



北京大学中国经济研究中心
China Center for Economic Research

讨论稿系列
Working Paper Series

C2026002

2026-01-22

结构方程的基本概念、核心步骤、技术难点与前沿进展*

李星宇 彭依菁 易君健

摘要

随着约简式研究局限性的展现和技术方法的进步,结构方程研究范式在经济学领域被重新重视起来。为了帮助更多研究者了解和使用结构方程,本文介绍了结构方程的基本概念和核心步骤。结合具体案例和已有研究,本文详细阐述了结构方程的模型构建、参数识别、参数估计、统计推断、模型验证和模型应用等步骤中的现有理论和方法,也指出了其中存在的技术困难。此外,本文还讨论了结构方程研究的前沿进展,即将约简式研究和机器学习与结构方程相结合,从而一定程度上解决模型构建、参数识别、参数估计和模型验证等方面困难。我们希望本文能够推进结构方程研究范式被更广泛地应用于探索社会经济发展的一般规律,从而为科学的最优政策设计提供严谨的分析基础。

结构方程的基本概念、核心步骤、技术难点与前沿进展*

李星宇 彭依菁 易君健

2026 年 1 月 21 日

摘要

随着约简式研究局限性的展现和技术方法的进步，结构方程研究范式在经济学领域被重新重视起来。为了帮助更多研究者了解和使用结构方程，本文介绍了结构方程的基本概念和核心步骤。结合具体案例和已有研究，本文详细阐述了结构方程的模型构建、参数识别、参数估计、统计推断、模型验证和模型应用等步骤中的现有理论和方法，也指出了其中存在的技术困难。此外，本文还讨论了结构方程研究的前沿进展，即将约简式研究和机器学习与结构方程相结合，从而一定程度上解决模型构建、参数识别、参数估计和模型验证等方面的困难。我们希望本文能够推进结构方程研究范式被更广泛地应用于探索社会经济发展的一般规律，从而为科学的最优政策设计提供严谨的分析基础。

*李星宇，浙江大学经济学院，邮政编码：310058，电子信箱：lixyecon@zju.edu.cn；彭依菁（通讯作者），北京大学国家发展研究院、中国经济研究中心，邮政编码：100871，电子信箱：yjpeng2023@nsd.pku.edu.cn；易君健，北京大学国家发展研究院、中国经济研究中心，邮政编码：100871，电子信箱：junjian@nsd.pku.edu.cn。作者感谢曹治、陈灯坤、胡诗云、江弘毅、林轶凡、吴周恒、颜玮、尹晔、张皓辰、周子焜、朱涵宇对本文提出的宝贵意见和建议。本文文责自负。

1 引言

作为社会科学的皇冠, 经济学对人类经济社会发展的贡献毋庸置疑. 然而当下经济学学科自身的发展却出现一些问题, 特别是经济学理论和经验数据分析之间存在脱节. 在应用微观经济学领域, 由于约简式 (reduced form) 研究的兴起, 研究者投入大量精力去追求内部效度. 这种以“原因的结果”为导向的因果推断范式虽然提升了研究的统计严谨性, 但是也弱化了经济学的探索精神. 我们的现实世界由已实现的均衡结果构成, 经济学家应当从这些均衡结果中发现各种有趣、困惑和重大的现象, 进而探索这些现象背后的原因、机制和一般科学规律, 最后用这些科学的规律指导政策实施. 基于这样的科学探索精神, 经济学不仅能够分析原因产生的后果, 也能够通过理论建模来解析“结果的原因”, 即现象背后深层的一般科学规律. 这一能力将使得经济学方法在预测问题中比单纯的统计方法更具优势: 在数据生成过程发生结构性变动时, 统计方法从历史数据中训练出的模型无法直接用于未来的预测; 而基于经济理论的建模方式能够将发生变化的结构与不变的结构分离开, 在不变的结构 (如偏好、技术和制度) 的大框架下, 用变量和/或参数描述变化的结构, 使得整体模型的未知参数可以从已有数据中估计, 并在此基础上通过反事实分析 (counterfactual analysis) 预测未来的均衡结果. 经济学如果能够充分发挥在预测方面的优势, 就能推动政策评估从事后分析向事前预测转型. 实现上述目标的核心在于基于经济理论的建模和数据分析, 这也就是结构方程 (structural equation) 范式. 一个具有代表性的基于结构方程的事前预测研究是 [McFadden et al. \(1977\)](#), 该研究使用随机效用模型在旧金山湾区捷运系统尚未建成时预测其客运需求, 而系统投入使用后的实际数据表明, 这些事前预测相当准确.

结构计量模型的思想可追溯至新古典经济学的兴起. 马歇尔的供需分析为后续构建基于代表性行为人最优化的宏观模型提供了雏形. 20 世纪 30 年代, 考尔斯委员会 (Carls Foundation) 主张通过结构方程研究刻画偏好、技术和制度的“基础参数” (primitive parameters), 并在此基础上预测政策变动下的新均衡. 然而, 结构方程研究范式在那个年代的局限性很快暴露出来: 受制于数据和算力, 研究者需要对模型施加较强的且难以检验的假定条件以保证可操作性, 但这也使得研究结果很大程度上依赖于这些假定, 在现实经济中的预测能力欠佳. 因此, 结构方程研究范式在此后很长一段时间没有受到重视.

与此同时, 计量经济学在宏观与微观领域走向分化. 宏观计量早期依赖向量自回归 (vector auto-regression)、协整分析 (co-integration analysis) 和误差修正模型 (error correction model) 等非结构化的时间序列分析方法. 随着真实商业周期 (real business cycle) 与动态随机一般均衡 (dynamic stochastic general equilibrium) 模型的兴起, 宏观计量的研究开始强调其经济学理论基础. 然而, 由于模型包含动态性、随机性、异质性和互动性导致形式非常复杂, 宏观模型通常难以实现计量经济学意义上的参数识别、估计和推断, 因而多在数值求解模型的基础上采用校准 (calibration) 的方式得到参数, 包括但不限于使用文献已有结果、借助理论和经验判断或匹配数据特征. 相比之下, 微观计量则沿约简式路径快速发展: 以随机对照实验 (randomized controlled trial) 和准实验设计 (quasi-experimental design) 为代表的研究范式强调在施加尽可能少的行为假定的情形下识别因果效应. 随机对照实验通过随机化设计, 从根本上避免了混杂因素和样本选择问题, 但是大量经济学研究问题不适合或者没有条件开展随机对照实验. 准实验设计则在给定观察数据 (observational data) 的条件下, 通过研究设计尽可能解决混杂因素和样本选择问题. 因为其模型设定简单且方法易于实施, 约简式研究范式受到研究者 (特别

是应用微观经济学研究者) 广泛青睐. 在我国, 2008 年后中国健康与养老追踪调查 (China Health and Retirement Longitudinal Study, CHARLS) 和中国家庭追踪调查 (China Family Panel Studies, CFPS) 等大规模微观数据库的建立推动了约简式研究方法更为广泛的应用, 《基本无害的计量经济学》等工具书的引入更强化了这一趋势.

近年来, 研究者也开始逐渐意识到约简式研究范式的一些局限性, 例如行为假定的经济学含义不够清晰、缺少对于深层机制和经济规律的分析、研究结果难以推广到存在结构变迁的场景等. 与此同时, 伴随着技术进步, 经验研究数据获取的难度大大降低, 计算机的算力迅速提升且更高效的算法不断涌现, 这意味着研究者已经有能力处理形式较为复杂且假定相对较弱的结构方程模型. 在这样的背景下, 结构方程研究范式开始在经济学经验研究中被重新重视起来, 特别是在那些需要还原基础参数、厘清经济机制、开展事前分析等场景当中. 在我国, 结构方程也逐渐被应用于经济学各个领域, 包括产业组织 (李国栋等, 2019; 冯笑和王楚男, 2022; 郭晓丹和王帆, 2024)、劳动与公共经济学 (杨继生和邹建文, 2021; 徐彤等, 2023; 易行健等, 2025)、城市与区域经济学 (刘修岩和李松林, 2017; 陈诗一等, 2019; 余森杰等, 2025)、国际贸易经济 (王子和周雁翎, 2019; 谭用、周洛竹和綦建红, 2024; 谭用等, 2024; 冯笑等, 2024) 等. 为了帮助研究者更好地了解和掌握结构方程研究范式, 进而为科学的最优政策设计提供严谨的分析基础, 本文回顾和讨论了结构方程的基本概念、核心步骤、技术难点与前沿进展.

一个完整的经济学经验研究 (empirical study), 无论采用约简式还是结构方程范式, 都应当包含 6 个步骤: 模型构建 (construction)、参数识别 (identification)、参数估计 (estimation)、统计推断 (statistical inference)、模型验证 (validation) 和模型应用 (application). 这 6 个步骤并不是相互割裂的, 而是有机地融合在一起. 因此研究者在开展经验研究时并不是机械地顺次执行每个步骤, 而是站在整体的视角, 在每个步骤中考虑其它步骤的需求或者对其它步骤进行修改, 以一种螺旋上升的方式完成研究. 除了约简式和结构方程这两大经验研究范式之外, 与经济学研究相关的量化分析方法还包括校准和机器学习 (machine learning). 本文将依次介绍结构方程研究的上述 6 个步骤. 对于每个步骤, 我们将基于已有理论和应用文献, 回顾现有理论和方法, 指出其中存在的困难, 介绍相关的前沿进展并展望未来研究方向. 与此同时, 我们也会关注结构方程与其它几种量化分析方法 (约简式、校准、机器学习) 的关系, 具体包括这 4 种量化分析方法的特征对比 (见表 1 的总结) 和结构方程与约简式、机器学习相结合的优势 (见表 2 的总结).

具体地, 本文的后续内容安排如下. 第 2 节概述结构方程的定义和基本步骤, 并以一个简单的供给-需求模型为例加以阐释. 第 3 节介绍了构建结构方程模型的难点和基本要素. 第 4 节给出了参数识别的严格定义, 并介绍了常见的参数识别方法及其局限性. 第 5 节介绍结构方程的模型求解和参数估计, 着重讨论了机器学习思想和算法对于复杂模型求解的帮助, 以及高维模型估计方法的最新进展. 第 6 节讨论针对结构方程模型的统计推断方法. 第 7 节讨论结构方程模型的验证和透明度分析. 第 8 节介绍了结构方程模型在经济分析中的应用. 第 9 节总结全文.

表 1: 几类常用量化分析方法的特征对比

	约简式	结构方程	校准	机器学习
模型构建	<ol style="list-style-type: none"> 从数据分布出发, 施加行为假定 参数有因果含义, 但是行为假定的经济学含义不清晰 模型形式相对简单, 有较为标准化的操作流程 	<ol style="list-style-type: none"> 从经济学理论出发, 以偏好、技术、制度等元素为基础参数 假定条件的经济学含义清晰且透明, 参数有经济学含义 需要根据具体研究问题和场景构建模型 	本质上也属于结构方程模型, 但是形式更加复杂, 通常包含动态性、随机性、异质性和互动性	<ol style="list-style-type: none"> 从数据特征出发, 以预测为目的, 构建纯粹的统计模型, 可视作函数逼近 模型参数通常不具有可解释性 有专门的模型选择方法 (如信息准则、交叉验证)
参数识别	<ol style="list-style-type: none"> 识别主要依赖于行为假定, 给定相应的行为假定成立, 识别过程简单 主要挑战在于说明识别假定的合理性 	<ol style="list-style-type: none"> 包括构建识别约束和证明可识别性 构建识别约束需要分析行为特征和经济变量之间的关系 严格证明可识别性具有技术挑战 	无计量经济学意义上的识别	不需要
参数估计	因为模型形式相对简单, 可以使用极大似然、最小二乘、广义矩方法等标准方法实现参数估计	<ol style="list-style-type: none"> 以标准的估计方法为基础 通常需要数值求解模型, 在复杂模型中计算量较大 通常需要与模拟方法结合 (如间接推断) 	校准: 借用已有数值、根据经验、匹配数据等, 不是计量经济学意义上的估计	<ol style="list-style-type: none"> 通常被称为模型训练, 表示为最小化损失函数的问题 仅使用训练集数据
统计推断	多数情形下可以使用基于估计量统计性质的标准方法	<ol style="list-style-type: none"> 可能需要结合具体情形推导渐近理论 在渐近理论的基础上需要解决数值计算问题 	无	<ol style="list-style-type: none"> 常以构建结果变量的预测集为目标 若方法已有渐近分布或非渐近误差上界结果, 则可用于构建预测集 如无 (非) 渐近结果, 可使用自助法或共形预测等数据驱动的方法
模型验证	报告不同模型、样本、方法等设定之下 的结果, 作为稳健性检验	使用新的样本和/或新的数据特征, 需要保持不可观测外生变量分布不变, 操纵可观测外生变量和参数	比较模型的数值模拟结果与现实经济数据	<ol style="list-style-type: none"> 利用不参与模型训练的测试数据集进行模型验证 常用指标: 预测的均方误差、分类错误率等
模型应用	<ol style="list-style-type: none"> 可以在较为简单的模型设定下得到“其它变量不变”或因果效应 可以在特征事实层面检验经济学理论 分析结果定量外推能力有限, 因此难以于事前的政策评估等情形 	<ol style="list-style-type: none"> 借助模型研究经济机制和社会福利 基于基础参数的稳定性开展反事实分析: 结构变迁后的预测、事前最优政策设计等 可以直接在模型层面检验经济学理论 	<ol style="list-style-type: none"> 解释经济现象和特征事实 预测未来或结构变迁后的经济变量 	<ol style="list-style-type: none"> 在稳定场景下的预测效果很好, 但是在迁移学习的前提条件不满足时, 难以用于发生结构变迁后的预测 可以与因果推断相结合

注 1: 正如 [Cocci and Plagborg-Møller \(2025\)](#) 第 2952 页介绍的, 文献中对于校准有狭义和广义两种定义, 狹义的定义指的是宏观经济学中以非计量的方式获取参数的做法, 而广义的定义是经典最小距离估计方法. 本文所讨论的校准采用狭义定义.

注 2: 为了方便, 在没有额外说明的情况下, 本文所称的机器学习均特指有监督学习 (supervised learning).

表 2: 结构方程与约简式和机器学习结合的优势

	结构方程+约简式	结构方程+机器学习
模型构建	1.约简式研究可以提供特征事实、变量相关性,有助于研究者选择更符合现实情况的模型形式 2.模型设定检验可以为模型构建提供具有统计严谨性的证据	1.机器学习领域的模型选择算法可以帮助研究者用数据驱动的方式选择最佳模型 2.生成式模型等机器学习方法可以为变量提供测量指标 3.无监督学习或生成式人工智能可以从数据中提取特征事实,为模型构建提供信息
参数识别	在随机实验中,通过实验设计得到的变量之间的独立性可以用作识别约束	
参数估计	刻画变量之间相关关系或因果关系的约简式估计量都可以用作间接推断的辅助回归统计量	1.用深度学习求解动态经济模型,可以避免高维状态变量给数值计算带来的维数诅咒问题 2.将神经网络、生成对抗网络等方法应用于参数估计,既能提高模型设定的一般性(允许存在非参数部分),又能保持参数模型的效率
统计推断		数据驱动的推断方法(如自助法和共形预测)可以帮助研究者在缺乏渐近理论结果的情形下实现对结构方程的统计推断
模型验证	1.在包含多个处理组的随机实验中,可以选取控制组或者其中一部分处理组作为模型验证的样本 2.一个能够被当前结构方程模型理性化的约简式模型可以提供用于结构方程模型验证的矩条件和统计量	1.基于测试数据集的验证方法能够有效地评估模型的样本外预测能力(即外部效度) 2.在缺乏随机分配的情形下,迁移学习的一些思想可以帮助研究者选取用于模型验证的数据集
模型应用		用结构方程进行最优政策设计时,可以借鉴政策学习算法的思想,在经济学理论的基础上更有效地利用数据

2 结构方程概述

本节简要介绍结构方程的定义和基本步骤,并以一个简单的供给-需求模型为例来辅助说明.

所有的计量经济学模型本质都是对经济数据的分布建模,而结构方程的核心特征在于在经济学理论的基础上对数据分布进行建模.在 Matzkin (2007) 第 5313 页模型的基础上稍作修改,我们将一个结构方程模型表示为如下的一般形式:

$$\mathcal{M}(Y_o, Y_u, X, \varepsilon; \theta, \eta) = 0, \quad (X, \varepsilon) \sim G. \quad (1)$$

在上面的模型中, $Y = (Y_o, Y_u)$ 是从模型内生的变量 (endogenous variable), 例如消费者和厂商的最优选择、市场均衡结果, 其中 Y_o 可观测, Y_u 不可观测; X 是可观测的外生变量 (exogenous variable), 例如禀赋、经济政策; ε 是不可观测的外生变量, 例如个体层面的偏好和生产技术的扰动或异质性; θ 是研究者关注的未知经济参数 (parameter), 刻画了偏好、技术、制度等基础经济元素, 它可以是有

限维, 例如效用函数和生产函数的参数, 也可以是无穷维, 例如效用函数和生产函数本身; η 是研究者不感兴趣的干扰参数, 它也可以是有限维或无穷维, 通常不具有经济学意义, 例如用于标准化的常数; $\mathcal{M}(\cdot, \cdot, \cdot, \cdot; \cdot, \cdot)$ 是一个已知的函数, 用一组方程刻画了变量 Y_o, Y_u, X, ε 之间的关系, 而这组方程的函数形式完全由参数 (θ, η) 控制; 最后, G 是外生变量 (X, ε) 的联合分布, 它可以具有参数形式 (parametric) 也可以是半参数 (semi-parametric) 或非参数 (non-parametric) 设定.¹ 需要说明的是, 我们这里所述的变量可观测性是从研究者视角出发的; 这些变量对于模型中的行为人而言是否可被观测, 取决于具体的模型设定. 除此之外, 我们分别用 \mathcal{Y}_o 和 \mathcal{X} 表示可观测变量 Y_o 和 X 的取值空间.

与结构方程模型相对应地, 约简式模型则从可观测的变量 (Y_o, X) 的分布 (特别是变量之间的关系) 出发, 施加必要的行为假定, 建立可观测变量的统计模型. 它可以表示为如下的一般形式:

$$(Y_o, X) \sim P_r(\gamma, \zeta), \quad (2)$$

其中 $P_r(\gamma, \zeta)$ 是一个由研究者关注的函数或参数 γ 和干扰函数或参数 ζ 共同决定的概率分布. 这里 P_r 的形式中蕴含了简化的行为假定, 例如回归模型中的函数形式和误差项的统计性质, 以及因果推断中的潜在结果分析框架, 其目的是为了保证函数或参数 γ 具有其它条件不变 (*ceteris paribus*) 或因果效应的含义. 值得强调的是, 约简式研究的目标不是为了刻画数据背后的经济机制并还原偏好、技术、制度等基础参数, 而是分析变量之间具有因果意义的关联, 例如其它条件不变时, X 对 Y_o 的影响. 因此约简式模型的建立没有也不需要从完整的经济学理论模型出发, 而是在满足问题分析要求和符合已知事实信息的基础上以尽可能简单的方式刻画所关注变量的分布 (关系). 这就意味着在通常情况下, 约简式模型中的行为假定是技术性的, 不具有明晰的经济学含义, 而且约简式模型的参数也不具有偏好、技术、制度等经济学解读.

机器学习模型本质上也是可观测变量 (Y_o, X) 的统计模型, 它将 X 视作预测变量, 将 Y_o 视作结果变量, 以用 X 预测 Y_o 为目的. 在这个意义上, 机器学习可以看作是对 Y_o 关于 X 的条件分布的逼近, 或者按照 [Hastie et al. \(2009\)](#) 第 2.2 节的说法, 是一种函数逼近. 那么机器学习的模型设定实质上划定了逼近函数类, 而模型参数则用于确定函数类里面一个具体的函数, 这里允许不同的参数对应函数类中的同一个函数. 因此, 机器学习模型的参数是纯计算性的, 不具有现实的解读.

注意到, 从结构方程模型 (1) 出发, 我们可以推导出可观测变量 (Y_o, X) 的联合分布, 这一联合分布由参数 θ, η 和分布 G 共同确定, 形式上可以写为

$$(Y_o, X) \sim P_s(\theta, \eta, G). \quad (3)$$

利用 (2)–(3) 式, 我们可以介绍约简式模型可理性化的概念. 对于一个约简式模型 (2), 如果存在参数和分布 (θ, η, G) , 使得结构方程模型 (1) 推导出的数据分布满足 $P_s(\theta, \eta, G) = P_r(\gamma, \zeta)$, 那么就称约简式模型 (2) 可以被结构方程模型 (1) 理性化 (rationalize).

如前所述, θ 是研究者关注的基础经济参数, 但在很多的实际应用中, 研究者仅仅获取 θ 的信息是不够的, 因为后续的参数估计、统计推断、模型验证等步骤可能需要参数 η 和分布 G 中的一些信息, 即便它们本身没有经济学含义. 为了表述的一般性, 我们用 $\psi = \Psi(\theta, \eta, G)$ 来表示研究者希望研究的参

¹ 在 [Matzkin \(2007\)](#) 第 5313 页的模型中, G 被写为函数 \mathcal{M} 的一个参数, 这一写法本身无法体现 “ G 是外生变量 (X, ε) 的联合分布” 这一信息, 需要添加额外说明. 为了让模型表达式能够展现完整的信息, 本文采用了 (1) 式的写法.

数, 这里 Ψ 是一个映射. 从 (3) 式可知, 在一个结构方程模型之下, 数据的联合分布 $P_s(\theta, \eta, \mathbf{G})$ 中包含了参数和分布 $(\theta, \eta, \mathbf{G})$ 的信息, 从而也包含了我们希望研究的参数 ψ 的信息. 参数的识别, 指的是通过 $P_s(\theta, \eta, \mathbf{G})$ 还原 ψ , 即构建从数据分布 \mathbf{P} 到真实参数 ψ_0 的映射: $\psi_0 = \mathcal{I}(\mathbf{P})$. 这样的映射不必然存在, 如果它存在, 就称参数 ψ 是可识别的. 对于一般的结构方程模型, 参数的可识别性以及如何识别参数是一个复杂的问题, 我们将在第 4 节详细讨论.

在识别的基础上, 我们可以对参数进行估计和统计推断. 通过识别, 我们将真实参数 ψ_0 表示为数据分布 \mathbf{P} 的函数, 而未知的数据分布 \mathbf{P} 可以用样本来估计. 经过“用样本替换总体”的过程, 我们就能得到真实参数 ψ_0 的估计量 $\hat{\psi}$. 样本的随机性使得估计量 $\hat{\psi}$ 含有估计误差, 如果我们能推导出估计误差的 (精确或渐近) 分布并用样本去估计该分布, 那么我们就能据此对参数进行统计推断. 当然, 具体的估计和推断过程会因模型而异, 而且其中也存在着不少技术难题, 我们将在第 5 节和第 6 节详细讨论.

完成了参数估计之后, 为了诊断模型设定和参数估计的可靠性, 我们可以使用新样本来验证此模型. 我们要求这个新样本没有被用于参数估计, 但是与用于参数估计的样本具有相同或相似的外生变量分布. 如果模型设定正确且参数估计准确, 那么我们估计出的模型应当能够很好的拟合新样本. 换言之, 模型在新样本上的拟合优度可以用来诊断模型的可信度. 在没有新样本的情况下, 我们也可以使用一些未被用于参数估计的矩条件来验证模型. 如果模型设定正确, 这些总体矩条件需要在理论上成立, 我们可以基于当前估计的参数计算这些总体矩条件对应的样本矩条件, 以此来验证模型. 在确认了一个可靠的模型后, 我们可以在此基础上计算新的经济变量和指标 (例如福利分析), 还可以通过反事实分析研究一些在现实中尚未发生的情形. 所谓反事实分析, 是指改变模型中部分参数的取值和分布的设定, 在新模型下模拟数据, 分析新均衡的性质. 在将反事实分析用于政策分析时, 这种尚未发生的情形既可以指现有政策的推广, 也可以指一个全新的政策.

接下来, 我们以一个商品的供给-需求模型为例来说明上述概念和步骤.

例 1: 设 R 和 Q 分别是某种商品的价格和数量, 我们考虑线性形式的供给曲线 (反供给函数) 和需求曲线 (反需求函数):

$$\text{供给: } \ln R = a \ln Q + bZ + U, \quad (4)$$

$$\text{需求: } \ln R = c \ln Q + V, \quad (5)$$

其中 Z 是该产品生产线的自动化程度, U 是不可观测的来自生产技术的随机冲击, V 是不可观测的来自消费者偏好的随机冲击. 我们可以将这一模型表示为 (1) 式的一般形式. 数据中观测到的价格 R 是 Q 是联立方程 (4) 和 (5) 的解, 因此是可观测的内生变量, 即 $Y_o = (R, Q)$; 本模型中没有不可观测的内生变量, 即 $\{Y_u\} = \emptyset$; 自动化程度 Z 是可观测的外生变量, 即 $X = Z$; 随机冲击 U 和 V 是不可观测的外生变量, 即 $\varepsilon = (U, V)$; 假设我们只关注需求的价格弹性 (elasticity), 那么 c 就是我们感兴趣的参数, 即 $\theta = c$; 而 a 和 b 都是干扰参数, 即 $\eta = (a, b)$; 函数 \mathcal{M} 的形式为

$$\mathcal{M}(Y_o, Y_u, X, \varepsilon; \theta, \eta) = \begin{bmatrix} \ln R - (a \ln Q + bZ + U) \\ \ln R - (c \ln Q + V) \end{bmatrix}.$$

外生变量 (Z, U, V) 的联合分布记为 G , 这里我们对 G 施加如下假定:

$$\mathbb{E}(ZV) = 0, \quad \mathbb{E}(ZU) \neq -b\mathbb{E}(Z^2). \quad (6)$$

这个简单的例子也说明, 一个计量经济学模型是不是结构方程模型, 不取决于模型形式的复杂程度, 而是取决于模型是否有经济学理论基础, 参数是否有经济学含义.

考虑均衡条件产品的需求量等于供给量, 求解联立方程 (4) 和 (5) 可得

$$\ln Q = -\frac{bZ}{a-c} + \frac{V-U}{a-c}, \quad (7)$$

$$\ln R = -\frac{bcZ}{a-c} + \frac{c(V-U)}{a-c} + V. \quad (8)$$

如果将 (7) 式和 (8) 式分别看作 $\ln Q$ 和 $\ln R$ 对 Z 的线性回归模型, 那么这两个回归就是约简式模型, 它仅仅刻画了可观测数据之间的相关关系, 其中的系数 $b/(c-a)$ 和 $bc/(c-a)$ 并不具有直接的经济学含义. 虽然回归方程 (8) 的系数 $bc/(c-a)$ 与回归方程 (7) 的系数 $b/(c-a)$ 之商是 c , 也就是需求价格弹性的倒数, 但是这一经济学含义也必须基于结构方程模型 (4)–(5) 才能得到.

根据 (7) 式和假定 (6) 可知

$$\mathbb{E}(Z \ln Q) = -\frac{b}{a-c} \mathbb{E}(Z^2) - \frac{1}{a-c} \mathbb{E}(ZU) \neq 0.$$

将模型 (5) 与假定 (6) 结合可得

$$0 = \mathbb{E}(ZV) = \mathbb{E}[Z(\ln R - c \ln Q)] = \mathbb{E}(Z \ln R) - c \mathbb{E}(Z \ln Q).$$

因为 $\mathbb{E}(Z \ln Q) \neq 0$, 所以

$$c = \frac{\mathbb{E}(Z \ln R)}{\mathbb{E}(Z \ln Q)}. \quad (9)$$

这样, (9) 式证明了当模型 (4)–(5) 满足假定 (6) 时, 参数 c 可识别, 并且给出了从数据分布到参数 c 的映射的表达式.

现在假定有一个独立同分布样本 $\{(R_i, Q_i, Z_i) : i = 1, \dots, n\}$, 其中每个观测 (R_i, Q_i, Z_i) 的分布都等于模型 (4)–(5) 在假定 (6) 之下蕴含的 (R, Q, Z) 的分布. 此外, 我们假定 $\text{Var}(Z_i V_i) = \omega^2 < \infty$. 基于 (9) 和样本模拟量方法, 我们可以得到参数 c 的估计量

$$\hat{c} = \left(\frac{1}{n} \sum_{i=1}^n Z_i \ln R_i \right) \Bigg/ \left(\frac{1}{n} \sum_{i=1}^n Z_i \ln Q_i \right).$$

利用大数定律和中心极限定理, 我们可以证明当 $n \rightarrow \infty$ 时, $\hat{c} \xrightarrow{\mathbb{P}} c$ 且 $\sqrt{n}(\hat{c} - c) \xrightarrow{d} \mathcal{N}(0, \nu^2)$, 其中

$$\nu^2 = \frac{\omega^2}{[\mathbb{E}(Z_i \ln Q_i)]^2}.$$

如果 $\mathbb{E}(Z^4)$ 有限, 那么渐近方差 ν^2 的一个相合估计量 (consistent estimator) 是

$$\hat{\nu}^2 = \left[\frac{1}{n} \sum_{i=1}^n Z_i^2 (\ln R_i - \hat{c} \ln Q_i)^2 \right] \Bigg/ \left(\frac{1}{n} \sum_{i=1}^n Z_i \ln Q_i \right)^2.$$

基于上述理论, 我们能够使用样本得到模型参数的估计量, 并在此基础上开展统计推断.

我们还可以寻找一个未参与模型估计的新样本 $\{(R'_i, Q'_i, Z'_i) : i = 1, \dots, m\}$, 用于验证模型设定和

参数估计量 \hat{c} 的准确性. 具体地, 如果模型设定 (4)–(5) 和假定条件 (6) 对于新样本也成立而且 \hat{c} 能够准确地估计 c , 那么统计量 \hat{E} 应当很接近于 0, 其中

$$\hat{E} = \frac{1}{m} \sum_{i=1}^m Z'_i (\ln R'_i - \hat{c} \ln Q'_i).$$

因此这一统计量可以用于模型的验证. 基于验证以后的模型, 我们可以忽略随机误差项, 通过积分求出需求曲线以下、均衡价格以上的部分的面积, 从而得到消费者剩余 (consumer surplus) 的度量. 这一分析可以拓展到数据中没有出现过的假想的均衡价格上 (例如价格管制或价格补贴的情形), 这就是反事实分析. \square

3 模型构建

虽然约简式研究和结构方程研究都需要以模型构建作为出发点, 但是构建结构方程模型的难度和挑战性远远高于约简式模型. 如前所述, 约简式模型是可观测数据的分布的模型, 因此在建模时我们只需要根据数据分布当中与研究问题相关的那些特征来选择适当的模型. 一方面, 构建模型所需要的很多分布特征可以直接或间接地从样本中获取, 例如对样本的描述性统计和数据挖掘; 另一方面, 统计学和计量经济学理论已经提供了大量可供选择的模型及对其适用场景的说明, 此外还有关于模型选择和模型设定检验的理论和方法. 这意味着, 约简式模型的构建可以遵循一套相对成熟且标准化的操作指南.

不同于约简式模型, 结构方程模型需要从研究问题的经济情境出发, 用经济学理论描述相关行为人的行为及其结果, 最终刻画可观测数据的分布. 这一过程包含如下难点. 第一, 我们的研究问题涉及的是现实中具体的经济行为和经济现象, 我们需要首先厘清这些行为和现象背后的经济学逻辑. 但是, 这些逻辑并不是显而易见的, 有时可能存在多套不同的逻辑, 它们都能够解释某一行为和现象. 第二, 在厘清逻辑的基础上, 我们需要对现实进行抽象化, 用一个经济学理论模型来描述我们希望研究的问题. 这就要求我们熟练掌握经济学建模方法, 并对现有的模型及其适用场景和特征有深入的理解. 有些问题在文献中已经存在多个理论模型, 我们需要根据问题本身的特征和不同模型的具体设定来选择最适合的模型. 有些问题无法被文献中现存的模型准确刻画, 我们需要对某个已有模型进行适当的调整和修改. 第三, 我们还需要对模型施加一定的假定条件, 例如效用函数的函数形式和随机扰动项的分布. 这些假定通常都不具有固定的模式, 而是需要我们根据文献惯例、经济事实、可操作性等要素来自行确定.

正是因为上述难点的存在, 结构方程建模目前并不存在 (可能将来也很难出现) 一套教科书式的操作指南, 而是需要研究者针对具体的研究问题构建最合适的模型. 具体而言, 给定一个研究问题, 研究者在构建模型时需要考虑如下要素.

1. **行为人及其相互关系.** 行为人 (agent) 是经济活动的参与者和经济行为的实施者, 包括自然人、家庭、厂商、银行、政府等, 它们的经济角色可能是消费者、生产者、资本提供者、金融中介、政策制定者等. 研究者在建立模型时需要确定模型中包含单个行为人还是多个行为人. 如果存在多个行为人, 研究者需要确定它们各自的角色和相互间的关系. 如果存在大量角色相同的行为人, 而且任何一个人的行为不会对其他人的行为产生影响, 那么这就是一个完全竞争的市场; 如果不同的行为人之间存在策略性互动, 即一个人的行为会受到其他人行为的影响, 那么这就是一个博弈的场景. 此外, 研究者还需

要考虑具有同一经济角色的不同行为人之间的异质性. 如果属于同一角色的不同行为人完全同质, 那么我们在建模时可以只设置一个代表性行为人. 如果属于同一角色的不同行为人具有异质性, 那么我们可能需要根据异质性来区分不同的行为人. 例如, 在 Dey and Flinn (2005) 中, 健康保险的覆盖状况会影响雇员的就业决策. 而如果将健康保险视作家庭层面的公共品, 考虑到家庭内部一方的工资与保险状态会影响另一方的劳动供给决策, 则模型可相应扩展为多人的搜索模型 (Dey and Flinn, 2008). 又如, 在市场结构设定方面, Drechsler et al. (2017) 等经验研究表明银行的存款市场是非完全竞争的. 基于这一事实, Chiu et al. (2023) 构建了一个寡头垄断的银行体系模型, 用以分析央行数字货币 (central bank digital currency) 如何通过为储户提供外部选择来影响银行的存款利率和信贷供给.

2. **时间.** 研究者需要确定模型刻画的经济活动发生在单个还是多个 (包括无穷个) 时间点. 如果模型涉及多期, 研究者需要确定时间计量单位 (例如日、月、年) 和计时连续性, 并明确各个经济活动的先后顺序. 多期模型还有一个非常本质的分类: 动态 (dynamic) 模型和静态 (static) 模型. 需要强调的是, 模型动态和静态的区分是由模型中经济行为和变量运动的特征决定的, 而不是由数据决定的. 如果经济行为和状态具有跨时期的依赖性, 即一期的行为和状态依赖于其它期的行为和状态, 那么就是动态模型; 否则就是静态模型. 例如在 Berry (1994), McFadden (1974) 和 Train (1999) 中, 虽然可以有多期数据, 但是消费者在每一期面临的环境相同, 只需选择能够最大化当期效用的产品或方案, 因此它是静态模型. 而在 Rust (1987) 模型中, 当期的发动机维修决策会影响未来的车辆行驶成本, 因而这是一个动态模型, 但是该模型的估计只需要使用维修车辆的行驶里程这一横截面数据. 在动态模型中, 我们需要设定不同时期的行为和状态的依赖关系的具体形式, 这被称为相应变量的运动法则 (law of motion), 它包括但不限于生成外生冲击的随机过程、资本的折旧过程、行为人的惰性和预期. 例如在 Attanasio et al. (2012) 的动态家庭决策模型中, 儿童入学与青少年劳动供给的选择的跨期关联: 虽然辍学务工能为家庭带来即时收入, 但接受教育不仅可以在当期提升家庭的效用, 也会提高未来的工资水平. 此外, 动态模型可以是有限期的 (例如 Keane and Wolpin, 1997; Nevo et al., 2016), 也可以是无限期的 (例如 Rust, 1987).

3. **随机性和信息.** 一个经济模型可以是完全确定的, 也可以包含随机变量. 这些随机性既可以是外生的, 例如随机的偏好冲击和技术冲击, 也可以是内生的, 例如博弈中的行为人选择混合策略 (mixed strategy). 如果一个多期的模型包含随机性, 那么研究者需要明确行为人的信息集 (information set) 随时间的演化过程. 一个时间点上的信息集里面包含了到这个时间点为止已经实现的随机变量, 它们的取值在此时是已知信息, 不再具有随机性. 需要注意的是, 不同的行为人可以具有不同的信息集, 这被称为信息不对称 (information asymmetry). 例如, Chiu et al. (2023) 中, 在有摩擦的分散市场阶段, 买方与卖方按照一定概率随机匹配; 由于买卖双方可能持有或接受不同的支付方式, 因此交易的达成具有随机性.

4. **选择集和策略集.** 选择集 (choice set) 是行为人在一次行动时所面临的所有可行选项的集合, 它可以是离散的也可以是连续的. 例如一个人选择是否进入劳动力市场, 那么选择集就是一个二元的离散集合; 一个人选择平均每日的工作时长, 那么选择集 (理论上) 就是 0 – 24 小时的连续统. 在很多模型中, 特别是包含随机性和动态性的模型中, 行为人会依据某些已知信息 (例如已实现的随机变量) 来决定自己的选择; 我们将这些已知信息称为状态变量 (state variable), 并把这样的决策过程称为策略

(strategy), 它由从状态变量的取值空间到选择集的映射构成. 例如一个人根据当前的工资水平是否超过某个阈值来决定是否进入劳动力市场, 那么其策略就是从 $[0, \infty)$ 到 $\{0, 1\}$ 的映射, 这被称为阈值策略 (cut-off strategy). 又如在完全信息的序贯博弈中, 每个参与者的策略包含了在自己所有的行动节点上的决策模式, 而每个行动节点上的决策模式又是从包含该节点所有历史的集合到下一步所有可行行动的集合的映射. 在约束条件之下所有可行的策略形成的集合被称为策略集, 它也是下面将要介绍的最优化问题的目标函数的定义域.

5. **决策目标.** 决策目标刻画了每个行为人行动的准则, 或者说他/她从策略集中选择最优策略的标准. 根据经济学理论, 决策目标通常表述为一个最优化问题, 而这一最优化问题的目标函数的定义域就是策略集. 一般而言, 消费者的决策目标是最大化效用, 生产者的决策目标是最大化利润. 如果模型含有随机性, 那么最优化的目标变为期望效用或利润. 如果模型是多期的, 那么最优化的目标变为各期(期望)效用或利润的贴现之和. 例如, Wang et al. (2022) 中银行的目标是最大化向股东支付的现金股利贴现和. 这里的多期模型还会衍生出一个新问题: 偏好是否具有时间一致性, 即行动者是否会随着时间的推移而偏离起始时期的最优策略路径. 在时间不一致偏好之下, 行为人在不同时期可能有不同的最优化目标, 研究者需要将其明确. 例如, Allcott et al. (2022) 在分析数字成瘾中, 假设用户在计划阶段与实际消费阶段的效用评估不一致: 当期的诱惑使其实际使用量超过了先前计划的最优水平. 最后, 在需要开展福利分析的时候, 研究者还需要设定社会计划者的决策目标, 以此作为社会最优的标准.

6. **均衡的定义.** 我们需要在具体的经济场景中给出均衡的具体定义. 在单个行为人的模型中, 均衡通常就是这个行为人的最优化问题的解, 即最优策略. 在包含生产者和消费者的局部均衡模型中, 竞争性均衡 (competitive equilibrium) 就是在生产者最大化利润和消费者最大化效用的策略之下使得供给量等于需求量的价格和这一价格下的销售量. 在静态一般均衡模型中, 瓦尔拉斯均衡 (Walrasian equilibrium) 由所有产品的价格、所有厂商的生产决策和所有家户的消费 (包括闲暇) 决策构成, 它们需要满足: 在禀赋和这组价格下每个家户的消费决策能够实现其效用最大化, 在这组价格下每个厂商的生产决策能够实现其利润最大化, 每个产品的总需求量都等于总供给量. 在动态随机一般均衡模型中, 理性预期均衡 (rational expectation equilibrium) 则是在瓦尔拉斯均衡的基础上, 让价格和决策都是状态变量的函数, 将家户和厂商的最优化目标改为期望效用和期望利润, 并要求家户和厂商最优决策形成的状态变量运动法则与其预期相等. 而在博弈模型中, 均衡是所有参与者的策略组合, 根据博弈场景的不同, 可以有不同的均衡概念, 例如纳什均衡 (Nash equilibrium)、子博弈完美均衡 (subgame perfect equilibrium)、贝叶斯纳什均衡 (Bayesian Nash equilibrium)、完美贝叶斯均衡 (perfect Bayesian equilibrium). 例如, Nevo et al. (2016) 通过一个局部均衡模型, 分析了用户在给定流量套餐和价格的条件下的上网流量的需求. 在 Wang et al. (2022) 中, 均衡存款利率与贷款利率由存款市场和贷款市场的逐期出清内生决定. 该文章在其定义 1 中将均衡定义为一组策略与分布的集合: 在该均衡下, 各银行在给定其他银行存贷款定价策略的条件下, 最优选择以最大化贴现后现金流; 家庭与企业在给定利率下分别最大化其效用与利润; 存款市场与贷款市场在每一期出清; 银行业状态分布的演化概率律与银行自身的最优决策保持一致. 虽然在不同场景中有不同的经济学含义, 但均衡在数学上的刻画是统一的, 即映射的不动点. 这种刻画方式也说明了, 一个经济学模型的均衡并不必然是唯一的, 而多重均衡的存在可能导致参数只能被部分识别, 我们将在第 4.4 小节讨论该问题.

7. 变量对于研究者的可观测性和变量的测量. 如果允许部分函数形式和分布是非参数设定的, 那么上述 6 类要素已经足以构成一个完整的经济学模型. 此时, 作为结构方程建模的重要步骤, 我们需要将经济学模型与观测数据匹配起来. 在经济学模型中, 可能存在一些变量可以被行为人观测到, 但无法被研究者观测到 (即没有出现在数据中), 例如行为人的某些异质性或不完全信息博弈中的私有信息. 对于这类变量, 通常的处理方式是将其视作随机变量, 假定其分布. 例如, 在 Berry et al. (1995) 中, 研究者无法观测到不同消费者偏好的异质性, 因此作者使用服从正态分布的随机系数来刻画消费者对于价格的反应. 对于那些可以在数据中被观测到的变量, 我们需要区分它是内生变量 (例如均衡结果) 还是外生变量 (例如初始状态变量), 并明确这些变量的测量方式, 因为这可能涉及到测量误差 (measurement error) 的问题.

8. 函数形式和分布假定. 一般而言, 为了确保模型的适用性, 我们需要尽可能少地对模型中的函数形式和随机变量分布施加假定, 但是这通常会导致模型的可操作性 (tractability) 下降, 例如最优解没有解析表达式或参数无法识别. 因此, 出于可操作性的考虑, 我们需要对函数形式和分布施加一些技术性假定. 例如在离散选择模型中, 如果假定随机效用函数中的误差项独立同分布地服从第一型极值分布, 那么条件选择概率就有解析表达式. 再如我们将在第 4 节中讨论的, 假定工具变量和误差项之间的正交性可以为参数识别提供帮助.

从上面的讨论中可以看出, 要想针对特定问题构建出合适的结构方程模型, 研究者需要充分了解相关的经济现实和经济理论, 并将二者有机结合起来. 在这一过程中, 约简式分析可以为结构方程建模提供帮助. 约简式分析能够反映可观测变量的边缘分布特征和它们之间存在的统计依赖关系及其形式, 这是现实经济的特征事实 (stylized fact) 的一部分. 这些特征事实可能蕴含了有关市场结构、行为人的激励、约束条件、变量特征和关系等方面的信息, 从而可以帮助我们确定构建结构方程模型所需的要素. 此外, 约简式分析还包括模型设定检验, 它可以用观测数据来证伪某个关于函数形式或分布的假设, 进而帮助我们排除一些错误的设定.

例如, Buchak (2024) 通过交错双重差分 (staggered difference in differences) 的结果发现, 在美国低收入的群体需要借助汽车贷款参与网约车市场. 因此, 在基于离散选择模型刻画网约车司机劳动供给时, 将购车决策、贷款决策纳入行为人的选择结构中. Nevo et al. (2016) 先通过回归分析验证了用户当前流量使用量越接近套餐额度时, 影子价格越高, 当期使用量越低. 这种前瞻性行为的经验证据促使作者构建了一个以 30 天账单为周期的动态模型. 在该设定下, 用户当天的最优上网量取决于剩余流量额度和套餐到期时间.

机器学习也能够为结构方程的模型构建提供帮助. 首先, 结构方程的模型构建一定程度上也可以被视为模型选择 (model selection) 问题. 机器学习领域已经有一套成熟的数据驱动的模型选择方法, 包括赤池信息准则 (Akaike information criterion)、贝叶斯信息准则 (Bayesian information criterion)、最小描述长度 (minimum description length)、结构风险最小化 (structural risk minimization) 等解析方法和交叉验证 (cross validation)、自助法 (bootstrap) 等样本再利用方法, 详见 Hastie et al. (2009) 第 7 章. 在结构方程研究中, 我们也可以借鉴这些方法, 充分挖掘数据中包含的信息, 以帮助我们寻找更加合理的模型设定. 其次, 结构方程模型中的一些变量可能只有理论上的定义, 但在现实中没有对应的测量数据, 例如劳动力市场中某个职位对某项技能的依赖程度. 机器学习中的生成式算法, 特别是当

前非常火热的以大语言模型为代表的生成式人工智能, 为这类变量的测量提供了新的可能. 研究者可以根据研究需要, 提供相应的基础数据和指令, 让生成式模型生成这些变量的测量指标. 再次, 在面对大量(高维度)经济数据时, 我们还可以考虑使用聚类(clustering)和主成分分析(principal component analysis)等无监督学习(unsupervised learning)或者生成式人工智能来从这些数据中提取我们所需的特征事实, 用于帮助我们选择模型.

例如, Handa et al. (2025) 借助大语言模型, 将海量的非结构化的用户和对话数据映射到 O*NET 数据库中职业与任务信息, 从而刻画不同职业对生成式人工智能的暴露程度. 类似的研究还包括从企业招聘数据(Hampole et al., 2025) 和统计局劳动力报告(Eloundou et al., 2024) 中构建人工智能暴露程度指标.

4 参数识别

4.1 可识别性

第 2 节已经介绍过, 参数识别本质上是构建从可观测数据的联合分布到参数的映射. 从直观上看, 要保证这样的映射存在, 不同参数取值之下的数据分布应当不同. 否则, 不同的参数取值对应同样的数据分布, 那么我们就无法区分是哪一组参数取值生成了当前的数据分布. 参照 Matzkin (2007) 定义 3.2, 我们下面给出参数可识别性的严格定义. 在结构方程模型(1)中, (Y_o, X) 是可观测数据, 其联合分布用 $P_s(\theta, \eta, G)$ 表示. 设 \mathcal{M} 是模型空间, 即与模型设定相容的所有 (θ, η, G) 的集合, 那么这一模型空间可以生成的所有分布记为

$$P_s(\mathcal{M}) = \{P_s(\theta, \eta, G) : (\theta, \eta, G) \in \mathcal{M}\}.$$

正如第 2 节所定义的, $\psi = \Psi(\theta, \eta, G)$ 是我们根据研究问题和目的确定的希望识别的参数, 那么参数空间为

$$\Psi(\mathcal{M}) = \{\Psi(\theta, \eta, G) : (\theta, \eta, G) \in \mathcal{M}\}.$$

简单起见, 这里仅考虑 ψ 是有限维的情形, 即 $\Psi(\mathcal{M}) \subset \mathbb{R}^K$. 定义从参数空间 $\Psi(\mathcal{M})$ 到分布空间 $P_s(\mathcal{M})$ 的对应 (correspondence) \mathcal{P} , 使得对于每个 $\psi \in \Psi(\mathcal{M})$,

$$\mathcal{P}(\psi, \mathcal{M}) = \{P_s(\theta, \eta, G) \in P_s(\mathcal{M}) : \Psi(\theta, \eta, G) = \psi\},$$

即 $\mathcal{P}(\psi, \mathcal{M})$ 表示参数 ψ 之下所有可能的数据分布的集合. 举个简单的例子, 设 \mathcal{M} 是所有形如

$$Y_o = X^T \psi + \varepsilon, \quad \mathbb{E}(\varepsilon | X) = 0$$

的线性回归模型形成的集合, 那么对于任意给定的参数 ψ_0 , 对应 $\mathcal{P}(\psi_0, \mathcal{M})$ 中包含了所有满足 $\mathbb{E}(Y_o | X) = X^T \psi_0$ 这一性质的 (Y_o, X) 的联合分布.

定义 1: 参数 $\psi \in \Psi(\mathcal{M})$ 在模型空间 \mathcal{M} 中可识别, 当且仅当对任意 $\psi' \in \Psi(\mathcal{M}) \setminus \{\psi\}$, 都有

$$\mathcal{P}(\psi, \mathcal{M}) \cap \mathcal{P}(\psi', \mathcal{M}) = \emptyset.$$

根据定义 1, 在上述线性回归模型的例子中, 一个给定的参数 ψ 可以被识别, 当且仅当对于任意 $\psi' \neq \psi$, 必有 $\mathbb{P}(X^\top \psi \neq X^\top \psi') > 0$. 注意到, 定义 1 给出的是单组固定参数的可识别性. 理论上, 只要生成数据的真实参数满足定义 1, 我们就可以用该数据的分布信息来识别这一真实参数. 但是, 在实际中我们事先并不知道参数空间 $\Psi(\mathcal{M})$ 中的哪一个元素是真实参数, 所以为了确保参数的可识别性, 我们要求定义 1 对一切 $\psi \in \Psi(\mathcal{M})$ 成立, 这一条件也被称为全局识别 (global identification, 参见 Lewbel, 2019 第 842 页的讨论). 换言之, 全局识别要求: 对于任意 $\psi, \psi' \in \Psi(\mathcal{M})$, 只要 $\psi \neq \psi'$, 就有 $\mathcal{P}(\psi, \mathcal{M}) \cap \mathcal{P}(\psi', \mathcal{M}) = \emptyset$. 在上面的线性回归模型的例子中, 参数的全局识别要求对于任意 $\psi, \psi' \in \Psi(\mathcal{M})$, 只要 $\psi \neq \psi'$, 就有 $\mathbb{P}(X^\top \psi \neq X^\top \psi') > 0$, 这事实上等价于矩阵 $\mathbb{E}(XX^\top)$ 满秩.

例 1 (续): 按照定义 1, 结构方程模型 (4)–(6) 中的供给曲线的参数 a 和 b 是不可识别的. 假定需求曲线 (5) 和外生变量 (Z, U, V) 的分布 G 不变. 任取 $\lambda \in (0, 1)$, 构造新的供给曲线如下:

$$\ln R = a' \ln Q + b' Z + U', \quad (10)$$

其中 $a' = \lambda a + (1 - \lambda)c$, $b' = \lambda b$, $U' = \lambda U + (1 - \lambda)$. 联立方程 (5) 和 (10) 可得

$$\begin{aligned} \ln Q &= \frac{V - U' - b' Z}{a' - c} \\ &= \frac{V - \lambda U - (1 - \lambda)V - \lambda b Z}{\lambda a + (1 - \lambda)c - c} \\ &= \frac{\lambda(V - U) - \lambda b Z}{\lambda(a - c)} \\ &= -\frac{b Z}{a - c} + \frac{V - U}{a - c}. \end{aligned}$$

这说明均衡数量与 (7) 相同. 将均衡数量代入原需求曲线 (5) 可知均衡价格也与 (8) 相同. 在上述操作中, 我们将模型参数从 (a, b, c) 改为 (a', b', c) , 但这并不改变可观测数据 (R, Q, Z) 的联合分布. 因此参数 (a, b) 不可识别. Lewbel (2019) 第 852 页也有类似的讨论. \square

4.2 完全信息和有限信息

容易看出, 定义 1 比较抽象, 在不施加额外限定条件的情形下, 我们很难直接按照定义 1 去证明具体模型中参数的可识别性. 此外, 在参数可以被识别的情形下, 定义 1 本身也不能够帮助我们构建出从数据分布 P 到参数 ψ 的映射, 也就难以进一步的参数估计提供帮助. 因此, 结构计量经济学理论提供了一些更具可操作性的识别方法 (见 Newey and McFadden, 1994 第 2.2 小节), 我们将其介绍如下. 按照集合 $\mathcal{P}(\psi, \mathcal{M})$ 是否为单点集, 这些识别方法可以分为完全信息 (full information) 和有限信息 (limited information) 两大类.

完全信息方法需要利用数据分布的全部信息, 因此要求模型是全参数模型, 即对任意 $\psi \in \Psi(\mathcal{M})$, 集合 $\mathcal{P}(\psi, \mathcal{M})$ 都是单点集. 在全参数模型中, 参数 ψ 能够完全决定数据的分布, 从参数空间到分布空间的对应 \mathcal{P} 退化为映射. 按照定义 1, 如果映射 $\psi \mapsto \mathcal{P}(\psi, \mathcal{M})$ 是单射, 那么参数 ψ 就是可识别的. 用 $\ln f(Y_o, X; \psi)$ 表示对数似然函数 (log-likelihood function), 其中 $f(\cdot, \cdot; \psi)$ 是分布 $\mathcal{P}(\psi, \mathcal{M})$ 的密度函数. 在参数可识别的情形下, 我们可以证明真实参数 ψ_0 是期望对数似然函数的唯一最大值点, 并由此

得到如下隐式的识别结果:

$$\psi_0 = \arg \max_{\psi' \in \Psi(\mathcal{M})} \mathbb{E} [\ln f(Y_o, X; \psi')] . \quad (11)$$

这一识别方法对应的估计量是极大似然估计量.

有限信息方法只利用数据分布的部分信息, 即数据分布的一些泛函 (functional), 因此不要求模型是全参数设定的, 也就是允许集合 $\mathcal{P}(\psi, \mathcal{M})$ 含有多个分布. 在这样的模型中, 参数 ψ 虽然不能完全决定数据的分布, 但是可以决定数据分布的某些泛函. 有限信息方法的核心就在于利用这些泛函与参数之间的依赖关系来识别参数. 令 \mathcal{F} 是以 $P_s(\mathcal{M}) \times \Psi(\mathcal{M})$ 为定义域的映射, 它满足对于任意 $\psi \in \Psi(\mathcal{M})$, 都有 $\mathcal{F}(\mathbf{P}; \psi) = 0$ 对一切 $\mathbf{P} \in \mathcal{P}(\psi; \mathcal{M})$ 成立. 换言之, 等式约束 $\mathcal{F}(\mathbf{P}; \psi) = 0$ 刻画了参数 ψ 和与它相容的分布的特征之间的关系. 我们可以证明, 如果对于任意 $\mathbf{P} \in P_s(\mathcal{M})$, 有且仅有唯一的 $\psi \in \Psi(\mathcal{M})$ 使得 $\mathcal{F}(\mathbf{P}; \psi) = 0$ 成立, 那么参数 ψ 在该模型中可以被识别. 在实际操作中, \mathcal{F} 有更加具体的形式, 由此得到如下三类有限信息识别方法.

第一类方法利用了数据的条件期望与参数之间的关系. 不失一般性, 我们将这里的条件期望限定为 Y_o 关于 X 的条件期望, 并假设 Y_o 是一维随机变量. 这一类方法要求参数 ψ 能够决定 Y_o 关于 X 的条件期望, 即存在某个已知函数 $m : \mathcal{X} \times \Psi(\mathcal{M}) \rightarrow \mathcal{Y}_o$, 使得只要数据分布 $\mathbf{P} \in \mathcal{P}(\psi, \mathcal{M})$, 就有 $\mathbb{E}(Y_o|X) = m(X; \psi)$. 在这种情形下, 如果对于任意 $\mathbf{P} \in P_s(\mathcal{M})$, 有且仅有唯一的 $\psi \in \Psi(\mathcal{M})$ 使得 $\mathbb{E}(Y_o|X) = m(X; \psi)$, 那么参数 ψ 就可以被识别. 根据条件期望的投影性质 (条件期望能够最小化误差方差), 真实参数 ψ_0 有如下隐式表示:

$$\psi_0 = \arg \min_{\psi' \in \Psi(\mathcal{M})} \mathbb{E} \left[(Y_o - m(X; \psi'))^2 \right] . \quad (12)$$

这一识别方法对应的估计量是最小二乘估计量.

第二类方法利用了总体矩与参数之间的关系. 这一类方法要求存在某个已知函数 $g : \mathcal{Y}_o \times \mathcal{X} \times \Psi(\mathcal{M}) \rightarrow \mathbb{R}^L$, 使得只要数据分布 $\mathbf{P} \in \mathcal{P}(\psi, \mathcal{M})$, 就有 $\mathbb{E}[g(Y_o, X; \psi)] = 0$. 在这种情形下, 如果对于任意 $\mathbf{P} \in P_s(\mathcal{M})$, 有且仅有唯一的 $\psi \in \Psi(\mathcal{M})$ 使得 $\mathbb{E}[g(Y_o, X; \psi)] = 0$, 那么参数 ψ 就可以被识别, 而且真实参数 ψ_0 可以隐式地表示为总体矩方程的唯一解:

$$\psi_0 = \arg \text{zero}_{\psi' \in \Psi(\mathcal{M})} \mathbb{E}[g(Y_o, X; \psi')] . \quad (13)$$

上式中的 $\arg \text{zero}$ 表示函数的零点. 这一识别方法对应的估计量是 (广义) 矩估计量.

第三类方法利用了分布的泛函与参数之间的关系. 这一类方法要求存在某个已知或可构造的泛函 $\mathcal{H} : P_s(\mathcal{M}) \rightarrow \mathbb{R}^L$ 和某个 (可能未知的) 函数 $h : \Psi(\mathcal{M}) \rightarrow \mathbb{R}^L$, 使得只要数据分布 $\mathbf{P} \in \mathcal{P}(\psi, \mathcal{M})$, 就有 $\mathcal{H}(\mathbf{P}) = h(\psi)$. 在这种情形下, 如果函数 h 是单射, 那么参数 ψ 就可以被识别, 而且真实参数 ψ_0 有如下隐式表示:

$$\psi_0 = \arg \text{zero}_{\psi' \in \Psi(\mathcal{M})} [\mathcal{H}(\mathbf{P}) - h(\psi')] . \quad (14)$$

注意到, 这类方法与第二类方法有以下两点区别: 第一, 分布的泛函是一个比总体矩更加一般的概念, 因为总体矩只是分布的线性泛函; 第二, 这里的分布的泛函 \mathcal{H} 不直接以参数为自变量, 而第二类方法中的总体矩的自变量中包括参数. 这一识别方法对应的估计量是经典最小距离估计量.

4.3 现有方法及其不足

上文给出的表达式 (11)–(14) 都只是形式上的识别结果, 对于具体的模型, 研究者需要完成构造识别约束 (identification restriction) 和证明可识别性这两个步骤. 构造识别约束指的是在前述各个方法中推导似然函数、推导条件期望函数、构造总体矩方程或者寻找分布的泛函, 而证明可识别性指的是证明从参数到似然函数是单射、证明条件期望方程有唯一解、证明总体矩方程有唯一解或者证明分布的泛函是参数的单射. 我们先介绍文献中几种常见的构造识别约束的方法.

在离散选择模型中, 内生变量 Y_o 只有有限个取值, 不妨记为 $\{1, \dots, J\}$, 因此 Y_o 关于外生变量 X 的条件分布可以完全地由分布列 $\{\mathbb{P}(Y_o = j|X; \psi) : j = 1, \dots, J\}$ 刻画, 我们不妨将这一分布列称为条件选择概率. 在这类模型中, 研究者可以借助条件选择概率得到似然函数, 并以此作为识别约束. 例如 Berry (1994), McFadden (1974) 和 Train (1999) 分别研究了用户对差异化产品、交通工具和钓鱼点的选择行为; Todd and Wolpin (2006) 和 Attanasio et al. (2012) 用动态模型分析了儿童入学与青少年劳动供给的家庭决策. 值得注意的是, 如果我们定义一组虚拟变量 $Y_o^{(j)} = \mathbb{1}\{Y_o = j\}$, 那么 $\mathbb{P}(Y_o = j|X; \psi) = \mathbb{E}(Y_o^{(j)}|X; \psi)$, 即 j 的条件选择概率等于 $Y_o^{(j)}$ 的条件期望. 因此在离散选择模型中, 我们也可以使用条件期望方程来作为识别约束.

在结构方程模型 (1) 中, 我们允许外生变量 (X, ε) 的各个分量之间存在相依性. 一些模型会额外假定可观测外生变量 X 的某个子向量 Z 和不可观测外生变量 ε 的某个子向量 δ 满足均值独立性或正交性条件, 即 $\mathbb{E}(\delta|Z) = 0$ 或者 $\mathbb{E}(Z\delta) = 0$. 此时也称变量 Z 关于扰动项 δ 具有外生性, 这有别于“(X, ε 是外生变量”这句话中所指的关于模型的外生性. 如果我们能在模型中将扰动项 δ 表示为可观测数据和参数的函数, 即 $\delta = \delta(Y_o, X; \psi)$, 那么 Z 和 δ 之间的正交性就蕴含了如下的矩方程:

$$\mathbb{E}[Z\delta(Y_o, X; \psi)] = 0, \quad (15)$$

它可以用作识别约束. 此时我们也可以把变量 Z 称作工具变量. 在 BLP 需求模型 (Berry et al., 1995) 中, 作者就使用供给端的冲击作为工具变量, 利用它和偏好冲击的正交性得到形如 (15) 的矩方程. Davis et al. (2014) 在研究城市集聚效应与内生增长时, 使用基于过往信息预测的土地租金作为工具变量, 利用工具变量和全要素生产率冲击的正交性, 得到了如上矩方程; Xiao (2020) 研究了影子银行对货币政策传导的影响, 在估计银行存款的需求参数时, 使用银行员工工资、固定资产开支、管理费用等成本变动作为工具变量, 利用其和不可观测的需求冲击的正交性, 得到了如上矩方程.

对于经验研究而言, 由于变量 Z 和 δ 都具有实际的经济含义, 因而研究者需要论证正交性条件 $\mathbb{E}(Z\delta) = 0$ 为何成立. 经济学理论和生活常识可以一定程度上解决这个问题, 例如供给端的冲击与偏好冲击应当不相关. 此外, 正如 Galiani and Pantano (2021) 第 5 节和 Todd and Wolpin (2023) 所介绍的, 目前结构方程领域的一个进展方向是与随机实验和准实验设计相结合. 在 (准) 实验中, 研究者可以通过随机控制等方法确保某些状态变量与模型中的另一些变量不相关甚至独立. 这样一来, 这些状态变量就可以用作工具变量, 使得矩方程 (15) 成立. 例如, Blundell et al. (2016) 假定不同出生世代个体的偏好保持不变, 其对英国税收与福利制度改革的差异化反应源于模型中的待估参数, 并通过教育选择、劳动供给以及资产和收入分布等矩条件加以识别.

上面介绍的几种构造识别约束的方法即使不需要似然函数、条件期望方程和矩方程的解析表达

式, 也要求研究者能以某种可操作的方式刻画这些函数或方程. 然而, 对于复杂的结构方程模型, 特别是具有随机性、动态性、互动性和异质性的模型, 这一要求不必然能够满足. 为此, 研究者选择从数据分布入手, 构造分布的一些泛函 \mathcal{H} , 如果能够利用经济学理论定性地得出这些泛函和参数之间的依赖关系, 那么就可以得到识别约束 $\mathcal{H}(\mathbf{P}) = h(\psi)$. 这一做法的好处是, 构造识别约束的过程本身不要求已知 h 的具体形式, 只需要论证 h 是非平凡的 (non-trivial), 即说明泛函 $\mathcal{H}(\mathbf{P})$ 的取值会受到参数 ψ 的影响, 因而大大降低了操作难度. 当然, 这一做法的代价就是提高了证明可识别性 (即证明 h 是单射) 和构建估计量的难度, 这一点将在下文展开讨论.

在经验研究中, 研究者通常基于经济学逻辑和理论来讨论某个 (些) 参数对某个 (些) 分布泛函的影响方向, 以此说明 h 非平凡. Wang et al. (2022) 选择了对模型基础参数具有敏感性的 10 个矩并定性讨论了它们与参数的关系. 例如, 当融资成本上升, 银行会更多地依赖存款而非借款发放贷款; 因此平均非准备金借款占存款的比例与银行持有非准备金的成本之间存在单调函数关系. 银行的固定运营成本则会影响两个矩: 平均净非利息费用占资产的比率和银行的平均杠杆率, 因为前者刻画了银行在常规映存贷款业务之外的运营成本, 后者中银行倾向于在更高的固定成本下以较低杠杆运作. Voena (2015) 将结构参数与辅助参数 (即分布的泛函) 联系起来. 例如, 妻子劳动参与率对家庭帕累托权重 θ 呈单调反应. 当丈夫拥有更大的决策权时, 单方离婚政策会改善家庭内对妻子的资源分配, 进而降低其劳动参与率. 因此, 政策实施前后劳动参与率的变化与家庭的帕累托权重之间存在依赖关系. Chen et al. (2023) 注意到在公司投资决策中, 固定成本投入的上限越大, 意味着投资在企业决策中越重要. 对应地, 进行一次性大额投入而非分笔投入的企业比例更高, 投资的序列相关性则下降. 由此建立了这些数据特征 (分布泛函) 与企业动态投资决策模型中的投资不可逆性参数之间的定性关系.

在构造了识别约束之后, 为了保证这些识别约束能够正确识别出参数, 我们必须证明可识别性. 这是一个具有挑战性的难点, 因为它本质上是对识别约束的解的唯一性的论证, 这涉及到对识别约束中的函数 $f, m, \mathbb{E}(g)$ 和 h 的性质的分析. 在一个结构方程模型中, 我们想要识别的参数通常是多维向量. 除了线性模型 (例如例 1) 这类极为特殊的模型之外, 识别约束中的函数 $f, m, \mathbb{E}(g)$ 和 h 都是关于参数的非线性多元函数, 而且大多数情形下没有解析表达式. 例如 Dey and Flinn (2005) 考虑健康保险和个体就业, 利用搜寻匹配议价模型, 构建了包含失业与在职搜索到达率、工资和健康保险分布等结构参数的非线性似然函数; Dey and Flinn (2008) 将该模型扩展至家庭层面, 分析了存在家庭成员间保险覆盖的多人决策, 而家庭内部的联合状态依赖与连续时间设定使得难以写出似然函数. 类似的例子还包括 Low and Pistaferri (2015) 对伤残保险与劳动供给的动态决策研究和 Adda and Dustmann (2023) 对不同职业工资增长来源的分析. 在这种情况下, 要证明识别约束的解的唯一性是较为困难的.

文献中已有的严格证明结构方程模型参数可识别性的结论绝大多数都是具体问题具体分析, 即针对特定的结构方程模型在相应的技术性假定之下证明了结构元素的可识别性. 之所以这里使用了“结构元素”而非“结构参数”的说法, 是因为文献中的很多结果都是非参数识别 (non-parametric identification). 注意到, 如果一个结构元素 (比如效用函数和生产函数) 能够被非参数识别, 那么将这些结构元素做适当的参数化 (parameterization) 之后, 其中的参数也必然能够被识别. 因此这些非参数识别结果可以看做比参数识别更强的结论. 但是, 这一性质仅在识别对象是结构元素时才成立, 一个不具有经济学含义的分布或函数能被非参数识别并不必然保证结构参数能够被识别. 例如在识别约束

$\mathbb{E}(Y_o|X) = m(X; \psi)$ 中, Y_o 关于 X 的条件期望函数显然可以被数据分布非参数地识别, 但是这并不能保证参数 ψ 的可识别性.

一些代表性的识别结果介绍如下: Guerre et al. (2000) 证明了在具有独立私人价值的第一价格密封拍卖当中, 出价者的私人价值的分布可以被非参数地识别; Magnac and Thesmar (2002) 证明了离散时间的动态离散决策过程 (包含 Rust, 1987 模型及其推广) 中的效用函数和贴现因子的可识别性; Berry and Haile (2014) 证明了在只有市场层面的数据时, 差异化产品市场的离散选择模型 (包含 BLP 需求模型及其推广) 中的需求、消费者福利变化、边际成本和成本冲击都可以被非参数地识别; Arcidiacono et al. (2016) 在连续时间的动态离散选择和博弈模型中证明了转移强度矩阵和效用函数的可识别性; Fox and Gandhi (2016) 证明了多元离散选择模型中刻画个体不可观测异质性的随机系数的分布可以被非参数地识别; Ackerberg et al. (2015) 研究了一些生产函数估计方法 (Olley and Pakes, 1996; Levinsohn and Petrin, 2003) 的识别性质; Gandhi et al. (2020) 给出了总产出生产函数的非参数识别结果. 除此之外, Matzkin (2013) 总结了更一般的一类模型的非参数识别性质, 包括形如 $Y = m(X, \varepsilon)$ 的外生解释变量模型, 形如 $Y_1 = m_1(Y_2, \varepsilon_1)$, $Y_2 = m_2(X_2, \varepsilon_2)$ 的内生解释变量递归模型, 以及形如 $Y_1 = m_1(Y_2, X_1, \varepsilon_1)$, $Y_2 = m_2(Y_1, X_2, \varepsilon_2)$ 的联立方程模型.

需要再次强调的是, 上述识别结果都或多或少地依赖一些技术性假定, 而很多技术性假定是难以甚至无法用数据来检验的, 这在一定程度上也限制了经验研究使用这些理论结果. 例如, Matzkin (2013) 讨论的联立方程模型识别包含了 Newey and Powell (2003) 研究的非参数工具变量模型, Newey and Powell (2003) 证明了结构方程中的条件期望函数可以被非参数地识别, 当且仅当内生变量是内生变量关于工具变量的条件分布的完全统计量 (complete statistic). 这一充分必要条件涉及到了条件分布的复杂特性, 因而在经验研究中难以被检验.

由于上述客观困难的存在, 大部分结构方程范式的经验研究只对可识别性进行非正式的讨论, 以启发式的方式说明参数与数据特征之间的关系. 上文介绍构建识别约束 $\mathcal{H}(\mathbf{P}) = h(\psi)$ 的做法时列举的研究就是这一类的例证. 其它例子还包括, Altonji et al. (2005) 将识别来源归因于排他性约束和函数形式假设; DellaVigna et al. (2012) 认为虽然参数是联合估计的, 但也能识别单个参数; Goettler and Gordon (2011) 指出需求参数和供给参数识别所对应的矩条件; Beraja et al. (2019) 认为再融资弹性通过经济衰退期数据识别; Fu and Gregory (2019) 认为不可观测的家庭收益参数由断点回归系数识别.

除此之外, 计量经济学理论文献中也出现了一些关于参数可识别性的假设检验方法, 但是它们难以用于检验复杂结构方程模型中的参数可识别性. 例如, Wright (2003) 通过检验总体矩函数关于参数的 Jacobi 矩阵不满秩来检验广义矩方法的识别不足; 该方法需要使用总体矩函数关于参数的 Jacobi 矩阵的解析表达式来构造统计量, 但在复杂的结构方程模型中, 总体矩函数本身就没有解析表达式, 遑论它的 Jacobi 矩阵; 此外, 因为 Jacobi 矩阵只能刻画总体矩函数在给定参数处的局部性质, 所以该检验方法只能检验局部的识别性, 即使它拒绝了所有 (局部) 识别不足的零假设, 也并不能保证矩方程有唯一解. 再如, Forneron (2024) 构造了矩函数关于参数的拟 Jacobi 矩阵, 并用它的特征值来检验矩条件模型的全局和局部识别. 但是, 该方法在计算拟 Jacobi 矩阵时需要首先计算参数的水平集. 我们在第 5 节中将会详细介绍, 在复杂的结构方程模型中, 计算广义矩方法估计量的目标函数和寻找其 (局部) 最小值已经相当困难, 因此计算参数的水平集 (它并不必然是连通集) 是一个更加困难的任务.

在本小节的最后, 我们简要地将约简式模型的参数识别与结构方程模型的参数识别做比较. 事实上, 约简式模型的参数识别也需要构造识别约束和证明可识别性这两个步骤. 一方面, 约简式模型的识别约束主要来自于行为假定, 正如我们在第 2 节介绍的, 这些行为假定通常不具有明确的经济学含义, 因而在约简式模型中论证识别约束的合理性要比在结构方程中困难. 另一方面, 约简式模型的函数形式通常比较简单, 因此在识别约束成立的情形下, 可识别性的证明较为简单, 有时甚至是显然的. 以 (15) 式的工具变量为例, 约简式模型通常会设定一个线性回归模型, 即 $\delta(Y_o, X; \psi) = Y_o - X^\top \psi$, 在这个模型中, 变量 δ 的经济学含义是不够清晰的, 所以要论证 Z 和 δ 之间的正交性 (也就是识别约束) 会比较困难. 但是, 得益于线性模型的简单形式, 在识别约束之下, 我们只需要确保矩阵 $\mathbb{E}(ZX^\top)$ 列满秩 (这一条件可以直接用数据检验), 就能够非常容易地用构造的方式证明参数 ψ 的可识别性. 相对应地, 在结构方程模型中, δ 必须有具体的含义 (如偏好的冲击), 因此它关于可观测变量和参数的函数形式 $\delta(Y_o, X; \psi)$ 通常是非线性且复杂的. 变量 δ 的明确含义使得我们可以相对容易地借助经济学理论来论证识别约束的合理性, 但是 $\delta(Y_o, X; \psi)$ 的复杂形式给在识别约束下严格证明参数可识别性带来了巨大困难.

4.4 部分识别

到目前为止, 我们讨论的识别都是点识别 (point identification), 也就是用数据分布唯一地确定参数. 上文介绍的参数识别的困难, 一定程度上来源于定义中的唯一性. 只要同一个数据分布可以由多个不同的参数生成, 那么参数就不可识别. 但参数不可点识别并不意味着数据分布完全不含有关于参数的任何有用信息. 事实上, 只要与一个数据分布相容的参数集合不等于整个参数空间, 那么这个数据分布就能确定参数集合的一个真子集, 此时称参数可以被部分识别. 我们可以将部分识别 (partial identification) 表示为分布空间到参数空间的一个对应 \mathcal{S} , 使得对每个数据分布 $P \in P_s(\mathcal{M})$:

$$\mathcal{S}(P) = \{\psi \in \Psi(\mathcal{M}) : P \in \mathcal{P}(\psi, \mathcal{M})\},$$

即 $\mathcal{S}(P)$ 是所有能够生成数据分布 P 的参数形成的集合. 当 $\mathcal{S}(P)$ 是单点集时, 就得到了点识别的结果.

文献中已有一些研究分析了几类特定的模型, 得到了部分识别的结果. 在一部分模型中, 由于数据不完整、内生性或多重均衡的存在, 参数无法被点识别, 因此研究者通过构造矩不等式等方法, 推导了参数的识别集合. 具体的例子包括含有区间取值协变量的二元选择模型 (Manski and Tamer, 2002), 含有内生解释变量的离散选择模型 (Chesher et al., 2013), 选择集存在不可观测异质性的离散选择模型 (Barseghyan et al., 2021), 涉及反事实预测的离散选择模型 (Manski, 2007), 具有多重均衡的静态同时行动有限博弈 (Tamer, 2003; Ciliberto and Tamer, 2009), 独立私人价值的拍卖模型 (Haile and Tamer, 2003), 多重独立社会网络的形成模型 (Sheng, 2020), 单个社会网络的形成模型 (de Paula et al., 2018). 对此更为详尽的介绍可以参见 Molinari (2020) 第 3 节. 除此之外, 一些经济学理论直接蕴含了不等式形式的识别约束, 例如研究者可以利用显示偏好 (revealed preference) 原理得到刻画行为人偏好的不等式组, 进而部分地识别行为人的偏好 (Afriat, 1967; Varian, 1982, 1983; Crawford and De Rock, 2014). 这一方法被应用于分析美国处方药保险市场中的药物品牌效应与保险选择 (Abaluck and Gruber, 2011; Ketcham et al., 2016; Abaluck and Gruber, 2016a,b).

5 参数估计

5.1 一般方法和难点

现在假定有一个从总体分布 $P \in \mathcal{P}(\psi_0, \mathcal{M})$ 中抽取的独立同分布样本 $\{(Y_{o,i}, X_i) : i = 1, \dots, n\}$, 即每个观测 $(Y_{o,i}, X_i)$ 都服从分布 P , 且它们相互独立. 这里假定样本独立同分布只是为了形式简洁, 本节的讨论对于满足一定条件的相依样本和/或非同分布样本也成立. 第 4 节介绍的参数识别是在总体中构建数据分布 P 到真实参数 ψ_0 的映射, 这是理论上的分析, 因为我们在现实中并不知道真实的数据分布, 而是只有从总体分布中抽取的样本, 它包含了我们所能获得的关于总体分布的全部信息. 参数估计就是将样本映射为参数估计量, 将其作为真实参数的近似, 以使我们获得关于真实参数的信息. 因为参数识别给出了真实参数与总体分布的依赖关系, 而样本又是总体分布在现实中的代表, 因此我们需要利用参数识别的结果来构建估计量.

在第 4.2 小节介绍的识别方法中, 如果优化问题 (11)–(12) 或方程 (13)–(14) 有解析解, 那么我们就能直接将真实参数表示为总体分布的一个泛函, 即 $\psi_0 = \mathcal{I}(P)$. 这种情形就是 Matzkin (2013) 定义的构造性识别 (constructive identification). 在参数可以被构造性识别且 \mathcal{I} 具有一些良好性质的情况下, 我们只需将总体分布 P 替换为样本的经验分布 \hat{P}_n , 即可得到参数估计量 $\hat{\psi} = \mathcal{I}(\hat{P}_n)$. 然而, 在一般形式的计量经济学模型 (当然也包括结构方程模型) 中, \mathcal{I} 不具有解析表达式, 我们也就无法基于构造性识别的结果得到上述估计量. 因此我们需要直接从第 4.2 小节讨论的 4 种识别方法的识别约束出发, 通过这些识别约束的样本模拟量 (sample analogue) 来构建相应的估计量. 这 4 种识别方法分别对应了如下 4 种估计量, 它们也是计量经济学中最常用的估计方法.

(1) 极大似然 (maximum likelihood, ML) 估计量. 将 (11) 式最大化问题的目标函数替换为其样本模拟量, 也就是将期望替换为样本均值, 即可构建出极大似然估计量:

$$\hat{\psi}_{\text{ML}} = \arg \max_{\psi' \in \Psi(\mathcal{M})} \frac{1}{n} \sum_{i=1}^n \ln f(Y_{o,i}, X_i; \psi'). \quad (16)$$

从直观上看, 样本似然函数衡量了在给定参数之下出现该样本的可能性, 而极大似然估计量就是找到一个参数以最大化出现当前样本的可能性. 与识别结果 (11) 对应地, 极大似然估计量是一种完全信息估计量, 因为似然函数包含了数据分布的完整信息. 除了极大似然估计量之外, 基于识别结果 (11) 的估计方法还包括贝叶斯方法, 考虑到贝叶斯方法天然地具有统计推断的性质, 我们将其放在统计推断部分的第 6.5 小节讨论. 需要注意的是, 贝叶斯方法所要求的参数识别与极大似然法完全相同.

(2) 最小二乘 (least squares, LS) 估计量. 将 (12) 式最小化问题的目标函数替换为其样本模拟量, 即可构建出最小二乘估计量:

$$\hat{\psi}_{\text{LS}} = \arg \min_{\psi' \in \Psi(\mathcal{M})} \frac{1}{n} \sum_{i=1}^n (Y_{o,i} - m(X_i; \psi'))^2. \quad (17)$$

给定参数 ψ , 函数 $m(X_i; \psi)$ 可以看作一个基于 X_i 的信息对 $Y_{o,i}$ 的拟合, 而 $Y_{o,i} - m(X_i; \psi)$ 就是在参数 ψ 之下的拟合残差. 最小二乘估计量就是寻找能够最小化残差平方和的参数. 当 $m(x; \psi) = x^T \psi$ 时, $\hat{\psi}_{\text{LS}}$ 就是我们非常熟悉的普通最小二乘估计量 (ordinary least squares, OLS).

(3) 广义矩方法 (generalized method of moments, GMM) 估计量. 识别结果 (13) 本质上是在求解总

体矩方程 $\mathbb{E}[g(Y_o, X; \psi)] = 0$, 为了保证参数可识别性, 矩的数量需要不少于参数的数量, 即 $L \geq K$. 在恰好识别 (即 $L = K$) 的情形下, 我们可以将总体矩方程替换为样本矩方程 $(1/n) \sum_{i=1}^n g(Y_{o,i}, X_i; \psi) = 0$, 此时样本矩方程中的方程数量等于未知数数量, 因此该方程以高概率有唯一解, 这个解就是真实参数的矩估计量 (moment estimator). 在过度识别 (即 $L > K$) 的情形下, 只要设定正确, 总体矩方程中包含 $L - K$ 个冗余方程, 因此总体矩方程仍然有解. 但是, 将总体矩方程替换为样本矩方程之后, 样本中包含的抽样误差将使得这 L 个方程无法同时成立, 即在过度识别情形下样本矩方程无解. 此时的解决方案是寻找参数以使得样本矩的某个范数最近接 0. 具体地, 选取 $L \times L$ 的权重矩阵 \widehat{W} , 它满足 $\widehat{W} \xrightarrow{\mathbb{P}} W$, 其中 W 对称且正定, 那么广义矩估计量定义如下:

$$\widehat{\psi}_{\text{GMM}} = \arg \min_{\psi' \in \Psi(\mathcal{M})} \left[\frac{1}{n} \sum_{i=1}^n g(Y_{o,i}, X_i; \psi') \right]^T \widehat{W} \left[\frac{1}{n} \sum_{i=1}^n g(Y_{o,i}, X_i; \psi') \right]. \quad (18)$$

注意到方程可以等价地转化为范数最小化问题, 因此广义矩估计量 $\widehat{\psi}_{\text{GMM}}$ 也包含了恰好识别情形下的矩估计量. 除了广义矩方法以外, 基于总体矩方程 $\mathbb{E}[g(Y_o, X; \psi)] = 0$ [即识别结果 (13)] 的估计量还包括经验似然 (empirical likelihood) 估计量. 经验似然方法由 Owen (1988, 1990) 提出, Qin and Lawless (1994) 将其拓展至基于矩方程来估计参数的问题中, 并证明了经验似然估计量与最有效的广义矩方法渐近等价. 此后, Newey and Smith (2004) 证明了经验似然估计量比广义矩方法具有更好的高阶性质.

(4) 经典最小距离 (classical minimum distance, CMD) 估计量. 在识别方法 (14) 中, 泛函 \mathcal{H} 的构造通常是直接与参数估计环节结合在一起的. 用 $\mathbf{S}_n = \{(Y_{o,i}, X_i) : i = 1, \dots, n\} \in (\mathcal{Y}_o \times \mathcal{X})^n$ 表示容量为 n 的独立同分布样本. 我们先构建一个统计量 $\widehat{\pi}_n : (\mathcal{Y}_o \times \mathcal{X})^n \rightarrow \mathbb{R}^L$, 使得当 $n \rightarrow \infty$ 时, $\widehat{\pi}_n(\mathbf{S}_n) \xrightarrow{\mathbb{P}} \pi_0$, 那么 π_0 必然是总体分布的一个泛函, 记为 $\pi_0 = \mathcal{H}(\mathbf{P})$, 这样便得到了 \mathcal{H} 的构造. 虽然我们在讨论参数可识别性时, 可以允许识别约束 $\mathcal{H}(\mathbf{P}) = h(\psi)$ 中的函数 h 形式未知, 但是在进行参数估计时, 我们至少需要能够在不同的 ψ 之下数值计算函数值 $h(\psi)$. 为了保证参数可识别性, 我们也需要 $L \geq K$. 与广义矩方法类似, 选取 $L \times L$ 的权重矩阵 \widehat{W} , 它满足 $\widehat{W} \xrightarrow{\mathbb{P}} W$, 其中 W 对称且正定, 那么经典最小距离估计量定义如下:

$$\widehat{\psi}_{\text{CMD}} = \arg \min_{\psi' \in \Psi(\mathcal{M})} [\widehat{\pi}_n(\mathbf{S}_n) - h(\psi')]^T \widehat{W} [\widehat{\pi}_n(\mathbf{S}_n) - h(\psi')]. \quad (19)$$

注意到该估计量也适用于恰好识别的情形.

在约简式模型中, 由于函数形式相对简单, 我们可以直接套用上述估计方法得到参数估计量. 然而, 由于模型的复杂性, 当我们把上述几类一般的估计方法应用于结构方程的参数估计时, 会存在以下难点. 第一, 估计量 (16)–(19) 中的最优化目标与参数之间的函数关系通常依赖于结构方程模型本身的最优解或者均衡, 因此要想估计参数, 我们首先需要能够在不同的参数之下 (至少数值地) 求解模型. 我们将在第 5.2 小节详细讨论模型求解方法. 第二, 在求解模型和计算最优化问题 (16)–(19) 的目标函数的过程中, 不确定性、不可观测的异质性和随机系数的存在或者诸如 h 的函数定义方式要求我们计算一些随机变量的期望或分布的泛函, 而它们通常没有解析表达式, 因此我们需要使用蒙特卡洛模拟 (Monte Carlo simulation) 来近似这些期望或泛函. 我们将在第 5.3 小节讨论基于模拟的估计方法. 第三, 因为常见的数值求根和数值最优化算法都需要进行迭代更新, 所以在结构方程模型和参数估计量没有解析解的情况下, 数值地计算出参数估计量至少包含两层嵌套的循环: 内层循环是在给定的参数之

下求解模型, 外层循环是寻找参数以求解估计量 (16)–(19) 定义式中的最优化问题. 这就意味着参数估计环节会带来巨大的计算负担. 机器学习的一些思想和算法可以帮助我们提高模型求解和参数估计的计算效率, 我们将在第 5.4 小节详细介绍. 第四, 如果参数 ψ 的维度 K 随着样本量的增加而增加, 那么前面讨论的估计方法可能不再适用, 我们将在第 5.5 小节讨论高维模型的估计问题.

5.2 模型求解

在第 4 节我们已经讨论过, 结构方程模型的参数识别首先要求刻画模型的解, 包括单人模型中最优化问题的解和多人模型中的均衡解. 对于参数估计, 仅仅刻画模型的解是不够的, 我们必须具体地计算出模型的解. 这是因为, 参数估计的过程需要计算似然函数或矩函数的值, 而似然函数和矩函数中又会涉及模型的内生变量, 它们的值需要求解模型才能得到. 早期理论文献和教科书中的经典模型为了给读者提供可操作化的结果并分析其经济学含义, 通常会通过设置特殊的函数形式和假定来确保模型有解析解. 但在基于结构方程模型的经验研究中, 为了贴合实际情形和提高结论的可推广性, 模型的设定会倾向于更加一般化, 这也导致绝大多数非线性结构方程模型的解都不存在解析表达式. 因此, 在本小节我们着重讨论结构方程模型的数值求解方法; 至于在模型求解过程中可能涉及到的数值积分问题, 我们将在第 5.3 小节详细讨论.

静态模型的求解相对容易. 在静态模型中, 行为人当期的效用只取决于当期的状态和选择, 不存在跨期依赖关系, 因此即使模型涉及多个时期, 我们也可以将其分解为若干个彼此独立的单期问题来求解. 在每个单期问题中, 行为人的选择是有限维向量, 因而其对应的最优化问题只涉及有限维自变量, 对应的均衡条件方程和/或不等式组只涉及有限个未知数. 当这些最优化问题和方程没有解析解时, 我们可以使用黄金分割搜索 (golden section search)、单纯形方法 (simplex method)、拟牛顿算法 (quasi-Newton algorithm) 等数值最优化方法和二分查找 (bisection)、割线法 (secant method)、牛顿-拉弗森算法 (Newton-Raphson algorithm) 等数值求根方法来求出它们的解. 这些数值算法已经发展得较为成熟, 对此更加详尽的介绍可以参见 Press et al. (2007) 的教材. 此外, 梯度下降法 (gradient descent) 及其衍生方法成为目前较为流行的数值最优化算法, 详见 Ruder (2017) 的综述. 注意到, 参数估计量 (16)–(19) 定义式中的最优化问题也只涉及有限维自变量, 因此上面提到的数值最优化方法也可以应用于求解估计量的外层循环.

动态模型的求解则比较困难. 在动态模型中, 行为人当期的状态和选择会影响未来的效用, 这使得不同期的选择存在跨期依赖关系, 无法被拆分开来单独考虑. 即使没有不确定性, 行为人的策略也必然是由所有期的选择构成的序列或过程. 而在现实的大多数情况下, 行为人无法在当期获知未来的状态 (包括动态博弈中对手在未来的选择), 所以研究者在构建动态模型时通常会同时引入不确定性. 在这种情况下, 行为人每一期的选择会依赖于当前已有的信息, 即每一期的决策模式是从状态空间到选择集的映射, 而所有期的决策模式 (映射) 共同构成了行为人的完整的策略. 这就意味着, 每个行为人面对的是一个随机动态最优化问题, 其自变量是前面所述的映射 (函数) 序列, 而多人模型中的均衡条件也是关于这些函数的方程或不等式.

对于离散时间的非随机动态最优化问题, 我们可以通过拉格朗日 (Lagrange) 乘数法得到差分方程组作为最优性条件; 对于连续时间的非随机动态最优化问题, 我们可以通过庞特里亚金 (Pontryagin)

最大值原理得到微分方程组作为最优性条件. 然而, 在一般的模型设定下, 这些差分或微分方程组并没有解析解. 此外, 拉格朗日乘数法和庞特里亚金最大值原理都不适用于随机动态最优化问题的求解. 因此, 文献中通常会将 (非随机和随机) 动态最优化问题用递归的方式写为函数方程的形式, 即离散时间情形下的贝尔曼 (Bellman) 方程和连续时间情形下的哈密尔顿-雅可比-贝尔曼 (Hamilton–Jacobi–Bellman, HJB) 方程. 如果动态最优化问题的每一期具有相同的结构, 那么函数方程的表示方式可以在形式上极大地简化原始问题, 例如在指数贴现下的无穷期 (期望) 效用最大化问题中, 如果单期效用函数形式和状态变量运动法则不随时间变化, 那么值函数也不随时间变化, 原始问题仅用一个关于这个值函数的函数方程就能刻画.

在离散时间情形下, 构建和求解贝尔曼方程的理论可见 [Stokey et al. \(1989\)](#) 的教材. 数值求解贝尔曼方程的基本方法是值函数迭代法 (value function iteration): 在有限期问题中, 从最后一期的值函数开始逆向归纳 (backward induction), 从后往前依次解出各期值函数; 在无穷期问题中, 如果贝尔曼方程右端是值函数的一个压缩映射, 那么从任意形式的值函数开始用贝尔曼方程进行迭代, 最终将收敛至不动点. 值函数迭代法的操作对象是值函数, 这就会涉及到如何表示函数的问题. 最直接的做法是将值函数的定义域 (即状态空间) 离散化为格点, 用这些格点处的函数值来代表整个函数. 给定我们要求的计算精度, 上述格点搜寻法 (grid searching) 所需要的格点数量与状态变量的维度呈现指数函数关系, 因此当状态变量是多维甚至高维时, 格点搜寻法会带来极大的甚至不可行的计算负担. 为了应对高维状态变量的问题, [Brumm and Scheidegger \(2017\)](#) 提出了自适应稀疏格点方法 (adaptive sparse grids), 极大地提高了计算效率. 除此之外, 机器学习的思想和算法也可以帮助我们高效地求解动态最优化问题, 我们将在第 [5.4](#) 小节详细讨论.

在连续时间情形下, 构建和求解 HJB 方程的理论可见 [Fleming and Rishel \(2012\)](#) 的教材. 数值求解 HJB 方程的基本方法是有限差分法 (finite difference methods, [Candler, 2001](#)), 其核心思想是用格点函数值表示值函数并用差分之商近似 HJB 方程中的偏导数. [Wang and Forsyth \(2010\)](#) 在研究连续时间均值-方差资产配置问题时, 以及 [Achdou et al. \(2022\)](#) 在研究连续时间收入和财富分布的 Aiyagari–Bewley–Huggett 模型时, 都使用了有限差分法求解 HJB 方程. 注意到, 当状态变量是多维甚至高维时, 基于格点的有限差分法同样也面临着计算负担过大的问题.

对于多人模型, 在求解了每个行为人的动态最优化问题之后, 还需要将个体最优解与均衡条件相结合, 进而求出均衡解. 如果我们按照上文讨论的数值方法求解了贝尔曼方程或 HJB 方程, 那么我们可以将这些最优解代入均衡条件求出动态经济模型的均衡解, 这种方法被称为全局解法. 与全局解法相对应的是局部解法, 多被用于异质性行为人的动态一般均衡求解, 其做法是先在去除某些随机性的条件下求出模型的稳态解 (steady state solution), 再在稳态解附近做扰动 (perturbation). 例如 [Reiter \(2009\)](#) 在状态空间下对稳态解进行了扰动, 而 [Auclert et al. \(2021\)](#) 则利用序列空间雅可比矩阵来求解离散时间异质性行为人模型, 这种基于序列空间的方法后续被拓展到连续时间模型中 ([Glawion, 2023; Bilal and Goyal, 2025](#)). 除此之外, 关于求解离散时间动态经济模型的更多讨论可以参见 [Maliar and Maliar \(2014\)](#), 而关于求解动态博弈模型的详细讨论可以参见 [Jorgensen et al. \(2007\)](#).

5.3 基于模拟的估计

在求解模型和计算最优化问题 (16)–(19) 的目标函数的过程中, 不确定性、不可观测的异质性和随机系数的存在或者诸如 h 的函数定义方式要求我们计算一些随机变量的期望或分布的泛函. 如果我们无法推导这些期望或泛函的解析表达式, 但是能够从相应的分布中抽样, 那么我们就可以使用蒙特卡洛模拟来近似这些函数或泛函. 我们把参数估计 (包括模型求解) 过程中涉及了蒙特卡洛模拟的方法统称为基于模拟的估计. 在第 5.1 小节介绍的 4 种估计方法基础上, 结构方程常用的基于模拟的估计方法可以分为两大类.

第一类是模拟极大似然、模拟最小二乘和模拟矩方法, 它们分别是极大似然、最小二乘和广义矩方法与蒙特卡洛模拟的结合. 因为核心思想是相同的, 这里我们仅以模拟矩方法 (simulated method of moments) 为例进行介绍, 模拟极大似然和模拟最小二乘完全可以类比. 模拟矩方法适用的情形是在广义方法估计量 (18) 中, 函数 g 或者其计算过程涉及没有解析表达式的期望 (或更一般的积分). 例如我们需要计算 $\mathbb{E}[\xi(U)]$, 其中 $U \sim Q$ 且 ξ 已知. 假设我们无法直接算出 $\mathbb{E}[\xi(U)]$, 但是能够生成服从分布 Q 的随机变量, 那么我们就可以用如下的蒙特卡洛方法来近似 $\mathbb{E}[\xi(U)]$: 从分布 Q 中抽取一个容量为 N_s 的独立同分布的蒙特卡洛样本 $\{U_1, \dots, U_{N_s}\}$, 并计算

$$\widehat{\mathbb{E}[\xi(U)]} = \frac{1}{N_s} \sum_{r=1}^{N_s} \xi(U_r).$$

对模拟矩方法的更多介绍可以参见 Carrasco and Florens (2002).

在实际应用中, 模拟矩方法的一个经典案例是 BLP 需求模型 (Berry et al., 1995) 的估计. 在该模型中, 工具变量和偏好冲击之间的正交性蕴含了一组矩条件, 在此基础上我们使用广义矩方法估计参数. 但是偏好冲击无法在数据中被直接观测到, 我们需要在每一组给定参数之下用可观测数据将其还原出来. 在数值计算偏好冲击的过程中, 我们需要对模型中的随机系数做积分, 这一步操作正是通过蒙特卡洛模拟实现.

第二类是间接推断 (indirect inference) 方法, 它是经典最小距离估计与蒙特卡洛模拟的结合, 也是经典最小距离能够被用于结构方程参数估计的几乎唯一途径. 这是因为, 在经典最小距离估计量 (19) 中, 结构方程模型的复杂性导致我们没有办法直接计算函数 $h(\psi)$, 只能借助于模拟方法. Gourieroux et al. (1993) 最早提出的间接推断方法只考虑了经典最小距离估计中一类比较特殊的 h (通过最大化问题定义的函数), 但间接推断的后续应用已经扩展到一般的函数 h 的情形, 因此我们在这里介绍一般情形.

在给定的参数 ψ 之下, 我们先模拟一个容量为 n 的样本 \mathbf{S}_n^\dagger . 一般而言, 可观测外生变量 X 的边缘分布不是我们感兴趣的元素, 而也通常不受到参数 ψ 的影响, 因此模拟样本 \mathbf{S}_n^\dagger 中的可观测外生变量直接使用原始样本的观测 $\{X_i : i = 1, \dots, n\}$. 我们只需要从与参数 ψ 相容的分布中抽取不可观测的外生变量 $\{\varepsilon_i^\dagger : i = 1, \dots, n\}$, 再基于每个外生变量 $(X_i, \varepsilon_i^\dagger)$ 求解模型, 解出可观测的内生变量 $Y_{o,i}^\dagger$. 这样, 一个模拟样本就是

$$\mathbf{S}_n^\dagger = \left\{ \left(Y_{o,i}^\dagger, X_i \right) : i = 1, \dots, n \right\}.$$

利用一个模拟样本, 我们可以计算出统计量 $\hat{\pi}_n(\mathbf{S}_n^\dagger)$. 独立地模拟 N_s 次样本, 我们就可以计算出 N_s 个统计量, 为了记号简便不妨将其记为 $\hat{\pi}_1^\dagger(\psi), \dots, \hat{\pi}_{N_s}^\dagger(\psi)$, 它们的平均值就可以作为 $h(\psi)$ 的近似. 由此

即可得到间接推断估计量:

$$\hat{\psi}_{II} = \arg \min_{\psi' \in \Psi(\mathcal{M})} \left[\hat{\pi}_n(\mathbf{S}_n) - \frac{1}{N_s} \sum_{r=1}^{N_s} \hat{\pi}_r^\dagger(\psi') \right]^T \hat{W} \left[\hat{\pi}_n(\mathbf{S}_n) - \frac{1}{N_s} \sum_{r=1}^{N_s} \hat{\pi}_r^\dagger(\psi') \right].$$

在实际应用中, 有几类常见的选取间接推断统计量 $\hat{\pi}_n$ 的方法. 一是直接使用一些样本矩作为 $\hat{\pi}_n$, 此时很多经验研究者也会将这种做法称为模拟矩方法, 如果严格按照定义, 这种说法也是正确的, 因为在仅使用不依赖于参数的矩条件的情形下, 广义矩方法和经典最小距离估计量可以写为相同的形式. 例如, 受东亚社会中“高结婚率、低生育”等典型事实的启发, Myong et al. (2021) 构建了一个考虑社会规范的生育与婚姻模型, 并采用间接推断对模型进行估计. 作者利用数据中的 30 个经验矩来识别并估计 10 个结构参数. 这些矩包括: 在 7 个教育水平分组下, 完成生育期女性的平均子女数量、无子率, 以及相应的男女结婚率等. 具体而言, 作者在给定参数的情况下, 对不同教育水平的女性进行随机模拟(每组十万名个体), 计算模拟数据中的经验矩, 并通过最小化模拟经验矩与原始样本经验矩之间的距离, 得到结构参数的估计值.

二是按照 Gourieroux et al. (1993) 对间接推断方法的原始定义, 构建辅助回归 (auxiliary regression), 并使用拟极大似然将其与参数联系起来. 这些辅助回归侧重于刻画变量之间的相关性, 一般不具有因果的解读. 例如, 在 Altonji et al. (2013) 分析职业收入变化的研究中, 作者刻画了多种离散变量(如就业保持、再就业与换工作)的转移概率, 以及连续变量(如工资、工时和收入)的动态过程与协方差结构. 该辅助模型包含大量回归系数与分布参数, 它们被拟似然函数加以汇总. 作者通过最小化真实数据与结构模型模拟数据的差异, 得到结构模型中 55 个参数的估计.

三是使用约简式方法中一些具有因果含义的估计量, 这种做法也是当前的经济学经验研究较为推崇的, 因为它将约简式范式和结构方程范式有机地结合了起来. 具体做法是, 在原始样本中利用约简式方法得到诸如双重差分 (difference in differences) 和断点回归 (regression discontinuity) 等具有因果含义的估计量, 同时在任意一组参数下求解结构方程模型并生成模拟样本, 在模拟样本中计算同类型的估计量, 通过最小化原始样本估计量和模拟样本估计量之间的距离, 得到参数的估计量. 例如, Guo et al. (2025) 提出“配给收入效应”用以解释子女数量-质量权衡效应. 该机制区分两类生育变化: 合意生育变化与非合意生育变化. 该分类无法直接从微观数据中观察, 而是由复杂的结构函数定义. 为此, 作者利用中国独生子女政策与双胞胎出生的准自然实验, 构建结合工具变量与分层变量的识别框架, 并通过两阶段回归分别识别合意与非合意生育增加对子代人口素质的因果效应. 这些通过约简式得到的明确经济含义的因果效应, 被作为间接推断中的经验矩条件, 用以估计结构模型中的偏好参数. 值得注意的是, 尽管选取具有因果含义的估计量作为间接推断统计量的做法备受欢迎, 目前并没有研究证明这种做法在理论上存在优势, 例如更容易保证参数识别或估计量有更好的渐近性质. 这一开放问题值得在今后被继续探索.

5.4 与机器学习的结合

机器学习 (machine learning) 作为由数据驱动并以预测为目标的一套方法, 其模型本身的可解释性较低, 这似乎与结构方程研究范式的目标相悖, 因而机器学习的模型通常不能直接当作结构方程模型来使用. 但是, 机器学习领域中有一些思想和算法, 它们的作用是提高机器学习方法的计算效率和稳健

性, 而这也是结构方程研究 (特别是模型求解和参数估计环节) 所需要的性质. 因此, 这些思想和算法可以为结构方程模型的求解和估计提供新的思路.

首先是深度学习 (deep learning) 对动态经济模型求解的帮助, 这也是计算经济学领域的一个前沿研究方向. 如第 5.2 小节所述, 动态最优化问题求解的困难在于目标函数的自变量本身也是函数 (或函数序列), 在递归的表示形式之下, 我们需要求出使得函数方程成立的值函数 (状态变量到目标函数最优值的映射) 和/或政策函数 (状态变量到最优选择的映射). 传统的基于格点搜寻的值函数迭代算法本质上是将状态空间离散化为格点, 用格点上的函数值来逼近整个值函数和政策函数. 为了确保逼近的精度, 每个维度上都必须有足够的格点数量. 因为值函数和政策函数都以状态变量为自变量, 且迭代的步骤中需要对控制变量进行搜索, 所以所需的总的格点数量是状态变量与控制变量数量之和的指数函数, 这就导致了维数诅咒 (curse of dimensionality) 问题. 除了第 5.2 小节介绍的自适应稀疏格点 (Brumm and Scheidegger, 2017) 等方法以外, 深度神经网络 (deep neural network) 也能用来解决上述维数诅咒问题.

正如 [Hastie et al. \(2009\)](#) 第 2 章所言, 有监督的机器学习的本质就是 (有效率的) 函数逼近. 深度神经网络是当前比较流行的机器学习或函数逼近方法, 它本质是多层函数的依次复合, 其中的每一层函数将上一层的输出做线性变换后, 再将同一个非线性函数点点地作用于其上作为输出. 设 x 为最初的输入向量, 则一个多层神经网络可以写为如下形式:

$$D(x) = B_L \tau(\cdots \tau(B_2 \tau(B_1 \tau(B_0 x + b_0) + b_1) + b_2) + \cdots) + b_L,$$

其中 L 是神经网络的层数, B_ℓ 和 b_ℓ 分别是参数矩阵和向量, τ 是将已知的一元函数 $\tau_0 : \mathbb{R} \rightarrow \mathbb{R}$ 点点地应用到向量上. [Hornik et al. \(1989\)](#) 已经证明, 神经网络理论上能够以任意精度逼近任何一个有限维的博雷尔 (Borel) 可测函数. 进一步地, [Barron \(1993, 1994\)](#) 给出了神经网络逼近和估计误差的上界, 由该结论可推知: 给定逼近误差, 神经网络所需要的参数个数随着自变量的个数的增长速度不超过线性增长, 这就避免了格点逼近和多项式逼近的维数诅咒问题.

如果我们难以直接求解某个未知函数, 但能够以可操作的方式对这个函数进行刻画, 那么我们就可以考虑用深度神经网络来逼近这个函数, 并利用对函数的刻画来构建损失函数 (loss function). 训练神经网络的过程本质上就是求解损失函数最小化问题, 得到的最优解就是在给定的神经网络模型空间中对未知函数的最佳逼近. 在偏微分方程数值求解领域, 已经出现基于上述思想用深度学习方法求解偏微分方程的方法 (例如 [Han et al., 2018; Sirignano and Spiliopoulos, 2018](#)). 注意到连续时间动态最优化问题的 HJB 方程本质上也是一个 (组) 偏微分方程, 因此将深度学习方法用于求解动态最优化问题也是顺理成章的.

用深度学习方法求解动态最优化问题的核心思想就是用深度神经网络逼近值函数或政策函数, 从而将求解动态最优化问题转化为训练深度神经网络的问题 ([Fernández-Villaverde et al., 2020; Kahou et al., 2021; Maliar et al., 2021; Duarte et al., 2024](#)). 在训练过程中, 损失函数可以通过动态最优化问题的最优性条件来构造, 例如将贝尔曼方程、欧拉方程或 HJB 方程左右两端的误差的范数作为损失函数. 因为神经网络的函数形式具有光滑性, 所以求解损失函数最小化问题可以通过随机梯度下降等方法实现, 这些方法在机器学习领域已有成熟的开发和应用. 与基于格点搜寻的值函数迭代相比, 上述方

法将每一步迭代需要计算量从指数量级减少到线性量级. 在深度神经网络逼近方法之外, Scheidegger and Bilionis (2019) 用高斯过程机器学习方法来求解高维随机动态规划问题.

深度神经网络逼近的思想可以被进一步地延伸到动态经济模型的均衡求解中. Azinovic et al. (2022) 提出了深度均衡网络方法, 用于近似求解带有异质性、不确定性和紧约束的经济模型的理性预期均衡. 其核心思想是用深度神经网络逼近均衡下的选择函数和价格函数, 并将所有均衡条件 (包括最优性条件) 的均方误差作为损失函数用以训练该神经网络.

除了动态经济模型的求解, 机器学习的思想和算法还可以直接应用于结构方程模型的参数估计. Geng et al. (2023) 将动态离散选择模型的估计看作逆强化学习 (inverse reinforcement learning) 问题, 在传统的嵌套不动点迭代-极大似然估计的基础上, 通过状态的加总, 降低了计算的复杂度. Kaji et al. (2023) 则将原本用于图像生成的生成对抗网络 (generative adversarial networks) 应用于结构方程模型, 通过生成器和判别器的博弈来得到模型参数的估计量, 其中的生成器是用结构方程模型来模拟数据, 使其尽可能与真实观测数据相似, 而判别器则尽可能将模型生成的数据与真实观测数据区分开. 前面讨论的都是全参数模型, 对于半参数和非参数模型, Chen et al. (2021) 和 Farrell et al. (2021) 提出用神经网络来逼近数据分布或数据分布中的某个未知函数, 进而实现对模型的估计. 进一步地, Farrell et al. (2025) 考虑了结构方程模型, 其中的结构参数是个体特征的 (非参数形式的) 函数, 以此刻画异质性; 他们用深度神经网络来逼近这个结构函数, 进而得到它的估计量, 并在此基础上估计具有经济学含义的有限维参数.

上述方法在使用结构方程的经验研究中也已经被应用. Fernández-Villaverde et al. (2023) 在连续时间框架下利用神经网络估计资产分布演化的状态转移方程, 从而在含金融摩擦的异质性家庭模型中刻画内生总体风险及“高杠杆”和“低杠杆”两种稳态. Fernández-Villaverde et al. (2025) 则考虑了离散时间版本, 引入零利率下限构成了一个高度非线性的异质性行为人新凯恩斯 (heterogeneous agent new Keynesian) 模型. Payne et al. (2024) 将搜寻与匹配模型表述为含有分布状态的高维偏微分方程, 并利用深度神经网络进行全局求解与结构参数估计. Ruiz et al. (2020) 扩展了传统离散选择模型对商品相关性的参数化设定, 借鉴自然语言处理中的嵌入方法和推荐系统中的矩阵分解, 估计了商品之间的替代性与互补性, 从而构建了一个包含多商品交互的序列化选择模型.

5.5 高维模型的估计

第 5.1 小节介绍的估计方法都建立在参数 ψ 的维度 K 有限且固定的假定之上; 此外, 基于有限信息的方法还需额外要求使用的矩 (或分布的其它泛函, 以下不再赘述) 的数量 L 也是有限且固定的. 然而, 高维模型违背了这些限定条件. 在理论上, 按照 Fan et al. (2011) 第 1.2 节的定义, 只要参数维度随着样本量的增加而发散, 即 $\lim_{n \rightarrow \infty} K(n) = \infty$, 那么该模型就被视作高维 (high dimensional) 模型; 特别地, 如果参数维度发散的速度比样本量的任意多项式函数还要快, 即 $K(n) = O[\exp(n^\alpha)]$ 且 $\alpha > 0$, 那么该模型被称为超高维 (ultra-high dimensional) 模型. 例如, 在博弈和社会网络问题中, 模型刻画了样本中的各个观测值之间的关系, 因而参数维度 K 会随着样本量 n 的增加而增加. 考虑到不同个体之间选择相互影响, 在一个有 N 个体的无向网络中, 潜在的链接方式达到 $2^{\binom{N}{2}}$ 种 (Graham, 2015). 在应用上, 对于给定的模型和数据集, 如果参数维度接近或大于样本量, 我们就将该模型视作高维模型; 如

果参数维度“远远”大于样本量, 我们就将该模型视作超高维模型. 例如在 Gentzkow et al. (2019) 使用的美国国会演讲记录文本数据中, 语料的维度远远大于语料的数量; 又如 Feng et al. (2020) 考虑了众多的定价因子, 其数量也可能大于时间序列的期数. 此外, 在使用有限信息方法识别参数的模型中, 识别参数所需要的矩的数量也可能随着样本量的增加而发散. 例如, 当参数是高维时, 我们需要数量不少于参数维度的矩方程才能识别这些参数, 这导致矩方程也是高维的; 又如, 当参数通过条件矩方程识别时, 我们可以使用缝合 (sieve) 方法将条件矩方程转换为一列无条件矩方程, 其数量随着样本量增加而发散 (Donald et al., 2003). 结构方程经验研究中也有使用大量的矩方程来识别参数的案例, 例如 Eaton et al. (2011) 研究法国企业国际贸易和 Altonji et al. (2013) 研究工资与收入动态时都使用了多至上千个矩方程 (样本量为几万量级). 因此, 我们在本小节中讨论的 (超) 高维模型既包括参数是 (超) 高维的情形, 也包括使用有限信息方法时使用的矩是 (超) 高维的情形.

对于高维模型, 文献已经证明, 只有在参数的维度和矩的维度以很慢的速度发散时, 极大似然估计量、广义矩方法估计量和经验似然估计量才具有良好的性质 (Portnoy, 1988; Donald et al., 2003; Chen et al., 2009; Candès and Sur, 2020). 具体地, 仅当参数和矩的维度以不快于样本量的速度发散时 (即 $\limsup_{n \rightarrow \infty} (\max\{K(n), L(n)\}/n) < 1$), 这些估计量才具有相合性; 仅当参数和矩的维度以比样本量的平方根更低阶的速度发散时 (即 $\lim_{n \rightarrow \infty} (\max\{K(n), L(n)\}/\sqrt{n}) = 0$), 这些估计量才具有渐近正态性. 然而, 随着数据可得性的提高和数据类型的丰富, 在现在的很多高维乃至超高维模型中, 参数和矩的维度的发散速度都远远超出了上面要求的慢速度. 这种情况下, 我们必须对传统的极大似然、广义矩和经验似然等估计方法加以修改, 才能满足 (超) 高维模型的估计需要.

文献在处理高维模型时通常的出发点是参数的稀疏性 (sparsity), 即假定高维参数中仅有极少的非零值. 在这一基础上, 估计高维模型的通常思路是使用惩罚较大参数的正则化 (regularization) 方法实现参数选择和/或参数压缩, 例如 Tibshirani (1996) 基于线性模型提出的 LASSO (least absolute shrinkage and selection operator) 和 Fan and Li (2001) 基于广义线性模型提出的 SCAD (smoothly clipped absolute deviation) 惩罚方法. 这些思想后续被应用于估计更加一般化的高维模型. Kwon and Kim (2012) 研究了带有 SCAD 惩罚的极大似然估计量在高维模型中的性质, 发现该估计量与仅需估计非零参数的先知 (oracle) 极大似然估计量渐近等价. Shi (2016) 考虑参数维度固定但矩方程数量发散的高维矩约束模型, 提出了将矩约束放松的经验似然估计量和进一步的纠偏方法. Chang et al. (2018) 考虑了参数和矩方程都是 (超) 高维的模型, 提出了惩罚经验似然估计量, 它同时对参数和经验似然的拉格朗日乘子施加了惩罚. Chang et al. (2023) 在 (超) 高维矩约束模型中允许存在误设的矩条件, 并将惩罚经验似然估计方法中对于参数的惩罚项改为对于矩的惩罚项, 从而实现参数估计.

6 统计推断

6.1 漐近正态推断

在计量经济学理论中, 极大似然估计量 (16)、最小二乘估计量 (17)、广义矩方法估计量 (18) 和经典最小距离估计量 (19) 在标准情形下的渐近性质 (即大样本性质) 已经有教科书级别的结论 (Newey and McFadden, 1994 第 3 节). 具体而言, 如果计量经济学模型和样本满足一些常规的技术性假定,

那么利用中心极限定理 (central limit theorem)、连续映射定理 (continuous mapping theorem) 和德尔塔 (Delta) 方法等工具即可推导出参数估计量的渐近正态性, 即 $\sqrt{n}(\hat{\psi}_e - \psi_0) \xrightarrow{d} \mathcal{N}(0, V_e)$, 其中 $e \in \{\text{ML, LS, GMM, CMD}\}$. 如果我们能得到渐近方差矩阵 V_e 的相合估计量, 那么我们就能利用渐近正态性去近似参数估计量的分布, 或者在此基础上构造渐近服从 χ^2 分布的统计量, 进而开展统计推断. 下面给出了上述 4 种估计量在标准情形下的渐近方差的形式.

(1) 极大似然估计量:

$$V_{\text{ML}} = [I(\psi_0)]^{-1},$$

其中

$$I(\psi) = \mathbb{E} \left[\left(\frac{\partial \ln f(Y_{o,i}, X_i; \psi)}{\partial \psi} \right) \left(\frac{\partial \ln f(Y_{o,i}, X_i; \psi)}{\partial \psi} \right)^T \right].$$

顺便指出, 这里的 $I(\psi_0)$ 被称为 (标准化的) 信息矩阵, 而它的逆矩阵 $[I(\psi_0)]^{-1}$ 是参数 ψ_0 的一切相合估计量的渐近方差的下界, 也被称为 Cramér–Rao 下界. 极大似然估计量的渐近方差之所以能够达到这一下界, 是因为它使用了分布的全部信息.

(2) 最小二乘估计量:

$$V_{\text{LS}} = [M(\psi_0)]^{-1} \Omega_m(\psi_0) [M(\psi_0)]^{-1},$$

其中

$$M(\psi) = \mathbb{E} \left[\left(\frac{\partial m(X_i; \psi)}{\partial \psi} \right) \left(\frac{\partial m(X_i; \psi)}{\partial \psi} \right)^T \right],$$

$$\Omega_m(\psi) = \mathbb{E} \left[\left(\frac{\partial m(X_i; \psi)}{\partial \psi} \right) \left(\frac{\partial m(X_i; \psi)}{\partial \psi} \right)^T (Y_{o,i} - m(X_i; \psi))^2 \right].$$

(3) 广义矩方法估计量:

$$V_{\text{GMM}} = \left[G(\psi_0)^T W G(\psi_0) \right]^{-1} G(\psi_0)^T W \Omega_g(\psi_0) W G(\psi_0) \left[G(\psi_0)^T W G(\psi_0) \right]^{-1},$$

其中 W 是权重矩阵 \widehat{W} 的概率极限,

$$G(\psi) = \mathbb{E} \left[\frac{\partial g(Y_{o,i}, X_i; \psi)}{\partial \psi^T} \right],$$

$$\Omega_g(\psi) = \mathbb{E} \left[g(Y_{o,i}, X_i; \psi) g(Y_{o,i}, X_i; \psi)^T \right].$$

(4) 经典最小距离估计量: 假定当 $n \rightarrow \infty$ 时, 统计量 $\hat{\pi}_n(\mathbf{S}_n)$ 具有渐近正态性, 即 $\sqrt{n}[\hat{\pi}_n(\mathbf{S}_n) - \pi_0] \xrightarrow{d} \mathcal{N}(0, \Omega_\pi)$, 那么

$$V_{\text{CMD}} = \left[H(\psi_0)^T W H(\psi_0) \right]^{-1} H(\psi_0)^T W \Omega_\pi W H(\psi_0) \left[H(\psi_0)^T W H(\psi_0) \right]^{-1},$$

其中 W 是权重矩阵 \widehat{W} 的概率极限, $H(\psi) = \partial h(\psi) / \partial \psi^T$.

如果极大似然、最小二乘或广义矩方法估计量的渐近方差定义式中的偏导数 $\partial \ln f(Y_{o,i}, X_i; \psi) / \partial \psi$, $\partial m(X_i; \psi) / \partial \psi$ 或 $\partial g(Y_{o,i}, x_i; \psi) / \partial \psi$ 存在解析表达式, 那么我们只需要将样本模拟量方法和代入 (plug-in) 方法结合就能得到渐近方差的估计量, 也就是将定义式中的期望替换为样本均值, 并将真实参数 ψ_0

替换为其估计量. 对于经典最小距离估计量, 如果统计量 $\hat{\pi}_n(\mathbf{S}_n)$ 的渐近方差矩阵 Ω_π 存在相合估计量 $\hat{\Omega}_\pi$ 且偏导数矩阵 H 存在解析表达式, 那么我们只需将渐近方差 V_{CMD} 定义式中的 Ω_π 和 ψ_0 替换为相应估计量, 即可得到渐近方差的估计量. Newey and McFadden (1994) 第 4 节讨论了这些的渐近方差估计量的相合性.

上文已经多次提到过, 在一般形式的结构方程模型中, 函数 f, m, g 或 h 通常没有解析表达式, 只能数值计算或近似. 因此, 在估计量具有渐近正态性的前提下, 为了估计渐近方差, 我们还需要用数值求导方法来近似上述偏导数. 对于经典最小距离估计量, 其渐近方差的估计要求我们首先找到统计量 $\hat{\pi}_n(\mathbf{S}_n)$ 的渐近方差矩阵 Ω_π 的相合估计量. 如果 $\hat{\pi}_n$ 具有比较常规的形式 (如样本矩), 那么我们可以借助现有统计和计量理论找到 Ω_π 的相合估计量. 否则, 我们可能需要使用自助法 (第 6.2 节) 等方法来估计渐近方差. 此外, 在实际问题中, 由于结合多个数据集等原因, 我们可能只能估计 Ω_π 的对角线元素, 而无法估计不同统计量之间的协方差. 在这种情况下, Cocci and Plagborg-Møller (2025) 提出了一个只利用 Ω_π 的对角线元素估计量的保守 (conservative) 方法, 用来估计 $\hat{\psi}_{\text{CMD}}$ 的函数的标准误和开展统计推断. 顺便一提, 虽然 Cocci and Plagborg-Møller (2025) 使用了校准 (calibration) 这一说法, 但是他们这里对于校准的定义比宏观经济学中的定义 (也是本文中的定义) 更加宽泛, 实际上就等同于经典最小距离估计量, 详见其第 2952 页的介绍和本文表 1 的注 1.

6.2 自助法

我们需要注意, 以上讨论的大前提是估计量具有渐近正态性, 而该性质的成立要求模型和样本满足相应技术性假定, 例如函数 f, m, g, h 关于参数可微和样本独立同分布等. 在很多结构方程模型特别是动态模型中, 函数 f, m, g, h 没有解析表达式, 而且它们的函数值对于参数的依赖关系高度复杂, 此时我们难以甚至无法研究这些函数关于参数的可微性, 那么估计量的渐近正态性就无法得到保证. 在这样的情形下, 上面介绍的基于渐近正态性的统计推断方法也就不适用.

此时我们可以考虑用自助法 (bootstrap) 来近似渐近分布, 而后进行统计推断. 假设原始样本 $\{(Y_{o,i}, X_i) : i = 1, \dots, n\}$ 独立同分布, 利用该样本计算出的原始估计量为 $\hat{\psi}$. 接下来我们从原始样本中抽取出一个自助法样本 $\{(Y_{o,i}^*, X_i^*) : i = 1, \dots, n\}$, 其中每个 $(Y_{o,i}^*, X_i^*)$ 都是从原始样本中独立且有放回地抽取. 使用这一自助法样本, 我们可以计算出一个自助法估计量 $\hat{\psi}^*$. 取足够大的整数 N_b , 将该操作独立地重复 N_b 次, 我们就能得到 N_b 个自助法估计量 $\{\hat{\psi}_1^*, \dots, \hat{\psi}_{N_b}^*\}$, 它们的经验分布即可作为原始估计量 $\hat{\psi}$ 的渐近分布的近似. 出于严谨性的考虑, 在使用自助法之前, 我们应当先证明原始估计量 $\hat{\psi}$ 的渐近分布存在, 且自助法估计量在给定原始样本时的条件分布也收敛到 $\hat{\psi}$ 的渐近分布, 这样才能确保自助法近似的可靠性.

在研究用户的上网套餐选择与使用行为时, Nevo et al. (2016) 采用了区块自助法 (block bootstrap) 计算模型的统计量及反事实分析的标准误. 具体地, 以用户为抽样单元, 有放回地抽取用户的完整 30 天上网记录, 使用构造的 1000 组重抽样样本重新计算距条件和权重.

自助法实质上是一种数据驱动的统计推断方法, 它在机器学习领域也有广泛应用. 我们在第 2 节已经介绍过, 机器学习的目标是预测结果变量, 模型的参数本身通常没有可解释性也不是研究者关注的重点, 因此针对机器学习的统计推断主要集中在构造结果变量的预测集 (prediction set) 上. 在构造预

测集的过程中, 研究者需要分析预测误差的性质, 而预测误差是样本内和样本外误差之和. 在一些特定情形中, 预测误差存在渐近或非渐近的理论结果. 例如, 当参数估计量性质良好且样本外误差被加总时, 预测误差可能具有渐近正态性; 再如, 对于特定的参数估计量和样本外误差分布族, 研究者可以推导出预测误差在概率意义上的上界. 在上述情况下, 关于预测误差的渐近或非渐近理论结果可以用于构建结果变量的预测集. 但是, 一些新近提出的机器学习算法可能还没有相应的误差分析理论, 此时研究者就只能借助于数据驱动的方法来实现统计推断. 除了自助法以外, 机器学习领域中数据驱动的统计推断方法还包括共形预测 (conformal prediction, [Vovk et al., 2009](#)) 等.

6.3 逼近误差

我们分别在第 [5.3](#) 和 [5.4](#) 小节中介绍了基于模拟的估计和用深度学习求解动态模型的方法, 它们本质上都是在参数估计环节的某些具体步骤中对一些无法直接求值的对象进行了逼近, 例如用蒙特卡洛模拟逼近分布的一个泛函或者用深度神经网络逼近动态最优化问题的解. 既然是逼近, 那么必定会存在逼近误差, 由此带来一个很自然的问题: 这些具体步骤中的估计误差是否会影响最终的参数估计量的渐近性质? 对于这个问题, 我们需要根据逼近误差的来源分两种情况讨论.

第一种情况是逼近误差可以由算法控制而不受限于样本. 此时, 理论上我们总是可以通过算法设计使得逼近误差在最终估计量的渐近分布中可以被忽略. 例如在基于模拟的估计方法中, 蒙特卡洛模拟对于目标值的逼近效果可以完全地由蒙特卡洛试验次数 (即模拟的数据集数量) 控制, 只要令蒙特卡洛试验次数远远大于样本量, 那么来自于模拟的逼近误差就不会影响最终估计量的渐近分布. 再如用神经网络逼近动态最优化问题的解, [Barron \(1993, 1994\)](#) 给出了神经网络的逼近误差的上界, 该上界的阶只取决于目标函数本身的特征、自变量的维度和训练时的采样数量, 而与参数估计所使用的样本量 n 无关, 因此我们总是可以通过把训练神经网络时的采样数量设置得足够大来保证神经网络的逼近误差不影响参数估计量的渐近分布. 当然, 在实际应用中, 出于计算负担的考虑, 我们不可能无限制地增加蒙特卡洛试验次数或者训练神经网络的采样数量. 对于基于模拟的估计方法, [Carrasco and Florens \(2002\)](#) 和 [Gourieroux et al. \(1993\)](#) 分别推导了在蒙特卡洛试验次数 N_s 固定时, 模拟矩方法和间接推断估计量的渐近分布: 这些估计量仍然具有渐近正态性, 只是渐近方差发生了变化.

第二种情况则是逼近误差受限于样本而不可无限制地减小. 此时, 我们在推导最终估计量的渐近分布时必须考虑到这些逼近误差的影响. 例如第 [5](#) 小节介绍的, 一些机器学习的思想和算法也可以直接应用于结构方程模型的参数估计. 在这类方法中, [Farrell et al. \(2021, 2025\)](#) 和 [Kaji et al. \(2023\)](#) 借助非渐近分析和半参数理论等工具分别分析了其提出的基于深度神经网络和生成对抗网络的估计量的渐近正态性. 这些研究也涉及到了半参数模型的一个重要问题: 当我们用深度神经网络等机器学习方法去估计无穷维干扰参数时, 由于正则化误差和过拟合的存在, 干扰参数估计量通常不具有 $1/\sqrt{n}$ 速度的相合性, 在这种情况下, 如何保证有限维主要参数的估计量仍然具有 $1/\sqrt{n}$ 速度的相合性和渐近正态性? 这也是以 [Chernozhukov et al. \(2018\)](#) 为开端的双重机器学习 (double machine learning) 文献关注和解决的问题.

6.4 高维模型的统计推断

第 6.1 小节介绍的基于渐近正态性的统计推断方法的核心在于使用中心极限定理等工具推导出估计量的渐近分布, 而标准的中心极限定理要求参数和矩的数量有限且固定, 同时样本量要趋于无穷. 在第 5.5 中我们已经提到过, 除非参数和矩的数量随着样本量以很慢的速度发散, 极大似然、广义矩方法和经验似然等估计量在高维模型中不再具有渐近正态性甚至不存在渐近分布. 这一现象背后的本质原因是中心极限定理等渐近理论工具在高维情形下失效. 因此, 为了解决高维模型的推断问题, 研究者需要研究高维模型估计方法的统计性质并在此基础上寻找适合的推断方法, 这一过程可能还需要使用一些非渐近 (non-asymptotic) 的分析工具 (Wainwright, 2019).

van de Geer et al. (2014) 研究了 LASSO 在高维线性回归模型中的统计性质并将分析拓展至广义线性模型, 这些统计性质可以用于构建渐近最优的置信域和检验. Sur and Candès (2019) 和 Sur et al. (2019) 分别研究了高维逻辑斯蒂 (Logistic) 回归中极大似然估计量和似然比检验的渐近性质. Zhu et al. (2020) 针对一般形式的高维模型, 提出了一个约束极大似然方法来进行假设检验. Chang et al. (2021) 则在 (超) 高维模型的惩罚经验似然估计量的基础上, 利用投影方法和经验似然比构建了参数低维分量的置信域, 并用边缘经验似然比构建了过度识别检验.

6.5 贝叶斯推断

截至目前, 本文的讨论都是在频率学派 (frequentist) 的视角下进行: 我们将计量模型的参数当作非随机的未知常量, 并用样本去推断关于这个常量的信息. 与频率学派相对应地, 贝叶斯推断 (Bayesian inference) 则将计量模型的参数也当作随机变量并假定它服从参数空间 $\Psi(\mathcal{M})$ 上的某个先验分布 (prior distribution), 即 $\psi \sim \Pi$. 令 ϖ 为分布 Π 的密度函数. 对于每个给定的参数 ψ , 数据分布与参数之间的依赖关系则类似于频率学派中的有限信息方法, 即设定数据 \mathbf{S}_n 关于参数的条件密度为 $p(\mathbf{S}_n|\psi)$, 它形式上就等于数据的联合似然函数. 特别地, 如果数据 \mathbf{S}_n 是独立同分布样本, 那么

$$p(\mathbf{S}_n|\psi) = \prod_{i=1}^n f(Y_{o,i}, X_i; \psi),$$

其中 f 的定义与 (11) 式中相同.

贝叶斯推断的核心是利用贝叶斯公式, 得到参数关于数据的后验分布 (posterior distribution), 即

$$q(\psi|\mathbf{S}_n) = \frac{\varpi(\psi)p(\mathbf{S}_n|\psi)}{\int_{\Psi(\mathcal{M})} \varpi(\psi')p(\mathbf{S}_n|\psi') d\psi'}.$$

注意到上式右端的分母只依赖于数据而与参数无关, 因此在忽略常数因子的情况下, 参数关于数据的后验分布可以完全地由上式右端的分子刻画, 即

$$q(\psi|\mathbf{S}_n) \propto \varpi(\psi)p(\mathbf{S}_n|\psi). \quad (20)$$

(20) 式的好处是, 它只涉及参数的先验密度函数和数据的联合似然函数, 无需计算积分. 进一步地, 只要 $\varpi(\psi)p(\mathbf{S}_n|\psi)$ 的函数形式已知, 我们就可以使用诸如 MH 算法 (Metropolis–Hastings algorithm) 和吉布斯抽样 (Gibbs sampler) 等具有普遍适用性的马尔科夫链蒙特卡洛 (Markov chain Monte Carlo, MCMC) 方法实现从后验分布 q 中抽样. 当抽样数量足够多时, 抽出的样本的经验分布就可以作为参

数关于数据的后验分布的良好近似, 我们可以基于它开展关于参数的统计推断.

我们在第 5.1 小节中介绍的频率学派的估计方法最终都通过求解某个最优化问题实现, 而我们在前面也已经讨论过求解这些最优化问题的困难. 贝叶斯推断则避开了优化问题的求解, 而是通过抽样来得到对于后验分布的近. 因此, 将参数当作非随机变量还是随机变量只是频率学派和贝叶斯的表面区别, 从深层逻辑看, 贝叶斯推断将频率学派下的最优化问题转化为了抽样问题, 从而在一些求解最优化问题存在困难的情形下变得有吸引力. 此外, 贝叶斯经验似然方法 (Lazar, 2003) 的提出, 使得贝叶斯推断不局限于有限信息的情形, 而是可以扩展到基于总体矩方程的有限信息情形. 我们在第 5.1 小节介绍过, 从总体矩方程出发, 频率学派也可以使用经验似然方法估计参数, 其中会涉及到经验似然函数. 贝叶斯经验似然就是将 (20) 式中的似然函数 $p(\mathbf{S}_n|\psi)$ 替换为了经验似然函数, 并发现传统的贝叶斯推断的性质在进行这样的替换后仍然成立. 对于高维模型, Chang et al. (2025) 提出了贝叶斯惩罚经验似然方法, 借助 MCMC 的优势 (相较于频率学派的惩罚经验似然方法) 进一步降低了计算复杂度.

在经验研究中, Li and Zheng (2009) 构建了存在竞标者的内生进入行为的一价采购拍卖模型, 并通过半参数贝叶斯方法对进入成本、竞标者私人成本分布以及不可观测异质性进行联合估计. 由于文中 (18) 式的出价密度具有由均衡条件内生决定的参数依赖支持集, 标准极大似然估计的正规化假设不再成立. 为此, 文章采用基于 MCMC 的贝叶斯推断方法, 在结构均衡约束下对模型进行估计, 并结合样本内与样本外的拟合表现, 在三种候选模型中选取了最优的一价采购拍卖模型.

7 模型验证

结构方程研究的本质特征是基于经济理论模型对观测数据建模, 它在经济学理论和数据分布之间搭建起一座桥梁, 并在此基础上对基础经济参数进行识别、估计和统计推断, 再将其应用于具体经济问题的分析. 这一本质特征决定了结构方程分析结果的可信度很大程度上依赖于模型设定 (包括假定条件), 即建模过程中使用的经济理论模型能否很好地刻画真实的数据生成过程. 虽然经济学目前已经发展出一套比较完整的理论体系, 但是在对特定经济场景和数据建模的过程中, 研究者仍然需要自行决定很多设定. 当使用参数化的效用函数或生产函数时, 研究者需要选取具体的函数形式, 例如科布-道格拉斯型或拟线性型. 当模型中涉及随机误差或随机系数时, 需要对误差项或系数的分布施加假定, 甚至设定具体的分布族, 例如正态分布或均匀分布. 从理想主义的角度来看, 研究者在给出这些设定时需要尽可能地从理论和数据中寻找证据, 用于论证这些设定的合理性. 但是在现实中, 研究者能够获取的信息是有限的, 通常不足以用于判断设定的合理性. 而且在更多时候, 研究者需要为了后续分析的可行性或简便性选择某个特定的函数形式或分布假定. 例如, 在离散选择模型中, 为了保证条件选择概率有解析表达式, 我们需要假定随机效用函数中的误差项独立同分布地服从第一型极值分布, 但是这一技术性假定同时也蕴含了无关选项之间的独立性, 而这一特性在很多现实应用中并不成立. 因此, 对模型设定和假定条件的质疑成为了威胁结构方程分析结果可信度的主要因素. 为了应对这一挑战, 文献中发展出两条解决路径: 一是设法论证模型的可靠性, 这称为结构方程模型的验证; 二是提升模型误设情形下分析结果的稳健性, 这称为结构方程的透明度 (transparency).

模型验证是对模型可靠性的评估和诊断, 这里的模型指的是代入了参数估计值的模型. 模型的可靠

性应当包含两个方面: 一是内部效度 (internal validity), 即代入了参数估计值的模型与现有数据的吻合程度, 这又进一步涉及模型设定是否与数据相符和参数估计是否准确等问题; 二是外部效度 (external validity), 即代入了参数估计值的模型能否被用于在新的场景下预测某些经济变量, 这涉及到模型的外推或样本外预测能力. 传统的经济学研究非常重视模型的内部效度, 并且提供了大量检验内部效度的方法, 这些方法本质上都是在检验模型在当前样本内的拟合效果. 例如, 在约简式研究中, 研究者会开展大量的稳健性检验, 即在不同的 (约简式) 模型设定、变量定义、样本选取、估计和推断方法之下重复主要分析, 如果得到的结果都非常类似, 就说明分析结果是稳健的. 因为这些稳健性检验大部分都在同一个样本上开展, 即使有子样本分析, 子样本的划分也不是随机的, 所以检验结果衡量的都是模型在样本内的拟合效果, 也就是内部效度.

然而, 一个具有内部效度的模型并不一定具有外部效度, 即一个模型在已有假设和环境下具有好的表现, 并不意味它在新的假设和环境下能继续保持好的表现. 如果我们希望在发生了结构性变迁的情形下分析和预测经济变量, 那么我们就需要一个具有较高的外部效度的模型, 以实现良好的样本外预测能力. 机器学习因为以预测为目标, 所以高度重视模型的样本外预测能力, 即外部效度或泛化 (generalization) 能力. 目前机器学习领域已经有一套标准的检验外部效度的方法: 首先将数据集随机地划分为训练集和测试集, 模型选择和参数估计只能使用训练集数据, 在得到估计的模型后, 再使用测试数据集诊断模型的样本外预测表现; 对于连续型结果变量, 最常用的诊断指标是预测的均方误差 (mean squared error) 和平均绝对误差 (mean absolute error); 对于离散型结果变量, 最常用的诊断指标是预测的平均错分误差 (mean misclassification error) 和平均偏差 (mean deviance). 对此的详细讨论可见 [Hastie et al. \(2009\)](#) 第 7.2 节. 反观经济学领域, 研究者对于模型外部效度的重视程度以及相应的检验方法远远没有达到机器学习的高度, 这也使得经济学研究结果在实际预测中表现欠佳.

不过上述状况已经有好转的趋势, 随着反事实预测需求增加和结构方程研究范式重新被重视, 研究者开始关注结构方程的模型验证和在结构变迁情形下的预测表现. 正如我们在第 1 节讨论的, 经济学理论基础使得结构方程范式在涉及结构性变迁的预测问题中比单纯的统计方法更具优势, 因为基于经济理论的建模方式能够将发生变化的结构与不变的结构分离开, 在不变的结构 (如偏好、技术和制度) 的大框架下, 用变量和/或参数描述变化的结构, 使得整体模型的未知参数可以从已有数据中估计, 并通过反事实分析预测未来的均衡结果. 如果结构方程研究能够以此为基础, 并进一步借鉴机器学习领域关于外部效度检验的思想和方法, 那么就能更好地发挥其在预测方面的优势. 下面我们介绍结构方程模型验证的现有方法.

在进行模型验证时, 研究者需要区分用于估计的数据集和用于测试的数据集. 基于测试数据集, 研究者可以检验模型的外部效度, 进而说明模型的可靠性. 用于测试的数据集需要与用于估计的数据集具有相同的不可观测外生变量的分布, 而可观测的外生变量则被允许在两个数据集中有不同的分布, 在有些情形下, 研究者甚至希望对可观测外生变量进行操纵 (manipulation) 以实现更为全面的验证. 测试数据集的选择通常有以下三种做法 ([Wolpin, 2007](#)).

第一, 利用在模型估计阶段尚未出现的全新数据进行验证, 这类情形通常源于经济环境发生了结构性变化 (regime shift). 此时, 测试数据在模型估计时是不可获得的. 例如, 在 [McFadden et al. \(1977\)](#) 等关于旧金山湾区捷运系统的研究中, 研究者在其尚未建成时构建随机效用模型 (random utility model)

预测其建成后的客流需求，并在系统投入运行后，使用实际客流数据检验模型对需求的预测能力。

第二，在可得的原始数据中预留部分样本作为测试集。当所有样本服从同一数据生成过程时，可以随机抽取部分样本进入测试集。在随机对照实验中，如果随机分配本身不改变偏好、技术、制度底层参数，则实验组与对照组可被视为来自相同分布。在这种情况下，研究者可以使用对照组（或实验组）对结构模型进行估计，再利用实验组（或对照组）对模型进行验证。例如，[Todd and Wolpin \(2006\)](#) 分析了墨西哥的 PROGRESA 项目对入学决策、劳动供给和生育行为的影响。该项目对 506 个社区中实施了随机化分组，其中 320 个社区立即实施补贴，186 个社区延迟两年实施。作者首先利用对照组数据估计动态离散选择模型，然后通过处理组数据进行样本外验证，将模型预测与随机控制实验结果进行对比，说明模型具有较小的预测误差。例如，模型预测项目实施一年后，6 至 11 岁女孩的入学率为 97.1%，与实际数据 98.5% 接近。另一个例子是 [Galiani et al. \(2015\)](#) 利用美国 Moving to Opportunity 项目估计了社区选择模型。实验将家庭随机分成三组。作者使用对照组和受限补贴组数据估计模型，并将无限制补贴组保留用于样本外验证。结果显示，模型对平均贫困率、学校质量、通勤距离等 6 个矩的预测和实际值相近。

第三，在缺乏随机分配的情形下，不同样本往往对应不同的激励机制或经济场景，其数据生成过程在整体上并不相同，导致无法实践上述做法。然而，在给定一组协变量后，不同样本的条件分布可能具有可比性。围绕训练集与测试集在给定协变量条件下分布一致的情形，机器学习领域中的迁移学习 (transfer learning) 已有大量讨论 ([Kouw and Loog, 2019](#); [Sugiyama and Kawanabe, 2012](#))。在结构方程模型中，模型的一些不可观测变量通常被假定在给定可观测变量下具有相同条件分布。因此，我们可以将这些可观测变量纳入模型当中，起到控制变量的作用。在经验研究中，研究者往往有意选择在可观测变量具有差异的数据作为测试集。例如 [Keane and Wolpin \(2007\)](#) 在对女性行为的研究中，将相对于其他五个州福利制度最不慷慨的德州用作保留样本。

值得一提的是，[Galiani and Pantano \(2021\)](#) 在第 5.2 节指出，当前结构方程模型验证中存在这样一种做法：研究者根据模型在“验证集”上的拟合表现不断修改和更新模型设定，并报告达到理想拟合效果的结果。这实质等同于利用全部数据进行模型估计，虽然可以提升模型在样本内的表现，却无法提供关于外部效度的有效信息。这一问题本质上触及到了模型选择和模型评估（也就是本文所称的模型验证）的区别：根据 [Hastie et al. \(2009\)](#) 第 222 页的定义，模型选择指的是估计不同模型的表现以选择出一个最佳模型，模型评估 (model assessment) 指的是在得到最终模型后在新的数据集上估计预测误差。按照这一定义，模型选择必须只在训练数据集上进行，在这一过程中可以进一步切分训练数据集（例如交叉验证），而模型评估（验证）必须只在测试数据集上进行。

在有些研究场景中，由于数据的限制，研究者可能无法获得单独的测试数据集用于模型验证。在这种情形下，如果参数估计方法是广义矩方法或经典最小距离这样的有限信息方法，那么研究者可以考虑使用未参与估计的信息（例如新的矩条件和统计量）来验证模型。以广义矩方法 (18) 为例，研究者需要找到一组新的函数 \tilde{g} ，使得 $\tilde{g} \neq g$ 而且在正确的模型设定之下有 $\mathbb{E}[\tilde{g}(Y_o, X; \psi)] = 0$ ，也就是说找到一组新的未用于参数估计的总体矩条件。基于参数估计量 $\hat{\psi}_{\text{GMM}}$ 和同一个样本，研究者可以计算这组新的总体矩的样本模拟量 $(1/n) \sum_{i=1}^n \tilde{g}(Y_{o,i}, X_i; \hat{\psi}_{\text{GMM}})$ ，如果它充分接近于 0，则可以为模型的可靠性提供一些支撑证据。注意到，这一做法本质上是一种矩约束检验，因此可以使用计量经济学领域的现有方法

得到严格的统计推断结果. 在实际操作中, 如果研究者能够找到一个可以被当前的结构方程模型理性化的约简式模型, 那么就可以使用这一约简式模型蕴含的矩条件和统计量来验证结构方程模型.

以上介绍的模型验证是在设法为模型可靠性提供更多的证据, 而近些年提出的结构方程模型透明度分析则是希望在模型误设的情况下, 为读者提供尽可能多的正确信息. Andrews et al. (2017) 研究了矩估计量对矩方程的敏感程度, 并提供了一个可行的测度方式. 如果一个结构方程模型通过矩方法识别和估计, 那么读者可以计算当矩方程发生偏离 (模型误设) 时估计量的偏误. Andrews et al. (2020a) 发现, 在结构估计中, 描述性统计量也可能包含关于结构参数的有用信息. 在模型误设时, 结构参数的估计量会存在偏误, 但是如果研究者同时提供了相关的描述性统计量, 那么读者可以借助这些描述性统计量中的信息来 (一定程度上) 修正偏误. Andrews et al. (2020b) 则给出了结构方程研究的透明度的一般化定义, 其核心是考察结构参数估计量和辅助统计量能够多大比例地提取出样本中关于真实参数的有用信息.

例 2: 我们将例 1 修改如下:

$$\text{供给: } \ln R = a \ln Q + bZ + U, \quad (21)$$

$$\text{需求: } \ln R = c \ln Q + dZ + V, \quad (22)$$

$$\text{假定: } b \neq 0, \quad \mathbb{E}(ZU) = \mathbb{E}(ZV) = \mathbb{E}(UV) = 0. \quad (23)$$

某研究者在假定 (23) 的基础上, 认为排他性约束 $d = 0$ 也成立, 因此按照例 1 的方式, 用工具变量 Z 在需求曲线中的外生性 $\mathbb{E}(ZV) = 0$ 作为矩条件, 识别了参数

$$c = \frac{\mathbb{E}(Z \ln R)}{\mathbb{E}(Z \ln Q)},$$

并基于独立同分布样本构建了估计量

$$\tilde{c} = \left(\frac{1}{n} \sum_{i=1}^n Z_i \ln R_i \right) \Bigg/ \left(\frac{1}{n} \sum_{i=1}^n Z_i \ln Q_i \right).$$

如果研究者假定的排他性约束 $d = 0$ 成立, 那么 \tilde{c} 是真实参数 c 的相合估计量.

现在考虑研究者模型误设的情形: 在真实的数据生成过程中, $d \neq 0$, 即排他性约束不成立. 求解联立方程 (21) 和 (22) 可得

$$\ln Q = \frac{V - U + (d - b)Z}{a - c}. \quad (24)$$

结合 (24) 式和假定 (23) 可得

$$\mathbb{E}(Z \ln Q) = \beta \mathbb{E}(Z^2), \quad \text{其中 } \beta = \frac{d - b}{a - c}. \quad (25)$$

在此情形下, 估计量 \tilde{c} 的概率极限为

$$\tilde{c} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}(Z \ln R)}{\mathbb{E}(Z \ln Q)} = c + \frac{d \mathbb{E}(Z^2)}{\mathbb{E}(Z \ln Q)} = c + \frac{d}{\beta}.$$

这说明一旦模型误设, \tilde{c} 不再是参数 c 的相合估计量, 其中的偏误是 d/β .

从上面的讨论可以看出, 给定大前提 (21)–(23), 读者是否相信研究者估计出的需求曲线参数 \tilde{c} , 取决于他/她是否相信研究者假定的排他性约束 $d = 0$. 研究者可以在报告核心参数估计量 \tilde{c} 的同时, 报

告其它的统计量以增加结构方程分析的透明度, 即在读者认为排他性约束不成立时, 也能从研究结果中得到尽可能多的有用信息. 在本例中, 研究者可以选择报告 $\ln Q$ 对 Z 的线性回归系数, 即工具变量一阶段回归结果. 由 (25) 可知, 一阶段回归系数的概率极限是 β . 因此读者可以从一阶段回归结果中获得关于 β 的相对准确的信息. 在排他性约束不成立的情形下, 如果读者对 d 的符号或者取值范围有一定的认识, 那么他/她就可以根据偏误公式得到对 \tilde{c} 的偏误的定性刻画. 更进一步地, 如果读者对 d 的取值有一个信念, 那么他/她甚至可以根据这一信念修正 \tilde{c} 的偏误. 相似的例子也见于 Andrews et al. (2020b) 第 2.2 节. \square

在经验研究中, 有一些研究者报告了 Andrews et al. (2017) 所述的敏感性分析的结果. Adda et al. (2017) 基于 743 个矩条件, 采用模拟矩方法识别家庭工资、职业选择、效用函数、婚姻状态等结构参数, 并在附录 B 中报告了敏感性检验. 结果表明, 任一参数变化 1%, 目标函数变化幅度超过 1%. 类似地, Chen et al. (2023) 在附录 E 中报告了结构参数对矩条件的敏感性. 例如, 当投资的序列相关性存在 10% 的误设时, 凸性调整成本参数的估计值下降约 0.059.

8 模型应用

经过估计和验证后的结构方程模型可以用作进一步探索现象背后的原因、机制和一般科学规律, 并指导政策实施. 具体而言, 结构方程模型有以下应用.

第一, 经济机制分析. 约简式通过估计不同设定下因果效应, 给出了不同经济机制的证据. 如果想要在同一个模型下区分不同经济机制的存在和相对大小 (尤其是考虑到非线性的影响渠道时), 结构方程模型提供了更为严格和准确的证据. 注意到 (1) 式的结构方程模型本身就是对经济机制的刻画, 其中的参数 θ 具有经济学意义. 通过显性考虑各种变量, 结构模型在构建时, 对不同经济机制的做出了阐述和定义. 在经过参数识别、估计、统计推断和模型验证之后, 我们可以获得关于结构参数 θ 的较为可靠的估计量 $\hat{\theta}$ (它必然是参数估计量 $\hat{\psi}$ 的一部分). 将 $\hat{\theta}$ 与必要的干扰参数估计量相结合, 我们即可得到一个完整的经济学模型, 进而分析各种经济机制. 例如 Wang et al. (2022) 建立了一个银行业的一般均衡模型, 在完成结构估计后, 利用该模型分析了银行的市场势力与货币政策的传导机制. Page (2018) 刻画了公司 CEO 对公司价值的四种影响机制: 风险厌恶、努力厌恶、外部选择和对董事会的影响. Allcott et al. (2022) 将数字成瘾背后的习惯养成 (habit formation) 和自我控制 (self control) 机制加以区分.

第二, 社会福利分析. 在一定条件下, 足量的约简式可以用作元分析 (meta-analysis), 也可以在不需要复杂建模的情形下做福利分析 (Chetty, 2009). 例如, Einav et al. (2010) 在分析保险市场时, 在不同价格点估计局部需求弹性, 进而恢复需求曲线. 而结构方程模型可以更加广泛的应用于福利分析, 直接计算福利水平, 进而进行模型比较. 根据古典需求理论 (Mas-Colell et al., 1995 第 3 章), 我们可以从效用函数出发计算支出函数, 再利用支出函数计算当经济状态 (价格) 发生变化时消费者的福利变化. 根据竞争性局部均衡理论 (Mas-Colell et al., 1995 第 10 章), 我们可以通过对需求曲线和供给曲线积分计算消费者和生产者剩余. 这些福利计算需要已知结构参数 θ 的取值, 因此在完成结构方程模型的估计和验证之后可以执行. 例如, Meghir et al. (2022) 估计了孟加拉国地区的迁移补贴政策使得总福

利增加 12.9 %. [Nevo et al. \(2016\)](#) 发现以三段式计价的上网套餐对总福利影响不大, 但会使得消费者的福利转移至生产者.

第三, 反事实分析. 反事实分析可以看做结构方程与因果推断的结合. 在验证模型具有预测和外推能力后, 研究者可以进行反事实分析并设计新的政策. 与约简式的因果推断分析相比, 结构方程的反事实分析的优势是可以估计不同于已有政策的效果, 即政策的事前评估 (*ex ante* evaluation), 这包括: (1) 外推政策到新的时间或人群. 例如, [Todd and Wolpin \(2006\)](#) 的教育补贴对三年级以上的学生发放. 作者通过模型参数估计了另一种政策设计——取消三至五年级的补贴, 将节省的资金转移至六至九年级. 结果显示, 女孩完成九年级的比例由 19.8% 上升至 26.2%. (2) 评估全新的政策设计. 例如, [Buchak \(2024\)](#) 研究了网约车市场的融资约束和劳动供给问题. 现有市场结构要求劳动者必须自有车辆才能注册为司机. 作者考虑另一种制度安排, 允许劳动者通过车辆共享或租赁平台参与市场. 反事实分析表明, 这种资本和劳动的分离会虽然会提高网约车市场规模, 但低收入群体福利大幅下降.

第四, 最优政策设计. 在反事实分析的基础上, 研究者可以使用具有高可信度的结构方程模型进行最优政策的设计. 例如, 如果以社会福利作为比较不同政策的标准, 那么研究者可以通过结构方程模型模拟不同政策下的社会福利, 寻找可行的能够最大化行为人福利的参数组合. 在机器学习中, 与这类问题相关的研究是政策学习 (policy learning), 可参见 [Athey and Wager \(2021\)](#).

第五, 检验经济学理论. 事实上, 第 7 节所讨论的模型验证在很大程度上也是对经济学理论的验证, 因为结构方程模型本身就是经济学理论模型. 当然, 模型验证本身更侧重于对模型所代表的经济学理论整体的检验. 如果我们想检验一个经济学理论中的一些特定要素, 那么我们也可以对针对这些要素做相应的模型设定检验; 特别地, 如果我们感兴趣的理论要素可以被模型中的一些参数刻画, 那么我们可以直接对这些参数做假设检验. 当同一个经济现象存在多种理论时, 我们可以考虑构建一个最一般化的模型将这些理论同时包含进去, 并用通过一些参数来确定具体的理论. 一个简单的类比就是在常替代弹性 (constant elasticity of substitution) 效用函数或生产函数中, 我们可以通过设定替代弹性的具体取值使其成为线性、柯布-道格拉斯或者里昂惕夫效用函数或生成函数. 在这个一般化的模型之下, 我们可以基于数据对这些参数进行估计和统计推断, 并由此判断哪个具体理论是最符合实际的. 例如, [Ketcham et al. \(2016\)](#) 和 [Abaluck and Gruber \(2011\)](#) 比较了不同的健康保险选择模型, 并计算了不同模型下的支付意愿 (willingness to pay) 和福利水平.

相比之下, 约简式研究虽然也经常用于检验经济学理论, 但是它难以实现模型层面的检验, 而是更多地通过理论推导出来的特征事实来检验理论. 用更加严谨的语言来表述就是, 结构方程研究有能力检验理论本身, 而约简式研究通常只能检验理论的必要条件, 这被称为证伪检验 (falsification test). 这一特点降低了约简式研究对理论的检验功效. 例如研究者通过某个理论推导出一些特征事实, 但是这些特征事实并不必然只在这个特定理论之下才成立, 可能有其它的替代理论也能推出同样的特征事实. 在这一情况下, 如果研究者只通过约简式研究来检验这些特征事实, 就可能出现理论不成立但检验仍然通过 (不拒绝) 的情形.

9 总结与展望

1932 年考尔斯委员会的成立标志着经济学经验研究新范式, 其核心思想是通过结构方程刻画人类行为与社会制度的基础参数. 结构方程对行为人与外部环境进行定义和建模, 其模型参数具有明确的经济学含义; 研究者通过分离方程中不变与可变的要素, 开展反事实推断、政策模拟和福利评估. 综合来看, 结构方程模型具有不可替代的优势. 首先, 它能够刻画行为人之间的相互作用, 识别同伴效应 (peer effect)、价格效应、溢出效应 (spillover effect) 与一般均衡效应等. 其次, 基于动态优化框架的结构模型具有跨时间的估计能力, 能够预测超出样本期的长期结果. 此外, 结构方程模型可以识别和估计不可观测但具有经济学意义的参数, 如边际成本、规模报酬、需求价格弹性、生产率等. 进一步地, 结构模型具备较强的外部效度与政策泛化能力, 既可用于政策的外推分析, 也能进行政策设计, 而约简式方法通常仅限于政策的事后评估. 同时, 结构方程框架还可用于比较竞争性理论模型. 例如, 通过构建数量竞争与价格竞争的市场模型, 检验哪种行为假设更符合数据表现. 更重要的是, 结构模型具有丰富的政策含义, 可广泛应用于成本-效益分析、福利测算、政策仿真和宏观预测等领域, 为经济政策评估和设计提供系统化工具.

当然, 结构方程研究也存在着一些局限和挑战: 模型复杂度较高、证明参数的可识别性较为困难、高维非线性环境下求解的计算负担较重、有效性依赖于大量行为与分布假设等. 近年来, 约简式方程和结构方程的结合增强了识别的严谨性, 深度学习与强化学习方法的引入则在一定程度上缓解了维数诅咒, 降低了计算复杂度, 为结构方程模型在政策评估和经济预测中的应用和拓展奠定了基础. 未来, 随着计量经济学与机器学习更多的结合, 结构方程研究有望在理论与经验研究之间搭建更稳固的桥梁, 进一步推动经济学由事后分析走向更多的事前预测. 因此, 我们建议广大经济学研究者: 在实际研究中应当根据问题性质、研究目标和数据特征, 合理选取研究范式并综合使用各种量化分析方法, 以科学的态度探索社会经济发展的一般规律.

受篇幅所限, 本文未能进一步讨论结构方程模型中的若干相关话题. 第一, 本文的主要注意力集中在参数模型和参数方法上. 尽管第 4.3 小节涉及了非参数识别问题, 但后文并未系统地讨论相应的非参数估计与推断方法. 对于半参数或非参数结构方程模型, 如果我们关注的元素可以被非参数地识别, 那么在此基础上我们可以使用计量经济学中已经发展成熟的非参数方法来实现参数估计和统计推断, 例如核 (kernel) 方法 (可参见 Bosq, 2012) 和缝合 (sieve) 方法 (可参见 Chen, 2007) 等. 特别地, 我们在第 5.4 小节介绍的神经网络本质上属于一种非线性的缝合方法. 当然, 结构方程模型本身的复杂性也会给这些非参数估计和推断方法带来困难, 其中有一些困难可以继续使用第 5–6 节讨论的方法解决, 例如用基于模拟的估计方法解决数值积分问题, 但是另外一些困难可能需要进一步研究解决方案.

第二, 本文讨论的结构方程模型都是完全结构式的模型, 但是在完全结构式和完全约简式的模型中间还存在着一类半结构式模型. 半结构式模型也具有一定的经济学理论基础, 但是根据研究需要和可操作性对模型形式进行了适当的简化. 一个例子是边际处理效应 (marginal treatment effect) 模型 (Heckman and Vytlacil, 2007), 它在因果推断的潜在结果模型基础上增加了样本选择方程, 这一选择方程的理论基础是随机效用最大化. 另一个例子是福利分析的充分统计量 (sufficient statistics) 方法 (Chetty, 2009; Kleven, 2021). 在统计学中, 某个参数的充分统计量的含义是: 给定这个统计量, 数据的

条件分布不再依赖于参数; 即这个统计量包含了样本中关于相应参数的全部信息. 这一概念在经济学福利分析中被借用并拓展: 在一个完全结构式的模型之下, 一些福利分析 (或更一般的经济学分析) 并不要求研究者获知全部基础参数, 而是只需要使用一些包含与福利分析相关的所有信息的统计量 (即充分统计量). 这样一来, 研究者可以只关注这些充分统计量, 从而规避一些由基础参数的识别和估计带来的技术性困难.

第三, 本文第 5.5 小节介绍的高维模型估计方法都要求模型对应的似然函数或矩函数具有解析表达式, 但是这一要求对于大部分结构方程模型都不成立. 对于含有高维参数的结构方程模型, 我们仍然需要在第 5.5 小节介绍的估计方法基础上处理由模型求解和数值积分等操作带来的技术性困难, 例如如何将高维模型估计方法与模拟方法相结合, 结合之后估计量具有何种统计性质. 这些开放问题值得在未来被进一步研究.

第四, 除了第 6.5 小节介绍的贝叶斯推断之外, 本文其余部分讨论的统计推断都要求大样本. 在基于设计的因果推断 (design-based causal inference) 中, 有一类常用的有限样本统计推断方法是随机化检验 (randomization test), 例如 Abadie et al. (2010) 中使用的安慰剂检验 (placebo test). 这类思想能否以及如何用于结构方程模型的有限样本统计推断, 也是一个值得研究的问题.

最后, 读者如果希望从更多的视角来了解结构方程, 还可以阅读 Low and Meghir (2017) 和 Galiani and Pantano (2021) 等综述.

参考文献

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Abaluck, J. and Gruber, J. (2011). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program. *American Economic Review*, 101(4):1180–1210.

Abaluck, J. and Gruber, J. (2016a). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program: Reply. *American Economic Review*, 106(12):3962–3987.

Abaluck, J. and Gruber, J. (2016b). Evolving choice inconsistencies in the choice of prescription drug insurance. *American Economic Review*, 106(8):2145–2184.

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *The Review of Economic Studies*, 89(1):45–86.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Adda, J. and Dustmann, C. (2023). Sources of wage growth. *Journal of Political Economy*, 131(2):456–503.

Adda, J., Dustmann, C., and Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, 125(2):293–337.

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review*, 8(1):67–77.

Allcott, H., Gentzkow, M., and Song, L. (2022). Digital addiction. *American Economic Review*, 112(7):2424–2463.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human Resources*, 40(4):791–821.

Altonji, J. G., Smith Jr, A. A., and Vidangos, I. (2013). Modeling earnings dynamics. *Econometrica*, 81(4):1395–1454.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2020a). On the informativeness of descriptive statistics for structural estimates. *Econometrica*, 88(6):2231–2258.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2020b). Transparency in structural research. *Journal of Business & Economic Statistics*, 38(4):711–722.

Arcidiacono, P., Bayer, P., Blevins, J. R., and Ellickson, P. B. (2016). Estimation of dynamic discrete choice models in continuous time with an application to retail competition. *The Review of Economic Studies*, 83(3):889–931.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

Attanasio, O. P., Meghir, C., and Santiago, A. (2012). Education choices in Mexico: Using a structural model and a randomized experiment to evaluate PROGRESA. *The Review of Economic Studies*, 79(1):37–66.

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2021). Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Azinovic, M., Gaegau, L., and Scheidegger, S. (2022). Deep equilibrium nets. *International Economic Review*, 63(4):1471–1525.

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945.

Barron, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Machine Learning*, 14(1):115–133.

Barseghyan, L., Coughlin, M., Molinari, F., and Teitelbaum, J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048.

Beraja, M., Fuster, A., Hurst, E., and Vavra, J. (2019). Regional heterogeneity and the refinancing channel of monetary policy. *The Quarterly Journal of Economics*, 134(1):109–183.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797.

Bilal, A. and Goyal, S. (2025). Some pleasant sequence-space arithmetic in continuous time. NBER Working Paper No. 33525.

Blundell, R., Costa Dias, M., Meghir, C., and Shaw, J. (2016). Female labor supply, human capital, and welfare reform. *Econometrica*, 84(5):1705–1753.

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: Estimation and prediction*. Springer Science & Business Media.

Brumm, J. and Scheidegger, S. (2017). Using adaptive sparse grids to solve high-dimensional dynamic models. *Econometrica*, 85(5):1575–1612.

Buchak, G. (2024). Financing the gig economy. *The Journal of Finance*, 79(1):219–256.

Candès, E. J. and Sur, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *The Annals of Statistics*, 48(1):27–42.

Candler, G. V. (2001). Finite-difference methods for continuous-time dynamic programming. In Marimon, R. and Scott, A., editors, *Computational Methods for the Study of Dynamic Economies*, chapter 8. Oxford University Press.

Carrasco, M. and Florens, J.-P. (2002). Simulation-based method of moments and efficiency. *Journal of Business & Economic Statistics*, 20(4):482–492.

Chang, J., Chen, S. X., Tang, C. Y., and Wu, T. T. (2021). High-dimensional empirical likelihood inference. *Biometrika*, 108(1):127–147.

Chang, J., Shi, Z., and Zhang, J. (2023). Culling the herd of moments with penalized empirical likelihood. *Journal of Business & Economic Statistics*, 41(3):791–805.

Chang, J., Tang, C. Y., and Wu, T. T. (2018). A new scope of penalized empirical likelihood with high-dimensional estimating equations. *The Annals of Statistics*, 46(6B):3185–3216.

Chang, J., Tang, C. Y., and Zhu, Y. (2025). Bayesian penalized empirical likelihood and Markov Chain Monte Carlo sampling. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(4):1127–1149.

Chen, H., Didisheim, A., and Scheidegger, S. (2021). Deep structural estimation: With an application to option pricing. arXiv preprint arXiv:2102.09209.

Chen, S. X., Peng, L., and Qin, Y.-L. (2009). Effects of data dimension on empirical likelihood. *Biometrika*, 96(3):711–722.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. J. and Leamer, E., editors, *Handbook of Econometrics Volume 6B*, chapter 76. Elsevier.

Chen, Z., Jiang, X., Liu, Z., Serrato, J. C. S., and Xu, D. Y. (2023). Tax policy and lumpy investment behaviour: Evidence from China's VAT reform. *The Review of Economic Studies*, 90(2):634–674.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chesher, A., Rosen, A. M., and Smolinski, K. (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics*, 4(2):157–196.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, 1(1):451–488.

Chiu, J., Davoodalhosseini, S. M., Jiang, J., and Zhu, Y. (2023). Bank market power and central bank digital currency: Theory and quantitative assessment. *Journal of Political Economy*, 131(5):1213–1248.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828.

Cocci, M. D. and Plagborg-Møller, M. (2025). Standard errors for calibrated parameters. *Review of Economic Studies*, 92(5):2952–2978.

Crawford, I. and De Rock, B. (2014). Empirical revealed preference. *Annual Review of Economics*, 6(1):503–524.

Davis, M. A., Fisher, J. D. M., and Whited, T. M. (2014). Macroeconomic implications of agglomeration. *Econometrica*, 82(2):731–764.

de Paula, Á., Richards-Shubik, S., and Tamer, E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, 86(1):263–288.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, 127(1):1–56.

Dey, M. S. and Flinn, C. J. (2005). An equilibrium model of health insurance provision and wage determination. *Econometrica*, 73(2):571–627.

Dey, M. S. and Flinn, C. J. (2008). Household search and health insurance coverage. *Journal of Econometrics*, 145(1-2):43–63.

Donald, S. G., Imbens, G. W., and Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117(1):55–93.

Drechsler, I., Savov, A., and Schnabl, P. (2017). The deposits channel of monetary policy. *The Quarterly Journal of Economics*, 132(4):1819–1876.

Duarte, V., Duarte, D., and Silva, D. H. (2024). Machine learning for continuous-time finance. *The Review of Financial Studies*, 37(11):3217–3271.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from French firms. *Econometrica*, 79(5):1453–1498.

Einav, L., Finkelstein, A., and Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *The quarterly journal of economics*, 125(3):877–921.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). Gpts are gpts: Labor market impact potential of llms. *Science*, 384(6702):1306–1308.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, J., Lv, J., and Qi, L. (2011). Sparse high-dimensional models in economics. *Annual Review of Economics*, 3(1):291–317.

Farrell, M. H., Liang, T., and Misra, S. (2021). Deep neural networks for estimation and inference. *Econometrica*, 89(1):181–213.

Farrell, M. H., Liang, T., and Misra, S. (2025). Deep learning for individual heterogeneity. arXiv preprint arXiv:2010.14694v3.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370.

Fernández-Villaverde, J., Hurtado, S., and Nuno, G. (2023). Financial frictions and the wealth distribution. *Econometrica*, 91(3):869–901.

Fernández-Villaverde, J., Marbet, J., Nuño, G., and Rachedi, O. (2025). Inequality and the zero lower bound. *Journal of Econometrics*, 249(C):105819.

Fernández-Villaverde, J., Nuno, G., Sorg-Langhans, G., and Vogler, M. (2020). Solving high-dimensional dynamic programming problems using deep learning. Unpublished working paper.

Fleming, W. H. and Rishel, R. W. (2012). *Deterministic and Stochastic Optimal Control*. Springer Science & Business Media.

Forneron, J.-J. (2024). Detecting identification failure in moment condition models. *Journal of Econometrics*, 238(1):105552.

Fox, J. T. and Gandhi, A. (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *The RAND Journal of Economics*, 47(1):118–139.

Fu, C. and Gregory, J. (2019). Estimation of an equilibrium model with externalities: Post-disaster neighborhood rebuilding. *Econometrica*, 87(2):387–421.

Galiani, S., Murphy, A., and Pantano, J. (2015). Estimating neighborhood choice models: Lessons from a housing assistance experiment. *American Economic Review*, 105(11):3385–3415.

Galiani, S. and Pantano, J. (2021). Structural models: Inception and frontier. NBER Working Paper No. w28698, Available at SSRN: <https://ssrn.com/abstract=3829396>.

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016.

Geng, S., Nassif, H., and Manzanares, C. A. (2023). A data-driven state aggregation approach for dynamic discrete choice models. In Evans, R. J. and Shpitser, I., editors, *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 647–657. PMLR.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Glawion, R. (2023). Sequence-space Jacobians in continuous time. Available at SSRN: <https://ssrn.com/abstract=4504829>.

Goettler, R. L. and Gordon, B. R. (2011). Does AMD spur Intel to innovate more? *Journal of Political Economy*, 119(6):1141–1200.

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118.

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7(1):465–485.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574.

Guo, R., Yi, J., Zhang, J., and Zhang, N. (2025). Rationed fertility: Treatment effect heterogeneity in the child quantity–quality tradeoff. *Journal of Political Economy*, 133(10):3349–3386.

Haile, P. A. and Tamer, E. (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy*, 111(1):1–51.

Hampole, M., Papanikolaou, D., Schmidt, L. D., and Seegmiller, B. (2025). Artificial intelligence and the labor market. NBER Working Paper No. 33509.

Han, J., Jentzen, A., and E, W. (2018). Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510.

Handa, K., Tamkin, A., McCain, M., Huang, S., Durmus, E., Heck, S., Mueller, J., Hong, J., Ritchie, S., Belonax, T., et al. (2025). Which economic tasks are performed with ai? evidence from millions of claude conversations. arXiv preprint arXiv:2503.04761.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics Volume 6B*, chapter 71. Elsevier.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

Jorgensen, S., Quincampoix, M., and Vincent, T. L., editors (2007). *Advances in Dynamic Game Theory: Numerical Methods, Algorithms, and Applications to Ecology and Economics*. Springer Science & Business Media.

Kahou, M. E., Fernández-Villaverde, J., Perla, J., and Sood, A. (2021). Exploiting symmetry in high-dimensional dynamic programming. NBER Working Paper No. w28981, Available at SSRN: <https://ssrn.com/abstract=3880214>.

Kaji, T., Manresa, E., and Pouliot, G. (2023). An adversarial approach to structural estimation. *Econometrica*, 91(6):2041–2063.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, 105(3):473–522.

Keane, M. P. and Wolpin, K. I. (2007). Exploring the usefulness of a nonrandom holdout sample for model validation: Welfare effects on female behavior. *International Economic Review*, 48(4):1351–1378.

Ketcham, J. D., Kuminoff, N. V., and Powers, C. A. (2016). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program: Comment. *American Economic Review*, 106(12):3932–3961.

Kleven, H. J. (2021). Sufficient statistics revisited. *Annual Review of Economics*, 13(1):515–538.

Kouw, W. M. and Loog, M. (2019). An introduction to domain adaptation and transfer learning. arXiv preprint arXiv:1812.11806v2.

Kwon, S. and Kim, Y. (2012). Large sample properties of the SCAD-penalized maximum likelihood estimation on high dimensions. *Statistica Sinica*, 22(2):629–653.

Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika*, 90(2):319–326.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903.

Li, T. and Zheng, X. (2009). Entry and competition effects in first-price auctions: Theory and evidence from procurement auctions. *The Review of Economic Studies*, 76(4):1397–1429.

Low, H. and Meghir, C. (2017). The use of structural models in econometrics. *Journal of Economic Perspectives*, 31(2):33–58.

Low, H. and Pistaferri, L. (2015). Disability insurance and the dynamics of the incentive insurance trade-off. *American Economic Review*, 105(10):2986–3029.

Magnac, T. and Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816.

Maliar, L. and Maliar, S. (2014). Numerical methods for large-scale dynamic economic models. In Schmedders, K. and Judd, K. L., editors, *Handbook of Computational Economics Volume 3*, chapter 7. Elsevier.

Maliar, L., Maliar, S., and Winant, P. (2021). Deep learning for solving dynamic economic models. *Journal of Monetary Economics*, 122:76–101.

Manski, C. F. (2007). Partial identification of counterfactual choice probabilities. *International Economic Review*, 48(4):1393–1410.

Manski, C. F. and Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, 70(2):519–546.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

Matzkin, R. L. (2007). Nonparametric identification. In Heckman, J. J. and Leamer, E., editors, *Handbook of Econometrics Volume 6B*, chapter 73. Elsevier.

Matzkin, R. L. (2013). Nonparametric identification in structural economic models. *Annual Review of Economics*, 5(1):457–486.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3(4):303–328.

McFadden, D., Talvitie, A., and Associates (1977). Demand model estimation and validation. Technical Report UCB-ITS-SR-77-9, Institute of Transportation Studies, University of California. Urban Travel Demand Forecasting Project, Phase I Final Report Series, Volume V.

Meghir, C., Mobarak, A. M., Mommaerts, C., and Morten, M. (2022). Migration and informal insurance: Evidence from a randomized controlled trial and a structural model. *The Review of Economic Studies*, 89(1):452–480.

Molinari, F. (2020). Microeconometrics with partial identification. In Durlauf, S., Hansen, L. P., Heckman, J. J., and Matzkin, R. L., editors, *Handbook of Econometrics Volume 7A*, chapter 5. Elsevier.

Myong, S., Park, J., and Yi, J. (2021). Social norms and fertility. *Journal of the European Economic Association*, 19(5):2429–2466.

Nevo, A., Turner, J. L., and Williams, J. W. (2016). Usage-based pricing and demand for residential broadband. *Econometrica*, 84(2):411–443.

Newey, W. K. and McFadden, D. L. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics Volume 4*, chapter 36. Elsevier.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Page, T. B. (2018). Ceo attributes, compensation, and firm value: Evidence from a structural estimation. *Journal of Financial Economics*, 128(2):378–401.

Payne, J., Rebei, A., and Yang, Y. (2024). Deep learning for search and matching models. Swiss Finance Institute Research Paper No. 25-05, Available at SSRN: <https://ssrn.com/abstract=4768566>.

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *The Annals of Statistics*, 16(1):356–366.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3 edition.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22(1):300–325.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, 33(3):649–665.

Ruder, S. (2017). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747v2.

Ruiz, F. J. R., Athey, S., and Blei, D. M. (2020). SHOPPER: A probabilistic model of consumer choice with substitutes and complements. *The Annals of Applied Statistics*, 14(1):1–27.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Scheidegger, S. and Bilionis, I. (2019). Machine learning for high-dimensional dynamic stochastic economies. *Journal of Computational Science*, 33:68–82.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, 88(5):1829–1858.

Shi, Z. (2016). Econometric estimation with high-dimensional moment equalities. *Journal of Econometrics*, 195(1):104–119.

Sirignano, J. and Spiliopoulos, K. (2018). DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364.

Stokey, N. L., Lucas Jr., R. E., and Prescott, E. C. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Sugiyama, M. and Kawanabe, M. (2012). *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. The MIT Press.

Sur, P. and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525.

Sur, P., Chen, Y., and Candès, E. J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probability Theory and Related Fields*, 175(1):487–558.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Todd, P. E. and Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review*, 96(5):1384–1417.

Todd, P. E. and Wolpin, K. I. (2023). The best of both worlds: Combining randomized controlled trials with structural modeling. *Journal of Economic Literature*, 61(1):41–85.

Train, K. E. (1999). Mixed logit models for recreation demand. In Herriges, J. and Kling, C., editors, *Valuing Recreation and the Environment*, chapter 4. Edward Elgar Publishing Limited.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica*, 50(4):945–973.

Varian, H. R. (1983). Non-parametric tests of consumer behaviour. *The Review of Economic Studies*, 50(1):99–110.

Voena, A. (2015). Yours, mine, and ours: Do divorce laws affect the intertemporal behavior of married couples? *American Economic Review*, 105(8):2295–2332.

Vovk, V., Nouretdinov, I., and Gammerman, A. (2009). On-line predictive linear regression. *The Annals of Statistics*, 37(3):1566–1590.

Wainwright, M. J. (2019). *High-dimensional Statistics: A Non-asymptotic Viewpoint*. Cambridge University Press.

Wang, J. and Forsyth, P. A. (2010). Numerical solution of the Hamilton–Jacobi–Bellman formulation for continuous time mean variance asset allocation. *Journal of Economic Dynamics and Control*, 34(2):207–230.

Wang, Y., Whited, T. M., Wu, Y., and Xiao, K. (2022). Bank market power and monetary policy transmission: Evidence from a structural estimation. *The Journal of Finance*, 77(4):2093–2141.

Wolpin, K. I. (2007). Ex ante policy evaluation, structural estimation, and model selection. *American Economic Review*, 97(2):48–52.

Wright, J. H. (2003). Detecting lack of identification in GMM. *Econometric Theory*, 19(2):322–330.

Xiao, K. (2020). Monetary transmission through shadow banks. *The Review of Financial Studies*, 33(6):2379–2420.

Zhu, Y., Shen, X., and Pan, W. (2020). On high-dimensional constrained maximum likelihood inference. *Journal of the American Statistical Association*, 115(529):217–230.

陈诗一、刘朝良、冯博 (2019):《资本配置效率、城市规模分布与福利分析》,《经济研究》第2期。

冯笑、王楚男 (2022):《寡头市场结构下中国进口汽车税费改革的福利效应分析》,《世界经济》第4期。

冯笑、王永进、戴冕 (2024):《本地化成本、不完全传递与最优关税——以中国汽车行业为例》,《经济学 (季刊)》第6期。

郭晓丹、王帆 (2024):《“双碳”目标下政府补贴、需求替代与减排效应——来自中国乘用车市场的证据》,《数量经济技术经济研究》第2期。

李国栋、罗瑞琦、谷永芬 (2019):《政府推广政策与新能源汽车需求: 来自上海的证据》,《中国工业经济》第4期。

刘修岩、李松林 (2017):《房价、迁移摩擦与中国城市的规模分布——理论模型与结构式估计》,《经济研究》第7期。

谭用、邱斌、叶迪、綦建红 (2024):《中国创新模式选择: 自主创新抑或技术引进?》,《经济研究》第4期。

谭用、周洺竹、綦建红 (2024):《不确定性与中国粮食分散进口: 结构估计与反事实研究》,《经济学 (季刊)》第2期。

王子、周雁翎 (2019):《结构模型在国际贸易研究中的应用》,《中国工业经济》第4期。

徐彤、殷越、王擎 (2023):《家庭住房需求弹性和最优房产税设计》,《经济研究》第8期。

杨继生、邹建文 (2021):《人口老龄化、老年人消费及其结构异质性——基于时变消费效用的分析》,《经济学动态》第11期。

易行健、李青塬、吴卫星、杨碧云 (2025):《老年家庭消费储蓄行为研究新进展》,《保险研究》第9期。

余森杰、杨伯燦、徐铭梽、王勇（2025）：《行政区经济、经济区经济到全国统一大市场——基于结构模型下量化最优补贴政策的分析》，《数量经济技术经济研究》第8期。