

基于典型相关分析法的我国邮电业与国民经济关系分析

刘程馨, 牟唯嫣, 代铁林

北京建筑大学理学院, 北京

收稿日期: 2022年7月22日; 录用日期: 2022年8月15日; 发布日期: 2022年8月24日

摘要

为了更好地研究我国邮电业的发展对国民经济状况的影响, 本文就2007~2019年我国国民经济及邮电业的相关数据, 利用R语言软件, 采用典型相关分析法分析了我国邮电业和国民经济之间的关系。结果表明, 我国邮电业和国民经济之间具有很强的正相关关系。因此, 邮电业的发展对国民经济水平的提升发挥着重要的推动作用, 应采取相关措施促进我国邮电业的持续发展。

关键词

邮电业, 国民经济, 典型相关分析

The Analysis of the Relationship between Post and Telecommunication Industry and National Economy Based on CCA

Chengxin Liu, Weiyan Mu, Tielin Dai

School of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: Jul. 22nd, 2022; accepted: Aug. 15th, 2022; published: Aug. 24th, 2022

Abstract

In order to better study the impact of the development of post and telecommunication industry on the national economic situation in China, according to the relevant data from 2007 to 2019, this paper uses R language software, and applies CCA method to analyze relationship between the national economy and post and telecommunication industry. The results show that there is a strong positive

correlation between post and telecommunication industry and the national economy in China. Therefore, the development of post and telecommunication industry plays an important role in promoting the improvement of the level of the national economic, and relevant measures should be taken to promote the sustainable development of post and telecommunications industry in China.

Keywords

Post and Telecommunication Industry, National Economy, CCA

Copyright © 2022 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

邮电业是国民经济中办理通信业务的社会生产部门，是国民经济的基础设施和社会发展的必要条件[1]。随着我国电子信息技术蓬勃发展，邮电业从人们的工作、生活及消费等方面对国民经济产生日益深刻的影响。因此，要在我国当前的经济状况下更好地发展邮电业，提高人民的生活质量和水平，充分了解我国邮电业和国民经济状况之间的关系是不可或缺的。

本文以 2007~2019 年我国邮电业及国民经济状况为例，以函件、包裹、移动电话用户和固定电话用户四个指标来衡量我国各年份的邮电业，采用第一产业产值、工业产值、建筑业产值以及第三产业产值四个指标衡量我国各年份的经济状况[2]。利用多元统计中的典型相关分析方法进行数据分析，从而客观、科学地认知我国邮电业指标和国民经济指标之间的相关性及组内的相关性。

2. 思想方法

针对我国邮电通讯业与经济的关系，已有不少学者进行了相关研究与分析，并取得了研究成果。康国栋等(2008)以 1985~2005 年我国 31 个省(市、自治区)的人均邮电业务量为例，将时间序列和空间差异方法相结合，分析得出了人均邮电业务量与人均 GDP 呈正相关关系[3]；李再扬与杨少华(2010)基于 2003-2008 年我国 31 个省(市、自治区)的电信业技术效率，采用 DEA 方法进行度量，并建立计量经济模型得到了其影响因素[4]；薛声家与王清(2010)基于 DEA 超效率模型，分析了我国 31 个省(市、自治区)电信业的效率及影响因素[5]；韩磊与窦彩兰(2010)同样运用 DEA 方法分析得到了我国电信业区域运营效率的排名顺序和层次结构，并利用 Tobit 回归得出了我国电信业发展的影响因素[6]；郭艳春与刘斐(2012)基于 2000~2011 年深圳、广州、东莞以及佛山四个城市的邮电业务情况，采用面板数据模型分析其影响因素，结果表明对邮电业的推动作用最大的因素是第三产业发展状况和区域开放度[7]；李清清与李厚彪(2018)构建了多元回归分析模型，并通过统计检验得出了移动电话年末用户数、总人口数量、公路里程、居民消费水平以及对外贸易总额 5 个对邮电业发展影响显著的因素[8]；刘珉慧等(2020)基于 1990-2019 年邮电业务数据，利用 Johansen 协整检验及误差修正模型检验了邮政业务量、电信业务量和人均 GDP 三个指标之间的关系，结果表明电信业务量与邮政业务量存在着长期均衡关系[9]。综合上述文献可以发现，典型相关分析方法鲜有研究者使用。因此本文采用典型相关分析法，对我国邮电业与经济发展之间的关系进行分析。

2.1. 数据的来源及说明

本文以函件(单位：亿件)、包裹(单位：万件)、移动电话用户(单位：万户)和固定电话用户(单位：万

户)作为衡量我国各年份的邮电业的指标,以第一产业产值(单位:亿元)、工业产值(单位:亿元)、建筑业产值(单位:亿元)以及第三产业产值(单位:亿元)作为衡量我国各年份的经济状况的指标。本文选取了2007~2019年以上各指标的数据,数据均来源于2021年中国统计年鉴[10],见表1。

Table 1. Output value of post and telecommunication industry and that of primary industry, industry, construction industry and tertiary industry

表 1. 邮电业与第一产业、工业、建筑业及第三产业产值情况

年份	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4
2007	69.5	9103	54,731	36,564	27,674	111,691	15,348	115,788
2008	73.6	7937	64,125	34,036	32,464	131,724	18,808	136,828
2009	75.3	7230	74,721	31,373	33,584	138,093	22,682	154,765
2010	74.0	6643	85,900	29,434	38,431	165,123	27,259	182,062
2011	73.8	6883	98,625	28,510	44,782	195,139	32,927	216,124
2012	70.7	6876	111,216	27,815	49,085	208,901	36,896	244,856
2013	63.4	6925	122,911	26,699	53,028	222,333	40,897	277,984
2014	56.1	6024	128,609	24,943	55,626	233,197	45,402	310,654
2015	45.8	4243	127,140	23,100	57,775	234,969	47,761	349,745
2016	36.2	2794	132,193	20,662	60,139	245,406	51,499	390,828
2017	31.5	2657	141,749	19,376	62,100	275,119	57,906	438,356
2018	26.7	2408	156,610	19,209	64,745	301,089	65,493	489,701
2019	21.7	2155	160,135	19,103	70,467	317,109	70,904	534,233

2.2. 典型相关分析的思想

典型相关分析研究两组变量间整体的线性相关关系,它是将每一组变量作为一个整体来进行研究[11]。典型相关分析借助主成分分析的思想,对每组变量以其线性组合的形式分别构造综合变量,而后考察两个综合变量的相关程度,这种相关程度用典型相关系数来进行描述。

设 $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ 和 $\mathbf{y} = (y_1, y_2, \dots, y_p)^T$, 是两个具有相关性的随机向量, 分别对两组变量构造综合变量 u_i, v_i , 并且每一个综合变量都为原变量的线性组合, 即

$$u_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{ip}x_p = \mathbf{a}^T \mathbf{x}$$

$$v_i = b_{i1}y_1 + b_{i2}y_2 + \dots + b_{ip}y_p = \mathbf{b}^T \mathbf{y}$$

考虑方差为 1 的 \mathbf{x}, \mathbf{y} 的线性函数 $\mathbf{a}^T \mathbf{x}$ 和 $\mathbf{b}^T \mathbf{y}$, 考察它们的相关系数。如果存在常向量 $\mathbf{a}_1, \mathbf{b}_1$, 使得

$$\rho(\mathbf{a}_1^T \mathbf{x}, \mathbf{b}_1^T \mathbf{y}) = \max \rho(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y})$$

$$\text{var}(\mathbf{a}_1^T \mathbf{x}) = \text{var}(\mathbf{b}_1^T \mathbf{y}) = 1$$

换言之 $\mathbf{a}_1^T \mathbf{x}, \mathbf{b}_1^T \mathbf{y}$ 是相关程度最大的一对典型相关变量, 则称 $\mathbf{a}_1^T \mathbf{x}, \mathbf{b}_1^T \mathbf{y}$ 为 \mathbf{x}, \mathbf{y} 的第一对典型相关变量, 以类似的方法可以求得第二对、第三对……使得各对典型相关变量之间是互不相关的。

3. 实证分析

我们使用 R 软件，对邮电业和国民经济数据进行多元统计分析。

3.1. 原始变量的相关系数与典型相关系数

首先，我们求样本系数矩阵，结果见表 2。结果显示，邮电业与国民经济指标间的相关系数有正有负，并且任意两个指标间的相关系数的绝对值都接近 1，这说明邮电业指标和国民经济指标有很强的相关性，同组内指标的相关性也很强。

Table 2. The sample correlation coefficient matrix

表 2. 样本相关系数矩阵

	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4
x_1	1.000	0.947	-0.869	0.899	-0.887	-0.895	-0.932	-0.963
x_2	0.947	1.000	-0.896	0.964	-0.911	-0.914	-0.941	-0.958
x_3	-0.869	-0.896	1.000	-0.972	0.994	0.992	0.984	0.964
x_4	0.899	0.964	-0.972	1.000	-0.975	-0.969	-0.972	-0.966
y_1	-0.887	-0.911	0.994	-0.975	1.000	0.989	0.984	0.970
y_2	-0.895	-0.914	0.992	-0.969	0.989	1.000	0.994	0.981
y_3	-0.932	-0.941	0.984	-0.972	0.984	0.994	1.000	0.995
y_4	-0.963	-0.958	0.964	-0.966	0.970	0.981	0.995	1.000

求得原始变量的相关系数之后，我们再进行典型相关分析。由于变量的单位不一致，我们用标准化之后的数据进行分析。得到第一典型相关系数为 0.99921968，它比邮电业指标和国民经济指标间的任一其他对的相关系数都要大，这充分体现了典型相关分析对两组变量间高维关系的有效浓缩。

3.2. 典型相关系数的显著性检验

为了确定要保留几对典型相关变量，我们有必要对典型相关系数进行显著性检验，考察典型相关系数是否显著不为 0。检验结果见表 3。

Table 3. Significance test for the canonical correlation coefficient

表 3. 典型相关系数的显著性检验

	r	Q	P
第一对典型相关变量	0.99921968	81.11218059	1.048196e-10
第二对典型相关变量	0.99098003	32.64610390	1.538818e-04
第三对典型相关变量	0.52934104	2.49824202	6.449506e-01
第四对典型相关变量	0.05855955	0.03465585	8.523187e-01

给定显著性水平 $\alpha = 0.05$ ，第一对、第二对典型相关变量的相关系数对应的相伴概率 P 值分别为 1.048196e-10 和 1.538818e-04，均远小于显著性水平 $\alpha = 0.05$ ，因此拒绝典型相关系数为 0 的原假设，认为第一对、第二对典型相关是显著的；而第三对、第四对典型相关变量的相关系数对应的 P 值分别为

6.449506e-01 和 8.523187e-01, 均大于显著性水平 $\alpha = 0.05$, 因此接受原假设, 认为第三对、第四对典型相关是不显著的。综上所述, 只有第一对和第二对典型相关是显著的。

3.3. 典型载荷

检验完成后, 我们来求典型相关载荷[12], 邮电业指标 x 和国民经济指标 y 的典型载荷分别见表 4 和表 5。

Table 4. The canonical loadings of x
表 4. x 的典型载荷

	\hat{u}_1	\hat{u}_2	\hat{u}_3	\hat{u}_4
x_1^*	-0.4103491	1.6066966	-3.103210	2.6948379
x_2^*	-0.5457285	0.5501079	8.556985	-3.8056044
x_3^*	0.6917840	1.8652805	-7.457607	-0.6065234
x_4^*	0.6097558	-0.3592558	-12.828155	0.4848087

Table 5. The canonical loadings of y
表 5. y 的典型载荷

	\hat{v}_1	\hat{v}_2	\hat{v}_3	\hat{v}_4
y_1^*	0.1621535	0.6878757	6.853337	1.252144
y_2^*	-0.4169586	1.0338773	-1.015479	-17.399498
y_3^*	-0.2731420	6.6106280	-12.537197	30.789045
y_4^*	1.5201848	-8.1857882	6.795978	-14.805694

根据典型相关载荷的结果可以得出, 邮电业指标的第一典型变量 \hat{u}_1 为:

$$\hat{u}_1 = -0.4103491x_1^* - 0.5457285x_2^* + 0.6917840x_3^* + 0.6097558x_4^*$$

它为邮电业各指标的线性组合, 在移动电话用户上的权数最大, 其次为固定电话用户。说明移动电话用户和固定电话用户为影响邮电业发展的主要指标, 它们在邮电业中占据了主导地位。

邮电业指标的第二典型变量 \hat{u}_2 为:

$$\hat{u}_2 = 1.6066966x_1^* + 0.5501079x_2^* + 1.8652805x_3^* - 0.3592558x_4^*$$

它近似地是函件和移动电话用户的加权求和, 仍然在移动电话用户上的权数最大。移动电话用户与固定电话用户的权数符号相反, 说明两变量之间存在一定的相互抑制作用。

来自国民经济指标的第一典型变量 \hat{v}_1 为:

$$\hat{v}_1 = 0.1621535x_1^* - 0.4169586x_2^* - 0.2731420x_3^* + 1.5201848x_4^*$$

它为国民经济各指标的线性组合, 在第三产业产值上的权数最大, 且其权数远高于第一产业、工业以及建筑业产值的权数。这表明第三产业产值相较于其他产业的产值对国民经济的影响更大, 是与国民经济相关联的主要指标。

来自国民经济指标的第二典型变量 \hat{v}_2 为:

$$\hat{v}_2 = 0.6878757x_1^* + 1.0338773x_2^* + 6.6106280x_3^* - 8.1857882x_4^*$$

它在第三产业产值上的权数最大, 其次是建筑业产值。

综合两对典型变量可知, 对邮电业产生最大影响的指标是移动电话用户, 对国民经济产生最大影响的指标是第三产业产值。

3.4. 典型变量解释原始变量方差的比例

下面我们计算两组典型变量解释两组原始变量方差的比例, 汇总计算结果可得表 6。

第一对典型相关变量能够比较全面地预测对应的那组原始变量, 来自邮电业指标的标准方差被第一个典型变量 u_1 的解释的方差比例为 0.92620846, 第一个典型变量 v_1 解释国民经济指标的比例为 0.950306675; 而来自邮电业指标的标准方差被对方组的第一个典型变量 v_1 解释的比例为 9.247635e-01, 国民经济指标的标准方差被对方组的第一个典型变量 u_1 解释的比例为 9.488242e-01, 这表明国民经济水平与邮电业发展有很强的关联性。

第二对典型相关变量的预测效果远次于第一对典型相关变量, 来自邮电业指标的标准方差被第二个典型变量 u_2 的解释的方差比例为 0.04522376, 第二个典型变量 v_2 解释国民经济指标的比例为 0.045275384; 而来自邮电业指标的标准方差被对方组的第二个典型变量 v_2 解释的比例为 4.441160e-02, 国民经济指标的标准方差被对方组的第二个典型变量 u_2 解释的比例为 4.446230e-02。第二对典型变量解释本组和对方组原始变量的方差比例均小于第一对典型相关变量, 解释性较第一对典型相关变量存在较大差距。

Table 6. The proportion of the variance of the original variable explained by the canonical variable

表 6. 典型变量解释原始变量方差的比例

指标	第一对典型相关变量		第二对典型相关变量		第三对典型相关变量		第四对典型相关变量	
	u_1	v_1	u_2	v_2	u_3	v_3	u_4	v_4
邮电业	0.92620846	9.247635e-01	0.04522376	4.441160e-02	0.00559996	1.569120e-03	0.02296782	7.876174e-05
国民经济	9.488242e-01	0.950306675	4.446230e-02	0.045275384	9.200588e-04	0.003283556	3.890056e-06	0.001134385

3.5. 典型变量之间的关系

计算邮电业指标与国民经济指标的典型变量得分(表 7 和表 8), 并绘制得分等值平面图(图 1)。

从得分情况来看, 随着年份的变化, 邮电业指标的第二个典型变量 u_2 的得分和国民经济指标的第二个典型变量 v_2 的得分均呈现逐年增长的趋势, 这说明邮电业指标和国民经济指标有强正相关关系。

从得分等值平面图中可以看出, 第一对典型相关变量 u_1 和 v_1 的散点近似在一条直线上, 两者之间呈现高度线性正相关关系, 散点图中没有离群点出现。这表明我国邮电业和国民经济发展之间的关系很稳定, 整体的变化趋势是非常平稳的。

Table 7. Canonical variable scores for post and telecommunication
表 7. 邮电业指标的典型变量得分

年份	u_1 得分	u_2 得分	u_3 得分	u_4 得分
2007	-1.1863467	-1.81146349	0.34248454	-1.96814578
2008	-1.0768271	-1.08771727	-0.95192132	0.07743987
2009	-1.0144568	-0.37406083	-0.17900264	1.03140706
2010	-0.8307984	0.11583652	-0.24295193	1.44243392
2011	-0.7221947	0.90692852	-0.06908357	0.72638629
2012	-0.4762605	1.38844316	-0.82239920	0.04281105
2013	-0.2204433	1.52633749	0.40050023	-1.31039867
2014	0.0659975	1.15591970	0.91040085	-1.08150023
2015	0.4624293	-0.04316522	0.45061212	0.28381363
2016	0.8371868	-0.71680919	0.98237269	1.04252001
2017	1.0219841	-0.52154473	1.97579927	0.35915748
2018	1.4598404	-0.14052603	-1.05769060	-0.15753814
2019	1.6798894	-0.39817865	-1.73912044	-0.48838651

Table 8. Canonical variable scores for national economy
表 8. 国民经济指标的典型变量得分

年份	v_1 得分	v_2 得分	v_3 得分	v_4 得分
2007	-1.19196786	-1.72964305	-0.18067330	-0.0518260
2008	-1.08588527	-1.11802653	0.48809802	-1.2433353
2009	-0.97563230	-0.57120226	-0.92327596	1.9851646
2010	-0.86239963	0.20301719	-0.82318767	0.1863091
2011	-0.69274235	1.10841234	-0.45635091	-1.1013468
2012	-0.47461224	1.32776180	0.08383825	-0.5777711
2013	-0.21050537	1.27383039	0.64256727	-0.4155037
2014	0.04134644	1.32734596	0.18288858	1.2497782
2015	0.45117094	0.02294182	1.48800997	0.8864289
2016	0.80796694	-0.73156509	1.87999374	0.3968440
2017	1.06476315	-0.57708720	0.18075492	-1.3694685
2018	1.37787718	-0.23206884	-1.76742407	-0.4052368
2019	1.75062037	-0.30371653	-0.79523885	0.4599634

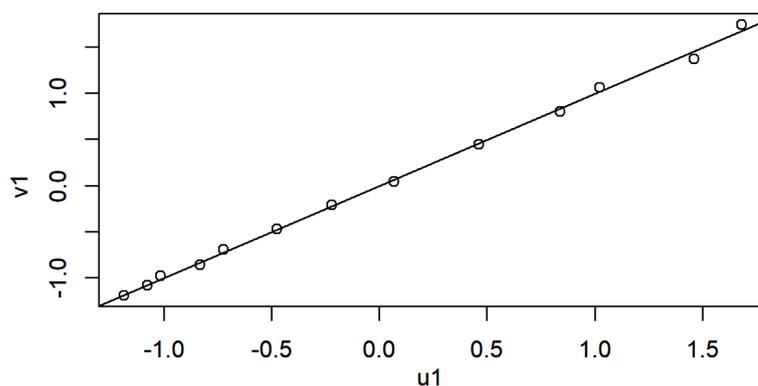


Figure 1. Diagram of the scores of the first pair of canonical correlation variables for the post and telecommunication industry and national economic data

图 1. 邮电业和国民经济数据第一对典型相关变量得分等值平面图

4. 结论

本文基于典型相关分析法,依据 2007~2019 年的数据,对我国的邮电业和国民经济的关系进行了统计分析。首先得到了原始变量之间的相关系数及典型相关系数,邮电业指标和国民经济指标之间有很强的相关性,同组内指标的相关性也很强。而后进行了典型相关系数的检验,检验结果表明,只有第一对和第二对典型相关变量是显著的。随后计算得出了邮电业指标和国民经济指标的典型相关载荷,并分析得出对邮电业产生最大影响的指标是移动电话用户,对国民经济产生最大影响的指标是第三产业产值。最后,计算了典型变量得分并绘制了典型相关变量得分等值平面图。结果表明,我国邮电业和国民经济发展呈高度的线性相关关系。因此,邮电业的发展对我国国民经济水平的提高具有深刻的影响,发挥着积极的推动作用。除此之外,经济社会的发展对邮电业的高速成长起着重要的支撑作用[7]。不同的国民经济水平会对邮电业的发展产生不同程度的影响。所以,我国应采取适合我国国情的更加有效的措施以促进邮电业的发展,从而推动国民经济水平的提升;此外,国民经济的增长也应更好地推动邮电业的持续发展。

参考文献

- [1] 成凯. 从邮电建设看宁夏省政府与国民政府的关系[J]. 宁夏大学学报: 人文社会科学版, 2020, 42(2): 116-123.
- [2] 费宇, 郭民之, 陈贻娟. 多元统计分析——基于 R [M]. 北京: 中国人民大学出版社, 2014.
- [3] 康国栋, 孙根年, 王美红. 我国邮电业发展的地域需求模型研究[J]. 统计与决策, 2008(16): 95-97.
- [4] 李再扬, 杨少华. 中国省级电信业技术效率: 区域差异及影响因素[J]. 中国工业经济, 2010(8): 129-139.
- [5] 薛声家, 王清. 基于 DEA 超效率模型的电信业效率及影响因素分析[J]. 北京邮电大学学报: 社会科学版, 2010, 12(5): 51-57.
- [6] 韩磊, 窦彩兰. 中国电信业区域经营效率和影响因素研究——基于两阶段的 DEA-Tobit 的方法的实证研究[J]. 消费导刊, 2010(2): 38-39, 41.
- [7] 郭艳春, 刘斐. 广东邮电业发展的需求效应分析[J]. 开放导报, 2012(4): 105-108.
- [8] 李清清, 李厚彪. 我国邮电业发展的影响因素分析[J]. 实验科学与技术, 2018, 16(2): 52-55.
- [9] 刘珉慧, 张辉, 张姝. 我国邮电通讯和经济发展关系探究[J]. 中国传媒大学学报(自然科学版), 2020, 27(3): 45-49.
- [10] 国家统计局. 中国统计年鉴 2021 [EB/OL]. <http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm>, 2022-07-06.
- [11] 何晓群. 多元统计分析[M]. 第四版. 北京: 中国人民大学出版社, 2015.
- [12] 吴喜之. 多元统计分析: R 与 Python 的实现[M]. 北京: 中国人民大学出版社, 2019.