

# 地理集聚与企业的自选择效应

——基于上、下游关联集聚和专业化集聚的比较研究\*

李瑞琴 孙浦阳

**内容提要:** 本文从企业异质性角度检验了中国企业的区位自选择效应,即高生产率的企业是否一定会选择在产业集聚程度较高的中心区。在衡量集聚程度的方式上,本文不仅考察了传统的专业化集聚,还首次考虑了上、下游产业关联的多样化集聚,使用嵌套 Logit 模型进行了实证检验,着重对区位自选择效应在两类集聚条件下的差异性进行了对比分析。研究发现,在专业化集聚程度较高的中心地区,企业的自选择效应较为明显,但对上、下游产业集聚程度较高的中心地区,企业的自选择效应则相对较弱。进一步的研究还发现,国外市场依赖程度较大的纯加工贸易型企业更为关注优越的地理位置,因此,尽管这类企业生产率相对较低,但在竞争效应尚可容忍的条件下,它们仍倾向于选择那些具有地理优势的中心地区,这说明这些地区企业生产率的分布是较为分散的,这类企业不存在明显的区位自选择效应。基于此,本文认为,为缩小地区经济发展的不平衡性,核心地区在加快产业转型升级的同时,要为那些生产率较低的加工贸易型企业的区位转移创造有利条件;外围地区则要重视开放、基础设施建设和上、下游产业链的搭建,逐渐培育具有地区特色的产业集群。

**关键词:** 异质性企业 自选择效应 专业化集聚 上、下游关联集聚

**作者简介:** 李瑞琴,中央财经大学国际经济与贸易学院党委副书记、副教授,100081;

孙浦阳,南开大学跨国公司研究中心、南开大学经济学院教授,300071。

**中图分类号:** F129.9 **文献标识码:** A **文章编号:** 1002-8102(2018)04-0114-16

## 一、引言

党的十九大报告指出,新时代中国社会的主要矛盾已经转化为人民日益增长的美好生活需要和不平衡不充分发展之间的矛盾。从中国地区经济发展的不平衡看,受自然、历史条件以及国家宏观经济政策的影响,中国工业制造业较为明显地集中在东部及沿海地区,从而造成了地区经济发展“量”的不平衡。与此相关的另一个更重要的问题是,选择进入产业集聚程度较高的中心地区

\* 基金项目:教育部人文社会科学青年基金项目“对外贸易、异质性企业的区位选择与中国的地区差异”(13YJC790076);教育部重点研究基地重大项目“外资政策自由化、产业动态演化及其竞争力研究”(16JJD790026)。

企业的生产率是否会更高?如果这种企业区位自选择效应存在,那么将会进一步带来中国经济地区间“质”的差异性,从而加剧了中国地区经济发展的不均衡性。更进一步地,企业的区位自选择效应是否与产业集聚的不同形式相关呢?较多文献考察的一般都是同产业的专业化集聚,但是包含上、下游关联的多样化集聚,企业的自选择效应是否也存在呢?此外,上述规律适用于中国所有类型的企业吗?例如,纯加工贸易型企业是否也存在区位自选择效应呢?对以上这些问题的研究,一方面可以深入探讨“新”新经济地理学基本理论在发展中国家的具体表现,另一方面也可以更进一步理解中国地区经济发展不平衡背后的深层次原因,并找到切实可行的方法;同时对区域内部的产业关联问题也将会有一个更清晰的认识。

针对企业区位自选择效应,国内外都有一些理论拓展和实证研究。Baldwin 和 Okubo(2006)发现自选择效应和集聚效应都是存在的,且在二者的共同作用下,生产率较低的企业既可为生存而扩散至竞争水平较低的外围区,也可受本地外部性和本地市场效应的影响继续留在中心区。Melitz 和 Ottaviano(2008)、Okubo 和 Forslid(2010)证明市场竞争激烈程度会反作用于企业进入该市场的选择,只有生产率较高的企业才能在中心地区生存。Behrens, Duranton 和 Robert-Nicoud(2014)认为自选择效应的存在的确容易高估集聚效应的作用。Okubo 和 Tomiura(2011,2012)采用日本的微观企业数据,利用密度函数图验证了中心地区平均生产率较外围地区更高。Combes, Duranton, Gobillon, Puga 和 Roux(2012)也证实了自选择效应和集聚效应的存在。国内梁琦等(2013)运用分位数回归的方法,证明异质性企业的区位自选择行为在中国显著存在,但集聚效应并不明显。王永进、张国峰(2016)以开发区企业为对象,认为集聚效应是开发区企业短期生产率优势的主要来源,而自选择效应则是长期生产率优势的主要来源。相反,余壮雄、杨扬(2014)基于格点搜索的 NLS 回归来识别集聚效应与自选择效应的方法,认为集聚效应是解释中国大城市生产率优势的基本原因,而自选择效应并不存在。

综上文献,我们发现:(1)现有研究对地区集聚程度的衡量主要基于专业化集聚,而没有太多涉及上、下游关联的多样化集聚,但是,现实中上、下游产业的配套能力对于一个地区企业竞争力的提升是非常重要的,本文认为对它们进行区别讨论,将会使这一领域的研究更深入。(2)现有文献在研究不同生产率企业的区位选择时,没有太多考虑企业类型的差异性,占中国贸易半壁江山的加工贸易企业,具有明显的地区偏向性,我们认为需要关注区位自选择效应在这类企业中的表现。(3)在实证的方法上,目前的研究大多采用了密度函数图、分位数回归等方法,研究的结论还存在一定争议。本文运用嵌套 Logit 模型进行实证研究。这一方法的好处是能够同时考虑地区特征和企业特征对企业区位选择的影响,并且可以对两种不同集聚形式的区位自选择效应进行比较研究。

本文其余部分安排如下:第二部分为理论假定及机制说明;第三部分为研究设计;第四部分是实证检验及结果分析、讨论和稳健性检验;第五部分是对全文的总结和政策建议。

## 二、理论假定及机制说明

以 Krugman 为代表的学者开创的新经济地理学,研究了影响产业集聚的“经济关联”(包括生活成本效应、生产成本效应和市场接近效应)和“知识关联”,与竞争效应和拥挤效应的内在作用机制。但是新经济地理学模型对上述问题的分析都假定企业是同质的,因此在面对市场中共同的集聚力和分散力时,所有企业的表现是相同的,这点与现实不太相符。而引入了企业

异质性假定的“新”新经济地理学,则进一步认为企业生产率的异质性会对企业的区位选择带来重要的影响。正如前文文献部分所述,由于产业集聚程度高的中心地区拥有更大规模的市场需求,受规模经济的影响,企业会倾向于选择中心地区。但是在产业集聚程度较高的中心地区,企业之间的竞争效应和拥挤效应往往也更强,只有生产率更高的企业才有较低的边际成本,面对高昂的进入门槛,只有它们有能力跨越这一门槛,应对中心地区的高成本,分享产业集聚给其带来的正外部性。而生产效率较低的企业则会得不偿失,无利可图,所以只有生产率更高的企业才会选择进入中心地区,存在区位自选择效应,这无疑会强化地区经济发展的不平衡性。由于“新”新经济地理学的研究也是基于垄断竞争的市场结构展开的,因此他们对集聚程度较高的中心地区的界定标准,就是基于同产业的专业化集聚。由此,可以得到本文的第一个假定。

假定 1:在专业化集聚程度较高的中心地区,只有生产率更高的企业才能进入,存在区位自选择效应。

产业的地区集聚通常可以分为同产业的专业化集聚和基于产业上、下游关联的多样化集聚,而往往上、下游产业的配套能力对地区企业竞争力的提升和区域经济发展具有重要作用,因此多样化集聚受到了很大的关注。诸如 CPVL 模型、FCVL 模型和 FEVL 模型等新经济地理学的垂直关联模型(Venables, 1996)指出,上游行业的集聚能够为下游企业带来质优价廉且更多差异化的中间产品,而下游产业的集聚能够为上游企业提供便捷的销售渠道等。因此,一个地区的上、下游产业的集聚程度也是吸引企业区位选择的重要决定因素。那么,在单纯的上、下游产业集聚的中心地区,企业是否也存在区位自选择效应呢?我们认为,尽管在上、下游产业高度集聚的中心地区,资源和生产要素成本也会在一定程度上提高,增加企业的进入成本,从而对企业的生产效率也会有比较高的要求。但是与专业化集聚不同的是,基于产业关联的上、下游企业更多体现的是一种企业间的合作和协作,企业之间往往并不会形成非常直接的竞争关系,它们之间的竞争和拥挤效应会相对较弱,尤其是对那些下游产业集聚程度较高的地区,往往会使上游企业的产品供不应求,从而对选择进入该地区的上游企业生产效率的要求不会太高,因此,企业区位自选择效应可能没有专业化集聚地区表现得那么强。由此,得到本文的第二个假定。

假定 2:在单纯的上、下游产业集聚程度较高的中心地区,企业区位自选择效应弱于专业化集聚地区。

上述分析并没有特别考虑外贸型企业,尤其是纯加工贸易型企业。而众多的研究都表明,这种类型的企业在中国都表现出了诸多的差异性(戴觅等,2011)。加工贸易型企业具有明显的大进大出特点,对国外市场的依赖性非常大,除关注地区要素成本外,这类企业的选址对地区的地理位置和交通的便利性更为敏感(黄玖立、李坤望,2006)。只要地区拥挤成本和竞争压力带来的要素成本上升尚可容忍,即使它们的生产率相对较低,也依然倾向选择那些地理位置优越的地区,而往往这些地区产业集聚程度又相对较高。这说明在这类中心地区,尽管企业的平均生产率相对较高,但企业生产率的分布是比较分散的。由此,得到本文的第三个假定。

假定 3:鉴于纯加工贸易型企业的特殊性,尽管它们生产率偏低,但由于它们对运输成本更敏感,那些地理位置比较优越的中心地区可能仍是它们的优选,因此纯加工贸易型企业不存在明显的区位自选择效应。

### 三、研究设计

#### (一) 计量模型的设定

一个企业是否选择某一个地区,主要取决于地区特征和企业特征,McFadden(1981,1984)提出的条件 Logit 模型(Conditional Logit Model, CL)在研究个体选择行为时,既考虑了个体的特征又考虑了选择方案的特征,因此比较适合研究企业的地理集聚。但是 McFadden 同时指出,这一模型必须满足独立不相关性假设,即两个备选方案发生的机会比率要独立于第三个备选方案,各备选项之间要相互独立,但现实中很多问题是不满足这一假定条件的,若坚持使用 CL 予以分析,其参数估计必将产生严重的不一致性。为此,McFadden 构建了嵌套 Logit 模型(Nested Logit Model, NL)。该模型把多个选择方案  $k$  分为  $n$  个组,每组内的选择方案不具有相关性,而组间则具有相关性。由此,在第  $n$  组方案  $k$  被个体  $i$  选中的概率可表示为:

$$P_{nk}^i = P_n^i P_{k/n}^i = \frac{\exp(z_n' \alpha + \tau_n I_n)}{\sum_{m=1}^N \exp(z_m' \alpha + \tau_m I_m)} \times \frac{\exp(x_{nk}' \beta_n / \tau_n)}{\sum_{l=1}^{K_n} \exp(x_{nl}' \beta_n / \tau_n)}, m \neq n, l \neq k_n \quad (1)$$

式(1)中,  $P_n^i$  表示第  $n$  组(省区分组)被个体  $i$ (某企业)选中的概率,  $P_{k/n}^i$  表示第  $n$  组第  $k$  个方案(某省区)被个体  $i$  选中的概率;  $I_n = \ln[\sum_{l=1}^{K_n} \exp(x_{nl}' \beta_n / \tau_n)]$  为组内值(Inclusive Value),衡量了第  $n$  组所涵盖的各方案的具体特征对该组影响的期望最大效用;  $\tau_n$  表示该效用的影响系数,称为不相似系数;  $x_{nk}'$  表示在第  $n$  组内对选择第  $k$  个方案产生影响的变量,即反映备选方案  $k$  的特征变量;  $z_n'$  表示对选择第  $n$  组产生影响的变量,反映个体  $i$  的特征变量。

本文将中国 30 个省区作为企业进行行政区划选择的基本单位,<sup>①</sup>但是企业在这些省区之间进行区位选择时,各个省区之间是不满足相互独立条件的,比如某一家企业在选择是去东部沿海经济发达省区,还是去中西部偏远省区时,它们之间的相关性,就远高于选择只属于东部沿海地区或中西部地区省份的相关性。这说明企业并不是把中国各省区看作一个层面,而是会把各个省区按照某种标准进行分类后再进行选择,即按照一种“区域-省”的树状分层结构来选择最终的厂址(Crozet 和 Mucchielli, 2004)。现有文献对中国省区的分类有按照“中心-外围”标准分为两类的;有按照各省地理位置的不同划分为华北、华东、华南、西北、西南、东北、华中七大区域的(余佩、陈继勇, 2012),还有综合考虑国家政策的差异性、地区经济发展水平和地理位置等划分为东部、中部、西部三大区域的。相比较而言,“中心-外围”是比较理想的方式,但本文对“中心-外围”的划分是基于行业层面的,因此针对不同年份的不同省区,可能对某一行业是中心地区,但对其他行业则可能为外围地区。可见,这种分类方式将导致在“中心-外围”两类地区下的较多省份是重复的,故这一分类标准在本文的研究中不可行。而七大区域的划分方法主要是基于地理位置的划分,几个区域间的经济发展水平和交通便利程度仍有相似性,不能满足独立不相关假设。因此,本

① 选取省作为基本的研究单位,主要是由于省是我国基本的行政区域单位之一,大多数经济指标都是在省级层面进行统计的,企业在进行行政区划选择时,首先决定的是省,然后再决定去该省具体的市县。当然,地区单位选择越细越有意义,但是根据本文选取的全国层面企业的数据看,每个企业可以选取的省级单位,已经有 30 个选择,如果考虑到更细的单位,如市层面时,则意味着一个企业有 280 多种选择,各选项过多导致采用嵌套 Logit 模型(非线性回归)运算极为困难的问题。在后续的研究中可以针对某些代表性省份的企业研究更细单位的区位选择问题。

文主要采用了东部、中部、西部三大区域的划分方法,但考虑到东部在经济发展水平和地理位置方面都明显优于中西部地区,因此后文还就两区域分类进行了进一步的稳健性检验。

本文的计量模型即为式(1),因变量为2000—2007年省*k*被某企业*i*选为投资目的地的情况。如果*k*被选中,因变量记为1,如果没有,则记为0。根据已有相关文献,本文选取的反映地区特征的控制变量 $x'_{jk}$ 主要有:专业化集聚指数(*EG*)和上、下游产业集聚指数(*EG\_up*和*EG\_down*):反映循环累积效应对企业区位选择的影响;“中心”和“外围”地区的虚拟变量(*Cluster*, *Cluster\_up*和*Cluster\_down*)和企业生产率(*TFP*)的交叉项:反映企业的区位自选择效应;市场规模(*MP*):用市场潜能<sup>①</sup>来表示(Harris, 1954),根据新经济地理学的基本理论,经济规模越大的地区,越有利于厂商实现规模经济,企业越愿意选择该地区;效率工资(*Ewage*):用当地工人平均工资除以人均工业总产值来表示(余佩、陈继勇, 2012),它既考虑了成本又考虑了生产率,用来衡量劳动力成本;技术水平(*Patent*):用各省区的专利授权数占全国的比重来反映(黄肖琦、柴敏, 2006);国外市场接近度(*FMA*):用国外市场接近度指数<sup>②</sup>来反映(黄玖立、李坤望, 2006)。根据新经济地理学和“新”新经济地理学的理论文献,影响企业区位选择的企业特征 $z'_n$ 主要有企业的生产率(*TFP*)和是否为贸易企业(*Trade*):为避免利用传统的OLS方法进行估计时产生的同时性偏差和样本选择性偏差,本文分别使用了Olley和Pakes(1996)以及Levinsohn和Petrin(2003)提出的两种半参数估计法来衡量*TFP*,分别为*TFP\_op*和*TFP\_lp*;由于会较多的涉及跨国的产品运输,贸易型企业将非常关注地区的地理和交通优势,因此企业是否为贸易企业也是影响其区位选择的重要因素。本文假定只要某一企业有出口或进口,*Trade*就记为1,相反则为0。<sup>③</sup>

(二)主要指标的界定

1. 专业化集聚指数(*EG*)和上、下游产业的多样化集聚指数(*EG\_up*和*EG\_down*)。本文使用*EG*指数(Ellison和Glaeser, 1997)来衡量某地区某行业的专业化集聚水平,该指数充分考虑了企业规模、区域差异以及产业组织等因素对产业集聚带来的影响,弥补了传统产业集聚测度方法的缺陷,使产业集聚能够在给定区域范围内进行跨行业、跨时间和跨地区的比较。本文以省区为单位来构建该指数,设*k*代表省,*r*为*k*省下面的市,*j*代表行业。受构建上、下游产业集聚指标时使用的投入产出表信息的限制,本文的行业定义在工业企业数据库中两位行业代码的基础上。因此,行业*j*在地区*k*的*EG*指数为:

$$EG_{jk} = \frac{G_{jk} - (1 - \sum_{r \in k} x_r^2) H_{jk}}{(1 - \sum_{r \in k} x_r^2) (1 - H_{jk})} \tag{2}$$

① 根据Harris(1954)的定义,国内市场潜能 $DMP_i = \sum_j \left( \frac{GDP_j}{D_{ij}} \right) + \frac{GDP_i}{D_{ii}} (i \neq j)$ ,其中 $GDP_i$ 和 $GDP_j$ 为第*i*和第*j*省的国内生产总值, $D_{ij}$ 为*i**j*两省省会城市间的最短公路里程, $D_{ii}$ 指标参照Redding和Venables(2004)的处理方式 $D_{ii} = \frac{2}{3} \sqrt{S_i/\pi}$ ( $S_i$ 为*i*省的陆地面积)。

② 国外市场接近度指数的计算公式为 $FMA_i = \begin{cases} 100 \times D_{ij}^{-1} \\ 100 \times [\text{Min}(D_{ij}) + D_{ii}]^{-1} \end{cases}, i \notin C, j \in C$ ,其中*C*表示所有沿海省份的集合。由于中国的出口主要是通过海洋运输实现的,因此本文就用该指标来反映地区的地理优势。

③ 本文还用企业的出口比例进行了再检验,与采用虚拟变量的回归结果没有太大差异。限于篇幅本文未报告结果,感兴趣的读者可以向作者索取。

式(2)中,  $G_{jk} = \sum_{r \in k} (s_{jr} - x_r)^2$  是  $k$  省  $j$  行业的 Gini 系数 ( $s_{jr}$  是某市  $r$  行业  $j$  的就业人数占该行业在  $k$  省就业总人数的比重;  $x_r$  为  $k$  省某市就业人数在  $k$  省就业总人数中所占的比重);  $H_{jk} = \sum_{i \in j, k} z_i^2$  是行业  $j$  在地区  $k$  的赫芬达尔系数 ( $z_i$  是行业  $j$  中企业  $i$  的就业人数占本行业在  $k$  省就业总人数的比重)。

衡量上、下游集聚程度的指标,借鉴 Javorcik(2004)的方法,本文将  $EG$  指数结合投入产出表的信息得到。若用  $\sigma_{jm}$  表示  $j$  行业使用  $m$  行业的产品产值在  $j$  行业总投入中所占的比重,企业  $i$  所属行业  $j$  在地区  $k$  上游产业的集聚指标  $EG_{up}$  可以表示为:

$$EG_{up_{jk}} = \sum_{m, m \neq j} \sigma_{jm} EG_{mk} \quad (3)$$

同理,若用  $\mu_{jm}$  表示  $j$  行业被  $m$  行业使用的产品产值在  $j$  行业总产值中所占的比重,则企业  $i$  所属行业  $j$  在地区  $k$  下游产业的集聚指数  $EG_{down}$  可以表示为:

$$EG_{down_{jk}} = \sum_{m, m \neq j} \mu_{jm} EG_{mk} \quad (4)$$

2. 对地区产业集聚程度的划分。若某年某省区某行业的  $EG(EG_{up}, EG_{down})$  值高于当年各省区该行业的平均  $EG(EG_{up}, EG_{down})$  值,则该年该省区为该行业(该行业的上、下游行业)聚集的中心地区,设定  $Cluster = 1 (Cluster_{up} = 1, Cluster_{down} = 1)$ , 否则该年该省区为该行业的外围地区,此时  $Cluster = 0 (Cluster_{up} = 0, Cluster_{down} = 0)$ ①。

### (三) 数据来源及说明

本文的数据主要来源于 2000—2007 年的《中国工业企业数据库》。由于指标多、样本大,数据库存在错漏和统计口径上的误差等。根据已有的研究(聂辉华等,2012),本文对样本做了以下筛选:剔除工业增加值和工业总产值为负或者为零的企业;剔除出口交货值为负的企业;剔除固定资产年均净值为负或者为零的企业;剔除职工人数小于 8 人的企业;剔除存活时间低于 1 年或高于 59 年的企业;考虑到能够自主进行区位选择的行业需为“松脚型”行业,本文只保留了工业制造业企业。此外,本文还根据 Brandt、Biesebroeck 和 Zhang(2012)的做法,先后利用企业代码、企业名称、法人姓名和邮编等信息对同一家企业进行识别。为计算行业上、下游的集聚程度,本文使用了 2007 年 135 个部门的《中国投入产出表》计算了某一个行业使用的上游产业和被下游产业使用的比例。省级经济变量来自相关各期的《中国工业经济统计年鉴》和《中国统计年鉴》。在估计企业生产率时,为避免因价格变化带来的影响,本文用各地区工业品出厂价格指数对工业增加值进行平减,用各地区固定资产投资价格指数对资本进行价格平减,平减指数均来自国研网数据库。在分样本检验中,为识别企业是否为贸易型企业,特别是其是否为纯加工贸易型企业,本文将《中国海关数据库》与《中国工业企业数据库》进行了匹配,为了保证能匹配到更多的企业,本文采用企业名称以及邮政编码和电话号码两种方法进行了匹配(田巍、余森杰,2013)。

① Ellison 和 Glaeser(1997)将某地区制造业的集聚程度分为三类:第一类是某省某行业  $EG > 0.05$ , 则该行业为所在省的高集聚(Highly Concentrated)行业;第二类是某省某行业  $EG$  满足  $0.02 \leq EG \leq 0.05$ , 则该行业为所在省的非常集聚(Very Concentrated)行业;第三类是某省某行业  $EG < 0.02$ , 则该行业为所在省的较低集聚(Not Very Concentrated)行业。由于这是依据美国相关数据的计算结果进行划分的,不具有一般性,另外,为避免阈值设定的人为性可能会影响实证结果的稳健性,本文采用正文所述的方法设定中心和外围地区。

#### (四) 内生性问题的说明及处理

正如文献部分所指出的,如果集聚效应和自选择效应同时存在,那么用全样本进行回归时,将不可避免地出现反向因果带来的内生性问题,从而影响结论的可信度。因此,本文仅保留了新建企业的样本,这样可以较好地剔除集聚效应的影响。此外,还可能存在的内生性问题是:一方面工业基础较好、企业集聚程度高的地区是工业制造业企业的优选之地;另一方面,这些企业的区位选择行为又会提升该地区的工业基础,从而也可能会出现反向因果带来的内生性问题。本文选择了所有地区特征解释变量的滞后期来处理这一内生性问题,尽管这种方法不尽完美,但在一定程度上能够缓解内生性导致的估计偏误(Stiebale, 2011)。

#### (五) 描述性分析

1. 中国制造业各行业的集聚程度。从图 1 显示的 2000—2007 年全国层面各制造业的空间 Gini 系数的均值可以看出,产业空间集聚程度比较高的行业主要集中在烟草制品业(16<sup>①</sup>),皮革、毛皮、羽毛(绒)及其制品业(19),文教体育用品制造业(24),石油加工、炼焦及核燃料加工业(25),通信设备、计算机及其他电子设备制造业(39)以及仪器仪表及文化、办公用机械制造业(40)。考虑到石油加工、炼焦及核燃料加工业对地区资源的严重依赖,而烟草制品业作为一个高利税的行业,可能受到当地政府更多的保护,所以在下文的实证检验中,本文将只选取其余的四个集聚程度较高的行业进行回归分析。

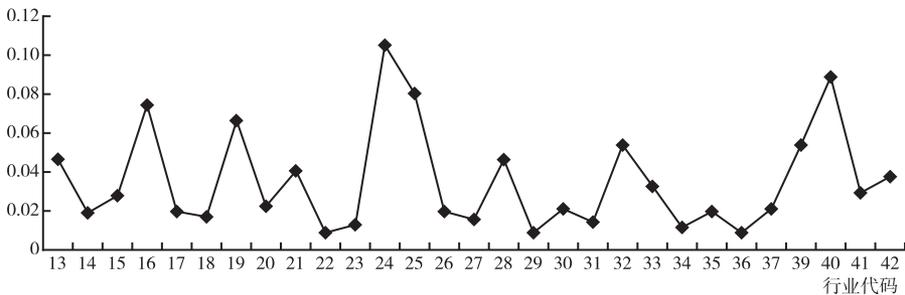


图 1 2000—2007 年中国各制造业行业的空间 Gini 系数均值

资料来源:根据 2000—2007 年《中国工业企业数据库》的相关数据整理计算而得。

2. 中心和外围地区企业 *TFP* 的比较分析。从表 1 可以看到,无论是专业化集聚还是多样化集聚,无论采用哪种方法衡量企业的平均 *TFP*,中心地区企业的 *TFP* 都显著较高。当然,由于同一个地区存在既是专业化集聚的中心地区又是上、下游产业集聚的中心地区的可能性,在简单的统计分析中还不能分离它们各自的影响,因此需进行规范的计量分析。

3. 不同类型企业的统计分析。从表 2 可以看到,纯加工贸易型企业与其他类型企业的 *TFP* 相比是低的,这与多数文献对中国出口企业生产率悖论的研究是一致的(戴觅等, 2011)。表 3 进一步分析了纯加工贸易型企业的地区分布,可以看出,无论哪种集聚方式,它们都主要集中在东部沿海地区。这说明这些企业对便捷的交通和优越的地理位置更敏感,并不非常关注该地区是否为该产业集聚的中心地区以及上、下游配套程度如何。

① 括号内为国民经济行业分类代码的前两位码,下同。

表 1 中心和外围地区企业 TFP 的统计分析

集聚类型	集聚程度	观测值	TFP <sub>OP</sub> 平均值	TFP <sub>OP</sub> 标准差	TFP <sub>LP</sub> 平均值	TFP <sub>LP</sub> 标准差
专业化集聚	中心地区	114596	3.6013	1.0873	6.2351	1.0964
	外围地区	67654	3.5609	1.1539	6.2206	1.0998
多样化集聚 (上游产业集聚)	中心地区	89198	3.5974	1.1325	6.2397	1.1194
	外围地区	93052	3.5497	1.1181	6.2202	1.1045
多样化集聚 (下游产业集聚)	中心地区	89198	3.5842	1.1312	6.2834	1.1326
	外围地区	93052	3.5457	1.1236	6.2076	1.1024

资料来源:根据 2000—2007 年《中国工业企业数据库》的相关数据整理计算而得。

表 2 纯加工贸易型企业与其他企业 TFP 的对比

企业类型	观测值	TFP <sub>OP</sub> 平均值	TFP <sub>OP</sub> 标准差	TFP <sub>LP</sub> 平均值	TFP <sub>LP</sub> 标准差
纯加工贸易型企业	7950	3.0740	1.3727	6.1490	1.0910
其他类型企业	174300	3.6007	1.1089	6.1843	1.4357

资料来源:根据 2000 年—2007 年《中国工业企业数据库》和《中国海关数据库》的相关数据整理计算而得。

表 3 纯加工贸易型企业主要的地区分布

集聚的分类	集聚程度	地区	观测值	占比(%)	地区	观测值	占比(%)
专业化集聚	中心地区	东部	4097	50.74	沿海	4088	50.63
		非东部	285	3.53	非沿海	294	3.64
	外围地区	东部	3483	43.13	沿海	3463	42.89
		非东部	210	2.60	非沿海	230	2.85
多样化集聚 (上游产业集聚)	中心地区	东部	3958	49.02	沿海	3953	48.95
		非东部	257	3.18	非沿海	262	3.24
	外围地区	东部	3622	44.85	沿海	3598	44.56
		非东部	238	2.95	非沿海	262	3.24
多样化集聚 (下游产业集聚)	中心地区	东部	5327	65.97	沿海	5299	65.62
		非东部	334	4.14	非沿海	362	4.48
	外围地区	东部	2253	27.90	沿海	2252	27.89
		非东部	161	1.99	非沿海	162	2.01

资料来源:根据 2000—2007 年《中国工业企业数据库》和《中国海关数据库》的相关数据整理计算而得。

## 四、实证检验及结果分析

### (一) 基准回归

基准回归中,我们将中国 30 个省分为东部、中部、西部三个区域,企业在确定了区域后再选择

所在区域的具体省份。表4列示了基于嵌套logit模型所得的五个基本回归结果。总的来看,这些回归结果的Wald chi2值都在1%的显著水平上通过检验,说明模型整体的回归结果较好。IIA也在1%的显著水平上通过检验,不相似系数也显著的不为1,说明嵌套logit的结果要好于条件logit。<sup>①</sup>

1. “中心-外围”地区与企业生产率的关系。在表4的(2)~(4)列中,本文分别加入了反映地区“中心”、“外围”特征的虚拟变量与企业TFP的交叉项。从回归结果来看, $Cluster \times TFP_{op}$ 、 $Cluster_{up} \times TFP_{op}$ 和 $Cluster_{down} \times TFP_{op}$ 的估计系数都为正,且在1%的水平上具有显著性,但是 $Cluster \times TFP_{op}$ 的回归系数明显大于其他两个。(5)列将三者同时放入进行回归,结果显示, $Cluster \times TFP_{op}$ 、 $Cluster_{up} \times TFP_{op}$ 都有较高的显著性,且 $Cluster \times TFP_{op}$ 估计系数数值更大,同时 $Cluster_{down} \times TFP_{op}$ 估计系数则没有了显著性,这说明集聚程度较高的中心地区,尤其是专业化集聚的地区,由于企业之间竞争和拥挤效应更强,只有生产率更高的企业才能进入,存在企业的区位自选择效应;而单纯的上、下游产业的集聚,因企业间的合作多于竞争,故对新进入企业生产率的要求并不是很高,尤其是下游产业集聚程度较高的地区,对上游企业产品的需求旺盛,反而对上游企业的生产效率降低了要求,因此上、下游集聚地区的自选择效应相比于专业化集聚地区较弱。

2. 其他控制变量对企业区位选择的影响。在表4(1)~(5)列的分层检验中,与基准组中部地区相比,东部地区Trade的估计系数为正,且在1%水平上显著,说明与中部地区相比,越是贸易型的企业越会选择东部地区。而西部地区Trade估计系数则不具有显著性,说明贸易型企业在中部、西部地区的选择上并不具有明显的偏好。主要的原因可能是相比于西部地区,中部地区虽然离沿海地区较近一些,但是一方面,中部地区交通等基础设施建设仍相对落后,另一方面,近些年国家对西部地区开发政策力度的加大,导致中部地区相比于西部的优势并不是特别明显,故也不是贸易型企业的优选,这可能在一定程度上佐证了中部成了“被遗忘的区域”的事实。此外,在以上五个回归结果中,国内市场潜能(MP)的估计系数只有(3)列和(5)列显著为正,其他则不显著,这可能说明尽管一个省区的经济发展潜力和规模越强,越会吸引企业进入,以更好地实现规模经济,但是在当前开放经济的条件下,外部市场需求作用往往不容忽视,这会导致国内市场潜能的作用不太稳健。效率工资(Ewage)的估计系数显著性较弱,说明一个省区劳动力成本对企业进入与否影响不大,对劳动力质量的要求可能会更关注。地区技术水平(Patent)的估计系数为正,都在1%水平上显著,说明省区技术水平的发展程度越高,越有利于企业的进入。国外市场接近度(FMA)的估计系数为负,也都在1%水平上显著,说明开放条件下,国外市场的接近度对企业的区位选择非常重要。三个反映集聚程度的变量EG、EG<sub>up</sub>和EG<sub>down</sub>基本都有正的显著性,证明了循环累积的集聚力对于企业区位选择的重要性。

## (二)分样本回归结果分析

为分析不同类型企业的区位自选择效应,这里将进行分样本检验。<sup>②</sup>在排除了纯加工贸易型企

① 限于篇幅,本文未报告CL和NL两种回归方式的结果对比,感兴趣的读者可以向作者索取。

② 本文将所有的样本分为非贸易型企业、非纯加工贸易型企业和纯加工贸易型企业三类,并对它们也做了回归,实证结果发现,前两类企业的回归结果没有太大的差异,主要原因在于这两类企业比较集中的东部沿海地区兼具优越的地理位置和较大的市场规模两个特点,高集聚区对企业的TFP要求都较高,而这两类企业也的确拥有较高的TFP。受篇幅所限,正文将其合并,只报告了它们与纯加工贸易型企业的回归结果。

表 4 基准回归结果

Pro	(1)	(2)	(3)	(4)	(5)
<i>lnMP</i>	0.0626 (0.5978)	0.0630 (0.5758)	0.3301 *** (2.7902)	0.1173 (1.0363)	0.3396 *** (2.9915)
<i>lnEwage</i>	-0.3015 *** (-2.7137)	-0.0562 (-0.4734)	-0.1399 (-1.1553)	-0.2695 ** (-2.2814)	0.0839 (0.6942)
<i>lnPatent</i>	1.3981 *** (14.5981)	1.5562 *** (17.0957)	1.4426 *** (14.1203)	1.4492 *** (14.5129)	1.4921 *** (16.2681)
<i>lnFMA</i>	-0.6777 *** (-11.4927)	-0.6872 *** (-12.7389)	-0.6730 *** (-11.2165)	-0.7206 *** (-11.5121)	-0.6237 *** (-12.2221)
<i>EG</i>	1.1481 *** (5.6896)	0.1596 (0.6293)	1.1182 *** (5.1778)	1.0742 *** (4.8794)	0.2566 (1.0326)
<i>EG<sub>up</sub></i>	1.0025 *** (3.4270)	1.1384 *** (3.8248)	0.2790 (0.8017)	1.0682 *** (3.4450)	0.3169 (0.9434)
<i>EG<sub>down</sub></i>	0.3696 (0.9081)	0.7785 * (1.7591)	0.4966 (1.1850)	0.7326 * (1.7094)	0.9020 ** (2.1475)
<i>Cluster × TFP<sub>op</sub></i>		0.2483 *** (8.3669)			0.2235 *** (8.2392)
<i>Cluster<sub>up</sub> × TFP<sub>op</sub></i>			0.1512 *** (8.3564)		0.1401 *** (8.0205)
<i>Cluster<sub>down</sub> × TFP<sub>op</sub></i>				0.0877 *** (3.2811)	0.0335 (1.2656)
type equation					
<i>east Trade</i>	1.4026 *** (9.3504)	1.3432 *** (8.5845)	1.3429 *** (8.5874)	1.3508 *** (8.6366)	1.3314 *** (8.5125)
<i>middle Trade</i>	0(基准组)				
<i>west Trade</i>	-0.1189 (-0.3735)	-0.0368 (-0.1145)	-0.0360 (-0.1118)	-0.0299 (-0.0927)	-0.0548 (-0.1705)
不相似系数					
<i>east_tau</i>	1.7899 *** (14.6019)	1.7967 *** (17.2912)	1.8040 *** (14.3212)	1.8125 *** (14.5719)	1.7449 *** (16.8714)
<i>middle_tau</i>	2.2549 *** (11.8234)	2.2855 *** (14.3245)	2.2847 *** (11.5812)	2.3050 *** (11.8499)	2.1868 *** (13.7414)
<i>west_tau</i>	1.5910 *** (8.1170)	1.5041 *** (9.6720)	1.6233 *** (7.9827)	1.6077 *** (8.0890)	1.4587 *** (9.2070)
观测值	183289	173581	173581	173581	173581
Log likelihood	-16569.853	-15409.372	-15396.105	-15446.933	-15356.353
Wald chi2	378.51	491.25	354.91	366.89	474.01
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000
LR test for IIA	402.37	433.38	367.44	393.67	383.12
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000

注:(1) \*、\*\*和\*\*\*分别代表10%、5%和1%的显著性水平;(2)括号内为估计系数的z值,下表同。

业后,表5中(1)~(4)列的结果显示,  $Cluster \times TFP_{op}$  和  $Cluster_{up} \times TFP_{op}$  仍然显著为正,  $Cluster \times TFP_{op}$  的估计系数大于  $Cluster_{up} \times TFP_{op}$ , 而  $Cluster_{down} \times TFP_{op}$  在(3)列中回归系数相对较小,(4)列中则没有显著性,再次证明区位自选择效应在专业化集聚地区的确表现得更为明显。同时,在排除了纯加工贸易型企业后,上述估计系数值比全样本下的相应值都要更大一些,在一定程度上间接说明了纯加工贸易型企业不存在明显的区位自选择效应。为确认这一结论,我们还用纯加工贸易型企业的样本做了进一步的验证。表5中(5)~(8)列为对纯加工贸易型企业的回归结果,可以发现只有当地市场技术水平 ( $\ln Patent$ ) 和国外市场接近度 ( $\ln FMA$ ) 这两个指标具有较高的显著性;  $Cluster_{up} \times TFP_{op}$  和  $Cluster_{down} \times TFP_{op}$  没有显著性,而  $Cluster \times TFP_{op}$  的估计系数则显著为负,这再次佐证了假定3的结论,对于这种特殊类型的企业,尽管生产率较低,但在市场要素成本尚可容忍的情况下,仍会倾向于选择地理位置更优越的东部沿海中心地区,的确不存在明显的区位自选择效应。

表5 分样本回归结果

Pro	除纯加工贸易型企业外的样本				纯加工贸易型企业样本			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln MP$	0.1568 (1.4194)	0.4480*** (3.6503)	0.2121* (1.8311)	0.4468*** (3.8664)	0.0339 (0.6892)	0.0826 (0.5798)	0.0360 (0.5578)	0.2633 (1.0633)
$\ln E_{wage}$	-0.0896 (-0.7521)	-0.1844 (-1.5064)	-0.3196*** (-2.6578)	0.0515 (0.4265)	0.6658 (0.5181)	-0.0680 (-0.0742)	0.7992 (0.6012)	-0.3518 (-0.3897)
$\ln Patent$	1.4883*** (16.6553)	1.3756*** (13.6000)	1.3860*** (13.8985)	1.4208*** (15.8610)	4.1664*** (3.7229)	3.8085*** (4.7516)	4.6454*** (3.6959)	3.7629*** (4.7940)
$\ln FMA$	-0.6762*** (-12.4160)	-0.6632*** (-10.8229)	-0.7186*** (-11.0518)	-0.6070*** (-11.9133)	-0.8236** (-2.0785)	-0.7493** (-2.5360)	-0.9397** (-2.1486)	-0.7735** (-2.5680)
$EG$	0.1365 (0.5418)	1.1209*** (5.2341)	1.0781*** (4.9188)	0.2496 (1.0141)	3.5926* (1.7837)	2.2823 (0.9825)	3.0767 (1.1199)	3.6794** (2.2670)
$EG_{up}$	1.1191*** (3.7655)	0.1949 (0.5544)	1.0580*** (3.3899)	0.2294 (0.6805)	-0.0761 (-0.0217)	1.5411 (0.6098)	0.0274 (0.0080)	1.4496 (0.6090)
$EG_{down}$	0.7215 (1.6241)	0.4513 (1.0717)	0.6867 (1.5882)	0.8424** (2.0058)	3.1997 (1.0629)	3.2775 (1.1524)	2.8522 (0.8112)	2.4140 (0.7898)
$Cluster \times TFP_{op}$	0.2706*** (8.6080)			0.2424*** (8.6389)	-0.4744*** (-3.0803)			-0.4542*** (-2.9208)
$Cluster_{up} \times TFP_{op}$		0.1575*** (8.5658)		0.1444*** (8.1851)		-0.2056 (-1.2727)		-0.2103 (-1.5371)
$Cluster_{down} \times TFP_{op}$			0.0920*** (3.3377)	0.0309 (1.1452)			-0.3036 (-1.5850)	-0.1213 (-0.8160)

续表 5

Pro	除纯加工贸易型企业外的样本				纯加工贸易型企业样本			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
type equation								
<i>east Trade</i>	1.3066*** (7.1317)	1.3133*** (7.1707)	1.3189*** (7.2012)	1.3007*** (7.1019)				
<i>middle Trade</i>	0(基准组)							
<i>west Trade</i>	0.1087 (0.3060)	0.1240 (0.3492)	0.1327 (0.3737)	0.0939 (0.2646)				
不相似系数								
<i>east_tau</i>	1.7892*** (16.9130)	1.8035*** (13.8033)	1.8172*** (13.9345)	1.7317*** (16.5501)	2.3516*** (3.4906)	1.9772*** (4.3933)	2.4987*** (3.4775)	2.0016*** (4.4033)
<i>middle_tau</i>	2.2415*** (13.8942)	2.2554*** (11.0821)	2.2823*** (11.2520)	2.1362*** (13.3555)	3.9733*** (3.1177)	3.3997*** (3.8590)	4.3549*** (3.1320)	3.4345*** (3.8722)
<i>west_tau</i>	1.4895*** (9.3402)	1.6324*** (7.6628)	1.6171*** (7.6915)	1.4431*** (8.9186)	2.2493** (2.1650)	1.8498** (2.4480)	2.4587** (2.2827)	1.9415** (2.5246)
观测值	166189	166189	166189	166189	7392	7392	7392	7392
Log likelihood	-14926.147	-14907.256	-14966.735	-14865.247	-425.948	-421.738	-425.308	-420.190
Wald chi2	432.47	300.36	306.71	417.87	27.11	41.63	27.97	39.70
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LR test for IIA	409.52	342.73	371.65	357.83	29.24	29.78	31.72	29.86
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

(三) 稳健性检验<sup>①</sup>

1. TFP 不同衡量方法下的稳健性检验。表 6 为基于 LP 方法衡量的 TFP 的检验结果,总的来看,各回归系数的符号和显著性基本没有太大的变化,说明前文的结论具有稳健性。

2. 不同区域分组下的稳健性检验。考虑到东部地区的经济发展水平明显优于中、西部地区,中、西部地区的差异性并不明显,本文再将中国分为东部和非东部两个地区进行稳健性检验。从表 7 的回归结果仍然可以发现,省区分类方式的调整并没有影响前文的相关结论,说明本文所得结论是非常稳健的。

① 受篇幅所限,以下的稳健性检验只报告了核心变量和主要的统计量。

表 6 基于 LP 方法衡量的 TFP 的稳健性检验结果

Pro	全样本				分样本	
					除纯加工贸易型企业外的样本	纯加工贸易型企业样本
	(1)	(2)	(3)	(4)	(5)	(6)
<i>EG</i>	0.1497 (0.5802)	1.1225 *** (5.1932)	1.0704 *** (4.8537)	0.2532 (1.0031)	0.2122 (0.8386)	0.3036 (0.2751)
<i>EG<sub>up</sub></i>	1.1502 *** (3.8362)	0.1936 (0.5514)	1.0632 *** (3.4207)	0.2346 (0.6869)	0.1188 (0.3390)	0.9330 (0.7198)
<i>EG<sub>down</sub></i>	0.7872 * (1.7623)	0.4910 (1.1726)	0.7737 * (1.8136)	0.9480 ** (2.2583)	0.9367 ** (2.1501)	1.0086 (0.9731)
<i>Cluster × TFP<sub>lp</sub></i>	0.1593 *** (8.5974)			0.1437 *** (8.5060)	0.1637 *** (8.9578)	-0.1097 ** (-1.9866)
<i>Cluster<sub>up</sub> × TFP<sub>lp</sub></i>		0.0905 *** (8.1566)		0.0837 *** (7.8251)	0.0918 *** (8.2100)	-0.1213 ** (-2.5174)
<i>Cluster<sub>down</sub> × TFP<sub>lp</sub></i>			0.0585 *** (3.6927)	0.0281 * (1.8027)	0.0296 * (1.6670)	-0.0230 (-0.4250)
观测值	173581	173581	173581	173581	166189	7392
Log likelihood	-15411.044	-15390.597	-15445.329	-15351.621	-14901.853	-1067.3787

表 7 基于两大区域分层的稳健性检验结果

Pro	全样本				分样本	
					除纯加工贸易型企业外的样本	纯加工贸易型企业样本
	(1)	(2)	(3)	(4)	(5)	(6)
<i>EG</i>	0.8059 ** (2.5653)	1.7364 *** (6.4153)	1.6324 *** (6.2514)	0.9192 *** (3.0601)	0.9083 *** (3.0071)	-0.4718 (-0.5770)
<i>EG<sub>up</sub></i>	1.8024 *** (5.4556)	1.0631 *** (2.7727)	1.8184 *** (5.2570)	0.9852 *** (2.6480)	0.8683 ** (2.3029)	2.1749 ** (2.0208)
<i>EG<sub>down</sub></i>	0.2814 (0.4687)	0.0719 (0.1251)	0.3432 (0.6004)	0.3911 (0.6632)	0.3590 (0.6049)	1.3182 (1.3744)
<i>Cluster × TFP<sub>op</sub></i>	0.3087 *** (8.4437)			0.2759 *** (8.2444)	0.3014 *** (8.4677)	-0.1926 * (-1.8906)
<i>Cluster<sub>up</sub> × TFP<sub>op</sub></i>		0.1336 *** (6.8763)		0.1280 *** (6.6597)	0.1356 *** (6.8659)	-0.0112 (-0.1102)
<i>Cluster<sub>down</sub> × TFP<sub>op</sub></i>			0.1288 *** (4.2489)	0.0666 ** (2.2365)	0.0685 ** (2.0593)	-0.2006 ** (-2.3282)
观测值	173581	173581	173581	173581	166189	7392
Log likelihood	-15592.186	-15549.264	-15612.029	-15519.733	-15014.636	-1083.727

## 五、结论及政策启示

本文基于企业异质性假定的“新”新经济地理学框架,研究了企业区位自选择效应在中国是否存在,即高生产率的企业是否一定会选择在集聚程度较高的中心区;在考虑集聚的形式上,通过使用投入产出表刻画出包含上、下游关联产业的多样化集聚,并与专业化集聚进行了对比性验证;此外还考查了纯加工贸易型企业是否存在区位自选择效应的问题。本文主要得出的结论是,在专业化集聚程度较高的地区,正如“新”新经济地理学理论所分析的那样,存在显著的区位自选择效应,即生产率较高的工业制造业企业将倾向选择在企业集聚程度较高的中心地区,而生产率相对较低的工业制造业企业则会选择在集聚程度较低的外围地区。但是与专业化集聚相比,在单纯的上、下游产业集聚程度的中心地区,这种区位自选择效应则相对较弱。主要原因在于,在同行业企业集聚程度较高的中心地区,企业之间的竞争效应和拥挤效应更大,只有生产率相对较高的企业,才能够克服中心地区激烈的市场竞争带来的巨大压力和较高的拥挤成本。而在单纯的上、下游企业集聚较高的中心地区,尽管也会导致地区要素成本的上升,但是企业之间的协作更重要,竞争则并不是那么的针锋相对,所以企业区位自选择效应体现的并不是特别强。另外,本文还发现,对于大多数的纯加工贸易型企业,尽管它们的生产率相对较低,但是鉴于其具有明显的外向型特征,它们往往倾向选择在具有地理优势且产业集聚程度较高的东部沿海地区,以克服高额的运输成本。因此,对于这种特殊类型的企业,区位自选择效应并不明显。这说明尤其是在东部沿海中心地区,尽管企业的平均生产率高于外围地区,但这些地区企业生产率的分布是比较分散的。本文的结论,针对不同的 *TFP* 衡量方法以及不同的区域分类方法都是非常稳健的。

由此可以发现,与外围地区相比,尤其是那些专业化集聚程度较高的中心地区,不仅存在地区经济“量”的差距,地区企业的生产效率还存在“质”的差距,这在某种程度上加剧了中国地区经济发展的不平衡。从地区分布看,中心地区主要分布于中国的东部沿海地区,外围地区则主要分布于中部、西部地区,究其原因主要在于,中国贸易渠道仍主要依赖于海上通道,相对比较单一;且中国对外贸易的伙伴主要是美国、日本、韩国等位于东南方向的国家,东部沿海地区凭借其优越的地理位置成了产业聚集地和中国主要的进出口地;而中部、西部地区交通、通信等基础设施水平较东部地区仍有较大差距,工业的自我发展能力和产业优势难以得到发挥。

为解决这一矛盾,引导和促进地区间“松脚型”产业的流动无疑是非常重要的。首先,对于东部沿海等产业集聚程度较高的中心地区,从长远来讲,生产要素价格的提高和企业竞争的加剧是必然的趋势,未来产业的转型升级是这些地区的重中之重。这些地区一方面应提升技术、人力资本等要素的积累,可以考虑重点发展战略型新兴产业、高技术产业,同时加快服务业,特别是金融业的发展;另一方面,在“一带一路”倡议下,大范围的区域协同合作将成为重要趋势,新市场的开拓,东部、中部、西部地区联动性的增强也将为东部地区产业的腾笼换鸟做好充分的准备,尤其为那些生产率较低的加工贸易型企业的转型或是区位转移创造有利条件。其次,对于那些产业集聚程度较低的中西部地区,一方面,“一带一路”倡议的提出将打开面向西北的中亚、西亚乃至欧洲的开放大门,改变地理的空间结构,使内陆地区有相当于过去沿海、沿边的条件,从而改变开放的约束条件,使开放型经济在内陆地区的实施成为可能,从而使中部、西部地区更贴近国际市场。这些地区要充分利用好“一带一路”倡议大力推进的有利契机,结合地区产业发展特色,拓展地区内制造业上、下游产业链和价值链的搭建和完善,加强产业协作,整合延伸产业链条,推进产业链上、下

游深度合作,逐渐培育形成优势互补、分工合理、布局优化的先进产业集群。另一方面,要加快本地区基础设施建设,构筑强大的铁路、公路、机场以及产业园区的强大网络,这有助于扩大产品供应和服务半径,降低物流等运输成本,为西部地区更好融入“一带一路”提供强大的支持。与此同时,“一带一路”倡议的推进主要以基础设施互联互通为先导和突破口,中部、西部地区可以以此为契机,通过打开路上通道,增强沿线经济带和城市群的连接性,形成横贯东中西、联结南北方的对外经济走廊,为东中西部地区间产业转移提供通道,并为要素集聚和产业发展创造良好条件,这些都是缩小中国地区间经济发展“量”和“质”的差距的可行途径。

参考文献:

1. 戴觅,余森杰,Maitra:《中国出口企业生产率之谜:加工贸易企业的作用》,北京大学中国经济研究中心讨论稿,2011年。
2. 黄玖立,李坤望:《对外贸易、地方保护和我国的产业布局》,《经济学(季刊)》2006年第3期。
3. 黄肖琦,柴敏:《新经济地理学视角下的FDI区位选择:基于我国省际面板数据的实证分析》,《管理世界》2006年第10期。
4. 梁琦,李晓萍,简泽:《异质性企业的空间选择与地区生产率差距研究》,《统计研究》2013年第6期。
5. 聂辉华,江艇,杨汝岱:《中国工业企业数据库的使用现状和潜在问题》,《世界经济》2012年第5期。
6. 田巍,余森杰:《企业出口强度与进口中间品贸易自由化:来自中国企业的实证研究》,《管理世界》2013年第1期。
7. 王永进,张国峰:《开发区生产率优势的来源:集聚效应还是选择效应?》,《经济研究》2016年第7期。
8. 余佩,陈继勇:《新经济地理学框架下跨国公司在华分层区位选择研究》,《世界经济》2012年第11期。
9. 余壮雄,杨扬:《大城市的生产率优势:集聚与选择》,《世界经济》2014年第10期。
10. Baldwin, R. E., & Okubo, T., Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting. *Journal of Economic Geography*, No. 6, 2006, pp. 323–346.
11. Behrens, K., Duranton, G., & Robert-Nicoud, F., Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, Vol. 122, No. 3, 2014, pp. 507–553.
12. Brandt, L., Biesebroeck, J. V., & Zhang, Y., Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*, No. 97, 2012, pp. 339–351.
13. Combes, P. P., Duranton, G., Gobillon, L., Puga, D., & Roux, S., The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica*, No. 6, 2012, pp. 2543–2594.
14. Crozet, M., Mayer, T., & Mucchielli, J. L., How Do Firms Agglomerate? A Study of FDI in France. *Regional Science and Urban Economics*, No. 34, 2004, pp. 27–54.
15. Ellison, G., & Glaeser, E., Geographic Concentration in U. S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, No. 105, 1997, pp. 889–927.
16. Harris, C. D., The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, Vol. 44, No. 4, 1954, pp. 315–348.
17. Javorcik, B. S., Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages. *American Economic Review*, Vol. 94, No. 3, 2004, pp. 605–627.
18. Levinsohn, J., & Petrin, A., Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, No. 70, 2003, pp. 317–342.
19. McFadden, D. L., Econometric Model for Probabilistic Choice. *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge: MIT Press, 1981.
20. McFadden, D. L., Econometric Analysis of Qualitative Response Model. *Handbook of Econometrics*, Vol. 1, No. 2, 1984, pp. 1395–1457.
21. Melitz, M. J., & Ottaviano, G. I. P., Market Size, Trade and Productivity. *Review of Economic Studies*, No. 75, 2008, pp. 295–316.
22. Okubo, T., & Forslid, R., Spatial Relocation with Heterogeneous Firms and Heterogeneous Sectors. *Research Institute of Economy, Trade and Industry*, 2010.

23. Okubo, T. , & Tomiura, E. , Productivity Distribution, Firm Heterogeneity and Agglomeration: Evidence from Firm-Level Data, Research Institute for Economics & Business Administration, Kobe University, 2011.
24. Okubo, T. , & Tomiura, E. , Skew Productivity Distributions and Agglomeration: Evidence from Plant-Level Data, Keio/Kyoto Joint Global COE Discussion Paper Series, 2012.
25. Olley, G. S. , & Pakes, A. , The Dynamics of Productivity in the Telecommunication Equipment Industry. *Econometrica*, No. 64, 1996, pp. 1263 – 1297.
26. Redding, S. , & Venables, A. J. , Economic Geography and International Inequality. *Journal of International Economics*, No. 62, 2004, pp. 53 – 82.
27. Stiebale, J. , Do Financial Constraints Matter for Foreign Market Entry? A Firm-Level Examination. *The World Economy*, Vol. 34, No. 1, 2011, pp. 123 – 153.
28. Venables, A. J. , Equilibrium Location of Vertically Linked Industries. *International Economic Review*, No. 37, 1996, pp. 341 – 359.

## Geographical Cluster and the Self-Selection Effect of the Firms: Comparative Analysis on Cluster of Upstream and Downstream Industries and the Specialized Cluster

LI Ruiqin (Central University of Finance and Economics, 100081)

SUN Puyang (Nankai University, 300071)

**Abstract:** This paper examines the self-selection effect in spatial selection by the heterogeneous firms, that is, whether the Chinese industrial firms with high productivity are inclined to select the core zone. We not only consider the traditional specialized cluster within the same industry, but also analyze the diversified cluster based on the upstream and downstream relationship. We also study the difference of the self-selection effect under these two kinds of clusters through empirical analysis based on the nested logit model. Our main conclusion is that more efficient firms are inclined to select the core zone of the specialized cluster, while the self-selection effect is weak when there is just the cluster of the upstream and downstream industries. Furthermore, we also find that processing trade firms pay more attention to the superior geographical position. Therefore, in the core areas with geographical advantages, the distribution of firm productivity is more dispersed, because the processing trade firms have low productivity. The self-selection effect is not obvious in processing trade firms. In light of this, while the core area is accelerating the transformation and upgrading of traditional industries, it is necessary to create favorable conditions for the location transfer of enterprises with low productivity. The periphery districts should attach importance to the openness, the infrastructure and the establishment of the upstream and downstream industrial chain in order to narrow the gap with the core zone.

**Keywords:** Heterogeneous Firms, Self-Selection Effect, Specialized Cluster, Cluster of the Upstream and Downstream

**JEL:** L22, F14

责任编辑:鲁 洲