

# 中国劳动力市场的职业结构变迁

李志远 吴磊

## 目录

附录 I 数据介绍 .....	1
附录 II 人口普查与 O*NET 数据匹配 .....	8
附录 III 职业特征分布的稳健性分析 .....	14
附录 IV 其他数据图表 .....	18
附录 V 附图 .....	28
参考文献 .....	29

## 附录 I 数据介绍

## (一) 人口普查数据

本文使用的就业数据源于 2000 年第五次全国人口普查微观调查数据、2010 年第六次全国人口普查微观数据以及 2015 年 1%人口抽样调查微观数据。其中，2000 年和 2010 年人口普查微观数据库均来自于人口普查中的长表数据。长表数据是通过在总人口中抽样 10%的住户填写特定具体信息得到的，如就业状况、迁移流动以及婚姻生育等。本文使用的微观数据库是在这一长表数据中再按照比例随机抽取部分个体而组成的，其中 2000 年的微观数据是按照长表数据的 10%抽样得到的，2010 年的微观数据是按照长表数据的 3%左右抽样得到的。因此，我们也可以将本文使用的 2000 年和 2010 年的微观抽样数据分别认为是第五次人口普查 1%的抽样数据和第六次人口普查 0.3%的抽样数据。本文有关 2015 年的微观人口数据则源自 2015 年开展的 1%人口抽样普查。这一人口抽样普查按照 1.55%在总人口中进行随机抽样，最终样本 2131 万人。本文使用的 2015 年的数据就是这 2131 万人中随机抽取的 200 万条数据。

考虑到抽样数据的质量可能会使得本文计算得出的就业结构存在一定偏差。因此，我们将本文的行业就业结构数据，与国家统计局公布的人口普查行业就业数据，以及王亚菲等（2021）基于《中国劳动统计年鉴》、《中国农村统计年鉴》以及国家统计局公布的就业人数核算得出的行业就业数据相互比对。表 I1 中的 Panel A-C 分别给出了 2000、2010 以及 2015 年的行业就业结构数据。从表中可以发现，基于本文数据计算的行业就业分布，与国家统计局人口普查，以及王亚菲等（2021）计算的结果相似，尤其与国家统计局公布的人口普查长表数据结果相近。

表 I 1 不同数据源的行业就业结构比对

	统计局数据 (1)	王亚菲等 (2021) (2)	本文数据 (3)
Panel A: 2000 年行业大类 (16 种)			
农、林、牧、渔业	64.376	56.469	64.390
采掘业	1.044	1.588	1.044
制造业	12.461	12.715	12.453
电力、煤气及水的生产和供应业	0.626	0.828	0.630
建筑业	2.684	2.505	2.687
地质勘查业、水利管理业	0.126	0.171	0.122
交通运输、仓储及邮电通信业	2.579	3.593	2.572

批发和零售贸易、餐饮业	6.690	8.588	6.694
金融、保险业	0.590	0.956	0.588
房地产业	0.231	0.289	0.231
社会服务业	2.151	2.410	2.151
卫生、体育和社会福利业	1.061	1.427	1.069
教育、文化艺术及广播电影电视业	2.558	4.574	2.552
科学研究和综合技术服务业	0.224	0.662	0.223
国家机关、政党机关和社会团体	2.352	3.224	2.346
其他行业	0.245	0.001	0.248

**Panel B: 2010 年行业大类 (20 种)**

农、林、牧、渔业	47.305	43.003	43.074
采矿业	1.149	1.716	1.225
制造业	17.216	15.598	19.289
电力、燃气及水的生产和供应业	0.707	0.949	0.710
建筑业	5.573	4.302	6.681
交通运输、仓储和邮政业	3.620	2.097	3.691
信息传输、计算机服务和软件业	0.627	1.929	0.646
批发和零售业	9.497	10.352	9.895
住宿和餐饮业	2.785	2.000	3.360
金融业	0.829	1.435	0.809
房地产业	0.686	0.636	0.683
租赁和商务服务业	0.699	2.034	0.739
科学研究、技术服务和地质勘查业	0.327	0.892	0.335
水利、环境和公共设施管理业	0.381	0.677	0.377
居民服务和其他服务业	1.977	1.622	2.193
教育	2.351	4.833	2.218
卫生、社会保障和社会福利业	1.189	1.923	1.151
文化、体育和娱乐业	0.462	0.403	0.475
公共管理和社会组织	2.619	3.597	2.448
国际组织	0.001	0.001	0.001

**Panel C: 2015 年行业大类 (20 种)**

农、林、牧、渔业	-	29.633	34.562
采矿业	-	1.464	1.067
制造业	-	16.380	18.446
电力、燃气及水的生产和供应业	-	1.065	0.757

建筑业	-	6.917	8.446
交通运输、仓储和邮政业	-	2.502	3.933
信息传输、计算机服务和软件业	-	2.296	0.820
批发和零售业	-	15.729	12.642
住宿和餐饮业	-	3.152	4.173
金融业	-	1.632	1.139
房地产业	-	1.105	0.992
租赁和商务服务业	-	3.688	1.155
科学研究、技术服务和地质勘查业	-	1.105	0.496
水利、环境和公共设施管理业	-	0.751	2.014
居民服务和其他服务业	-	2.236	3.936
教育	-	4.668	1.502
卫生、社会保障和社会福利业	-	2.294	0.239
文化、体育和娱乐业	-	0.402	0.468
公共管理和社会组织	-	2.979	3.213
国际组织	-	0.001	0.000

注：为了与国家统计局公布的人口普查长表的行业就业占比相对应，本表展示的行业分类主要基于普查数据对应年份使用的国民经济行业分类与代码《GB/T 4754-1994》、《GB/T 4754-2002》和《GB/T 4754-2017》。列（1）中的数据主要源于国家统计局公布的第五次和第六次人口普查长表数据。由于2015年是1%的抽样调查，统计局并未公布相应总体数据，因此列（1）Panel C 中的数据缺失，以“-”表征。列（2）中的数据来自王亚菲等（2021）分行业的就业核算，其通过下述等式计算得到：全国就业人数=城镇单位就业人数+私营企业和个体就业人数+除乡村私营企业和个体外的乡村就业人数。具体的数据细节和测算过程可以参照对应文献。列（3）中的数据源自本文使用的微观抽样数据。

为了更直观的观察本文数据结果与其他数据源的差异，笔者绘制了行业维度就业占比偏差的密度图。如图I1所示。其中，图I1（a）给出了本文数据与统计局长表数据的偏差分布。可以发现大部分行业的就业占比偏差极小，整体偏差聚集在1处，且基本呈现正态分布。这反映出基于本文数据计算的劳动力在行业层面的分布较为准确，与国家统计局公布的长表数据极其相近。图I1（b）给出了本文数据与王亚菲等（2021）宏观数据的偏差分布。可以发现，本文计算得出的大部分行业就业占比偏低，仅少部分行业就业占比偏高。这在数据中表现为本文计算的农业相关就业占比较高，以及其他行业较低的现象。这一点可能是由于王亚菲等（2021）计算时使用的城镇单位就业人数对在城镇从事农业的劳动者本就存在统计遗漏导致的。最后，由于行业之间的比重是相对的，因此本文数据相较王亚菲等（2021）数据在农业相关行业计算得出的较高的就业比重，会导致在其他行业中就业比重相对偏低。

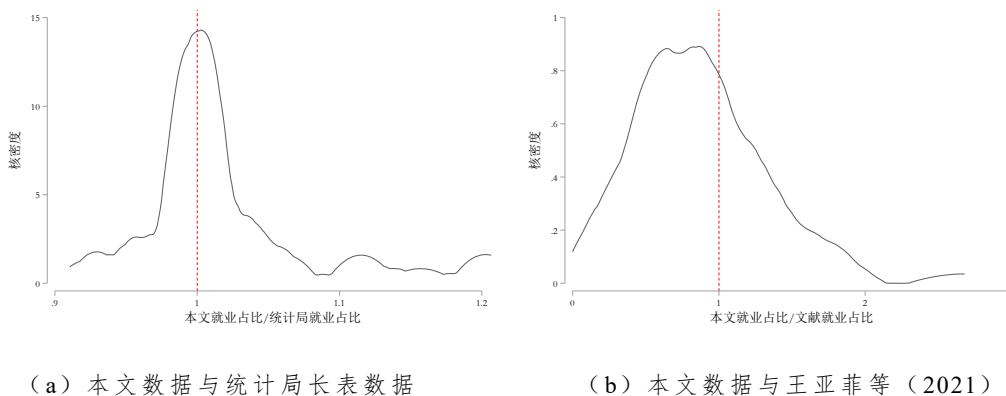


图 1 1 行业就业占比偏差的密度图

注：(a) 中的横轴为本文数据与统计局长表数据的行业就业占比偏差，其计算公式为：行业就业占比偏差=本文数据计算的就业占比/统计局长表数据计算的就业占比。纵轴为密度。  
 (b) 中的横轴为本文数据与王亚菲等(2021)的行业就业占比偏差，其计算公式为：行业就业占比偏差=本文数据计算的就业占比/王亚菲等(2021)计算的就业占比。纵轴为密度。

表 I 2 给出了 2000、2010 以及 2015 年的职业大类就业结构数据。可以发现本文数据计算得出的职业就业结构与国家统计局公布的宏观数据结果相近，尤其是 2000 和 2010 年。<sup>①</sup>

表 I 2 不同数据源的就业结构对比

	统计局数据			本文数据		
	2000	2010	2015	2000	2010	2015
	(1)	(2)	(3)	(4)	(5)	(6)
国家机关、党群组织、企事业单位负责人	1.668	1.811	1.954	1.667	1.729	1.737
专业技术人员	5.703	6.965	11.695	5.702	6.931	7.996
办事人员和有关人员	3.097	4.414	9.524	3.088	4.302	4.444
商业、服务业人员	9.177	16.499	24.714	9.171	17.767	27.883
农、林、牧、渔、水利业生产人员	64.460	47.281	28.312	64.478	43.042	34.429
生产、运输设备操作人员及有关人员	15.828	22.931	23.414	15.829	26.114	23.349
不便分类的其他劳动者	0.066	0.098	0.388	0.066	0.115	0.163

注：列(1)和列(2)的数据主要源于国家统计局公布的第五次和第六次人口普查，列(3)的数据主要源于《2016年中国劳动统计年鉴》。列(4)-(6)的数据主要源于本文使用的微观数据库。

此外，考虑到国家统计局公布了更广维度的职业-行业就业结构，以及更细维度的职业就业结构，如 2000 年职业-行业就业结构数据，以及 2010 年小类职业的就业结构数据。因此，笔者进一步使用本文的微观数据与之进行比对。图 I 2 给出了普查数据与抽样数据在职业-行业和职业小类这两大维度上的对比结果。可以

<sup>①</sup> 2015 年中部分职业就业结构的偏差可能是因为人口普查的劳动统计与《中国劳动统计年鉴》统计口径的差异，如《中国劳动统计年鉴》定义的就业人口需与单位具有长期稳定的关系，而人口普查仅考虑调查前一周的工作关系(岳希明，2005)。

看出本文使用的微观抽样数据与统计局公布的长表数据计算得出的各行各业的就业占比基本一致。主要表现为图 I 2(a)中总体职业-行业对子的偏差聚集在 1 处,整体呈现正态分布态势;以及图 I 2(b)中总体职业小类的偏差聚集在 1 处,整体表现出了正态分布。

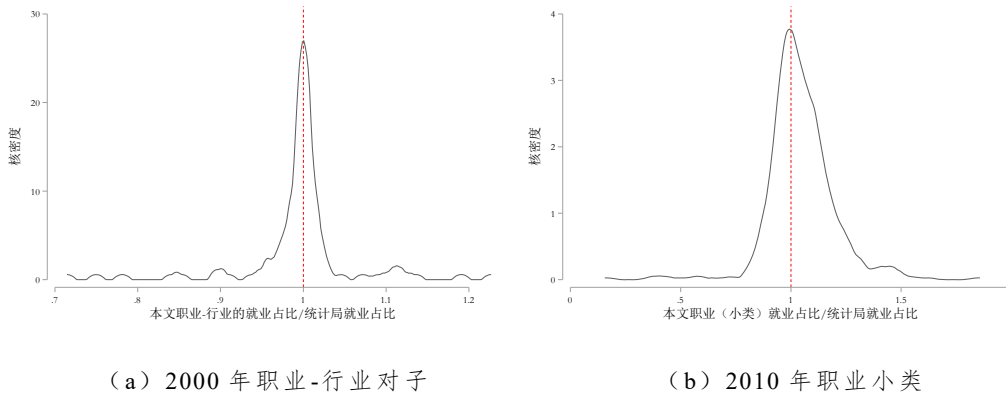


图 I 2 普查数据与抽样数据对比

注:(a)横轴中的职业(大类)-行业(大类)就业占比=(在某一行业以及某一职业中就业的人数/总体就业人数)\*100%,笔者通过分别计算统计局公布的职业(大类)-行业(大类)就业占比以及本文的微观抽样数据,从而得到(a)横轴中的就业结构偏差=本文职业(大类)-行业(大类)就业占比/统计局职业(大类)-行业(大类)就业占比。(b)中的职业小类就业占比=(在某一职业小类中的就业人数/总体就业人数)\*100%,这里的职业小类即为《职业分类与代码》GB/T6565-2009中409项职业类目的口径。笔者通过分别计算统计局公布的职业(小类)就业占比以及本文的微观抽样数据,从而得到(b)横轴中的就业结构偏差=本文职业(小类)就业占比/统计局职业(小类)就业占比。

## (二) O\*NET 数据库以及指标构建

O\*NET 数据库作为美国劳工部组织开发的职位分析数据,描绘了美国近 1000 种职业的工作任务、工作活动、工作环境和兴趣。这为我们从技能角度归纳总结劳动要素调整的方向提供了统一的量化指标。O\*NET 数据库自 2003 年以来,每年均会发布新的版本,更新内容涉及职业类别和职业特征。本文的正文部分主要使用 2020 年公布的 O\*NET25.0 版本进行分析。并且,考虑到使用不同版本将可能影响本文观察到的劳动要素调整方向,为此我们将在附录的后续部分使用不同的 O\*NET 版本度量调整方向。在度量技能特征时,本文主要借助 O\*NET 相关指标,参照 Acemoglu and Autor (2011)和 Cortes and Pan (2018)的方法,构建不同的技能指数,如体力能力(Physical Ability)、认知能力(Cognitive Ability)、重复性认知(Routine Cognitive)、重复性体力(Routine Manual)、协同要求(Interaction)、竞争要求(Competition)以及教育要求(Required Education)。

具体的构建方法如表 I 3 所示。其中,体力能力(Physical Ability)主要依据每个职业对一般体力活动、处理和移动物体这两项活动的打分。以此来刻画需要进行体力劳动的生产活动,如货物搬运、机器操作等。认知能力(Cognitive Ability)

主要依据每个职业对书面理解能力、数学推理能力、演绎推理、归纳推理这四项活动的打分。以此来刻画需要思维能力的生产活动，如思考和解决问题，抽象化和具象化思维理解等。重复性认知（**Routine Cognitive**）依据每个职业对重复相同任务、任务的精准和准确性和结构化工作这三项活动的打分，以此来刻画职业分工活动中呈现出的重复性思维的生产活动，如秘书、律师等的工作。重复性体力（**Routine Manual**）具体依据每个职业对工作进度由机器决定、控制机器的能力以及重复性动作这三项活动的打分，以此来刻画职业分工活动中呈现出的重复性体力的生产活动，如机器设备操作员等的工作。协同要求（**Interaction**）主要依据每个职业对与他人接触、团队合作、人际关系和社会感知这四项活动的打分。以此来刻画需要与他人沟通交流以完成任务的生产活动，如报刊记者、公关部门等。竞争要求（**Competition**）主要依据每个职业对竞争程度的打分。以此来刻画竞争考核压力强的工作活动，如金融投资从业者等。教育要求（**Required Education**）主要依据每个职业对所需教育水平的等级，以此来刻画所需的人力资本投入。

表 1 3 不同技能指数的构建方法

技能指数	细分特征	来源	典型职业
体力能力	i) 一般体力活动	Cortes and Pan (2018)	土木工程建
	ii) 处理和移动物体		筑施工人员
认知能力	i) 阅读理解	Cortes and Pan (2018)	航空工程技 术人员
	ii) 数学推理		
	iii) 演绎推理		
	iv) 归纳推理		
重复性体力	i) 由设备速度决定工作节奏	Acemoglu and Autor (2011)	水泥、石灰、 石膏及其制 品制造人员
	ii) 控制机器和流程		
	iii) 花时间进行重复动作		
重复性认知	i) 重复同样任务的重要性	Acemoglu and Autor (2011)	校对员
	ii) 准确或精确的重要性		
	iii) 结构化与非结构化工作		
协同要求	i) 与他人接触	Cortes and Pan (2017)	社会工作专 业人员
	ii) 团队合作		
	iii) 建立和维持人际关系		
	iv) 社会知觉		
竞争要求	i) 竞争力的重要性	依据问卷直接构建	证券专业人 员
教育要求	i) 所需教育水平	依据问卷直接构建	西医医师

注：表中最后一列中的典型职业，仅列出了部分对应技能指数排名较高的职业。

考虑到各个指标的原始数值并不具备经济意义，仅表示了重要性打分。因此我们参考文献中的处理办法，对研究中使用的每个细分特征指标进行标准化处理。得到均值为 0 标准差为 1 的职业特征。然后对标准化的细分指标进行简单平均计算，得出最终的体力能力、认知能力、重复性认知、重复性体力、协同要求和竞争要求的测度。有关教育要求的测度，由于问卷中直接以学历分类变量展示，因此本文按照其原始分类在后续分析中直接使用。

### (三) 美国社区调查数据

为具象化我国劳动要素 2000-2015 年在职业层面调整规模的大小，本文借助 2000, 2010 和 2015 年的美国社区调查数据 (American Community Survey, ACS) 计算同时期美国劳动要素在职业层面的调整规模。美国社区调查数据作为美国人口普查局实施的年度抽样调查，详细记录了人口特征、社会特征、经济特征和住房特征。



## 附录 II 人口普查与 O\*NET 数据匹配

### (一) 数据匹配过程

为了将人口普查数据库与 O\*NET 数据库相匹配从而得到职业特征的度量,我们主要依据《职业分类与代码》中列出的职业名称和职业描述,以及 O\*NET 数据中的职业列表和描述进行匹配。考虑到 O\*NET 数据库每年均会定期更新数据库,更新内容涉及职业和各项技能特征。但是有关职业类目的变化并非是每年的,而是依循四个版本的职业标准: O\*NET-SOC2000, O\*NET-SOC2006, O\*NET-SOC2009 和 O\*NET-SOC2010。为匹配便捷,本文后续的匹配均是将 O\*NET-SOC2010 中的美国职业分类与中国职业分类作对应,随后再借助 O\*NET-SOC 不同版本之间的对应表,得到 O\*NET 不同版本与中国职业的关系。接下来我们将详细阐述具体的匹配过程,图 II 1 也大致的展示了匹配的整个流程,以供参考。

就手动将中国职业和美国职业在细分小类层面进行匹配,即将中国职业小类(3 位码或者 5 位码)与美国的职业小类(8 位码)进行匹配。由于中美职业分类的依据不同:美国依据技能水平和专业领域,而中国则首要依据专业领域,因此两国职业种类数存在较大差异。其中,人口普查统计口径下的职业类目仅 400 多种,美国 O\*NET 的职业分类则有近 1000 种。所以一个中国职业可能对应某一个美国职业,也可能对应多个美国职业,甚至也可能无法找到能与之对应的美国职业。我们接下来将具体阐述这三种对应关系以及数据处理的过程。

首先,我们依据《职业分类与代码》职业名称和职业描述,在 O\*NET 数据库中寻找能与之较为精准对应的职业,从而确定第一种关系:一个中国职业可能对应的某一美国职业。<sup>①</sup>以《职业分类与代码》GB/T6565-2009 版本为例,中国的国际商务人员(2-15)对应美国的报关员(Customs Brokers, 13-1199.03),中国的印刷操作人员(8-52)在 O\*NET 数据库中是美国的印刷操作员(Printing Press Operators, 51-5112.00),中国的电影放映人员(8-44)对应美国的电影放映员(Motion Picture Projectionists, 39-3021.00)。其中,有且仅能确定这种对应关系的中国职业共有 123 种,占《职业分类与代码》种类的 30.07%,涉及 O\*NET 中职业种类的 8.57%。

其次,我们基于《职业分类与代码》职业描述中涉及的细分职业,搜寻 O\*NET 数据库中可以与之对应的美国职业,从而确定第二种关系:一个中国职业可能对应的多个美国职业。由于我国人口普查中对职业的统计口径更粗,因此中美职业大部分呈现第二种关系。以《职业分类与代码》GB/T6565-2009 版本为例,中国

<sup>①</sup> 由于我们的数据匹配目标是为了在 O\*NET 数据库中找到能和中国职业对应的美国职业,因此我们在 O\*NET 数据库中进行搜寻匹配时均是“有放回”的搜寻。即某一 O\*NET 的美国职业可能出现在第一种对应关系中,也可能出现在第二种对应关系中。

的包装人员(9-91)对应美国的包装和灌装机操作员(Packaging and Filling Machine Operators and Tenders, 51-9111.00)以及手工包装工人(Packers and Packers, Hand, 53-7064.00),中国的兽医、兽药技术人员(1-75)对应美国的兽医技术员(Veterinary Technologists and Technicians, 29-2056.00)以及兽医(Veterinarians, 29-1131.00)。其中,有且仅能确定这种对应关系的中国职业共有 259 种,占《职业分类与代码》种类的 63.33%;涉及 O\*NET 中职业种类的 94.65%。

最后,基于上面两步的匹配,我们将还未确定匹配关系的中国职业归类到第三种对应关系:无法找到能与之对应的美国职业。以《职业分类与代码》GB/T6565-2009 版本为例,共计 13 个中国职业。主要是军工类职业有关,如枪炮制作人员(6-84),装甲车辆装试人员(6-83)等。

基于上述的匹配步骤,我们得以将中国职业与美国职业对应匹配,从而获得相应美国职业对应的各个特征,如工作任务、工作活动、工作环境和兴趣等。考虑到中美职业存在的三种对应关系,因此在描绘某一职业特征时,中国职业可能得到唯一确定的数值测度,也可能有多个数值测度,或者没有数值测度。为此,我们对于有多个数值测度的取简单平均值。对于没有数值测度的,我们对职业归属的中类或者大类内的所有职业的数据测度取简单平均值来赋值。

得到历年人口普查微观数据中职业相对应的特征测度后,我们得以观察特定年份内,劳动力在不同职业间的分布。为进一步观察特定职业及其所表征的知识技能和分工活动在不同年份间的调整,仍需将不同年份的职业标准进行统一。具体的,2000 年、2010 年和 2015 年微观人口普查中的细分职业分别对应使用国家标准《职业分类与代码》的 GB/T6565-1999 版本、GB/T6565-2009 版本以及 GB/T6565-2015 版本。《职业分类与代码》的 GB/T6565-1999 版本和 GB/T6565-2009 主体结构相同,变化较小;而 GB/T6565-2009 版本和 GB/T6565-2015 版本之间出现了较大的职业标准变动:对职业大类、中类和小类都进行了调整,尤其是新增了 30 多项职业小类。为了得到统一的职业分类标准,我们参照官方职业分类对照表,将 2009 版职业分类代码与 2015 版职业分类代码进行对应和统一,最终得到 438 项细分职业。其中,因为职业标准变动而归并和拆分的职业共有 77 项,产生的新职业共有 114 项。

得到新的历年统一的职业分类表之后,我们进一步将其与我们之前手动匹配得到的历年职业特征数据结合。具体的,我们首先将 2010 年中国职业特征匹配到“统一的职业分类表”,保留可以一一精准匹配的职业;随后我们将 2015 年中国职业特征匹配到“统一的职业分类表”,保留可以一一精准匹配的职业。通过将这两步得到的数据合并以及删除重复值之后,我们可以得到“统一的职业分类表”及其对应的各个职业特征。

同时,有关中国职业对应的不同 O\*NET 版本的特征,我们主要借助和 O\*NET-



的误差。一方面,中美两国在职业数量和职业分类标准上存在差异,会因为无法实现精准的一一对应而错误度量中国的职业特征。另一方面,中美两国在技术和经济发展速度上的差异,会使得跨国数据的匹配存在误差。因此,本文先后采用 2014 年中国动态劳动力调查数据、2022 年中国职业技能发展数据库,以及 Chat-GPT4.o 中对中国职业各个特征的测度,将其与我们手工匹配得到的职业特征进行相关性分析。其中,2014 年中国动态劳动力调查数据直接给出了劳动者对所在职业特定特征的打分,代表了中国劳动者的技能测度。2022 年中国职业技能发展数据库基于我国就业招聘市场上最为常见的 500 多个工作岗位,给出了相应的特征测度,反映了中国最具代表性的职业所需的技能特征。Chat-GPT4.o 作为人工智能技术驱动的自然语言处理工具,其可以依据大量的语料库对职业名称进行分析,并对各个特征进行测度,代表了基于工作内容的最一般性的职业特征,其指标测度与劳动力市场的国别无关。<sup>①</sup>

表 II 1 给出了本文的指标测度与这三种数据源的相关性分析的结果。列(1)-列(3)中用来对比的指标和数据分别来自于王林辉等(2022),香帅工作室公布的 2022 年中国职业技能发展数据库,以及 Chat-GPT 4.o。可以发现无论是何种数据源和测度方法,都与我们手工匹配得到的职业层面特征显著正相关,且各指标的相关系数均维持在 0.5-0.9 左右。这一显著的正相关性以及较高的相关系数表明,美国的职业特征测度系统同样也适用于评判中国职业的特征,具有一致性。

表 II 1 多个数据源的对照检验

指标	CLDS-2014 (1)	2022 年中国职业技能发展数据库 (2)	Chat-GPT4.o (3)
认知能力	0.6001***	0.8992***	0.6579***
体力能力	0.4185***	0.7974***	0.5839***
社交能力	0.6326***	0.8249***	0.7916***
阅读能力	0.4592***	-	0.6671***
重复性体力	-	0.9056***	0.7790***
重复性脑力	-	0.7364***	0.3350***

注:考虑到 CLDS 以及中国职业技能发展数据库中样本的限制,相关性分析主要在三维码的职业层面展开。每一行的系数代表斯皮尔曼等级相关系数,上标\*\*\*, \*\*和\*分别代表在 1%, 5%和 10%水平上显著。由于 2014 年的 CLDS 和 2022 年中国职业技能发展数据库提供的职业特征有限,因此部分指标无法进行相关性分析的交叉检验,以“-”来表征缺失值。

<sup>①</sup> 本文参考 Chen et al., (2024)的方法,在 Chat-GPT 4.o 中询问的语句范式为:“请假装你是人力资源专家,你可以基于下面的问题[某一特征/能力对职业的重要性],对下面列出的职业在相应问题上进行打分吗?其中,1 为不重要,2 为有点重要,3 为重要,4 为非常重要,5 为极其重要。请给出你的打分。具体需要你进行打分的职业是:[所有职业小类]。”同时,为了确保 Chat-GPT 得出的指标测度可比,笔者在构建 Chat-GPT 指标测度时,具体使用的细分指标与附表 I 3 一致,向 Chat-GPT 4.o 中询问的语句范式中[某一特征/能力对职业的重要性]也均与 O\*NET 数据库中的问卷一致。

为了更为直观的展示各个指标在职业层面的相关关系,本文基于任意三位码职业对应计算得出的本文指标测度,与其他数据源的测度绘制了散点图,如图II 2-1到附图II 2-6所示。我们可以直观地发现明显的线性正相关性。由此表明本文整理得到的数据及其测度在中国的适用性以及一致性。

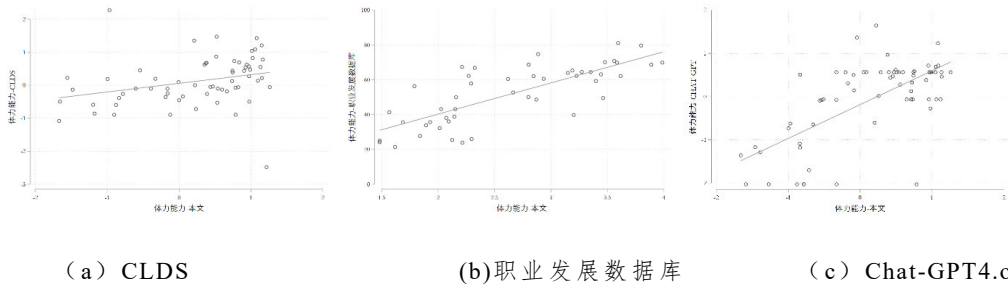


图 II 2-1 体力能力

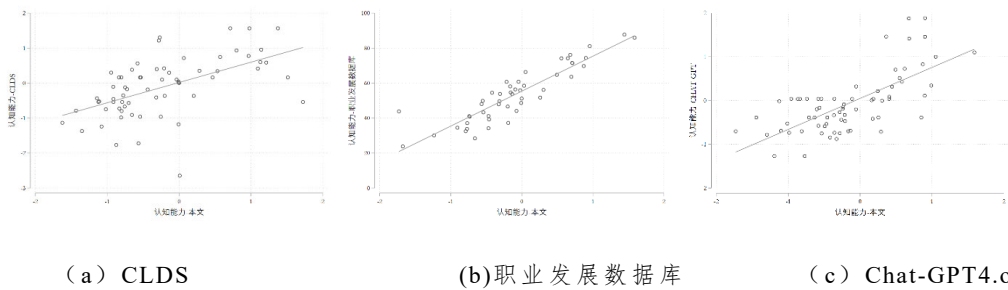


图 II 2-2 认知能力

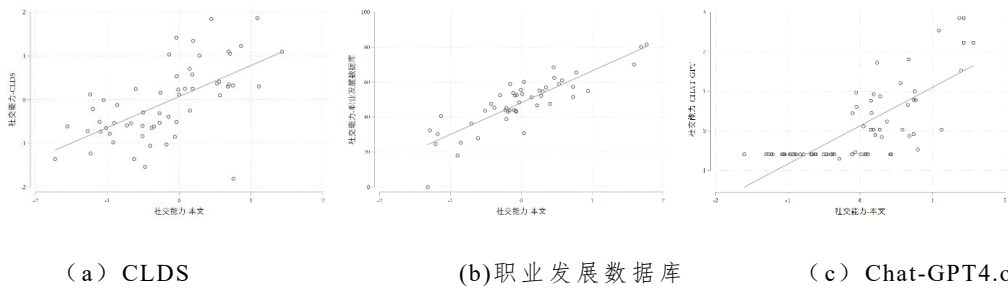


图 II 2-3 社交能力

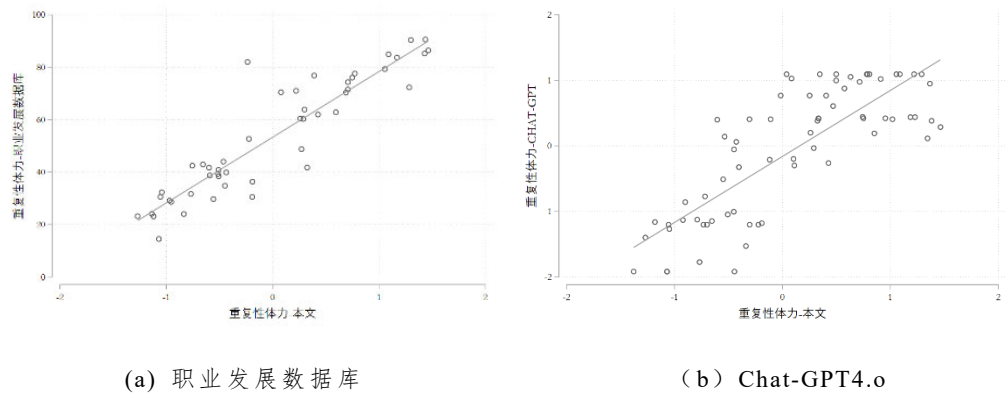


图 II 2-4 重复性体力能力

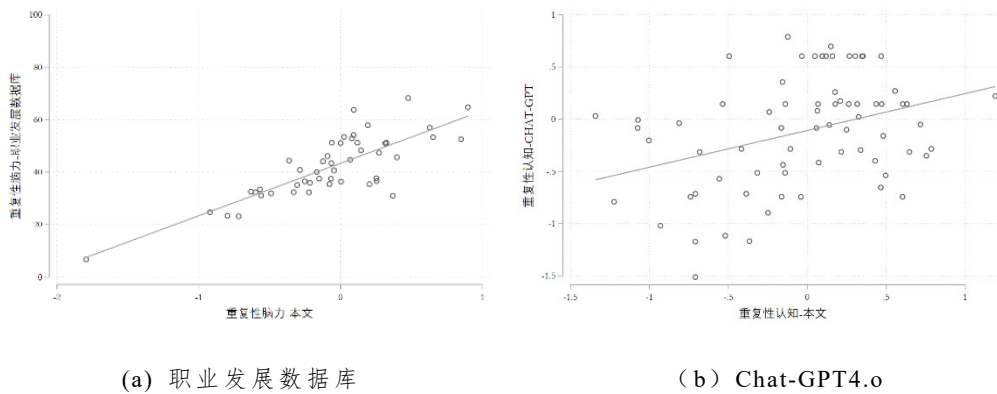


图 II 2-5 重复性认知能力

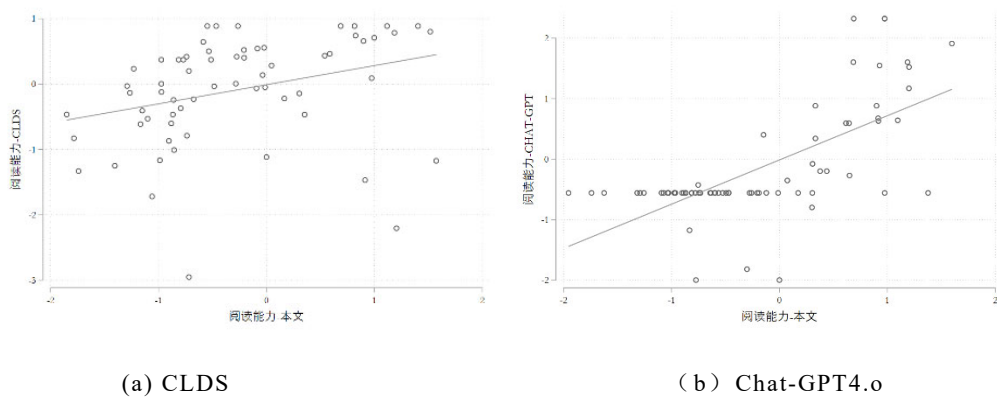


图 II 2-6 阅读能力

### 附录 III 职业特征分布的稳健性分析

当我们讨论职业就业占比变化带来的总体任务倾向的调整时,使用不同版本的 O\*NET 度量职业特征可能会影响我们观察到的技能分布。为此,除了正文中使用的 2020 年发布的 O\*NET 25.0 之外,笔者分别采用 2000、2010 和 2015 年更新的 O\*NET 版本来度量职业特征,也即 O\*NET 5.0、O\*NET 15.1 和 O\*NET 20.0。具体的,本文首先(1)考虑在给定某一 O\*NET 版本时,不同时期职业就业结构变化所产生的技能分布,以及其变化的分布是否有所差异;其次(2)使用微观抽样调查对应年份公布的 O\*NET 版本来度量该年份的技能特征。在考察技能分布在不同 O\*NET 版本下是否有较大差异时,我们主要依据两个统计指标:其一,以 0 为分界点对应的累积就业占比在不同 O\*NET 版本下是否有较大差异;其二,随机占优(Stochastic Dominance),即 2000-2015 年间的技能分布变化的方向在不同 O\*NET 版本下是否有较大差异。前者“0 线”对应的累积就业占比即为某一技能特征表现为较低水平的劳动者占比,可用于判断整体就业市场中各项技能的大致分布。后者包括一阶随机占优(First Order Stochastic Dominance)和二阶随机占优(Second Order Stochastic Dominance),可分别用于判断各项技能分布的均值水平和离散程度在样本期初和期末的变化。

#### (一) 某一版本下的技能分布变化

首先是考察在给定某一 O\*NET 版本时,不同时期职业就业结构变化所产生的技能分布,以及其变化的特征是否有所差异。具体的,我们绘制了 O\*NET 5.0、O\*NET 15.1、O\*NET 20.0 以及 O\*NET 25.0 的职业度量下,2000、2010 和 2015 年职业就业结构所反映的技能分布。图 III 1 中的(a)-(f)给出了我国职业就业占比变化所带来的认知能力、体力能力、重复性认知、重复性体力、协同要求以及竞争要求的分布变化。其中,以 2015 年的技能分布为例,低认知能力在就业市场中的占比为 70%~80%,低体力能力在就业市场中的占比为 30%,低重复性认知在就业市场中的占比为 60%~70%,低重复性体力在就业市场中的占比为 30%~40%,低协同能力在就业市场中的占比为 60%~70%,低竞争能力在就业市场中的占比为 60%。这一“0 线”对应的累积就业占比在不同版本间较为相近,并未因为 O\*NET 版本的不同而产生较大差异。<sup>①</sup>

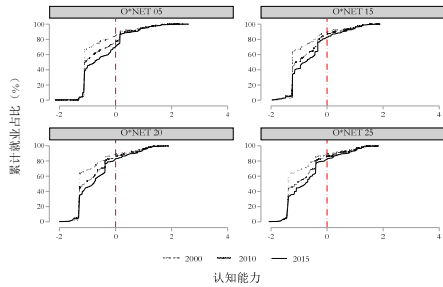
随后,我们观察不同技能分布形态的变化在不同 O\*NET 版本下是否有所差异。图 III 1 (a) 和 (b) 给出了认知能力和体力能力的分布变化。其中,这 15 年

<sup>①</sup> 图 III 1 (d) 展示的竞争要求这一技能分布中,任意年份“0”线对应的累积概率在 O\*NET 05 版本中均低于其他版本,这主要是由于这一 O\*NET 版本中有关“竞争要求”的度量指数存在大量缺失值。笔者在数据处理中主要使用大类职业的平均竞争要求去填补该类别下小类职业的缺失值。因此,可能会使该版本下“0”线表征的技能水平与其他版本存在较大差异。

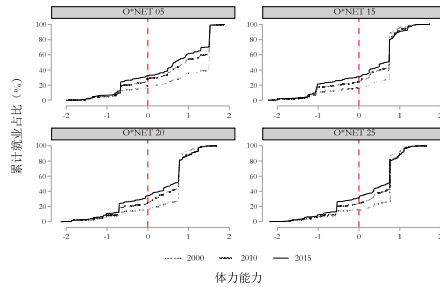
间有关认知能力的累积分布函数逐渐右移,样本期末的就业市场(2015)一阶随机占优于样本期初的就业市场(2000),这意味着劳动者更大概率从事认知能力要求更高的工作;有关体力能力的分布函数逐渐左移,样本期初的就业市场(2000)一阶随机占优于样本期末的就业市场(2015),这意味着劳动有更高的概率从事体力能力要求更低的工作。

当我们比较认知能力和体力能力在重复性工作上的不同要求的分布变化时,结果如图III1(c)和(d)所示。从分布函数的形态来看,这15年间有关重复性认知能力的累积分布函数逐渐右移,样本期末的就业市场(2015)一阶随机占优于样本期初的就业市场(2000),这意味着劳动者更大概率从事重复性认知能力要求更高的工作;有关重复性体力能力的累积分布函数逐渐左移,样本期初的就业市场(2000)呈现二阶随机占优于样本期末的就业市场(2015)的形态,这意味着就业市场在这一指标上的平均水平并未产生很大变化,更多表现为劳动者相较从前更为分散的从事了具备不同重复性体力能力的工作。

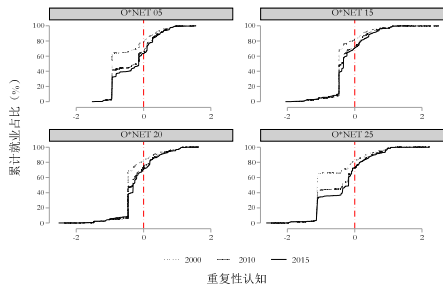
此外,有关竞争要求和协同要求的分布变化结果如图III1(e)和(f)所示。一方面,从分布函数的形态来看,这15年间有关竞争要求和协同要求的累积分布函数均逐渐右移,样本期末的就业市场(2015)一阶随机占优于样本期初的就业市场(2000),这意味着劳动者更大概率从事竞争要求和协同要求都更高的工作。



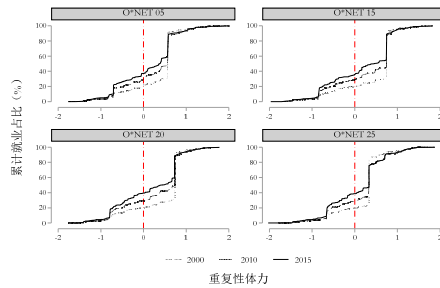
(a) 认知能力



(b) 体力能力



(c) 重复性认知



(d) 重复性体力



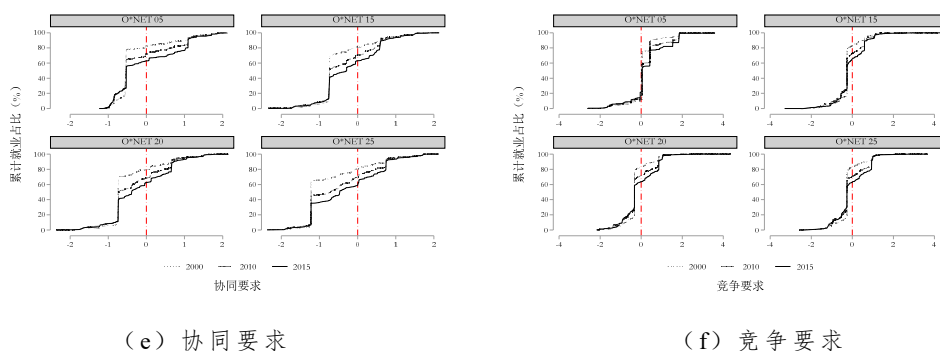
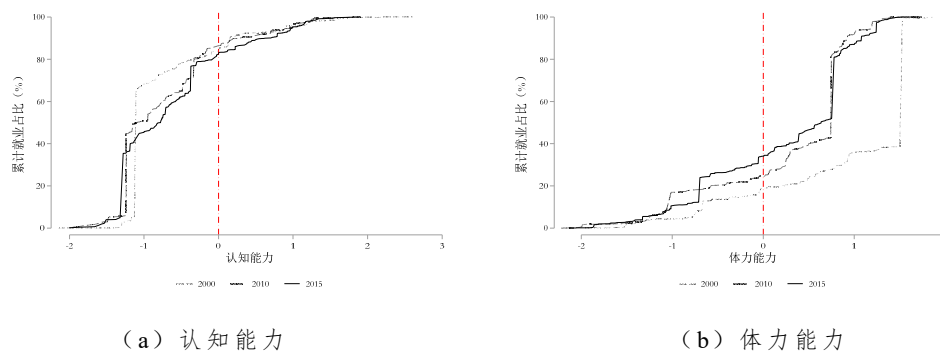


图 III 1 O\*NET 不同版本下的技能分布

(二) 不同版本下的技能分布变化

本文将使用微观抽样调查对应年份公布的 O\*NET 版本来度量该年份的技能特征，如使用 2000 年公布的 O\*NET 5.0 版本度量 2000 年职业就业结构带来的总体技能特征变化，使用 2010 年公布的 O\*NET 15.1 版本度量 2010 年职业就业结构带来的总体技能特征变化，以及使用 2015 年公布的 O\*NET 20.0 版本度量 2015 年职业就业结构带来的总体技能特征变化。

图 III2 (a) - (f) 分别展示了采用这一对应 O\*NET 测度时认知能力、体力能力、重复性认知、重复性体力、协同要求以及竞争要求的分布及其变化。直接观察图 III2 (a) - (b) 可以发现：虽然在这 15 年间劳动者逐渐更多的从事强认知能力和弱体力能力的工作，但是体力型劳动者的就业占比在历年均保持在 70% 以上，而认知型劳动者基本保持在 20%。当我们对比重复性认知和重复性体力这两大特征的分布时，我们可以发现劳动者更大概率从事重复性认知能力要求更高的工作；或是劳动者相较从前更为分散的从事了具备不同重复性体力要求的工作。其中，重复性体力能力等于或低于 0 时的累积就业占比上升了 20%，进一步结合图 III2(b) 中的发现，这意味着虽然就业市场中体力能力仍旧占据绝大部分比例，但是劳动者正在逐渐朝着具有一定精密度的体力生产转型。此外，有关竞争要求和协同要求的分布变化结果如图 III2 (e) 和 (f) 所示。我们可以发现，在这 15 年间，劳动者更大概率在从事竞争要求或者协同要求很高的工作。



(a) 认知能力

(b) 体力能力

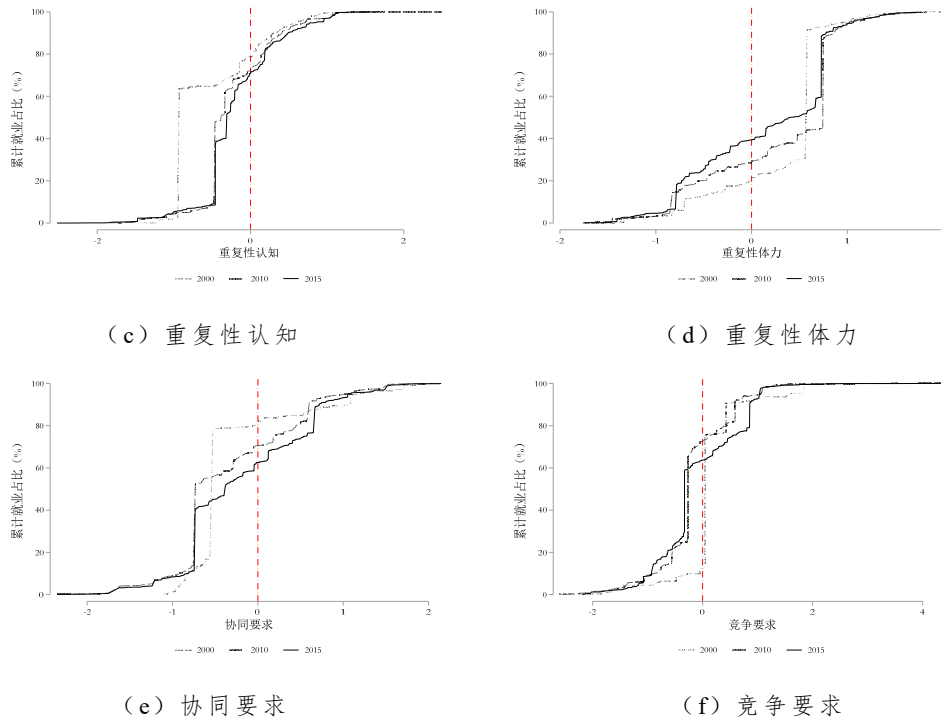


图 III 2 O\*NET 不同版本下的技能分布

据上述的分析可知，无论是在给定某一 O\*NET 版本时，不同时期职业就业结构变化所产生的技能分布，还是使用微观抽样调查对应年份公布的 O\*NET 版本来度量该年份的技能特征，均不会影响笔者从职业就业结构调整中观察到的整体就业市场任务倾向的变化。

## 附录 IV 其他数据图表

## (一) 职业内行业间和职业内产业间

除了劳动要素在职业间的调整外，劳动要素在职业内的调整亦是理解我国劳动要素配置的关键一环。本文试图结合职业内劳动者在不同生产单位，如三位码的行业和三次产业间的调整进行具体分析。从三位码行业以及三次产业维度观察职业内劳动力调整主要是归因于职业属性与行业属性存在交叉。这一点可以从任一职业内劳动者能够参与的行业和产业数量观察得到。为此，笔者将所有职业分别按照该职业内劳动者涉及的行业种类数、涉及的行业种类数占总行业数的占比、主要行业就业占比和主要产业就业占比的分位数排序，并相应给出了 1th-99th 分位数下的对应指标。表 IV 1 分别展示了 2000、2010 和 2015 年职业内的劳动者在行业 and 三次产业层面的分布情况。其中，以 2015 年为例，大部分职业中劳动者涉及的行业种类为 36 种（位于 50 分位数处），占行业总数的 9.97%；仅在少部分职业中的劳动者可以参与多样化的行业生产活动，即位于 99 分位数处职业的种类为 333 种，占行业总数的 92.15%。结合主要行业的就业占比来看，职业内的劳动者相对分散的参与各个行业的生产，位于 50 分位数处的就业占比为 41.21%。这意味着特定职业的技能可能适用于多个行业。

但是职业属性可能和三次产业一一对应，对比同样位于 50 分位数处的主要产业的就业占比，位于 50 分位数处的就业占比已经高达 93.21%。这意味着特定职业的技能仅适用于三次产业中的一种，与产业属性是高度对应的。劳动要素在职业内基本无法实现从第一产业调整至第二、第三产业，仅能通过换职业来实现换产业，这意味着三次产业间的调整可能与职业间调整存在相似性。

表 IV 1 职业内劳动者所属行业和产业的分布情况

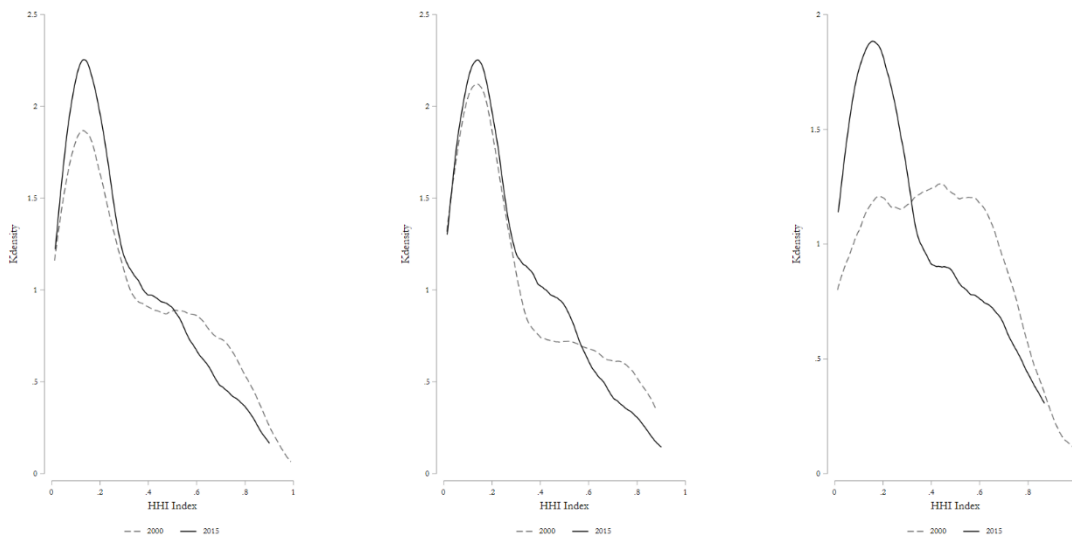
分位数	涉及的行业	涉及的行业	主要行业	主要产业
	种类	种类占比 (%)	就业占比 (%)	就业占比 (%)
<b>Panel A:2000</b>				
1th	13	4.42	4.59	48.51
20th	40	13.56	21.64	73.53
40th	70	23.73	36.10	87.91
50th	85	28.81	45.72	93.27
60th	105	35.59	54.41	95.46
80th	178	60.34	75.93	97.82
99th	293	99.31	93.66	99.82
<b>Panel B:2010</b>				
1th	9	2.32	5.58	41.60

20th	39	9.85	18.17	68.96
40th	63	15.91	27.60	86.78
50th	77	19.44	32.77	90.82
60th	100	25.25	41.84	94.42
80th	173	43.69	63.16	97.25
99th	389	98.23	91.09	99.46
Panel C:2015				
1th	3	0.91	5.22	48.28
20th	18	4.99	22.11	75.66
40th	28	7.70	32.98	89.43
50th	36	9.97	41.21	93.21
60th	46	12.74	50.00	95.20
80th	89	24.54	67.90	98.34
99th	333	92.15	92.69	100.00

注:行业和三次产业分类分别参考三位码行业、三大产业在国民经济行业分类 GB/T 4754-2002 标准中的划分。其中,涉及的行业种类占比=(从事职业 $o$ 的劳动者所从事行业的种类/行业总数) $\times 100$ ,主要行业(产业)指,职业 $o$ 的劳动者在从事的各个行业(产业)中相应就业人数最大的行业,主要行业(产业)就业占比=(职业 $o$ 在主要行业(产业)中的就业人数/职业 $o$ 的总就业人数) $\times 100$ 。

为了描绘劳动要素在任一职业内的行业间流动方向,我们首先观察劳动要素的流动方向是表现为相对集中还是分散。我们具体采用赫芬达尔指数(HHI)刻画职业内行业的集中度,以此来测度劳动力在职业内行业间调整集中度的变化(Blom et al., 2021)。图IV 1(a)展示了所有职业2000-2015年间赫芬达尔指数的分布变化。我们可以发现:职业内的行业结构在2000-2015年间趋于分散,且主要表现为较高集中度职业的数量减少和较低集中度的职业数量的显著增加。这一现象表明劳动者依据特定职业的知识和技能,所能参与的生产活动范围在拓宽。

进一步的,考虑到不同职业在经济生产中重要性差异可能会影响到劳动要素的配置方向。因此我们基于职业在这15年间就业占比呈现出的上升和下降态势将其分类为就业占比上升的职业和就业占比下降的职业,并相应绘制其赫芬达尔指数的分布图,分别如图IV 1(b)和(c)所示。我们可以发现无论是在经济活动中的重要性是上升还是下降,两类职业的行业结构都逐渐分散,相应技能所适用的生产范围都在拓宽。而这两类职业的区别在于,呈就业占比下降的职业无论是在2000年还是2015年,其职业赫芬达尔指数的分布都较上升职业更为扁平,尤其是在初始年份2000年时的差异尤其明显。这一方面表明呈现下降态势的职业所表征的知识技能适用的生产活动范围更窄,另一方面也暗示了较窄的行业覆盖度可能是劳动要素从这类职业中流出的原因。



(a) 所有职业 (b) 就业占比上升的职业 (c) 就业占比下降的职业

图 IV 1 2000-2015 年职业赫芬达尔指数 (HHI) 分布变化

(二) 非农职业结构反映的技能调整

考虑到 2000-2015 年覆盖了我国城镇化的关键时期, 农民就业占比巨幅下降。这一调整可能主导了我国就业市场大部分的结构变化和反映出来的技能分工调整。因此, 我们在这一小节中将剔除第五大类职业: 农、林、牧、渔业生产及辅助人员, 并对我国劳动力市场的技能需求调整做再一步的考察。总体来看, 剔除农民后的劳动力市场中各项技能分工的分布差异都在缩小, 并且有些技能的调整方向上也与整体就业市场不同。

图 IV 2 展示了我国非农劳动力在 2000、2010 和 2015 年各项工作技能上的分布变化。具体的, 如图 IV 2 (a) 所示, 认知能力在样本期初和期末的随机占优性完全翻转: 2000 年的就业市场一阶占优于 2015 年的就业市场, 这意味着劳动者更大概率从事认知能力更低的工作, 且认知能力等于或低于 0 时的概率提升了 5% 左右。而在图 IV 2 (b) 中, 体力能力的随机占优特征并不明显, 劳动者相对更少的从事体力能力要求高的工作, 这一点与整体就业市场的变化一致。同时, 当我们比较认知能力和体力能力的概率分布时, 我们发现从事认知型和体力型工作的劳动力的绝对占比略微调整, 认知型工作占比上升, 整体维持“3-7 分”或者“4-6 分”的格局。

比较图 IV 2 (c) 和 (d) 中的重复性认知能力和重复性体力能力可以发现: 重复性认知能力和重复性体力能力的累计分布函数均左移: 样本期初的就业市场 (2000) 呈现一阶随机占优于样本期末的就业市场 (2015) 的形态, 这意味着劳动者更少的从事重复性认知和重复性体力强度高的工作, 与整体就业市场的分布变化是不同的。并且, 重复性认知和体力能力等于或低于 0 时的概率均上升了 5%。综合两个维度的分析, 我们可以得到非农劳动者更多的在从事需要一定思维能力

和复杂劳动的生产活动中。

最后，我们对图 IV 2 (e) 和 (f) 中的竞争要求和协同要求的分布变化。其中，有关竞争要求的分布变化与整体就业市场相同，且竞争要求等于或低于 0 时的概率下降了近 10%。这意味着除农民以外的劳动力市场竞争性增强。而有关协同要求的调整则非常轻微，协同要求的分布并没有发生很大变化。

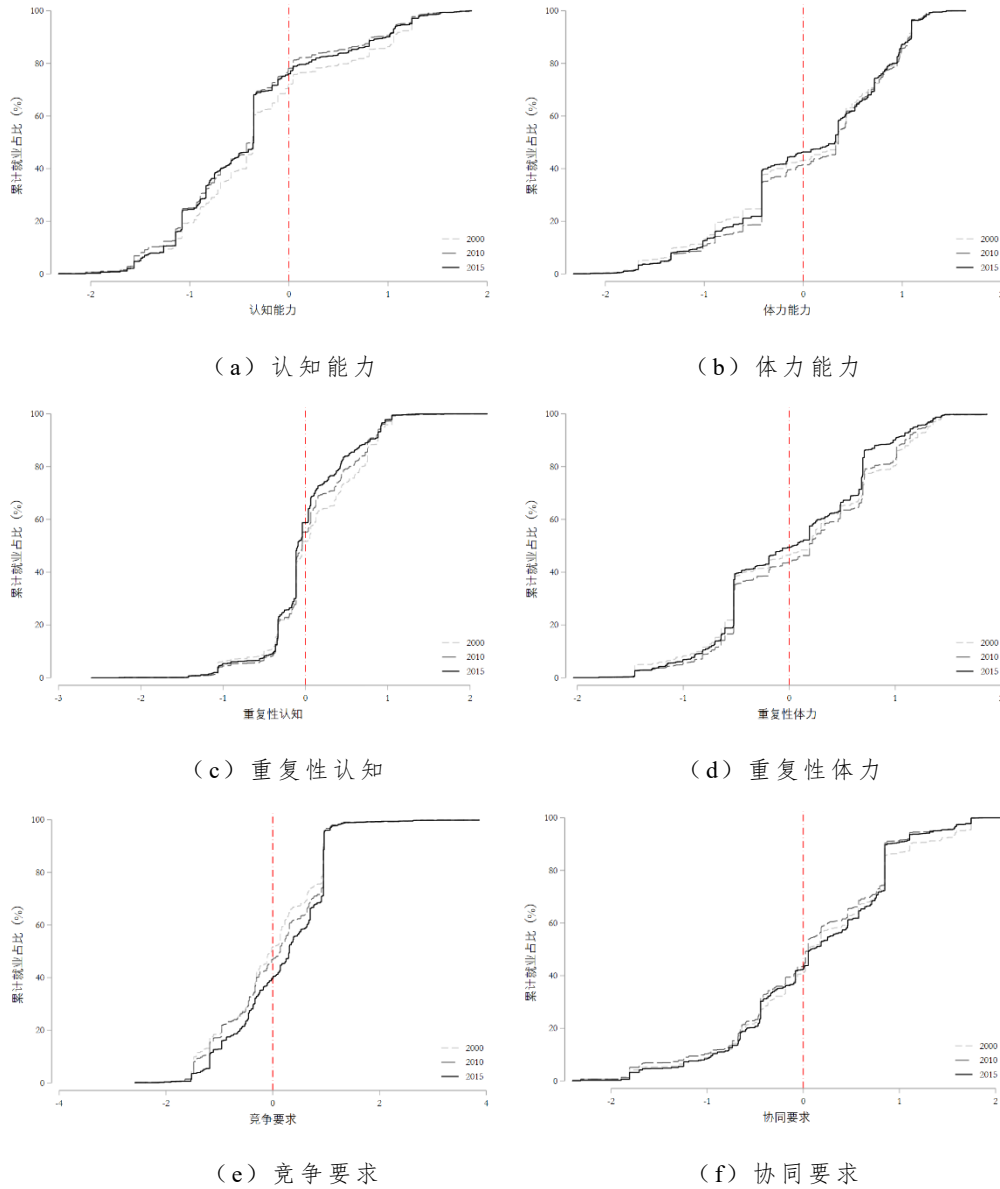


图 IV 2 2000-2015 年非农劳动力工作技能分布变化

(三) 年龄、队列和时期效应

在本文使用的截面数据下存在三种效应：年龄效应，队列效应以及时期效应。其中年龄效应指：随个体生理年龄变化有关的生理和心理差异所产生劳动力市场表现的差异。队列效应指：同一时期出生的群体，由于共同经历的同时代历史事件，在或长或短的时间内所产生的劳动力市场表现的差异。时期效应指：外部环境在某一个时期发生了变化，如社会经济行情等，从而使得不同时期的劳动力市

场表现存在差异。该影响涉及该时期内所有年龄段的群体。在具体分析时,如果我们试图在简单的回归方程中通过分别控制年龄、时期和队列的分类变量从而拆分出年龄效应,队列效应以及时期效应,将会出现共线性问题。简单的回归方程表达式如下,

$$y_{ijk} = \mu + Age_i + Period_j + Cohort_k + \varepsilon_{ijk} \quad (A1)$$

其中  $y_{ijk}$  表征第  $i$  个年龄段,第  $j$  个时期,第  $k$  个队列下的劳动力市场特征,  $Age_i$  为第  $i$  个年龄效应  $Period_j$  为第  $j$  个时期效应,  $Cohort_k$  为第  $k$  个队列效应。 $Age = Period - Cohort$  这一关系,会产生完全共线性,使得我们无法在回归式中准确拆分出三种效应。

为解决同一维度中出现的完全共线性问题,我们借助多层模型 (Multilevel Model) 将截面数据中隐含的年龄效应拆分出来 (Yang et al., 2014; Yang and Land, 2013; 吴晓刚和李晓光, 2021)。首先我们将数据整理成 APC (Age-Period-Cohort) 标准形式,将样本涉及的劳动年龄 15-64 岁,每 5 个年龄为一组分成 10 组。结合关注的 3 个时期: 2000, 2010 和 2015, 我们可以借助这一  $10 \times 3$  的维度计算各个格点表征的劳动力市场特征: (1) 可选职业种类占比 = (该格点在岗劳动者所涉及的职业种类数/该时期整体劳动者所从事职业种类的最大值)  $\times 100$ ; (2) 可选职业-行业种类占比 = (该格点在岗劳动者所涉及的职业-行业种类数/该时期整体劳动者所从事职业-行业种类的最大值)  $\times 100$ 。(3) 该格点在岗劳动者在体力能力、认知能力、重复性认知、重复性体力、竞争要求和协同要求上的平均水平。

其次,具体的估计模型主要分为两层,第一层为个体层面模型,在这一层中将年龄效应视同低层次的固定效应来估计:

$$y_{ijk} = \beta_{0jk} + \beta_1 Age_i + \varepsilon_{ijk} \quad (A2)$$

其中,  $y_{ijk}$  表征第  $i$  个年龄段,第  $j$  个时期,第  $k$  个队列下的劳动力市场特征,具体的劳动力市场特征的指标按照上述阐述的构建。 $\beta_{0jk}$  为模型的截距项,  $Age_i$  代表第  $i$  个年龄效应。 $\varepsilon_{ijk}$  为误差项。

第二层模型主要考虑时期和队列的随机效应,具体的数学表达式如下:

$$\beta_{0jk} = \gamma_0 + Period_{0j} + Cohort_{0k} \quad (A3)$$

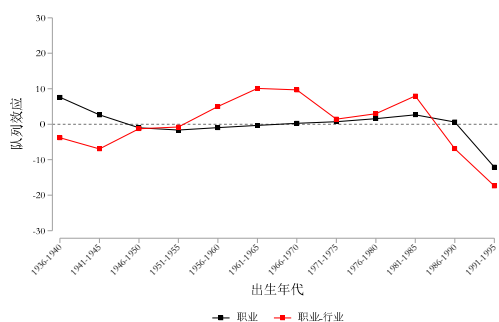
其中,  $\gamma_0$  为模型的截距项,  $Period_{0j}$  代表第  $j$  个时期效应,  $Cohort_{0k}$  代表第  $k$  个队列效应。最后我们可以得到多重线性模型的最终形式如下

$$y_{ijk} = \gamma_0 + Period_{0j} + Cohort_{0k} + \beta_1 Age_i + \varepsilon_{ijk} \quad (A4)$$

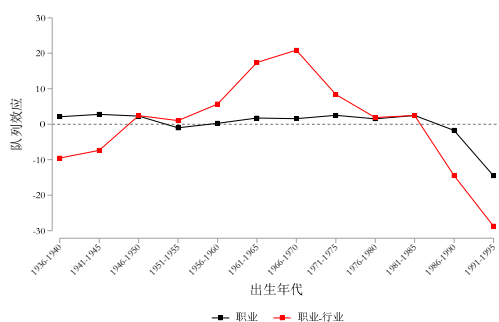
笔者通过式 (2) - (3) 的两层嵌套模型,将得出的年龄效应、时期效应以及队列效应,并绘制成图形。

图 IV 3 展示了分教育水平的可选职业或职业-行业种类的队列效应。从不同队列面临的就业机会来看,职业-行业的范围变动要比职业的范围变动更为明显,这一点与我们在年龄效应中的观察一致。从整体就业市场,图 IV 3 (a), 我们发

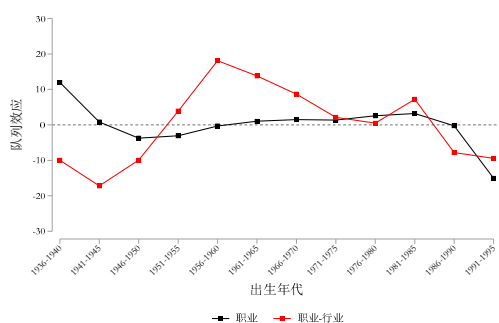
现：20 世纪 50 年代到 70 年代，以及 20 世纪 80 年代初出生的劳动者相对会比其他劳动者拥有更为广泛的行业选择，在就业市场中的各行各业均占据一席之地。而 1985 年之后出生的一代面临的职业，或职业-行业种类均要比其他代际少。由于低教育水平劳动者占比较高，因而这一整体特征可能主要与低教育水平劳动者面临的时代浪潮有关。如图 IV 3 (b) 所示，在低教育水平劳动者中，20 世纪 50 年代-20 世纪 80 年出生的具有显著更广的就业范围。这可能是由于这代人正经历了我国有企业体制改革、中国加入 WTO 等重要经济历史事件，大批劳动力或被动或主动的开始流动到其他职业和行业中去。而 20 世纪 80 年代之后出生的低教育水平劳动者所涉及的职业或者行业种类迅速下降。一定程度上反映了人力资本在当下就业市场的重要性。图 IV 3 (c) 中展示的是中等教育水平劳动者的队列效应。对于这部分学历的劳动者，其变化特征与低教育水平类似。值得关注的是图 IV 3 (d) 中展示的是高等教育水平劳动者的队列效应。我们发现：队列效应是随着出生年份的增加而逐渐增强的，这可能也从另一个角度反映出人力资本随着时代的发展，在就业市场中越来越重要。



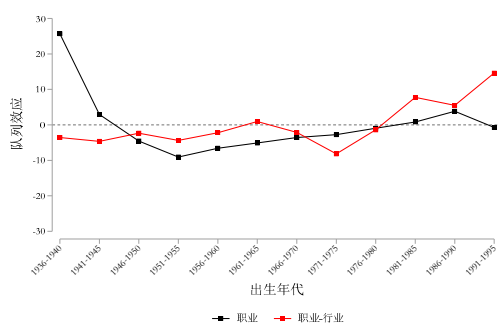
(a) 总体劳动力



(b) 低教育水平



(c) 中等教育水平



(d) 高等教育水平

图 IV 3 可选职业或职业-行业种类的队列效应 (分教育水平)

图 IV 4 展示了分教育水平的可选职业或职业-行业种类的时期效应。从图 IV 4 (a) 中，我们可以发现我国整体就业市场中 2010 年是相对来说劳动者就业机会较多，企业生产活动较丰富的年份。而对于不同年龄劳动者而言，这一“黄金时期”有所不同。低教育劳动者的“黄金时期”为 2000 年，中等教育劳动者为



2010年,而高等教育劳动者在2000年和2010年均表现出了较为丰富的就业种类。

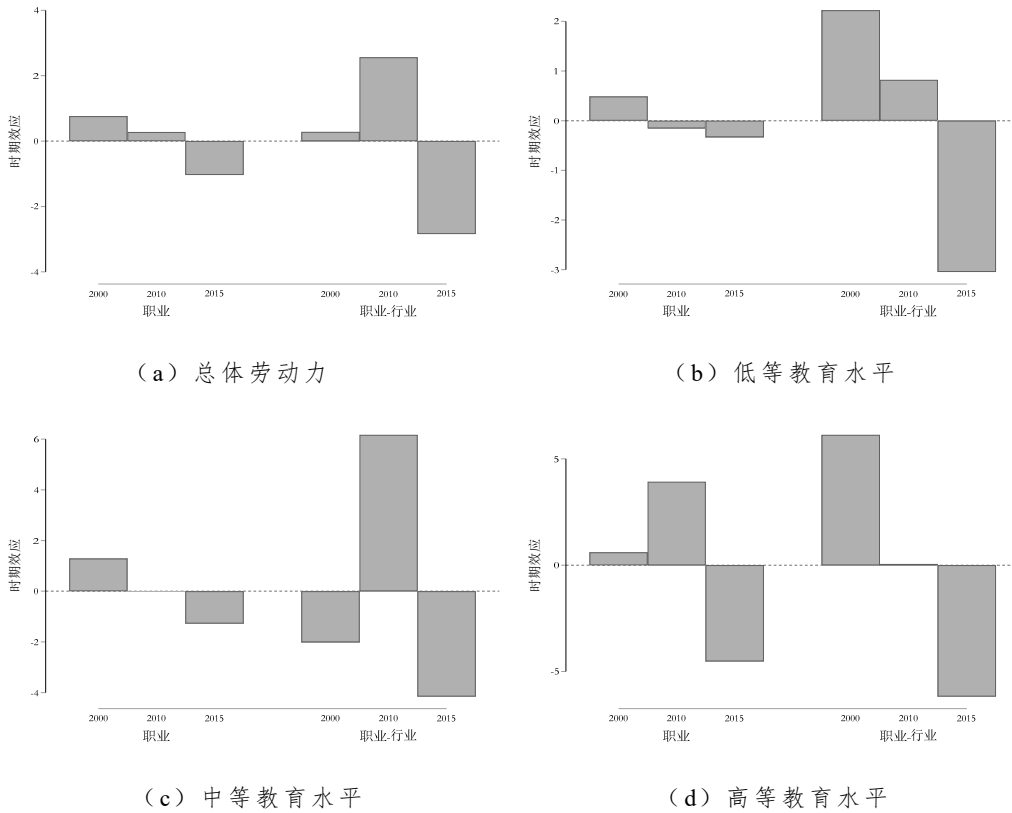


图 IV 4 可选职业或职业-行业种类的时期效应(分教育水平)

图 IV 5 (a) 给出了技能特征的队列效应,我们可以发现非常有趣的队列分工变化:即 20 世纪 50 年代-20 世纪 70 年代的劳动者主要参与简单体力生产,而较少进行脑力、协同性和竞争性要求比较高的工作。而 20 世纪 70 年代以后的劳动者更多参与脑力、协同性和竞争性要求比较高的工作,而较少参与简单体力生产。这反映了我国代际间的分工变化,与我们在年龄效应处观察到的“年龄分工”得以相互补充。图 IV 5 (b) 给出了技能特征的时期效应,可以发现有关简单体力生产相关的时期效应在下降,有关脑力、协同性和竞争性的效应在上升。这一点也与笔者在正文中发现的整体技能分布的变化相一致。

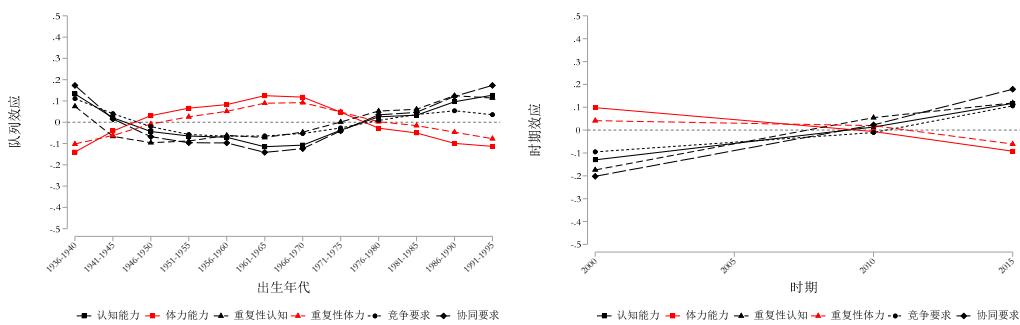


图 IV 5 技能特征的队列效应和时期效应

## (四) 人力资本就业结构

表 IV 2 展示了使用 2000 年和 2010 年的女性社会地位调查数据,2012-2018 年中国家庭追踪调查数据的结果。通过这两项补充数据计算得出的职业所需教育水平与在岗劳动者教育水平。我们可以在表中得到和正文一致的发现:其一,对教育水平要求较低的职业(初中学历及以下)恰能得到教育水平相匹配的劳动者,历年低教育水平劳动者就业占有所有教育水平劳动者的绝大部分,基本适配。其二,对教育水平要求为中等的职业(高中学历),存在“人力资本缺乏”的问题,这类对知识储备和技术有一定要求的职业目前我国仍然主要以低教育水平劳动者为主,历年均超过其他教育类别的劳动者。反映了我国整体中职教育劳动尚存在较大缺口。同时,值得关注的是,在这类职业中,高等教育水平劳动者的比例存在逐年上升趋势,反映了由于中职教育劳动者的稀缺而高等教育职业需求不足,高等教育劳动者对这部分缺口补足。其三,教育水平要求较高的职业(大专及以上学历)整体表现出适配度逐渐提升的趋势。因此,总的来说,用其他微观数据计算得到的发现与本文使用的微观人口普查抽样数据得到的结果是一致的,表明本文数据结果的可靠性和稳健性。

表 IV 2 职业所需教育水平与在岗劳动者教育水平

年份	低等教育 (%)	中等教育 (%)	高等教育 (%)
<b>Panel A: 职业要求低等教育</b>			
2000*	64.90	29.70	5.40
2010*	85.54	12.45	2.01
2012	91.93	7.24	0.83
2014	91.33	7.47	1.20
2016	90.80	7.79	1.40
2018	89.89	8.15	1.96
<b>Panel B: 职业要求中等教育</b>			
2000*	79.68	15.90	4.43
2010*	46.93	30.73	22.35
2012	64.87	21.45	13.68
2014	62.28	22.09	15.64
2016	60.43	21.63	17.94
2018	54.86	23.68	21.46
<b>Panel C: 职业要求高等教育</b>			
2000*	19.36	42.64	38.00
2010*	15.62	27.72	56.66

2012	24.82	25.54	49.64
2014	36.21	23.60	40.18
2016	44.83	19.51	35.66
2018	29.92	19.69	50.39

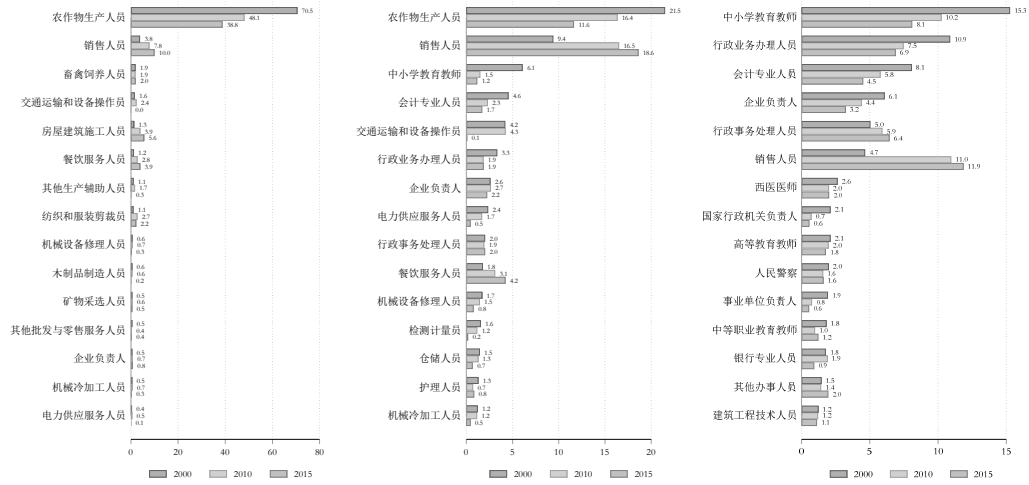
注：标\*的 2000 和 2010 年的数据分别来自于 2000 年和 2010 年的女性社会地位调查数据，2012-2018 年的数据来自于中国家庭追踪调查数据。

我们还可以结合不同教育水平劳动者所选职业的具体类目来进一步验证人力资本可能并未流向亟需人力资本的生产活动。具体的，我们选取低、中、高教育水平劳动者在 2000 年时各自就业占比排名前十五的职业，并绘制其在 2000、2010 和 2015 年三个年份的就业占比变动，结果如图 IV 6 (a) - (c) 所示。从职业就业占比变动来看，各教育组别内排名前十五的职业就业累积占比均表现了下降的趋势，不同劳动力市场中的职业集聚度下降。如低教育水平组前十五的就业占比从 86.31% 下降到了 65.37%，中等教育水平组从 65.17% 下降到了 47.06%，高教育水平组从 67.14% 下降到了 52.83%。虽然低、中教育水平组别的下降特别显著，但是考虑到低教育水平劳动者在 2015 年前十五的职业占比和仍然达到了 65.37%，该类劳动力市场中的职业集聚度仍然较高。

进一步观察不同教育组内具体的职业类目，如图 IV 6 (a) 显示，在低教育水平的劳动力市场中，劳动者的职业集聚在农作物生产中，即便销售人员或者餐饮服务人员的就业占比逐步提升，但是这一转变幅度仍然有限，各自在 2015 年占比分别为 9.95% 和 3.89%。在中等教育水平的劳动力市场中，如图 IV 6 (b) 显示，这部分劳动者相应具备一定知识技能，流动方向开始转向基础服务业和制造业生产，如销售人员容纳了最多的劳动力。与此同时，像中小学教育教师、会计专业人员这类需要一定人力资本和技术支撑的职业占比在下降。在高等教育水平的劳动力市场中，如图 IV 6 (c) 显示，其劳动力分布主要以提供公共服务的专业人员为主，如教师、公务员、医生、警察、银行人员等，而涉及市场化生产的技术岗位，如建筑工程技术人员等容纳的劳动要素相对来说更少。从该劳动力市场内部的结构变化来看，销售人员取代中小学老师，成为了聚集更多劳动力的职业。

这一调整方向表明：虽然不同类型劳动力市场中，劳动要素在职业层面的分布逐渐分散，但是教育水平较低的就业市场仍然主要从事农作物生产。教育水平中等的就业市场劳动要素开始转向基础服务业和工业制造就业，该市场内的劳动要素从农作物生产转向销售人员。而在高教育水平的就业市场中，劳动者的调整有限，分布以公共部门的服务为主，市场化技术岗位次之。该市场内的劳动要素从教师转为销售人员。这一调整再次表明人力资本，尤其是高等教育劳动力，可能并未流向并未流向亟需人力资本的生产活动。进一步结合我国人力资本供给逐

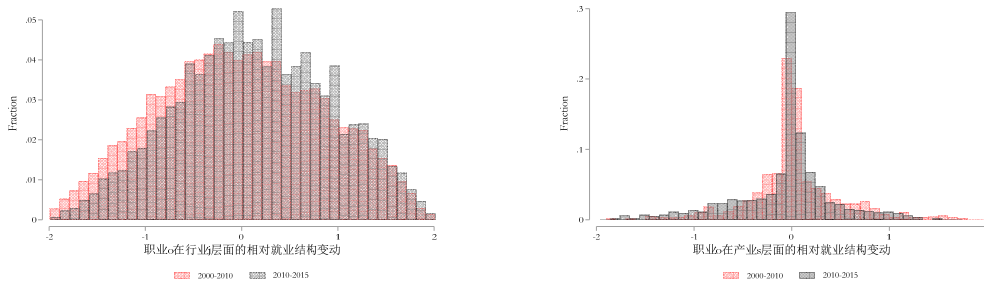
年提升的背景,这一人力资本的流向实际上也暗示了我国技术性岗位有效需求的不足。



(a) 低等教育水平 (b) 中等教育水平 (c) 高等教育水平  
图 IV 6 不同教育水平劳动者排名前十五的职业结构变动

### 附录V 附图

图V1 (a) 和 (b) 分别展示了 2000-2010 以及 2010-2015 年间职业内的劳动要素在行业、三次产业间产生的相对就业结构变动。整体来看,职业内劳动要素在行业层面的结构调整非常普遍,但是在产业层面的调整则十分有限。由此可见,从事特定职业的劳动者可能可以相对较为灵活的转换从事的行业和生产的产 品,但是较难在农业、工业和服务业的三次产业之间自由流动,这一点也应证了我们在行业和产业数量中的推测,即特定职业所需的知识技能可能适用于多个行业,但是无法应用在不同的产业中。



(a) 职业内的行业结构调整

(b) 职业内的产业结构调整

图 V 1 2000-2015 年职业内行业 and 产业的就业结构变动

注: (a) 图中的横轴为职业  $o$  在行业  $j$  层面相对就业占比变动的程度 ( $Re\_d_{ot}^j$ ), 其依照正文公式计算得出。纵轴为职业数占比。(b) 图中的横轴为职业  $o$  在产业  $s$  层面相对就业占比变动的程度 ( $Re\_d_{ot}^s$ ), 其依照正文公式计算得出。纵轴为职业数占比。分布图 (a) 和 (b) 中暂不包括新进入的职业分工或者消失的职业分工。

## 参考文献

- [1] Acemoglu, D., and D.H. Autor, “Skills, Tasks and Technologies: Implications for Employment and Earnings”, *Handbook of Labor Economics*, 2011, 1043-1171.
- [2] Cortes, P., and J. Pan, “Occupation and gender”, *The Oxford handbook of women and the economy*, 2018, 425-452.
- [3] Chen, Y., H. Fang, Y. Zhao and Z. Zhao, “Recovering Overlooked Information in Categorical Variables with LLMs: An Application to Labor Market Mismatch”, No. w32327. *National Bureau of Economic Research*, 2024.
- [4] 吴晓刚、李晓光, “中国城市劳动力市场中教育匹配的变迁趋势——基于年龄、时期和世代效应的动态分析”, 《中国社会科学》, 2021 年第 2 期, 第 102-122 页。
- [5] 王亚菲、贾雪梅、王春云, “中国行业层面就业核算研究”, 《统计研究》, 2021 年第 12 期, 第 3-18 页。
- [6] 王林辉、胡晟明、董直庆, “人工智能技术、任务属性与职业可替代风险: 来自微观层面的经验证据”, 《管理世界》, 2022 年第 7 期, 第 60-79 页。
- [7] Yang, Y. and K. C. Land, *Age-period-cohort analysis: New models, Methods, and Empirical Applications*. Taylor & Francis, 2013.
- [8] Yang, Y. W. Fu, and K. Land, “A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models”, *Sociological Methodology*, 2014, 34(1), 75-110.
- [9] 岳希明, “我国现行劳动统计的问题”, 《经济研究》, 2005 年第 3 期, 第 46-56 页。

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处。