an University of Technology, Xi'an 710054, China)

Abstract: Based on the nonparametric local linear estimation method, Panel data fixed effect nonparametric models for rural residents' consumption and net income are established respectively, and an empirical test is carried out with data in 11 cities of Shanxi Province from 2001 to 2019. The results show that the fitting error of nonparametric Panel data models are significantly smaller than that of parametric Panel data models. Meanwhile, the dynamical fluctuation of the marginal consumption tendency for the nonparametric models is more reasonable than the constant marginal consumption tendency for the parametric models. It can provide some reference for relevant government departments to formulate consumption policies in different regions.

Keywords: Panel data model; tendency consumption; disposable income; kernel function

0 引言

目前,我国经济增长已由高速增长阶段进入中高速增长阶段. 消费作为经济增长的基础,其对经济的拉动作用十分明显,但由于我国居民消费增长相对较为缓慢,特别是农村地区的居民消费,因此研究如何提高居民消费能力具有重要意义. 针对消费问题,目前大多数学者是采用协整理论、ELES 模型、误差修正模型、线性回归分析等方法进行实证探索. 例如:张玉周^[1]和汪旭辉等^[2]利用协整理论及误差修

收稿日期: 2021-09-17

基金项目: 国家自然科学基金青年基金(11801438);山西省高等学校教学改革创新项目(J2021795)

第一作者: 张慧芳(1982—),女,硕士,讲师,研究方向为应用概率统计.

正模型对我国数年的消费与收入数据进行了实证分析;王攀娜^[3]和黄飞娜^[4]利用线性回归模型研究了我国城镇居民收入与消费间的关系;邱慧等^[5]利用灰色理论对山西省主要消费支出进行了灰色关联性分析.1992年,Mátyás等^[6]提出了面板数据(Panel data)模型,由于该模型能有效地利用已知数据信息区别个体间的差异,因此受到学者的关注.2014年,田萍等^[7]利用 Panel data 模型对我国居民消费与经济增长的区域差异进行了定量分析.2016年,武新乾等^[8]利用 Panel data 模型分析了我国城镇居民收入与消费的区域差异性.在上述研究中,研究者均使用了对模型限定条件较严的参数估计方法,因此容易导致模型的假设与实际不符,即出现估计偏差较大的现象.由于非参数模型无需设定具体的模型形式,能有效避免模型的误设问题,因此本文利用非参数估计方法建立了一种非参数固定效应 Panel data 模型,并利用该模型对山西省 11 个地市的农村居民收入与消费的差异进行了实证研究.

1 非参数 Panel data 模型的建立

1.1 数据预处理

在《山西省统计年鉴》中选取 2001—2019 年山西省 11 个地市的农村居民人均消费支出和可支配收入数据,并且将这 11 个地市按北、中、南部分为 3 组,划分方式如表 1 所示.为有效避免价格等因素的影响,以及防止经济数据的异方差干扰,在分析面板数据之前对所选定的数据进行了对数化预处理.

表 1 区域划分情况

地区	覆盖城市
北部地区	大同 朔州 忻州
中部地区	吕梁 太原 晋中 阳泉
南部地区	临汾 长治 晋城 运城

1.2 非参数 Panel data 模型的建立

Panel data 模型包含固定效应模型和随机效应模型.由于本文将山西省 11 个地市的数据分为北、中、南部 3 个地区,其个体影响只限于相应的横截面,满足固定效应模型仅以样本自身效应为条件进行推理的特点,因此本文选用非参数单因素固定效应 Panel data 模型.该模型的表达式为:

$$y_{it} = \alpha_i + f(x_{it}) + \varepsilon_{it} \quad (i = 1, 2, \dots, N; t = 1, 2, \dots, T). \quad (1)$$

其中: y_{it} 是被解释变量,为对数化预处理后的消费支出; x_{it} 是解释变量,为对数化预处理后的可支配收入; $f(x)$ 为未知光滑函数; ε_{it} 是随机干扰项,满足 $E(\varepsilon_{it} | x_{it}) = 0$, $E(\varepsilon_{it}^2 | x_{it}) = \sigma_\varepsilon^2$; N 是某地区的地市数; T 是各地市的数据个数.

1.2.1 非参数估计

当 α_i 是固定效应时,式(1)中的未知函数 $f(x)$ 虽不能被识别,但该函数的导数 $f'(x)$ 可被识别.在某固定点 x 附近对函数 $f(x_{it})$ 进行泰勒展开可得:

$$y_{it} = \alpha_i + f(x) + f'(x)(x_{it} - x) + \gamma_{it}, \quad (2)$$

其中 γ_{it} 是泰勒公式余项及 ε_{it} 的和.

由式(2)可得 $\frac{1}{T} \sum_{t=1}^T y_{it} = \frac{1}{T} \sum_{t=1}^T [\alpha_i + f(x) + f'(x)(x_{it} - x) + \gamma_{it}]$, 即

$$y_{i.} = \alpha_i + f(x) + f'(x)(x_{i.} - x) + \gamma_{i.}. \quad (3)$$

由式(2)和式(3)得:

$$y_{it} - y_{i.} = f'(x)(x_{it} - x_{i.}) + \mu_{it}, \quad (4)$$

其中 $\mu_{it} = \gamma_{it} - \gamma_{i.}$. 由式(4)可得:

$$\sum_{i=1}^N \sum_{t=1}^T \mu_{it} = \sum_{i=1}^N \sum_{t=1}^T (y_{it} - y_{i.}) - \sum_{i=1}^N \sum_{t=1}^T f'(x)(x_{it} - x_{i.}). \quad (5)$$

最小化式(5)的加权残差平方 $\sum_{i=1}^N \sum_{t=1}^T [(y_{it} - y_{i.}) - f'(x)(x_{it} - x_{i.})]^2 K(\frac{x_{it} - x}{h})$ 可得 $f'(x)$ 的估计量,为:

$$\hat{f}'(x) = \frac{\sum_{i=1}^N \sum_{t=1}^T K\left(\frac{x_{it} - x}{h}\right)(y_{it} - y_{i\cdot})(x_{it} - x_{i\cdot})}{\sum_{i=1}^N \sum_{t=1}^T K\left(\frac{x_{it} - x}{h}\right)(x_{it} - x_{i\cdot})^2}, \quad (6)$$

其中 $K(\cdot)$ 是核函数, h 是窗宽, $x_{i\cdot} = \frac{1}{T} \sum_{t=1}^T x_{it}$, $y_{i\cdot} = \frac{1}{T} \sum_{t=1}^T y_{it}$.

1.2.2 窗宽选择

在非参数局部线性估计^[9]中,窗宽 h 选取过大会造成回归函数的估计过分光滑,窗宽选取过小会使回归函数的估计很不光滑,因此选取合适的窗宽对非参数估计至关重要.在此,本文采用交叉验证方法^[10]来选择最佳的窗宽,其具体方法为:首先对于各局部的观测点 $X = X_i$,从样本中去掉该观测点 (X_i, Y_i) ;其次在 $X = X_i$ 处将剩余的 $n - 1$ 个观测点做局部核权估计,然后比较平方拟合误差 $cv(h) = n^{-1} \sum_{i=1}^n (Y_i - \hat{f}_{n,-i}(X))^2 W(X)$,其中 $\hat{f}_{n,-i}(\cdot)$ 是去掉第 i 个观测点后回归函数 $f(\cdot)$ 的非参数估计;当拟合误差达到最小时,选择此时的窗宽作为最佳窗宽.

为了获得最佳窗宽,利用 MATLAB7.0 软件对式 $cv(h) = n^{-1} \sum_{i=1}^n (Y_i - \hat{f}_{n,-i}(X))^2 W(X)$ 进行编程,并在程序中分别代入北、中、南部地区的相应收入、消费数据.由此得到的北、中、南部地区的非参数 Panel data 模型的 CV 曲线见图 1.由图可以看出,最佳窗宽分别为: $h_{北} = 0.3517 (cv_{北} = 7.638 \times 10^{-4})$, $h_{中} = 0.2135 (cv_{中} = 6.631 \times 10^{-4})$, $h_{南} = 0.1735 (cv_{南} = 7.645 \times 10^{-4})$.

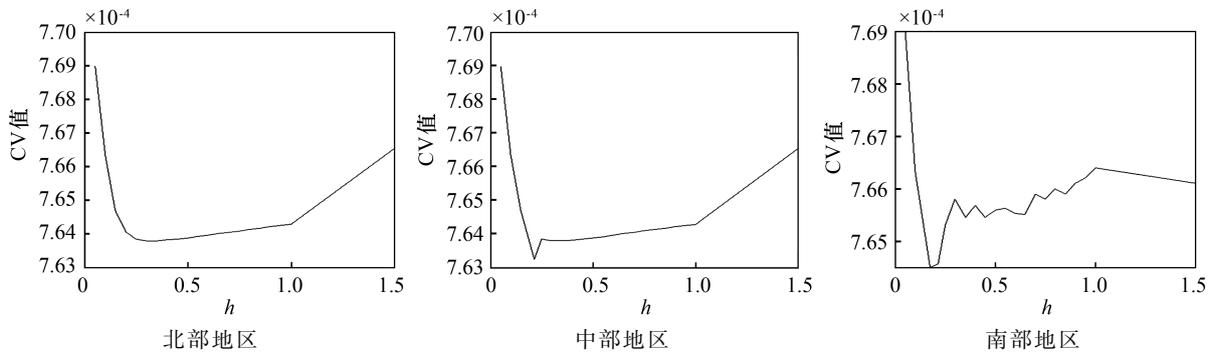


图 1 CV 曲线图

1.3 模型的比较

为了验证非参数模型的有效性,首先用 EVIEWS10.0 软件分别建立北、中、南部地区的 Panel data 参数模型,具体如下:

1) 北部地区农村居民的人均消费与可支配收入的 Panel data 固定效应参数模型为:

$$LN\hat{C}A_{it} = -1.526164 + \alpha_i^* + 1.141857LYNA_{it}, \quad i = 1, 2, 3; \quad t = 1, 2, \dots, 19. \quad (7)$$

$(t = -7.138885) \quad (t = 44.95064)$

模型(7)中: α_i^* 为各地市消费差异估计结果,1.141857 为边际消费倾向,-1.526164 为北部地区农村居民的平均自发消费水平, $LN\hat{C}A_{it}$ 为对人均消费支出进行数字化预处理后的估计, $LYNA_{it}$ 为对数字化预处理后的人均可支配收入, i 分别对应北部的 3 个市区; t 分别对应 2001、2002、2003、...、2019 年.

2) 类似于公式(7),中部地区相应的固定效应参数模型为:

$$LN\hat{C}A_{it} = -0.473745 + \phi_i^* + 1.016867LYNA_{it}, \quad i = 1, 2, 3, 4; \quad t = 1, 2, \dots, 19. \quad (8)$$

3) 南部地区相应的固定效应参数模型为:

$$LN\hat{C}A_{it} = -1.958582 + \gamma_i^* + 1.182883LYNA_{it}, \quad i = 1, 2, 3, 4; \quad t = 1, 2, \dots, 19. \quad (9)$$

由模型(7)~(9)可得出山西省北、中、南部各地区农村居民平均自发消费水平、边际消费倾向以及

各地市对相应地区自发消费水平的偏离程度,见表 2.

表 2 北、中、南部地区的边际消费倾向、平均自发消费水平及偏离程度

维度	北部地区		中部地区		南部地区	
平均自发消费水平	-1.526 164		-0.473 745		-1.958 582	
边际消费倾向	1.141 857		1.016 867		1.182 883	
各地市的 偏离程度	忻州	0.274 662	太原	-0.050 498	临汾	-0.075 381
	朔州	-0.208 928	吕梁	0.174 536	运城	0.068 668
	大同	-0.065 735	阳泉	-0.028 356	长治	0.000 942
			晋中	-0.095 681	晋城	0.005 770

为比较本文建立的参数 Panel data 模型与非参数 Panel data 模型的估计效果,选择绝对误差 MAE 和均方根误差 RMSE 作为评价指标进行评价. MAE 和 RMSE 的计算公式为:

$$MAE = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T |y_{it} - \hat{y}_{it}|, \quad RMSE = \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \hat{y}_{it})^2 \right)^{\frac{1}{2}},$$

其中 \hat{y}_{it} 是 y_{it} 的估计. 由 MAE 和 RMSE 的计算公式得到的相应误差结果见表 3.

表 3 固定效应下参数与非参数模型的 MAE 和 RMSE 值

指标	北部地区		中部地区		南部地区	
	参数模型	非参数模型	参数模型	非参数模型	参数模型	非参数模型
MAE	0.183 3	0.033 5	0.115 5	0.027 2	0.201 5	0.031 3
RMSE	0.230 1	0.050 9	0.135 0	0.029 0	0.210 1	0.030 5

由表 3 可以得出,北、中、南部地区的非参数模型的 MAE 分别占参数模型相应 MAE 的 18.2%、23.5%、15.5%,北、中、南部地区的非参数模型的 RMSE 分别占参数模型相应 RMSE 的 22.1%、21.4%、14.5%. 由此可知,非参数 Panel data 模型的拟合效果优于参数 Panel data 模型,所以选择非参数 Panel data 模型对居民消费问题进行研究更为合理.

2 边际消费的变化特征及分析

边际消费 $f'(x)$ 不但能体现消费支出占单位可支配收入的比重,还能体现不同时期消费支出受可支配收入影响的程度. 利用 MATLAB7.0 软件对式(6)进行计算得到山西省北、中、南部各地区不同时间的边际消费后,对各地区的边际消费倾向进行了拟合,结果见图 2.

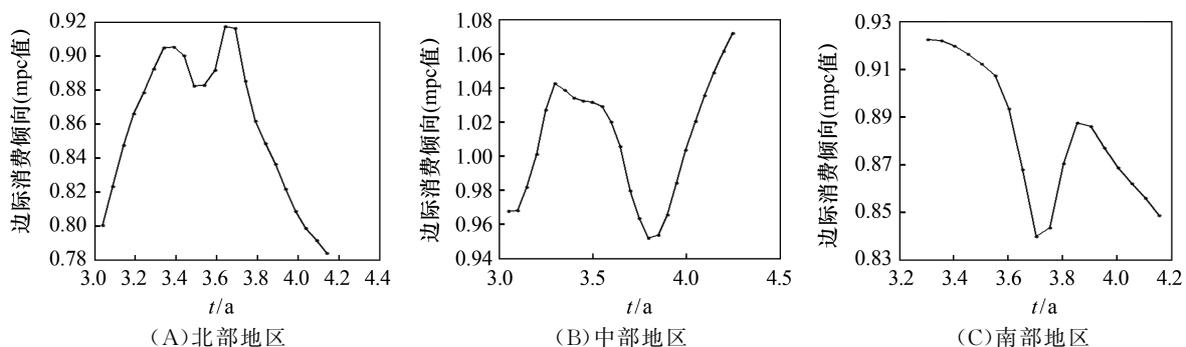


图 2 固定效应非参数 Panel data 模型下的边际消费倾向拟合图

由图 2(A)可以看出,北部地区的居民边际消费倾向出现两个波动周期. 其中:2001—2010 年为第 1 个波动周期,该周期初始时边际消费呈持续快速增长趋势,当边际消费达到 0.905 1 后,出现短暂下降趋势;2011—2019 年为第 2 个波动周期,该周期初始时边际消费呈现短暂增长趋势,但随后呈持续下降

的趋势.由图 2(B)可以看出,中部地区的农村居民边际消费倾向可划分为 3 个阶段.其中:2001—2005 年为第 1 个阶段,该阶段边际消费呈持续上升趋势;2006—2013 年为第 2 个阶段,该阶段边际消费出现了先缓慢下降后快速下降趋势;2014—2019 年为第 3 个阶段,该阶段边际消费在达到低谷期后出现了持续增加趋势.由图 2(C)可以看出,南部地区的农村居民边际消费倾向可划分为 3 阶段:2001—2007 年为第 1 个阶段,该阶段边际消费呈持续下降趋势;2008—2012 年为第 2 个阶段,该阶段边际消费出现了一定幅度的回升;2013—2019 年为第 3 个阶段,该阶段边际消费再次出现了持续下降趋势,但相对于第 1 阶段来说其下降速度有所减缓.

3 结论

本文基于非参数估计理论,在固定效应下根据 2001—2019 年山西省北部、中部、南部地区的农村居民人均消费支出和可支配收入数据建立了非参数 Panel data 模型.实证分析结果显示:①非参数 Panel data 模型比传统的参数 Panel data 模型能够更准确地反映山西省农村居民人均消费支出和可支配收入间的关系;②山西省北、中、南部地区的农村居民表现出不同的边际消费倾向变化,其主要原因是北、中、南部地区的经济发展水平不同,进而导致农村居民主要收入不同所致.本文研究结果可为政府相关部门制定不同区域的消费政策提供参考.本文建立的非参数 Panel data 模型是利用局部线性估计方法对未知函数进行估计的,计算量较大,因此在后续的研究过程中我们将优化未知函数的估计方法,以给出更为高效的估计方法.

参考文献:

- [1] 张玉周.我国城乡居民收入与消费关系的比较分析[J].统计与决策,2013(22):95-98.
- [2] 汪旭辉,顾晶.中国农村居民消费与收入关系的实证分析[J].北京工商大学学报:社会科学版,2009,24(1):58-63.
- [3] 王攀娜,徐博韬.城镇居民消费与可支配收入关系的实证分析[J].会计之友,2013(5):55-60.
- [4] 黄飞娜.中国居民消费与经济增长的关系:基于 1978—2005 年数据的实证研究[J].中国集体经济,2011(3):19-20.
- [5] 邱慧,黄解宇,李珏.山西省农村居民收入与消费结构的灰色关联分析[J].数学的实践与认识,2018(12):283-286.
- [6] MÁTYÁS L, SEVESTRE P. The econometrics of Panel data: handbook of theory and applications[J]. The Economic Journal, 1994,104(422):177-179.
- [7] 田萍,武新乾,梅倩倩,等.我国城乡居民消费与经济增长区域差异性的实证分析[J].数学的实践与认识,2014,44(22):89-95.
- [8] 武新乾,司福宁,田萍,等.基于非参数面板数据模型的我国城镇居民消费与收入关系的区域差异性实证分析[J].数理统计与管理,2016,35(4):587-593.
- [9] HARDLE W, TSYBAKOV A. Local polynomial estimators on the volatility function in nonparametric autoregression[J]. Journal of Econometric, 1997(81):223-242.
- [10] 徐晔,黄文.互联网技术对中国宏观经济的动态效应及传导机制研究:基于 DSGE 模型的实证分析[J].江西师范大学学报(自然科学版),2020,44(4):358-365.