



# 非线性时变因子模型下中国区域价格水平的俱乐部收敛检验

陈钊钊, 叶 露

(浙江理工大学经济管理学院, 杭州 310018)

**摘 要:** 区域价格水平收敛性检验对商品市场是否存在区域分割具有重要研究意义。为研究中国区域价格水平收敛情况, 首先基于非线性时变因子模型对中国 1999—2017 年的居民消费价格指数进行  $\log t$  回归和俱乐部收敛检验, 发现全国范围内不存在 CPI 趋同, 中国商品市场存在区域分割, 且形成了 3 个趋同俱乐部和 1 个发散组。然后, 利用各俱乐部的空间分布和 Ordered Probit 模型对俱乐部形成机制进行探索, 发现趋同俱乐部的形成与地理位置有关, 距离越近的省市, 越容易出现 CPI 趋同, 商品房平均销售价格和国际旅游外汇收入对俱乐部的转移影响显著。最后, 对比最长距离法的聚类结果可知, 基于  $\log t$  回归的聚类结果是稳健的。

**关键词:** 区域价格水平; 居民消费价格指数; 非线性时变因子模型;  $\log t$  回归; 俱乐部收敛; 聚类

中图分类号: F127

文献标志码: A

文章编号: 1673-3851 (2020) 08-0351-10

## Club convergence of China's regional price level based on a nonlinear time-varying factor model

CHEN Chuanchuan, YE Lu

(School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** Convergence test of regional price level is of great research significance to the research of regional segmentation in the commodity market. To study the convergence of regional price level in China,  $\log t$  regression and club convergence test are carried out for China's consumer price index (CPI) from 1999 to 2017 based on a nonlinear time-varying factor model firstly. The results show that the CPI is not convergent nationwide, China's commodity market has been segmented by regions, and 3 convergence clubs and 1 divergence group have been formed in China. Secondly, the club formation mechanism is explored based on the space distribution of each club and Ordered Probit model, and it is found that the formation of convergence club is related to geographical location. This means CPI convergence is more likely to occur for close regions, and the average sales price of commercial housing and foreign exchange earnings from international tourism have a significant effect on the transfer of the clubs. Last, comparison of the clustering results of the longest distance method indicates that the clustering result based on  $\log t$  regression is robust.

**Key words:** regional price level; consumer price index (CPI); non-linear time-varying factor model;  $\log t$  regression; club convergence; clustering

近年来, 区域价格水平收敛性的研究已成为经济统计领域的研究热点。收敛的概念源于数学, 经

济分析中的收敛一般是指地区间或国家间的收入随时间推移存在趋同的趋势。随着收敛理论的快速发

收稿日期: 2019-11-18 网络出版日期: 2020-05-27

基金项目: 国家自然科学基金项目(71901195); 浙江理工大学基本科研业务费专项项目(2019Q091)

作者简介: 陈钊钊(1997—), 女, 浙江绍兴人, 本科生, 主要从事经济统计方面的研究。

通信作者: 叶 露, E-mail: zjwzajyl@126.com

展,目前有以下几种收敛性概念受到关注: $\sigma$ 收敛,指不同地区之间收入或消费的标准差随时间推移而逐渐下降<sup>[1]</sup>;绝对 $\beta$ 收敛,指经济增长率只与发展初始水平相关,随着时间的推移,经济增长率趋同<sup>[2]</sup>;条件 $\beta$ 收敛,指允许各个经济体之间存在异质性的条件下,每个地区都收敛于各自的稳定状态<sup>[3]</sup>;俱乐部收敛,指结构特征相似且初始收入水平相同区域的人均收入在长期发展中趋同<sup>[4]</sup>。

价格水平收敛性的深入研究对区域市场一体化的检验有重要意义。苏桔芳等<sup>[5]</sup>指出,价格水平的收敛性不仅影响政府与人民生活水平相关的政策,且与资源配置的有效性息息相关;若各个区域实行单一货币,那么不收敛的价格水平将意味着区域商品市场存在分割。王磊等<sup>[6]</sup>指出中国地区间相对价格水平的收敛性是反映中国区域市场是否存在分割状况的重要指标。Hoeberichts等<sup>[7]</sup>通过检验1960—2015年奥地利、比利时等11个经济与货币联盟成员国家的价格水平分散度来确定欧洲一体化的进程。本文试图利用中国31个省市居民消费价格指数来分析区域价格水平的收敛性,进而研究中国区域商品市场是否存在分割现象。

中国区域价格水平由于受多个因素的影响,其结果是趋于收敛还是发散备受争议<sup>[8-10]</sup>。由于中国实行单一货币制度,且改革开放后中国加快了经济一体化进程,生产要素流动相对自由,中国区域价格水平趋于收敛;但又由于中国幅员广阔,区域物质条件和经济发展存在差异,价格水平可能趋于发散。对于判断价格水平的收敛性,国内外学者大多采用聚类分析<sup>[11]</sup>、面板数据回归<sup>[12-13]</sup>、时间序列方法<sup>[14-16]</sup>和空间计量模型<sup>[17-21]</sup>等方法。但是大多数研究方法都是在参数同质的假设下,由于中国各省市在环境体制等方面有差异,同质性假定显然存在不足。Phillips等<sup>[22]</sup>提出了非线性时变因子模型,该模型考虑了可能的时间路径和个体异质性,克服了之前研究方法同质性假设的缺点。因此本文将运用非线性时变因子模型对中国1999—2017年居民消费价格指数进行 $\log t$ 回归,研究中国区域价格水平是否存在全局性收敛,进而检验俱乐部收敛状态及影响因素,分析各俱乐部的相对转移曲线。最后,为检验结果的稳健性,对比了基于 $\log t$ 回归的聚类方法与最长距离法的聚类方法得到的结果。

## 一、文献综述

通过居民消费价格指数(Consumer price

index,CPI)的统计数据讨论区域相对价格水平收敛的方法很多。如Sonora<sup>[14]</sup>收集了1982—2000年墨西哥34个城市的CPI,使用面板单位根检验的方法发现区域价格趋同。Bilgili<sup>[15]</sup>采用单位根检验探讨了土耳其18个城市的相对价格趋同问题,发现18项消费价格指数中有13项趋于一致。杨茜<sup>[21]</sup>利用空间计量模型对中国各省市2006—2013年的平均CPI进行空间分布研究,发现CPI越接近的省市更容易收敛。但是,以上研究方法存在一些缺陷:聚类分析的类别需要研究者自行设定,可能缺乏科学性;面板数据回归对识别个体效应有局限性;时间序列方法(单位根检验、协整分析)缺乏与增长理论的内在联系;空间计量模型主要用于研究绝对趋同,而对条件趋同的验证较少<sup>[23]</sup>。

Phillips等<sup>[22]</sup>提出了非线性时变因子模型来表示转型期经济的行为,该模型考虑了可能的时间路径和个体异质性,它既有共同的部分,又有各自的具体组成部分。非线性时变因子模型的核心内容包括两部分: $\log t$ 回归和划分趋同俱乐部的聚类算法。国内外学者应用非线性时变因子模型对环境、房价、股票等区域进行了大量研究。Fritsche等<sup>[24]</sup>利用非线性时变因子模型对欧盟15个成员国1960—2006年的人均收入和生产率进行分析,结果显示人均收入存在3个俱乐部,生产率存在2个俱乐部。Glushenkova等<sup>[25]</sup>对1990—2010年40个国家和96种商品、服务的半年度微观价格水平数据进行分析,发现零售价格存在趋同俱乐部。钞小静等<sup>[26]</sup>采用非线性时变因子模型对各地区经济增长质量的差距进行检验,发现东部呈发散趋势,中西部呈条件收敛趋势。贺祥民等<sup>[27]</sup>抓住地区间的时变特征,利用非线性时变因子模型发现地区间环境效率存在俱乐部收敛现象。安勇等<sup>[28]</sup>选取中国35个城市2005—2015年的房屋销售价格数据,发现无论是整体层面还是东中西部三大地带,其城市房价均不存在收敛性。赵胜民等<sup>[29]</sup>采用非线性时变因子模型对A股市场的股票和ETF价差进行收敛性检验,发现个股间存在俱乐部收敛。

## 二、CPI的统计特征

本文选取1999—2017年31个省市年度CPI作为研究区域价格水平收敛性的指标,数据均取自于国家统计局。CPI是在一段时间内衡量一组代表性消费商品和服务项目的价格水平随时间变动的相对数,可用来反映人民生活负担的变化情况和通货膨胀程度。

### (一) 平稳性检验

为避免伪回归,首先对CPI的平稳性进行检验。对于原序列,ADF单位根检验的 $P$ 值为0.12,因此需先对原序列进行处理。将各省市区CPI(上年为100)换算成以1998年为基期的数据,再根据Phillips等<sup>[22]</sup>的建议,将CPI取对数,最后进行HP滤波处理,剔除数据中可能存在的周期性,得到处理后的数据。

用Eviews软件对处理后的数据进行平稳性检验。首先检验31个省市区是否存在共同的单位根,由LLC检验统计量的 $P$ 值为0,因此拒绝存在共同单位根的原假设。然后检验各省市区是否存在单位根,由Fisher-ADF和Fisher-PP检验统计量的 $P$ 值均为0,因此拒绝各省市区存在单位根的原假设,即处理后的数据是平稳数据,可被用于之后的分析。

### (二) CPI 描述性分析

#### 1. 时间序列数据

处理后的时间序列数据按方差从小到大排序得到的箱线图如图1所示。从图中可以看出,各省市区的时间序列数据不存在异常值,并且中位数在0上下震荡。各省市区的方差相差较小,说明数据分布集中,分布特征相似。这也反映了数据的平稳性,说明处理后的数据可以应用于后续研究。但是,从均值上看,正负性不一致,具体数值大小也存在差异。CPI反映了消费水平,CPI越高的省市区通常来说消费支出更大,因此均值为正的省市区一般比均值为负的省市区消费支出更大。

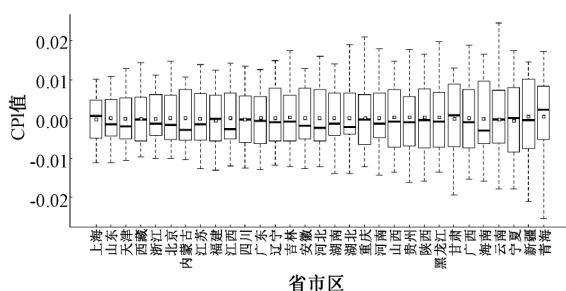


图1 各省市区CPI时间序列数据的箱线图

#### 2. 截面数据

为了分析CPI与年份之间的关系,图2描述了1999—2017年各截面数据的箱线图,可以看出各年的数据分布相差较大。相比于时间序列数据而言,各年份CPI的方差和平均值都相差很大,数据分布比较分散,结构特征差异明显。各年份的均值波动很大,反映了CPI的一个特点:稳定性较差,每年甚至每月都会产生波动。均值为正的年份相比于均值为

为负的年份来说,消费水平更高。CPI除了反映消费水平,还可以作为观察通货膨胀水平的指标。2002—2007年CPI均值一直为负值,而2008年突然转为正值,反映了2008年通货膨胀水平上升,经济出现较大的波动。各年份CPI的方差和平均值相差很大,反映了数据随时间明显变化的特征,因此下文采用非线性时变因子模型对中国31个省市区的价格水平收敛性进行研究。

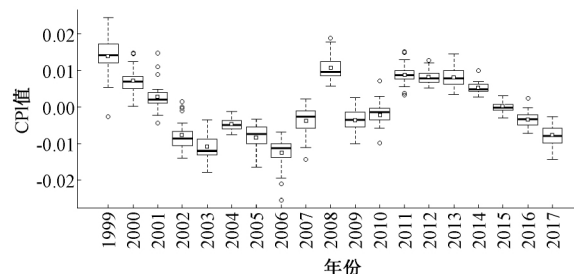


图2 1999—2017年CPI截面数据的箱线图

### 三、非线性时变因子理论模型

Phillips等<sup>[22]</sup>提出的非线性时变因子模型检验由两个核心部分组成:logt回归和聚类算法。logt回归是一种基于简单回归的收敛性检验方法,用来检验趋同的原假设。基于logt回归,Phillips等<sup>[22]</sup>构造了一种新的聚类算法来划分趋同俱乐部。非线性时变因子模型结合了聚类分析的特征,且允许各种趋向于收敛的可能转换路径,因此这种假定参数异质的时变方法更适合中国CPI收敛的研究。

#### (一) 非线性时变因子模型

各省份CPI表示为 $X_{it}$ ,其中: $i=1,2,\dots,31$ , $t=1,2,\dots,19$ ,对应年份1999,...,2017。对其建立时变单因子模型:

$$X_{it} = \delta_{it}\mu_t,$$

其中: $\mu_t$ 表示 $X_{it}$ 中的共同部分, $\delta_{it}$ 是会随时间变化的异质性部分。在本文的研究框架下, $\mu_t$ 可理解为各省市区CPI的共同趋势部分, $\delta_{it}$ 代表每一个省市区 $i$ 在时刻 $t$ 所占的份额。将 $\delta_{it}$ 分解为两部分:不随时间变化的均值和方差随时间变化的异质性误差项。 $\delta_{it}$ 可表示为半参数形式(即模型中部分结构已知,部分未知):

$$\delta_{it} = \delta_i + \sigma_i \xi_{it} L(t)^{-1} t^{-\alpha},$$

其中: $\delta_i$ 为不随时间变化的稳定值; $\xi_{it}$ 是均值为0,方差为1的白噪声过程; $\sigma_i$ 是特质性的规模参数; $L(t)$ 是设定的一个缓慢变化的函数,如 $L(t) = \log(t)$ ,一般当 $t \rightarrow \infty$ 时,则 $L(t) \rightarrow \infty$ ; $\alpha$ 为收敛速度,当 $\alpha \geq 0$ 时, $\delta_{it}$ 收敛于 $\delta_i$ 。因此可以将面板数据

的收敛性检验转化为对以下假设的检验:

$$\begin{cases} H_0: \delta_{it} = \delta_i, \text{ 且 } \alpha \geq 0 \\ H_A: \text{对所有的 } i, \delta_{it} \neq \delta_i \text{ 或者 } \alpha < 0, \end{cases}$$

原假设表示所有省市区的价格水平收敛,而备择假设表示有些省市区未收敛。为刻画各省份的收敛路径,可考虑相对转移参数:

$$h_{it} = \frac{X_{it}}{\frac{1}{N} \sum_{i=1}^N X_{it}} = \frac{\delta_{it}}{\frac{1}{N} \sum_{i=1}^N \delta_{it}},$$

其中:  $N$  表示横截面单元总数,本文中为  $N = 31$ 。若所有省市区趋同,  $\delta_{it}$  将收敛于  $\delta_i$ , 则  $h_{it} \rightarrow 1$ ,  $h_{it}$  的横截面方差趋于零。由此,用来衡量个体发散程度的  $t$  时刻截面方差  $D_t = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2$ , 当  $t \rightarrow \infty$  时,  $D_t \rightarrow 0$ , 记年份为 1999 的数据为  $D_1$ 。以上性质可用于检验收敛的原假设以及划分趋同俱乐部。

## (二) logt 回归

logt 回归是用来检验所有城市经济收敛的原假设  $H_0$  和存在发散组的备择假设  $H_A$  的一种回归方法,具体步骤如下:

a) 构建横截面方差比  $D_1/D_t$ , 衡量时刻  $t$  个体相对于初始时刻(即  $t = 1$ )的发散程度,借此来考察个体收敛程度的动态变化。

b) 进行回归:

$$\log\left(\frac{D_1}{D_t}\right) - 2\log L(t) = \hat{a} + \hat{b}\log(t) + u_t,$$

其中:  $\hat{a}$  是常数项估计值;  $\hat{b}$  为  $\log(t)$  的回归参数估计值;  $L(t)$  是缓慢变化的函数,本文设  $L(t) = \log(t+1)$ 。根据 Phillips 等<sup>[22]</sup> 基于模拟试验的建议,对  $t = 5, 6, \dots, 19$  进行回归。在回归过程中,由于随机扰动项  $u_t$  存在序列相关,因此需要计算  $\hat{b}$  的异方差自相关稳健标准误(HAC)下的  $t$  统计量  $t_{\hat{b}}$ 。

c) 使用异方差自相关稳健标准误(HAC)对  $\hat{b}$  进行单侧  $t$  检验,若  $t_{\hat{b}}$  小于  $-1.65$ , 则拒绝收敛的原假设,反之,则接受原假设。

## (三) 俱乐部收敛聚类算法

俱乐部收敛聚类算法分为四个步骤:

a) 排序。以最近时间点的数据大小进行排序,即将 31 个省市区按 2017 年的 CPI 从大到小排序,如果数据存在较大的波动,可以根据平均值来排序。

b) 建立核心组  $G_{k^*}$ 。以排序最高的省市区为基准,每次加入一个省市区进行 logt 回归,组成备选核心组,当备选核心组的  $t_{\hat{b}}$  首次小于  $-1.65$  时停止运算。假设排序最高的前  $k$  个省市区满足这一要

求,记为  $t_{\hat{b}}(k)$ , ( $2 \leq k < N$ )。按照以下准则,挑选核心组成员数  $k^*$ :

$$\begin{cases} k^* = \operatorname{argmax}\{t_{\hat{b}}(k)\} \\ \text{s.t. } \min\{t_{\hat{b}}(k)\} > -1.65, \end{cases}$$

前  $k$  个成员共组成  $(k-1)$  个备选核心组,选出最大的  $t_{\hat{b}}$  值,即  $\max\{t_{\hat{b}}(k)\}$ , 其对应的备选组就是核心组。若  $k^* = N$ , 则说明 31 个省市区的价格水平收敛。当  $k=2$  时,  $\min\{t_{\hat{b}}(k)\}$  大于  $-1.65$  不能满足,则去掉排序最高的省市区,从第二个省市区重复上述步骤,若第二个省市区也对  $k=2$  不能满足条件  $\min\{t_{\hat{b}}(k)\}$  大于  $-1.65$ , 则继续去掉这个省市区,从第三个省市区重复上述步骤。倘若接下来的省市区都不满足  $\min\{t_{\hat{b}}(k)\}$  大于  $-1.65$ , 则说明 31 个省市区的价格水平发散,不存在趋同俱乐部。按照上述步骤选出核心组  $G_{k^*}$ 。

c) 确定俱乐部成员。核心组  $G_{k^*}$  形成后,每次从核心组的补集中选择一个成员加入到核心组  $G_{k^*}$ , 然后进行 logt 回归得到相应的  $t_{\hat{b}}$ , 如果  $t_{\hat{b}}$  超过某一临界值  $c$  ( $c$  称为过滤尺度,用来衡量聚类的保守程度,  $c$  值越大,俱乐部成员越少,本文取  $c = -1.30$ ), 则把该成员加入到核心组,称为趋同俱乐部的一员。反之,则不能称为俱乐部的一员。对核心组外的每个省市区进行检验,可以得到第一个俱乐部,再对该俱乐部内的所有成员进行 logt 回归以确保  $t_{\hat{b}}$  大于  $-1.65$ 。

d) 停止规则。形成第一个俱乐部后,对剩余所有的个体进行 logt 回归,如果得到的  $t_{\hat{b}}$  大于  $-1.65$ , 则说明余下的所有个体形成了另一个趋同俱乐部,停止程序。否则,对余下的所有个体重复前三个步骤,研究该组内是否还存在俱乐部,直到无法找到趋同俱乐部为止。如果到最后,还有一些个体未被包含在任何俱乐部中,则说明这些个体是发散的。

## 四、俱乐部收敛实证检验

### (一) 俱乐部收敛实证结果

利用非线性时变因子模型,对中国 31 个省市区 1999—2017 年的年度 CPI 进行收敛性实证检验。首先验证 CPI 是否存在全国性收敛,对全国 31 个省市区进行 logt 回归,得到  $t_{\hat{b}}$  为  $-2.39$ , 小于  $-1.65$ , 在 5% 的显著性水平下拒绝趋同的原假设,说明 1999—2017 年 CPI 不存在全国性收敛。为了进一步研究样本区间内是否存在俱乐部收敛现象,本文按照 Phillips 等<sup>[22]</sup> 给出的聚类算法进行计算,检验过程及结果如表 1 所示,表中给出了各省市区回归后的  $t_{\hat{b}}$  值。

表1 基于  $\log t$  回归的聚类结果

排序	省名	俱乐部 1		俱乐部 2		俱乐部 3
		第一步	第二步	第一步	第二步	第一步
1	上海市	—	-1.74	—	-2.48	—
2	北京市	-2.24	核心	—	—	—
3	浙江省	0.26	核心	—	—	—
4	天津市	0.57	核心	—	—	—
5	广东省	-0.13	-0.13	—	—	—
6	海南省	-1.29	-1.23	—	—	—
7	江西省	-1.50	-0.62	—	—	—
8	山东省	-1.68	-1.30	-2.21	核心	—
9	江苏省	—	-1.31	0.89	核心	—
10	西藏	—	-1.91	-1.74	-1.74	-2.86
11	福建省	—	-0.54	—	—	—
12	辽宁省	—	-1.79	—	-2.69	-3.33
13	吉林省	—	-1.25	—	—	—
14	湖北省	—	-1.29	—	—	—
15	重庆市	—	-1.00	—	—	—
16	湖南省	—	-1.42	—	-0.46	—
17	河北省	—	-1.86	—	-3.05	-2.10
18	安徽省	—	-1.67	—	-1.15	—
19	广西	—	-1.46	—	-0.68	—
20	内蒙古	—	-2.30	—	-2.47	-0.15
21	四川省	—	-1.81	—	-1.11	—
22	河南省	—	-1.47	—	-2.43	-0.33
23	贵州省	—	-0.53	—	—	—
24	黑龙江省	—	-2.39	—	-4.25	0.05
25	陕西省	—	-2.95	—	-5.03	0.22
26	新疆	—	-3.42	—	-2.40	-0.17
27	云南省	—	-0.64	—	—	—
28	山西省	—	-1.98	—	-2.20	-0.15
29	宁夏	—	-2.51	—	-1.80	-0.32
30	甘肃省	—	-2.53	—	-3.34	-0.54
31	青海省	—	-1.82	—	-2.66	-0.88
俱乐部		-1.62		-0.63		-0.88
剩余		-2.96		-2.94		—

首先对 31 个省市按 2017 年 CPI 从大到小排序。从排序情况可以看到,上海、北京、浙江、天津等东部地区省市排名比较靠前,而宁夏、甘肃、青海等西部地区省市排名较靠后。得到最后一期的排名后,对各省市的 CPI 进行  $\log t$  回归并寻找核心组。由于上海排名最高,因此以上海为基准。将排名第二的北京与上海进行  $\log t$  回归,得到的  $t_b$  为 -2.24,小于 -1.65,拒绝原假设,所以去掉上海,以北京为基准。将北京和浙江进行  $\log t$  回归,  $t_b$  为 0.26,大于 -1.65,接受原假设。继续加入天津,  $t_b$  为 0.57,以此类推,加入广东、海南、江西。加入山东时,  $\log t$  回归的  $t_b$  为 -1.68,小于 -1.65,核心组

算法停止。选出最大的  $t_b(3) = 0.57$ ,对应的省市为天津市,因此第一个核心组由北京、浙江、天津组成。

核心组形成后,进行俱乐部成员的筛选。将上海加入核心组,  $\log t$  回归的  $t_b$  为 -1.74,小于过滤尺度( $c = -1.30$ ),因此不能将上海加入核心组。类似地,将广东放入核心组进行  $\log t$  回归,得到的  $t_b$  大于 -1.30,说明广东可以加入核心组。以此类推,每次加一个省市到核心组进行  $\log t$  回归,若回归得到的  $t_b$  大于过滤尺度,则该省市可以加入核心组,反之,则不能加入。继续利用  $c = -1.30$  这个门槛值,可以将海南、江西、福建、吉林、湖北、重庆、贵州、云南加入核心组,由此得到北京、浙江、天津、广东、海南、江西、福建、吉林、湖北、重庆、贵州、云南为第一趋同俱乐部。对第一俱乐部内的所有省市进行回归,  $t_b = -1.62$ ,大于 -1.65,接受原假设。对第一俱乐部外的 19 个省市进行  $\log t$  回归,得到  $t_b$  为 -2.96,拒绝收敛的原假设,因此继续寻找该组内是否存在趋同俱乐部。

对剩余的 19 个省市重复上述步骤。检验结果显示核心组由山东和江苏组成。利用过滤尺度  $c = -1.30$ ,将湖南、安徽、广西、四川加入该核心组,因此第二俱乐部由山东、江苏、湖南、安徽、广西、四川组成。对第二俱乐部内的所有省市进行回归,  $t_b = -0.63$ ,大于 -1.65。除去第一、第二俱乐部的省市,对剩余 13 个省市进行  $\log t$  回归,  $t_b$  为 -2.94,因此对剩余的 13 个省市重复上述步骤。在第三轮中,河北、内蒙古、河南、黑龙江、陕西、新疆、山西、宁夏、甘肃、青海形成第三俱乐部,该俱乐部的回归  $t_b$  值为 -0.88,大于 -1.65,上海、西藏、辽宁未进入任何趋同俱乐部。

由俱乐部收敛结果可知,中国 1999—2017 年 CPI 存在 3 个趋同俱乐部和 1 个发散组,说明目前中国区域商品市场存在分割。表 2 整理了各俱乐部收敛的聚类结果,在第一俱乐部以北京市、天津市、浙江省为核心组,包含了 12 个省市。第二俱乐部以山东和江苏为核心组,包含了 6 个省市。发散组由上海、西藏和辽宁组成。图 3 描绘了每年第一、第二、第三俱乐部 CPI 均值的变化,从 2013 年开始,各俱乐部 CPI 均值呈下降趋势。2015—2017 连续三年,第一俱乐部的 CPI 均值大于第二俱乐部,第二俱乐部的 CPI 均值大于第三俱乐部。为了探讨俱乐部趋同的影响因素,下文将考虑从地理位置角度展开研究。

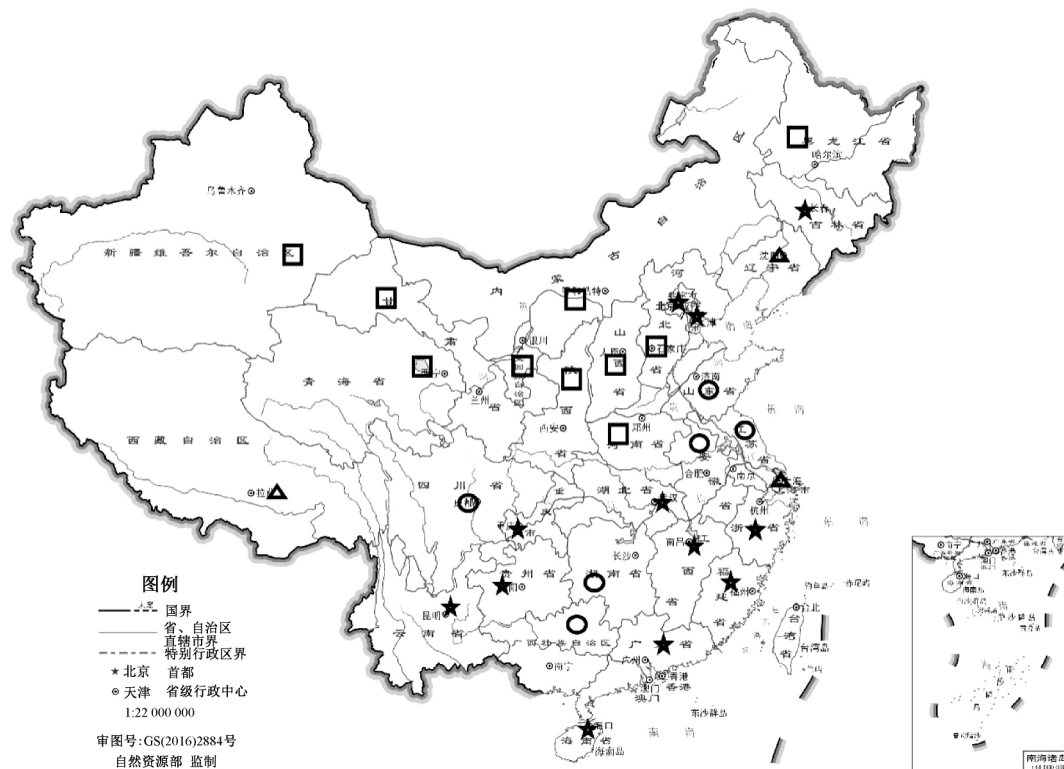
表2 俱乐部收敛结果

第一俱 乐部	第二俱 乐部	第三俱 乐部	发散 组
北京市	山东省	河北省	上海市
浙江省	江苏省	河南省	辽宁省
天津市	湖南省	陕西省	西藏自治区
广东省	安徽省	山西省	—
海南省	四川省	甘肃省	—
江西省	广西壮族自治区	青海省	—
福建省	—	黑龙江省	—
吉林省	—	内蒙古自治区	—
湖北省	—	宁夏回族自治区	—
重庆市	—	新疆维吾尔自治区	—
贵州省	—	—	—
云南省	—	—	—

### (二) 俱乐部的空间分布

Lee<sup>[30]</sup>指出观察到的价格差异很大程度上取决于地理位置,地理位置越近的省市区,价格水平越相近,CPI也更容易趋同。为了直观验证这一观点是否适用中国价格水平情况,将聚类结果在地图上进行描述。

图4展示了中国大陆各俱乐部的空间分布,其中五角星代表第一俱乐部,圆圈代表第二俱乐部,矩形代表第三俱乐部,三角形代表发散的3个省市区。

图4 中国大陆各俱乐部空间分布图<sup>①</sup>

### (三) 俱乐部形成机制

对俱乐部形成机制的探索,本文选用 Ordered Probit 模型对影响因素进行建模。参考王磊等<sup>[6]</sup>,选取了商品房平均销售价格(元/平方米)、境内目的

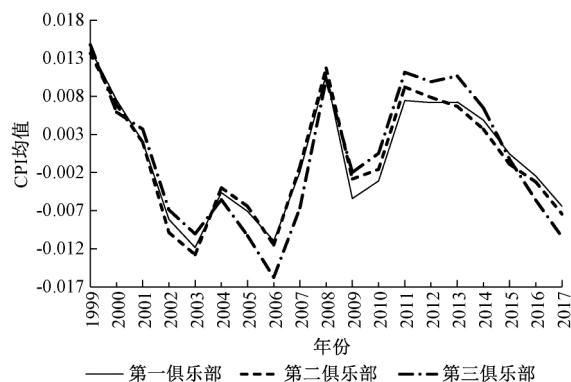


图3 各俱乐部 CPI 均值的时间序列图

第三俱乐部出现了大面积的聚集,这些省市区都集中在中国的北部。南部大部分省市区为第一俱乐部和第二俱乐部,且第一俱乐部比第二俱乐部的聚集省份更多。从总体来看,各俱乐部的空间分布呈现出地理位置越近的省市区 CPI 更容易趋同的趋势。例如,北部大部分为第三俱乐部,但北京和天津却属于第一俱乐部;第二俱乐部虽然未出现大面积的聚集,但相邻的广西和湖南,安徽、江苏和山东都聚在了一起。因此,各俱乐部的空间分布在一定程度上说明了地理位置越靠近的省市区越容易出现 CPI 趋同。

地和货源地进出口总额(千美元)、国际旅游外汇收入(百万美元)。以上数据均来自中国统计局。为了

① 该图基于国家测绘地理信息局标准地图服务网站下载的审图号为 GS(2016)2884 号的标准地图制作,底图无修改。

消除量纲的影响,各数值均作了标准化处理。由计算得,2017 年第一俱乐部 CPI 均值大于第二俱乐部,第二俱乐部大于第三俱乐部。对落入第一俱乐部的省份赋值为 3,落入第二俱乐部的省份赋值为 2,落入第三俱乐部的省份赋值为 1。表 3 给出了 Ordered Probit 模型中各影响因素的参数估计值,以及各俱乐部的边际效应估计结果。商品房平均销售价格和国际旅游外汇收入对俱乐部的转移显著为正,即商品房平均销售价格增加一个单位,则该省份

所属俱乐部向 CPI 均值高的俱乐部转移概率就越大。同样,国际旅游外汇收入增加一个单位,则该省份所属俱乐部向 CPI 均值高的俱乐部转移概率越大。从边际效应可以看出,商品房平均销售价格增加,则在第一俱乐部的可能性显著增加,但在第三俱乐部的可能性显著降低。境内目的地和货源地进出口总额增加,则在第一俱乐部的可能性显著降低。国际旅游外汇收入增加,则在第一俱乐部的可能性显著增加。

表 3 Ordered Probit 模型的估计结果

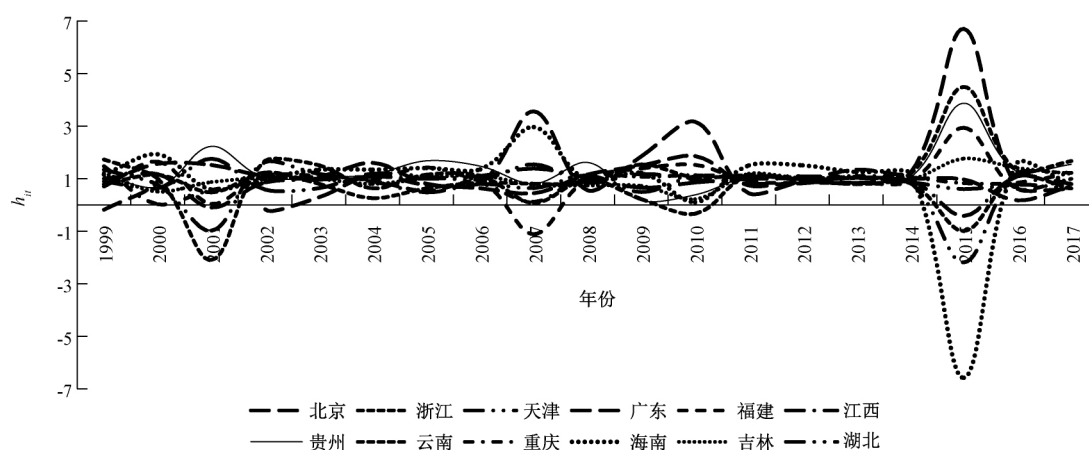
变量	参数估计	边际效应		
		第一俱乐部	第二俱乐部	第三俱乐部
商品房平均销售价格	2.12*	0.76**	-0.38	-0.38**
	(0.08)	(0.03)	(0.20)	(0.04)
境内目的地和货源地进出口总额	-0.90	-0.32*	0.16	0.16
	(0.11)	(0.08)	(0.22)	(0.12)
国际旅游外汇收入	1.81*	0.65*	-0.33	-0.32
	(0.09)	(0.07)	(0.20)	(0.14)

注:圆括号内数值为  $P$  值,\*、\*\*分别表示 10%和 5%的显著性水平。

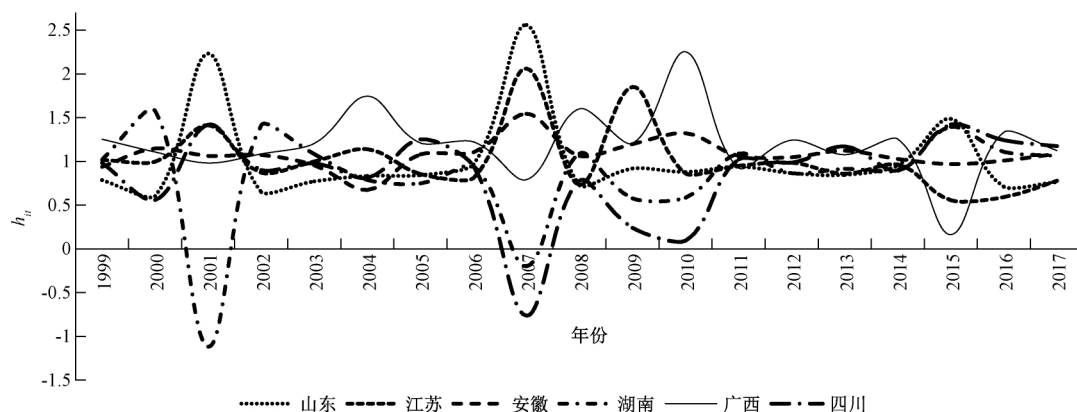
#### (四)各俱乐部的异质性转移动态

由于各省市区间可能存在区域异质性,虽然各趋同俱乐部的省市区最终趋于收敛,但 CPI 在收敛

过程中可能会有不一样的路径。为了探究各俱乐部内的成员是否存在异质性转移动态,图 5 展示了各俱乐部成员的相对转移曲线,即  $h_{it}$  的轨迹。



(a) 第一俱乐部



(b) 第二俱乐部

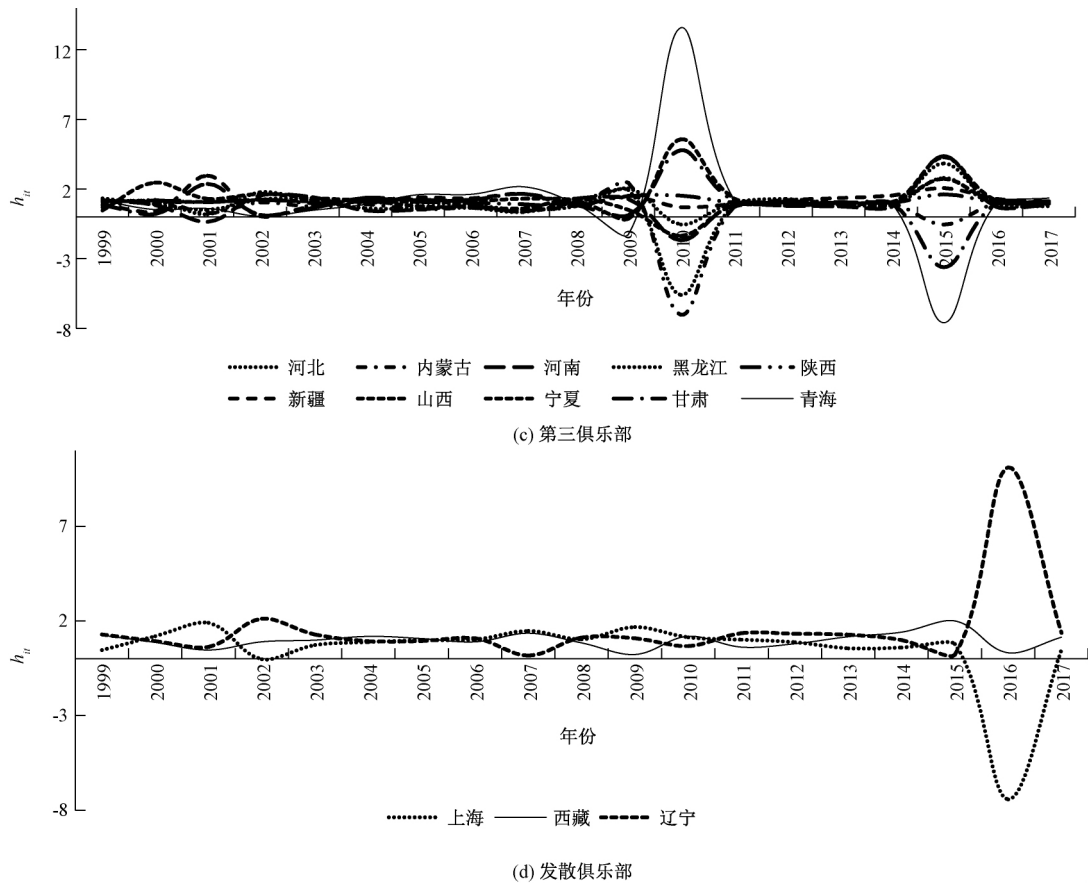


图5 各俱乐部的相对转移曲线

从图 5(a) 第一俱乐部成员的转移路径可看出, 虽然最终这 12 个省市趋于收敛, 但路径却是截然不同的。如北京的起点要明显低于其他 11 个省市, 之后在 2001、2007、2010 年和 2015 年出现了大幅度的波动, 虽然北京的转移路径与其他省市截然不同, 但所有省市在 2016 年趋于收敛; 不止北京出现了大幅度波动, 其他 11 个省市也均表现出不同程度的波动, 尤其是在 2015 年, 北京和广东出现了两极分化的现象, 但在 2016 年, 所有省市都从 2015 年的大波动突然趋于收敛。此外, 第一俱乐部的成员除了在 2001、2007、2010 年和 2015 年出现大波动外, 其余年份都呈现收敛的状态。

从图 5(b) 可见, 与第一俱乐部相比, 第二俱乐部成员的相对转移路径变化更大。除了在之前提到的四个年份出现大波动外, 第二俱乐部的 6 个省市在其他年份均有波动。总体来说, 在 2011 年之前, 各省市的路径变化很大, 尤其是广西一直处于上下起伏的状态; 2011 年后各省市趋于收敛, 虽然在 2015 年出现波动, 但 2016 年又趋于收敛。将安徽和湖南进行对比, 安徽一直处于比较平缓的状态, 而湖南波动很大, 两者之间的转移路径截然不

同, 但最终在 2017 年两者的轨迹线几乎重合。

从图 5(c) 可见, 第三俱乐部各省市的转移路径十分相似, 要么同时出现波动, 要么处于高度收敛的状态。在 2010 年之前各省市都呈现收敛状态, 2010 年出现巨大波动, 2011 年又迅速收敛, 2015 年出现波动, 又在 2016 年高度收敛。

从图 5(d) 可见, 发散俱乐部的转移路径与收敛俱乐部完全不同。收敛俱乐部的成员无论转移路径波动有多大, 在 2016 年都会出现收敛趋势, 而发散俱乐部刚好相反。在 2015 年之前, 3 个省市的转移路径相差很小, 呈现收敛的趋势, 但是 2015 年辽宁和上海突然出现大的波动, 到 2017 年为止, 3 个省市最终呈现发散的状态。

非线性时变因子模型得到广泛应用的原因在于, 该模型可处理平稳数据, 也可处理有单位根非平稳的数据, 允许面板数据中的个体存在随时间变动的异质性。从以上收敛俱乐部和发散俱乐部的转移路径可看出, 无论是哪种俱乐部, 成员间都存在异质性转移动态, 显示了各省市间存在区域异质性。此外, 收敛俱乐部的成员可以展示出完全不一样的转移路径, 但在靠近样本结束期时是趋于收敛的。



## 五、logt 回归聚类与最长距离法聚类的比较

为检验俱乐部收敛结果的稳健性,本文将对基于 logt 回归的聚类方法与最长距离的聚类方法得到的结果,并分析两种方法各自的优缺点。表 4 是利用最长距离的聚类方法对 31 个省市区 CPI 进行聚类的结果,为了便于与俱乐部收敛结果作比较,表 4 呈现了四类结果。

表 4 最长距离法的聚类结果

第一类	第二类	第三类	第四类
天津市	山西省	甘肃省	北京市
辽宁省	海南省	青海省	上海市
江苏省	云南省	宁夏回族自治区	西藏自治区
安徽省	贵州省	新疆维吾尔自治区	—
江西省	陕西省	—	—
河南省	黑龙江省	—	—
湖南省	内蒙古自治区	—	—
重庆市	广西壮族自治区	—	—
河北省	—	—	—
吉林省	—	—	—
浙江省	—	—	—
福建省	—	—	—
山东省	—	—	—
湖北省	—	—	—
广东省	—	—	—
四川省	—	—	—

将表 2 中的第一俱乐部与表 4 的第一类进行比较,发现浙江省、天津市、广东省、江西省、福建省、吉林省、湖北省、重庆市这 8 个省市区是重合的,类似地,将第二俱乐部与第二类、第三俱乐部与第三类、发散组与第四类进行比较,重合的省市区整理成表 5。

表 5 俱乐部收敛和最长距离法聚类结果相同的省市区

第一俱乐部 VS 第一类	第二俱乐部 VS 第二类	第三俱乐部 VS 第三类	发散组 VS 第四类
浙江省	广西壮族自治区	甘肃省	上海市
天津市	—	青海省	西藏自治区
广东省	—	宁夏回族自治区	—
江西省	—	新疆维吾尔自治区	—
福建省	—	—	—
吉林省	—	—	—
湖北省	—	—	—
重庆市	—	—	—

通过比较发现,在全国 31 个省市区中,共有浙江、天津、广东、江西、重庆、福建、湖北、吉林、广西、甘肃、青海、宁夏、新疆、上海和西藏 15 个省市区出现重合,说明用两种聚类方法得到的结果有很大的相似性,这表明本文基于 logt 回归的聚类方法得到

的结果是稳健的。虽然用两种聚类方法得到的结果相似度很高,但基于 logt 回归的聚类算法更有优势。首先,传统统计学上的最长距离聚类分析需要研究者事先设定分为几类,但在设定类别时是不知道结果的,需要不停地试直到出现想要的结果。其次,传统统计学的聚类方法有多种距离计算方式,具体选择哪种方式是未知的,需要将结果都做出来再选择自己想要的。总的来说,最长距离的聚类方法由于主观选择性太多导致结果的科学性不高,而基于 logt 回归的聚类算法则更客观,虽然过滤尺度可以自己选择,但它的约束条件很多,所以最终的结果差异很小。

## 六、结 语

从中国 31 个省市区 1999—2017 年年度 CPI 数据的描述性统计结果来看,中国各省市区间存在区域商品价格异质性,且时变性特征较为显著,故本文采用 Phillips 等<sup>[22]</sup>提出的非线性时变因子模型对中国 CPI 进行俱乐部收敛研究。研究发现:全国范围内存在 3 个趋同俱乐部和 1 个发散组,且各趋同俱乐部成员间存在异质性转移动态。

首先,通过 logt 回归发现 CPI 不存在全国性收敛,由俱乐部收敛聚类算法可以得出,中国 CPI 存在 3 个趋同俱乐部和 1 个发散组,说明目前中国区域商品市场存在分割。发散组由上海、西藏、辽宁 3 个省市区组成,由于地理、历史、开放程度等多方面原因,使得上海经济增长迅速,而西藏、辽宁经济增长相对落后,导致这 3 个省市区被排除在俱乐部外。从俱乐部收敛结果的空间分布图可以看出中国北部出现大面积趋同,南部相邻的省市区特别容易聚在一起,说明俱乐部的趋同与地理位置有关,距离越靠近的省市区,越容易出现 CPI 趋同。

其次,由各趋同俱乐部成员的相对转移曲线看出收敛路径截然不同,说明各省市区间存在异质性。第一俱乐部的成员在 2001、2007、2010 年和 2015 年出现大波动,其余年份都趋于收敛。第二俱乐部成员的路径差异很大,各省市区一直处于上下波动的状态。相比之下,第三俱乐部成员的转移路径要相似很多,要么同时出现波动,要么处于高度收敛的状态。无论收敛俱乐部成员的转移路径波动有多大,在靠近样本结束期时都趋于收敛,而发散俱乐部刚好相反。在 2015 年之前,发散俱乐部的成员呈现收敛趋势,但在靠近样本结束期时突然出现大波动,呈现发散状态。

最后,为检验俱乐部收敛的稳健性,对比了基于  $\log t$  回归的聚类方法与最长距离的聚类方法得到的结果。两者得到的聚类结果高度重合,相比之下,由于最长距离的聚类算法有更多主观选择性,基于  $\log t$  回归的聚类算法更客观,得出的结果更具有科学性和可信度。

本文的创新点及贡献主要有如下三点:第一,利用非线性时变因子模型对中国居民消费价格指数进行实证分析,得出全国范围内存在 3 个趋同俱乐部和 1 个发散组。第二,利用各俱乐部的空间分布和 Ordered Probit 模型得出俱乐部形成的影响因素,趋同俱乐部的形成与地理位置有关,且商品房平均销售价格和国际旅游外汇收入对俱乐部的转移影响显著。第三,比较了基于  $\log t$  回归的聚类方法与最长距离法的聚类方法得到的结果,发现基于  $\log t$  回归的聚类算法更客观。

#### 参考文献:

- [1] Sala-i-Martin X. The classical approach to convergence analysis[J]. The Economic Journal, 1996, 106(437): 1019-1036.
- [2] Barro R J, Sala-i-Martin X. Convergence[J]. Journal of Political Economy, 1992, 100(2):223-251.
- [3] Mankiw N G, Romer D, Weil D N. A contribution to the empirics of economic growth[J]. The Quarterly Journal of Economics, 1992, 107(2):407-437.
- [4] Baumol W J. Productivity growth, convergence, and welfare: What the long-run data show [J]. The American Economic Review, 1986, 76(5):1072-1085.
- [5] 苏栎芳,蔡经汉,王世杰.中国区域相对价格水平收敛性实证研究[J].华东交通大学学报,2010,27(3):118-124.
- [6] 王磊,王康.地区特征与中国地区间相对价格水平收敛性[J].山西财经大学学报,2016,38(8):12-24.
- [7] Hoeberichts M, Stokman A. Why price-level dispersion went up in Europe after the financial crisis [J]. The World Economy, 2018, 41(3):913-925.
- [8] 杨宝臣,张亮.我国东部地区第三产业发展收敛性分析[J].河北工程大学学报(社会科学版),2007,24(3):9-12.
- [9] 鄂永健.中国地区间价格水平差距趋于收敛还是发散:基于省级面板数据的单位根检验[J].经济评论,2007,147(5):113-117.
- [10] 王磊.国内地区间价格水平差异收敛性再检验[J].统计与信息论坛,2011,26(2):59-64.
- [11] Mamatzakis E, Matousek R, Vu A N. What is the impact of problem loans on Japanese bank productivity growth? [J]. Financial Markets, Institutions & Instruments, 2019, 28(2):213-240.
- [12] Bhat F, Yasmin E. Economic growth and convergence across Indian states [J]. Indian Journal of Economics and Development, 2018, 14(3):456-463.
- [13] 刘方.中国区域消费价格水平差异研究:基于面板门槛模型的分析[J].经济经纬,2013(2):7-12.
- [14] Sonora R J. City CPI convergence in Mexico [J]. Review of Development Economics, 2005, 9(3):359-367.
- [15] Bilgili F. City price convergence in Turkey with structural breaks [J]. International Journal of Economics and Financial Issues, 2016, 6(3):933-941.
- [16] 王建林.“二元经济”的瓶颈与和谐社会战略的协整检验[J].统计与决策,2010(11):75-78.
- [17] Lv K, Yu A, Gong S, et al. Impacts of educational factors on economic growth in regions of China: A spatial econometric approach [J]. Technological and Economic Development of Economy, 2017, 23(6): 827-847.
- [18] Chatterjee T. Spatial convergence and growth in Indian agriculture: 1967-2010 [J]. Journal of Quantitative Economics, 2017, 15(1):121-149.
- [19] 吴玉鸣.中国省域经济增长趋同的空间计量经济分析[J].数量经济技术经济研究,2006,23(12):101-108.
- [20] 钟海燕.四川经济增长的空间计量模型分析[J].西南民族大学学报(人文社科版),2018,39(9):118-121.
- [21] 杨茜.我国省域居民消费价格指数影响因素的空间计量分析[D].杭州:浙江财经大学,2016:20-30.
- [22] Phillips P C B, Sul D. Transition modeling and econometric convergence tests[J]. Econometrica, 2007, 75(6): 1771-1855.
- [23] 邹薇,周浩.经济趋同的计量分析与收入分布动态学研究[J].世界经济,2007(6):81-96.
- [24] Fritsche U, Kuzin V. Analysing convergence in Europe using the non-linear single factor model[J]. Empirical Economics, 2011, 41(2):343-369.
- [25] Glushenkova M, Kourtellis A, Zachariadis M. Barriers to price convergence [J]. Journal of Applied Econometrics, 2018, 33(7):1081-1097.
- [26] 钞小静,任保平,许璐.中国经济增长质量的地区差异研究:基于半参数个体时间异质模型的检验[J].江西财经大学学报,2016,103(1):10-20.
- [27] 贺祥民,赖永剑.基于非线性时变因子模型的地区环境效率俱乐部收敛分析[J].软科学,2017,31(3):103-106.
- [28] 安勇,王拉梯.中国城市房价收敛性及其驱动因素[J].经济问题探索,2015(12):45-50.
- [29] 赵胜民,闫红蕾.A股市场统计套利风险实证分析[J].管理科学,2015,28(5):93-105.
- [30] Lee I. On the determinants of price dispersions in Asia [J]. Korea Trade Review, 2014, 39(2):139-151.

(责任编辑:陈丽琼)