

# 基于Lasso的商业保险影响因素分析

孙维一, 赵明清

山东科技大学数学与系统科学学院, 山东 青岛

收稿日期: 2021年11月13日; 录用日期: 2021年12月8日; 发布日期: 2021年12月15日

## 摘要

随着我国保险业的迅速发展, 商业保险在我国国民经济和居民生活中的作用日益提高, 对商业保险的研究也愈加重要。本文利用Lasso回归和固定效应模型考察了商业保险的影响因素, 研究发现: 物价指数、人均可支配收入、人均消费支出、进出口贸易总额、老龄化程度对商业保险的影响不显著, 而少儿化程度在10%水平上对商业保险有着一定的负向作用, 在校学生数在5%的水平上对商业保险有着一定的正向作用, 国内生产总值、就业人数和医疗卫生人数均在1%的水平上对商业保险有着一定的正向作用。最后, 对商业保险的发展提出了若干可行性建议。

## 关键词

商业保险, Lasso回归, 固定效应模型, 影响因素

# Analysis of Influencing Factors of Commercial Insurance Based on Lasso

Weiye Sun, Mingqing Zhao

School of Mathematics and System Science, Shandong University of Science and Technology, Qingdao Shandong

Received: Nov. 13<sup>th</sup>, 2021; accepted: Dec. 8<sup>th</sup>, 2021; published: Dec. 15<sup>th</sup>, 2021

## Abstract

With the rapid development of China's insurance industry, commercial insurance plays an increasingly important role in China's national economy and residents' lives, and the research on commercial insurance is becoming more and more important. In this paper, Lasso regression and fixed effects model are used to investigate the influencing factors of commercial insurance. It is found that price index, per capita disposable income, per capita consumption expenditure, total import and export volume and aging level have no significant impact on commercial insurance, while the level of youth at 10% has a certain negative effect on commercial insurance, and the level of EDU

development at 5% has a certain positive effect on commercial insurance. The economic development level, labor force level and medical and MED level of 1% have a certain positive effect on commercial insurance. Finally, some feasible suggestions on the development of commercial insurance are put forward.

## Keywords

Commercial Insurance, Lasso Regression, Fixed Effect Model, Factors Affecting

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,我国深化金融市场改革,社会经济快速发展。其中,保险业作为现代金融市场的重要支柱之一,发展势头迅猛,政府对保险业的监管也日渐完善,出多项方案确保保险市场平稳运行。目前,我国保险市场规模先后超过德国、法国、英国和日本,成为全球第二大保险市场,为全球最重要的新兴保险市场大国。随着保险业的不断发展,越来越多的问题开始浮出水面,探索影响我国保险发展的因素就显得更为紧迫和重要。

## 2. 文献综述

世界各国的保险业都以高于经济增长的速度在发展,其背后的驱动因素一直是理论界研究的一个重要问题。在国外, Skipper (1999)认为经济增长是影响保险需求最重要的因素[1]; Meglena (2000)认为宏观经济波动会影响消费者风险暴露程度,进而影响其对保险的购买策略[2]; Kugler 和 Ofoghi (2005)考察了经济增长和保险业增长之间表现出来的短期和长期动态关系,发现保险业增长率不论在短期还是长期都与 GDP 增长率显著相关[3]; Browne 和 Kim (1993)认为消费者教育水平越高,风险意识也越强,对保险需求有正面的推动作用[4]; Park 和 Lemaire (2012)研究发现良好的法律环境会正面影响保险业的发展,而市场集中度会负面影响保险需求[5]。在国内,袁成(2015)对中国整体保险、人寿保险和非人寿保险的需求研究表明,收入水平、保险市场发展水平和市场化水平是共同的影响因素;受教育程度、社会保障养老发展水平、子女抚养比和老年抚养比主要影响对人寿保险的需求,通货膨胀主要影响对非寿险的需求[6];李东辉(2007)在研究经济合作和发展中国家寿险消费的决定因素,发现寿险需求的收入弹性显著为正,需求随着受抚养人数和教育水平的增加而增加,随着预期寿命和社会保障支出的增加而减少[7];张宗军(2020)研究发现不同类型、不同区域国家保险的需求差异性根本上源于其工业化、城镇化和老龄化程度,而不是 GDP [8];孙祁祥(2010)研究了 67 个国家 1995~2007 年的面板数据,验证了经济发展程度对保险增长具有显著的正向作用,而不同国内生产总值对保险增长的作用程度不同[9];张连增(2011)对我国 30 个省、市、自治区(除西藏外) 1997~2008 年度面板数据进行了研究,认为人口结构和抚养率问题对人身保险有着正向影响[10];皎亮(2016)对我国 30 个省份 2000~2015 年的财险需求变动进行了分析,发现了经济增长对财险需求存在显著的非线性影响,同时经济增长和城镇化水平的差异也是造成不同省份保险需求发展不平衡的重要原因[11];李晓宇(2012)运用最小二乘估计商业保险发展对经济增长有着极大的正相关性,人身险和财险保费收入的增加能极大地减少损失同时促进经济的快速增长[12];南永清

(2020)考察了商业保险对居民消费行为的影响效应及其作用机制,发现以商业保险密度和保险强度对居民消费有着正向影响[13];锁凌燕(2015)使用 2004~2013 年中国 31 个省市自治区的面板数据对影响商业健康保险发展失衡的原因进行了实证分析,发现医疗卫生资源、预期寿命、教育水平等因素对商业健康保险的发展也十分关键[14];高立飞(2021)使用 probit 模型研究人口流动对居民商业保险需求的影响,发现其促使居民购买商业医疗保险和商业养老保险的概率均显著提高[15]。

目前,国内外学者大多都使用面板数据模型进行保险的影响因素分析,而针对商业保险的研究,多以其作为解释变量分析对居民消费需求和经济增长的影响,如文献[12]、[13],对其影响因素进行分析的较少,且多使用 Probit 模型和面板数据模型进行分析,如文献[14]、[15]。由于复杂的经济体系和社会环境影响,不同时期、不同经济环境下得出的有关结论可能有所差异,且在研究中往往忽略多重共线性的影响。因此,本文结合当前我国商业保险迅速发展和多变的保险环境现状,在用 Lasso 回归筛选因素指标的基础上,用面板数据模型对商业保险的影响因素进行分析,并根据分析结果提出若干可行性建议。

### 3. 模型介绍

本文在对商业保险的影响因素进行分析时,首先利用 Lasso 回归进行指标筛选以克服指标间的多重共线性问题。Lasso 回归是由 Tibshirani 于 1996 年提出的一种压缩估计,其数学描述为[16]

$$\arg \min \left\{ \sum_{i=1}^n \left( y_i - \alpha - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (1)$$

其中:  $x_{ij}$  为解释变量,  $y_i$  为被解释变量;  $\sum_{i=1}^n \left( y_i - \alpha - \sum_{j=1}^p \beta_j x_{ij} \right)^2$  为残差平方和,  $\alpha$  为截距项,  $\beta_j$  为回归系数;  $\lambda \sum_{j=1}^p |\beta_j|$  为惩罚项,  $\lambda$  为惩罚参数,其值越大,删除的变量越多,反之删除的越少。

本文在指标筛选的基础上,利用面板数据模型对商业保险的影响因素进行分析。面板数据模型的一般形式为[17]

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

其中  $y_{it}$  为被解释变量,  $x_{it} = (x_{it}^1, x_{it}^2, \dots, x_{it}^p)$  为解释变量向量;  $\alpha_{it}$  为截距项,  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  为回归系数向量,  $\varepsilon_{it}$  为误差项;  $p$  表示解释变量的个数,  $i$  表示个体,  $t$  表示时间,  $N$  表示个体数量,  $T$  表示时间长度。

一般来说,面板数据模型常见的有两类:一类为混合估计模型,即所有横截面和个体在不同时期的斜率和截距都相等;另一类为变截距模型,即截距项在不同的个体上有所不同,又可以分为固定效应模型和随机效应模型,前者在每一个截面具有不同的截距项,但这些截距项不会随着时间而变化,而后者在每一个截面具有不同随机截距项。由于商业保险市场的多变性和商业保险的迅速发展导致不同时期的截距很难相等,所以在面板数据模型的选择上,本文选择变截距模型进行分析。

## 4. 实证分析

### 4.1. 指标选取及数据来源

商业保险的保费收入(PI)是衡量商业保险发展水平最直观的指标,因此选其作为商业保险影响因素分析的被解释变量。商业保险的影响因素通常有人口、经济等,居民作为商业保险的购买者,其特征会对商业保险有重要影响,而经济发展同样会对商业保险的发展起到重要支持作用,故本文选取人口总数

(POP)、就业人数(EMPL)、少儿化程度(CHD)和老龄化程度(OLD)作为人口方面的指标; 选取国内生产总值(GDP)、物价指数(CPI)、人均可支配收入(DPI)、人均消费支出(PCE)、进出口贸易总额(TIAE)和通货膨胀率(INF)作为经济方面的指标。另外, 还选取了医疗卫生人数(MED)、在校学生数(EDU)和城镇人口比例(URB)分别作为教育、医疗和城市化方面的指标。选取的指标如表 1 所示。

**Table 1.** Selected indicators and their meanings

**表 1.** 选取指标及其含义

指标类型	指标	符号	备注
人口	人口总数	POP	
	就业人数	EMPL	
	少儿化程度	CHD	15 岁以下人口数量与 15 至 65 岁人口数量的比值
	老龄化程度	OLD	65 岁以上人口数量与 15 至 65 岁人口数量的比值
经济	国内生产总值	GDP	
	物价指数	CPI	
	人均可支配收入	DPI	
	人均消费支出	PCE	
	进出口贸易总额	TIAE	
	通货膨胀率	INF	
医疗	医疗卫生人数	MED	每万人中医疗卫生人数
教育	在校学生数	EDU	每千人中本专科在校学生数
城市化	城镇人口比例	URB	城镇人口占总人口的比值

本文选取全国 31 个省(市、自治区) 2005~2018 年的面板数据进行建模分析, 其来源是 EPS 平台和中国统计年鉴。由于保费收入、人口总数、国内生产总值、就业人数、物价指数、人均可支配收入、人均消费支出、进出口贸易总额、医疗卫生人数、在校学生数的数值较大, 为避免异方差的影响, 并保证数据的正态性, 对这些指标进行对数处理。

## 4.2. Lasso 降维

本文对选取的 13 个指标用 Lasso 回归进行降维, 结果如表 2 所示。当  $\lambda$  由大到小变化时, Lasso 估计可筛选出越来越多的指标进入模型, 比如: 当  $\lambda > 599.2123$  时, 由于惩罚力度过大, 所有指标的系数均为 0; 当  $\lambda = 599.2123$  时, 指标人均可支配收入首先进入模型; 当  $\lambda$  降到 342.8913 时, 指标医疗卫生人数又进入模型。

**Table 2.** Lasso regression variables

**表 2.** 拉索回归变量

Knot	Lambda	s	L1-Norm	EBIC	R-sq	Entered/removed
1	599.2123	2	0.13424	-373.807	0.163	Added DPI
2	342.8913	3	0.73966	-844.848	0.7001	Added MED

Continued

3	236.3417	4	1.06039	-1133.24	0.8408	Added GDP
4	112.2815	6	1.43397	-1572.44	0.9397	Added PCEEDU
5	53.34279	5	1.68765	-1805.55	0.963	Removed GDP
6	48.60396	6	1.77064	-1815.46	0.9642	Added OLD
7	44.28612	7	1.98747	-1826.09	0.9655	Added GDP
8	30.5247	8	2.75615	-1868.06	0.9689	Added EMPL
9	15.91561	9	3.96715	-1907.71	0.9718	Added TIAE
10	10.97001	10	4.48436	-1919.54	0.9729	Added CPI
11	2.47595	11	6.4533	-1941.78	0.9745	Added CHD
12	0.81076	12	6.92413	-1937.16	0.9746	Added POP
13	0.55883	13	7.02466	-1931.3	0.9746	Added URB
14	0.31978	14	7.18528	-1925.53	0.9746	Added INF

图 1 为 Lasso 回归求解过程的路径图, 直观反映了不同指标回归系数的变化过程。当  $\lambda = 0$  时, 不存在惩罚项, 此时 Lasso 估计等价于最小二乘估计。

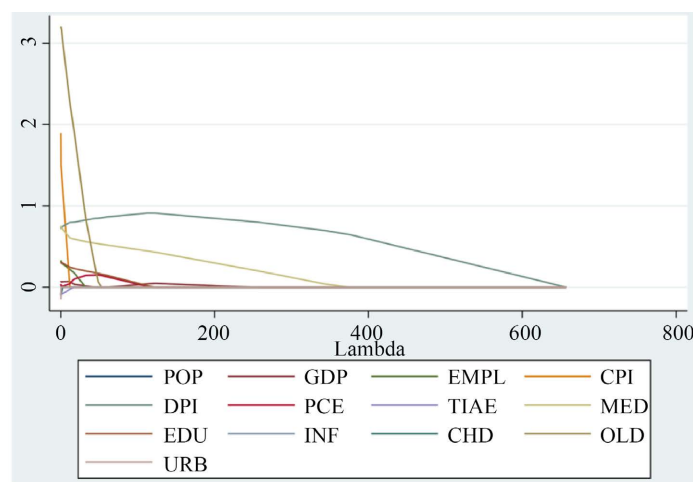


Figure 1. Lasso regression path

图 1. 拉索回归路径

文本采用 10 折交叉验证的方法选择最优的  $\lambda$ , 如图 2 所示。打星号处  $\lambda = 2.0555795$ , 即为可使 MSPE 最小化的  $\lambda$  值。

表 3 是降维后的回归情况, 其中第二列为最优  $\lambda$  对应的 Lasso 回归的估计结果, 可以看出人口总数、通货膨胀率和城镇人口比例通过 Lasso 回归后被剔除; 第三列是用 Lasso 进行变量筛选后的 OLS 回归结果。两列的数值差别不大, 说明指标经过 Lasso 筛选后受多重共线性的影响很小, 不容易受异质性影响, 使估计较为稳定。

53	5.2116433	0.01651613	0.00121646
54	4.7486553	0.01648028	0.00121959
55	4.3267978	0.01645114	0.00122289
56	3.942417	0.01643112	0.00122859
57	3.5921836	0.01641608	0.00123396
58	3.2730639	0.01640492	0.00123901
59	2.9822939	0.01639657	0.00124371
60	2.7173551	0.01639037	0.00124814
61	2.4759528	0.01638595	0.00125233
62	2.255996	0.0163829	0.00125624
63	2.0555795	0.01638093	0.00125989 *
64	1.8729675	0.01638305	0.00126336
65	1.7065782	0.01638886	0.00126668
66	1.5549705	0.01639616	0.00126926
67	1.4168312	0.01640628	0.00127025
68	1.2909638	0.01641602	0.00127113
69	1.1762782	0.0164265	0.00127209
70	1.0717809	0.01643845	0.00127369
71	0.9765668	0.01645197	0.00127515
72	0.8898113	0.01646534	0.00127669
73	0.81076291	0.01647942	0.0012784

Figure 2. 10 fold cross verification results (part)  
图 2. 10 折交叉验证结果(部分)

Table 3. Regression of Lasso dimension reduction  
表 3. 拉索降维后的回归

Selected	Lasso	Post-est OLS
EMPL	0.2953526	0.3115293
CHD	-0.0324545	-0.0972736
OLD	3.0978678	3.2569414
GDP	0.0675413	0.0644618
CPI	1.3127682	1.5309542
DPI	0.7435855	0.7334780
PCE	0.0134543	0.0112672
TIAE	-0.0836843	-0.0932533
MED	0.7227370	0.7433331
EDU	0.2989148	0.3028934

### 4.3. 模型选择

本文使用 Hausman 检验在固定效应模型和随机效应模型中选择, 如表 4 所示, chi2 值为 124.22, P 值为 0, 所以可以拒绝使用随机效应模型的原假设。

### 4.4. 结果分析

固定效应模型的参数估计结果如表 5 所示, 由于  $R^2$  为 0.975, 所以模型的拟合效果良好。

物价指数对商业保险有着不显著的负向作用, 人均可支配收入、人均消费支出、进出口贸易总额、老龄化程度对商业保险有着不显著的正向作用。

**Table 4.** Hausman test  
**表 4.** 豪斯曼检验

Variable	FE	RE	Difference	sqrt(diag(V_b-V_B))
	(b)	(B)	(b-B)	S.E.
EMPL	0.32705	0.3100553	0.0169947	0.0072827
CHD	-0.082663	-0.4710282	0.3883652	0.0905461
OLD	3.203845	2.936344	0.2675006	0.1965339
GDP	0.0669863	0.0553316	0.0116547	.
CPI	1.994288	1.450754	0.5435336	.
DPI	0.7197498	0.6624854	0.0572645	0.0081205
PCE	0.0416554	0.1127406	-0.0710852	.
TIAE	-0.0936961	-0.0512486	-0.0424474	0.0109342
MED	0.7560655	0.6885826	0.067483	0.0144555
EDU	0.2980343	0.2184758	0.0795585	0.0254603
chi2 (13)		124.22		
Prob > chi2		0.0000		

少儿化程度在 10% 的水平上显著, 系数为-0.5339, 说明对商业保险有着一定的负向作用。目前, 少儿的成长是我们一直关注的问题, 各个家庭都愿意给自己的孩子一个更好的生活, 而目前抚养孩子需要的成本较高, 且孩子的健康安全问题的相对成年人和老年人来讲相对好一些, 所以更多的家庭将资金投入到了孩子的发展之中, 购买商业保险的需求就没那么明显, 所以少儿化程度对商业保险就产生了负向影响。在校学生数在 5% 的水平上显著, 系数为 0.1243, 对商业保险有着一定的正向作用。随着教育水平的提高, 消费者风险意识也得到了提高, 教育程度可以改变风险态度, 而商业保险正好可以降低和规避风险, 因此教育水平越高, 对商业保险的购买需求就越高, 从而对商业保险就产生了正向影响。

国内生产总值、就业人数和医疗卫生人数在 1% 的水平上显著。国内生产总值的系数为 0.2347, 对商业保险有着一定的正向作用。商业保险作为现代金融市场的重要支柱之一, 它不仅能促进金融市场的发展, 同样也受到金融市场的影响, 所以国内生产总值的提高, 同样会带动商业保险的发展, 进而对商业保险产生正向影响。就业人数的系数为 0.2358, 对商业保险就有着一定的正向作用。就业人数增多意味着更多的人有能力去购买商业保险, 同时在劳动的过程中注定会出现一定的风险, 所以就业人数的提高对商业保险就产生了正向影响。医疗卫生人数系数为 0.5978, 对商业保险有着一定的正向作用。消费者购买商业保险, 有很大一部分是为了自身健康和意外作准备的, 而医疗卫生人数的提高, 令消费者更加相信在自身受到伤害和健康出现状况时, 可以得到最有效的治疗, 而商业保险还可以降低其所要承担的风险, 进而促进商业保险的发展。

对比固定效应模型与 Lasso (OLS) 的估计结果可以发现, 物价指数和进出口贸易总额对商业保险的影响在估计中的方向相反。物价指数在 Lasso (OLS) 估计中为正, 而在固定效应模型中为负: 在经济社会中, 物价指数的升高往往会对经济造成一定的影响, 这同样会对商业保险起着一定的抑制作用, 所以固定效应模型所得到的负向影响更符合实际; 进出口贸易总额在 Lasso (OLS) 估计中为负, 而在固定效应模型中为正: 进出口贸易总额体现一个地区的进出口水平, 进出口贸易总额越高的省份其经济发展也相对较快, 这给了商业保险一个很好的推动力, 因此对商业保险产生了正向的影响, 符合实际情况。其它指标在估

计中对商业保险的影响方向一致。

**Table 5.** Parameter estimation of fixed effect model

**表 5.** 固定效应模型参数估计

VARIABLES	Premium	VARIABLES	Premium
EMPL	0.2358*** (3.03)	DPI	0.2258 (1.16)
CHD	-0.5339* (-1.93)	PCE	0.0801 (0.60)
OLD	0.4868 (1.09)	TIAE	0.0075 (0.37)
GDP	0.2347*** (3.22)	MED	0.5978*** (8.67)
CPI	-0.0921 (-0.12)	EDU	0.1243** (2.00)
	Constant		0.7452 (0.18)
	Observations		465
	R-squared		0.984
	Numberof province		31

\*\*\*P < 0.01, \*\*P < 0.05, \*P < 0.1.

## 5. 结论与启示

本文利用 2004~2018 年的 31 个省(市、自治区)的省际面板数据在利用 Lasso 进行指标筛选的基础上, 利用固定效应模型进行了商业保险的影响因素分析, 结果表明, 物价指数、人均可支配收入、人均消费支出、进出口贸易总额、老龄化程度对商业保险的影响不显著, 而少儿化程度在 10% 水平上对商业保险有着一定的负向作用, 在校学生数在 5% 的水平上对商业保险有着一定的正向作用, 国内生产总值、就业人数和医疗卫生人数均在 1% 的水平上对商业保险有着一定的正向作用。基于以上结论, 本文就商业保险发展提出以下建议:

### 1) 进一步采取有效措施刺激居民对商业保险的购买需求

政府要着力完善劳动力就业市场, 增加就业机会, 挖掘新增就业岗位, 切实稳定地提升居民收入水平。深化教育、卫生改革, 进一步降低消费者因上学、看病等因素导致的大额支出, 让更多的人有学可上, 有病可医, 同时加大对教育和医疗方面的重视, 降低居民因教育、医疗、消费等因素导致的大额支出, 使居民可支配收入得到更好的利用。完善商业保险市场制度, 宣传和展示商业保险给居民带来的益处。

### 2) 发展新兴商业保险产品和服务

在社会发展多元化的背景下, 需要开发多层次、个性化的商业保险产品和服务。特别是就少儿化问题和老龄化问题上, 扩展商业少儿保险和养老保险产品和服务, 例如可以保障孤寡老人生活和居住问题的养老保险、减轻少儿用品和教育费用的少儿保险等。



### 3) 使用科技力量完善商业保险市场

结合大数据、云计算等技术使保险服务更加精准可靠, 提高风险保障, 同时构建网络保障体系来保障商业保险市场。

## 参考文献

- [1] Skipper, H.D. (1997) Foreign Insurers in Emerging Markets: Issues and Concerns Centre for Risk Management and Insurance. Georgia State University, Atlanta.
- [2] Meglena, J. (2000) Background Risk Demand for Insurance and Choquet Expected Utility Preferences. *The Geneva Paper on Risk and Insurance Theory*, **25**, 7-28.
- [3] Kugler, M. and Ofoghi, R. (2005) Does Insurance Promote Economic Growth? Evidence from the UK. University of Southampton, Southampton.
- [4] Browne, M.J. and Kim, K. (1993) An International Analysis of Life Insurance Demand. *Journal of Risk and Insurance*, **60**, 616-634. <https://doi.org/10.2307/253382>
- [5] Park, S.C. and Lemaire, J. (2012) The Impact of Culture on the Demand for Non-Life Insurance. *ASTIN Bulletin: The Journal of the IAA*, **42**, 501-527.
- [6] Yuan, C. and Jiang, Y. (2015) Factors Affecting the Demand for Insurance in China. *Applied Economics*, **47**, 4855-4867. <https://doi.org/10.1080/00036846.2015.1037437>
- [7] Li, D.H., Moshirian, F., Nguyen, P. and Wee, T. (2007) The Demands for Life Insurance in OECD Countries. *Journal of Risk and Insurance*, **74**, 637-652. <https://doi.org/10.1111/j.1539-6975.2007.00228.x>
- [8] 张宗军, 令涛. 工业化、城镇化、老龄化与保险发展——理论分析与实证检验[J]. 保险研究, 2020(8): 30-46.
- [9] 孙祁祥, 郑伟, 锁凌燕, 何小伟. 市场经济对保险业发展的影响: 理论分析与经验证据[J]. 金融研究, 2010(2): 158-172.
- [10] 张连增, 尚颖. 中国人口老龄化对人身保险市场发展的影响分析——基于省际面板数据的经验分析[J]. 保险研究, 2011(1): 46-53.
- [11] 咬亮, 程凯. 经济增长、城镇化与财险需求的非线性关系研究——基于面板平滑转换回归模型 PSTR 的检验[J]. 保险研究, 2016(10): 48-62.
- [12] 李晓宇. 商业保险发展对经济增长影响的实证分析[J]. 中国商贸, 2012(33): 234-235.
- [13] 南永清, 贺彭皓, 周勤. 商业保险对居民消费影响研究——基于中国省级面板数据的经验证据[J]. 保险研究, 2020(3): 23-40.
- [14] 锁凌燕, 完颜瑞云, 陈滔. 我国商业健康保险地区发展失衡现状及原因研究[J]. 保险研究, 2015(1): 42-53.
- [15] 高立飞, 王国军. 人口流动对居民商业保险需求的影响研究——基于 CGSS2017 数据的实证分析[J]. 河北经贸大学学报, 2021, 42(2): 85-91.
- [16] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [17] 白仲林. 面板数据的计量经济分析[M]. 天津: 南开大学出版社, 2008.