

海外市场、制度环境与本土集聚^{*}

张一力 周 康 张俊森

内容提要: 在研究企业的空间分布时, 一个突出的难题在于我们不得不将企业分配至特定的行政区域内, 这将空间上的点(企业)转化成了平面上不均匀的黑盒子(行政区), 进而破坏了集聚经济的原始状态。借助于新近获取的经纬度坐标数据, 本文在连续空间上开展数值模拟进而克服了这一局限。模拟结果显示中国出口企业在空间分布上呈现出一种新的集聚模式——出口目的地的本土集聚, 即出口到相同海外市场的企业在中国境内呈现出显著的集中倾向。随后我们从出口障碍的视角解释了这一现象, 实证结果显示: 出口目的地的制度环境越差, 出口至该市场的企业在中国境内的集聚程度越高; 另一方面, 当目的国的出口障碍变弱时(如与中国存在相同的意识形态或民族语言), 出口至该市场的企业在中国境内的集聚程度显著降低。这些证据表明出口障碍对中国出口企业的空间分布产生了显著影响。本文的研究既为中国出口企业的空间分布提供了一个新的解释, 也揭示了出口企业挑战高难度出口市场的一种生存法则。

关键词: 本土集聚 制度环境 出口障碍 空间统计

一、引 言

企业的空间分布并非是随机的, 事实上, 企业在一些区域的集聚早已成为中国经济的典型特征。对此, 大量文献对集聚的经济效应进行了研究(范剑勇, 2006; 张海峰和姚先国, 2010; 孙浦阳等, 2013; 包群等, 2012)。概括而言, 相关研究大致呈现出三个特征: 第一, 大多基于产业集聚的理论框架, 尤其注重对集聚溢出效应的考察; 第二, 对于集聚的动力来源, 现有文献大多接受马歇尔的经济外部性假说, 即集聚产生的行业内与行业间的本地经济效应推动了企业在特定区域的相对集中; 第三, 在研究方法上, 现有研究往往事先将企业分配于特定的行政区域内(地级市、县级区等), 这将空间上的点(企业)转化成了平面上不均匀的黑盒子(行政区), 破坏了集聚经济的原始状态, 也造成了可塑性面积单元问题(Openshaw, 1984)。^①

尽管我们并不质疑行业内以及行业间的经济效应对企业集聚的影响, 但本文的研究超越了行业集聚。我们的研究对象是出口企业, 考察的是出口至相同海外市场的企业是否在中国本土存在集中倾向(后文简称为出口目的地本土集聚)。^② 如果存在, 如何解释这一经济现象。因此,

* 张一力, 温州大学商学院, 邮政编码: 325019, 电子信箱: yili66@163.com; 周康(通讯作者), 香港中文大学经济学系博士研究生, 电子信箱: zhoukanghk@gmail.com; 张俊森, 香港中文大学经济学系, 电子信箱: jszhang@cuhk.edu.hk。作者感谢两位匿名审稿人以及彭国华、李佩源的宝贵意见, 特别感谢陈潜的诸多技术性帮助; 张一力感谢国家自然科学基金项目“企业家集群网络形成机制及其对区域经济增长影响的研究”(71273193)、浙江省哲学社会科学重点研究基地立项课题“海外温商区域性产业集群网络演化”(15JDWZ01YB)的资助。文责自负。

① 可塑性面积单元问题是指由于人为划分集聚单位具有随意性, 集聚的测量通常会随着集聚单位的大小以及划分方法的不同而改变(Openshaw, 1984)。在相关研究中, 运用非常广泛的EG指数便存在这一问题。

② 类似于行业集聚, 出口目的地的本土集聚指的是拥有相同海外市场的企业相互集中, 这是作者提炼的一个描述性术语。请注意, 这个概念强调的是出口企业本身的集聚, 而非出口企业海外销售市场的集聚。在本文中, 我们混用了出口目的地本土集聚、本土集聚以及出口目的地聚类分布这三个术语。

本文对集聚的考察关注了中国出口企业空间分布的三个最基本问题: 集聚模式、集聚强度以及集聚形成的原因。选择这样一个研究视角有三层原因。首先, 尽管中国的出口企业数量众多且在国民经济中扮演了极其重要的角色, 然而现有文献很少关注出口企业的空间分布。第二, 与非出口企业相比, 出口企业无疑是特殊的, 他们的成功经营高度依赖于海外市场, 因此海外市场层面的因素, 如出口障碍、汇率变动等对出口企业至关重要。第三, 尽管出口企业具有特殊性且日益增长的文献考察了企业集聚对企业出口行为的影响(包群等, 2012; 佟家栋和刘竹青, 2014), 然而, 现有文献很少能够超越行业集聚研究出口企业的集聚程度并基于出口企业的特殊性分析出口企业集聚发生的原因。

为此, 本文超越行业集聚研究了中国出口企业的目的地本土集聚, 即出口到相同海外市场的企业在中国境内是否呈现出显著的集中倾向。为了考察这一新的集聚模式, 我们借鉴了 Duranton & Overman(2005) (之后简称 DO) 的数值模拟方法。采用这一研究方法有两个原因: 首先, DO 方法提供了一个很好的方法论, 虽然这一方法最初被用来研究行业集聚问题, 但它的适用范围并不局限于行业集聚研究本身, 将它从行业集聚文献引入国际贸易文献, 可以从研究方法上丰富国内的贸易文献; 第二, 我们最为看重 DO 方法的假设检验优势, 如果考察中国是否存在一种新的集聚类型, 我们应该有一个基准(baseline)作为参照, DO 方法基于模拟的假设检验满足了这一要求。适当的调整 DO 方法用以分析中国出口企业的空间分布, 可以从技术上加深我们对中国出口企业空间分布的理解。该方法的其他优势还包括: 不需要人为地将企业分配至特定行政区域内考察集聚。因此, 本文对出口目的地本土集聚的考察是在连续空间上进行的。^① 此外, DO 方法对经纬度坐标数据的精度要求非常高, 本文使用的经纬度数据精确至邮政区(见第三部分数据说明), 因此, 本文在研究数据上对现有文献具有一定贡献。

需要强调的是, 虽然本文研究的出口目的地本土集聚有别于产业理论所强调的行业与产业集聚, 但我们并不质疑行业集聚对于出口(或者非出口)企业空间分布的解释力。本文的目的是为中国出口企业的空间分布增加一个新的解释视角, 尽管这一视角是出口企业所特有的。考虑到现有文献大多从行业与产业的角度解释企业的空间分布, 一个新的解释视角(出口目的地)对于理解中国出口企业的空间布局无疑是有帮助的。此外, 虽然本文的研究视角有别于行业(产业)集聚, 但行业集聚很可能会干扰本文对出口目的地本土集聚的检验。为此, 在检验出口目的地本土集聚时, 我们借鉴了 Ramos & Moral-Benito(2013)的研究方法控制了行业集聚的干扰。

基于 DO 方法的数值模拟显示中国的出口企业在空间分布上存在一种新的集聚模式——出口目的地的本土集聚。随后, 我们从克服出口障碍的视角解释了这一现象, 即出口到相同海外市场的企业很可能通过空间集中, 加快彼此间信息的传递从而产生溢出效应, 克服出口障碍。我们的解释得到了实证证据的有力支持, 我们发现当出口目的地制度环境更差时, 出口至该市场的企业在中国国内的集聚程度更高, 而当出口目的地与中国具有相同的语言与意识形态等有助于缓解出口障碍的共同特征时, 出口至该市场的企业的集聚程度就显著下降。

本文可能的边际贡献体现在三个方面。首先, 丰富的经济地理文献从行业与产业的视角解释了企业的集中现象, 然而, 鲜有文献专门研究出口企业的空间分布, 本文的研究从克服出口障碍的视角为中国出口企业的空间分布提供了一个新的解释, 这加深了我们对中国出口企业空间分布的理解。第二, 本文在数据与研究方法上对现有文献存在贡献。基于全国代表性的经纬度坐标数据, 我们在连续空间上检验了中国的出口企业是否存在出口目的地本土集聚。检验方法来自于行业集聚文献中近来发展起来的 DO 数值模拟法, 从而将行业集聚的研究方法引入国际贸易文献, 增加了

^① Duranton & Overman(2005) 详细论述了 DO 方法在假设检验方面的优势。

贸易文献研究方法的多样性。^① 第三,在出口市场上,如何获取出口目的地的出口相关信息一直是出口企业面临的重要出口障碍之一,然而,鲜有文献为出口障碍影响企业行为(尤其是空间分布)提供经验证据。本文发现的出口目的地本土集聚实际上为出口障碍影响企业行为提供了指示性的间接证据,并为集聚通过信息共享这一机制产生溢出效应提供了指示性的间接证据,进而揭示了出口企业挑战高难度出口市场的一种生存法则。

本文的结构安排如下:第二部分是文献综述;第三部分介绍了本文使用的主要数据以及出口目的地本土集聚的研究方法;第四部分对出口目的地本土集聚的原因进行了实证研究并予以解释;第五部分是稳健性讨论;最后是结论性评述。

二、文献综述

与本文相关的第一支文献检验并解释了企业的地理集聚。Ellison & Glaeser(1997)基于自然优势、行业溢出与随机选择效应较早地研究并解释了集聚现象,他们发现在美国基本上所有制造业行业都存在本地化倾向。使用类似的研究方法,Devereux et al.(2004)发现英国本地劳动力市场与行业间的垂直分工是集聚发生的主要推动力。Duranton & Overman(2005)改进了先前的研究方法,基于详实的企业位置信息,他们突破了地域限制,从而在连续空间上而非将企业分配至特定的行政区域内研究了行业集聚,他们发现在英国超过一半的行业存在本地化集聚。Ellison et al.(2010)的研究同样具有代表性,基于对美国制造业企业集聚动力的追踪,以及通过对 Marshall 外部性的三个理论来源的系统性研究,他们发现供给—需求链是集聚产生的主要推动力。这些文献大多基于行业与产业的视角,由于研究对象并不局限于出口企业,他们对于企业集聚的解释并不强调出口目的地的重要性。

近来,融合集聚经济与企业出口行为的研究日益增长。Cassey & Schmeiser(2013)使用 Moran's I 指数发现俄罗斯出口至相同海外市场的企业存在区域集中现象。他们对集聚的检验基于 89 个俄罗斯联邦区域(类似于中国的省),这种方法无法考察跨地区的集聚。此外,因为联邦区域在人口、面积等方面存在显著差异,这使得测量的集聚结果很难跨地区比较。因为数据的限制,他们的研究也没有考察行业集聚的潜在干扰。此外,Koenig et al.(2010)对法国以及 Choquette & Meinen(2011)对丹麦的研究均发现出口目的地的本土聚类分布存在显著的出口溢出效应。Ramos & Moral-Benito(2013)同样使用 DO 方法研究了西班牙出口企业的空间分布,他们发现出口至相同海外市场的西班牙出口企业在空间上相互集中并考察了这一集聚对企业出口动态与贸易福利的影响。在国内研究方面,佟家栋和刘竹青(2014)与刘竹青和周燕(2014)分别从融资约束与契约执行两个方面发现地理集聚显著提升了企业出口概率,因为研究的集聚类型与本文不同,因此,他们对集聚的研究并不是出口目的地特殊性的。

此外,对于出口目的地本土集聚的解释,有别于现有的诸多研究,我们强调了出口障碍的影响。在出口市场上,如何获取出口目的地的出口相关信息一直是出口企业面临的主要挑战之一。然而,鲜有文献为出口障碍影响企业行为(尤其是空间分布)提供经验证据。Gould(1994)以及 Rauch & Trindade(2002)的间接证据显示移民网络通过缓解出口障碍促进了双边贸易,潜在的机制很可能是移民网络降低了出口信息的获取成本。与之类似,双边信任同样可能通过降低出口信息障碍进而促进双边贸易。Guiso et al.(2009)研究了信任对欧洲跨国贸易的促进作用,Xing & Zhou(2018)发现信任关系对于中国国内的跨省贸易也存在显著影响。本文发现的出口目的地本土集聚意味

^① 仅就我们掌握的文献,袁海红等(2014)是唯一使用 Duranton & Overman(2005)研究方法考察中国集聚的研究,他们研究了北京市的行业集聚,我们研究的则是具有全国代表性的出口企业目的地本土集聚。

着:出口至相同海外市场的企业很可能通过相互集中加强了彼此间目的地特殊性的信息交流与分享,进而缓解了出口企业面临的出口障碍。这实际上为出口障碍影响企业行为提供了指示性的间接证据,并为集聚通过信息共享这一机制产生溢出效应提供了指示性的间接证据。

三、出口目的地本土集聚的研究方法

(一) 数据说明

为了研究中国企业的出口目的地本土集聚,在这一部分,我们使用了三套微观数据。分别为企业层面的中国海关数据、规模以上的中国工业企业数据以及中国邮政区域经纬度坐标数据。其中基于专业的GIS工具,专业的技术人员为我们整理提供了中国邮政区域标准的WGS84(World Geodetic System 1984)经纬度信息数据,这套数据的最大特点是经纬度坐标十分详细。我们以安徽省安庆市怀宁县为例,2006年,高河镇的邮政编码为246121,与之对应,我们获取了高河镇精确至小数后6位的经纬度坐标。

出口目的国信息来自海关进出口统计数据库,它是中国海关总署产品层面交易的月度数据,我们在年度层面进行了汇总。此外,为了控制行业结构,我们也使用了中国工业企业数据。这套数据来自中国国家统计局的制造业企业年度调查,它收录了2002—2006年每年约230000家制造业企业的主要生产信息,覆盖了中国全部国有企业和年销售额500万元以上的非国有制造业企业。其中包含了我们最为关心的企业所属行业以及企业所在地的邮政编码。我们以年份与企业名称为唯一识别符将这套数据与海关数据匹配,并将手工整理的邮政区域经纬度坐标赋值给相应年份的相同邮政区域,以开展这一部分的研究工作。

(二) 出口目的地本土集聚的检验

为捕捉出口企业的目的地本土集聚,本文借鉴Duranton & Overman(2005)的数值模拟方法。与先前的研究方法相比这一方法具有两点突出的优势,首先,它克服了行政区域边界问题,即在连续空间上而非将企业分配至特定的行政区域内考察集聚,从而避免了可塑性面积单元问题(Openshaw, 1984)。更为重要的是,虽然DO方法最初被用来研究行业集聚,但它的适用范围并不局限于行业集聚研究本身。如果检验中国是否存在出口目的地本土集聚,我们应该有一个基准作为参照,DO方法通过将真实的企业空间分布与蒙特卡罗(Monte Carlo)模拟结果进行对比,可以定量地考察企业真实的空间分布对随机分布的偏离,从而在统计意义上测量特定置信度下的集聚程度。

与Duranton & Overman(2005)不同的是,本文研究的是出口目的地本土集聚而非行业集聚。基于比较优势理论,进口国不需要与本国资源禀赋相同的产品,这意味着简单地考察出口目的地本土集聚可能会受到行业集聚的干扰。为此,我们借鉴了Ramos & Moral-Benito(2013)的研究方法,控制了出口企业的行业结构。在下文的分析中,我们将具体讨论行业结构的控制方法。

本文研究的时间跨度为2002—2006年。^①下面我们以2006年为例对研究方法进行说明。鉴于只考虑出口企业数量大于30家的海外市场,我们剔除了9个目的国,对剩余的161个出口市场进行研究。首先,我们将工业企业数据与海关数据根据企业的中文名称进行了匹配,从而获得出口企业的邮政编码与行业种类;随后,我们将邮政编码的经纬度坐标赋值给邮政编码覆盖区的所有出口企业。由此,对于任意出口企业,我们获得了相对准确的企业经纬度坐标信息、企业所属的行业种类以及出口目的地信息。对出口目的地本土集聚的检验需要分国家单独进行,我们这里仅以2006年的哥斯达黎加为例说明检验方法。我们的思路是先对DO方法的检验逻辑进行非技术介绍

^① 在第五部分的稳健性讨论中,我们拓展了数据年限。

(分四步) 然后再单独讨论技术细节。

第一步, 定位出口到哥斯达黎加的所有企业, 一共有 650 家。为方便后文的解释, 我们假设这 650 家出口企业分布在 L 个邮政区域内, 且全部来自于 A、B 两个行业, 其中 A 行业 M 家, B 行业 N 家, 则 $M + N = 650$; 进一步假设 $M : N = 8 : 2$ 。

第二步, 有条件地为企业随机选址, 并在选址时控制行业集聚。首先, 所谓“有条件”指的是每一次随机选址时, 将这 650 家企业随机分配在这 L 个邮政区内, 而不是全国的任意邮政区。这些地区已经出现了出口到哥斯达黎加的企业, 所以不可能是无法产生出口企业的地区, 换言之在为随机选址时, 企业只会被分配到那些现实中存在向哥斯达黎加出口可能的邮政区。其次, 控制行业集聚, 这是指在每一次随机模拟(选址)时, 我们都确保 A、B 行业的企业比重为 8:2, 进而控制住行业集聚。如果 A 行业的出口企业大量集中在珠三角, 那么随机模拟的出口目的地本土集聚就会因为行业集聚的干扰变的很强。然而, 我们对目的地本土集聚的非技术定义是真实的企业分布比多次模拟的结果更加集中。因此, 当模拟的出口目的地本土集聚越强时, 通过假设检验进而认为该市场存在本土集聚的难度就越高。这类似于对行业集聚推高目的地本土集聚施加了一个惩罚。换言之, 如果行业集聚造成了虚假的出口目的地本土集聚, 我们控制行业集聚的方法便拉高了模拟的集聚水平, 进而水涨船高, 提高了出口目的地本土集聚的判断门槛, 从而控制行业集聚的干扰。关于这一方法更详细的技术性讨论, 可以参考 Ramos & Moral-Benito (2013)。

第三步, 将企业的空间分布转化成二维坐标系中关于企业之间距离的密度图。对于出口到哥斯达黎加的所有企业, 计算任意两家企业之间的距离, 并将距离的密度图绘制在二维坐标系中(如图 1 右)。横轴是任意两家企业之间的距离, 纵轴代表密度。进而将出口到哥斯达黎加的所有企业的空间分布转化成二维平面上依赖于任意两家企业之间距离的密度分布图。如果将任意两家企业之间的距离称为距离对, 那么密度线上(500, 0.00064)的坐标点表示在出口到哥斯达黎加的所有企业中, 彼此之间相距 500 千米的距离对占(任意两家企业之间)所有距离对的 0.064%。

第四步, 多次模拟, 构建置信区间。重复步骤二的模拟(随机选址)1000 次, 从而画出 1000 张步骤三描述的企业密度分布图进而构建置信区间并与真实的密度分布图进行比较。对于任意的两两企业间距离, 如果真实的企业分布密度高于 950 次模拟的密度(一共模拟 1000 次, 本文的置信度为 95%), 我们就认为在这个距离上, 出口到哥斯达黎加的企业存在超越行业集聚的出口目的地本土集聚, 反之则认为不存在出口目的地本土集聚。^① 因此检验出口目的地的本土集聚高度依赖于出口至该市场的企业的真实空间分布以及模拟的空间分布。

现在, 我们讨论一些技术细节:

首先, 在步骤二中, 为何要有条件地为企业随机选址, 并控制行业集聚? 检验出口目的地本土集聚面临两个技术难题。第一个技术难题是: 在随机模拟企业分布时, 无条件的随机选址并非是检验企业集聚的合理基准。因为企业一方面会有意避开山川湖泊, 另一方面也更青睐于人口更加集聚、基础设施更加优良的城市。这意味着企业的分布总会呈现出一定的集聚倾向。Ellison & Glaeser(1997) 的镖盘隐喻颇为生动: 只要企业的选址被完全随机, 企业真实的空间分布总会偏离随机分布。这就好比向镖盘投掷飞镖, 因为镖盘上每个位置被飞镖击中的概率并不是相等的, 飞镖(企业)不可能落在镖盘上的“湖泊”区域, 但更可能落在镖盘上的“人口集中区”。因此, 哥斯达黎加案例中的镖盘并非全国所有地区, 而是已经出现了以哥斯达黎加为出口目的地的 L 个邮政区,

^① 请注意, 这里只是在介绍出口目的地本土集聚的检验逻辑。事实上, 这里描述的集聚是在单个距离上的, 我们还需要将这种点检验推广为全局检验, 即寻找到一条线使得在全部距离上, 只有 50 个模拟的企业分布密度点高于这条线。下文在讨论如何确定包迹线时会详细阐述这一技术问题。

因为这些地区已经产生了向哥斯达黎加出口的企业,那么这些区位显然不会是无法产生向哥斯达黎加出口的企业。另一个技术难题在于,由于特定国家的进口需求往往来自某些特定行业。那么特定出口市场的集聚很可能只是行业集聚的另一种表现形式,这将干扰我们对出口目的地本土集聚的检验。因此,我们在检验出口目的地本土集聚时需要控制行业集聚的干扰。换言之,本文发现的出口目的地本土集聚是在控制了行业集聚干扰之后依然存在的集聚模式。

其次,如何绘制步骤三中的密度分布图?在绘制密度分布图之前,我们需要计算出口到该市场所有企业的双边距离(即任意两家企业之间的距离)。在计算企业双边距离时,由于英国的地域范围有限,Duranton & Overman(2005)使用了欧氏距离(Euclidean distance),考虑到中国地域辽阔,我们使用了半正矢距离公式(haversine formula)计算了企业间的双边距离。^①随后使用高斯核密度估计获得出口到特定海外市场的企业双边距离密度图。具体为,对于国家 c 在任意距离 d 上的核密度估计函数 $\hat{K}_c(d)$ 其表达形式如下:

$$\hat{K}_c(d) = \frac{2}{n_c(n_c - 1)h} \sum_{i=1}^{n_c-1} \sum_{j=i+1}^{n_c} f\left(\frac{d - d_{ij}}{h}\right)$$

其中 n_c 是出口到 c 国的企业数量, h 为通过最小化均方积分误差(MISE)获得的最优带宽(Silverman, 1986), f 是高斯核函数。需要说明的是,如果出口到 c 国的企业数量为 N ,则需要计算的企业双边距离数量将呈指数增长,这是一项富有挑战性的计算工作。为此,我们使用了 Ellison et al. (2010) 的简化方法,如果某一出口市场的企业数量超过 2500 家,则从该市场的所有出口企业中随机抽取 2500 家。^②

图 1 分别展示了目的地为也门共和国与哥斯达黎加的企业在中国本土的两两之间的距离的分布图。他们的出口企业数量非常接近,分别为 465 家与 650 家。注意到也门的密度分布更加左偏,说明在短距离上累积的密度远大于哥斯达黎加,这意味着出口到也门的企业在中国境内更加集中。事实上,在出口到也门的所有企业两两之间的距离对中,61% 的企业彼此相距不超过 1000 千米,而对于出口到哥斯达黎加的企业,这一比例为 55%。

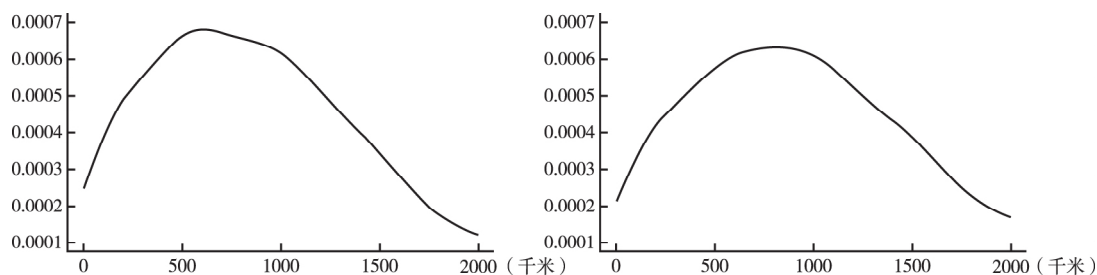


图 1 也门共和国(左)与哥斯达黎加(右)出口企业的本土两两企业距离分布核密度

出口到也门的企业距离对在 600 千米处出现了波峰,这意味着大量出口企业彼此相邻的距离为 600 千米;而哥斯达黎加的分布图在 900 千米处才出现波峰,这与哥斯达黎加出口企业两两距离对分布的中位数(约 1000 千米)十分接近,随后我们将会发现以哥斯达黎加为出口目的地的企业,其实并不存在本土集聚。

另外,如何构建步骤四中的置信区间,进而开展假设检验?我们使用 95% 的置信水平构建了

① 与欧氏距离(Euclidean distance)不同,半正矢公式(haversine formula)计算的是地球表面两点间的球面距离,它考虑了曲面问题。

② 由于出口企业数量众多,本文对计算力的要求超过了 Duranton & Overman(2005)对英国以及 Ramos & Moral-Benito(2013)对西班牙的研究。虽然尽可能优化运算程序,并设计了并行结构,但 24 个国家在数值模拟时的硬件要求依然超过我们的运算能力。

出口目的地本土集聚与分散的两个检验。下面,我们结合也门与哥斯达黎加的实例进行说明。图2 两幅图中均有两条虚线和一条实线。其中实线为真实的企业距离对的密度线,即图1 的短距离(1000 千米)形式;两条虚线为上下包迹线,由关于距离的两个函数 $global_K_c(d)$ 和 $global_K_c(d)^c$ 确定。这两条包迹线(虚线)可以分别用来检验出口目的地本土集聚与出口目的地本土分散,其中 c 代表出口目的国。

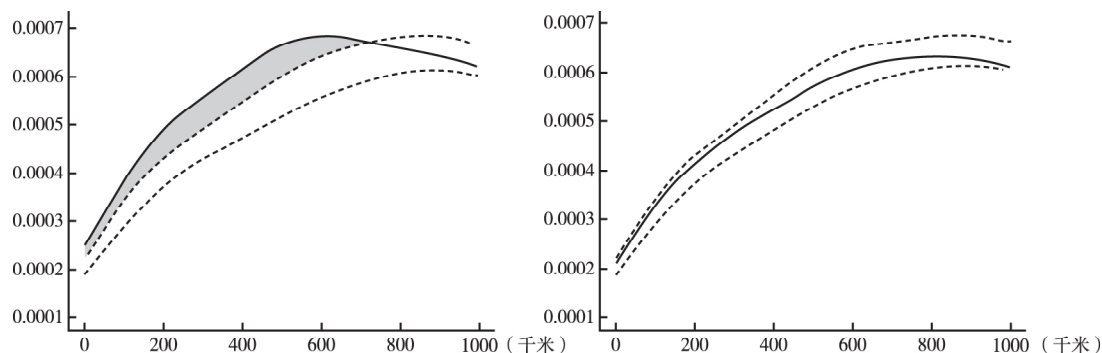


图2 也门共和国(左)与哥斯达黎加(右)出口企业的目的地集聚与分散

为确定上下包迹线,我们先要确定局部集聚和分散阈值 $local_K_c(d)$ 与 $local_K_c(d)^c$ 。其确定方式如下。首先,在任意距离 d 上($d \in [0, 1000]$ 且取整数)将前文步骤二与步骤三中模拟得到的1000个随机分布密度值在各自距离 d 上按升序排列。然后,找出相应距离 d 上各自模拟密度的95%分位点,从而确保在任意距离 d 上,只有不超过50次模拟的密度高于局部集聚阈值 $local_K_c(d)$,则该95%分位点便是距离 d 上的集聚阈值 $local_K_c(d)$ 。因为我们在区间 $[0, 1000]$ 内的任意整数距离 d 上都进行了密度排序并找出了95%分位点,我们连接这些点便可以得到一条曲线,进一步将这条曲线向上水平移动从而保证在整个区间 $[0, 1000]$ 上,一共只有50个密度点位于这条曲线以上,那么这条曲线便是图2(左)以及图2(右)中的上包迹线,其函数表达式便是 $global_K_c(d)$ $d \in [0, 1000]$ 。^①对于出口目的地 c ,在任意距离 d 上,如果出口企业的真实双边距离的密度值(图2左或右中的实线)高于图2左或图2右的上包迹线,我们就认为在距离 d 上,出口市场 c 存在本土集聚。与Duranton & Overman(2005)类似,只要某个出口市场在任意距离 d 上出现集聚,我们就认为该市场存在本土集聚。以也门共和国为例,上包迹线代表了任意距离上1000次随机模拟的98.72%分位点,这些分位点保证了全局距离上95%的模拟值位于该点以下。^②下包迹线的确定方法与上包迹线类似,它确保了 $[0, 1000]$ 全局距离上只有5%的模拟密度位于该函数代表的曲线以下,这条下包迹线便是图2左以及图2右两条虚线中更低的一条。对于出口市场 c ,由于在所有距离上出口企业双边距离的密度的总和 $\hat{K}_c(d) = 1$ 。那么出口企业在一定距离上的集聚则意味着企业在另外距离下的分散。因此,只有当企业的真实分布线位于下包迹线以下并且真实分布线不会超过上包迹线,我们才认为存在出口目的地本土分散。图2中,也门存在明显的本土集聚,所以不可能存在本土分散,集聚范围从0一直持续到约700千米。哥斯达黎加不存在集聚,但因为实线一直位于下包迹线上方所以也不存在本土分散。

对于集聚程度我们以图2中阴影部分表示。当存在集聚时,其含义为95%的置信度水平上,真实分布线对上包迹线的总偏离。不存在集聚便以0表示。根据Duranton & Overman(2005),其

① 由于无法避免的离散性距离问题,在全局距离上,并非所有的国家都能找到集聚(分散)阈值使得95%(5%)的模拟值位于该阈值之下。当这种情况发生时,我们使用线性插值法寻找最合适的阈值点。

② 在每一个距离 d 上的分位数更大,是因为我们在连接各个距离上的95%分位点之后,又向上移动直线以获得上包迹线,从而将局部95%的置信度上升为全局95%置信度。

函数表达式为:

$$\Gamma_c \equiv \sum_{d=0}^{d=1000} \max\{\widehat{K}_c(d) - \overline{K}_c(d) \mid \rho\}$$

使用以上方法,我们检验了 2002—2006 年出口企业超过 30 家的所有出口目的地的本土集聚。其中,在 2006 年 161 个主要出口目的国中,109 个国家出现了出口目的地本土集聚(200 千米以内的集聚),布基纳法索、卢森堡与玻利维亚 3 个国家存在出口目的地本土分散。至此,我们证明了仅就中国而言,大多数目的地在短距离内存在本土集聚,并且这种集聚并非来自行业集聚。

表 1 报告了出口目的地本土集聚排名前 10 位的海外市场。一个有趣的发现是这 10 个国家全部来自经济文化发展相对落后的非洲(4 个)、亚洲(4 个)与南美洲(2 个),这些国家的市场规模相对较小,制度环境相对较差,甚至正在遭受(2006 年)社会动荡或者战后重建。我们该如何理解出口国的制度环境特征与本土集聚的关系?这一问题我们将在下一部分集中讨论。

表 1 2006 年出口目的地本土集聚排名前 10 位的海外市场

排序	出口目的地	集聚指数	企业数量	排序	出口目的地	集聚指数	企业数量
1	博茨瓦那	0.003178	38	6	荷属安地列斯群岛	0.003300	44
2	文莱	0.003219	116	7	利比里亚	0.002524	66
3	吉尔吉斯共和国	0.003923	91	8	加那利群岛	0.003058	31
4	利比亚	0.003890	467	9	苏里南	0.002480	82
5	也门共和国	0.002736	471	10	伊拉克	0.002630	230

注:这里报告的是 1000 千米内的集聚指数,企业数量是将海关企业、工业企业与邮政编码经纬度坐标匹配后的样本数。

四、出口目的地本土集聚的解释

前文的研究证实了中国存在出口目的地本土集聚,这一部分我们对典型事实进行简单提炼,获得直觉推论,引导实证研究,进而对出口目的地本土集聚进行解释。首先,我们来对比也门共和国与哥斯达黎加这两个市场的目的地本土集聚。虽然在数量上,出口到这两个市场的企业数非常接近,分别为 450 家与 650 家,但就集聚程度而言,出口到也门的企业显然呈现了更强的出口目的地本土集聚,前者的集聚 Γ 值(500 千米)约为 0.027,而后者并不存在本土集聚。一个可能的解释是向这两个市场出口的企业面临不同的出口障碍。根据世界银行的“全球治理指标”,也门在全球 216 个国家(地区)的制度环境排名中,位于第 194 位,而哥斯达黎加的排名则处于 84 位。^①通过这个典型事例,我们做出如下推论。

推论:出口目的地制度环境越差,出口至该市场企业面临的出口障碍越强,为了克服出口障碍,出口企业通过相互集中加强了彼此间的信息交流与分享,最终形成了以出口目的地为类别的本土集聚。

进一步,我们对比了卢森堡与古巴的出口企业分布。出口到这两个国家的企业数量同样非常接近,分别为 72 家与 120 家。在制度环境的全球排名上,卢森堡高达 15 位,而古巴则排在了 156 位。与我们的预期一致,出口到卢森堡的企业在空间分布上并不集聚;但令我们意外的是,出口到古巴的企业同样没有表现出显著的集聚趋势。我们认为古巴与中国共同的意识形态可能在一定程

^① 这里使用的指标为法制环境(rule of law),后文会详细介绍。

度上降低了出口障碍,进而弥补了不良的制度环境。由此我们推测双边贸易国共同的社会意识形态可以在一定程度上弥补出口市场不良的制度环境,缓解出口企业面临的出口障碍,进而降低出口企业的目的地本土集聚。包括中国在内,普遍承认的社会主义国家共有 5 个;此外,共同的语言同样可能降低出口障碍进而降低本土集聚。包括中国大陆在内,汉语使用比例超过 9% 的国家(地区)共有 6 个。虽然样本有限,但基于国家体制与语言环境等的推测却建立在本文的核心推论基础上,可以作为间接证据检验本文核心推论的可靠性。为此,我们在下文的回归分析中同时加入了国家体制与语言的虚拟变量。

为了检验出口目的国的制度环境对本土集聚的影响,下面我们构建计量模型进行实证分析。由于影响集聚的因素错综复杂,我们在回归分析中逐步控制了影响集聚的国别因素、企业所在地的区位因素以及企业异质性。我们的实证研究包括三个部分。首先,出口目的地本土集聚是一种宏观经济现象,为此基于双边引力模型 Frankel & Romer(1999),我们构建以下计量方程检验本文的基本推论:

$$\Gamma_{ct} = \alpha + \beta_1 Quality_{ct} + \beta_2 Ideology_{ct} + \beta_3 Language_{ct} + Coucontrol_{ct} + \varepsilon_{ct} \quad (1)$$

其中,因变量是前文计算的集聚指标 Γ ,我们对它进行了标准化处理。 t 表示年份, $Quality_{ct}$ 是本文最关心的制度环境变量, $Ideology_{ct}$ 代表国家体制虚拟变量,我们对社会主义国家取 1,反之取 0。 $Language_{ct}$ 代表语言虚拟变量,当汉语在 c 国的使用比例超过 9% 时取 1,否则为 0。因为国家层面的意识形态、语言、距离等变量不随时间变化,我们没有在回归模型中加入国家层面的固定效应。 $Coucontrol$ 是国家层面的 5 个控制变量,包括: 1) 是否与中国相邻,当目的国是中国的邻国时,该变量取 1,否则为 0; 2) 北京到出口目的地的距离; 3) 目的国的人口规模; 4) 2002 年不变价的人均 GDP; 5) 向 c 国出口的企业数量。 ε_{ct} 为随机干扰项。

下面我们介绍这一部分使用的数据。除了前文计算的集聚指数,其余数据有 4 个来源。国家控制变量中, GDP 与人口数据来自世界银行; 出口到特定国家的企业数量由作者根据相应年份的中国海关进出口数据统计而得; 语言、邻国与国家间的距离变量来自 CEPII。最后,我们最关注的核心变量制度环境来自 Kaufmann et al. (2009) 的“全球治理指标”(worldwide governance indicators),这套系列数据由世界银行(World Bank)与布鲁金斯研究所(Brookings Institution)合作编写。Kaufmann 等自 1996 年便持续构建世界各国(地区)的政治治理指标以反映世界各国(地区)的治理水平。他们提供了有关政府治理的 6 个指标,包括政府责任(voice and accountability)、政治稳定/无暴力冲突(political stability/no violence)、政府效率(government effectiveness)、监管质量(regulatory quality)、法治水平(rule of law)以及反腐(control of corruption)。这套数据在国外相关研究中运用广泛,同时也是世界银行“治理是重要的”这一论断的判断依据,得分越高说明该国的制度环境质量越高;反之则说明该国的制度环境越差。这 6 个指标的相关性极高,在表 2 上表的回归结果中,我们以法治水平(rule of law)作为制度环境的代理变量,其他 5 个指标的回归结果是类似的(见表 2 下半部分)。

如前所述,本文研究 200 千米内的出口目的地本土集聚,这使本文的被解释变量被右截取。在处理这一受限被解释变量时我们使用 Tobit 回归模型。表 2 第 1 列汇报了(1)式的估计结果(没有控制随时间变化的国别因素:目的国人口、人均 GDP 以及出口企业数)。与本文的推论一致,相比于制度环境更好的目的国,当目的国的制度环境更差时,出口至该市场的企业在中国本土的集聚程度更高,而相同的意识形态则显著降低了本土集聚。在第 2 列中我们加入了国家控制变量,我们发现,制度环境与意识形态对本土集聚的影响依然显著且系数的符号保持不变,共同的国家语言也同样显著降低了该市场的本土集聚,国家相邻性对本土集聚的影响并不显著,而国家间的距离与本土集聚显著负相关,这也是符合直觉的,因为获取更远目的地的出口信息的难度更大,因此向该市场出口的企业更依赖于通过本土集聚分享出口信息进而克服出口障碍。

表 2 制度环境对出口目的地本土集聚的回归结果: 国家层面证据

因变量: 出口目的地本土集聚	(1)	(2)	(3)	(4)	(5)
法治水平	-0.077 ^{***} (-2.63)	-0.155 ^{**} (-2.30)	-0.166 ^{**} (-2.33)	-0.143 ^{**} (-2.19)	-0.066 [*] (-1.66)
国家体制	-0.724 ^{***} (-3.00)	-0.530 ^{***} (-2.72)	-0.339 ^{***} (-2.92)	-0.538 ^{***} (-2.84)	-0.295 ^{**} (-2.40)
语言	-0.317 [*] (-1.75)	-0.493 ^{**} (-2.19)	-0.495 ^{***} (-4.56)	-0.397 [*] (-1.73)	-0.209 [*] (-1.71)
是否邻国	0.121 (0.36)	0.206 (0.55)		0.135 (0.37)	-0.156 [*] (-1.65)
国家距离	-0.207 ^{**} (-2.14)	-0.321 ^{***} (-3.07)	-0.339 ^{***} (-3.27)	-0.374 ^{***} (-3.04)	-0.172 ^{***} (-3.52)
随时间变化的国别因素	否	是	是	是	是
年份固定效应	是	是	是	是	是
R ²	0.010	0.027	0.034	0.029	0.125
观测值	690	659	612	659	659

使用其他 5 个制度环境指标的回归结果

测量制度质量的变量	政府责任	反腐	政治稳定	政府效率	监管质量
制度质量	-0.175 ^{***} (-3.39)	-0.135 ^{**} (-2.43)	-0.153 ^{***} (-3.03)	-0.133 [*] (-1.96)	-0.204 ^{**} (-2.44)
国家体制	-0.719 ^{***} (-3.33)	-0.444 ^{**} (-2.16)	-0.336 [*] (-1.82)	-0.496 ^{**} (-2.48)	-0.642 ^{***} (-3.33)
语言	-0.618 ^{***} (-2.62)	-0.445 [*] (-1.95)	-0.464 ^{**} (-2.08)	-0.445 [*] (-1.92)	-0.454 ^{**} (-1.98)
国家控制变量	是	是	是	是	是
年份固定效应	是	是	是	是	是
R ²	0.028	0.025	0.026	0.025	0.028
观测值	659	659	659	659	659

注: 括号中为基于怀特稳健标准误的 t 统计量或者 z 统计量; ***、**、* 分别表示在 1%、5%、10% 的水平上显著。probit 回归中, 基于原始(非标准化)的被解释变量获得二值因变量, 且估计结果为解释变量的边际效应。没有汇报常数项。

在第 3 列的回归结果中, 我们剔除了相邻国家。邻国的出口障碍往往更低, 我们发现制度环境系数的绝对值出现了轻微的提升, 这符合预期。在第 4 列的回归结果中, 我们考察了更短距离的集聚(100 千米), 此时制度环境系数的绝对值出现了轻微的下降。一个可能的解释是, 在更小的地域空间内, 出口到相同市场的企业间联系更加紧密, 缓解了出口障碍, 进而降低了制度环境对集聚的影响。

然而, 本文计算的集聚指数 I 不但是典型的截尾数据, 而且存在大量零值, 这造成了一个潜在的担忧, 因为不管是零膨胀泊松回归还是零膨胀负二项回归都无法同时解决截尾与零值问题。一个可供选择的处理办法是, 将所有在 200 千米内集聚的出口市场定义为本土集聚并赋值为 1, 反之则为 0, 再使用 probit 模型进行估计, 从而将右截尾且存在大量零值的连续数据转化成二值离散数据。

(被解释变量只包括 0 和 1)。这样操作的一个合理性在于: 我们发现对于出口到任意海外市场的企业, 如果在 200 千米范围内不存在本土集聚, 那么在更远的距离上也基本不会出现集聚。因此, 我们选取 200 千米为分界线, 将 200 千米以内集聚的目的地简单认定为存在本土集聚; 而将 200 千米依然不集聚的目的地认定为不存在本土集聚。结果汇报于表 2 第 5 列, 可以发现回归结果同样支持了本文推论: 出口目的地的制度环境越差, 出口至该市场的企业在中国境内集聚的概率越高, 而相同的意识形态和语言则显著降低了出口目的地本土集聚的发生概率。

Kaufmann et al. (2009) 提供了衡量制度环境的 6 个指标, 在表 2 的下半部分, 我们汇报了使用其他 5 个指标的回归结果, 可以发现, 核心变量的显著性与符号与前文的回归结果(使用法治水平作为制度的代理变量)是类似的。至此, 本文的推论得到了实证结果的初步支持。

下面我们简要讨论第一部分实证研究的潜在偏误及处理办法。在实证研究的第一部分, 有两个潜在问题值得注意。首先, 因为实证研究在国家层面展开, 我们无法控制企业所在地的基础设施以及当地劳动力市场的供给情况, 而这些被遗漏的因素可能密切关系到企业的区位选择, 因而影响了本文计算的本土集聚(我们的被解释变量)。而在企业选址之后, 当地的劳动力市场状态与基础设施条件都可能影响企业选择进入哪些出口市场(进而影响本文最关心的解释变量, 制度环境)。因此, 与当地劳动力市场相关的遗漏变量很可能造成本文最关心的回归系数(β_1)是有偏的。另一个潜在的担忧来自于企业层面的遗漏变量。一方面, 基于异质性贸易理论, 企业基于其技术效率(TFP)自选择进入出口市场;^①另一方面, 基于经济地理文献, 企业的技术水平(TFP)同样影响了企业的区位选择, 效率最高的企业最有可能选择区域经济中心。如果技术效率(TFP)最高的企业进入了制度最差的出口市场, 这类企业同时又选择了基础设施条件最优越的区位进而形成集聚, 这种自选择会造成第二部分回归结果(β_1)的高估, 这种高估来自于遗漏变量——技术效率(TFP)。

为了处理地区以及企业层面主要遗漏变量可能造成的估计偏差, 我们引入了企业维度, 将实证分析由宏观逐步推进到微观, 进而控制出口企业所在地区的区位因素以及企业异质性。首先将前文计算的出口目的地本土集聚指数赋值给当年出口到该国的所有出口企业, 并将其他国家层面的宏观变量也同样赋值给出口到该市场的所有企业。^②这样, 我们便获得了微观企业样本。在使用微观样本进行回归分析时, 便可以控制地区层面和企业层面的主要遗漏变量。加入企业维度的回归方程建立如下:

$$\Gamma_{cft} = \alpha + \beta_1 Quality_{cft} + \beta_2 Ideology_{cft} + \beta_3 Language_{cft} + Coucontrol_{cft} + Admcontrol_{cft} + Fircontrol_{cft} + \varepsilon_{cft} \quad (2)$$

其中 f 代表企业, 其他的变量与方程(1)一致。例如 $Quality_{cft}$ 指 t 年出口到 c 国的企业 f 所对应的出口目的地制度环境。 $Admcontrol$ 是地区层面的两个控制变量, 包括企业所在地级市的年从业人员总数(单位从业人员数、私营和个体从业人员数之和)和基础设施状况(以年铁路、公路、水运以及航空运输的总运量衡量), $Fircontrol$ 是企业层面的三个控制变量: 使用嵌入生存概率的 Olley & Pakes(1996) 两步法计算的全要素生产率、政府补贴与融资约束。其中, 政府补贴以及地区层面的控制变量均取自然对数。企业层面的控制变量来自 2002—2006 年中国工业企业数据库; 地区控制变量来自相应年份的《中国城市统计年鉴》。

在表 3 第 1 列中, 我们首先加入了地区控制变量, 我们发现不良的制度环境依然对出口目的地

① 因为这些企业最具有成本优势挑战这些高难度的出口市场。

② 当我们在企业层面进行回归分析时, 虽然向同一个目的国出口的企业具有相同的集聚指数(被解释变量)以及其他国家层面的变量, 但这些企业与企业层面的变量(全要素生产率、政府补贴与融资约束)并不相同, 也正因此我们控制住了企业层面的遗漏变量, 且不会出现大量回归样本所有变量都是相同的情况。

本土集聚产生了显著的促进作用,而共同的意识形态以及共同的语言均显著降低了出口目的地本土集聚。这再次印证了前文的发现:出口障碍显著影响了中国出口企业的空间分布。第 2 列中,我们进一步加入了企业控制变量,发现制度环境、意识形态与语言对集聚的影响程度出现了轻微的下降。在第 3 列中,同样剔除了相邻国家,发现核心指标保持稳健。第 4 列中,再次考察更短距离内的本土集聚,与宏观研究类似,制度环境与语言文化对本土集聚的影响似乎出现了轻微下降,但是意识形态的影响似乎增强了。

在第 5 列中,我们处理了另一个潜在的担忧——出口企业是否会因为围绕出口港口的集聚而呈现出口目的地本土集聚的假象。我们参考了黄玖立和徐旻鸿(2012)选择了 39 个 2002 年吞吐量在 100 万吨以上的港口城市作为核心港口,并整理了这 39 个核心港口的经纬度坐标,随后调用了前文使用的邮政编码经纬度坐标,计算了出口企业所在的邮政区域到最近港口城市的半正矢球面距离,作为控制变量(取对数),以隔离港口集聚可能的干扰。我们发现,出口企业与最近港口的距离对出口目的地本土集聚的影响并不显著,而核心变量的显著性依然良好。

表 3 制度环境对出口目的地本土集聚的回归结果:地区与企业层面证据

因变量: 出口目的地本土集聚	(1)	(2)	(3)	(4)	(5)
	地级市	地级市 + 企业	剔除邻国	100KM 内集聚	控制港口
法治水平	-0.221 *** (-109.07)	-0.218 *** (-62.68)	-0.248 *** (-58.92)	-0.159 *** (-49.01)	-0.218 *** (-62.65)
国家体制	-0.602 *** (-145.68)	-0.598 *** (-84.88)	-0.178 *** (-10.25)	-0.722 *** (-108.94)	-0.597 *** (-84.62)
语言	-1.175 *** (-293.03)	-1.099 *** (-157.44)	-1.092 *** (-167.04)	-1.022 *** (-148.31)	-1.099 *** (-157.41)
到港口距离					0.001 (1.27)
国家控制变量	是	是	是	是	是
地级市控制变量	是	是	是	是	是
企业控制变量	否	是	是	是	是
年份虚拟变量	是	是	是	是	是
R ²	0.157	0.152	0.155	0.163	0.152
观测值	741139	246631	221839	246631	246353

注: 括号中为基于怀特稳健标准误的 t 统计量; ***, **, * 分别表示在 1%、5%、10% 的水平上显著; 同时控制了行业与企业所有制类型虚拟变量; 没有汇报常数项。

五、数据拓展与稳健性讨论

如前所述,本文对出口目的地本土集聚的检验高度依赖于邮政区域层面的经纬度坐标数据以及 DO 研究方法。为了考察本文核心结论的稳健性,在这一部分,我们基于行政区域定义了新的出口目的地本土集聚指数进而开展稳健性分析。

首先,我们定义如下式的出口目的地本土集聚指数:

$$\Gamma_{ct} = \frac{\max\{Firm_{cpt}\}}{\sum_p Firm_{cpt}}$$

其中 c 和 t 分别代表出口目的地与年份, p 代表地级市, $Firm$ 代表企业数量, $\max\{Firm_{cpt}\}$ 表示(t

年) 向 c 国出口的所有地级市中企业数量最多的一个, $\sum_p Firm_{cpt}$ 表示 (t 年) 向 c 国出口的企业总数。因此, 我们对目的地本土集聚的定义衡量了向特定国家出口的企业集聚在最核心的一个地级市的强度, 即对于任意目的国, 如果出口企业在一个地级市集聚的程度越高, 我们便认为出口至该市场的企业具有更高的目的地本土集聚。

表 4 制度环境对出口目的地本土集聚的回归结果: 2000—2008 年

Panel A						
因变量: 出口目的地本土集聚	(1)	(2)	(3)	(4)	(5)	(6)
测量制度质量的变量	政府责任	反腐	政治稳定	政府效率	监管质量	法治水平
制度质量	-0.2358*** (0.0444)	-0.4431*** (0.0542)	-0.0945** (0.0407)	-0.4686*** (0.0508)	-0.3313*** (0.0473)	-0.3813*** (0.0472)
国家体制	-0.5874*** (0.2076)	-0.3566* (0.2102)	-0.2963 (0.2094)	-0.3715* (0.2028)	-0.5511*** (0.2067)	-0.4216** (0.2039)
语言	-1.1325** (0.5541)	-0.7847 (0.5440)	-1.0646* (0.5550)	-0.6644 (0.5535)	-0.8519 (0.5438)	-0.8846 (0.5482)
国家控制变量	是	是	是	是	是	是
R ²	0.9642	0.9649	0.9638	0.9651	0.9644	0.9646
观测值	1408	1400	1397	1407	1392	1406
Panel B: 控制年份固定效应						
测量制度质量的变量	政府责任	反腐	政治稳定	政府效率	监管质量	法治水平
制度质量	-0.1129*** (0.0386)	-0.2226*** (0.0499)	-0.0288 (0.0355)	-0.2406*** (0.0487)	-0.1455*** (0.0418)	-0.1880*** (0.0434)
制度	-0.4181** (0.1741)	-0.3085* (0.1729)	-0.2853 (0.1757)	-0.3167* (0.1685)	-0.4028** (0.1746)	-0.3403** (0.1686)
语言	-0.6302 (0.5138)	-0.4620 (0.5083)	-0.5994 (0.5128)	-0.4037 (0.5132)	-0.5115 (0.5035)	-0.5132 (0.5103)
国家控制变量	是	是	是	是	是	是
年份虚拟变量	是	是	是	是	是	是
R ²	0.9736	0.9738	0.9735	0.9739	0.9737	0.9737
观测值	1408	1400	1397	1407	1392	1406

注: 由于 Kaufmann et al. (2009) 没有估算 2001 年的制度环境指数, 表 4 的回归结果不包含 2001 年; 国家控制变量包括: 是否与中国相邻、从北京到目的国的距离、目的地国的人口、不变价的人均 GDP 以及向该国出口的企业数量。括号中为怀特稳健标准误; **、*、* 分别表示在 1%、5%、10% 的水平上显著; 没有汇报常数项。

基于行政区域定义集聚的另一个优势是我们可以拓展本文的研究数据。如前所述, 在本文的第三与第四部分, 我们一共使用了七套数据。由于邮政区域的经纬度坐标数据以及工业企业数据时间年限的限制, 本文研究的时间跨度为 2002—2006 年。然而一旦(通过定义新的目的地本土集聚指数) 绕开 DO 方法, 我们便可以利用更长的时间序列(2000—2013 年) 检验本文结论的稳健性。

利用新定义的目的地本土集聚,我们使用混合最小二乘法重新估计了等式1,结果汇报于表4。在上表(Panel A)中,测量制度环境的6个指标均与目的地本土集聚呈现出了高度显著的负相关性。说明与制度环境更好的目的地相比,当制度环境更差时,出口至该市场的企业在中国境内更加集中于一个地区(地级市),即存在更强的目的地本土集聚。此外,与中国相同的意识形态与民族语言似乎显著降低了目的地本土集聚,这与上一部分的回归结果一致。在表4的下表(Panel B)中,我们控制了年份固定效应。可以发现测量制度环境的六个变量依然与目的地本土集聚存在显著的负相关性,即目的国制度环境越差,出口至该市场的企业在中国本土的集聚程度越高。

表4的稳健性分析使用了四套数据,分别为海关进出口企业数据(计算集聚指数)、制度环境指数、世界银行国家数据(包括目的国经济水平与人口)、CEPII数据(包括贸易距离等),时间跨度为2000—2008年。为了尽可能地利用样本信息,我们进一步将本文研究的时间跨度延长至2013年。其中,由于Kaufmann et al.(2009)只提供了截至2008年的制度环境指数,我们使用任意相邻三年的简单平均法推算出2009—2013年的制度环境指数。^①结果(没有汇报)依然显示制度环境与集聚存在显著且稳定的负相关性。

此外,如前所述,本文实证工作的开展,很大程度上依赖于出口企业的位置信息。我们的位置信息来自于邮政区的经纬度坐标数据。为考察本文实证结果的稳健性,我们也搜集并使用了来自于腾讯地图API端口的中国县级行政区域经纬度坐标数据,并进一步将县级行政区域的经纬度坐标赋值给所在县级区域内的所有出口企业,进而开展与本文第三、第四部分相同的数值模拟与回归分析。结果显示,出口目的地本土集聚依然存在,实证研究结果同样显示出口障碍越强企业的本土集聚程度越高。^②通过基准分析以及这些稳健性讨论我们认为,本文第三、第四部分的实证结果具有良好的稳健性。

六、结论性评述

本文基于2002—2006年详实的微观企业数据与手工整理的中国邮政区域经纬度坐标数据,使用Duranton & Overman(2005)的数值模拟方法,在控制行业结构干扰的基础上考察了中国出口企业空间分布的一个基本问题——是否存在出口目的地本土集聚。得益于详实的微观数据,我们克服了现有文献在研究集聚问题时普遍存在的行政区域限制,并证实中国确实存在出口目的地本土集聚;随后,我们使用计量研究方法对这一本土集聚现象的推动力进行了研究。我们发现当海外市场的制度环境越差时,出口至该市场的企业在中国本土的空间集聚越强,此外共同的意识形态与民族语言也对出口目的地本土集聚产生了显著影响。为了考察结果的稳健性,我们将实证研究从宏观逐步推进到微观开展了多层次的分析;控制了港口对集聚的潜在干扰,拓展了研究年份(2000—2013年),并使用县级行政区域的经纬度数据讨论了稳健型。这些工作均得到了与基准研究一致的结论,即出口目的地的制度环境越差,出口至该市场的企业在中国境内集聚的概率越高,而相同的意识形态和语言则显著降低了出口目的地本土集聚的发生概率。

本文的研究既为中国出口企业的空间分布提供了一个新的解释,也揭示了中国出口企业挑战高难度出口市场的一种生存法则。此外,本文也具有一定的政策寓意。首先,仅就出口促进政策而言,不同地区在引导当地企业进入出口市场时不但应该考虑出口企业面临的冰山成本,

^① Kaufmann et al.(2009)也没有估算2001年的制度环境指数,我们根据2000年与2002年的数据进行了推算。

^② 由于篇幅限制这一部分的结果没有报告,感兴趣的读者可以向作者索取。

还应考虑到中国企业普遍面临的出口信息障碍。从这个角度讲,引导相同出口市场的企业开展交流与合作具有现实意义。另一方面,不同地区在发展集聚经济,建立产业集群的过程中,分类别引导出口企业聚集具有意义。此外,对于微观企业而言,挑战高难度出口市场的企业在选址时,更接近相同目的地出口商不失为一种克服出口障碍的更优选择。虽然企业间的激烈竞争是市场经济的基本生存法则,但本文发现至少在现阶段,在出口市场上中国企业间的合作共赢机会依然普遍存在。

参考文献

- 包群、邵敏、Ligang Song 2012 《经济集聚、行业集中与中国企业出口模式的差异性》,《管理世界》第12期。
- 范剑勇 2006 《产业集聚与地区间劳动生产率差异》,《经济研究》第11期。
- 黄玖立、徐旻鸿 2012 《境内运输成本与中国的地区出口模式》,《世界经济》第1期。
- 刘竹青、周燕 2014 《地理集聚、契约执行与企业的出口决策——基于我国制造业企业的研究》,《国际贸易问题》第9期。
- 孙浦阳、韩帅、许启钦 2013 《产业集聚对劳动生产率的动态影响》,《世界经济》第3期。
- 佟家栋、刘竹青 2014 《地理集聚与企业的出口抉择: 基于外资融资依赖角度的研究》,《世界经济》第7期。
- 袁海红、张华、曾洪勇 2014 《产业集聚的测度及其动态变化——基于北京企业微观数据的研究》,《中国工业经济》第9期。
- 张海峰、姚先国 2010 《经济集聚、外部性与企业劳动生产率》,《管理世界》第12期。
- Cassey, Andrew, and Schmeiser, Katherine, 2013, “The Agglomeration of Exporters by Destination”, *Annals of Regional Science*, 51 (2), 495—513.
- Choquette, Eliane, and Meinen, Philipp, 2011, “The Channels of Export Spillover”, EARIE Conference Working Paper.
- Devereux, Michael, Griffith, Rachel, and Simpson, Helen, 2004, “The Geographic Distribution of Production Activity in the UK”, *Regional Science and Urban Economics*, 35 (5), 533—564.
- Duranton, Gilles, and Overman, Henry, 2005, “Testing for Localization Using Micro-geographic Data”, *Review of Economic Studies*, 72 (4), 1077—1106.
- Ellison, Glenn, and Glaeser, Edward, 1997, “Geographic Concentration in US Manufacturing Industries: A Dartboard Approach”, *Journal of Political Economy* 105 (5): 889—927.
- Ellison, Glenn, Glaeser, Edward, and Kerr, William, 2010, “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns”, *American Economic Review*, 100 (3), 1195—1213.
- Frankel, Jeffrey, and Romer, David, 1999, “Does Trade Cause Growth?”, *American Economic Review*, 89 (3): 379—399.
- Gould, David, 1994, “Immigrant Links to the Home Country: Empirical Implications for U. S. Bilateral Trade Flows”, *Review of Economics and Statistics*, 76 (2), 302—316.
- Guiso, Luigi, Sapienza, Paola and Zingales, Luigi, 2009, “Cultural Biases in Economic Exchange?”, *Quarterly Journal of Economics*, 124 (3): 1095—1131.
- Kaufmann, Daniel, Kraay, Aart, and Mastruzzi, Massimo, 2009, “Governance Matters VIII: Aggregate and Individual Governance Indicators 1996—2008”, World Bank Working Paper.
- Koenig, Pamina, Mayneris, Florian, and Poncet, Sandra, 2010, “Local Export Spillovers in France”, *European Economic Review*, 54 (4), 622—641.
- Olley, Steven, and Pakes, Ariel, 1996, “The Dynamics of Productivity in the Telecommunications Equipment Industry”, *Econometrica*, 64 (6), 1263—1297.
- Openshaw, Stan, 1984, “The Modifiable Areal Unit Problem”, *Concepts and Techniques in Modern Geography*.
- Ramos, R., and Moral-Benito, Enrique, 2013, “Agglomeration Matters for Trade”, Banco de España Working Paper, No. 1316.
- Rauch, James, and Trindade, Vitor, 2002, “Ethnic Chinese Networks in International Trade”, *Review of Economics and Statistics*, 84 (1), 116—130.
- Silverman, Bernard, 1986, *Density Estimation for Statistics and Data Analysis*, CRC Press.
- Xing, Weibo, and Zhou, Li-An, 2018, “Bilateral Trust and Trade: Evidence from China”, *World Economy*, 41 (8), 1918—1940.

Overseas Market , Institutional Environment and Origin Agglomeration

ZHANG Yili^a , ZHOU Kang^b and ZHANG Junsen^b

(a: Business School , Wenzhou University; b: Department of Economics , Chinese University of Hong Kong)

Summary: The location distribution of exporters is far from random , even in a relatively narrow region within a country. In contrast , the tendency of establishments to cluster within some areas has been a well-perceived economic phenomenon in China. For example , in 2006 , out of all 2932 counties where export firms were mostly concentrated , over 77% of the firms were clustered in the top 100 counties. More interestingly , more than 3500 of these firms were gathered in Shenzhen Baoan , a district less than 400 square kilometers in size. Motivated by these facts , in this paper , we step back from the thorough industrial organization literature to explore the location pattern of export firms by raising simple questions: do firms export to the same geographically concentrated counties in China relative to overall distribution? If this is the case , how can the phenomenon be explained?

A major issue arises when answering these questions: we must ex ante distribute firms into specific administrative units and then rely on administrative units to examine the firms' location pattern. This approach transforms spatial spots (firms) into uneven black boxes (administrative units) and thus artificially distorts the attributes of agglomeration. To overcome this challenge , we use the distance-based approach proposed by Duranton & Overman (2005) (DO approach) . This approach has been widely used for agglomeration studies but is rarely present in the Chinese literature due to a lack of data availability. Relying on detailed and unique coordinate data , we apply this approach to test destination-specific agglomeration , controlling for industry clusters in China. The main findings suggest that firms selling to the same destinations appear to be geographically concentrated , revealing the existence of a novel location pattern of export firms in terms of spatial distribution in China , referred to as export agglomeration by destination.

To explain this phenomenon , we analyze typical facts and empirically exploit the effects of export barriers in shaping the destination-specific location pattern. Our main findings suggest that firms selling to countries with worse business regulations and larger gaps in ideology and language than China tend to cluster significantly more in specific regions. Additionally , we conduct various robustness analyses and obtain consistent empirical evidence that destination-specific export barriers significantly affect export firms' location decisions.

Our work makes three marginal contributions to the present literature. First and foremost , a large body of economic geography literature examines the geographic agglomerations of firms from the perspective of industrial concentration. However , little literature has focused on the spatial distributions of export firms. We provide a novel technical explanation for the location distribution of China's export firms from the perspective of overcoming export barriers.

Second , our research approach and unique data contribute to the Chinese literature. Based on recently available nationwide coordinate data , this paper tests the effects of destination-specific agglomeration on continuous space and consequently extends beyond the restrictions placed on administrative units. Additionally , considering that the DO approach was initially developed to study the location patterns of industries , we also contribute to the trade literature by introducing a novel industrial organization research method.

Finally , one of the major export barriers that exporters face in export markets is the difficulty of obtaining destination-specific export information. However , little literature aims to provide empirical evidence of the effects of export barriers on firm behaviors (particularly from the perspective of spatial distribution in terms of exporters) . The main findings in this paper provide indicative indirect evidence that export barriers can affect the agglomeration behavior of firms , and that agglomeration affects spillovers through the information-sharing mechanism. Consequently , they reveal a survival rule for exporters looking to conquer tough export markets.

Keywords: Origin Agglomeration; Institutional Environment; Export Barriers; Spatial Statistics

JEL Classification: F14 , D21 , R12

(责任编辑: 林 一) (校对: 王红梅)