

纵向数据的中介效应分析*

方杰¹ 温忠麟^{**2} 邱皓政³

(¹ 广东财经大学应用心理学系, 广州, 510320)

(² 华南师范大学心理学院 / 心理应用研究中心, 广州, 510631)

(³ 台湾师范大学管理学院, 台北)

摘要 目前中介效应检验主要是基于截面数据, 但许多时候截面数据的中介分析不适合进行因果推断, 因而需要收集历时性的纵向数据, 进行纵向数据的中介分析。评介了基于交叉滞后面板模型、多水平模型和潜变量增长模型的纵向数据的中介分析方法及其四个发展。第一, 中介效应随时间变化, 如连续时间模型、多层时变系数模型。第二, 中介效应随个体变化, 如随机效应的交叉滞后面板模型和多层自回归中介模型。第三, 中介模型的整合, 如交叉滞后面板模型与多水平模型整合为多层自回归中介模型。第四, 中介检验方法的发展, 建议使用 Bootstrap 和贝叶斯法进行纵向数据的中介分析。总结出一个纵向数据的中介分析流程并给出相应的 *Mplus* 程序。随后展望了纵向数据的中介分析的拓展方向。

关键词 纵向数据 中介效应 交叉滞后面板模型 多水平模型 潜变量增长模型

中介 (mediation) 是社会科学研究中重要的方法学概念。如果自变量 X 通过某一变量 M 对因变量 Y 产生一定影响, 则称 M 在 X 和 Y 之间起中介作用, 此时称 M 为中介变量 (见图 1)。近三十年来, 中介研究备受关注, 分析方法和实际应用都得到长足发展。然而目前中介效应分析主要涉及截面数据 (cross-sectional data), 基于截面数据中介分析的最大不足就是不适合进行因果推断。如果中介过程中至少有两个变量之间的影响作用是历时性的, 需要进行纵向 (追踪) 中介研究 (温忠麟, 2017), 即间隔一段时间对研究对象重复测量自变量、中介变量和因变量, 获得历时性的纵向数据 (longitudinal data), 进行纵向数据的中介分析。近年来, 纵向数据的中介分析已经成为中介研究的新热点 (邱皓政, 2017; MacKinnon, 2008; Preacher, 2015)。本文

在介绍了交叉滞后面板模型、多水平模型和潜变量增长模型的纵向数据的中介分析方法后, 讨论了这三类方法的发展; 接着, 提出一套纵向数据的中介分析流程; 然后对相关问题进行了讨论和拓展。

1 传统的纵向数据的中介分析方法

1.1 基于交叉滞后面板模型的中介分析

交叉滞后面板模型 (cross-lagged panel model, CLPM) 又称为自回归中介模型 (autoregressive mediation models), 它假设变量之间的影响都是历时性的, 同一时间点的测量没有直接的影响关系 (见图 1) (Cole & Maxwell, 2003)。根据重复测量的次数不同, 交叉滞后面板模型可分为两种。一种是至少重复测量三次的二阶滞后 CLPM ($X_1 \rightarrow M_2 \rightarrow Y_3$, 见图 1(a)), 另一种是只需要重复测量二次的

* 本研究得到国家自然科学基金项目 (31771245) 和国家社会科学基金项目 (17BTJ035) 的资助。

** 通讯作者: 温忠麟, E-mail: wenzl@snu.edu.cn

DOI:10.16719/j.cnki.1671-6981.20210431

一阶滞后 CLPM ($X_1 \rightarrow M_2$ 和 $M_1 \rightarrow Y_2$, 见图 1(b)) (Cole & Maxwell, 2003; Wang & Zhang, 2020)。假设变量已经标准化, 二阶滞后 CLPM 的纵向中介效应 (见图 1(a), *Mplus* 程序见附录) 可表示为

$$X_{t+1} = \beta_X X_t + \varepsilon_{X(t+1)} \quad (1)$$

$$M_{t+1} = \beta_M M_t + aX_t + \varepsilon_{M(t+1)} \quad (2)$$

$$Y_{t+2} = \beta_Y Y_{t+1} + bM_{t+1} + c'X_t + \varepsilon_{Y(t+2)} \quad (3)$$

模型中的每个变量既依赖于前因变量的先前测量水平, 又依赖于此变量的先前测量水平。每个变量对前因变量的回归称为滞后效应 (lag effect) (用系数 a 、 b 和 c' 表示), 每个变量前后时间点的效应称为自回归效应 (autoregressive effect) (分别用系数 β_X 、 β_M 和 β_Y 表示)。系数 ab 表示纵向中介效应, c' 表示直接效应, ε 表示残差, 下标 t 、 $t+1$ 和 $t+2$ 表示重复测量的时间点。一阶滞后 CLPM (见图 1(b)) 只需要将公式 (3) 改为:

$$Y_{t+1} = \beta_Y Y_t + bM_t + c'X_t + \varepsilon_{Y(t+1)} \quad (4)$$

Maxwell 和 Cole (2007, 2011) 用数学推导证明, 如果将纵向中介数据提取某次 (如第二次) 重复测量的数据形成截面数据 (X_2, M_2, Y_2), 则截面数据的中介分析的中介效应、直接效应和中介效应量 (ab/c) 结果相对于纵向中介分析结果而言, 都将产生明显的偏差, 偏差的原因是分析的模型与假设的模型设定不同, 即历时性影响当作即时性影响分析了。另外, 如果公式 (1)~(3) 去掉自回归效应, 仅保留滞后效应, 称为序列设计 (sequential design), 即在时间点 t 进行自变量 X_t 的测量, 时间点 $t+1$ 进行中介变量 M_{t+1} 的测量, 时间点 $t+2$ 进行因变量 Y_{t+2} 的测量的纵向研究设计 (Cole & Maxwell, 2003)。Mitchell 和 Maxwell (2013) 用数学推导证明, 序列设计的纵向中介分析结果并不比截面数据的中介分析结果好, 在部分情况下甚至还更差。

1.2 基于多水平模型的中介分析

如果将纵向 (追踪) 数据看成是重复测量嵌套

于个体的两层数据结构 (层 1 为重复测量点 t , 层 2 为观测个体 j), 则可以使用多水平模型 (multilevel model, MLM) 进行中介分析 (Kenny et al., 2003)。以自变量 X_{ij} 、中介变量 M_{ij} 和因变量 Y_{ij} 都是层 1 变量 (即变量随测量时间变化) 为例, 中介效应可表示为公式 (5)~(7) (Huang & Yuan, 2017):

测量水平 (层 1):

$$M_{ij} = r_j + aX_{ij} + \varepsilon_{Mij} \quad (5)$$

$$Y_{ij} = \delta_j + c'X_{ij} + bM_{ij} + \varepsilon_{Yij} \quad (6)$$

个体水平 (层 2):

$$r_j = r + \mu_{Mj} \quad \delta_j = \delta + \mu_{Yj} \quad (7)$$

系数 ab 表示中介效应, c' 表示直接效应。 r_j 和 δ_j 表示层 1 回归方程的截距, ε_{Mij} 和 ε_{Yij} 表示层 1 回归方程的残差; r 和 δ 表示层 2 回归方程的截距, μ_{Mj} 和 μ_{Yj} 表示层 2 回归方程的残差。各层残差都假设服从正态分布。但公式 (5)~(7) 是固定效应的 MLM 模型, 此模型设定变量之间的影响 (a 、 b 和 c') 不随个体 j 变化。因此更一般地, 可以将系数 a 、 b 和 c' 设为随机效应 (a_j 、 b_j 和 c'_j), 让系数随个体 j 的变化而变化 (见公式 (8)~(10)),

测量水平:

$$M_{ij} = r_j + a_j X_{ij} + \varepsilon_{Mij} \quad (8)$$

$$Y_{ij} = \delta_j + c'_j X_{ij} + b_j M_{ij} + \varepsilon_{Yij} \quad (9)$$

个体水平:

$$r_j = r + \mu_{Mj} \quad \delta_j = \delta + \mu_{Yj} \quad a_j = a + \mu_{aj} \quad b_j = b + \mu_{bj} \quad c'_j = c' + \mu_{c'j} \quad (10)$$

中介效应的均值为 ab 加上 a_j 和 b_j 的协方差, 直接效应的均值为 c' (Kenny et al., 2003)。

1.3 基于潜变量增长模型的中介分析

潜变量增长模型 (latent growth model, LGM) 是在结构方程模型的视角下定义变量随时间的发展趋势。对应于一个 LGM, 每个变量有一个初始水平 a 和一个变化率 S 作为潜变量 (邱皓政, 2017)。自变量 LGM、中介变量 LGM 和因变量 LGM 结合在

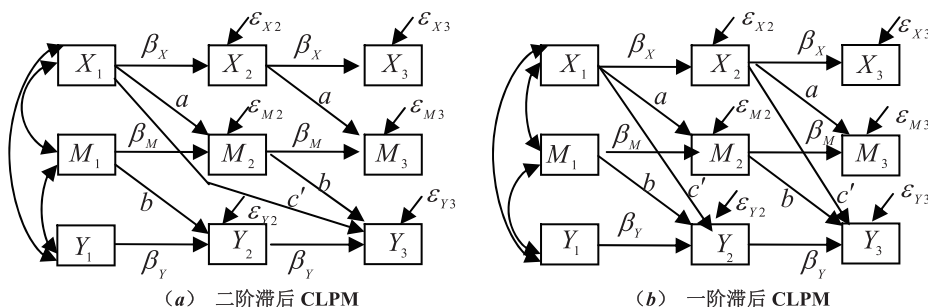


图 1 基于 CLPM 的纵向数据的中介模型

一起，建立基于 LGM 的中介模型，包括初始水平的中介($\alpha_X \rightarrow \alpha_M \rightarrow \alpha_Y$)和变化率的中介($S_X \rightarrow S_M \rightarrow S_Y$) (见图 2, Mplus 程序见附录) (Cheong et al., 2003)。基于 LGM 的中介模型有下列方程，前面 3 个描述初始水平之间的关系，后面 3 个描述变化率之间的关系 (μ 为潜变量均值、 ζ 为残差项)：

$$a_X = \mu_{Xa} + \zeta_{Xa} \quad (11)$$

$$a_M = \mu_{Ma} + r_1 a_X + \zeta_{Ma} \quad (12)$$

$$a_Y = \mu_{Ya} + r_2 a_M + r_3 a_X + \zeta_{Ya} \quad (13)$$

$$S_X = \mu_{XS} + \zeta_{XS} \quad (14)$$

$$S_M = \mu_{MS} + a S_X + \zeta_{MS} \quad (15)$$

$$S_Y = \mu_{YS} + c' S_X + b S_M + \zeta_{YS} \quad (16)$$

初始水平的中介效应为 $r_1 r_2$ ，直接效应为 r_3 ；变化率的中介效应为 ab ，直接效应为 c' 。残差均假设服从正态分布且相互独立。Von Soest 和 Hagtvet (2011) 指出，如果研究目标不是初始水平的中介效应，为了能有效控制因变量的初始状态 α_Y ，建议增加 α_Y 对 S_Y 的回归，将公式 (16) 变为 $S_Y = \mu_{YS} + c' S_X + b S_M + r_4 \alpha_Y + \zeta_{YS}$ 。O'Laughlin 等 (2018) 用实际的纵向数据同时进行截面中介分析和基于 LGM 的中介分析，发现两种分析结果截然相反，与 Maxwell 和 Cole (2007) 的结果一致。

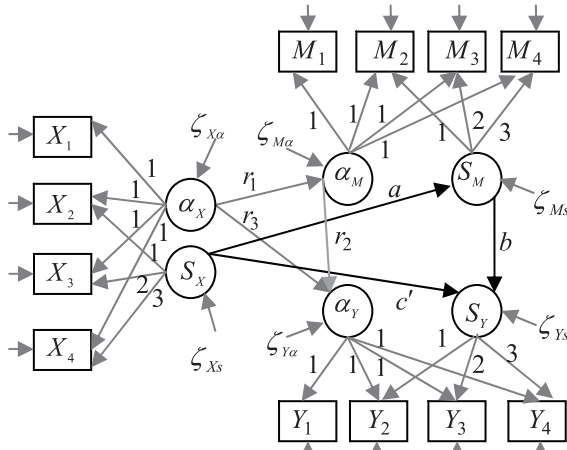


图 2 基于 LGM 的纵向数据的中介模型 (改编自 Mackinnon, 2008)

2 纵向数据的中介分析方法的发展

2.1 考察系数随个体变化的随机效应

2.1.1 基于随机效应的 CLPM 的中介分析

Wu 等 (2018) 指出二阶滞后 CLPM (图 1 (a)) 的一个不足是将系数 (a 、 b 、 c' 、 β_X 、 β_M 和 β_Y) 设定为固定效应 (常数)，即认为系数不存在个体差异。Wu 等人提出将公式 (1) ~ (3) 中的系数设定为可

随个体 j 变化的随机效应 (a_j 、 b_j 、 c'_j 、 β_{Xj} 、 β_{Mj} 和 β_{Yj})，和公式 (10) 一样，随机效应都可以写成一个常数加上残差项的形式，假设残差服从多元正态分布，利用随机效应 (random effect) 的 CLPM 来考察随个体变化的中介效应 (Mplus 程序见附录)。具体地，Wu 等人使用 Mplus 软件将随机效应当成潜变量，利用 EM (expectation maximization) 算法估计出每个随机效应的均值和方差。中介效应的均值为随机系数 a_j 、 b_j 的均值 (a 和 b) 的乘积加上 a_j 和 b_j 的协方差。直接效应的均值为随机系数 c'_j 的均值 c' 。

Wu 等人用模拟研究比较了随机效应与固定效应的二阶滞后 CLPM (图 1 (a)) 在中介效应分析中的表现。结果发现，当系数 a 、 b 、 c' 、 β_X 、 β_M 和 β_Y 都是固定值时，两种模型的结果一致；当系数存在随机效应时，无论随机效应是正态分布还是非正态分布，结果一致发现，随机效应的 CLPM 的中介分析在相对偏差、置信区间覆盖率、统计功效和第 I 类错误率上都显著优于固定效应时。Wu 等人建议，如果随机效应的 CLPM 的 AIC (akaike information criteria) 和 BIC (bayesian information criteria) 指标小于固定效应，就使用随机效应的 CLPM 进行纵向中介效应分析。

随机效应的 CLPM 存在如下不足 (Wu et al., 2018)。第一，只能将滞后效应和自回归效应设定为随机效应。如果变量没有标准化，公式 (1) ~ (3) 所示回归方程的常数项就无法设为随机效应。第二，仅适用于显变量且是连续变量的情况。

2.1.2 基于多层自回归中介模型的中介分析

Zhang 等 (2018) 另辟蹊径，将 MLM 和 CLPM 结合在一起，使用多层自回归中介模型 (multilevel autoregressive mediation models, MAMM) 进行中介分析 (见图 3)，方程如下：

测量水平：

$$X_{(t+1)j} = l_{Xj} + \beta_{Xj} X_{tj} + \varepsilon_{Xj(t+1)} \quad (17)$$

$$M_{(t+1)j} = l_{Mj} + \beta_{Mj} M_{tj} + a_j X_{tj} + \varepsilon_{Mj(t+1)} \quad (18)$$

$$Y_{(t+1)j} = l_{Yj} + \beta_{Yj} Y_{tj} + b_j M_{tj} + c'_j X_{tj} + \varepsilon_{Yj(t+1)} \quad (19)$$

其中的参数包括常数项 l_{Xj} 、 l_{Mj} 和 l_{Yj} ，自回归效应 β_{Xj} 、 β_{Mj} 和 β_{Yj} ，滞后效应 a_j 、 b_j 和 c'_j 。和公式 (10) 一样，它们都可以写成一个常数加上残差项的形式。我们感兴趣的路径效应是 $a_j = a + \mu_{aj}$ 、 $b_j = b + \mu_{bj}$ 和 $c'_j = c' + \mu_{cj}$ 。Zhang 等人用贝叶斯方法进行纵向中介效应检验，其中，纵向中介效应的均值为 ab 与

μ_a 和 μ_b 的协方差之和,直接效应的均值为 c' 。

Cain 等 (2018) 用模拟研究比较了截面数据的中介模型、一阶滞后 CLPM (图 1 (b))、序列设计和 MLM (公式 (8)~(10)) 和多层自回归中介模型在纵向数据的中介分析中的表现。结果表明,在第 I 类错误率和统计功效上,多层自回归中介模型的表现最优。多层自回归中介模型的不足之一在于需要大数据量。Zhang 等人的模拟研究发现,当重复测量至少 5 次 ($t \geq 5$) 且有 50 人以上 ($j \geq 50$) 时,可以得到准确的 ab 值;如果要得到准确的中介效应值,则需要 $t \geq 10$ 且 $j \geq 100$ 或者 $t \geq 20$ 且 $j \geq 50$ (Zhang et al., 2018)。

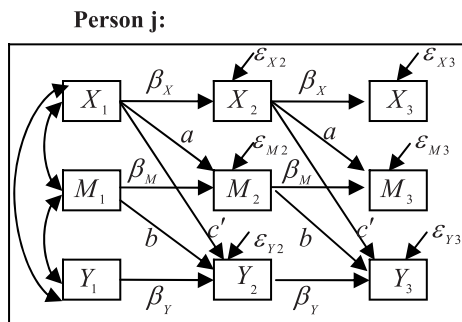


图3 多层自回归中介模型 (改编自 Zhang et al., 2018)

值得注意的是,多层自回归中介模型和随机效应的 CLPM 的中介效应虽然都能考察系数随个体变化的随机效应 (即个体差异),但二者至少有三处明显不同。首先,最大的不同在于随机效应的 CLPM 是单层的,模型更简单,更易于操作和理解。其次,多层自回归中介模型有两个前提条件 (Zhang et al., 2018),一是变量间关系具有跨时间的稳定性 (以 β_X 为例, X_2 对 X_1 和 X_3 对 X_2 的自回归系数相等),二是测量间的时间间隔相等,但随机效应的 CLPM 没有这样的前提条件。第三,多层自回归中介模型可以将常数项 l_{Mj} 和 l_{Yj} 设为随机效应,但随机效应的 CLPM 不可以这样做。

2.2 考察随时间变化的中介效应

2.2.1 连续时间模型的中介分析

有研究者指出,交叉滞后面板模型 (见图 1) 得到的纵向中介效应大小会随着滞后时间 Δt 的变化而变化 (Cole & Maxwell, 2003; Maxwell et al., 2007, 2011),但交叉滞后面板模型一次只能得到某一个滞后时间 Δt 的纵向中介效应,只能了解有限的信息,无法阐释纵向中介效应大小是如何随滞后时间而变化的,因此又被称为离散时间模型 (discrete time model) (Deboeck & Preacher, 2016; Voelkle et

al., 2012)。Deboeck 和 Preacher (2016) 使用连续时间模型 (continuous time model) 来解决离散时间模型的不足。

连续时间模型的基本原理是,假设 Δt 无限趋近于 0,连续时间模型可用随机微分方程表示为

$$\frac{dx(t)}{dt} = Ax(t) + G \frac{dW(t)}{dt} \quad (20)$$

$x(t)$ 是 3×1 的矩阵,数字 3 表示在时刻 t 测得的自变量、中介变量和因变量。 A 是 3×3 的自回归和滞后效应矩阵 (自回归效应见矩阵的对角线)。 A 表明自回归和滞后效应是独立于滞后时间。 G 是常数, $\frac{dW(t)}{dt}$ 是由维纳过程 (Wiener process, 是一种连续时间随机过程,也称为布朗运动) 产生的残差项 (Deboeck & Preacher, 2016; Voelkle et al., 2012)。连续时间模型使用和离散时间模型相同的数据进行纵向中介分析,估计出独立于滞后时间的自回归和滞后效应矩阵 A ,然后计算出特定滞后时间 Δt 的条件下的纵向中介效应 $e^{A \times \Delta t}$,实现了一次分析得到滞后时间取值范围 $[t_{lower}, t_{upper}]$ 下所有可能的纵向中介效应,相比离散时间模型能获得更多有用信息。接着以横坐标为 Δt 、纵坐标为纵向中介效应画图,用图形直观地描绘出纵向中介效应大小随滞后时间而“连续”变化的规律,并能发现滞后时间取何值时,纵向中介效应取得最大值 (Deboeck & Preacher, 2016)。值得注意的是,连续时间模型有两个前提条件,一是变量间关系具有跨时间的稳定性,二是连续时间模型仅适用于一阶 CLPM (见图 1 (b)) (Deboeck & Preacher, 2016; Wang & Zhang, 2020)。

2.2.2 多层时变系数模型的中介分析

Huang 和 Yuan (2017) 指出基于 MLM 的纵向数据的中介分析 (公式 (5)~(7)) 没有考虑变量之间的关系 (如回归系数 a 和 b) 随时间变化的效应,因此,提出变量之间的关系随时间变化的多层时变系数模型 (multilevel time-varying coefficient model) 或称动态多层中介模型 (dynamic multilevel mediation model) (公式 (21)~(23)),

层 - 1:

$$M_{ij} = r_j(t) + a(t)X_{ij} + \varepsilon_{Mij} \quad (21)$$

$$Y_{ij} = \delta_j(t) + c'(t)X_{ij} + b(t)M_{ij} + \varepsilon_{Yij} \quad (22)$$

层 - 2:

$$r_j(t) = r(t) + \mu_{Mj} \quad \delta_j(t) = \delta(t) + \mu_{Yj} \quad (23)$$

用函数 $a(t)$ 、 $b(t)$ 和 $c'(t)$ 替代原模型中 (公式 (5)~(7)) 恒定的系数 a 、 b 和 c' ,表示变量之

间的关系随时间变化,是时间 t 的函数。 $a(t)b(t)$ 表示中介效应, $c'(t)$ 表示直接效应。Huang 和 Yuan 采用惩罚样条 (penalized spline) 法来建模时间函数 (如 $a(t)$ 和 $b(t)$), 即用多项式函数 (样条) 来逼近时间函数, 并加入惩罚项来惩罚函数复杂性 (即函数曲线的粗糙程度), 使函数更加光滑 (惩罚样条法的介绍详见唐文清等, 2020)。接着将时间函数带入公式 (21)~(23), 用贝叶斯法进行纵向数据的中介效应分析。

Huang 和 Yuan (2017) 用模拟研究比较了多层时变系数模型与 MLM (公式 (5)~(7)) 在中介效应分析中的表现。结果发现当中介效应不随时间变化时, 两种模型的结果一致; 当中介效应随时间变化时, 多层时变系数模型的中介分析在相对偏差、均方误和可靠区间覆盖率上都显著优于 MLM 的分析结果。多层时变系数模型的不足在于需要重复测量的次数较多, Huang 和 Yuan (2017) 建议收集至少重复测量 10 次以上的数据用于多层时变系数模型分析。

值得注意的是, 多层时变系数模型和连续时间模型虽然都能考察中介效应随时间变化的规律, 但二者存在本质区别。多层时变系数模型考察的是 $X_t \rightarrow M_t \rightarrow Y_t$ 的中介效应且没有控制自回归效应, 而连续时间模型考察的是 $X_t \rightarrow M_{t+1}$ 和 $M_t \rightarrow Y_{t+1}$ 的中介效应且控制了自回归效应, 只有连续时间模型反映了自变量、中介变量和因变量之间的历时性影响关系。

2.3 中介效应显著性检验方法的比较

纵向数据的中介效应的显著性检验方法主要在潜变量增长模型 (LGM) 下开展。Cheong (2011) 用模拟研究比较了数据正态情况下, Sobel 法和乘积分布法在基于 LGM 的变化率的中介效应 (见图 2 的粗黑箭头) 分析中的表现。结果表明, 乘积分布法的统计功效优于 Sobel 法。Koo 等 (2016) 进一步用模拟研究比较了数据非正态情况下, Sobel 法、乘积分布法和偏差校正的 Bootstrap 法在基于 LGM 的变化率的中介效应分析中的表现。结果表明, 数据的非正态性对纵向中介效应分析的影响不大, 偏差校正的 Bootstrap 法表现最优, 乘积分布法次之, Sobel 法最差。Koo 等 (2016) 的不足在于方法比较没有涉及未校正的 Bootstrap 法。因为在截面数据的中介分析方法比较中发现, 偏差校正的 Bootstrap 方法在某些条件下会高估第 I 类错误

率, 且置信区间覆盖率表现不如未校正的 Bootstrap 方法, 建议使用未校正的 Bootstrap 法进行中介效应分析 (Falk & Biesanz, 2015; Fang et al., 2019; Fritz et al., 2012)。另外, 张沥今等 (2019) 指出贝叶斯方法在估计 LGM 模型时具有更方便处理缺失值、更易于处理数据转换的情况、更有效估计复杂模型、更适用于小样本等诸多优势, 但贝叶斯法在 LGM 中介分析中的表现如何, 还有待进一步深入研究。

3 纵向数据的中介效应分析方法的综合讨论

纵向数据的中介效应分析一定能进行历时性因果 (温忠麟, 2017) 推论吗? 答案是不一定。只有使用固定系数和随机系数的 CLPM、多层自回归中介模型和连续时间模型才能进行历时性因果推论, 因为这三个模型体现了历时性因果推论所必须的测量时间点先后关系, 同时考虑了自回归效应 (MacKinnon, 2008; Preacher, 2015)。

MLM 和多层时变系数模型在建模过程中, 没有考虑变量之间影响的先后顺序, 其中介分析 ($X_t \rightarrow M_t \rightarrow Y_t$) 本质上与截面数据的中介分析无异, 要得出因果推论还要有统计以外的理据 (温忠麟, 2017)。Cain 等 (2018) 更明确地将纵向数据的 MLM 中介分析称为多层截面中介 (multilevel cross-sectional mediation) 分析。值得注意的是, 近来, 研究者建议使用多层结构方程模型 (multilevel structural equation modeling, MSEM) 进行中介分析, 因为 MSEM 克服了 MLM 假设没有测量误差的不足, 但 MSEM 的建模过程仍然没有考虑变量之间的先后关系, 因此还是难以进行历时性影响的因果推论 (Zhang & Phillips, 2018)。

LGM 本是为了估计变量的成长曲线或轨迹而生的。因此, LGM 只是用纵向数据建立了各个变量本身的增长模型, 在此建模过程中, 变量之间的影响关系 (初始水平的中介 ($\alpha_X \rightarrow \alpha_M \rightarrow \alpha_Y$) 和变化率的中介 ($S_X \rightarrow S_M \rightarrow S_Y$)) 并没有时间先后的设计 (见图 2), 因而其中介分析本质上还是截面中介分析, 要得出因果推论也还要有统计以外的理据 (O'Laughlin et al., 2018)。

已有研究者论证了 LGM 与 MLM 在处理纵向数据时具有等价性, LGM 模型中的潜变量对应于 MLM 中的随机效应 (王孟成, 毕向阳, 2018; Curran, 2003)。但 LGM 能对纵向数据的中介模型进行更为

灵活地拓展。例如,图2所示的LGM模型是以发展过程的同质性假设和连续性假设(即单阶段性,如线性增长)为前提的,如果发展趋势同时存在异质性和多阶段性,则建议使用多阶段混合增长模型(piecewise growth mixed model)(王婧等,2017)。

4 纵向数据的中介分析的流程图

面对一个纵向数据的中介效应分析任务,研究者应当如何进行呢?根据前面的讨论,我们总结出一套纵向数据的中介效应分析流程(见图4)如下:

1. 首先判断因果影响是历时性的还是即时性的,以决定是否需要进行历时性因果推论。如果是,则进入步骤2; 如果否,则使用LGM或MLM进行分析。
2. 是否需要考察时变效应。如果是,则使用连续时间模型进行分析; 如果否,则进入步骤3。
3. 是否要考察个体差异。如果是则使用随机效应的交叉滞后面板模型(CLPM)或多层自回归中介模型进行分析, 如果否,则使用固定效应的CLPM进行分析。

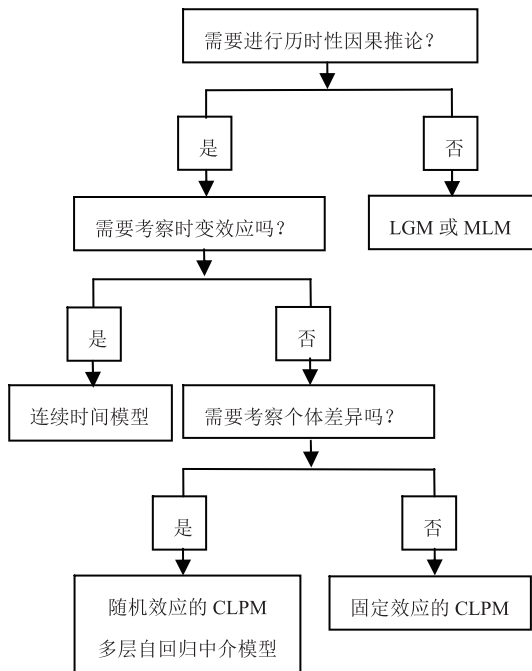


图4 纵向数据的中介分析流程

5 小结

纵向数据的中介效应分析能有效避免截面(横断)研究存在的估计偏差问题,为揭示变量间的因果方向和中介机制提供有力证据等优点,已经成为中介研究的新热点(MacKinnon, 2008; Preacher, 2015)。目前,纵向数据的中介分析方法的发展有

四条主线,第一条主线是关注时间在纵向数据的中介分析中的作用(时变效应),如连续时间模型和多层时变系数模型。第二条主线是关注个体间发展的差异,CLPM采用设置随机效应或与多水平模型相结合来考察个体间发展的差异,MLM使用层2回归方程的残差来反映个体间差异,LGM使用初始水平和变化率的方差来反映个体间增长的差异。第三条主线是模型的整合,突出表现在CLPM和MLM整合为多层自回归中介模型。研究者还指出,下一步应考虑将多层自回归中介模型、随机效应的CLPM各自与连续时间模型进行整合,以便于同时考察中介的时变效应(Wu et al., 2018; Zhang et al., 2018)。第四条主线是中介效应检验的方法,多层自回归中介模型和多层时变系数模型建议使用贝叶斯法进行中介分析,LGM建议使用Bootstrap法进行中介分析。

当前,纵向数据的中介分析还处于发展和完善阶段。例如,本文仅涉及单层中介效应,Zhang和Phillips(2018)将多层自回归中介模型推广到多层中介的情况,使用多层自回归中介模型进行2-2-1多层中介分析(自变量和中介变量都在顶层)。Zhang等(2018)还将多层自回归中介模型推广到有调节的中介效应的分析中。又如,本文的CLPM没有考虑测量误差,Zhang和Yang(2020)建议用单指标(即多个指标的得分的均值)潜变量的固定系数的CLPM来控制测量误差。再如,本文仅涉及被试群体同质的情况,Huang等(2019)将多层时变系数模型拓展到群体异质的情况,提出了潜类别(latent class)的多层时变系数模型(Huang et al., 2019)。另外,除了本文涉及的纵向数据的中介分析模型外,还有状态空间模型(state space model)(Gu et al., 2014)、指数衰减模型(exponential decay model)(Fritz, 2014)、潜变计分模型(latent change score model)(Preacher, 2015)等。方法的进步给研究者提供了一个深入理解和应用纵向数据的中介效应的机会,相信随着纵向数据的中介分析方法研究的深入,会不断增加我们对纵向数据的中介问题的理解。

注:附录于https://www.researchgate.net/profile/Jie_Fang34/research下载。

参考文献

- 唐文清,张敏强,方杰.(2020).时变效应模型及在密集追踪数据分析中的应用.《心理科学》,43(2),488-497.
- 王婧,唐文清,张敏强,张文怡,郭凯茵.(2017).多阶段混合增长模型的方

- 法及研究现状. *心理科学进展*, 25(10), 1696-1704.
- 王孟成, 毕向阳. (2018). *潜变量建模与 Mplus 应用: 进阶篇*. 重庆: 重庆大学出版社.
- 温忠麟. (2017). 实证研究中的因果推理与分析. *心理科学*, 40(1), 200-208.
- 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.
- Cain, M. K., Zhang, Z. Y., & Bergeman, C. S. (2018). Time and other considerations in mediation design. *Educational and Psychological Measurement*, 78(6), 952-972.
- Cheong, J. W., Mackinnon, D. P., & Khoo, S. T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(2), 238-262.
- Cheong, J. W. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 195-211.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112(4), 558-577.
- Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38(4), 529-569.
- Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 61-75.
- Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 24-38.
- Fang, J., Wen, Z. L., Hau, K. T. (2019). Mediation effects in 2-1-1 multilevel model: Evaluation of alternative estimation methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(4), 591-606.
- Fritz, M. S. (2014). An exponential decay model for mediation. *Prevention Science*, 15(5), 611-622.
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47(1), 61-87.
- Gu, F., Preacher, K. J., & Ferrer, E. (2014). A state space modeling approach to mediation analysis. *Journal of Educational and Behavioral Statistics*, 39(2), 117-143.
- Huang, J., & Yuan, Y. (2017). Bayesian dynamic mediation analysis. *Psychological Methods*, 22(4), 667-686.
- Huang, J., Yuan, Y., & Wetter, D. (2019). Latent class dynamic mediation model with application to smoking cessation data. *Psychometrika*, 84(1), 1-18.
- Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8(2), 115-128.
- Koo, N., Leite, W. L., & Algina, J. (2016). Mediated effects with the parallel process latent growth model: An evaluation of methods for testing mediation in the presence of nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 32-44.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12(1), 23-44.
- Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioral Research*, 46(5), 816-841.
- Mitchell, M. A., & Maxwell, S. E. (2013). A comparison of the cross-sectional and sequential designs when assessing longitudinal mediation. *Multivariate Behavioral Research*, 48(3), 301-339.
- O'Laughlin, K. D., Martin, M. J., & Ferrer, E. (2018). Cross-sectional analysis of longitudinal mediation processes. *Multivariate Behavioral Research*, 53(3), 375-402.
- Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, 66, 825-852.
- Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, 17(2), 176-192.
- von Soest, T., & Hagtvet, K. A. (2011). Mediation analysis in a latent growth curve modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 289-314.
- Wang, L., & Zhang, Q. (2020). Investigating the impact of the time interval selection on autoregressive mediation modeling: Result Interpretations, effect reporting, and temporal designs. *Psychological Methods*, 25(3), 271-291.
- Wu, W., Carroll, I. A., & Chen, P. Y. (2018). A single-level random-effects cross-lagged panel model for longitudinal mediation analysis. *Behavior Research Methods*, 50(5), 2111-2124.
- Zhang, Q., Wang, L. J., & Bergeman, C. S. (2018). Multilevel autoregressive mediation models: Specification, estimation, and applications. *Psychological Methods*, 23(2), 278-297.
- Zhang, Q., & Phillips, B. (2018). Three-level longitudinal mediation with nested units: How does an upper-level predictor influence a lower-level outcome via an upper-level mediator over time. *Multivariate Behavioral Research*, 53(5), 655-675.
- Zhang, Q., & Yang, Y. (2020). Autoregressive mediation models using composite scores and latent variables: comparisons and recommendations. *Psychological Methods*, 25(4), 472-495.

Mediation Analysis of Longitudinal Data

Fang Jie¹, Wen Zhonglin², Chiou Hawjeng³

(¹ Department of Applied Psychology, Guangdong University of Finance & Economics, Guangzhou, 510320)

(² Center for Studies of Psychological Application & School of Psychology, South China Normal University, Guangzhou, 510631)

(³ College of Management, Taiwan Normal University, Taipei)

Abstract Over the past 30 years, most efforts on testing for mediation have been based on cross-sectional data, which may not get causal inference. A possible solution for this could be to collect longitudinal data and perform a longitudinal mediation analysis. There are three causal arrows in a simple mediation model for analyzing a system of causality. If there is at least one causal arrow where the effect arises sometime after the cause, a longitudinal mediation design will be necessary for effectively observing the causation. There are three types of longitudinal mediation analysis approaches: (1) cross-lagged panel model (CLPM); (2) multilevel mediation model (MLM); (3) latent growth mediation model (LGM).

There are four types of development of longitudinal mediation analysis. First, time-varying effect of mediation effect is tested. Continuous time models (CTM) would illustrate how mediating effects vary as a function of lag. Multilevel time-varying coefficient model (MTVCM) can capture direct and indirect effects over time. Second, individuals-varying effect of mediation effect is investigated. Random-effects cross-lagged panel model (RE-CLPM) and Multilevel autoregressive mediation model (MAMM) should be adopted to analyze longitudinal mediation. Third, during integration between different longitudinal mediation models, the outstanding performance is the integration of CLPM and MLM into MAMM. Fourth, the method testing mediation analysis is compared. Bayesian method should be adopted in mediation analysis of MAMM and MTVCM. Bootstrap method should be adopted in mediation analysis of LGM.

In the present study, we propose a procedure to analyze longitudinal mediation analysis. The first step is to decide whether it is necessary to make a causal inference. If the aim of research is to make a causal inference, then proceed with the second step. Otherwise, LGM or MLM should be adopted to analyze longitudinal mediation. In the second step, we decide whether it is necessary to test time-varying effect of mediation effect. If the aim of research is to test the time-varying effect of mediation effect, CTM should be adopted to analyze longitudinal mediation. Otherwise, proceed with the third step. The third step is to decide whether it is necessary to investigate the individuals-varying effect of mediation effect. If the aim of research is to investigate the individuals-varying effect of mediation effect, RE-CLPM model or MAMM should be adopted to analyze longitudinal mediation. Otherwise, CLPM should be adopted to analyze longitudinal mediation.

Key words longitudinal data, mediation effect, cross-lagged panel model, multilevel model, latent growth model

附录：纵向数据的中介效应分析的 *Mplus* 程序

程序 1：固定效应的 CLPM 的中介分析

```
DATA:          File = 1.txt;
VARIABLE:      names = x1-x4 m1-m4 y1-y4; lx, m, y 都重复测量四次
ANALYSIS:      bootstrap is 5000;
MODEL:         x2 on x1;    x3 on x2;    x4 on x3;    m2 on m1 x1;    m3 on m2 x2;    m4 on m3 x3;
               y2 on y1 m1;    y3 on y2 m2 x1;    y4 on y3 m3 x2    x1 with m1 y1;    m1 with y1;
MODEL INDIRECT: y3 ind x1;    y4 ind x2;
OUTPUT:        CINTERVAL (BOOTSTRAP)  SAMPSTAT  STDYX;
```

注：本程序不考虑变量间关系具有跨时间的稳定性。如果要保持变量间关系的稳定性，可增加限制性的语句，如 $x2 \text{ on } x1(\beta_x)$; $x3 \text{ on } x2(\beta_x)$; $x4 \text{ on } x3(\beta_x)$; $!(\beta_x)$ 使 $x1-x4$ 的自回归系数 β_x 都相等。

程序 2：随机效应的 CLPM 的中介分析

```
DATA:          File = 1.txt;
VARIABLE:      names = x1-x4 m1-m4 y1-y4;
ANALYSIS:      type=random;
MODEL:         x2 on x1;    x3 on x2;    x4 on x3;    m2 on m1;    m3 on m2;    m4 on m3;
               y2 on y1;    y3 on y2;    y4 on y3;    x1 with m1 y1;    m1 with y1;
               b1| y2 on m1; b2| y3 on m2;    b3| y4 on m3;    !设置随机效应
               a1| m2 on x1; a2| m3 on x2;    a3| m4 on x3;    c1|y3 on x1;    c2|y4 on x2;
               [a1] (am1); [b2] (bm1); a1 with b2 (cova1b2);    !随机效应 a1 和 b2 的均值为 am1 和 bm1, 协方差为 cova1b2
               [a2] (am2); [b3] (bm2); a2 with b3 (cova2b3);
model constraint: new (mab1 mab2);
               mab1=am1*bm1+cova1b2; mab2=am2*bm2+cova2b3;    !求中介效应
```

注：本程序将自回归效应设为固定效应（避免收敛问题），仅将滞后效应设为随机效应，且不考虑变量间关系具有跨时间的稳定性。如果要保持变量间关系的稳定性，可增加限制性的语句，如 $a1| m2 \text{ on } x1$; $a1| m3 \text{ on } x2$; $a1| m4 \text{ on } x3$;

程序 3：基于 LGM 的中介效应分析（仅呈现与程序 1 不同的部分）

```
MODEL:         IX SX | x1@0 x2@1 x3@2 x4@3;
               IM SM | m1@0 m2@1 m3@2 m4@3;
               IY SY | y1@0 y2@1 y3@2 y4@3;
               IM ON IX;    IY ON IX IM;
               SM ON SX;    SY ON SX SM;
               IX WITH SX;    IM WITH SM;    IY WITH SY;
MODEL INDIRECT: IY IND IM IX;    SY IND SM SX;
```

注：IX 表示变量 X 的初始水平，SX 表示增长速度，以此类推。