arXiv:2304.02270v2 [stat.ME] 4 Jul 2023

具有分类工具变量的非可忽略无应答数据的可验证可识别性条件

Kenji Beppu[a] 和 Kosuke Morikawa[a]

[a]大阪大学工程科学研究生院，日本大阪

文章历史信息
编于2023年7月6日

摘要
我们研究一种模型识别问题，其中结果变量存在非随机缺失值。进行统计推断需要确保模型具有可识别性，从而获得具备一致性和渐近正态性等理论上合理性质的估计量。近年来，研究者强调结合结果模型中的完备性条件，利用工具变量或影子变量实现模型可识别。本文阐明工具变量为分类变量时，完备性条件与模型可识别性之间的关系。我们首先证明，当结果变量和工具变量均为分类变量时，这两个条件等价。然而，当结果变量或工具变量中有一个是连续变量时，即使对于简单模型，完备性条件也未必成立。因此，我们提出一个充分条件，以保证具有单调似然性质的模型可识别；当完备性条件难以确立时，这一条件尤其有用。利用观测数据，我们表明，对于许多实际模型，所提出的条件很容易检验，并通过数值实验和真实数据分析说明其用途。

关键词
非随机缺失；非可忽略缺失；可识别性；工具变量；指数族

1. 引言

人们日益广泛地开始利用所有可用数据，而这些数据可能明确甚至隐含地包含缺失值，例如因果推断（Imbens 和 Rubin，2015）和数据整合（Hu、Wang、Li 和 Miao，2022；Yang 和 Kim，2020）。对于此类数据集，恰当分析缺失数据对于纠正由缺失造成的选择偏倚不可或缺。近年来，基于随机缺失（MAR）假设的缺失数据分析（Little 和 Rubin，2019年）已逐渐成熟（Kim 和 Shao，2021；Robins、Rotnitzky 和 Zhao，1994）。尽管模型可识别性是构建渐近理论最基本的条件之一，但去除随机缺失假设会使统计推断变得极为困难，尤其是在模型识别方面（Miao、Ding 和 Geng，2016）。基于不可识别模型进行估计，可能会得到多个完全相同的

电子邮箱：k.morikawa.es@osaka-u.ac.jp

模型拟合结果。已有多位研究者探讨了非随机缺失（MNAR）条件下模型识别的充分条件。

构建观测似然包括两个分布：(R) 响应机制和 (O) 结局分布（Kim 和 Shao，2021）。Miao 等人（2016）针对 (R) 的逻辑斯蒂、Probit 和 Robit（ t-分布的累积分布函数）模型，以及 (O) 的正态分布和 t （混合）分布，考虑了识别条件。Cui、Guo 和 Yang（2017）假设 (R) 采用逻辑斯蒂、Probit 和互补对数-对数模型，(O) 采用广义线性模型。这些研究高度依赖于 (R) 和 (O) 的模型设定。Wang、Shao 和 Kim（2014）引入了一种称为工具变量或影子变量的协变量，并证明使用工具变量可以显著放宽对 (R) 和 (O) 的条件。例如，(O) 只需满足单调似然性质，该性质涵盖多种模型，例如广义线性模型。Tang、Little 和 Raghunathan（2003）以及 Miao 和 Tchetgen（2018）借助工具变量，在不对 (R) 提出任何假设的情况下，推导出了模型可识别性的条件。Miao、Liu、Tchetgen 和 Geng（2019）在 (R) 满足完备性条件这一假设下，进一步放宽了该假设（D'Haultfœuille，2010，2011）。例如，具有连续协变量的广义线性模型满足完备性条件。据我们所知，在 (R) 中使用工具变量并在 (O) 中使用完备性条件的这种组合，是模型识别最一般的条件，并已被众多研究所采用（Yang、Wang 和 Ding，2019；Zhao 和 Ma，2022）。

通常，关于（O）的假设依赖于完整数据的分布，而这一分布无法通过观测数据检验。近年来，研究者开始对观测结果模型或应答者结果模型（O'）建模，而非采用（O），以放宽主观假设（Miao 等人，2019年；Riddles、Kim 和 Im，2016）。然而，同时采用（R）和（O'）的观测似然涉及积分，使得识别问题难以处理。Morikawa 和 Kim（2021）以及 Beppu、Morikawa 和 Im （2021）证明，对于（R）采用逻辑斯蒂模型、对于（O'）采用广义线性模型时，可以显式计算该积分，并推导出识别条件。对于一般的响应机制和应答者结局分布，模型识别仍是一个未解决的问题。此外，当工具变量为吸烟史和性别等分类变量时，完备性条件并不成立。例如，Ibrahim、Lipsitz 和 Horton （2001年）研究了康涅狄格州儿童的心理健康，并使用父母对儿童精神病理状况的报告作为二元工具变量。

<escape>本文中，我们考虑了一个关于 (R) 和 (O') 的工具变量识别问题，其中工具变量满足单调似然比性质。需要注意的是，尽管我们的模型设定与 Wang等人(2014) 相似，但我们可以利用观测数据检验 (O') 的有效性，例如，采用赤池信息准则（AIC）和贝叶斯信息准则（BIC）等信息准则。此外，我们可以使用半参数/非参数方法对 (O') 和 (R) 进行建模。</escape>

本文其余部分的组织如下。第 2 节介绍符号并定义模型可识别性。第 3 节推导所提出的识别条件。我们在第4节通过一项有限的数值研究展示可识别性的影响。此外，第 5节介绍真实数据的应用。最后，在第6节总结结论性意见。所有技术性证明均置于附录中。

2. 基本设定

2.1. 观测似然

设 $\{\boldsymbol{x}_i, y_i, \delta_i\}_{i=1}^n$ 是来自 $(\boldsymbol{x},\ y,\ \delta)$ 分布的独立同分布样本，其中 $\boldsymbol{x}$ 是完全观测的协变量向量，y 是存在缺失的结果变量，δ 是 y 的响应指示变量：当 y 被观测（缺失）时取 1（0）。我们分别使用通用记号 $p(\cdot)$ 和 $p(\cdot \mid \cdot)$ 表示边缘密度和条件密度。例如，$p(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 的边缘密度，而 $p(y \mid \boldsymbol{x})$ 是给定 $\boldsymbol{x}$ 时 y 的条件密度。我们对非随机缺失（MNAR）响应机制 $P(\delta = 1 \mid \boldsymbol{x},\ y)$ 建模，并考察其可识别性。观测似然定义为

$$\prod_{i:\delta_i=1} P(\delta_i = 1 \mid y_i, \boldsymbol{x}_i) p(y_i \mid \boldsymbol{x}_i) \prod_{i:\delta_i=0} \int \{1 - P(\delta_i = 1 \mid y, \boldsymbol{x}_i)\}\, p(y \mid \boldsymbol{x}_i) dy. \quad (1)$$

如果（1）中的参数可识别，我们就称该模型是可识别的，这等价于 P（$\delta = 1 \mid y,\ \boldsymbol{x}$）$p$（$y \mid \boldsymbol{x}$）中的参数可识别。即使对于由矩条件定义的估计量等半参数模型，这一识别条件也至关重要（Morikawa 和 Kim，2021）。然而，简单的模型也很容易变得不可识别。例如，Wang等人（2014）中的例1介绍了一个不可识别模型，其中结果模型服从正态分布，响应机制为逻辑斯蒂模型。

还有一种表达 y 与 $\boldsymbol{x}$ 之间关系的方式。对 $p(y \mid \boldsymbol{x})$ 建模的一个缺点是，它对完整数据的分布而非观测数据的分布作出了主观假设。换言之，如果我们对 $p(y \mid \boldsymbol{x})$ 作出假设并确保其可识别性，就无法利用观测数据验证这些假设。相比之下，通过对 $p(y \mid \boldsymbol{x}, \delta = 1)$ 建模可以克服这一问题，因为 $p(y \mid \boldsymbol{x}, \delta = 1)$ 是观测数据的结果模型，我们可以使用 AIC 和 BIC 等序数信息准则检查其有效性。因此，我们对 $p(y \mid \boldsymbol{x}, \delta = 1)$ 建模，并考虑第 3 节中的识别条件。下文中，我们假设两个参数模型 $p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})$ 和 $P(\delta = 1 \mid \boldsymbol{x}, y; \boldsymbol{\phi})$，其中 $\boldsymbol{\gamma}$ 和 $\boldsymbol{\phi}$ 分别是结果模型和响应模型的参数。尽管我们的方法需要两个参数模型，但可识别模型的类别非常广泛。例如，除逻辑斯蒂模型之外，它还可以包括 $p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})$ 的半参数结果模型和一般响应模型 $P(\delta = 1 \mid \boldsymbol{x}, y; \boldsymbol{\phi})$，正如例3.7中所讨论的。

2.2. 估计

我们提出一种基于参数模型的参数估计方法 $p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})$ 和 $P(\delta = 1 \mid \boldsymbol{x}, y; \boldsymbol{\phi})$。设 $\hat{}\boldsymbol{\gamma}$ 为 $\boldsymbol{\gamma}$ 的最大似然估计量。观测似然（1）导出关于 $\boldsymbol{\phi}$ (Kim 和 Shao，2021）的平均得分方程：

$$\sum_{i=1}^n \left\{ \delta_i \frac{\partial \log \pi(\boldsymbol{x}_i, y_i; \boldsymbol{\phi})}{\boldsymbol{\phi}} - (1 - \delta_i) \frac{\int \partial \pi(\boldsymbol{x}_i, y; \boldsymbol{\phi}) / \partial \boldsymbol{\phi} \cdot p(y \mid \boldsymbol{x}) dy}{\int \{1 - \pi(\boldsymbol{x}_i, y; \boldsymbol{\phi})\} p(y \mid \boldsymbol{x}) dy} \right\} = 0$$

其中 $\pi(\boldsymbol{x}, y; \boldsymbol{\phi}) = P(\delta = 1 \mid \boldsymbol{x}, y; \boldsymbol{\phi})$。根据贝叶斯公式 $p(y \mid \boldsymbol{x}) \propto p(y \mid \boldsymbol{x}, \delta = 1)/\pi(\boldsymbol{x}, y; \boldsymbol{\phi})$，平均得分可写为

$$\sum_{i=1}^{n} \{\delta_i s_1(\boldsymbol{x}_i, y_i; \boldsymbol{\phi}) + (1 - \delta_i) s_0(\boldsymbol{x}_i; \boldsymbol{\phi})\} = 0,$$

其中

$$s_1(\boldsymbol{x}, y; \boldsymbol{\phi}) = \frac{\partial \log \pi(\boldsymbol{x}, y; \boldsymbol{\phi})}{\partial \boldsymbol{\phi}}, \; s_0(\boldsymbol{x}; \boldsymbol{\phi}) = -\frac{\int s_1(\boldsymbol{x}, y; \boldsymbol{\phi}) p(y \mid \boldsymbol{x}, \delta = 1) dy}{\int \{1/\pi(\boldsymbol{x}, y; \boldsymbol{\phi}) - 1\} p(y \mid \boldsymbol{x}, \delta = 1) dy}.$$

为了计算 $s_0(\cdot)$ 中的两个积分，我们可以使用分数插补（Kim，2011）。正如 Riddles 等人(2016) 所述，EM算法也同样适用。

3. 可识别性

3.1. 可识别性的定义

回想一下，式（1）中的可识别性条件针对的是 $P(\delta = 1 \mid y, \boldsymbol{x}) p(y \mid \boldsymbol{x})$ 中的参数。如第2.2节所示，条件密度 $p(y \mid \boldsymbol{x})$ 可表示为 $p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})$，而 $P(\delta = 1 \mid \boldsymbol{x}, y; \boldsymbol{\alpha}, \boldsymbol{\phi})$ 则可由贝叶斯公式表示。因此，利用该公式，这些模型中的可识别性问题转变为 $\varphi(y, \boldsymbol{x}; \boldsymbol{\phi}, \boldsymbol{\gamma})$ 中参数的可识别性，其中

$$\varphi(y, \boldsymbol{x}; \boldsymbol{\phi}, \boldsymbol{\gamma}) = \frac{p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})}{\int p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})/\pi(\boldsymbol{x}, y; \boldsymbol{\phi}) dy}. \tag{2}$$

严格来说，识别条件 $\varphi(y, \boldsymbol{x}; \boldsymbol{\phi}, \boldsymbol{\gamma}) = \varphi(y, \boldsymbol{x}; \boldsymbol{\phi}', \boldsymbol{\gamma}')$ 以概率 1 意味着 $(\boldsymbol{\phi}^\top, \boldsymbol{\gamma}^\top) = (\boldsymbol{\phi}'^\top, \boldsymbol{\gamma}'^\top)$。通常，式 (2) 分母中的积分没有闭式解，这使得推导可识别性的充分条件颇具挑战。Morikawa 和 Kim（2021）发现，响应模型和结果模型采用逻辑斯蒂模型与正态分布的组合时，积分具有闭式解，并据此推导出模型可识别性的充分条件。Beppu 等人（2021）将模型扩展到结果模型属于指数族、而响应模型仍为逻辑斯蒂模型的情形。然而，当响应机制较为一般时，即使是正态分布这类简单的结果模型也可能不可识别。

例3.1. 假设受访者结果模型为 $y \mid (\delta = 1, x) \sim N(\gamma_0 + \gamma_1 x, 1)$，响应模型为 $P(\delta = 1 \mid x, y) = \Psi(\alpha_0 + \alpha_1 x + \beta y)$，其中 Ψ 是一个已知的分布函数，且（2）中的积分存在；那么，该模型是不可识别的。例如，不同的参数化方式 $(\alpha_0, \alpha_1, \beta, \gamma_0, \gamma_1) = (0, 1, 1, 0, 1)$ 和 $(\alpha_0', \alpha_1', \beta', \gamma_0', \gamma_1') = (0, 3, -1, 0, 1)$ 会产生相同的观测似然值。

最近，已有研究提出了广泛适用的充分条件。假设协变量 $\boldsymbol{x}$ 有两个分量，$\boldsymbol{x} = (\boldsymbol{u}^\top, \boldsymbol{z}^\top)^\top$，使得

(C1) $\boldsymbol{z} \perp\!\!\!\perp \delta \mid (\boldsymbol{u}, y)$ and $\boldsymbol{z} \not\perp\!\!\!\perp y \mid (\delta = 1, \boldsymbol{u})$.

协变量 $\boldsymbol{z}$ 被称为工具变量（D'Haultfœuille，2010）或影子变量（Miao 和 Tchetgen Tchetgen，2016）。Miao 等人(2019年）通过结合工具变量和完备性条件，推导出了模型可识别性的充分条件：

(C2) 对于所有平方可积函数 $h(\boldsymbol{u}, y)$, $E[h(\boldsymbol{u}, y) \mid \delta = 1, \boldsymbol{u}, \boldsymbol{z}] = 0$ 几乎必然意味着 $h(\boldsymbol{u}, y) = 0$ 几乎必然。

引理 3.2 (Miao 等人提出的识别条件 Miao 等人(2019年))。 在条件 *(C1)* 和 *(C2)* 下，联合分布 $p(y, \boldsymbol{u}, \boldsymbol{z}, \delta)$ 是可识别的。

尽管完备性条件适用于一般模型且实用，但一个包含分类工具变量的简单模型并不满足完备性条件。

例 3.3 （分类工具变量下违反完备性）. 假设 $y \mid (\delta = 1, u, z)$服从正态分布 $N(u + z, 1)$，且工具变量 z为二元变量，取值为 0 或 1。该分布不满足完备性条件，因为条件期望 $E[h(u, y) \mid \delta = 1, u, z] = 0$当 $h(u, y) = 1 + y - u - (y - u)^2$时。

Example 的一个重要含义3.3是，当工具变量为分类变量时，工具变量不再是模型识别的证据。针对具有离散工具变量的模型发展识别条件对于应用而言十分重要（Ibrahim 等人，2001年）。我们分别讨论两种情况：（i）两者 y 和 $\boldsymbol{z}$ 均为分类变量；（ii）受访者结果模型具有单调似然比性质。

当所有变量 y 和 $\boldsymbol{z}$均为分类变量时，该模型可以是完全非参数的。定理 3.4表明，在这些条件下，完备性条件与可识别性条件等价。关于此类完全非参数模型的估计，请参见 Riddles 等人 (2016的附录 2。

定理 3.4。 当 y 和 $\boldsymbol{z}$ 均为分类变量时，在条件 *(C1)* 下，联合分布 $p(y, \boldsymbol{u}, \boldsymbol{z}, \delta)$ 可识别，当且仅当条件 *(C2)* 成立。

正如引理 3.2所证明的，条件 (C2) 通常足以保证模型的可识别性，但定理3.4 还揭示了当 y 和 $\boldsymbol{z}$ 均为分类变量时，该条件也是必要的。

接下来，我们考虑另一种情形（ii）的识别条件。令 $\mathcal{S}_y$ 为随机变量 y的支撑集。我们假设以下四个条件：

（C3）响应机制是

$$P(\delta = 1 \mid y, \boldsymbol{x}; \boldsymbol{\phi}) = P(\delta = 1 \mid y, \boldsymbol{u}; \boldsymbol{\phi}) = \Psi\{h(\boldsymbol{u}; \boldsymbol{\alpha}) + g(\boldsymbol{u}; \boldsymbol{\beta})m(y)\}, \quad (3)$$

其中 $\boldsymbol{\phi} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta})^\top$, $m : \mathcal{S}_y \to \mathbb{R}$ 和 $\Psi : \mathbb{R} \to (0, 1]$ 是已知的连续严格单调函数，且 $h(\boldsymbol{u}; \boldsymbol{\alpha})$ 和 $g(\boldsymbol{u}; \boldsymbol{\beta})$ 分别是 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$的已知单射函数。

（C4）密度函数或质量函数 $p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})$ 是可识别的，且其支撑集不依赖于 $\boldsymbol{x}$。

（C5）对于所有 $\boldsymbol{u} \in \mathcal{S}_{\boldsymbol{u}}$，存在 $\boldsymbol{z}_1$和 $\boldsymbol{z}_2$，使得 $p(y \mid \boldsymbol{u}, \boldsymbol{z}_1, \delta = 1) \neq p(y \mid \boldsymbol{u}, \boldsymbol{z}_2, \delta = 1)$，且 $p(y \mid \boldsymbol{u}, \boldsymbol{z}_1, \delta = 1)/p(y \mid \boldsymbol{u}, \boldsymbol{z}_2, \delta = 1)$ 是单调的。

(C6)

$$\int \frac{p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})}{\Psi\{h(\boldsymbol{u}; \boldsymbol{\alpha}) + g(\boldsymbol{u}; \boldsymbol{\beta})m(y)\}} dy < \infty \quad \text{a.s.}$$

条件（C3）意味着随机变量 $\boldsymbol{z}$ 起到工具变量的作用。条件（C4）是 $p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})$ 的可识别性，可通过观测数据进行检验。条件（C5）假设结果模型具有单调似然性质，这一性质也曾用于Wang等人(2014)对完整数据的研究中。

条件（C6）是使（1）定义良好的必要条件。本质上，它与Morikawa 和 Kim(2021) 的定理 3.1（I1）是相同的条件。当 y 的支撑集有限时，该条件总是成立。然而，当y 是连续变量时，必须仔细验证该条件。对于受访者结果模型服从正态分布的情况，以下命题 3.8 给出了有用的充分条件。

在条件（C3）–（C6）下，我们得到了所需的识别条件。

定理 3.5. 若条件（*C1*）和（*C3*）–（*C6*）成立，则参数$(\boldsymbol{\phi}^\top, \boldsymbol{\gamma}^\top)^\top$ 可识别。

我们给出一个满足条件（C5）的结果模型示例。

示例 3.6 （满足（C5）的模型）**.** 设指数族中的密度函数为

$$p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma}) = \exp\left(\frac{y\theta - b(\theta)}{\tau} + c(y; \tau)\right),$$

其中 $\theta = \theta(\eta)$、 $\eta = \sum_{l=1}^{L} \eta_l(\boldsymbol{x})\kappa_l$, $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_L)^\top$ 和 $\boldsymbol{\gamma} = (\tau, \boldsymbol{\kappa}^\top)^\top$。则密度比变为

$$\frac{p(y \mid \boldsymbol{u}, \boldsymbol{z}_1, \delta = 1)}{p(y \mid \boldsymbol{u}, \boldsymbol{z}_2, \delta = 1)} \propto \exp\left(\frac{\theta_1 - \theta_2}{\tau} y\right),$$

其中 $\boldsymbol{x}_i = (\boldsymbol{u}, \boldsymbol{z}_i)$ 和 $\theta_i = \theta\{\sum_{l=1}^{L} \eta_l(\boldsymbol{x}_i)\kappa_l\}$, $i = 1,2$。因此，密度比是单调的。

例3.7 （满足（C6）的模型）。 在实际应用中，假设受访者结果模型服从正态分布通常是合理的。着眼于结果模型的尾部，我们为具有一般响应机制的模型提供了一个检验（C6）的充分条件。

命题 3.8。 假设观测分布 $p(y \mid \boldsymbol{x}, \delta = 1)$ 是正态分布 $N(\mu(\boldsymbol{x}; \boldsymbol{\kappa}) \sigma^2)$，响应机制为 *(3)*，其中 $m(y) = y$ 且$g(\boldsymbol{u}; \boldsymbol{\beta}) = \beta$，并且严格单调递增函数 Ψ 满足以下条件：

$$^{\exists} s \in (0, 2) \text{ s.t. } \liminf_{z \to -\infty} \Psi(z) \exp(|z|^s) > 0. \tag{4}$$

那么，该模型满足（*C6*）。

条件（4）很容易检验。例如，对于逻辑斯蒂和 Robit 函数，该条件成立，但对于 Probit 函数则不成立。根据命题 3.8，可以利用样条及其他非参数方法，基于观测数据估计 $\mu(x; \boldsymbol{\kappa})$，从而使用非常灵活的模型。此外，我们也可以用非参数方法估计响应机制，因为这不会对 $h(\boldsymbol{u}; \boldsymbol{\alpha})$ 的函数形式施加任何限制。

4. 数值实验

我们通过比较弱可识别模型和强可识别模型，展示数值实验中的可识别性效应。我们准备了四种情景 S1–S4：

S1：（结果变量：正态分布，响应变量：逻辑斯蒂）$[y \mid u, z, \delta = 1] \sim N(\kappa_0 + \kappa_1 u + \kappa_2 z, \sigma^2)$，$\text{logit}\{P(\delta = 1 \mid u, y; \boldsymbol{\alpha}, \beta)\} = \alpha_0 + \alpha_1 u + \beta y$，$u \sim N(0, 1^2)$，以及 $z \sim B(1, 0.5)$，其中 $(\kappa_0, \kappa_1, \sigma^2)^\top = (0.3, 0.4, 1/\sqrt{2^2})^\top$ 且 $(\alpha_0, \alpha_1, \beta)^\top = (0.7, -0.2, 0.29)^\top$。S2：（结果变量：正态分布，响应变量：柯西）$[y \mid u, z, \delta = 1] \sim N(\kappa_0 + \kappa_1 u + \kappa_2 z, \sigma^2)$，$P(\delta = 1 \mid u, y; \boldsymbol{\alpha}, \beta) = \Psi(\alpha_0 + \alpha_1 u + \beta y)$，$u \sim \text{Unif}(-1, 1)$，以及 $z \sim B(1, 0.7)$，其中 $(\kappa_0, \kappa_1, \sigma^2)^\top = (-0.36, 0.59, 1/\sqrt{2^2})^\top$，$(\alpha_0, \alpha_1, \beta)^\top = (0.24, -0.1, 0.42)^\top$，且 Ψ 是柯西分布的累积分布函数。S3：（结果变量：伯努利分布，响应变量：Probit）$[y \mid u, z, \delta = 1] \sim B(1, p(u, z; \boldsymbol{\kappa}))$，$P(\delta = 1 \mid u, y; \boldsymbol{\alpha}, \beta) = \Psi(\alpha_0 + \alpha_1 u + \beta y)$，$u \sim N(0, 1^2)$，以及 $z \sim N(0, 1^2)$，其中 $p(u, z; \boldsymbol{\kappa}) = 1/\{1 + \exp(-\kappa_0 - \kappa_1 u - \kappa_2 z)\}$，$(\kappa_0, \kappa_1, \kappa_2)^\top = (-0.21, 3.8, 1.0)^\top$，$(\alpha_0, \alpha_1, \beta)^\top = (0.4, 0.39, 0.3)^\top$，且 Ψ 是标准正态分布的累积分布函数。S4：（结果变量：具有非线性均值结构的正态分布，响应变量：柯西分布或逻辑斯蒂分布）$[y \mid u, z, \delta = 1] \sim N(\mu(\boldsymbol{x}) 0.52)$，$P(\delta = 1 \mid u, y; \boldsymbol{\alpha}, \beta) = \Psi(\alpha_0 + \alpha_1 u + \beta y)$，$u \sim \text{Unif}(-1, 1)$，以及 $z \sim B(1, 0.5)$，其中 $\mu(\boldsymbol{x}) = z + \cos(2\pi u) + \exp(z + u)$，$(\alpha_0, \alpha_1, \beta)^\top = (0.1, -0.2, 0.3)^\top$，且 Ψ 是柯西分布或逻辑斯蒂分布的累积分布函数。

在 S1 和 S2 中，可以通过改变参数 κ_2 来调整可识别性的强度，因为 $\kappa_2 = 0$ 表明该模型根据例3.1是不可识别的。另一方面，根据定理 3.5，我们可以验证 S3 和 S4 中的模型是可识别的。例如，在 S4 中，我们可以看到，检验条件 (C3) 和 (C4) 对于该设定而言是直接的，而 (C5) 和 (C6) 则分别由例 3.6 和命题3.8保证。通过 S3 和 S4，我们可以确认，即使在结果变量为离散变量以及均值结构复杂的情况下，也能分别成功进行推断。

我们生成了 1,000 个相互独立的蒙特卡洛样本，并使用两种方法计算了 $E[y]$ 和 β 的两种估计量：分数插补（FI）和完全案例（CC）估计量，后者仅使用完全观测的数据。$E[y]$ 的估计量通过使用估计的响应模型的标准逆概率加权方法计算（Riddles 等人，2016）。对于情景 S1–S3，我们使用了设定正确的模型；但对于情景 S4，我们使用了非参数模型，因为假设复杂的均值结构已知并不现实。我们使用 R 软件包 'crs' 来估计受访者结果模型；该软件包专门用于对分类变量和连续变量协变量的混合进行非参数样条回归（Nie 和 Racine，2012）。响应模型通过第 2.2节中讨论的方法进行估计。

表1报告了 S1–S4 中 95% 置信区间的偏差、均方根误差（RMSE）和覆盖率。在所有情景中，CC 估计量均存在显著偏差，覆盖率也远低于 95%；而当模型确实可识别时，FI 估计量表现良好。当 κ_2 在 S1 和 S2 中较小时，正如预期，FI 的方差估计表现不佳，但点估计尚可。S4 的结果表明，即使采用非参数均值结构，模型仍可识别，而且两种响应模型的估计结果几乎相同。

表 1. S1–S4 的结果：报告偏差、均方根误差（RMSE）以及 95% 置信区间的覆盖率（CR，%）。CC：完整案例；FI：分数插补。

情景	参数	κ_2	方法	Bias	RMSE	CR
S1	$E[y]$	1.0	CC	0.053	0.066	73.5
			FI	0.000	0.043	95.4
		0.5	CC	0.039	0.053	80.9
			FI	-0.001	0.059	97.1
		0.1	CC	0.034	0.049	83.0
			FI	0.021	0.136	99.8
	β	1.0	FI	0.001	0.163	95.2
		0.5	FI	0.003	0.330	98.6
		0.1	FI	-0.146	0.865	100
S2	$E[y]$	1.0	CC	0.146	0.152	5.7
			FI	-0.004	0.051	94.8
		0.5	CC	0.130	0.136	7.7
			FI	-0.008	0.086	86.2
		0.1	CC	0.127	0.133	9.4
			FI	-0.007	0.105	92.4
	β	1.0	FI	0.008	0.148	95.4
		0.5	FI	0.044	0.365	100
		0.1	FI	0.033	0.448	100
S3	$E[y]$	–	CC	0.100	0.102	0.3
		–	FI	0.001	0.022	95.3
	β	–	FI	-0.023	0.279	95.0
S4	$E[y]$	–	CC（逻辑斯蒂）	0.341	0.355	5.4
		–	FI（逻辑斯蒂）	0.005	0.079	95.4
		–	CC（柯西）	0.296	0.312	10.7
		–	FI（柯西）	0.007	0.080	94.3
	β	–	FI（逻辑斯蒂）	0.006	0.050	94.7
		–	FI（柯西）	0.011	0.063	93.8

5. 真实数据分析

我们分析了 2139 名 HIV阳性患者的数据，他们参加了艾滋病临床试验组研究175（ACTG175；Hammer、Katzenstein、Hughes、Gundacker、Schooley、Haubrich、Henry、Lederman、Phair、Niu 等人（1996））。在本分析中，我们选取其中接受齐多夫定（ZDV）单药治疗的 532 名患者进行分析。令每个 y、 x_1和x_2 分别表示 96 ± 5 周时、基线时以及 20 ± 5 周时的 CD4细胞计数， x_3 表示基线时的 CD8细胞计数，而 z 表示性别。结局存在缺失，观测率为 60.34%，而所有协变量均有观测。为使估计稳定且简便，我们对所有数据进行了标准化。我们认为， z （性别）是合理的工具变量，因为该信息属于生物学特征，会影响 CD4 值，但对响应概率影响甚微。

患有轻度 HIV 疾病的患者往往 CD4细胞计数较高；因此，可以认为结局的缺失与病情严重程度有关

，并预期结局缺失值对应的 CD4细胞计数会低于应答者。因此，我们考虑了五种不同的非随机缺失响应模型：

$$P(\delta = 1 \mid x_1, x_2, x_3, y) = \Psi(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \beta y),$$

其中， Ψ表示逻辑斯蒂函数，或柯西分布或 t 自由度为 $v\,(=\ 2、5、10)$的分布函数。定理3.5 和命题3.8保证了采用这五种响应模型的所有模型均可识别，即使工具变量 z 是离散的。根据上述关于缺失值的猜想，预计 β 的符号为负。我们假设受访者的结果服从正态分布，且均值结构为非参数形式，并使用“crs”R 程序包进行估计，方法与第4节情景 S4 中所述相同。图1 所示的残差图和计算得到的 R^2值（$=\ 0.453$）表明，假设的受访者结果分布拟合良好。表2 报告了估计参数及其标准误，标准误由 1,000 次自助抽样计算得出。五种响应模型的结果几乎相同。这表明响应机制对响应模型的选择具有稳健性。尽管由于 β 的估计标准误较大，我们无法判断数据是非随机缺失（MNAR）还是随机缺失（MAR），但点估计为负，符合预期。这一结果与Zhao、Wang 和 Shao(2021)的结果一致。

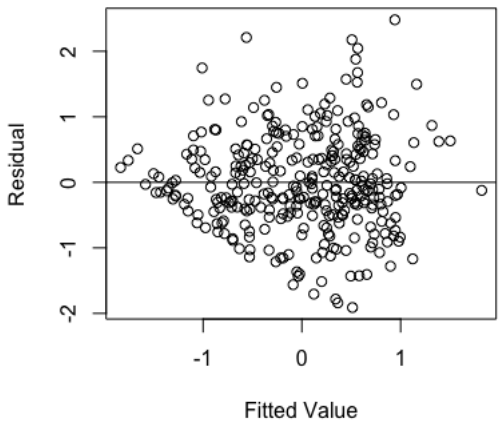


图1. ACTG175数据中受访者结局变量的残差图。

表 2. 估计参数：报告目标参数的估计值和标准误。Logistic 和 Cauchy 分别表示使用逻辑斯蒂分布和柯西分布对响应机制进行分数插补。T_v：自由度为 $v(=2,\ 5,\ 10)$ 的 t 分布函数。

参数	模型	估计值	SE	参数	模型	估计	SE
α_0	逻辑斯蒂	0.464	0.104	α_1	逻辑斯蒂	0.125	0.156
	柯西	0.417	0.260		柯西	0.108	0.139
	T_2	0.341	0.081		T_2	0.091	0.113
	T_5	0.306	0.069		T_5	0.082	0.102
	T_{10}	0.295	0.066		T_{10}	0.080	0.099
α_2	逻辑斯蒂	0.255	0.192	α_3	逻辑斯蒂	0.093	0.107
	柯西	0.244	0.207		柯西	0.083	0.097
	T_2	0.196	0.148		T_2	0.069	0.079
	T_5	0.169	0.126		T_5	0.062	0.070
	T_{10}	0.160	0.120		T_{10}	0.060	0.068
β	逻辑斯蒂	-0.032	0.314	$E[y]$	逻辑斯蒂	276.70	13.476
	柯西	-0.030	0.387		柯西	276.51	14.107
	T_2	-0.027	0.235		T_2	276.57	13.437
	T_5	-0.021	0.203		T_5	276.61	13.271
	T_{10}	-0.019	0.194		T_{10}	276.63	13.217

6. 结论

本文提出了一种针对使用受访者结果模型和响应模型的模型的新识别条件。尽管我们的方法要求设定这两个模型，但借助工具变量，模型可以具有非常一般的形式。如第4节情景 S4 中所述，受访者结果模型中的均值函数可以是非参数的，而响应模型可以是除逻辑斯蒂模型之外的任意严格单调函数。即使工具变量是类别型的，我们的条件也能保证模型可识别，而这不在以往识别条件的适用范围内。使用我们方法的另一项优势是，可以利用观测数据轻松验证识别条件。

然而，我们的方法存在一些局限。首先，根据条件（C5），受访者的结果模型需要满足单调似然性质。例如，我们的框架无法处理混合模型。其次，必须事先指定工具变量。迄今为止，已有一些关于寻找工具变量的研究（Zhao 等人，2021），但目前仍没有公认的标准方法。

资助

第二作者的研究得到了面向地震数据研究创新的地震学文部科学省项目（STAR-E）的资助，资助编号为 JPJ010217。

参考文献

Beppu, K., Morikawa, K., and Im, J. (2021), “针对非可忽略缺失结局且具有可验证可识别性条件的数据插补,” *arXiv*预印本 *arXiv:2204.10508*，。

Cui, X., Guo, J., and Yang, G. (2017), “具有参数化非可忽略缺失数据机制的广义线性模型的可识别性与估计,” Computational Statistics & Data Analysis, 107, 64–80. D’ Haultfœuille, X. (2010), “一种处理内生选择的新工具变量方法,” Journal of Econometrics, 154(1), 1–15. D’Haultfœuille, X. (2011), “关于非参数工具变量问题中的完备性条件,” Econometric Theory, 27(3), 460–471. Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M. 等 (1996), “一项比较核苷类单药疗法与联合疗法在CD4细胞计数为每立方毫米200至500的HIV感染成人中疗效的试验,” New England Jo urnal of Medicine, 335(15), 1081–1090. Hu, W., Wang, R., Li, W., and Miao, W. (2022), “个体数据与汇总数据半参数融合中的悖论及其解决方案,” arXiv预印本 arXiv:2210.00200, 。 Ibrahim, J. G., Lipsitz, S. R., and Horton, N. (2001), “ 利用辅助数据估计结局非随机缺失时的参数,” Journal of the Royal Statistical Society: Series C (Applied Statistics), 50(3), 361–373. Imbens, G. W., and Rubin, D. B. (2015), 《统计学、社会科学与生物医学科学中的因果推断》 Cambridge University Press. Kim, J. K. (2011), “用于缺失数据分析的参数化分数插补,” Biometrika, 98(1), 119–132. Kim, J. K., and Shao, J. (2021), 《处理不完整数据的统计方法》 CRC press. Little, R. J., and Rubin, D. B. (2019), 《含缺失数据的统计分析》, 第793卷 John Wiley & Sons. Miao, W., Ding, P., and Geng, Z. (2016), “非可忽略缺失数据下正态分布模型与正态混合模型的可识别性,” Journal of the American Statistical Association, 111(516), 1673–1683. Miao, W., Liu, L., Tchetgen, E. T., and Geng, Z. (2019), “利用影子变量对非可忽略缺失数据进行可识别性分析、双重稳健估计及半参数效率理论,” arXiv预印本 arXiv:1509.02556, 。 Miao, W., and Tchetgen, E. T. (2018), “非可忽略协变量数据缺失情况下的可识别性与推断,” Statistica Sinica, 28(4), 2049. Miao, W., and Tchetgen Tchetgen, E. J. (2016), “关于非随机缺失且具有影子变量时双重稳健估计量的各种形式,” Biometrika, 103(2), 475–482. Morikawa, K., and Kim, J. K. (2021), “非可忽略无应答数据的半参数最优估计,” The Annals of Statistics, 49(5), 2991–3014. Nie, Z., and Racine, J. S. (2012), “crs软件包：适用于连续和分类预测变量的非参数回归样条,” R Journal, 4(2). Riddles, M. K., Kim, J. K., and Im, J. (2016), “针对非可忽略无应答的倾向评分调整方法,” Journal of Survey Statistics and Methodology, 4(2), 215–245. Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), “部分回归变量并非总能观测到时的回归系数估计,” Journal of the American statistical Association, 89(427), 846–866. Tang, G., Little, R. J., and Raghunathan, T. E. (2003), “非可忽略无应答情况下的多变量缺失数据分析,” Biometrika, 90(4), 747–764. Wang, S., Shao, J., and Kim, J. K. (2014), “利用工具变量对非可忽略无应答进行可识别性分析与估计,” Statistica Sinica, 24, 1097–1116. Yang, S., and Kim, J. K. (2020), “调查抽样中的统计数据整合：综述,” Japanese Journal of Statistics a nd Data Science, 3, 625–650. Yang, S., Wang, L., and Ding, P. (2019), “混杂因素非随机缺失时的因果推断,” Biometrika, 106(4), 875–888. Zhao, J., and Ma, Y. (2022), “无需估计非可忽略缺失机制的通用估计程序,” Journal of the American Statistical Association, 117(540), 1916–1930.

Zhao, P., Wang, L., 和 Shao, J.（2021），《针对非可忽略无响应数据的充分降维与工具变量搜索》，*Bernoulli*，**27**，930–945。

附录 A. 技术证明

我们首先给出一个技术结果，以证明定理3.4。

引理 A.1. 设 a、b 和 c 为任意正实数。假设 r_1 和 r_2 为满足以下条件的正实数

$$-\frac{ab}{a+b} < \frac{r_1^2 - r_2^2}{r_1^2 r_2^2} < c. \tag{A1}$$

那么，存在 $0 < \pi_j^{(k)} < 1\,(j = 1、2、3;\ k = 12)$ 使得

$$\sum_{j=1}^{3} \pi_j^{(1)} = r_1^2, \quad \sum_{j=1}^{3} \pi_j^{(2)} = r_2^2, \tag{A2}$$

and

$$\frac{1}{\pi_1^{(1)}} - \frac{1}{\pi_1^{(2)}} = a, \quad \frac{1}{\pi_2^{(1)}} - \frac{1}{\pi_2^{(2)}} = b, \quad \frac{1}{\pi_3^{(1)}} - \frac{1}{\pi_3^{(2)}} = -c. \tag{A3}$$

引理*A.1*的证明。利用极坐标系，我们将 $\pi_j^{(k)}$ $(j = 1,\ 2,\ 3;\ k = 1,\ 2)$ 转换为

$$\begin{aligned}
(\sqrt{\pi_1^{(1)}}, \sqrt{\pi_2^{(1)}}, \sqrt{\pi_3^{(1)}}) &= r_1(\sin\phi_1 \cos\phi_2, \sin\phi_1 \sin\phi_2, \cos\phi_1), \\
(\sqrt{\pi_1^{(2)}}, \sqrt{\pi_2^{(2)}}, \sqrt{\pi_3^{(2)}}) &= r_2(\sin\psi_1 \cos\psi_2, \sin\psi_1 \sin\psi_2, \cos\psi_1),
\end{aligned}$$

其中 $0 < \phi_1,\ \phi_2,\ \psi_1,\ \psi_2 < \pi/2$ 为确保 $\pi_j^{(k)}$ $(j = 1,\ 2,\ 3;\ k = 1,\ 2)$ 满足（A2）。由（A3）和双角公式可得

$$\begin{aligned}
&r_1^2(1-\omega_1)(1+\omega_2) - r_2^2(1-\omega_3)(1+\omega_4) \\
&= -\frac{a r_1^2 r_2^2}{4}(1-\omega_1)(1+\omega_2)(1-\omega_3)(1+\omega_4),
\end{aligned} \tag{A4}$$

$$\begin{aligned}
&r_1^2(1-\omega_1)(1-\omega_2) - r_2^2(1-\omega_3)(1-\omega_4) \\
&= -\frac{b r_1^2 r_2^2}{4}(1-\omega_1)(1-\omega_2)(1-\omega_3)(1-\omega_4),
\end{aligned} \tag{A5}$$

$$r_1^2(1+\omega_1) - r_2^2(1+\omega_3) = \frac{c r_1^2 r_2^2}{2}(1+\omega_1)(1+\omega_3), \tag{A6}$$

w这里 $\omega_1 = \cos\ 2\phi_1$，$\omega_2 = \cos\ 2\phi_2$，$\omega_3 = \cos 2\psi_1$，以及 $\omega_4 = \cos\ 2\psi_2$。令 $\omega_2 = \omega_4$

以及方程（A4）和（A5）可得

$$r_1^2(1-\omega_1)-r_2^2(1-\omega_3)=-\frac{ar_1^2r_2^2}{4}(1-\omega_1)(1+\omega_2)(1-\omega_3),$$
$$r_1^2(1-\omega_1)-r_2^2(1-\omega_3)=-\frac{br_1^2r_2^2}{4}(1-\omega_1)(1-\omega_2)(1-\omega_3).$$

固定 $\omega_2=1-2a/(a+b)$ 后，上述方程可简化为一个共同方程。

$$r_1^2(1-\omega_1)-r_2^2(1-\omega_3)=-\frac{r_1^2r_2^2ab}{2(a+b)}(1-\omega_1)(1-\omega_3), \tag{A7}$$

维持该条件 $-1<\omega_2<1$。接下来需要证明存在 $-1<\omega_3<1$满足（A6）和（A7）。将方程（A7）关于 ω_1求解，可得

$$\omega_1=1-\frac{r_2^2(1-\omega_3)}{r_1^2+r_1^2r_2^2ab(1-\omega_3)/\{2(a+b)\}}. \tag{A8}$$

将（A8）代入（A6，得到关于ω_3的以下二次方程:

$$f(\omega_3)=\left(\frac{r_1^2r_2^4ab+cr_1^4r_2^4ab}{2(a+b)}-\frac{cr_1^2r_2^4}{2}\right)\omega_3^2-\left(\frac{r_1^4r_2^2ab}{a+b}+cr_1^4r_2^2\right)\omega_3$$
$$+\left(\frac{r_1^2r_2^2ab\left(2r_1^2-r_2^2-cr_1^2r_2^2\right)}{2(a+b)}+\frac{cr_1^2r_2^4}{2}+2r_1^4-2r_1^2r_2^2-cr_1-4r_2^2\right)=0.$$

由（A1）可得

$$f(1)=r_1^2\left(2r_1^2-2r_2^2-2cr_1^2r_2^2\right)<0,\quad f(-1)=2r_1^2\left(r_1^2-r_2^2+\frac{r_1^2r_2^2ab}{a+b}\right)>0,$$

这意味着方程 $f(\omega_3)=0$ 在开区间 $(-1,1)$ 内至少存在一个 ω_3 的解。

□

最后，我们借助引理 3.4 证明定理 A.1。

定理 *3.4* 的证明。 不失一般性，我们将 $\boldsymbol{u}$ 设为一个固定向量，因为以下证明对每个 $\boldsymbol{u}$都成立。令分类变量 y和 z 分别取值于 $\{1,2,\ldots,p\}$ 和 $\{1,2,\ldots,q\}$。我们分别讨论以下三种情形，以证明模型可识别性蕴含完备性条件 (C2)：(i) $p=2$，(ii) $p=3$，以及 (iii) $p\geq 4$，因为“若”部分已由引理3.2证明。

当 $p=2$ 时，条件 (C1) 意味着由 $p(y=j\mid\delta=1,\ z=i)$ 构成的 $q\times 2$ 矩阵在其 $(i,\ j)$ 阶元素 $(i=1,\ 2;\ j=1\ldots,\ q)$ 的秩为 2。因此，可识别模型始终满足完备性条件 (C2)。

对于 $p\geq 3$ 的情形，我们必须证明，当完备性条件被违反时，模型将变得不可识别。完备性条件遭到违反表明存在一个非零向量 $(h_1,\ \ldots,\ h_p)$，使得对于 $z=1,\ \ldots,\ q$，我们

have

$$E[h_y \mid \delta = 1, z] = \sum_{y=1}^{p} h_y p(y \mid \delta = 1, z) = 0. \tag{A9}$$

$(h_1, \cdots, h_p)$ 中的元素并非全都具有相同的符号，且将该向量乘以任意常数都不会影响上述方程。回想一下，模型不可识别意味着，对于某个 $y \in \{1, \ldots, p\}$，存在 $\pi_y^{(1)} \neq \pi_y^{(2)}$，满足 $\sum_{y=1}^{p} p(y \mid \delta = 1, z)/\pi_y^{(1)} = \sum_{y=1}^{p} p(y \mid \delta = 1, z)/\pi_y^{(2)}$。现在，当完备性条件被违反时，我们构造一个不可识别模型。

当 $p = 3$时，不失一般性，我们假设 $h_1 > 0$, $h_2 > 0$, 且 $h_3 < 0$满足条件 $\sum_{y=1}^{3} h_y p(y \mid \delta = 1, z) = 0$对所有 $z \in \{1, \ldots, q\}$均成立。利用引理 A.1，结合 $a = h_1$, $b = h_2$, $c = -h_3$和 $r_1 = r_2 = 1$，我们推导得:

$$\frac{1}{\pi_1^{(1)}} - \frac{1}{\pi_1^{(2)}} = h_1, \quad \frac{1}{\pi_2^{(1)}} - \frac{1}{\pi_2^{(2)}} = h_2, \quad \frac{1}{\pi_3^{(1)}} - \frac{1}{\pi_3^{(2)}} = h_3,$$

其中 $\sum_{j=1}^{3} \pi_j^{(1)} = \sum_{j=1}^{3} \pi_j^{(2)} = 1$。将 h_1、h_2 和 h_3 代入 $\sum_{y=1}^{3} h_y p(y \mid \delta = 1, z) = 0$，表明该模型不可识别。

<escape>最后，我们考虑 $p \geq 4$的情形。假设 $h_y\,(y = 1, \ldots, p)$满足(A9)。在$(h_1, \cdots, h_p$中，我们分别选取三个符号为正、正和负的元素，并将其定义为 a, b和 $-c$其中 $a, b, c > 0$，并且 λ将其设定得足够大，以确保</escape>

$$\lambda > 2 \max\left\{\frac{a+b}{ab}, \frac{1}{c}\right\}. \tag{A10}$$

为便于记号表示，我们将 $(h_1, \cdots, h_p) = (h_1, \cdots, h_{p-3}, a, b, -c)$。当符号组合为负、负、正时，证明的其余部分类似。选取的 λ, $0 < \pi_y^{(k)} < 1\,(y = 1, \ldots, p-3;\ k = 1, 2)$ 已确定得足够小，以满足

$$\left(1 - \sum_{y=1}^{p-3} \pi_y^{(1)}\right)\left(1 - \sum_{y=1}^{p-3} \pi_y^{(2)}\right) \geq \frac{1}{2}, \quad \sum_{y=1}^{p-3} \pi_y^{(1)} < 1, \quad \sum_{y=1}^{p-3} \pi_y^{(2)} < 1, \tag{A11}$$
$$\frac{1}{\pi_y^{(1)}} - \frac{1}{\pi_y^{(2)}} = \lambda h_y, \quad \text{for } y = 1, \ldots, p-3.$$

此外，我们定义 r_1 和 r_2 为

$$r_1^2 = 1 - \sum_{y=1}^{p-3} \pi_y^{(1)}, \; r_2^2 = 1 - \sum_{y=1}^{p-3} \pi_y^{(2)}. \tag{A12}$$

通过这些步骤确定变量，可由（A10）、（A11）得出，并且

(A12) 条件（A1）在 $a=\lambda a$、 $b=\lambda b$ 和 $c=\lambda c$ 下得到满足：

$$\frac{r_1^2-r_2^2}{r_1^2r_2^2} \leq 2(r_1^2-r_2^2) \leq \frac{2}{c}c < (\lambda c),$$
$$-\frac{(\lambda a)(\lambda b)}{(\lambda a)+(\lambda b)} < -\frac{ab}{a+b}\frac{2(a+b)}{ab} = -2r_1^2r_2^2\frac{1}{r_1^2r_2^2} \leq -\frac{1}{r_1^2r_2^2} < \frac{r_1^2-r_2^2}{r_1^2r_2^2}.$$

因此，通过应用引理A.1，我们证明存在 $\pi_{p-2}^{(k)}$、 $\pi_{p-1}^{(k)}$和$\pi_p^{(k)}(k=1,\ 2)$，使得

$$\sum_{y=p-2}^{p}\pi_y^{(1)}=r_1^2,\ \sum_{y=p-2}^{p}\pi_y^{(2)}=r_2^2,$$
$$\frac{1}{\pi_{p-2}^{(1)}}-\frac{1}{\pi_{p-2}^{(2)}}=\lambda a,\quad \frac{1}{\pi_{p-1}^{(1)}}-\frac{1}{\pi_{p-1}^{(2)}}=\lambda b,\quad \frac{1}{\pi_p^{(1)}}-\frac{1}{\pi_p^{(2)}}=-\lambda c.$$

条件（A9）表明，构造的 $\pi_y^{(k)}(y=1,\ \ldots,\ p;\ k=1,\ 2)$ 满足$\sum_{y=1}^{p}\pi_y^{(k)}=1$ 对于 $k=1,\ 2$，且对于任意 $z\in\{1,\ldots,q\}$,

$$\sum_{y=1}^{p}\left(\frac{1}{\pi_y^{(1)}}-\frac{1}{\pi_y^{(2)}}\right)p(y\mid\delta=1,z)=\lambda\sum_{y=1}^{p}h_yp(y\mid\delta=1,z)=0.$$

因此，该模型不可识别。

□

定理*3.5*的证明。考虑 y 为连续变量的情形，因为当 y 为离散变量时，只需将积分改为求和即可。为简化讨论，我们考虑 $\mathcal{S}_y=\mathbb{R}$的情形。令 $\boldsymbol{u}$ 为一个固定值。由于 h 和 g 是单射函数，只需证明 $\alpha:=h(\boldsymbol{u};\ \boldsymbol{\alpha})$ 和 $\beta:=g(\boldsymbol{u};\ \boldsymbol{\beta})$ 的情形。因此，我们的目标是证明

$$\frac{p(y\mid\boldsymbol{x},\delta=1;\boldsymbol{\gamma})}{\int p(y\mid\boldsymbol{x},\delta=1;\boldsymbol{\gamma})\Psi\{\alpha+\beta m(y)\}^{-1}dy}=\frac{p(y\mid\boldsymbol{x},\delta=1;\boldsymbol{\gamma'})}{\int p(y\mid\boldsymbol{x},\delta=1;\boldsymbol{\gamma'})\Psi\{\alpha'+\beta' m(y)\}^{-1}dy},$$

蕴含 $\alpha=\alpha'$、 $\beta=\beta'$ 和 $\boldsymbol{\gamma}=\boldsymbol{\gamma'}$。对上述等式两边关于 y 积分，可得分母相等。因此，我们有 $p(y\mid\boldsymbol{x},\delta=1;\boldsymbol{\gamma})=p(y\mid\boldsymbol{x},\delta=1;\boldsymbol{\gamma'})$；由 (C4) 可得 $\boldsymbol{\gamma}=\boldsymbol{\gamma'}$。

接下来，我们考虑 β的可识别性。取满足 (C5) 的 $\boldsymbol{z}_1$ 和 $\boldsymbol{z}_2$，我们证明，

$$\int\frac{p(y\mid\boldsymbol{u},\boldsymbol{z}_1,\delta=1;\boldsymbol{\gamma})}{\Psi\{\alpha+\beta m(y)\}}dy=\int\frac{p(y\mid\boldsymbol{u},\boldsymbol{z}_1,\delta=1;\boldsymbol{\gamma})}{\Psi\{\alpha'+\beta' m(y)\}}dy, \tag{A13}$$
$$\int\frac{p(y\mid\boldsymbol{u},\boldsymbol{z}_2,\delta=1;\boldsymbol{\gamma})}{\Psi\{\alpha+\beta m(y)\}}dy=\int\frac{p(y\mid\boldsymbol{u},\boldsymbol{z}_2,\delta=1;\boldsymbol{\gamma})}{\Psi\{\alpha'+\beta' m(y)\}}dy, \tag{A14}$$

意味着 $\beta = \beta'$。由（A13）和（A14）可得

$$\begin{aligned}&\int K(y;\alpha,\alpha',\beta,\beta')p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1;\boldsymbol{\gamma})dy\\&= \int K(y;\alpha,\alpha',\beta,\beta')p(y \mid \boldsymbol{u},\boldsymbol{z}_2,\delta=1;\boldsymbol{\gamma})dy = 0,\end{aligned} \tag{A15}$$

其中 $K(y;\ \alpha,\ \alpha',\ \beta,\ \beta')\ = \Psi^{-1}\{\alpha+\beta m(y)\} - \Psi^{-1}\{\alpha'+\beta' m(y)\}$。接下来需分两步证明 (A15) 蕴含 $\beta = \beta'$：

步骤 I. 我们证明函数 $K(y;\ \alpha,\alpha',\beta,\beta')$ 在$\beta \neq \beta'$时仅发生一次符号变化。假设 $\beta \neq \beta'$。由于函数$m(\cdot)$ 和 $\Psi(\cdot)$ 的单射性，方程 $K(y;\ \alpha,\alpha',\beta,\beta')\ = 0$ 仅有一个满足 $m(y^*)\ = (\alpha-\alpha')/(\beta'-\beta)$ 的解$y^* \in \mathcal{S}_y$。这意味着 $K(y)$ 仅发生一次符号变化。

步骤 II. 我们证明当 $\beta = \beta'$ 时，方程 (A15) 不成立。不失一般性，根据步骤 I，我们考虑以下情形：$K(y;\ \alpha,\alpha',\beta,\beta')\ < 0\,(y < y^*)$，且 $K(y;\ \alpha,\alpha',\beta,\beta')\ > 0\,(y > y^*)$，并且 $p(y \mid \boldsymbol{u},\boldsymbol{z}_2,\delta=1)/p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)$ 单调递增。令 c 为密度比的上界。

$$c := \sup_{y<y^*} \frac{p(y \mid \boldsymbol{u},\boldsymbol{z}_2,\delta=1)}{p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)}.$$

根据 $K(y;\ \alpha,\ \alpha',\ \beta,\ \beta')$ 中所示的性质，我们有A15)，可得

$$\begin{aligned}0 &= \int K(y;\alpha,\alpha',\beta,\beta')p(y \mid \boldsymbol{u},\boldsymbol{z}_2,\delta=1)dy\\&= \int_{-\infty}^{y^*} K(y;\alpha,\alpha',\beta,\beta')\frac{p(y \mid \boldsymbol{u},\boldsymbol{z}_2,\delta=1)}{p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)}p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)dy\\&\quad + \int_{y^*}^{\infty} K(y;\alpha,\alpha',\beta,\beta')\frac{p(y \mid \boldsymbol{u},\boldsymbol{z}_2,\delta=1)}{p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)}p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)dy\\&\geq \int_{-\infty}^{y^*} cK(y;\alpha,\alpha',\beta,\beta')p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)dy + \int_{y^*}^{\infty} cK(y;\alpha,\alpha',\beta,\beta')p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)dy\\&= c\int K(y;\alpha,\alpha',\beta,\beta')p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1)dy = 0,\end{aligned}$$

其中，不等式由 c的定义得出。这导致密度比$p(y \mid \boldsymbol{u},\ \boldsymbol{z}_2,\ \delta=1)/p(y \mid \boldsymbol{u},\ \boldsymbol{z}_1,\ \delta=1)$在 $\mathcal{S}_y$上为常数，因此，$p(y \mid \boldsymbol{u},\ \boldsymbol{z}_2,\ \delta=1)\ = p(y \mid \boldsymbol{u},\ \boldsymbol{z}_1,\ \delta=1)$在 $\mathcal{S}_y$上成立。这与(C5)矛盾，因此 $\beta = \beta'$。

最后，由于 Ψ 的严格单调性，可知积分

$$\int \frac{p(y \mid \boldsymbol{u},\boldsymbol{z}_1,\delta=1;\boldsymbol{\gamma})}{\Psi\{\alpha+\beta m(y)\}}dy,$$

相对于 α是单射。因此，方程（A13）意味着 $\alpha = \alpha'$。

□

P*命题的证明*3.8。由假设（4）可知，存在

M， $C > 0$ 使得

$$\begin{aligned}
&\int \frac{p(y \mid \boldsymbol{x}, \delta = 1; \boldsymbol{\gamma})}{\Psi\{h(\boldsymbol{u}; \boldsymbol{\alpha}) + g(\boldsymbol{u}; \boldsymbol{\beta}) m(y)\}} dy \\
&\propto \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \frac{(y - h(\boldsymbol{u}; \boldsymbol{\alpha}) - \beta\mu(\boldsymbol{x}, \boldsymbol{\kappa}))^2}{\beta^2 \sigma^2}\right\} \frac{1}{\Psi(y) \exp(|y|^s)} \exp(|y|^s) dy \\
&\leq \int_{-\infty}^{-M} \exp\left\{-\frac{1}{2} \frac{(y - h(\boldsymbol{u}; \boldsymbol{\alpha}) - \beta\mu(\boldsymbol{x}, \boldsymbol{\kappa}))^2}{\beta^2 \sigma^2}\right\} C \exp(|y|^s) dy + C < \infty,
\end{aligned}$$

其中 $0 < s < 2$。最后一个方程中的第一项和第二项分别由条件（4）以及对 Ψ的递增假设得出。 □