

社会资本的劳动力市场效应估算*

——关于内生性问题的文献回溯和研究策略

陈云松 范晓光

提要：内生性问题是社会科学定量分析面临的重要挑战之一。如何通过不同的模型识别策略来消减内生性问题，判定因果关系，成为了近来西方社会资本研究领域的前沿议题。本文围绕社会资本的劳动力市场效应，从“使用社会关系”、“使用了的社会资本”和“可使用的社会资本”这三条主线，重点回溯了相关研究中的内生性问题和模型识别策略。最后，本文提出今后社会资本实证研究中模型设置、数据搜集纵向跨度、横向跨度、外生变量挖掘、研究假定说明和复制性研究等六大努力方向。

关键词：社会资本 内生性 因果关系 劳动力市场

一、引言

“社会资本”因为定义含混不清而长久以来颇受争议（Portes, 1998; Durlauf, 1999; Manski, 2000; 边燕杰, 2004; 张文宏、李沛良、阮丹青, 2004; 罗家德、赵延东, 2005: 100-142）。本文所涉及的社会资本仅指向个人层面，而非集体层面的信任、社会规范等等。我们把社会资本理解为“嵌入于一种社会结构中的可以在有目的的行动中涉取或动员的资源”（Nan Lin, 2001: 29）。由于“嵌入性资源”是一个宽泛的定义，这就使得社会学家对个体层面的社会资本与曼斯基（Manski）和杜尔拉夫（Durlauf）等经济学家对社会资本的理解得以兼容。比如，莫维（Mouw）在一篇重要的文献回溯里就指出，同群效应(peer effects)或邻里效应(neighborhood effects)引起的社会互动（social interaction）都可以被“归入社会资本的定義”（Mouw, 2006）。

* 本文初稿曾投“社会网络与社会资本研究专业委员会 2010 年青年论坛”（山东大学社会学系，2010 年 4 月）。作者感谢牛津大学统计学系 Tom Snijders 教授，社会学系 Peter Hedström 和 Nan Dirk de Graaf 教授，中山大学社会学与社会工作系梁玉成副教授、中国人民大学社会学系王卫东副教授、西安交通大学社会学系张顺副教授、新加坡南洋理工大学经济学系吴桂英助理教授、美国西北大学政治学博士候选人章奇、牛津大学经济学系博士候选人李志赞的批评和建议。当然，文责自负。

就此而言，凡是能影响个体行为或结果的某个参照群体（reference group）的特征、行为或结果，都可以被视为嵌入性资源（类似定义参见 Durlauf, 1999 以及 Portes, 1998），也即社会资本。本文采用这一宽泛的社会资本概念，是因为它可以提供了更为丰富的文献。这些文献资料中有关因果推断识别的策略，为我们如何在今后的经验研究中纠正估计偏误提供了经验和教训。

在劳动力市场领域，社会资本理论的主要论点就是：嵌入性资源（embedded resources），也即社会网络以及社会网所附带的信息、声望、权力等等，能够显著地影响着个体劳动力市场结果（如就业、职业流动、工资、职业声望、工作满意度和岗位晋升等等）。尽管在研究旨趣、分析传统和方法论上存在差异，社会学家和经济学家们长期以来一直关注着社会资本在劳动力市场的作用。在过去几十年中，东西方学界积累了大量有关社会资本和劳动力市场结果的经验文献（社会学回顾性文献参见 Granovetter, 1995/1974；N. Lin, 1999 和 Mouw, 2003；经济学回顾性文献参见 Angrist & Krueger, 1999；Ioannides & Loury, 2004；关于社会资本理论模型的文献参见 Montgomery, 1991；Topa, G., 2001；Calvó-Armengol & Zenou, 2005 等）。这些研究基本都是基于调查数据而非实验数据，因此，它们都必须面对内生性的问题，也即回归模型中的遗漏变量与解释变量相关，而导致参数估计值的偏误，影响因果推断。注意，这里的因果关系，是基于反事实框架（counterfactual framework）的。

长期以来，社会学经验研究对内生性问题关注较少。不过，在一个相对较新的社会资本研究中，莫维（Mouw, 2003）批判性地对以往有关社会资本劳动力市场效应的社会学研究作了梳理。他认为友谊形成过程中的“自选择”（self-selection）^①至少可以部分地取代社会资本的论理解释。尽管他主要关注的仅仅是众多内生性偏误来源之一的“自选择”问题，但他的研究启发了社会学家进一步专注类似的内生性问题。更新的回溯性研究则开始系统性反思社会资本研究中更多类型的内生性问题，并大量借鉴计量经济学、心理学等领域的最新成果（Mouw, 2006）。

内生性问题对于社会学分析而言，它仅仅是一个微不足道的议题还是非常重要的问题？是一种普遍存在的现象还是现有研究的某个细枝末节？我们认为，正如莫维所言，如果社会资本对劳动力市场结果的影响甚微，那么“或许劳动力市场经济模型可以放心地忽略嵌入

① 所谓“自选择”，也即社会关系的形成是选择性的，物以类聚，人以群分。如果劳动力市场结果的差异可以用某种非观测的选择因素来解释，那么社会资本效应的参数估计就可能被高估，甚至根本无从证实（例如，朋友对我的影响，既可以归因于朋友的社会资本，也可以归因于我的某种特质促使我选择某一类朋友）。

性 (embeddedness), 因为它对模型的解释力毫无影响”(Mouw, 2003)。这里最核心的问题在于, 倘若没有解释力, 社会资本的概念和视角将被沦为“仅仅是模棱两可词藻的教训”(Manski, 2003)。如果社会学的研究兴趣在于确定某种机制或因果关系, 而非局限于描述自变量与应变量之间的统计相关, 那么这些内生性偏误的来源应该在社会学分析中被正式提出并加以解决。正是因为大量经验研究对社会资本的内生性问题重视不足, 导致迄今为止, 社会资本究竟有无因果作用的问题, 很大程度上依然悬而未决。

本文以社会资本的劳动力市场效应研究为例, 回溯了大约 80 余篇从 1980 年代至今专注于劳动力市场的实证研究, 详述它们在内生性问题上的处理方式。本文的其他内容如下: 第二部分简要介绍内生性偏误的主要来源和应对策略。第三部分对有关社会资本的劳动力市场的已有文献进行总体性回溯, 讨论内生性问题的应对策略。第四部分提出今后经验研究中的六大努力方向。

二、内生性偏误及模型识别策略举要^①

在微观计量经济学的术语中, 模型就是对一系列观察值概率分布的设置 (specification), 而结构 (structure) 则是对该分布参数的设置。假如两个及以上模型指向相同的联合概率分布, 那么这两个结构是等相同的, 即无法识别。尤为重要的是, 当模型估计的目标是为了评估主解释变量的因果效应系数时, 一个重要的前提条件是同构的替代模型应被剔除 (Cameron & Trivedi, 2005: 20-30)。一般意义而言, 可以把识别某个结构理解为排除竞争性解释存在的可能性。重要的是, 实证研究中会碰到许多让模型无法被识别的问题来源 (例如, 方程设定错误和测量误差所导致的内生性问题等)。限于篇幅, 本文主要解释方程设定错误中的四个方面。

(一) 遗漏变量偏误

在不同的学科中, 遗漏变量偏误经常被称作未观察到偏误 (unobservable bias)、隐藏偏误 (hidden bias)、紊乱问题 (confounding problem) 抑或未观察到的异质性问题 (unobserved heterogeneity)。在社会资本文献中, 普遍存在遗漏变量偏误的可能。在我们分析“使用社

^① 具体的模型和模型识别策略, 参见陈云松、范晓光 (2010)。

会关系”（using contacts）相对于“不使用社会关系”在工作方面的因果效应时，如果难以测量的个体“口才”是一个被遗漏的变量，那么估计参数显然是有偏的。其原因是：语言技能不仅影响个体找到和找不到关系人的机会，而且还可能影响工资水平——因为较高的语言技能本身就代表了较强的能力。再如，如果我们要分析“使用了的社会资本”的效应（used social capital，例如熟人、帮助者的职业声望），同样我们有充分的理由怀疑那些具有更好口才的求职者能够找到社会地位更高的帮助者。因此，当口才这一变量被从模型中遗漏，熟人的职业声望和被访者的工资间的关系则可能是虚假的。另外一个例子是分析网络资源，也即“可使用的社会资本”（accessed social capital）的个体效应。如果我们要分析被访者朋友圈的平均地位对于被访者个人收入的影响，那些我们同样必须考虑被访者的口才本身就与朋友的平均地位相关。一旦如此，被估计的社会网的因果效应就会是有偏的。在经验研究中，有四种主要的方案用以缓解遗漏变量偏差问题。第一是利用非传统数据作为遗漏变量的替代（proxy）以控制潜在的遗漏变量。第二种方法依赖于同一个被访者的自变量和因变量在时间维度上的差异，运用固定效应模型去消除时间固定的（time-invariant）的非观测因素。第三种应是利用非时间性的(non-time)组内差异去估计固定效应模型，这尤其适用于遗漏变量是组别水平上的异质性（group-level heterogeneity）。它实际是第二种方法的变种。第四种处理遗漏变量偏误的方法是使用工具变量（Instrumental Variable）——假若我们能发现并使读者相信某个外生因素与误差项无关但又和解释变量高度相关，那么，它就可能是一个可以帮助我们解决内生性问题的排除限定（exclusion restriction）。

（二）自选择偏误

自选择问题实际是一种特殊的遗漏变量偏误，它主要指主解释结果/因变量是在某种程度上被个人选择所决定（Cameron & Trivedi, 2005:546）。最典型的例子就是“物以类聚、人以群分”。比如，我们分析学校平均成绩对学生个体成绩的影响，就不能简单地进行回归分析。这是因为，学生家长会选择学校，因此学校的质量本身就与家长的观念、智力、收入、教育背景等等有关。而这些家庭因素，通常又会影响学生的成绩。要进一步理解“选择”，我们可以把分析的社会现象解析为两个过程。过程一是学校内“近朱者赤”的效应发挥作用的过程，也是我们最关心的研究内容；而过程二则是择校过程，这个

过程中会有很多因素也同时影响学生的成绩。和一般性的遗漏变量偏误相比，自选择问题的实质就是——两个过程的非观测因素相互关联。在实证研究中，主要有三种校正自选择偏误的方法。第一，随机分配（Random Assigning to Treatment）是解决问题的最佳途径。其中，实验和自然实验（natural experiment）设计显然是理想的方法，因为它们可以设计出随机分配。第二种校正自选择偏误的方法是匹配（matching）。这种方法被广泛应用于干预效应研究。不过，这一方法的前提假设是，自选择完全由观察到的变量所导致。罗森鲍姆等（Rosenbaum & Rubin, 1983）是该方法的首创者，并将其命名为倾向评分方法（PSM）^①。其直观的理念是，用尽量相似的控制组成员和对照组成员来进行比较以减少偏误。第三种方法是建立选择过程的正式模型，建立基于选择模型和实质模型的联立方程组是一种校正自选择偏误的简明方法。内生转换模型（Endogenous switching regression models）就是方法之一。当然，因为自选择问题也可以被看作是一种遗漏变量偏误的特殊类型，因此理论上我们也可以通过纳入更多的控制变量、代理变量和固定效应模型等传统方法来处理该问题。

（三）样本选择偏误

当对因变量的观察仅仅局限于某个有限的非随机样本时，样本的选择偏误就容易产生。一个典型的例子就是估计 MBA 培训班对于受训者收入的影响。问题在于，受训者本身作为分析对象，就是一个非随机的整体。例如，个体可能会决定是否去参加 MAB 课程，做这个决定受制于个人的很多非观测因素。如果热衷于交际的人更乐意参加这样的培训，那么我们分析这个培训班的因果效应时，就会使用热衷于交际人员所占比例过高的样本。一旦热衷交际这个特征本身就影响个人收入，那么简单回归就会出现偏误。被广泛运用来解决样本选择问题的方法是赫克曼（Heckman, 1976）的二阶段（two-stage）方法。其步骤具体如下：第一步，利用总体样本里所有的观察值，对个体的选择接受干预的概率进行预测，计算出每一个观察值的反向 Mills 比率。在 logit 回归中，反向 Mill 比率实际就是非响应概率（不选择受干预）。第二步，利用参与干预的样本进行回归，而同时把反向 Mills 比率作为控制变量以获得一致估计量。

（四）联立性偏误

① 倾向评分是指基于个体的可观察特征而进行某种选择的条件概率。运用 PSM 的主要步骤有：1) 用 logit 和 probit 方程对选择建模，并且把估计到的概率作为个人倾向得分；2) 为每个选择了的被访者匹配一个非常接近的不选择、或不参与的人。接近标准就是倾向得分；3) 比较匹配对子的结果变量。

内生性的另外一个重要来源是联立性问题 (simultaneity)。其本质就是解释变量连带地由因变量决定, 也即双向因果关系。例如, 在网络资源效应的研究中, 我们可能将个体当前职业声望回归到被调查者社会网络的平均职业声望或者工资。然而, 由于社会网的声望均值显然同时也受到被访者个体声望或者工资的影响, 用简单的OLS回归模型就会产生联立性偏误。总体而言, 增加排除限制(exclusion restrictions), 也即寻找工具变量是解决联立性问题最佳方法。外生的自然现象或者随机实验对部分样本产生的冲击往往被用来作为工具变量。这种方法就是与苏特温特 (Soetevent, 2006) 所称道的“偏总体试验” (partial-population experiments) 或莫菲特 (Moffit, 2001) 所建构的“偏好选择干预” (preference changing intervention)。更新的文献提出一种所谓的“条件方差限定” (conditional variance restrictions) 的识别内生效应方法。该方法由格莱塞尔等人 (Glaeser et al., 1996) 开创, 但由格拉哈姆 (Graham, 2008) 将其正式模型化。

三、相关文献中的内生性问题及其应对

根据对社会资本不同的理论旨趣和操作性定义, 现存的有关文献可以分为三种类型 (N. Lin, 1999; Mouw, 2003)。(1)“使用社会关系” (Using Contacts)。此类研究主要涉及运用社会关系或社会网络在个体劳动力市场结果的直接干预效应 (Treatment effects)。(2)“使用了的社会资本” (Used Social Capital)。它所指的是来自某个或某些帮助者的、且被明确使用了的社会资本数量和质量。典型的研究就是探究某个帮助者的资源 (如职业声望或收入) 是否与被帮助者的求职结果相关联。(3)“可使用的社会资本” (Accessed Social Capital) 或“网络社会资本” (Network Social Capital)。这类文献主要关注被访者“可涉取”的社会资源数量和质量, 也即一种潜在的资源总和。正如沃尔克等 (Volker & Flap, 1999) 所强调的那样, “可涉取”的资源并非意味着这些社会关系就一定会被使用。另外, 可使用的社会资本能够从两个维度加以度量: 一是直接的资源测量 (如职业关系人的特性, 网络的中介组织特性), 另外就是网络的结构测量 (如网络规模, 网络密度或范围和结构洞)。

(一) “使用社会关系” 的因果效应

以往的文献很少专门研究“使用社会关系”本身对劳动力市场结果的影响，但在该问题在涉及社会关系的文献中或多或少都会被提及（文献评述参见Granovetter, 1995/1974; N. Lin, 1999; Marsden & Gorman, 2001; Mouw, 2003）。从现有的研究结果看，“使用社会关系”究竟有无作用并无定论。表1为此类研究的总体概括。

从表1可见，尽管有一小部分文献（Campbell & Rosenfeld, 1985; Simon & Warner, 1992; Petersen, Trond et al., 2000）支持“使用社会关系”与工作结果之间存在正向关联，而更多的研究却发现没有显著效应，甚至有负向效应。尤为重要的是，那些发现正向显著性关联的研究，除了巴耶尔等（Bayer et al., 2005）的研究，其他的不是基于小样本，就是基于单个公司研究抑或某类特定的帮助者，而且估计出的参数值也都不大（Mouw, 2003）。同时，表1也显示，相关的研究从80年代早期起步，经过不断深入，数据的丰富性和估计模型的合理性在进一步提高。值得注意的是，在内生性问题和模型识别策略上可谓泾渭分明。因为2003年前的绝大多数研究都把“使用社会关系”视为外生性的，基本以OLS模型和截面数据(cross-section data)^①为主。因此，从因果分析的角度看，2003年后的研究发现更具说服力和可靠性。

在2003年后的新近文献中，有两项研究采用了固定效应模型，借此排除在时间维度上恒定不变的干扰因子，以解决内生性问题。例如，莫维对一个1990年代末搜集的美国面板数据(NLSY1994-1998)进行了固定效应分析（Mouw, 2003）。他发现，“使用社会关系”和工资（对数）之间并没有显著相关。而库格勒等（Kugler, Adriana, 2003）基于另一时期(1981-1982)的NLSY数据进行一阶差分分析（first difference analysis），结果表明，一旦工作部门被控制，无论是OLS还是差分模型，经过员工推荐获得工作对于工资就没有了显著关系。不过，此类方法也有一个问题。未被观察因子可能是随时间变化的。倘若事实确实如此，固定效应方程的估计值就仍然会有偏误。尤其是，如果求职者对他原先通过社会关系获得的工作不满，那么他将有可能转向其他的求职办法，也即个人异质性、个人求职偏好等等就会发生变化。

表1 “使用社会关系”的干预效应研究回顾

① 虽然 Volker 等收集了一个二期面板数据，但他们只是对两期数据分别采用了 OLS 模型。参见 Volker & Flap, 1999。

(是否“使用社会关系”的哑变量作为自变量)

研究	因变量	数据与模型	研究发现
Lin, Vaugh & Ensel 1981 (美)	初始/现在工作地点	两个截面数据 OLS	不显著
Campbell & Rosenfeld 1985 (美)	现在工资	截面数据 OLS	正向显著
Bridges & Villemez 1986 (美)	现在工资	截面数据 OLS	不显著
Wielgosz & Carpenter 1987 (美)	工作满意度	截面数据 OLS	不显著
De Graaf & Flap 1988 (德荷美)	现在工作工资/地位 初始工作地点	两个截面数据 OLS	不显著 负向显著
Marsden & Hurlbert 1988 (美)	现在工资	截面数据 OLS	不显著
Simon & Warner 1992 (美)	初始工资	截面数据 OLS	负向显著
Vecchio 1995 (美)	工作不满意度	截面数据, 多层线性回归	不显著
Korenman & Turner 1996 (美)	现在工资	截面数据 OLS	不确定
Bian & Ang 1997 (中、新)	现在工作地点	两个截面数据 OLS	中: 正向显著 新: 不显著
Volker & Flap 1999 (东德)	89年工资/地位 初始工作地点	两个截面数据 OLS	不显著
Elliot 1999 (美)	现在工资	截面数据 OLS	不显著
Smith 2000 (美)	现在工资	截面数据 OLS	负向显著
Peterson et al. 2000 (美)	初始和现在工资	截面数据 OLS	正向显著
Mouw 2003 (美)	工资/满意度 /持续雇佣	截面/面板 OLS/固定效应	不显著
Bentolila et al. 2003 (美欧)	现在工资	截面 工具变量	负向显著
Kugler 2003 (美)	现在工资	面板 OLS/一阶差分	不显著
Bayer et al. 2005 (美)	工资/收入/持续雇佣	截面 固定效应(配对)+ 工具变量	正向显著
Delattre & sabatie 2007 (法)	现在工资	面板 OLS/工具变量/切换回 归	负向显著
梁玉成 2010 (中)	初职收入	截面数据 PSM匹配+内生转换模型	正向显著

工具变量方法则可以解决这一问题。邦托利阿等 (Bentolila & Suarez, 2003) 使用两个工具变量去解决“使用社会关系”进行求职的内生性问题: 年长的兄弟数目和联邦就业率。其理由是: 一个人的兄弟越多越年长, 他得到的社会关系网络存量和使用关系的可能性就越高。与此同时, 兄弟数目被认为是外生性的, 而且在员工生产率上没有直接作用。至于就业率, 这是因为根据措帕 (Topa, 2001) 的研究, 大部分有价值的工作关系仅仅来自那些有工作的人。所以, “有用的”关系应该与地方就业率有正向相关, 而且也与使用社会关系呈正相关。然而, 联邦就业率并不直接作用于个体工作结果 (当国家的相关属性被控制)。运用这两个工具变量, 他们发现, 由于工人职业和能力的不匹配, 使用社会关系反而导致低工资。当然, 工具变量的外生性是无法验证的。而类似家庭成员数目和联邦就业率等等因素, 都很难说是“完全”外生的因素。

为了解决可能存在的自选择偏误问题, 德拉特 (Delattre) 等比较了三种模型 (OLS, Heckman & Robb二阶段法, 以及内生性切换回归 endogenous switching regression) 的结果, 以探究“使用社会关系”对于工资的直接作用。他们发现, 内生性切换模型较为合适。这是因为它不仅可以检定社会关系的选择内生性, 而且还可以根据使用关系与否对工资的其他解释因子进行取舍 (Delattre & Sabatie, 2007)。例如, 他们发现那些利用网络关系求职的人, 只有公司或工作特性(如同同期限, 工作技能水平, 公司规模, 所有的这些因素都对工资有正向效应) 解释小时工资, 而性别和个体属性(如年龄, 种族)都对此没有任何影响。他们校正选择偏误后发现, 使用社会网络对工资的效应是负向的。对此的解释是, 遗漏变量本身可能与工资呈负相关, 或者, 尽管使用关系在信息传播上是高效的, 但对于工资则影响微乎其微。

在近期西方学者的研究中, “使用社会关系”的正向效应仅仅被巴耶尔等 (Bayer & Topa, 2005) 的波士顿研究间接证明。这项研究运用固定效应模型和匹配方法 (matched pairs method), 发现两个住在同一个人口普查区的人要比两个住在同一个地区 (block group) 但不在同一个人口普查区的人在同一地区工作要多三分之一的可能性。而且, 前者比后者具有更高的工资、收入和更长的雇佣期限。如果我们相信使用社会关系在前一情形 (住居更为靠近) 中可能更大, 那么这个研究可以间接地证明“使用社会关系”是劳动力市场结果的重要决定因素。

梁玉成(2010)的实证分析是国内社会资本研究的一篇重要文献,也是第一篇运用恰当的模型识别策略系统处理自选择偏误的实证研究。基于“2009年中国大城市社会网与求职调查(JSNET2009)”数据,他首先使用倾向性得分匹配的方法,发现尽管和OLS估计量相比,使用社会资本的收入效应略有减少,但仍然统计显著。据此他指出网络资源的“质量”可能是一个容易被遗漏的可观测变量。接下来,他针对PSM方法的强假设,用内生切换模型方法来进一步纠正基于非观测因素的自选择所导致的估计偏误。研究结果表明,总体上自选择可能没有带来大的偏误,使用社会资本仍然具有正向干预效应。但是,如果以1995年为界划分样本,则95年前后两个子样本的分析结果都证明了自选择的存在。饶有兴味的是,在95年之前,个人能力与和使用社会关系正相关,而1995年之后则负相关。作者给出的解释是,1995年之后由于“杀熟”现象,导致社会关系的真正使用价值大减。这样,有能力的人在95年之后会倾向于不是用社会关系。梁玉成(2010)的理论贡献在于明确指出,自选择本身是一个复杂的过程,社会关系的使用,“不仅与行动者所拥有的社会关系的质量和数量有关,还应该受到与关系运作的宏观社会环境的影响”。这是一个事关进一步厘清自选择作为一个“黑箱过程”的重要出发点。

(二)“已使用了的社会资本”的效应

该研究主线的中心是林南的社会资源理论(N. Lin, 1999)。也就是说,使用的社会资源越多或者质量越高,被访者的劳动力市场的结果就越好。目前有大量的经验研究支持这一观点(包括 Lin Nan, 1981; Lin Nan et al., 1981; Marsden & Hurlbert, 1988; De Graaf & Flap, 1988; Wegener, 1991; Bian & Ang, 1997; Bian, 1997; Völker & Flap, 1999; Mouw, 2003),具体参见表2。从表中可见,此类研究对于社会资本的度量,主要就是帮助者的职业声望。至于工作结果,职业声望和工资使用的最多。马登等(Marsden & Hurlbert, 1988)对此做了一些扩展,使分析对象还具有了更多类型的工作结果。

如表2所示,绝大多数研究似乎都证明“使用了的社会资本”对于个人的劳动力市场结果具有正向关系(尽管其中的一些并非统计学意义上显著)。这与林南的文献回顾(N. Lin, 2001)基本一致。然而,这所有的研究,都基于截面分析和OLS模型。因此,遗漏变量和选择

性问题在这些研究中都未能得以解决。^①2003年前的文献，也并非对可能存在的内生性问题毫无关注。但解决的方法仅仅是在模型中纳入更多的控制变量——诸如把受访者的早期劳动力市场结果（上一个工作，或者第一份工作）来作为控制变量。^②林南等人（Nan Lin et al., 1981）在1981年的一项研究中，也曾经考虑到社会关系的选择性。他们试图说明基于某些可观察变量的“人以群分”的选择现象在他们的数据中并不明显。尽管这看似有一定道理，但对于非观察因子所导致的自选择则于事无补。2003年，莫维把“社会趋同性”（social homophily）作为与林南的社会资源理论分庭抗礼的替代性解释。为此，他重复了马登等（Marsden & Hurlbert, 1988）基于1977年底特律市民调查研究。在回归分析中，他建立了一个表示被访者和工作帮助者之间职业是否相同的哑变量。当仅仅对与帮助者职业不同的求职者进行回归时，原来在总样本显著的社会资本效应消失了。据此莫维总结道“有利于社会资源视角的证据在很大程度上是帮助者与求职者间的同职信息流的假象”（Mouw, 2003）。^③

除了一般性遗漏变量和自选择问题，样本选择是这里研究必须面对的另外一个重大问题。例如，马登发现较高社会地位的个体较少利用社会关系（Marsden, 2001）。那么，如果我们把工作结果对“已使用了的社会资本”进行回归，就会使用一个“被删节样本”（truncated sample）而样本中多为社会地位较低的人，从而可能引发偏误估计。正如1988年马登等人所强调的，在估计社会资本在劳动力市场的因果效应时，我们总有充分的理由怀疑，导致人们偏向于使用社会关系的某种未被测量因子很可能与回归方程中的误差项相关（Marsden & Hurlbert, 1988）。于是，他运用赫克曼二阶段法，用logit模型去预测个体使用社会关系的可能性，而后将使用社会关系可能性的预测值作为一个控制项纳入到主体模型。韦格纳（Wegener）和较近的林南也运用了类似的方法以消解样本选择偏误（Wegener, 1991; Nan Lin, 2004）。

-
- ① 因为大部分相关的调查问卷明确地要求受访者提供具体工作协助者的信息，如“当工作协助者帮助你获得现在的工作，他的职业是什么？”，该时间序列是明晰的，因此不存在互为因果关系的问题。换言之，联立性（simultaneity）问题将不会存在。
- ② 如前所述，该方法虽合理但远非完美。这是因为利用早先时期（之前的工作或第一份工作）工作结果去“代理”（proxy）未被观察的个体因子，会在模型中引进了大量额外的未被观察因子。
- ③ 根据 Mouw (2003) 在其个人主页公开的 STATA 执行文件，本文作者发现其对“同职”的操作化是错误的。不过，莫维对自选择的关注富有相当的洞见。

不过，正如伍德里奇（Wooldridge, 2006: 620）强调的，要真正解决样本选择问题，主体模型中的解释因子群应该是选择模型的解释因子群的完全子集。也就是说，出现于主体模型中的全部自变量应该都被纳入选择方程。更为重要的是，我们应也在选择模型中至少有一个特殊的解释变量，而它不能出现在主体模型之中。对于社会资本研究而言，也就是要有一个变量只影响个人对使用社会关系倾向，而不影响工作结果。如果没有这个特殊变量，赫克曼方法就实际很难解决样本选择问题。遗憾的是，在林南2004年的研究中，出现在选择模型而非工作结果模型中的变量是与教育相联的工作经验和父亲工作的单位级别（Nan Lin, 2004）。在马登1988年的研究中这个特殊变量则是工作经验（Marsden & Hurlbert, 1988）。然而，人们有充分的理由去怀疑，这些变量（如工作经验）应该同样也出现在实质（substantive）模型中。^①也即，排除限定并非完美，因此在这些应用中还存在估计偏误。

表2 以往研究中利用的社会资本的效应
(社会关系的声望作为自变量)

研究	因变量	数据与模型	研究发现
Lin, Vaugh & Ensel 1981 (美)	初始/现在在工作地位	两个截面数据 OLS	正向显著
Lin, Ensel & Vaugh 1981 (美)	初始/现在在工作地位与收入	两个截面数据 OLS	正向显著
Marsden & Hurlbert 1988 (美)	现在的工作地位 现在的工资	截面 OLS (Heckman二阶段法)	正向显著 不显著
De Graaf & Flap 1988 (德荷美)	初始/现在在工作地位 初始/现在工资	两个截面数据 OLS	正向显著 不显著
Requena 1991 (西班牙)	现在收入/地位	截面数据 OLS	不显著
Wegener 1991 (西德)	现在的地位	截面 OLS (Heckman二阶段法)	正向显著
Bian & Ang 1997 (中 新)	现在的地位	截面数据 OLS	正向显著
Bian 1997 (中)	现在的地位	截面数据 OLS	正向显著

^① 例如，劳动经济学中著名的Mincer 工资方程就把经验作为预测工资的主要解释变量之一。

Volker & Flap 1999 (东德)	初始/1989 工作地位 1989 工资	两个截面数据 OLS	正向显著 不显著
Smith 2000 (美)	现在工资	截面数据 OLS	正向显著
Mouw 2003 (美)	现在的地位	截面数据 OLS (复制研究)	不显著
Lin 2004 (中)	现在的级别/声望/工资	截面 OLS (Heckman二 阶段法)	正向显著
Ao 2007 (美)	现在的地位	截面数据 OLS	正向显著

(三) “可使用的社会资本”的效应

在以往的研究中，“可使用的社会资本”有各种不同的指称。例如，杜尔拉夫和莫维直接称之为“社会资本”，莫维在2003年的研究中则用的是“网络社会资本”，国内学者赵延东称之为“所占有的社会资本”。一般而言，“可使用的社会资本”概念离范德加奇等(Van der Gagg, & Snijders, 2004)所给出的定义最接近：“对个体的个人化社会网络成员自身所有的资源的积聚，它由于这些关系历史的原因可能有益于个体”。虽然有不同的术语，包括信息、权力、行为、决策和网络结构等各种不同的测量指标，这些概念都代表了嵌入在某一给定的自我中心(ego-centric)网络的潜在社会资源的总和。

杜尔拉夫(Durlauf, 2002)检验了在两个不同的假设下如何去识别“可使用的社会资本”的因果效应：一是把“可使用的社会资本”预先定义为小组特征，另一则定义为小组结果。个体行为/结果就是此类研究的因变量。正如他所指出的，基于第一个假设的社会资本模型就是曼斯基所谓的外生社会互动效应，而基于第二个假设的就是社会资本的内生社会互动效应。需要注意的是，这里“外生效应”、“内生效应”的术语并非意味着社会资本变量被假定为外生或者内生，而是沿用了著名经济学家曼斯基关于社会互动的经典论文(Manski, 1993)里的概念。^①我们可用一个例子来说明二者的区别：以劳动力市场为例，假若求职者的收入受到朋友平均收入的影响，我们称之为内生互动效应，即收入影响收入；当他的收入受到朋友平均教育水平的影响时，就是外生效应，即其他因素影响收入。根据曼斯基和杜尔拉夫的二分法，本文所回顾的研究也分成两个方面，分别见表3和表4。

^① 有时候这里的“外生效应”也被称为“情境效应”(contextual effects)。

1. “可使用的社会资本”的外生效应

这些文献中，社会资本的度量是小组的特征，而非小组的结果。另外，如表3所示，该研究主要关心三类社会资本变量。第一类是对嵌入性资源的直接测量（板块1）。这类“可使用的社会资本”常常被操作化为嵌入于整个社会网络的社会资源（如教育获得）的小组均值。第二类（板块2）则是结构性度量，如就被访者社会网络而言的网络规模、网络多样性和网络跨度、达高性等等。最后，第三种则是前两者的混合型（板块3）。

表3 估计可使用社会资本的外生效应

研究	自变量	因变量	数据 & 模型	发现
1. 直接的资源测量				
Lin & Dumin 1986 (美)	网络达高性/跨度	地位	截面数据 OLS	正向显著
Volker & Flap 1997 (东德)	网络达高性	地位	截面数据 OLS	正向显著
Flap & Volker 2001 (荷兰)	特定目标有关的社会资本	工作满意度	截面数据 OLS	正向显著
Sorensen 2001 (美)	经理队伍工作雇用期标准差的历年平均值	个人职业流动率(退出)	面板 逐段固定风险 模型	正向显著
Mouw 2003 (美)	朋友的平均教育水平、具有稳定工作的比例	工资	截面数据 OLS	正向显著
Weinberg et al 2004 (美)	邻里特征和到单位距离	每年工作时间	面板 固定效应	正向显著
Bayer & Ross 2006 (美)	居住地到单位距离、邻里贫困率和教育水平	就业、工作周数及每周工作时间	截面数据 OLS+工具变量	正向显著
Dujardin & Goffette-Nagot 2006 (法)	居住在贫困邻里	失业率	截面数据 工具变量且建立自选择模型	正向显著
2. 网络结构测量				
Burt 1992	结构洞	晋升比率	截面数据 OLS	正向显著

(美)				
Meyerson 1994 (瑞典)	外在关系数 结构洞	收入	截面数据 OLS	正向显著 不显著
Podolny & Baron 1997 (美)	网络规模	晋升	截面数据 OLS	正向显著
Erickson 1996 (加)	关系资源的多样化程度	工作自主权	截面数据 OLS	正向显著
Kim 2001 (美)	网络约束(密度) 网络桥(多样化) 高教育程度者的数量 名人的数量	年收入	截面数据 OLS	不显著 不显著 正向显著 正向显著
Korpi 2001 (瑞典)	失业者的社会网规模	就业机会	截面 Cox 模型 (Heckman二阶段)	正向显著
Munshi 2003 (墨 西哥)	移民网络大小	就业机会和工资	面板数据 固定效应+工具变 量	正向显著
Yakubovich 2005 (俄罗斯)	关系强度	就业机会	截面数据 固定效 应 (受试者内设计)	正向显著
Beaman 2008 (美)	网络规模 有长期工作的熟人数量	就业机会和工资	面板数据 固定效应	负向显著 正向显著
Hellerstein et al. 2008 (美)	网络连通性	在同一企业工作	截面数据 配对方法	正向显著
Yunsong Chen 2009 (中)	同村外出打工人数	农民工打工收入	工具变量 + Heckman 二阶段 法	正向显著
3. 资源+ 网络				
Campbell et al. 1986 (美)	网络多样化、规模和密度的 因子得分 教育和声望的均值/最大值	工作地位 和家庭收入	截面数据 OLS	正向显著
Lai et al. 1998 (美)	帮助者地位; 网络达高 性; 跨度; 网络规模	工作地位	截面数据 结构方程	正向显著
Flap & Boxman	以关系强度为权重的资源	收入	截面数据 OLS	正向显著

2001 (荷)	总和; 网络结构特征			
Oreopoulos 2003 (加)	邻里规模; 结构类型; 邻里贫困率	教育获得	自然实验	不显著
Seibert et al. 2001 (美)	弱关系; 结构洞; 信息量; 资源量	收入和职业成就	结构方程	正向显著
王卫东 2006 (中)	社会资本总量 (含直接资源和网络结构 因子)	个人资本 (财富权力声望)	结构方程	正向显著

如板块 1 所示, 林南等和沃尔克等用资源的达高性(在社会网络中关系的最高职业声望得分)作为一种“可使用的社会资本”的度量, 并证明了它对个体职业声望的正向作用(Lin & Dumin, 1986; Völker & Flap, H., 1999)。弗拉普(Flap)考察了与特定目标(goal-specific)有关的社会资本如何影响工作满意度。他们发现, “可使用的社会资本”并有各自使用的情境。索伦森(Sørensen)利用 1953-1988 年 24 个广告市场中电视台的数据, 证明了经理队伍中的差异性的历史影响经理人个体的职业流动率。在该研究中, 社会资本被定义为经理队伍中工作雇用期(tenure)标准差的历年平均值。他采用逐段固定风险模型(piecewise-constant), 发现那些属于处于差异性较大经理队伍中的经理更可能退出团队、另谋高就。基于两个不同的横截面数据集, 莫维也发现了平均教育获得水平、拥有稳定工作的朋友比例对被访者的工资有正向效应。然而, 他指出, 由于自选择偏误和遗漏变量偏误的可能, 我们无法断言这些研究中的相关反映的就是“可使用的社会资本”的因果效应, 还是相似的个体倾向于彼此联合的社会趋同性, 甚至相反的因果关系(Flap & Volker, 2001; Mouw, 2003)。

为了解决这些内生性问题, 固定效应模型、对选择过程建模和工具变量等方法在经济学相关文献中得到了不少应用。譬如, 魏因贝里等(Weinberg et al., 2004)探究了一系列邻里特征以及邻里到工作单位的距离对于年度工时的作用。他们利用 1979 年全国青年纵向调查(NLSY79)中被访者每年一次所记录的街道地址去匹配来自 1990 年人口普查中相关被访者的邻里社会特征。在这个独特的面板数据基础上, 他们利用固定效应模型进行回归。研究结果发现, 和相应的 OLS 估计相比, 固定效应模型对于邻里就业率和个体就业的相关度的估计

量减少了2/3。基于1999年法国人口普查数据，杜雅尔丁（Dujardin, 2007）等考察了个体失业是如何被生活于贫困邻里的因素所影响的。为了避开困扰因果推断的内生性组群的形成问题，他们同时评估了一个失业率的probit模型和另一邻里选择（贫困或非贫困）的probit模型。结果显示，联立probit模型的社会资本效应估计量比单纯的probit估计量要小，不过仍然是显著的。

在该类研究中，对内生性问题处理最为全面的可能是巴耶尔和罗斯（Bayer & Ross, 2006）对作用于就业和就业时间的内生性邻里效应的研究。首先，他们基于邻里住房价格与邻里特性间的单调(monotonic)关系，将某一特定的邻里住房价格当作对邻里未被观察值的代理变量。其次，为解决个体层面的未被观察值或自选择因素，他们用与被访者具有相同个人特征的其他被访者的平均邻里特征来作为被访者邻里特征的工具变量。其理由是，既然相似个体选择相似的邻里，那么基于类似的人的平均邻里特征（也即这个工具变量）应该与被访者的邻里特征相关。而同时他们假设个体的可观测特征与未被观察因子无关，这样这个工具变量也就必然与个人异质性无关。运用来自1990年波士顿（每十年一次）人口普查的准确微观数据，他们发现邻里贫困率和平均邻里教育水平的直接效应的工具变量估计值要比运用OLS所得到的估计值大。其原因可能是自身能力或者就业前景较差的人，会主动选择比较“好”的邻里来弥补自身的不足。

接下去我们来看板块2。这里的文献主要把“可使用的社会资本”定义为社会网络结构。早期的研究关注网络规模、多样化和许多其它反映网络所能涉取的信息与影响力等各种社会资本度量（Meyerson, 1994; Erickson, 1996）。大部分研究结果表明“可使用的社会资本”和各种工作结果之间存在正相关。在克服内生性问题的估计策略方面，2000年以来的一些文献值得我们的关注。例如，波多利内(Podolny)等基于来自某一高科技工程与制造公司的数据，说明了拥有一个大而分散的信息关系网络能够有效促进组织内部的晋升。考虑到参与调查本身就是选择性的，他们应用了赫克曼样本选择模型（Meyerson, 1994; Erickson, 1996）。类似地，科皮（Korpi, 2001）在考察瑞典失业者的网络规模对求职的影响时，也利用赫克曼方法解决了可能的自我选择问题及样本选择问题。他分析提出，之所以要用选择模型，是因为失业者会考虑到与某些特定类型的朋友讨论工作问题，这样就具有自选择特征。同时，研究中所使用样本都具有从失业到就业的历史，

而这个样本本身可能招致样本选择问题。固定效应模型和配对方法也曾被用来估计“可使用的社会资本”的效应。例如，雅库伯维奇（Yakubovich, 2005）采用对每个被访者做多次观察的受试者内（within-subject）方法，证实了关系强度对俄罗斯工人就业的正向效应。

不少研究将多个方法综合在一起去处理内生性问题。孟希（Munshi, 2003）分析了墨西哥移民网大小对于在美国劳动力市场上工资获得的影响。他利用面板数据排除了不随时间变化的个体未被观察因素，同时使用移民在墨西哥老家社区的降雨量作为移民网规模的工具变量。墨西哥社区的降雨量显然和美国劳动力市场无关，但却与本社区移民到美国的数量有关。他发现，工具变量估计量要比OLS估计量要大。其解释就是：因为存在不少已经获得了相当财富而返回墨西哥的“回流移民”（return migrants），使得测量到的移民网规模大于实际有效的网络规模。贝尔曼（Beaman, 2007）研究了难民网大小是如何影响他们在美国劳动力市场的结果。为了避免难民选择居住城市这一内生性问题，他巧妙地利用了难民由安置机构以随机的方式在各城市间加以分派这个自然实验。另外，他运用面板数据解决了跨城市和个体固定效应问题。结合这两种策略，他发现在同年（或早一年）安置的难民网络规模与他们的就业和小时工资有负面相关。但是，更早的难民网中具有长期稳定工作的人越多，则新难民的有关劳动力市场结果越好。陈云松（Chen, 2009）分析了中国同村庄的外出打工人数对村民打工收入的影响。考虑到外出打工决策本身具有选择性以及个人能力差异的影响，作者用村庄遭受的自然灾害作为工具变量，用家庭劳动力人数作为赫克曼二阶段方法中的排除限定，将两者结合起来使用，发现工具变量估算值网络效应比简单回归要大。

板块3中的研究主要是对资源度量和结构度量的效应同时进行估计（Campbell et al., 1986; Lai et al., 1998; Flap & Boxman, 2001）。在早期的研究中，对内生性问题的关注有限。这些研究主要运用结构方程模型（SEM）来分析众多社会资本变量之间的相互关系。例如，Lai et al. (1998) 对于纽约大都会地区就业男性的研究，模型中既包括“已使用了的社会资本”，也包括“可使用的社会资本”。使用几乎相同的方法，西伯特（Seibert, 2001）等聚焦于“可使用的社会资本”，将社会资本概念化为网络结构（如网络规模，弱连带的结构洞）和可使用的信息、可使用的资源等等多元度量。结构方程模型的结果显示社会资本对晋升、工资和职业满意度具有正向效应。王卫东（2006）利用

结构方程模型对 2003 年中国城市综合调查数据进行分析，发现中国城市居民的社会网络资本对于网络主体的个人资本有着明显的正向影响。该文是国内第一篇对社会资本进行正式量化的重要文献。不过，尽管结构方程模型能够解决一些测量误差和潜变量问题，但是它的本质是证实性研究（confirmatory modeling）而非解释性建模（exploratory modeling）。这样，它往往是“对潜在的因果结构做出无法检验和依赖于反复推测的假定”（Mouw, 2006），对内生性问题解决无能为力。

值得一提的是最新的这类研究开始借助于自然实验的方法。在加拿大多伦多，出生于1963-1970年的人，在儿童时代曾被政府的“住房建造计划”随机分配居住在多伦多大都会地区内的邻里。奥伦普洛斯（Oreopoulos, 2002）利用这个自然实验来解决邻里选择问题。他分析了当时邻里的不同规模、结构类型（住宅高楼和平房的差异）和贫困率对于这些人成年后社会经济地位指标（如劳动收入，失业可能性和福利分享）的影响。不过，他发现并没有显著性差异。这可能意味着，如果人们不选择邻里或居住地，社会学家原本期望的邻里社会资源和网络结构对于人的工作影响很小。

2. “可使用的社会资本”的内生效应

如表4所示，我们把内生社会互动效应研究分为非线性模型（板块1）和线性均值模型（板块2）。这其中的区别就在于，板块1中的因变量不是连续性变量而是反映个体工作选择的哑变量，而面板2恰恰相反。这是因为，logit或probit模型的非线性可以回避线性均值模型所导致的“映射问题”（指线性均值模型的估计量无法区别外生和内生效应，详见Manski 2003）。因此，板块1中研究的估计偏误主要是常规性的遗漏变量偏误和自选择偏误。而板块2中的研究还同时受到“映射问题”的困扰。

在板块1中，早期的研究对可能的遗漏变量偏误和自选择并没有太多留意。例如，埃布尔等（Abell et al., 1995）考察了英国工人失业是如何被地区失业率影响的。他们直接对混合横截面（pooled cross-section）数据进行logit回归，并发现了显著相关。凯斯等人（Case & Katz, 1991）可能是第一个在同群效应研究中系统考虑了潜在估计偏误的学者。他们指出，当邻里结果作为主解释变量时候，邻里的平均家庭背景应该被纳入到模型中去作为控制变量。尽管如此，这些早期的研究无法解决未被观察的邻里层次、家户层次和个体层次的遗漏因

子。在这些研究的基础上，更多的复杂方法在近来的研究中被加以采用。

赫斯特罗姆等（Hedström et al., 2003）基于瑞典的面板数据，用Cox比例风险模型来研究邻里失业水平如何影响个体就业。为了控制邻里的不随时间变化的干扰因子和居民对于居住地的选择，他们的模型包括大量观察因子和接近700个哑变量——考虑了每一个邻里的固定效应。虽然这种直接控制邻里固定效应的方法可能使得估计值下偏，但他们发现即使是用了这个“过度控制”的方法，邻里对于个人就业的影响仍然是显著的。除了固定效应方法，工具变量方法也被这类研究用来解决内生性问题。孔特雷亚斯（Contreras）等借鉴了巴耶尔的方法，用房价来控制邻里未被观察因素，用和被访者具有同样个人可观察特征者的平均邻里特征来作为邻里特征的工具变量。研究结果表明，邻里的非农业就业对玻利维亚妇女获得非农业工作具有正面影响（Contreras et al., 2007）。莫兰（Maurin）等则考察了法国邻里的劳动力市场参与是如何影响母亲进入职场的决定（Maurin & Moschion, 2009）。为了消解内生性问题，他们运用邻里平均的儿童性别混合度（sex mix）去作为邻里劳动力市场参与的工具变量。这个工具变量成立的理由是，法国家庭的儿童性别混合度影响母亲是否工作，但和家庭择邻无关。研究结果证明，邻里就业和母亲个体就业息息相关。

在板块2中，联立性问题和映射问题往往同时存在。不少研究只解决了联立性问题，但是由于采取的线性均值模型，对映射问题无能为力。比如，马尔毛罗希（Marmaros）等利用达特默斯学院宿舍随机分配，发现室友的平均工资和个人工资之间存在很大的正相关（Marmaros & Sacerdote, 2002）。然而，他们承认由于“映射问题”，他们无法将估计系数直接解释为内生性同群效应的实际大小。同样的，为了估计合作者平均收入对个体收入的影响，什维德科（Shvydko, 2007）对面板数据进行固定效应分析以消解遗漏变量偏误和自选择问题。然而，他同样无法把内生性效应和外生性效应相分离，因为“映射问题”并未被解决。

表 4 估计嵌入的社会资本的内生性效应

研究	自变量	因变量	数据 & 模型	发现
非线性模型				

Sprengers et al 1988 (荷)	地位的加权和 (关系强度作为权数)	地位	4期面板数据 OLS	正向显著
Case & Katz 1991 (美)	邻里空闲(idle)率 (不就业)	个人空闲 (idle)	截面数据OLS	正向显著
Abell et al 1995(英)	地区失业率	个人就业	混合截面数据 Logit	负向显著
Hedström et al 2003 (瑞)	邻里平均就业率	个人就业	面板 固定效应	正向显著
Contreras et. al. 2007 (玻利维亚)	邻里平均就业率; 邻里平均非农工作率	个人就业 妇女非农就业	截面数据 IV+Proxy	正向显著
Maurin & Moschion 2008 (法)	邻里其他母亲就业	被访母亲就业	截面 工具变量	正向显著
线性均值模型				
Marmaros & Sacerdote 2002 (美)	室友平均工资	个人工资	截面 自然实验	正向显著
Laschever 2005 (美)	战友平均工资 战友平均失业率	个人工资 个人失业	截面 多重参照 法	正向显著
Shvydko 2007 (美)	公司同事平均收入	个人收入	面板 固定效应	正向显著
Grodner et al 2007 (美)	平均小组工作时间	工作小时数	截面 工具变量	正向显著

然而，一些研究通过创新性方法论，成功地从外生性效应中区分出内生性效应。例如，拉榭佛(Laschever, 2005)通过多重组群(或多重参照小组 MRG)的模型识别方法，分析了二战老兵战友社会网对于个体就业的效应。假定子总体 AB 属于 A 和 B 两个组群(如 AB 是 A 和 B 的重叠部分)，在 A 和 B 中其余的人分别称为 RA 和 RB。注意到在 RA 中的个体虽不认识 RB 中的任何一人，但 RA 和 RB 共有一个交往的中间群——AB。因此，RA 的结果仅仅会间接影响 RB 的结果。由此可见，一旦我们能确认 RA 对于 RB 存在一个影响，那这个影响必然是通过中间人 AB 所传播出去的 RA 对于 AB 的内生效应。在他们所使用的二战老兵样本中，他们既有一个二战期间服役的军事单位网，又有一个 1930 年人口普查的邻里网络。因此，通过分析彼

此不认识的战友和邻居之间就业变量的关联，就可以确认老兵社会网对于个体就业的内生效应。借助类似的策略，格罗德内等（Grodner et al., 2007）分析了工人个体劳动时间和组群平均工作时间之间的关系。由于在该研究中的“组群”是通过因子分析所测得的表示“社会经济距离”非常接近的个体，那么这就意味着每位被访者都有一个不同的组群。和拉榭佛的研究一样，他们也发现了基于工作时间的内生效应。

四、 解决内生性问题的研究策略

前文介绍和回溯了社会资本研究中内生性问题的来源和模型识别的方法。归根结底，要获得偏误更小的估计量，必须把新的信息引入模型分析之中。但我们应该注意到，高质量的数据和精巧的研究设计，才是未来社会资本实证研究值得注意的努力方向。因为如果我们仅仅掌握了相对复杂的模型和识别策略，而没有好的数据和研究设计，就会巧妇难为无米之炊。因此，能够带来和使用更大信息的数据和研究设计才是解决社会资本分析内生性问题的根本前提。^①

总体来看，能够部分解决内生性的方法主要是固定效应方程和倾向性评分匹配，而理论上可以全部解决内生性的方法则包括工具变量以及自然实验。针对这些具体的模型识别策略，我们从数据搜集、研究设计等角度提出今后社会资本实证分析中应努力的几个方向。

其一，在模型设置中，应尽可能将有可能成为遗漏变量的因子纳入模型。正如前文所提及的，内生性偏误的本质就是和自变量、因变量都有关的变量被从模型中遗漏。我们认为，在社会资本研究中，常见的容易被遗漏的变量包括两个大类。第一类我们称为“应免型”的遗漏变量。比如，在以工资收入等为因变量的分析中，城市、行业类型、工作经验等等变量，往往既与工资有关，也与被访者使用社会资本或个人关系网络的特征息息相关，如果不加控制，必然成为遗漏变量。更重要的是，这类变量容易收集，常规问卷中也往往都有测量。

^① 本文主要立足于模型识别提出相应的研究策略。事实上，内生性还有一个重要来源就是测量误差，会导致估计量偏低。除了传统的工具变量等方法外，改进变量测量方法也是克服测量误差引发的内生性的应有之义。如边燕杰教授支持的“2009年中国大城市社会网与求职调查（JSNET2009）”调查中所使用的信息、人情测量方法，就是减少社会资本测量测量误差的一个重要方法。当然，这不在本文讨论范围之内。

也正是从这个角度，我们称之为“应免型”遗漏变量，也即应该尽量避免其被从模型中遗漏。第二类我们称为“难免型”遗漏变量。比如，个体能力（口才、社会交际等）、性格、劳动力市场质量等等，这些因素也往往与我们关心的自变量与因变量同时相关。但要获得这些因素的精确测量，不是在调查问卷中加一个问题就能解决的。很多因素，我们基本没有好的测量基准和指标。像个人能力这样的因素，在教育回报研究中就一直是一个难以解决的内生性来源——无论是教育水平还是智商 IQ 指数，都无法替代。对于这类“难免”被遗漏的因子，我们应该设法通过采取高级的模型识别方法来解决。

其二，加强社会网数据搜集的纵向跨度，以获得面板数据（panel data），进行固定效应模型分析。过去我们的社会资本研究主要依赖社会网络的截面数据，往往以年为时间单位。譬如，2003年中国综合调查（CGSS2003）仅提供2003年度被访者的社会网络。而我们应该认识到，个人的社会网络在时间跨度上往往是有变化的。一个简单的理由就是，人的交往对象和交往层面不会一成不变。因此，如果能对被访者进行更长时间跨度上的跟踪研究，将使得今后的社会资本实证分析中可以应用面板数据，估算固定效应模型。当然，这可能对数据搜集提出了过高的要求。一个麻烦的问题在于，不少被访者的社会关系网络可能是相当稳定的。比如，有可能五年之内，一个人的核心讨论网不会发生任何变化。这实际上就使得面板数据在社会资本研究中的用武之地有了很大的限制。尽管如此，我们仍然可以尝试在3-5年的跨度上对同一被访者进行跟踪。在没有获得面板数据之前，我们无法明确获知人们的社会网络究竟随着时间发生多大程度的变化。

其三，扩展社会网络数据横向跨度，以获得工具变量。过去的社会网络数据，其横向涉及的关系层级基本是两级。以CGSS2003的“社会交往”题器为例，数据收集仅仅局限于两层数据——被访者和被访者的朋友（该题器运用提名法要求被访者提供不超过5个朋友或熟人的工作、教育、职业等信息）。而在今后的设计中，我们可以对“朋友”的配偶、父母和朋友的朋友进行数据收集。这样就构成一个三层数据体系。也即被访者（第一层）——被访者的朋友（第二层）——被访者的朋友的其他社会关系（第三层）。在具体问题设置中，注意加上一个“您是否与朋友的配偶、父母或朋友熟识”的问题。对于不熟识第三层社会关系的被访者来说，第三层社会关系所附带的社会资源本质上可谓第二层社会关系（也即“朋友”）提供了一个工具

变量。也即，第三层作为第二层的工具变量。该策略的提出理由是：第三层变量显然和第二层有关，但如果我们能确保第三层的朋友的社会关系和被访者不直接认识（譬如，被访者会听说自己的朋友的某些社会关系人的信息），那么第三层变量就会仅仅通过影响第二层变量来影响第一层变量。

其四，加强外生性数据的收集和挖掘，以获得工具变量。我们把实证研究中可能具有价值的工具变量分为两个大类。第一类是绝对外生变量。例如，孟希（Munshi, 2002）、陈云松（Chen, 2009）等研究中使用了降雨量、自然灾害等绝对外生数据。所谓绝对外生，也就是说通过控制一些基本的变量，可以确保这些变量对于模型的外生性。譬如，假若我们控制了地区、省份甚至城市，那么可以相信各个城市、村庄的降雨量或者自然灾害是完全随机的。在过去的问卷调查中，设计者有意无意地会提供一个值得挖掘的外生数据。例如CGSS2003问卷就涉及到兄弟姐妹的数量、子女的性别等等。实际上，诸如彩票号码、河流数目等等，都曾在相关的经典工具变量文献里被使用过，往往能够对于社会资本研究中内生性问题的解决起到作用。第二类是群体层面外生变量。一些在省、州、国家等群体层面上的变量，往往通过影响自变量来影响因变量。这些可能的工具变量虽然在外生性程度上较弱，但在实证分析中仍不失为对标准OLS方法的补充。对于社会资本和社会网分析而言，自然实验的方法相对困难。这是因为人们的社会网形成需要一定的时间和过程。而作为社会学家，人工干预被访者的社会网从研究伦理和可行性的角度均是不允许的。但国家、政府层面上的外生政策干预，则可能带来自然实验，可以用工具变量模型来进行分析。实际上，工具变量分析本身往往需要“灵感式”的逻辑推导，一些看似和研究无关的变量，反而就有可能成为解决内生性问题的奇兵。

其五，向读者认真说明模型所需要的假设。不管是简单的OLS模型还是更复杂的固定效应模型、倾向得分匹配或者工具变量方法，都需要一定的假设作为前提。即便我们有了更多信息量的数据，使用了更复杂的识别策略和模型，我们一定要清醒地认识到，内生性将是社会学家永远的达摩克利斯之剑。这是因为，模型识别策略只能在一定的假设下才能保证估计量是无偏的。比如，固定效应模型的使用中，我们必须假设干扰因子在时间跨度上不变。但是，人的性格、能力等观测不到的因子，往往是可以变化的。又如，倾向得分匹配方法可以

解决自选择问题，但它依赖于大样本，而且假设内生性的来源是数据中提供的可观测变量。再如，工具变量方法看似完美，但工具变量本身的外生性是无法验证的，而且还有弱工具变量、局部样本效应等等诸多问题，需要在分析中作详细说明。归根结底，对于内生性问题不存在一个一劳永逸的方法。任何针对内生性问题进行的模型识别，都应该向读者清清楚楚地说明这个方法所需要的工作假定，让读者知道，只有在这些假定成立的情况下，估计量才是无偏的。至于读者是否接受这个假定，我们无法确认。因此，在实证研究中，对自己的分析结果保持高度的警觉和谦虚，是非常必要的。梁玉成（2010）在使用倾向得分匹配方法时，就非常清楚地指出了该方法的局限性。

其六，倡导复制性研究。复制性研究（replication）是社会科学不断积累进步的重要方法。也即，我们应该尽可能对前人的同样数据，用不同的模型进行复制研究。这本质上，是一个对过去的数据进行挖掘的过程。比如，80年代的研究中，固定效应模型和工具变量方法还未被社会学家认识，因此可能存在很好的数据，但只是用OLS这样的简单回归进行了分析。如果能针对这样的数据进行二次分析并进行对比，那么我们就从内生性的角度验证以往研究成果的可信度。

参考文献：

- 边燕杰，2004，《城市居民社会资本的来源及作用：网络观点与调查发现》，《中国社会科学》第3期。
- 陈云松、范晓光，2010，《社会学定量分析中的内生性问题：测估社会互动的因果效应研究综述》，《社会》第4期。
- 梁玉成，2010，《社会资本和社会网无用吗》，《社会学研究》第5期。
- 罗家德、赵延东，2005，《社会资本的层次及其测量方法》，李培林、覃方明主编《社会学：理论与经验》（第一辑），北京：社会科学文献出版社。
- 王卫东，2006，《中国城市居民的社会网络资本与个人资本》，《社会学研究》第3期。
- 张文宏、李沛良、阮丹青，2004，《城市居民社会网络的阶层构成》，《社会学研究》第6期。
- Abell, Peter, Robert Crouchley & Colin Mills 1995, "An Exploration of Self Employment in Great Britain: In Search of Social Capital." *unpublished manuscript*.
- Angrist, Joshua D. & Alan B. Krueger 1999, "Empirical Strategies in Labor Economics." in Orley Ashenfelter & David Cardn (eds.), *Handbook of Labor Economics III*. Amsterdam: North-Holland.
- Bayer, P., S. Ross & G. Topa 2005, "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," NBER Working Paper 11019.

- Bayer, P. & S. Ross 2006, "Identifying Individual and Group Effects in the Presence of Sorting: A Neighborhood Effects Application." Cambridge, United States: National Bureau of Economic Research. NBER Working Paper 12211.
- Beaman, Lori A. 2007, "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the U.S." mimeo Northwestern University.
- Bentolila, S., C. Michelacci & J. Suarez 2003, "Social Networks and Occupational Choice." London School of Economics, Centre for Economic Performance, mimeo.
- Bian, Yanjie & Soon Ang 1997, "Guanxi Networks and Job Mobility in China and Singapore." *Social Forces* 75(3).
- Bian, Yanjie 1997, "Bringing Strong Ties Back in: Indirect Ties, Network Bridges, and Job Searches in China." *American Sociological Review* 62(3).
- Calvó-Armengol & Zenou 2005, "Job Matching, Social Network and Word-of-mouth Communication." *Journal of Urban Economics* 57.
- Cameron & Trivedi 2005, *Microeconometrics*. New York: Cambridge University Press.
- Campbell, K. E., P. V. Marsden, & J. S. Hurlbert 1986, "Social Resources and Socioeconomic Status." *Social Networks* 8.
- Campbell, K. & Rosenfeld R. 1985, "Job Search and Job Mobility: Sex and Race Differences." *Research in the Sociology of Work* 3.
- Case A. & Katz L. 1991, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." NBER Working Paper 3705.
- Contreras, Dante, Diana Kruger, Marcelo Ochoa & Daniela Zapata 2007, "The Role of Social Networks in the Economic Opportunities of Bolivian Women," RES Working Papers 3240, Inter-American Development Bank, Research Department.
- Delattre & Sabatie 2007, "Social Capital and Wages: An Econometric Evaluation of Social Networking's Effects." *Labour* 21(2).
- Dujardin, Claire & Florence Gpffette-Nagot 2007, "Neighborhood effects, public housing and unemployment in France." Working paper.
- Durlauf, S. 1999, "On the Empirics of Social Capital." *Economic Journal* 112.
- Durlauf, N. 1999, "The Case against Social Capital." *Focus* 20(3).
- Erickson, Bonnie H. 1996, "Culture, Class, and Connections." *American Journal of Sociology* 102(1).
- Flap, Henk & Beate Volker 2001, "Goal Specific Social Capital and Job Satisfaction," *Social Networks* 23.
- Flap, Henk & Ed Boxman 2001, "Getting Started: the Influence of Social Capital on the Start of the Occupational Career." In N. Lin, K. Cook & R.S. Burt(eds.), *Social Capital: Theory and Research*,

- New York: Aldine De Gruyter.
- Glaeser, Edward L. , Bruce Sacerdote & Jose A. Scheinkman, 1996, "Crime and Social Interactions." *The Quarterly Journal of Economics* 111(2).
- Graham, Bryan S. 2008, "Identifying Social Interactions through Conditional Variance Restrictions." *Econometrica* 76(3).
- Granovetter, Mark 1995/1974, *Getting a Job: A Study of Contacts and Careers*, Cambridge, MA: Harvard University Press.
- Grodner, Andrew & Thomas J. Kniesner 2007, "Labor Supply with Social Interactions: Econometric Estimates and Their Tax Policy Implications." IDEAS Working Paper.
- Heckman, J. 1976, "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and A Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5(4).
- Hedström, Peter, Ann-Sofie Kolm & Yvonne Åberg 2003, "Social Interaction and unemployment." Working paper.
- Ioannides, Yannis M. & Linda Datcher Loury 2004, "Job information networks, neighborhood effects and inequality." *Journal of Economics Literature* 42.
- Korpi, Thomas 2001, "Good Friends in Bad Times? Social Networks and Job Search among the Unemployed in Sweden." *Acta Sociologica* 44.
- Kugler, Adriana 2003, "Employee Referrals and Efficiency Wages." *Labour Economic* 10.
- Lai, G., Nan Lin & S. Leung 1998, "Network Resources, Contact Resources, and Status Attainment." *Social Networks* 20(2).
- Laschever, Ron 2005, "The Doughboys Network: Social Interactions and the Labor Market Outcomes of World War I Veterans." mimeo, Northwestern University.
- Lin, N. 1999, "Social Networks and Status Attainment." *Annual Review of Sociology* 25.
- 2001, *Social Capital: A Theory of Structure and Action*. Cambridge: Cambridge University Press.
- Lin, Nan & Mary Dumin 1986, "Access to Occupations through Social Ties." *Social Networks* 8.
- Lin, Nan 2001, *Social Capital: A Theory of Social Structure and Action*. New York: Cambridge University Press.
- 2004, "Job Search in Urban China: Gender, Network Chains, and Embedded Resources." In Henk Flap and Beate Volker (eds.), *Creation and Returns of Social Capital: A New Research Program*. London and New York: Routledge.
- Lin, Nan, John Vaughn & Walter M. Ensel 1981, "Social resources and occupational status attainment." *Social Forces* 59(4).
- Lin, Nan, Walter M. Ensel & John C. Vaughn 1981, "Social Resources and Strength of Ties: Factors in

- Occupational Status Attainment.” *American Sociological Review* 46(4).
- Manski C. 1993, “Identification of endogenous social effects: the reflection problem.” *Review of Economic Studies* 60.
- 2000, “Economic Analysis of Social Interactions.” *Journal of Economic Prospective* 14(3).
- Marmaros, David & Bruce Sacerdote 2002 “Peer Effects in Occupational Choice.” *European Economic Review* XLVI.
- Marsden, Peter 2001, “Interpersonal Ties, Social Capital, and Employer Staffing Practices.” In Nan Lin, Karen Cook & Ronald S. Burt (eds.), *Social Capital: Theory and Research*. Hawthorne. NY: Aldine de Gruyter.
- Marsden, Peter, V. & Elizabeth Gorman 2001, “Social Networks, Job Changes, and Recruitment.” In Berg & Kalleberg (eds.), *Sourcebook on Labor Markets: Evolving Structures and Processes*. New York: Kluwer Academic/Plenum Publishers.
- Marsden, Peter & Jeanne Hurlbert 1988, “Social Resources and Mobility Outcomes: A Replication and Extension.” *Social Forces* 66.
- Maurin, Eric & Julie Moschion 2009, “The Social Multiplier and Labour Market Participation of Mothers.” *American Economic Journal: Applied Economics*, American Economic Association, 1(1).
- Meyerson, Eva M. 1994, “Human Capital, Social Capital and Compensation: The Relative Contribution of Social Contacts to Managers' Incomes.” *Acta Sociologica* 37.
- Moffit. Robert. 2001, “Policy Interventions, Low-Level Equilibria and Social Interactions.” In S. N. Durlauf & P. H. Young(eds.), *Social Dynamic*. Boston, MA: MIT Press.
- Montgomery, J.D. 1991, “Social Networks and Labour-Market Outcomes: Toward an Economic Analysis.” *American Economic Review* 81(5).
- 1992, “Job Search and Network Composition: Implications for the Strength-of-weak-ties Hypothesis.” *American Sociological Review* 57.
- Mouw, T. 2003, “Social Capital and Finding A Job: Do Contacts Matter?” *American Sociology Review* 68(6).
- Mouw, Ted 2006, “Estimating the Causal Effect of Social Capital: A Review of Recent Research.” *Annual Review of Sociology* 32.
- Munshi, Kaivan 2003, “Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market.” *Quarterly Journal of Economics* 118(2).
- De Graaf, N.D. & H.D. Flap 1988, “With A little Help from My Friends: Social Capital as An Explanation of Occupational Status and Income in the Netherlands, the United States and West Germany.” *Social Force* 67.

- Petersen, Trond, Saporta Ishak & Marc-David L. Seidel. 2000, "Offering A Job: Meritocracy and Social Networks." *American Journal of Sociology* 106.
- Philip, Oreopoulos 2002, "Do Neighbourhoods Influence Long-term Labour Market Success? A Comparison of Adults Who Grew Up in Different Public Housing Projects." *Statistics Canada, Analytical Studies Research Paper* 185.
- Portes A. 1998, "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24.
- Rosenbaum, Paul R. & Donald B. Rubin 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1).
- Seibert, Scott E., Maria L. Kraimer & Robert C. Liden 2001, "A Social Capital Theory of Career Success." *Academy of Management Journal* 44.
- Shvydko, Tetyana 2007, "Interactions at the Workplace: Peer Effects in Earnings." *Job Market Paper*, Department of Economics and Carolina Population Center, University of North Carolina at Chapel Hill.
- Simon, Curtis J. & John T. Warner 1992, "Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure." *Journal of Labor Economics* 10(3).
- Soetevent. A. 2006, "Empirics of the Identification of Social Interactions: An Evaluation of the Approaches and Their Results." *Journal of Economic Surveys* 20(2).
- Topa. G. 2001, "Social Interactions, Local Spillovers and Unemployment." *Review of Economics Studies* 68.
- Van der Gagg, M. & Snijders, T. 2004, Proposals for the measurement of individual social capital. in Henk Flap and Beate Volker(eds.), *Creation and Returns of Social Capital: A New Research Program*. London and New York: Routledge.
- Völker, Beate, and Henk Flap 1999, "Getting Ahead in the GDR: Social Capital and Status Attainment under Communism." *Acta Sociologica*, 42.
- Wegener, Bernd 1991, "Job Mobility and Social Ties: Social Resources, Prior Job, and Status Attainment." *American Sociological Review*, 56(1).
- Weinberg, B., P. B. Reagan & J. J. Yankow 2004, "Do Neighborhoods Affect Hours Worked? Evidence from Longitudinal Data." *Journal of Labor Economics* 22(4).
- Wooldridge, J. M. 2006, *Introductory Econometrics: A Modern Approach. (3rd Edition)* Thomson South-Western.
- Yakubovich V. 2005, "Weak Ties, Information, and Influence: How Workers Find Jobs in A Local Russian Labor Market." *American Sociological Review* 70(3).
- Chen, Yunsong 2009, "Network Effects among Rural-to-Urban Migrants in China." Working Paper

2009-06, Department of Sociology, University of Oxford.

作者单位: 牛津大学社会学系、Nuffield 学院 (陈云松)
复旦大学社会学系、浙江省社会科学院 (范晓光)
责任编辑: 杨 可

文章来源:《社会学研究》2011 年第 1 期

中国社会学网 www.sociology.cass.cn