

《心理学报》审稿意见与作者回应

题目：缺失数据的结构方程建模：全息极大似然估计时辅助变量的作用

作者：王孟成 邓俏文

第一轮

审稿人 1 意见：

该论文通过蒙特卡洛模拟的方法分析了在采用全息极大似然估计进行缺失值建模时辅助变量的作用。考虑的因素包括：共缺机制、共缺率、相关程度、辅助变量数目、及样本量。已有研究没有系统考虑共缺机制和共缺率，因此，本研究是对已有研究很重要的补充，具有一定的学术价值。但如果发表于《心理学报》上，还需要对某些细节做进一步的思考和完善。

我有三点主要建议和意见供作者考虑：

意见 1：我个人感觉该模拟研究考虑的因素过多，造成研究结果难以很好地呈现在有限的篇幅之内。我想这也是为什么作者把很多结果都以附表的形式附在文章的后面。在一个模拟研究比较忌讳考虑过多的因素，研究者一般考虑最重要的几个因素，其它因素设置为固定值。本研究最重要的因素应该是共缺机制、共缺率、辅助变量的数目，因为已有研究没有考虑过这些因素。其他因素如样本量是不是可以固定为“大样本”情况比如 500？另外，辅助变量的数目是不是可以只考虑 1 和 2 的情况？共缺率的四种情况是不是过于接近，能否只考虑 0.10 和 0.20 两种情况？因为 0.05 的缺失在实际数据处理中是可以忽略的。如果这样设计的话，模拟设计将会大大简化，模拟的组合将大大减少，每个因素是如何影响的将会很清晰，结果也容易解释和呈现，对实际应用的指导意义才好阐述。

回应：感谢评审人提出的意见和建议。您的这些考虑和建议都非常专业，说到我们心坎上了。我们在研究设计以及结果整理和论文撰写的过程中也都有同样的顾虑。但经过对国外研究的梳理我们发现，要想达到我们的研究设想、解决前人研究未解决的问题必须要考虑多种条件，在衡量这些条件不同水平的组合后，最终确定了本研究所保留的研究条件及组合，进而造成了研究结果呈现的不便。尽管存在这些困难，我们希望通过细致的描述可以让读者能够抓住本研究的重点和主要发现。

另外，关于您提到的固定样本量的建议，我们认为：如果在研究设计时只考虑 $N=500$ 时的情况，应该也是可行的，但我们通过设置不同的样本量（综合前人的研究，提出 4 个样本量 100, 200, 500, 1000）发现，样本量在不同条件下确实会对结果产生影响，所以我们还是考虑报告不同样本量条件下的结果。同样地，对于您提出辅助变量数的建议，我们也选择报告 3 种数目的结果。

关于您的第三个问题（对共缺率的考虑），可能由于我们没有交代清楚。这里的缺失率指共缺率。共缺指同一个体的数据在研究变量上有缺失，在辅助变量上也有缺失，

共缺率则是共缺频数在样本中的比例（见第七页标蓝处）。这意味着 0.05 的共缺率所代表的是缺失率的最小值，实际存在缺失率（非共缺率）可以很大。

意见 2: 模拟情况 1 为辅助变量不缺失的情况，不可能有 6 种缺失机制组合，只有 Y 对应的两种缺失机制；也不存在共缺率的概念，因为只有 Y 缺失。文中的附表 1, 2, 和 6 中对情况 1 有 6 种缺失机制是不正确的。希望作者仔细考虑。

回应: 感谢评审人细心的指正，非常抱歉我们在文章中没有交代清楚，这里的表达方式造成评审人的误解。确实，辅助变量不缺失情况下的 6 种缺失机制组合中只有 Y 变量存在缺失，所以应在上述表格中的备注处添加相关说明。再次感谢您的认真，我们已在相应表格的备注处做了修改。

意见 3: 某些部分（具体见下面）阐述不够清楚，希望在有限的篇幅内把重要的内容清楚地表达出来。

回应: 非常感谢您的反馈和意见。我们会逐条修改。

其它意见:

意见 4: 第一部分“引言”和第二部分“问题的提出”逻辑不够清楚，显得比较紊乱，内容应该重新组织。这两部分其实应该合并为“引言”或者“问题的提出”。这部分可以先对缺失值处理的重要性和目前相关研究的不足进行总结，然后提出本研究的目的。我个人认为，文中的 1.1 和 1.2 和 1.3 应该专门放在一部分“缺失值处理方法”。也就是说，第一部分为“引言”，第二部分为“缺失值处理方法”，第三部分“研究设计”，然后为“研究结果”“结论和讨论”。

回应: 衷心感谢您提出的问题与建议。针对您的建议，我们对文章内容进行了认真研读，由于方法模拟类的文章涉及的专业术语比较多，而且不容易懂，所以我们才顺着研究思路分为引言和问题提出两部分，并且在问题提出之前对缺失值处理方法的相关内容进行了介绍，这也就导致您所说的引言和问题提出分为了两部分，而且 1.1、1.2、1.3 穿插在引言部分的紊乱问题。为了方便读者明白研究目的的同时更容易读懂本文，我们认为这样的逻辑比较适合对该领域不甚熟悉的读者。不过，我们也得承认，您的建议是有道理的，所以为了让逻辑更加清晰，我们对 1.1 的内容进行了修改使之更加精炼。

意见 5: 本文主要关注全息极大似然估计对缺失值的处理，所以应该对这个方法进行较详细的描述。文中只是给出公式 1 和 2，并没有对公式中的每个项或符号的含义进行任何说明，这是不合适的。希望给出更多的基本信息，让第一次看到这个方法的研究者也可以大致看懂。

回应: 感谢您的建议。我们已经在公式下方对相关内容作出了补充（见第 5 页，标红）。

具体如下：其中， k 为变量个数， Y_i 为数据 i 的分数值， μ 和 Σ 分别为总体均值和协方差

矩阵。这个公式最重要的部分是马氏距离 $(Y_i - \mu)^T \Sigma^{-1} (Y_i - \mu)$ ，似然函数值的大小主要受这部分值的大小影响。公式 2 中的 k_i 为完全数据的个数，其他符号所代表的意义和公式 1 一样。

意见 6: 该研究设定辅助变量和研究变量的相关程度为 0.3（低）和 0.6（中高），为什么同时也设定辅助变量和 X 的相关程度为 0.2 和 0.5？为什么不固定辅助变量与 X 之间的相关程度？

回应: 感谢您提出的这个问题，这确实是一个让人容易迷惑的地方。我们在一开始研读这类型研究的时候也出现同样的疑问。后来我们发现，在结构方程模型中，低相关或者中等偏高相关实际包含辅助变量与研究变量（包括 X 与 Y 变量）的相关，不仅仅是辅助变量与 Y 变量的相关，所以不但不需要控制辅助变量和 X 的相关程度，从而研究辅助变量与 Y 的相关程度，而且需要设置辅助变量与 X 的相关程度随辅助变量与 Y 变量的相关程度的高低而变化。因为在结构方程模型中，如果辅助变量与 Y 变量呈低水平的相关，由于 X 与 Y 之间的关系固定，辅助变量也会和 X 变量存在较低的相关。同理，如果辅助变量与 Y 变量呈高水平的相关，由于 X 与 Y 之间的关系固定，辅助变量也会和 X 变量存在较高的相关。

另外，类似设置同样出现在以下的模拟研究中：

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 6, 330–351.

Enders, C. K. (2008). A note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling*, 15, 434–448.

意见 7: 模拟数据的生成过程需要较详细的说明，不能只给出软件。

回应: 在本研究中，关键部分不仅仅是对研究设计或说变量类型的考虑上，很重要一部分也是模拟数据的生成问题，所以我们能理解您对数据生成过程说明的要求。虽然模拟数据的生成的基本要素我们已经在文中的研究假设模型部分详细说明（见 3.1 模拟研究设计 8-10 页），具体如下：

在参考前人研究(Enders & Peugh, 2004; Hardt et al., 2012)的基础上，本研究考虑如下两组相关水平：低相关设为 $\rho_{xz} = .2$, $\rho_{yz} = .3$ 和中等偏高相关为 $\rho_{xz} = .5$, $\rho_{yz} = .6$ 。所有变量均值为 0，方差为 1(服从正态分布)，固定因子方差为 1，因子负荷为 .70，残差方差为 .51。存在多个辅助变量时，综合 Hardt 等(2012, $r_{zs} = .1$ 或 $r_{zs} = .5$)与 Enders 和 Peugh(2004, $r_{zs} = .3$)的研究，本研究设定辅助变量之间的相关系数为 .4。

与先前类似研究一致(Enders, 2008)，本研究采用逻辑回归生成缺失数据。具体来说，当 Y 观测变量的缺失机制为 MAR 时，设定 Y 的缺失与辅助变量 Z 有关，斜率参数为正，即 Z 的值越大，Y 的缺失率越大。当辅助变量 Z 的缺失机制为 MAR 时，设定 Z 的缺失

与 X 观测变量有关，斜率参数设为负，即 X 观测变量的值越大，Z 的缺失率越小。当辅助变量 Z 的缺失机制为 MNAR 时，Z 的缺失与自身相关，设定斜率参数为负，即 Z 的值越大，Z 的数据缺失率越小。这样，当共缺组合形式为 MAR-MNAR 时，辅助变量可为研究变量提供更多的信息，同时可减少研究变量与辅助变量的缺失都是由辅助变量的缺失造成的可能性。

尽管在第 8-10 页对本模拟研究的模型设定及相关变量（相关系数、缺失机制组合等），但读者在阅读这些描述的时候依然不清楚数据究竟是怎么生成的。所以，为了更清楚说明数据生成的程序在软件中是如何表述的，我们在附录中增加了一个本研究模拟数据的 Mplus 语句，见附录 10。

意见 8: 估计的参数也应该给出。因为本文主要研究各个因素对参数估计结果准确性的影响，所以需要估计的模型参数有哪些需要明确指出。

回应: 感谢您的建议。我们对您的这个问题有两个方面的理解。第一：文中没有明确描述本次模拟研究需要估计的模型参数，尽管在结果部分一直在谈论因子负荷和回归系数的估计偏差和覆盖率。经过认真考虑，我们承认这方面确实没有明确的说明，所以衷心感谢您的耐心指出。我们已经在文中结果部分补充了关于这方面的说明：

综合前人的研究 (Collins et al., 2001; Enders, 2008)，本研究主要考虑因子负荷和回归系数的估计值。

另外，对您提出的这个问题，我们的第二种理解是文中的表格或附表没有给出相应的参数值。对于这个问题，我们也认真考虑过。但由于表格空间的有限性，最后在呈现结果的时候，我们只给出了偏差的结果。为了便于读者对需要估计的参数有个整体的把握，我们添加了一个表格，把需要估计的模型参数、理论值、估计值、标准差和标准误等都呈现出来了，见附表 9。

意见 9: P.10 给出了模拟情况 2 和 3 中控制的因素，对情况 1 中控制的因素没有说明。情况 1 为辅助变量不缺失的情况，不可能有 6 种缺失机制组合，只能有 Y 对应的两种缺失机制；也不存在共缺率的概念，因为只有 Y 缺失。文中的附表 1 和 2 中有 6 种缺失机制是不正确的。

回应: 感谢您再次提出这个问题，显然这个问题也很容易让人混淆，也很抱歉我们一开始没有添加更多的说明。确实，模拟情况 1 即辅助变量不缺失情况下的 6 种缺失机制组合中只有 Y 变量存在缺失，而此种模拟情况是由模拟情况 2（辅助与研究变量共缺）设定辅助变量不缺失得到的，或许如果一不小心跳过了这句话，就会对这种模拟的结果描述产生误解。我们已在文中相应位置和相应表格的备注处做了修改。

文中的补充：第一种模拟条件由第二种模拟条件设定辅助变量不缺失得到，其中的缺失机制组合形式实质上只是研究变量的缺失机制，但这两种条件下研究变量的缺失率相同。（见 3.1.5 数据生成部分）

表格备注添加的说明：辅助变量不缺失时，各种缺失组合机制指辅助与研究变量共缺条件下对应的但只有研究变量存在缺失的情况。

意见 10: P.10: 公式 3 和 4 的几种简化形式中的估计值应该是“平均估计值”吧？为什么偏差 SE 和偏差 SD 结果是接近的？我们知道标准误和标准差是不等的，特别是样本量大的时候，为什么它们的结果会接近？偏差 100 是公式 4 的简化，并不是公式 3 的简化，而偏差 SE 和偏差 SD 是和公式 4 不同的概念，所以不能称为公式 3 和 4 的简化。这 3 个偏差各自的优劣是什么？请对这些评价标准进行正确清楚的描述和解释。

回应: 感谢您提出的问题。关于您的第一个问题：由于本文中 X 因子和 Y 因子各有 3 个指标，所以计算 X 因子和 Y 因子的估计值、标准误、标准差都是用平均估计值。已在结果评价指标中对相应的内容进行了修改。

文中的修改：其中因子负荷的结果(偏差与覆盖率)由对应结果求平均数得到(除以自由估计的因子负荷的个数)。

关于您的第二个问题：由于本文模拟的次数达到 5000 次，所以偏差 s_{E} 和偏差 s_{D} 结果是接近的。但是像您一开始提到的本次模拟研究考虑的因素比较多，这样会导致结果呈现的不便。所以，通过与前人研究的比较，我们打算选用最常用的偏差指标，即偏差 s_{E} 和偏差 100，同时统一表格呈现的偏差指标为偏差 s_{E} 。

关于您的第三个问题，感谢您的细心指正，确实，偏差 100 是公式 3 和 4 的简化，偏差 s_{E} 和偏差 s_{D} 在这里没有给出数学公式。相应的补充与说明我们已经在文中相应位置进行了修改（也见 3.2 结果评价指标部分），具体如下：

文中的修改：比较常用的估计偏差的指标是标准偏差(Standardized Bias; Collins et al., 2001; Enders & Gottschall, 2011)，标准偏差 = (平均估计值 - 理论值) / 平均标准误(以示区分，记为偏差 s_{E})。如果偏差 s_{E} 等于 -.50，意味着该平均估计值处于理论值 -.50 个标准误的位置。Collins 等(2001)指出偏差 s_{E} 小于 .40 为无偏估计，后来的研究也采用此标准(Enders & Gottschall, 2011)。但有研究质疑以 .40 作为判断标准的适切性(Graham, 2009)，因此当偏差出现的情况较多时，本研究同时采用另一偏差指标(偏差 100)作为结果评价指标。

意见 11: 覆盖率怎么计算的？含义到底是什么？从文中得不出相关信息。另外，在文中的结果部分，对覆盖率结果的解释也非常少。

回应: 这确实是个需要认真说明的问题。在模拟研究中，覆盖率表示每次重复模拟计算所得结果等于/接近真值的比例，类似于参数区间估计(频率论)上的置信区间(confidence interval)：区间包含的真值。例如，覆盖率为 .95，意味着 1000 次的抽样中(模拟计算)，有 950 次得到的估计值包含真值，为此我们在 3.2 结果评价标准增加了如下内容：

在模拟研究中，覆盖率表示每次重复模拟计算所得结果等于/接近真值的比例，类似于参数区间估计（频率论）上的置信区间（confidence interval）：区间包含的真值。例如，覆盖率为.95，意味着 1000 次的抽样中（模拟计算），有 950 次得到的估计值包含真值。前人的研究认为覆盖率小于.90 是不可接受的(Collins, Schafer & Kam, 2001; Enders & Peugh, 2004)，本研究也采用.90 作为覆盖率是否可接受的标准。

意见 12: P.11: 文中说“本研究中因子负荷的结果有对应结果求平均值得到”，请问是求谁的平均值，除以因子负荷的个数还是重复次数 5000？请在文中说明。

回应: 感谢您的问题，非常抱歉我们没有交代清楚。由于本文中 X 因子和 Y 因子各有 3 个指标，所以计算 X 因子和 Y 因子的估计值、标准误、标准差都是用平均估计值。已在结果评价指标中对相应的内容进行了修改。

文中的修改：其中因子负荷的结果(偏差与覆盖率)由对应结果求平均数得到(除以自由估计的因子负荷的个数)。

意见 13: 4.1 部分: 辅助变量不缺失时，结果有 6 种缺失机制，应该是不正确的。另外，随着“样本量的增大，单辅助变量的优势突出”从附表 1 得不到支持。附表 1 中可以看出相反的结论：辅助变量越多，偏差越小。

回应: 感谢您再次指出这一问题确实，辅助变量不缺失情况下的 6 种缺失机制组合中只有 Y 变量存在缺失，而此模拟情况是由模拟情况 2（辅助与研究变量共缺）设定辅助变量不缺失得到的，或许如果一不小心跳过了这句话，就会对这种模拟的结果描述产生误解。所以我们不仅应该在文中添加更多的说明，还要在上述谈到的表格中的备注处添加相关说明。类似的回答见对问题 2 和问题 9 的回应。

对于“随着样本量的增大，单辅助变量的优势突出”这一问题，感谢您的细心指正。的确，我们对这一结果的表达欠妥。在这里我们是想表达：对于回归系数的参数估计而言，随着样本量的增大，纳入单个辅助变量依然是有益的。

从附表 1 所得到的结论：辅助变量越多，偏差越小。这一结论对于因子负荷的结果来说是正确的，但对于其他参数并不适用。我们已将相关内容修改为：

总的来说，样本量越大，因子负荷和回归系数的偏差值越小(结果见附表 1 和附表 2)。对于小样本($n = 100$)，辅助变量越多，回归系数的偏差越大。对于回归系数的参数估计而言，随着样本量的增大，纳入单个辅助变量依然是有益的。另外，辅助变量越多($n = 100$ 除外)，X 因子负荷的偏差越小。

意见 14: 为什么在表 2 用偏差 SE,而在表 3 中用偏差 100? 表 3 中 CAR 应该是 MCAR, NAR 应为 MNAR。

回应：感谢您的细心研读，表 3 中的表述已改正。至于结果呈现部分是偏差_{S.E.}还是偏差₁₀₀。经过认真的思考和检查，我们已对相应的表格补充统一的结果：指标偏差_{S.E.}。

意见 15：附表 6 中辅助变量不缺失时也不应该有 6 种共缺机制。所以，在 p.13 对第一种和第二种模拟条件下的结果比较是有问题的。

回应：再次感谢您的耐心指正，也为我们描述不够清晰这一问题而导致您的多次疑惑感到抱歉。对于辅助变量不缺失时的 6 种缺失机制的表述问题见问题 2、问题 9 或问题 13 的回应。在本文中，由于第一种模拟条件由第二种模拟条件设定辅助变量不缺失得到，即第一种模拟条件中，辅助变量不缺失，也没有所谓的辅助变量的缺失机制。但第一种模拟条件由第二种模拟条件得到，他们的 Y 变量的缺失率是一样的，即是说，通过第一种模拟条件和第二种模拟条件的比较，可以比较 Y 变量缺失率相同的时候，纳入辅助变量与不纳入辅助变量的差别。所以这里缺失机制组合的表述是为了便于与其他模拟条件的比较。

意见 16：为什么在结果中仅考虑 X 对应的因子负荷的估计偏差，而不考虑 Y 对应的因子负载偏差？

回应：感谢您留意到这一问题。在文中，当使用 X 因子负荷表述时，所指的是“X 的因子负荷”，如果使用“因子负荷”的表述则指 X 和 Y 的因子负荷。为了让读者更加容易理解这一差异，我们已在文中相应位置进行补充说明：

在此模拟条件下，X、Y 因子负荷的覆盖率都在可接受范围内。在高相关、MAR-MAR 组合条件下，纳入 5 个辅助变量时，X 因子负荷的覆盖率比其他辅助变量数目条件下的稍高(差异在 0.01-0.02 之间，结果见附表 4)。Y 因子负荷不见此特点。(4.2 辅助与研究变量共缺：建模时纳入辅助变量部分)

意见 17：P.16：当谈到辅助变量数目的影响时，作者说“当辅助变量与研究变量共缺时，纳入多于一个辅助变量的效果更好，尤其是对于 MAR 机制的研究变量”，而我从表 2 中的结果并不能看出这一点。比如，MR=0.15 时，对 MAR-MAR 机制，纳入 5 个辅助变量的估计偏差是最大的。

回应：感谢您提出的这个问题，这里的表述确实容易让人产生混淆，容许我们进行更正。修改的内容如下：

本研究发现，当辅助变量与研究变量存在共缺时，对于 MAR-MAR 组合机制，纳入单个辅助变量是有益的；对于 MAR-MCAR 或 MAR-MNAR 组合机制，纳入多于一个辅助变量的效果更好。

审稿人 2 意见：

研究中存在几个问题需要作者进一步澄清或改进：

意见 1： 作者在第 7 页研究目的的第三条指出，有必要在研究变量与辅助变量的共缺机制情形下就结构方程模型的参数估计问题进行探讨。在后续的统计模拟中也进行了相应的实验设计和模拟。然而，在第 13-17 页的讨论与结论中，作者没有指出在实际数据分析中，什么时候会出现模拟中考察的共缺机制，即能否告诉读者：如何判断实际数据的数据缺失机制是否满足模拟设计下的共缺机制情况？

回应： 您这个问题提的非常专业，关于缺失值机制的检验问题一直都是该领域研究的难点，目前对其的研究也不是很多。尽管我们在本研究中设计了几种缺失值机制的组合，但是并未涉及如何判断实际研究中如何检验其机制的问题，我们也没有打算这么做，因为这个问题超出了本研究所打算解决的问题。但是，我们在讨论部分结合三种缺失值机制，给出了应用研究者可以根据自己的研究简单判断的缺失值机制的说明。

意见 2： 作者无需在修改稿中补充新的模拟实验，但请回答以下问题：模拟研究设计中的样本量为何最大取为 1000？Mplus7.0 完成模拟实验的效率如何？比如，当样本量为 2000, 3000 时，Mplus7.0 做相关问题研究的能力是否受到限制？

回应： 感谢审稿人提出这个有趣的问题。

关于您的第一个问题：样本量的选取主要考虑到心理学研究中样本量并不是很大，通常在 200-1000 之间，类似的样本量也见本领域的相关研究(e.g., Enders & Bandalos, 2001; Enders & Peugh, 2004)。

关于您的第二个问题：Mplus 是款功能强大的潜变量建模软件，其提供的潜变量建模方法均可使用 Mplus 进行模拟研究。Mplus7.0 在模拟和分析传统 SEM 时的效率还是比较高的，像本研究的模型基本上在 3 分钟以内。如果模拟和分析混合模型（mixture model）时间相对比较久，通常在十几分钟到几个小时，如果需要特定拟合参数可能会需要数天。

第三个问题：Mplus 对样本量没有限制，对变量数有限制。

意见 3： 第 10 页公式 3 的表达式存在笔误。

回应： 感谢您细心的指正。公式 3 出现的问题确实是在插入公式的时候出现了错误。在

文中我们已经进行了更正，更正为：
$$\bar{\hat{\phi}} = \sum_{k=1}^K \hat{\phi}_k / K \quad (\text{公式 3})$$

审稿人 3 意见：

文章具有一定意义，但存在以下问题：

意见 1： 文章版面过长，且方法过于复杂和数学化，似乎投到数学志杂更为合适。

回应： 感谢您的建议。我们已尽力避免复杂的数学公式，而更多的采用文字描述。心理统计确实比较尴尬，心理学杂志的审稿人觉得我们做的很数学化，而数学志杂的审稿人可能又觉得我们做的很简单。

意见 2： 英文文献格式不够规范，未完全遵循 APA 格式。例如：

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.

回应： 感谢审稿人提出的问题。已在参考文献处进行修改。包括：

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.

Thoemmes, F., & Rose, N. (2013). *Selection of auxiliary variables in missing data problems: Not all auxiliary variables are created equal*. Technical Report Technical Report R-002, Cornell University.

意见 3： 文章在表述中存在一些问题：一是研究结果多以表格形式呈现，其实有些也可以用图的形式表现（这样较为直观）；二是讨论部分，须与研究结果对应起来。

回应： 感谢审稿人提出的建议。我们在数据分析时也考虑采用图的形式呈现结果，但由于考虑的条件较多，图形呈现不易。我们已对本文的结果与讨论进行补充与重新整理，具体见文中标红部分。

意见 4： 结构方程模型的应用与数据形态有关，但是这篇文章没有对这方面有讨论。

回应： 感谢审稿人提出的问题。本文主要关注的问题是辅助变量与共缺机制、共缺率对模型参数估计的影响，数据形态不是主要的关注对象，所以本文的模拟设置设定数据形态为标准正态分布。另外，文中不足与展望部分已增加对数据形态的展望，具体如下：

另外，本研究模拟的数据服从正态分布，而实际研究中数据完全满足正态性的情况相对较少，以后的研究可以考虑数据非正态情况下的估计结果。

意见 5： 文章结论认为：纳入多于一个的辅助变量进行分析，参数估计的结果更好。但是，在讨论与结论部分却认为：随着共缺率的增多，单个辅助变量的参数估计差别也不大。因此，缺失数据研究中，缺失率的影响有待进一步的研究。前后显得有些矛盾。

回应：感谢您提出的问题。对于“纳入多于一个的辅助变量进行分析，参数估计的结果更好。”这一表述，由于我们的疏忽导致一些混淆，我们已在文中进行了更正。修改为：对于 MAR-MAR 组合机制，纳入单个辅助变量是有益的；对于 MAR-MCAR 或 MAR-MNAR 组合机制，纳入多于一个辅助变量的效果更好。

另外，对于共缺率的讨论与结论部分，经过我们更加认真的思考，修改为：奇怪的是，在大样本量(N=500 或 1000)、MAR-MAR 组合机制条件下，随着共缺率的增大，单个辅助变量时的参数估计的偏差总体呈增大的趋势，但并非与共缺率同步增大。因此，缺失数据研究中，缺失率的影响有待进一步的研究。

意见 6：文章结论过于概论，所得结论仅是基于特定条件之下，所给出的三条结论应该注明特定条件。

回应：感谢您提出的建议。我们理解您说的结论过于概论，指关于缺失机制、相关程度和辅助变量数三段中得到的结论。经过认真的思考，我们已对必要的内容进行补充与修正。

缺失机制这一段补充的内容如下：所以，当辅助与研究变量共缺时，如果采用纳入辅助变量的方法进行缺失数据分析，不用因为辅助变量的缺失机制为 MNAR 而有过多的顾虑，因为共缺组合机制为 MAR-MNAR 或 MCAR-MNAR 所得到的结果可能比 MAR-MCAR 或 MCAR-MCAR 要好。

相关程度这一段修正的内容如下：因此，根据本研究结果，当辅助变量与研究变量的相关达到.2-.3 时，即可考虑纳入该辅助变量对缺失数据进行建模分析，尤其是当共缺组合机制为 MAR-MCAR 时。

辅助变量数这一段修正的内容如下：本研究发现，当辅助变量与研究变量存在共缺时，对于 MAR-MAR 组合机制，纳入单个辅助变量是有益的；对于 MAR-MCAR 或 MAR-MNAR 组合机制，纳入多于一个辅助变量的效果更好。

意见 7：关于 FIML 和 MI 的比较，本文仅是引用文献来佐证 FIML 的优势，缺乏实证分析，更为恰当的做法是作者先要有前期基础来证明 FIML 比 MI 更为优势才来进行本研究。

回应：感谢您的建议。为了更清楚表明我们选用 FIML 进行缺失数据分析的理由，我们在文中进行了如下的补充：

运用 MI 处理缺失数据时辅助变量的效用还没有一致的结论，尤其是在实际数据处理中(Mustillo, 2012)。综合以上 FIML 和 MI 特点的比较，以及本文主要研究与辅助变量相关的内容，本研究只关注 FIML 处理缺失数据时辅助变量的相关问题。

通过文中对先前类似研究的整理中也可以发现这一问题：Collins, Schafer 和 Kam (2001)的模拟研究侧重采用 MI 对单个不缺失的辅助变量进行线性回归分析；Graham (2003)的研究通过 SCM 对单个不缺失的辅助变量进行 FIML/SEM 分析。这些模拟研究

均发现不缺失的辅助变量与研究变量高度相关时，纳入辅助变量能够改善模型参数估计。Mustillo (2012)的研究在回归模型中通过 MI 处理缺失值，探究辅助变量类型、研究变量的缺失率与缺失机制的关系，该研究发现纳入辅助变量对参数估计没有明显的改善。然而 Collins, Schafer 和 Kam (2001) 指出，即使纳入与缺失变量无关的辅助变量，得到最坏的结果也是中性的，并不会恶化参数估计。而当与缺失变量相关的辅助变量被忽略时，均值、方差、回归估计会产生实质性的偏差。

意见 8: 研究考虑了偏差和覆盖率，作者详尽地给出了偏差有关公式，但有关覆盖率公式并没有详细给出或作出较为详尽的说明。

回应: 感谢您的问题。这确实是个需要认真说明的问题。在模拟研究中，覆盖率表示每次重复模拟计算所得结果等于/接近真值的比例，类似于参数区间估计（频率论）上的置信区间（confidence interval）：区间包含的真值。例如，覆盖率为.95，意味着1000次的抽样中（模拟计算），有950次得到的估计值包含真值，为此我们在3.2结果评价标准增加了如下内容：

在模拟研究中，覆盖率表示每次重复模拟计算所得结果等于/接近真值的比例，类似于参数区间估计（频率论）上的置信区间（confidence interval）：区间包含的真值。例如，覆盖率为.95，意味着1000次的抽样中（模拟计算），有950次得到的估计值包含真值。前人的研究认为覆盖率小于.90是不可接受的(Collins, Schafer & Kam, 2001; Enders & Peugh, 2004)，本研究也采用.90作为覆盖率是否可接受的标准。

意见 9: 结构方程需要考虑的模型较多，本研究仅是研究特定模型之下缺失数据结构方程建模时采用全息极大似然估计时辅助变量的作用，并未全面考虑辅助变量间、辅助变量与研究变量、其它变量与辅助变量及研究变量相关时共缺率对于结构方程建模参数估计的影响，结论条件阐述不够全面。进行本文研究，应该事先有一个较为全面的条件交待。

回应: 您的这一观点我们非常赞同。通常 SEM 都比较复杂，涉及的变量较多，变量间的关系也比较多，所以模拟研究只能对特定模型（研究中常见的模型）进行研究，同时模拟研究所设定的模型也比较简单，比如这里我们所设计的模型只有2个潜变量，每个潜变量只有3个测量指标。关于您的这个问题，我们在3.1模拟研究设计部分已经做了相应介绍。

意见 10: 作者讨论了样本量对模型纳入辅助变量时结构方程建模全息极大似然估计的影响，但有两点需要作出说明：一是与缺失率、缺失机制等因素如何影响估计，作者讨论不够充分；二是样本量影响结构方程建模全息极大似然的估计，未体现在结论中，显得不足。

回应: 感谢您提出的建议。经过认真的思考，我们综合以上两点对样本量这一部分的讨论进行了补充。

补充内容为：结合本研究的结果，对于回归系数偏差的参数估计(偏差_{SE}、偏差₁₀₀)，当辅助与研究变量呈低相关，共缺率为.20，辅助变量数为3个的时候，如果样本量为200或500，在“MCAR-”组合机制条件下得到的偏差值最大；而当样本量为1000时，在“MCAR-”组合机制条件下得到的偏差值最小。控制相关程度、共缺率、辅助变量数不变的情况下，结果表明“MCAR-”组合机制条件下得到的偏差值随着样本量的增大而减小(见附表3)。对于Y因子负荷偏差的参数估计(偏差_{SE}、偏差₁₀₀)，相同条件下，如果样本量为100、500或1000，在“MCAR-”组合机制条件下得到的偏差值最小。控制相关程度、共缺率、辅助变量数不变的情况下，结果表明“MCAR-”组合机制条件下得到的偏差值随着样本量的增大而减小。

再次感谢您的耐心审阅与提出的宝贵建议。

第二轮

修改稿在表述方面比初稿要好很多。可以看出作者认真考虑了各位评审老师的意见和建议并仔细做了相应的修改。针对我提出的绝大多数意见，作者都给出了很清楚的回应，在此表示感谢！以下几点建议供作者参考：

意见 1:对初稿的意见 1：“我个人感觉该模拟研究考虑的因素过多，造成研究结果难以很好地呈现在有限的篇幅之内”。虽然作者解释了为什么要包括众多因素，但我仍然认为本文中同时考虑的因素过多。在这一点上，可能我和作者的观点不太一致。这个研究非常有价值，完全可以分成几个模拟研究来做。

回应：感谢您非常中肯的建议,在今后的研究中我们会接受您的这个建议。

意见 2:对初稿的意见 2：“模拟情况 1 为辅助变量不缺失的情况，不可能有 6 种缺失机制组合，只有 Y 对应的两种缺失机制；也不存在共缺率的概念，因为只有 Y 缺失。文中的附表 1, 2, 和 6 中对情况 1 有 6 种缺失机制是不正确的。希望作者仔细考虑。”在作者的回应中也承认只有 Y 变量存在缺失。但为什么在修改稿中附表 1, 2, 和 6 的表头仍有 6 种机制？破折号后面的内容是辅助变量的缺失机制，既然辅助变量不缺失，怎么会有破折号后面的内容？我认为表头只能有 2 种机制：Y: MCAR 和 Y: MAR。所以，表头

N	MR	AUX	MAR- MAR	MAR- MCAR	MAR- MNAR	MCAR- MAR	MCAR- MCAR	MCAR- MNAR
---	----	-----	-------------	--------------	--------------	--------------	---------------	---------------

应改为：

N	MR	AUX	MAR	MCAR
---	----	-----	-----	------

回应：感谢您再次提出这个问题。像您描述的一样，我们在第一次回应的时候也承认只有 Y 变量存在缺失，事实也确实如此。但在修改稿中我们并没有像您给出的建议那样（您这样修改是正确的）对附表 1, 2 和 6 的表头进行修改，是因为模拟情况 1 的程序由模拟情况 2 通过设定辅助变量不缺失得到，是为了方便读者进行两种情况结果的对比。但为了不误导读者，我们将表格里的字母进行了“虚化”，并在表格的备注处都进行了如下说明：由于辅助变量不缺失，各种缺失组合机制指辅助与研究变量共缺条件下对应的只有研究变量存在缺失的情况。

意见 3.文中“标准偏差=(平均估计值-理论值)/平均标准误(以示区分，记为偏差 *S.E.*)”，请对这个公式中的平均估计值和平均标准误进行解释和说明。

回应：感谢您的指正，非常抱歉我们没有交代清楚。由于本文中 X 因子和 Y 因子各有 3 个指标，所以计算 X 因子和 Y 因子的指标的估计值、标准误、标准差时都是用平均值。例如，求 X 因子的指标的平均估计值的时候，是把 X 因子的三个指标的估计值加起来再除以 3；求 X 因子的指标的标准误的时候，是把 X 因子的三个指标的标准误加起来再除以 3；求 Y 因子的指标的估计值或标准误也是如此。

意见 4.作者在文中提到：“...但有研究质疑以.40 作为判断标准的适切性(Graham, 2009)，因此当偏差出现的情况较多时，本研究同时采用偏差₁₀₀(Bias₁₀₀)作为结果评价指标。”但修改稿的结果中几乎没有用到偏差₁₀₀，是否考虑删除？

回应：感谢您的问题。在初稿的表格中:表 3, 附表 5 和附表 7 的结果都是呈现偏差₁₀₀的结果，后来根据修改意见，我们认为统一呈现偏差_{S.E.}的结果会更清晰，同时由于表格数据比较多，就不再呈现偏差₁₀₀的结果了。但是，为了让读者清楚我们是两种偏差都考虑到的，在修改稿的表格中（附表 9），我们把本文中考虑到的所有的结果评价指标（偏差₁₀₀，偏差_{S.E.}，覆盖率）都呈现出来了。在此，为了不混淆读者，我们在文中结果部分进行了如下补充：由于版面限制本研究只呈现了部分结果($n=500$)，更多的结果(其他样本量与偏差₁₀₀的结果)可与作者联系获得。

意见 5. P.16, 倒数第二行，ML 应该为 FIML。

回应：感谢您的指正，已对此进行了修改。

意见 6. P.23, 在解释覆盖率时，作者说“覆盖率为.95，意味着 1000 次的抽样中（模拟计算），有 950 次得到的估计值包含真值。”我认为这里可能有误。估计值怎么可能包含真值？应该为“估计区间”或“置信区间”吧？只有区间才能说包不包含真值。

回应：感谢您的问题。您的观点是正确的，只有区间才能说包不包含真值。也很抱歉我们没有充分表达出我们的意思。在模拟研究中，覆盖率表示每次重复模拟计算所得结果等于/接近真值的比例，类似于参数区间估计（频率论）的置信区间（confidence interval）：区间包含的真值。但在这里置信区间所允许的犯错误的概率（即显著性水平）不是固定的，显著性水平=1-覆盖率。所以，当覆盖率为.95 时，意味着在抽样 1000 次（模拟计算）得到的结果所组成的区间中，有 950 次得到的估计值在总的区间中包含了

真值，此时犯错误的概率为.05。再如，当覆盖率为.89 时，意味着在抽样 1000 次（模拟计算）得到的结果所组成的区间中，有 890 次得到的估计值在总的区间中包含了真值，此时犯错误的概率为.11。

文中相应的部分修改为：在模拟研究中，覆盖率表示每次重复模拟计算所得结果等于/接近真值的比例，类似于参数区间估计（频率论）的置信区间（confidence interval）：区间包含的真值。但这里的置信区间所允许的犯错误的概率（即显著性水平）不是固定的，显著性水平=1-覆盖率。所以，当覆盖率为.95 时，意味着在抽样 1000 次（模拟计算）得到的结果所组成的区间中，有 950 次得到的估计值在总的区间中包含了真值，此时犯错误的概率为.05。再如，当覆盖率为.89 时，意味着在抽样 1000 次（模拟计算）得到的结果所组成的区间中，有 890 次得到的估计值在总的区间中包含了真值，此时犯错误的概率为.11。

第三轮

意见 1. 作者针对我第二轮所提意见和建议，逐个给出了清楚恰当的回答。多谢！

回应：非常感谢您对本文的辛勤付出。

意见 2. 关于该修改稿，我没有其他问题。只有一个小小的建议：以后再做模拟研究时，不要在一个研究中同时考虑太多因素，某些因素完全可以取固定值。

回应：感谢您中肯的建议，在今后的研究中我们一定会接受您建议。

主编终审

综合各个方面的意见，考虑到应该使得学报的资源（包括篇幅）与效应（包括读者群和读者能利用的信息），学报主编建议作者在审就的定稿的基础上，压缩到一万字以内，即保留论述和实验部分，具体推演还可以请有兴趣的读者和作者做深入沟通（实际上，对于不熟悉推演的读者，仅凭目前的篇幅，也难以完全明白），在学报发表。如果作者觉得，删减有伤原文主题，或者拒绝删减，只好请您再投其他刊物，我们可以将审稿的全部过程，包括学报主编的最后判断，一并转过去，以利清晰情况。

回应：已按主编要求对论文进行了删减，已将正文字数压缩到万字以内（不含参考文献和英文摘要）。