

《心理学报》审稿意见与作者回应

题目：作答时间与反应依赖关系建模：基于双因子模型视角

作者：郭小军, 柏小云, 罗照盛

第一轮

审稿人 1 意见：

作者发现以往研究在使用分层模型 (HM) 对被试作答时间和反应进行联合建模时, 并不能很好地解释作答时间和作答反应间的依赖关系。为此, 本文基于双因子模型的视角, 认为分属作答时间和反应的题目分别测量两个局部因子、作答时间和反应两个维度上的所有题目测量一个全局因子, 并由此提出双因子分层模型 (Bi-HM)。研究还通过一个模拟研究比较 Bi-HM 和 HM 的参数返真性, 通过一个实证研究探究 Bi-HM 的真实表现。论文选题比较新颖, 研究结果可为实践者在对被试作答时间和反应进行联合建模时提供参考。另外, 研究设计合理, 结果完备, 解读准确。然而, 论文在“内容完备性”和“表述规范性与准确性”等方面还存在较多问题。

意见 1: 表 1 呈现了作答时间与反应的不同关系类型, 请对部分类型下的两种情况做具体区分和解释。比如, 对于类型二和类型三, 它们各自所包含的两种情况是否有区别? 如果有, 区别在哪里?

回应: 感谢审稿专家的意见, 我们已经增加不同类型的表现, 以区分不同类型的差异。

项目参数 λ 和被试参数 η 可以联合反映被试在五种不同作答时间与反应之间的关系, 具体结果如表 1 所示。从表中可以看到不同类型的表现: 在类型一中, 作答时间与作答准确率是相互独立的, 它们之间没有影响; 在类型二中, 被试更注重准确率, 会花费更多的作答时间来提高准确率, 这类似速度与准确率权衡实验强调准确率而忽视时间; 类型三的被试则更注重作答速度, 因此会通过牺牲准确率来提高作答速度或降低作答时间, 这和速度与准确率权衡实验强调作答速度而忽视准确率的表现类似 (郭小军等, 2019); 类型四表现为最佳作答模式, 被试能够同时保持快速和准确地作答; 最后一种类型五与类型四表现完全相反, 被试既作答拖拉又准确率低。

意见 2: “2.2 参数估计与模型识别”部分, 作者将 Maximum Likelihood Robust (MLR) 翻译为“最大似然加权法”不准确。

回应: 感谢审稿专家的批评与指正, 我们已经将 Maximum Likelihood Robust (MLR) 翻译为稳健极大似然。

意见 3: “3.1 数据生成与分析”部分, 关于被试参数和题目参数的生成方式, 请注明参考出处。另外, 建议呈现评价指标 Bias 和 MSE 的计算公式。

回应: 感谢审稿专家的意见, 我们已经进行相应的补充。

(1) 我们已经在文中补充各参数相应的参考文献。

(2) 对于模拟结果采用 MSE (Mean squared error) 和平均 Bias 进行评价, MSE 和 Bias 值越小意味参数的估计值与模拟值越接近, 结果越佳。

$$\text{MSE}(\hat{\xi}) = \frac{\sum_{r=1}^R \sum_{k=1}^K (\hat{\xi} - \xi)^2}{R * K} \quad (7)$$

$$\text{Bias}(\hat{\xi}) = \frac{\sum_{r=1}^R \sum_{k=1}^K (\hat{\xi} - \xi)}{R * K} \quad (8)$$

其中 $\hat{\xi}$ 和 ξ 分别为估计值和模拟值， R 为重复次数， K 为被试量或测验项目量。

意见 4: “3.2 结果”部分，请对结果描述进行重新梳理：（1）建议重点强调 Bi-HM 和 HM 在不同条件下的 Bias 和 MSE 的变化趋势比较，而非对数字做简单描述；（2）目前是将样本量作为参考点描述 Bi-HM 和 HM 的变化趋势，建议增加题目难度参数 b_j 与作答时间强度参数 β_j 的相关（0、0.4 和 0.8）和测验长度（30 和 60）维度上，两种模型的参数返真性对比描述；（3）建议增加 Bi-HM 和 HM 在不同条件下参数返真性的变化图，提供可视化的结果。

回应: 感谢审稿专家的意见，我们已经修改了内容的描述并增加了变化图。其中变化图见正文，内容描述修改如下：

对 Bi-HM 和 HM 的各参数的估计结果，通过矩阵图和表格分别进行了呈现，以更好的反映其趋势和特点。在矩阵图中，每一行的图形对应 HM 和 Bi-HM 的 MSE 和 Bias 值，每一列的图形对应不同被试量和测验项目量的组合；在每一个行列交叉的图形中，横坐标为 HM 和 Bi-HM 对应的项目参数或被试参数，纵坐标为 MSE 或 Bias 的值，不同线段对应不同的项目相关 $\rho_{a\beta}$ 水平。

图 2 和表 2 呈现了 HM 和 Bi-HM 的各项目参数的估计结果。从图表中可以看出，整体而言，Bi-HM 的各参数估计结果较好，而 HM 的参数估计结果相对较差。在相同被试量（ $N=500$ 或 1000 ）上，Bi-HM 和 HM 的各自项目参数的 MSE 值趋向水平相似，Bias 波动较小。随着被试量从 500 增加到 1000，Bi-HM 的各项参数的 MSE 值都呈现下降趋势，特别是反应模型的区分度参数 α 的 MSE 值从 0.028 降低到 0.011，难度相关参数 d 的 MSE 从 0.02 降低到 0.009， λ_1 的 MSE 从 0.021 降低到 0.009，作答时间模型的各项参数有轻微降低，且各参数的最大 MSE 值为 0.007；所有项目参数的 Bias 值在接近 0 的范围内波动。HM 的各项参数的 MSE 值也随着被试量的增加而略微下降，但是其反应模型的 α 参数的 MSE 从 0.123 降低到 0.083，MSE 值仍然较高，估计结果较差，作答时间模型的 α 参数的 MSE 和 Bias 值也都明显偏高，MSE 值始终在 0.096 附近，Bias 值则在 0.19 附近；其他项目参数的 MSE 值和 Bias 值都相对偏小，估计较好。在不同项目相关 $\rho_{a\beta}$ 水平上，HM 的各项参数的 MSE 和 Bias 值保持稳定，Bi-HM 的各项参数的 MSE 值也保持不变，虽然 Bi-HM 的 Bias 值有所波动，但差异非常微小。

HM 和 Bi-HM 的被试参数的估计结果如图 3 和表 3 所示。从图表中可以看出，在相同测验项目量（ $m=30$ 或 60 ）上，Bi-HM 和 HM 间的各被试参数的 MSE 值随着被试量变化趋势接近，Bias 值虽然趋势不同，但是差异较小。随着测验项目量从 30 增加到 60，Bi-HM 的各被试参数的 MSE 值都明显降低，特别是能力参数 θ 的 MSE 值从 0.157 降低到 0.086，速度与准确率权衡参数 η 的 MSE 值从 0.057 降低到 0.024；HM 被试参数的 MSE 值随着测验项目量增加而有所降低，虽然被试参数能力参数 θ 的 MSE 从 0.283 降低到 0.194，但值仍然较高，估计偏差较大。相对 Bi-HM 和 HM 被试参数的 MSE 值，两个模型的被试参数的 Bias 值则都在 0 附近波动。此外，Bi-HM 和 HM 的各被试参数的 MSE 值在不同项目相关 $\rho_{a\beta}$ 水平上趋于稳定，特别是 Bi-HM 的被试参数的波动完全重叠，Bias 值则呈现随机波动趋势，但是差异较小。

意见 5: 建议在模拟研究部分增加 Bi-HM 和 HM 两种模型的运行效率对比结果。

回应: 感谢审稿专家的意见，文中表述可能还不够清晰，我们已经进一步修改相关表述。

（1）在运行效率方面，我们强调的是基于 Mplus 的 MLR 估计效率要远高于贝叶斯估

计（比如 Openbugs 或 jags）。同时，相对复杂的数学基础和统计软件使用，Mplus 等结构方程模型软件通用性更强，这更有利于本文提出的 Bi-HM 的推广与应用。

(2) 我们对 Bi-HM 和 HM 在不同条件的运行时间进行了统计，结果如表 1 所示。从表中可以看出，随着被试量和测验项目数量的增加，HM 的平均运行时间由 2 秒左右增加到 11 秒多，而 Bi-HM 从 29 秒多增加到约 306 秒。相比之下，贝叶斯估计运行时间可能需要几十分钟甚至数小时，而基于 Mplus 的 Bi-HM 估计只要 5 分钟左右，其运行效率已经非常高。

表 1 不同条件 HM 和 Bi-HM 的平均运行时间（秒）

模型	$\rho_{b\beta}$	N=500				N=1000			
		m=30		m=60		m=30		m=60	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
HM	0	2.05	0.10	6.22	0.35	3.33	0.18	11.15	0.71
	0.4	2.05	0.10	6.52	0.47	3.25	0.15	11.20	0.72
	0.8	2.03	0.08	6.32	0.39	3.42	0.20	11.26	0.74
Bi-HM	0	29.23	4.05	160.78	17.79	84.31	11.50	303.17	15.18
	0.4	29.96	3.76	153.25	13.34	63.84	7.18	306.98	14.74
	0.8	30.09	3.50	157.72	17.35	84.58	11.07	307.73	15.01

注：HM=van der Linden 的分层模型；Bi-HM=双因子分层模型。

(3) 经过慎重考虑，我们认为在模拟研究部分增加 Bi-HM 和 HM 两种模型的运行效率对比结果可能不太恰当。原因在于，HM 是嵌套于 Bi-HM，Bi-HM 需要估计的参数明显要多于 HM。因此，在相同参数估计方法下，Bi-HM 的运行时间肯定是要高于 HM（如表 1 所示）。此外，我们也认同考虑模型的运行效率是非常重要的，并打算在后续研究中比较 Bi-HM 在贝叶斯估计与 MLR 两种参数估计方法下的运行效率。

意见 6: “4.2 分析”部分，建议补充模型拟合指数的评判标准，即“越小表示模型拟合效果越好”。而且，建议在文字描述部分补充“将对数似然 LL 作为拟合评价指标”（也即表 4 中呈现的指标需逐一在“4.2”部分阐述清晰）。

回应: 感谢审稿专家的意见，我们已经补充 LL 拟合评价指标，并进行了统一描述。

不同模型优劣采用 LL (Loglikelihood)、AIC (Akaike's information criterion)、BIC (Bayesian information criterion) 和 SABIC (Sample-Size Adjusted BIC) 四个模型拟合指数进行评价。在不同模型比较中，四个模型拟合指数值越小，该模型拟合效果越好，也越倾向选择该模型为最优模型。

意见 7: 图 2、3 和 4，请将数字字体统一为 Times New Roman，将文字统一为宋体。图 3 的相关系数显著性用规范方式呈现，图 3 的相关系数需统一字号，建议对图 3 做进一步的美观化处理。

回应: 感谢审稿专家的意见，我们已经对文中所有图形全部进行了修改，并进行了美观化处理，具体见文中图形。

意见 8: 实证研究采用《瑞文标准推理测验》，请在“4.3 结果”部分增加“Bi-HM 对瑞文推理测验结果的心理分析”的内容，以此凸显 Bi-HM 在心理学研究中的优势，而非简单地以数据驱动方式探讨 Bi-HM 和 HM 在实证数据拟合上的区别。

回应: 感谢审稿专家的意见，我们已经增加 Bi-HM 对瑞文推理测验结果的心理分析，以凸显 Bi-HM 的优势。

在心理和教育测验中，有许多因素会影响被试者的作答过程和结果。其中，被试的能力水平会影响其作答的准确性，作答速度则会影响到其完成测验项目所需的时间。同时，作答速度和准确性之间还存在一种权衡关系。这些影响因素在不同类型的测验中普遍存在，例如能力测验、人格测验等。为了更好地考虑这些影响，文中提出了 Bi-HM。下面以《瑞文标准

推理测验》为例，具体探讨 Bi-HM 在这一测验中的心理学分析，以凸显该模型的特点。从被试作答过程的总体心理来看，Bi-HM 的大部分项目的 λ_1 值为正值（除难度偏低项目 1、3、15、25 和 37 为负值），而 λ_2 在所有项目上均为正值（图 5），结合被试的速度与准确率权衡能力 η 的正负，可以得出被试在作答过程中主要包括两种作答心理状态：一种是被试通过花费更多的时间以提高准确率的作答心理（表 1 类型二），另一种是牺牲准确率以降低作答时间的作答心理（表 1 类型三）。针对具体的项目作答心理，根据 λ_1 和 λ_2 参数分布选择了两个典型项目（项目 55 和 37），结果如图 6 所示。图中不同直线代表速度与准确率权衡能力 η 为正或负的被试在作答过程中的作答趋势。从图 6（a）可以看出，速度与准确率权衡能力 η 为正（+）的被试在项目 55 上更倾向于花费更多时间以提高准确率（表 1 类型二），而速度与准确率权衡能力 η 为负（-）的被试在该项目上更倾向于牺牲准确率以降低作答时间（表 1 类型三）。从图 6（b）可以看出，速度与准确率权衡能力 η 为负（-）的被试在项目 37 上既注重准确率，又注重速度（表 1 类型四），而速度与准确率权衡能力 η 为正（+）的被试在该项目上的作答速度慢且准确率低（表 1 类型五）。因此，不同项目上的作答时间与反应之间存在不同类型的关系，进而对被试的作答准确率与时间产生不同影响。

意见 9：“5 总结和展望”部分，作者提及“在模拟研究中，发现基于 MLR 的 MPLUS 程序能有效估计 Bi-HM 的参数并且参数估计结果不受题目参数相关水平的影响，这与 Molenaar 等人(2014)结果一致”。但是，Molenaar 等人(2014)使用的是广义线性因子模型“Generalized linear factor model”，请核实引用是否有误。

回应：感谢审稿专家的批评与指正，我们已经对文章进行了相应修改。在 Molenaar 等人（2015）的研究中，对比了贝叶斯估计和 Mplus 的估计结果，其中贝叶斯估计考虑了项目参数间的相关，而 Mplus 估计则未考虑项目参数间的相关。结果发现贝叶斯估计结果与 Mplus 的估计结果是一致的，这表明 Mplus 忽视项目参数间的相关并不会对“Generalized linear factor model”参数估计结果产生影响。因此，我们强调的是 Bi-HM 模型在 Mplus 估计中忽视项目参数间的相关性的估计结果与 Molenaar 等人（2014）结果是一致的。

模拟研究发现，基于 MLR 的 Mplus 程序能有效估计 Bi-HM 的各参数，并且参数估计结果不受项目参数间的相关水平的影响。

意见 10：部分参考文献的年份不一致。比如，5 总结和展望部分（Molenaar et al., 2014），而在参考文献中则是 Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. (2015)…。请通篇检查修改，并检查 APA 格式引用是否准确。

回应：感谢审稿专家的批评与指正，Molenaar 等人文章最早 online 为 2014，但是正式出版是 2015，同时我们已经对文中所有参考文献进行了检查。

意见 11：请重新通读全文并做修改，因为存在多处表述不通顺或不准确，例如：

（1）“摘要”部分：“基于双因子模型视角，测验的作答时间与反应可视为测量了局部因子的两个题组，而所有项目测量了被试的速度与准确率权衡的一般能力或全局因子”；

（2）“1 引言”部分：存在多处“的地得”不分的情况；“作答时间与反应数据可以视为两个不同组测验”？数据被视为测验？“作答时间项目测量…作答反应项目测量…”？什么是作答时间项目？什么是作答反应项目？“测验项目的作答时间与反应数据等同于局部因子的两个维度”？数据等同于维度？

（3）“2.1 模型构建”部分：“先介绍 van der Linden（2007）HM”？；“被试 i 在项目 j 的作答反应为两参数 logistic 模型”？另外，如果 2PLM 采用公式（1）描述的话，那么 d_j 就不是难度参数，而代表区分度和难度的乘积；“在 HM 的第二层为被试与项目参数间的相关”？准确地讲，应该是“被试参数间的相关和项目参数间的相关”；

（4）“3.2 结果”部分第 5 行，“降低到和 0.009”？

回应：感谢审稿专家的意见，我们已经对全文进行了检查与修改。

（1）基于双因子模型视角，测验的作答时间与反应分别测量了被试速度和能力两个局部因子，而作答时间与反应又综合测量了被试的速度与准确率权衡的一般能力或全局因子。

(2) 在心理与教育测验中, 不仅可以采集被试在测验项目上的作答时间, 还可以采集被试的作答反应。从双因子模型的视角看, 测验项目的作答时间与反应可以视为测量了不同的局部因子。具体而言, 测验的作答时间测量了被试的速度特质, 而测验的作答反应测量了被试的能力特质。同时, 被试在作答测验过程中还会受到时间与准确率的综合影响 (Davison, Semmes, Huang, & Close, 2012; Bolsinova & Tijmstra, 2017), 即一般潜在特质或全局因子, 或速度与准确率权衡能力 (Chen, De Boeck, Grady, Yang, & Waldschmidt, 2018)。因此, 这样的测验结构非常符合双因子模型的特点, 为解决测验作答时间与反应依赖关系的联合分析提供了全新视角。

(3) “2.1 模型构建” 语言表述已进行了重新修改, 比如“先介绍 van der Linden (2007) 提出的 HM”、“作答反应模型为两参数 logistic 模型, 可以表示为”、“ d_j 为项目难度相关参数”、“在 HM 的第二层为被试参数间的相关与项目参数间的相关”

(4) 我们已经根据意见 4 对模拟结果的描述重新进行描写, 具体见正文。

.....

审稿人 2 意见:

论文“作答时间与反应依赖关系建模: 基于双因子模型视角”, 基于双因子模型的假设, 对作答时间和反应的关系进行建模, 具备一定的合理性和创新性。通读文稿, 表达流畅、简洁, 逻辑清晰, 结构完整, 是一篇较高质量的学术论文。

意见 1: 参数生成范围的依据是什么?

回应: 感谢审稿专家的意见, 对文中的参数生成的依据我们已经列出相应的参考文献。

意见 2: 根据 BI-HM 生成模拟数据, 因此参数估计结果表明 BI-HM 的结果优于 HM, 是不是应该有 HM 生成的数据进行两个模型参数估计的结果对比? 另外, 有没有一般性的方法, 并不是依据这两个模型生成数据的方法, 来进行模型参数估计分析呢?

回应: 感谢审稿专家的意见。

(1) 我们的模拟研究的核心目的是检验基于 Mplus 估计方法对 Bi-HM 参数估计的准确性。在这个过程中, 我们同时考虑了忽视作答时间与反应依赖关系对 HM 参数估计结果的偏差情况。然而, 基于 HM 模拟数据, 并将其与 Bi-HM 进行对比估计, 会偏离我们研究的主要目的, 因为 HM 本身相对于 Bi-HM 存在一些局限性。此外, 类似的单向模拟研究思路在以往的研究中已经广泛存在 (Lee & Ying, 2015; Man, Haring, Jiao, & Zhan, 2019; Meng, Tao, & Chang, 2015)。因此, 我们经过慎重考虑认为添加 HM 模拟部分可能不太恰当。

Lee, Y.-H., & Ying, Z. (2015). A Mixture Cure-Rate Model for Responses and Response Times in Time-limit Tests. *Psychometrika*, 80(3), 748–775.

Man, K., Haring, J. R., Jiao, H., & Zhan, P. (2019). Joint modeling of compensatory multidimensional item responses and response times. *Applied Psychological Measurement*, 43, 639–654.

Meng, X.-B., Tao, J., & Chang, H.-H. (2015). A conditional joint modeling approach for locally dependent item responses and response times. *Journal of Educational Measurement*, 52(1), 1–27.

(2) 目前还没有一种通用的模拟方法来考虑作答时间与反应的依赖关系, 但这可以作为未来进一步研究的方向, 以综合比较不同模型的估计偏差。

意见 3: 3.1 部分“项目区分度参数 a_j 和项目作答时间区分度参数 α_j 服从左截尾的正态分布 $N(0,1)I(0,)$ ” 写作有误?

回应: 谢谢审稿专家的意见。经过仔细检查, 确认写作无误。在截尾分布中, 通常使用 $I(a,$

b)符号表示,其中 a 表示左截尾,即取值大于 a , b 表示右截尾,即取值小于 b ,当缺失某一部分时可以省略。在软件 jags 和 openbugs 等中也采用这种模式进行标识。

意见 4: 4.2 部分实证数据的分析汇总为什么“固定项目水平的速度与准确率权衡参数 λ_{1j} 和 λ_{2j} 为 1”,这和表 1 依据两者的关系说明反应类型是“重视作答速度,忽视反应正确性”,这个约束是不是违背瑞文推理测验的初衷,因此最后的结果并不好?

回应: 谢谢审稿专家的意见,我们已经补充增加该模型的原因。当 Bi-HM 的“ λ_{1j} 和 λ_{2j} 为 1”时,被试作答时间与反应的关系只存在两种,分别为 η 为正 (+)时“忽视作答速度,重视准确率”,以及 η 为负 (-)时“牺牲准确率,提高作答速度”。

同时,已有研究从项目水平和被试水平分别探讨了作答时间与反应的依赖关系 (Bolsinova et al., 2017b; Meng et al., 2015; Ranger et al., 2012)。因此,为了进一步探讨 Bi-HM 模型的性能,将增加固定项目参数 λ_{1j} 和 λ_{2j} 为 1 的双因子分层模型 (Fix-Bi-HM) 来拟合数据。

意见 5: 图 2 部分描述“基于参数估计结果对两个模型作答时间的残差通过取绝对值再计算项目水平的平均绝对残差”,根据这句话,得到的是两种模型下每个项目的作答时间的平均绝对残差,得到项目和对应的平均绝对残差值,是如何能够反映在图 2 中呢?图 2 的横纵坐标分别是 HM 和 BI-HM?似乎应该是两个模型得到的每个项目的作答时间的绝对值的散点?另外,参数估计的结果是以什么作为标准的?

回应: 谢谢审稿专家的意见。

(1)我们分别计算了 HM 和 Bi-HM 两个模型在项目水平上的作答时间的平均残差和残差方差,并以 HM 的值为纵轴,以 Bi-HM 的值为横轴,绘制散点图,同时添加了对角虚线方便识别点的位置,两个模型的每个项目的值构成一个坐标点的位置。如果两个模型的值完全相等,点将会落在对角线上;如果 HM 的值较大,点会落在对角线的左边或上方;如果 HM 的值较小,点则会落在对角线的右边或下方,例如 A(1,1)、B(2,1)和 C(1,2)是三个点的坐标分别对应上述三种情况。因此,我们可以通过 HM 和 Bi-HM 在散点图中的点位置来判断残差的大小以及拟合的优劣程度。

(2)参数估计的结果是以 Mplus 的估计结果为标准。本研究的模拟和实例数据均采用 Mplus 的 MLR 估计方法对 HM 和 Bi-HM 两个模型进行参数估计。

意见 6: 图 3 对应的文字中说“同时从 λ_1 和 λ_2 的直方图可以发现大部分项目的 λ_1 值为正值,而 λ_2 在所有项目上均为正值,”可是图中 λ_1 的直方图表明其为负值。图三包含了很多信息,但是缺少必要的标目和说明,使得读者不易读懂图形的意思。

回应: 谢谢审稿专家的批评与指正。

(1)在实例数据中, λ_1 和 λ_2 分别对应 60 个项目参数估计值,因此 λ_1 的直方图表明部分项目是存在负值的,比如项目 1、3、15、25 和 37 均为负值。

(2)图三我们已经增加了图形说明,以便读者更易读懂图形的意思。在相关矩阵图中,对角线的直方图为各参数的直方图分布,下三角为各参数的两两散点关系图,上三角为各参数的两两相关系数。

第二轮

审稿人 1 意见:

针对审稿人意见,作者不仅认真回复,而且在稿件中做了相应修改。当前的版本只有一

个小的问题，建议“小修后发表”。

补充的 Bias 和 MSE 计算公式目前呈现在模拟研究目的和“3.1 数据生成和分析”之间，从逻辑结构上看可能不太合适，可以新增一个小的章节——“评价指标”。

回应：感谢审稿专家的建议，我们已经在模拟研究目的后面增加了一个小节“3.1 评价指标”，并将补充的 Bias 和 MSE 计算公式和解释置于此节中。

编委意见：

当前的版本只有一个小的问题：补充的 Bias 和 MSE 计算公式目前呈现在模拟研究目的和“3.1 数据生成和分析”之间，从逻辑结构上看可能不太合适，可以新增一个小的的章节——“评价指标”。

回应：感谢编委专家的建议，我们已经在模拟研究中重新增加了一个小节“3.1 评价指标”，具体见正文。

主编意见：

针对分层模型存在的不足，本论文提出了双因子分层模型，并通过了模拟研究和实证研究对该模型的表现进行了检验。该论文的选题具有一定新颖性，研究设计合理，数据处理过程规范，获得的研究结果具有较为重要的理论和实际应用价值。