arXiv:2302.01897v1 [stat.ME] 3 Feb 2023

政治学理论的量化：识别、解释与结构方法的作用

Nathan Canen 和 Kristopher Ramsay

摘要。最优秀的政治学实证研究会明确界定所关注的实质性参数，提出一组能够保证其可识别性的假设，并采用恰当的估计量。我们主张将严谨的理论明确融入这一过程至关重要，并着重阐述这样做的优势。通过将理论结构融入实证策略，研究人员能够量化竞争性机制的效应，考察新政策的事前效应，将研究发现外推至新环境，估计特定于模型的理论参数，评估理论模型的拟合程度，并检验旨在解释同一现象的竞争模型。为介绍这一方法论，我们概述结构估计，包括正式定义、实施建议、实例以及与其他方法的比较。

关键词。定量方法、形式理论、识别、结构方法、反事实、研究设计。

日期：2023年2月6日。**Canen：**休斯敦大学及 NBER 研究经济学家。**Ramsay：**普林斯顿大学。我们感谢 Scott Ashworth、Andy Eggers、Sean Gailmard、Mike Gibilisco、Bobby Gulotty、Gleason Judd、Amanda Kennard、Korhan Kocak、Monika Nalepa、V´ıtor Possebom、Joseph Ruggiero、Kevin Song、Tara Slough、Dustin Tingley、Scott Tyson，以及芝加哥大学形式理论应用研讨会的与会者提出宝贵意见。本文的思想经过多年在不同场合和会议中的多次交流，不断得到完善和塑造。

1. 引言

所有定量实证分析都基于模型。反事实源自模型，参数的识别及其估计值的解释也是如此。事实上，理论假设是连接估计值与其解释的桥梁（Lundberg 等人，2021）。这些假设是否有效，以及选择采用何种设定，必然取决于实证策略与底层理论之间的联系，无论这种联系是由形式化分析还是定性分析所驱动。因此，仔细思考底层模型至关重要。

我们基于这一观察提出一个实证分析框架。由于所有我们关心的实证量都是相对于某个理论模型定义的，因此任何研究者都可以使用这一框架。它并不取决于是否采用“结构式”或“简化式”方法。

我们的框架首先要定义一个实证目标，该目标取决于研究问题和理论兴趣。接着，研究者需说明其基础建模假设和设定。这些要素结合起来，使研究者能够解释其发现，并确保其关注参数的统计识别。最后，研究者提出一个估计量，以便根据样本估计所需参数。

为了说明我们的框架为何有用，请看下面这个例子。假设一位研究者掌握了政党 p 在选区 d 的得票率数据，以及一项选区层面的特征 X_d（例如平均受教育程度、选区女性占比等），并决定进行以下线性回归：

$$\log(vote\ share_{p,d}) = \alpha + \beta X_d + \varepsilon_{p,d}. \tag{1.1}$$

这位研究者应在何种情况下将 β 解释为：（i）结果变量与 X_d之间线性最佳拟合的斜率；（ii） X_d 对 p得票率影响的（半）弹性；（iii）一项提高 X_d的政策所产生的平均处理效应；（iv）具有特征 X_d 的选民对政党 p的偏好；还是（iv）一项反事实政策对 X_d的影响，该政策会提高 X_d？在讨论这项回归的结果时，文献中出现了所有这五种解释（Bartels 和 Zaller，2001；Gerber，1998；Nadeau 和 Lewis-Beck，2001；Hansford 和 Gomez，2010），但每种情况下的量都是相对于不同的基础模型定义的。此外，这些理论量无法与（1.1）的估计值区分开来。然而，在（i）至（v）之间作出的选择可能会带来重大影响：每种解释都可能导向截然不同的理论、政策和福利结论。从这个意义上说，分析从根本上依赖于理论模型。

受此启发，本文第一部分详细阐述了一个分析框架，该框架仔细考量其模型。这个框架只是承认

所有感兴趣的参数都是相对于某个模型定义的。也就是说，不存在非理论的或无模型的统计学。定量分析归根结底都是模型。允许进行统计分析的假设是否有效或有用，以及应采用哪种设定，本质上都取决于理论。

在本文的第二部分，从第 5 节开始，我们将讨论实证方法，供希望识别并估计相对于政治形式理论模型定义的参数的研究者使用。示例包括偏好（如意识形态）、福利衡量指标，或支配行为主体行为的参数（如在决定是否参战或参加抗议时策略替代性的程度）。这种方法通常称为 结构方法，或者较宽泛地称为 结构估计。它以形式化的数学理论为统计模型奠定基础，而该理论以所关注的政治现象为依据。例如，它可以将立法投票的空间模型作为统计模型的基础，用于估计立法者偏好；也可以利用国家间的战争对抗模型，估计民主政体带来的军事回报。

结构方法旨在识别并估计相对于形式模型具有明确解释的参数（见 Keane，2010）。这使其有别于其他方法。然而，与普遍看法相反，结构方法不一定需要大量计算，也不一定是非线性的，尽管它们也可以如此。它们不一定要源自博弈论模型。结构方法与实验之间并不存在内在冲突，无论是实验室实验、随机实验还是准自然实验。事实上，结构研究会利用所有这些类型的数据，正如我们在第 6 节所述。

然而，从受研究主题启发的模型出发，结构式方法至少有四项主要优势。首先，结构参数具有清晰的因果解释和理论解释，这源于其形式化的理论基础。这意味着，对这些参数进行识别和解释所需的假设是明确的。其次，通过明确界定理论领域，有可能恢复那些在其他情况下无法识别（或无法观测）的参数。例如意识形态或偏好，这使研究人员能够回答那些原本无法进行经验量化的问题。[1] 第三，可以将估计出的效应外推到新的情境中，而外推范围由基础理论的适用范围决定，从而能够开展多种反事实实验和分析。最后，模型本身的评估也很重要。分析者可以提出这样的问题：理论模型对数据的拟合度如何？这个模型是否比其他模型表现更好？模型能够很好地解释哪些观察结果，又有哪些解释得不好？

[1]对偏好进行非理论估计的困难在Abramson 等人（2022）关于联合实验的研究中有详细讨论。

模型在样本内预测是否有效？在样本外预测是否有效？

这些益处对作为一门实证与理论学科的政治学具有重大价值。一旦认识到所有定量分析都以模型为基础，理论结构（即由理论驱动的假设）与识别便会成为互补关系；此时也就能明白，致力于实验传统和简化式传统的人士如何与注重实证的理论学者开展富有成效的交流。

其次，为实证分析提供理论结构，会促使研究者关注机制而非效应，从而有助于联系现实世界。以政策为导向的实证研究者往往对结构性研究持怀疑态度，因为后者依赖若干理论假设。人们普遍认为，基于看似最低限度或简单假设得出的结果更有用。这种不安可能源于对形式模型所依据的行为假设和均衡假设的怀疑。然而，如果目标是开展能够为政策提供参考的研究，那么探究因果效应背后的机制就十分重要。由于缺乏与行为或均衡效应的关联，简化式方法在识别改变结果变量的有效途径，以及理解政策变化和可扩展的反事实所带来的后果方面，能力有限（Al-Ubaydli 等人，2017年；Heckman，2008年）。但我们并非对世界如何运作一无所知；理论能够提供洞见，舍弃这些洞见似乎是一种浪费。对于数据匮乏的问题和环境而言，这种浪费尤为可惜。因此，关注政策影响的研究者尤其应当有兴趣采用结构方法。

最后，也许对政治学而言最重要的是，结构方法能够促进累积科学。我们已经看到，简化式研究难以产生累积知识，这在很大程度上是因为我们不了解特定研究结果的外部效度。仅用简化式方法检验一般理论，往往会使我们陷入一系列又一系列相互矛盾的发现：在某处得到肯定结果，在另一处却得到否定结果，如此循环不休；下文将讨论其中一些例子。这并不意味着任何一个简化式结果都是错误的。问题在于我们对外部效度的理解不足。原则上，尽管实践中未必总是如此，以理论为基础的结构模型有助于积累知识，因为它们旨在具有外部效度。这样，许多学者便能够构建模型、比较不同模型，并对其加以修改或整合，从而具体探究政治中的某些特征，而这些特征是简化式研究因其构建方式所限而无法发现的。

得益于这些优势，结构方法在政治学和政治经济学中的地位日益突出，并在各个领域得到广泛应用。例如，它已广泛用于美国政治领域，DW-Nominate（Poole and Rosenthal，1984；Bonica 等人，2013）便是一个突出的例子；在比较政治学领域，它被用于研究学习民主的作用（Abramson and Montero，2020）、选民在不同情境下对信息的运用（例如，Kendall etal.，2015；Cruz et al.，2020），以及议会民主制中的联盟形成（例如，Diermeier et al.，2003）；在国际关系领域，它被用于研究冲突与内战（例如，Signorino，1999；Crisman-Cox andGibilisco，2018；Kenkel andRamsay，2021）等。

然而，许多人发现结构方法难以实施和评估。这是因为统计程序可能是量身定制的，而且分析者往往关注的是评估整个模型，或仅仅量化理论参数，而不是检验某个参数是否显著。这意味着，与其说是 t 检验，不如说是更类似于机器学习中常见方法的工具——例如非嵌套模型检验、似然比检验、联合显著性检验和交叉验证——往往更有用。此外，标准误描述的是某些量的不确定性，而非用于拒绝一个从未提出过的原假设的标准。第6节列举了学者们运用统计技术估计模型参数的多种方式，并提出研究者、编辑和审稿人可以如何评估论文中此类方法的应用是否恰当。下文还将举出政治学和政治经济学现有研究中的许多实例。

2. 无处不在的模型

我们首先从一个多少有些显而易见的观察出发：所有定量实证分析都始于一个模型。我们也可以将这些称为理论、理论模型，或理论的模型。无论如何，研究都以一组未经检验的假设为前提，这些假设涉及世界如何生成数据，以及哪些内容会被观察到。为明确说明这一点，请看以下常见的统计模型。

2.1. 线性模型

在经典线性回归模型中，研究者假设结果变量 $Y_i \in \mathbb{R}$ 与外部变量（例如协变量）$X_i \in \mathbb{R}^n$ 之间存在线性关系，如下：

$$Y_i = X_i'\beta_0 + \varepsilon_i, \tag{2.1}$$

其中，ε_i 是研究者无法观测到的随机变量，$\beta \in \mathbb{R}^n$ 是参数。此外，假设 $\mathbb{E}X_iX_i'$ 可逆，并且当（Y_i，X_i）为独立同分布时，$\mathbb{E}[\varepsilon_i \mid X_i] = 0$。

线性设定与关于不可观测变量分布的假设，以及 X_i 的总体对应量共同构成了模型。研究者假设该模型成立，随后旨在估计参数的取值，记为 β_0，它生成了观测数据(Y_i, X_i)。因此，统计模型由2.1及其假设定义，而研究者希望从数据中了解关注参数 β_0 。

2.2. Logit

在Logit模型中，研究者观测到一个二元结果，$Y_i \in \{0,1\}$ 和 $X_i \in \mathbb{R}^n$，同时假设：

$$Y_i = 1\{X_i'\gamma_0 + \varepsilon_i \geq 0\}, \tag{2.2}$$

其中，$1\{.\}$ 表示一个指示变量（即，若括号中的条件成立，则取值为1；否则取值为0）。

同样，该模型完全由关于（X_i，ε_i）总体对应量的假设所定义。更确切地说，ε_i假设服从（标准）逻辑斯蒂分布，其累积分布函数为 $\Lambda(\cdot)$，因此 $P(Y_i = 1 \mid X_i) = \Lambda(X_i'\gamma)$，其中 $\mathbb{E}X_iX_i'$ 可逆。

研究者可能希望了解生成这些数据的 γ_0 值。模型对 γ的许多取值都有定义，但关注参数 γ_0是其中一个特定取值，它会产生(Y_i, X_i)的观测分布。

2.3. 潜在结果框架中的平均处理效应

在潜在结果框架中，研究人员考虑这样一种情形：某种处理（记作 $D_i \in \{0,1\}$）施加于目标单位，且每个单位都有两个潜在结果，

$$Y_{Di} = \begin{cases} Y_{1i} \text{ if i is treated} \\ Y_{0i} \text{ if i is untreated.} \end{cases} \tag{2.3}$$

D_i 对 i 的因果效应为

$$\tau_i = Y_{1i} - Y_{0i}. \tag{2.4}$$

模型的定义满足以下条件：$D_iY_{1i} - (1 - D_i)Y_{0i}$ 可观测，处理分配独立于潜在处理效应 $(Y_{1i} - Y_{0i} \perp D_i)$ (例如通过随机化实现)，且不存在溢出效应。(Cunningham，2021，第140页)。

例如，研究者可能会对平均处理效应（ATE）感兴趣，$\tau_{ATE} = \mathbb{E}\tau_i$。

2.4. 立法机构中的意识形态投票

考虑政治学中一类重要的模型：多维意识形态投票。研究者观察一系列唱名表决 $t = 1, \ldots\ldots T$后，希望测量政治人物的意识形态。每位政治人物 i 具有意识形态 $\gamma_i \in \mathbb{R}^n$，并根据替代政策$x_t$ 是否能带来高于现状政策（q_t）的效用，在每次唱名表决中选择投“赞成”（$Y_{i,\ t} = 1$）或“反对”（$Y_{i,\ t} = 0$）。

政治人物 i 对政策 x_t 的偏好被建模为随机效用，由确定性部分 $U(x_t,\ \gamma_i)$ 和随机部分 $\varepsilon_{i,\ x,\ t}$构成。因此，政治人物 i 在以下条件下会在唱名表决 t 中投“赞成”：

$$U(x_t, \gamma_i) + \varepsilon_{i,x,t} \geq U(q_t, \gamma_i) + \varepsilon_{i,q,t}. \tag{2.5}$$

如果 $U(k_t, \gamma_i)$ 是关于 $k_t \in \{q_t, x_t\}$ 和 ε_i 的二次函数 $\|k_t - \gamma_i\|^2$，且 $\cdot$ 服从正态分布，那么这就是 Clinton 等（2004）的模型（不含政党效应）、Heckman 和 Snyder（1997）、Rivers（2003）等人提出的模型。如果 $U(k_t, \gamma_i)$ 是高斯函数，且 $\varepsilon_{i,\cdot,t}$ 服从正态分布，那么这就是 DW-Nominate（Carroll 等，2009；Boche 等，2018）。

这四个例子体现了任何实证研究的本质。研究者的目标是根据随机向量（此处为$(Y_i$、X_i、$D_i)$）的实现值，了解潜在参数（β_0、γ_0、$\{\gamma_i\}_{i=1}^n$ 或 τ_{ATE}）的取值，同时以一个模型为依据（即一项设定和一系列假设，用以刻画数据中变量之间可能关系的一个“类别”）。

从最一般的形式来看，实证（统计）模型是一组关于可观测变量的概率分布，这些分布根据关于世界如何生成数据的理论推导而来；所有可能的模型都由参数值索引，且真实分布属于这一类。只有在定义模型之后，统计识别才有意义。它正是统计（理论）命题所预示的逻辑结论，而模型就是该命题的假设。

3. 理论模型与识别

术语识别在实证研究中有多种用法（Lewbel，2019）。然而，正式定义只有一种，它与定义良好的模型的最低要求有关。

形式上，模型是一个三元组（Θ，γ，$\mathbb{P}_X$），其中 Θ 是不可观测参数的集合，γ 是一个映射 $\gamma:\ \theta \to \mathbb{P}_X$，其中 $\mathbb{P}_X$ 是关于可观测随机变量的所有联合分布的集合 X。

定义 3.1。模型 $(\Theta,\ \gamma,\ \mathbb{P}_X)$ 当且仅当对于每个 $(\theta,\ \tilde{\theta}) \in \Theta^2,\ \gamma(\theta) = \gamma(\theta)$ 当且仅当 $\theta = \theta$ (Athey and Haile，2002)。

因此，θ 相对于模型而言是可识别的，即存在一个唯一的值能够合理解释数据分布 $P_X \in \mathbb{P}_X$。[2]

在上面的线性示例中，这要求 β (即 β_0) 只有一个取值能使方程 (2.1) 中的模型生成可观测变量 $(Y_i,\ X_i)$的分布。

3.1. 线性模型中的识别

考虑第 2.1 节中的模型。研究者希望了解真实的 β_0。根据模型所述的假设，我们可以写出：

$$\beta_0 = \left(\mathbb{E}X_iX_i'\right)^{-1}\mathbb{E}X_iY_i. \tag{3.1}$$

如我们所见，(3.1) 的左边是关注参数，右边则是可观测数据分布（即 (Y_i, X_i)的分布）的已知函数。给定该模型，如果研究者知道可观测变量的分布（包括 $\mathbb{E}X_iY_i, \mathbb{E}X_iX_i'$），就能确定 β_0 的确切值。也就是说，没有其他β 值能生成该数据分布。因此，β_0 是可识别的。

请注意，如果 $\mathbb{E}X_iX_i'$ 不可逆（即 X_i 存在多重共线性），那么 β_0 将不可识别。事实上，β 可能有多个取值都能生成相同的数据。[3] 请注意，多重共线性是数据生成总体模型的问题，而非样本的问题。样本在识别中不起作用；参数是否可识别，取决于数据生成过程理论的性质，以及在假设的无限样本中观察到的内容。

[2]还有其他类型的识别，例如部分识别或集合识别，其思路相似，但不要求该关系以等式成立。关于实证分析中使用的其他识别形式的综述，参见 Lewbel (2019。[3]例如，如果 $Y_i = \alpha + \beta_0 X_i + \varepsilon_i$，其中 $Var(X_i) = 0$，那么该模型将产生与 (Y, X) 相同的联合分布，$Y_i = \tilde{\alpha} + \varepsilon_i$其中 $\tilde{\alpha} = \alpha + \beta_0\mathbb{E}X_i$。

3.2. Logit

我们现在重新讨论第2.2节。在给定的假设下，

$$\begin{aligned} \mathbb{E}[Y_i \mid X_i] &= \Lambda(X_i'\gamma_0) \\ \Lambda^{-1}(\mathbb{E}[Y_i \mid X_i]) &= X_i'\gamma_0 \\ \mathbb{E}X_iX_i'\gamma_0 &= \mathbb{E}X_i\Lambda^{-1}\mathbb{E}[Y_i \mid X_i] \\ \gamma_0 &= (\mathbb{E}X_iX_i')^{-1}\mathbb{E}X_i\Lambda^{-1}(\mathbb{E}[Y_i \mid X_i]). \end{aligned} \tag{3.2}$$

同样，右侧是 $(Y_i,\ X_i)$ 的矩的已知函数，因此 γ_0 同样是可识别的。

相比之下，假设 ε_i 服从逻辑斯蒂分布，但尺度参数为 $\sigma \neq 1$。那么，模型 $Y_i = 1\{X_i'\gamma_0/\sigma > \varepsilon_i/\sigma\}$ 将生成与上文相同的 $(Y_i,\ X_i)$ 分布。因此，观察 $(Y_i,\ X_i)$ 分布的研究者无法知道这些数据来自第一个模型（参数为 γ_0），还是第二个模型（参数为 γ_0/σ）。因此，如果 ε_i 的分布未知，γ_0 就无法被识别。

3.3. 潜在结果与平均处理效应

在潜在结果框架中，我们可以将观测到的 Y_i

$$Y_i = Y_{i0} + (Y_{i1} - Y_{i0})D_i + \sum_{j \neq i} \rho_{ji} D_j, \tag{3.3}$$

其中，前两项定义了来自 i 的观测结果变量，第三项则是处理对其他 j 个观测值的溢出效应。假设不存在溢出效应，即对于 j 和 i，有 $\rho_{ji} = 0$，则我们有

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i0}|D_i = 1] + \mathbb{E}[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0].$$

后一个项是选择效应，当 $Y_{1i} - Y_{0i} \perp D_i$ 时，该效应为零。同质性和期望算子的线性告诉我们

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{i1} - Y_{i0}|D_i = 1] = \mathbb{E}[\tau_i].$$

如果无溢出效应假设不成立，那么就会有许多（$\rho,\ \tau$）组合产生相同的观测差异 Y_i，而该模型将无法识别。

3.4. 意识形态的识别

请参阅Rivers (2003，其中对一般维度的偏好识别进行了细致讨论并给出了证明 n 以及二次偏好 $U(\cdot)$，相关内容见

第2.4节。虽然可以运用与前几节类似的直觉，但由于存在非线性以及参数数量众多，证明更加微妙（例如，n 每位政客对应一个，另有 x_t，q_t 每次唱名表决对应一个）。

与此同时，Canen 等人（2022）对一维 DW-Nominate 中的识别进行了细致讨论，并探讨了二维 DW-Nominate 中识别面临的挑战。（参见该文的附录 B。）

3.5. 关于识别的重要信息

在讨论统计识别这一概念时，值得关注几个要点。

首先，识别总是相对于某个模型而言的。也就是说，不可能存在一种不依赖理论或仅凭数据就成立的识别论证。这是因为模型使我们能够定义感兴趣的参数，并要求我们明确说明其背后的假设。

事实上，如果改变理论建模假设，就无法保证关注参数仍然可识别。例如，假设我们修改第 2.1 节中的方程（2.1 ），在右侧加入一个新的随机变量。那么，原来的 β_0 可能无法识别。同样，如果我们将外生性假设改为 β_0 ，那么 $\mathbb{E}[\varepsilon_i \mid X_i] \neq 0$也可能无法识别。无法保证某个参数在不同模型中仍然可识别。

其次，识别始终关乎总体对应量，而绝不关乎样本。假设研究者知道了她所观测变量中所有可能想了解的信息（例如第 $(Y_i,\ X_i)$ 节 3.1-3.2中提到的完整分布）。用上面的符号表示，她知道 P_X ，因此也知道可观测数据的所有矩。识别要探究的是：研究者能否根据这个已知分布推断出参数的真实值？换句话说，她能否唯一确定究竟是哪些参数生成了我们观测到的数据？如果不能，那么这个模型就是“定义不明确”的：即使在理想世界中，你也无法知道究竟是哪些参数生成了这些数据。

当然，我们并非生活在这样的理想世界中。我们观测不到 P_X，只能观测到 $(Y_i,\ X_i)$ 的一个样本。然而，在考虑这样的样本之前，必须先讨论识别：如果即使知道真实总体，我们都无法找出参数的真实值，那么仅凭一个样本，又怎能指望估计出接近 θ_0 的值呢？

最后，在任何实证研究中，研究者的目标都是识别其关注参数。在许多模型中，有些参数可能可识别，而另一些则不可识别。在这种情况下，研究者只需确保其关注参数可识别。

如果有些参数对于回答研究问题毫无用处，那么并非所有参数都必须可识别。[4]

3.6. 估计

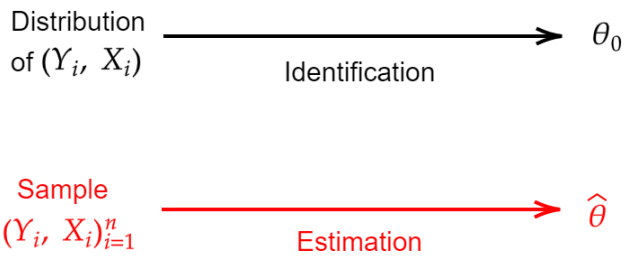
估计是从样本中量化理论参数的次级过程。参数 θ 的估计量 θ_0 是观测样本的一个函数。在上面的线性示例中，估计量是样本（Y_i，X_i）$_{i=1}^{n}$的任意函数。仅此而已。线性模型中常见的一种估计量是普通最小二乘估计量（OLS），但它并非唯一选择。例如，可以使用极大似然估计量（MLE），也可以使用 X_i的第一个观测值作为估计量。不同的估计量会因其假设而表现得“更好”或“更差”。OLS常因其在线性模型中的优良性质而被使用，但这些性质在其他模型中并不成立。

我们对这一定义中的两个重要方面作些说明。第一，估计量总是相对于某个参数定义的。否则，估计量要量化什么呢？第二，估计量基于样本，而非总体。它是实现数据的函数，而非识别的“理想世界”。因此，估计只能在识别之后进行。毕竟，只有在我们知道 θ_0 具有统计识别性之后，估计它才有意义。否则（也就是说，即使我们掌握了所有希望了解的可观测数据，也无法确定 θ_0 ），估计量就不可能对 θ_0作出良好的近似。图 1 展示了这些概念。

作者在论述中常常混淆识别与估计。[5] 这通常不会影响对实证结果的解释，但可能会让人困惑：哪些假设意味着模型和可观测总体矩足以计算某个参数，哪些条件则是样本能够为该参数提供可靠估计所必需的。

[4]一个例子是，当误差项的方差不同于 1 时，Probit模型并不可识别。然而，平均偏效应 $APE = \int \frac{\partial \mathbb{E}[Y_i|X_i=x_i]}{\partial x_i} dF_X$，其中 F_X表示 X_i的边际分布，是可识别的。参见 Marshak(1953)，以及 Heckman(2010) 的讨论。[5]例如，参见 Burke等人 (2015，第581页)。一种常见的此类错误是说“估计一个普通最小二乘法模型”。这混淆了正在估计的估计量（普通最小二乘法）、模型（线性）及其目标参数（β_0 第2.1节中的参数）。

图1. 识别与估计



注：模型规定参数如何生成（Y_i，X_i）的分布。识别所要探究的是，我们如何从（Y_i，X_i）的分布中得知 θ_0。识别关乎总体特征。当我们有了样本（即（Y_i，X_i）分布的实现值）时，就会尝试估计 θ_0。

4. 理论、解释与外推

（或者，《无理论推断的局限》）

前面的章节表明，实证研究的技术要素需要模型。[6] 这与一种广泛流行的观点相冲突：认为“让数据自己说话”，或者可以进行“数据驱动”的分析。人们很容易认为自己的研究不需要假设，在解释结果时也不受理论约束。遗憾的是，事实并非如此。理论模型无论是什么样的，除了为定量实证分析提供技术基础之外，还赋予分析以意义和适用范围。

4.1. 因果关系是一项理论主张

关于理论模型深层作用的早期讨论，至少可以追溯到 Koopmans （1947）提出的一点：没有理论，甚至连变量都无法测量。这一点也适用于因果主张。

“可信度革命”对社会科学的实证研究产生了巨大影响。尤其是，它促使人们更加重视统计识别，也强调了实证研究必须谨慎。然而，它广受欢迎也带来了一个副作用：人们热衷于寻找“因果”效应，却很少关注估计目标的含义（Lundberg 等人，2021）。因果关系，无论是从均衡模型还是 Rubin 因果模型中推导出来，都只有相对于理论模型才存在。

例如，回到Logit模型和方程（3.2），

$$\gamma_0 = (\mathbb{E}X_iX_i')^{-1}\mathbb{E}X_i\Lambda^{-1}(\mathbb{E}[Y_i|X_i]) \tag{4.1}$$

[6]本节标题取自 Wolpin（2013）一书，我们从中汲取了灵感。

其中， $\Lambda(\cdot)$是累积分布函数，且 X_i 和 Y_i 是随机变量。

这是一个将数字映射为数字的公式，本身并无内在含义。如上所示，它可以从一个具有逻辑斯蒂分布函数的潜变量模型推导出来。在这一语境下， Y_i 是一个指示变量，当 $X_i'\gamma_0 \geq \varepsilon_i$ 时取值为1，否则取值为0。γ_0 等于一个系数，该系数描述 x 变化一个单位会导致结果变量的对数优势发生变化。

不过，至少还有两种不同的方式可以得出这个公式。假设一位研究者对个体选择感兴趣，并考虑这样一个决策问题：一个行为主体必须在两个备选项之间作出选择。自然的第一步是考虑一种决策理论，其中（i）行为主体选择某个备选项的概率随该备选项的价值单调变化；（ii）满足这样一项原则：新增备选项不会改变在两个给定备选项之间作出选择的相对可能性。后者是无关选项统计独立性的一种形式。Luce （1959）表明，这种决策理论与Logit模型之间存在公理化的理论联系。具体而言，如果备选项的价值由 $v（x） = u（x） + \varepsilon$给出，且 ε 服从I型极值分布，那么，这种通过理论推导得出的选择规则与概率效用最大化相一致，并且等价于Logit。这里， x是属性特征，而 γ_0 是它们对备选项价值评估的影响。也就是说，该参数表示 x 对决策者选择备选项时效用的影响。

在围绕有争议商品展开的策略竞赛中，Kenkel 和 Ramsay (2021的分析表明，标准 Tullock 模型采用对努力边际收益和边际成本的自然参数化后，会生成与 Logit 概率模型等价的均衡策略。也就是说，他们的策略性理论框架蕴含着同一个方程。在这种情况下， γ_0 是观测因素带来的努力边际收益。

这个例子与引言中的第一个例子一样，意在说明：每当我们想要走过从总体参数到意义的这段路径时，理论模型就会发挥另一项重要作用；Wolpin (2013对此作了细致阐述。理论模型可以是统计模型、Rubin 因果模型，也可以是博弈均衡。无论哪种情况，γ_0 的含义都源自基础理论框架的应用。

理解结构方法的一个准确方式，是将其目标视为依据实质性的理论模型来识别和估计模型特定参数。这可能包括偏好（例如意识形态）、福利衡量指标、支配行为主体行为的参数（例如在是否发动战争的决策中策略替代性的程度），

或者是依赖于明确定义的模型的政策评估（例如实施一项此前从未实施过的政策）。

这不仅仅是一个哲学观点。在政治学中，一个突出的例子是关于极端事件（例如鲨鱼袭击或大学橄榄球比赛结果）是否会影响选民行为的讨论（Achen 和 Bartels，2017年）。有时，正向的研究发现会被解读为选民非理性的证据——也就是说，面对独立事件，选民为什么要改变对政客的看法？[7] 不过，我们可以借助理论退一步，为我们解读这场争论提供指引；Fowler 和 Montagnes （2022）采用的正是这种方法。具体而言，极端事件会影响选民行为这一现象完全可能与贝叶斯学习相一致（例如，极端事件让选民了解到政客在相关事件中的行为）——参见Ashworth 等人（2018）。因此，脱离理论的线性模型估计值不足以告诉我们究竟识别出了什么。相反，研究者必须说明，假设理性选民也能从鲨鱼袭击中了解现任者的类型，是否合理。

4.2. 外部效度与外推属于理论问题

当前统计学研究也关注许多统计分析所面临的外部效度与外推问题（Findley 等人，2021；Egami 和 Hartman，2022；Hartman，2021）。明确运用实质性理论可以解决这两个问题。对于外部效度与外推，研究人员想知道的是，研究结果自然适用于哪些总体？

理论模型的构建旨在明确表征一类定义清晰的事件。例如，关于发达民主国家选举的结构模型适用于发达民主国家，关于国家间战争的结构模型适用于国家间战争，而关于国会投票的结构模型则适用于国会投票。结构模型的估计值也是其基础理论的深层参数。这使研究者能够将这些参数值用于相关但全新的环境中。当估计的量是理论框架中的关键基础要素时，外推便有了意义。

5. 结构方法：入门

前文各节阐述了开展并解读任何实证分析所需的内容。这些论点或主张都不限于结构式研究。现在我们

[7]参见 Graham 等人(2022) 和 Fowler 和 Montagnes(2022)，了解近期研究成果及这场辩论的概况。

概述结构方法本身。为此，有必要重新审视结构分析在形式上的（历史）定义，并将其与当下“结构式”和“简化式”这两个术语的用法联系起来。

5.1. “结构式”与“简化式”方法：一种错误的二分法

结构方法往往与简化式方法相对照。后者如今与双重差分（DiD）和回归不连续设计（RDD）等研究设计联系在一起，而前者则更多见于采用形式模型的论文中。[8] 这导致形成了泾渭分明的阵营，研究人员往往专攻其中一种方法，鲜有交叉。有时，这种区分还与对立情绪相关（参见Angristand Pischke（2010）），因为人们认为二者势不两立。

然而，无论从定义还是历史上的用法来看，这种区分都是错误的。为了考察结构方法，首先有必要界定与之相对应的“简化式”方法。以下这个国际关系领域中理查森军备模型的简单示例，最能清楚地说明其定义（Richardson，2012年；DunneandSmith，2007）。

假设有两个行为主体 $i = 1, 2$，他们选择要保障多少武器 $Y_i \in \mathbb{R}_+$。此外，假设每个行为主体都有一个特征 $X_i \in \mathbb{R}$（为简单起见，设为单维）。理查森模型的一种版本提出了以下关于共同军备决策的模型：

$$\begin{aligned} Y_1 &= \alpha_1 + \beta_1 Y_2 + \gamma_1 X_1 + \varepsilon_1 \\ Y_2 &= \alpha_2 + \beta_2 Y_1 + \gamma_2 X_2 + \varepsilon_2, \end{aligned} \tag{5.1}$$

因此，$\Theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2)'$ 是参数，而 $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ 是均值为零、且与 $X = (X_1, X_2)'$ 相互独立的不可观测变量。经验研究者观察到 $(Y_i, X_i)_{i=1,2}$。

方程（5.1）及其相关假设称为结构模型。这是因为它们概述了研究者提出的原始模型，清楚地说明这些关系如何取决于基本的，或称结构参数，Θ。

不过，可以通过对方程组求逆，将同一个模型（5.1）改写，使所有因变量 $Y = (Y_1, Y_2)'$ 都位于表达式左侧。这样

[8] 然而，这类定义并不精确。随机对照试验、回归不连续设计、双重差分等研究设计也需要定义一个理论模型（我们为什么要进行改变 X？），而这会引入更多假设（例如，对 X_i 进行随机分配，或满足平行趋势）。随后，每个模型都可以识别一个关注参数（例如，政策平均效应），并使用特定的估计量进行估计（例如普通最小二乘法、非参数方法等）。

可得：

$$
\begin{aligned}
Y_1 &= \frac{\alpha_1 + \beta_1\alpha_2}{1-\beta_1\beta_2} + \frac{\gamma_1}{1-\beta_1\beta_2}X_1 + \frac{\beta_1\gamma_2}{1-\beta_1\beta_2}X_2 + \frac{\varepsilon_1 + \beta_1\varepsilon_2}{1-\beta_1\beta_2} \\
Y_2 &= \frac{\alpha_2 + \beta_2\alpha_1}{1-\beta_1\beta_2} + \frac{\beta_2\gamma_1}{1-\beta_1\beta_2}X_1 + \frac{\gamma_2}{1-\beta_1\beta_2}X_2 + \frac{\varepsilon_2 + \beta_2\varepsilon_1}{1-\beta_1\beta_2},
\end{aligned} \tag{5.2}
$$

在保持关于 ε 和 X的相同假设下。方程（5.2）称为简化式方程，因为它们将结构模型化简了。事实上，“输出”变量现在位于等式左侧，而“输入”变量则都位于右侧。因此，第一条方程中 $\frac{\gamma_1}{1-\beta_1\beta_2}$ 对应 X_1 的系数，就是 X_1 对 Y_1的简化式效应。[9]这是我们所知唯一对“简化式”的~~正式~~定义，至少可追溯至20世纪40年代考尔斯委员会的研究。

请注意，关注 X_1 对 Y_1 的效应的研究者，无需识别所有 Θ：她只需识别简化式参数 $\frac{\gamma_1}{1-\beta_1\beta_2}$，而不必将 γ_1 与 β_1 和 β_2 分别识别出来，正如（~~5.1）中~~所设定的那样。同时，其因果解释并非来自简化式表达式，而是来自模型本身。

因此，视研究者关注的问题而定，他们可以只依赖模型的简化式，而非其结构对应形式。这正是如今这种术语用法的由来。不过，对于其他问题，研究者可能会关注（β_1, β_2）的具体取值，例如，这就需要证明这些参数能够分别识别。

(5.1) 与 (5.2) 的根本区别在于，前者是理论的数据生成过程，后者则是一种遵循理论结构的记账方式，但将内生变量排列在左侧，外生变量排列在右侧。将简化式写作 (5.2)，并不取决于研究者采用结构式方法还是简化式方法。相反，在 (5.1) 与 (5.2) 之间作出选择，取决于是否要展示模型中外部变量与内部变量之间的依赖关系，以及识别参数 Θ 时可能面临的挑战。由于主体 1 和 2 相互作用，彼此的特征会影响各自的选择，进而影响对手的选择。

正如我们所看到的，结构式和简化式的传统定义之间存在着深刻的联系。更精确的区别在于，结构性研究旨在识别并估计该模型的基本参数——即（整个或部分）向量 Θ。在这种情况下，作者必须完整地设定模型，其中包括（5.1）以及进一步的假设（例如，随机向量 ε 的均值为零）。随后，作者在该模型下证明这些参数的识别性，并可以进一步讨论估计量。与此同时，对总效应感兴趣的作者

[9]按照 Heckman (2010的术语， Y 可以表示为“外部”变量的函数，即 $(X,\ \varepsilon)$。当 $(X,\ \varepsilon)$ 相互独立时，我们可以说等式左侧是因变量，右侧是自变量。

只需识别并估计 X_1 对 Y_1 的参数，该参数由 γ_1、 β_1和 β_2构成。但在此过程中，他们仍然假设原有的统计（以及形式）模型，只是从数据中提取不同的信息。

5.2.“结构式”一词的现代用法

结合这个例子来看，结构估计这一术语用于描述研究者希望识别并估计模型特定参数（即相对于一个清晰阐述的理论模型及其相关假设所定义的参数）。当然，这一做法是由他们的研究问题所驱动的。[10]

正如我们所见，结构方法的定义并未要求使用非线性模型或计算密集型方法；有些通过结构方法估计的模型是线性的，可以用简单方法（例如普通最小二乘法（OLS））进行估计（Heckman and Snyder，1997；Smith，1989）。此外，实验设计或准实验设计也可以帮助识别结构模型中的参数。例如，只要谨慎地将随机对照试验（RCT）纳入形式模型，就可以用它来识别选民的信念和偏好（例如，Kendall et al.，2015；Cruz et al.，2020）。关于这一主题的深入讨论，参见Toddand Wolpin（2020）。结构方法也可以采用断点和准实验方法。例如，Martin and Yurukoglu （2017年）明确利用福克斯新闻频道的随机排列顺序，帮助识别对志趣相投的新闻的偏好。

因此，Keane (2010) 写道，“简化式方法与结构式方法的区别不在于假设的数量（如上面的模型所示），而在于它们对这些假设的明确程度以及各自的目标。”

结构式论文往往旨在识别不同于非结构式论文的参数（如意识形态、偏好、生产率等，而非政策的平均效应），并依赖于对研究者用来解释其估计结果的理论环境进行形式化描述。因此，相较于政治学实证研究的大多数工作，结构模型要求研究人员明确提出其设想中的形式理论模型并概述假设，清晰界定目标参数，说明如何将其映射到数据，并解释只要原始理论模型成立，其参数如何在反事实情境和新环境中保持原有解释。

这样一来，研究人员便可以利用其估计来回答一些重要问题；若非如此，这些问题将无法得到解答。我们将在下一节讨论这一点。

[10]有些研究者认为，“结构式估计”这一术语并不恰当，因为正如我们上文所见，任何估计只有在先设定某种（结构）模型之后才有意义。不过，我们仍采用最符合该术语在文献中用法的定义。

5.3 感兴趣的问题类型

研究者是否应该使用结构方法？这从根本上取决于所关注的问题。下面，我们将阐述我们对哪些类型的研究问题最适合采用结构方法的看法。在我们看来，这类问题分为三种。

第一类是旨在量化理论参数的研究问题（即只有相对于某个理论模型才有定义的参数）。例如，DW-Nominate 及其他用于估计国会议员意识形态的指标（Poole and Rosenthal，1984；Bonica 等人，2013；Heckman和Snyder，1997；Canen等，2020、2022，以及许多其他研究），或受众成本（Crisman-Cox and Gibilisco，2018；Kurizaki and Whang，2015）。意识形态即偏好，只有相对于一个界定偏好和个人行为的模型（在 DW-Nominate 中，即对法案的投票）才有意义；而受众成本只有在无法观测时才具有重要意义。这类问题还包括衡量政治抗议活动的策略互补性或替代性（例如 Cantoni 等人，2019）、政客对连任顾虑的重视程度（Iaryczower 等人，2022）、沿网络提供公共品时溢出效应的程度（Acemoglu 等人（2015））、根据汇总数据估计选民偏好（Iaryczower 等人，2022；Rekkas，2007；Ujhelyi 等人，2021）、志趣相投的新闻偏好在推动媒体倾向方面的作用（Gentzkow and Shapiro（2010））、取决于模型的福利效应（例如消费者剩余），等等。在每种情况下，目标都是以只能借助结构化方法实现的方式，估计一个理论上重要的参数。[11]

第二类问题是反事实情景或假设性的政策干预，它们明确要求将结果外推至新环境。这类问题包括将政策外推至可观测数据之外（例如，哥伦比亚或巴西的竞选捐款限额——见Gulzar 等人，2021；Avis 等人，2022 ——是否会在其他地方产生类似效果）。外推需要一些信息，让研究者相信其结果在其他地方同样成立。研究者认为适用于这两种情境的理论模型，正是发挥这一作用的工具。

同样，结构方法尤其适合评估从未实施过或未被观察到的政策。例如，如果数据中从未出现过最低工资上调，就无法评估最低工资上调对少数群体投票行为的效应。若没有形式模型，就难以估计此类政策的效应，因为数据并不涵盖这种情况。相反，只有借助与之紧密结合的理论，才能评估此类政策：

[11]此外，还可以运用理论来帮助衡量难以观测的行为，例如美国国会中立法者之间的联系，如Battaglini 等人 (2021)；Canen 等人 (2022)，或者游说中的竞争如何影响政策批准（例如，Kang(2016）以及政府合同（Cox(2022)。

是外推其他已观察到的上调（例如，基于不同的参数化），还是在策略环境中重新计算均衡（在反事实情境下，简单外推现有数据可能无法得到均衡）。[12]

这类练习可以在Iaryczower 和 Shum(2012年) 中找到。在对美国最高法院的分析中，他们利用反事实模拟，将现存法院的表现与一个反事实法院进行比较；在后者中，要作出对被告不利的裁决，必须获得全体大法官的一致同意。改变规则所带来的后果并不显而易见，因为大法官掌握的信息性质以及他们的政治偏见，都会影响其投票和法院集体决策的质量。我们知道，在由偏好和能力各异的成员组成的集体决策环境中，多数规则并不总是优于一致同意规则。然而，美国最高法院的结果表明，与简单多数规则相比，一致同意规则会导致更高的错误概率。只有借助结构模型，才能进行此类分析。

第四，结构技术对于区分针对同一类事件的竞争性理论或模型非常有用。在政治学的许多领域，我们对同一现象都有若干模型。结构估计可视为统计模型校准的一种形式；让模型与数据相互作用，能够促使研究者提出有关模型拟合度的问题，并比较不同模型的表现。这类研究的一个典型例子是 Francois 等人 (2015)。在那篇文章中，对权力分享模型进行量化，使其能够与若干替代理论直接比较，例如权力集中的“大人物”理论，并对各种理论与数据的拟合程度作出定量判断。另一个例子见于 Kang 和 Miller (2022)；作者能够估计并比较以下因素在解释政府合同竞争为何如此有限方面所起的作用：(i) 信息，(ii) 卖方特征，(iii) 行政障碍，以及 (iv) 腐败。

[12]Heckman（Heckman(2008年）对政策评估问题的分类很好地概括了这类观点，而这些问题贯穿于实证研究之中。他将这类问题分为三大类。第一类是评估历史干预对可观测结果影响的问题。对于这类问题，现有的简化式方法或许就足够了，因为可能无需在统计模型本身之外另行定义形式模型（即，为识别处理效应所需的设定和模型假设）。第二类问题关注在一种环境中实施的干预能否推广到另一种环境。显然，这类问题需要比第一类问题更多的理论假设。具体而言，必须明确一种环境中的理论模型有哪些部分在另一种环境中仍保持不变。最后一类问题是预测数据中从未观测到的干预。这通常需要一个更全面、更细致的理论模型。

6. 从零开始构建结构模型

诚然，前文以非常抽象的方式讨论了“结构方法”。为此，我们将概述我们认为可以遵循的步骤，以结构化方式构建并估计模型。在此过程中，我们谨提醒读者，是否采用结构方法，首先取决于其研究问题。

6.1. 第一步：研究问题和关注参数是什么?

是否有必要构建一个完全设定的结构模型，取决于我们感兴趣的对象。因此，构建任何结构模型的第一步，都是谨慎界定目标参数。

例如，我们可能会问，是什么决定了联盟在战争中的效能？这个问题意味着行动者的特征与战争结果之间存在某种关系。具体而言，哪些特征会影响作战力量的效能？哪些属性会使战争作战变得更加昂贵？一个国家可以期待其伙伴和自身付出多少努力?

请注意，目标参数的定义比“政策 Z 的效应”或“因果效应”更为具体。符合这些描述的参数有无数个。我们鼓励研究者具体说明：我们研究的是政策 Z 对哪个总体的影响?我们关心的是平均效应、对最贫困者的效应，还是对最富裕者的效应？或者是某项政策对农村社区与城市社区的效应？此外，我们想要考察哪种因果效应？是局部均衡还是一般均衡，抑或是在固定某些行动的情况下考察？是否考虑外部性？等等。只有在结合政治学家的研究问题明确界定这些内容之后，研究才能有所进展。

我们再次强调，无论采用经验方法还是理论方法，这一步都是相同的。

6.2 第二步：什么样的（形式化）模型最能近似经验设定?

明确目标之后，研究者接下来需要定义其研究的形式化环境，在这个环境中，第一步中的参数才具有意义。这包括该环境中有哪些行动者、他们的行为是否具有策略性、信息结构（哪些主体知道什么、又有哪些信息并不知晓）、时序（例如，该环境是静态的还是动态的）等等。

这一步的最终目标是让形式模型根据底层模型参数，对可观测数据作出预测。例如，在研究政治

国会极化时，形式模型会根据政客的意识形态（例如 DW-Nominate，见第 2.4 节），预测其对某项法案投赞成票或反对票的决定；而另一些模型还进一步允许这一决定取决于政党纪律参数（例如 Clinton 等人，2004； Canen 等人，2020、 2022）。有些模型设定二次偏好（例如 Heckman and Snyder，1997），另一些则采用高斯分布（DW-Nominate）。尽管如此，所有这些模型都会根据模型特定参数生成可观测预测，正如我们在第 2.4 节所展示的那样。

关于模型环境的选择，最好以理论、数据、制度知识和内部一致性等作为指导。例如，将最高法院大法官在决策过程中进行审议的作用，近似为大法官彼此充分了解对方偏好的环境（毕竟他们共事数十年），是否合理？还是假设他们掌握私人信息更为贴切（例如 Iaryczower 等人，2018）？数据中是否存在某种模式，表明私人信息是更合适的假设——例如，结果变量的相关结构有所不同？或者，包含私人信息的理论模型是否会产生更符合原始数据模式的均衡结果？

有时，这一选择会比较简单，因为参数可以从应用更广泛的理论模型中推导出来。例如，Acemoglu 等人（2015）研究哥伦比亚各市镇的公共品供给。他们关注这类溢出效应的程度：市镇可以选择付出成本来提供公共品，也可以搭便车，享受其他市镇提供的公共品。这一决策被建模为网络上的策略性选择，其均衡映射为利用结果变量（公共品供给数据）、网络（市镇的地理邻近关系）以及其他市镇层面的数据进行估计提供依据。

在第4.1节提到的联盟战争作战示例中，Kenkel 和 Ramsay（2021）将危机的经典议价模型与团队之间的竞赛函数结合起来。均衡分析表明，当双方各有多个成员时，联盟赢得战争的概率服从广义 Logit 分布；而当战争发生在两个国家之间时，该概率则服从简单的 Logit 分布。

由于结构式研究会将形式模型与数据相结合，因此，对环境作出的此类选择尤为重要。即使某种环境足以识别关注参数（见下文第三步），不考虑这些特征也可能导致拟合效果或实证表现变差（见下文第四步）。

6.3 第三步：要使参数（给定数据后）可识别，模型需要满足哪些条件？

到这一步，研究人员应当已经熟悉研究目标，以及模型对环境的总体近似。他们也应该已经得到了

模型关于参数的预测：例如，联系可观测数据且仅取决于参数的均衡关系。前两个步骤与其他任何实证研究中的步骤相似。结构方法与形式理论的区别体现在第三步。

形式模型仅仅得出能够映射到目标参数的理论预测，已经不够了。现在，我们必须证明感兴趣的参数是否可识别。由于各个形式模型可能彼此不同，因此没有明确的操作步骤。不过，我们会概述一些现有路径，同时牢记第3节中的定义。

通常，研究者会利用模型对可观测结果的预测，并将这些预测写成其他可观测数据的函数。例如，在存在策略互动的模型中，均衡结果（如是否发动战争）将是参数以及博弈参与者特征的函数。结果变量和特征都是已知的，我们必须为不可观测变量和参数找到可识别的条件。完整证明的例子包括Canen 等人(2020，2022）；Kawai 等人(2021）等。

在此过程中，研究者往往会发现需要施加额外约束：例如，矩条件、归一化（分布冲击的均值必须为零）、独立性约束（例如，个体类型必须独立于其他可观测变量），或者参数化（冲击的分布设为逻辑斯蒂分布，或偏好设为可观测特征的线性函数）。[13] 有时，选择也可能是出于易处理性和计算方面的考量，这也是逻辑斯蒂分布广受欢迎的原因。再次强调，这种选择是否合理，取决于（i）在该情境下是否看起来合理（也就是说，行为主体类型是否可能独立于政策？）；（ii）它是否会改变某个参数的解释；（iii）它是否会遗漏数据的关键特征。

判断某个参数是否无法识别，往往更为容易。简单来说，就是数据中的信息不足以区分模型中各参数所起的作用。例如，如果两个参数在模型中总是以乘积的形式出现，那么尽管二者不同，它们在观测上却发挥着相同的作用。它们无法分别识别，因此需要进行归一化。[14]

[13]我们注意到，归一化可能会改变可识别参数的解释：例如，某种归一化乍看之下似乎无关紧要，但可能会改变某个反事实效应参数的解释——参见 Aguirregabiria 和 Suzuki(2014)。[14]即使识别失败，模型仍然可能提供有用信息。例如，多重均衡意味着在给定相同可观测数据的情况下存在多个解。因此，在这种情形下，参数可能无法被确定。尽管如此，人们仍可能发现，无论处于哪种均衡，结果变量都是正的。文献将此类情形纳入部分识别的研究范畴——这可能是由于数据或模型不完整所致，但仍然可能有用。本文不综述此类方法，读者可参阅 Molinari (2020)。

因此，研究者需要针对所研究的问题选取合理的假设，以实现识别。如何衡量某个假设是否“合理”，取决于理论和研究问题。有些识别限制从根本上无法检验，其中包括对不可观测变量分布的结构设定。这类似于工具变量模型中常见的“排除”限制。不过，鉴于已有的理论研究，或定性、定量证据，这些限制可能是合理的。另一些假设则很容易检验。例如，数据中是否存在特定的相关性。

在我们贯穿全文的例子中，即第4.1节讨论的战争中联盟的效力，我们知道模型是可识别的，因为Logit模型是可识别的，这一点已在第3.2节证明。

6.4. 第四步：将模型应用于数据。

现在，研究者已经从形式上证明了其关注参数的统计识别。他们认为，自己的形式模型抓住了问题的关键方面，并且进一步的假设是合理的。现在，他们已准备好将模型应用于数据。

因此，现在该选择一个合适的估计量了——也就是说，选择一个对于其问题而言具有合理性质（例如，一致性和渐近正态性）的估计量。这种选择本质上取决于最终得到的模型。如果模型是线性的且可加可分（例如第2.1节中的线性模型），那么普通最小二乘法就是一个合适的候选方法。这是因为它有诸多优点：它具有一致性，有闭式解，渐近正态；并且在同方差性条件下，它在线性无偏估计量中具有最小方差。

然而，大多数结构模型最终并非线性且可加可分。对于非线性模型，具体选择取决于非线性的结构。例如，如果用于识别的均衡条件可以写成矩条件，那么一个候选估计量就是广义矩估计量（GMM）。在正则条件下，这类估计量将是一致且渐近正态的（Hansen，1982）。当矩条件的数量多于参数数量时，可以赋予权重，使其在一类合理的估计量中达到有效。GMM估计量是半参数的，因为它只使用数据的矩（期望），而不使用数据的完整分布。如果研究者没有对模型中不可观测变量施加参数分布假设，这种做法就很方便。

然而，在其他模型中，研究人员可能已经施加了进一步的参数假设：要么是为了识别，要么是因为这些假设有助于捕捉数据中的重要模式（例如厚尾），或有助于处理选择问题或缺失数据

（例如 Heckman（1979））或因为这些假设对理论研究很重要（Canen 等人，2020采用正态分布冲击，从而保证满足均衡唯一性的单调风险率条件）。在这种情况下，极大似然估计量是理想的选择。第 2.2 节中的 Logit 示例就是如此；这一点也适用于第 4.1 节中的主要示例，因为其中已施加逻辑斯蒂分布假设。此外，在一类合理的估计量（渐近无偏）中，MLE 达到了克拉美-罗界所规定的最低方差。

不过，极大似然估计的计算量可能非常大。当必须对多维不可观测变量进行积分时，就会出现这种情况，而这类积分可能没有闭式解。在这种情况下，模拟方法或间接推断可能会很方便：研究者尝试多种不同的参数组合，分别对每种组合模拟模型，并检验哪些组合对数据的拟合效果最佳。统计学教材对各种估计量的性质以及许多其他内容都有很好的综述，例如 Wooldridge(2010)。有了这些估计量，研究人员便可通过适当的推断方法来检验其基本假设，例如，利用这些估计量的渐近分布进行假设检验。

6.5. 第五步：结构模型的评估

至此，研究者已经获得了其模型（参数）的估计值，以及其他目标的估计值，包括反事实分析的结果。研究者可能会考虑如何评估其结构模型——包括是否需要进行修改，或是否有改进的方法。下面概述一些思路。为此，不妨回顾一下在线性回归模型这一较简单情形下，这类分析是如何进行的，详见第（2.1）节。

在经典线性回归模型中，通常会进行四种不同类型的验证。第一，作者可以提供样本内拟合的度量（例如，R^2）。第二，研究者可以提供样本外拟合的度量（例如，样本外预测，或针对估计时留出的样本计算均方误差）。第三，作者可以检验自己的模型是否优于竞争模型（例如，是否拒绝替代模型而支持自己的模型）。这可以通过估计嵌套或非嵌套替代模型，并检验它们是否优于基准模型来实现（例如，检验数据是否拒绝替代模型提出的系数，或替代模型对数据的拟合是否较差）。第四，作者可以检查数据中是否缺少某些关键要素：例如，数据可能未能捕捉尾部事件。

这个类比很有用，因为在结构性分析中，也可以进行完全相同类型的验证。

首先，研究人员可以给出样本内拟合的度量。当模型是非线性的，这可以包括将模型的矩（均值、方差等）与观测数据进行拟合比较。

其次，研究人员可以使用样本外拟合度量。例如，Francois 等人（2022）估计了一个关于派系及中国共产党中央委员会组织方式的模型，使用了包括政治人物地理来源在内的个体层面的政治人物数据。他们关注委员会成员遴选中的激励因素，但由于这些选择往往不公开，他们必须依靠理论来弥合这一缺口。他们使用了1921—2012年的数据。随后，他们用模型估计值预测2017年第十八次党代会的构成。另一个例子是，Canen 等人（2022）将模型对均衡社会关系的预测与数据中观察到的情况进行比较。他们的估计不包括有针对性的（成对）关系。

这类似于机器学习中使用的度量。但由于许多结构模型是非线性的，均方误差并非恰当的比较统计量。相反，应使用与模型相一致的度量，例如，在能够计算似然时使用似然比。

第三种方法是针对替代模型进行检验。常见的情形是检验关注的效应是否存在（即某个参数或平均效应是否等于零）。例如，在内战冲突中，可以检验策略替代性是否为零。这可以使用针对嵌套模型的方法来完成：或者对相关参数是否为零进行假设检验，或者进行似然比检验。

在其他情形中，替代模型本质上是非嵌套的。例如，Francois 等人(2015研究了非洲各国政府的内阁席位分配，以检验内阁席位如何在不同族群之间进行策略性分配。在该模型中，独裁者在决定如何分配内阁席位时，会预见到可能发生的政变或革命。与之竞争的理论是“大人物”权力理论，该理论认为非洲的独裁者拥有不受约束的决策能力。后一种理论会产生不同于Francois 等人(2015)战略共享模型的内阁席位分配组合。为检验这类非嵌套模型，作者分别估计每个模型，并进行 Vuong 检验（Vuong，1989）。Vuong 检验特别适合检验一个参数模型是否比一个非嵌套模型拟合得更好：它通过恰当地比较各模型的似然，考虑不同的自由度，并检验哪个模型的似然最大（也就是说，按照 Kullback-Leibler 意义，该模型的分布最有可能更接近真实分布）。Shi (2015)的近期研究作出了新的贡献：当模型可能接近嵌套时，该方法具有更好的性质。

7. 结构方法的局限

毋庸赘言，结构估计的优势显然也伴随着代价。许多人指出，约化式方法难以将其估计目标与所关注的理论量联系起来；结构估计虽避免了这一问题，却需要将行为假设和均衡假设直接纳入统计模型。因此，结构估计实际上是向读者提出一个命题：如果你认为所假定的模型是良好或有用的，那么这些估计就会告诉你现实世界对这些理论量意味着什么。如果你不认同行为假设和均衡假设，那么这些量在现实世界中就几乎没有意义。

关于随机变量参数形式和效用函数函数形式等的各种假设，可能影响统计检验的规模和功效，也可能影响参数的识别。另一个顾虑是行为模型是否恰当。在形式理论发展成熟的领域，例如空间投票或危机谈判，可能鲜有争议。在其他领域，理论分歧可能很大，或者完全缺乏理论。理论学家之间对于我们应当如何看待风格化模型，也存在分歧。在这些情况以及许多其他情况下，采用约化式实证策略是合理的，也能有所贡献。不过，我们强调，在这些情形中，简化式方法仍然依赖于各自存在局限的不同模型。

简化式方法与结构方法也可能相辅相成。例如，结构模型可以将一个极具可信度的因果效应作为估计时需要匹配的矩条件，或作为模型评估的手段。也就是说，基于完整的结构模型，可以计算某个变量的平均处理效应，并查看它是否与通过准实验方法估计出的结果相符。参见 Todd and Wolpin(2020，了解关于结合此类方法的讨论。

通过阐明理论与估计之间的联系，结构方法为政治学提供了重要洞见，并让我们了解政治理论需要在哪些方面改进。但这一益处并非没有代价。许多理论模型高度风格化，难以推广；而有时，推广后的模型对现实世界可能发生的情况几乎没有任何限制。此外，数据限制可能要求作出更多假设，从而削弱结构性研究对底层理论的阐释力；其中一些假设可能仅仅是为了让计算成本保持在可承受范围内。尽管如此，结构方法至少将研究聚焦于理论与实证分析之间的对话，在这一过程中，双方相互推动，共同增进我们对政治世界的实质性理解。

8. 结论

实证研究不应受方法论阵营（例如，实验与理论、结构式与简化式）的划分左右，而应以推导出识别并恰当估计关注参数所需的最少假设为基础，无论该参数是什么。有时，某个研究问题可能需要一个经过完整推导并通过实证估计的均衡模型（即结构模型），但有时则不需要。有时，准自然实验研究设计或许就足够了，但很多时候并非如此。有时，形式模型可以嵌入实验设计之中，但在另一些情况下，这可能行不通。我们应当恰当地运用理论和实证方法，以充分研究本领域广泛而有趣的问题。

本文综述了在政治学中日益得到应用的结构方法。通过识别并估计模型特定参数，结构方法使研究人员能够探讨不同于约化式分析的一系列问题。这些问题可能涉及偏好（例如意识形态极化的变化）、福利（例如信息变化对均衡政策的影响），或需要将结论外推至新环境的问题。然而，无论是否采用结构方法，任何实证研究都需要明确界定关注参数和（统计）模型，并仔细阐述识别与估计方法。这意味着实证过程的每一步都需要理论。

参考文献

Abramson, S. F., K. Koc.ak, and A. Magazinnik (2022). 我们从联合实验中能了解到哪些选民偏好？《美国政治科学杂志》，即将出版。Abramson, S. F. and S. Montero (2020). 关于增长与民主的学习。《美国政治学评论》*114*(4), 1195–1212. Acemoglu, D., C. Garc´ıa-Jimeno, and J. A. Robinson (2015). 国家能力与经济发展：一种网络方法。《美国经济评论》*105*(8), 2364–2409. Achen, C. H. and L. M. Bartels (2017, August). 现实主义者的民主：为何选举无法产生回应型政府。普林斯顿大学出版社。Aguirregabiria, V. and J. Suzuki (2014). 动态模型中市场进入与退出的识别和反事实。*QuantitativeMarketing and Economics 12*(3), 267–304. Al-Ubaydli, O., J. A. List, and D. L. Suskind (2017, May). 我们能从实验中学到什么？理解实验结果可扩展性的威胁。《美国经济评论》*107*(5), 282–286. Angrist, J. D. and J.-S. Pischke (2010). 实证经济学中的可信度革命：更好的研究设计如何消除计量经济学中的欺诈。*Journal of EconomicPerspectives24*(2), 3–30. Ashworth, S., E. Bueno de Mesquita, and A. Friedenberg (2018). 关于选民理性的学习。《美国政治科学杂志》*62*(1), 37–54. Athey, S. and P. A. Haile (2002). 标准拍卖模型的识别。*Economet-rica70*(6), 2107–2140. Avis, E., C. Ferraz, F. Finan, C. Varj˜ao, et al. (2022). 金钱与政治：竞选支出限制对政治进入和竞争的影响。*American EconomicJournal: Applied Economics14*(4), 167–199. Bartels, L. M. and J. Zaller (2001). 总统投票模型：重新审视。*PS: PoliticalScience and Politics34*(1), 9–20. Battaglini, M., E. Patacchini, and E. Rainone (2021, 09). 内生社会互动与未观测网络。*The Review of Economic Studies89*(4), 1694–1747. Boche, A., J. B. Lewis, A. Rudkin, and L. Sonnet (2018). 新版 voteview.com：保存并延续 Keith Poole 为国会议员、学生和观察者搭建的基础设施。*Public Choice 176*, 17–32. Bonica, A., N. McCarty, K. T. Poole, and H. Rosenthal (2013, September). 为什么民主未能减缓不断加剧的不平等？*Journalof Economic Perspectives 27*(3), 103–124. Burke, M., S. M. Hsiang, and E. Miguel (2015). 气候与冲突。*AnnualReview ofEconomics7*(1), 577–617.

Canen, N., M. O. Jackson, and F. Trebbi (2022). 社会互动与立法活动。《欧洲经济学会杂志》. Canen, N., C. Kendall, and F. Trebbi (2020). 分解极化。《计量经济学》*88*(3), 1197–1233. Canen, N. J., C. Kendall, and F. Trebbi (2022). 政党如何推动美国极化：1927-2018。技术报告，加利福尼亚大学伯克利分校。 Cantoni, D., D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2019). 抗议活动作为策略博弈：来自香港反威权运动的实验性证据。《经济学季刊》*134*(2), 1021–1077. Carroll, R., J. B. Lewis, J. Lo, K. T. Poole, and H. Rosenthal (2009). 通过参数自助法衡量DW-Nominate理想点估计中的偏差与不确定性。《政治分析》*17*(3), 261–275. Clinton, J., S. Jackman, and D. Rivers (2004). 唱名表决数据的统计分析。《美国政治学评论》*98*(2), 355–370. Cox, C. (2022). 为政府拨款而游说。《兰德经济学杂志》. Crisman-Cox, C. and M. Gibilisco (2018). 受众成本与战争与和平的动态。《美国政治科学杂志》*62*(3), 566–580. Cruz, C., P. Keefer, J. Labonne, and F. Trebbi (2020). 让政策发挥作用：选民对竞选承诺的反应。加利福尼亚大学洛杉矶分校. Cunningham, S. (2021). 《因果推断：混音带》. 耶鲁大学出版社. Diermeier, D., H. Eraslan, and A. Merlo (2003). 政府组建的结构模型。《计量经济学》*71*(1), 27–70. Dunne, J. P. and R. P. Smith (2007年1月). 第28章：军事军备竞赛的计量经济学。载于 T. Sandler and K. Hartley (编), 《国防经济学手册》，第2卷《国防经济学手册》，第913–940页。爱思唯尔。 Egami, N. and E. Hartman (2022年10月). 外部效度的要素：框架、设计与分析。《美国政治学评论》, 1–19. Findley, M. G., K. Kikuta, and M. Denly (2021). 外部效度。《政治学年度评论》*24*(1), 365–393. Fowler, A. and B. P. Montagnes (2022). 区分假阳性与真实结果：无关事件与选举的案例。第 Forthcoming 页。 Francois, P., I. Rainer, and F. Trebbi (2015). 非洲如何分享权力？《计量经济学》*83*(2), 465–503. Francois, P., F. Trebbi, and K. Xiao (2022). 非民主国家的派系：来自中国共产党的理论与证据。技术报告，未刊稿，英属哥伦比亚大学。

Gentzkow, M. and J. M. Shapiro (2010). 是什么推动了媒体倾向？来自美国日报的证据。《计量经济学》*78*(1), 35–71. Gerber, A. (1998年6月). 使用工具变量估计竞选支出对参议院选举结果的影响。《美国政治学评论》*92*(2), 401–411. Graham, M. H., G. A. Huber, N. Malhotra, and C. H. Mo (2022). 无关事件与投票行为：运用开放科学原则进行的重复研究。 Gulzar, S., M. R. Rueda, and N. A. Ruiz (2021). 竞选捐款限额能否遏制政治中的金钱影响？《美国政治科学杂志》. Hansen, L. P. (1982). 广义矩估计量的大样本性质。《计量经济学：计量经济学会杂志, 1029–1054. Hansford, T. G. and B. T. Gomez (2010年5月). 估计选民投票率的选举效应。《美国政治学评论》*104*(2), 268–288. Hartman, E. (2021). 推广实验结果。载于 D. P. Green and J. N. Druckman (编), 《实验政治学进展》, 第385–410页。剑桥：剑桥大学出版社。 Heckman, J. J. (1979). 样本选择偏差作为设定误差。《计量经济学：计量经济学会杂志, 153–161. Heckman, J. J. (2008). 计量经济学因果关系。《国际统计评论》*76*(1), 1–27. Heckman, J. J. (2010). 在结构式方法与项目评估方法之间搭建桥梁以评估政策。《经济文献杂志》*48*(2), 356–98. Heckman, J. J. and J. M. Snyder (1997). 属性需求的线性概率模型：立法者偏好估计的实证应用。《兰德经济学杂志》*28*. Iaryczower, M., G. Lopez-Moctezuma, and A. Meirowitz (2022). 职业前景顾虑与选举问责的动态。《美国政治科学杂志》. Iaryczower, M., S. Montero, and G. Kim (2022). 代表失灵。技术报告，美国国家经济研究局。 Iaryczower, M., X. Shi, and M. Shum (2018). 言辞会造成阻碍吗？集体决策中审议的效应。《政治经济学杂志》*126*(2), 688–734. Iaryczower, M. and M. Shum (2012). 法院中的信息价值：求准、求精。《美国经济评论》*102*(1), 202–237. Kang, K. (2016). 能源行业游说的政策影响力与私人收益。*TheReview of Economic Studies83*(1), 269–305.

Kang, K. 和 R. A. Miller（2022）。默认获胜：为什么政府采购中的竞争如此之少？*The Review of Economic Studies89*(3)，1495–1556。Kawai, K.、Y. Toyama 和 Y. Watanabe（2021）。选民投票率与偏好聚合。《美国经济学杂志：微观经济学》*13*(4)，548–86。Keane, M. P.（2010）。结构式方法与非理论计量经济学方法。《计量经济学杂志》*156*(1)，3–20。Kendall, C.、T. Nannicini 和 F. Trebbi（2015）。选民如何回应信息？来自一项随机化竞选活动的证据。《美国经济评论》*105*(1)，322–53。Kenkel, B. 和 K. W. Ramsay（2021）。军事联盟的有效力量：统一的理论与实证模型。未发表手稿。*https://bkenkel. com/files/structwar. pdf*。Koopmans, T. C.（1947）。无理论测量。*The Review of Economics andStatistics29*(3)，161–172。Kurizaki, S. 和 T. Whang（2015）。识别国际争端中的受众成本。《国际组织》*69*(4)，949–980。Lewbel, A.（2019）。识别动物园：计量经济学中识别的含义。《经济文献杂志》*57*(4)，835–903。Luce, R. D.（1959）。《个体选择行为：理论分析》，纽约：*John Willey and Sons*。Inc. Lundberg, I.、R. Johnson 和 B. M. Stewart（2021年6月）。你的估计目标是什么？定义目标量将统计证据与理论联系起来。《美国社会学评论》*86*(3)，532–565。Marshak, J.（1953）。用于政策制定和预测的经济测量。*Studies in Econo-metricMethod*，1–26。Martin, G. J. 和 A. Yurukoglu（2017）。有线电视新闻中的偏差：说服与极化。《美国经济评论》*107*(9)，2565–99。Molinari, F.（2020）。具有部分识别的微观计量经济学。*Handbookof Econo-metrics7*，355–486。Nadeau, R. 和 M. S. Lewis-Beck（2001年2月）。美国总统选举中的全国性经济投票。《政治学杂志》*63*(1)，159–181。Poole, K. T. 和 H. Rosenthal（1984）。《1968—1980年美国总统选举》：空间分析。《美国政治科学杂志》*28*(2)，282–312。Rekkas, M.（2007）。竞选支出对多党选举中选票的影响。《经济学与统计学评论》*89*(3)，573–585。Richardson, L. F.（2012年11月）。《军备与不安全：战争起因与根源的数学研究》。Literary Licensing, LLC。

Rivers, D.（2003）。多维空间投票模型的识别。打字稿。斯坦福大学。Shi, X.（2015）。非退化 Vuong 检验。《数量经济学》*6*(1)，85–121。Signorino, C. S.（1999年6月）。战略互动与国际冲突的统计分析。《美国政治学评论》*93*(2)，279–297。Smith, R. P.（1989）。军事支出模型。《应用计量经济学杂志》*4*(4)，345–359。Todd, P. E. 和 K. I. Wolpin（2020）。鱼与熊掌兼得：将随机对照试验与结构建模相结合。《经济文献杂志》。Ujhelyi, G.、S. Chatterjee 和 A. Szab´o（2021）。一个也不选：世界上最大的民主国家中的抗议投票。《欧洲经济学会杂志》*19*(3)，1936–1979。Vuong, Q. H.（1989）。用于模型选择和非嵌套假设的似然比检验。《计量经济学：计量经济学会杂志》，307–333。Wolpin, K. I.（2013）。《无理论推断的局限》。麻省理工学院出版社。Wooldridge, J. M.（2010）。《横截面数据与面板数据的计量经济分析》。麻省理工学院出版社。