在线实地实验研究进展*

翁茜 李栋

摘要:在线实地实验是 21 世纪以来基于计算机技术和互联网的发展所兴起的控制性实验研究方法,逐渐在多个社会科学领域和政策评估实践中得到广泛应用。本文首先介绍在线实地实验的设计与开展原则,继而阐述在线实地实验与其他数据收集方式相比存在的优势和潜在的问题。然后,本文梳理劳动经济学领域开展在线实地实验研究的最新进展,同时总结在中国背景下开展的初步研究进展,以期为经济学者提供新的研究思路,并为更好地利用各类中国在线平台开展实验研究提供经验借鉴。

关键词:实地实验 在线实验 实验经济学 劳动经济学

一、引言

进行因果推断日益成为经济学经验研究的重点,即通过比较接受某种干预的处理组(the treatment group)与未接受该干预的控制组(the control group)之间结果变量的差别,从而识别干预的平均效应 (treatment effects)。因果推断的关键在于构建适当的反事实的情形(counterfactual),其中控制性实验 (controlled experiments)由于能够通过在研究对象中随机分配干预条件来直接构建控制组,已经成为现有构建反事实情形方法中最为逻辑清晰、简明可信的一种方法(Harrison & List,2004)。该方法以人为实验对象(即被试),通过给予激励并实施控制性实验条件,比较实验对象在不同实验条件下的行为和决策来开展相关问题研究。实验不仅是一种研究设计方法,也是创造和收集数据的一种方法。

根据 Harrison & List(2004)提出的六个实验基本要素(即被试群体的性质、被试带入到实验中的信息的性质、实验中涉及的商品的性质、实验任务或交易规则的性质、实验收益的性质以及被试参与实验时所处环境的性质),控制性实验可被划分为传统实验室实验(conventional laboratory experiments)、人工实地实验(artefactual field experiments)、框架实地实验(framed field experiments)和自然实地实验(natural field experiments)。其中,自然实地实验是在研究问题发生的真实环境中,以本来就在这个环境中进行决策的主体为实验对象,在被试不知道自己被研究者观察的"自然"状态下,对具体研究问题做出行为和决策的控制性实验方法。而传统的实验室实验的被试通常为标准被试(通常为大学生),在知道自己被研究者观察的情形下,被给予小额的物质(通常是金钱)或非物质激励,在可以反映研究问题的抽象任务和明确的实验指令下做出行为和决策。人工实地实验将传统的实验室实验中的被试换为"真实的人";框架实地实验进一步将"实验场地"搬到真实的环境中,但被试仍然知道其处于被研究者观察的状态中。相较于其他三种实验类型,自然实地实验因其研究对象、环境、任务、信息和收益等方面均更为真实而具有更高的外部有效性(external validity)①,因而被

^{*} 翁茜、李栋,中国人民大学劳动人事学院,邮政编码:100872,电子邮箱:qian.weng@ruc.edu.cn,ld2015@ruc.edu.cn。基金项目:中国人民大学科学研究基金项目(中央高校基本科研业务费专项资金项目)(19XNB012)。感谢北京师范大学经济与工商管理学院何浩然教授的建议和评论,感谢审稿专家和编辑的建议,作者文责自负。

①尽管有研究者倾向于认为实验室实验对于决策环境具有更强的可控性(Falk & Heckman, 2009),但 Al-Ubaydli & List(2015)证明,当实验室实验的潜在被试参与实验的可能性存在系统性差异时,自然实地实验由于被试不知道在被观察而对究竟谁参与实验具有更强的可控性。

行为和实验经济学领域外的学者广泛地接受,已在劳动、环境、产业组织等经济学重要子领域以及管理学、营销学、政治科学等其他社会科学领域和政策评估实践中得到大量应用。

在线实地实验(online field experiments)是以在线平台或社区的用户为研究对象、以社区运转的软件为介质,通过互联网实施操控来开展的自然实地实验(Chen & Konstan, 2015)。① 在线实地实验得益于计算机技术和互联网的发展,兴起于 21 世纪初期,并在最近十年得到了迅猛发展。可供研究者开展在线实地实验的平台包括但不限于在线劳动力市场、在线招聘平台/招聘网站、在线社区、社交网络、小额信贷网站、在线教育平台、电子商务网站、门户网站等。这些平台并非为研究目的而创建,但其所具有的特点为开展实验提供了有效保证。例如,在线劳动力市场可以轻松地向特定用户支付报酬从而保证实验激励的可操作性,筛选出拥有多个账户的用户从而保证实验被试的唯一性和独立性,并有效防止被试之间相互交流从而保证实验决策的匿名性(Horton et al, 2011)。

本文首先介绍在线实地实验的设计与开展原则,继而阐述在线实地实验与其他数据收集方式相比存在的优势和潜在的问题。接下来,本文梳理 21 世纪以来劳动经济学领域^②开展的在线实地实验研究^③的最新进展,同时总结在中国背景下开展的初步研究进展,以期为经济学者提供新的研究思路,并为更好地利用各类中国平台开展在线实地实验研究提供经验借鉴。

二、在线实地实验的设计与开展

在线实地实验由于在被试群体、实验任务和实验环境的性质等实验基本要素方面具有的"在线"属性,从而表现出与同类线下实验相区别的特征。类似地,在线实地实验也拥有独特的依托网络技术的实验于预手段。本文首先简述在线实地实验的基本操作流程,然后分析其特征,最后介绍其于预手段。

(一)基本操作流程

在线实地实验通常操作简便,研究者以用户身份或在运营者协助下介入网站的正常活动,将参与这些活动的用户作为潜在被试,利用网站规则、功能和网站支持的计算机工具等对被试进行干预,并为他们提供合适的收益激励,收集被试的决策和行为及其他特征。需要格外注意的是,为了保证实验的"自然"属性,研究者应深入了解网站的规则、功能、用户群体和一般活动,以免被用户发现异常从而影响实验结果。此外,实验可能存在的道德伦理问题也需要研究者审慎考虑。④

由于实验的具体操作流程因其发生的网站类型不同而存在较大差异,本节以在线劳动力市场上开展的实验为例,介绍该类实验的基本操作流程。与在线劳动力市场中真实工作的劳动者招募模式相同,研究者亦通过在市场上发布工作广告招募被试。广告通常真实地描述了实验任务和收益,但不会说明这是实验,而是由网站用户自行选择是否参与"工作"。一旦用户选择接受"工作",研究者会将他们随机分配到不同的实验局中,并通过向不同实验局的被试呈现在诸如实验指令、薪酬方案或视觉刺激上不同的界面而实现干预。研究者也会通过该界面向被试发布指令和提供信息,记录被试决策,并决定被试收益。被试完成实验任务(通常是完成工作或做出选择)后提交任务并获得报酬,如同完成该市场上的其他任何工作一样(Horton et al,2011)。

①需要说明的是,也有一些实验将实验室搬到了互联网上,通过互联网开展的传统实验室实验不包含在本文的综述范围内。在线人工或框架在线实地实验也不包含在本文的综述范围内。

②选择对劳动经济学领域的在线实地实验进行梳理是出于该领域包含的研究主题丰富、技术干预手段涵盖广泛、可供研究的网站和平台众多、与人们的日常生活息息相关等考虑。劳动经济学领域其他实验类型的研究综述可参见 List & Rasul(2011)和 Charness & Kuhn(2011)。

③在电子商务网站(例如 eBay、亚马逊等)和门户网站(例如雅虎、谷歌等)上开展的实地实验不包含在本文的梳理范围之内,因为这些实验通常为评估新想法、指导产品开发或改进界面设计而非科学研究而开展,尽管其设计和实施方法与科学研究类似。

④与其他实验类型类似,在线实地实验开展之前,需要通过研究者所在机构的学术伦理审查,以保证实验的道德性。学术伦理审查在西方学术界和业界已经是非常惯常的做法,有成熟的标准和流程,但在中国仍未得到广泛开展。目前,一些大学已经成立了学术伦理委员会,可望在未来发挥更大的作用。

(二)实验基本要素的特征

根据 Harrison & List(2004)提出的实验的六个基本要素,本文发现在线实地实验在被试群体的性质、实验任务的性质和实验环境的性质方面与线下实地实验表现出明显差异,因其"在线"属性而存在明显的特征。

1. 被试招募便捷高效,被试数量多且多样化。在线实地实验的被试通常为网站、在线平台或社区的用户,通过自然选择或抽样的方式进行招募(Chen & Konstan,2015)。自然选择指研究者借助网站规则,以用户身份开展正常的在线活动(例如通过在线劳动力市场上发布任务,通过在招聘网站上进行职位招聘等),感兴趣的用户自主选择参与到活动中,成为实验的被试。例如,Flory et al (2015)在招聘网站上发布职位广告,将有意向的近 9000 名求职者作为被试。这种招募方式无须网站运营者的协助,操作便捷高效,因此是大部分在线实地实验采用的招募方式。抽样则指的是研究者在明确抽样框(通常是某网站的全部用户信息)的情况下,使用科学抽样方法生成样本,或直接将总体作为被试。例如,Gee(2019)在"领英"(LinkedIn)的协助下使用抽样方法,选取了符合要求的230 万求职者作为被试。由于该招募方式对抽样框有严格要求,研究者通常需要得到网站运营者的支持,因此需要花费一些额外成本。

无论何种招募方式,被试均具有数量多且多样化的特点。被试招募方式决定网站的注册用户均可能成为实验被试,如果实验在一些大网站开展,则被试数量可高达百万级别,线下实验通常难以企及。由于网站用户在成长背景、技能水平、社会环境等方面存在巨大差异,呈现出多样化特点,使得一些问题的研究成为可能。例如,Lyons(2015)为探究国籍多样性对团队生产率的影响而在 oDesk平台上招募来自不同国家的劳动者完成编程任务,这样的被试要求在线下实验中通常很难满足。

- 2. 实验任务与在线背景紧密结合。虽然在线实地实验的具体任务随研究主题而变化,但是均与适合探究该主题的实验场所依托的网络服务紧密联系,并借助该实验场所的主要职能和运行规则设计和开展,在线进行全部流程。例如,数据录入(Gilchrist et al,2016)、图文转译(Chen & Horton,2016)、翻译(Liu et al,2014)等工作是在线劳动力市场实验中常见的任务,而需要被试"在场"完成的任务(例如货物搬运、面对面交流等)则不适合于在线实验中开展。
- 3. 在线环境具有强非介入性。在线实地实验的环境通常允许更强的非介入性(unobtrusiveness),也即允许研究者实现更隐蔽的实验干预。例如,在线劳动力市场的劳动者在面对研究者设置的真实任务和激励而付出真实努力时,通常不会怀疑自己正在被观察(Horton et al,2011)。相应地,为保证自然实地实验的属性(Harrison & List,2004),大部分实验放弃了向被试做知情同意的说明。研究者担心被试如果了解实验和清楚自己处在被监控状态下,他们的行为可能会因此发生改变(List,2008)。

(三)技术干预手段

在线实地实验的干预手段也与线下实地实验存在较大差异, Chen & Konstan (2015) 将在线实地实验特有的技术干预方法总结为四类:通过邮件与短信发送定制信息、修改网站界面、编写自动程序或脚本、添加附加组件。以下结合最新的研究成果,依次介绍劳动经济学领域在线实地实验常用的技术干预手段。

1. 发送电子邮件、短信和站内信等定制信息。网络技术允许研究者向不同网站用户发送差异化的定制信息来实现干预。这种定制化信息通常通过发送电子邮件和短信或站内信等方式实现(Chen & Konstan,2015)。研究者根据用户的网站账号或电子邮件地址等特异性信息识别被试,采用分层随机(根据用户关键个人特征)或简单随机的方式将被试分配到不同实验局,通过上述三种手段向不同实验局的被试发送除关键干预信息外完全相同的内容,并继续通过网站账号追踪被试在接受干预后的决策和行为。通常,实验者以网站用户身份即可完成上述操作,借助网站规则和功能引入干预措施并收集干预结果。例如,Flory et al(2015)使用雇主账户在招聘网站上面向美国 16 个城市的求职者发布职位广告,向有意向的求职者随机发送除薪酬方案外信息全部相同的电子邮件,通

过对比职位申请完成情况发现,当提升薪酬方案的竞争性时,男性和女性都表现出规避倾向,但女性的规避倾向更明显。

- 2. 修改网站界面。在线实地实验另一种常用的技术干预手段是修改网站界面。研究者为同一网站页面设计了包含不同信息的多种界面,通过向不同实验局的网站用户显示包含不同形式或内容的界面来实现干预。如果网站要求用户登录后才能使用服务,研究者可以将用户的网站账号作为特异性信息,对被试进行识别、随机干预和决策追踪。对于不需要用户登录即可浏览的网站界面,Chen & Konstan(2015)介绍了对话跟踪、IP 地址和信息记录程序(cookies)三种技术手段实现随机干预的目的。无论是否要求用户登录,采用修改网站界面干预手段的研究者通常需要网站运营者的支持,对网站规则进行局部调整才能实现。例如,Gee(2019)在"领英"的支持下更改网站页面设置,根据网站账号将被试随机分为能否在招聘广告页面上看到该工作实时申请人数信息两组。研究发现,添加实时申请人数信息将使求职者开始和完成申请的概率提升 $1.9\%\sim3.6\%$,且提升女性的概率大于男性。
- 3. 编写自动程序或脚本。编写自动程序或脚本是网络技术提供的另一种干预方法。研究者通过编写自动程序或脚本,为网站用户提供新的功能,通过调整程序的适用范围或面向不同用户的程序功能来实现干预。该方法通常借助网站账号识别和追踪被试,因为这些程序或脚本提供的服务一般需要用户登录后才可使用。与修改网站界面类似,使用该方法的研究者通常也需要网站运营者的密切配合,协助研究者将实验所需的程序或脚本写入网站规则。例如,Horton(2017)通过在线劳动力市场 oDesk 采用一个自动匹配求职者的程序向随机选取的雇主推荐求职者,而其余雇主仍需自己搜索适合的求职者。研究发现,基于算法的求职者推荐使职位空缺填充率提高了 20%,且未发生推荐导致挤出未被推荐者的效应。

三、在线实地实验的优势和潜在问题

本节通过分析在线实地实验与其他数据收集方式(观测和调查、线下实地实验)相比存在的优势和潜在的问题,进而讨论在线实地实验的重要性、可行性及其研究结果的外部有效性。

(一)优势

相比于观测和调查的数据收集方式,在线实地实验具有所有实验方法共有的能够有效缓解内生性问题的优势。而相比于线下实地实验的数据收集方式,在线实地实验的优势体现在干预的随机性更强、可获得大量样本的成本更低、获得数据的准确性更高以及能够研究新的问题。

1. 内生性问题的有效缓解。通过观测或调查方式收集的数据通常在社会经济现象发生后由研究者被动收集,不可避免地存在干预条件(关键自变量)与不可观测的被试特征相关,导致在识别干预条件对结果变量的因果影响时存在偏误,也即内生性问题。观测和调查数据分析中的内生性问题通常可通过固定效应模型、双重差分、工具变量、断点回归、结构模型等计量经济学方法来缓解,但上述方法的适用条件通常有严苛限制,经验分析中很难完全满足。

实验则能够根据研究问题来设计和操控干预条件从而主动创造数据。具体而言,实验既能独立变动每种干预条件的不同水平(直接操控),又能对被试进行随机分组保证其可观测特征不存在统计上的显著差别(间接操控)。直接操控能够保证从干预结果的差别上直接识别出干预条件水平差别的影响,间接操控能够保证多种干预条件不同水平的组合(即实验局)的随机分配与被试特征①的正交,因而能在最大程度上缓解甚至解决内生性问题(Friedman & Sunder,1994)。因此,普通最小二乘法或非参数估计方法即可胜任实验数据的因果推断需要,不仅降低了对模型函数形式和误差项分布假设的要求,而且使得结果解释更加简单直观。另外,基于 List & Rasul(2011)的讨论,对于许多不可观测的个人或组织特征(如偏好、动机),在有激励的实验环境中可以更加准确地被测量,从而减少了测量误差。

①如果样本量足够大(统计上要求每个实验局至少包含 30 个独立的观测值),随机化也能保证不可观测的被试特征在各实验局之间的均衡。

例如,传统上对于补偿性工资差异的估计,通常采用雇员的工资和工作特征等观测数据,基于特征定价模型(Rosen,1974,1986)将工资对工作特征进行回归。由于可能存在不可观测的随时间变化的雇员或雇主的异质性特征,且这些特征与工作特征相关,即使采用固定效应模型估计出的工作特征的系数也可能存在偏误。然而,向求职者随机分配研究者关心的工作特征和不同工资的成对组合,并让求职者选出心仪组合的实验方法能够有效缓解内生性问题。 $Mas\ \&\ Pallais(2017)$ 考察了求职者对传统工作安排(朝九晚五、每周 40 小时、办公室工作)和弹性工作安排间的偏好,发现大多数求职者不愿为时间弹性牺牲薪酬;平均愿意放弃 8%的薪酬以换取在家工作;平均愿意放弃 20%的薪酬以避免雇主临时安排工作。

- 2. 干预的随机性更强。首先,在线实地实验允许在实验前更好地将被试随机分组。研究者如能得到网站运营者的协助,则可以获得用户的登记数据和在网站上的历史行为数据,有助于揭示线下实地实验难以获得的被试特征,为保证更多维度的被试特征在各实验局之间的均衡提供了便利。其次,在线实地实验允许在实验后更好地处理随机性未得到良好执行而带来的问题。例如,线下实地实验干预条件分配的随机性可能受到被试系统性退出实验损耗(attrition)和干预溢出(spillover)等问题的影响。系统性退出实验损耗指不同干预条件的被试未完成实验的比例存在系统性差异,因此会导致干预结果的差别并不能完全归因于干预条件的差别。在线实验环境可以记录用户的完备行为档案,利用这些资料可以有效帮助研究者理解并采用统计方法处理这种损耗而校正干预效果(Parigi et al,2017)。干预溢出通常指由于干预组与控制组被试发生互动而导致控制组被试错误地接受干预的情形。Muise & Pan(2019)指出,由于多数网站不支持被试间的线上交流(如招聘网站),或这种交流很容易被监测到(如 MTurk 的劳动者交流论坛),因此研究者可以控制干预溢出的发生。
- 3. 获得大量样本的成本更低。线下实地实验的实施成本通常是非常高昂的,样本量因此受到限制。例如,Shearer(2004)用于考察计件工资和固定工资的激励作用差异的实地实验中仅雇用了 9 名工人完成种树任务,每人工作 120 天。调查数据由于需要一个个受访者进行调查,不存在规模经济,通常也需要花费高额成本。而在线数据的收集依托网站,通常仅需要支付一个固定成本,甚至可以在大平台中几乎毫无成本地收集大量数据(Horton & Tambe,2015),这使得在线实验的样本量通常比同类线下实验大得多。例如,He et al(2019)通过电子邮件和招聘网站站内信向近 10 万求职者发送了工作广告,Gee(2019)在"领英"的协助下获得了 230 万求职者的数据。
- 4. 所获得数据的准确性更高。网站或在线平台在用户正常活动和决策时就会将诸多数据记录下来。以在线劳动力市场为例,劳动者的受雇历史、历史薪酬水平、劳动供给时间和工作表现等信息均会在他们正常的活动中被记录下来。比起大多数调查数据依赖于受访者自报告,线下实验数据依赖于人工观测记录,在线实地实验的数据显然具有更高的准确性。例如,Barach & Horton(2019)在探究是否向雇主展示劳动者的历史薪酬信息对其录用可能性的影响时,获得了来自在线劳动力市场自动记录的劳动者过往工作历史和薪酬,而这些信息在其他情况下可能被瞒报或误报。
- 5. 能够研究新的问题。由于在线环境拥有线下不存在的运行机制和功能,从而可为仅用线下数据难以回答的问题提供了研究可能。首先,线下环境通常只能研究不同干预条件对已雇用员工工作表现的影响,而难以考察对求职者进入工作决策的影响,而在线环境使后者成为可能,例如上文提到过的 Flory et al(2015)和 Gee(2019)等一系列的实验。其次,网站或在线平台通常存在作为消减信息不对称手段的声誉系统。Benson et al(2019)就探究了这种声誉系统的有效性。通过在 MTurk上创建被赋予好评、差评和无评论的雇主账号各 12 个,并用这些账号发布几乎完全相同的工作任务,研究者发现声誉更高的雇主吸引劳动者参与工作的速度更快,证明了声誉系统的有效性。除此之外,网站或在线平台还可以运用计算机算法提升传统工作的效率,例如上文提到过的 Horton(2017)的实验。

(二)潜在问题

尽管通过在线实地实验方式收集数据拥有众多优势,其重要性不言而喻,但也存在一些潜在问题,例如,样本选择、被试和任务的局限性以及网站运营者协助难等问题。

- 1. 样本选择。这是在线实地实验最突出的问题(Horton & Tambe, 2015)。一方面,如果无法取得网站或在线平台的全部用户的数据库(这通常很难),那么几乎不可能对实验总体的特征有比较准确的把握;另一方面,参与在线活动本质上是一种自主选择的结果,很难确定线上、线下市场参与者的差异程度和这些差异的影响。而比较线上研究数据和抽样规则明确的调查样本数据的描述性统计结果的相似程度是缓解这一问题的一种可能方式。
- 2. 被试和任务的局限性。由于网络活动的参与者至少要掌握基本的上网技能,拥有网络连接设备,这对被试的知识技能水平和经济状况做了隐性限制。这就意味着,如果研究者关注的是低收入、低受教育水平群体,或者希望在网络普及率和人均知识技能水平偏低的欠发达国家开展研究,应当慎重考虑在线用户的代表性和研究结果的外部有效性。此外,在线环境也限制了实验任务的选择,任务既需要适合在线开展,也需要符合开展场所的主要职能和运行规则。这也意味着,需要被试"在场"完成的任务不适合于通过在线实验中开展(Horton et al,2011)。
- 3. 网站运营者协助难。在需要网站或平台协助的情况下,说服网站运营者配合完成实验干预或向研究者共享数据通常并不容易。Horton & Tambe(2015)指出,共享数据对网站运营者的成本可能包括潜在的信息泄露危险、将数据转换为研究用数据的成本等,而协助干预措施增加的成本可能包括对用户体验的影响、增加网站相关员工的工作量等,并指出通常只有使网站运营者相信研究会为公司带来一些好处时,这种协助才可能发生。研究者在网站现场使用数据并掌握一定的数据提取和操控技术是缓解这一问题的一种可能方式。

四、在线实地实验在劳动经济学中的应用进展

本节以解释劳动力市场表现在不同群体间存在差异的原因、缓解劳动力市场中的信息不对称问题以及深入探究激励措施对工作搜寻和匹配的影响为例,回顾 21 世纪以来劳动经济学领域开展的在线实地实验研究的最新进展,展现在线实地实验如何更好地实现对劳动经济学理论的检验。

(一)解释劳动力市场结果在不同群体间存在差异的原因

1. 歧视。Altonji & Blank(1999)研究指出,雇主歧视是造成不同群体间劳动力市场表现差异的一个重要原因。由于使用观测数据直接测度歧视非常困难,早期的经验研究常将群体间劳动力市场表现(工资或劳动参与率等)差距拆分为一个可解释的(个体可观测特征导致的生产率差异)和一个不可解释的差距,将后者视作歧视来间接测度(Oaxaca,1973;Blinder,1973)。上述方法存在的问题是,由于劳动者的生产率可能还存在不可观测的个体差异,或难以获得影响劳动力市场表现的所有市场变量数据,歧视可能被高估或低估(Azmat & Petrongolo,2014)。

与上述方法不同,近些年来兴起的审计实验(audit studies)可直接测度歧视。研究者令除所属人口群体以外的简历信息和其他特征完全相同的劳动者向同一雇主求职,度量雇主对不同群体求职者的面试通知率、雇佣率或工资差异来直接测度雇佣过程中的歧视(Neumark,2018)。通讯实验(correspondence studies)更进一步,在工作申请中投递纸面化(通过邮寄或传真)或电子化(通过网站或电子邮件)的虚构简历,然后收集基于电话或邮件的面试通知作为雇主反馈。通讯实验由于不可控因素少和边际成本低,近年来被广泛应用于在一国多个城市开展实验以获取大规模并具有代表性的样本。电子化的通讯实验属于在线实地实验,被用于研究性别、年龄、种族、犯罪记录、失业时间、性取向等维度是否存在歧视。

在性别维度,Riach & Rich(2006)在英国劳动力市场上向雇主发送成对的虚构简历,发现在"女性职业"如秘书岗位中存在对男性的歧视,在"男性职业"如工程师岗位中,存在对女性的歧视,在两个"中性职业"注册会计师和电脑分析程序员岗位中,也存在对男性的歧视。Petit(2007)在法国金融

部门开展的类似实验显示,25岁的女性相比同龄的男性在申请高技能行政岗位和雇主提供长期劳动合同的岗位时受到严重歧视,而在 37岁单身且无子女和 37岁已婚有三个子女的求职者中,并未发现对女性的歧视。

在年龄维度,Neumark et al(2019)在向 4 万多份工作岗位投递年龄有差异且求职者临近退休年龄的虚构简历后发现,美国雇主存在针对年老女性的年龄歧视。Baert et al(2016)除变动求职者年龄外,还为年长求职者额外的工作经历随机匹配了从事本专业工作、从事专业外工作或未工作等职业经历,发现比利时雇主对年长劳动者的年龄歧视几乎全部来自从事专业外工作的年长求职者。

在种族维度,Oreopoulos(2011)在加拿大开展通讯实验后发现,相比于英文名字的求职者,拥有国外工作经验或拥有印度、巴基斯坦、中国或希腊名字的求职者在多种职位上都得到更少的面试通知,且注明语言流利、拥有跨国公司经验、出身名校或积极参与课外活动等都未能降低这种歧视。Agan & Starr(2017)使用类似方法显示种族信息,并借助美国雇主询问求职者犯罪记录的传统显示犯罪记录信息发现,有犯罪记录的人和黑人会比无犯罪记录的人和白人获得更低的面试通知率。特别地,研究者发现美国禁止"询问求职者犯罪记录"的政策使无犯罪记录的黑人受到了更大的歧视。

在失业时间维度,Kroft et al(2013)通过变动虚构简历中的失业时间发现,失业(特别是失业的前八个月内)时间越长,获得美国雇主面试通知的可能性越低;当地劳动力市场越紧张,雇主们越会使用失业期长度作为未观察到的生产力信号。Eriksson & Rooth(2014)发现,当期的短暂失业和过往的长期失业记录对美国雇主的面试通知率都没有影响,但当期超过 9 个月的失业记录显著降低了面试通知率。

在性取向维度,Tilcsik(2011)通过为虚构的毕业生安排校园同性恋组织参与经历暗示性取向,发现美国部分州的雇主对男同性恋求职者的面试通知率显著降低,这种地区差异在一定程度上反映了地区舆论和立法对同性恋的歧视程度。Weichselbaumer(2015)在使用 Tilcsik(2011)方法的基础上,利用德国劳动力市场要求申报家庭成员情况的传统,通过申报伴侣性别进一步增强了性取向信号,发现慕尼黑的雇主对已婚和未婚女同性恋求职者的面试通知率都较低,而柏林的雇主对女同性恋和异性恋求职者的面试通知率无显著差异。

另外,在线环境允许研究者进行深入的政策研究,评估诸如性别和种族平权政策等旨在消除群体间劳动力市场表现差异的政策的实际效果。Leibbrandt & List(2018)在招聘网站上发布工作广告后,将有意向的求职者随机分配到是否有明确的种族就业机会均等声明的环境中,发现该声明降低了少数族裔的工作申请率,因为他们担心其被雇用只是为了象征性地表示种族平等,而非因为他们的能力。Ibañez & Riener(2018)结合线上线下手段发布招聘广告,通过随机变动申请前或后告知性别平权行动宣言,发现申请前告知使女性申请率增加,并超过男性申请率的降低,而最优秀的男、女性申请率并未下降。两项研究对政策作用的评价存在的差异值得深思,可能与实验实施环境(Leibbrandt & List 在美国实施,Ibañez & Riener 在哥伦比亚实施)、平权政策内容均有关。这些研究结果为实施消除群体分隔现象的政策建议提供了科学依据。

2. 个人偏好。近年来,经济学家开始考察来自劳动力市场供给方的心理特质和偏好等因素对群体间(特别是不同性别的劳动者)劳动力市场表现差异的解释能力。这是因为,不同的风险和竞争偏好可能导致劳动者自主选择进入不同的行业和职位工作;不同的谈判偏好导致从事相同行业和职位的劳动者可能获得的薪资和职业发展路径不同;不同的社会偏好(利他、公平、信任、互惠、合作等)导致在相同工作场所中劳动者的表现不同(Azmat & Petrongolo,2014)。而这些心理特质和偏好难以用观测数据精确度量。因此,大量经济学实验研究(主要来自于实验室实验)尝试测度性别偏好差异。Croson & Gneezy(2009)及 Niederle(2016)回顾相关研究发现,男、女性在风险偏好、竞争偏好和社会偏好方面存在稳健的差异:女性比男性更厌恶风险,更厌恶竞争,更不愿直言想法,且具有更强的社会偏好特异性。

实地实验尝试在真实劳动场景中验证从实验室中获得的结论是否仍然有效,在线实地实验在此 — 138 — 基础上增加了数据收集和干预实施的便利,出现了一批关注群体间偏好差异的研究。例如,Flory et al(2015)和 Samek(2019)均探究了性别竞争偏好差异,分别通过对近 9000 名和 35000 名求职者随机发布工作内容相同但薪酬方案竞争程度不同的招聘广告发现,高竞争性薪酬方案使两性均表现出规避倾向,但女性受到的影响更大。Leibbrandt & List(2015)则关注性别工资协商偏好的差异,其把近 2500 名对招聘网站上发布的工作广告有意向的求职者随机分配到是否明确说明工资可协商的合同环境中发现,男性更可能在未明确说明协商可能性时进行工资协商,且男性更愿意申请未明确说明工资可协商环境下的工作。以上研究在大样本和真实环境中证实了实验室实验的研究结论,为性别劳动力市场表现差距的一种可能解释提供了更令人信服的证据。

(二)缓解劳动力市场中的信息不对称问题

劳动力市场存在严重的信息不对称(拉斯穆森,2017),例如雇主无法从简历中观察到劳动者的生产率,无法确定计时工资下劳动者的努力程度(鲍哈斯,2018)都是劳动力市场信息不对称的典型例证。为了缓解这个问题,理论与经验研究进行了大量探索,传统劳动经济学研究的回应是统计性歧视(Aigner & Cain,1977)和效率工资理论(Katz,1986)。在线环境给予研究者一个新的选择:某些信息不再缺失,或雇主可以在支付低成本的情况下将缺失的信息补充,如通过雇主工作评价、历史薪酬信息、员工推荐或平台推荐补充劳动者的生产率信息,通过雇主声誉、其他申请者的工作接受率或申请数量补充雇主或职位信息,进而探究雇主和劳动者的行为会发生哪些变化。

研究者发现,通过在线劳动力市场上补充或遮蔽劳动者的劳动生产率信息,可以有效改变雇主行为。例如,Pallais(2014)在雇佣结束后为部分欠缺工作经验的员工撰写了详尽的工作表现评价;Pallais & Sands(2016)则通过老员工推荐雇用了一批劳动者;Barach et al(2019)更改系统设置引入了平台担保(即如雇主在雇用后对受担保劳动者不满意,平台可承担其前两周的工资)或推荐劳动者的制度。以上研究均发现,补充劳动者的劳动生产率信息会提升其被雇用的可能性。Barach & Horton(2019)则从反面探究遮蔽该类信息的影响,他们通过系统设置使占样本一半的雇主无法查阅劳动者的历史薪酬信息,发现在这种情况下历史薪酬低的劳动者被雇用的可能性增大。

研究者还发现,通过向劳动者提供有关雇主或职位的社会信息也会改变劳动者行为。例如, $Coffman\ et\ al(2016)$ 对一半被录用者在工作录用通知邮件末尾补充前一年被录用者的实际工作接受率信息,Gee(2019)则更改招聘网站系统设置,使部分求职者可见工作的实时申请人数。两项研究均发现,补充其他劳动者的决策信息增加了工作申请/接受率。

此外,研究者还特别关注了在线平台普遍采用的声誉系统作为信息补充手段的效果。例如,Benson et al(2019)在 MTurk 上创建了分别被赋予好评、差评和无评论的雇主账号,用以发布相同工作,发现高声誉雇主吸引劳动者参与工作的速度更快。虽然声誉信息有效改变了决策行为,但 Filippas et al(2018)发现劳动力市场上的公开声誉评价系统存在严重的声誉"膨胀"现象。研究者要求某在线劳动力市场上的雇主同时在公开和秘密的两个声誉系统中为雇员评分和评价,当 28.4%的雇主秘密反馈不会再雇用某劳动者时,却公开反馈了 $4\sim5$ 星好评(5 星最高)。

(三)深入探究激励措施对工作搜寻和匹配的影响

雇主提供的工作职位空缺和求职者之间如何匹配是影响劳动力市场有效运行的另一个重要因素。众多关于工作搜寻与匹配的理论研究关注诸如工资等激励措施在这一过程中的作用,发展出随机搜寻(random search,例如 Diamond, 1982; Mortensen, 1982; Pissarides, 1985)和有目的搜寻(directed search,例如 Moen, 1997; Menzio et al, 2016)两种理论。但是,采用观测数据进行的经验研究难以保证除激励外的其他工作特征完全相同,因此难以获得激励措施与求职者申请数或申请率之间的因果关系(Holzer et al, 1991; Banfi & Villena-Roldán, 2019; Marinescu & Wolthoff, 2020)。而采用实地实验方法进行的经验研究虽然能够缓解内生性问题,但由于开展机会难得而使得实验实施环境通常比较特殊,体现在研究对象(例如 Deserranno(2019)仅在乌干达小额贷款组织中招募社区健康促进员)或招聘工作职位(例如 Dal Bo et al(2013)仅在墨西哥招募公务员)等方面。

在线实地实验提供了更加可靠和具有说服力的证据。在线实地实验不仅可以保证除激励措施外不同的工作职位的其他工作特征完全相同,而且可以确保被试背景和工作职位所在行业等实验实施环境的一般性和多样性。因此,不同职位间的申请数或申请率的差异就可以被归因于激励措施的差异,而且研究结果也具有更高的外部有效性。在线实地实验的另一贡献在于将金钱以外的非物质激励也纳入了实验,探究为社会做出有益贡献等是否能吸引更多求职者申请。例如,Hedblom et al (2019)在招聘网站上随机选择是否向求职者展示有关企业社会责任的信息并随机给予高低不等的时薪级别,发现提高时薪和展示企业社会责任的信息均能提高申请率。Flory et al (2015)则在发布职位广告后,向有意向的求职者随机发送除薪酬方案外信息全部相同的电子邮件,研究发现,当薪酬方案的竞争性提升时,劳动者的申请率下降。

五、基于中国环境的在线实地实验进展及展望

随着互联网在中国的普及以及各类网站和在线平台的创立,过去十年,基于中国环境开展的在 线实地实验也有诸多发展。本节通过介绍现有研究对在线实地实验发展做出的贡献以及存在的不 足,提出对未来研究的展望。

(一)国内研究总结

总体而言,由于中外网站和在线平台的类型和功能相似,因此基于中国环境开展的在线实地实验的研究领域和方法与国外的研究类似。然而,由于中国存在独特的制度、文化或问题,且一些网站运营者对与研究者合作持欢迎态度,这为关注中国问题、讲好中国故事创造了条件。下文通过归纳在劳动经济学、传播学、营销学等领域开展的研究进而总结它们的贡献。

- 1. 关注中国特有的制度和问题。在中国开展的一些在线实地实验关注了中国特有的制度、文化或问题,丰富了研究主题。对于歧视进行测度的中国研究与国外研究类似,大都采用通讯实验的方法,而关注的主题除与国外研究类似的性别歧视(Zhou et al,2013;Maurer-Fazio & Lei,2015;葛玉好等,2018)和民族歧视(Maurer-Fazio,2012)外,还包括计划生育政策等具有中国特色的主题。例如,He et al(2018)基于中国 2016 年"计划生育"政策的变化和求职者在政策前后生育二孩可能性的变化,在"前程无忧"招聘网站上完成通讯实验,发现因生育可能带来的预期家庭责任增加使得女性收到的面试通知率显著降低,但男性不受影响,且年龄增长导致的女性生育概率增加会进一步加剧雇主歧视。
- 2. 与网站运营者的深度合作。在中国开展的一些在线实地实验由于得到了网站运营者的深度合作和密切支持而得以顺利开展,并获得具有外部有效性的结论。例如,He et al(2019)就中国劳动者对非物质激励的支付意愿展开了研究,通过与中国某大型在线招聘平台合作,向超过 10 万潜在求职者随机发送弹性工作条件和薪酬不同而其他条件相同的工作广告,发现有弹性的工作吸引了更多求职者申请,申请者愿意为工作弹性损失 10%的工资,而基于近 200 万求职者的观测数据的分析却未发现类似结论。

营销学领域的实验主要关注社会关系和传播对营销结果的影响,通常在新型社交媒体(新浪微博、微信等)上开展。例如,Gong et al(2017)与某媒体公司合作,通过控制电视节目是否在新浪微博上宣传且请有影响力的微博用户(大 V)转发宣传,发现新媒体和大 V 转发显著增加了节目收视率:收视率因公司微博宣传增加 77%,因微博大 V 转发额外增加 33%,因转发详细播出信息额外增加 57%。微博大 V 的转发还通过增加公司微博的粉丝数间接增加了收视率。Yang et al(2016)与某服装公司微商合作,在微信上通过向新老客户展示商品描述、顾客反馈等发展商业关系,或展示旅游、美食、音乐推荐等发展朋友关系,发现与客户建立商业或朋友关系均提升了销售额,与新客户发展朋友关系可协助发展商业关系,但与老客户同时发展商业和朋友关系却使上述正向影响消失。

除此之外,P0 et al(2020)对 P2P 信贷还款依从行为进行了考察。研究者完整追踪了 2012 名中国大学生在某 P2P 信贷网站上的贷款和还款过程,通过调整提醒借款人还款的短信内容发现,提及

按时还款的积极期望显著增加了还款可能性,强调违约的消极后果仅在短期增加还款可能性,是否披露借出者身份(姓氏)信息对还款行为无影响。

尽管上述基于中国环境的研究为在线实地实验的发展做出了贡献,但是,现有研究仍存在着在线实验研究缺乏①、应用领域比较单一等问题,与网站运营者的合作也存在广阔空间。因此,这也是未来研究可望发展的方向。

(二)未来研究展望

Lucking-Reiley(1999)较早地借助在线环境开展实地实验的研究,彼时研究者还只能通过论坛、电子邮件艰难地开展实验。随着 21 世纪以来劳动、招聘、社交、教育等各类在线平台的出现,实地实验的应用场景和可探究的问题都出现了爆炸式的拓展。虽然国内外网络环境不完全一致,但上述研究使用的各类在线平台和工具在中国同样得到了迅猛发展,如与 Facebook 功能类似的微博、和领英功能近似的智联招聘等,实施环境不应该成为在中国开展此类研究的限制。截至 2018 年底,中国网民规模已达到 8 29 亿,互联网普及率达 59 6 % (中国互联网信息中心,2019),在线生活方式的逐渐普及为在中国巨量且更具人口代表性的互联网用户中开展在线实地实验创造了得天独厚的环境。

本文认为,基于中国环境的在线实地实验在未来大有可为,并提出以下几个可能的研究方向。

1. 拓宽应用领域,深入研究话题。目前基于中国环境开展的在线实地实验大多集中在劳动经济学领域,其他领域的应用相对不足,仍需继续拓展。国外研究除本文中提到的主要领域之外,还包含拍卖(Katkar & Reiley, 2006)、产业组织(Lewis & Reiley, 2014))等经济学子领域和传播学(Muise & Pan, 2019)、心理学(Parigi et al, 2017)、营销学(Johnson et al, 2016)、政治科学(Bond et al, 2012)等其他社会科学领域。相关研究已经证明,在线实地实验可以被广泛应用在各学科的研究中。国内学者可以借鉴上述做法,并积极探索在中国网络的合适环境中进行应用的可能性,用以解决更多领域的问题。

此外,部分基于中国环境开展的现有研究,需要更好地与前沿理论结合,继续深入探索,而不能是仅仅更换一个被试群体。以目前国内应用在线实地实验较多的劳动经济学研究为例,大多数研究仍关注性别歧视等传统话题,而国外研究已深度探索性别偏好差异等新问题,为前沿理论提供了经验证据。因此,研究者应结合理论前沿,利用在线实地实验的优势,为各领域的深入探索提供助益。

- 2. 继续深化研究者与在线环境的结合。当前基于中国环境开展的在线实地实验大多数仅以用户的身份进行实验,而部分得到了网站运营者的合作,未来的发展方向是继续加强与网站运营者合作开展实验,甚至是自己创立网站进行实验。与网站运营者合作虽然有不少困难和成本,但如果研究者能关注在线平台企业同样关注的兼具现实和理论意义的重要问题,则常常能够获得意想不到的支持,这也是前述许多国内外研究得以开展的基础。此外,Centola(2010,2011)更进一步地建立了完全由研究者运营的健康类社区网站开展行为和传播领域的研究。这样的网站显然能实现更加丰富的干预,探索更深入的问题。而芝加哥和哈佛大学的研究者则自己成立了公司,用于更好地进行对劳动力供给行为的研究。虽然上述设想会增加实验成本,但随着在线实地实验在中国继续发展,相信研究者会寻求与在线环境的更深度结合。
- 3. 立足中国实际,关注中国问题。国内外的制度、文化和环境差异决定了我们不能一味模仿国外研究。作为中国学者,应该与中国环境更密切地结合,关注中国问题。在这方面,现有部分研究已为我们做出了示范(参见 He et al,2018)。这些中国问题既是帮助我们理解现实政策的切入点,也是国外环境所不具备的且能做出研究贡献的机会,因而应该被更好地关注和把握。

总之,不断发现、不断创新的敏锐思维,对时代发展潮流的良好把握,以及对新兴研究环境、方法

①据我们所知,除上述研究外,已经公开为工作论文的研究还包括 Liu et al (2014) 在众包网站"任务中国" (Tasken) 考察物质激励对任务质量的影响; Ong & Wang(2015)和 Ong(2016)在婚恋网站上开展的考察性别、收入和受教育程度对异性关注程度影响的通讯实验; Kizilcec et al(2017)在中国慕课(MOOCs)探究价值肯定对学习行为和结果的影响; Morgan et al(2018)在购物网站(淘宝)上对有限注意力假说的验证等。

和工具的有效利用,都为中国经济学能在短时间内获得跨越式发展提供了契机,并提供了更多对中国重要经济学问题进行探索的可能性。

参考文献:

- 鲍哈斯,2018:《劳动经济学》,中国人民大学出版社。
- 葛玉好 邓佳盟 张帅,2018:《大学生就业存在性别歧视吗?——基于虚拟配对简历的方法》,《经济学(季刊)》第 4 期。 拉斯穆森,2017:《博弈与信息:博弈论概论》,中国人民大学出版社。
- 中国互联网信息中心,2019:《第 43 次〈中国互联网络发展状况统计报告〉》,国家互联网信息办公室网站,http://www.cac.gov.cn/2019-02/28/c_124175677.htm。
- Agan, A. & S. Starr(2017), "Ban the box, criminal records, and racial discrimination: A field experiment", *Quarterly Journal of Economics* 133(1);191-235.
- Aigner, D. J. & G. G. Cain(1977), "Statistical theories of discrimination in labor markets", ILR Review 30(2):175-187.
- Altonji, J. G. & R. M. Blank (1999), "Race and gender in the labor market", In: O. C. Ashenfelter & R. Layard (eds), Handbook of Labor Economics, Vol. 3, Elsevier...
- Al-Ubaydli, O. & J. A. List(2015), "Do natural field experiments afford researchers more or less control than laboratory experiments?", American Economic Review 105(5):462-466.
- Azmat, G. & B. Petrongolo(2014), "Gender and the labor market: What have we learned from field and lab experiments?", Labour Economics 30:32-40.
- Baert, S. et al(2016), "Getting grey hairs in the labour market: An alternative experiment on age discrimination", Journal of Economic Psychology 57:86-101.
- Banfi, S. & B. Villena-Roldan(2019), "Do high-wage jobs attract more applicants? Directed search evidence from the online labor market", *Journal of Labor Economics* 37(3):715—746.
- Barach, M. et al(2019), "Steering in online markets: The role of platform incentives and credibility", Forthcoming in *Management Science*.
- Barach, M. & J. J. Horton(2019), "How do employers use compensation history? Evidence from a field experiment", Working Paper, http://john-joseph-horton.com/research.
- Barron, J. M. et al(1985), "Employer search: The interviewing and hiring of new employees", *Review of Economics* and Statistics 67(1):43-52.
- Barron, J. M. et al(1987), "Employer size: The implications for search, training, capital investment, starting wages, and wage growth", *Journal of Labor Economics* 5(1):76-89.
- Belot, M. et al(2017), "How wage announcements affect job search behaviour: A field experimental investigation", European University Institute, Working Paper.
- Benson, A. et al(2019), "Can reputation discipline the gig economy? Experimental evidence from an online labor market", Management Science, https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2019.3303.
- Blinder, A. S. (1973), "Wage discrimination: Reduced form and structural estimates", *Journal of Human Resources* 8(4):436-455.
- Bond, R. M. et al(2012), "A 61-million-person experiment in social influence and political mobilization", *Nature* 489 (7415):295-298.
- Centola, D. (2010), "The spread of behavior in an online social network experiment", Science 329(5996):1194-1197.
- Centola, D. (2011), "An experimental study of homophily in the adoption of health behavior", *Science* 334(6060): 1269—1272.
- Charness, G. & P. Kuhn(2011), "Lab labor: What can labor economists learn from the lab?", In: O. C. Ashenfelter & R. Layard(eds), Handbook of Labor Economics, Vol. 4A, Elsevier.
- Chen, D. L. & J. J. Horton(2016), "Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts", *Information Systems Research* 27(2):403-423.
- Chen, Y. & J. Konstan(2015), "Online field experiments: A selective survey of methods", Journal of the Economic Science Association 1(1):29-42.
- Coffman, L. C. et al (2016), "Can social information affect what job you choose and keep?", American Economic Journal: Applied Economics 9(1):96-117.
- 142 —

- Croson, R. & U. Gneezy(2009), "Gender differences in preferences", Journal of Economic Literature 47(2):448-474.
- Dal Bo. E. et al(2013), "Strengthening state capabilities: The role of financial incentives in the call to public service", Quarterly Journal of Economics 128(3):1169-1218.
- Deserranno, E. (2019), "Financial incentives as signals: Experimental evidence from the recruitment of village promoters in Uganda", American Economic Journal: Applied Economics 11(1):277-317.
- Diamond, P. A. (1982), "Wage determination and efficiency in search equilibrium", Review of Economic Studies 49 (2):217-227.
- Du, N. et al(2020), "Prosocial compliance in P2P lending: A natural field experiment", Management Science 66(1):315-333.
- Eriksson, S. & D. Rooth(2014), "Do employers use unemployment as a sorting criterion when hiring ?Evidence from a field experiment", *American Economic Review* 104(3):1014—1103.
- Faberman, R. J. & G. Menzio(2018), "Evidence on the relationship between recruiting and the starting wage", Labour Economics 50:67-79.
- Falk, A. & J. J. Heckman(2009), "Lab experiments are a major source of knowledge in the social sciences", *Science* 326(5952):535-538.
- Filippas, A. et al(2018), "Reputation inflation", Paper presented at the *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 483-484.
- Flory, J. A. et al(2015), "Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions", *Review of Economic Studies* 82(1):122-155.
- Friedman, D. & S. Sunder(1994), Experimental Methods: A Primer for Economists, Cambridge University Press.
- Gee, L. K. (2019), "The more you know: Information effects on job application rates in a large field experiment", Management Science 65(5):2077-2094.
- Gilchrist, D. S. et al(2016), "When 3+1>4: Gift structure and reciprocity in the field", Management Science 62 (9):2639-2650.
- Gong, S. et al(2017), "Tweeting as a marketing tool: A field experiment in the TV industry", Journal of Marketing Research 54(6):833-850.
- Harrison, G. W. & J. A. List(2004), "Field experiments", Journal of Economic Literature 42(4):1009-1055.
- He, H. et al(2018), "Labor market discrimination against family responsibilities: A correspondence study with policy change in China", Working Paper, https://www.econ.iastate.edu/files/events/files/li-labormktdiscrimination.pdf.
- He, H. et al (2019), "Do workers value flexible jobs? A field experiment on compensating differentials", NBER Working Paper, No. 25423.
- Hedblom, D. et al(2019), "Toward an understanding of corporate social responsibility: Theory and field experimental evidence", NBER Working Paper, No. 26222.
- Holzer, H. J. et al(1991), "Job queues and wages", Quarterly Journal of Economics 106(3):739-768.
- Horton, J. J. (2017), "The effects of algorithmic labor market recommendations: Evidence from a field experiment", Journal of Labor Economics 35(2):345-385.
- Horton, J. J. et al(2011), "The online laboratory: Conducting experiments in a real labor market", *Experimental Economics* 14(3):399-425.
- Horton, J. J. & P. Tambe(2015), "Labor economists get their microscope: Big data and labor market analysis", *Big Data* 3(3):130-137.
- Ibañez, M. & G. Riener(2018), "Sorting through affirmative action: Three field experiments in Colombia", Journal of Labor Economics 36(2):437-478.
- Johnson, G. A. et al(2016), "When less is more: Data and power in advertising experiments", *Marketing Science* 36 (1):43-53.
- Katkar, R. & D. Reiley(2006), "Public versus secret reserve prices in ebay auctions: Results from a pokemon field experiment", *Economic Analysis and Policy* 6(2):1-25.
- Katz, L. F. (1986), "Efficiency wage theories: A partial evaluation", NBER Macroeconomics Annual 1:235-276.
- King, G. et al(2014), "Reverse-engineering censorship in China; Randomized experimentation and participant observation", Science 345(6199):891.

- Kizilcec, R. F. et al(2017), "Towards equal opportunities in MOOCs: Affirmation reduces gender & social-class achievement gaps in China", Paper presented at the Fourth (2017) ACM Conference on Learning at Scale, pp. 121—130.
- Kroft, K. et al(2013), "Duration dependence and labor market conditions: Evidence from a field experiment", Quarterly Journal of Economics 128(3):1123-1167.
- Leibbrandt, A. & J. A. List(2015), "Do women avoid salary negotiations? Evidence from a large-scale natural field experiment", *Management Science* 61(9):2016-2024.
- Leibbrandt, A. & J. A. List(2018), "Do equal employment opportunity statements backfire? Evidence from a natural field experiment on job-entry decisions", NBER Working Paper, No. 25035.
- Lewis, R. A. & D. H. Reiley(2014), "Advertising effectively influences older users: How field experiments can improve measurement and targeting", *Review of Industrial Organization* 44(2):147-160.
- List, J. A. (2008), "Informed consent in social science", Science 322(5902):672.
- List, J. A. & I. Rasul (2011), "Field experiments in labor economics", In: O. C. Ashenfelter & R. Layard (eds), Handbook of Labor Economics, Vol. 4A, Elsevier.
- Liu, T. X. et al(2014), "Crowdsourcing with all-pay auctions: A field experiment on Tasken", Management Science 60(8):2020-2037.
- Lucking-Reiley, D. (1999), "Using field experiments to test equivalence between auction formats: Magic on the internet", American Economic Review 89(5):1063-1080.
- Lyons, E. (2015), "Team production in international labor markets: Experimental evidence from the field", *American Economic Journal*: *Applied Economics* 9(3):70-104.
- Marinescu, I. & R. Wolthoff(2020), "Opening the black box of the matching function: The power of words", *Journal of Labor Economics* 38(2):535-568.
- Mas, A. & A. Pallais (2017), "Valuing alternative work arrangements", American Economic Review 107(12):3722-3759.
- Maurer-Fazio, M. (2012), "Ethnic discrimination in China's internet job board labor market", *IZA Journal of Migration* 1(1): Article No. 12.
- Maurer-Fazio M. & L. Lei(2015), "'As rare as a panda': How facial attractiveness, gender, and occupation affect interview callbacks at Chinese firms", *International Journal of Manpower* 36(1):68-85.
- Menzio, G. et al (2016), "Directed search over the life cycle", Review of Economic Dynamics 19:38-62.
- Moen, E. R. (1997), "Competitive search equilibrium", Journal of Political Economy 105(2):385-411.
- Morgan, J. et al(2018), "Location still matters: Evidence from an online shopping field experiment", *Journal of Economic Behavior & Organization* 146:43-54.
- Mortensen, D. T. (1982), "Property rights and efficiency in mating, racing, and related games", American Economic Review 72(5):968-979.
- Muise, D. & J. Pan(2019), "Online field experiments", Asian Journal of Communication 29(3):217-234.
- Neumark, D. (2018), "Experimental research on labor market discrimination", *Journal of Economic Literature* 56 (3):799-866.
- Neumark, D. et al(2019), "Is it harder for older workers to find jobs? New and improved evidence from a field experiment", *Journal of Political Economy* 127(2):922-970.
- Niederle, M. (2016), "Gender", In: J. Kagel & A. E. Roth(eds), Handbook of Experimental Economics, Princeton University Press.
- Oaxaca, R. (1973), "Male-female wage differentials in urban labor markets", International Economic Review 14(3):693-709.
- Ong, D. (2016), "Education and income attraction: An online dating field experiment", *Applied Economics* 48(19): 1816—1830.
- Ong, D. & J. Wang(2015), "Income attraction: An online dating field experiment", Journal of Economic Behavior & Organization 111:13-22.
- Oreopoulos, P. (2011), "Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes", *American Economic Journal*: *Economic Policy* 3(4):148-171.
- Pallais, A. (2014), "Inefficient hiring in entry-level labor markets", American Economic Review 104(11):3565-3599.
- Pallais, A. & E. G. Sands (2016), "Why the referential treatment? Evidence from field experiments on referrals",
- 144 —

- Journal of Political Economy 124(6):1793-1828.
- Parigi, P. et al(2017), "Online field experiments: Studying social interactions in context", Social Psychology Quarterly 80(1):1-19.
- Petit, P. (2007), "The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector", *Labour Economics* 14(3):371-391.
- Pissarides, C. (1985), "Short-run equilibrium dynamics of unemployment, vacancies, and real wages", American Economic Review 75(4);676−690.
- Riach, P. A. & J. Rich(2006), "An experimental investigation of sexual discrimination in hiring in the English labor market", The B. E. Journal of Economic Analysis & Policy 5(2):1-22.
- Rosen, S. (1974), "Hedonic prices and implicit markets: Product differentiation in pure competition", *Journal of Political Economy* 82(1):34-55.
- Rosen, S. (1986), "The theory of equalizing differences", In: O. C. Ashenfelter & R. Layard (eds), Handbook of Labor Economics, Vol. 1, Elsevier.
- Samek, A. (2019), "Gender differences in job entry decisions: A university-wide field experiment", *Management Science* 65(7);3272-3281.
- Shearer, B. (2004), "Piece rates, fixed wages and incentives: Evidence from a field experiment", Review of Economic Studies 71 (2):513-534.
- Tilcsik, A. (2011), "Pride and prejudice: Employment discrimination against openly gay men in the United States", American Journal of Sociology 117(2):586-626.
- Weichselbaumer, D. (2015), "Testing for discrimination against Lesbians of different marital status: A field experiment", Industrial Relations: A Journal of Economy and Society 54(1):131-161.
- Yang, S. et al(2016), "The role of business and friendships on WeChat business: An emerging business model in China", Journal of Global Marketing 29(4):1-14.
- Zhou, X. et al(2013), "Gender discrimination in hiring: Evidence from 19,130 resumes in China", SSRN Working Paper, No. 195840.

Research Progress on Online Field Experiments

WENG Qian LI Dong (Renmin University of China, Beijing, China)

Abstract: Online field experiment is a controlled experimental research method emerged with the development of computer technology and the Internet in the 21st century. It has been widely applied in various fields in social science and policy evaluation practices. This paper first describes the design and implementation principles of online field experiments, and then discusses the advantages and potential challenges compared with other methods of data collection. This paper also reviews the latest research progress on online field experiments in labor economics, and summarizes the research progress on experiments conducted in the context of China. This paper aims to introduce a new approach to economic research and to provide references for a better use of Chinese platforms for experimental studies.

Keywords: Field Experiment; Online Experiment; Experimental Economics; Labor Economics

(责任编辑:刘洪愧) (校对:刘新波)