

**本科毕业设计[论文]**

**基于改进PageRank算法的企业商业网络地位对全要素生产率影响的研究**

院 系 管理学院

专业班级 信管1801班

姓 名 卢冠文

学 号 U201816009

指导教师 李晓艳

2022年 5 月 31 日

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密 囗 ，在 年解密后适用本授权书。

2、不保密 囗 。

（请在以上相应方框内打“√”）

作者签名： 年 月 日

导师签名： 年 月 日

摘 要

供应链在实体企业当中占据着至关重要的位置，优秀的供应链管理能够帮助企业极大程度降低生产成本，提升企业生产运作效率，从而获取更高的收入和利润；而稳定、可靠的供应商以及客户，能够帮助企业降低供应链风险，优化企业生产运输链条，确保供需关系平衡，避免供应商或者客户违约导致合同无法按期完成、资金链断裂等问题，还可以在稳定的供应关系中互惠互利，使得中小企业依靠供应链获得更低的融资成本。

现代商业网络让企业之间的关系变得更加复杂，企业之间的联系更加紧密，企业的全要素生产率也在随着技术进步、企业结构改革而逐步提升。本文基于2009-2019年上市公司前五大供应商和前五大客户关系构建供应商—客户的商业网络，用改进后的PageRank算法量化企业在网络中的地位以及影响力，同时使用LP法对制造业上市公司的全要素生产率进行了估计，并考察企业在商业网络的影响力对于企业全要素生产率的影响。研究发现，我国商业网络中企业数量和企业之间联系整体呈现逐年增加趋势，但整体网络密度没有过大的变化，说明网络中的企业数量在增加时，企业间的联系保持较为稳定的状态。实证结果表明，企业网络的影响力会对全要素生产力产生影响，且企业在网络地位的提高能够促进全要素生产率的增长。

**关键词：**商业网络；加权PageRank；全要素生产率；LP法

Abstract

Supply chain plays a crucial role in entity enterprises. Excellent supply chain management can help enterprises greatly reduce the cost of production, improve production efficiency and thus obtain higher income and profits. Besides, stable and reliable suppliers and customers contribute to lower supply chain risks, optimize the production and transportation chain, ensure the balance between supply and demand, and avoid suppliers' or customers' default, which may results in the contract cannot be completed on schedule and capital chain repute. Enterprises can also be mutually beneficial in a stable supply relationship, helping SMEs obtain lower financing costs.

Modern business networks have made the relationship between enterprises more complex and the connection between enterprises closer, at the same time, the total factor productivity of enterprises is gradually improving with technological progress and enterprise structure reform. This paper constructs a supplier-customer business network based on the top five suppliers and top five customers relationships of listed companies from 2009 to 2019, quantifies the position of enterprises in the network and their influence by the improved PageRank algorithm, Additionally, estimates the total factor productivity of manufacturing-listed companies using the LP method, and examines the influence on the total factor productivity of enterprises status in the business network. The study find that the overall number of enterprises in China's business network and the ties between enterprises showed a trend of increasing year by year, but the network density did not change excessively, indicating that the ties between enterprises remained relatively stable when the number of enterprises in the network was increasing. The empirical results show that the influence of enterprise network will have an impact on total factor productivity, and the increased influence of companies in the network can contribute to the growth of total factor productivity.

**Key Words：**Business Network;Weighted PageRank;Total Factor Productivity;LP method

**目 录**

[摘 要 I](#_Toc105358874)

[Abstract II](#_Toc105358875)

[1 绪论 1](#_Toc105358876)

[1.1 研究背景与意义 1](#_Toc105358877)

[1.1.1 研究背景 1](#_Toc105358878)

[1.1.2 研究意义 2](#_Toc105358879)

[1.2 国内外研究现状及发展趋势 2](#_Toc105358880)

[1.2.1 网络节点重要性排序相关文献综述 2](#_Toc105358881)

[1.2.2 全要素生产率相关文献综述 5](#_Toc105358882)

[1.3 研究内容与研究方法 7](#_Toc105358883)

[1.3.1 研究内容 7](#_Toc105358884)

[1.3.2 研究方法 8](#_Toc105358885)

[2 企业网络地位度量 9](#_Toc105358886)

[2.1 商业网络构建 9](#_Toc105358887)

[2.2 节点重要性排序方法选择 10](#_Toc105358888)

[2.3 PageRank算法应用到商业网络解析 12](#_Toc105358889)

[2.3.1 算法基本原理 12](#_Toc105358890)

[2.3.2 商业网络地位计算及分析 13](#_Toc105358891)

[3 企业全要素生产率度量 16](#_Toc105358892)

[3.1 全要素生产率度量方法 16](#_Toc105358893)

[3.2 数据来源与TFP估计 18](#_Toc105358894)

[4 企业网络地位与全要素生产率实证分析 20](#_Toc105358895)

[4.1 研究假设 20](#_Toc105358896)

[4.2 数据来源及变量选取 20](#_Toc105358897)

[4.3 变量描述性统计 21](#_Toc105358898)

[4.4 模型的参数估计以及结果分析 22](#_Toc105358899)

[4.5 稳健性检验 24](#_Toc105358900)

[5 结论与展望 27](#_Toc105358901)

[致谢 29](#_Toc105358902)

[参考文献 30](#_Toc105358903)

[附录 33](#_Toc105358904)

1 绪论

1.1 研究背景与意义

1.1.1 研究背景

2022年的3月5日，李克强在十三届全国人民代表大会五次会议上作《政府工作报告》，多次提及供应稳定问题。在制造业方面，政府工作报告指出，需要提高制造业企业的核心竞争力，强化原材料和关键零部件的攻击保障，依靠龙头企业进行保链稳链工程，提高产业链、供应链的安全性和稳定性。可见，供应链稳定是是制造业提升竞争力、实现数智化转型的重要条件。在这一大背景下，大多企业选择抱团取暖，增强与主要客户的合作程度，在信息、资源、技术上进行了适当共享，打造牢固的利益共同体。目前，已有大量学者在客户集中度这一指标出发，考察企业与主要客户的依赖程度对于企业自身财务等指标所产生的影响；另外，还有部分学者研究供应链金融对中小企业融资产生的影响（张楚秦，2019），但度量供应链金融化程度的指标主要是通过财务报表的数据进行大致预测，只有较少学者从构建供应链网络、量化网络中企业信息（宋华、卢强，2017）这一角度出发对企业进行深层次的分析。

随着通信技术的发展以及国家对中小企业的扶持力度的增大，现代商业网络相较于传统的商业关系而言，企业数量更多，企业之间的关系更加复杂，企业与上游供应商以及下游客户之间会因为原材料和产成品供应、服务的提供等而产生不同程度的关联，在这种关联中也带来了企业之间信息的共享、技术互补等行为，大多数企业都希望能够拥有影响力强、生产运作稳定的供应商，因为其在一定程度上确保了自身良好的发展前景，减小供应链风险，更容易提升自身企业的规模以及市场影响力。且对于小型企业而言，优质的供应商甚至可以一定程度上帮助自身企业缓解融资约束，通过供应链金融获得更低的融资成本。而对在网络中有较高影响力的企业而言，他们能够利用自身企业的优势寻找更优质的客户，降低自身信息的不对称性，从而提升自身生产效率。

全要素生产率是衡量一个国家、行业甚至一个企业中对于资源的开发和利用效率的指标，他代表了企业产出中不能由要素投入解释的剩余部分，这部分与企业的技术、生产管理能力等指标息息相关，而企业较高的影响力或许能够给企业带来更好的技术和规模效应。基于以上猜测，本文用上市企业前五大客户和供应商搭建商业网络，并基于PageRank算法量化企业在商业网络中的影响力以及基于道格拉斯生产函数对全要素生产率进行估计，发现企业网络地位与全要素生产力之间存在的关联。

1.1.2 研究意义

1）理论意义

在商业网络以及供应链网络的度量上，国内学者仅用出度、入度、中介中心度等指标来量化网络指标，但这一类指标忽略了企业之间交易的金额，但企业的度和中心性并不能很好地反应企业在网络中的地位。尽管也有学者使用过PageRank算法量化企业网络地位（包群、但佳丽，2021），但事实上并未把企业之间交易金额（即网络中边的权重）参与到企业影响力的计算当中。因此，本文以PageRank算法为切入点，参考国内外学者对该算法的改进方式，对企业的网络地位进行了度量，并实证分析企业影响力与全要素生产力之间可能存在的关联，在对于网络地位的量化提供新思路的同时，尝试去寻找企业网络地位对企业可能产生的影响。

2）现实意义

受到2020年疫情冲击影响，大量企业由于疫情冲击，企业内部生产停滞，货物运输受阻，进而在供应链上产生一系列连锁反应，大量的企业需要重新寻求合适的供应商和客户，而在这一过程中，也是企业获取新信息和新资源、谋求更好发展的一个机遇，通过对上市企业网络的分析，对于企业寻找合适的合作伙伴，发现提升自身全要素生产率的新途径具有指导意义。

1.2 国内外研究现状及发展趋势

1.2.1 网络节点重要性排序相关文献综述

在复杂网络中，对于节点的重要性排序一直是一个备受关注的问题，通过构建网络并对其中节点的网络地位和影响力进行量化，能够从中挖掘许多有实用价值的信息。而在对于网络中节点的重要性度量上，不同的网络类型也有不同的度量方法。

首先，较为广泛使用且简单易懂的度量方法是用度中心性。该方法聚焦于节点的邻居数目，认为一个节点的度（包括出度和入度）越高，代表着其在网络中能够影响的节点数量越多，因此该节点的重要程度也就越大。但该方法所使用的信息也很有限，没有考虑到节点的网络位置、节点之间路径等信息。

接近中心性、介数中心性和KatZ中心性等方法则更多地聚焦于节点之间连通路径去对节点的重要程度排序，该类排序方法较为适用于交通网络、通信网络等等，能够通过该方法寻找到网络中传播路径短、对交通运输、信息传播等方面有重要影响力的节点。

上述度量方法更侧重于使用节点的邻居数量对节点重要性排序产生的影响，而部分网络中，节点邻居的重要性同样也会对节点产生较大影响。比如在网页搜索中，一个网页在大型网络中有超链接时，其被浏览的机会大大增加，基于此思路，国内外学者研究提出了HITs、LeaderRank以及PageRank等算法并逐步加以改进，成功应用到了许多现实网络当中。

HITs算法会同时考虑到链入节点和链出节点的影响，以网页为例，该算法认为，一个重要的网页，一方面是该网页中的内容质量高，许多网页中会存在该网页的超链接；另一方面是网页中许多优质网页的超链接，能够作为一个中转站帮助用户跳转到各种网站。

PageRank算法是于1999年由斯坦福大学博士生Larry Page和Sergey Brin（L. Page, S. Brin，1999）开发并在斯坦福大学内部使用，最早是用于搜索引擎的排序模块，主要原理是把通过评估网页超链接的质量和数量来为每个网页分配相对的重要性分数，以此作为网页排名的要素。该算法认为，当存在一个网页A中有一个网页B的链接时，B获得了A对其提供了重要性的“得分”，该得分会受到网页A的得分以及网页A中链接数量的影响。算法中得分的计算会逐次迭代直到分值趋于稳定，在迭代结束后各网页的得分会与其实际的影响力相匹配。该算法在当时能够有较为优异的排序效果，并在后来被Larry Page成功用应到Google的搜索引擎当中。

该算法在以下两种情况会出现异常。其一是当一个节点不存在任何出链（Dead Ends问题），其二是当一个节点的出链指向自己时（Spider Traps问题），PageRank计算结果会与事实不符，针对这两种问题，PageRank算法原式子基础上引入了一个阻尼因子，每一个节点都有（）的概率随机跳转到其他节点，其现实意义是用户浏览任意网页时都有可能跳出当前网页去随机浏览其他网页，从而确保了入度为0的网页也可以被访问到，并把网络构成的矩阵从邻接矩阵变为不可约矩阵，最终得到了一个为大众认可的PageRank算法如方程1-1所示：

其中指节点的PageRank值，指节点的出度。但该算法在后来也逐渐被发现了几个重要的缺点：

1）新网页的PageRank值一般较低（刘齐,黄树成，2022）。主要原因在于PageRank计算时网页的重要性会受到其入链网页数量和入链网页的PageRank值的影响，但旧网页一般能够有更多的网页链入进去，进而导致旧网页在计算时本身占据着一定的优势。

2）PageRank算法仅适用于有向无加权网络，且忽略了节点本身的影响。现实中很多网络都是有向带权网络，比如运输网络中，节点之间的边可以包含运输路程、时间等信息，且不同节点所对应城市的发达程度、交通便捷程度都不同。因此，传统的PageRank算法难以应用到现实生活中的其他网络来进行网络地位排序。

3）PageRank缺少对出链的考虑，在该算法中自身网页的出链数量并不会影响该网络的重要性，但现实中带有多个出链的网页同样也可能存在较高的重要性。

针对上述特点，不同学者在不同场景使用该算法时提出了不同的改进方式。在网页重要性计算中，Xing（Xing W, Ghorbani A A，2004）等人认为，一个网页的受欢迎程度会同时受到两个方面的体现，一是有许多网页存在该网页的超链接（即有较高的入度），二是该网页存在许多网页的链接（网页有较高的出度）。因此，他提出了用网页入度比例和出度比例来解决传统式子仅考虑节点入度的问题，使得源网页会分配更多的PageRank值给入度和出度更高的目标网页。而对于新网页不友好的问题，PS Yu（Yu P S， Li X，2004）等人提出了一个TimedPageRank算法，把时间也作为一个权重因子，让时间越久的网页得到的权重越低，从而削弱了旧网页因为存在时间长产生的影响，并让新出现的在网络中影响力较高的网页更容易得到较高的PageRank值。国内学者将上述两个改进融合，在Xing和Ghorbani提出的方程基础上加入了一个衰减参数（王崝,鞠时光，2008），让旧网页能够在计算中下沉而高质量网页上升。

除此之外，还有不少学者将PageRank应用到其他网络中。PanPan Zhang等人利用世界投入产出表的数据构建了世界投入产出网络（Zhang P, Wang T, 2022），并以此网络为基础对于传统PageRank提出的改进，主要是针对该算法无法考虑到边的权重，且原算法没有考虑节点附带的先验信息的缺点，提出了一个适用于有向加权网络的PageRank算法。在使用了改进的PageRank算法后，其计算结果更能反应世界投入产出网络中各地区部门的真实地位。此外，其使用各个地区部门（即网络中各个节点）的总附加值作为节点的先验信息，使得PageRank计算中能够把更多信息纳入考虑，帮助高质量的节点在迭代中更快速地上升。

除此之外，PageRank算法假设随机跳转过程中所有节点都有相同的跳转概率，而部分网络中不同节点本身就存在着不同的跳转概率，LeaderRank算法主要在这个思路基础上对网络中添加一个背景节点来与网络中所有节点都有一个链入和链出的边，从而确保网络是一个强连通网络，且访问背景节点的概率与自身节点的出度成反比。该算法相较于PageRank而言收敛速度更快，且对于随机跳转问题在一定程度上进行了优化。

在后面的章节中，本文将会基于节点重要性排序的研究现状，结合商业网络的特点以及实际情况，选择相对合适的方法来量化企业的网络地位。

1.2.2 全要素生产率相关文献综述

全要素生产率指的是排除了各种要素投入以外得到的经济增长的贡献率，其中可能包括企业技术水平的提升、企业内部管理甚至是企业所在的制度和环境等因素（郭庆旺,贾俊雪，2005）。

全要素生产率测算一般是基于生产函数来完成的。Cobb等人最早提出了道格拉斯生产函数模型（Cobb C W, Douglas P H，1928），Tinbergen将该模型纳入到经济问题，进一步研究了生产率与生产函数之间的关系。全要素生产率的测度方法大体可分为两种层面：宏观层面和微观层面。早期由于数据的缺少、大环境的生产能力相对底下，更多的学者集中关注在国家、行业这一类宏观层面，通过测度可以大致发现不同国家之间技术水平、生产管理能力之间的差异，其中代表性的计算方法为索罗增长理论（SOLOW R M，1957）和随机前沿法（SFA）。随机前沿模型由Meeusen（MEEUSEN W, VAN DEN BROECK J V，1977）等人提出，该模型主要思路是把全要素生产率的变动归纳为两大方面的变动：技术效率以及技术进步率，从而使得计算结果更贴近于经济增长以及实际生产。随后Kumbhakar（Kumbhakar S C, Lovell C，2000）等人结合超越对数生产函数，在随机前沿模型的技术上，把纯技术效率以及规模效率变化率这两大部分也纳入全要素生产率的计算范畴。在非参数方法上，数据包络分析法也得到了推广，其中包括Charnes（Charnes A, Cooper W W，1978）等人提出的DEA模型以及Caves提出的DEA-Malmquis指数模型（Caves R E，1982）。

而在企业层面上，很多在宏观层面应用较为广泛的估计方法可能会造成较大误差，国内对于企业的全要素生产率估计方法的文献较少。当前较为广泛使用的是OP法和LP法。鲁晓东和连玉君对多种半参数和参数估计方法进行了总结（鲁晓东,连玉君，2012），并结合我国上市企业数据，对OP法进行了改进建立了以下估计模型：

式中个各变量解释如下表1-1所示：

表1-1 OP法各变量解释

|  |  |  |
| --- | --- | --- |
| 变量符号 | 变量类型 | 变量含义 |
| i | 虚拟变量 | 企业代号 |
| t | 时间变量 | 不同的年份 |
|  | 连续变量 | i企业t年度的固定资产对数 |
|  | 连续变量 | i企业t年度的员工规模对数 |
|  | 连续变量 | i企业从注册到t年总存续年数 |
|  | 虚拟变量 | i企业t年是否为国有企业 |

表1-1（续表）

|  |  |  |
| --- | --- | --- |
| 变量符号 | 变量类型 | 变量含义 |
|  | 虚拟变量 | i企业t年度是否参加出口活动 |
| reg | 虚拟变量 | 地区虚拟变量 |
| ind | 虚拟变量 | 行业虚拟变量 |

钱雪松等（钱雪松,康瑾，2018）借鉴了黎文婧（黎文靖，胡玉明，2012）等人的研究，应用道格拉斯生产函数计算上市企业全要素生产率的指标，具体计算公式如下：

其中各变量解释如下表1-2所示：

表1-2 钱雪松计算方法中各变量解释

|  |  |  |
| --- | --- | --- |
| 变量符号 | 变量类型 | 变量含义 |
| i | 虚拟变量 | 企业代号 |
| t | 时间变量 | 不同的年份 |
| j | 虚拟变量 | 行业代号 |
|  | 连续变量 | i企业t年度的营业收入对数 |
|  | 连续变量 | i企业t年度的固定资产净值对数 |
|  | 连续变量 | i企业t年度员工规模对数 |
|  | 连续变量 | i企业t年度的中间投入对数 |

1.3 研究内容与研究方法

1.3.1 研究内容

本文研究企业在商业网络中的影响力大小是否会对制造业上市公司全要素生产率产生影响，若产生影响，该影响是正面还是负面，进而为企业提高全要素生产率提供新的可能性。每一章的具体内容如下：

第一章为绪论，对本文所研究内容的背景、意义进行简要阐述，并描述了当前国内外研究的现状以及本文使用的研究内容和方法。

第二章在理论层面描述了本文对PageRank算法改进的应用以及可行性，应用改进PageRank算法计算商业网络各个节点的PageRank值。

第三章详细介绍了现有的几种主流的全要素生产率度量方式，给出了本文实际采用LP法进行全要素生产率度量的过程以及数据描述性统计。

第四章为本文的实证部分，给出了实证研究模型、控制变量的选择，并分析了实证结果，最后对于结果进行了稳健性检验。

第五章对本文研究进行总结，给出结论和展望。

1.3.2 研究方法

1）文献研读法

根据研究的目的或者课题，通过认真搜集和学习相关文献、理论，从而全面、正确地了解所研究的问题。通过研究国内外关于PageRank算法的文献，加深对于该算法的认识和了解，从而帮助自己更好地把该算法应用到商业网络当中。

对于全要素生产率的测算，从最开始的道格拉斯生产函数相关文献进行研读学习，逐步了解各种测算方式的优缺点，找出最终适合本研究的全要素生产率测算方式。

2）实证分析法

本文借鉴Panpan Zhang等人的改进PageRank算法模型，选用CSMAR数据库中上市公司前五大供应商及客户数据来对企业网络地位与全要素生产率进行实证分析。

2 企业网络地位度量

本章主要是对商业网络的构建以及计算PageRank的过程和公式进行介绍，并给出企业PageRank值相关的描述性统计。

2.1 商业网络构建

本文使用CSMAR数据库中2009-2019年所有上市企业的前五大供应商数据以及前五大客户数据来构建有向加权网络。具体地，首先把前五大供应商以及前五大客户数据按照年份进行划分，在同一个年份的企业数据会在同一个网络中；网络中的节点为每个上市公司、其供应商以及客户。举一个简单的例子，假设某年度网络中的上市企业、其供应商及客户如表2-1所示。

表2-1 2009-2019年客户-供应商网络指标

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 上市公司名称 | 供应商名称 | | 供应商采购金额 | | 客户名称 | 客户销售金额 |
| A | B | 100 | | C | | 300 |
| D | A | 150 | | C | | 400 |

此时网络中共计有4条边，分别是B→A，A→C，A→D，D→C，其构成的商业网络如图2-1所示。以此类推可以初步得到2009年-2019年共计11个商业网络图。

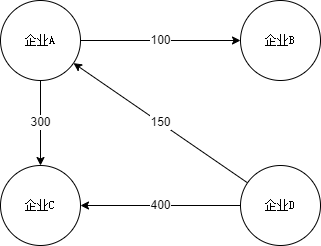


图2-1 示例数据构成的商业网络

在实际构建各年度商业网络前，需要对数据审阅并做预处理工作。首先把报表类型为2的数据删除（仅保留合并报表的数据）；部分企业披露的前五大供应商名称为“合计”、“前五大供应商”、“前五大客户合计”等，对于这一类合计的数据直接剔除；剔除供应商采购额以及客户销售额为空值的数据；其次，在CSMAR数据库中下载上市企业基本信息表，并与前五大供应商数据以及前五大客户数据进行匹配，从而让上市企业代码变为上市企业名称，此工作主要目的是使用企业名称来构建商业网络；第三步是处理部分模糊的数据。部分企业为了避免披露主要供应商给竞争对手，以及部分保密单位不适合披露的原因，样本数据中许多的供应商和客户名称为“第一名”、“单位1”等模糊数据，不同企业的“第一名”供应商并不一定是同一个企业，但在构建商业网络时，由于节点是用企业名称确定的，同样的名称会被认为是同一个企业，这样就导致了两个没有交集的企业之间可能被网络中认为有同样的一个“第一名”的供应商。为避免这一问题，同时最大程度地保留节点和边的数据，本文对于所有模糊的供应商和客户名称都给了唯一标识，具体做法：首先人工查看供应商和客户的数据，筛选出模糊的数据样本对应的企业名称，对于供应商模糊的数据，在相应企业名称前加上上市股票代码以及“供应商”字样，比如企业代码为600309的第一名供应商名称为“单位一”，本文将其更改为“600309供应商单位一”，同理，对于模糊的前五大客户数据，在其名称前面添加股票代码+“客户”的前缀。处理后最终剩下79288条上市公司前五大供应商数据以及106244条上市公司前五大客户数据。

2.2 节点重要性排序方法选择

在前面国内外研究现状中已经对知名度较高的几种重要性排序算法进行了阐述，本节主要针对在商业网络这一应用场景讨论更为合适的排序方法。

在选择排序方法中，首先我们需要明确本文的网络的类型。本文使用的是所有上市企业前五大供应商和前五大客户数据构建的有向加权网络，该网络在一定程度上反映了企业间基于原材料、产成品以及商业服务所形成的复杂的贸易行为。在现实生活中，一个企业尤其是制造业企业，由于生产产品的复杂和多样性使得企业需要和多个供应商合作，但并非所有供应商都占据着同等地位，核心零部件的供应商必然对企业能够产生更大的影响，且当企业的供应商和客户是较为优质，影响力较高的公司时，企业自身也能够有一定的收益，比如更低的供应链风险、依托供应链得到更好的信用凭证，进而得到更低的融资成本等等。因此，度中心性、KatZ中心性等指标并不能很好地反映商业网络的实际情况。

而PageRank和LeaderRank算法都有综合考虑邻居的数量以及对应节点的影响力。以PageRank为例，该算法认为，一个网页的重要性会受到两方面的影响，即其他网页指向该网页的超链接数量以及对应网页的重要性。当应用场景更改到本文的商业网络而言，一个企业的在网络中的地位会受到企业的供应商数量以及供应商在网络中的地位两个方面的影响。而现实生活中，供应商的的确确对企业能对企业的生产水平、生产计划等方面产生重大影响。一个优质、稳定的供应商不仅能保证企业企业的准时生产和较低的库存水平，还能大幅度降低供应风险和现金流风险，帮助企业降低融资成本等等。可见，用供应商的数量和供应商影响力来度量企业的网络地位，事实上也是能够一定程度上反应企业的供应链稳定水平。

而LeaderRank事实上可以看做是对于PageRank算法的一个改进方向。该算法设置了背景节点去代替PageRank算法中随机跳转的情况，能够一定程度上避免随机跳转概率的取值问题。而在商业网络中，由于本文使用的是2009-2019年的数据，每一个年份数据单独形成一个网络。因此随机跳转概率可以通过不同年份之间企业供应商和客户更改情况来计算得出，随机跳转概率是存在一定的意义的，因此在商业网络中PageRank或许是更适合的排序方法。

此外，PageRank算法也已经在许多网络中得到了应用。比如PanPan Zhang等人把该算法进行改进，使得算法能应用到了世界投入产出网络中，并且证明了使用该改进方法更能准确反映出真实情况。而世界投入产出网络与商业网络事实上也有一定的相似之处，他们都是有向带权网络，且边的权重都是代表着贸易货物的价值。其次，包群等人也提出过用传统PageRank算法去度量商业网络中节点的网络地位。因此，本文选择参考PanPan Zhang等人对算法的改进思路，使用针对有向加权网络的改进PageRank算法，对于随机跳转概率根据历史数据进行估计，从而计算得到企业的网络地位水平。

2.3 PageRank算法应用到商业网络解析

2.3.1 算法基本原理

前面提到，传统的PageRank算法仅考虑了链入节点的数量以及该节点的PageRank值，且存在着几个较为突出的缺点。本文所构建的商业网络是一个有向加权网络，参考国内外学者对于PageRank缺点的改进，以及结合本文商业网络的特点，使用以下改进PageRank算法计算商业网络中各个节点的影响力：

式（2-1）中代表t年企业i的PageRank值，是网络中所有节点的集合，指边的权重，即供应商销售给企业货物的价值。是一个0-1变量，当节点与节点之间存在边时，，否则为0。为节点指向所有边权重之和，为节点的出度。是一个阻尼因子，每个节点都有的概率随机跳转到其他节点，在PageRank计算中能够确保网络中入度为0的企业能够被访问到，同时避免没有出度的节点出现PageRank沉积而结果异常问题。是一个调整因子，能够调整式子中权重的相对重要性。为网络中节点的总数。PageRank计算是一个迭代算法，对每个节点赋予初始PR值后，程序会进行迭代运算，只到两次迭代之间PR变化量小于阈值，或者迭代次数到达上限时，才会给出各节点最终的PR值。

而作为一个阻尼因子，其显示含义是任何一个节点都有一定概率跳出当前已有的边，将其PageRank值分给网络中其他节点。而应用到商业网络中，可以理解为企业有的概率放弃与当前已有的客户进行交易，转而寻求网络中新的合作伙伴。在本文中可以利用2009年-2019年的数据分别求出2010-2019年每一年企业放弃继续与现有客户合作的概率。具体而言，以计算2010年企业放弃与现有客户合作的概率为例，首先对2010年数据进行筛选，仅保留在2009年和2010年均披露了数据的样本，以此作为2010年总样本数；其次对总样本进行遍历，分别与2009年数据进行比较，查看上市企业2010年的供应商或者客户是否保持一致，并计算保持一致的数量。将每个年度的总样本数和不变供应商、客户的样本数分别加总，阻尼因子=（不变供应商样本数+不变客户样本数）/总样本数。每个年度的数据情况如表2-2所示。

表2-2 各年度企业更换主要供应商和客户情况

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 年份 | 总样本数 | 未更换供应商/客户样本数 | 更换了供应商/客户的样本数 | 当年度企业保持合作关系的比例 |
| 2010 | 13496 | 7553 | 5943 | 0.559647303 |
| 2011 | 18163 | 10458 | 7705 | 0.575785938 |
| 2012 | 24660 | 13475 | 11185 | 0.546431468 |
| 2013 | 31170 | 19956 | 11214 | 0.640230991 |
| 2014 | 17388 | 11986 | 5402 | 0.689325972 |
| 2015 | 21375 | 12041 | 9334 | 0.563321637 |
| 2016 | 22908 | 17145 | 5763 | 0.748428497 |
| 2017 | 26936 | 21243 | 5693 | 0.788647164 |
| 2018 | 30562 | 23862 | 6700 | 0.78077351 |
| 2019 | 31429 | 24541 | 6888 | 0.780839352 |
| 合计 | 238087 | 162260 | 75827 | 0.68151558 |

从表中可以看到，当企业与供应商和客户的交易增加的大环境下，企业更换供应商和客户的概率在逐渐降低，越来越多的企业倾向于与已有的供应商和客户保持一个相对稳定的合作关系。本文由于在后面实证中使用的是2009年-2019年的面板数据，为了确保年份之间的计算方式一致性，本文对于的取值上使用合计值计算，即在所有计算中均取值为0.68，代表企业平均有68%的概率在与现有的合作企业保持继续合作的关系。在的取值上，认为商业网络中供应商客户的数量以及与客户交易的规模同样重要，因此取0.5。在迭代前给每个节点赋予PageRank初始值为1/，最大迭代次数为100，当两次迭代之间PageRank变化总值小于0.00001时，迭代结束。

2.3.2 商业网络地位计算及分析

在前一小节中已经对PageRank算法应用方式进行了描述，本节主要分析应用中存在的问题以及改进方式。在实际进行PageRank计算中，有以下几种主流算法：

第一种是用二维数组形式存储网络的节点和权重（谢月，2012），具体做法是首先构造一个n\*n的零矩阵NetworkMatrix，当存在一个边由a指向b，权重为5时，则令NetworkMatrix[index[a]][index[b]]=5，以此类推。这种存储方式适用于边数相对较多的网络中，可以直接使用矩阵运算得出PageRank结果，且算法自由性高，不局限于传统的PageRank算法。

第二种是使用python中的Networkx库（Breja Manvi,Bhatia Himanshi，2022），该库可自主生成图（graph），且内置了PageRank函数，可以直接输入网络图、阻尼系数、迭代次数以及迭代结束阈值来自动计算pagerank值，但缺点是改进PageRank算法难以应用在该库中，需要手动修改库中的函数。

第三种是使用pygraph库生成网络，并手动写PageRank计算代码。pygraph同样可以生成有向加权网络，而手动完成PageRank代码也可以按照实际需求对PageRank算法进行改进。

本文选择采用pandas库的DataFrame构建商业网络，并手动设置PageRank函数，使用Networkx库的PageRank算法作为本算法的一种特例进行算法准确度验证。具体做法是首先对每个公司（即网络中的节点）进行逐一编号，然后在DataFrame中构建三个列：FromNodeSerial、ToNodeSerial以及Weight，分别存储网络中的源节点企业序号、目标节点企业序号以及对应边的权重。这样做的主要原因是本文构建的商业网络的密度相对较小，使用矩阵存储会有过多的0值，而本文最大的网络有25000以上个节点，矩阵的空间占用较高，且进行矩阵运算时时间复杂度也较高，整体计算效率很低。本文结合商业网络的特点，对部分多次重复计算的数据直接存储，且使用dataframe存储仅存储两个节点之间的边，对于权重为0的边不会占用内存，在空间和时间上都达到较优的结果。关键代码在附录一展示。表2-3罗列了各年度PageRank值的描述性统计数据。

表2-3 2009-2019年PageRank值描述性统计

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 年份 | 节点数 | 均值 | 标准差 | 最小值 | 最大值 | 上市企业均值 | 非上市企业均值 |
| 2009 | 6096 | 1 | 0.471 | 0.786 | 7.138 | 1.135 | 0.969 |
| 2010 | 8427 | 1 | 0.432 | 0.804 | 4.722 | 1.223 | 0.955 |
| 2011 | 11438 | 1 | 0.511 | 0.756 | 5.587 | 1.459 | 0.918 |
| 2012 | 17786 | 1 | 0.588 | 0.683 | 6.431 | 1.948 | 0.862 |
| 2013 | 18712 | 1 | 0.590 | 0.680 | 5.530 | 1.993 | 0.859 |
| 2014 | 15496 | 1 | 0.607 | 0.638 | 5.861 | 2.419 | 0.826 |
| 2015 | 20359 | 1 | 0.606 | 0.632 | 5.880 | 2.580 | 0.822 |

表2-3（续表）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 年份 | 节点数 | 均值 | 标准差 | 最小值 | 最大值 | 上市企业均值 | 非上市企业均值 |
| 2016 | 22328 | 1 | 0.606 | 0.631 | 5.012 | 2.631 | 0.821 |
| 2017 | 25833 | 1 | 0.604 | 0.630 | 4.615 | 2.631 | 0.821 |
| 2018 | 25781 | 1 | 0.605 | 0.629 | 5.122 | 2.673 | 0.820 |
| 2019 | 27795 | 1 | 0.608 | 0.630 | 6.742 | 2.661 | 0.820 |

研究发现，在企业数量随时间逐渐变多的同时，上市企业的PageRank均值也在逐年增大，且每年都高于非上市企业，这说明着我国上市企业在供应链网络中长期占据主导地位，且其影响力逐年增强，这也解释了为什么许多上市企业作为供应链中的龙头企业能够帮助中小企业能够有效缓解中小企业融资约束问题，原因或许也在于供应链中核心的企业所带来的影响力以及其商誉能够成为中小企业融资时一大信用担保。此外，对于供应链金融的研究，或许也可以从网络构建以及网络地位这一角度去评估中小企业的供应链所能提供的担保能力。

3 企业全要素生产率度量

3.1 全要素生产率度量方法

测度企业全要素生产率（TFP）常用的方法有最小二乘法估计（OLS）、固定效应方法（FE）、OP法和LP法等等，这些方法事实上都是基于C-D函数，即道格拉斯生产函数。该函数的方程如下：

在公式（3-1）中，代表企业在年的总产出，式子中用A来指代全要素生产率，L是劳动投入，K代表资本投入。当等式两边分别取对数时，能够得到线性形式的公式如下：

其中的y、l、k为公式（3-2）中Y、L、K的对数值，此时的残差即为全要素生产率的对数值。后续介绍的全要素生产率估计方法均基于公式（3-2）展开。

1）OLS估计法

尽管直接对公式（3-2）回归并预测残差即可得到估计的全要素生产率的值，但许多学者对该估计值的准确性存在着质疑，主要问题在于两个方面，其一是同时性偏差，主要体现在企业也具有主观能动性，决策者能够观测到其自身的投入产出效率，并针对存在的问题调整要素投入，从而提高自身的生产效率，最终导致残差项事实上与回归项之间存在一定的关联。针对这一问题，可以把公式（3-2）中的残差项拆分为两个部分,分别为和，具体公式如下所示：

是企业可以观测到的自身决策并影响企业对要素投入的决策部分，而才是真正的残差项。学者们对拆分后得到的残差项使用了不同的处理方法，进而有了不同的全要素生产率估计方式。

除了同时性偏差外，第二个大问题是样本的选择性偏差。样本的选择性偏差产生的主要原因是企不同的企业在受到了生产率冲击后，会有不同的退出市场可能性。总体而言，对于有较为丰厚的资本的企业来说，即便面对较为严重的冲击，其并不会因此而导致无法继续生产进而退出市场，能够依靠存量资金让企业能够存货下来，即企业退出市场的概率与企业自身的资本存量有着反比的关系。当企业对于未来市场环境保持乐观态度时，留在市场的期望收益会高于退出市场，而这一现象的存在会让资本项的系数估计值偏低。因为此问题的存在，OLS估计值说服力相对不足。

2）固定效应方法（FE）

固定效应方法假设是公式（3-3）中企业可观测到的在同一企业中不会随着时间而发生改变，而企业之间是会存在差异的。此时可以直接在面板数据中把企业设为虚拟变量并进行固定效应回归，从而解决其中的内生性问题。但该方法仅适用于面板数据，且有较多的信息都无法在该方法中充分利用，所估计出来的全要素生产率与真实世界可能存在一定的偏差

3）OP法

OP法是一种半参数估计的方法，由Olley等人（Olley G S, Pakes A,1996）提出的。该方法的主要特点是考虑了生产率冲击的情况，并使用企业的投资额作为其代理变量。具体而言，首先将资本与投资额的关系表示如下：

其中K是企业资本存量，代表t期的投资额。当企业预测未来产出效率较高时，会结合本期资本存量情况考虑提高当期的投资额，即存在一个最优的投资函数：

对应的反函数为：

把反函数带入到公式（3-3），得

其中即为资本投入对生产率的贡献。定义

把对即可得出的估计值，代入到公式（3-7），得

对公式（3-9）的估计可得出劳动投入的一致无偏系数α，定义，则有

至此可得到生产函数中所包含的系数的估计值，全要素生产率的对数值即为公式（3-2）的残差估计量。

总的来说，虽然OP法对于OLS估计中存在的缺点有一定的优化作用，但是其使用的代理变量为企业投资，当企业投资额为0时，OP法会舍弃该类样本，最终可能会导致样本数量太少的问题。

4）LP法

针对上述的投资额为0导致的样本缺失问题，Levinsohn（James Levinsohn,Amil Petrin，2003）等人用中间投入作为代理变量的方法，而中间投入除了从企业财报直接获取外，还有较多学者通过不同的方法对此进行估算，能够有效避免OP法中存在的样本缺失问题。

基于此，本文在全要素生产率的估计上主要使用LP法和OP法，LP法测算得到的数据作为回归模型的主要数据，使用OP法得到的数据进行稳健性检验。

3.2 数据来源与TFP估计

本文选择2009-2019年的制造业上市公司作为研究样本，公司财务数据来自CSMAR数据库，并从中经网统计数据库获取分地区价格指数。价格指数主要用于平减企业的工业增加值、资本存量、总产出等。对于北交所数据以及上市不满一年的数据均做剔除处理。参考鲁晓东的方法，本文使用的模型如下：

对于各变量的解释已经在前文给出，其中工业增加值、固定资产以及中间投入等数据并无法直接从企业财务报表和年报中中直接获取，本文参考余淼杰(余淼杰,金洋,张睿,2018）、朱沛华（朱沛华,陈林,2020）的计算方法，对固定资产、中间投入等指标的计算公式定义如下：

1）固定资本存量 = 固定资产原值 + 在建工程 + 工程物资 + 固定资产清理

2）工业总产值 = 营业收入 + 期末存货净额 - 期初存货净额

3）中间投入 = 工业总产值 \* 营业成本 / 营业收入 - 应付职工薪酬 - 当期固定资产折旧

4）工业增加值 = 工业总产值 - 中间投入

在代理变量方面，选择用中间投入来充当。企业的地区数据按省进行划分；退出行业的虚拟变量取值规则如下：在统计年间，若企业从制造业退出到其他行业时，则取值为1，企业一直在制造业以及企业中途加入制造业的情况下，取值为0。基于此条件下，最终的全要素生产率描述性统计如表3-1所示。

表3-1 2009-2019年全要素生产率描述性统计

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | 观测值 | 均值 | 标准差 | 最小值 | 最大值 |
| 2009 | 1124 | 14.384 | 1.054 | 8.499 | 19.927 |
| 2010 | 1396 | 14.6 | 1.038 | 6.505 | 18.576 |
| 2011 | 1581 | 14.681 | 1.005 | 9.318 | 18.822 |
| 2012 | 1674 | 14.573 | 1.004 | 8.823 | 18.948 |
| 2013 | 1706 | 14.595 | 1.029 | 10.252 | 18.676 |
| 2014 | 1794 | 14.616 | 1.039 | 8.998 | 18.736 |
| 2015 | 1932 | 14.579 | 1.001 | 9.044 | 18.685 |
| 2016 | 2131 | 14.645 | 0.991 | 8.746 | 18.748 |
| 2017 | 2413 | 14.804 | 0.941 | 9.031 | 18.846 |
| 2018 | 2481 | 14.841 | 0.978 | 9.772 | 18.801 |
| 2019 | 2635 | 14.824 | 0.992 | 9.123 | 19.013 |

从表中数据可观测得知，近年我国制造业上市公司数量逐年以较为平稳的速度增加，但整体而言，制造业的全要素生产率分布没有太大的变化，均值和极值都在一个较小的区间内浮动，说明估计的数据相对合理，整个制造业的全要素生产率也稳定在一个相对合理的区间。

4 企业网络地位与全要素生产率实证分析

本章是企业网络地位对全要素生产率的影响实证分析，首先会根据前文内容以及已有的理论基础提出假设，然后将介绍实证过程中使用到的变量的相关定义，进行实证分析并对结果进行稳健性检验。

4.1 研究假设

企业的在商业网络中的网络地位同时受到供应商数量、供应商销售金额、供应商网络地位的影响。当一个企业有较高的网络地位，即企业的供应商数量较多、供应商供货量足够高、供应商是网络中重要的供应商时，企业能够在供应链上拥有更强的保障，存在更小的发生缺货损失的可能性。可靠的供应链在降低企业供应链风险的同时，也能够促进企业之间深度合作，供应链双方可能会发生交叉持股、技术互补、资源共享等行为，这些都能够一定程度上降低企业的融资约束和融资成本，促进企业内部的资源配置的优化、技术进步以及技术效率的提高，进而提高全要素生产率。基于此猜想，本文作出以下假设：

H0：企业网络地位对于企业全要素生产率产生影响且影响作用为正。

H1：企业网络地位对于企业全要素生产率产生影响且影响作用为负。

4.2 数据来源及变量选取

企业网络地位数据、全要素生产率已在前面给出，控制变量相关的财务数据来源于CSMAR数据库，利息支出数据来源于Wind数据库。

被解释变量为全要素生产率，记为TFP\_LP，是由LP法进行测算得到的，具体测算方式已在第三章给出。

解释变量为企业网络地位，记为PR，通过改进PageRank算法中迭代计算得出。具体计算方式已在第二章给出。

控制变量：

1）企业总资产对数值Size，通常情况下，规模较大的企业能够更容易实现专业化分工，降低企业生产制造成本，从而提升企业生产率水平。

2）企业存续年龄对数值Age，企业存续时长越长，对相关领域了解程度会较高，且更有可能形成错综复杂的网络关系，从而对全要素生产率产生影响。

3）托宾Q值TobinQ，托宾Q值能够反馈企业的未来投资机会，而全要素生产率也被部分学者解读为投资效率，因此控制托宾Q值来限制企业投资的影响。

4）资产负债率Lev，已有学者研究发现企业的财务杠杆对全要素生产率存在倒U型关系（李冠吾，2019），适当的贷款能够较好地提高企业全要素生产率。

5）固定资产占比Fixs，企业固定资产占比较高时，不同的固定资产折旧比率会对企业全要素生产率产生不同影响；且固定资产可能会对技术水平的进步产生限制。

6）资产收益率ROE，财务状况能够反映企业资金方面的经营成果，良好的收益率是企业健康发展并长期经营的基础，且当企业拥有较好的财务状况时，能够帮助吸纳更多的资金以及资源，促进员工积极性，进而提高进步效率。

具体各变量类型、变量符号以及变量含义如下表4-1所示。

表4-1 变量定义

|  |  |  |
| --- | --- | --- |
| 变量类型 | 变量符号 | 变量含义 |
| 被解释变量 | TFP\_LP | 企业全要素生产率 |
| 解释变量 | PR | 企业PageRank值 |
| 控制变量 | Size | 企业总资产对数值 |
| Age | 企业年龄对数值 |
| TobinQ | 托宾Q |
| Lev | 总负债/总资产 |
| Fixs | 固定资产/总资产 |
| Roa | 资产收益率，即净利润/平均资产总额 |

4.3 变量描述性统计

本文选取了在2009-2019年度内行业分类为制造业的上市企业样本，使用2012年证监会行业分类代码对行业进行分类。在筛选样本过程中，本文剔除了以下样本：

1）上市不到一年或已经退市的企业；

2）在北交所发行股票的上市公司；

3）数据不全的上市公司。

最终本文得到了公司年度样本20867个。对于部分存在空值的数据，本文选择采用均值填充的方式进行处理。表4-2给出了主要变量的描述性统计。描述性统计显示，企业全要素生产率（TFP\_LP）均值为14.68，标准差为1.008，极小值和极大值相差超过10，说明企业全要素生产率在样本期内存在较大的差异。企业的网络地位（PR）的标准差为0.821，与全要素生产率的标准差较为接近；均值仅有2.305而最大值为6.742，最小值为为0.629，说明样本当中不同企业的网络地位差异巨大，这或许与企业规模、企业所属细分行业等原因相关。样本的主要变量均呈现较大幅度波动的特点。

表4-2 变量描述性统计

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) |
| 变量名称 | 观测值 | 均值 | 标准差 | 最小值 | 最大值 |
| TFP\_LP | 20,867 | 14.68 | 1.008 | 6.505 | 19.01 |
| PR | 20,867 | 2.305 | 0.821 | 0.629 | 6.742 |
| Size | 20,867 | 21.85 | 1.252 | 13.08 | 27.56 |
| Age | 20,867 | 2.799 | 0.368 | 0.693 | 4.127 |
| TobinQ | 20,867 | 2.302 | 3.408 | 0.684 | 192.7 |
| Lev | 20,867 | 0.433 | 0.722 | -0.195 | 63.97 |
| Fixs | 20,867 | 0.224 | 0.148 | 0 | 0.902 |
| Roa | 20,867 | 0.0433 | 0.232 | -14.59 | 20.79 |

4.4 模型的参数估计以及结果分析

本文使用的数据为非平衡面板数据，即不同年份中上市企业的样本存在一定的区别。下面会对主流的三种面板数据回归模型进行介绍并给出本文所使用的回归模型。

混合回归模型是一种把面板数据当成截面数据的相对极端的一种情况，即认为所有个体在不同时期都会有完全相同的回归方程，其回归方程可写为

其中y为被解释变量，x为解释变量，ε为残差，i和t分别表示个体以及时间。这样可以把2009-2019年所有数据放在一起来直接做一个OLS回归。

而固定效应模型的基本假设是不同个体（或不同时期）间存在显著差别，此时可在前面混合回归模型中添加时间或个体的虚拟变量，从而获得考虑了时间或个体效应的回归结果。当然，可以同时加入时间、个体、行业等多个虚拟变量来控制多个层面的效应。

随机效应模型主要应用在对于每个个体的效应都与其解释变量不相关，因此不使用个体的虚拟变量来解释，而是选择把误差项拆分为，来进行解释。

在本文中，不同时间、不同上市企业之间必然存在差别，因此混合回归模型不在本文的考虑范围之内，而在固定效应模型和随机效应模型的选择上，通常使用豪斯曼检验来确定。豪斯曼检验的原假设H0:与解释变量均不相关（即认为随机效应模型更适用）。本研究中豪斯曼检验的统计值为486.44，p值为0.000，强烈拒绝原假设，因此本文将使用固定效应模型，对于所有连续变量在1%和99%分位进行缩尾处理。论文的基准模型如下：

模型估计结果如下表4-3所示：

表4-3 固定效应模型回归结果

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
| VARIABLES | TFP\_LP | TFP\_LP |
|  |  |  |
| PR | 0.0171\*\* | 0.0172\*\*\* |
|  | (2.23) | (2.93) |
| Size |  | 0.3951\*\*\* |
|  |  | (19.65) |
| Age |  | -0.0016 |
|  |  | (-0.02) |
| TobinQ |  | -0.0192\*\*\* |
|  |  | (-3.18) |
| Lev |  | 0.4187\*\*\* |
|  |  | (6.28) |
| Fixs |  | -1.1380\*\*\* |
|  |  | (-13.25) |
| Roa |  | 2.1816\*\*\* |
|  |  | (18.62) |
| Constant | 14.1785\*\*\* | 5.9203\*\*\* |
|  | (627.84) | (13.68) |
|  |  |  |
| Observations | 20,867 | 20,867 |

表4-3（续表）

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
| VARIABLES | TFP\_LP | TFP\_LP |
| Number of cd | 2,636 | 2,636 |
| Adjusted R-squared | 0.139 | 0.406 |
| Prob > F | 0.000 | 0.000 |

Robust t-statistics in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

表4-3中第（1）列为直接把网络地位（PR）与全要素生产率（TFP\_LP）进行回归的结果，第（2）列为添加了控制变量以后的固定效应回归结果。可以看到回归方程的F统计量对应的P值均为0.000，在1%的水平下拒绝回归系数全为0的假设，说明固定效应模型整体上非常显著，即全要素生产率与企业的网络地位之间存在显著的相关关系。对比第一列和第二列，在添加了控制变量后模型的调整R方从0.139提高到0.406，说明加入了控制变量以后模型整体的解释能力有了明显的提升。在模型的解释变量中，第（2）列PR的系数为0.0172，t值为2.93，在1%的显著性水平下拒绝回归系数为0的原假设，表明上市企业在网络中的影响力对全要素生产率（TFP）的影响显著，且该影响为正面影响，在其它条件都不变的情况下，上市公司的网络影响力每提升一点，全要素生产率平均会上升约1.72%。

4.5 稳健性检验

为了使本文的研究结论更加可靠，需要进行稳健性检验。本文稳健性检验主要使用替换变量法。

1）替换自变量

对于商业网络的构建，前文主要使用供应商与客户的交易金额作为边的权重，这样做的主要原因在于能够让网络中不同的边之间可以直观对比看出边的权重大小，但这样不可避免的忽略了企业之间交易对企业自身的重要程度。因此，在稳健性检验中，本文把边的权重更改为上市企业与供应商/客户交易额占全年采购/销售的比重，如此可以避免不同产业之间的差异，能够直观看出供应商对企业的影响力大小。通过更改网络构建规则后，用新的网络地位指标与全要素生产率回归的结果如表4-4的第（1）列所示。

回归结果显示，当使用供应商采购额占比、客户销售额占比作为边的权重时，其与全要素生产率的回归仍旧显著，与前文结论一致。

2）更改网络构建规则

在构建商业网络时提到，对于“第一名”、“单位一”等企业名称较为模糊的数据，本文是选择将其保留并更改为唯一标识，从而让网络能够保留更多的信息，但这些信息对于网络是否存在较大影响，甚至导致网络中各企业地位的度量变得不真实的问题需要进一步探讨。因此，在这一节将商业网络中所有的模糊企业名称对应的样本删除，仅保留准确的企业名称和企业信息样本，以此获得一个较小的样本来检验网络地位是否仍会对全要素生产率产生显著影响。回归结果在表4-4的第（2）列展示。此时样本数量从原来的20867个变为仅有6953个，且样本中上市企业数量从原来的2636变为1649个，但回归结果依旧显著，说明尽管剔除了许多模糊数据时，剩余的数据仍然能够证明企业的网络地位会对全要素生产率产生正向影响。

3）更换因变量

前面提到，全要生产率有多种测度方法，各种测度方法均有一定的优缺点。其中主流的方法有两种：OP法和LP法，这里使用OP法计算得到的全要素生产率代替前面使用LP方法得到的数据，与之形成对照。参考Olley and Pakes的做法，选择用投资额作为代理变量进行估计，其中投资额估计值=期末固定资本存量 - 期初固定资本存量 + 当期固定资产折旧。表4-4的第（3）列展示了改变使用OP法估计全要素生产率与网络影响力回归得到的结果。当使用OP法测度企业的全要素生产率时，仍能获得与前文相一致的结论。

表4-4 稳健性检验回归结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| VARIABLES | TFP\_LP | TFP\_LP | TFP\_OP |
|  |  |  |  |
| PR | 0.0151\*\* | 0.0342\*\*\* | 0.0166\*\*\* |
|  | (2.56) | (3.69) | (2.83) |
| Size | 0.3952\*\*\* | 0.3912\*\*\* | 0.1843\*\*\* |
|  | (19.65) | (11.08) | (10.16) |
| Age | -0.0016 | -0.0705 | 0.0515 |
|  | (-0.02) | (-0.46) | (0.64) |
| TobinQ | -0.0192\*\*\* | -0.0500\*\*\* | -0.0053 |
|  | (-3.18) | (-6.13) | (-0.91) |

表4-4（续表）

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| VARIABLES | TFP\_LP | TFP\_LP | TFP\_OP |
| Lev | 0.4186\*\*\* | 0.2935\*\*\* | 0.3155\*\*\* |
|  | (6.28) | (2.72) | (4.78) |
| Fixs | -1.1380\*\*\* | -0.7474\*\*\* | -1.1993\*\*\* |
|  | (-13.25) | (-5.64) | (-15.07) |
| Roa | 2.1811\*\*\* | 2.4996\*\*\* | 2.0121\*\*\* |
|  | (18.62) | (12.64) | (18.99) |
| Constant | 5.9231\*\*\* | 6.2233\*\*\* | 7.9999\*\*\* |
|  | (13.68) | (7.76) | (20.40) |
|  |  |  |  |
| Observations | 20,867 | 6,953 | 20,867 |
| Number of cd | 2,636 | 1,649 | 2,636 |
| Adjusted R-squared | 0.406 | 0.358 | 0.297 |
| Prob > F | 0.000 | 0.000 | 0.000 |

Robust t-statistics in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

可以看到，上述三个稳健性检验通过不同方式替换了自变量和因变量以后，均能够保持与第四节的实证检验相同的结论，说明企业网络地位对全要素生产率的影响的实证分析是比较稳健和可靠的。

5 结论与展望

本文从网络的角度出发，利用2009年-2019年所有上市企业前五大供应商和前五大客户数据构建网络，并以此为基础，综合对比了各种节点重要性排序算法， 基于PageRank算法的研究进展与商业网络的特点，将改进的PageRank算法应用到商业网络当中，并根据数据实际情况，估算了企业在网络中随机跳转概率，最终度量了各企业在网络中的地位及影响力，并将此指标与制造业企业的全要素生产率数据进行了实证分析，通过对模型回归结果的解读，得到了如下结论：

1）对于制造业企业而言，其在供应链中的影响力在一定程度上会对企业全要素生产率产生影响。

2）当企业在商业网络中拥有较为稳定的供应商、较高的影响力时，企业会有更高的全要素生产率，即企业网络地位与全要素生产率存在正相关关系。

由于时间和精力有限，本文仍存在进一步研究和拓展的空间，主要包括以下几个方面：

在网络地位的测算层面，本文使用的PageRank算法在计算企业网络地位时实际上仅考虑了各个节点入度的影响，没有把企业的客户（即节点出度）纳入到网络地位的影响因子中，而在现实当中，部分重要的客户或许对企业而言也是存在着一定的影响力。尽管有学者已经提出了同时考虑入度和出度来计算企业影响力的方案，但该方案并不适用于有向加权网络，如果能够设计一个同时考虑出度、入度以及边的权重的PageRank算法，并且能够明确客户和供应商分别对企业网络地位影响力的大小，或许对于网络影响力的测度会更加精准、有说服力。另一方面，由于缺少非上市企业的财务数据，本文默认所有节点都是类似的，即当企业选择更换主要供应商时，本文默认所有节点都存在相同的概率成为该企业的主要供应商，但事实上更大的企业显然更容易受到青睐，因此理论上大企业有更高的概率成为主要供应商，但由于许多中小企业没有强制要求信息披露，所有并没有可靠的数据来给网络中每个节点提供一个先验信息。

除此之外，相较于当前主流的用供应链集中度（即直接把前五大供应商采购额占比相加）指标，从网络地位去考虑供应链上企业的交互明显能够得出更多更有用的信息。同理，从网络角度出发可以重新审视许多个体、企业甚至行业之间的关联，比如构建引文网络、专利引用网络、借贷网络等等，或许可以发现更多有研究意义和现实意义的内容。

致谢

昨日清秋，忽而今夏。转眼间大学四年的求学路已经走到了最后，或许当这一份毕业设计论文完成以后，便到达了我四年求学生涯的终点。细数着仅剩无多的本科时光，心中充满着不舍和感恩。

首先，诚挚地感谢李晓艳老师在本科四年中对我的教育和帮助，老师能够急学生所急，耐心地解答我在学业上的困惑，为人亲和，亦师亦友，在我迷茫的时候也非常有幸得到老师给出的许多宝贵的建议，在毕业论文编写的过程中也能及时指出我的错误和问题，她积极乐观的生活态度和严谨的治学之道是我很好的学习榜样。

其次，感谢左月华老师在选题和论文研究方向上提供的帮助。在毕业设计过程中难免碰到瓶颈和研究预期与实际不一致的情况，左老师能够给予我很多创新的研究思路，从而让我的毕业设计过程变得顺利。

此外，感恩大学四年来认识的同学和朋友，正是因为他们的出现，我的大学生活变得更多姿多彩；感谢和我一同参加企业运营仿真竞赛的同学，通过比赛让我学习到了很多实用的技能，也在比赛中收获到了令人欣喜的成绩；特别感谢住在我对面寝室的王盛琛、熊浩然，他们不仅在日常学习生活中教了我许多宝贵的经验和知识，在毕业设计的行文、构思等方面都愿意与我耐心地交流并解答我的问题。

最后，感谢我的爸爸妈妈，从小到大一直鼓励我独立自强，对于我做的很多决定都会予以理解和支持，让我即使独自一人来到武汉，也能快速适应生活，找到自己大学四年的方向；感谢我的姐姐，站在过来人的角度对我未来的选择上给了许多很好的意见和建议，让我少走弯路，始终向前。

如果说大学四年是一部很长的电影，那么我希望我能够从头再看一遍其中的故事情节，再次身临其境感受自己过去的一次次快乐和痛苦。我常常会去想，要是让我重新回到大一，我会怎么度过这四年呢？或许有很多很好、很完美的答案，可惜我再也回不去了。这部电影我注定只能谢幕离场，结束以后便是人生新的篇章，我没有办法改写过去的电影情节，但我希望，下一场电影，我能够演得更加出色，更加美满。

参考文献

1. L. Page, S. Brin, R. Motwani, T. Wdinograd, The PageRank Citation Ranking: Bringing Order to the Web, Technical Report 1999-66, Stanford
2. 刘齐,黄树成.基于主题相似度改进的PageRank算法研究[J].计算机与数字工程,2022,50(01):45-48+60.
3. 史铭茗.加权PageRank算法研究综述[J].软件导刊,2013,12(02):30-32.
4. Xing W, Ghorbani A A. Weighted PageRank algorithm[C]// Conference on Communication Networks & Services Research. IEEE, 2004.
5. 王崝,鞠时光.基于时间维加权TimedWPR算法[J].计算机工程与设计,2008(12):3001-3004.DOI:10.16208/j.issn1000-7024.2008.12.011.
6. Zhang P, Wang T, Yan J.PageRank centrality and algorithms for weighted, directed networks[J]. Physica A: Statistical Mechanics and its Applications, 2022, 586.
7. 任晓龙,吕琳媛.网络重要节点排序方法综述[J].科学通报,2014,59(13):1175-1197.
8. 吴淑燕,许涛.PageRank算法的原理简介[J].图书情报工作,2003(02):55-60+51.
9. 鲁晓东,连玉君.中国工业企业全要素生产率估计:1999—2007[J].经济学(季刊),2012,11(02):541-558.DOI:10.13821/j.cnki.ceq.2012.02.013.
10. Bernard A B, Moxnes A. Networks and Trade[J]. Annual Review of Economics, 2018, 10(1).
11. Cobb C W , Douglas P H . A Theory of Production[J]. American Economic Review, 1928, 18(Supplement):139-165.
12. SOLOW R M.Technical change and aggregate production function[J].Review of economics of statistics,1957,39:312-320.
13. MEEUSEN W, VAN DEN BROECK J V.Efficiency estimation from Cobb-Douglas production functions with composed error[J].International economic review,1977,18(2):435-444.
14. Kumbhakar S C, Lovell C. Stochastic Frontier Analysis: Incorporating Exogenous Influences on Efficiency[M].2000.
15. Charnes A, Cooper W W, Rhodes E. Measuring the efficiency of decision making units[J]. European Journal of Operational Research, 1978, 2(6):429-444.
16. Caves R E. Multinational Enterprise and Economic Analysis[M]. Cambridge University Press, 1982.
17. 张兆芹,张少华,饶灿.供应链金融对中小企业融资约束的影响研究[J].商业经济,2022(02):89-91+113.DOI:10.19905/j.cnki.syjj1982.2022.02.030.
18. 钱雪松,康瑾,唐英伦,曹夏平.产业政策、资本配置效率与企业全要素生产率——基于中国2009年十大产业振兴规划自然实验的经验研究[J].中国工业经济,2018(08):42-59.DOI:10.19581/j.cnki.ciejournal.2018.08.003.
19. 宋华,卢强.什么样的中小企业能够从供应链金融中获益?——基于网络和能力的视角[J].管理世界,2017,No.285(06):104-121.DOI:10.19744/j.cnki.11-1235/f.2017.06.009.
20. 包群,但佳丽.网络地位、共享商业关系与大客户占比[J].经济研究,2021,56(10):189-205.
21. Yu P S, Li X, Liu B. On the temporal dimension of search[C]// Alternate track papers & posters of the 13th international conference. 2004.
22. 郭庆旺,贾俊雪.中国全要素生产率的估算:1979—2004[J].经济研究,2005(06):51-60.
23. 黎文靖，胡玉明. 国企内部薪酬差距激励了谁[J]. 经济研究,2012,(12):125-136.
24. 谢月. 网页排序中PageRank算法和HITS算法的研究[D].电子科技大学,2012.
25. 李蓬实.供应链网络模型研究综述[J].物流科技,2014,37(04):46-49.DOI:10.13714/j.cnki.1002-3100.2014.04.013.
26. Breja Manvi,Bhatia Himanshi,Juneja Dollie. Social Network for Game of Thrones[J]. Journal of Cases on Information Technology (JCIT),2022,23(4).
27. Olley G S, Pakes A. The Dynamics of Productivity in the Telecommunications Equipment Industry[J]. Econometrica, 1996.
28. James Levinsohn,Amil Petrin. Estimating Production Functions Using Inputs to Control for Unobservables[J]. The Review of Economic Studies,2003,70(2).
29. 余淼杰,金洋,张睿.工业企业产能利用率衡量与生产率估算[J].经济研究,2018,53(05):56-71.
30. 朱沛华,陈林.工业增加值与全要素生产率估计——基于中国制造业的拟蒙特卡洛实验[J].中国工业经济,2020(07):24-42.DOI:10.19581/j.cnki.ciejournal.2020.07.013.
31. 杨彬,康慕宁.基于概念的权重PageRank改进算法[J].情报杂志,2006(11):70-72.
32. Yuan Y, Yan J, Zhang P . Assortativity measures for weighted and directed networks[J].2021.
33. 张琨,李配配,朱保平,胡满玉.基于PageRank的有向加权复杂网络节点重要性评估方法[J].南京航空航天大学学报,2013,45(03):429-434.DOI:10.16356/j.1005-2615.2013.03.001.
34. 李冠吾.企业杠杆对企业生产率的影响[D].山东科技大学,2019.DOI:10.27275/d.cnki.gsdku.2019.000714.

# 附录

附录一：PageRank计算关键代码

network2009 = pd.DataFrame({  
 **'Weight'**: pur2009[**'供应商采购额'**].tolist() + cus2009[**'客户销售额'**].tolist(),  
 **'FromNodeSerial'**: pur2009[**'供应商代号'**].tolist() + cus2009[**'上市公司代号'**].tolist(),  
 **'ToNodeSerial'**: pur2009[**'上市公司代号'**].tolist() + cus2009[**'客户公司代号'**].tolist()  
})  
*# 检查是否所有节点都有出度，对于没出度的节点将会把其pagerank平分给网络中所有节点*nooutnode = np.ones(len(nodes2009), dtype=int)  
**for** i **in** range(0, len(network2009)):  
 nooutnode[network2009[**'FromNodeSerial'**][i]] = 0  
FinalNode = []  
**for** i **in** range(0, len(nooutnode)):  
 **if** nooutnode[i] == 1:  
 FinalNode.append(i)  
 **else**:  
 **continue***# 计算每个节点的出度*outdegree = np.zeros(len(nodes2009), dtype=int)  
weightout = np.zeros(len(nodes2009), dtype=float)  
**for** i **in** range(0, len(network2009)):  
 outdegree[network2009[**'FromNodeSerial'**][i]] += 1  
 weightout[network2009[**'FromNodeSerial'**][i]] += network2009[**'Weight'**][i]  
迭代计算pagerank值代码：  
  
**class** PRIterator:  
 \_\_doc\_\_ = *'''计算一张图中的WPR值'''* **def** \_\_init\_\_(self, network):  
 self.damping\_factor = 0.68 *# 阻尼系数,即α* self.Weight\_factor = 0.5 *# 调整权重的相对重要性* self.max\_iterations = 100 *# 最大迭代次数* self.min\_delta = 0.00001 *# 确定迭代是否结束的参数,即ϵ* self.mydf = network  
  
 **def** page\_rank(self):  
 graph\_size = len(nodes2009)  
  
 **if** graph\_size == 0:  
 **return** {}  
 *# 给每个节点赋予初始的PR值* page\_rank = np.full(graph\_size, 1 / graph\_size, dtype=float)  
  
damping\_value = (1.0 - self.damping\_factor) / graph\_size *# 公式中的(1−α)/N部分*flag = **False  
for** i **in** range(0, self.max\_iterations):  
 change = 0  
 rank = np.zeros(graph\_size, dtype=float)  
 **for** j **in** range(0, len(self.mydf)):  
 sub1 = (1 - self.Weight\_factor) / outdegree[self.mydf[**'FromNodeSerial'**][j]]  
 sub2 = self.Weight\_factor \* self.mydf[**'Weight'**][j] /  
weightout[self.mydf[**'FromNodeSerial'**][j]]  
rank[self.mydf[**'ToNodeSerial'**][j]] += self.damping\_factor \* page\_rank[self.mydf[**'FromNodeSerial'**][j]] \* (sub1 + sub2)  
*# 无出度的节点把pagerank值平分给所有节点*add\_rank = 0  
**for** l **in** range(len(FinalNode)):  
 add\_rank += self.damping\_factor \* page\_rank[FinalNode[l]] / graph\_size  
**for** k **in** range(0, graph\_size):  
 rank[k] += add\_rank + damping\_value  
 change += abs(page\_rank[k] - rank[k])  
 page\_rank[k] = rank[k]  
print(change)  
**if** change < self.min\_delta:  
 flag = **True** print(**"finished in %d iterations!"** % (i + 1))  
 **break  
if** flag:  
 print()  
**else**:  
 print(**"finished out of 100 iterations!"**)  
**return** page\_rank



**本科生毕业设计（论文）任务书**

|  |  |
| --- | --- |
| 题 目 | **基于改进PageRank算法的企业商业网络** |
|  | **地位对全要素生产率影响的研究** |

（任务起止日期：2021年11月2日～2022年6月5日）

|  |  |
| --- | --- |
| 院 系 | **管理学院** |
| 专业班级 | **信息管理与信息系统201801班** |
| 姓 名 | **卢冠文** |
| 学 号 | **U201816009** |
| 指导教师 | **李晓艳** |

教研室（系、所）负责人 2021年10月28日审查

院（系）负责人 2021年11月2日批准

|  |
| --- |
| 课题内容：  通过企业前五大供应商和前五大客户数据构建出上市公司的商业网络，并使用LP法估计2009-2019年期间企业全要素生产率，研究企业在商业网络中的影响力大小是否会对制造业上市公司全要素生产率产生影响，若产生影响，该影响是正面还是负面，进而为企业提高全要素生产率提供新的可能性。 |
| 课题任务要求：  从网络角度研究企业供应商和客户的关联； 使用PageRank等指标量化企业在商业网络中的地位以及影响力； 研究企业影响力对全要素生产力可能产生的影响以及分析影响机制。 |
| 主要参考文献（由指导教师选定）：  [1]L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, Technical Report 1999-66, Stanford [2]W. Xing, A.A. Ghorbani, Weighted PageRank algorithm, in: A.A. Ghorbani (Ed.), Proceedings of the 2nd Annual Conference on Communication [3]Zhang P ,Wang T,Yan J.PageRank centrality and algorithms for weighted, directed networks[J]. Physica A: Statistical Mechanics and its Applications, 2022, 586.  [4]鲁晓东,连玉君.中国工业企业全要素生产率估计:1999—2007[J].经济学(季刊),2012,11(02):541-558.DOI:10.13821/j.cnki.ceq.2012.02.013. [5]Cobb C W , Douglas P H . A Theory of Production[J]. American Economic Review, 1928, 18(Supplement):139-165.  [6]钱雪松,康瑾,唐英伦,曹夏平.产业政策、资本配置效率与企业全要素生产率——基于中国2009年十大产业振兴规划自然实验的经验研究[J].中国工业经济,2018(08):42-59. |
| 同组设计者： |
| 指导教师签名：  年 月 日 |