Advances in Applied Mathematics 应用数学进展, 2023, 12(5), 2429-2436
Published Online May 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.125245

Hans汉斯

中国语言多样性与生物多样性的建模分析

王春霞

青岛大学数学与统计学院，山东 青岛

收稿日期：2023年4月28日；录用日期：2023年5月21日；发布日期：2023年5月30日

摘　要

中国的汉语方言多样性和生物多样性具有区域相关关系。生物多样性和汉语方言多样性之间的关联性可能是由于一些公共因素的影响，本文验证了公共因素与中国各省的植被覆盖率和面积有关，建立结构方程模型综合地分析在去除公共因素影响后，由多指标构建的生物多样性和汉语方言多样性这两个潜变量之间的关系，并采用Lasso、MCP惩罚参数方法和Amos基于指定搜索方法进行模型选择。研究结果表明在去除面积、植被等公因子的影响后两者之间呈微弱的负相关关系。因此，我们要维护好自然环境，促进生物多样性与语言多样性的协同发展。

关键词

汉语方言多样性，少数民族语言，生物多样性，结构方程模型

Modeling and Analysis of Language Diversity and Biodiversity in China

Chunxia Wang

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Apr. 28th, 2023; accepted: May 21st, 2023; published: May 30th, 2023

Abstract

There is a regional correlation between the diversity of Chinese dialects and biodiversity in China. The correlation between biodiversity and Chinese dialect diversity may be due to the influence of some public factors. This paper verifies that public factors are related to the vegetation coverage and area of various provinces in China. A structural equation model is established to comprehensively analyze the relationship between the two latent variables, biodiversity and Chinese dialect diversity, constructed by multiple indicators after removing the influence of public factors. And using Lasso, MCP penalty parameter method, and Amos based on specified search method for model

文章引用: 王春霞. 中国语言多样性与生物多样性的建模分析[J]. 应用数学进展, 2023, 12(5): 2429-2436.
DOI: 10.12677/aam.2023.125245

selection. The research results show that there is a weak negative correlation between the two after removing the impact of common factors such as area and vegetation. Therefore, we need to maintain the natural environment and promote the coordinated development of biodiversity and linguistic diversity.

Keywords

Diversity of Chinese Dialects, Minority Languages, Biodiversity, Structural Equation Model

1. 引言

语言多样性与生物多样性在世界范围内呈区域相关性[1]。全球现存近 7000 种语言里，超过 4800 种位于生物多样性高度丰富的地区[2]。近代工业化、城市化、人口迁移和文化冲击等加剧了生物多样性的丧失，同时也引发语言的衰落与消亡[3]。语言学家应用生物与环境的相互作用原理，分析生物多样性和语言多样性的关联及形成机制、语言兴亡变化的影响因素，形成新的交叉学科——语言生态学[4]。语言生态学中，语言－生物多样性的相关性分析多基于环球或洲际的国家[5]、岛屿[6]、生物多样性热点地区的语种[2]。中国幅员辽阔、地形复杂、气候多样，具备较高的生物多样性。汉语又是世界上使用人口最多的语言各地汉语方言在词汇、语音和语法上的差异也体现了高度的多样性。张等人分析了中国 33 个省级行政区动植物种类，与数百个汉语方言音类、音值、词汇、语法和句法的表达种类等数据，验证了中国生物多样性与汉语方言多样性的地域相关性[7]。

关于语言、生物多样性的相关性分析，目前多数研究停留在语言指标与生物指标一对一式的计算和检验上[2] [7]。一方面，两种多样性都是抽象的、综合性的概念，都通过多种包含噪声与误差的外在指标度量和体现[8]，也都不等同于任何单一指标本身。另一方面，语言多样性与生物多样性指标层面上的正相关，也有气候、面积等宜居条件、客观因素共同作用的影响，并不完全体现语言与生态的互动机制。本文基于这两方面的考虑，利用结构方程模型[9]，以语言多样性、生物多样性及共性影响因素为潜变量，将指标间一对一的相关性计算，升级为多对多的同步建模，得出更细致的关联性刻画。为避免结论矛盾的结构模型，我们参照多种模型选择方法，有助于深化对生物多样性和语言多样性的关系及形成机制的认识，有助于协助和促进生态保护、保护生物多样性和语言多样性。

2. 研究设计

本文利用 box-cox 变换对不符合正态性要求的变量删除，本章数据来源于《Investigation on the Relationship between Biodiversity and Linguistic Diversity in China and Its Formation Mechanism.》(Zhang, X. Public Health 2022)，经数据处理后发现少数民族语言种数和入声的分化区域组合种数不符合正态性要求。因此，用汉语方言片数、汉语方言语音值变化的总种数、203 个汉语方言词汇的不同表达方式总数、102 个语法词、词法和句法的不同汉语方言表达方式总数、“爸爸”的汉语方言表达方式数、8 个亲属称谓重叠式的区域组合种数这 6 个变量来衡量潜变量汉语言多样性，动物种数、植物种数、生物多样性指数用来衡量潜变量生物多样性，所有的 9 个变量用来衡量公共因素，即潜变量公因子。

2.1. 模型构建

生物多样性和汉语言多样性有区域相关性，有多种衡量指标，两者之间的关系可能存在植被覆盖率、面积等公因子的影响。基于以上结论，本文提出在去除公因子影响下，生物多样性影响汉语言多样性的假设。通过上述度量指标构建生物多样性、汉语言多样性以及共性因素这三个潜变量，研究在公因子约束下的生物多样性与汉语言多样性的关系，结构方程模型的表达式如下：

$$\eta = \Gamma\xi + \zeta \tag{1}$$

$$X = \Lambda_y\xi + \varepsilon \tag{2}$$

$$Y = \Lambda_x\eta + \delta \tag{3}$$

其中 $\eta = (\eta_1, \Delta)'$，$\xi = (\xi_1, \Delta)'$，η_1 和 ξ_1 分别代表汉语言多样性和生物多样性这两个潜变量，结构系数矩阵 Γ 中的元素表示内生潜变量 η 为外源潜变量 ξ 的线性函数，Δ 是公因子。Y 为 6 × 1 维向量，X 为 3 × 1 维向量，Y 和 X 分别是衡量语言多样性和生物多样性的观察标识，Λ_y、Λ_x 为系数矩阵，ζ、δ 和 ε 是误差因子向量。

2.2. 模型拟合检验

以上模型的拟合检验结果(见表 1)表明，卡方/自由度为 1.06，符合数值小于 3 的要求，p 值为 0.376，大于 0.05，表明模型是可被接受的，相异性指标 RMSEA 等于 0.048，小于理想标准 0.08，相似性指标 CFI 和 TLI 均大于 0.9，达到理想数值的要求。这表明模型与数据的拟合效果良好，因此可开展关于生物多样性与汉语言多样性关系之间的深入研究，并得到模型路径的分析结果(见表 2)和结构方程模型的路径图如图 1 所示。

Table 1. Model fit test
表 1. 模型拟合度检验

检验指标	模型数值	检验标准	结论	参考标准
P 值	0.376	>0.05，不显著	模型通过	
卡方/DF	1.06	<3，优秀	拟合良好	Hayduck, 1987 [10]
CFI	0.997	> 0.8 可接受；>0.9 拟合良好	拟合良好	Bagozzi & Yi, 1988 [11]
TLI	0.994	>0.8 可接受；>0.9 拟合良好	拟合良好	Bagozzi & Yi, 1988
RMSEA	0.048	<0.08，优秀；<0.1 可接受	拟合良好	

Table 2. Model path analysis results
表 2. 模型路径分析结果

假设	UnStd.	SE.	z-value	P 值	Std
生物多样性 > 汉语言多样性	−0.157	0.594	−0.264	0.792	−0.155

表 2 所示的检验结果表明：在刨除公因子影响后，生物多样性与汉语言多样性的 Std (标准化载荷)为−0.155，P 值为 0.792，统计结果不显著。由此可得，在去除公因子的影响后，生物多样性与汉语言多样性之间无显著相关关系。

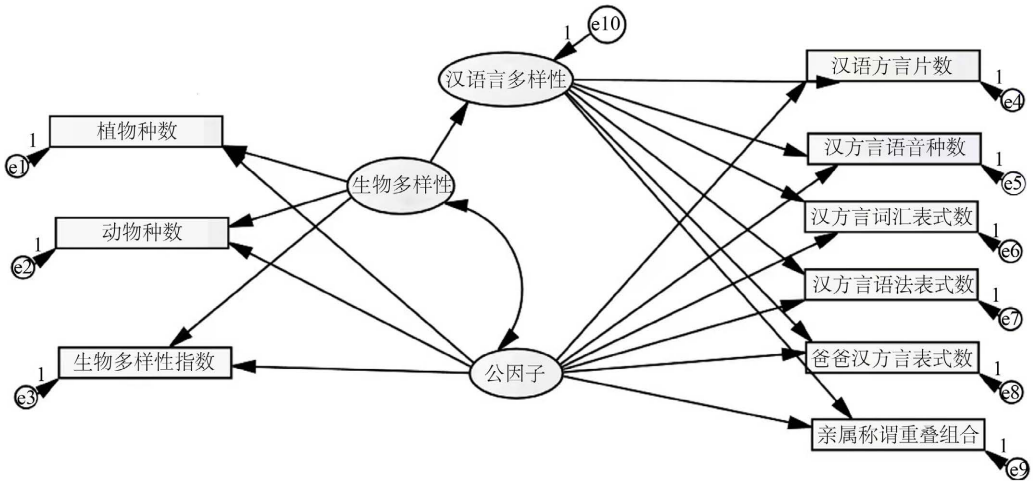

Figure 1. Structural equation model path map
图 1. 结构方程模型路径图

3. 模型选择

由 3 个潜变量，9 个观测变量构建的母体 CFA 模型如图 2 所示。本文分别用 Lasso [12] [13]、MCP [14] 方法和 Amos 基于指定搜索方法[15]寻找最佳模型惩罚生物多样性、汉语言多样和公因子三个因子之间的相关系数 ϕ_{12}、ϕ_{13}、ϕ_{23}，最终确定理想的拟合模型。

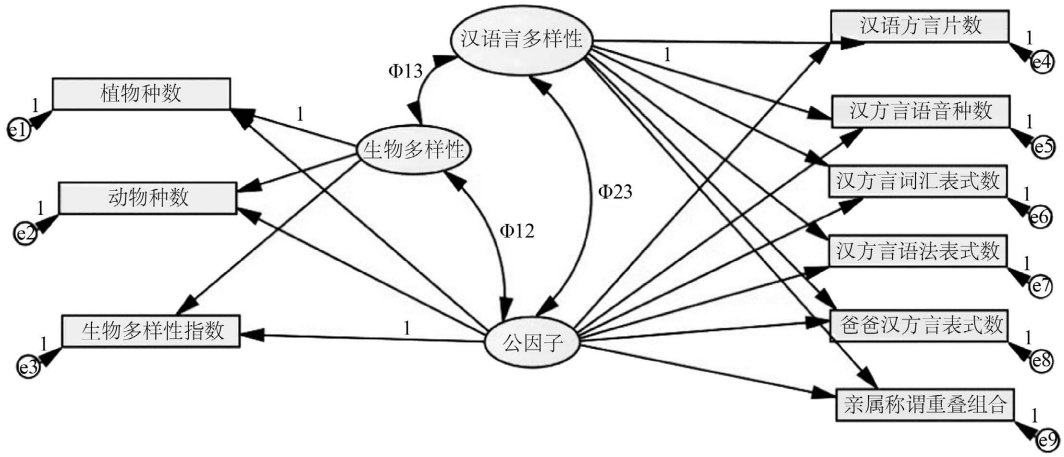

Figure 2. Maternal model
图 2. 母体模型

3.1. 基于 Lasso 正则化的模型选择

设定要测试惩罚值的数量为 20，每个模型的惩罚数依次增加 0.01。使用 Lass 估计并按 RMSEA 和 BIC 指标排序选出前 4 名的模型，模型检验的 P 值均大于 0.05，这表明模型在可接受的范围内(见表 3，带*的模型为本文选定的模型)，三个参数惩罚的轨迹图如图 3 所示。

由表 3 可知，在 Lasso 估计下，模型 1~3 的公因子与生物多样性的相关性均为正，与汉语言多样性不相关，且生物多样性与汉语言多样性不相关，与提出的假设不一致。模型 4 表示在公因子与汉语言多样性不相关，与生物多样性相关性为正的情况下，生物多样性与汉语言多样性的相关系数为负，与提出

的假设相一致。

Table 3. Fitting information of the model under lasso estimation
表 3. Lasso 估计下模型的拟合信息

模型	λ	RMSEA	BIC	P 值	生物多样性－汉语言多样性	汉语言多样性－因子	生物多样性－公因子
1	0.07	0.03	415.19	0.49	0.00	0.00	0.32
2	0.15	0.04	415.59	0.50	0.00	0.00	0.21
3	0.02	0.05	418.46	0.50	0.00	0.03	0.38
4*	0.04	0.05	418.49	0.50	−0.04	0.00	0.37

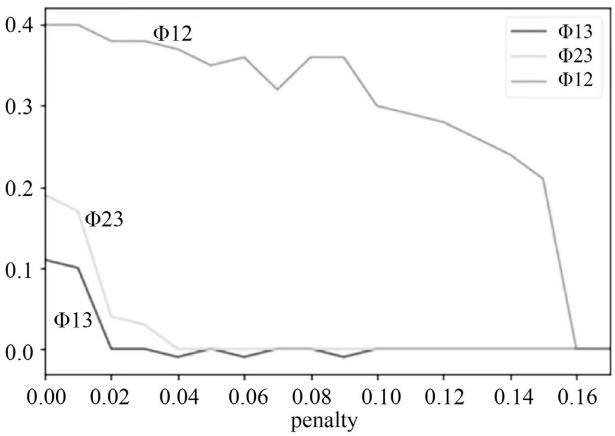

Figure 3. Trajectory of three penalized parameters under Lasso penalty
图 3. Lasso 惩罚下三个被惩罚参数的轨迹图

3.2. 基于 MCP 正则化的模型选择

设定要测试惩罚值的数量为 20，每个模型的惩罚数依次增加 0.01。使用 MCP 估计并按 RMSEA 和 BIC 指标排序选出前 4 名的模型，模型检验的 p 值均大于 0.05，这表明模型在可接受的范围内(见表 4，带*的模型为本文选定的模型)，三个参数惩罚的轨迹图如图 4 所示。

由表 4 可知，在 MCP 估计下，模型 1-2 均表示在公因子与生物多样性的相关性为正，以及与汉语言多样性不相关的情况下，生物多样性与汉语言多样性不相关，这与提出的假设不一致。模型 3 和 4 表示，在公因子与汉语言多样性的相关性为 0，以及与生物多样性相关性为正的情况下，生物多样性与汉语言多样性的相关性都为负；模型 3 和 4 均与提出的假设一致。

Table 4. Fitting information of the model under MCP estimation
表 4. MCP 估计下模型的拟合信息

模型	λ	RMSEA	BIC	P 值	生物多样性－汉语言多样性	汉语言多样性－因子	生物多样性－公因子
1	0.15	0.03	415.17	0.51	0.00	0.00	0.33
2	0.21	0.04	413.19	0.41	0.00	0.00	0.24
3*	0.09	0.05	418.48	0.49	−0.02	0.00	0.41
4*	0.11	0.05	418.50	0.48	−0.01	0.00	0.41

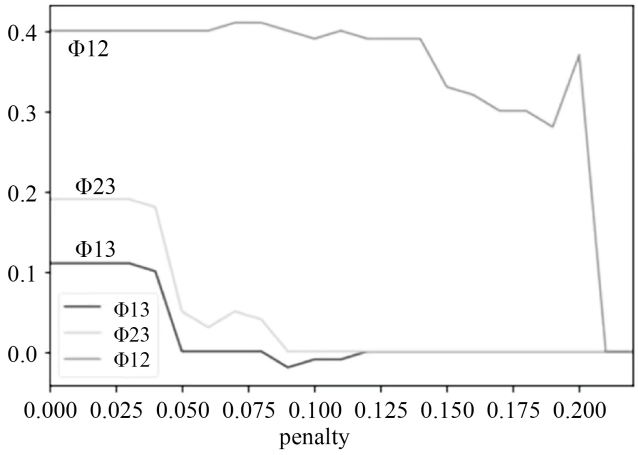

Figure 4. Trajectory of three penalized parameters under MCP penalty
图 4. MCP 惩罚下三个被惩罚参数的轨迹图

3.3. Amos 基于指定搜索方法寻找最佳模型

Amos 使用验证性的搜索方法寻找最佳模型，主要惩罚生物多样性、汉语言多样性和公因子三个因子之间的关系，计算出模型的拟合信息值，以便选出拟合较好的模型，根据 BCC 准则只将排名前 4 的模型显示在表中(见表 5，带*的模型为本文选定的模型)。其中 BCC0 在 0~2 之间表示没有证据表明该模型不是 K-L 最优模型；在 2~4 之间表明该模型不是最优模型的证据不明显；大于 4 表示有证据表明该模型不是最优模型。BIC0 在 0~2 之间表明该模型不是 K-L 最优模型的概率很小；在 2~6 之间表示有证据表明该模型不是最优模型，大于 6 则表明有很强的证据表明该模型不是最优模型。

由表 5 可知，在 Amos 验证性搜索下，模型 4 的 BCC 值为 2.95，在合理范围内，而 BIC 值为 3.23，超出模型为最优模型界限，BCC 值和 BIC 值均超出了模型拟合良好的范围，模型 4 不是最优模型。模型 1、模型 2、模型 3 的 BCC 值和 BIC 都在合理范围内，且 P 值都大于 0.05，表示模型可被接受。

模型 1 表示在提取公因子的影响后，生物多样性和语言多样性的相关性为 0，与提出的假设不一致。模型 2、3 分别表示，公因子与汉语言多样性相关性分别为 0.41 和 0，与生物多样性相关系数为分别为 0 和 0.47 时，生物多样性与汉语言多样性的相关系数都为负，即控制公因子影响后，生物多样性与汉语言多样性的相关性为负，与提出的假设一致。由此可得，Lasso、MCP 和 Amos 选出的模型是一致的，选出的模型均表示在去除公因子影响下，生物多样性影响语言多样性。

Table 5. Models selected under Amos confirmatory search
表 5. Amos 验证性搜索下选出的模型

模型	BCC0	BIC0	C/DF	P 值	生物多样性 - 汉语言多样性	汉语言多样性 - 因子	生物多样性 - 公因子
1	0.00	0.00	1.03	0.42	0.00	0.05	0.41
2*	0.01	0.01	1.03	0.42	−0.03	0.00	0.41
3*	1.74	1.74	1.12	0.32	−0.15	0.47	0.00
4	2.95	3.23	1.08	0.36	−0.11	0.19	0.40

4. 中国植被覆盖率等因素与公因子的区域相关关系

为验证公因子与园林草湿地总面积、森林覆盖率和各省总面积等的区域相关系数，本文采用 Spearman

相关系数对中国 30 个省份的面积和森林覆盖率等因素与公因子进行区域相关分析，本章使用的数据来源于 2021 中国统计年鉴[16]。计算结果显示：当 P 值 < 0.05 时，公因子与园林草湿地总面积的相关系数 > 0.5，与森林覆盖率和各省总面积的相关系数 < 0.5，公因子与各省人口密度的相关系数为−0.405，与光照时长和各省 GDP 的 P 值 > 0.05 (见表 6)。

Table 6. Spearman correlation coefficient between common factor scores and total area of garden grass wetlands, forest coverage, and provincial area
表 6. 公因子得分与园林草湿地总面积、森林覆盖率和省域面积等的 Spearman 相关系数

数值	园林草湿地覆盖率	森林覆盖率	各省总面积	各省人口密度	光照时长	各省GDP
Spearman 相关系数	0.572	0.429	0.478	−0.405	−0.041	0.017
检验 P 值	0.001	0.017	0.008	0.027	0.799	0.927

结果表明公因子与植被覆盖率中的园林草湿地覆盖率、森林覆盖率呈显著的区域正相关，与各省总面积也呈显著的区域正相关，且公因子与园林草湿地总面积的相关性大于与森林覆盖率和各省总面积的相关性。公因子与各省人口密度有显著的区域负相关关系，与各省的 GDP 和光照时长无区域相关关系。由此可得，公因子与植被覆盖率和各省面积有关。

5. 讨论与建议

本文通过结构方程建模和模型选择后，结果均表明在去除公因素的影响，生物多样性与汉语言多样性之间成微弱的负相关关系。园地、林地、草地和湿地为生物和人类提供了良好的栖息地，其覆盖率的高低和各省面积的大小影响着生物的繁衍和语种的传播，这些因素同时影响着生物多样性和汉语言多样性，在去除这些因素的影响后，生物多样性与汉语言多样性为争夺有限的资源而存在着竞争与对抗关系，但这个关系在中国不是主流。

自然资源和环境气候影响着民族和语言的分布，对动植物的生存和繁衍起着重要的作用，是语言多样性与生物多样性同时依赖的生存条件，同时影响着生物和语言多样性。随着科技的进步，人类为了追求更高的经济效益，造成了严重的资源浪费和环境污染，对人类自身和动植物的生存环境造成了毁灭性影响。生态环境的恶化导致的生物多样性锐减，同时又制约着经济的发展和语言多样性。因此，当今社会的发展要顺应自然，在改造自然、利用自然资源带给人们多元化生活的同时，更要注重保护生态环境，节约资源，促进生态的可持续发展。

参考文献

[1] Upadhyay, R.K. and Hasnain, S.I. (2017) Linguistic Diversity and Biodiversity. *Lingua*, **195**, 110-123. https://doi.org/10.1016/j.lingua.2017.06.002

[2] Gorenflo, L.J., Romaine, S., Mittermeier, R.A., *et al.* (2012) Co-Occurrence of Linguistic and Biological Diversity in Biodiversity Hotspots and High Biodiversity Wilderness Areas. *Proceedings of the National Academy of Sciences*, **109**, 8032-8037. https://doi.org/10.1073/pnas.1117511109

[3] Sutherland, W.J. (2003) Parallel Extinction Risk and Global Distribution of Languages and Species. *Nature*, **423**, 276-279. https://doi.org/10.1038/nature01607

[4] Li, J., Steffensen, S.V. and Huang, G.W. (2020) Rethinking Ecolinguistics from a Distributed Language Perspective. *Language Sciences*, **80**, Article ID: 101277. https://doi.org/10.1016/j.langsci.2020.101277

[5] Gorenflo, L.J. and Romaine, S. (2021) Linguistic Diversity and Conservation Opportunities at UNESCO World Heritage Sites in Africa. *Conservation Biology*, **35**, 1426-1436. https://doi.org/10.1111/cobi.13693

[6] Couto, H.H. (2014) Ecological Approaches in Linguistics: A Historical Overview. *Language Sciences*, **41**, 6-25.

https://doi.org/10.1016/j.langsci.2013.08.001

[7] Zhang, X., Bu, Z., Ju, H. and Jing, Y. (2022) Investigation on the Relationship between Biodiversity and Linguistic Diversity in China and Its Formation Mechanism. *International Journal of Environmental Research and Public Health*, **19**, Article No. 5538. https://doi.org/10.3390/ijerph19095538

[8] 万本太, 徐海根, 丁晖, 刘志磊, 王捷. 生物多样性综合评价方法研究[J]. 生物多样性, 2007, 15(1): 97-106.

[9] Muthén, B. (1984) A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators. *Psychometrika*, **49**, 115-132. https://doi.org/10.1007/BF02294210

[10] Fox, J. and Hayduk, L.A. (1987) Structural Equation Modeling with LISREL: Essentials and Advances. *Canadian Studies in Population*, **14**, 257-259. https://doi.org/10.2307/3341309

[11] Bagozzi, R.P. and Yi, Y. (1988) On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, **14**, 33-46. https://doi.org/10.1007/BF02723327

[12] Garrido, M., Hansen, S.K., Yaari, R. and Hawlena, H. (2021) A Model Selection Approach to Structural Equation Modelling: A Critical Evaluation and a Road Map for Ecologists. *Methods in Ecology and Evolution*, **13**, 42-53. https://doi.org/10.1111/2041-210X.13742

[13] Jacobucci, R., Grimm, K.J. and Mcardle, J.J. (2016) Regularized Structural Equation Modeling. *Structural Equation Modeling*: *A Multidisciplinary Journal*, **50**, 736-736. https://doi.org/10.1080/10705511.2016.1154793

[14] Huang, P.H., Chen, H. and Weng, L.J. (2017) A Penalized Likelihood Method for Structural Equation Modeling. *Psychometrika*, **82**, 329-354. https://doi.org/10.1007/s11336-017-9566-9

[15] Byrne, B.M. (2016) Structural Equation Modeling with AMOS. Taylor and Francis, Abingdon. https://doi.org/10.4324/9781315757421

[16] 中华人民共和国国家统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2021.