

DOI: 10.16538/j.cnki.fem.2018.03.010

# 工业企业数据库在企业经济行为研究中的应用

## ——视角、合并与拓展

肖利平<sup>1,2</sup>, 许巍峰<sup>2</sup>

(1. 武汉大学 经济发展研究中心, 湖北 武汉 430072; 2. 武汉大学 经济与管理学院, 湖北 武汉 430072)

**摘要:** 中国工业企业数据库已成为微观企业行为研究中最重要数据来源之一。本文基于该数据库的数据结构系统性地梳理了该数据库在企业经济行为方面的最新应用, 并整理了基于该数据库与其他微观数据库合并数据的研究, 分析发现: 基于该数据库的研究主要集中在企业生产率、创新、出口等传统领域, 但也开始关注企业出口持续、企业内工资差距以及生存能力等新视角; 许多文献开始基于合并数据进行拓展研究, 这类研究更多关注企业出口二元边际、出口产品质量、对外直接投资等问题。同时, 本文梳理了该数据库2007年后的数据情况, 指明后续年份数据的使用问题、解决方法及应用方向。在数据处理方面, 现有研究存在样本过宽匹配、平减指数混乱、出口企业识别不严谨等新问题。在研究视角方面, 今后可更多地着眼于企业生产率的性别差异、创新风险、研发平滑以及企业内工资差距等问题。

**关键词:** 工业企业数据库; 企业经济行为; 研究视角; 数据处理

**中图分类号:** F270 **文献标识码:** A **文章编号:** 1001-4950(2018)03-0137-16

### 一、引言

微观企业数据在企业行为研究中具有举足轻重的地位, 数据量的多少、数据质量的好坏直接影响这个研究所能达到的高度。在众多微观企业数据库中, 中国工业企业数据库因其在样本数量、统计指标等方面的显著优势而得到了广泛应用。基于规模以上工业统计口径的工业企业数据库始于1998年, 由国家统计局根据企业提交的季报和年报建立, 覆盖了39个二位数行业的全部国有及主营收入在500万元以上的非国有工业企业<sup>①</sup>, 以制造业为主, 也包括“采掘业”及

**收稿日期:** 2017-08-02

**基金项目:** 教育部人文社会科学重点研究基地重大项目“后发展经济学理论研究”(14JJD790009); 国家社会科学基金青年项目“基于地区差异的追赶绩效理论模型与经验分析研究”(12CJL037); 武汉大学马克思主义指导团队项目

**作者简介:** 肖利平(1977—), 女, 武汉大学经济发展研究中心、经济与管理学院副教授(通讯作者); 许巍峰(1992—), 男, 武汉大学经济与管理学院, 硕士研究生。

<sup>①</sup>关于工业企业数据库基本信息的介绍可详见《中国工业企业数据库的使用现状和潜在问题》(聂辉华等, 2012), 本文不再赘述。

“电力、燃气及水的生产和供应业”。该数据库的主要优势在于：(1)样本数量大。样本企业数目从1998年的16万余家涨至2013年的34万余家，16年观测值达435余万个。(2)样本覆盖全。根据2004年第一次全国经济普查年报数据，工业企业数据库当年全部样本企业的主营业务收入约占全国工业企业主营业务收入的91.06%<sup>①</sup>，可较充分地反映我国工业企业的整体发展状况。(3)统计指标多。数据库涵盖上百项统计指标，包括行业代码、工业增加值、新产品产值等企业基本信息和财务信息。(4)时间跨度长。数据库持续更新，目前已更新至2013年<sup>②</sup>，形成了完整的面板数据，便于观测企业行为的动态变化。(5)较易于合并。数据库提供了企业代码、名称、电话及所在地等信息，通过将其与其他微观数据库合并，可以进一步拓展研究的视角。

鉴于工业企业数据库的上述优势，利用工业企业数据库的研究文献近年呈不断增长趋势，至今基于该数据库的各种文献已达1 300余篇，尤其是自2012年以来，每年新增近百篇中英文文献。其中，外文文献达200余篇，刊载于*American Economic Review*等重要期刊<sup>③</sup>。可见，目前学界对工业企业数据库的潜在利用价值形成了一定的共识，并产出了一批高质量的研究成果。其中，许多文献聚焦于企业TFP，例如Heiseh和Klenow(2009)利用1998—2005年工业企业数据库数据，研究资源错配对中国制造业企业TFP的影响程度；杨汝岱(2015)利用1998—2009年数据测算并分解制造业企业TFP，探究其增长来源。另一些文献则关注企业出口、避税等问题，例如Liu和Lu(2015)利用1998—2007年数据，识别企业投资与出口行为的因果关系，揭示其内在机制；Cai和Liu(2009)利用2000—2005年数据，探究市场竞争对中国工业企业避税行为的影响；Gao和Van Biesebroeck(2014)利用1998—2007年数据库中的发电部门企业数据，考察中国电力部门改革对企业劳动力雇佣以及中间品投入使用的影响。除了上述代表性研究，利用该数据库展开的企业经济行为研究也出现了诸如企业内工资不平等、创新活动的高风险等新视角。因此，有必要全面梳理新近的研究成果。当然，工业企业数据库的应用范围不只局限于微观企业行为的研究，其也被加总处理后用于研究产业、地区层面的问题。由于这类研究在一定程度上丧失了该数据库作为微观数据库的优势，同时为了区别于企业管理行为，本文将主要关注其在微观企业经济行为研究中的典型应用，整理相关文献的研究视角及其与其他微观数据库的合并运用。本文的系统整理，可能让未来研究者能够更全面、深入地认识该数据既有的使用视角与未来可能的拓展方向。

工业企业数据库的应用日益广泛，该数据库本身存在的各种问题也不断凸显，一些研究开始关注数据库的使用现状与问题。聂辉华等(2012)较为全面地介绍数据库的基本情况与使用现状，重点探讨了数据库中存在的样本匹配、指标缺失、测度误差等问题。Brandt等(2012, 2014)则对数据库的处理方法进行了详细说明与整理，他们提出的方法已成为处理该数据库的代表性方法。但是，随着利用该数据库的文献日益丰富，人们在使用工业企业数据库的过程中对该数据库产生了一些新的认识，也遇到了一些新的问题，例如样本精确匹配、平减指数构造等。此外，随着许多领域研究的不断深入，单以该数据库数据难以进一步开展研究，越来越多的文献开始将该数据库与其他数据库进行合并研究，但目前还未见文献系统整理近年这些基于合并数据的研究。因此，本文希望通过甄别这些新问题及可能的解决思路，对聂辉华等(2012)和Brandt等(2012, 2014)在数据库问题和数据处理方面做出一些有益的补充。

①经济普查中主营收入数据来自中国统计局网站《第一次全国经济普查主要数据公报(第二号)》，工业企业数据库中主营收入来自笔者计算。

②虽然数据库现已更新至2013年，但现有研究多使用1998-2007年数据，主要因为工业企业数据库2008年后数据存在样本大量遗漏、重要指标缺失、统计口径调整等问题。详见本文“四、数据库样本时间范围的拓展”。

③利用工业企业数据库的重要外文期刊文献，代表性的有Song等(2011, *American Economic Review*)、Hsieh和Klenow(2009, *Quarterly Journal of Economics*)、Yu(2015, *The Economic Journal*)、Manova和Yu(2016, *Journal of International Economics*)。

基于工业企业数据库在微观企业经济行为中的相关研究,本文揭示了工业企业数据库的主要研究视角、拓展应用及使用中存在的一些问题,并提出了一些未来进一步研究的新方向。相比于现有文献,本文可能的贡献在于:(1)基于工业企业数据库的数据结构与企业经济行为研究范畴的匹配,整理了近年基于该数据库在企业经济行为研究领域内的主要文献,发现当前研究不仅关注企业生产率、出口决策与规模等传统领域,也将目光投向了诸如企业出口持续、企业内工资差距、生存能力等新视角;(2)指出工业企业数据库存在的应用局限,并首次梳理基于工业企业数据库与其他微观数据库的合并数据研究,发现此类研究更多关注企业出口二元边际、产品质量以及对外直接投资等问题,这种合并数据研究将是未来研究的一个重要方向;(3)首次整理了工业企业数据库2007年后的数据情况,系统性地提出现有较多文献仅利用2007年前数据的主要原因及应用后续年份数据所需解决的问题及方案,并指出新数据在未来研究中所能应用的方向;(4)从数据整理、变量测度以及变量初加工三个方面,指出了一些现有文献在数据处理方面的新问题,包括样本匹配、指标测度、名义变量平减、出口企业识别等,并尝试提出可行的解决方案;(5)提出了未来进一步研究的一些新思路、新视角,认为研究者可更多着眼于企业生产率的性别差异、创新风险、研发平滑以及企业内工资差距等问题。

## 二、数据库的应用视角

企业经济行为一直是微观企业数据库的研究焦点,其涵盖了企业的“求利”和“求德”行为,即企业经济行为不仅包含企业的投融资、要素投入、生产创新、贸易以及产品销售策略等生产经营行为,也包含企业的商业信用、社会责任等社会行为。工业企业数据库提供了大量指标数据,使得全方位地研究企业经济行为成为了可能。

与CSMAR中国股票市场交易数据库、世界银行的中国企业调查数据等其他微观企业数据库不同,工业企业数据库涵盖了全部国有工业企业和大中型非国有工业企业,样本量大,且拥有自1998年起逐年连续的工业企业数据,能够用来不间断地观测我国制造业的动态发展。同时,该数据库提供了大量企业研发、收入、学历构成等多品类的数据,有助于深入研究企业内部的具体问题。

表1显示了工业企业数据库的主要指标,包括一些企业识别信息、资产负债存量指标、投入产出流量指标,以及一些其他重要指标。我们既可以基于销售收入或工业增加值、职工数、固定资产等指标测算企业生产率,并探究企业内外部不同因素对其生产行为的影响,也能够基于出口交货值,研发投入、新产品产值、无形资产等数据研究企业的出口贸易行为和创新行为。囿于篇幅以及现有文献较多关注的视角,并兼顾企业生产特征,本文将主要从企业的生产行为出发,依次梳理企业的生产行为、创新行为、贸易行为、雇佣行为和企业的生存与成长。研究发现,工业企业数据库在微观企业经济行为领域中的运用,早期研究偏重于企业生产率、资源错配以及对熊彼特创新理论、新新贸易理论的检验等领域,近年研究除了继续关注上述问题外,更多探讨企业创新水平、出口持续、薪酬水平、工资差距、生存能力等问题。

### (一)企业的生产行为

在目前利用工业企业数据库的企业生产行为相关研究中,企业生产率最受关注,许多文献利用该数据库测算企业的生产率并考察企业内、外部不同因素对其生产率的影响。我国现行的制造业口径与国际产业分类具有一定的可比性,因此基于该数据库的研究结果可以和其他研究很好地进行横向比较。

企业生产率的准确衡量是这类研究的重点,目前主要有单要素生产率、全要素生产率(TFP)两种方法。相比而言,全要素生产率更能体现企业的技术水平以及生产过程中涉及的知

表1 中国工业企业数据库指标结构

类型	子 类	基本指标 <sup>①</sup>	应用视角
识别信息	名称地址信息	法人代码,企业名称,地区代码,街道、居委会,成立时间,电话、传真、邮箱	生存能力
	基本经营信息	注册类型,国有控股情况,行业类别,产业活动数、主要产品,营业状态,隶属关系	生产行为
存量指标	资产类	总资产,流动资产(总计、年平均余额),固定资产(总计、净值、原价),无形资产,存货,应收账款	创新行为,商业信用
	负债类	总负债,流动负债,长期负债	负债结构,商业信用
	权益类	实收资本(国有、集体、外资、港澳台、法人、个人资本金),所有者权益	股权结构,融资渠道
流量指标	投入类	职工数,应付工资、福利,财产保险费、养老保险费、医疗保险费、失业保险费、住房公积金,职工教育费	雇佣行为,社会责任
		投入(营业投入、制造投入、管理投入、中间投入),费用(期间费用、广告费等)	生产行为
		投资(长期、短期),利息支出,折旧(本年、累计),研发投入	投融资行为,创新行为
	产出类	收入(总收入、主营业务收入、营业外收入、其他收入、产品销售收入),产值(工业总产值、新产品产值、工业销售产值)、工业增加值,出口交货值,利润(总利润、营业利润、产品销售利润、其他业务利润)	生产行为,经营绩效,创新行为,贸易行为
税收与补贴	所得税、增值税,进项税额、销项税额,销售税额,政府补贴	税收负担,避税行为,政府补贴	
其他重要指标	仅有2004年数据	职工数的分解指标(总计、女性)—教育水平:初中以下、高中、专科、本科、研究生—技能水平:技师、高级技师,中级工、高级工—职称:低级、中级、高级职称	雇佣行为
		特殊支出(差旅费、排污费),筹资、投资、经营现金(总计、流入、流出)	投融资行为
		集团母公司、母公司代码	集团化经营
		是否有工会、工会人数、工会经费,计算机数、微机数	雇佣行为,创新行为

资料来源:基于1998—2013年中国工业企业数据库整理。

识、制度、管理等方面的综合能力(鲁晓东和连玉君,2012)。TFP的测算,主要分为参数法、非参数法以及半参数法三类方法。基于工业企业数据库的早期文献多侧重于用参数或非参数方法测算企业TFP,例如谢千里等(2008)发现1998—2005年企业TFP年均增长率达到10.26%,Brandt等(2012)利用非参数的Tornqvist指数法测算,发现1998—2007年我国企业TFP年均增长率为7.96%。在不断总结和比较各种不同估算方法后,半参数方法由于能很好地解决传统方法中存在的内生性和样本选择问题而备受重视(鲁晓东和连玉君,2012),例如杨汝岱(2015)就基于半参数的OP、LP方法,发现1998—2009年我国制造业企业TFP增速在2%—6%之间,年均增长3.83%。

在准确测算生产率的基础上,许多文献致力于探讨生产率的影响因素,并将焦点集中在企业出口决策、市场扭曲及经济集聚等方面。(1)出口对企业生产率的影响。新新贸易理论认为出口企业生产率高于内销企业,对此,许多文献利用工业企业数据库进行了检验,有些研究验证了这一假说,而另一些研究则反对这一观点,认为我国制造业出口企业存在“生产率悖论”。前一类文献认为出口企业生产率更高,并强调出口学习效应对企业生产率的提升作用(Park等,

①表1所涵盖的指标并未覆盖工业企业数据库全部指标,且表中所有指标并非1998—2013年持续存在。

2010),尤其是对处于中高技术产业的新出口企业(Du等,2012)以及出口前有持续研发投入的企业(Dai和Yu,2013)。后一类研究认为我国内销企业生产率更高(Lu等,2010),原因可能在于我国出口企业中存在大量低效率的加工贸易企业(李春顶等,2010)。近期文献也提出了一些导致这一悖论产生的其他原因,可能是因为存在大量低效率的纯出口企业(高宇,2014),也可能是由于多数研究未考虑到企业所在城市这一区位因素,从而无法控制地方保护、产业集聚的影响(Yang和He,2013)。(2)市场扭曲、资源错配对企业生产率的影响。要素市场扭曲(盖庆恩等,2015)与资源错配(聂辉华和贾瑞雪,2011)均会抑制企业生产率的提高。自Hsieh和Klenow(2009)开创性地提出用TFP的离散程度来衡量资源配置效率,考察资源错配同TFP之间的关系以来,多数文献基于他们的思路进行深化研究,例如龚关和胡关亮(2013)通过改进规模报酬不变的假设,并用投入要素边际产出的离散程度衡量资源配置效率,发现如果资本与劳动要素有效配置,1998年中国制造业TFP会上升57.1%,2007年会上升30.1%,远小于Hsieh和Klenow(2009)测算的结果。也有研究者利用加成率异质性衡量资源配置效率,并考虑市场扭曲对企业行为的影响,发现如果消除资本市场扭曲,制造业企业TFP会提高57.79%,其中31.46%来自于资源有效配置,26.32%来自于企业更替,如果消除劳动市场扭曲,企业TFP会提高33.12%,其中资源有效配置贡献了11.42%(盖庆恩等,2015)。(3)经济集聚对企业生产率的影响。企业可以从本行业集群中受益(地方化经济),也能从多行业集聚中获益(城市化经济),已有文献利用工业企业数据库数据,发现地方化经济会通过技术效率改善来促进企业TFP增长(范剑勇等,2014),且地方化经济对生产率的促进作用会随着企业雇员受教育水平与中间投入比重的提高而增强,而城市化经济则无显著影响(Hashiguchi和Tanaka,2014)。此外,经济集聚也会产生规模效应与拥挤效应,有文献证实了我国主要工业城市中拥挤效应的存在,并认为自2003年起,经济集聚对制造业企业TFP主要表现为拥挤效应(周圣强和朱卫平,2013),对我国整体企业TFP产生负向影响,但在西部地区仍表现为集聚效应(李晓萍等,2015)。

除了上述主要因素,在利用工业企业数据库的研究中,许多其他影响企业生产率的因素也受到了关注,这不仅包括传统的研发投入、外商直接投资(FDI)、融资约束等因素,如一些研究认为FDI会通过内资企业从上游外资企业购买高质量中间品及设备来提升内资企业生产率(Xu和Sheng,2011),企业内部资金也会影响企业生产率的提高,特别是对流动性不足的企业,充足的现金流有助于企业投资于诸如高风险、高不确定性的研发活动等有利于生产率提升的项目(Chen和Guariglia,2011),也包括近来颇受关注的企业更替、企业收购等新因素,例如一些研究认为企业更替会通过市场竞争效应提升存活企业生产率(毛其淋和盛斌,2013),且来自于OECD国家的收购会对企业TFP的增长产生更大、更持久的促进作用(Kamal,2015)。

## (二)企业的创新行为

企业创新行为研究,主要从创新投入强度和创新能力出发,探讨不同因素对它们的影响。企业创新投入可以用工业企业数据库中的研发支出衡量。早期研究主要是检验熊彼特假说,探究市场竞争、企业规模与创新的关系,但并未得出一致结论(Jefferson等,2006;聂辉华等,2008)。近年来,也有文献重新检验该假说,但更多文献从金融发展、企业技术能力、出口参与等其他视角探讨企业创新的影响因素,发现企业自身的技术能力、出口参与以及资产流动性会分别通过技术推动、需求拉动以及充足企业内部资金来促进企业研发(周亚虹等,2012),而金融体系的“所有制歧视”“规模歧视”和金融发展的滞后则会通过影响企业融资约束以及融资渠道的选择对集体和私人性质的民营企业研发投入产生负向效应(张杰等,2012)。

另一方面,企业创新产出能够反映创新投入的绩效和企业的创新能力。许多文献用新产品产值来衡量企业创新产出,较多关注集聚经济和政策环境的影响,认为集聚经济会显著影响企

业产品创新,但在企业生命周期的不同阶段,不同类型的聚集经济起主导作用,年轻企业创新更多受益于城市化经济,而成熟企业创新更多受益于地方化经济(董晓芳和袁燕,2014),不过也有文献认为只有城市化经济对中国企业创新具有促进作用,而地方化经济没有显著正效应(Zhang,2015)。另一些研究则发现,2004年的增值税改革会通过降低企业税负来激励企业产品创新(Howell,2016),政府补贴政策也有利于企业创新,尤其是对民营企业的创新绩效(杨洋等,2015),但可能只有适度的补贴才具有促进效应,高额补贴反而会因企业“寻补贴”投资而抑制企业产品创新(毛其淋和许家云,2015)。

此外,还有文献同时运用创新投入、产出来更全面地刻画企业创新,例如Fang和Ge(2012)利用工业企业数据库中的工会相关数据,发现中国企业工会会通过“民主管理”“职工代表大会制度”等途径来与管理层协作,实现生产率提升等共同目标,有利于企业创新投入与产出。

### (三)企业的贸易行为

我国作为出口贸易大国,如何保障出口平稳增长一直是国内外文献讨论的焦点。基于工业企业数据库,研究者可根据出口交货值识别出口企业并进行研究,目前主要的研究集中在出口参与和出口规模两方面。企业出口参与会受企业内、外部双重因素的影响。其中,外部因素包括政策环境、地理集聚等,比如地理集聚会对企业出口参与产生激励作用,但地理集聚通过融资渠道对企业出口参与的激励作用也许仅存在于民营企业(佟家栋和刘竹青,2014),生产性补贴也会促进企业进行出口(苏振东等,2012),而最低工资标准和劳动力成本的提升则会通过削弱我国企业在国际市场上的比较优势来降低企业出口概率(Gan等,2016)。内部因素除了包含生产率、要素投入等基本因素外,还有融资约束、固定资产投资等其他企业自身特征,例如企业融资状况改善有助于企业承担进入进出口市场的固定成本从而提高企业进行出口的概率(阳佳余,2012),企业固定资产投资也会通过促进企业生产率的提高来持续地刺激企业出口参与(Liu和Lu,2015)。

企业出口规模能够反映企业出口参与的绩效,因此也吸引了许多研究的关注。基于工业企业数据库的出口规模影响因素研究多用出口交货值来表示企业出口规模,其影响因素也涉及企业内、外部双重因素。外部因素包括产业政策、劳动力成本等,例如提高增值税退税率有助于弥补增值税退还制度不完善的负面影响,促进企业出口,且退税率每增加1%,企业出口量会增加13%(Chandra和Long,2013),而劳动力成本和最低工资的提高则会对企业出口规模产生负向影响(Gan等,2016)。内部因素主要涉及企业融资条件、社会责任等企业自身特征,例如外商收购会由于显著改善目标企业的融资条件而促进企业出口(Wang和Wang,2015),企业承担社会责任也会通过提高企业在筹资、获得政府扶持等方面的便利性来促进企业出口(赵永亮和高颖欣,2012)。

此外,还有文献考察了企业出口持续时间的影响因素,发现最低工资上涨可能会通过逼迫企业提高固定资产投资来提升生产率,从而降低企业出口风险,正向促进企业出口持续(赵瑞丽等,2016),而融资约束恶化则会增加企业出口风险,不利于企业出口持续(李宏兵等,2016)。

### (四)企业的雇佣行为

企业的雇佣规模和薪酬水平是企业生产投入的重要因素,对就业与收入问题的探讨也是工业企业数据库的重要应用方向,现有文献主要围绕企业雇佣规模、薪酬水平与工资差距进行研究。围绕企业雇佣规模,近年文献发现政府补贴(张伯伟和沈得芳,2015)、外商收购(Wang和Wang,2015)和企业出口贸易增加(陈昊和刘骞文,2014)均会促进企业雇佣规模的扩张,且政府补贴能够提高企业女性员工的雇佣规模(张伯伟和沈得芳,2015),而企业出口规模的增加则因为随之增高的生产交易成本和工作强度使得企业不愿雇佣女性员工,从而不利于就业公平

(陈昊和刘骞文,2014)。也有研究聚焦于具体行业的就业问题,发现2002年的电力部门重组使得电力部门劳动雇佣数下降了7%,但这种作用存在滞后效应(Gao和Van Biesebroeck,2014)。

除了关注劳动力雇佣规模,很多文献也基于工资、福利费以及2004年的学历构成等数据,关注劳动力的薪酬差异及其提升路径。工资不平等,包括企业间和企业内两个研究维度。关于企业间工资差距,一些研究发现企业间生产率差距是企业间工资差距的主要原因(杨继东和江艇,2012),企业出口也会因为工资溢价现象而扩大出口与内销企业间的工资差距(李静和彭飞,2012),相反,外商直接投资则会通过技术溢出效应缩小内外资企业间的工资差距(周云波等,2015)。关于企业内的工资差距,涉及技术不平等、性别歧视等多个方面,一些研究用技术工人工资溢价来衡量工资差距,发现出口贸易会造成企业内技术与非技术工人间的工资差距,这种差距会随着企业出口强度的增加而扩大(陈波和贺超群,2013);市场经济下,劳动生产率的性别差距会在一定程度上造成企业内,尤其是高技术企业内工资的性别差距(陈国强和罗楚亮,2016)。关于提升薪酬的途径,有研究发现企业出口由于出口学习效应而能够提升员工薪酬,但这种提升作用会随着出口强度的增加先上升后下降,其可能原因在于我国存在大量低效率的加工贸易企业(史青,2013),同时,外商收购,尤其是来自于OECD国家的投资也能通过管理技术的引入来降低劳动的无效率从而提升企业平均工资(Wang和Wang,2015;Kamal,2015),而融资约束则会通过影响企业生产率和就业规模来降低员工收入(文磊等,2015)。

#### (五)企业的生存与成长

随着我国经济步入转型期,如何提高企业的生存能力及发展质效也成为近年研究的焦点。通过企业成立时间、消失时间以及营业状态等信息,研究者可对样本企业,尤其是国有企业进行久期分析。研究发现,一些因素有利于企业生存能力的提升,例如加强产权保护与判决执行效率等制度质量的提升能降低私营企业生存风险(Che等,2011),市场化程度深化等制度障碍的减少也有助于降低新成立企业生存风险,使之与在位企业间的生存风险差异不断减小(Chang和Wu,2014),企业内部的创新活动,尤其是研发投入也有助于延长企业存活时间(Zhang和Mohnen,2013)。另一些因素则不利于企业生存,如外资进入会加剧行业内竞争而增大同行业内资企业的生存风险(包群等,2015),甚至过多的创新活动也会由于带来过高风险而不利于企业存活(Howell,2015)。

与生存风险相对应的是,良好的经营绩效有助于企业的健康成长。一些文献基于工业总产值、工业增加值来研究企业的产出规模,发现市场潜力会扩大地方保护程度较高区域企业的产出规模,有利于企业成长(杨汝岱和朱诗娥,2015)。但由于产出指标无法直接体现企业当期销售获得收入的情况,多数文献选择基于企业利润、收入进行分析,发现出口企业比非出口企业利润率低,且出口强度与企业利润率成反比,其可能原因在于我国加工贸易企业一直处于产品价值链的底部,难以实现出口学习效应,且对出口市场的较大依赖也削弱了其议价能力(苏振东和洪玉娟,2013),国际市场潜力也会对低生产率企业的利润率产生不利影响,但国内市场潜力则有利于国内企业获取利润(Ouyang等,2015)。

显然,工业企业数据库还有很多值得研究的方向。已有文献围绕企业社保参与(Rickne,2013)、企业避税(Cai和Liu,2009)以及资产增长(Guariglia等,2011)等话题进行研究,囿于篇幅这里不再深入探讨。

### 三、同其他数据库的合并运用研究

虽然中国工业企业数据库指标众多、应用方向广泛,但仍然存在诸如出口、创新等品类指标单一而无法对企业出口方式、出口产品以及创新模式进行深入地刻画和研究,缺少对外直接

投资等数据而无法拓展研究边界等诸多局限。随着微观数据库数量的不断增加、种类的逐步丰富以及合并方法的日趋完善,近年许多研究者开始选择将工业企业数据库与其他微观数据库合并,以此丰富基于工业企业数据库的应用广度和深度。目前,中国海关进出口数据库和境外投资企业(机构)名录是与工业企业数据库合并最多的微观数据库,主要用于深化扩展对企业出口贸易行为的研究和企业对外直接投资行为的研究。

#### (一)企业出口贸易行为扩展研究:与海关进出口数据库合并

中国海关进出口数据库完整记录了进出口货物、企业、海关等月度信息。仿照Upward等(2013)的方法,研究者可借助企业名称、邮政编码及电话号码等识别信息匹配两种数据库,更准确地识别出口企业<sup>①</sup>,并构造出“企业产品—出口方式—国家”数据集。基于该合并数据,研究者不仅能够利用更优质、细致的数据深入研究企业出口决策、规模,还能够围绕企业出口产品,拓展研究企业出口产品的种类、质量及价格等领域。

合并数据提供了更优质的企业数据,并可以进一步地基于企业出口方式筛选出加工贸易企业,极大地推动了企业出口贸易行为的研究,其中一个焦点就是检验“生产率悖论”,并探索导致其可能存在的原因。一些文献验证了企业生产率与出口的正相关关系(Yu等,2013),发现出口商品种类越多,这种正效应越大(胡翠等,2015)。另一些文献则利用合并数据证实了我国出口企业存在“生产率悖论”,认为加工贸易企业是导致该悖论的重要原因(戴觅和余淼,2014),而且,加工出口的低固定成本以及产业政策支持导致了这种加工贸易企业的大量存在(Dai等,2016)。

合并数据提供的出口产品种类数据让工业企业数据库应用于企业出口增长的二元边际也成为了可能,许多研究开始关注出口增长的集约边际与扩展边际。近年文献主要围绕影响出口二元边际的因素进行分析,发现人民币升值、适量政府补贴会促进企业扩展出口市场,但汇率波动会对企业出口扩展边际产生负向影响(田朔等,2015),过大规模的政府补贴会抑制企业高端出口能力,不利于企业出口的扩展边际(张杰和郑文平,2015)。进口中间品投入则与企业出口的二元边际正相关,企业不断增加的进口中间品投入由于其内含的高技术和质量而有助于产品升级,从而提升企业在出口市场中的广度与深度(Feng等,2016),而由于出口企业更长的运送时间导致的更严格的信贷约束,则不利于企业出口的二元边际(Feenstra等,2014)。

合并数据还提供了出口产品的质量和价格数据。围绕出口产品质量,一些研究认为企业生产率、研发效率的提升有助于企业出口产品质量的提升(施炳展和邵文波,2014),市场竞争整体上也会促进企业出口产品质量提升,但它并未对国有企业产生促进作用,而政府补贴则会抑制产品质量提升,但对外资企业无抑制作用(张杰等,2015)。围绕出口产品价格,已有研究关注了汇率变动的影响,发现人民币实际汇率升值会对出口产品价格产生负面影响,且实际有效汇率每升值10%会带来出口产品价格0.63%的下跌,并且这种效应会随着企业生产率的提升以及产品在多产品企业中的地位的提升而增强(许家云等,2015)。

除了上述研究领域,利用合并数据也可以继续分析企业生产、创新行为、贸易模式选择等问题,这些分析是从国际贸易角度展开的。一些研究揭示了贸易自由化(Hu和Liu,2014)、进口中间品与最终产品的关税减免(Yu,2015)对企业生产率的正向作用,但这种正向作用也许仅存在于复杂产品生产商,对简单产品生产商则表现为负效应(Yu等,2013)。也有研究围绕企业研发行为,发现中间品关税的下降会促进企业研发投入,且主要是对生产技术的投入而非新产品的研发(田巍和余淼杰,2014)。从贸易模式选择出发,一些研究发现信贷约束会使得企业更

<sup>①</sup>工业企业数据库能够帮助海关数据库剔除纯粹的贸易中间商,从而更准确地识别出口企业;海关数据库能够帮助工业企业数据库筛选出通过贸易中间商间接出口的生产企业,从而更准确地识别出口企业。

多选择来料加工与加工贸易模式,不利于企业追求更高增加值与利润的产业活动(Manova和Yu,2016)。

#### (二)企业对外直接投资行为研究:与境外投资企业(机构)名录合并

境外投资企业(机构)名录提供了1980年以来所有中国对外直接投资(OFDI)企业的信息。已有文献利用它与工业企业数据库的合并数据识别具有对外直接投资的企业,研究企业对外直接投资的影响因素及其经济效应。(1)企业对外直接投资的影响因素。企业创新、生产率、出口强度等体现企业竞争优势的因素均会促进企业进行对外投资,其中,新产品占比与出口强度的促进效果最大,而体现企业竞争劣势的债务利息率则不利于企业进行对外投资的概率(葛顺奇和罗伟,2013)。行业竞争、外资企业比重等产业结构特征以及政府扶持则会显著影响企业对外直接投资规模,其中,行业竞争、政府激励政策能有效促进企业对外投资,企业内部的人力资本也能显著促进企业对外投资,而外资企业比重则会产生负向影响(Wang等,2011)。(2)企业对外直接投资的经济效应。企业对外投资促进了企业出口参与和出口强度,且有助于降低企业退出出口市场的风险,延长企业出口持续(毛其淋和许家云,2014),其也会通过逆向技术溢出来促进企业生产率提升,且这种提升作用会随着投资活动的持续而增强,但该效应仅存在于具有研发活动的企业(齐亚伟,2016)。

#### (三)其他企业经济行为研究:与其他微观数据库合并

中国工业企业数据库还可与其他微观数据库合并,进行多样化的企业经济行为研究,这也是未来工业企业数据库合并应用的趋势之一。已有研究利用科技型中小企业技术创新基金、国家专利数据库等数据,围绕创新补贴政策效果等方面进行研究(Guo等,2016)。利用中国城镇住户调查、世界银行投资环境调查等数据,研究者可以对企业社会责任行为、员工社会保障等方面进行更深入的研究(周浩和汤丽荣,2015)。也有研究者结合一些特殊的微观数据进行分析,如黄小勇(2015)利用2002—2007年黄金资源广告中标名单与工业企业数据库,分析企业投放黄金资源广告对其销售收入的影响。

近年来,越来越多的研究者开始利用合并数据进行研究,这不仅丰富了研究视角,扩展了基于工业企业数据库进行研究的广度与深度,极大地提升了工业企业数据库的利用价值。但是,工业企业数据库的数据指标也存在缺乏小微制造业企业、农业与服务业企业样本的缺陷,极大地限制了该数据库在小微企业,尤其是服务业企业经济行为研究中的应用,有待更全面、系统的数据库与之整合。

### 四、数据库样本时间范围的拓展

除了和其他数据库进行横向合并,工业企业数据库在纵向时间范围上也在不断延伸,目前已经更新到2013年。然而,Brandt等(2014)指出工业企业数据库2008年数据存在大约30%的规模以上企业样本遗漏,2009年数据缺少诸如应付工资、分类资本金等重要指标,2010年数据存在固定资产有较多极端值等严重错误的问题,且该年数据质量也被学界广泛质疑,同时自2011年起,工业企业数据库的统计口径发生了变动,由原先的“全部国有及年主营业务收入达500万元及以上的非国有工业企业”变为“年主营业务收入达2000万元及以上工业法人企业”,导致数据库样本覆盖范围发生了较大变化。因此,虽然工业企业数据库已更新至2013年,但由于样本大量遗漏、重要指标缺失、数据质量较差、统计口径变动、合并方法复杂等一系列问题,而1998—2007年数据的完整性及权威性也已得到学界的广泛认同与传播,现有文献仍大多使用1998—2007年数据进行研究。

然而,随着时间推移,该数据距今已有10年,显然继续使用2007年之前的数据会极大地限

制工业企业数据库的应用和价值。因此,今年以来已有部分文献开始探索合并方法,将2008—2013年数据纳入使用(王贵东,2017;谭语嫣等,2017)。对于2008年样本缺失的问题,可以通过中国第三次经济普查数据库进行补充,使用该方法能够很好的补充该年份的缺失样本,但也需注意,受限于经济普查数据库的指标量,使用该方法得到的数据库仅能应用于大多数基础研究(Brandt等,2014)。对于统计口径问题,已有文献建议采用最新的“规模以上”标准,即年主营业务收入达2000万元及以上,并以2011年为基准年进行统一(王贵东,2017)。在实际合并时,也需注意1998—2013年国民经济行业分类标准存在2次变动(2002年、2011年),须进行行业分类的统一。而对于2010年固定资产指标存在较多极端值的问题,虽然技术上可使用中位数法进行一定修复,但由于数据质量问题仍旧建议剔除该年数据(王贵东,2017)。

虽然经过上述方法可以将工业企业数据库的时间跨度延展至2013年,但受限于指标的缺失,研究范围也有所限制。从表2可以看出,利用2008—2013年工业企业数据库仍能够围绕企业出口贸易行为、雇佣行为、生存与成长、税收负担、政府补贴等方面进行研究,但无法对企业的生产行为<sup>①</sup>、创新行为进行深入地探讨。并且即便在可研究领域中,也存在部分指标不连续的问题,如职工数、应付工资、政府补贴等指标均存在不同程度的缺失。当然,随着工业企业数据库的不断更新,数据量的不断扩充,工业企业数据库在这些可研究领域的意义将逐步凸显。

表2 中国工业企业数据库重要指标更新情况

应用视角	核心变量	主要指标	1998—2007年	2008年	2009年	2010年	2011年	2012年	2013年
生产行为	生产率	工业总产值	√	√	√	√	√	√	√
		工业增加值	√ <sup>②</sup>			√			
		固定资产	√	√	√	√	√	√	√
		本年折旧	√				√	√	√
		中间品投入	√						
创新行为	创新	研发投入	√ <sup>③</sup>			√			
		新产品产值	√ <sup>④</sup>		√	√			
出口贸易行为	出口	出口交货值	√	√	√	√	√	√	√
雇佣行为	雇佣规模	职工数	√	√	√	√		√	√
	薪酬水平	应付工资	√	√			√	√	√
		应付福利	√	√					
生存与成长	生存能力	成立时间	√	√	√	√	√	√	√
		营业状态	√	√	√	√	√	√	√
	经营绩效	营业利润	√	√	√	√	√	√	√
		主营业务收入	√	√	√	√	√	√	√
其他重要领域	税负及避税	所得税	√	√	√	√	√	√	√
		增值税	√	√	√	√	√	√	√
	政府补贴	政府补贴	√	√			√	√	√
其他重要指标	分类资本金(国有、集体、外资、港澳台、法人、个人)		√	√		√	√	√	√
	费用(财务、管理、销售) <sup>⑤</sup>		√	√	√	√	√	√	√

资料来源:基于1998—2013年中国工业企业数据库整理。

①由于2008年后工业增加值、中间品投入存在较大缺失,因而2008年后无法使用LP、OP等目前主流的半参数方法测算企业TFP。

②工业增加值缺少2004年数据,可依据会计准则估算:工业增加值=工业总产值-中间品投入+增值税。

③研发投入仅有2001、2005—2007年数据。

④新产品产值缺少2004年数据。

⑤费用中的销售费用仅有1998—2003、2011—2013年数据,其他费用有1998—2013年数据。

在因受数据覆盖限制而无法进行深入探讨的领域,2008—2013年工业企业数据库仍旧能够继续应用。对于企业的创新行为,可通过国家专利数据库与工业企业数据库合并的方式,利用合并数据,从企业创新行为的产出端进行研究。对于企业的生产行为,虽然由于中间品投入、工业增加值的缺失而无法使用目前较为主流的LP、OP等半参数方法进行企业TFP的测算,但仍可以利用DEA-Malmquist等非参数方法测算企业TFP,或基于现有数据测算企业的劳动生产率或近似TFP。

## 五、数据处理中的新问题

虽然已有文献对数据库的潜在问题进行了较为充分的讨论,但随着研究的日益丰富,近年研究仍在数据整理、变量测度和变量加工等环节存在一些有待解决的数据处理新问题。

### (一)数据库的基本整理

在工业企业数据库中,由于难以找到一个识别每个企业的唯一特征进行编码,在利用该数据库展开研究时,每一个学者首先都会面临数据的基础整理问题。现有文献大多基于Brandt等(2012)提出的“序贯识别法”整理数据,但近年来的不少研究也暴露出该方法存在的一些问题。(1)企业识别问题。利用企业代码与名称进行精确匹配会识别出“过多”的企业,应采用“交叉匹配”与人工识别结合的方法进行精确匹配(聂辉华等,2012)。(2)样本错配。在利用“邮编+行业代码+主要产品+县名+开业年份”作为其他信息进行匹配时,忽略了在同一地区的不同企业同时开始从事某一行业的情况,从而导致样本错配,对此可以采用“地区代码+电话+开工年份”作为其他信息进行匹配,以避免这种匹配错误的产生(杨汝岱,2015)。(3)行政区划代码变动。未考虑我国近年来对县、地区划分的大规模调整,而县域调整会使得地级区划代码发生较大变化,影响了该指标匹配的准确性,可借鉴Bao等(2013)的方法,将历年行政区划代码统一。(4)开业年份问题。部分样本存在开业时间不合理或缺失的情况,如开业年份大于样本年份、开业年份为10至100等,如果该样本有其他年份开业时间数据,可考虑使用最新一年的合理数据替代,对于部分开业年份为10至100的样本,由于多数样本为漏写19导致,故可使用“1900+开业年份”替代。由此可见,现有的数据库处理方法仍存在优化空间,未来应不断改进与完善数据库的处理方法,以获得更优质的研究数据。

### (二)主要指标测度

现有研究中存在分歧的主要指标是生产率、劳动收入和创新。如前所述,在利用工业企业数据库进行的企业生产率研究中,多使用OP与LP的半参数方法测算企业TFP,但这些方法存在“函数相关性”等问题。所以,对企业TFP测算方法的改进也是今后研究的重点。未来研究可借鉴Akerberg等(2015)的测算方法,将劳动力因素纳入到投资或中间品需求函数中,以解决“函数相关性”问题。对于劳动收入,数据库中各项福利费指标存在不同年份的缺失,使得现有文献对员工收入的衡量存在一定的差异。有文献直接使用本年应付工资总额来衡量劳动力成本,也有文献基于连贯性原则,使用一直连续记录的工资、雇员补贴和失业保险来衡量员工收入,相比而言,较多文献使用本年应付工资与本年应付福利总额来衡量员工收入。在未来研究中,研究者应尽量避免衡量标准混乱导致结论差异的问题,采用相对一致的衡量方法。此外,有许多文献发现工业企业数据库中劳动报酬占增加值的份额与宏观统计数据存在较大冲突,认为工业企业报告的劳动报酬可能存在低估(Hsieh和Klenow,2009;Brandt等,2014)<sup>①</sup>,因此,可参考Hsieh和Klenow(2009)提出的方法,将企业的劳动报酬比重调整至国民核算中的份额,以获得较为合理的劳动收入数据。

<sup>①</sup>关于工业企业报告的劳动报酬存在低估的原因可详见“Challenges of working with the Chinese NBS firm-level data”(Brandt等,2014)。

此外,企业创新的度量一直是一个颇受争议的问题,在工业企业数据库研究中也不例外。现有文献主要从研发投入与研发产出两个维度衡量企业创新,但无论是衡量企业研发投入的研发支出、研发人员全时当量,还是衡量研发产出的专利、新产品产值均存在一定缺陷,也导致研究结果的不可比。研发支出难以反映企业人力资本投入以及对新技术的消化吸收,大部分研发活动也不能直接带来商业价值。新产品产值等研发产出指标虽能克服上述问题,但也无法体现企业创新活动中的商标权、著作权以及生产小工艺或技术的改进等创新产出。还有些文献使用无形资产增量来反映企业创新,认为它能够体现更多的关于企业研发活动的信息,但诸如土地使用权、特许权等与企业研发活动无关的数据也被计算在无形资产之内。所以在实际研究中,研究者应当明确各指标存在的缺陷,根据具体的研究问题挑选合适的代理变量。值得提醒的是,国家统计局在2004年的《第一次经济普查方案》中对新产品产值的界定进行了修订,将企业自行研发且未经政府认定,从投产日起一年内的新产品纳入到统计范畴内,这一修订无疑增加了企业上报新产品产值的“活动空间”,可能在一定程度上夸大新产品产值。

### (三)变量的初步加工

在变量的初加工环节,一个存在较大分歧的问题是名义变量的平减。企业经济行为研究,不可避免地涉及许多名义变量,需要剔除物价因素的干扰,然而现有文献对于平减方法并未达成共识。例如,对工业总产值、增加值等名义变量,所用的平减指数就有分行业工业品出厂价格指数、分地区工业品出厂价格指数以及Brandt等(2012)提出的分行业投入、产出平减指数。相比而言,Brandt等(2012)提出的平减指数构造方法能更为细致地对名义变量进行平减,但仍存在改良空间。从产出平减指数来看,该方法主要基于各企业的真实产出与名义产出比值的变动,来估算1998—2003年间四位数行业层面的产出平减指数,对于2004年后的产出平减指数仍旧使用两位数行业出厂价格指数,这样在面板数据中就存在不同年份数据平减的行业层次不一致,对此一个替代方法是利用分行业总产出环比价格指数对产出变量进行平减,以避免行业层次不同的问题。在投入平减指数的处理上,Brandt等(2012)仅使用2002年投入产出表来计算中间投入品价格指数,未考虑样本期间内投入产出结构的变化,今后可以参考杨汝岱(2015)的做法,使用1997年、2002年和2007年投入产出表对不同年份区间的投入平减指数进行计算。

除了数据平减,变量初加工的另一个焦点问题是如何识别出口企业。许多文献直接根据工业企业数据库中的出口交货值来识别出口企业,但在对工业企业数据库与海关进出口数据库进行匹配后,有文献发现2000—2005年有15.22%的出口交货值为零的企业在海关数据库中有出口记录(钱学锋,2015)。这是因为,海关数据库中记录的是贸易企业而非生产企业,有一部分生产企业并不会直接出口,而是通过贸易公司间接出口,因而出口交货值大于零但没有出口记录的企业并不一定是非出口企业,因此单以出口交货值来判定企业是否出口是不合理的。在未来的相关研究中,应当在基于出口交货值识别的基础上考虑该企业是否在海关数据库中存有出口记录,另外也可以结合企业上一期的出口交货值是否为零来判断。

## 六、数据库的未来应用方向

基于工业企业数据库及合并数据的研究,已经涵盖了较为丰富的研究领域,涉及宏观层面的政策环境、集聚经济,微观层面的企业投资、工资差距、就业公平、产品质量等,但仍然有一些颇具研究价值的视角可以切入,这些视角依然集中在企业生产行为、创新行为、雇佣行为等方面,但不同的是可以做更深层次和更精细的研究。

### (一)企业生产行为的深化研究

企业生产率的研究已经非常丰富,但依然存在可以进一步拓展和深化的方向。(1)加工贸

易与生产率。如前所述,我国存在大量低效率的加工贸易企业,可能是导致我国出口企业存在“生产率悖论”的主要原因,因而在出口与生产率的研究中应当借助海关进出口数据库,识别企业出口贸易类型,考虑加工贸易企业产生的影响。(2)纯出口企业与生产率。近年来,优惠政策、贸易中介等因素导致大量生产率低下的新企业选择进入出口市场而成为纯出口企业(高宇,2014),在对出口问题进行研究时也应当注意到纯出口企业所造成的影响,这是现有文献所缺乏考虑的。(3)进口行为与生产率。现有文献较少讨论企业进口行为,而企业进口与生产率也存在正向联系,未来研究中,应当注意进口对企业生产率的影响,并且在对企业的区分中也应考虑进口因素,不应简单的将企业区分为出口企业与非出口企业(钱学锋,2015)。(4)国有企业的生产率。一直以来,国有企业被认为是低效率的“代名词”,但近期有文献发现,央企的生产率仅次于外资企业,且国有企业隶属等级越高,生产效率越高(方明月,2014),可借助数据库中的隶属关系指标,深入国有企业内部,区分不同隶属等级的国有企业,探讨不同隶属等级国有企业的效率演变及来源。(5)多元化经营与生产率。利用2004年的所属母公司代码可以识别出哪些企业属于同一集团公司,目前利用该指标研究制造业企业集团化经营的文献不多,可以围绕集团化经营与非集团化经营企业、集团企业内部关联等视角进行研究,并进一步探究企业的多元化经营。(6)企业生产率的性别差异。在生产率的研究中,对企业生产率的性别差异研究仍较为缺乏,尤其缺少对生产率性别差异的产生原因与缩小方法的探讨。

## (二)企业创新行为的拓展研究

关于企业创新的研究一直是一个热点,未来的研究也可考虑下面三个方向。(1)研发平滑与创新。目前关于企业研发平滑的研究并不多见,且主要从企业现金持有的角度进行思考,仅有少数文献从营运资本等其他角度进行分析,未来应当从其他具有较低调整成本和较高变现能力的资本切入进行研究。(2)房价上涨与创新。在企业创新活动的影响因素研究中,房价上涨也是重要影响因素之一,它会通过企业融资约束的缓解以及资源的重新配置来对工业企业研发活动产生正、负两方面的影响(Miao和Wang,2012),虽然机制明确,但利用微观企业数据,研究房价快速上涨如何影响工业企业研发活动的文献还是比较匮乏。(3)创新活动的负面效应。现有文献大多只考虑创新活动的正向影响,并未考虑过多研发活动所带来的不确定性与风险,所以,未来研究应考虑企业创新活动可能造成的负面影响,估计不同行业的最优创新投入水平。

## (三)企业雇佣行为的拓展研究

在企业雇佣行为研究中,现有文献多关注企业间的工资差距,而较少研究企业内工资差距,尤其是仅仅局限于利用上市公司数据探讨高管层薪酬差距,而缺少对企业内普通员工间工资差距的探讨。普通员工间的工资差距,包括学历差距、技术差距与性别差距等不同层次。关于工资技术差距,可借鉴陈波和贺超群(2013)的方法测算技术工人工资溢价,以此衡量企业内工资差距,从企业内工资不平等的原因与经济效应两方面进行讨论。此外,在工资的性别差异方面,可依据学历、技术资质等数据进一步区分女性群体中的技术与非技术工人,考察女性群体内的工资差距。需要指出的是,工业企业数据库只提供了2004年的学历构成相关数据,因此,当前阶段只能利用截面数据进行分析,或借助企业学历构成短期内不变的假设,将截面数据扩展为面板数据,显然,基于这两种方法得到的结论都是不够严谨的,应考虑利用更丰富的数据资料进行研究或采用更好的方法对该截面数据进行处理。

当然,还有许多其他研究视角值得我们去探索。利用企业排污费、养老保险费、住房公积金等数据,我们能够从环境、社会保障等角度进行研究;利用企业的地理位置信息,我们能够识别企业所处区位,考察设立经济开发区等政策的效果,也可以研究企业的区位选择、空间分布以及迁址等问题。

除了本文梳理的企业经济行为研究,工业企业数据库也可应用于管理行为领域的研究。例如,基于职工教育费、各项保险费研究企业的人力资源投入,基于缴税数据研究非上市企业的避税行为。同时,基于2004年独有的不同教育水平、技能水平的职工数数据可研究企业的人才结构及其影响,基于工会人数、经费以及集团公司数据研究企业内部组织结构及其影响。

企业的经济行为一直受到极大的关注,如何提高实体经济尤其是民营企业的生产创新活力也是近年来国内研究所关注的重点。本文较为全面地梳理了近年基于工业企业数据库及其合并数据在企业经济行为领域中的主要研究,并整理了该数据库的数据结构和问题,在研究视角、数据处理和未来应用方向上得出了一些启示。未来研究可以进一步地从企业雇佣的性别差异、多元化经营、出口贸易方式等新视角探讨其对企业生产率的影响,更多关注研发平滑、房价上涨对企业创新的影响以及创新活动本身给企业带来的不确定性、高风险等常被忽视的负面效应,还可考察企业内工资不平等,尤其是不同性别、学历、技术工人间的工资差距。同时,与其他微观数据库的合并也将成为工业企业数据库未来应用的主要方向之一,工业企业数据库将成为基础性数据库,为其他微观数据库提供各类基本信息指标,并通过数据合并广泛应用于企业经济与管理行为的研究。

### 主要参考文献

- [1]包群,叶宁华,王艳灵.外资竞争、产业关联与中国本土企业的市场存活[J].经济研究,2015,(7):102-115.
- [2]陈波,贺超群.出口与工资差距:基于我国工业企业的理论与实证分析[J].管理世界,2013,(8):6-15.
- [3]陈国强,罗楚亮.劳动生产率与工资决定的性别差距——来自我国工业企业数据库的经验研究[J].经济学动态,2016,(8):38-52.
- [4]盖庆恩,朱喜程,名望,等.要素市场扭曲、垄断势力与全要素生产率[J].经济研究,2015,(5):61-75.
- [5]葛顺奇,罗伟.中国制造业企业对外直接投资和母公司竞争优势[J].管理世界,2013,(6):28-42.
- [6]龚美,胡关亮.中国制造业资源配置效率与全要素生产率[J].经济研究,2013,(4):4-15,29.
- [7]胡翠,林发勤,唐宜红.基于“贸易引致学习”的出口获益研究[J].经济研究,2015,(3):172-186.
- [8]李晓萍,李平,吕大国,等.经济集聚、选择效应与企业生产率[J].管理世界,2015,(4):25-37,51.
- [9]毛其淋,盛斌.中国制造业企业的进入退出与生产率动态演化[J].经济研究,2013,(4):16-29.
- [10]毛其淋,许家云.政府补贴对企业新产品创新的影响——基于补贴强度“适度区间”的视角[J].中国工业经济,2015,(6):94-107.
- [11]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[J].世界经济,2012,(5):142-158.
- [12]施炳展,邵文波.中国企业出口产品质量测算及其决定因素——培育出口竞争新优势的微观视角[J].管理世界,2014,(9):90-106.
- [13]苏振东,洪玉娟.中国出口企业是否存在“利润率溢价”——基于随机占优和广义倾向指数匹配方法的经验研究[J].管理世界,2013,(5):12-34,46.
- [14]苏振东,洪玉娟,刘璐瑶.政府生产性补贴是否促进了中国企业出口?——基于制造业企业面板数据的微观计量分析[J].管理世界,2012,(5):24-42,187.
- [15]谭语嫣,谭之博,黄益平,等.僵尸企业的投资挤出效应:基于中国工业企业的证据[J].经济研究,2017,(5):175-188.
- [16]王贵东.中国制造业企业的垄断行为:寻租型还是创新型[J].中国工业经济,2017,(3):83-100.
- [17]许家云,佟家栋,毛其淋.人民币汇率变动、产品排序与多产品企业的出口行为——以中国制造业企业为例[J].管理世界,2015,(2):17-31.
- [18]杨汝岱.中国制造业企业全要素生产率研究[J].经济研究,2015,(2):61-74.
- [19]杨洋,魏江,罗来军.谁在利用政府补贴进行创新?——所有制和要素市场扭曲的联合调节效应[J].管理世界,2015,(1):75-86.
- [20]周浩,汤丽荣.市场竞争能倒逼企业善待员工吗?——来自制造业企业的微观证据[J].管理世界,2015,(11):135-144.

- [21]周云波, 陈岑, 田柳. 外商直接投资对东道国企业间工资差距的影响[J]. 经济研究, 2015, (12): 128–142.
- [22]Akerberg D A, Caves K, Frazer G. Identification properties of recent production function estimators[J]. *Econometrica*, 2015, 83(6): 2411–2451.
- [23]Chandra P, Long C. VAT rebates and export performance in China: Firm-level evidence[J]. *Journal of Public Economics*, 2013, 102: 13–22.
- [24]Chang S J, Wu B. Institutional barriers and industry dynamics[J]. *Strategic Management Journal*, 2014, 35(8): 1103–1123.
- [25]Dai M, Maitra M, Yu M J. Unexceptional exporter performance in China? The role of processing trade[J]. *Journal of Development Economics*, 2016, 121: 177–189.
- [26]Feng L, Li Z Y, Swenson D L. The connection between imported intermediate inputs and exports: evidence from Chinese firms[J]. *Journal of International Economics*, 2016, 101: 86–101.
- [27]Gan L, Hernandez M A, Ma S. The higher costs of doing business in China: Minimum wages and firms' export behavior[J]. *Journal of International Economics*, 2016, 100: 81–94.
- [28]Guo D, Guo Y, Jiang K. Government-subsidized R&D and firm innovation: Evidence from China[J]. *Research Policy*, 2016, 45(6): 1129–1144.
- [29]Howell A. Firm R&D, innovation and easing financial constraints in China: Does corporate tax reform matter?[J]. *Research Policy*, 2016, 45(10): 1996–2007.
- [30]Hu A G, Liu Z N. Trade liberalization and firm productivity: Evidence from Chinese manufacturing industries[J]. *Review of International Economics*, 2014, 22(3): 488–512.
- [31]Kamal F. Origin of foreign direct investment and firm performance: Evidence from foreign acquisitions of Chinese domestic firms[J]. *World Economy*, 2015, 38(2): 286–314.
- [32]Liu Q, Lu Y. Firm investment and exporting: Evidence from China's value-added tax reform[J]. *Journal of International Economics*, 2015, 97(2): 392–403.
- [33]Manova K, Yu Z H. How firms export: Processing vs. ordinary trade with financial frictions[J]. *Journal of International Economics*, 2016, 100: 120–137.
- [34]Wang J, Wang X. Benefits of foreign ownership: Evidence from direct investment in China[J]. *Journal of International Economics*, 2015, 97(2): 325–338.
- [35]Yu M J. Processing trade, tariff reductions and firm productivity: Evidence from Chinese firms[J]. *The Economic Journal*, 2015, 125(585): 943–988.

## **The Application of the Database of “Chinese Annual Survey of Industrial Firms” to Enterprise Economic Behavior Research: Perspectives, Integration and Extension**

Xiao Liping<sup>1,2</sup>, Xu Weifeng<sup>2</sup>

(1. *Center for Economic Development Research, Wuhan University, Wuhan 430072, China*; 2. *School of Economics and Management, Wuhan University, Wuhan 430072, China*)

**Summary:** The database of “Chinese Annual Survey of Industrial Firms” has become one of the most important data sources in micro research on enterprise economic behaviors. Based on the matching between the data structure of this database and the research category of enterprise economic behaviors, by systematically reviewing the latest application in research on enterprise economic behaviors, such as production, innovation, international trade, employment, survival and growth of enterprise, this paper points out that this database has some limitations, such as the limited index types and the singleness of

some category indexes, impeding the further study in some fields. In this view, this paper argues that more and more micro-databases will be used to be merged with the database of “Chinese Annual Survey of Industrial Firms”, and using the merged data to study will be one of the most important directions in future research. Therefore, it reviews the latest applications of the merged data for the first time and puts forward some relative new perspectives in future extended application of this database.

Some findings of this paper are as follows: firstly, current researches based on this database mostly focus on enterprise productivity, innovation, export decision and export scale, and other traditional fields, but also begin to pay attention to some new perspectives such as export continuity, wage inequality, and viability of enterprises; in order to solve the problem of limitations of the database of “Chinese Annual Survey of Industrial Firms”, more and more literature tries to merge this database with other micro-databases, such as “China Customs Import and Export Database”, “Chinese Oversea Enterprise List”, and expands and deepens studies based on the merged data; these kinds of research concern more about the dual margins of export, the quality of exports, foreign direct investment and so on; secondly, due to the problems such as sample missing, a lack of important indexes, statistic scope adjustment and poor data quality in the following years’ data, most of existing researches only use the data between 1998 and 2007; however, as time goes by, continuing to use the short data before 2007 severely limits the application value of this database, and hence some researches have begun to explore the methods of adding the following years’ data to solve this limitation; this paper points out the problems of using the following years’ data and provides some feasible solutions; furthermore, it sorts out the data structure of this database after 2007 and points out the directions in future research using this new data, namely using the data from 2008 to 2013, researchers can still continue their research on some important fields, such as enterprise export trade, employment, survival and growth, tax burden and government subsidies, and further based on the merged data, they can also explore enterprise production, innovation and other economic behaviors; at the same time, with the continuous updating of the database, the significance of the database will gradually highlight; thirdly, in the aspect of data processing, in three major aspects including data arrangement, variable measurement and initial processing of variables, this paper points out that there are some new problems in the existing literature, such as over-matching of samples, biased index measurement, misusing of deflator, and loose identification of export firms; also, this paper tries to provide some feasible solutions to address these problems; fourthly, in the aspect of research areas, future research can go deeper on the exploring of enterprise production, innovation and employment behaviors, and pay more attention to the effects of different types of import and export behaviors, diversified business and subordinate levels on their productivity, and further explore the issues of gender differences in productivity, innovation risk, R&D smoothing, and wage inequality within an enterprise.

**Key words:** Chinese Annual Survey of Industrial Firms; enterprise economic behavior; research area; data processing

(责任编辑: 子 文)