心理学探新 2021,Vol.41,No.1,69-75
PSYCHOLOGICAL EXPLORATION

贝叶斯多组比较——渐近测量不变性*

宋琼雅 张沥今 潘俊豪
(中山大学心理学系,广州 510006)

摘 要:在行为科学研究领域中,检验测量工具的测量不变性是进行群体差异比较的前提。目前,多组验证性因子分析(多组 CFA)方法被广泛用于检验测量不变性,但是它对跨组等值的限制过于严格,在实际应用中常常存在大量局限。贝叶斯渐近测量不变性方法基于贝叶斯思想的优良特性,放宽了传统多组 CFA 方法对跨组差异的严格限制,避免了传统方法的问题,具有较高的应用价值。文章详细介绍了贝叶斯渐近测量不变性方法的原理及优势,同时通过实例展示了渐近测量不变性方法在 Mplus 软件中的具体分析过程。

关键词:贝叶斯方法;多组验证性因子分析;渐近测量不变性

中图分类号:B841.2 **文献标识码**:A **文章编号**:1003-5184(2021)01-0069-07

1 引言

组间差异比较在社会科学领域中有着广泛的应用,包括跨文化、跨区域、跨群体及跨时间点的研究等(如,Shorey, Allan, Cohen, Fite, Stuart, & Temple, 2018)。在这类研究中,研究者进行跨组比较的常常是构念(Contrust),即不能直接被观测的潜变量(Latent variable),如态度、能力等,而用于反映潜变量的可直接观测的变量通常被称为外显变量(Manifest variable)。基于潜变量的多组比较中,测量工具的测量不变性(Measurement invariance, MI)是进行组间差异比较的前提条件。测量不变性是指测量工具所测量的构念不会因施测群体的改变而改变(van de Schoot, Kluytmans, Tummers, Lugtig, Hox, & Muthén,2013)。以潜变量框架为基础的结构方程模型(Structural Equation Modeling, SEM)为测量不变性的检验提供了切实可行的方法,测量不变性的检验主要依赖于 SEM 的测量模型部分,即验证性因子分析(Confirmatory Factor Analysis, CFA)模型,该模型反映了外显变量与潜变量之间的关系。

目前已经发展出多种检验测量不变性的 SEM 方法,包括多组 CFA(Raju, Laffitte, & Byrne, 2002),多层 CFA(Jak, Oort, & Dolan, 2013),多层因子混合模型(Kim, Cao, Wang, & Nguyen, 2017)和对齐法(Asparouhov & Muthén, 2014)等。其中,多层 CFA 和多层因子混合模型方法均是基于多层结构建立测量不变性模型。多层 CFA 以组的平均测量模型为组间模型,检测不同组的模型与平均模型的差异;多层因子混合模型则是根据模型的相似性,将群组归类为不同的潜在类别,通过检测不同的潜在类别模型以评估群组间测量不变性。两种方法关注的都是多群组参数对测量不变性的违反程度以及相似群组的划分等,并不会关注每个组的模型。对齐法主要依据旋转的方式获得最优测量恒定模型,这种方法比较适宜进行群组间因子均值的比较。具体方法的理论介绍和比较可以参考 Kim 等(2017)的相关讨论。

多组 CFA 是社会科学领域最为常用的方法(Raju, Laffitte, & Byrne, 2002),其与贝叶斯渐近测量不变性方法都是基于每个具体组的模型进行组间比较逐步确定测量不变性,二者具有原理上的相似性,贝叶斯渐近测量不变性方法可以视为多组 CFA 与贝叶斯方法的结合(Kim et al., 2017)。本文将针对传统的多组 CFA 和贝叶斯渐近测量不变性进行深入地讨论。

1.1 多组验证性因子分析

多组 CFA 建模需要对每个组别分别建立一个 CFA 模型,再依次施加模型参数跨组相等的限制,以此检验测量工具的测量不变性,参照 Schmit 和 Kuljanin(2008),其建模步骤按顺序包括:(1)形态不变性(Configural invariance)模型,要求各组具有相同的因子结构,即潜变量和条目间的对应关系跨组一致;(2)载荷不变性(Metric invariance)模型,指不同群组的因子载荷跨组相等;(3)截距不变性(Scalar invariance)模型,指条目截距跨组相等。只

* **基金项目**:国家自然科学基金项目(31871128),教育部人文社会科学研究规划基金项目(18YJA190013)。
通讯作者:潘俊豪,E-mail:panjunh@mail.sysu.edu.cn。

有当模型满足载荷和截距不变性,因子均值的跨组比较才是有意义的,否则可能导致有偏的估计结果;(4)误差方差不变性(Error - variance invariance)模型,指条目误差方差跨组等值。Little 和 Card (2013)指出误差方差不变性通常不需要被满足,因为误差方差部分包括了外显变量的测量误差,严格限制其跨组相等可能会导致对结构模型参数的有偏估计。

多组 CFA 方法在每施加一种测量不变性限制后,都需要通过模型拟合比较来检验施加限制后模型拟合是否显著变差,在满足了当前不变性限制的基础上才能够进行下一种跨组不变性的检验。针对这种嵌套模型的拟合比较可以采用似然比检验(log - likelihood ratio test, Maslowsky, Jager, & Hemken, 2015),但由于似然比检验容易受到样本量的影响,可能无法获得准确的结论,一些研究者提出了更加有效的模型拟合比较标准,如 $|\Delta CFI| \leqslant 0.01$ (Little & Card,2013)、$|\Delta RMSEA| \leqslant 0.03$ 作为检验模型无明显差异的稳健方法(Rutkowski & Svetina,2014)。

多组 CFA 方法思路简洁且易于掌握,得到了广泛的应用。在多组 CFA 模型满足跨组截距不变性的基础上,研究者还可以进一步进行结构不变性的检验,包括对因子均值、因子方差/协方差的跨组比较(Little & Card,2013)。目前流行的建模软件也都支持多组 CFA 建模,便于研究者应用该方法。

1.2 多组验证性因子分析的局限

尽管多组 CFA 方法存在着诸多优势,但在实证研究中,多组 CFA 方法对于跨组不变性的限制往往过于严苛(要求载荷、截距严格跨组相等),这使得添加限制条件后的模型更容易被拒绝(Asparouhov & Muthén,2014;Kim et al. ,2017)。Marsh 等(2018)指出在实证研究中截距不变性模型几乎不会被满足,且这种现象在组数较多的情况下更为严重,极大地影响了研究的后续分析。

在实证研究中,当载荷或截距的跨组不变性不被满足时,研究者通常可以采用事后模型修正(Post - hoc model modification)方法,结合理论和修正指数(Modification index)的建议释放一些违背跨组不变性的参数以改善模型拟合情况,即建立部分测量不变性(Partial measurement invariance)模型。但是这种方法存在着一些局限:(1)事后模型修正方法每释放一个参数都需要进行模型拟合比较,修正过程较为繁琐;(2)通过修正指数发现的违背测量不变性的参数可能源于抽样变异性(Asparouhov & Muthén,2014),同时,参数的释放还会受到研究者主观性的影响,这增加了一类错误率,降低了结论的可重复性;(3)在部分测量不变性模型的基础上进行组间比较,容易导致对跨组因子均值差异的有偏估计(Marsh et al. ,2018)。

方法学家一直在探索具有更高适用性的不变性检验方法。其中,Muthén 和 Asparouhov(2013)提出的贝叶斯渐近测量不变性(Bayesian approximate Measurement Invariance)方法,较好地解决了因传统方法限制过于严格而产生的问题。遗憾的是,由于其发展时间较短,应用研究者对于这种建模思路并不熟悉、对相关软件操作也不够了解。本文将对该方法进行详细地介绍与总结,阐释其原理与优势。同时以实例分析演示其建模步骤、评价标准,并与传统多组 CFA 进行对比。希望本文能够为应用研究者提供新的多组建模思路,解决采用传统方法建模时难以克服的问题。

2 渐近测量不变性

2.1 方法原理

在结构方程建模中,除了传统的频率学派估计方法(如,极大似然估计),贝叶斯估计方法近年来也愈加流行(张沥今,陆嘉琦,魏夏琰,潘俊豪,2019)。相比于传统方法在小样本、非正态数据和复杂模型等情况下容易出现模型无法识别、参数估计有偏等问题,贝叶斯 SEM 对模型或数据的极端情况更为宽容(李锡钦,2011)。

从统计学思想的角度来看,频率学方法将未知参数视为常数,而贝叶斯方法则将其视为随机变量,结合样本数据和先验信息得到未知参数的后验分布(Muthén & Asparouhov,2012)。因此贝叶斯方法能够灵活地纳入先验信息,有效的先验信息可以带来更准确的参数估计(Yuan & MacKinnon,2009)。研究者对先验信息准确性的把握体现在先验分布的方差大小中,先验分布的方差越小,对未知参数后验分布的影响越大。

鉴于贝叶斯 SEM 表现出的优良特性,越来越多的研究者致力于其的发展和应用(van de Schoot, Winter, Ryan, Zondervan - Zwijnenburg, & Depaoli,2017),进而衍生出了一系列新的建模思路。例如,传统 CFA 方法在建模时,常常将交叉载荷或测量误差相关严格限制为 0,但是这种严格的限制在实际应用中很难得到满足(图 1a)。基于贝叶斯方法结合先验信息的特性,Muthén 和 Asparouhov(2012)创造性地指出:如果为交叉载荷和测量误差相关提供均值为 0、方差极小的先验分布,就可以允许这些参数在 0 附近波动,放宽对模型的严格限制(图 1b)。而放宽模型限制后发现的显著不为 0 的交叉载荷或测量误差相关,可以被视为由抽样变异性导致的、不需要理论

解释的冗余参数(nuisance parameters)。这种放宽模型限制的思想自提出后就得到了深入的研究与推广(Lu, Chow, & Loken, 2016; Pan, Ip, & Dubé, 2017等)。Muthén 和 Asparouhov(2013)更进一步将其应用到了多组结构方程建模中,发展出渐近测量不变性方法。

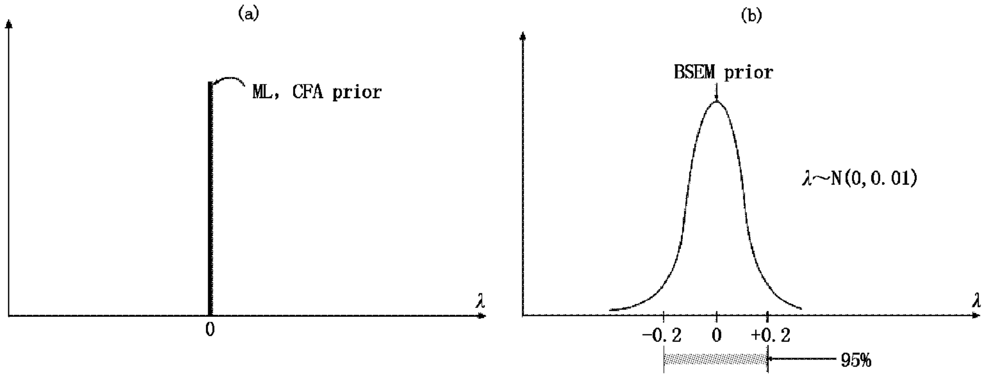

图 1　传统方法和贝叶斯方法的比较(Muthén & Asparouhov, 2013)

注:图 1(a)为传统方法中严格限制交叉载荷或测量误差相关参数为 0;图 1(b)为贝叶斯方法通过对参数提供均值为 0 的小方差先验分布,就可以允许其在 0 附近波动。如通过对参数提供 $N(0,0.01)$ 的先验分布,相当于在数据标准化的前提下,允许参数有 95% 的概率落在(−0.2,0.2)之间。

渐近测量不变性方法通过对载荷和截距的跨组差异提供均值为 0、方差极小的先验分布,就可以允许载荷和截距存在微小的跨组差异,放宽传统方法严格的限制。Muthén 和 Asparouhov(2013)指出,渐近测量不变性方法本质上认为数据存在着诸多微小、随机的跨组差异,这些差异可能是互为相反方向、可以相互抵消的,因此不影响对结构模型参数的准确估计。

渐近测量不变性方法相比于传统方法有着诸多优势:(1)避免了传统方法限制过于严格而容易导致模型拟合过差的问题;(2)在一次估计中就可以放宽对所有参数的限制,不需要进行多次事后模型修正,避免了修正时重复利用同一批数据可能导致的更高的一类错误率、有偏的参数估计等问题,可以作为一种检验违背测量不变性的参数的工具(Muthén & Asparouhov, 2013);(3)通过提供不同的小方差先验分布可以检验数据违背不变性的程度(Cieciuch, Davidov, Algesheimer, & Schmidt, 2017),如检验提供 $N(0,0.01)$ 和 $N(0,0.05)$ 的先验下模型的拟合情况等。

模拟研究显示,当载荷或截距参数存在着较小的跨组差异时,渐近测量不变性方法在模型拟合和参数估计中表现都很好(Kim et al., 2017; van de Schoot et al., 2013),且贝叶斯方法的框架使其更具灵活性,模型拟合评估也更为有效(Kim et al., 2017; Schmitt & Kuljanin, 2008)。

2.2　分析步骤

渐近测量不变性方法自提出以来,就得到了广泛的发展与研究,但由于其发展时间较短,具体的分析流程和评价标准尚未规范,本文基于渐近测量不变性方法的一些应用研究和模拟研究结果(Fong & Ho, 2014; Kim et al., 2017),总结出了较为完整、全面的分析步骤。

2.2.1　形态不变性

渐近测量不变性方法首先需要建立形态不变性模型,在模型拟合良好的前提下才能进行后续分析。模型拟合评估指标包括后验预测 p 值(Posterior Predictive p - value, PPp 值)和后验预测区间(Posterior Predictive interval),后验预测 p 值指理论模型生成数据的统计检验量(如,卡方值)大于样本数据的统计检验量的比例,后验预测区间则指样本数据的统计检验量和理论模型生成数据的统计检验量的差异值的 95% 可信区间。当 PPp 值在 0.5 左右,且后验预测区间包括 0 时,说明模型拟合良好。其中,PPp 值在 0.5 左右的判断标准较为主观,模拟研究显示 PPp 值大于 0.1 且后验预测区间包括 0 时通常认为模型拟合良好(Meghan & Zhang, 2018; Muthén & Asparouhov, 2012)。

此外,在形态不变性模型中,还可以对测量误差项的方差协方差矩阵提供合适的逆 Wishart 先验分布,允许测量误差相关在 0 附近波动以放宽限制(Fong & Ho, 2014)。

2.2.2　载荷和截距不变性

渐近测量不变性方法可以通过对载荷和截距的跨组差异提供均值为 0、方差极小的先验分布(建议采用 0.01 的方差,即在数据被标准化处理的前提下,允许跨组差异有 95% 的变异落在(−0.2,0.2)之间),在一次估计中放宽对所有载荷和截距跨组

的严格限制。如果此时模型拟合较差,说明数据不能满足载荷和截距跨组不变性,而如果模型拟合较好还需要进一步进行敏感性分析。

2.2.3 敏感性分析

van de Schoot 等(2013)及 Kim 等(2017)均建议在采用渐近测量不变性分析时,进行敏感性分析(Sensitivity Analysis)以避免先验方差的选取受到研究者主观性的影响。

具体来说,研究者需要先确定预设先验(predetermined prior)方差,建议采用 0.01(Kim et al., 2017)。在对载荷和截距的跨组差异提供 $N(0, 0.01)$的先验分布后可以得到模型和数据的拟合程度,如 M*plus* 中提供的 PP*p* 值、偏差信息准则(Deviance Information Criterion, DIC)和贝叶斯信息准则(Bayesian Information Criterion, BIC)。这一步骤即为载荷和截距不变性模型的检验。

在模型拟合良好的前提下,对跨组差异提供方差更大的先验分布,通常为 $N(0, 0.05)$(Kim et al., 2017; van de Schoot et al., 2013)。比较含有两种先验分布的模型:如果结果显示先验分布为 $N(0, 0.01)$的模型拟合更好、或两个模型拟合没有显著差异,则可以认为模型满足跨组载荷和截距不变性;否则说明对模型施加跨组不变的限制是错误的,导致了模型拟合明显变差,即模型不满足跨组不变性(Kim et al., 2017)。

在模型拟合比较中,Kim 等(2017)建议采用 DIC 指标。因为 BIC 指标总是会选择先验方差为 0.05 的模型,这可能是由于方差 0.05 允许存在的跨组差异相对更大,通常会发现更多显著违背不变性的参数,而 DIC 则会惩罚过于复杂的模型。在分析中,当两个模型 $|\Delta \mathrm{DIC}| > 7$ 时,则有足够的证据认为 DIC 更小的模型更优,否则可以认为两个模型没有明显差异(Spiegelhalter, Best, Carlin, & Van Der Linde, 2002)。

在满足载荷和截距不变性的基础上,研究者可以进行因子均值的比较。尽管渐近测量不变性方法目前已经可以满足大部分实证研究的需求,但还没有合适的先验分布能够放宽对测量误差方差/协方差跨组不变的严格限制,因此还无法支持组间方差/协方差不变性的检验。

2.3 部分渐近测量不变性

渐近测量不变性方法假设载荷和截距参数中存在着诸多微小的组间差异,此时可以实现准确的参数估计。但是如果数据中一些载荷或截距参数存在着较大的组间差异,$N(0, 0.01)$的小方差先验可能会压缩这些参数真实的组间差异,进而造成对跨组因子均值差异的有偏估计(Muthén & Asparouhov, 2013)。因此研究者们建议在发现显著违背测量不变性的参数后,可以采用部分渐近测量不变性方法对这些参数进行自由估计,以实现更准确的跨组因子均值比较(van de Schoot et al., 2013)。即对这些参数采用 Mplus 默认的无信息先验分布进行自由估计,就可以建立部分渐近测量不变性模型(Muthén & Asparouhov, 2013)。相比于传统方法,部分渐近测量不变性方法能够在一次估计中检测到所有违背不变性的参数,极大地降低了事后模型修正方法的工作量,且避免了研究者主观地选择需要修正的参数。

但是部分渐近测量不变性模型依然存在着与传统方法相同的两个问题:(1)需要释放的参数可能来源于抽样变异性;(2)释放违背不变性的参数会对结构模型参数的估计产生怎样的影响尚不确定(Schmitt & Kuljanin, 2008)。因此研究者也需要结合理论假设,建立更符合理论意义的模型。

3 示例:贝叶斯渐近测量不变性分析

本文数据来源于大学生创新训练项目(国家级项目):社会支持对职业决策困难的影响机制研究。为了检验不同性别在主要变量社会支持和职业决策困难上反应的差异,本研究采用传统的多组 CFA 和贝叶斯渐近测量不变性两种方法进行建模,展示渐近测量不变性方法的具体分析步骤,并与传统方法进行对比。研究采用 M*plus*8.0 软件(代码见附录)进行建模,数据来源于 353 名本科大学生,平均年龄为 19.55 ± 1.45 岁(range: 17 ~ 24),其中男生89 名。

3.1 多组验证性因子分析

本研究中社会支持量表(叶悦妹,戴晓阳,2008)各条目的偏度在 −1.107 ~ −1.43 之间,峰度在 −0.705 到 1.534 之间,职业决策困难量表(李娜,2009)的各个条目的偏度在 −0.756 到 0.906 之间,峰度在 −1.053 到 0.920 之间,偏度绝对值小于 2 且峰度绝对值小于 7(West, Finch, & Curran, 1995),因此条目均基本符合正态分布,且由于数据中不含缺失值,可以采用极大似然法进行模型估计。此外,考虑到职业决策困难题目过多,测量模型拟合较好,且最终目的是考察两性别组的因子均分的差异,因此对于量表的题目采用随机打包法进行了条目打包(Item parceling,吴艳,温忠麟,2011)。在多组 CFA 模型比较中,如果 $|\Delta \mathrm{CFI}| < 0.01$ 可以认为两个模型的拟合无明显差异(Little & Card, 2013)。

对于社会支持的多组 CFA 分析发现,在部分截距不变性模型(见表 1)的基础上,比较两组因子均值差异为:女生在主观支持、客观支持和支持利用度三个维度上得分均显著高于男生(标准化均值差

异：*Diff* = 0.389, p = 0.002；*Diff* = 0.566, p < 0.001；*Diff* = 0.410, p = 0.003)。对于职业决策困难的分析则发现，在截距不变性模型(见表1)的基础上，女生和男生没有表现出显著差异。

表1　多组验证性因子分析模型拟合及模型比较

模型	χ^2	*df*	CFI	TLI	SRMR	RMSEA	90% CI of RMSEA	ΔCFI	Pass?
社会支持									
形态不变性	494.817	230	0.921	0.907	0.064	0.081	[0.071,0.091]		
载荷不变性	521.450	244	0.917	0.908	0.075	0.080	[0.071,0.090]	−0.04	No
部分载荷不变性	512.170	242	0.920	0.910	0.072	0.080	[0.070,0.089]	−0.01	Yes
截距不变性	560.015	256	0.909	0.904	0.083	0.082	[0.073,0.091]	−0.11	No
部分截距不变性	526.292	253	0.919	0.913	0.076	0.078	[0.069,0.088]	−0.01	Yes
职业决策困难									
形态不变性	88.949	48	0.964	0.946	0.043	0.070	[0.046,0.092]		
载荷不变性	95.865	54	0.963	0.951	0.050	0.066	[0.044,0.088]	−0.01	Yes
截距不变性	97.452	60	0.967	0.960	0.050	0.059	[0.037,0.080]	+0.04	Yes

3.2　贝叶斯渐近测量不变性

由上述分析可知数据服从正态分布且不含缺失值，根据贝叶斯测量不变性模型的建模步骤进行分析，不同先验分布下模型的|ΔDIC|<7，则可认为模型满足敏感性分析。此外，对于Mplus中"DIFFERENCE OUTPUT"中显示的违背跨组不变性的参数，研究者需要提供无信息先验以释放该参数，构建部分测量不变性模型。

结果显示，针对不同性别在社会支持量表的反应(见表2)，在满足敏感性分析的基础上，比较两组因子均值发现女生在三个维度上得分均显著高于男生(标准化均值差异：*Diff* = 0.394, p = 0.004；*Diff* = 0.443, p < 0.001；*Diff* = 0.559, p < 0.001)。在职业决策困难的渐近测量不变性分析中发现，在满足敏感性分析(见表2)的基础上未发现性别在该量表反应上的显著差异。

表2　贝叶斯渐近测量不变性分析——模型拟合及模型比较

模型	PP*p*	95% C.I.[1]	BIC	DIC	ΔDIC
社会支持					
形态不变性	0.624	[−86.681,62.040]	14723.613	13131.384	
载荷和截距不变性 $N(0,0.01)$	0.664	[−90.916,60.739]	14766.800	13113.782	
载荷和截距不变性 $N(0,0.05)$	0.674	[−90.851,57.596]	14757.127	13111.530	−2.252
部分测量不变性	0.681	[−93.554,58.890]	14764.332	13112.122	
职业决策困难					
形态不变性	0.552	[−42.964,37.873]	8257.405	7672.218	
载荷和截距不变性 $N(0,0.01)$	0.562	[−44.122,40.016]	8276.149	7678.028	
载荷和截距不变性 $N(0,0.05)$	0.587	[−44.386,40.079]	8272.110	7681.004	2.976

注：[1]95% C.I.为后验预测区间。

3.3　方法比较

从上述分析中可以发现，与Muthén和Asparouhov(2013)及van de Schoot等(2013)的研究结果一致，渐近测量不变性方法对因子均值差异的估计结果与传统方法相差不大。

传统的多组CFA方法对模型限制过于严格，且模型比较和修正过程都更为繁琐。此外，传统方法中现有的评价标准仍然存在着较多争议(Little & Card, 2013)，容易受到研究者主观性的影响。而渐近测量不变性方法操作简单，通过放宽对组间差异的严格限制也避免了传统方法容易导致模型拟合过差的问题，敏感性分析也可以帮助研究者避免先验信息主观性的影响，得到更可靠的结果。

3.4　其他应用

在社会科学研究领域中，一个有意义的研究应保证在各国家中测量的构念保持一致(Cieciuch, Davidov, Algesheimer, & Schmidt, 2017)。21题目的世界价值观量表(21 - item Portrait Value Questionnaire)是测量人们价值观的工具，具有世界范围的普适性，经常被应用于跨文化研究(Schwartz, 1992)。Cieciuch 等(2017)依托公开数据库 European Social Survey(ESS)在 2002 - 2012 年采集的来自 15 个主要欧洲国家的 6 轮数据进行传统的多组 CFA 和贝叶斯渐近测量不变性分析。结果发现,传统多组 CFA 能够较好拟合形态不变性模型,但是无法支持后续的载荷不变性和截距不变性模型,对于跨文化研究者的后续的因子相关分析可能产生影响。而贝叶斯载荷和截距渐近测量不变性模型能够支持对变换的开放性和自我提升两个高级维度并支持部分国家的自我超越(12 个)和保守性(10 个)维度。同时,Cieciuch 等人(2017)也通过敏感性分析发现贝叶斯渐近测量不变性模型在 0.01,0.05 和 0.1 三种先验方差下结果都基本一致,模型较为稳定。

4 总结与展望

基于贝叶斯渐近测量不变性方法的优良特性,该方法也越来越受到应用研究者的欢迎。该方法被应用于焦虑、抑郁的性别不变性检验(Fong & Ho, 2014)等应用研究中。随着计算机技术的飞速发展,数据的获取和存储都发生了巨大的变化,跨区域、跨文化和跨时间的研究也不断增加,因此考察不同群体对测量构念理解的同质性也变得十分重要。希望本文的介绍能帮助研究者更快地掌握这一方法,解决实际数据分析中遇到的一些障碍。

但是作为一种建模工具,贝叶斯渐近测量不变性方法不可避免地存在着一定的适用范围和局限:首先,这种方法在组数多时计算耗时更长,软件运行效率相对较低;其次,该方法还未能明确如何处理不能满足载荷和截距不变性的模型。研究者们提出了诸如部分渐近测量不变性(Muthén & Asparouhov, 2013)、拓展的对齐法等诸多解决办法(Marsh et al.,2018),但这些方法都还存在着许多亟待发展的地方,未来还需要更多的模拟研究和实证研究来填补这些空白。最后,尽管渐近测量不变性方法能够完成对载荷和截距的跨组不变性检验,也可以用于后续的因子均值比较,但是目前还不能检验误差方差/协方差不变性。

除了渐近测量不变性方法,目前存在的其他检验测量不变性的诸多方法(如,多层 CFA,多层因子混合模型,对齐法等),它们虽然都有各自较为独特的优势,但适用范围相对更窄,且存在一定的局限。针对现有不同方法的优劣比较依然有待探究,此外,这种方法间的比较也可以进一步扩展到结构不变性层面,包括进行因子均值,因子方差/协方差的不变性检验。

参考文献

李锡钦.(2011).*结构方程模型:贝叶斯方法*(蔡敬衡,潘俊豪,周影辉 译).北京:高等教育出版社.

李娜.(2009).*大学生职业决策困难问卷修编及其特点研究*(硕士论文).西南大学.

吴艳,温忠麟.(2011).结构方程建模中的题目打包策略.*心理科学进展*,*19*(12),1859 - 1867.

叶悦妹,戴晓阳.(2008).大学生社会支持评定量表的编制.*中国临床心理学杂志*,*16*(5),465 - 468.

张沥今,陆嘉琦,魏夏琰,潘俊豪.(2019).贝叶斯结构方程模型及其研究现状.*心理科学进展*,*27*(11),1812 - 1825.

Asparouhov, T., & Muthén, B. (2014). Multiple - group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495 - 508.

Cieciuch, J., Davidov, E., Algesheimer, René., & Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European social survey. *Sociological Methods & Research*, *47*(4), 665 - 686.

Fong, T., & Ho, R. (2014). Testing gender invariance of the hospital anxiety and depression scale using the classical approach and bayesian approach. *Quality of Life Research*, *23*(5), 1421 - 1426.

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(2), 265 - 282.

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 524 - 544.

Little, T. D., & Card, N. A. (2013). *Longitudinal structural equation modeling*. The Guilford Press.

Lu, Z. H., Chow, S. M., & Loken, E. (2016). Bayesian factor analysis as a variable - selection problem: Alternative priors and consequences. *Multivariate Behavioral Research*, *51*(4), 519 - 539.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., et al. (2018). What to do when scalar invariance fails: The extended alignment method for multi - group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524 - 545.

Maslowsky, J., Jager, J., & Hemken, D. (2015). Estimating and interpreting latent variable interactions: A tutorial for applying the latent moderated structural equations method. *International Journal of Behavioral Development*, *39*(1), 87 - 96.

Meghan, K. C., & Zhang, Z. Y. (2018). Fit for a Bayesian: An evaluation of PP*p*and DIC for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 1 - 12.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313 – 335.

Muthén, B., & Asparouhov, T. (2013). *BSEM measurement invariance analysis: Mplus Web Note*17. http://www.statmodel.com/examples/webnotes/webnote17.pdf.

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods*, *22*(4), 687 – 704.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*(3), 517 – 529.

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large – scale international surveys. *Educational & Psychological Measurement*, *74*(1), 31 – 57.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210 – 222.

Schwartz, S. H. (1992). "Universals in the content and structure of values: Theory and empirical tests in 20 countries." In P. Z. Mark (Ed.), *Advancesin experimental social psychology* (pp. 1 – 65). New York: Academic Press.

Shorey, R. C., Allan, N. P., Cohen, J. R., Fite, P. J., Stuart, G. L., & Temple, J. R. (2019). Testing the factor structure and measurement invariance of the conflict in Adolescent Dating Relationship Inventory. *Psychological Assessment*, *31* (3), 410.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583 – 639.

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with scylla and charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 770.

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan – Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217 – 239.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56 – 75). Thousand Oaks, CA: Sage Publications, Inc.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301 – 322.

Bayesian Multiple – group Analysis: Approximate Measurement Invariance

Song Qiongya　Zhang Lijin　Pan Junhao

(Department of Psychology, Sun Yat – sen University, Guangzhou 510006)

Abstract: In the behavioral science, the comparison of multiple groups under the latent variable framework is popular. Measurement invariance (MI) is a pre – requisite for such multiple – group comparison. Multiple – group Confirmatory Factor Analysis (Multiple – group CFA) is the most commonly used approach for testing measurement invariance. In traditional multi – group CFA, strict invariant constraints are imposed on measurement parameters across groups. However, due to the complexity of modeling multi – group data, these strict constraints are unrealistic in real data analysis and can easily lead to poor model fitting. In fact, scalar invariance is almost unachievable in practice. The Bayesian approximate MI proposed by Muthén and Asparouhov (2013) compensates for these limitations to some extent by providing a zero – mean, small – variance prior for the differences in measurement parameters. It allows for small differences between groups and avoids the problems caused by strict restrictions in classical method, such as poor model fitting, awkward model modifications and higher Type I error rate. These strengths make this new approach a more suitable method in the practical research. This paper introduced the principles and advantages of the approximate MI approach by comparing it with the classical multi – group CFA method. Besides, a real data set was analyzed to demonstrate the validity and application of this approach using Mplus.

Key words: Bayesian estimation; multiple – group structural equation modeling; approximate measurement invariance