

附录 1

考虑到基于网络招聘数据可能存在样本偏差,即该平台数据库只涵盖了通过其发布招聘信息的企业,且这些企业也可能通过其他线上平台或者线下方式招聘劳动力。为了确保本文使用的样本数据能客观反映总体特征,将智联招聘企业数据库、本文使用的样本数据和其他数据的企业用工需求状况和企业特征进行对比,尝试找出存在偏差的维度,使得本文的结论更加客观。通过观察不同数据样本的特征可知:①劳动力受教育程度特征(附表1),智联招聘数据库和本文使用的样本数据对招聘的劳动力的平均受教育年限为15.18年和14.31年,多为大专及以上学历,分别与《非你莫属》招聘数据(刘毓芸和程宇玮,2020)的平均受教育年限15.75年和Kuhn和Shen(2015)所用厦门人才网中平均受教育年限14.35年较为接近,但高于《中国人力资本报告2020》中劳动者平均受教育年限10.15年,这主要是由于前三者都是企业层面的用工需求数据,学历较高的劳动力在企业招聘竞争中更具有优势,而《中国人力资本报告2020》里面包含了受教育程度较低的劳动者,更多反应的是全国劳动力市场的平均水平,智联招聘数据库中缺乏教育程度为小学(受教育年限为6年)的样本也可能使得平均水平略高。②企业规模特征(附表2),按照国家统计局规定的工业企业规模划分标准,将企业从业人员数2000及以上划为大型企业,2000以下划为中小型企业,将2014和2015年本文使用数据样本和中国工业企业数据库进行对比。附表2显示,中小型企业均占比例较大,且数值较为接近,说明大部分制造业企业为中小型企业。③企业的行业分布(附图1),将2014年本文使用的样本数据、智联招聘数据库与工业企业数据库的各行业企业数量进行对比,结果显示,企业在各行业的分布数量在各数据库中表现一致,不存在偏差。④企业的所有制特征,通过观察样本观测期的企业类型发现,非国有企业占比在本文使用的最终数据样本、中国工业企业数据库中平均分别为99%和98%,说明大部分制造业企业为非国有企业。

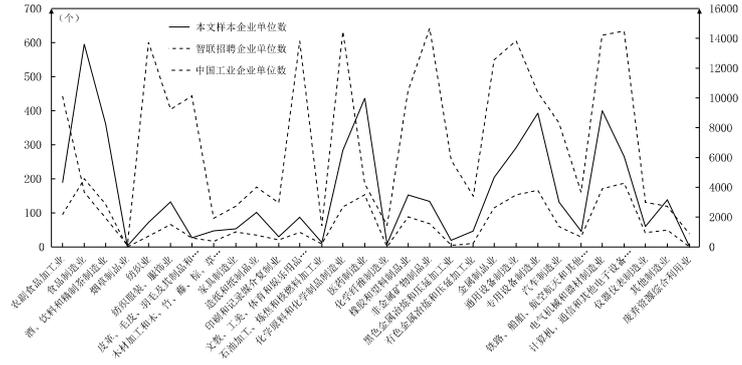
综上,通过对本文所用样本、智联招聘数据库以及其他现有研究的样本之间的比较,可以说明本文选择的样本数据一定程度上对总体具有较好的代表性,基于该数据样本得到的实证研究结论可信度高。

表1 不同数据库的劳动力特征对比

		厦门人才网	《中国人力资本报告2020》	《非你莫属》数据	智联招聘数据库	本文使用样本数据	
劳动力特征	教育程度(年)	14.35	10.15	15.75	15.18	14.31	
	工作经验占比(%)	经验不限	/	/	/	41.38	38.41
		经验1~10年	/	/	/	57.74	59.18

表2 2014-2015年不同数据库的制造业企业特征对比

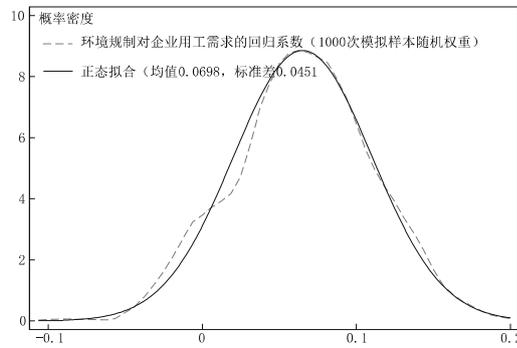
		使用样本数据		中国工业企业数据
企业特征	大型企业占比(%)	2014	3.51	2.37
		2015	6.69	1.79
	中小型企业占比(%)	2014	96.49	97.63
		2015	93.31	98.21



附图 1 不同数据库的制造业企业的行业分布对比 (2014 年)

附录 2

由于本文使用的智联招聘数据样本可能存在样本抽样偏差,导致因样本与总体的差异问题致使方程的估计系数结果有偏、不一致,故参考刘毓芸和程宇玮(2020)的方法,以正文的表3中的列(4)为基准回归,通过对样本数据随机再抽样方法重新评估估计系数,将上述过程重复模拟1000次,分别生成1000个 β_1 的样本估计系数 $\hat{\beta}_1$,画出这1000个估计系数 $\hat{\beta}_1$ 的概率密度分布图,以此作为 β_1 的真值分布。将正文基准回归结果表3中的 β_1 与其进行比较,判断样本偏差问题是否存在。如果基准回归系数会与 $\hat{\beta}_1$ 的真值存在显著差异,说明样本确实存在样本偏差问题;如果基准回归系数会与 $\hat{\beta}_1$ 的真值不存在显著差异,说明样本不存在样本偏差问题。附图2为随机再抽样的检验结果,即1000次模拟的估计系数分布。观察可知,该经验分布的标准差为0.0451,分布的均值为0.0698。正文的基准回归结果表3中,在控制企业和年份固定效应的情况下(列(4)),估计系数约为0.0694,说明二者之间不存在显著差异,因此,本文采用的数据样本不存在严重的样本选择偏差问题。



附图2 对观测值生成1000次随机抽样的回归系数与基准回归系数对比

附录3

附表3 政府环境规制对制造业企业用工需求的影响:工具变量估计

解释变量	被解释变量: <i>er</i>	被解释变量: <i>job</i>	被解释变量: <i>er</i>	被解释变量: <i>job</i>	被解释变量: <i>er</i>	被解释变量: <i>job</i>
	第一阶段回归	第二阶段回归	第一阶段回归	第二阶段回归	第一阶段回归	第二阶段回归
	(1)	(2)	(3)	(4)	(5)	(6)
<i>avger</i>	1.0244*** (0.0305)				1.0185*** (0.0310)	
<i>inversion</i>			1.0975*** (0.1982)		0.3264* (0.1741)	
<i>er</i>		0.2501*** (0.0618)		0.7966** (0.3783)		0.2559*** (0.0965)
观测值	5973	5973	5973	5973	5973	5973
Kleibergen -Paap Wald rk F 统计 量		1130.485		30.668		574.237
Kleibergen -Paap rk LM 统计量		388.862***		27.668***		407.798***
Hansen J 统计量		/		/		2.334 (0.1266)
Hausman 统计量		12.72 (0.0792)		3.42 (0.8437)		13.40 (0.0629)

注: Kleibergen-Paap Wald rk F 统计量为弱识别检验; Kleibergen-Paap LM 统计量为不可识别检验; Hansen J 统计量为过度识别检验; Hausman 统计量为内生性检验。

附录4

附表4 依据行业污染密集度划分的制造业行业

重污染行业（14个）	轻污染行业（17个）
农副食品加工业、食品制造业、酒、饮料和精制茶制造业、纺织业、皮革、毛皮、羽毛及其制品和制鞋业、造纸和纸制品业、石油加工、炼焦和核燃料加工业、化学原料和化学制品制造业、医药制造业、化学纤维制造业、非金属矿物制品业、黑色金属冶炼和压延加工业、有色金属冶炼和压延加工业、金属制品业	烟草制品业、纺织服装、服饰业、木材加工和木、竹、藤、棕、草制品业、家具制造业、印刷和记录媒介复制业、文教、工美、体育和娱乐用品制造业、橡胶和塑料制品业、通用设备制造业、专用设备制造业、汽车制造业、铁路、船舶、航空航天和其他运输设备制造业、电气机械和器材制造业、计算机、通信和其他电子设备制造业、仪器仪表制造业、其他制造业、废弃资源综合利用业、金属制品、机械和设备修理业

注：虽然《重点排污单位名录管理规定（试行）》文件中将印刷和记录媒介复制业、专用设备制造业和汽车制造业中的少量四位数代码子行业列入重点监管行业，占该部分占整个行业比重低，且这三个行业的污染排放强度偏低，故本文将其纳入轻污染行业范畴。