

中国省际间人力资本流动原因的隐变量分析

刘娟

(首都经济贸易大学 统计学院, 北京 100026)

摘要: 自改革开放以来中国发生了大规模的省际间人力资本流动, 在人力资本流动的原因中还存在着一些不能直接测量的因素如自然生存条件的影响、传统文化的影响、社会发展水平的影响等, 如何对这些影响进行分析是本文研究的主要问题。因此, 笔者引入多个方程的因果关系模型, 即建立包含隐变量的结构方程模型 (SEM), 评估影响人力资本流动的各个因素及其与各个隐变量之间的联系, 从而确定影响人力资本流动的原因。

关键词: 人力资本流动; 隐变量分析; 结构方程;

中图分类号: C631.42 **文献标识码:** A **文章编号:** 1000-4149 (2007) 02-0035-05

Latent Variable Analysis of Provincial_Level Migration of Human Capital in China

LIU Juan

(School of Statistics, Capital University of Economics and Business, Beijing, China, 100026)

Abstract: Since early 1980s, the first stage of China economic reform, China has witnessed large_scale migration of human capital. There are many reasons for people to move from one province to another. This paper mainly looks at some implicit factors such as individuals' natural living condition, their traditional cultural values, etc, by using SEM so as to find the relationship between these factors in terms of their impact on individuals' intention of migration.

Keywords: migration of human capital; latent variable analysis; SEM

自上个世纪 90 年代以来中国发生了大规模的省际间人力资本流动, 有些学者利用 Logistic 回归模型对产生人力资本流动的原因进行了分析, 但是这些方法有些局限性: 一是这些方法仅单独在一个维度中讨论其相应指标的影响力, 没有充分考虑各个综合指标之间的相互作用, 从而不能客观地反映现实各个因素对人力资本流动影响程度的强弱; 二是这些方法仅对能够测量的经济发展水平指标如人均 GDP、对外开放度等进行分析, 但是对人力资本流动的原因中存在的一些不能直接测量的因素如自然生存条件的影响、传统文化的影响、社会发展水平的影响等不能进行分析。而这些影响因素对人力资本是否流动以及流入地的选择等可能会有决定性的作用。

鉴于人力资本流动的复杂性, 笔者试图引入多个方程的因果关系模型, 即建立包含隐变量的结构方程模型, 以便充分提取原始显变量的信息, 评估影响人力资本流动的各个因素及其与各个隐变量之间的联系, 从而确定影响人力资本流动的原因。

收稿日期: 2006-07-13

作者简介: 刘娟 (1963-), 女, 北京人, 经济学博士, 首都经济贸易大学统计学院教授, 主要从事统计方法应用和调查研究。

结构方程模型（SEM）是一种既考虑因子内部结构，同时又注重因子之间的因果关系的多变量测量模型。其思想起源于路径分析概念。有人称其为联立方程模型、因果模型等等。结构方程模型发展过程中较大的一个突破就是发展了隐变量的概念，并且融合了因子分析（factor analysis）以及路径分析（path analysis）两种统计技术^[1]。目前普遍应用于社会学、经济学、心理学等多种学科和社会调查领域。对于包含隐变量的结构方程模型，可利用 LISREL（Linear Structural Relationships）软件实现。

LISREL 方法的原理是，用样本数据对所设定的模型参数进行估计，再根据这些参数估计来重建方差协方差，然后尽可能地将重建的方差协方差矩阵与观测的方差协方差矩阵相匹配，它们的匹配程度就决定模型拟合数据的程度。常用的拟合指数主要有：绝对拟合指数、相对拟合指数和简约拟合指数。

一、模型设定

结构方程应用的是否成功关键要看其是否有理论基础，即结构方程主要应用于验证已知的理论模型。因为结构方程模型主要是一种证实性技术，而不是一种探测性技术^[2]，这就是说，尽管结构方程模型分析中也涉及到一些探测性的因素，但研究者主要是通过应用结构方程模型来确定一个特定模型是否合理，而这个特定模型就应该是在已有的理论基础之上建立起来的。

正因如此，根据有关人力资本流动原因的理论和有关分析结论，我们设定影响人力资本流动的因素为：生存资源禀赋、经济活力程度、社会发展水平和科学文化教育资源。由于这些因素不是用一个指标就能直接测量出来的，而是需要通过许多相关的指标共同来反映，因此这些因素是研究中的隐变量。

选择何种指标来测量这些隐变量，是研究的第一个问题，也是非常关键的问题。对于上述的每一个隐变量都可从不同的角度采用多个变量来反映，如经济活力程度隐变量可从经济发展规模、吸引内外资能力以及经济发展潜力等多个方面来评价，而每个方面又可以采用多个有关的指标来测量。而如果直接采用现有的各种指标来测量这些隐变量，这个模型中的测量变量就会多达几十个甚至上百个，模型就会变得非常复杂也很难识别，并且有些需要的指标数据很难获取。因此本文决定利用一些评价指标从多个角度来测定隐变量，这些评价指标来自于中国科学院可持续发展研究组发布的《2002 年中国可持续发展战略报告》（科学出版社）^[3]，报告中采用了 219 个指标建立了可持续发展指标体系，对 45 个评价指数进行定量描述，我们从中选取有关的评价指数进行分析。该指数体系中的每个指标来源于 1997~ 2000 年的各种统计年鉴和有关数据，所以与我们研究的时间范围正好吻合。

对隐变量选取的评价指数和每个评价指数所包含的指标内容如表 1 所示：

表 1 隐变量选取的评价指数和指标内容		
隐变量	评价指数	包含的指标
生存资源禀赋 (ξ_1, η_1)	土地资源指数 (x_1, y_1)	人均耕地、耕地质量、耕地占全国份额 人均水资源、水资源密度 光合有效辐射、年平均降水、年均霜日
	水资源指数 (x_2, y_2)	
	气候资源指数 (x_3, y_3)	
经济活力程度 (ξ_2, η_2)	经济吸引指数 (x_4, y_4)	外资占全国份额、外资占本地 GDP 比例、进出口总额占全国份额、外贸依存度、单位面积货源周转量、每万人邮电业务总量、千人拥有电话数、百户拥有个人电脑数、GDP 占全国份额、人均 GDP、GDP 密度、GDP 的增长率
	基础设施能力 (x_5, y_5)	
	经济规模指数 (x_6, y_6)	
社会发展水平 (ξ_3, η_3)	生活质量指数 (x_7, y_7)	城乡人均收入水平、城乡人均消费支出、恩格尔系数等 城镇失业率、贫困发生率 赡养比、社会保障覆盖率
	社会安全指数 (x_8, y_8)	
	社会保障指数 (x_9, y_9)	
科学文化教育资源 (ξ_4, η_4)	教育投入指数 (x_{10}, y_{10})	教育经费占 GDP 比例、教育经费占全国份额 万人中等学校在校学生数、在校大学生数、拥有中等、大学教师数 万人拥有科技人员数、R&D 经费占 GDP 比例、地方科技事业费、科技三费占财政支出比例等
	教育规模指数 (x_{11}, y_{11})	
	科技资源指数 (x_{12}, y_{12})	

表 1 中的 x_i 与 y_i 分别代表人力资本流出地各项指数（即外生变量）和人力资本流入地的各

项指数 (即内生变量), ξ_i , η_i 分别表示外生隐变量和内生隐变量。

根据以上变量的选取, 我们构建了本研究的路径图 (由于篇幅所限略)。鉴于本研究的目的是验证并量化外生隐变量对内生隐变量 (人力资本流动因素) 的影响, 而外生变量之间的关系不在本次研究范围之内, 同时考虑到生存资源禀赋对人力资本流动中的经济活力程度、社会发展水平和科学文化教育资源的影响不大, 所以它们之间的路径关系省去。

二、模型识别性

至今为止没有一个能保证结构方程模型收敛的充要条件。但有一些充分或必要条件可以为我们构造模型提供方便。本研究的模型识别性主要是从两个方面考虑: 测量部分和结构部分。按照模型识别的两指标法则: 每个因子有两个或两个以上的指标; 每个指标只测量一个隐变量; 每个隐变量至少有另一个隐变量与之相关。只要满足以上条件, 模型就是可识别的 (但不是必要条件)。模型的测量部分满足以上条件, 因此是可识别的。非递归因果模型都是可识别的, 结构部分也是可识别的。另外, 给每个隐变量规定了测量单位, 即指定每一个隐变量的任意一个指标变量的负载为 1。

对于测量模型的前两个条件很好判断, 我们建立的模型已经满足要求, 但是对每个隐变量至少有另一个隐变量与之相关就不能直接判别, 而需要进行验证性的因子分析模型^[4]。首先对内生隐变量 (即流入地的隐变量) 进行分析, 结果如图 1 和表 2 所示。

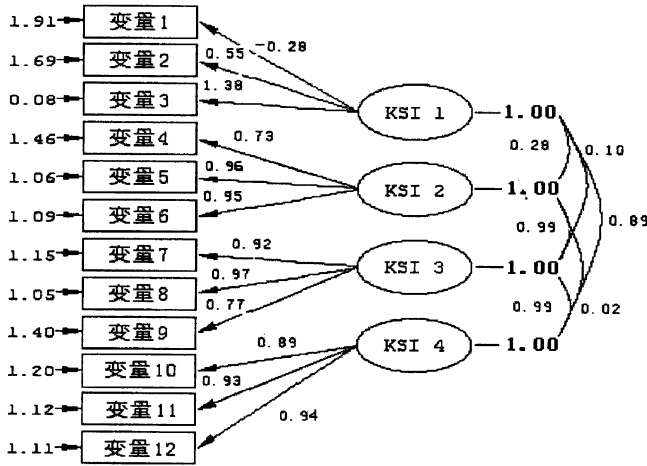


图 1 内生隐变量的因子分析模型图

表 2 内生隐变量的参数估计结果

	KSI 1 (η_1)	KSI 2 (η_2)	KSI 3 (η_3)	KSI 4 (η_4)
KSI 1 (η_1)	1.00			
KSI 2 (η_2)	0.28 (0.10) 2.71	1.00		
KSI 3 (η_3)	0.10 (0.09) 1.09	0.99 (0.06) 18.02	1.00	
KSI 4 (η_4)	0.02 (0.09) 0.26	0.89 (0.06) 14.90	0.99 (0.05) 18.78	1.00

注: 每个参数度对应三个值, 第一个是 (未标准化的) 参数估计值, 第二个 (括号内的) 是标准误, 第三个是 t 值。

从表 2 可知 KSI 1 与 KSI 2 显著相关, KSI 2 与 KSI 3、KSI 4 都显著相关, KSI 3 与 KSI 4 也显著相关。因此满足了每个隐变量至少有另一个隐变量与之相关的条件。从验证性的因子分析模型的模拟结果看, $\chi^2 (173) = 67.19$, RMSEA= 0.047, NNFI= 0.95, CFI= 0.97, 说明模型拟和的相当好。

另外，从隐变量与每个测量变量的因子载荷和显著性检验来看（即与隐变量参数估计输出结果类同的输出结果，这里省略），每个隐变量与每个测量变量都有显著的因子关系，即我们所选的评价变量都能比较好地测量每个评价变量。

以同样的方法再对外生隐变量（即流出地的隐变量）进行分析（表略），可知 KSI 1 与 KSI 2 显著相关，KSI 2 与 KSI 3、KSI 4 都显著相关，KSI 3 与 KSI 4 也显著相关。因此也满足每个隐变量至少有另一个隐变量与之相关的条件。同样从验证性的因子分析模型的模拟结果看， $\chi^2(173) = 69.86$ （自由度= 48），RMSEA= 0.048，NNFI= 0.94，CFI= 0.95，说明模型拟和的相当好。

另外，从隐变量与每个测量变量的因子载荷和显著性检验来看（即与隐变量参数估计输出结果类同的输出结果，这里省略），每个隐变量与每个测量变量都有显著的因子关系，即我们所选的评价变量都能比较好地测量每个评价变量。

三、估计方法和模型评价

虽然本研究用的数据是来自中国科学院可持续发展研究组发布的《2002 年中国可持续发展战略报告》中有关的评价指数，但是仍然要确定人力资本的流出地和流入地。由于结构方程中没有直接使用人力资本流入率或流出率的测量指标，所以就要求在研究中的样本数据应该是从流出地*i* 流向流入地*j* 的流出率比较高的流出与流入地区，因此本文根据第五次人口普查中 1%抽样调查原始数据，并根据人力资本流动选择指数公式^①计算出各地区的流动选择指数，当流动选择指数大于 100 时，说明地区*i* 人力资本选择流向地区*j* 的流动选择强度超过全国平均水平。以此作为标准进行流出地与流入地的选择确定，最后样本量确定为 173 个。

对模型的评价目前还没有统一的指标，通常评价拟合指数的原则有： χ^2 除以自由度在 2~ 5 之间，RMSEA 在 0.08 以下（越小越好），NNFI 和 CFI 在 0.9 以上（越大越好），所拟合的模型是一个“好”模型^[5]。

本文最终模型中拟合指数为： $\chi^2(173) = 484$ （自由度= 241），RMSEA = 0.079，NNFI= 0.79，CFI= 0.82，虽然 NNFI 和 CFI 没有达到 0.9 以上，但是 χ^2 检验和 RESEA 都符合好模型的标准，因此模型拟合的结果还是可接受的。

从观测变量参数的显著性看（见表 3），有两个参数的 *t* 值小于 1.96，不显著。但在最终模型中保留这部分不显著参数，其原因是为了模型的对称性以及它们的作用在理论上认为确实存在。造成这种数据和理论不完全吻合的原因可能是由于我们用于分析的样本量偏少或指标选取上还有待完善，这一点需要后续研究来证实。综合以上各个方面的评价，我们认为可以接受最终模型为研究用的模型。

表 3 观测变量参数的显著性检验结果

变量	1	2	3	4	5	6	7	8	9	10	11	12
Y	- 0.39	1.00	1.10	1.00	1.50	1.45	1.00	1.06	0.84	1.00	1.06	1.05
	(0.07)		(0.30)		(0.27)	(0.27)		(0.14)	(0.13)		(0.15)	(0.15)
	- 2.21		3.68		5.47	5.39		7.62	6.25		6.90	6.84
X	- 0.49	1.00	1.14	1.00	1.04	0.11	1.00	0.51	0.96	1.00	0.90	1.13
	(0.19)		(0.30)		(0.21)	(0.23)		(2.34)	(2.89)		(0.24)	(0.32)
	- 2.62		3.74		4.90	4.84		0.22	0.33		3.80	3.58

注：每个参数度对应三个值，第一个是（未标准化的）参数估计值，第二个（括号内的）是标准误，第三个是 *t* 值。

① 人力资本流动选择指数:

$$I_{ij} = \frac{M_{ij}}{(P_i/P_t)(P_j/((P_t - P_i)) \sum_j} K$$

式中， I_{ij} 为地区*i* 的人力资本选择流向地区*j* 的选择指数， M_{ij} 为地区*i* 选择流向地区*j* 的人力资本数， P_i 为地区*i* 的人力资本总量， P_j 为地区*j* 的人力资本总量， P_t 为全国人力资本总量，*K* 为一常数（通常以 1 或 100 表示，本文以 100 表示）。如果 I_{ij} 大于 100，说明地区*i* 人力资本选择流向地区*j* 的流动选择强度超过全国平均水平； I_{ij} 越大，表示地区*i* 人力资本选择流向地区*j* 的流动选择强度越大或流入地*j* 吸引流出地*i* 流出人力资本流入本地的作用越强。

最终模型的路径图如图 2 所示:

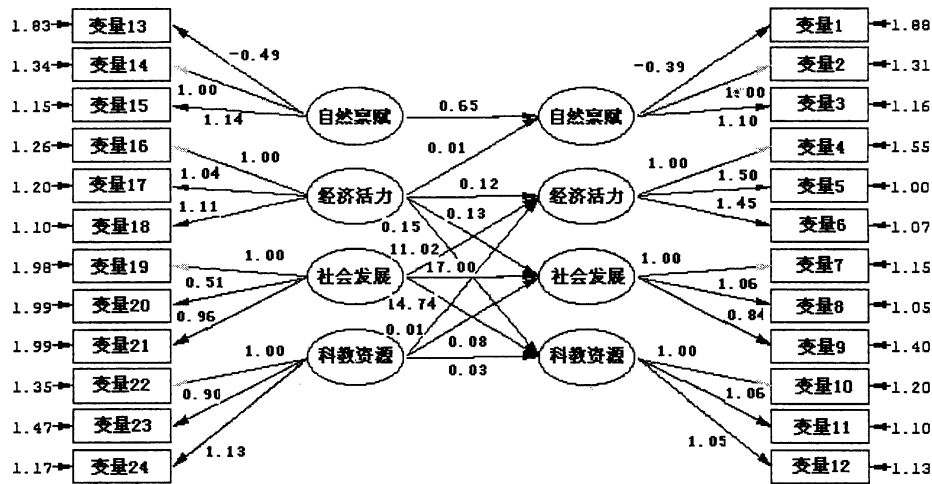


图 2 人力资本流动原因分析路径图

四、结果分析与结论

由于结构方程模型可识别的必要条件之一是为模型中的每个隐变量建立一个测量尺度^[9]。建立测量尺度的常用方法是将隐变量的观测标识中任何一个的因子负载 λ 设定为一常数, 通常为 1。因此我们在该模型中的每个隐变量设定一个观测标识的因子负载为 1, 在该隐变量中的其他观测标识的因子负载是相对于因子负载为 1 来测定的。

从隐变量观测标识的因子负载来看, 只有生存资源禀赋中的土地资源指数一个观测标识的因子负载为负的, 其余全部为正向的。土地资源指数的因子负载为负, 说明土地资源越贫乏的地区, 人力资本越倾向于流出, 虽然流入地的土地资源指数因子负载也为负值, 但是流入地的土地资源指数因子负载小于流出地的因子负载。此外, 大部分流入地隐变量观测标识的因子负载大于相应的流入地的隐变量观测标识的因子负载, 显示了流入地的各项观测标识相对于流出地来说, 都有更好的表现。

我们进行结构方程模型分析的主要目的是要观察各隐变量之间的关系, 即人力资本流出地的各隐变量是如何影响流入地的隐变量, 从而了解人力资本流动的各种动因。从反映隐变量之间关系的 Γ 系数可以观察隐变量的关系 (见表 4)。首先根据参数估计的显著性检验来确定哪些隐变量之间存在着显著影响作用。流出地生存资源禀赋 $\rightarrow$ 流入地生存资源禀赋, 流出地经济活力程度 $\rightarrow$ 流入地经济活力程度、流出地经济活力程度 $\rightarrow$ 流入地社会发展水平和流出地经济活力程度 $\rightarrow$ 流入地科技文化教育水平等是显著的, 而其他的隐变量之间没有显著影响作用。

表 4 反映隐变量之间关系的 Γ 系数

	自然禀赋 ξ_1	经济活力 ξ_2	社会发展 ξ_3	科教资源 ξ_4
自然禀赋 η_1	0.65 (0.08) 7.84	0.01 (0.05) 0.24	—	—
经济活力 η_2	—	0.12 (0.04) 3.12	11.02 (9.56) 1.15	0.01 (0.04) 0.36
社会发展 η_3	—	0.13 (0.05) 2.54	17.00 (14.72) 1.15	0.08 (0.06) 1.35
科教资源 η_4	—	0.15 (0.05) 3.13	14.74 (12.77) 1.15	0.03 (0.05) 0.52

注: 每个参数度对应三个值, 第一个是 (未标准化的) 参数估计值, 第二个 (括号内的) 是标准误, 第三个是 t 值。
(下转第 34 页)

加强与世界各国工会的联合与合作；积极推进劳动法制建设，配合政府劳动执法监察工作，同时对政府部门的工作进行监督。

5. 关注社会道德构建，从社会价值观和道德意识层面，为改善劳动者就业质量提供支持

近年来，我国政府已开始采取类似西方干预主义阶段的措施来化解劳资矛盾，但效果并不明显，重要原因之一就是缺少社会观念和意识的支持。在“重资本轻劳动”和利润崇拜观念的支配下，无论是政府还是企业，对保护劳工权益措施的执行力度都会大打折扣。因此，要在社会意识层面构建“劳动创造价值”和“以人为本、尊重劳动”的社会价值观，形成全社会共同关注劳动者就业质量的氛围，形成

政府、企业、劳动者与社会意识的良性互动，共同促进就业质量的提升。

参考文献：

[1] 郑功成. 经济全球化下的强资本弱劳工格局及其对我国的影响. 劳动经济与劳动关系, 2005, (5).
[2] 国际劳工组织. 劳工市场主要指标. 北京：中国劳动社会保障出版社, 2005
[3] 常凯. WTO、劳工标准与劳工权益保障. 中国社会科学, 2002, (1).
[4] 韩德强. 碰撞——全球化陷阱与中国现实选择. 北京：经济管理出版社, 2000
[5] 曾湘泉、李丽林. 我国劳动力市场中的就业政策支持. 中国人民大学学报, 2003, (1).
[6] 深圳市社会保障局、深圳市政府政策研究室. 深圳应推进企业履行社会责任. 经济前沿, 2006, (2).

[责任编辑 崔凤垣]

(上接第 39 页)

因此可以得到以下几点结论：

第一，流出地的生存资源禀赋条件是导致人力资本流动的主要原因，而在生存资源禀赋条件中起到“推力”作用的又主要是土地资源的缺乏，即大量的剩余劳动力是人力资本流动的最主要因素。

第二，经济活力程度是人力资本流动的又一重要原因，这种影响不仅体现在流出地对流入地的经济活力程度之间，而且还对流入地的社会发展水平和科学教育水平也有影响。这是因为经济活力程度与社会发展水平、科学教育水平之间是有相互关系的，而这种关系不是我们研究关心的问题，所以忽略了。

第三，无论是流出地的社会发展水平还是科学教育水平对流入地的各项水平都没有显著影响，说明流出地的社会发展水平和科学教育水平都不是引起人力资本流动的主要原因。为什么会有这样的结果，其实并不奇怪，因为在人力资本流动中只有初等教育程度的人力资本占总流动人数的 3/4，即绝大部分流动的人力资本是来自于仅受过初中及以下教育的农村劳动力，他们流动的最主要原因：一是人多地少，有剩余的劳动力；二是为了追求更多的经济利益，能为自己及家人带来更多的收入。从笔者对人力资本流动原因的分析中可知^①，接受过高等教育的人力资本流动的原因与初等素质人力资本流动的原因有很大差别：经济原因不再是高等素质人力资本流动的显著原因，而教育科学发展水平是影响他们流动的根本因素，这些问题有待进一步的研究。

参考文献：

[1] 黄芳铭. 结构方程模式 [M]. 北京：中国税务出版社, 2005. 2.
[2] 郭志刚. 社会统计分析方法——SPSS 软件应用 [M]. 北京：中国人民大学出版社, 1999. 341.
[3] 中国科学院可持续发展研究组. 2002 年中国可持续发展战略报告 [M]. 北京：科学出版社, 2003. 255~ 285.
[4] 候杰泰. 结构方程模型及其应用 [M]. 北京：教育科学出版社, 2004. 30.
[5] 同 [4].
[6] 同 [2].

[责任编辑 崔凤垣]

^① 刘娟. 人力资本流动理论与实证. 博士论文第四章, 2006。