

十余年来陕西经济发展轨迹及特征分析

——基于时序全局主成分的分析方法

曹 飞

(西安电子科技大学, 陕西 西安 710071)

摘 要:通过介绍主成分分析方法,选取陕西省 1995—2010 年第一产业增加值、第二产业增加值、工业增加值、第三产业增加值、固定资产投资、地方财政一般预算收入、人均地区生产总值、社会消费配零售总额、粮食产量、城镇居民人均可支配收入、农民人均纯收入 11 个指标数据,利用时序全局主成分分析方法动态描绘陕西经济发展历程。结果表明,动态轨迹与客观实际能很好地吻合,这一结果同时也为促进陕西经济进一步的快速发展提供了参考依据。

关键词:时序分析;全局主成分;经济发展;动态研究

中图分类号: O 212

文献标志码: A

文章编号: 1008-7192(2013)01-0027-05

An Analysis of the Economic Development Trajectories and Features of Shaanxi Province in Last Decates

——Based on the time series analysis and all around PCA

CAO Fei

(Xidian University, Xi'an 710071, China)

Abstract: This paper makes a comprehensive introduction of principal component analysis, and illustrates dynamically the economic development of Shaanxi province by means of the time series analysis and all around PCA. 11 index data of Shaanxi province from 1995 to 2010 have been chosen, including the added value of the primary industry, the added value of the secondary industry, the manufacture added value, the added value of the tertiary industry, the fixed assets investment, the budget revenue of local finance, the regional GDP, the total retail sale of consumer goods, the grain production, the per capita disposable income of urban residents and the net per capita income of rural residents. It turns out that as the concise dynamic trajectory goes with the reality identically, it can be regarded as a reference to further the economic development.

Key words: time series analysis; all around PCA; economic development; dynamic research

区域经济的发展水平是多因素(指标)综合作用的结果,因此,衡量一个地区的经济发展水平,应通过综合分析找出影响区域经济发展的主

要因素,为实现区域经济的均衡发展提供理论依据。为了能综合反映区域经济的发展水平,本文根据 2011 年《陕西省统计年鉴》统计资料,选取

收稿日期:2012-09-13

作者简介:曹 飞(1974-),男,陕西榆林人,西安电子科技大学人文学院副教授,博士,主要从事区域经济方面的研究。

了反映区域经济发展的九项主要指标,采用主成分分析法对陕西省 1995—2010 年的经济发展水平进行了综合评价。

一、主成分分析法理论介绍

1. 主成分分析法的基本思想^[1]

主成分分析是利用降维的思想,将多个变量转化为少数几个综合变量(即主成分),其中每个主成分都是原始变量的线性组合,各主成分之间互不相关,从而这些主成分能够反映始变量的绝大部分信息,且所含的信息互不重叠。

采用这种方法可以克服单一的经济指标不能真实反映地区的经济发展情况的缺点,引进多方面的经济指标,但又将复杂因素归结为几个主成分,使得复杂问题得以简化,同时得到更为科学、准确的区域经济发展信息。

2. 主成分分析法代数模型

假设用 p 个变量来描述研究对象,分别用 $X_1, X_2 \cdots X_p$ 来表示,这个变量构成的 p 维随机向量为 $X = (X_1, X_2 \cdots X_p)'$ 。设随机向量 X 的均值为 μ ,协方差矩阵为 Σ 。对 X 进行线性变化,考虑原始变量的线性组合:

$$\begin{cases} Z_1 = \mu_{11}X_1 + \mu_{12}X_2 + \cdots \mu_{1p}X_p \\ Z_2 = \mu_{21}X_1 + \mu_{22}X_2 + \cdots \mu_{2p}X_p \\ \cdots \cdots \cdots \\ Z_p = \mu_{p1}X_1 + \mu_{p2}X_2 + \cdots \mu_{pp}X_p \end{cases}$$

主成分是不相关的线性组合 $Z_1, Z_2 \cdots Z_p$, 并且 Z_1 是 $X_1, X_2 \cdots X_p$ 的线性组合中方差最大者, Z_2 是与 Z_1 不相关的线性组合中方差最大者, \cdots, Z_p 是与 $Z_1, Z_2 \cdots Z_{p-1}$ 都不相关的线性组合中方差最大者。

3. 主成分分析法基本步骤

第一步:设估计样本数为 n ,选取的经济指标数为 p ,则由估计样本的原始数据可得矩阵 $X = (x_{ij})_{n \times p}$,其中 x_{ij} 表示第 i 地区的第 j 项经济指标数据。

第二步:为了消除各项经济指标之间在量纲化和数量级上的差别,对指标数据进行标准化,得到标准化矩阵(系统自动生成)。标准化后的数据在原变量前加 z 表示。

第三步:根据标准化数据矩阵建立协方差矩阵 R ,是反映标准化后的数据之间相关关系密切程度的统计指标,值越大,说明有必要对数据进行主成分分析。其中, $R_{ij}(i, j = 1, 2, \cdots, p)$ 为原始变量 X_i 与 X_j 的相关系数。

第四步:根据协方差矩阵 R 求出特征值、主成分贡献率和累计方差贡献率,确定主成分个数。解特征方程 $|\lambda E - R| = 0$, 求出特征值 $\lambda_i (i = 1, 2, \cdots, p)$ 。因为 R 是正定矩阵,所以其特征值 λ_i 都为正数,将其按大小顺序排列,即 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \geq 0$ 。特征值是各主成分的方差,它的大小反映了各个主成分的影响力。主成分 Z_i 的贡献率

$$W_i = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}, \text{ 累计贡献率为 } \frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j}。 \text{ 根据选取主成分个数的原则,特征值要求大于 1 且累计贡献率达 } 80\% \sim 95\% \text{ 的特征值 } \lambda_1, \lambda_2, \cdots, \lambda_m \text{ 所对应的 } 1, 2, \cdots, m (m \leq p), \text{ 其中整数 } m \text{ 即为主成分的个数。}$$

主成分个数提取原则为主成分对应的特征值大于 1 的前 m 个主成分。如果特征值小于 1,说明该主成分的解释力度还不如直接引入一个原变量的平均解释力度大,因此一般可以用特征值大于 1 作为纳入标准。其优点是它确定的权数是基于数据分析而得到的指标之间的内在结构关系,不受主观因素的影响,而得到的综合指标(即主成分)之间彼此独立,减少信息的交叉,使得分析评价结果具有客观性和准确性^[2]。

第五步:建立初始因子载荷矩阵,解释主成分。因子载荷量是主成分 Z_i 与原始指标 X_i 的相关系数 $R(Z_i, X_i)$,揭示了主成分与各经济之间的相关程度,利用它可较好地解释主成分的经济意义。

第六步:计算地区经济综合评分函数 F_m ,并进行降序排列:

$$F_m = W_1Z_1 + W_2Z_2 + \cdots + W_iZ_i$$

二、陕西省经济发展时序全局主成分分析

1. 评价指标体系的建立

依照指标能客观、系统、全面反映地区经济

发展水平的原则,结合具体研究的问题,并考虑到数据的可获得性,本文选取了反映经济发展的 11 项主要指标:第一产业产值(x_1)、第二产业产值(x_2)、工业增加值(x_3)、第三产业产值(x_4)、固定资产投资(x_5)、地方财政一般预算收入(x_6)、人均国民产值(x_7)、社会消费配零售总额(x_8)、粮食产量(x_9)、农民人均纯收入(x_{10})、城镇居民人均可支配收入(x_{11})(表 1)。

表 1 陕西省 1995—2010 年基本经济指标原始数据

年份	x_1 (亿元)	x_2 (亿元)	x_3 (亿元)	x_4 (亿元)	x_5 (亿元)	x_6 (万元)	x_7 (元)	x_8 (亿元)	x_9 (万吨)	x_{10} (元)	x_{11} (元)
1995	217.27	441.67	377.91	377.91	324.33	513 011	2 965	398.87	913.40	963	3 310
1996	250.58	514.27	439.66	450.99	372.00	676 022	3 446	474.87	1 217.30	1 165	3 810
1997	255.42	567.25	477.35	540.93	424.10	841 178	3 834	552.37	1 044.40	1 285	4 001
1998	266.92	607.82	500.26	583.66	544.89	933 309	4 070	601.89	1 303.10	1 406	4 220
1999	254.57	681.43	549.01	656.64	619.27	1 064 033	4 415	657.80	1 081.60	1 456	4 654
2000	258.22	782.58	629.88	763.20	745.85	1 149 711	4 968	725.64	1 089.10	1 470	5 124
2001	263.63	878.82	706.62	868.17	850.66	1 358 109	5 511	809.31	976.61	1 520	5 484
2002	282.21	1 007.56	819.51	963.62	974.63	1 502 934	6 161	907.64	1 005.60	1 596	6 331
2003	302.66	1 221.17	1 006.92	1 063.89	1 278.72	1 773 300	7 057	1 010.95	968.40	1 676	6 806
2004	372.28	1 553.10	1 306.50	1 250.20	1 544.19	2 149 586	8 638	1 162.80	1 160.36	1 867	7 492
2005	435.77	1 951.36	1 650.63	1 546.59	1 982.04	2 753 183	10 674	1 331.35	1 139.50	2 052	8 272
2006	484.81	2 452.44	2 094.02	1 806.36	2 610.22	3 621 295	12 840	1 542.37	1 041.90	2 260	9 268
2007	592.63	2 986.46	2 544.42	2 178.20	3 642.13	4 752 398	15 546	1 837.25	1 067.91	2 645	10 763
2008	753.72	3 861.12	3 274.57	2 699.74	4 851.41	5 914 750	19 700	2 317.11	1 150.90	3 136	12 858
2009	789.64	4 236.42	3 501.25	3 143.74	6 553.39	7 352 704	21 947	2 699.67	1 131.40	3 438	14 129
2010	988.45	5 446.10	4 558.97	3 688.93	8 561.24	9 582 065	27 133	3 195.67	1 164.90	4 105	15 695

2. 主成分分析过程

首先对上述 16 年包含了 11 项评价指标的数据按全局主成分分析的要求进行整理,形成时序性立体数据表,再利用 SPSS17.0 统计分析软件进行全局主成分分析,得到 KMO 取样适当性度量、Bartlett 球形检验的统计参数值(表 2)。从表 2 可看出,KMO 检验值为 $0.759>0.5$,表明分析指标之间有较多的共同因素,数据适合主成分分析,Bartlett 球形检验中的显著性小于 $0.01(p=0.00)$,表明 Bartlett 球形检验拒绝单位相关的原假设,数据适合全局主成分分析。

表 2 KMO 与 Bartlett 检验结果

KMO 取样适度性检验		0.759
近似卡方分布		744.983
Bartlett 的球形度检验	自由度	55.
	显著性	0.000

其次,计算相关的系数矩阵(略)及其特征

值、贡献率和累计贡献率(表 3)^[3]。结果中每一新因子对应方差所占的百分比表示该变量在所

表 3 方差分解主成分提取分析 %

成份	初始特征值			提取平方和载入		
	合计	方差百分比	累积百分比	合计	方差的百分比	累积百分比
1	10.004	90.945	90.945	10.004	90.945	90.945
2	0.937	8.518	99.463			
3	0.039	0.358	99.821			
4	0.014	0.130	99.951			
5	0.003	0.029	99.979			
6	0.001	0.012	99.992			
7	0.001	0.005	99.997			
8	0.000	0.003	100.000			
9	4.894E-5	0.000	100.000			
10	1.904E-6	1.731E-5	100.000			
11	4.260E-7	3.873E-6	100.000			

有变量中的相对地位以及对系统综合评价的贡献,称为贡献率。从表 3 得到的特征值相关矩阵可知,第一个特征值的贡献率达到 90.945%,满足选取主成分个数指标的大于等于 85%的要求,且第二主成分的特征值小于 1,因此不再提取主成分。故选择一个主成分即可,且这两个主成分

基本上保留了原来指标 $x_1、x_2、x_3、x_4、x_5、x_6、x_7、x_8、x_{10}、x_{11}$ 的信息。

提取方法:主成分分析。

由于提取了一个主成分,所以不需要进行旋转,利用各主成分载荷值除以主成分特征值的平方根,得到特征向量(表 4)。

表 4 因子载荷矩阵表及特征向量

项目	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
主成分因子载荷	0.997	0.999	0.998	0.998	0.991	0.997	0.999	0.998	0.273	0.998	0.991
主成分特征向量	0.3152	0.3158	0.3155	0.3155	0.3133	0.3152	0.3158	0.3155	0.086	0.3155	0.3133

3. 各年份经济发展综合价值及时序分析

从表 4 可见,第一主成分与粮食产量(x_9)相关系数为 0.273,但粮食产量的重要性在第一产业产值及人均国民生产总值都有体现,因此影响不大。由于提取了一个主成分,所以不再进行加权计算。

$$F = F_1 = 0.3152 \times zx_1 + 0.3158 \times zx_2 + 0.3155 \times zx_3 + 0.3155 \times zx_4 + 0.3133 \times zx_5 + 0.3152 \times zx_6 + 0.3158 \times zx_7 + 0.3155 \times zx_8 + 0.086 \times zx_9 + 0.3155 \times zx_{10} + 0.3133 \times zx_{11}$$

其中, ZX_i 是原始变量 X_i 的标准化,其均值为 0,方差为 1。具体计算结果如表 5 所示。

表 5 陕西省 1995—2010 年基本经济发展综合价值(F)及年度绝对增幅(ΔF)

年份	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10
时序	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
F	-3.16	-2.616	-2.582	-2.202	-2.247	-2.045	-1.925	-1.6	-1.272	-0.492	0.234	1.011	2.254	4.043	5.175	7.422
ΔF		0.544	0.034	0.38	-0.045	0.202	0.12	0.325	0.328	0.78	0.726	0.777	1.243	1.789	1.132	2.247

三、结 语

利用表 5 的数字得图 1,可以更直观地反映陕西省经济发展的时序过程。从图 1 可见,1995—2010 年陕西经济发展的时序大致可以分为三个阶段,主要是“九五”、“十五”、“十一五”时期。

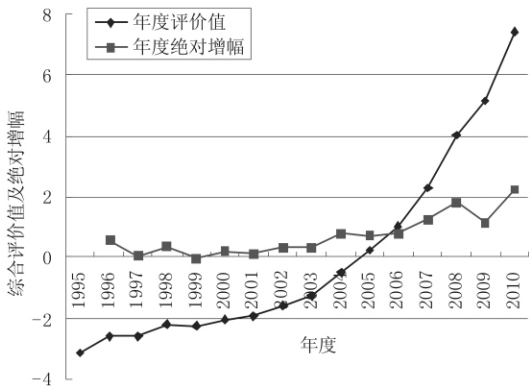


图 1 陕西省 1995—2010 经济发展
综合评价值及年度绝对增幅

1995—1999 年(主要是“九五”时期)是陕西省的缓慢发展期,各年份的综合评价得分达 -3.16~-2.247,经济增长幅度也在 0.4 以下,平均年绝对增幅为 0.228。这主要是由于 20 世纪 90 年代末和 21 世纪初陕西省高投入、高消耗、高污染、低效益的产业层次和结构主导的粗放增长方式,附加值低、可持续性差。另外国有企业效益低下,非公有制经济还没有充分发展起来,基础设施相对薄弱,产业结构不合理,地域分工不明显,这些都制约了该省经济发展。

2000—2005 年(主要是“十五”时期)是陕西经济发展的较快时期。各年份的综合评价得分达 -1.92~0.234,经济健康、持续发展,年均绝对增长幅度 0.423,这主要得益于国家西部大开发战略支持和市场经济持续发展,国有企业改革取得的成果。“十五”时期,陕西经济总量迈上新台阶,生产总值年均增长 11.5%,高出“九五”时期年均增长速度 0.7 个百分点;经济结构调整实

现新突破,工业经济的主导地位进一步显现;六大特色产业增势强劲,四大基地建设步伐加快,特色鲜明的区域经济发展格局初步形成;国有大企业的股份制改革稳步推进,城乡居民生活达到新水平。

2006—2011 年(“十一五”时期)是陕西省经济持续、健康、快速发展的时期。各年份的综合评价得分达 1.011~7.422,年绝对增幅在 0.7 以上,平均增幅为 1.258,为最好最快的时期。“十一五”时期,全省综合实力大幅提升,人民生活显著改善,可持续发展能力不断增强。和“十五”时期相比,“十一五”经济数据揭示的不仅是量的全面扩张,更是质的科学提升。具体体现在:实力增强,GDP 在全国位次前移 3 位,财政收入增量创历史最多时期;质量提高,人均生产总值实现历史新突破,人均生产总值提前两年完成目标,突破 4000 美元;速度加快增速创历史最快和持续时间最长时期,生产总值增速保持 13% 以上,年均增长 14.8%,比“十五”时期加快 2.9 个百分点;百姓富裕居民收入增量为历史最高,城镇居民可支配收入与农村居民人均纯收入显著增加;工业更强主导地位进一步凸现,工业占经济总量比重超过“十一五”规划 43% 的目标;固定资产投资增长迅速,“十一五”时期,全省累计完成投资

达 2.6 万亿元,相当于“十五”时期的 4 倍,年均增长 34.4%,比“十五”时期的 19.8% 加快 14.6 个百分点;三次产业结构明显变化,区域经济发展的协调性进一步增强,关中经济区规划上升为国家战略,率先发展步伐加快,陕南通过发展循环经济促进突破发展的成效越来越凸现,陕北在“三个转化”战略的带动下,实现了新的跨越,成为全省经济重要增长极。

通过上述分析说明,以主成分分析方法所绘出的曲线对陕西经济发展的描述与实际情况基本吻合。使用时序全局主成分分析对描述陕西省经济发展动态过程是成功的。在结合历史的基础上,对该曲线的理解可以增加决策者的系统知识,提高其洞察力,对发展规划和相关政策措施的制订提供理论支持^[4]。在社会主义和谐社会理论和科学发展观的指导下,陕西省应充分利用好国家对西部支持的优惠政策,不断推进产业结构调整,充分发挥资源优势 and 区位优势,实现人均 GDP 和地方财政一般预算收入增加;通过完善就业结构与加快农村剩余劳力转移,增加城乡居民收入,扩大内需;通过优化投资结构,改善投资环境,实现良好的投资效益;并最终促进经济持续、快速、健康发展。

参 考 文 献

- [1]朱星宇,陈勇强. SPSS 多元统计分析方法及应用[M]. 北京:清华大学出版社,2011:241.
- [2]邱东. 多指标总体评价方法的系统分析[M]. 北京:中国统计出版社, 1991:72-89.
- [3]薛薇. 基于 SPSS 的数据分析[M]. 北京:中国人民大学出版社, 2006:359-377.
- [4]乔峰,姚俭. 时序全局主成分分析方法在经济发展动态描绘中的应用[J]. 数理统计与管理, 2003,22(2):1-5.