Three Essays on Development Microeconomics

by

Reem El Cheikh Taha

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 10, 2025

Keywords: Fast Internet, Research Production, Nuclear Power Plants, Local Labor Markets, Automation, Social Security Disability Applications

Approved by

Duha T. Altindag, Chair, Associate Professor of Economics
R. Alan Seals, Associate Professor of Economics
Chris Vickers, Associate Professor of Economics
Michael Stern, Associate Professor of Economics

Abstract

In the first chapter, I investigate the impact of faster Internet access on research production in African universities, resulting from the staggered rollout of submarine cables to 12 coastal African countries in the late 2000s. When a university gains access to the fast Internet, the number of publications produced by the researchers affiliated with that university increases by about 65 percent. Web access to existing research and improved communication between researchers appear to be the mechanisms for increased production. In the second chapter, we estimate the local economic effects of U.S. commercial nuclear power plants (NPPs) using a differences-in-differences strategy. Our control group consists of locations where plant construction was planned but ultimately canceled. We find that NPP construction significantly increases local employment and wages, with effects concentrated in the construction and public utilities sectors. However, these gains dissipate once construction concludes, and the plant becomes operational, as operational employment requirements are minimal. We find no significant spillover effects on neighboring towns, and commercial operations do not meaningfully impact broader labor market outcomes. In the third chapter, we examine the relationship between SSDI applications and the spread of automation technologies. Using confidential data on SSDI applications at the commuting zone level, we estimate the effect of automation exposure on application rates across age and gender groups from 2005 to 2019. Our findings suggest that SSDI application rates for the 18–64 age group decline with greater automation exposure. This effect is more pronounced in the 35–54 and 55–64 age groups.

Acknowledgments

I would first like to express my deepest gratitude to Dr. Duha T. Altindag and Dr. R. Alan Seals for their support and guidance throughout my time at Auburn University. Your invaluable mentorship and the opportunities you provided have been instrumental in shaping my research. Without you, none of the papers in this dissertation would have been possible.

I am also thankful to Dr. Chris Vickers and Dr. Michael Stern for serving on my committee and for their insightful feedback and support. Additionally, I extend my appreciation to Dr. Samir Huseynov for his time and effort in serving as the University Reader.

This achievement would not have been possible without the unconditional love and sacrifices of my parents, who gave me everything without asking for anything in return. Thank you for your endless support and encouragement throughout this journey.

To my husband, I am grateful for your continuous support, patience, and encouragement. Your presence by my side has been a constant source of strength and motivation, pushing me forward even during the most challenging times.

Last but not least, to my daughter, Mila, I hope this accomplishment makes you proud. You are my greatest inspiration.

The research reported herein was performed pursuant to a grant from the US Social Security Administration (SSA) funded as part of the Retirement and Disability Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the Federal Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or

usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACS	American Community Survey
CZ	Commuting Zone
EGY	Egypt
FOA	Fiber Optic Association
ICT	Information and Communication Technology
IFR	International Federation of Robotics
ITU	International Telecommunication Union
IV	Instrumental Variables
NPP	Nuclear Power Plants
NRC	Nuclear Regulatory Commission
OLS	Ordinary Least Squares
O*NET	Occupational Information Network
SA	South Africa
SC	Synthetic Control
SOC	Standard Occupation Classification
SSDI	Social Security Disability Insurance
THE	Times Higher Education
TMI	Three Mile Incident

Chapter 1

The Impact of Fast Internet on Research Production: Evidence from Africa

1.1 Introduction

Africa is under-represented in international research and development scenes (World Economic Forum 2017, Okeke et al. 2017, North et al. 2020).[1] The dearth of scientific research in Africa could be contributing to the low quality of life for Africans and possibly affecting the rest of the world. Some suggest that the persistent lack of African research reduces the richness of intellectual contributions toward hard problems worldwide.[2] The paucity of research activity could also stand in the way of economic prosperity and labor market opportunities since universities, as producers of knowledge and innovation, are key to regional economic development and cluster formation of firms.[3] They attract important companies to the area creating more jobs, launching and incubating entrepreneurship, licensing, and selling their knowledge to other businesses (Wolfe 2005, Reveiu and Dardala 2013, Baily and Montalbano 2017, Rosowsky 2022). Therefore, it's important to understand the underlying reasons for low research in Africa (Sawyerr 2004, World Bank 2014, Kumwen et al. 2017, Gurib-Fakim 2022).

Beginning in the late 2000s, the rollout of submarine fiber-optic cables connected the African coast to European and Asian broadband networks (Mauldin 2017).[4] The submarine connections led to a remarkable increase in data transmission capacity and drastically reduced the transmission time and cost. The gap in Internet usage rates between African countries and more developed countries began to close, as Africa's international bandwidth increased by more than 100 times

[1] For example, although Africa comprises 17% of the world's population (Hwenda et al., 2022), research in Africa accounts for less than 1% of the global research output (Duermeijer et al., 2018).
[2] Africa accounts for 25% of the global disease burden, yet it contributes to less than 3% of clinical research, which disadvantages everyone since medical interventions should be tested in diverse populations (World Economic Forum 2017, Hwenda et al. 2022).
[3] Hausman (2022) shows that university knowledge spillovers derive from the agglomeration of related industries. Several of the foremost innovative and entrepreneurial areas in the U.S. cluster around local universities such as Silicon Valley, Route 128, and the Research Triangle of North Carolina (National Research Council 2013, Hausman 2022, Baily and Montalbano 2017).
[4] By the end of 2005, a new submarine cable arriving on the African shores of Egypt, Algeria, and Tunisia delivered a new terabit system which was more than 32 times the capacity of the previous system, creating a shock to the Internet traffic network in those countries. (https://www.lightreading.com/ethernet-ip/alcatel-fujitsu-build-sea-me we-4/d/d-id/601119).

from 2007 to 2017.[5] In this paper, I estimate the impact of access to high-speed broadband on research production at universities in the coastal countries of Africa. I exploit cross-country variation in arrival times of submarine fiber-optic cables and the capacity of the country's backbone network to distribute Internet data. The first source of identifying variation comes from the differential timing in the arrival times of the submarines to a country.[6] Figure (1) depicts the share of countries with at least one submarine cable over time. Before 2005, none of the countries were connected to modern submarine cables, but by the late 2000s, Egypt, Algeria, and Sudan had gained access to high-speed broadband. Although Sub-Saharan countries were not connected before 2009, there were twelve African countries with at least one submarine cable by 2012. The second source of identifying variation is a university's location within the country relative to the location of the local Internet backbone network. When a country is connected to the Internet by submarine cables, only those who are near the country's local backbone network benefit from higher-speed Internet.[7] Appendix Figure (1A) presents the locations of universities relative to the Internet backbone network in Egypt. In this context, the "lucky" universities that were constructed near a country's backbone network gain access to higher-speed Internet when the country is connected to a submarine cable than other universities that are farther away from the local backbone network. This feature of the identification strategy enables me to study the impact of access to fast Internet within a country.

The increase in Internet speed and capacity may support university researchers in two ways. First, it could lower the cost of communication among researchers. For example, with fast broadband, telepresence, and virtual environments such as audio/videoconferencing and email become more viable tools of communication (Onwubiko 2012, Kim et al. 2009, Butler et al. 2008, Butler 2007, Barjak 2006).[8] Increased access to broader information types provides the second

[5] According to the Africa Bandwidth Maps, International bandwidth is the capacity of international connection between countries to transmit Internet traffic.
[6] According to the 2006 African tertiary institutions connectivity survey, universities in countries connected with at least one submarine cable had higher Internet capacity than universities located in unconnected African countries.
[7] Fiber-optic cable networks provide high-speed data connectivity but have limited reach to populations outside urban and suburban areas (ITU 2020). Many people still do not have potential access to fiber networks because they are living beyond the reach of the Internet backbone network (Commonwealth Telecommunications Organization 2012). According to Hjort and Poulsen (2019), submarine cables increased Internet speed in Africa by around 35% in locations connected to the backbone network (within 0.5 km radius from the backbone network) relative to unconnected ones.
[8] According to Cisco's Global IP Traffic Forecast and Methodology reports, Internet voice application traffic, such as Skype, increased from 0.096 Petabyte (PB) in 2005 to 2 PB in 2012 in Africa and the Middle East. Similarly,

mechanism through which fast Internet may improve research production. For example, researchers could more easily access online databases and websites outside the network, and this enhances the capability of learning about other people's work (Onwubiko 2012, Imhonopi and Urim 2011, Kim et al. 2009).[9]

To perform my empirical analysis, I compile a data set that includes information about each country's arrival timing of submarine cables and the research activity in each university in that country. Once the first modern submarine cable arrived in Africa, Internet connectivity spread unevenly across the continent. Over my analysis period, which spans years from 1999 to 2019, I compare the research output of African universities located near the Internet backbone network to those farther away. I measure university research output with data from SciVal, an online bibliometric tool that uses the publications in the Scopus database. The database reports different bibliometric measures, such as the number of publications, awards, citation counts, publication counts in the top-cited journals, co-authorship types, and the number of publications in different fields.

Using a staggered Difference-in-Differences (DD) design, I show that access to fast Internet increases research production in a university. The number of publications by researchers at a university close to the Internet backbone network increases by 65 percent after a submarine cable is connected to that country relative to that in other universities in the same country but farther away from the backbone network. All major academic fields exhibit increases following high-speed Internet connectivity, with medicine experiencing the highest increase. The effect could be driven by the increase in both access to knowledge and communication between researchers. Although broadband rollout boosted the number of publications, the findings show that this increase does not have a clear impact on the publications' quality.

I implement several robustness checks. First, in an event study analysis, I show that the treated universities (those that obtained access to fast Internet) were no different from the untreated ones in the same country in terms of the number of their research outputs before they were treated. This rules out the possibility of pre-trends in terms of the outcome. The event study analysis also

Internet video calling and webcam traffic increased from 0.111 PB in 2005 to 94 PB in 2012 in Africa and the Middle East.

[9] Cisco's Global IP Traffic Forecast and Methodology reports show that consumer web, email, and data traffic increased from 4.18 PB in 2005 to 56 PB in 2012 for Africa and the Middle East.

indicates that the effect does not take place immediately but with a lag. This is probably because research production and the peer review process take a while.

I also explore some concerns related to the potentially endogenous government funding, researchers' movement between universities, and the construction or re-location of universities. Governments might prioritize providing extra funding to universities having access to fast Internet because the returns of additional funding would be much larger there. Researchers also might have the incentive to relocate to universities with faster Internet access creating a problem in the identification strategy. Additionally, universities could try to move their operations to places that are closer to the backbone network of their country to benefit from the faster Internet speed. In this case, my findings would likely be an overestimate of the true relationship. To address these concerns, first, I use a shorter post-treatment period since increasing government funding and reallocation of researchers take time. Second, I re-estimate my regression using only the universities that were present prior to the arrival of submarine cables or the construction of the backbone networks in the country. The results from these exercises suggest that these external shocks are not the primary drivers of the main results.

In the identification strategy, the treatment and comparison groups are determined by their proximity to the existing Internet backbone network. If a positive impact on the number of publications was found for universities located farther from the backbone network, the validity of the results would be called into question. I find that the effect of fast Internet on research production gets smaller with increasing the distance from the Internet backbone network.

Countries in the sample might have built their Internet backbone network selectively, such that it is close to population centers. As a result, the location of the Internet backbone network may have also coincided with other amenities. Under this scenario, the results could be driven by the proximity to infrastructure other than the Internet backbone network. To alleviate this concern, I include the universities' proximity to the electricity and road networks of the country in the main regression. The point estimates obtained from this exercise indicate that the number of publications does not differ significantly between universities near main roads or power stations versus farther away after versus before the post-treatment periods.

In the rest of the paper, Section 1.2 discusses the literature review. Section 1.3 details the data used in this paper. In Section 1.4, I discuss the empirical analysis, the main results, several robustness checks, and potential mechanisms. Section 1.5 concludes.

1.2 Literature Review

My paper contributes to two lines of research. The first includes studies on how the development of broadband infrastructure impacts research production and collaboration.[10] Some of the past papers that investigated this issue focused mainly on the early adoption of the Internet in developed countries. For example, Rosenblat and Mobius (2004) show that distant US collaboration increased by 30 percent during the rise of the Internet. Agrawal and Goldfarb (2005) find that earlier versions of Internet connectivity raised the total number of published papers by increasing research collaboration between connected US universities by 85 percent. Agrawal and Goldfarb (2008) show that the adoption of the Internet amplified the role of middle tier universities in the production of high-quality research. Exploiting the variation in Internet adoption time and publications in seven top electrical engineering journals from 1981 to 1991, Agrawal and Goldfarb (2008) find that Internet connection increased multi-institutional collaboration by 40 percent. Pachi et al. (2012) argue that free access to the Internet increased scientific production in Brazilian institutions.[11] Also, Reed and Xu (2020) use the population of Internet users in a panel of developed and developing countries to document that a 10 percent increase in broadband subscription in a country increases the number of citable documents by 1.2 percent. Other past output as a function of Internet penetration in developing countries. Sooryamoorthy (2016) shows that the Internet in South Africa contributes to knowledge production.[12] To encourage the growth of scientific research in South Africa, the author emphasizes the necessity for a faster communication infrastructure. Many previous papers showed that there is a positive relationship between ICT infrastructures such as email use and research productivity and communication between co-authors. Sooryamoorthy (2016), Ynalvez and Shrum (2011), and Duque et al. (2009) find that email use related to research has a positive significant impact on co-authorship and production of publications in national and/or foreign journals in South Africa, the Philippine, and Chile respectively. Sooryamoorthy et al. (2007), using a survey of researchers from academic and research institutions in India, show that access to ICTs increased email use from 6% to 86% and

[10] Previous papers used surveys, measurement indexes, models, and different measures of the Internet to study its impact on research production.
[11] Using Fisher's ideal price index, Pachi et al. (2012) find that Internet connections accounted for more than 90% of the variation in the number of Ph.D. theses in three Brazilian institutions.
[12] Sooryamoorthy (2016) uses different models to examine the relationship between the use of the Internet and the publication productivity of academics in universities and scientists in research institutes from a sample population of the South African scientific system. The models for the productivity of all articles and total research productivity explain a variance ranging from 38% and 54%.

that over 70% of the researchers used emails to maintain a professional relationship. According to these papers, the net effect of the Internet on researchers seems to be positive. Okafor et al. (2011) show that the utilization of Internet services improved research output by around 79% in two universities in southwestern Nigeria. This paper contributes to the previous literature by providing evidence on the impact of fast Internet on 12 developing countries using quasi-random access to high-speed broadband within a country.[13]

More in general, this paper connects to a large strand of literature investigating several consequences of geographical expansion of infrastructure, ranging from employment (Zuo 2021, Moretti 2021, Ford and Seals 2021, Hjort and Poulsen 2019, Akeman et al. 2015, Kline and Moretti 2014), firm performance (Chen et al. 2020), education, and student performance (Vigdor et al. 2014, Dettling et al. 2018), political violence, mobilization, confidence in government, and voting behavior (Manacorda and Tesei 2020, Guriev et al. 2021, Falck 2014, Pierskalia and Holienbach 2013), social capital (Geraci et al. 2022), and health (Amaral-Garcia et al. 2021, McDool et al. 2020).

1.3 Data

Data on university research output are obtained from SciVal, a research intelligence tool from Elsevier that builds on the abstracts and citations found in Scopus.[14] SciVal provides comprehensive access to the research performance of over 20,000 research institutions from 230 nations worldwide.[15]

The number of publications offers an objective and consistent measure of research production across universities.[16] However, the number of publications has some limitations for use. First, it

[13] This paper uses the same identification strategy used by Hjort and Poulsen (2019). They show that the gradual arrival of submarine cables to Africa significantly increased the employment rate in connected areas when the fast Internet becomes available.
[14] Scopus is the largest abstract and citation database of peer-reviewed publications. SciVal is a new application programming interface (API) based on the publication database Scopus. Therefore, it covers more outcomes than other bibliometric analysis tools and allows downloading a variety of metrics aggregated at the level of institutions and year. SciVal has been used in many previous papers (Guseva et al. 2022, Abusaada and Elshater 2021, Horenberg et al. 2020, Avanesova and Shamliyan 2019, Caglar and Gurel 2019, Patelli et al. 2017, Ekpo et al. 2016).
[15] According to the Scival metric guidebook, SciVal uses the Scopus database from the 1990s onwards so that the numbers displayed in SciVal are based on consecutive uninterrupted years of data reflecting the current trends. 16
[16] This outcome could be further refined to denote publication types such as articles and reviews, books or book chapters, and conference abstracts or papers.

excludes working papers, so it reflects the research activity only with a lag.[17] Second, the number of publications varies across fields due to the nature of the work being performed. For instance, the creative output in fields such as art, music, drama, or theater could not be measured by the number of published works. Finally, the number of publications does not necessarily reflect the quality of the research output, which might be indicated by the journal where the study is published.[18] Previous papers showed that there is an alarming growth of publishing in predatory journals, especially in developing countries (Machacek and Srholec 2021, Xia et al. 2015, Shen and Bjork 2015).[19]

I use several measures of quality for each publication. As common in previous literature, I consider citation count and the probability of receiving an award or grant as measures of quality (Berkes and Nencka 2021, Moretti 2021). Second, I identify publications in the top ten and between the top ten and top 25 percent of the most cited journals indexed by Scopus. I test whether fast Internet increases the number or share of these higher-quality publications. Lastly, I use predatory publication count by manually matching publishing journals' names from Scopus with Beall's list for each institution in the sample. Machacek and Srholec (2017) map the penetration of predatory journals into Scopus. They show that, by 2015, three percent of the total indexed documents in Scopus come from predatory journals listed on Beall's website.

To test for potential mechanisms behind the increase in publication production, I look at the type of co-authored publications of each research output. Publications are categorized as an outcome of one of the following four groups: single authorship, institutional, national, and international co-authored publications. I estimate whether submarine cables' arrival increases knowledge access through single authorship research outcomes or increases communication between authors through raising co-authored publications.

Second, using the Telegeography submarine cables map, I construct a dataset of all submarine cables connecting Africa. I link each cable to the landing stations where it was connected and use the date of service as the "treatment year". In countries that are connected to multiple cables at different points in time, I identify the country-level connection year as the year when the first

[17] The process of publishing a paper takes time, as the publishing cycle varies between different fields. For example, the refereeing process typically takes a long time (Berk et al. 2017, Card and DellaVigna 2013, Ellison 2002).
[18] To address some of these issues, I show the impact of the fast Internet on different scientific fields, publications' quality, and on the number of predatory research outputs. (Section 1.4)
[19] Predatory journals publish papers without performing promised quality checks for issues such as plagiarism or ethical approval in return for a fee paid by the author.

modern cable arrived at the country's shore. Since information on the last-mile infrastructure is unavailable, I follow Hjort and Poulsen (2019) by defining a location as treated if it is near the country's Internet backbone network. The map of Africa's Internet backbone network before submarine cables' arrival is available from the Infrapedia map.[20] To identify universities that are near the Internet backbone network, I use the university's coordinates (latitude and longitude) to calculate the distance between the university's location and the country's backbone network. The sample excluded landlocked countries (except Ethiopia) since their international connectivity via their neighbors is unclear. Thus, data includes information on 12 countries (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania) spanning the period from 1999 to 2019.

For university characteristics, data could be compiled for only two countries: Egypt and South Africa. I construct university-level covariates from the Ministry of Higher Education of each country, universities' websites, and other sources. Characteristics acquired include the number of enrolled students, student-to-faculty ratio, graduation rate, number of faculty members with and without research responsibilities, and the number of schools within each university. University rankings are obtained from the Times Higher Education World University Rankings' annual publication.

Table (1) presents summary statistics and basic descriptions of all the outcome variables for the full sample in Panel (A), as well as the outcome variables, university characteristics, and some area characteristics for the subsample (Egypt and South Africa) in Panel (B) for the overall period 1999-2019.[21] In total, I have data for 317 universities in 12 countries. This produces a sample of 5,519 observations.[22] A university is identified as treated if it is within five kilometers radius of the country's Internet backbone network.[23]

Panel (A) shows the average number of publications is 130 per university in a year, and most of these publications are articles and reviews. Treated universities have higher levels of research

---

[20] According to the AfTerFibre's map and Africa Telecoms Infrastructure 2014 report, there are few backbone segments announced or built during the period of submarine cables arrival, especially in the countries in the sample (The report accessed via https://manypossibilities.net/2015/01/africa-telecoms-infrastructure-in-2014/, and the AfTerFibre's map accessed via https://afterfibre.nsrc.org/). The map from Infrapedia shows the dates of service of each of the backbone cables. Africa's backbone networks map prior to the arrival of the submarine cables is used to measure connectivity.

[21] Area characteristics, which are population and average years of education in the city where the university is located, are collected from Global data lab for Egypt and South Africa via https://globaldatalab.org.

[22] Some universities in my sample were built during the sample period.

[23] See Section (1.4.1) for further details about the difference between treated and untreated universities.

production compared to untreated ones. Around 33 percent of the research production is published in the top 25 percent of the most-cited journals indexed by Scopus. Most of the publications are an outcome of some type of co-authorship between researchers. In Online Appendix Section (A2), I show that the number and types of publications, types of co authorships, citation counts, and the quality of the publications of universities near compared to farther away from the backbone network just before the first country get its first submarine cable are not statistically different.

The subsample includes 773 observations for two countries, Egypt and South Africa, for which I have detailed information on the universities. Panel (B) shows the average number of publications for the subsample is 474 per university in a year, which is higher than the average for the full sample.[24] Similar to Panel (A), most of the publications are articles and reviews, and an outcome of co-authorship. Around 38 percent of the research output is published in the top 25 percent of the most-cited journals indexed by Scopus. The average university has an enrollment greater than 55,000 students, and the majority of the faculty members in the sample are researchers. Relative to the comparison group, treated universities are more likely to attract students and hire new faculty members. In addition, treated universities have a higher probability of getting a world rank.[25] However, they have a similar number of schools. Cities, where untreated universities are located, seem to have higher population density than cities where treated universities are established. Yet, both have the same average of years of education, around 8 years for adults aged twenty and above. A difference-in-difference approach does not require balance in characteristics at baseline, yet, in Online Appendix Section (A2), I check whether the observable characteristics, such as the number of students, faculty members, and schools, are similar between treatment and comparison groups. In Online Appendix Figure (2A), standardized differences between universities located near and those farther away from the backbone network are depicted, conditional on university and country

[24] The subsample, Egypt and South Africa, is different from the rest of the African countries in different aspects. Several previous reports showed that over the years South Africa and Egypt have led the continent by producing about half of the total publications in Africa (Sooryamorthy 2021, Kigotho 2022). According to Sooryamorthy (2021), this unequal contribution to publications could be explained in terms of many factors such as GDP, percentage of GDP allocated to research and development, political stability, and others. On the other hand, Egypt and South Africa are considered hubs due to their good Internet backbone network (Bannerman 2021). Thus, according to the African e-Connectivity Index, South Africa and Egypt are the top-ranked African countries in terms of Internet quality (Mitchell 2021). In addition, Egypt has the highest number of submarine cables compared to the rest of the countries in Africa.

[25] A university getting a world rank in the paper means that the university appeared in the annual publication of university rankings by the Times Higher Education (THE) magazine as one of the top universities in the world.

x year fixed effects. For some characteristics, such as the number of students enrolled and student to faculty ratio, the point estimates seem to be large, but not statistically different.

Figure (2) depicts the evolution of the number of publications for the universities in the sample. Until the mid-2000s, the number of publications was stable in the African universities in the sample.[26] It started increasing greatly during the same period of the arrival of submarine cables to the African shores as shown in Figure (1). According to the International Telecommunication Union (ITU) data, in 2004, there were only 0.5 fixed broadband subscriptions per 100 inhabitants in the sample. By the end of 2018, this figure had jumped to 28. Appendix Figure (2A), which plots the number of fixed broadband subscriptions per 100 inhabitants over time, indicates that submarine cable coverage might have driven broadband penetration on the African continent enabling digital growth.[27]

1.4 Empirical Analysis

1.4.1 Empirical Specification

To obtain the impact of submarine cables connectivity on the research output of universities at a given year and location, I follow a similar specification as Hjort and Poulsen (2019) and estimate the following regression:

$$(1.1) \quad y_{ict} = \beta_0 + \beta_1 \left(Near\ Internet\ Network_i \times Post\ Submarine\ Cable_{ct}\right) + \beta_2 X_{it} + \mu_{ct} + \gamma_i + \epsilon_{ict}$$

where y_{ict} stands for one of the outcome variables for university i in country c at year t. The indicator $Near\ Internet\ Network_i$ equals one if the university i is located within five kilometers from the country c's Internet backbone network, and zero otherwise.[28]

[26] Number of scientific papers with at least one African author more than quadrupled from 12,500 to more than 52,000 papers from 1996 to 2012 (Schemm 2013). Also, publications in most African countries increased from 2 to 20 times between 2001 and 2018, whereas publications in the highest productive countries grew at different rates (Sawahel 2022). Duremeijer et al. (2018) report that Africa had the strongest growing production and that the period between 2012 and 2018 saw almost 40% growth in the published research. ICT resources, open free and low-cost access to peer review literature, improved research infrastructure and more might be the factors contributing to the increasing trend in the production of publications (Schemm 2013).

[27] Data is collected from the International Telecommunication Union (ITU), a specialized agency of the United Nations for information and communication technologies (accessed via https://www.itu.int/).

[28] Hjort and Poulsen (2019) use a five hundred meters distance from the Internet backbone network to define connectivity, yet they show that the point estimates remain significant well beyond the five hundred meters radius used to define connected locations. They mention that they chose a conservative radius based on copper-cable technologies and that their empirical strategy might underestimate the true effect of the fast Internet. They compare the connectivity of residential buildings in their paper, while I compare the connectivity of buildings in a campus environment. Residential buildings could be connected through copper wires or multimode fiber optic cables, where

$Post\ Submarine\ Cable_{ct}$ is a dummy variable that takes the value of 1 if at least one modern submarine Internet cable has arrived country c's shores at time t. X_{it} is a vector of time-varying university covariates, including the number of students enrolled, the student-to-faculty ratio, the number of schools, and the number of faculty members with and without research responsibilities. The country × year fixed effect, μ_{ct}, is included to isolate time-varying changes across countries. The γ_i represents university fixed effect that would account for any permanent differences across universities. The standard errors are clustered at the country × year level.[29] The coefficient of interest, β_1 , estimates the within-country impact of high-speed broadband on research production by comparing universities located within five kilometers of the Internet backbone network relative to those that are farther after the arrival of submarine cables to the country.

1.4.2 Results

Submarine Cables' Arrival and Internet Use:

Before showing the impact of the fast Internet on research production, Online Appendix Section (A3) details how submarine cables' arrival affects Internet use in Africa. I show, in Online Appendix Table (1A), that individuals living within five kilometers of the country's

both are used for short-distance connections (FS community 2022, Cleerline 2019, Gannon 2016, Photonics Media 2015). According to Fiber Optic Solutions, multimode cables are used for short distances such as residential buildings, and running over 3,000 feet (equivalent to 914.4 meters) causes signal distortion resulting in unclear and incomplete data transmission (accessed via http://www.fiber-optic-solutions.com/fast-fiber-optic-cable-speed.html). As stated by the Fiber Optic Association (FOA), single-mode fiber optic cables are used to connect buildings in the campus environment to the outside world (campus backbone) (accessed via https://www.thefoa.org/tech/ref/OSP/design.html). These cables are not manufactured in lengths more than five kilometers (accessed https://www.thefoa.org/tech/ref/basic/basics.html). Two optical fiber cables could be connected for extra length in two different ways: with splices or connectors. Connectors and splices affect the performance of all fiber optic cables. According to Fiber Optic Cabling Solutions, a signal loss could occur when the optical signal passes through each splice and connector. The more splices and connectors you have in a segment, the greater the loss on the line (accessed via https://www.cables-solutions.com/tag/single-mode-fiber). The total amount of the loss depends on the types, quality, and number of connectors and splices according to Fiber optic network products (accessed via https://www.fiberopticshare.com/factors-that-limit-optical-transmission-distance.html). There are always at least two connectors per fiber segment. It is important to note that even the highest quality connectors can get dirty. Dirt and dust, for example, could completely obscure a fiber light wave and cause high losses (accessed via https://www.cablinginstall.com/testing/article/16466088/the-importance-of-measuring-fiber-loss-and-distances). In addition, according to a report by World Bank & Knowledge Consulting Ltd. (2021), small campuses (having 3,000 as the average number of students, two medium, and three large buildings) in Africa need five kilometers of fiber optic cables, while medium and large campuses need ten and twenty kilometers of cables, respectively to be connected to the outside world. Small campuses account for about 94% of all higher education institutions in Africa. In section (1.4.3), I show that the results are robust to varying the radius used to define connectivity.

[29] Results are robust to clustering on the university level, to multiway clustering on university and year, university and country, country and year, as well as to wild-t bootstrap clustering at the year and country level.

Internet backbone network are more likely to use the Internet at least once per week compared to individuals living farther away.

Submarine Cables' Arrival and the Number of Publications:

The main results from estimating Eq. (1.1) are presented in Table (2). Column (1) reports the results for the full sample where no controls are included.[30] In column (1), universities near the Internet backbone network produce around 87 publications more than farther universities after the arrival of submarine cables. This impact is about 65 percent of the mean of the outcome, and it is around 102 percent relative to the mean of the outcome of the comparison group.[31] Results in columns (2) and (3) pertain to the subsample, including Egypt and South Africa only.

The point estimates in column (2) show that treated universities have 213 more publications compared to universities farther away. This impact corresponds to 45 percent of the mean of the number of publications, and around 94 percent relative to the mean of the outcome of the comparison group.[32] The results in column (3), where the set of controls is included, are very similar to the point estimates presented in column (2). In Online Appendix Table (2A), I use the number of submarine cables as a continuous variable indicating the number of submarine cables arriving at a country's shores. I find a 20 percent increase in the number of publications with the arrival of each submarine cable. Results from the subsample show that treated universities in Egypt and South Africa produce 33 more publications with each submarine cable that arrives on the country's shores, which corresponds to a 7 percent increase in the number of publications.

[30] All countries in the sample are included (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania).

[31] I re-estimate Eq. (1.1) using the log of the number of publications. To transform the outcome variable into logs, I have to transform it into a log plus a constant in order not to lose observations from universities with zero publications in some years. The point estimate from using log of publications plus one is 0.004 and the standard error is 0.043. The treatment size and significant changes with arbitrary rescaling the number of publications. Recently, Chen and Roth (2023) have argued that one can obtain a treatment effect of any magnitude simply by rescaling the units of the outcome before taking the log-like transformation. Mullahy and Norton (2024) show that transforming non-negative outcome variables into a log plus a constant raises many concerns and doubts. Also, McConnell (2023) argues that a policy or intervention can raise outcomes for the treatment group using one functional form but may show the precise opposite using a different functional form, even with the same data, the same sample period, and the same control variables.

[32] Re-estimating Eq. (1.1) on publication per researcher level in the subsample (EGY & SA) shows that universities near the Internet backbone network have 4.4 percent higher publications per researcher per year, relative to the mean of the outcome, compared to the ones farther away after the arrival of submarine cables. However, this impact is considered to be statistically small. The number of publications in the subsample (EGY & SA) increases by around 1236, and on average we have 213 faculty members with research responsibility in the subsample, so publications per researcher increase by around 6.

To investigate whether the subsample (Egypt and South Africa) is driving the results, I re-estimate Eq. (1.1) excluding the subsample.[33] The results are presented in Online Appendix Table (3A). In column (3), universities near the Internet network backbone show a 70% increase in their number of publications after the arrival of high-speed broadband compared to universities farther away. The results from column (3) are very similar to the results from the full sample in column (1) showing that Egypt and South Africa are not alone responsible for the impact of high-speed broadband arrival.[34]

1.4.3 Robustness Checks and extensions

Test for Parallel Trends:

Note that the causal interpretation of the estimates presented in Table (2) is based on the "parallel trends" condition, which means that universities near the Internet backbone network would have trended in parallel fashion to farther away universities in the absence of submarine cables. To investigate whether the condition of parallel trends is satisfied, event study regression in the following form is estimated:

$$(1.2)\quad y_{ict} = \Sigma_{\substack{r=-10 \\ r\neq -1}}^{10} \alpha_r\,(Time\ Relative\ to\ Submarine\ Cable\ Arrival\ r_{ict} \times Near\ Internet\ Network_i) + \mu_{ct} + \gamma_i + \epsilon_{ict}$$

The number of publications is considered the outcome variable.[35] $Time\ Relative\ to\ Submarine\ Cable\ Arrival\ r_{ict}$ is a vector including several dummy variables for years relative to the time of the entry of submarine cables. Specifically, in Eq. (1.2), I include dummies for 1, 2, 3 to 10+ years after the arrival of submarine cables, as well

[33]For this exercise, all countries except Egypt and South Africa are included.
[34] The impact of the fast Internet on research production in each country is discussed in detail in section (1.4.3).
[35] I also conduct an event study analysis for all the other outcome variables presented in Panel A of Table (1). The treated universities are not different from the untreated ones before the arrival of the submarine cables for all the outcome variables except the number of single authorship publications. Further, I construct an aggregate measure of the research output by first standardizing the following 11 outcome variables from Panel A of Table (1): publications, articles, books and book chapters, conference papers, international, national, and institutional co authored papers, single authorship publications, publications in the top 25% journal percentiles, citation count, and the number of awards. Then, I add all these standardized variables together to create an aggregate measure of research output, and I perform an event study analysis for this aggregate measure. The results show that universities near the Internet backbone network were no different from universities farther away in terms of the constructed aggregate measure of research before treatment.

as dummies for years before submarine cables' rollout.[36] If the estimated impact of high-speed broadband on publications was trending upward even before submarine cables' arrival, this would imply that universities near the Internet backbone network were positively selected on publication trends and that post-fast Internet differences are unlikely to reflect the causal impact of submarine cables' rollout.

Figure (3) presents the coefficients of the relative time estimated in Eq. (1.2) for the full sample. The excluded category is One Year before the arrival of submarine cables. Therefore, the coefficient estimates of the indicators in Time Relative to Submarine Cable Arrival are interpreted relative to that year. The solid bars represent the point estimates, and the capped lines are the 95 percent confidence intervals. The coefficients of the yearly time dummies before fast Internet access are not statistically significant and close to zero. This graphical evidence shows the validity of the parallel trend assumption. After the submarine cables' rollout, publication activity diverges. The findings show that the number of publications by a university's faculty increases about three to four years after access to high-speed broadband.[37]

Alternative Measure for Being Near the Internet Backbone Network:

In this section, I investigate the robustness of the baseline results to the radius used to define locations near the Internet backbone network. I group universities into smaller different distance bins, and estimate the following regression:

$$(1.3) \qquad Y_{ict} = \beta_0 + \beta_1 (Distance\ 0\ to\ 0.5\ km_i \times Post\ Submarine\ Cable_{ct}) + \beta_2 (Distance\ 0.5\ to\ 2\ km_i \times Post\ Submarine\ Cable_{ct}) + \beta_3 (Distance\ 2\ to\ 3.5\ km_i \times Post\ Submarine\ Cable_{ct}) + \beta_4 (Distance\ 3.5\ to\ 5\ km_i \times Post\ Submarine\ Cable_{ct}) + \beta_5 (Distance\ 5\ to\ 6.5\ km_i \times Post\ Submarine\ Cable_{ct}) + \beta_6 (Distance\ 6.5\ to\ 8\ km_i \times Post\ Submarine\ Cable_{ct}) + \beta_7 (Distance\ 8\ to\ 9.5\ km_i \times Post\ Submarine\ Cable_{ct}) + \mu_{ct} + \gamma_i + \epsilon_{ict}$$

Figure (4) plots the relationship between fast Internet and the number of publications by distance to the Internet backbone network. Universities that are within five kilometers of the

[36]As an example, for a submarine cable arriving in a country in 2006, the first year of the submarine cable arrival dummy in 2006 is equal to 1, the second year of the submarine arrival dummy in 2007 is equal to 1, and so on. For the same country, the dummy for two years before submarine arrival is equal to 1 in 2004.

[37] The impact of high-speed broadband on research production seems to increase with time which might be due to an increase in the number of journals or an increase in the faculty's communication network as shown in section (1.4.4).

Internet backbone network experience an increase in the number of publications after fast Internet arrival compared to universities located at 9.5 km and farther away from the backbone network.[38] After the five kilometers radius, the relationship between submarine cables' arrival and the number of publications disappears, providing empirical support for my decision to restrict the effect of high-speed broadband to universities within five kilometers for the main analysis.

In Online Appendix Table (4A), I exclude from the sample all universities located more than ten, twenty, and fifty kilometers from the Internet backbone network.[39] The results show that the point estimates are similar to those obtained from Table (2).

Using Pre-existing Universities Only:

Next, I analyze the robustness of the main results to the inclusion of only the universities founded in the 1990s, i.e., before the entry of fast Internet to the country. I do this to alleviate concern about the possibility that universities might be choosing their locations to be closer to the country's Internet backbone network and experience higher-speed Internet as a result. I re estimate Eq. (1.1) using a sample of 185 universities built a long time before the arrival of submarine cables or the introduction of their country's Internet backbone network. Appendix Table (1A) shows that pre-existing universities near the network produce around 53 percent more publications than universities farther away. The results are similar to the main impact found in Table (2), suggesting that endogenous re-location of universities is not driving the results. In addition, I run another regression to check whether submarine cable arrival has any effect on the location of the newly built universities with the Internet backbone network.[40] The point estimates in Online Appendix Table (5A) show a null impact of high-speed broadband access on the new universities' distance from the country's backbone network. Thus, the endogenous construction of new universities is also not the primary driver of the main results.[41]

---

[38] Using the F-test, I check the equality of the first four coefficients as well as the equality of all the coefficients. I fail to reject the significant joint equality of the first four coefficients, $\beta_1, \beta_2, \beta_3, and\ \beta_4$ in Eq. (1.3), yet I could reject the null hypothesis of the equality of all the seven coefficients.

[39] Online Appendix Figure (3A) reports the number of universities by their distance from the Internet backbone network. Around 85% of the universities in the sample are located within ten kilometers of the Internet backbone network.

[40] Universities that are built after the submarine cables' arrival.

[41] Choosing the right location is one of the key factors that determine the success and profitability of institutions (Tropea 2019, Coombes et al. 1991). Public universities in Africa rely heavily on government funding, which is determined by the number of students, while private universities depend on students' tuition fees to cover their costs.

External shocks:

One of the concerns is that the point estimates could be picking up not only the impact of fast Internet but also the impacts of some potential external shocks such as researchers' reallocation and government funding. I believe that government funding in Africa is not sufficient enough to change research production, as well as it does not depend on the number of publications produced by universities. Most African public universities rely heavily on limited government funding to cover their expenditures, including research (Ishengoma 2018).[42] Public universities charge low tuition fees, but this is usually only a fraction of the total economic cost. They rely heavily on limited government funding, so they struggle to cover costs or diversify their revenue streams (Kolade 2022, Nganga 2022). Government funding for public universities in Africa is generally declining, leading to many challenges (Ishengoma 2018). Governments in Africa have several priorities and obligations to achieve, using their insufficient resources, thus such universities are struggling financially since they are fully funded by the government and incapable of internally generating adequate revenues (Kolade 2022, Ishengoma 2018). The insufficient funding of research is identified as one of the major obstacles to research productivity in Africa (Dhalla and Guirguis 2018, Bonfoh 2016, Kollmann et al. 2015). Africa has sustained public investment in higher education at around 0.78% of GDP over the past 15 years (Ishengoma 2018, Wolhuter et al. 2014).[43] In addition, Baro et al. (2017) show that most of the research conducted by academic faculty in Nigeria is self-financed through their low salaries. Public funding to higher education institutions in Africa depends mainly on the number of students (Kolade 2022, Wahab 2022, World Bank 2010). This lack of research funding is also contributed by a sharp increase in enrollments and a decrease in per-student spending across the continent, favoring allocations to teaching over research (Ishengoma 2018).

In both cases, higher student enrollment would lead to higher funding. Thus, the need for huge space, access to infrastructures such as transportation and communication networks, and other universities' location are not the only factors to determine the right location of a new campus. Attracting potential students and employees by locating in areas with high primary and secondary enrollment is taken into consideration when opening a new campus in Africa (World Bank 2010). As shown in Panel (B) of Table (1), the average years of education for adults aged twenty and above is the same in areas where treated and untreated universities are located.

[42] Relying on government funding and underfunding are common even in countries with strong economies such as South Africa (Ishengoma 2018).

[43] This share drops to 0.63% in the poorest countries of Africa, and most of the African countries spend less than 1% of their GDP, on average, on research and development (Ishengoma 2018). According to the OECD's 2018 "Education at a Glance" report, public spending on higher education in the United States is 1.3% of GDP, and it is equal to public spending in Switzerland and the United Kingdom.

In the absence of data on government funding and researchers' allocation at the university level in Africa, shortening the post-treatment period to five or three years could partially address this concern since an increase in government spending or reallocation of researchers takes time. As shown in columns (2) and (3) of Appendix Table (2A), the impact of submarine cables' arrival on the number of publications is still largely significant after restricting the post-period to five and three years, respectively. The results show that external shocks such as government funding and reallocation of researchers might not change much the impact of high-speed broadband on research production in Africa.

Impact on Each Country Separately:

The empirical strategy used identifies the treatment effect by exploiting the variation in the timing of the arrival of submarine cables. Unlike in classical difference-in-differences designs, the treatment and comparison groups are not fixed. Put differently, universities become treated at different times. This setting is referred to as staggered DD or two-way fixed effect estimation. To investigate this issue, I estimate simple DD regressions for each country separately. In these regressions, there are clearly defined treatment and comparison groups, as well as pre- and post-treatment periods. The results, displayed in Appendix Figure (3A), indicate that the arrival of submarine cables significantly increased the number of publications in all the countries in the sample except for Libya. Libya was connected to its first modern cable in 2011 when its first civil war took place. The civil war affected Libya negatively in different aspects, and one of them is probably through research activity. Online Appendix Figure (4A) shows the number of publications in each country separately by the proximity of the universities from the Internet backbone network. The shape of the graph in many countries is virtually similar for universities near vs farther away from the Internet backbone network before the arrival of submarine cables. The gap in the number of publications between near and farther away universities starts to increase and widen further over time after the arrival of the fast Internet in many of the illustrated countries. Appendix Figure (3A) and Online Appendix Figure (4A) show also that there is a large variation in the impact across countries. There are several reasons behind this variation. First, there are differences in the number and capacity of submarine cables connected to each country's shore as shown in Appendix Figure (4A).[44] Second, submarine cables alone cannot improve Africa's connectivity challenges (Walker 2018). The national Internet backbone network is the main cable connecting other cables and linking them to the international traffic in a country. Telecommunication infrastructure disruptions stand as one of the barriers to rapid Internet expansion (Foster et al. 2021, Mannisto et al. 1997). Although many countries have taken major steps to upgrade and maintain their backbone network, some countries in Africa are still using less than 20% of the available capacity of submarine cables (Chattopadhyaya 2019, Grigorik 2015). Low quality and inadequate maintenance leave the Internet backbone network exposed to backhoe accidents, water, and rodent damage, and limits access to the Internet (Ankilu 2022).[45]

Infrastructure Other than the Internet:

In many developing countries, the Internet backbone network runs parallel to other infrastructures such as roads and electricity (Agrawal, Galasso, and Oettl 2017).[46] If universities near such infrastructure saw higher publication growth over time, irrelevant of whether they were also near the Internet backbone network, it would be possible that the impact of fast broadband on research production is misreported. Using Africa's power stations and road network maps, I calculate the distance between a university and the nearest power station or road network. I estimate a version of Eq. (1.1) that interacts separately *Post Submarine Cables* with an indicator for whether a university is within five kilometers of power stations and road network, respectively. When it is included, the estimated effect of *Near Internet Network* × *Post*

---

[44] For example, Egypt received more than ten submarine cables, while Algeria and Cameroon received only two modern cables during the sample period. In addition, the capacity of submarine cables differs substantially. According to the submarine cable networks, submarine cable capacities vary from less than 2Tbps (FALCON,SEACOM) to more than 20Tbps (HAWK, SEA-ME-WE 5, AAE-1), and some submarine cable systems have been upgraded several times over the years.

[45] For example, Haitham Zahran, the vice president of EMEA Subsea Cable Systems, said that the reason why places like South Africa and Egypt are considered hubs is that they have good backbones (Bannerman 2021). According to the African Tertiary institution connectivity survey 2006 report, Morocco and most North African institutions have a higher capacity for connection because of the presence of good national backbones (Echezona et al. 2010). According to the African e-Connectivity Index, South Africa is the top-ranked African county in terms of Internet quality connectivity followed by Egypt (Mitchell 2021). 46 Broadband networks could be seen in a similar way to the develop

[46] Broadband networks could be seen in a similar way to the development of roads, railways, and electricity networks. They are considered amenities in this context. Laboratory facilities, public libraries, and other amenities near the road and electricity networks could help researchers and scientists access information and validate their experiments.

Submarine Cables is essentially unchanged as shown in Appendix Table (3A), and the estimated coefficients on road and electricity coverage are insignificant.

Additional Robustness Checks:

In the Online Appendix section (A4), I conduct several additional robustness checks. First, a recent line of research highlights the deficiencies in staggered DD designs with two-way fixed effects when the treatment effects are heterogenous (Goodman-Bacon 2021, Callaway and Sant'Anna 2021, Sun and Abraham 2020, de Chaisemartin and D'Haultfœuille, 2020). To deal with this concern, in Online Appendix Figure (5A), I show that the results are robust to using the Interaction-weighted estimator proposed by Sun and Abraham (2020). Second, in Online Appendix Table (6A), I show that the baseline results are robust to excluding or including any of the control variables. Next, I restrict the sample to three and five years before and after the treatment for each country in Online Appendix Table (7A). Following the arrival of high-speed Internet, universities near the Internet backbone network still show a significant increase in the number of publications compared to universities located farther away. Also, I trim the sample of universities, found before the arrival of submarine cables, in the full sample to drop the outliers. In Online Appendix Table (8A), I show that, after dropping universities with more than 100 publications from the sample, universities near the Internet backbone network increase their number of publications by around 55% compared to universities located farther away after the submarine cables' arrival.

1.4.4 Mechanisms and Heterogeneity

In this section, I consider two potential mechanisms that could explain the observed increase in the number of publications. The emergence of information and communication technologies (ICT), especially the Internet, might have facilitated communication between authors. Consequently, new patterns of co-authorship might have been developed.[47] Imhonopi and Urim (2011) and Barjak (2006) argue that the use of emails, knowing when and where conferences are held, joining chat rooms, and video conferencing are useful mediums of communication between scientists. Iaria et al. (2018) show that reduced international scientific cooperation due to World War I led to a decline in the productivity of scientists who relied on foreign research from

[47] The Internet has enhanced the possibility of communication between people in different parts of the world.

abroad. Carillo et al. (2013), Ductor (2014), Aldieri et al. (2017), and Aldieri et al. (2020) find a positive relationship between research co-authorship patterns and the performance of academic institutions. Besides, the Internet is an easy way to gain knowledge through access to news, debates, electronic books, journals, and publications that are not available in local libraries. Frank Werthvein, a physicist at UC, said that after the arrival of a series of high-speed fiber optic cables to the University of California in 2015, scientists working with data collected by CERN (Switzerland) could use these experimental data and run programs from remote locations at significant cost savings (Markoff, 2015). Berkes and Nencka (2021) show that Carnegie libraries increased patenting by approximately 7 to 11 percent due to increased access to scientific knowledge. Moreover, Bryan and Ozcan (2020) find that open access journals cause patents to cite academic knowledge much more frequently. I also investigate whether certain fields or types of publications are more affected by the arrival of fast broadband.

To test the impact of high-speed broadband on communication between researchers, I consider three types of co-authored publications: international, national, and institutional coauthored publications. I estimate analogs of the baseline model with different types of coauthored publications as the outcome variable. Columns (1)-(3) in Table (3) show the results of this analysis. The point estimates indicate that co-authored publications increased in treated universities relative to untreated ones after the arrival of submarine cables. For example, highspeed connection increases international and national co-authored publications by more than 75 percent. The impact on institutional co-authored publications is smaller yet significantly large in magnitude relative to the mean. In column (4), I perform a similar exercise using single authorship publications as the outcome variable. The estimated coefficients from Table (3) provide evidence that the fast Internet changes the practice of science by offering greater knowledge and facilitating communication between researchers.

In Online Appendix Section (A5), I discuss a potential concern of whether the effect found in Table (2) reflects higher real production of publications or higher collaboration between co-authors. The overall yearly publication production may remain unchanged, yet the increase in the number of co-authors per publication creates a risk of potential duplication of papers in my sample. To deal with this issue, I eliminate the double-counted papers as proposed in Online Appendix Section (A5). The findings from this exercise, presented in Online Appendix Table

(9A) show that high-speed broadband not only increases co-authorship but also increases the real production of publications.

In Online Appendix Section (A6), I examine whether the effect of the arrival of submarine cables on research output varies depending on the type or field of publication. Online Appendix Table (10A) suggests a positive and significant impact of the high-speed broadband on all types of publications, with a generally larger effect on books and book chapters. The point estimates, presented in Online Appendix Table (11A), show that the arrival of submarine cables has the largest effect, a 90 percent increase, on medical publications. The impact on the other fields ranges between a 40 to 55 percent increase.

1.4.5 Publications' Quality

Having addressed the effect of high-speed broadband on publications' quantity, I turn to examine its effect on quality. In this section, I re-estimate Eq. (1.1) using different measures of publication quality. Tables (4) and (5) include the parameter estimates from this exercise for four measures of quality: the probability of getting an award, total citation count, the number of publications in predatory journals, and publications count and share in the most-cited journals indexed by Scopus.[48] Columns (1) and (2) of Table (4) show that the probability of getting an award and the number of citations increase by more than 50 percent for publications produced by the treated universities after fast Internet access.

A positive and significant impact on the number of publications in the most-cited journals is observed in columns (1)-(3) in Table (5), yet the larger impact was on publications in lower quality journals. Columns (4)-(6) show the effect on the share of publications in the most-cited journals. The share of publications to the top 10 percent of the most-cited journals produced by the treated universities drops by around four after submarine cables' arrival, a reduction of 33 percent. The point estimates in column (6) show that the quality of research output published by universities near the Internet backbone network is lower relative to universities farther away. In

[48] The probability of getting an award is the probability of receiving a grant award from major funding organizations across the UK, USA, and Australia. Grants in SciVal are assigned to institutions based on their Scopus affiliation at the time of the grant being awarded. Citation count is the total number of citations in a year that the university's publications received.

column (7), I match the publications' journals with Beall's list of predatory journals.[49] The parameter estimates show significant growth in publications in predatory journals after the rollout of high-speed broadband. The findings in Tables (4) and (5) suggest that while high speed Internet has increased the number of publications, there is no clear impact on the quality of these publications. The Internet could affect publications' quality negatively in different ways. First, Carrell and Simoni (2017) show that in the recent publishing environment, it's very easy to get a publication. They discuss that more access to publishers has led to more papers being published, and higher demand for reviewers which caused lower access to quality reviewers, more low-quality submissions, and so a decline in the quality of published manuscripts. Second, some imprecise scientific articles are published in what seems to be peer-reviewed journals but are open-access publishers that accept whatever is submitted for a fee (Clarke 2022). Previous papers showed that there is a substantial increase in publishing in predatory journals, especially in developing countries (Machacek and Srholec 2021, Xia et al. 2015, Shen and Bjork 2015). Column (7) in Table (5) shows a significant growth in publications in predatory journals after the rollout of high-speed broadband. Third, Sarewitz (2016) argues that the Internet makes it much easier to find supporting papers, no matter whether good or not, so it's not surprising that citation rates are increasing. Scientists might use some online work as a foundation for their projects and cite it. It is becoming so difficult for scientists to evaluate the quality of papers when searching for literature (Besse-Lototskaya and Zwietering 2021).

1.4.6 Other Outcomes

As an extension, I explore whether other university characteristics are also affected by fast Internet access. I estimate analogous models of the baseline difference-in-difference regression in Eq. (1.1). For this analysis, the outcome variables are the rate of graduation and the probability of getting a world rank in a year. The results, in Table (6), indicate that submarine cables' rollout does not affect the graduation rate, while the probability of getting a world rank is higher for the treated universities.

[49] This list was also used by Mouton and Valentine (2017) showing that over 4,000 South African papers are published in 48 predatory journals with the largest increase in 2011. This quality measure is discussed in more details in section (1.3).

1.5 Conclusion

In this paper, I exploit a major change in Internet speed due to submarine cable rollout to the African coast in the late 2000s to investigate the effect of broadband access on the research production of universities in Africa. The identification strategy depends on the variation in the timing of submarine cables' arrival in a country and the fact that only the institutions that are located close to the existing network of a country benefit from the submarine cables. The causal effect of broadband on research production is identified under the condition that universities near the country's Internet backbone network (treated) are credible counterfactuals to those universities that are farther away (untreated).

I provide multiple pieces of evidence that support this assertion. First, I show that conditional on university and country-year fixed effects, there are no statistically significant differences in characteristics between universities near vs. farther away from their country's Internet backbone network. I also show that patterns in the number of publications in treated vs. untreated universities were similar prior to their access to high-speed broadband. In other words, treated and untreated universities followed a "parallel trend" in the absence of the treatment.

The main findings, which remain unchanged in several robustness checks, indicate that broadband access increases the number of publications by 65 percent. Like any identification strategy, strong assumptions have been made to estimate the effect of fast Internet on the number of publications. The effect could be smaller, which is the case for some of the countries in the sample according to Appendix Figure (3A). Increased access to knowledge on the web and easier communication between co-authors are potential mechanisms for research production gains. I also find that high-speed broadband increases research output in many different fields. However, I do not find clear evidence of the impact on the quality of the published works.

The results motivate the need for a policy solution to bridge the equal diffusion of high speed Internet across regions to connect more people. Over the years, there have been many local and international efforts by governments, telecommunication companies, and multinational organizations to improve Internet infrastructure networks in Africa. Delivering higher quality and stable connections across different areas in Africa through expanding the Internet network backbone could be a way to facilitate the growth of science and so highering the standards of living in Africa.

Table (1): Summary Statistics and Descriptions of the Variables

Variable	Description	Mean (Std. Dev.)	Treatment Group (Std. Dev.)	Comparison Group (Std. Dev.)
Panel A: Full Sample				
Publications	The number of publications the university indexed in Scopus	130.479 (332.668)	147.396 (373.557)	85.443 (176.407)
Articles and reviews	The number of published articles and reviews	107.531 (269.799)	122.545 (303.839)	67.562 (136.505)
Books and book chapters	The number of published books or book chapters	4.237 (16.414)	4.972 (18.180)	2.281 (10.079)
Conference papers	The number of published conference papers	15.191 (44.755)	15.565 (47.748)	14.196 (35.572)
Publications in Top 25% Journal Percentiles	The number of publications of a university that have been published in the 25% most-cited journals	42.419 (136.038)	49.550 (154.918)	31.961 (121.097)
International Co-authorship	The number of publications that have at least one international coauthor	56.809 (157.991)	64.498 (177.922)	36.341 (81.059)
National Co-authorship	The number of publications with all coauthors from the same country	31.958 (73.913)	35.175 (83.672)	23.394 (35.641)
Institutional Co-authorship	The number of publications with all the co-authors from the same institution	27.093 (73.823)	31.002 (82.146)	16.635 (42.959)
Single Authorship	The number of publications which are single authorship	14.766 (46.218)	16.939 (51.581)	8.981 (26.356)
Citation Count	Total citations received by publications of a university	2127.427 (6,565.768)	2,529.296 (7,525.450)	1,057.56 (2,354.916)
Awards	=1 if a university receives a grant award	0.032 (0.176)	0.041 (0.200)	0.006 (0.081)
Total universities	The number of universities	317	230	87
N	The number of observations	5,519	4,012	1,507

Table (1) Continued				
Variable	Description	Mean (Std. Dev.)	Treatment Group (Std. Dev.)	Comparison Group (Std. Dev.)
Panel B: Subsample (EGY & SA)				
Publications	The number of publications the university indexed in Scopus	474.16 (698.528)	605.413 (799.066)	226.836 (334.027)
Articles and reviews	The number of published articles and reviews	389.895 (564.488)	500.330 (647.455)	181.799 (251.994)
Books and book chapters	The number of published books or book chapters	17.980 (38.497)	23.318 (43.828)	7.922 (22.442)
Conference papers	The number of published conference papers	51.281 (79.132)	60.695 (81.471)	33.541 (71.363)
Publications in Top 25% Journal Percentiles	The number of publications of a university that have been published in the 25% most-cited journals	176.842 (310.157)	233.055 (362.205)	70.918 (115.156)
International Co-authorship	The number of publications that have at least one international coauthor	205.508 (343.831)	265.693 (397.699)	92.101 (153.3)
National Co-authorship	The number of publications with all coauthors from the same country	98.806 (136.991)	123.445 (159.188)	52.377 (55.809)
Institutional Co-authorship	The number of publications with all the co-authors from the same institution	99.174 (148.394)	125.631 (166.999)	49.321 (84.899)
Single Authorship	The number of publications which are single authorship	72.142 (103.375)	92.453 (116.519)	33.869 (54.954)
Citation Count	Total citations received by publications of a university	8688.950 (14849.67)	11550.96 (17419.55)	3295.981 (4484.918)
Awards	=1 if a university receives a grant award	0.086 (0.281)	0.114 (0.319)	0.034 (0.18)
Total universities	The number of universities	56	36	20
N	The number of observations	773	505	268

Table (1) Continued

Variable	Description	Mean (Std. Dev.)	Treatment Group (Std. Dev.)	Comparison Group (Std. Dev.)
Panel B: Subsample (EGY & SA)				
Students Enrolled	The number of enrolled students	55,643.17 (71,817.09)	66,446.33 (82,241.8)	35,286 (38,842)
Graduation Rate	The percentage of students who received a degree or diploma	18.221 (6.358)	18.858 (6.245)	17.022 (6.407)
Student to Faculty Ratio	The number of enrolled students divided by the total number of faculty members	35.887 (25.495)	38.689 (29.330)	30.607 (14.587)
Faculty with Research Responsibilities	The number of faculty members whose primary effort is in research	1,236.322 (1,419.932)	1,409.727 (1,576)	909.57 (985.786)
Faculty without Research Responsibilities	The number of faculty members who don't have research responsibilities	390.859 (538.129)	439.754 (601.626)	298.723 (375.518)
Schools	The number of schools within a university (ex. School of Medicine)	10.543 (5.609)	10.980 (6.079)	9.72 (4.491)
Top Rank University	=1 if a university is ranked among global universities	0.069 (0.255)	0.093 (0.29)	0.026 (0.159)
Population	population in the city where the university is located (in millions)	5.573 (2.595)	4.976 (2.721)	6.381 (2.172)
Education Years	Mean years of education of adults aged 20+	8.182 (1.386)	8.189 (1.476)	8.173 (1.258)
Total universities	The number of universities	56	36	20
N	The number of observations	773	505	268

Note: The unit of observation is university-year. The full sample includes 12 countries (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania) with 5,519 observations. There are 773 observations in the subsample which includes only two countries (Egypt and South Africa). Treated is the group of universities located within 5 kilometers of the country's Internet backbone network, while the comparison group consists of universities that are farther away.

Table (2): Impact of Fast Internet on The Number of Publications

	(1)	(2)	(3)
	All Countries	Egypt & South Africa only	
	Number of Publications		
Near Internet Network × Post Submarine Cable	86.849*** (10.619)	209.793*** (32.397)	213.110*** (31.773)
N	5,519	773	773
Controls	No	No	Yes
Country × Year FE	Yes	Yes	Yes
University FE	Yes	Yes	Yes
Mean of outcome	130.479	474.16	474.16

* p<0.1 ** p<0.05 *** p<0.01

Note: The table reports the difference in difference estimates from Eq. (1). Column (1) includes all the countries in the sample (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania). In columns (2) and (3), the sample is restricted to include only the subsample (Egypt and South Africa). Column (3) includes some universities' characteristics as controls: the total number of students enrolled, total faculty members with and without research responsibilities, student-to-faculty ratio, and the number of schools.

Table (3): The Impact of Fast Internet on Co-authorship

	(1)	(2)	(3)	(4)
	International Co-authorship	National Co-authorship	Institutional Co-authorship	Single Authorship
Near Internet Network × Post Submarine Cable	43.456*** (6.727)	25.388*** (2.949)	14.969*** (1.377)	3.120*** (0.637)
N	5,519	5,519	5,519	5,519
Controls	No	No	No	No
Country × Year FE	Yes	Yes	Yes	Yes
University FE	Yes	Yes	Yes	Yes
Mean of outcome	56.809	31.958	27.093	14.766

* p<0.1 ** p<0.05 *** p<0.01

Note: The Table reports the point estimates from Eq. (1) using collaboration as the outcome. Columns (1) – (4) includes all the countries in the sample (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania).

Table (4): The Impact of Fast Internet on the Quality of Publications

	(1)	(2)
	Awards	Citation Count
Near Internet backbone Network × Post Submarine Cable	0.055***	1110.621***
	(0.008)	(164.670)
N	5,519	5,519
Controls	No	No
Country × Year FE	Yes	Yes
University FE	Yes	Yes
Mean of outcome	0.032	2124.427

* p<0.1 ** p<0.05 *** p<0.01

Note: The table reports the point estimates from Eq. (1) using some measures of publications' quality as the outcome variables. Columns include all the countries in the sample (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania).

Table (5): The Impact of Fast Internet on the Number and Share of Publications in the Most-Cited Journals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Number of Publications			Share of Publications			
	top 10% most-cited journals	between 10% and 25% of most-cited journals	under 25% of most-cited journals	top 10% most-cited journals	between 10% and 25% of most-cited journals	under 25% of most-cited journals	Publications in Predatory Journals
Near Internet backbone Network × Post Submarine Cable	6.625***	12.671***	67.554***	-4.471**	-1.308	5.780**	1.503***
	(1.669)	(2.174)	(8.551)	(1.822)	(1.876)	(2.904)	(0.341)
N	5,519	5,519	5,519	4,808	4,808	4,808	5,519
Controls	No	No	No	No	No	No	No
Country × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
University FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of outcome	18.964	23.633	87.881	13.422	17.938	68.639	1.741

* $p<0.1$ ** $p<0.05$ *** $p<0.01$

Note: The table reports the point estimates from Eq. (1) using the number or share of publications in the most-cited journals as the outcome variables. Columns include all the countries in the sample (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania).

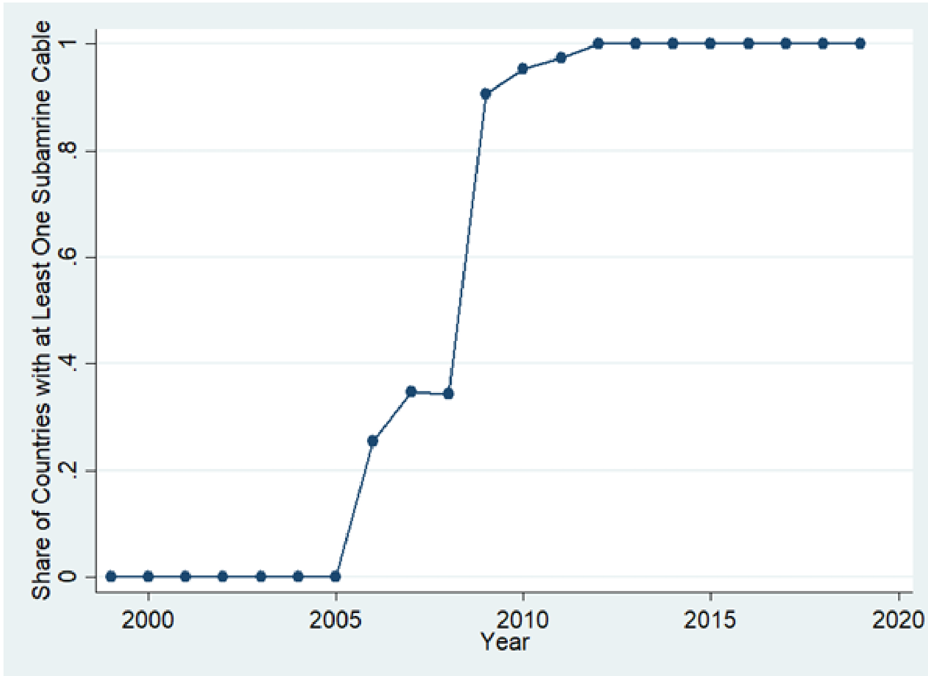
Table (6): The Impact of Fast Internet on Graduation Rate and World Rank

	(1)	(2)
	Graduation Rate	World Rank
Near Internet backbone Network × Post Submarine Cable	0.582	0.099***
	(0.470)	(0.034)
N	773	773
Country × Year FE	Yes	Yes
University FE	Yes	Yes
Mean of outcome	18.221	0.069

* p<0.1 ** p<0.05 *** p<0.01

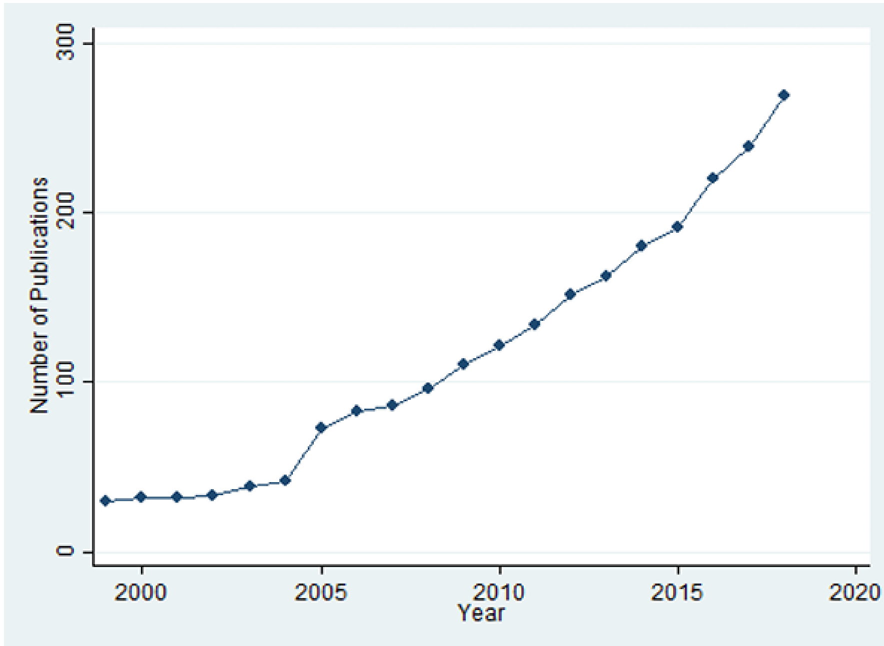
Note: In this table, I use the subsample including only 2 countries (Egypt and South Africa).

Figure (1): The Share of Countries in the Sample with At Least One New Submarine Cable Connection



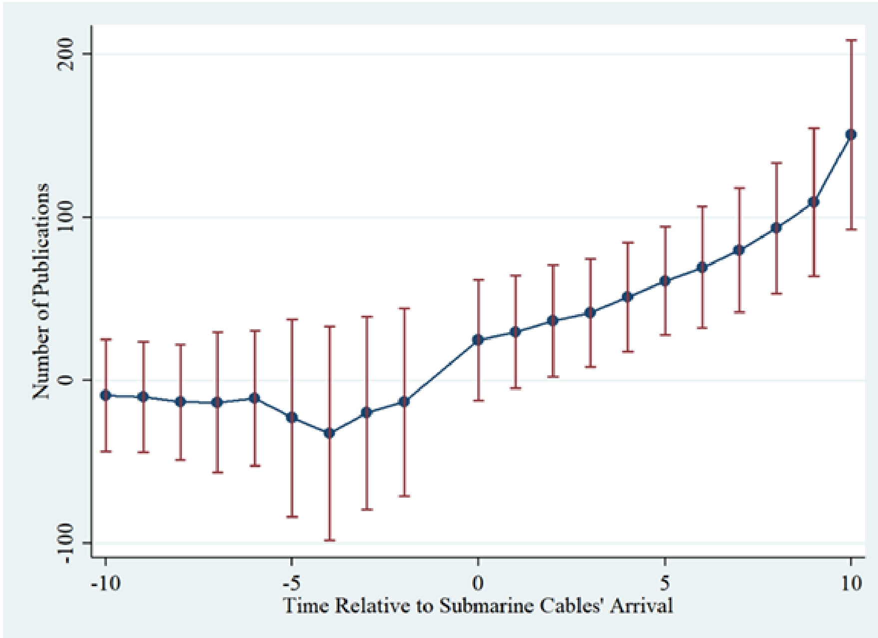
Note: This figure shows the share of countries in the sample with at least one modern submarine cable over years. The countries included are Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania.

Figure (2): Over Time Variation in The Number of Publications



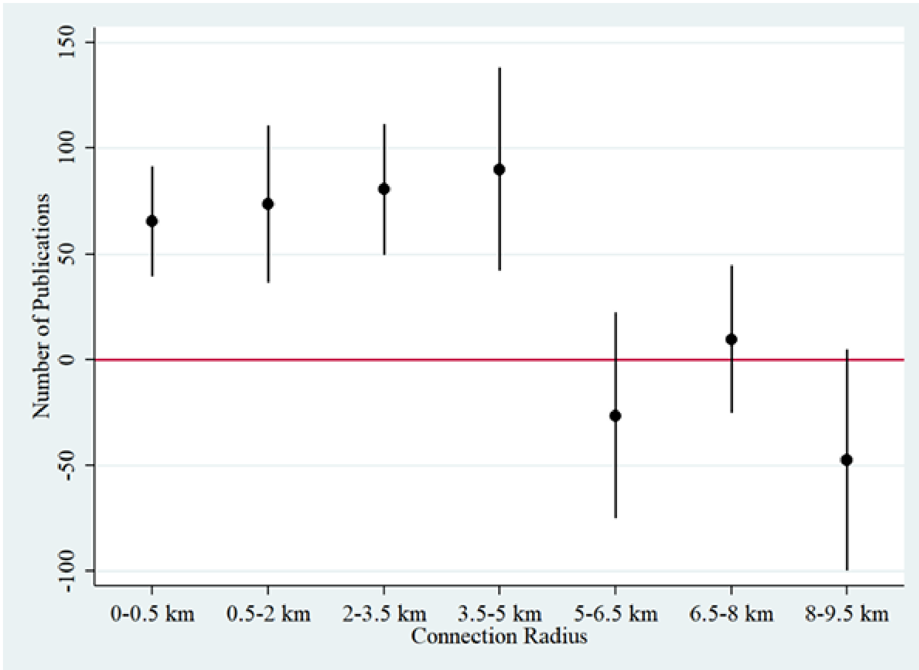
Note: This figure presents the evolution of the number of publications over years. The countries included are Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania.

Figure (3): Event Study Analysis



Note: The outcome is the number of publications. The blue line shows the point estimates of the indicators in *Time Relative to Submarine Cable Arrival* in Eq. (2). The red lines contain the 95% confidence intervals. The horizontal axis indicates the year relative to the time of the submarine cable entry. One year Before the submarine cable arrival is the omitted category. The countries included are Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania.

Figure (4): The Impact of Fast Internet on the Number of Publications by Distance



Note: This figure plots the coefficients from estimating Eq. (3) using different connection radii to define connectivity. The black lines contain 95% confidence intervals.

Chapter 2

Local Labor Market Effects of Nuclear Power Plants

2.1 Introduction

The 1954 Atomic Energy Act integrated commercial nuclear power into US domestic energy policy and granted the federal government regulatory control over the construction and operation of private reactors (Lemov, 1964). By the early 1970s, commercial nuclear reactors had sprung up across the US. Despite greater efficiency, reliability, and lower emissions, the trajectory of commercial nuclear power was disrupted by the March 1979 Three Mile Island (TMI) accident, which heightened public skepticism and elicited political backlash toward commercial nuclear energy (Walker, 2004).[50] More recently, the discourse surrounding nuclear power has focused on the rising costs associated with it and the challenges of maintaining its cost-competitiveness as a source of clean energy in the future (Lovering, Yip, and Nordhouse, 2016; Eash-Gates et al., 2020).

While nuclear power is a controversial geopolitical issue, proximity to a Nuclear Power Plant (NPP) tends to heighten the intensity of the host community's opinions toward them. On the one hand, locals are most concerned about nuclear accidents because they are at the greatest risk of an intense impact compared to neighboring areas.[51] For example, our analysis (presented in Appendix Table A.4) of the respondents to a CBS News/New York Times Poll in April 1979 suggests that individuals living near an NPP were more aware of TMI, less inclined to support building more NPPs, and believed similar accidents were more likely in the future, indicating a stronger distaste for NPPs compared to those living farther away.[52] Other research suggests that residents near prospective NPP sites generally show less acceptance and are swayed by emotional identification and social trust rather than rational cost-benefit analysis (Guo & Ren, 2017). On the other hand,

[50] Europe and Japan experienced strong public reactions to the 1986 Chornobyl disaster in Ukraine (Plokhy, 2018) and the 2011 Fukushima disaster in Japan (Lochbaum et al., 2014). For example, Mehic (2023) shows that the radioactive fallout from Chernobyl caused a major increase in the Green Party's vote share in the elections following the disaster.

[51] For example, immediately after the Chernobyl meltdown, the city of Pripyat (about three miles from the plant) was completely evacuated, and the area within the thirty-mile radius of the plant was declared the "exclusion zone," which remains uninhabited even today.

[52] These data, available through the Inter-university Consortium for Political and Social Research at the following link: https://doi.org/10.3886/ICPSR07819.v1, include personal information of about one thousand randomly selected Americans as well as data on their tendency to support the construction of NPPs and knowledge of TMI. We regressed these outcomes on an indicator that takes the value of one if the individual lives by an NPP and personal attributes. Appendix Table A.4 presents the results from these regressions.

NPPs are advertised to bring economic benefits to host communities, such as job creation, tax revenue, and infrastructure support.[53] These benefits may improve the locals' acceptance of nuclear projects. For example, for rebuilding existing plants, acceptance in China is positively related to perceived local economic benefits and environmental concerns (Wang et al., 2020).

This paper examines the effects of NPP construction and operation on local employment and wages. It is crucial to study this topic for at least two reasons. First, analyzing the local labor market impacts of NPPs helps determine whether economic benefits accrue to locals. Secondly, given the substantial upfront investment needed to build and maintain NPPs, often located in remote, rural areas, the investigation of their local impact could enhance our understanding of the agglomeration effects associated with significant investments, as documented by other researchers (e.g., Greenstone et al. 2010).

Before a nuclear reactor becomes operational, it undergoes several developmental stages initiated by a commercial energy company's public announcement of its intention to build one. After vetting the designs and plans, the US Nuclear Regulatory Commission (NRC) issues a construction license. Post-construction, NRC grants a commercial operation license to the reactor if the grid connection is successful and additional test results are favorable. However, companies may cancel plans at any stage. By 2019, 102 out of 294 planned reactors were canceled due to factors like changes to energy demand forecasts, cost increases, and political issues. From publicly available sources, we compiled a comprehensive data set for US commercial nuclear reactors, capturing their milestones such as the dates of the announcement, issuance of construction and commercial operation licenses, and cancellation.

Site selection for NPP construction is guided by NRC mandates, which emphasize geographic and seismological factors for public safety. These guidelines ensure low radiation doses, consideration of natural and man-made hazards in plant design, adequate security, and identification of site-specific challenges for emergency planning.[54] Consistent with the NRC regulations, we show that counties with larger bodies of water, situated away from population centers, with lower elevations are more likely to be selected/considered for NPP construction

[53] See https://www.nei.org/advantages/jobs: "The nuclear energy industry is a powerful engine for job creation. The U.S. nuclear energy sector directly employs nearly 100,000 people in high-quality, long-term jobs. This number climbs to 475,000 when you include secondary jobs. We cannot afford to lose nuclear jobs by closing plants, nor can we afford to miss out on thousands of jobs that building new reactors will create." Also see, https://www.energy.gov/ne/articles/advantages-and-challenges-nuclear-energy.

[54] See https://www.nrc.gov/reading-rm/doc-collections/cfr/part100/index.html.

(regardless of whether the construction was ultimately completed) compared to the rest of the counties that were never considered for an NPP.[55] We also show that the considered and never-considered counties differ significantly in terms of a number of socio-economic dimensions.

Because of these pre-existing differences, rather than studying the counties in the whole nation, in our empirical analysis, we focus on a subsample of counties: those selected for an NPP. Some of them were "treated" in that they had a phase of NPP construction, while others were not treated, e.g., included among the "controls," because the energy companies canceled their plans before the construction began. Within this group of counties, we show that the physical and socio-economic attributes of the treated and untreated counties are similar, indicating that the counties considered for an NPP form a homogenous group. The risk of cancelation of an NPP project was almost non-existent before the mid-1970s. In fact, in only one control group county (out of 34), the NPP project was scrapped in the pre-1970s. To estimate the impact of having an NPP in the county as opposed to *fully expecting* but ultimately not getting one, we include in our sample only the counties in which its first NPP construction began in 1974 or later. We also exclude counties whose NPP constructions started in 1979 to avoid the potentially confounding influence of TMI. Later in the paper, we show that our results are robust to using the complete set of counties.

The baseline results show that NPP construction increases the county's employment-to-population ratio by approximately 10 percentage points, or 20 percent above the mean, with no significant impact from commercial operation. Per capita wages rise significantly during the construction phase as well. This wage impact is substantial in percentages, about 50%, but because of the low baseline wages in selected counties, the level increase is $8,500 (in 2019 dollars). In some, but not all, of the regressions, we find that commercial operations of NPPs increase wages in the county. Several additional tests reveal that the positive effect of NPP construction on employment and wages is robust to the estimation sample, the regression specification, and estimators that eliminate bias arising from treatment effect heterogeneity. However, we cannot say the same for the impact of NPP commercial operations, which is sensitive to some of these changes. Event study analyses verify that the control and treated counties follow a parallel trend in the pre-treatment period. We find results similar to the baseline findings when we estimate the impact of NPPs using an alternative synthetic control strategy described in the appendix.

[55] Throughout the paper, we refer to the counties where an NPP's construction is planned and announced as "selected" or "considered" counties.

NPPs influence some income components other than wages. For example, an NPP construction reduces transfers to individuals from the government (such as the welfare programs), possibly because some residents become ineligible as their wages increase during the construction phase. Our results point to a slight (3.5%) increase in total disposable income in a county due to NPP construction, which is smaller than the rise in wages and salaries partly because of decreases in other income components.

Our analysis of the sectoral impacts suggests that NPPs improve the labor market outcomes only in industries that are directly related to them. During the construction phase (typically lasts ten to fifteen years in our sample period), employment and earnings in the construction sector improve. The government sector also benefits during the first handful of years of NPP construction. Our analysis of local government finances supports this finding: revenues and expenditures of county, municipal, and township governments rise by 10-15% during NPP's construction. A completed and operational NPP appears to improve labor market prospects only in the public utilities sector. The other industries in the NPP-hosting counties do not benefit from the NPPs. Moreover, the growth in employment and wages in the largest (highest population) counties adjacent to the treated counties is not different from that of their counterparts neighboring the control counties, suggesting no economic spillover effects from NPPs. We detect only an economically negligible impact in the commuting zone where an NPP is erected only during construction. These results reinforce the idea that the labor market benefits of NPPs are local, short-lived, and concentrated in directly related industries.

Our paper contributes to the literature on place-based policies (e.g., see Greenstone et al., 2010; Neumark & Simpson, 2015). However, our context differs from Greenstone et al. (2010), who study agglomeration effects associated with large industrial plants, where the losers from the siting competition between municipalities create natural controls. Hence, in Greenstone et al. (2010), the control units expected treatment only with some probability. Our identification strategy is underpinned by control units, where treatment was *fully expected* – i.e., economic decisions after the announcement but before cancellation were likely predicated on the expected treatment – precisely the same way as the actually treated units. This feature mitigates the risk that the (potentially unobservable) investments undertaken by the economic agents (such as the municipalities competing for the large industrial plants) to manipulate the outcome confound the estimates of the treatment effects.

Most of the existing economic research on the nuclear energy sector studies either the environmental or health effects of NPPs (e.g., Freese et al., 2023; Jarvis et al., 2022; Neidell, Uchika, and Veronisi, 2021; Kawaguchi & Yukutake, 2017; Roth & Jaramillo, 2017; and Davis and Hausman 2016) or their impact on the housing market (e.g., Bauer et al. 2017; Fink & Stratmann, 2015; and Davis 2011). Some, such as Davis and Wolfram (2012), investigate the deregulation of markets using the nuclear power market as a case study. Hausman (2014) shows that this deregulation improved nuclear reactor safety. Despite the variety of research on nuclear energy, an overlooked question in this line of research, except for some economic impact studies that use Keynesian multipliers (e.g., Lewis 1986), is the effects of NPPs on the labor market. We are unaware of any other paper on this topic except for Ando (2015), who uses the SC method to case study the impact of eight NPPs established in the 1970s and 1980s in Japan. His paper demonstrates highly heterogeneous effects, ranging between zero to sixty percent, of NPPs on per capita income levels in the long run. Therefore, the issue of quantifying the labor market effects of NPPs is far from being settled. In fact, Ando (2015, p. 84) explicitly encourages others to study other NPP sites. We accomplish this in our paper.

Furthermore, besides its focus on the US NPPs, our paper differs from Ando (2015) in several ways. First, our analysis was able to differentiate the impact of the construction of NPPs from that of their operation. This is an important distinction as constructing an NPP requires a vastly greater amount of labor input than a fully operational NPP and thus may have a differential impact. Second, thanks to access to data, we are able to investigate components of total personal income (wages, transfers from the government, and so on) and employment and income generated in different sectors. More importantly, while in the SC methodology, the short-run pre-treatment outcomes of the treated and synthetic control units are matched, it is unclear whether the lack of investments/preparations (that went on in the locations where an NPP was planned to be built) shifted the path of the synthetic control units in the *long run*. In this context, our empirical design, which compares the NPP counties to the control group counties that were fully expecting an NPP, could be an improvement over Ando (2015), who estimates the differences between the NPP municipalities and the nearby coastal municipalities that do not have NPPs and never expected to have one.

Our estimates also enable researchers and policymakers to contrast the labor market impacts of NPPs with other renewable energy investments. For example, the effect of NPPs appears to be similar to solar electric plants in that they significantly enhance employment during their construction but not as much after they start operating (Fabra et al., 2023). The work of others, such as Costa and Veiga (2021) and Gonçalves, Rodrigues, and Chagas (2020), indicates short-lived employment effects for wind farms. In their examination of hydropower plants, Felipe et al. (2017) also point to short-lived (<15 years) economic impacts.[56] These convergent findings imply that sustained local economic development may not be achieved with such projects.

In the next section, we cover the fundamental aspects of nuclear regulation concerning site selection, construction, and operation, followed by the descriptions of our data set and empirical methodology in sections 2.3 and 2.4. Results are presented in section 2.5. The details of our alternative synthetic control strategy are discussed in the appendix. In section 2.6, we conclude with a comparison to the relevant literature and a discussion of the policy relevance of our work.

2.2 Background

2.2.1 Nuclear Power Regulation

The Nuclear Regulatory Commission (NRC) oversees US commercial NPPs.[57] A firm can sell energy generated in its NPP only after going through NRC's long, multi-step regulatory process. First, the NRC must issue a license to construct the NPP based on an assessment of preliminary safety and design information. Once construction is complete, the NRC evaluates the final design and other operational factors and determines whether to grant an operating license to the NPP.[58] Partly because NRC provides numerous opportunities for the public to weigh in with their opinion before issuing construction and operation permits, the licensing processes are typically extended. Even after obtaining the operating license and despite much scrutiny, an NPP cannot start commercial operations. For this, the nuclear reactor must be connected to the electricity grid for

---

[56] Felipe et al. (2017) show that hydropower construction in Brazil can boost GDP by 10% at its peak and significantly raise tax revenues. Yet, these effects often vanish soon after construction ends, with public revenues and economic activity reverting to near pre-construction levels within 15 years for hydropower projects.

[57] The U.S. federal government gained authority over commercial nuclear reactor construction and operation largely through the Atomic Energy Act of 1954. Before this act, the development and control of nuclear energy were strictly under federal domain, primarily for military applications. It delegated the authority to regulate the safety and licensing of nuclear facilities to the Atomic Energy Commission, which was later replaced by the Nuclear Regulatory Commission (NRC) through the Energy Reorganization Act of 1974.

[58] A construction permit and an operating license are essentially combined into one license by the NRC's alternative licensing system, which was established in 1989. In either case, the NRC must give its approval before the applicant can proceed with the construction and operation of a nuclear power plant (Duke Energy 2012; U.S. Nuclear Regulatory Commission (NRC) 2020).

the first time for initial testing (International Atomic Energy Agency, 2012). Upon successful results, the NPP owner company takes control of it from the contractors that built it and begins to sell the energy produced.[59]

Figure 5 displays the timeline of construction (blue bars) and commercial operation (red) licenses for reactors granted by the NRC. The construction of nuclear reactors in the US started in 1954 with the Shippingport Atomic Power Station in Pennsylvania. Most nuclear power reactors in the United States acquired construction licenses between 1960 and 1978. With TMI, the year 1979 marks the start of a hiatus in the issuance of construction permits that would last about three decades. As a result of this incident, public concern over NPP safety has grown significantly, and the NRC tightened and stepped up its regulatory oversight. Consequently, several nuclear reactor construction projects were canceled, although they were in progress (US Energy Information Administration, 2017). The figure also shows that there is a downward trend in the number of reactors that started their commercial operations after TMI.[60] This decline is because of the hiatus in the new constructions and increased regulatory standards.

The effect of the post-TMI period's tighter regulations can also be seen in Figure 6, which depicts the number of years between the announcement of a nuclear reactor and the commencement of power generation and sales. Nuclear reactors initiated in the 1950s to early 1970s began commercial operations within ten years of announcement, while those announced in the mid-to-late 1970s, around the time of TMI, took about 15 years. Reactors from the latter period likely faced stricter and changing standards.[61]

2.2.3 Site Selection of Nuclear Power Reactors

The map in Figure 7 depicts the geographic distribution of the NPPs in the US. When determining the location of their NPPs, energy companies, besides their profit motives, must consider Part 100 of NRC Regulations Title 10, which lays out several considerations.[62] These include proximity to abundant water resources, such as rivers, lakes, or the ocean, the distance

---

[59] The licenses are valid for 15 years and can be renewed for another 10-15 years. The combined license is valid for 40 years with the option to renew for an additional 20 years Duke Energy (2012).
[60] As of 2019, there were 93 nuclear reactors that could commercially operate. See https://www.nrc.gov/ reading-rm/doc-collections/nuregs/contract/cr2907/v25/index.html.
[61] However, Komanoff (1981) argues that regulatory costs of reactor construction had started to rise in the U.S. prior to TMI.
[62] See https://www.nrc.gov/reading-rm/doc-collections/cfr/part100/full-text.html.

between the site and the population centers, and the availability of transportation and space for NPP waste. The area's topography and the soil's bearing capacity are considered because of the substantial weight and size of the equipment used in NPP construction. The geology and the seismology of the region are additional critical considerations for the site selection of NPPs. For example, evaluation of the site for earthquake recurrence rates, fault geometry, slip rates, landslides, seismically induced floods, and water waves is necessary, as these could affect the nuclear plant's safety.

We tested whether the NRC's site selection rules matter in practice by estimating the probability that a county is ever selected (in other words, considered) as an NPP site as a function of variables that mimic the NRC's regulation criteria. Specifically, we estimate the equation depicted below:

$$(2.1) \qquad P(Considered_c = 1) = f(X_c\beta + u_c),$$

where the unit of observation is a county, indexed by c. The outcome variable *Considered* takes the value of one if an NPP's construction is announced in the county. For all 132 counties where an NPP's construction is publicly declared (regardless of whether the plant is eventually built), *Considered* equals one.[63] For the remaining 2,920 counties in the nation, it equals zero. The vector *X* comprises proxies of the NRC criteria in Title 10. Specifically, they are the water-to-land ratio in the county, the average precipitation and temperature, the distance of the county to the nearest major earthquake (5.5 or greater on the Richter scale), population density, the highest point in the county in feet, and the distance of the county to the nearest highly populated city (defined as one of the top 100 most populated cities in 1960) in the country as well as interactions of these variables.

We estimate Equation (2.1) with a logit regression. The marginal effects of the variables are presented in column 1 in Panel A of Table 7. The coefficient of the water-to-land ratio in the county is positive and statistically significant, indicating that counties with relatively more water bodies are more likely to be considered for NPP construction. In line with the NRC's regulations, counties considered for an NPP are less dense population-wise, located away from high population centers and seismically active areas compared to other counties. The estimate of the elevation of the

[63] We explain in the next section that plans to build an NPP can be canceled, and consequently, a considered county may end up with no NPP.

highest point in the county is negative, suggesting that NPP sites are chosen to be in less mountainous regions. The point estimates of the average temperature and precipitation have the expected signs, though they are statistically insignificant. These results in column 1 in Panel A of Table 7 imply that NRC's regulations are followed when choosing NPP construction sites. In other words, the counties selected as potential NPP sites have significantly different geographic characteristics than those in the rest of the country.

We also investigate whether there are differences in other pre-existing attributes between the counties deemed favorable for NPP construction versus other counties. Specifically, we tested whether the considered counties' and others' socio-economic characteristics in the 1950 and 1960 censuses were different using a propensity score matching method (PSM), in which the propensity scores are estimated by Equation (2.1). The distributions of the propensity scores of the considered and never-considered counties are given in Appendix Figure A.5. There is a significant overlap in the propensity scores of the considered counties and others. We matched each considered county with its nearest neighbor in terms of propensity scores. We found a control county within a caliper of 0.01 in propensity scores for each considered county. Appendix Figure A.6 displays the differences in covariates in the matched versus unmatched samples. Matching improved the covariate balance.

The results obtained from this PSM analysis are presented in column 1 of Panel B in Table 7. Each row is an estimate from a separate regression. For example, the row "Percentage of Blacks (1950-1960)" shows the PSM estimate of the difference in this outcome between the considered counties versus the others. Specifically, it indicates that there were about 0.7 percentage points more blacks in the selected counties compared to not-considered counties, which is statistically insignificant at the conventional levels. Similarly, there are no significant differences between the considered counties and their counterparts in the shares of older people and the share of owner-occupied houses. However, there are notable differences. For example, the residents in the selected counties are more educated and are more likely to be foreign-born than other counties. The employment-to-population ratio and the average income are also higher. Housing prices in the considered counties are higher than never-considered counties. Taken together, the results in Table 7 suggest that the counties selected for NPP construction have significantly different characteristics than those of other locations.

2.2.2 Plans versus Reality of NPP Construction

When a utility company applies for a license to construct an NPP in a county, it publicly announces its plans early in the planning process. These announcements are typically verbal commitments in press releases or annual reports and do not constitute legal obligations. The utility company can cancel its NPP construction plans anytime. For example, the Pacific Gas and Electric company announced the construction of the Bodega Bay nuclear reactor in Sonoma County in California in 1958, but the project was canceled in 1964, before the start of construction (Walker, 1990). During the 1970s and 1980s, as demand growth slowed, regulatory requirements tightened, and public opposition intensified, a flurry of NPP projects were canceled or suspended. Wisconsin's Tyrone 2 was scrapped in 1974 because of insufficient demand growth and local resistance. Virginia Electric & Power (VEPCO) shelved Surry 3 and 4 in 1977 after significant drops in power demand, then abandoned North Anna 4 in 1980 with the project just 4% complete. Several other initiatives met similar fates around that time, including Sterling, Jamesport 1 and 2, and Forked River 1, each undone by mounting expenses, licensing uncertainty, and community protests. Throughout the 1980s, the pattern persisted: Callaway 2, Hope Creek 2, Bailly 1, and Shearon Harris 3 and 4 were all discontinued under escalating financial and regulatory pressures. Multi-reactor endeavors such as Washington Nuclear 4 and 5, Hartsville B1 and B2, Yellow Creek 1 and 2, Marble Hill 1 and 2, and Midland 1.2 had collectively amassed billions in sunk costs before being canceled. Even nearly finished projects, such as Bellefonte 1 and 2, were shelved.[64] In sum, as explained by the US Department of Energy (1983), the primary reasons for canceling planned nuclear units include a significant downward revision in the anticipated future electricity demand, unexpected increases in capital requirements or other construction costs, delays in licensing due to new regulations, and political reasons.

As a result of the cancelation of construction projects, 34 of the 132 counties considered for an NPP never had one. The map in Figure 7 displays the locations of these counties. We analyzed the observable characteristics of the counties where the NPP construction plans are canceled by estimating the regression below:

$$(2.2) \qquad P(Canceled_c = 1) = f(X_c\beta + u_c),$$

[64] A summary of each NPP cancellation can be found here: https://www.powermag.com/interactive-map-abandoned-nuclear-power-projects/.

This equation is identical to Equation (2.1), except for the outcome variable, which is replaced by a dummy that takes the value of one if the county ultimately did not have an NPP due to a cancelation. Other variables are the same as those in Equation (2.1). Note that only the 132 considered counties enter into this regression. If cancellation is not a result of the physical and geographic attributes of the county (proxied by vector *X*), then the estimates of β from Equation (2.2) should be all statistically not different from zero.

The marginal effects from Equation (2.2) are shown in column 2 of Panel A in Table 7. In contrast to column 1, which shows several differences in physical, geological, and atmospheric characteristics between the considered and not-considered counties, column 2 shows that all counties that were once considered for an NPP, regardless of whether the construction plans were quashed or not, are alike in these attributes. We also tested whether the counties whose NPP projects were canceled versus those that ultimately had an NPP constructed differ in socio-economic characteristics using the PSM method described above. The estimates in column 2 in Panel B of Table 7 are much smaller than those in column 1, close to zero in most cases, and statistically insignificant at the conventional levels. These findings suggest that there are no statistically significant differences within the subsample of considered counties, and the probability of cancellation is unrelated to the pre-existing socio-economic characteristics of the considered counties.

2.3 Data

The data used in this paper come from several sources. We obtain the annual county-level economic outcomes, including employment, wages and salaries, and personal income components from the Regional Economic Accounts (REAs) published by the Bureau of Economic Analysis (BEA). These data are available on the BEA website since 1969. We chose 2019 as the last year in our sample period to avoid the influence of the global health crisis in 2020.

We hand-collected the NPP data from several sources, such as the National Technical Information Service, the Nuclear Regulatory Commission, and the Office of Scientific and Technical Information websites. The resulting data set includes the names and locations of the NPPs and the years when their construction was announced, their construction licenses were obtained, their construction plans were canceled (if ever canceled), and their commercial operation licenses were secured.[65]

This information allows us to create our study's two essential variables, the first of which is *Construction*.[66] This variable indicates the years in which an NPP's construction occurs in the county. Specifically, we generated this dummy variable to take the value of one starting with the year when a construction license was granted to an NPP. It continues to equal one in the years until the commercial operation license is procured or the NPP's construction is canceled if the project is ever canceled. The other variable of interest, *Commercial Operation*, is constructed similarly. It indicates the years in which an NPP is operational in the county, and it turns from zero to one in the year when the NPP's commercial license was granted.[67]

To see the construction of these variables more clearly, consider the hypothetical NPPs presented in Table 8. In 1960, NPP A's construction in County X was announced, but the plan was canceled in 1962. We classify this county as a considered but untreated county and accordingly group it among the control counties in the DD strategy or donor counties in our SC strategy. For county X, both *Construction* and *Commercial Operation* dummies always equal zero. NPP B's construction in County Y was declared in 1965, and the plan was implemented in 1967. However, before the NPP could obtain its commercial license, the project was canceled in 1970. Thus, for county Y, the *Construction* indicator takes the value of one only between 1967 and 1970, and the Commercial Operation variable is always zero. In another county, Z, NPP C's construction started in 1973 and continued without cancellation until 1980, when an operation license was granted to the NPP. County Z's *Construction* (*Commercial Operation*) indicator equals one between 1973 and 1980 (in 1980 and after).[68] A county may have multiple nuclear reactors, and their construction

---

[65] Our data set includes information on all but six nuclear reactors. These reactors are Crystal River 4 (Citrus, FL), Perryman 1 and 2 (Baltimore, MD), Ravenswood (Queens, NY), and Somerset 1 and 2 (Bristol, MA). We excluded them from our analysis because we were unable to locate information about their milestone dates.

[66] In a paper that investigates the determinants of cancellation of NPP projects, Berndt and Aldrich (2016) compiled a data set of most of the commercial nuclear reactors in the US. We compared our data to theirs (which are available at the following link: https://www.openicpsr.org/openicpsr/project/100127/version/V1/view) and found that our data set is more comprehensive. For example, Berndt and Aldrich's (2016) data do not include 69 reactors, some of which are purposefully left out of their analysis, such as the "Turn-Key" reactors and others. This list of reactors is available upon request. Berndt and Aldrich (2016) were focused on siting outcomes, whereas we are focused on the economic effects of those siting outcomes.

[67] None of the reactors that are in our main estimation sample (those that started construction between 1974 and 1978) have been decommissioned. Out of the 187 reactors that have ever been constructed 41 had been decommissioned.

[68] Except for only one county, Wake (NC), the Commercial Operation variable turns to one immediately after the Construction variable becomes zero.

and commercial operation periods may overlap. Such an example is county W in Table 8. In this county, where two NPPs were ultimately constructed, one reactor's (D2) construction was still ongoing in 1972 when the other (D1) started operating. D2's construction was completed in 1975. As a result, the *Construction* and *Commercial Operation* variables are equal to one between 1972 and 1975 for this county.

2.4 Empirical Framework

We estimate the impact of the construction and operation of an NPP using a difference-in-differences (DD) framework. Table 7 shows that the geographic and socio-economic differences are large and significant for the counties considered/selected for an NPP relative to the rest of the counties in the nation. At the same time, it shows that the counties considered for an NPP form a homogenous group. Specifically, counties where an NPP was planned but never built have similar physical and socio-economic attributes. Given these results, in our empirical strategy, we defined the counties with a phase of NPP construction as the "treated," and those counties considered for/selected as an NPP site but never got one as the "control." Specifically, we focus on the counties where the first NPP construction started in 1974, 1975, 1976, 1977, or 1978. We exclude counties whose construction started pre-1974 because the cancelation of NPP plans in a county was not a significant risk before 1974.[69] Thus, our estimates can be interpreted as the impact of an NPP in a county relative to another county expecting to have an NPP. We also exclude from our sample the counties where NPP construction started in 1979. Our analysis sample includes 23 and 34 counties in the treatment and control groups, respectively. Later, we show that our baseline results do not change appreciably if we use the complete set of counties with NPPs (regardless of whether they were built before or after 1974).

Table 9 presents the means of the yearly outcome variables in the pre-treatment period (1969-1973) for various groups of counties. Columns 1, 2, and 3 show the means for the treated counties (those in which an NPP was constructed), control counties (those considered for an NPP but never got one), and the rest of the US. The stars indicate whether the average in a sample is statistically different from the one in column 1. We can infer from Table 9 that the control group is similar to

[69] This is likely because of NRC's stricter regulations in the 1970s (see Figure 6). Only one out of 63 nuclear reactor plan cancellations occurred before the 1970s. This nuclear plant was planned to be built in Sonoma county in CA. It was canceled in 1964.

the treated group in terms of employment-to-population ratio and per capita wages in the pre-treatment period. On the contrary, the rest of the counties differ significantly from the treated counties in total personal income and wages.

Our baseline estimation equation is depicted below:

$$(2.3) \qquad Y_{ct} = \beta_1 Construction_{ct} + \beta_2 Commercial\ Operation_{ct} + \mu_c + \tau_t + u_{ct},$$

where the unit of observation is a county-year (ct). Yearly observations from 23 treated counties and all 34 counties in the control group enter the regressions. Thus, the number of observations in these analyses is 2,907 (= 57 counties × 51 years). We consider two outcomes: the employment-to-population ratio (the percentage of the employed in the whole population) and per capita wages and salaries. The variables of interest are the *Construction* and *Commercial Operation* indicators. The former (latter) takes the value of one if a nuclear reactor is being built (producing and selling electricity) in county c in year t (see the Data Section and Table 8). The regressions include county fixed effects (μ) and year dummies (τ). Finally, *u* stands for the error term. The standard errors are clustered at the county level.

2.5 Results

2.5.1 Baseline Estimates

Estimates obtained from Equation (2.3) are presented in Table 10. In column 1, the outcome is the employment-to-population ratio, which measures the percentage of the county population that is employed. Specifically, the employment-to-population ratio is defined as one hundred times the number of employed divided by the total county population. The outcome in column 2 (3) is the natural log (level) of per capita wages and salaries. The coefficient of the *Construction* dummy in column 1 indicates that when a nuclear reactor is being built in a county, the employment-to-population ratio in that county goes up by about eleven percentage points. This corresponds to a 23% increase from the mean employment-to-population ratio in our sample (48.4 percentage points). On the other hand, the estimate of *Commercial Operation* is not statistically different from zero. Column 2 shows that the log of per capita wages in a county goes up during the NPP construction and its commercial operations. The *Construction* (*Commercial Operation*) estimate implies a 47% (25%) increase in per capita wages.[70] The inference from column 3, where the outcome is the level of per capita wages, is similar to that from column 2, except that *Commercial Operation* is statistically insignificant.

The large effect of NPPs may stem from individuals outside the county relocating to the NPP-hosting county for employment opportunities in the NPP project rather than the local population filling these jobs. If this is the case, then we should observe an increase in the total population in the county. However, when we estimate Equation (2.3) with county population (in one thousand people) as the outcome, we find that neither the construction nor operations of NPPs significantly increase it. The coefficients (and standard errors) of the *Construction* and *Commercial Operation* indicators from this regression were 28.2 (26.3) and 61.0 (124.0). These estimates may suggest that NPP projects do not necessarily attract outsiders as permanent residents into the NPP-hosting area.

2.5.2 Robustness Checks

Results presented in Table 10 are obtained from regressions that use counties where an NPP was built between 1974 and 1978 for the first time. We checked whether these results are robust to including the complete set of counties with an NPP. That is, in addition to the counties that enter into Table 10 regressions, in this exercise, we use observations from other counties that had an NPP constructed before 1974. Note that none of the NPPs built between 1974–1978 had been decommissioned by 2019. Thus, the *Commercial Operation* variable in the baseline sample remains one after it turns to one. However, because some NPPs built before 1974 were decommissioned, the *Commercial Operation* dummy contains greater variation, such as 0-1-0. Results from this robustness check are presented in Appendix Table A.5. They show that the construction of an NPP increases the employment-to-population ratio and the per capita wages in the county. However, the operation of an NPP does not have a statistically significant impact on these outcomes.

Four counties (Burke, GA, Callaway, MO, Cherokee, SC, and Wake, NC) in our estimation sample experienced two separate construction phases, the second one in the 2010s. In other words, additional NPP construction projects started in these counties after the 1974–1978 period. We

[70] We computed these figures by $100 \times (\exp(\text{Estimate}) - 1)$.

reran our baseline regressions without these counties to test whether our baseline results were sensitive to excluding these counties from the sample. The results in Appendix Table A.6 are virtually identical to those in Table 10.

We also experimented with controlling county-specific time trends in Equation (2.3). Including linear or quadratic time trends did not alter our estimate of the effect of the construction of NPPs on wages and employment (Panel A of Appendix Table A.7). Conversely, when county-specific time trends are included in the regressions, the estimate of *Commercial Operation* becomes statistically insignificant.

The baseline specification does not include any control variables except for the county fixed effects and year dummies. Although earlier analysis suggests that there are no discernable differences in socio-economic predetermined covariates between the treated and control counties, those with certain attributes may take on differential paths that coincide with the NPPs. To investigate whether our baseline estimates are subject to such confounding due to covariates, we estimate equation (2.3) with time-varying demographic county attributes and the interactions of the time-invariant predetermined (1960 values) county characteristics with the time variable.[71] The estimates are in Panel B of Appendix Table A.7. They are similar to but slightly smaller than our baseline estimates. For example, the point estimates of the *Construction* variable in the employment rate and log wage regressions are 7.6 and 0.29, while they were 11.4 and 0.38 in Table 10. However, these estimates are within a one-standard deviation window of each other. The coefficients of the *Commercial Operation* variable are statistically not different from zero.[72]

State-specific policies and local public/political pressure may bias our results by creating unobserved differences in how NPPs are regulated, constructed, and operated across regions. Without controlling these factors, estimated economic effects may reflect policy-driven variations rather than the true impact of NPPs. To guard against this possibility, we estimated equation (2.3), including state-by-year fixed effects on the right-hand side. Results, shown in Panel C of Appendix Table A.7, are similar. The point estimates are slightly larger than those in Table 10.

[71] The time-varying covariates are the proportions of Blacks, women, and individuals 65 and older in the county. The time-invariant attributes are the proportions of foreign-born individuals and the population with at least a high school degree, the median rental price, the share of the houses occupied by the owners, and the median family income in the county.

[72] To alleviate the concern that the NPP project cancelations are due to exits of major manufacturing sites in the local area, changes in the expected number of residents moving in/out of the area, and the operation of other energy sources, we re-estimated equation (2.3) controlling for proxies of these potential confounders. The results shown in Panel D of Appendix Table A.7 are almost the same as those in Table 10.

Large-scale, significant accidents such as in TMI, Chernobyl, and Fukushima may influence the public's perceptions regarding nuclear power (see Appendix Table A.4). Consequently, the nuclear power authorities and energy companies may slow down the construction and commercial operations of the NPPs. This may introduce a bias to the estimates. To avoid such bias, we estimated Equation (2.3), excluding observations post-Fukushima (2011), post-Chernobyl (1986), and post-TMI (1979), separately. For example, in the first set of regressions, we only include observations between 1969 and 2010. The results, which are presented in Appendix Table A.8, are similar to those presented earlier in that the *Construction* variable is always positive and significant in the employment and wage regressions, and the *Commercial Operation* dummy is positive but not always statistically significant at the conventional levels.[73]

NPP intervention sizes may vary as some projects contain more than one reactor.[74] For example, more than one reactor was built in 20 of the 23 treated counties in our sample.[75] We investigated whether constructing and operating one versus multiple reactors in an NPP has differential impacts by estimating a modified version of Equation (2.3), where the *Construction* and *Commercial Operation* variables are replaced with dummy variables that indicate the construction and operation of the first, second, and third or further reactors.[76] The estimates, presented in Appendix Table A.9, show that the construction of the 1st reactor in a county increases the employment rate by about six percentage points (12% from the mean) and per capita wages by about 30%. The point estimates of the construction of second and third reactors are also positive and large but not statistically significant. On the contrary, the commercial operations of the first and second reactors have virtually zero effect on the employment-to-population ratio and wages. However, operational three or more reactors appear to increase both employment and wages. Given the magnitudes and standard errors of the coefficients, it is hard to comment on the effects of NPP size with certainty. This analysis points to labor market benefits from the construction of the first reactor. The construction of large NPPs with several reactors could elevate employment

[73] None of the NPPs in our sample, i.e., those whose construction started between 1974 and 1978, began commercially operating before 1979. Therefore, we do not include the *Commercial Operation* dummy in the 1969-1979 regressions.
[74] The number of reactors in an NPP is also positively correlated with the energy capacity of the plant. For a sub-sample of the NPPs, we have information on the electricity generation capacities. We regressed these capacities on the number of reactors in the NPP. According to the results of this regression, each reactor increases NPP's energy generation capacity by about 750 MWs.
[75] In 3 counties, there is only one reactor. In 15 counties, there are 2 reactors. In 4 counties, the number of reactors is 4, and only in one county, there are 5 reactors.
[76] In 16 counties, the construction of two reactors was initiated simultaneously. In such cases, we designate the reactor that completes construction first as the "first reactor."

and wages as well. However, the operations of only large NPPs with three or more nuclear reactors appear to improve the labor market conditions.

Next, we conduct a falsification test in which the outcomes of the control group counties are compared to other counties that have never been considered for an NPP but had characteristics similar to the treated counties. We implement this test in two stages. First, using the propensity scores obtained from Equation (2.1) (see Table 7), we find never-treated counties from the rest of the nation that are closest in propensity scores to each treated county. Second, assuming the never-treated counties we identified in the first stage were treated the same way as the actually treated county, we estimate our baseline regression in Equation (2.3).[77] We expect null effects because we are comparing the outcomes of counties that were not treated to the control group. The results from this analysis, presented in Appendix Table A.10, verify this. The coefficients of the *Construction* and *Commercial Operation* variables are statistically insignificant in all regressions.

Our identification relies on the exogeneity of the cancelations of NPP projects before their construction begin. Some reasons for cancellations listed in the US Department of Energy (1983), such as unexpected increases in capital costs, new regulations, and political reasons, are unlikely to be related to local labor market conditions. However, a critic may argue that our baseline results are observed as an artifact of cancelations due to a downward revision in the anticipated future electricity demand in the area. Specifically, suppose that an NPP was built in a county around which an increase in energy demand was expected, and other counties did not get an NPP because no significant increase in energy usage was forecasted. Also, suppose that these forecasts were realized in the future, such that treated counties required more energy due to, say, industrial growth in the region, causing an increase in employment and wages. In control counties, no growth in employment and wages was realized, as predicted. In this scenario, our variables of interest would overestimate the true effects of NPPs.

To address this concern, we studied the history of cancellations due to energy usage forecasts. According to the US Department of Energy (1983), between 1966 and 1972, the anticipation was a consistent annual energy usage growth of around 7 percent or higher in regional summer peaks.

[77] To be more clear, suppose the 34 control group counties are labeled X1, X2,..., X34 and the 23 treated counties as Z1, Z2,..., Z23. For each treated county, Z1-Z23, we find a county from the rest of the US (i.e., not including X's and Z's) whose propensity score is closest to the treated county. Let's label these counties Y1, Y2,..., Y23. We assume Y1, Y2,..., and Y23 are treated the same way (i.e., the construction and commercial operation start dates are the same) as Z1, Z2,..., and Z23, respectively. Then, in the second step, we estimate our models using the Z counties and X counties.

This prompted utility companies to strategize expansive construction initiatives to nearly double the installed generating capacity within a decade. In 1973, the Arab oil embargo led to significant spikes in fossil fuel costs for utilities. Consequently, electricity prices surged, contributing to a widespread economic downturn nationwide. As a result, there were noticeable reductions in the immediate growth of electricity demand and, consequently, revisions in the long-term projections for load growth. Thus, it appears that the downturn in energy demand was due to global macroeconomic conditions, not because of local economic activity.

Nevertheless, we obtained information about the reasons for the cancellation of each NPP project in our sample. 22 out of 62 canceled nuclear reactor projects were called off at least partly because of the revision of the energy demand forecast. To investigate whether our baseline results are obtained due to the inclusion of these counties, we implemented our estimation strategy without the control counties where an NPP project was canceled due to the energy demand forecast. That is, in this exercise, our control group in the DD framework only consisted of the counties where the NPP was canceled due to reasons other than future energy demand forecasts. As a result, the number of counties in the control group was 15 (as opposed to 34 in the baseline). The estimates obtained from this robustness analysis are presented in the top panel of Appendix Table A.11, which are not appreciably different from the baseline results. This indicates that our main results are free from a bias caused by the future energy demand.[78]

In a similar fashion, we ran our regressions excluding counties whose NPP projects were canceled due to changes in the anticipated construction costs. This exclusion left us with fifteen control counties. The results in the second panel of Appendix Table A.11 are not much different from those in Table 10. For example, the coefficient of the *Construction* dummy is about 11 (8.3) in the employment-to-population ratio (per capita wages and salaries) regressions. The estimate of the *Commercial Operation* indicator is not statistically significant in these regressions. We also experimented with excluding the control counties whose NPP plans were scrapped due to either revisions in the energy demand forecast or construction cost updates. Regressions using the remaining ten control counties and the treated counties yielded similar effects described in Table 10. See the third panel in Appendix Table A.11.

[78] An analysis of the control counties suggests that the employment and wage outcomes of counties whose NPP projects were canceled due to energy demand forecast revisions followed a path similar to that of other control counties.

Although these robustness analyses show that our estimates are not sensitive to the group of control counties or the period considered, one may be concerned that the cancelation of plans was realized due to county counties' unobserved characteristics correlated with our outcomes. To mitigate this concern, we estimated Equation (2.3) excluding *all* control counties from the sample. That is, in this regression, only the treated counties where an NPP's siting was not canceled enter. Note that the identifying variation in this setting comes only from the timing and length of the construction. This specification again results in similar, albeit slightly smaller, estimates compared to Table 10. For example, the bottom panel of Appendix Table A.11 shows that the coefficient of *Construction* is 7.7 (with a standard error of 3.9) in the employment-to-population ratio regression. The corresponding estimate of the *Commercial Operation* dummy is 5.9 (7.8). Similarly, *Construction*'s estimate is 6.9 (3.6) for the per capita wages and salaries outcome, while that of *Commercial Operation* is 3.1 (4.1).

2.5.3 Parallel Trends

The identification of causal effects with our empirical methods hinges on the "parallel trends" condition, which requires the treated and the control counties to move in the same direction prior to the implementation of the treatment.[79] This condition is typically evaluated with an event study analysis, where time dummies relative to the start of the event are included in the specification. In our case, there are two separate events: construction and commercial operations. We implemented event study analyses by combining them into one (*Treat* = *Construction* + *Commercial Operation*) and estimating the equation below:

$$(2.4)\quad Y_{ct} = PreTreat\ Period_{ct}\Gamma_0 + PostTreat\ Period_{ct}\Gamma_1 + \mu_c + \tau_t + u_{ct}$$

where *PreTreat Period* and *PostTreat Period* vectors contain dummies that indicate the year relative to the start of the construction period. The *PreTreat Period* vector contains dummies, such as *Two Years Before, Three Years Before*, etc., to indicate the years two, three, and so on, years before Treat turns from zero to one. Similarly, the *PostTreat Period* includes indicators for *One Year After Construction*, *Two Years After*, and so on, for the first, second, and further years after the *Treat* variable turns one. Thus, the first few post indicators correspond to the

[79] An analysis of the raw means of the outcomes suggests that this is the case (Appendix Figure A.7)

construction period. In our sample, construction periods range between 8 and 16 years, with an average of 11 years. The comparison category is one year before construction started. Recall that no NPPs in our sample have been decommissioned during the period we consider.

The results of the event study analyses are presented in Figures 8 and 9, which illustrate the point estimates and 95% confidence bands for the employment-to-population ratio and per capita wage outcomes. -5 and 20 in these figures stand for “5 years before or earlier” and “20 years after or later.” Zero (0) indicates the first year of construction. The plot points and capped vertical lines are the point estimates and 95% confidence bands. Both figures show similar patterns. The differences in employment and wages between the treated and control counties are statistically insignificant before an NPP’s construction starts, and employment and wages in the treated counties increase significantly during the first few years of NPP construction. Although they remain positive, the point estimates diminish and become statistically insignificant over time, most visibly about 15 years after the commencement of an NPP construction. This span roughly corresponds roughly to the average construction period in our sample (11 years). These findings suggest that the treatment and control groups followed statistically indistinguishable trends before the start of construction, suggesting that they may constitute the counterfactual for one another. The analysis also indicates that the positive employment and wage effects of NPPs are most noticeable in the first handful of years of NPP construction and then disappear over time.[80]

2.5.4 Two-Way Fixed Effects and Effect Heterogeneity

A line of research has pointed out that the two-way fixed effects (TWFE) models could be problematic when treatment effects are heterogeneous across cohorts or over time (for example, see Goodman-Bacon, 2021; Athey and Imbens, 2022; and Roth and Sant’Anna, 2023). Since we have used TWFE models as our baseline estimation strategy, in this section, we investigate whether using alternative estimators that address the limitations of the TWFE model alters our conclusions.

[80] Estimating the event studies additionally controlling for covariates (as in Panel B of Appendix Table A.7) leads to similar inference. The results are available if requested.

Sun and Abraham (2021) show that with treatment effect heterogeneity, the coefficients on leads and lags of treatment in traditional TWFE event studies can be contaminated by effects from other periods, leading to misleading estimates of dynamic treatment effects. They propose an estimator that eliminates such biases. The results obtained from this method are presented in Figures 10 and 11, where the time axis (vertical axis) shows years relative to the start of the construction. The pattern of estimates is almost identical to TWFE event-study analyses in Figures 8 and 9. There are no pre-trends in either employment or wage outcomes in the treated counties before the construction of the NPPs. The impact of NPPs peaks in about 5 years, after which it starts to dissipate. Figure 10 suggests no employment effects after about 15 years from the initiation of NPP construction. Figure 11 shows that the impact of NPPs persists after two decades, albeit at a lower rate than the peak impact.

A recent paper by Chaisemartin and D'Haultfœuille (2023) studies TWFE models in multiple-treatment cases, similar to our context, where counties are treated with two treatments: NPP construction and commercial operation. They show that the TWFE estimates of one treatment could additionally be contaminated by the effects of other treatments included in the regression due to the presence of non-zero contamination weights and potential heterogeneity in the other treatments' effects. Their proposed alternative estimator isolates the effect of a given treatment by leveraging comparisons between groups whose treatment status changes while keeping other treatments constant. In our study, this amounts to, for example, comparing counties that initiate an NPP construction but not commercial operations to similar counties where neither treatment status changes, ensuring that the impact of operations does not confound the estimated effect of the construction.

The results obtained from the Chaisemartin and D'Haultfœuille (2023) estimator are depicted in Appendix Figures A.8 and A.9 for the employment-to-population outcome and A.10 and A.11 for the log per capita wages outcome. The event considered in Appendix Figures A.8 and A.10 is the construction of an NPP. The effect of NPP construction on employment and wages is highest in the first few years following the commencement of the project and then declines over the years. The impact is negligibly small after about 14-15 years. The cumulative effect of NPP construction on employment (log wages) is 11.73 (0.49), with a standard error of 6.84 (0.22) in 14 years. These estimates are remarkably similar to our baseline coefficients. In contrast, the picture for commercial operations is completely different. Appendix Figures A.9 and A.11 show that an operational NPP reduces the employment-to-population ratio by about ten percentage points and the per capita wages by about 25%. These results suggest that the labor market prospects in a county improve during the construction of an NPP. However, a commercially operational NPP does not necessarily have a positive impact.

2.5.5 Alternative Synthetic Control Strategy

We additionally implement a synthetic control (SC) strategy, which may be a good fit for our setting for many reasons, as described in (Abadie, 2021). First, SC is appropriate for analyzing outcomes from aggregate entities, such as counties. Second, the abrupt alteration in construction and operations triggered by TMI, together with the temporal variation in construction and operation phases, furnishes a rich treatment structure well-suited for the SC methodology. Also, in most cases, an NPP likely represents a "large intervention" relative to the size of the local economy, in which SC is more likely to detect an effect. Finally, because the treatment intensity, i.e., the magnitude of construction and operations, may vary greatly across counties, the best counterfactual counties may differ across treated counties. Thus, a case study-like comparison could be appropriate.

We implemented the SC method, as described in Abadie et al. (2010), separately for each of the 23 treated counties whose NPP constructions commenced between 1974 and 1978. The details of our SC strategy, several robustness checks, and the findings are discussed in detail in the self-contained Appendix Section B. The results from this analysis can be visualized for each county in figures like Appendix Figures B.1-B.4. Appendix Figures B.1 and B.2 show the results for Burke County, GA, where the construction of two reactors took place between 1974 and 1987. They were both operational afterward. Decades later, in 2012, construction of an additional reactor was initiated in Burke. Appendix Figure B.3 illustrates the results from Tishomingo County, MS, where the Tennessee Valley Authority started an NPP construction in 1978. However, the project was canceled in 1984 due to escalated construction costs and the decrease in projected energy demand.[81] Appendix Figure B.4 presents the results from Cherokee County, SC, whose NPP project commenced in 1977 but was scrapped in 1983 due to economic

[81] See the NY Times coverage in 1984 at this link: https://www.nytimes.com/1984/08/30/business/tva-cancels-4-reactors-knoxville-tenn-aug-29.html).

and regulatory challenges.[82] In all of these examples, the labor market prospects in these counties relative to their synthetic controls improve at varying degrees during the construction period.[83] There is evidence of some improvement in employment and wages during commercial operations, but it is smaller.

We aggregate and summarize the results from the SC analyses of 23 treated counties in Appendix Table B.4. They are almost identical to our baseline results. For example, during the construction of NPP in a county, the employment-to-population ratio in that county increases by about 10.2 percentage points. The commercial operations of an NPP do not impact employment significantly. We also find that per capita wages and salaries go up both during the construction and operation of an NPP. However, the effect of commercial operations on wages is not robust. In summary, consistent with our DD estimates, the results from the SC strategy indicate labor market benefits of the construction, but not the commercial operation, of an NPP.

2.5.6 Extensions

Impact of NPPs on Total Income and Income Components Other Than Wages:

While the largest component of personal income is wages and salaries, individuals may earn from other sources, such as dividends and interest or the government's transfers. According to the Bureau of Economic Analysis (BEA), total personal income is the sum of wages, supplements to wages, proprietors' income, dividends, interest, and rental income, and personal transfers from the government minus contribution for government social insurance.[84] Since NPPs increase wages, they probably affect some of these other income sources. For example, individuals with relatively low wages are entitled to payments from the welfare programs. If wages and salaries increase due to NPP construction, then locals' earnings from welfare programs may decrease.

---

[82] A second effort to re-initiate the construction of an NPP in Cherokee County also failed in 2017. See the local TV coverage at this link: https://www.youtube.com/watch?v=1kQv6jLIAaE.

[83] In the canceled reactors, other endeavors took advantage of the infrastructure built. For example, portions of Abyss, a 1989 James Cameron movie that won an Academy Award in 1990, were filmed in the reactor containment building of the NPP in Cherokee, SC (https://www.telegraph.co.uk/films/0/abyss-how-james-camerons-underwater-thriller-almost-killed). Similarly, in the late 1980s, NASA began constructing an establishment to build rocket motors in close proximity to the canceled NPP site in Tishomingo, MS. This initiative was dead-ended as Congress stopped funding in 1993 (https://www.nytimes.com/1993/11/29/us/hit-twice-town-feels-misled-by-us.html). Our analyses show that neither of these projects impacted employment in these counties.

[84] BEA also makes a residential adjustment to account for income earned by those who live out of the county.

To investigate the impact of NPPs on total personal income and its components, we re-estimated Equation (2.3) with these outcomes. Results are presented in Table 11. Components of total personal income are the outcomes in columns 1-6, and the results for the total per capita personal income are given in column 7. In column 1, for presentation purposes, we reprint the per capita wage results. The outcome in column 2 is the per capita supplements to wages. These include benefits of employment typically paid by employers (such as health insurance, retirement contributions, and so on). Estimates in column 2 show that both the construction and operation of NPPs increase these supplements to wages. Results in columns 3 and 4 reveal that NPPs have no impact on proprietors' income, dividends, interest, or rental income in the county. As we predicted, individuals' income from governmental transfers (such as welfare programs) decreases with NPPs (column 5). Column 6 shows that NPP construction increases the contributions of the firms to social insurance programs. This is likely because firms' contributions to programs such as the Social Security fund are proportional to wages. Finally, in column 7, the outcome is total per capita personal income. This column indicates that the construction and operation of NPPs appear to increase total income by a small amount. The effect size is about 3.5% for construction and 8.5% for operations.

Impact of NPPs on Employment and Earnings by Industry:

The Bureau of Economic Analysis (BEA) provides data on employment and total earnings for each industry in each county. Using these data, we explore whether there are sectoral differences in the effect of NPPs' construction and commercial operation. Note that BEA shifted to the North American Industry Classification System (NAICS) from the Standard Industrial Classification (SIC) industry codes in 2000. To avoid biases that may arise because of this change, our analysis in this section focuses on the period spanning from 1969 to 2000. In addition, when some sectors are too small in a county, the BEA does not provide employment and income statistics for those industries. Therefore, such counties are omitted from the regressions in this section.

Results obtained from Equation (2.3) are presented in Tables 12 and 13. The outcome in Table 12 is the employment-to-population ratio in various industries, measuring the percentage of the county population employed in agricultural services, forestry, and fishing; manufacturing; mining; construction; transportation; public utilities; wholesale and retail trade; finance, insurance, real estate; and government enterprises. Table 13 showcases the results from the per capita total income outcome in these industries. Not surprisingly, estimates in these tables show that the employment-to-population ratio and earnings in the construction sector exhibit a large increase in a county during the construction of an NPP. The construction of an NPP does not influence employment or earnings in other sectors.[85] Consistent with our earlier results, Tables 12 and 13 show that when the NPP starts producing and selling energy, employment and earnings in the construction industry decrease. The commercial operation of NPPs significantly increases employment and earnings in the public utilities sector. Wholesale trade also appears to benefit from commercially operating NPPs, albeit to a smaller extent. Although statistically insignificant, the coefficients are large for the employment and earnings in the government sector, suggesting that this sector could also benefit from NPPs.

To study the industry-specific effects of NPPs in more detail, we estimate event study specifications (equation 2.4) using employment and earnings by industry as the outcomes. The results from the construction, public utilities, and government sectors are given in Figure 12 (employment rate) and Figure 13 (per capita earnings). The first graphs in these figures show that employment and earnings in the construction sector increase significantly during the typical construction phase of an NPP, with these effects gradually fading around 15 years after the project's initiation. The labor market outcomes in the public utilities sector remain largely unchanged in the early years following the start of an NPP project. However, both employment and earnings in public utilities begin to rise approximately a decade later, aligning with the transition from construction to commercial operation. Additionally, our analysis reveals a notable surge in government sector employment and earnings during the construction phase of NPPs. We did not detect any meaningful effects of NPPs on other industries.

Taken together, these findings suggest that NPPs positively impact local labor market outcomes, primarily through employment and earnings increases in construction and public utilities and, to an extent, government sectors. The most significant labor market effects occur during the construction phase, with sharp increases in employment and earnings in the construction sector. Once the NPP becomes operational (on average, 11 years in our sample), these effects

[85] In a separate analysis, using data from the County Business Patterns, we find that the number of manufacturing establishments is not impacted by NPPs. The coefficient (and standard error) of the Construction variable was -10.81 (54.69). The estimate for Commercial Operation was 127.95 (104.19).

disappear.[86] However, immediately after employment and earnings rise in the public utilities sector, reflecting the long-term workforce needs for plant operations and utility services. The overall evidence suggests that NPPs drive labor market improvements through direct employment effects rather than stimulating broader business expansion in other industries.

Spillover Impacts of NPPs to Neighboring Areas:

As explained above, due to the NRC regulations, NPPs must be constructed in remote, population-sparse areas. At the same time, electricity is hard to store, and some electricity is wasted when delivered over long distances.[87] Given these constraints and the profit motivation, firms are likely to locate their NPPs in regions close to places where there is or will be high energy demand. Thus, it is plausible that neighboring regions may benefit from the operations of NPPs. Moreover, it is similarly plausible that some workers choose to reside in adjacent counties and commute to their NPP jobs. In this section, we test whether the construction and operation of NPPs have spillover impacts on the labor markets of the neighboring counties.

To implement this analysis, we compare the economic outcomes of the most populated neighboring counties of the treated counties (those that have an NPP construction) to the outcomes of the most populated neighboring counties of the controls (those considered for an NPP at one point, but did not get one because of cancellations). That is, using data from these most populated neighboring counties, we estimate regressions like Equation (2.3), where the treatment variables *Construction* and *Commercial Operation* indicate whether an NPP is being built in the next-door county and whether an NPP is operational in the neighboring county.

The results, presented in Table 14, show that the point estimates of *Construction* and *Commercial Operation* are small and statistically insignificant for employment-to-population ratio and per capita wages and salaries outcomes. This means that the employment and wages of the big neighbors of the counties with an NPP, which are presumably targeted by the energy companies, did not grow faster or slower than other comparable big counties in employment and wages. This

[86] Typically, NPPs worldwide are constructed in 6 to 8 years on average (Moreira, Gallinaro, and Carajislecov 2017; Carajislecov, 2011). However, in our sample, the construction periods ranged between 8 and 16 years, with a mean of 11 years, largely due to regulatory changes and heightened public opposition following the Three Mile Island incident.
[87] When choosing a site for a nuclear power plant, utilities balanced the cost of closer, safeguarded sites near populated areas against the expense of high-voltage lines for remote, rural locations (Berndt and Aldrich, 2016).

finding indicates that there is no apparent spillover impact or positive externalities of NPPs on neighboring counties.

If an NPP influences only the county in which it is built but not the surrounding counties, then its impact on employment and wages in the extended labor markets, i.e., commuting zones (CZ), should be smaller than baseline estimates. This is because, under this hypothesis, the impact of the NPP in the county where it is constructed would be diluted within the whole CZ. On the other hand, if NPPs affect the labor market in other counties in the same CZ as much as the NPP-hosting county or more, then estimates from the CZ analysis would be similar to or greater than the baseline estimates obtained from the county where the NPP is constructed.

To test this hypothesis, we estimate a version of Equation (2.3), modified in such a way that the unit of observation is a CZ-year. The wage and employment outcomes are the population-to-employment ratio and per capita wages and salaries in a CZ (computed as the population-weighted outcomes of the counties in a CZ). The treatment variables *Construction* and *Commercial Operation* take the value of one if an NPP is being constructed and operational in any county in the CZ.[88] The results in Appendix Table A.12 show smaller effects than the baseline estimates reported in Table 10 of NPPs on employment and wages in the CZ. The upshot is that the construction of an NPP in a CZ may increase the employment-to-population ratio and wages in that CZ only by a small amount, and there is no effect from the commercial operations to speak of. These results reinforce the idea that NPPs have no significant spillover effects in neighboring areas.

Impact of NPPs on Local Government Finances:

Our examination of the sectoral impacts of NPPs revealed that NPPs improve earnings and employment in the government sector (Tables 12 and 13, and Figures 12 and 13). This could be because of the improvement of the financial positions of the local governments through taxes and other revenues, as NPPs elevate taxable wages and incomes in a county. The additional

[88] We used the 1990 definitions of CZs, as described in (Autor and Dorn, 2013). There are 741 CZs in the US. 22 of them experienced an NPP construction between 1974 and 1978. The control group/donor pool contains 22 other CZs. We implemented the same SC strategy as our county analysis.

government revenues could also be used to provide public goods for the residents of the NPP-hosting counties.

To investigate these possibilities, as a final extension analysis, we estimate the impact of NPPs on local government finances using data obtained from the Census of Governments and the Annual Survey of State and Local Government Finance (CoG), which was compiled and described by Pierson, Hand, and Thompson (2015).[89] The censuses are conducted every five years, in those that end with 2 or 7. In the off-census years, a sample of government organizations is surveyed. The data set is an unbalanced annual panel of all governmental units. These units include general-purpose governments, such as county governments, municipalities, and townships, and special-function units, like school districts, fire protection authorities, and housing development agencies.

One complexity associated with this investigation is the irregular organization of local governments across the country, as their functions and responsibilities vary from one location to another. For example, it is possible that in one city, the municipal government provides fire protection, while in another city, the municipal government does not offer any fire protection because the county government or a fire protection authority provides that service. Thus, the funding and expenses of a specific government could vary significantly relative to others.

To overcome this challenge, we run regressions similar to equation (2.3) controlling for government organization fixed effects instead of county fixed effects as depicted below:

$$(2.5) \qquad Y_{ict} = \beta_1 Construction_{ct} + \beta_2 Commercial\ Operation_{ct} + \mu_i + \tau_t + u_{ict},$$

where the unit of observation is a government organization-year.[90] Y_{ict} stands for expenditures or revenues of governmental unit *i* located in county *c* in year *t*. For example, the municipality of the City of X, located in County A, is a separate governmental unit from the county government of County A and the X City school district. The *Construction* and *Commercial Operation* dummies take the value of one if an NPP is being constructed and operational in the county. To study the within-governmental organization effects of NPPs, we control for governmental unit fixed effects (μ_i). Year dummies are denoted by τ. We cluster the standard errors at the county level. In the

[89] The source of these data is: Urban Institute. 2019. State and Local Finance Initiative Data Query System (SLF-DQS). Accessible from https://slfdqs.taxpolicycenter.org/index.cfm.
[90] We did not aggregate the spending and revenues of all the governmental organizations to the county level because the data set is an unbalanced panel.

regressions, we include only the observations of government units serving the counties selected for an NPP.

The results are shown in Table 15. In columns 1.3 of Panel A, the outcomes are the log of total revenues. In column 1, we pool together all types of governmental organizations, i.e., both general purpose (county, municipal, and township) and special function (schools districts and alike) government organizations enter the regression. The coefficient of the *Construction* variable is 0.14, indicating that the total revenue of a government unit in a county goes up by about 15% while an NPP is being built in that county. Columns 2 and 3 suggest that general-purpose rather than special-function governments mainly drive this increase. In column 4, the outcome is the log of the per capita revenues. That is, we divide the revenues by the number of residents served by the governmental unit. Only county, municipal, and township governments can be included in this regression because the population served by the special function units was not reported in the data source. The estimates reveal that per-person revenues increase during an NPP's construction period. The coefficient of *Commercial Operation* is positive and significant in the total revenue per capita regression, suggesting that an operational NPP improves the financial position of a local government by about 25%. However, an operating NPP does not appear to increase per capita revenues.

Results are very similar when we use expenditures as the outcome (Panel B). The construction of an NPP in one county raises the per capita expenditures of the general-purpose governments in that county by 10-15%. Special-function governmental organizations are not impacted. The effect of commercial operations is similar in magnitude but not estimated with statistical precision.

In Panel C of Table 15, we examine the expenditures by function. We use log per capita spending outcomes in the general-purpose governments. We categorized the expenses in the CoG similar to Jerch, Kahn, and Lin (2023).[91] Results show that NPPs impact neither education nor public safety spending. However, the construction and commercial operation of an NPP increases general administrative expenditures. This increase likely reflects the effort required to manage,

[91] Education expenditures (column 1) include spending on all levels of education. Government administration expenditures (column 2) are made up of spending on local government finance, general public buildings, and central staff services. Public safety spending (column 3) includes expenses related to police, fire protection, correctional facilities, and the judiciary. The public assistance category (column 4) contains spending on public welfare, public housing, hospitals, health, and employment security administration. In column 5, the outcome is the spending on public works. These are expenditures on sewer and water delivery, parking, parks and recreation, and the like. The other expenditure variable in column 6 is the residual.

regulate, and support the operation of a nuclear power plant since NPPs may necessitate higher levels of oversight, infrastructure support, safety measures, and community relations. We also detect a reduction in local government's public assistance expenditures with the commercial operations of NPPs. Finally, our results indicate a 15% increase in public works spending during the construction phase. This could be because the construction of an NPP drives higher demand for local infrastructure and services, requiring increased spending on roads, sewer and water systems, public amenities, and environmental compliance.

2.6 Summary, Discussion, and Conclusion

We compile a comprehensive dataset of all US commercial nuclear reactors that covers significant events, such as public announcements, construction and commercial operation licensing, and shutdowns. These data straddle the regulatory shifts post-Three Mile Island (TMI) and their impact on reactor construction and operations. We observe a clear pattern of decreasing initiation of commercial nuclear construction operations over time. The diverse circumstances of counties with either completed or canceled nuclear power plant (NPP) constructions offer a unique natural experiment to examine the effects of NPPs on local economies.

We use a difference-in-differences framework to assess the impact of NPPs on local economies. Counties where planned NPP projects were scrapped exhibit physical and socio-economic characteristics comparable to those with completed NPPs during the pre-construction period. Our econometric investigation reveals that while the construction of NPPs leads to a significant temporary surge in employment and wages, this effect persists post-construction only to a limited extent. In addition, these benefits are limited to certain sectors, such as the construction and public utilities industries. These results are consistent across various specifications, robustness checks, and alternative estimators, like the synthetic control method and those that eliminate biases that may arise due to treatment effect heterogeneity. The confirmation of parallel trends lends credibility to our causal interpretations.

We find the impact of NPP construction on wages is substantial, about 50%. However, because of the low baseline wages (e.g., the mean is about $17,000), this increase translates into a level change of approximately $8,500 per person in 2019 dollars. Despite the profound impact of NPP construction on wages, their effect on disposable income is modest, e.g., about 4% or $1,200 during the construction period. We also find that NPPs improve the financial position of local governments, with about a 10-15% increase in revenues and expenditures. The increase in government expenditure is mainly due to higher spending on government administration and public works projects that are related to NPP construction.

The lack of a large long-run effect on local labor markets is likely because the labor input required to operate an NPP, approximately 500-700 workers, is small relative to a large manufacturing facility, such as Tesla's Freemont Factory, which employs over 20,000 workers.[92] NPPs also differ from other types of large manufacturing facilities, like the Million Dollar Plants studied by Greenstone et al. (2010), as the uranium and specialized equipment required to generate electricity are generally sourced far away from the plant site.[93]

One may be concerned about the possibility that the short-term benefits of NPP construction we documented do not accrue to the local residents of the NPP-hosting towns, as energy companies could be using labor and capital inputs imported from other regions of the country. While this is probable, our analysis demonstrates that there are no spillover effects of NPPs on neighboring counties, and an economically negligible impact on the extended labor markets at the commuting zone level provides counterevidence. However, even if this is the case, e.g., energy companies use labor imported from completely different areas, our estimates would constitute an upper bound for the local labor market impact of NPPs.

The novel treatment structure, characterized by the staggered timing of NPP construction (and cancellations) across the US, enables us to contribute to the analysis of the economic impact of large-scale energy investments. Consistent with the findings of Fabra et al. (2023), Costa and Veiga (2021), Gonçalves, Rodrigues, and Chagas (2020), Felipe et al. (2017), who show only short-term employment effects of the installation of solar energy panels, wind energy farms, and hydropower plants, our results suggest that NPP construction does not guarantee long-term employment and wage growth for the locals who reside in NPP site counties.

Our findings align with Ando (2015), who identifies heterogeneous income effects of NPPs in Japan, as well as economic impact studies in the UK, France, and South Korea, which report high employment multipliers during construction but varying operational effects depending on national

[92] See https://www.latimes.com/business/story/2024-01-11/tesla-boosts-pay-for-fremont-factory-workers- that-the-uaw-wants-to-unionize and https://www.energy.gov/ne/articles/advantages-and-challenges- nuclear-energy.

[93] Employment levels of these establishments are much larger on average than NPPs, as Greenstone et al. (2010) show that the average manufacturing plant required about 3 million hours of labor input per year. See Table 13 of their paper.

policies and labor market conditions.[94] While NPP job creation appears to follow a general pattern, the persistence and scale of these effects appear to depend on country-specific labor policies, fiscal incentives, and the integration of nuclear power into the broader economy. Understanding the variation in these factors is crucial for designing effective labor policies and maximizing the economic benefits of nuclear investments.

The non-persistent nature of NPP-induced employment and wage effects have important implications for the future of US nuclear power policy.[95] Emerging small modular reactors (SMRs) may reshape some of these dynamics. Because SMRs rely on modular, off-site manufacturing, the on-site construction workforce can be smaller and construction timelines shortened, potentially diminishing the scale of local job booms associated with large conventional NPP projects. However, SMRs could spur a different pattern of employment gains, shifting demand from large onsite construction crews to specialized roles in reactor design, engineering, and manufacturing within a more geographically dispersed supply chain. Further, SMRs' ability to add capacity incrementally, along with built-in safety features that reduce operational staffing needs, may present novel trade-offs for local economic development. While these innovations might lessen the long-run, place-based employment footprint, they also open the door for new high-tech industries and potentially more stable, albeit redistributed, job opportunities. Future research should examine whether SMRs alter the local economic benefits and risks observed with conventional nuclear plants, particularly considering evolving regulatory frameworks and supportive government policies.

[94] For example, see the 2018 report by the Nuclear Energy Agency at this link: https://www.oecd-nea.org/upload/docs/application/pdf/2019-12/7204-employment-nps.pdf, and a 2023 report commissioned by UK's Nuclear Industry Association at this link: https://www.niauk.org/wp-content/uploads/2023/01/Delivering-Value_Economic-Impact-Civil-Nuclear.pdf.

[95] For example, both the Biden administration and the new Trump administration have established policy positions to bolster domestic nuclear power production. For an example, see the White House briefings at this link: https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/05/29/fact-sheet-biden-harris-administration-announces-new-steps-to-bolster-domestic-nuclear-industry-and-advance-americas-clean-energy-future/. Trump's executive order "Unleashing American Energy" can be found at this link: https://www.whitehouse.gov/presidential-actions/2025/01/unleashing-american-energy/.

Table 7
Differences Between and Within NPP versus Never Considered Counties

Panel A: Differences in Geographical Characteristics

	(1)	(2)
	Considered vs Never Considered	Treated vs Control
Log of distance of nearest major earthquake (1900-1960)	-0.022***	0.019
	(0.006)	(0.066)
Log of distance of nearest most populated city in the country	-0.021***	-0.01
	(0.005)	(0.050)
Average Precipitation (1901-2000)	0.009	-0.008
	(0.008)	(0.095)
Average Temperature (1901-2000)	-0.012	0.032
	(0.071)	(0.872)
The highest point (in Ft)	-0.006***	0.046
	(0.003)	(0.038)
Population per square mile	-0.016***	0.089
	(0.008)	(0.181)
Water to land area ratio	0.017***	-0.023
	(0.005)	(0.088)
Pseudo R-Squared	0.064	0.074

Columns 1 and 2 in Panel A show the marginal effects obtained from estimating Equations (2.2) and (2.1) with logit. The outcome in column 1 (2) is an indicator that takes the value of one if a county is selected for an NPP (the NPP construction project in the county is canceled). The number of observations in columns 1 and 2 of Panel A is 3,040 and 132. All counties in the US (only the counties selected for an NPP siting) enter into the regression in column 1 (2). Average precipitation and temperature are divided by 100, and the highest point and population per square mile are divided by 1000. White standard errors are reported in the parentheses. . ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

Table 7
Differences Between and Within NPP versus Never Considered Counties

Panel B: Differences in Socio-economic Characteristics

	(1)	(2)
	Considered vs Never Considered	Treated vs Control
Percentage of Blacks (1950-1960)	0.676	0.080
	(1.469)	(3.309)
Percentage of 65 years and older population (1950-1960)	-0.321	-0.210
	(0.251)	(0.496)
Percentage of population 25+ with HS degree or higher (1960)	2.929***	0.086
	(0.984)	(2.120)
Percentage of Foreign Born (1960)	1.004***	-0.215
	(0.290)	(0.580)
Percent of farms in land area (1950-1960)	-6.926***	1.024
	(2.430)	(5.107)
Log of Median of House Value (1950-1960)	0.240***	0.009
	(0.031)	(0.072)
Log of Median of House Rent (1950-1960)	0.115***	0.014
	(0.028)	(0.063)
Percentage of Owner-Occupied Houses (1950-1960)	0.176	0.349
	(0.883)	(1.947)
Employment Rate (1950-1960)	1.060***	0.567
	(0.351)	(0.700)
Log of Median Family Income (1950-1960)	0.165***	-0.017
	(0.032)	(0.069)

The number of observations in column 1 (2) of Panel B ranges between 2,820 and 3039 (129 and 132) depending on the availability of Census data. Panel B presents the estimates obtained from the propensity score matching method, where the propensity scores are estimated with Equations (2.1) and (2.2) (Panel A). Each cell represents an estimate for a separate outcome listed in the rows. White standard errors are reported in the parentheses. . ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

Table 8
Construction of the Variables of Interest with Hypothetical Examples

NPP	County	Announce Year	Cancel Year	Constr. Lic.	Comm. Op. Lic.	Constr. Dummy	Comm. Op. Dummy
A	X	1960	1962	NA	NA	Always=0	Always=0
B	Y	1965	1970	1967	NA	=1 (1967-70)	Always=0
C	Z	1972	NA	1973	1980	=1 (1973-80)	=1 in 1980&after
D1	W	1962	NA	1966	1972	=1 (1966-75)	=1 in 1972& after
D2	W	1965	NA	1968	1975		

The table presents hypothetical examples for constructing the variables of interest in our study.

Table 9
Summary Statistics of the Outcomes in the Pre-treatment Period (1969-1973)

	(1)	(2)	(3)
	Treated	Control	Rest of the US
Employment-to-Population Ratio	40.785	39.595	40.977
Per capita wages and salaries	12.615	12.77	10.821***
Per capita personal income	24.582	25.765	23.447*
Total observations	115	170	14791

The table presents the means of the outcome variables for samples of counties between 1969 and 1973. In column 1 are the counties where an NPP is constructed. Column 2 shows the averages for the counties in which the construction of an NPP was announced but canceled. Column 3 presents the means for the rest of the counties in the US. ***, **, and * in columns 2.3 indicate that the p-values are smaller than 0.01, 0.05, and 0.1, respectively, for the null hypothesis that the mean is equal to that in column 1.

Table 10
The Impact of Nuclear Power Plants on Employment and Wages

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	11.454**	0.383***	8.411**
	(4.608)	(0.117)	(3.229)
Commercial Operation	3.314	0.224**	3.543
	(3.842)	(0.096)	(2.195)
N	2907	2907	2907
Mean of the outcome	48.376	16.844	16.844

The table presents the estimates obtained from Equation (2.3). The outcome in column 1 is the Employment-to-population ratio, defined as (total employment/population)*100. The outcome in columns 2 (3) is the natural logarithm (level) of per capita real wages and salaries (in 2019 dollars, $1,000s). The means of the outcome variables are at the bottom. In parentheses are the standard errors clustered at the county level. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

Table 11
The Impact of Nuclear Power Plants on Income Components

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Wages & Salaries	Suppl. to Wages & Salaries	Propri. Inc.	Div., Int.& Rental Inc.	Transfers	Contr. to Gov. Soc. Ins.	Total Personal Inc.
Constr.	8.41**	1.78**	-0.03	0.27	-0.31***	0.91**	1.22*
	(3.23)	(0.67)	(0.21)	(0.27)	(0.10)	(0.37)	(0.71)
Comm. Op.	3.54	1.31**	-0.01	0.43	-0.56**	0.27	3.06*
	(2.20)	(0.50)	(0.47)	(0.53)	(0.23)	(0.28)	(1.57)
N	2907	2907	2907	2907	2907	2907	2907
Outcome Mean	16.84	3.74	3.08	6.28	5.63	2.45	35.67

All outcomes are per capita. Outcomes in columns 1: Wages and Salaries, 2: Supplements to Wages and Salaries, 3: Proprietors Income, 4: Dividends, Interest, and Rental Income, 5: Personal Transfers, 6: Contributions for Government Social Insurance, 7: See notes in Table 10.

Table 12
The Impact of Nuclear Power Plants on Employment by Industry

	(1)	(2)	(3)	(4)	(5)
	Agr., Forest., & Fishing	Manuf.	Mining	Constr.	Transp. & Pub. Util.
Constr.	0.010	-0.038	0.006	5.328**	0.173
	(0.037)	(0.385)	(0.066)	(2.444)	(0.162)
Comm. Op.	0.104	-0.207	-0.089	-4.328*	1.262***
	(0.080)	(0.765)	(0.116)	(2.578)	(0.408)
N	1657	1779	1650	1814	1675
Outcome Mean	0.588	8.708	0.291	3.863	2.125

	(6)	(7)	(8)	(9)	(10)
	Whole. Trade	Ret. Trade	Fin., Ins. & Real Est.	Services	Gov. & Gov. Enter.
Constr.	0.078	0.053	-0.125	0.301	3.771
	(0.064)	(0.144)	(0.103)	(0.240)	(2.737)
Comm. Op.	0.191*	-0.087	0.113	1.714*	2.963
	(0.108)	(0.309)	(0.248)	(0.869)	(1.862)
N	1801	1813	1796	1694	1824
Outcome Mean	1.574	7.174	2.51	9.388	7.210

Outcomes are the proportions of employment in the industries in the county population. Industries in column 1: Agricultural services, forestry, and fishing, 2: Manufacturing, 3: Mining, 4: Construction, 5: Transportation and public utilities, 6: Wholesale Trade, 7: Retail Trade, 8: Finance, Insurance and real estate, 9: Other services, 10: Government and government enterprises. See notes in Table 10.

Table 13
The Impact of Nuclear Power Plants on Earnings by Industry

	(1)	(2)	(3)	(4)	(5)
	Agr., Forest., & Fishing	Manuf.	Mining	Constr.	Transp. & Pub. Util.
Constr.	0.012	-0.091	-0.008	4.417**	0.169
	(0.017)	(0.257)	(0.044)	(1.918)	(0.154)
Comm. Op.	0.030	-0.016	-0.051	-3.356*	1.637***
	(0.026)	(0.451)	(0.074)	(1.846)	(0.510)
N	1657	1779	1650	1814	1675
Outcome Mean	0.196	5.336	0.171	2.478	1.476

	(6)	(7)	(8)	(9)	(10)
	Whole. Trade	Ret. Trade	Fin., Ins. & Real Est.	Services	Gov. & Gov. Enter.
Constr.	0.043	0.001	-0.017	0.153	3.042
	(0.044)	(0.043)	(0.054)	(0.244)	(2.090)
Comm. Op.	0.245**	-0.030	0.184	1.003	2.255
	(0.119)	(0.124)	(0.191)	(0.731)	(1.403)
N	1801	1813	1796	1694	1824
Outcome Mean	0.919	2.082	0.752	3.49	3.711

Outcomes are earnings in 1,000 real (base year: 2019) dollars. Industries in column 1: Agricultural services, forestry, and fishing, 2: Manufacturing, 3: Mining, 4: Construction, 5: Transportation and public utilities, 6: Wholesale Trade, 7: Retail Trade, 8: Finance, Insurance and real estate, 9: Other services, 10: Government and government enterprises. See notes in Table 10.

Table 14
The Impact of Nuclear Power Plants on Neighboring Counties

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	0.610	0.005	0.495
	(0.846)	(0.024)	(0.568)
Commercial Operation	-0.509	-0.033	0.280
	(1.741)	(0.051)	(1.617)
N	2907	2907	2907
Mean of the outcome	52.685	20.829	20.829

Observations include the highest populated counties adjacent to the treated and control counties. The table presents the estimates obtained from a specification similar to Equation (2.3), where the *Construction* and *Commercial Operation* variable equals one if an NPP is being constructed and commercially operating in the neighboring county. See notes in Table 10.

Table 15
Impact of NPPs on Local Government Finances

Panel A: Revenues

	(1)	(2)	(3)	(4)
Sample	All Gov. Units	County, City, and Townships	Special Function Units	County, City, and Townships
Outcome	Total Revenues	Total Revenues	Total Revenues	Per capita Revenues
Construction	0.137**	0.168***	0.122	0.121***
	(0.062)	(0.045)	(0.112)	(0.039)
Commercial Op.	0.193	0.227*	0.191	0.101
	(0.154)	(0.116)	(0.216)	(0.071)
N	71278	24425	46853	24405

Panel B: Expenditures

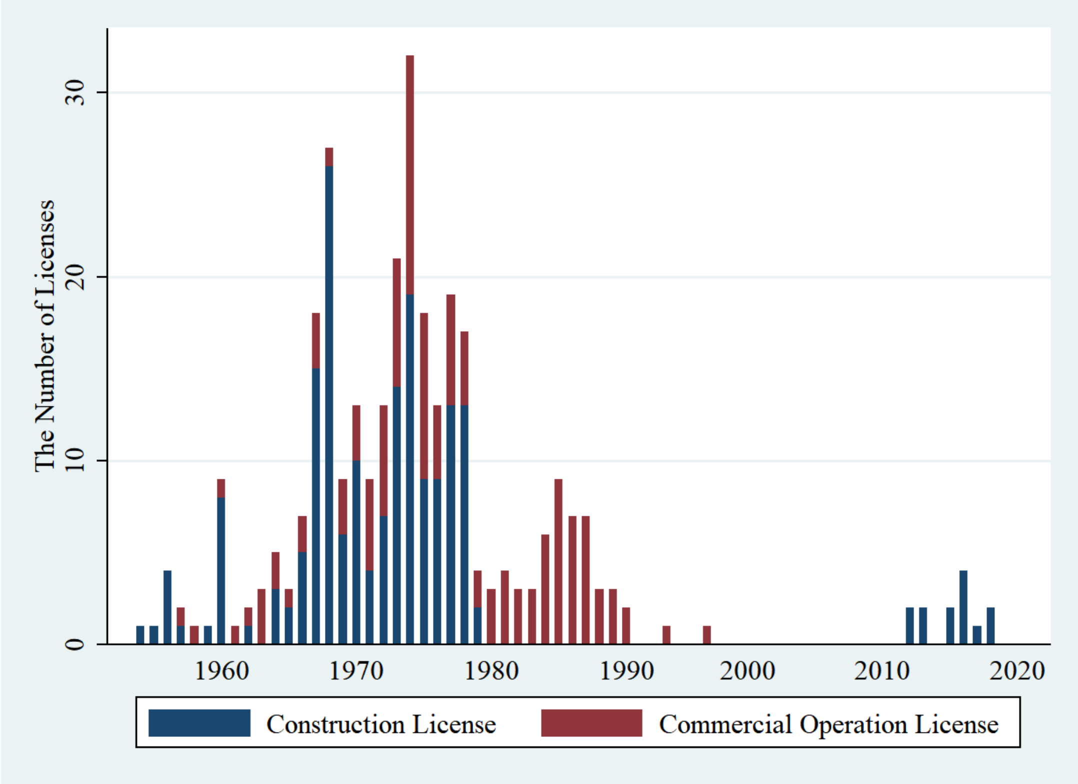
	(1)	(2)	(3)	(4)
	All Gov. Units	County, City, and Townships	Special Function Units	County, City, and Townships
	Total Expenditure	Total Expenditure	Total Expenditure	Per capita Expenditures
Construction	0.118**	0.135***	0.112	0.101***
	(0.058)	(0.050)	(0.103)	(0.036)
Commercial Op.	0.163	0.235*	0.133	0.103
	(0.167)	(0.124)	(0.237)	(0.062)
N	71278	24425	46853	24405

Panel C: Per Capita Expenditures by Function (County, City, and Township governments)

	(1)	(2)	(3)	(4)	(5)	(6)
	Education	Gov. Admin.	Public Safety	Public Assistance	Public Works	Other
Construction	0.032	0.439**	0.125	0.196	0.150*	-0.134
	(0.110)	(0.198)	(0.141)	(0.187)	(0.079)	(0.136)
Commercial Op.	-0.021	0.494**	-0.099	-0.990**	0.056	0.233
	(0.121)	(0.198)	(0.166)	(0.435)	(0.123)	(0.143)
N	24405	24405	24405	24405	24405	24385

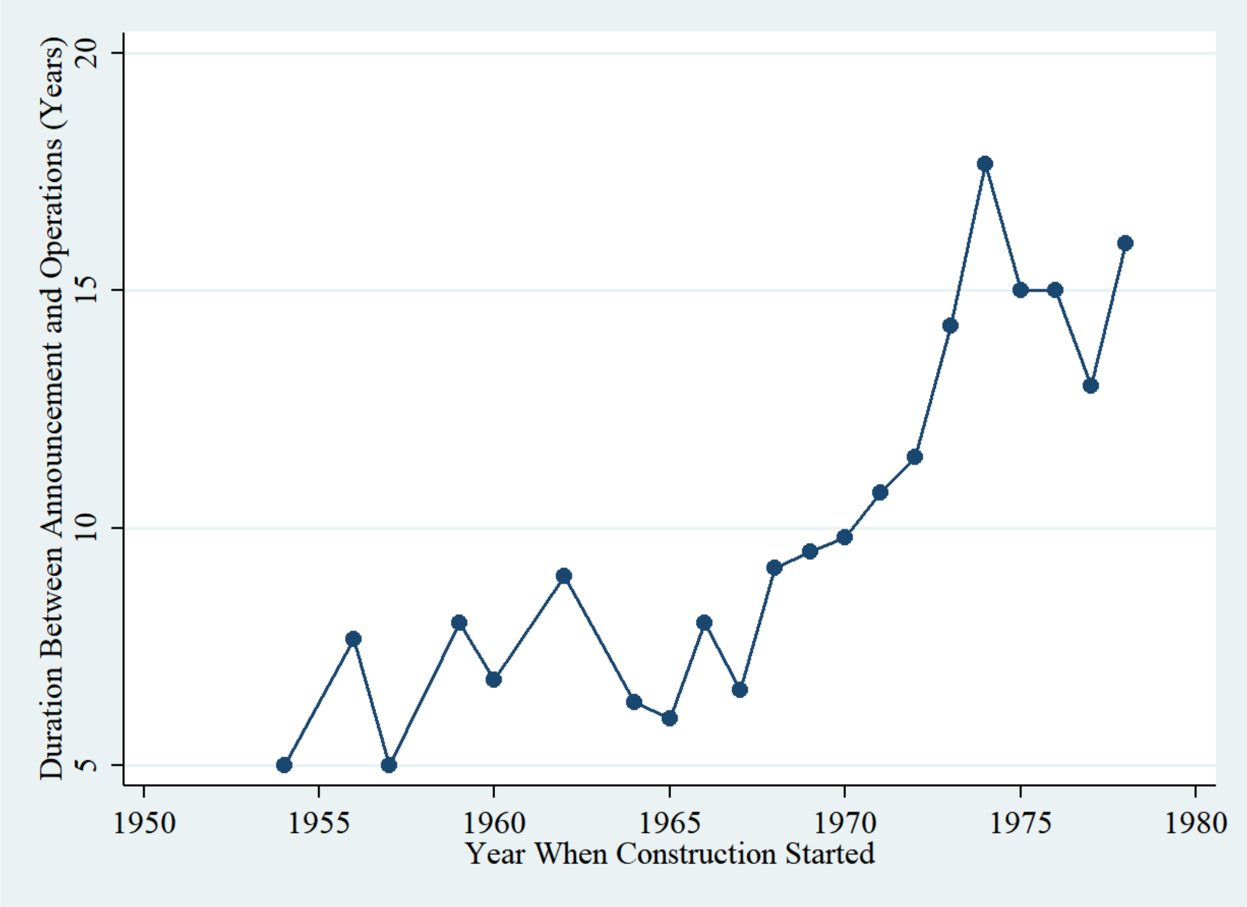
Estimates are obtained from equation (2.5). The unit of observation is a local government unit-year. Outcomes are in logs. See notes in Table 10.

Figure 5
The Number of Nuclear Reactor Construction Licenses Awarded Over Time



The blue (red) bars indicate the number of construction (commercial operation) licenses the NRC granted to nuclear reactors to start construction (selling electricity generated in the reactor).

Figure 6
Time Between the Year of Announcement and Issuance of Commercial License

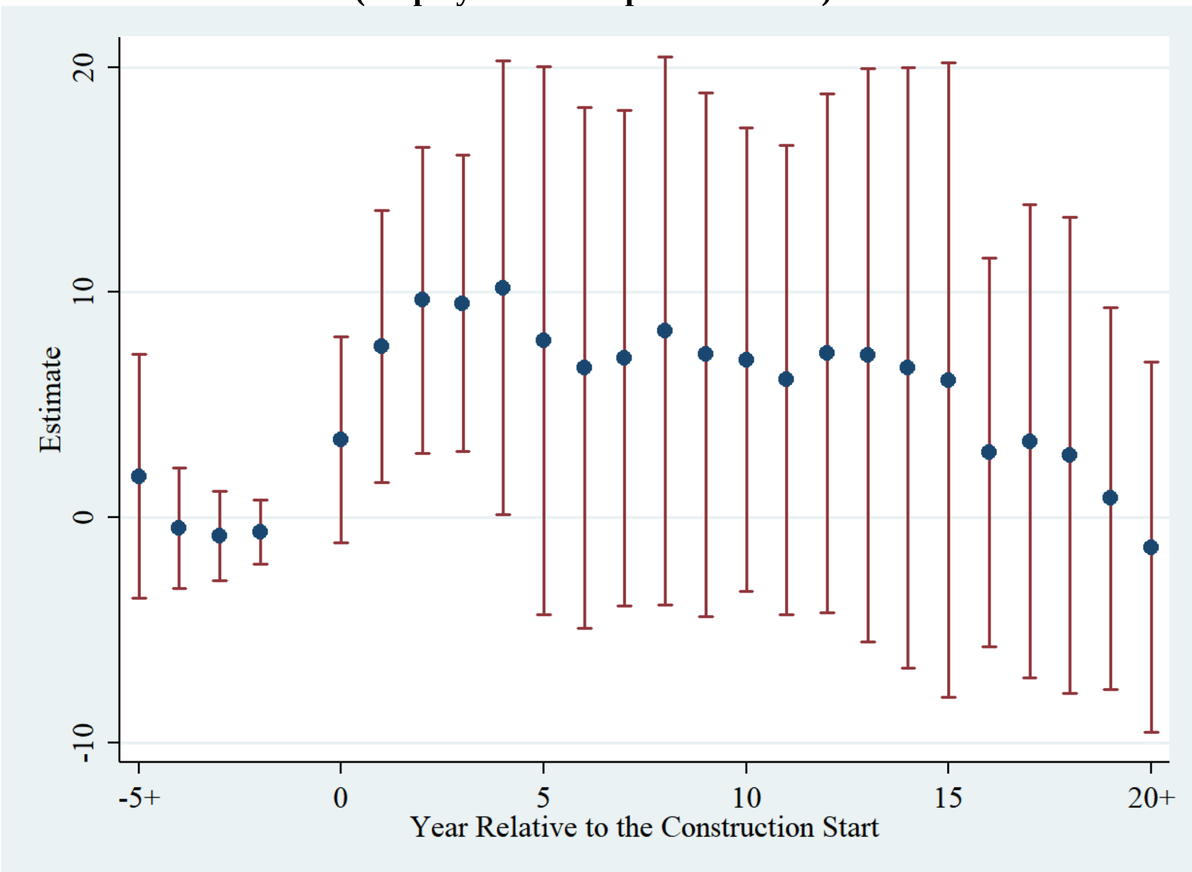


The figure includes the reactors that ultimately commercially operate. The horizontal axis is the year of the start of construction. The vertical axis is the average number of years between the years of the announcement of the nuclear reactors and the years in which they obtained commercial operation licenses.

Figure 7
Nuclear Power Plant Site Selection

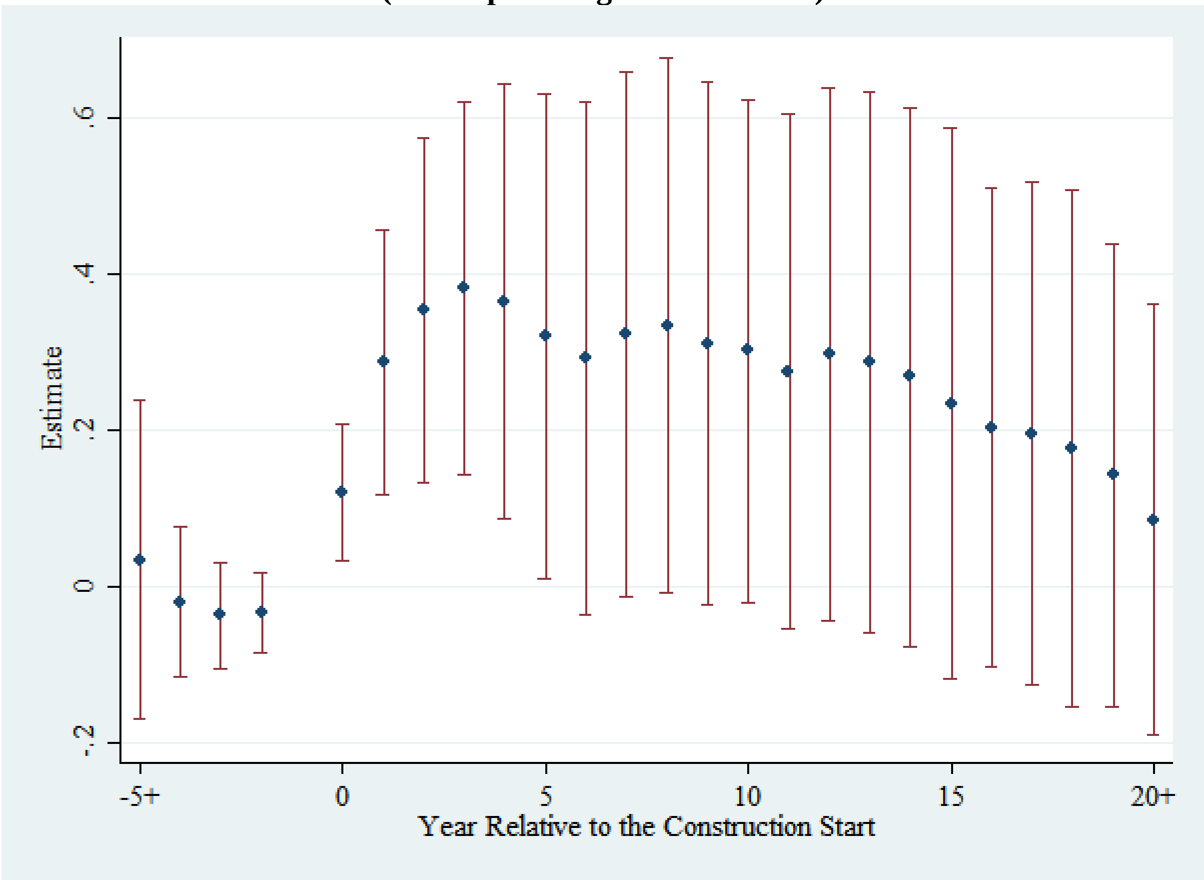
Counties that are marked with dark and light gray are selected for the construction of a nuclear power plant. In 98 counties (dark gray), the construction was at least partially completed. In 34 counties (light gray), the construction plans were canceled.

Figure 8
Event Study Analysis
(Employment-to-Population Ratio)



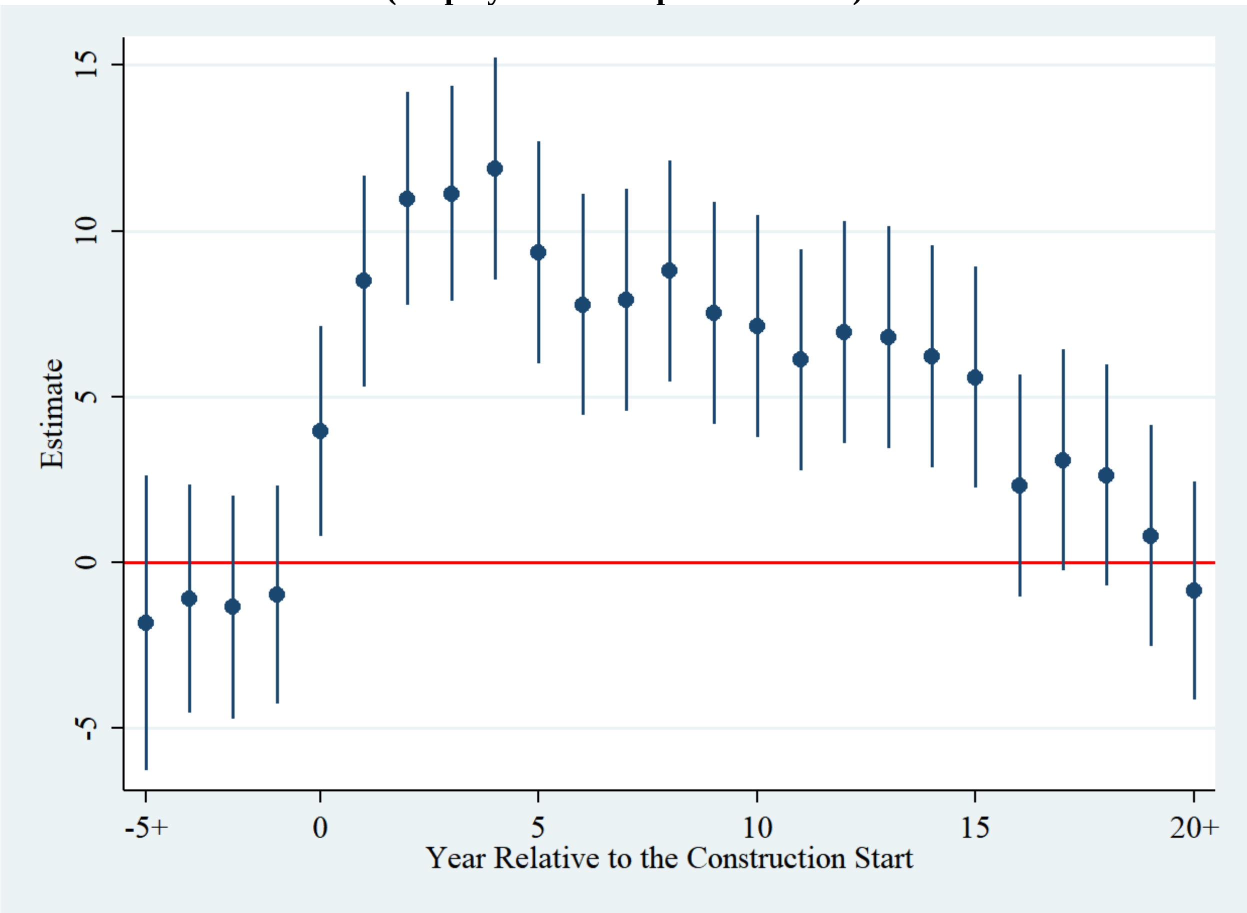
8
The outcome is the employment-to-population ratio. Estimates are obtained from equation (4). Points (capped lines) represent the point estimates (95% confidence intervals).

Figure 9
Event Study Analysis
(Per Capita Wages and Salaries)



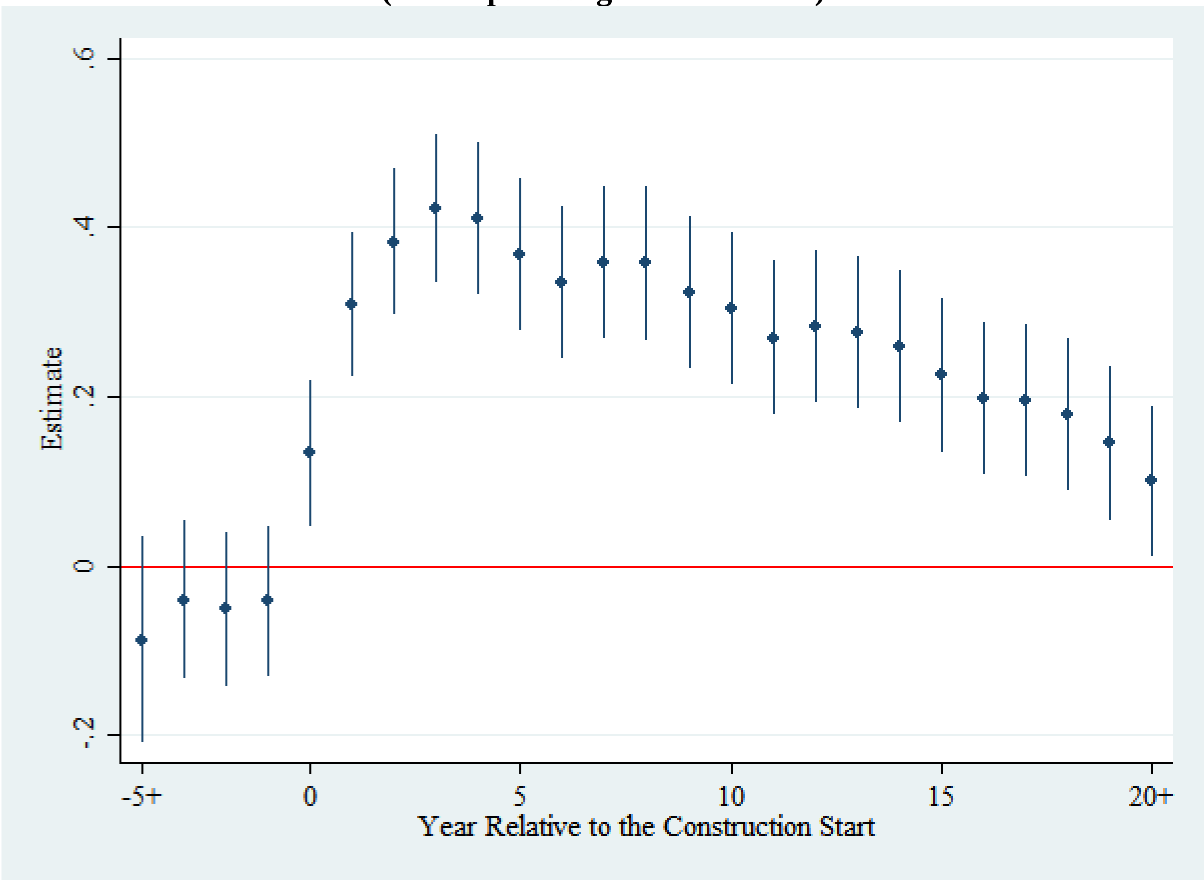
The outcome is the log of per capita wages and salaries. Estimates are obtained from equation (4). Points (capped lines) represent the point estimates (95% confidence intervals).

Figure 10
Event Study Analysis Using the Sun and Abraham (2021) Estimator
(Employment-to-Population Ratio)



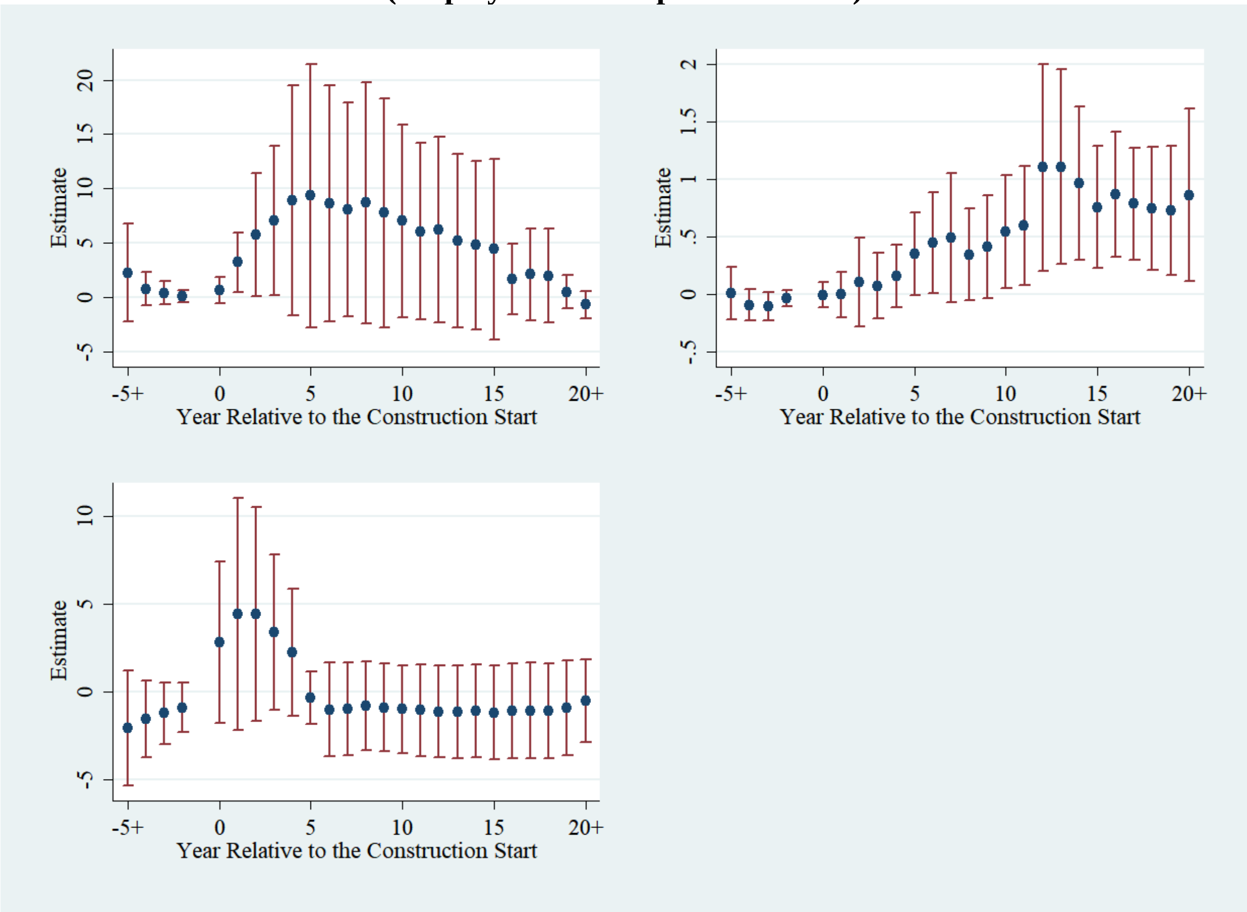
Event study estimates from the estimator in Sun and Abraham (2021). 0 indicates the first year of construction/operation, 1 is the second year, and so on. The outcome is the employment-to-population ratio.

Figure 11
Event Study Analysis Using the Sun and Abraham (2021) Estimator
(Per Capita Wages and Salaries)



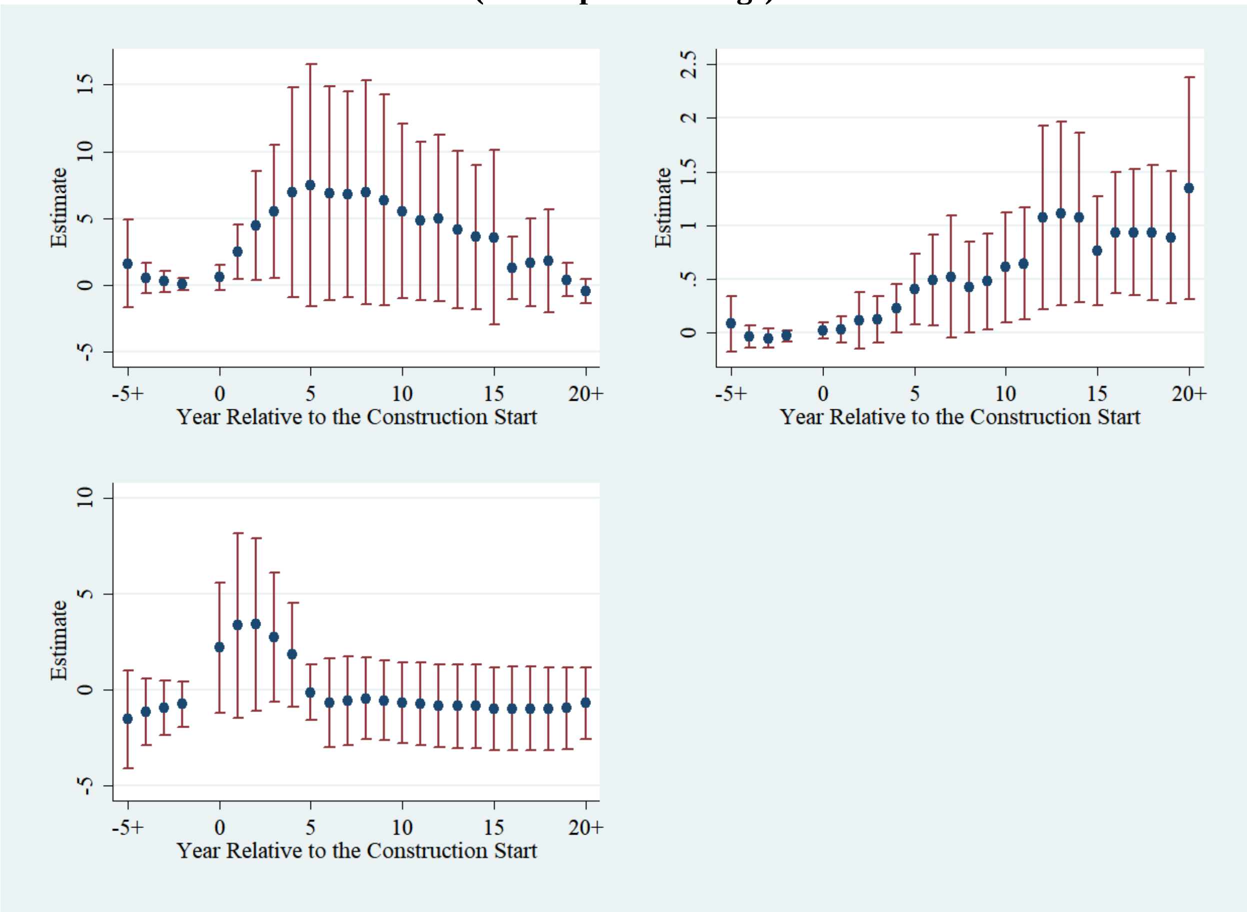
Event study estimates from the estimator in Sun and Abraham (2021). 0 indicates the first year of construction/operation, 1 is the second year, and so on. The outcome is log per capita wages and salaries.

Figure 12
Industry-Specific Event Study Analyses
(Employment-to-Population Ratio)



The outcome is the employment-to-population ratio in the industry. Estimates are obtained from equation (4). Points (capped lines) represent the point estimates (95% confidence intervals).

Figure 13
Industry-Specific Event Study Analyses
(Per Capita Earnings)



The outcome is the log of per capita earnings in the industry. Estimates are obtained from equation (4). Points (capped lines) represent the point estimates (95% confidence intervals).

Chapter 3

Automation and the Decline in Social Security Disability Insurance Applications

3.1 Introduction

SSDI application rates generally followed an upward trend from the early 1990s until an unexpected and sharp reversal began in 2010.[96] Various hypotheses have been put forward to explain the abrupt change in applications, including improved economic conditions, the transition of Baby Boomers out of disability-prone age brackets, enhanced support for dis abled individuals in the workforce, modifications in claim processing, and a lack of program awareness among eligible individuals.[97] This study explores the relationship between SSDI applications and the proliferation of robotics and other automation technologies.

Recent research shows that automation accounts for most of the change in the US wage structure since 1980 (Acemoglu and Restrepo, 2022b, 2024; Acemoglu et al., 2024). Automation could impact disability uptake by displacing workers and lowering wages or reducing injury-prone tasks (e.g., see Gihleb et al., 2022). The predicted impact on SSDI applications is therefore ambiguous, as automation may create slack conditions in the labor market, increasing uptake, or create safer work conditions, decreasing uptake. This paper presents estimates on the relationship between exposure to automation technologies and SSDI application rates.

Using confidential data on SSDI applications measured at the commuting-zone level, we estimate the effect of various measures of "automation" exposure on changes in per-capita SSDI applications by age group (18–64, 18–34, 35–54, and 55–64) and age group × sex. The extent of automation in the economy is measured via employment shares, which follows Autor and Dorn (2013), and industrial robot exposure, which is based on the measures used in Gihleb et al. (2022) and Acemoglu and Restrepo (2020). Long- and stacked-differences specifications are estimated using the 2005–2019 to estimate the effects of automation exposure on changes in the SSDI application-to-population ratio.

The regression equation of interest is estimable by ordinary least squares (OLS), but the adoption of automation technologies can be influenced by other factors that affect firms' demands for labor within a commuting zone. If these factors were observable, we could account

[96] See https://www.ssa.gov/oact/STATS/table6c7.html.
[97] e.g., see https://www.ssa.gov/policy/docs/briefing-papers/bp2019-01.html.

for them. However, many of these factors are likely unobserved and, as a result, an alternative estimation strategy is needed. To circumvent these empirical identification issues, the study employs an instrumental variables estimation strategy that follows Autor and Dorn (2013) for the employment-share measures and Gihleb et al. (2022) and Acemoglu and Restrepo (2020) for the measure of industrial robot exposure. For the former, we develop a Bartik-type instrument from the 1990 routine employment share to predict automation related employment shares in 2005.[98] For the latter, we leverage the fact that European countries led the US in terms of robot adoption and, therefore, industry-specific adoptions across France, Denmark, Finland, Italy, Germany, Norway, Spain, Sweden, and the United Kingdom are used as an instrument to predict US industrial robot exposure in 2005. For both the automation-related employment-share measures and the robot-exposure measure, the instruments are relevant in the first stage and contend that they affect SSDI outcomes via their impact on the suspected endogenous variable (i.e., the automation measures).

Application rates for the 18- to 64-year-olds decrease with greater automation expo sure. These effects are not statistically significant with OLS but become significant with 2SLS. The estimates for 18–64 year-olds, however, mask heterogeneity in the effects of automation exposure across the 18–34, 35–54, and 55–64 age groups. The 35–54 age group shows the largest negative effect, while the 55–64 age group also shows significant but smaller negative effects.

The study's findings for the link between industrial robot exposure and SSDI application rates align with those from the analysis using the automation-related employment-share measures. One difference in the estimates of industrial robot exposure's impact and that of the employment-share measures is that the negative effects are statistically significant, regardless of whether OLS or 2SLS is used to estimate the parameters of interest. The similarities include null effects for the 18–34 age group and a negative and statistically significant effect for the 35–54 and 55–64 age groups. Moreover, industrial robot exposure has the largest effect on the 35–54 age group, which is also consistent with the findings for the employment-share measures.

3.2 Literature Review

3.2.1 Disability Applications

[98] See, for example, Goldsmith-Pinkham et al. (2020); Borusyak et al. (2022)

Since the 1970s, the SSDI program has undergone dramatic changes, mainly influenced by shifts in beneficiary demographics and a labor market with increasing female participation (Liebman, 2015). As baby boomers began to reach ages with higher SSDI risk, concerns arose about the program's solvency and policies (Autor and Duggan, 2006). Maestas et al. (2021) find that the Great Recession led to nearly one million additional SSDI applications, of which 41.8 percent were awarded benefits, resulting in over 400,000 new beneficiaries who constituted 8.9 percent of all new beneficiaries during the recession. However, an unexpected large decrease in SSDI claiming began as the deleterious employment effects of the Great Recession began to wane. As of 2022, approximately 9 million individuals received $12.6 billion in SSDI payments and the program is expected to remain solvent through the long-run 75-year projection period (US Social Security Administration, 2023).

3.2.2 Automation and Tasks

General Automation technologies:

The increasing penetration of robots and technology in the labor market has raised concerns about the future of employment and wages (Johnson and Acemoglu, 2023). Acemoglu and Restrepo (2022b) reveal that automation-driven task displacement accounts for 50–70 percent of changes in the US wage structure over the past four decades, particularly affecting routine-task workers. Their study also shows that automation significantly reshapes wage structures, increasing the college premium while reducing real wages for less educated workers.

Autor and Dorn (2013) exploit differential exposure to routine tasks across local labor markets to identify the effects of automation technology adoption on labor-market outcomes. By combining past values of a commuting-zone-industry employment share with industry-level routine occupational share, the authors create a shift-share-type instrumental variable (e.g., see Goldsmith-Pinkham et al., 2020). They find that automation (particularly computers) increased wages for high-skilled workers and substituted for lower-skill workers in routine tasks. This also increased demand for service-sector jobs and reallocated lower-skilled labor to those jobs.

Bratsberg et al. (2022) study the Norwegian labor market over approximately the same sample period as this study. Individuals in occupations with higher routine task intensity (RTI) scores in 2003 were significantly less likely to remain employed and more likely to receive a disability pension or die by 2019. The study found that a standard-deviation increase in RTI score was associated with a 6.7 percent higher mortality rate for men and a 5.5 percent higher rate for women.

Acemoglu and Restrepo (2022a) show that aging populations promote from demand for automation technologies, particularly as older workers are often less suited for physically demanding tasks. This scenario leads to a capital substitution effect, where labor-saving automation technologies become more prevalent because of the relative scarcity and higher wage demands of younger workers. Identifying an effect of automation on disability claiming requires taking account of these market forces, as aging populations will have both higher levels of disability and automation.

The literature on the health effects of automation risk has mixed results from labor markets around the world. Lordan and Stringer (2022) find small, negative effects of automation risk on the mental health of Australian workers. Blasco et al. (2024) also find negative effects on mental health for French workers, but the effect sizes are larger than those reported in Lordan and Stringer (2022). In Blasco et al. (2024), anxiety over the labor market is the primary mechanism through which the effect operates. Cheng et al. (2021) study a nationwide survey from Taiwan and find that people employed in jobs with a high likelihood of automation tended to experience lower levels of job control, greater job insecurity, and a higher prevalence of work-related injuries and illnesses. In contrast, those in positions with a low probability of automation faced greater psychological and physical demands, along with a higher incidence of burnout.

Industrial Robots:

A growing body of literature examines the effects of industrial robot exposure on employment, wages, and worker health across different regions and market sectors. Acemoglu and Restrepo (2020) analyze the effect of the increase in industrial robot usage on US local labor markets. They show that robots may reduce employment and wages, and that the local-labor-market effects of robots can be estimated by regressing the changes in employment and wages on robot exposure. Dauth et al. (2021) estimate the effect of industrial robots on employment, wages, and the composition of jobs in German labor markets. They find that the adoption of industrial robots had no effect on total employment in local labor markets specializing in industries with high robot usage. Robot adoption led to job losses in manufacturing that were offset by gains in the business service sector. Acemoglu et al. (2020) study the firm-level implications of robot adoption in France and find that adopters experienced significant declines in labor shares and the share of production workers in employment and increases in value added and productivity and overall employment. However, the employment expansion comes at the expense of competitors, leading to a net negative effect on employment.

The literature on the health effects associated with industrial robot penetration is less mixed than for automation shares and automation risk. Gunadi and Ryu (2021) investigate the health impacts of increased industrial robot use in US cities. They find that a 10 percent increase in robots per 1,000 workers corresponds to a 10 percent decrease in poor health, work disability, and job quitting for health reasons among this group. This improvement is partly attributed to a shift away from physically demanding tasks. Gihleb et al. (2022) find that an increase of one standard deviation in industrial robot use reduces work-related injuries by approximately 1.2 per 100 full-time workers in the US and Germany. They report an annual injury cost reduction of $1.69 billion (2007 dollars) from 2005-2011, mainly due to decreased injury rates in manufacturing. However, US mortality and mental health data, shows a countervailing effect, as industrial robot penetration into a labor market significantly escalates substance abuse-related deaths and negatively impacts mental health. O'Brien et al. (2022) also found that robot penetration increases the rate of overdose deaths in the United States. However, Gihleb et al. (2022) find no consequential effects on suicide rates, suggesting that labor-market pressures and anxiety, exacerbated by robot penetration, are primarily responsible for the observed impacts.

3.3 Data

3.3.1 SSDI Applications

The dependent variables used in this study are population-weighted SSDI application counts across commuting zones and years by (a) age group (18–64, 18–34, 35–54, and 55–64) and (b) age group × sex. These data were provided by the SSA and are confidential.[99] In some instances, application counts from particular CZs are suppressed because of an insufficient number of applications in these locations. The missing information varies by (a) age group and (b) age group × sex. When sex is ignored in the application counts, the data cover 86 percent of CZs

[99] All results involving the data provided by SSA must be approved by the SSA's Disclosure Review Board (DRB) for external presentation. The estimates reported herein have obtained the DRB's approval.

(643 out of 722) for the 18–64 age group;[100] 87 percent for the 18–34 age group, 95 percent for the 35–54, and 94 percent for the 55–64 age group. When examining application rates by sex and age group, the shares of CZs that are covered fall to 77, 78, 89, and 86 percent for the 18–64, 18–34, 35–54, and 55–64 age groups, respectively.[101] The CZs with missing SSDI application data tend to be located in the Mountain West and West North Central Midwest, and these locations have relatively small populations.

To adjust the raw application counts for population size, we use total population counts from the US Census's intercensal tables, which provide county-year population statis tics by age and sex. Using the county-year population counts, the populations of counties within a CZ were then summed.[102] The outcome measure, termed the application-to population ratio, is calculated as follows:

$$(3.1) \qquad Apps^{*}_{a,s,c,t} = \frac{Apps_{a,s,c,t}}{Pop_{c,t}} \times 100$$

The terms a, s, c, and t index age groups, sex, commuting zones, and years, respectively. The variable $Apps^{*}_{a,s,c,t}$represents the SSDI application-to-population ratio for age group a and sex s in commuting zone cd at year t; $Apps_{a,s,c,t}$ is the raw count of SSDI applications for age group a and sex s in commuting zone c at year t; and $Pop_{c,t}$ is the population count in commuting zone cd at year t. In our econometric analysis, we estimate separate regressions for the 18–64, 18–34, 35–54, and 55–64 age groups with and without taking sex into account.

Table 17 reports the averages (column 1) and standard deviations (column 2) of the SSDI application-to-population ratios for each age grouping. In addition, we decompose the standard deviation for each outcome variable into "within" and "between" components (column 3 and 4). The ratios for males and females combined are presented in Panel A, and those for males and females separately are shown in Panels B and C, respectively. Regardless of sex, the SSDI

[100] Using 1990 commuting zones, the entire US has 741 commuting zones. Our analysis is restricted to the continental US, which results in the exclusion of 19 commuting zones located in Alaska and Hawaii.
[101] The 18- to 64-year-old outcome is created by summing the application counts for the 18–34, 35–54, and 55–64 age groups (and then dividing by the commuting zone's population). The SSA redacted data for commuting zones with applications below the threshold count required for external presentation, and the redactions vary by age group as well as age group × sex. Thus, when computing the measure for 18- to 64-year olds, the numbers of observations are less than those for the subgroups, as a commuting zone with a missing value for only one of the three age groups results in a missing value for the application rate for all age groups combined. The extent of redaction is greater for the 18–34 age group than the 35–54 and 55–64 age groups.
[102] The county-year population statistics were downloaded from the US Census's intercensal tables (2000– 2009) and Vintage 2020 Population estimates (2010–2019).

application-to-population ratios vary considerably across the age groups. For example, the ratio for the 35–54 age group is almost three times larger than that for the 18–34 age group and 1.5 times than that for the 55–64 age group. Moreover, the variation between commuting zones is greater than the within-commuting-zone-variation by a factor of 1.6 to 2.4. These patterns hold for males and females (see columns 3 and 4).

In Figure 15, we present the 2005 geographic variation (across the continental US) in the SSDI application-to-population ratio (Panel A) along with its change between 2005 and 2019 (Panel B) for the 18–64 age group. The application-to-population ratios tend to be highest in the south, but commuting zones in southern states, such as North Carolina, West Virginia, Tennessee, South Carolina, Arkansas, and Kentucky, experienced relatively larger reductions than commuting zones located in the Mountain West, West North Central, and Middle Atlantic regions. In total, about 75 percent of commuting zones across the United States experienced reductions in the application-to-population ratios.

Figures 16, 17, and 18 present maps analogous to those shown in Figure 15 separately for each age group: 18–34, 35–54, and 55–64. Across the three figures, there is a similar concentration in 2005 of higher application-to-population ratios in the southern states. However, the evolution of the application-to-population ratios between 2005 and 2019 varies considerably across the three age groups. For the 35–54 age group, the reduction in the ratio spans the vast majority of the US, with only around 3 percent of commuting zones experiencing upticks in their ratios. The share of commuting zones experiencing upticks in their application-to-population ratios is 54 percent for the 18–34 age group and 72 percent for the 55–64 age group. The patterns in Figures 16 and 18 indicate that it is the 35–54 age group (i.e., Figure 17) driving the patterns observed in Figure 15.

3.3.2 Measuring Automation

Automation-Related Employment Shares:

We use three different employment-share variables to measure the extent of automation in the labor market. The measures are based on data from the Occupational Information Network (O*NET) and the American Community Survey (ACS). The computation of the three employment shares follow Autor and Dorn (2013), who connect the routine employment share to the hollowing out of middle-skill occupations in the US. Their measure identifies occupations in the top 1/3 of the 1950 routine task intensity distribution, and then employment in routine-intensive occupations as well as employment in general are totaled by commuting zone and year. We perform the same calculations but use variables from O*NET instead of the Department of Transportation (DOT) and rely on the 1990 task-intensity distributions instead of those from 1950.

The three employment-share variables are based on the routine task intensity measure from Deming (2017), which has two components. The first relates to the importance of performing repetitive tasks, and the second captures the extent to which a job is automated. The first employment share measure uses the composite measure, whereas the second and third employment-share variables use the individual components of the composite measure (i.e., repetition and automation). The three measures are referred to as the routine-, repetition-, and automation-intensive employment shares.

To perform the calculations, we first follow the approach commonly used in the literature and convert the measures based on the O*NET data, which are ordinal (typically 1-5 or 1-7 scales), to 0-10 scales (e.g., see Deming, 2017). This allows for identifying occupations in the top 1/3 of the variable's distribution. The occupation-identifier is then linked to the 2005-2019 ACSs via the 2018 Standard Occupation Classification (SOC) codes. For the occupations identified as being intensive in routine, repetition, and automation tasks, the number of workers employed in these occupations is counted as well as the number of workers employed overall in each CZ-year. The ratio of these employment counts yields the three employment-share variables used in our regression analysis. Formally, the employment shares are computed as follows:

$$(3.2) \qquad Empsh_{ct} = \sum_{j=1}^{J} Emp_{c,j,t} \ .1 \left[TaskInt_{j,2005} > TaskInt_{2005}^{p66}\right] \sum_{j=1}^{J} (Emp_{c,j,t})^{-1}$$

The terms c, j, and t index commuting zones, occupations, and time periods, respectively. The variable $Empsh_{ct}$ measures the share of workers employed in occupations in the top 1/3 of the automation measure's task-intensity distribution in commuting zone cs in year t; $Emp_{c,j,t}$ is employment in occupation j in commuting zone cs at time t; and $1 \left[TaskInt_{j,2005} > TaskInt_{2005}^{p66}\right]$ is an indicator function that identifies occupations in the 2005 top 1/3 of the task intensity distribution in 2005. The three variables that comprise Empshcs,t include the routine-intensive employment share ($RSH_{c_s,t}$), the repetition-intensive employment sha($RTSH_{c_s,t}$), and the automation-intensive employment share ($ASH_{c_s,t}$).[103]

Table 16 presents the top 10 occupations ranked for each of the automation-related employment shares. In some cases, there is overlap across the measures in terms of rankings. For example, the bookkeeping, accounting, and auditing clerks occupations are ranked #1, #1, and #8, respectively, across the routine, repetition, and automation intensity measures. However, travel agent and insurance claims occupations tend to be "automation intensive" and less "repetition intensive". By contrast, other occupations tend to be more "repetition intensive" than "automation intensive", such as tellers, pharmacists, and tax preparers.

In Figure 19, we present the initial routine-, repetition-, and automation-intensive employment shares across commuting zones for the year 2005 (Panels A, C, and E) as well as the change in each measure between 2005 and 2010 (Panels B, D, and F). In terms of the initial geographic distribution, no clear pattern emerges, as the heat map reveals both relatively high and relatively low employment shares in the same geographic regions.

The changes in these variables, however, vary considerably between 2005 and 2010, as the automation-intensive employment share becomes much more pervasive across commuting zones than the routine- and repetition-intensive employment shares. Similar to the initial geographic distributions, there is more of a pattern than the initial distribution, as Heartland disproportionately experiences increases between 2005 and 2010 (Panels B and D). For the automation-intensive employment share (Panel F), its increase is widespread, affecting each region across the US.

Industrial Robot Exposure:

The International Federation of Robotics (IFR) has collected data on industrial robot usage across countries and industries since the early to mid-1990s. The IFR provides industry year information on robot usage for various European countries dating back to its inception, but the

[103] The computation of the employment shares requires two separate aggregations. The first is to the PUMA year level, and the second is from the PUMA-year to the CZ-year level. Frequency weighting is applied in the first aggregation using the *perwt* variable from Ruggles et al. (2023). For the second aggregation, we incorporate the PUMA-commuting zone crosswalk from Autor and Dorn (2013), and sum the number of workers employed overall as well as those in occupations intensive in routine, repetitive, and general automation tasks across PUMAs within a CZ-year. Because there is uncertainty regarding the geographic location of survey respondents, we use the allocation factors from Autor and Dorn (2013), which are applied as importance weights in the aggregation from the PUMA-year to the CZ-year level.

inclusion of North America began in 2004. The IFR combined robot usage in the US, Canada, and Mexico into a single category from 2004 to 2010 before reporting separate statistics for each country beginning in 2010. Because our study period uses data reported before and after the reporting change, we use the North American statistics for the 2004 – 2010 period and then those for the United States thereafter. For the 2004 – 2010 period, the US accounts for over 90 percent of the North American market, making the use of North American robot usage a viable proxy (e.g., see Acemoglu and Restrepo, 2020). Measurement error is then inherent in the robot exposure measure used in our regression analysis, which results in attenuation bias. As such, an estimation strategy is needed to remove bias resulting from mismeasurement of the key explanatory variable. We return to this discussion in Section 3.4.3.

The IFR provides information on robot usage for nonmanufacturing industries (agriculture, forestry and fishing, mining, utilities, construction, education, research and development, and services) as well as disaggregated industries within the manufacturing sector (food and beverage, textiles, wood and furniture, paper and printing, plastics and chemicals, minerals, basic metals, metal products, industrial machinery, electronics, automotive, ship building and aerospace, and "other" manufacturing). In total, there are 19 industries in the IFR classification.

The analysis includes two measures of industrial robot exposure. The first is from Gihleb et al. (2022) and the second follows Acemoglu and Restrepo (2020). The measure from Gihleb et al. (2022) requires two data sources: the robot usage data from IFR and industry-commuting employment statistics from the 1990 Census. Acemoglu and Restrepo (2020) rely on the same data sources but also incorporate an industry-level output growth measure from the EU KLEMS and the world supply of industrial robots from the IFR. The measure from Gihleb et al. (2022) is defined as follows:

$$(3.3) \qquad RobotExp_{c,t}^{GGSW} = \sum_{i=1}^{I} l_{i,c,1990} \frac{M_{i,t}}{E_{i,1990}}$$

The GGSW superscript indicates that the measure is from Gihleb et al. (2022). The variable $M_{i,t}$ is the operation stock of autonomous robots in industry i in year t, and $E_{i,1990}$ is employment in industry i in 1990. The exposure measure is created by projecting the robots per worker to commuting zones via multiplication by the 1990 share of workers employed in industry i (i.e., $l_{i,c_d,1990}$).

The second measure, which is from Acemoglu and Restrepo (2020), is defined as follows:

$$(3.4) \qquad RobotExp^{AR}_{c,(t_0,t_1)} = \sum_{i=1}^{I} (l_{i,c,1990} \,.\, APR_{i,(t_0,t_1)})$$

in which

$$(3.5) \qquad APR_{i,(t_0,t_1)} = \frac{M_{i,t_1} - M_{i,t_0}}{E_{i,1990}} - g_{i,(t_0,t_1)} \frac{M^{w}_{i,t_0}}{E_{i,1990}}$$

The AR superscript indicates that the measure is from Acemoglu and Restrepo (2020). The variable $APR_{i,(t_0,t_1)}$ represents the adjusted robot-penetration ratio for industry i between time periods t_0 and t_1, which is projected to commuting zones via multiplication by the share of workers employed in industry i located in commuting zone c in 1990 (i.e., $l_{i,c,1990}$). In equation 5, the first term is the change in the operational stock of autonomous industrial robots in commuting zone c between time periods t_0 and t_1 relative to employment for industry i in 1990, and the second term is the growth rate of output for industry i between time periods t_0 and t_1 multiplied by the global stock of robots, M^{w}_{i,t_0} , for industry i at time t_0 relative to employment for industry i in 1990.

In Figure 20, we present the initial 2005 geographic distribution of robots per 1,000 workers (Panel A) and its changes between 2005 and 2010 (Panel B). Initially, robot usage is concentrated in the upper Midwest, but the usage in industrial productions expands in all areas. The expansion between 2005 and 2010 is greatest in the upper Midwest and surrounding areas.

3.4 Econometric Methodology

Our empirical strategy relies on both ordinary OLS and 2SLS estimation to compute the parameters of interest. In what follows, we outline our empirical approach for studying the relationship between the employment-share variables and SSDI applications (Section 3.4.2) as well as our strategy for empirically examining the link between exposure to industrial robots and SSDI applications (Section 3.4.3).

3.4.1 Dependent Variables

We transform SSDI outcomes presented in Table 17 into “long” or “stacked” differences for the purposes of our econometric analysis, which are based on differences in $Apps^{*}_{a,s,c,t}$ between a starting (t_0) and ending (t_1) period. We compute the following separately for each age-sex group:

$$(3.6) \qquad Apps^*_{c,(t_0,t_1)} = Apps^*_{c,t_1} - Apps^*_{c,t_0}$$

The starting and ending periods for our long and stacked differences specifications vary across the automation measures. When examining the employment-share measures and the robot-exposure measure from Gihleb et al. (2022), 2005 and 2019 are the starting and ending periods, respectively. For the stacked-differences specifications, there are two starting (s_1 and s_2) and ending (e_1 and e_2) periods. In our primary specification, $t_0^{s_1} = 2005$ and $t_0^{s_2} = 2010$ are the two starting periods and $t_1^{e_1} = 2010$ and $t_1^{e_2} = 2019$ are the two ending periods. The application of Acemoglu and Restrepo (2020)'s measure relies on 2004 and 2016 as the starting and ending years, respectively, for the long-differences specifications. For the stacked-differences specifications, the starting and ending years of the first "stack" are 2004 and 2007, respectively, and the starting and ending years of the second "stack" are 2013 and 2016, respectively.[104]

3.4.2 Automation-Related Employment Shares

The structural regression equation for the long-differences specifications is

$$(3.7) \qquad Apps^*_{c_s,(t_0,t_1)} = \alpha_0 + \alpha_1 \, Empsh_{c_s,t_0} + X'_{c_s,t_{-1}} \Lambda + \phi_s + \epsilon_{c_s,(t_0,t_1)}$$

The variable $Apps^*_{c_s,(t_0,t_1)}$represents the dependent variables defned in equations (3.6) for commuting zone c_s (s indexes the state in which the commuting zone is located) between a starting period, t_0 , and an ending period, t_1 . The employment-share variables, which are defned in equation (3.2), are represented by the variable $Empsh_{c_s,t_0}$, which is for commuting zone c_s at the starting period, t_0 . Separate models are estimated for the three employment share variables: the routine-intensive employment share (RSH_{c_s,t_0}), the repetition-intensive employment share ($RTSH_{c_s,t_0}$), and the automation-intensive employment share (ASH_{c_s,t_0}).

The vector of control variables, $X'_{c_s,t_{-1}}$, is from Autor and Dorn (2013), whose study includes controls for the college-to-noncollege population, the ratio of immigrants to the noncollege

[104] Our findings are insensitive to starting the analysis in 2004 versus 2005. Moreover, our findings are qualitatively robust to different starting and ending years across the two stacks. In the specification we present in this report, we use the early and late parts of the sample. The use of 2004-2007 and 2013-2016 also removes the influence of the Great Recession. We note that our results are even stronger if we examine, for example, 2004-2010 and 2010-2016 as the two stacked differences.

population, the share of workers employed in manufacturing, the unemployment rate, the share of the population that is female and employed, the share of the population 65 or older, and the share of noncollege workers earning a real wage below the minimum wage that will prevail over the next decade (2000-2010). In addition, we include two additional control variables: the "China Shock" from Autor et al. (2013) and an SSDI processing efficiency variable from Kearney et al. (2021). Each of the variables in $X'_{c_s,t_{-1}}$ is measured for commuting zone c_s in the year 2000, except the "China Shock" variable and the SSDI processing variable. The variable capturing rising import competition from China is the change in Chinese import exposure between 1990 and 2000, and the SSDI processing variable is measured in 2003 (the frst year it is available). The inclusion of ϕ_s, which is a set of state dummy variables, means that the estimate for the parameter of interest, α1, is identifed based on temporal variation within states and across CZs. The error term $\epsilon_{c_s,(t_0,t_1)}$, captures predictors of $Apps^*_{c_s,(t_0,t_1)}$ not held constant.

When estimating the stacked-differences specifications, equation (3.7) is altered as follows:

$$(3.8) \quad Apps^*_{c_s,(t_0^{s_i},t_1^{e_i})} = \beta_0 + \beta_1\, Empsh_{c_s,t_0^{s_i}} + \beta_2 D_{(t_0^{s_i},t_1^{e_i})} + X'_{c_s,t_{-1}^{s_i}}\,\Theta + \phi_s + \epsilon_{c_s,(t_0^{s_i},t_1^{e_i})}$$

Equation (3.8) adds new superscripts to the time indices, t0 and t1, as well as an additional right-hand-side variable, $D_{(t_0^{s_i},t_1^{e_i})}$. The superscripts s_i and e index the different starting and ending years used in the stacked-differences specifications. The inclusion of $D_{(t_0^{s_i},t_1^{e_i})}$ accounts for differences in the application-to-population ratio across the different periods. Thus, the stacked-differences estimates are based on within-period variation instead of variation over the entire time horizon, as is the case in the long-differences specifications.

We include two stacked differences, 2005–2010 and 2010–2019, in our empirical specifications. In effect, a panel of commuting zones with two time periods (one for each "stack") is formed for the dependent variables. Thus, each commuting zone has two observations. The first is for the 2005-2010 period and the second is for the 2010–2019 period. The variable $Empsh_{c_s,t_0^{s_{2005}}}$ is assigned to the first period (2005–2010) and $Empsh_{c_s,t_0^{2010}}$ is assigned to the second period (2010–2019). The variables in $X'_{c_s,t_{-1}^{s_i}}$ are the same as those in equation (3.7), except the 1990 version of the control variables are linked to the first period (2005–2010) and the 2000 version of the control variables are linked to the second period (2010–2019). For the "China Shock," we use the change between two different points in time: the change in Chinese import penetration from 1990–2000 is linked to the 2005–2010 difference, and the analogous change between 2000 and 2007 is linked to the 2010–2019 difference.[105] The 2003 values for the SSDI processing time control variable are linked to the 2005–2010 difference period, and the 2008 values are linked to the 2010–2019 difference period.

When equations (3.7) and (3.8) are estimated via OLS, $\hat{\alpha}_1$ and $\widehat{\beta_1}$ have causal interpretations if the following are true: (i) the automation proxies and population-adjusted SSDI applications are not jointly determined; (ii) after conditioning on the full set of control variables, the automation measures are uncorrelated with factors in the error term that also afect SSDI outcomes; and (iii) the automation-related variables are measured without error. Simultaneity bias is unlikely given that automation technologies generally afect economic outcomes with a delay. However, it is more likely that unmeasured variables in the error term or measurement error could bias our OLS estimates. Given these concerns, an alternative estimation strategy that circumvents these problems is needed.

We employ an instrumental variables (IV) estimation strategy to measure the causal effect of interest. For the three automation-related employment-share variables, we use an already-established instrument from Autor and Dorn (2013), who examine the effects of changes in the routine employment share on changes in service-sector employment via OLS and IV estimation. The instrument uses historical differences in industrial composition across commuting zones as a source of plausibly exogenous information with which to identify the causal effect of interest. We use the same instrument for $Empsh_{c_s,t_0}$ but make one change. The task-intensity distribution used in our cases is from 1990, whereas Autor and Dorn (2013) use the 1950 distribution.[106] The instrument has two components, both measured 15 years prior to the start of our sample period. The first is the industrial composition of commuting zones in 1990, and the second is the national structure of occupations across industries in 1990. The product of these separate components forms the instrument:

$$(3.9) \quad \widetilde{RSH}_{c_s,1990} = \sum_{i=1}^{I} (E_{i,c_s,1990} \times R_{i,-c_s,1990}$$

[105] The data for Chinese import penetration into US commuting zones are from the replication package associated with Autor et al. (2013), which is available at the following link: https://www.openicpsr.org/openicpsr/project/112670/version/V1/view.

[106] Our use of 1990 instead of 1950 is due to the time periods of our study relative to Autor and Dorn (2013). Their study covers a long time span, 1950-2005. Our sample period spans 2005 through 2019.

The variable $\widetilde{RSH}_{c_s,1990}$ is the routine employment share in commuting zone c in 1990, which is the instrument for $Empsh_{c_s,t_0}$; $E_{i,c_s,1990}$ is the employment share of industry i in commuting zone cs in 1990; and $R_{i,-c_s,1990}$ is the routine occupation share among workers in industry i across all commuting zones except commuting zone cs in 1990. When estimating the long diferences specifcations, $\widetilde{RSH}_{c_s,1990}$ enters the first-stage regression equation as defned in equation 3.9, but $\widetilde{RSH}_{c_s,1990}$ and its interaction with $D_{(t_0^{s_i},t_1^{e_i})}$ comprise the instruments in the first-stage regression equation when estimating the stacked-differences specifications.

In the context of Autor and Dorn (2013), using equation (3.9) as instrument allows them to identify the quasi-permanent component of the routine employment share's impact on different outcomes, as the instrument would affect the long-run component but likely have no relationship with short-term fluctuations in outcomes. The same logic applies to our case. One would expect the routine employment share across occupations and industries in 1990 to be a powerful predictor of the routine-, repetition-, and automation-intensive employment shares in 2005. Indeed, this is shown to be the case in Section 3.5.1.

3.4.3 Industrial Robot Exposure

To estimate the relationship between industrial robot exposure and the change in the SSDI application-to-population ratio, both OLS and IV estimation are used. The structural regression equation for the long-differences specifications is

$$(3.10) \qquad Apps^*_{c_d,(t_0,t_1)} = \gamma_0 + \gamma_1\, RobotExp_{c_d,t_0} + X'_{c_d,t_{-1}}\,\Psi + \phi_d + \epsilon_{c_d,(t_0,t_1)}$$

The subscript indexing the broader location of commuting zone c differs between the studies of Acemoglu and Restrepo (2020) from Autor and Dorn (2013). The former study identifies the effects via variation across commuting zones within US Census divisions, which is indexed in equation (3.10) with a *d*, rather than variation across commuting zones within states, which is indexed with an s. Each of the variables in equation (3.10) are defined above. The variables in $X'_{c_d,t_{-1}}$ differ from those in equations (3.7) and (3.8). In particular, we hold constant the full set of control variables from Acemoglu and Restrepo (2020), which includes US Census division dummies (i.e., ϕ_d), the natural logarithm of the population, the unemployment rate, the shares of the population who are female, Asian, Black, Hispanic, White, over age 65, did not go to college, completed some college, graduated with a college degree or professional degree, completed a master's or doctorate degree, and employed in manufacturing in general as well as light manufacturing, and the share of females employed in manufacturing relative to total manufacturing employment. Each of these variables are measured in 1990, as in Acemoglu and Restrepo (2020). The "China Shock" and the SSDI processing time controls are held constant and are defined in Section 3.4.2.

We augment equation (3.10) and estimate stacked-differences specifications analogous to those described in Section 3.4.2. In particular, we add the variable $D_{(t_0^{s_i}, t_1^{e_i})}$ to equation (3.10) and compute the stacked-differences estimates analogous to those described in equation (8) but with the right-hand-side variables shown in equation (3.10) held constant. When estimating the stacked-differences specifications, the 1990 version of the control variables held constant in equation (3.10) are linked to the first stack (either 2005-2010 or 2004-2007) and their 2000 versions are linked to the second stack (either 2010-2019 or 2013-2016).

We study the relationship between exposure to industrial robots via OLS and IV estimation. The same concerns of bias apply to robot exposure as applied to estimating the effects of the employment-share variables. As a result, we follow Acemoglu and Restrepo (2020) and Gihleb et al. (2022) and employ an IV estimation strategy that leverages the fact that European countries began adopting autonomous robots in production prior to their use in production throughout the US.

The instrument employed when using the robot-exposure measure from Gihleb et al. (2022) is defined as

$$(3.11) \qquad RobotExp_{c_d,t_0}^{IV,GGSW} = \sum_{i=1}^{I} l_{i,c_d,1970} \frac{M_{i,t_0}^{EU}}{E_{i,1990}^{EU}}$$

The variable $l_{i,c_d,1970}$ is the share of workers employed in commuting zone c_d in 1970; M_{i,t_0}^{EU} is the operation stock of autonomous industrial robots in industry i at time t_0 in nine European countries, which includes France, Denmark, Finland, Italy, Germany, Norway, Spain, Sweden, and the United Kingdom; and $E_{i,1990}^{EU}$ is employment in industry i across the aforementioned European countries in 1990. When estimating the stacked-diferences specifcations, M_{i,t_0}^{EU} is altered to $M_{i,t_0^{s_i}}$, which links the 2005 value to the starting period of the first difference (2005-2010) and the 2010 value to the starting period of the second difference (2010-2019).

When applying the measure from Acemoglu and Restrepo (2020), our instrument again relies on the previously described European countries. The measure is defined identically to the measure for the US, but for the 1994-2004 period. Formally, the instrument is

$$(3.12) \qquad RobotExp^{IV,AR}_{c,(1994,2004)} = \sum_{i=1}^{I} l_{i,c,1990} \ . APR^{EU}_{i,(1994,2004)}$$

When estimating the stacked-differences specifications, the instrument defined in equation (3.12) is included along with its interaction with $D_{\left(t_0^{s_i}, t_1^{e_i}\right)}$.

3.5 Results

We present the findings from our econometric analysis in two subsections. The first focuses on the automation-related employment shares (i.e., routine, repetition, and automation), and industrial robot exposure is the subject of the second subsection.

3.5.1 Automation-Related Employment Shares and SSDI Applications

In Table 18, we present OLS and 2SLS estimates for the relationship between the three automation-related employment share measures and the change in the SSDI application-to population ratio over the 2005-2019 period. The table includes eight columns, with the odd numbered columns containing the OLS estimates and the even-numbered columns showing the 2SLS estimates. The table is separated into three panels, each of which focuses on a particular employment-share variable. The estimates for the 18–64 age group are shown in columns 1 and 2, the 18–34 age group in columns 3 and 4, the 35–54 age group in columns 5 and 6, and the 55–64 age groups in columns 7 and 8. In the final row of the table, we provide the unconditional average of the dependent variable employed in each specification. From the table, we find little evidence of statistically significant relationships between the employment-share variables and the SSDI outcomes when using OLS estimation. In fact, we only find statistically significant effects for 35–54 age groups in Panels A and B. However, when estimating the parameters of interest with 2SLS, we find robust evidence of a statistically significant relationship between the employment-share variables and the SSDI outcomes, particularly for 35–54 and 55–64 age groups. In absolute value, the coefficients estimated via 2SLS are roughly twice as large as those estimated via OLS. The instrument is relevant in the first-stage regression, which is supported by the statistical significance of the coefficient on the instrument as well as the relatively large KP F -statistics.

In terms of interpretation, we use the standard deviations of employment-share variables to evaluate the magnitude of the effects shown in Table 18. For each employment share measure, the standard deviation is approximately three. From column 2 in Panel A, the −0.0069 coefficient when multiplied by three becomes approximately −0.021. Thus, a standard-deviation increase in the routine-intensive employment share corresponds to ap proximately a −0.021 percentage-point reduction in the application-to-population ratio for the 18–64 age group. Given that the application-to-population ratio fell by 0.083 per centage points over the 2005-2019 period, the 0.021 percentage-point reduction implied by the point estimate suggests that about 25 percent of the observed decline in the application to-population ratio for 18- to 64-year-olds could be explained by increases in the routine intensive employment share. The point estimate for the automation-intensive employment share is even larger (column 2, Panel C). The coefficient estimate when multiplied by 3 becomes −0.046, which explains 52 percent of the decline in the application-to-population ratio. The other point estimates in Table 18 can be evaluated by multiplying the coefficient estimate by 3 and then dividing by the unconditional change in the application-to-population ratio between 2005 and 2019 for each age group.

In Table 19, we repeat the analysis shown in Table 18 separately for males (odd numbered columns) and females (even-numbered columns). OLS and 2SLS estimates are presented for each age group (columns 1 and 2, 3 and 4, 5 and 6, and 7 and 8) and employment-share measure (Panels A, B, and C). Again, we find that 2SLS estimates are considerably larger than those produced when using OLS estimation. The estimates for males and females are, in large part, consistent with each other. One deviation is for the 55–64 age group. We find strong evidence of a negative effect of each employment-share variable among females, but find no evidence of a statistical link when focusing on the male application-to-population ratio.

In Table 20, we present stacked-difference estimates. The table layout is identical to that used in Table 18. Using the stacked-difference specifications, we tend to find statistically significant effects using both OLS and 2SLS. The only exception is the 18–34 age group for whom we find limited evidence of a statistically significant link between the automation related employment shares and the SSDI outcomes. Similar to the results in Table 18, the 2SLS estimates are at least twice as large as those based on OLS estimation. For the most part, the stacked-difference estimates are consistent with those based on the long-differences specifications. Using the standard deviation of the employment-share variables (≈ 3).

Table 21 examines separately the effects of the employment-share variables on males and females using the stacked-differences estimation strategy. Although there are exception, the OLS estimates tend to be statistically insignificant in the majority of cases, but 2SLS estimates are, for the most part, statistically significant. Relative to the long-difference estimates (See Table 19), the stacked-differences estimates suggest that the 35–54 age group is driving the aggregate estimates in columns 1 and 2, as we find little evidence of a statistical relationship for the 18–34 age group and relatively small effects for the 55–64 age group.

3.5.2 Exposure to Industrial Robots and SSDI Applications

We now examine how a specific technology that tends to be purely labor saving affect SSDI outcomes. In Table 22, we present overall estimates for each age group using the measure proposed by Gihleb et al. (2022). The table is organized into two panels: Panel A presents the long-differences estimates, and Panel B presents the stacked-difference estimates. OLS and 2SLS estimates are presented in each column, and the columns give estimates for males (odd numbered) and females (even numbered) in particular age groups: columns 1 and 2 (18–64), columns 3 and 4 (18–34), columns 5 and 6 (35–54), and columns 7 and 8 (55–64). We note the similarities in the coefficient estimates between the sexes as well as the stability of the estimates when using OLS and 2SLS estimation. In general, the long- and stacked differences estimates are consistent with each other, but the stacked-differences estimates tend to be smaller than those estimated via the long-differences specifications.

Using a one-unit change for evaluating the coefficient estimates is problematic due to the uncommonness of values of 1 or more in the data. Therefore, we, again, use the variables' standard deviations to assess the effect sizes, which are around 0.33 for the long differences specifications and 0.72 for the stacked-differences specifications. When we use the values to evaluate the coefficients, for example, in column 1 from Table 22, we find similarly sized effects. For example, multiplying the −0.0355 coefficient (the 2SLS estimate in Panel A) by 0.33 yields a −0.0112 percentage-point change in the application-to-population ratio. Likewise, the product of 0.72 and the −0.0138 coefficient (the 2SLS estimate in Panel B) is a −0.0099 percentage-point change in the SSDI outcome. These estimates imply that–when compared relative to the sample mean for the dependent variable (last row in Table 22), a standard-deviation increase in robots per 1000 workers explains about 12 percent of the decline in the SSDI application-to-population ratio between 2005 and 2019.

We repeat the analysis from Table 22 separately for males and females in Table 23. The estimates for males and females within each age group are similar when estimating via OLS and 2SLS. The OLS and 2SLS estimates, likewise, are similar and statistically different from zero at conventional levels. One discrepancy between the long- and stacked-differences estimates is the statistically significant negative effects shown in Panel A and null effects shown in Panel B.

Lastly, in Table 24, we use the robot-exposure measure advanced by Acemoglu and Restrepo (2020). The table presents estimates for males and females combined (Panel A), males only (Panel B), and females only (Panel C). In columns 1-4, we present the long-differences estimates, and columns 5–8 show the stacked-difference estimates. The vast majority of the estimated coefficients presented in Table 24 are statistically different from zero at conventional levels. We rely on the standard deviations of the robot-exposure measures to interpret the magnitudes of the effects. For the long-differences estimates, the standard deviation of the robot-exposure measure is 1.75, and it is 1.00 for the stacked-differences specifications. Thus, if we multiply the -0.0058 coefficient in column 1 (2SLS estimate from Panel A) by 1.75, the estimates imply a reduction in the application-to-population -0.01 percentage points. The implied impact is smaller for the stacked-differences specifications, as the point estimate (i.e., the 2SLS estimate from column 5 from Panel A)–after multiplying by 1–is about half the size of the long-differences estimates. Thus, the robot-exposure measure captures about 12 percent of the observed decline in the overall SSDI application-to-population ratio when using the long-differences estimates and about 6 percent when using the stacked-differences specification.

3.6 Discussion

While the literature focuses more on the negative effects of automation on lower-skilled workers labor market outcomes, this paper indicates a robust, negative effect of automation on SSDI claiming. Insofar as SSDI claiming rates proxy for the health of the workforce, our results are congruent with much of the literature on automation and worker health. Although further inquiry into the mechanisms underpinning the negative effect on SSDI applications is needed, it appears that the replacement of more dangerous and injury-prone tasks with machine labor (e.g., see Gihleb et al., 2022) has dominated the negative psychological effects of job destruction (e.g., see O'Brien et al., 2022). More research on different worker populations (e.g., race/ethnicity) is warranted to better identify the relationship between automation-type-technologies and workers' health outcomes.[107]

3.7 Conclusion

We study the effects of both broad and narrow measures of automation technology on disability claiming in the United States. We find that the broader measure of automation, which uses employment shares of routine tasks, is negatively related to SSDI claiming. However, the negative relationship is not as robust as we find when we examine the effect of industrial robot penetration on SSDI applications. Automation accounts for about a third of the drop in SSDI application-to-population ratio.

There are several obstacles to establishing a credible causal relationship between automation and SSDI claiming. First, omitted-variable bias is a concern. Automation technology is likely correlated with unobserved variables that also affect SSDI applications. Second, our measures of automation are imperfect, as we follow the literature and project national industry and occupational statistics related to automation onto regional labor markets called commuting zones. Third, the passage of time is required, as the effect of any automation technology requires time to spread through the economy. Lastly, the combination of the first three estimation issues could be compounded by large movements in aggregate economic activity, such as the Great Recession, which is in the middle of our sample period.[108]

We attempt to address each of these concerns. With respect to timing issues, we use long- and stacked-differences models to both allow automation time to diffuse through the labor market and minimize the impact of short-run disruptions in aggregate economic activity on estimates. To identify the causal relationship between automation and SSDI claiming, we implement a shift-share-type estimation strategy (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022). This strategy is similar to that of prominent studies in the economics automation literature, primarily the work of Autor and Dorn (2013) and Acemoglu and Restrepo (2020).

[107] Further analysis of SSDI claiming rates for racial/ethnic sub-populations may require aggregation to larger geographic areas, as many commuting zone observations by race were censored due to privacy concerns.
[108] We do not consider the COVID-19 period in our data.

Recent research on technological change, within a task framework of production (e.g., see Acemoglu et al., 2024), shows that automation shifts tasks away from certain groups of workers, which also affects the wage distribution. The confluence of these two labor-market changes makes it more difficult to identify a specific mechanism that would affect SSDI claiming–that is, does automation affect SSDI by replacing dangerous tasks and/or by changing the wage structure? By using two measures of automation, one broad and one specific to industrial robots, we hoped to gain insight into the underlying mechanisms driving the reduction in SSDI claiming. In future research, we plan to investigate the extent to which automation has replaced human labor in more dangerous and/or physical tasks versus the destruction of jobs that might incentivize SSDI applications.

Table 16: Top 10 Detailed Occupations for the Automation-Related Employment Shares

	Routine-Intensive	Repetition-Intensive	Automation-Intensive
Ranking	(1)	(2)	(3)
#1	Bookkeeping, Accounting, and Auditing Clerks	Bookkeeping, Accounting, and Auditing Clerks	Extruding, Forming, Pressing, and Compacting Machine Setters, Operators, and Tenders
#2	Insurance Claims and Policy Processing Clerks	Gambling Cage Workers	Travel Agents
#3	Eligibility Interviewers, Government Programs	Radiation Therapists	Library Technicians
#4	Medical Records Specialists	Reservation and Transportation Ticket Agents and Travel Clerks	Insurance Claims and Policy Processing Clerks
#5	Health Information Technologists and Medical Registrars	Eligibility Interviewers, Government Programs	Accountants and Auditors
#6	Tire Builders	Payroll and Timekeeping Clerks	Budget Analysts
#7	Loan Interviewers and Clerks	Brokerage Clerks	Sawing Machine Setters, Operators, and Tenders, Wood
#8	Budget Analysts	Tax Preparers	Bookkeeping, Accounting, and Auditing Clerks
#9	Atmospheric and Space Scientists	Tellers	Financial and Investment Analysis
#10	Library Technicians	Pharmacists	Financial Risk Specialists

Notes: The table presents the top 10 occupations for the routine-, repetition-, and automation-intensive employment shares generally defined in equation (2).

Table 17: Summary Statistics for SSDI Applications by Age Group

		Standard Deviation		
	Average	Overall	Between	Within
	(1)	(2)	(3)	(4)
Panel A: Males and Females				
18-64 Year Olds	0.6098	0.2089	0.1921	0.0837
18-34 Year Olds	0.1035	0.0424	0.0395	0.0215
35-54 Year Olds	0.2988	0.1227	0.1118	0.0551
55-64 Year Olds	0.1971	0.0594	0.0565	0.0273
Panel B: Males				
18-64 Year Olds	0.3229	0.1158	0.1038	0.0530
18-34 Year Olds	0.0523	0.0233	0.0214	0.0130
35-54 Year Olds	0.1531	0.0663	0.0581	0.0333
55-64 Year Olds	0.1106	0.0343	0.0313	0.0168
Panel C: Females				
18-64 Year Olds	0.3003	0.0992	0.0931	0.0373
18-34 Year Olds	0.0538	0.0212	0.0205	0.0105
35-54 Year Olds	0.1521	0.0591	0.0547	0.0256
55-64 Year Olds	0.0885	0.0275	0.0254	0.0138

Notes: The table presents sample means, standard deviations, and the within and between components of the standard deviations for the four dependent variables. The sample sizes vary in Panel A: 9,755 observations for the 18-64 year olds; 9,846 observations for the 18-34 year olds; 10,685 observations for the 35-54 year olds; and 10,586 observations for 55-64 year olds. The sample sizes vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board.

Table 18: Routine, Repetition, and Automation Intensive Employment Shares and the SSDI Application-to-Population Ratio: OLS and 2SLS Estimates, Long Differences, 2005-2019

	Age Group							
	18-64		18-34		35-54		55-64	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Routine-Intensive Employment Share								
$RSH_{c,2005}$	−0.0018	−0.0069***	0.0004	0.0008	−0.0021**	−0.0055***	−0.0001	−0.0020***
	(0.0013)	(0.0026)	(0.0004)	(0.0007)	(0.0009)	(0.0018)	(0.0004)	(0.0008)
First Stage Coeff.	–	0.6016***	–	0.6016***	–	0.6062***	–	0.6069***
KP F-Statistic	–	65.1875	–	65.2252	–	67.0934	–	66.9569
Panel B: Repetition-Intensive Employment Share								
$RTSH_{c,2005}$	−0.0009	−0.0077***	0.0006	0.0009	−0.0020**	−0.0061***	0.0005	−0.0023***
	(0.0012)	(0.0029)	(0.0004)	(0.0008)	(0.0009)	(0.0020)	(0.0004)	(0.0008)
First Stage Coeff.	–	0.5432***	–	0.5432***	–	0.5459***	–	0.5464***
KP F-Statistic	–	76.3211	–	76.3632	–	79.6606	–	79.3729
Panel C: Automation-Intensive Employment Share								
$ASH_{c,2005}$	−0.0017	−0.0142***	0.0002	0.0016	−0.0012	−0.0116***	−0.0006	−0.0044***
	(0.0015)	(0.0052)	(0.0004)	(0.0016)	(0.0009)	(0.0039)	(0.0005)	(0.0015)
First Stage Coeff.	–	0.2938***	–	0.2938***	–	0.2853***	–	0.2847***
KP F-Statistic	–	45.4623	–	45.4796	–	45.7129	–	44.8765
N	623	623	626	626	687	687	679	679
$\overline{Apps^{*}_{c,(2005,2019)}}$	−0.0826	−0.0826	−0.0038	−0.0038	−0.0931	−0.0931	0.0151	0.0151

Notes: The table presents OLS and 2SLS long-differences estimates based on equation (7). For each age group, we present the OLS and 2SLS estimates side-by-side separately for 18-64, 18-34, 35-54, and 55-64 year-olds. For the specifications estimated via 2SLS, we report the coefficient on the instrument (defined in equation 9) and the Kleibergen-Paap Wald rk F-statistic. The sample sizes, N, vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board. Each specification includes the control variables held constant in Autor and Dorn (2013)'s study, which include the college-to-noncollege population, the ratio of immigrants to the noncollege population, the share of workers employed in manufacturing, the unemployment rate, the share of the population who is female and employed, the share of the population 65 or older, and the share of noncollege workers workers earning a real wage below the minimum wage that will prevail over the next decade (2000-2010). Each of these variables is measured in 2000. We also hold constant the "China Shock" from (Autor et al., 2013), which is measured as the change in Chinese import exposure between 1990 and 2007 and the SSDI processing efficiency measure from Kearney et al. (2021) for the year 2003. We report standard errors clustered at the state level in parentheses. *, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 19: Routine, Repetition, and Automation Intensive Employment Shares and Per-capita SSDI Applications: Males and Females Separately, OLS and 2SLS Estimates, Long Differences, 2005-2019

	Age Group							
	18-64		18-34		35-54		55-64	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Routine-Intensive Employment Share								
OLS	-0.0010	-0.0007	0.0003^{*}	0.0001	-0.0013^{**}	-0.0008^{**}	-0.0001	-0.0001
	(0.0008)	(0.0007)	(0.0002)	(0.0002)	(0.0006)	(0.0004)	(0.0002)	(0.0003)
2SLS	-0.0030^{**}	-0.0039^{***}	0.0005	0.0002	-0.0029^{***}	-0.0026^{***}	-0.0005	-0.0016^{***}
	(0.0014)	(0.0015)	(0.0004)	(0.0005)	(0.0010)	(0.0009)	(0.0004)	(0.0005)
First Stage Coeff.	0.6015^{***}	0.6015^{***}	0.6016^{***}	0.6016^{***}	0.6030^{***}	0.6030^{***}	0.6014^{***}	0.6014^{***}
KP F-Statistic	63.3168	63.3168	63.3654	63.3654	65.5689	65.5689	65.5885	65.5885
Panel B: Repetition-Intensive Employment Share								
OLS	-0.0006	-0.0002	0.0004^{**}	0.0002	-0.0013^{**}	-0.0007^{*}	0.0003	0.0002
	(0.0008)	(0.0007)	(0.0002)	(0.0003)	(0.0006)	(0.0004)	(0.0002)	(0.0002)
2SLS	-0.0033^{**}	-0.0043^{***}	0.0005	0.0003	-0.0033^{***}	-0.0029^{***}	-0.0006	-0.0017^{***}
	(0.0016)	(0.0016)	(0.0004)	(0.0005)	(0.0011)	(0.0010)	(0.0004)	(0.0006)
First Stage Coeff.	0.5417^{***}	0.5417^{***}	0.5418^{***}	0.5418^{***}	0.5442^{***}	0.5442^{***}	0.5431^{***}	0.5431^{***}
KP F-Statistic	73.1754	73.1754	73.2304	73.2304	77.2897	77.2897	77.1331	77.1331
Panel C: Automation-Intensive Employment Share								
OLS	-0.0003	-0.0014^{*}	0.0002	-0.0001	-0.0005	-0.0007	0.0000	-0.0006^{**}
	(0.0008)	(0.0008)	(0.0002)	(0.0002)	(0.0005)	(0.0005)	(0.0003)	(0.0003)
2SLS	-0.0060^{**}	-0.0078^{***}	0.0010	0.0005	-0.0061^{***}	-0.0054^{***}	-0.0011	-0.0032^{***}
	(0.0029)	(0.0026)	(0.0009)	(0.0009)	(0.0023)	(0.0017)	(0.0008)	(0.0009)
First Stage Coeff.	0.3014^{***}	0.3014^{***}	0.3014^{***}	0.3014^{***}	0.2903^{***}	0.2903^{***}	0.2918^{***}	0.2918^{***}
KP F-Statistic	45.6387	45.6387	45.6574	45.6574	45.5121	45.5121	46.9524	46.9524
N	559	559	562	562	639	639	621	621
$\overline{Apps^{*}_{s,c,(2005,2019)}}$	-0.0664	-0.0029	-0.0593	-0.0012	-0.0207	-0.0016	-0.0353	0.0175

Notes: The table presents OLS and 2SLS long-differences estimates based on equation (7) separately for males (odd-numbered columns) and females (even-numbered columns). The sample sizes, N, vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board. Each specification includes the control variables held constant in Autor and Dorn (2013)'s study, which include the college-to-noncollege population, the ratio of immigrants to the noncollege population, the share of workers employed in manufacturing, the unemployment rate, the share of the population who is female and employed, the share of the population 65 or older, and the share of noncollege workers workers earning a real wage below the minimum wage that will prevail over the next decade (2000-2010). Each of these variables is measured in 2000. We also hold constant the "China Shock" from (Autor et al., 2013), which is measured as the change in Chinese import exposure between 1990 and 2007 and the SSDI processing efficiency measure from Kearney et al. (2021) for the year 2003. We report standard errors clustered at the state level in parentheses. *, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 20: Routine, Repetition, and Automation Intensive Employment Shares and Per-capita SSDI Applications: OLS and 2SLS Estimates, Stacked Differences, 2005-2010 and 2010-2019

	Age Group							
	18-64		18-34		35-54		55-64	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Routine-Intensive Employment Share								
$RSH_{c,2005}$	-0.0035^{**}	-0.0083^{*}	-0.0002	-0.0011	-0.0026^{**}	-0.0055^{**}	-0.0008^{**}	-0.0017^{**}
	(0.0016)	(0.0044)	(0.0004)	(0.0010)	(0.0010)	(0.0028)	(0.0004)	(0.0008)
First Stage Estimates								
Main Effect Coeff.	–	0.5804^{***}	–	0.5804^{***}	–	0.5845^{***}	–	0.5840^{***}
Coeff. on Interaction Term	–	-0.1124^{**}	–	-0.1124^{**}	–	-0.1094^{**}	–	-0.1093^{**}
KP F-Statistic	–	21.9941	–	21.9968	–	22.5637	–	22.4685
Panel B: Repetition-Intensive Employment Share								
$RTSH_{c,2005}$	-0.0041^{**}	-0.0099^{*}	-0.0004	-0.0021^{*}	-0.0030^{***}	-0.0062^{*}	-0.0008^{**}	-0.0016^{*}
	(0.0016)	(0.0052)	(0.0004)	(0.0013)	(0.0010)	(0.0032)	(0.0004)	(0.0009)
First Stage Estimates								
Main Effect Coeff.	–	0.5424^{***}	–	0.5424^{***}	–	0.5453^{***}	–	0.5449^{***}
Coeff. on Interaction Term	–	-0.1812^{***}	–	-0.1812^{***}	–	-0.1770^{***}	–	-0.1771^{***}
KP F-Statistic	–	30.5157	–	30.5187	–	31.6373	–	31.4715
Panel C: Automation-Intensive Employment Share								
$ASH_{c,2005}$	0.0003	-0.0153^{**}	0.0004	0.0003	0.0007	-0.0120^{**}	-0.0008^{*}	-0.0043^{***}
	(0.0015)	(0.0076)	(0.0004)	(0.0017)	(0.0009)	(0.0053)	(0.0005)	(0.0015)
First Stage Estimates								
Main Effect Coeff.	–	0.2343^{***}	–	0.2343^{***}	–	0.2295^{***}	–	0.2304^{***}
Coeff. on Interaction Term	–	0.0181	–	0.0181	–	0.0174	–	0.0171
KP F-Statistic	–	15.1533	–	15.1554	–	14.8478	–	14.9314
N	1,266	1,266	1,268	1,268	1,388	1,388	1,361	1,361
$\overline{Apps^{*}_{c,(2005,2019)}}$	-0.0826	-0.0826	-0.0038	-0.0038	-0.0931	-0.0931	0.0151	0.0151

Notes: The table presents OLS and 2SLS stacked-differences estimates based on equation (8). For each age grouping, we present the OLS and 2SLS estimates side-by-side separately for 18-64, 18-34, 35-54, and 55-64 year-olds. For the specifications estimated via 2SLS, we report the coefficient on the instrument (defined in equation 9) and the Kleibergen-Paap Wald rk F-statistic. The sample sizes, N, vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board. Each specification includes the control variables held constant in Autor and Dorn (2013)'s study, which include the college-to-noncollege population, the ratio of immigrants to the noncollege population, the share of workers employed in manufacturing, the unemployment rate, the share of the population who is female and employed, the share of the population 65 or older, and the share of noncollege workers workers earning a real wage below the minimum wage that will prevail over the next decade. We link the 1990 version of each of these variables to the starting year of the first difference in the stack (i.e. 2005), and the 2000 version of these variables is linked to the starting period of the second difference in the stack (i.e. 2010). For the share of workers earning below the minimum wage that will prevail in the decade, the relevant decade for the starting period of the first difference is 1990-2000 and the decade relevant to the starting period of the second difference is 2000-2010. We hold constant the "China Shock" from (Autor et al., 2013), which is measured as the change in Chinese import exposure between two points in time. We link the change between 1990 and 2000 to the starting period of the first difference and the change between 2000 and 2007 to the starting year of the second difference. We report standard errors clustered at the state level in parentheses. Lastly, we hold constant the SSDI processing efficiency measure. The 2003 value is linked to the first difference in the stack, and the 2008 value is linked to the second difference in the stack. *, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 21: Routine, Repetition, and Automation Intensive Employment Shares and Per-capita SSDI Applications: Males and Females Separately, OLS and 2SLS Estimates, Stacked Differences, 2005-2010 and 2010-2019

	Age Group							
	18-64		18-34		35-54		55-64	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Routine-Intensive Employment Share				
OLS	−0.0025**	−0.0010	−0.0002	−0.0000	−0.0017***	−0.0009**	−0.0006**	−0.0002
	(0.0010)	(0.0007)	(0.0002)	(0.0002)	(0.0006)	(0.0004)	(0.0003)	(0.0002)
2SLS	−0.0051*	−0.0032*	−0.0008	−0.0004	−0.0032**	−0.0023*	−0.0011**	−0.0006*
	(0.0026)	(0.0019)	(0.0006)	(0.0005)	(0.0016)	(0.0012)	(0.0005)	(0.0003)
First Stage Estimates								
Main Effect Coeff.	0.5765***	0.5765***	0.5765***	0.5765***	0.5817***	0.5817***	0.5795***	0.5795***
Coeff. on Interaction Term	−0.1125**	−0.1125**	−0.1125**	−0.1125**	−0.1116**	−0.1116**	−0.1100**	−0.1100**
KP F-Statistic	21.1402	21.1402	21.1437	21.1437	22.1461	22.1461	22.0142	22.0142
				Panel B: Repetition-Intensive Employment Share				
OLS	−0.0031***	−0.0009	−0.0003	−0.0000	−0.0020***	−0.0009**	−0.0008***	−0.0000
	(0.0009)	(0.0007)	(0.0002)	(0.0002)	(0.0006)	(0.0004)	(0.0002)	(0.0002)
2SLS	−0.0069**	−0.0033	−0.0016**	−0.0007	−0.0037**	−0.0024*	−0.0014**	−0.0002
	(0.0032)	(0.0022)	(0.0008)	(0.0006)	(0.0018)	(0.0014)	(0.0007)	(0.0004)
First Stage Estimates								
Main Effect Coeff.	0.5384***	0.5384***	0.5384***	0.5384***	0.5433***	0.5433***	0.5417***	0.5417***
Coeff. on Interaction Term	−0.1818***	−0.1818***	−0.1818***	−0.1818***	−0.1799***	−0.1799***	−0.1788***	−0.1788***
KP F-Statistic	29.2231	29.2231	29.2270	29.2270	30.8400	30.8400	30.7101	30.7101
				Panel C: Automation-Intensive Employment Share				
OLS	0.0002	0.0001	0.0002	0.0002	0.0003	0.0003	−0.0003	−0.0005**
	(0.0010)	(0.0006)	(0.0002)	(0.0002)	(0.0006)	(0.0004)	(0.0003)	(0.0002)
2SLS	−0.0070	−0.0074**	0.0004	−0.0001	−0.0063**	−0.0053**	−0.0017*	−0.0025***
	(0.0044)	(0.0032)	(0.0009)	(0.0008)	(0.0031)	(0.0022)	(0.0010)	(0.0007)
First Stage Estimates								
Main Effect Coeff.	0.2420***	0.2420***	0.2420***	0.2420***	0.2312***	0.2312***	0.2343***	0.2343***
Coeff. on Interaction Term	0.0167	0.0167	0.0167	0.0167	0.0189	0.0189	0.0164	0.0164
KP F-Statistic	14.8690	14.8690	14.8708	14.8708	14.9907	14.9907	15.1521	15.1521
N	1,137	1,137	1,139	1,139	1,301	1,301	1,256	1,256
$\overline{Apps^{*}_{s,c,(2005,2019)}}$	−0.0664	−0.0029	−0.0593	−0.0012	−0.0207	−0.0016	−0.0353	0.0175

Notes: The table presents OLS and 2SLS stacked-differences estimates based on equation (8) separately for males (odd-numbered columns) and females (even-numbered columns). For the 2SLS estimates, we report the coefficient on the instrument (defined in equation 9), its interaction with the difference-period indicator (see equation (7), and the Kleibergen-Paap Wald rk F-statistic. The sample sizes, N, vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board. Each specification includes the control variables held constant in Autor and Dorn (2013)'s study, which include the college-to-noncollege population, the ratio of immigrants to the noncollege population, the share of workers employed in manufacturing, the unemployment rate, the share of the population who is female and employed, the share of the population 65 or older, and the share of noncollege workers workers earning a real wage below the minimum wage that will prevail over the next decade. We link the 1990 version of each of these variables to the starting year of the first difference in the stack (i.e. 2005), and the 2000 version of these variables is linked to the starting period of the second difference in the stack (i.e. 2010). For the share of workers earning below the minimum wage that will prevail in the decade, the relevant decade for the starting period of the first difference is 1990-2000 and the decade relevant to the starting period of the second difference is 2000-2010. We hold constant the "China Shock" from (Autor et al., 2013), which is measured as the change in Chinese import exposure between two points in time. We link the change between 1990 and 2000 to the starting period of the first difference and the change between 2000 and 2007 to the starting year of the second difference. We report standard errors clustered at the state level in parentheses. Lastly, we hold constant the SSDI processing efficiency measure. The 2003 value is linked to the first difference in the stack, and the 2008 value is linked to the second difference in the stack. *, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 22: Exposure to Industrial Robots and Per-capita SSDI Applications: OLS and 2SLS Estimates, Long Differences (2005-2019) and Stacked Differences (2005-2010, 2010-2019)

	Age Group			
	18-64	18-34	35-54	55-64
	(1)	(2)	(3)	(4)
		Panel A: Long Differences Estimates		
OLS	-0.0389^{***}	-0.0085^{***}	-0.0215^{***}	-0.0089^{**}
	(0.0080)	(0.0016)	(0.0040)	(0.0038)
2SLS	-0.0355^{***}	-0.0066^{***}	-0.0224^{***}	-0.0065
	(0.0092)	(0.0015)	(0.0038)	(0.0051)
First Stage Estimates				
First Stage Coeff.	0.1776^{***}	0.1776^{***}	0.1775^{***}	0.1775^{***}
KP F-Statistic	1,247.8025	1,247.7057	1,236.5783	1,237.8741
N	623	628	689	681
		Panel B: Stacked Differences Estimates		
OLS	-0.0169^{***}	-0.0005	-0.0122^{***}	-0.0042^{**}
	(0.0034)	(0.0010)	(0.0020)	(0.0019)
2SLS	-0.0138^{***}	0.0010	-0.0109^{***}	-0.0039
	(0.0050)	(0.0007)	(0.0026)	(0.0025)
First Stage Estimates				
Main Effect	0.1626^{***}	0.1626^{***}	0.1625^{***}	0.1626^{***}
Coeff. on Interaction	0.2850^{***}	0.2850^{***}	0.2850^{***}	0.2850^{***}
KP F-Statistic	4,626.7748	4,626.7956	4,617.6992	4,629.3804
N	1,266	1,268	1,388	1,361
$\overline{Apps^{*}_{c,(2005,2019)}}$	-0.0826	-0.0038	-0.0931	0.0151

Notes: The table presents OLS and 2SLS long-differences estimates based on equation (10) in Panel A, and OLS and 2SLS stacked-differences estimates based on equation (8 are presented in Panel B. For each age grouping, we present the OLS and 2SLS estimates side-by-side separately for 18-64, 18-34, 35-54, and 55-64 year-olds. For the specifications estimates via 2SLS, we report the coefficient on the instrument (defined in equation (11)) and the Kleibergen-Paap Wald rk F-statistic. The sample sizes, N, vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board. Each specification includes the control variables held constant in Acemoglu and Restrepo (2020)'s study, which includes US Census division dummies (i.e. ϕ_d), the natural logarithm of the population, the "China Shock" from Autor et al. (2013), the shares of the population who are female, Asian, Black, Hispanic, White, over age 65, did not go to college, completed some college, graduated with a college degree or professional degree, completed a masters or doctorate degree, and employed in manufacturing in general as well as light manufacturing, and the share of females employed in manufacturing relative to total manufacturing employment. When estimating the long-differences specifications, the 1990 version of these variables is used. By contrast, when estimating the stacked-differences specifications, we link the 1990 and 2000 versions of these variables to the first and second periods in the stack, respectively. We also hold constant the "China Shock" from (Autor et al., 2013) and the SSDI processing efficiency variable from Kearney et al. (2021). See the notes from Tables 3 and 5 for details on the Chinese import penetration and SSDI processing efficiency controls. We account for these variables in the same way that we do when estimating the models focused on the employment-share automation measures. We report standard errors clustered at the state level in parentheses. *, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 23: Exposure to Industrial Robots and Per-capita SSDI Applications: Males and Females Separately, OLS and 2SLS Estimates, Long Differences (2005-2019) and Stacked Differences (2005-2010, 2010-2019)

	Age Group							
	18-64		18-34		35-54		55-64	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Long Differences Estimates								
OLS	-0.0213^{***}	-0.0180^{***}	-0.0044^{***}	-0.0044^{***}	-0.0118^{***}	-0.0098^{***}	-0.0049^{*}	-0.0040^{***}
	(0.0059)	(0.0032)	(0.0010)	(0.0009)	(0.0027)	(0.0018)	(0.0028)	(0.0013)
2SLS	-0.0213^{***}	-0.0146^{***}	-0.0039^{***}	-0.0029^{***}	-0.0134^{***}	-0.0093^{***}	-0.0037	-0.0027
	(0.0063)	(0.0035)	(0.0011)	(0.0007)	(0.0023)	(0.0019)	(0.0036)	(0.0017)
First Stage Estimates								
First Stage Coeff.	0.1776^{***}	0.1776^{***}	0.1776^{***}	0.1776^{***}	0.1776^{***}	0.1776^{***}	0.1776^{***}	0.1776^{***}
KP F-Statistic	1,243.1879	1,243.1879	1,243.1220	1,243.1220	1,245.8722	1,245.8722	1,244.1930	1,244.1930
N	561	561	564	564	641	641	623	623
Panel B: Stacked Differences Estimates								
OLS	-0.0085^{***}	-0.0082^{***}	0.0001	-0.0006	-0.0065^{***}	-0.0057^{***}	-0.0023^{*}	-0.0019^{**}
	(0.0023)	(0.0013)	(0.0005)	(0.0006)	(0.0012)	(0.0008)	(0.0012)	(0.0008)
2SLS	-0.0068^{**}	-0.0067^{***}	0.0009^{*}	0.0002	-0.0059^{***}	-0.0050^{***}	-0.0020	-0.0019^{*}
	(0.0033)	(0.0019)	(0.0005)	(0.0004)	(0.0016)	(0.0011)	(0.0016)	(0.0010)
First Stage Estimates								
Main Effect	0.1625^{***}	0.1625^{***}	0.1625^{***}	0.1625^{***}	0.1626^{***}	0.1626^{***}	0.1626^{***}	0.1626^{***}
Coeff. on Interaction	0.2851^{***}	0.2851^{***}	0.2851^{***}	0.2851^{***}	0.2850^{***}	0.2850^{***}	0.2850^{***}	0.2850^{***}
KP F-Statistic	4,634.0076	4,634.0076	4,634.0831	4,634.0831	4,633.5938	4,633.5938	4,624.9458	4,624.9458
N	1,137	1,137	1,139	1,139	1,301	1,301	1,256	1,256
$\overline{Apps^{*}_{s,c,(2005,2019)}}$	-0.0664	-0.0029	-0.0593	-0.0012	-0.0207	-0.0016	-0.0353	0.0175

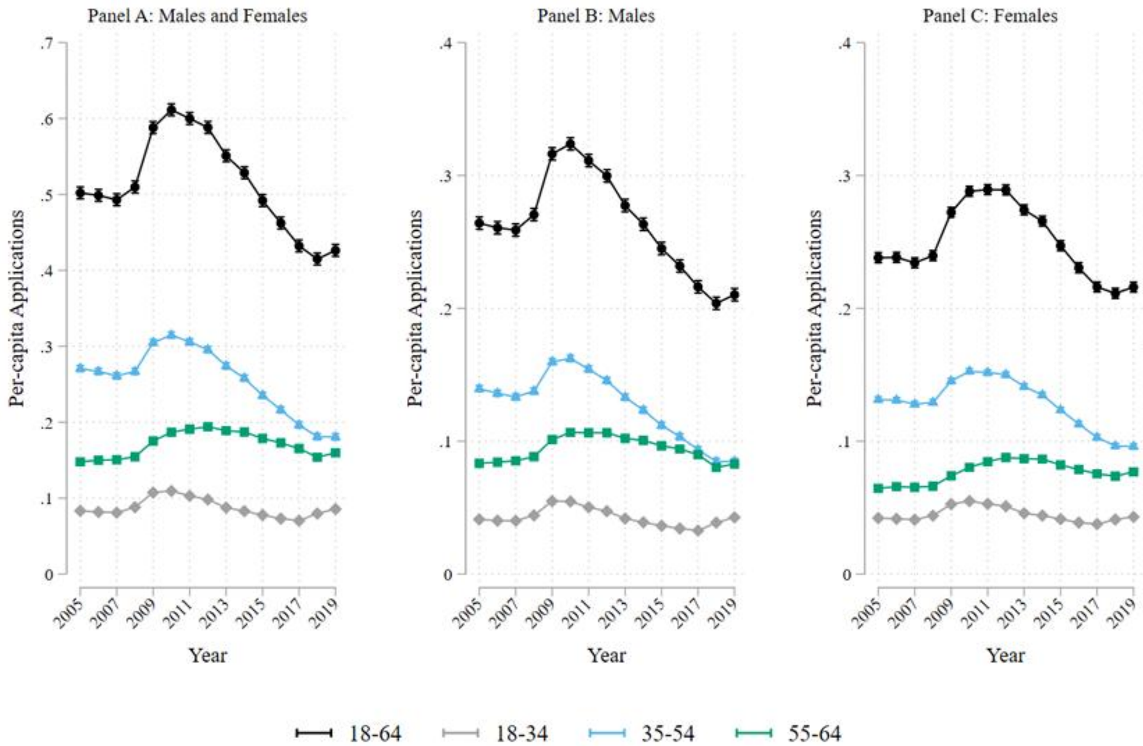
Notes: Separately for males (odd-numbered columns) and females (even-numbered columns), the table presents OLS and 2SLS long-differences estimates based on equation (10) in Panel A, and OLS and 2SLS stacked-differences estimates described in Section 4.3 in Panel B. For each age grouping, we present the OLS and 2SLS estimates side-by-side separately for 18-64, 18-34, 35-54, and 55-64 year-olds. For the specifications estimated via 2SLS, we report the coefficient on the instrument (defined in equation (11) and the Kleibergen-Paap Wald rk F-statistic in Panel A and also report the coefficient on the interaction effect between the instrument and the difference-period indicator variable. he sample sizes, N, vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board. Each specification includes the control variables held constant in Acemoglu and Restrepo (2020)'s study, which includes US Census division dummies (i.e. ϕ_d), the natural logarithm of the population, the "China Shock" from Autor et al. (2013), the shares of the population who are female, Asian, Black, Hispanic, White, over age 65, did not go to college, completed some college, graduated with a college degree or professional degree, completed a masters or doctorate degree, and employed in manufacturing in general as well as light manufacturing, and the share of females employed in manufacturing relative to total manufacturing employment. When estimating the long-differences specifications, the 1990 version of these variables is used. By contrast, when estimating the stacked-differences specifications, we link the 1990 and 2000 versions of these variables to the first and second periods in the stack, respectively. We also hold constant the "China Shock" from (Autor et al., 2013) and the SSDI processing efficiency variable from Kearney et al. (2021). See the notes from Tables 3 and 5 for details on the Chinese import penetration and SSDI processing efficiency controls. We account for these variables in the same way that we do when estimating the models focused on the employment-share automation measures. We report standard errors clustered at the state level in parentheses. *, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 24: Exposure to Industrial Robots and Per-capita SSDI Applications, OLS and 2SLS Estimates, Long Differences (2004-2016) and Stacked Differences (2004-2007, 2013-2016)

	Long Differences, 2004-2016				Stacked Differences, 2004-2007 and 2013-2016			
	18-64	18-34	35-54	55-64	18-64	18-34	35-54	55-64
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Males and Females								
OLS	-0.0051^{***}	-0.0015^{***}	-0.0020^{**}	-0.0016^{**}	-0.0045^{***}	-0.0008^{***}	-0.0023^{***}	-0.0013^{*}
	(0.0017)	(0.0005)	(0.0009)	(0.0006)	(0.0013)	(0.0003)	(0.0007)	(0.0007)
2SLS	-0.0058^{***}	-0.0014^{***}	-0.0027^{***}	-0.0016^{***}	-0.0046^{***}	-0.0006^{**}	-0.0022^{***}	-0.0017^{***}
	(0.0016)	(0.0005)	(0.0008)	(0.0006)	(0.0015)	(0.0003)	(0.0008)	(0.0006)
First Stage Coeff.	1.3422^{***}	1.3422^{***}	1.3431^{***}	1.3429^{***}	0.6476^{***}	0.6476^{***}	0.6481^{***}	0.6480^{***}
Coeff. on Interaction	–	–	–	–	0.0974^{***}	0.0974^{***}	0.0974^{***}	0.0974^{***}
KP F-Statistic	2,293.2306	2,293.7997	2,329.0867	2,323.7103	4,271.1866	4,271.8920	4,351.5636	4,318.4567
N	625	627	692	685	1,246	1,248	1,381	1,362
Panel B: Males								
OLS	-0.0030^{***}	-0.0010^{***}	-0.0011^{*}	-0.0008^{*}	-0.0014	-0.0003^{*}	-0.0012^{**}	0.0000
	(0.0011)	(0.0002)	(0.0006)	(0.0004)	(0.0010)	(0.0002)	(0.0005)	(0.0005)
2SLS	-0.0037^{***}	-0.0011^{***}	-0.0016^{***}	-0.0009^{**}	-0.0016	-0.0003^{*}	-0.0012^{**}	-0.0002
	(0.0009)	(0.0002)	(0.0005)	(0.0004)	(0.0010)	(0.0001)	(0.0005)	(0.0005)
First Stage Coeff.	1.3408^{***}	1.3408^{***}	1.3427^{***}	1.3423^{***}	0.6472^{***}	0.6472^{***}	0.6479^{***}	0.6477^{***}
Coeff. on Interaction	–	–	–	–	0.0967^{***}	0.0967^{***}	0.0975^{***}	0.0973^{***}
KP F-Statistic	2,240.2580	2,240.9812	2,306.9613	2,295.4873	4,217.8029	4,218.6427	4,292.4579	4,258.9084
N	557	559	644	630	1,119	1,121	1,293	1,254
Panel C: Females								
OLS	-0.0022^{***}	-0.0005^{*}	-0.0009^{**}	-0.0008^{***}	-0.0029^{***}	-0.0005^{**}	-0.0011^{***}	-0.0013^{***}
	(0.0007)	(0.0003)	(0.0004)	(0.0002)	(0.0005)	(0.0002)	(0.0003)	(0.0003)
2SLS	-0.0022^{***}	-0.0003	-0.0011^{***}	-0.0008^{***}	-0.0029^{***}	-0.0003^{*}	-0.0010^{**}	-0.0016^{***}
	(0.0007)	(0.0003)	(0.0004)	(0.0002)	(0.0006)	(0.0002)	(0.0004)	(0.0002)
First Stage Coeff.	1.3408^{***}	1.3408^{***}	1.3427^{***}	1.3423^{***}	0.6472^{***}	0.6472^{***}	0.6479^{***}	0.6477^{***}
Coeff. on Interaction	–	–	–	–	0.0967^{***}	0.0967^{***}	0.0975^{***}	0.0973^{***}
KP F-Statistic	2,240.2580	2,240.9812	2,306.9613	2,295.4873	4,217.8029	4,218.6427	4,292.4579	4,258.9084
N	557	559	644	630	1,119	1,121	1,293	1,254

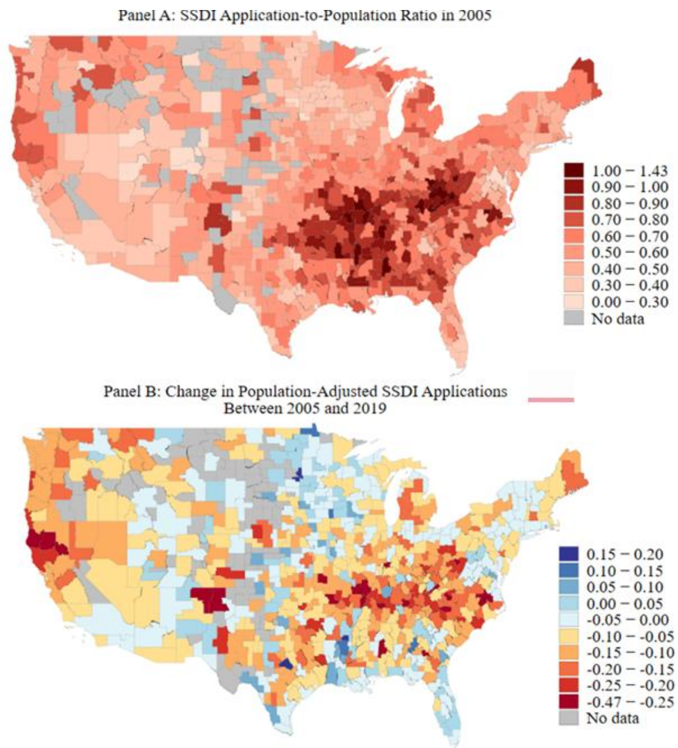
Notes: The table presents OLS and 2SLS long-differences estimates based on equation (10) in columns 1-4, and OLS and 2SLS stacked-differences estimates described in Section 4.3 are presented in columns 5-8. Estimates are presented for males and females combined in Panel A; males only in Panel B; and females only in Panel C. For the specifications estimated via 2SLS, we report the coefficient on the instrument (defined in equation (12) and the Kleibergen-Paap Wald rk F-statistic in Panel A and also report the coefficient on the interaction effect between the instrument and the difference-period indicator variable. he sample sizes, N, vary across the age groups due to the suppression of data from commuting zones with application counts below the threshold for external presentation determined by SSA's Disclosure Review Board. Each specification includes the control variables held constant in Acemoglu and Restrepo (2020)'s study, which includes US Census division dummies (i.e. ϕ_d), the natural logarithm of the population, the "China Shock" from Autor et al. (2013), the shares of the population who are female, Asian, Black, Hispanic, White, over age 65, did not go to college, completed some college, graduated with a college degree or professional degree, completed a masters or doctorate degree, and employed in manufacturing in general as well as light manufacturing, and the share of females employed in manufacturing relative to total manufacturing employment. When estimating the long-differences specifications, the 1990 version of these variables is used. By contrast, when estimating the stacked-differences specifications, we link the 1990 and 2000 versions of these variables to the first and second periods in the stack, respectively. We also hold constant the "China Shock" from (Autor et al., 2013) and the SSDI processing efficiency variable from Kearney et al. (2021). See the notes from Tables 3 and 5 for details on the Chinese import penetration and SSDI processing efficiency controls. We account for these variables in the same way that we do when estimating the models focused on the employment-share automation measures. We report standard errors clustered at the state level in parentheses. *, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.

Figure 14: SSDI Application-to-Population Ratio, 2005-2019



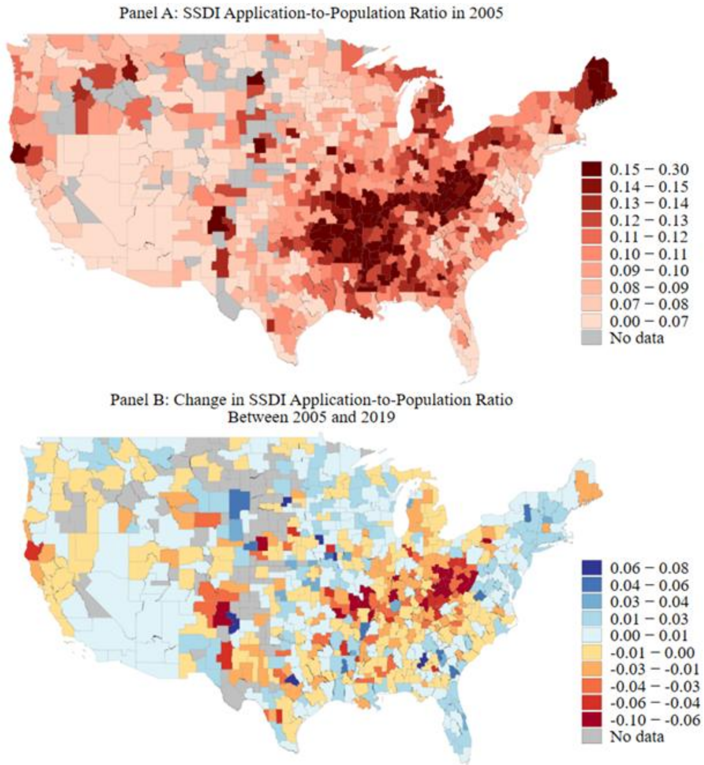
Notes: For the 2005-2019 period, the figure plots predicted values from a regression of SSDI application-to-population ratio, as defined in equation 1, in the commuting zone on a constant, year dummies, and state dummies. The regression estimates are weighted by the commuting zone's share of the US population in 2000. The year-by-year estimates are presented separately in Panels A (males and females combined), B (males only), and C (females only) for the 18–64, 18–34, 35–54, and 55–64 age groups.

Figure 15: Geographic Variation in Population-Adjusted SSDI Applications, 18–64 Year Olds



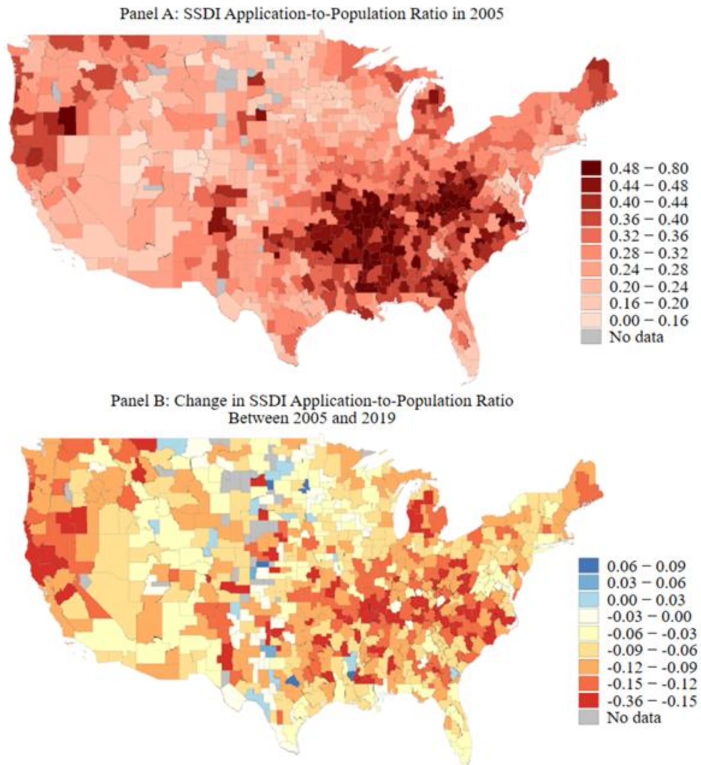
Notes: For 18–64 year-olds, the figure presents heat maps of the SSDI application-to-population ratio, as defined in equation 1, in 2005 (Panel A) as well as its change between 2005 and 2019 (Panel B). In Panel A, lighter red colors indicate a lower application-to-population ratios, and darker red colors indicate a greater prevalence of per-capita applications. Commuting zones for which insufficient applications were submitted to allow for external presentation are shown in gray. In Panel B, we use a diverging color scheme, in which light to dark blue indicate positive changes in application rates, and yellow to red indicate negative changes in application rates. The positive changes are greater when the color is a darker blue. Likewise, negative changes are greater when the color is a darker red.

Figure 16: Geographic Variation in Population-Adjusted SSDI Applications, 18–34 Year Olds



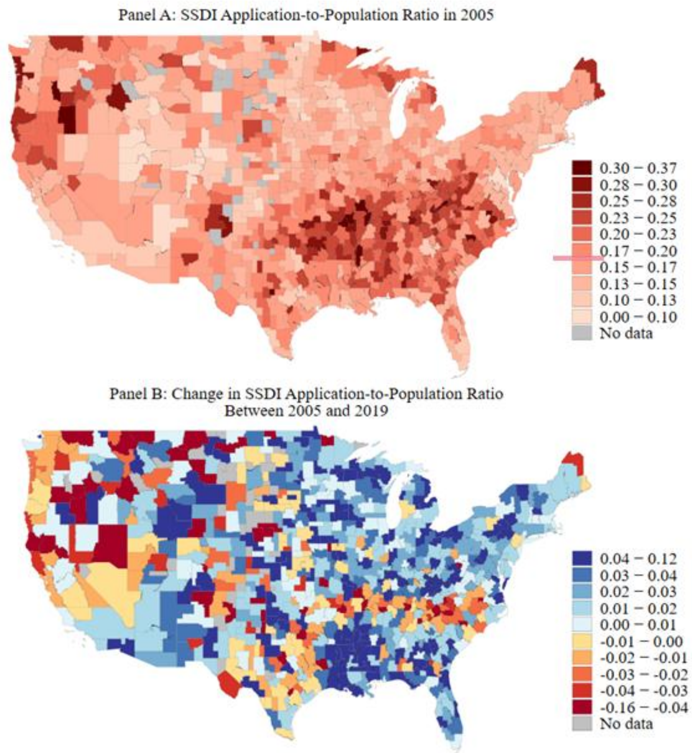
Notes: For 18–34 year-olds, the figure presents heat maps of the SSDI application-to-population ratio, as defined in equation 1, in 2005 (Panel A) as well as its change between 2005 and 2019 (Panel B). In Panel A, lighter red colors indicate a lower application-to-population ratios, and darker red colors indicate a greater prevalence of per-capita applications. Commuting zones for which insufficient applications were submitted to allow for external presentation are shown in gray. In Panel B, we use a diverging color scheme, in which light to dark blue indicate positive changes in application rates, and yellow to red indicate negative changes in application rates. The positive changes are greater when the color is a darker blue. Likewise, negative changes are greater when the color is a darker red.

Figure 17: Geographic Variation in Population-Adjusted SSDI Applications, 35–54 Year Olds



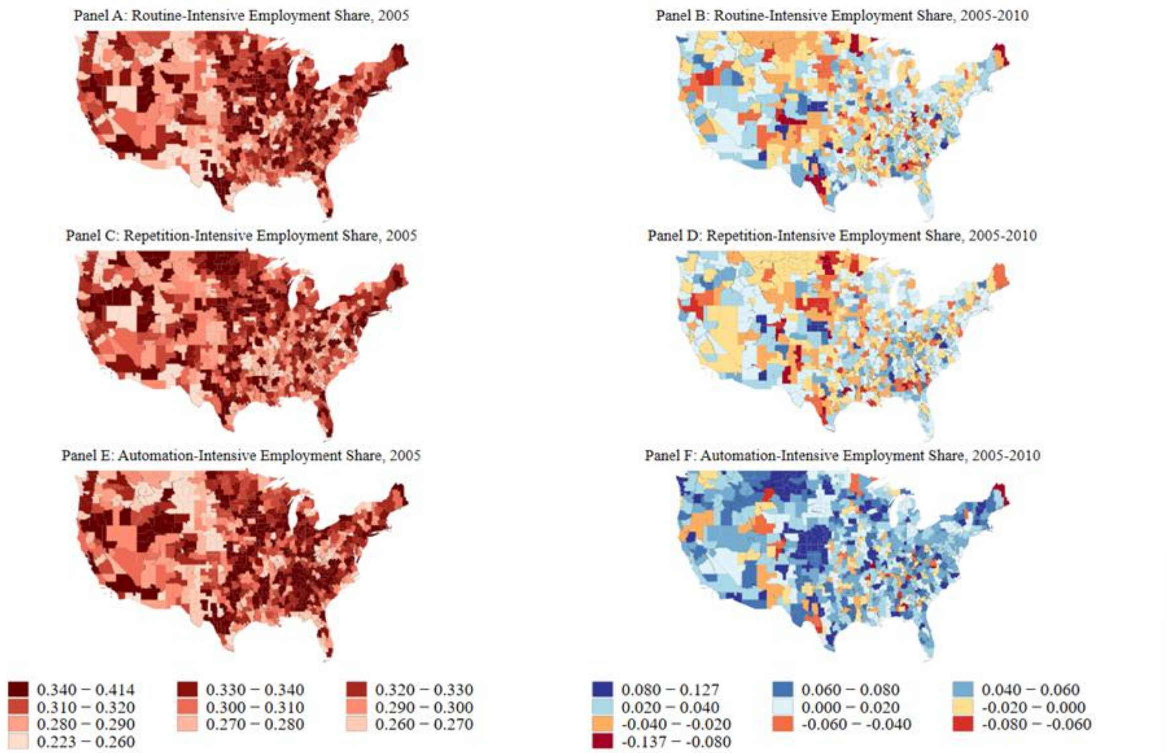
Notes: For 35–54 year-olds, the figure presents heat maps of the SSDI application-to-population ratio, as defined in equation 1, in 2005 (Panel A) as well as its change between 2005 and 2019 (Panel B). In Panel A, lighter red colors indicate a lower application-to-population ratios, and darker red colors indicate a greater prevalence of per-capita applications. Commuting zones for which insufficient applications were submitted to allow for external presentation are shown in gray. In Panel B, we use a diverging color scheme, in which light to dark blue indicate positive changes in application rates, and yellow to red indicate negative changes in application rates. The positive changes are greater when the color is a darker blue. Likewise, negative changes are greater when the color is a darker red.

Figure 18: Geographic Variation in Population-Adjusted SSDI Applications, 55–64 Year Olds



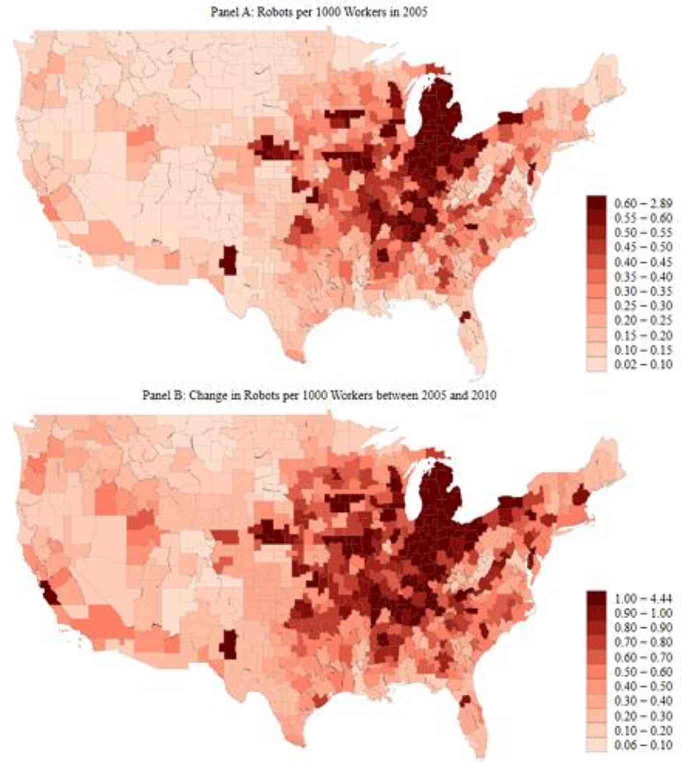
Notes: For 55–64 year-olds, the figure presents heat maps of the SSDI application-to-population ratio, as defined in equation 1, in 2005 (Panel A) as well as its change between 2005 and 2019 (Panel B). In Panel A, lighter red colors indicate a lower application-to-population ratios, and darker red colors indicate a greater prevalence of per-capita applications. Commuting zones for which insufficient applications were submitted to allow for external presentation are shown in gray. In Panel B, we use a diverging color scheme, in which light to dark blue indicate positive changes in application rates, and yellow to red indicate negative changes in application rates. The positive changes are greater when the color is a darker blue. Likewise, negative changes are greater when the color is a darker red.

Figure 19: Geographic Variation in the Routine, Repetition, and Automation Intensive Employment Shares



Notes: The figure presents heat maps for the automation measures described in Section 3.2.1 for 2005 (Panels A, C, and E) as well as the changes in the measures between 2005 and 2010 (Panels B, D, and F). Panels A and B focus on the routine-intensive employment share; Panels C and D on the repetition-intensive employment share; and Panels E and F on the automation-intensive employment share.

Figure 20: Geographic Variation in Exposure to Industrial Robots



Notes: The figure presents heat maps for the automation measures described in Section 3.2.2 for 2005 (Panel A) as well as the changes in the measure between 2005 and 2010 (Panel B).

References

Abadie, Alberto. 2021. “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects.” Journal of Economic Literature 59 (2): 391–425.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program.” Journal of the American Statistical Association 105 (490): 493–505.

Abusaada, Hisham, & Elshater, Abeer. (2021). “Competitiveness, Distinctiveness and Singularity in Urban Design: A Systematic Review and Framework for Smart Cities.” Sustainable Cities and Society 68, pp. 102782.

Acemoglu, D., F. Kong, and P. Restrepo (2024). “Tasks At Work: Comparative Advantage, Technology and Labor Demand.” Publisher: National Bureau of Economic Research.

Acemoglu, D., C. Lelarge, and P. Restrepo (2020, May). “Competing with Robots: Firm Level Evidence from France.” AEA Papers and Proceedings 110, 383–388.

Acemoglu, D. and P. Restrepo (2020, June). “Robots and Jobs: Evidence from US Labor Markets.” Journal of Political Economy 128 (6), 2188–2244.

Acemoglu, D. and P. Restrepo (2022a, January). “Demographics and Automation.” The Review of Economic Studies 89 (1), 1–44.

Acemoglu, D. and P. Restrepo (2022b). “Tasks, automation, and the rise in us wage inequality.” Econometrica 90 (5), 1973–2016. Publisher: Wiley Online Library.

Acemoglu, D. and P. Restrepo (2024). “Automation and Rent Dissipation: Implications for Wages, Inequality, and Productivity.” Technical report, National Bureau of Economic Research.

Agrawal, A. & Avi Goldfarb. (2005). “How Do Communication Costs Affect Scientific Collaboration? Exploring the Effect of Bitnet.” University of Toronto.

Agrawal, Ajay, & Avi Goldfarb. (2008). “Restructuring Research: Communication Costs and the Democratization of University Innovation.” American Economic Review 98 (4), pp. 1578-90.

Agrawal, A., A. Galasso, & A. Oettl. (2017). “Roads and Innovation.” The Review of Economics and Statistics 99(3), pp. 417–34.

Aldieri, L. & Kotsemir, M., & Vinci, C.P. (2017). “The Impact of Research Collaboration on Academic Performance: An Empirical Analysis for Some European Countries.” Socio-Economic Planning Sciences 62, pp. 13-30.

Aldieri, L., Kotsemir, M., & Vinci, C.P. (2020). “The Effects of Collaboration on Research

Performance of Universities: an Analysis by Federal District and Scientific Fields in Russia.” Journal of the Knowledge Economy 11, pp. 766-787.

Amaral-Garcia, S., Nardotto, M., Propper, C., & Valletti, T. (2021). “Mums Go Online: Is the Internet Changing the Demand for Healthcare?” The Review of Economics and Statistics 104(6), pp. 1157-1173.

Ando, Michihito. 2015. “Dreams of Urbanization: Quantitative Case Studies on the Local Impacts of Nuclear Power Facilities Using the Synthetic Control Method.” Journal of Urban Economics 85 (January):68–85. https://doi.org/10.1016/j.jue.2014.10.005.

Ankilu, Masahudu. (2022). “Africa Connectivity Outlook: 2022 and Beyond - African Eye Report.” African Eye Report. https://africaneyereport.com/africa-connectivity-outlook-2022-and-beyond/

Autor, D. H. and D. Dorn (2013). “The growth of low-skill service jobs and the polarization of the US labor market.” American economic review 103 (5), 1553–1597. Publisher: American Economic Association.

Autor, D. H., D. Dorn, and G. H. Hanson (2013, October). “The China Syndrome: Local Labor Market Effects of Import Competition in the United States.” American Economic Review 103 (6), 2121–2168.

Autor, D. H. and M. G. Duggan (2006). “The growth in the social security disability rolls: a fiscal crisis unfolding.” Journal of Economic perspectives 20 (3), 71–96. Publisher: American Economic Association.

Avanesova, Anna A., & Shamliyan, Tatyana A. (2019). “Worldwide Implementation of Telemedicine Programs in Association with Research Performance and Health Policy.” Health Policy and Technology 8(2), pp.179-191.

Baily, M. N., & Montalbano, N. (2017). “Clusters and Innovation Districts: Lessons from the United States Experience.” The Brookings. https://www.brookings.edu/research/clusters-and innovation-districts-lessons-from-the-united-states-experience/

Bannerman, Natalie. (2021). “Land and Sea: The Role of Terrestrial in Subsea Networks.” Capacity Media. https://www.capacitymedia.com/article/29otdn0ylukn01p0q169t/news/land and-sea-the-role-of-terrestrial-in-subsea-networks

Barjak, Franz. (2006). “Research Productivity in the Internet Era.” Scientometrics 68, pp. 343 360. 31 Baro, E. E., Bosah, G. E., & Obi, I. C. (2017). “Research Funding Opportunities and Challenges A Survey of Academic Staff Members in Nigerian Tertiary Institutions.” The Bottom Line 30(1), pp. 47-64.

Bauer, Thomas K, Sebastian T Braun, and Michael Kvasnicka. 2017. "Nuclear Power Plant Closures and Local Housing Values: Evidence from Fukushima and the German Housing Market." Journal of Urban Economics 99:94–106.

Berk, Jonathan B., Campbell R. Harvey, & David Hirshleifer. (2017). "How to Write an Effective Referee Report and Improve the Scientific Review Process." Journal of Economic Perspectives 31(1), pp. 231-44.

Berkes, Enrico, & Nencka, Peter. (2021). "Knowledge Access: The Effects of Carnegie Libraries on Innovation." Available at SSRN: http://dx.doi.org/10.2139/ssrn.3629299
Berndt, Eric, and Daniel P Aldrich. 2016. "Power to the People or Regulatory Ratcheting? Explaining the Success (or Failure) of Attempts to Site Commercial US Nuclear Power Plants: 1954–1996." International Journal of Energy Research 40 (7): 903–23.

Besse-Lotoskaya, Anna, & Zwietering, Marcel. (2021). "Open Access Publishing versus Quality." Open Science channel WUR Library and Wageningen Data Competence Center (WDCC). https://weblog.wur.eu/openscience/open-access-publishing-versus-quality/ Bonfoh, Bassirou. (2016). "Challenges of Research Funding in Africa." SciDev. https://www.scidev.net/afrique-sub-saharienne/opinions/challenges-of-research-funding-in africa/

Blasco, S., J. Rochut, and B. Rouland (2024, January). "Displaced or depressed? Working in automatable jobs and mental health." Industrial Relations: A Journal of Economy and Society, irel.12356.

Borusyak, K., P. Hull, and X. Jaravel (2022, January). "Quasi-Experimental Shift-Share Research Designs." The Review of Economic Studies 89 (1), 181–213.

Bratsberg, B., O. Rogeberg, and V. Skirbekk (2022). "Technology-induced job loss risk, disability and all-cause mortality in Norway." Occupational and Environmental Medicine 79 (1), 32–37. Publisher: BMJ Publishing Group Ltd.

Brunner, Eric J., and David J. Schwegman. (2022). "Windfall revenues from windfarms: How do county governments respond to increases in the local tax base induced by wind energy installations?" Public Budgeting and Finance 42(3), pp. 93-113.

Bryan, K.A., & Ozcan, Yasin, (2020). "The Impact of Open Access Mandates on Invention." Review of Economics and Statistics 1(45). Butler, D., Butler, R., & Rich, J. (2008). "The Equalizing Effect of the Internet on Access to Research Expertise in Political Science and Economics." Political Science and Politics 41(3), pp. 579-584.

Butler, R. (2007). "JRI, JF, and the Internet: Coauthors, New Authors, and Empirical Research." The Journal of Risk and Insurance 74(3), pp. 713-737.

Caglar, Musa, & Gurel, Sinan. (2019). "Impact Assessment Based Sectoral Balancing in Public R&D Project Portfolio Selection." Socio-Economic Planning Sciences 66, pp. 68-81.

Callaway, B., & P. H. Sant'Anna. (2021). "Difference-in-Differences with Multiple Time Periods." Journal of Econometrics 219 (1), pp. 101-122

Carillo, Maria, Papagni, Erasmo, & Sapio, Alessandro. (2013). "Do Collaborations Enhance the High-quality Output of Scientific Institutions? Evidence from the Italian Research Assessment Exercise." The Journal of Socio-Economics 47, pp. 25-36.

Card, David, & Stefano DellaVigna. (2013). "Nine Facts about Top Journals in Economics." Journal of Economic Literature 51(1), pp. 144-61.

Carrell, D.T., and Simoni, M. (2017). "Easier Ways to Get A Publication: The Problem of Low Quality Scientific Publications." Andrology 6(1), pp. 1-2.

Chattopadhyaya, Suvesh. (2019). "Submarine Cables for Africa – A Close Look at 2019-21." Submarine Networks. https://www.submarinenetworks.com/en/insights/submarine-cables-for africa-a-close-look-at-2019-21 32

Chen, Shiyi, & Liu, Wanlin, & Song, Hong. (2019). "Broadband Internet, Firm Performance, And Worker Welfare: Evidence and Mechanism." Economic Inquiry. 58 (3), pp. 1146-1166.

Chen, Jiafeng. & Roth, Jonathan. (2023). "Logs with Zeros? Some Problems and Solutions." The Quarterly Journal of Economics. 10.1093/qje/qjad054. Cisco. (2008). "Global IP Traffic Forecast and Methodology, 2006-2011."

Cheng, W., L. Pien, and Y. Cheng (2021, February). "Occupation-level automation probability is associated with psychosocial work conditions and workers' health: A multilevel study." American Journal of Industrial Medicine 64 (2), 108–117.

Costa, Hélia, and Linda Veiga. "Local labor impact of wind energy investment: An analysis of Portuguese municipalities." Energy Economics 94, 105055.

Cisco White Paper. Cisco. (2013). "Cisco Visual Networking Index: Forecast and Methodology, 2012-2017."

Cisco White Paper. Clarke, Isobelle. (2022). "How Fake Science Websites Hijack Our Trust in Experts to Misinform and Confuse." The Conversation. https://theconversation.com/how-fake-science-websites-hijack our-trust-in-experts-to-misinform-and-confuse-189730

Cleerline. (2019). "Single Mode vs. Multimode Fiber Optic Cables". https://cleerlinefiber.com/2019/03/19/singlemode-vs-multimode-fiber-optic-cables/

Coombes, M. G., Storey, D. J., Watson, R., & Wynarczyk, P. (1991). “The Influence of Location Upon Profitability and Employment Change in Small Companies.”Urban Studies 28(5), pp. 723 734.

Commonwealth Telecommunications Organization. (2012). “The Socio-Economic Impact of Broadband in Sub-Sahara Africa: The Satellite Advantage.” London: Commonwealth Telecommunications Organization.

Davis, Lucas, and Catherine Hausman. 2016. “Market Impacts of a Nuclear Power Plant Closure.” American Journal: https://doi.org/10.1257/app.20140473. Applied Economics 8 (2): 92–122.

Davis, Lucas W. 2011. “The Effect of Power Plants on Local Housing Values and Rents.” Review of Economics and Statistics 93 (4): 1391–1402.

Davis, Lucas W, and Catherine Wolfram. 2012. “Deregulation, Consolidation, and Efficiency: Evidence from US Nuclear Power.” American Economic Journal: Applied Economics 4 (4): 194 225. https://doi.org/10.1257/app.4.4.194.

Dauth, W., S. Findeisen, J. Suedekum, and N. Woessner (2021, December). “The Adjustment of Labor Markets to Robots.” Journal of the European Economic Association 19 (6), 3104–3153.

de Chaisemartin, C. and X. D’Haultfœuille. (2020). “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” American Economic Review 110(9), pp. 2964-2996.

Deming, D. J. (2017). “The growing importance of social skills in the labor market.” The Quarterly Journal of Economics 132 (4), 1593–1640. Publisher: Oxford University Press.

Dettling, Lisa J., Goodman, Sarena, & Smith, Jonathan. (2018). “Every Little Bit Counts: The Impact of High-Speed Internet on the Transition to College.” The Review of Economics and Statistics 100(2), pp. 260-273.

Dhalla, K. A., & Guirguis, M. (2018). “Barriers and Incentives For Conducting Research Amongst the Ophthalmologists in Sub-Sahara Africa.” PLoS ONE, 13(10).

Ductor, Lorenzo. (2014). “Does Co-Authorship Lead to Higher Academic Productivity?” Oxford Bulletin of Economics & Statistics 77(3), pp. 385-407

Duermeijer, Charon, Amir, Mohamad, and Schoombee, Lucia. (2018). “Africa Generates Less Than 1% of the World’s Research; Data Analytics Can Change That.” Elsevier Connect. https://www.elsevier.com/connect/africa-generates-less-than-1-of-the-worlds-research-data analytics-can-change-that

Duke Energy. 2012. “NRC New Nuclear Licensing Process.” https://nuclear.duke energy.com/2012/01/17/nrc-new-nuclear-licensing-process.

Duque, R. B., Shrum, W. M., Barriga, O., & Henríquez, G. (2009). “Internet Practice and Pofessional etworks in Chilean Science: Dependency or Progress?”. Scientometrics 81, pp. 239 263.

Eash-Gates, Philip, Magdalena M. Klemun, Goksin Kavlak, James McNerney, Jacopo Buongiorno, Jessika E. Trancik. (2020). “Sources of Cost Overrun in Nuclear Power Plant Construction Call for a New Approach to Engineering Design.” Joule 4(11), pp. 2348-2373.

Echezona, R. I., & Ugwuanyi, C. F. (2010). “African University Libraries and Internet Connectivity: Challenges and the Way Forward.” Library Philosophy and Practice 1.

Ekpo, Ernest U., Hogg, Peter, & McEntee, Mark F. (2016). “A Review of Individual and Institutional Publication Productivity in Medical Radiation Science.” Journal of Medical Imaging and Radiation Sciences 47(1), pp. 13-20.

Ellison, G. (2002).” Evolving Standards for Academic Publishing: A q-r Theory.” Journal of Political Economy 110(5), pp. 994-1034.

Fabra, Natalia, Gutiérrez Chacón, Eduardo and Lacuesta, Aitor, and Ramos, Roberto. (2023). “Do renewable energies create local jobs?” Banco de Espana Working Paper No. 2307.

Falck, Oliver, Robert Gold, and Stephan Heblich. (2014). “E-lections: Voting Behavior and the Internet.” American Economic Review, 104 (7), pp. 2238-65.

Faria, F. A. M. de, Davis, A., Severnini, E., & Jaramillo, P. (2017). “The local socio-economic impacts of large hydropower plant development in a developing country.” Energy Economics, 67, 533–544. https://doi.org/10.1016/j.eneco.2017.08.025

Fink, Alexander, and Thomas Stratmann. 2015. “U.S. Housing Prices and the Fukushima Nuclear Accident.” Journal of Economic Behavior & Organization 117 (September):309–26. https://doi.org/10.1016/j.jebo.2015.07.005.

Ford, George S., & Seals, Richard A. (2021). “The Rewards of Municipal Broadband: An Econometric Analysis of the Labor.” Telecommunications Policy 45(8).

Foster, Vivian, Comini, Niccolo, & Sharada, Srinivasan. (2021). “Improving Data Infrastructure Helps Ensure Equitable Access for Poor People in Poor Countries.” World Bank Blogs. https://blogs.worldbank.org/opendata/improving-data-infrastructure-helps-ensure-equitable access-poor-people-poor-countries

Freese, Lyssa M, Guillaume P Chossière, Sebastian D Eastham, Alan Jenn, and Noelle E Selin. 2023. “Nuclear Power Generation Phase-Outs Redistribute US Air Quality and Climate-Related Mortality Risk.” Nature Energy, 1–12.

FS Community. (2021). “Multimode Fiber Types: OM1 vs OM2 vs OM3 vs OM4 vs OM5.” https://community.fs.com/blog/advantages-and-disadvantages-of-multimode-fiber.html

Gannon, Mary. (2016). “Multimode Fiber Optic Cable or Single: What’s the Difference?” Wire & Cable Tips. https://www.wireandcabletips.com/should-i-use-single-or-multi-mode-fiber-optic cable/

Gazni, Ali, Sugimoto, Cassidy & Didegah, Fereshteh. (2012). “Mapping World Scientific Collaboration: Authors, Institutions, and Countries.” Journal of the American Society for Information Science and Technology 63(2), pp. 323-335.

Geraci, Andrea, Nardotto, Mattia, Reggiani, Tommaso, & Sabatini, Fabio. (2022). “Broadband Internet and social capital.” Journal of Public Economics 206. Goodman-Bacon, A. (2021). “Difference-in-Differences With Variation in Treatment Timing.” Journal of Econometrics 225(2), pp. 254-277.

Gihleb, R., O. Giuntella, L. Stella, and T. Wang (2022). “Industrial robots, workers’ safety, and health.” Labour Economics 78, 102205. Publisher: Elsevier.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). “Bartik instruments: What, when, why, and how.” American Economic Review 110 (8), 2586–2624. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Gonçalves, S., T.P. Rodrigues, and A.L.S. Chagas. (2020). “The impact of wind power on the Brazilian labor market.” Renewable and Sustainable Energy Reviews 128, 109887.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings.” Journal of Political Economy 118 (3): 536–98.

Grigorik, Ilya. (2015). “A Primer on Latency and Bandwidth.” O’Reilly. Retrieved November 19, 2022, from https://www.oreilly.com/content/primer-on-latency-and-bandwidth/

Gunadi, C. and H. Ryu (2021). “Does the rise of robotic technology make people healthier?” Health Economics 30 (9), 2047–2062. Publisher: Wiley Online Library.

Guo, Yue, and Tao Ren. 2017. “When It Is Unfamiliar to Me: Local Acceptance of Planned Nuclear Power Plants in China in the Post-Fukushima Era.” Energy Policy 100 (January):113–25. https://doi.org/10.1016/j.enpol.2016.10.002.

Gurib-Fakim, Ameenah, & Signee, Landry. (2022). “Investment in Science and Technology Is Key to an African Economic Boom.” Brookings. https://www.brookings.edu/blog/africa-in focus/2022/01/26/investment-in-science-and-technology-is-key-to-an-african-economic-boom/

Guriev, Sergei, Melnikov, Nikita, & Zhuravskaya, Ekaterina. (2021). “3G Internet and Confidence in Government.” The Quarterly Journal of Economics 136(4), pp. 2533-2613. 34

Guseva, Anna I., Kalashnik, Viacheslav M., Kaminskii, Vladimir I., & Kireev, Sergey V. (2022). “Key performance indicators of Russian universities for 2015–2018: Dataset and Benchmarking Data.” Data in Brief 40.

Hausman, Catherine. 2014. “Corporate Incentives and Nuclear Safety.” American Economic Journal: Economic Policy 6 (3): 178–206. https://doi.org/10.1257/pol.6.3.178.

Hausman, N. (2022). “University Innovation and Local Economic Growth”. The Review of Economics and Statistics 104(4), pp. 718-735.

Hjort, Jonas, & Poulsen, Jonas. (2019). “The Arrival of Fast Internet and Employment in Africa.” American Economic Review 109 (3), pp. 1032-79.

Horenberg, Frank, Lungu, Daniel A., & Nuti, Sabina. (2020). “Measuring Research in the Big Data Era: The Evolution of Performance Measurement Systems in the Italian Teaching Hospitals.” Health Policy 124(12), pp.1387-1394.

Hwenda, L., Sidibe, M., & Makanga, M. (2022). “The African Medicines Agency: The Key to Unlocking Clinical Research in Africa.” The Lancet Global Health 10(8), pp. e1088–e1089.

Iaria, Alessandro, Schwarz, Carlo, & Waldinger, Fabian. (2021). “Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science.” The Quarterly Journal of Economics 133(2), pp. 927-991.

Imhonopi, D. & Urim, U. M. (2011). “The Impact of Internet Services on the Research Output of Academic Staff of Selected State Universities in South-Western Nigeria.” International Journal of Information and Communication Technology 8(1), pp. 9-20.

International Atomic Energy Agency. 2012. Electric Grid Reliability and Interface with Nuclear Power Plants. IAEA.

Ishengoma. J. M. (2018). “Financing Public Universities and Research in Sub-Saharan Africa: Challenges and Prospects in the Context of Sustainable Development Goals (SDGs).” RUFORUM 15, pp. 51-61.

ITU. (2020). The Last-mile Internet Connectivity Solutions Guide Sustainable connectivity options for unconnected sites. https://www.itu.int/en/ITU-D/Technology/Pages/LMC/LMC-Home.aspx

Jarvis, Stephen, Olivier Deschenes, and Akshaya Jha. 2022. “The Private and External Costs of Germany’s Nuclear Phase-Out.” Journal of the European Economic Association 20 (3): 1311–46.

Jerch, R., Kahn, M. E., & Lin, G. C. (2023). "Local public finance dynamics and hurricane shocks." Journal of Urban Economics, 134, 103516. https://doi.org/10.1016/j.jue.2022.103516

Johnson, S. and D. Acemoglu (2023). "Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity." Hachette UK.

Kawaguchi, Daiji, and Norifumi Yukutake. 2017. "Estimating the Residential Land Damage of the Fukushima Nuclear Accident." Journal of Urban Economics 99 (May):148–60. https://doi.org/10.1016/j.jue.2017.02.005. 29

Kearney, M. S., B. M. Price, and R. Wilson (2021). "Disability Insurance in the Great Recession: Ease of Access, Program Enrollment, and Local Hysteresis." Volume 111, pp. 486–490. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Kigotho, Wachira. (2022). "Egypt Leads Research Field for the Second Year Running." University World New. https://www.universityworldnews.com/post.php?story=20220515185506702

Kim, E. Han, Morse, Adair, & Zingales, Luigi (2009). "Are Elite Universities Losing Their Competitive Edge?" Journal of Financial Economics 93, pp. 353-381.

Kline, Patrick, & Moretti, Enrico. (2014). "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." The Quarterly Journal of Economics 129(1), pp. 275-331. 35

Kollmann, T. R., Bortolussi, R., & MacDonald, N. E. (2015). "MicroResearch—Finding Sustainable Solutions to Local Health Challenges in East Africa." The Journal of Infection 71(1), pp. S97–S100.

Kolade, Seun. (2022). "Nigeria's Universities Can Find Funds and Produce Job Creators: Here's How." The Conversation. https://theconversation.com/nigerias-universities-can-find-funds-and produce-job-creators-heres-how-190155

Komanoff, Charles. 1981. Power Plant Cost Escalation: Nuclear and Coal Capital Costs, Regulation, and Economics. New York, USA: Van Nostrand Reinhold.

Kumwenda S, Niang EHA, Orondo PW, William P, Oyinlola L, Bongo GN, & Chiwona B. (2017). "Challenges Facing Young African Scientists in Their Research Careers: A Qualitative Exploratory Study." Malawi Med J. 29(1), pp. 1-4.

Lemov, Michael R. 1964. "State and Local Control Over the Location of Nuclear Reactors Under the Atomic Energy Act of 1954." NYUL Rev. 39:1008.

Lewis, Pamela M. 1986. “The Economic Impact of the Operation and Closure of a Nuclear Power Station.” Regional Studies 20 (5): 425–32. Lochbaum, David, Edwin Lyman, and Susan Q Stranahan. 2014. Fukushima: The Story of a Nuclear Disaster. New Press, The.

Liebman, J. B. (2015). “Understanding the increase in disability insurance benefit receipt in the United States.” Journal of Economic Perspectives 29 (2), 123–150. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418.

Lordan, G. and E.-J. Stringer (2022). “People versus machines: The impact of being in an automatable job on Australian worker’s mental health and life satisfaction.” Economics & Human Biology 46, 101144. Publisher: Elsevier.

Lovering, Jessica R., Arthur Yip, and Ted Nordhaus. (2016). “Historical construction costs of global nuclear power reactors.” Energy Policy 91, pp. 371-382.

Machacek, Vit., & Srholec, Martin. (2017). “Predatory Journals in Scopus.” The Economic Institute of the Czech Academy of Sciences. Macharia, Joel. (2014). “Internet Access Is No Longer a Luxury”. The Africa Renewal Magazine. https://www.un.org/africarenewal/magazine/april-2014/internet-access-no-longer-luxury

Maestas, N., K. J. Mullen, and A. Strand (2021, July). “The effect of economic conditions on the disability insurance program: Evidence from the great recession.” Journal of Public Economics 199, 104410.

Manacorda, Marco, & Tesei, Andrea. (2020). “Liberation Technology: Mobile Phones and Political Mobilization in Africa.” Econometrica 88(2), pp. 533-567.

Markoff, J. (2015). “Research Scientists to Use Network Much Faster Than Internet.” The New York Times. https://www.nytimes.com/2015/08/01/science/research-scientists-to-use-network-much-faster than-internet.html

Mauldin, Alan. (2017). “Submarine Cable Frequently Asked Questions: Submarine Cables 101.” TeleGeography. https://www2.telegeography.com/submarine-cable-faqs-frequently-askedquestions.

McConnell, Brendon. (2023). "What's Logs Got to do With it: On the Perils of log Dependent Variables and Difference-in-Differences." Papers 2308.00167, arXiv.org, revised Aug 2023.

McDool, Emily, Powell, Philip, Roberts, Jennifer, & Taylor, Karl. (2021). “The Internet and Children’s Psychological Wellbeing.” Journal of Health Economics 69, pp. 1022-1074. Mitchell, Jason. (2021). “Which Country Tops the African e-Connectivity Index?” Investment Monitor. https://www.investmentmonitor.ai/tech/africa-connectivity-index-2021

Mehic, Adrian. 2023. “The Electoral Consequences of Environmental Accidents: Evidence from Chernobyl.” Journal of Public Economics 225:104964.

Moretti, Enrico. (2021). "The Effect of High-Tech Clusters on the Productivity of Top Inventors." American Economic Review 111 (10), pp. 3328-75

Mouton, Johann, & Valentine, Astrid. (2017). "The Extent of South African Authored Articles in Predatory Journals." S Afr J Sci. https://sajs.co.za/article/view/3995 36

Mullahy, J. & Norton, E.C. (2024). "Why Transform Y? The Pitfalls of Transformed Regressions with a Mass at Zero." Oxf Bull Econ Stat, 86: 417-447. https://doi.org/10.1111/obes.12583

National Research Council (US) Committee on Competing in the 21st Century. (2013). "Best Practice in State and Regional Innovation Initiatives; Wessner CW, editor. Best Practices in State and Regional Innovation Initiatives: Competing in the 21st Century." Washington (DC): National Academies Press (US 3, Universities as Innovation Drivers. Available from: https://www.ncbi.nlm.nih.gov/sites/books/NBK158819/

Neidell, Matthew, Shinsuke Uchida, and Marcella Veronesi. 2021. "The Unintended Effects from Halting Nuclear Power Production: Evidence from Fukushima Daiichi Accident." Journal of Health Economics 79 (September):102507. https://doi.org/10.1016/j.jhealeco.2021.102507.

Neumark, David, and Helen Simpson. 2015. "Place-Based Policies." In Handbook of Regional and Urban Economics, 5:1197–1287. Elsevier.

Nganga, Gilbert. (2022). "As University Shutdowns Loom, Students Face Funding Cuts". University World News. https://www.universityworldnews.com/post.php?story=20221102065458l5

Ngari, Leah, and Petrack, Shira Aliza. (2019). "Internet Infrastructure in Africa." Empower Africa. https://www.empowerafrica.com/internet-infrastructure-in-africa/

North, M. A., Hastie, W. W., & Hoyer, L. (2020). "Out of Africa: The Underrepresentation of African Authors in High-Impact Geoscience Literature." In Earth-Science Reviews 28, pp.103262.

O'Brien, R., E. F. Bair, and A. S. Venkataramani (2022). "Death by robots? Automation and working-age mortality in the United States." Demography 59 (2), 607–628. Publisher: Duke University Press.

OECD. (2018). "Education at a Glance 2018: OECD Indicators." https://doi.org/10.1787/eag 2018-en.

Okafor, E., Imhonopi, D., & Moses Urim, U. E. (2011). "Utilisation of Internet Services and Its Impact on Teaching and Research Outputs in Private Universities in South-Western Nigeria." International Journal of Emerging Technologies and Society 9(2), pp.135-151.

Okeke, I. N., Babalola, C. P., Byarugaba, D. K., Djimde, A., & Osoniyi, O. R. (2017). “Broadening Participation in The Sciences Within and from Africa: Purpose, Challenges, and Prospects.” CBE Life Sciences Education 16(2).

Onwubiko, CPC. (2012). “Impact of the Internet on Research Effort of Academics at Abia State University, Uturu, (ABSU).” University of Nebraska - Lincoln.

Pachi, C. G. d. F., Yamamoto, J. F., Costa, A. P. A. d., & Lopez, L. F. (2012). “Relationship Between Connectivity and Academic Productivity.” Scientometrics 93, pp. 265-78.

Patelli, Aurelio, Cimini, Giulio, Pugliese, Emanuele, & Gabrielli, Andre. (2017). “The Scientific Influence of Nations on Global Scientific and Technological Development.” Journal of Informetrics 11(4), pp.1229-1237.

Photonics Media. (2015). “Fiber Optics: Understanding the Basics” https://www.photonics.com/Articles/Fiber_Optics_Understanding_the_Basics/a25151

Pierson, K., Hand, M. L., & Thompson, F. (2015). “The Government Finance Database: A common resource for quantitative research in public financial analysis.” PLoS ONE, 10(8), e0130119. https://doi.org/10.1371/journal.pone.0130119

Plokhy, Serhii. 2018. Chernobyl: The History of a Nuclear Catastrophe. Basic Books.

Reveiu, A., & Dardala, M. (2013). “The Role of Universities in Innovative Regional Clusters. Empirical Evidence from Romania.” Procedia - Social and Behavioral Sciences 93, pp. 555-559.

Rosenblat, T. S., & Mobius, M. M. (2004). “Getting Closer or Drifting Apart?” The Quarterly Journal of Economics 119(3), pp. 971-1009.

Rosowsky, David. (2022). “The Role Of Research At Universities: Why It Matters.” Forbes. https://www.forbes.com/sites/davidrosowsky/2022/03/02/the-role-of-research-at-universities why-it-matters/?sh=6abe6d936bd5

Roth, Michael Buchdahl, and Paulina Jaramillo. 2017. “Going Nuclear for Climate Mitigation: An Analysis of the Cost Effectiveness of Preserving Existing US Nuclear Power Plants as a Carbon Avoidance Strategy.” Energy 131:67–77.

Ruggles, S., S. Flood, M. Sobek, D. Backman, A. Chen, G. Cooper, S. Richards, R. Rogers, and M. Schouwiler (2023). IPUMS USA: Version 14.0.

Sarewitz. Daniel. (2016). “The pressure to publish pushes down quality”. Nature 533, pp.147. Sawahel, Wagdy. (2022). “Science Output Rising, But Some Countries’ Yields Still Low.” University World News. https://www.universityworldnews.com/post.php?story=20220214141713369 Sawyerr, Akilagpa. (2004). “African Universities and the Challenge of Research Capacity Development.” Journal of Higher Education in 2(1), pp. 213-242.

Schemm, Ylann. (2013). “Africa Doubles Research Output Over Past Decade, Moves Towards a Knowledge-Based Economy.” 3BL Media. https://www.3blmedia.com/news/africa-doubles research-output-over-past-decade-moves-towards-knowledge-based-economy

Shen, Cenyu, & Bjork, B. (2015). “Predatory’ Open Access: A Longitudinal Study of Article Volumes and Market Characteristics.” BMC Medicine 13.

Sooryamoorthy, R. (2016). “Producing Information: Communication and Collaboration in the South African Scientific Community.” Information, Communication & Society 19(2), pp. 141 159.

Sooryamoorthy, R. (2021). “Science in Africa: Contemporary Trends in Research”. Journal of Scientometric Research 10(3), pp. 366-372.

Sooryamoorthy, R., Duque, R. B., Ynalvez, M. A., & Shrum, W. (2007). “Scientific Collaboration and The Kerala Model: Does The Internet Make A Difference?” Journal of International Development 19(7), pp. 982-996.

Sun, Liyang and Abraham, Sarah. (2021). “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” Journal of Econometrics 225(2), pp. 175-199.

Tropea, Anthony. (2019). “Four Ways Location Affects Profits.” Forbes. https://www.forbes.com/sites/forbesrealestatecouncil/2019/12/13/four-ways-location-affects profits/?sh=657e441085a8

US Department of Energy. 1983. “Nuclear Plant Cancellations: Causes, Costs, and Consequences.” Washington DC (United States): Energy Information Administration, Office of Coal, Nuclear, Electric and Alternate Fuels. https://www.osti.gov/biblio/6211281.

US Energy Information Administration. 2017. “Most U.S. Nuclear Power Plants Were Built between 1970 and 1990.” 2017. https://www.eia.gov/todayinenergy/detail.php?id=30972.

US Nuclear Regulatory Commission. 2022. “Backgrounder on Three Mile Island.” https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html. 30 Process.”

US Nuclear Regulatory Commission (NRC). 2020. “Backgrounder on Nuclear Power Plant Licensing https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/licensing process-fs.html.

US Social Security Administration (2023). Annual Statistical Report on the Social Security Disability Insurance Program, 2022.

Vigdor, J.L, Ladd, H.F. Ladd, & Martinez E. (2014). “Scaling The Digital Divide: Home Computer Technology and Student Achievement.” Economic Inquiry 52 (3), pp. 1103-1119. 38

Wahab, Adesina. (2022). “University Education in Nigeria: Funding Challenges and Controversy.” Vanguard. https://www.vanguardngr.com/2022/02/university-education-in nigeria-funding-challenges-and-controversy/

Walker, J. Samuel. 1990. “Reactor at the Fault: The Bodega Bay Nuclear Plant Controversy, 1958 1964: A Case Study in the Politics of Technology.” Pacific Historical Review 59 (3): 323–48. https://doi.org/10.2307/3640755.

Walker, J Samuel. 2004. Three Mile Island: A Nuclear Crisis in Historical Perspective. Vol. 41. Univ of California Press.

Wang, Yu, Jibao Gu, and Jianlin Wu. 2020. “Explaining Local Residents’ Acceptance of Rebuilding Nuclear Power Plants: The Roles of Perceived General Benefit and Perceived Local Benefit.” Energy Policy 140 (May):111410. https://doi.org/10.1016/j.enpol.2020.111410.

Wolfe, David A.. (2005). "10. The Role of Universities in Regional Development and Cluster Formation". Creating Knowledge, Strengthening Nations: The Changing Role of Higher Education, edited by Glen A. Jones, Patricia McCarney and Michael L. Skolnik, Toronto: University of Toronto Press, pp. 167-194.

Wolhuter, C. C., Kangumu, B., & Mungongi, F. (2014). “Higher Education in Africa: Survey and Assessment.” Higher Education Forum 11, pp. 91-104.

World Bank. (2010). “Financing Higher Education in Africa.” Directions in Development; Human Development. World Bank. World Bank. (2014). “Improving the Quality and Quantity of Scientific Research in Africa.” World Bank Publications.

World Bank and Knowledge Consulting Ltd.. (2021). “Feasibility Study to Connect all African Higher Education Institutions to High-Speed Internet”. World Bank Publications - Reports 36044, The World Bank Group.

World Economic Forum. (2017). “There Are Not Enough Scientists in Africa. How Can We Turn This Around?” https://www.weforum.org/agenda/2017/05/scientists-are-the-key-to-africas future/

Xia, J. Xia, J., Harmon, J. L., Connolly, K. G., Donnelly, R. M., Anderson, M. R., & Howard, H. A. (2015). “Who Publishes in “Predatory” Journals?” Journal of the Association for Information Science and Technology, 66(7), pp. 1406-1417.

Xu, Xu, & Reed, Markum. (2021). “The Impact of Internet Access on Research Output - A Cross-Country Study.” Information Economics and Policy 56(C).

Ynalvez, M. A., & Shrum, W. M. (2011). “Professional Networks, Scientific Collaboration, and Publication Productivity in Resource-constrained Research Institutions in a Developing Country.” Research Policy 40(2), pp. 204-216.

Zuo, George W. (2021). “Wired and Hired: Employment Effects of Subsidized Broadband Internet for Low-Income Americans.” American Economic Journal: Economic Policy 13(3), pp. 447-82.

Appendix Table (1A): Using Subsample of the Pre-existing Universities Only

	(1)	(2)	(3)
	All Countries	Egypt and South Africa	
	Number of Publications		
Near Internet network × Post Submarine Cable	84.004***	222.158***	213.902***
	(8.688)	(32.971)	(34.000)
N	4,256	587	587
Controls	No	No	Yes
Country × Year FE	Yes	Yes	Yes
University FE	Yes	Yes	Yes
Mean of outcome	156.625	588.39	588.39

* $p<0.1$ ** $p<0.05$ *** $p<0.01$

Note: The Table reports the point estimates from Eq. (1) restricting the sample to universities constructed before the arrival of submarine cables or the construction of the Internet backbone network. Column (1) includes all the countries in the sample (Algeria, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Libya, Morocco, Nigeria, South Africa, Sudan, and Tanzania). There are 211 universities existing from the 1990s in the full sample. In columns (2) and (3), the sample is restricted to include only 2 countries (Egypt and South Africa). There are 32 universities existing before the arrival of submarine cables or the construction of the Internet backbone network in Egypt and South Africa alone. Column (3) includes additionally some universities' characteristics as controls: the total number of students enrolled, total faculty members with and without research responsibilities, student-to-faculty ratio, and the number of schools.

Appendix Table (2A): Using a Shorter Post-treatment Period

	(1)	(2)	(3)
	All Countries		
	Number of Publications		
	Full Sample Period	Restricting to 5 years after treatment	Restricting to 3 years after treatment
Near Internet network × Post Submarine Cable	86.849***	46.697***	35.017***
	(10.619)	(6.324)	(5.016)
N	5,519	3,435	2,828
Controls	No	No	No
Country × Year FE	Yes	Yes	Yes
University FE	Yes	Yes	Yes
Mean of outcome	130.479	81.963	70.21

* p<0.1 ** p<0.05 *** p<0.01

Note: Column (1) shows the main results from Table (2). In columns (2) and (3), I re-estimate Eq. (1) restricting the post period to five and three years, respectively, after the arrival of the submarine cables.

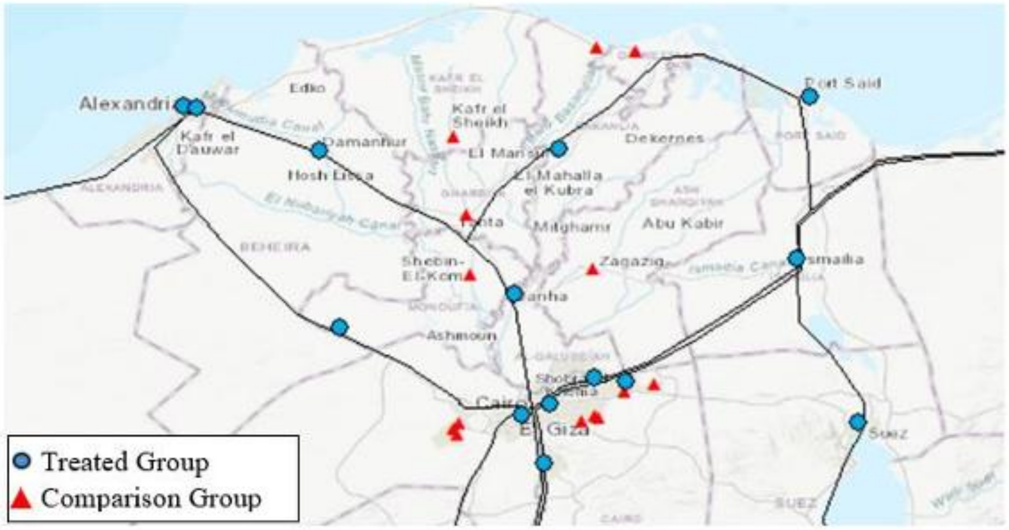
Appendix Table (3A): Fast Internet and Publications, Including Placebo Treatments

	(1)	(2)	(3)
	All Countries		
	Number of Publications		
Near Internet network × Post Submarine Cable	86.336***	85.804***	85.566***
	(11.116)	(10.614)	(11.082)
Near road network × Post Submarine Cable	4.098		2.023
	(9.460)		(9.691)
Near electric power × Post Submarine Cable		6.171	6.078
		(5.187)	(5.262)
N	5,519	5,519	5,519
Controls	No	No	No
Country × Year FE	Yes	Yes	Yes
University FE	Yes	Yes	Yes
Mean of outcome	130.479	130.479	130.479

* p<0.1 ** p<0.05 *** p<0.01

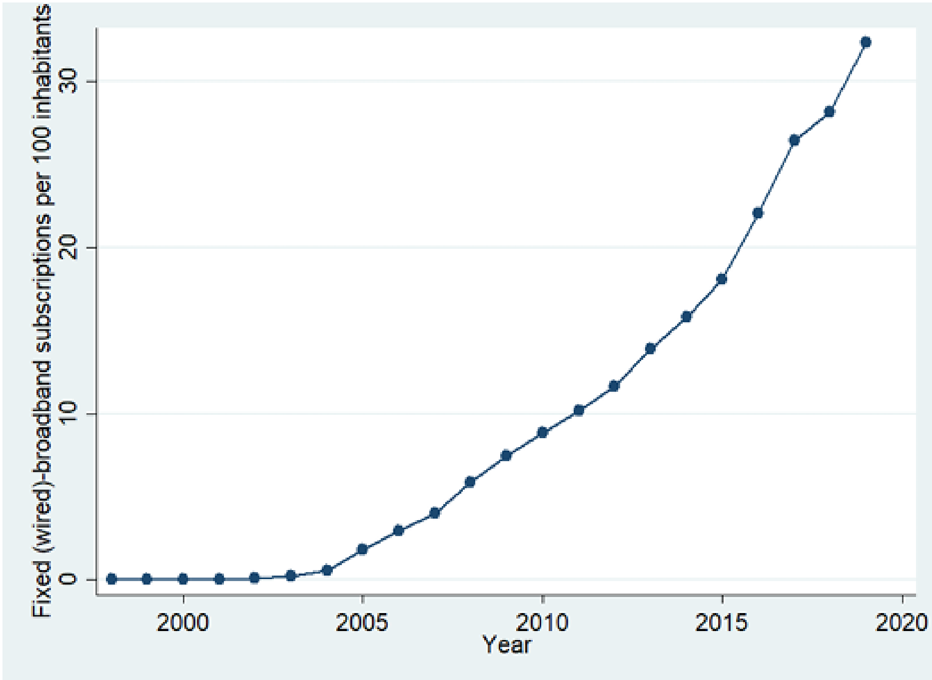
Note: Column (1) reports the point estimates from Eq. (1) adding the interaction term of *Near road network × Post Submarine Cable arrival*. Column (2) shows the results from Eq. (1) adding the interaction term of Near power station × Post Submarine Cable arrival. In column (3), all three interaction terms are included.

Appendix Figure (1A): Map of the Internet backbone Internet Network in Egypt



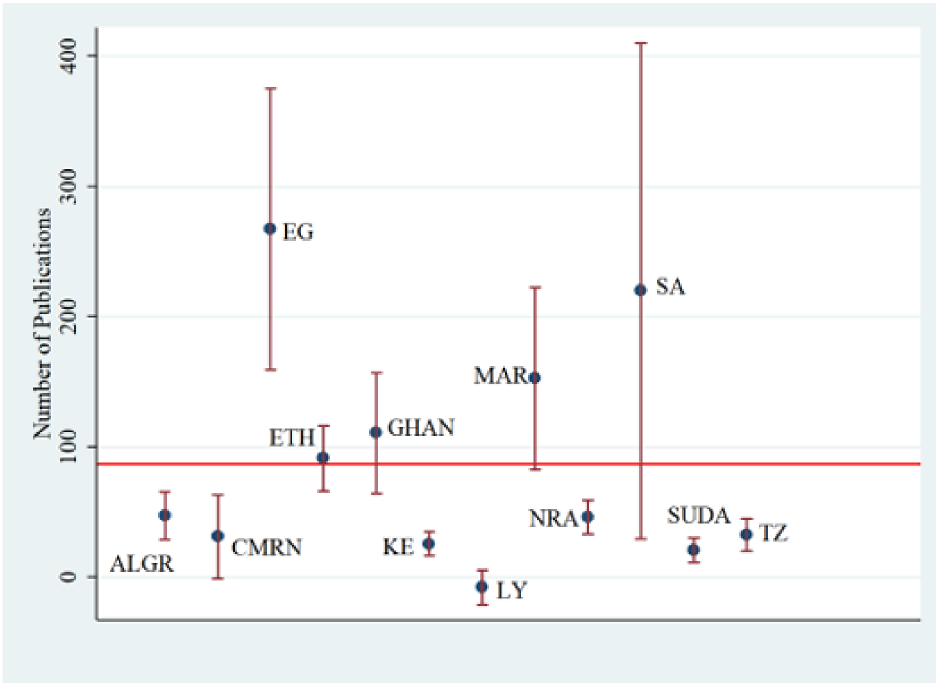
Note: The map depicts the locations of some of the treated and untreated universities from the Internet backbone network in Egypt. Blue dots correspond to the universities that are within five kilometers of the Internet backbone network (treated universities). Red dots correspond to the universities that are more than five kilometers away from the country's network.

Appendix Figure (2A): Over Time Variation in The Number of Fixed Broadband Subscriptions per 100 Inhabitant



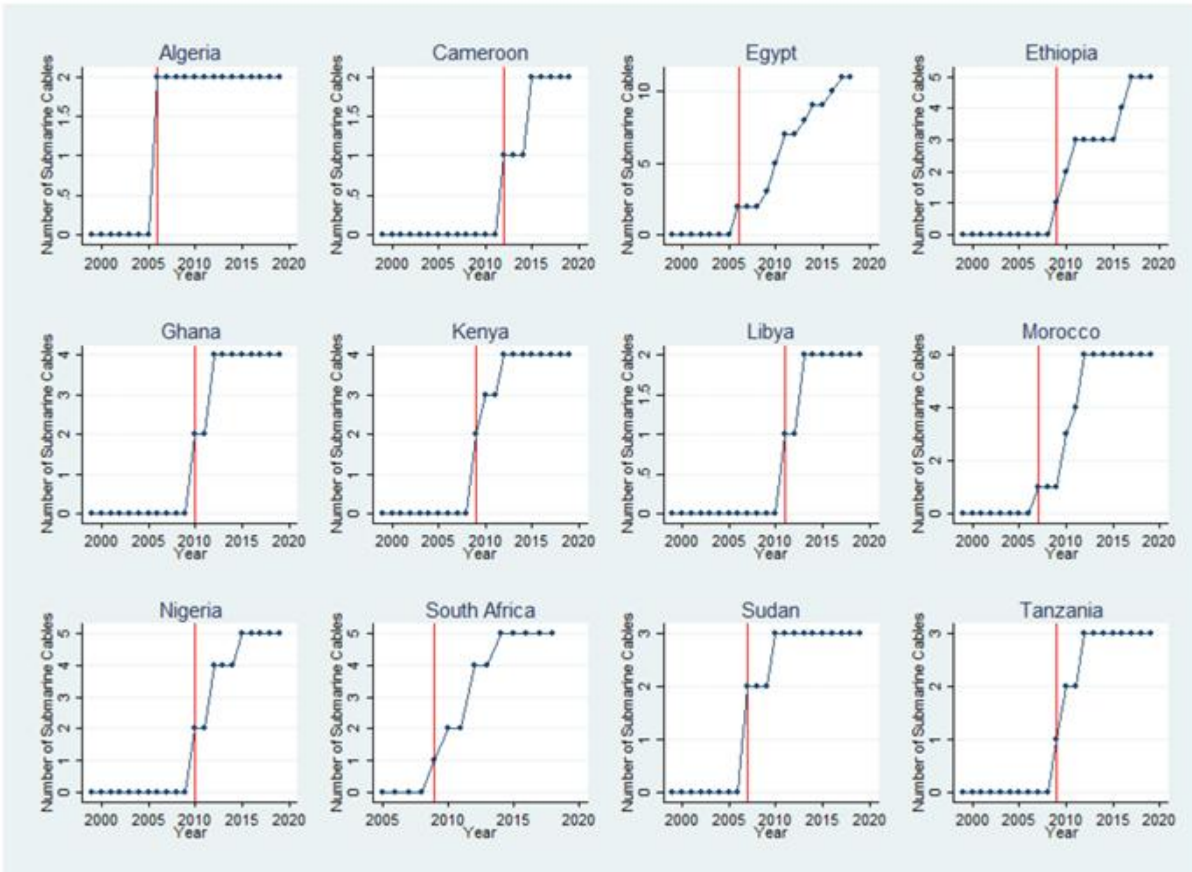
Note: The figure presents the evolution of the number of fixed broadband subscriptions per 100 inhabitants over years in the sample.

Appendix Figure (3A): The Impact of Fast Internet on The Number of Publications for Each Country Separately



Note: Blue dots represent point estimates of the impact of fast Internet for each of the 12 countries in the full sample using Eq. (1). The red line is the average impact of fast Internet on the number of publications obtained in Table (2). The symbols shown on the graph are as follows ALGR for Algeria, CMRN for Cameroon, EG for Egypt, ETH for Ethiopia, GHAN for Ghana, KE for Kenya, LY for Libya, MAR for Morocco, NRA for Nigeria, SA for South Africa, SUDA for Sudan, and TZ for Tanzania. The red lines contain 95% confidence intervals.

Appendix Figure (4A): The Variation in the Arrival Times of Submarines to Each Country



Note: The figure presents the number of modern submarine cables arriving in each country in the sample from 1999 to 2019.

Appendix 2: Chapter 2: Additional Analyses and Robustness Checks

Table A.4
Attitudes Toward NPPs in Locations With versus Without an NPP

	=1 if the respondent is aware of the TMI accident	=1 if the respondent approves of building more NPPs	=1 if the respondent believes more accidents like TMI are likely
Lives nearby an NPP	0.033***	-0.068*	0.061
	(0.010)	(0.038)	(0.037)
Female	-0.012	-0.223***	0.098***
	(0.009)	(0.036)	(0.035)
Non-white	-0.008	0.005	0.205***
	(0.027)	(0.062)	(0.060)
Age 30-44	0.019	0.06	-0.029
	(0.014)	(0.045)	(0.046)
Age 45-60	0.009	0.136***	-0.058
	(0.016)	(0.047)	(0.047)
Age 60 plus	0.014	0.016	0.031
	(0.020)	(0.064)	(0.063)
No. Adults in the HH	0.001	-0.012	0.005
	(0.006)	(0.020)	(0.021)
High School Graduate	0.031*	0.057	-0.068
	(0.019)	(0.049)	(0.046)
College Graduate	0.043***	0.116*	0.059
	(0.017)	(0.060)	(0.058)
Republican	-0.011	0.055	0.012
	(0.014)	(0.039)	(0.040)
Constant	0.930***	0.573***	0.505***
	(0.025)	(0.077)	(0.076)
N	1020	899	909

The table presents the OLS estimates obtained from the analysis of the CBS News/New York Times Poll in April 1979 data. The *Lives nearby an NPP* variable indicates whether the individual reports that s/he lives close to an NPP. Besides the listed personal attributes, the regressions also include city fixed effects (three-digit area telephone codes). Robust standard errors are shown in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

Table A.5
The Impact of Nuclear Power Plants on Employment and Wages
(Including all counties selected for an NPP)

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	4.620***	0.195***	3.683***
	(1.722)	(0.050)	(1.303)
Commercial Operation	-1.042	0.032	0.187
	(1.173)	(0.034)	(0.810)
N	6732	6732	6732
Mean of the outcome	48.655	17.407	17.407

Replication of Table 10 using the complete set of counties with an NPP.

Table A.6
The Impact of Nuclear Power Plants on Employment and Wages
(Excluding the counties that had an additional construction phase post-1979)

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	13.878**	0.448***	9.649**
	(5.765)	(0.135)	(3.852)
Commercial Operation	5.994	0.296***	4.744**
	(4.468)	(0.103)	(2.344)
N	2703	2703	2703
Mean of the outcome	48.172	16.695	16.695

Replication of Table 10 excluding observations from the four counties with construction phases in 2010s.

Table A.7
The Impact of Nuclear Power Plants on Employment and Wages
(With Additional Controls)

Panel A: Controlling for County-Specific Time Trends

	(1)	(2)	(3)	(4)	(5)	(6)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	9.961**	0.336***	7.501**	9.964**	0.336***	7.504**
	(3.830)	(0.105)	(2.960)	(3.832)	(0.106)	(2.961)
Commercial Op.	-2.562	0.004	-2.440	-2.510	0.005	-2.394
	(4.123)	(0.077)	(2.681)	(4.096)	(0.076)	(2.660)
N	2907	2907	2907	2907	2907	2907
County Trends	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic

Replication of Table 4 with county-specific time trends.

Panel B: Including Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	7.628***	0.288***	6.110**	7.632***	0.288***	6.112**
	(2.805)	(0.089)	(2.334)	(2.805)	(0.089)	(2.334)
Commercial Op.	-1.907	0.086	0.306	-1.900	0.086	0.311
	(3.479)	(0.080)	(2.479)	(3.476)	(0.080)	(2.478)
N	2804	2804	2804	2804	2804	2804

Replication of Table 10 with the additional control variables. In all regression, the percentages of blacks, people above 65, and women in the county population are included. In columns 1-3 (4-6), the following variables interacted with a linear (quadratic) time trend are additionally controlled for: the proportions of foreigners in 1960 and the population with at least a high school degree in 1960 in the county, median rental price in 1960, the share of the houses occupied by the owners in 1960, and the median family income in 1960.

Panel C: Including State-by-Year Fixed Effects

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	13.009*	0.410***	9.705**
	(6.574)	(0.137)	(4.710)
Commercial Op.	-0.189	0.189**	2.572
	(3.433)	(0.086)	(2.071)
N	2244	2244	2244

Replication of Table 10 with state-by-year fixed effects.

Table A.7 Continued

Panel D: Controlling for Manufacturing Establishments, Population Change, and Other Energy-Producing Plants

	(1)	(2)	(3)
	Emp. Rate	Log PC Wages	PC Wages
Construction	11.469**	0.380***	8.395**
	(4.657)	(0.116)	(3.218)
Commercial Operation	3.098	0.220**	3.396
	(4.130)	(0.098)	(2.374)
Other Utility Plants (=1 if exists)	4.435	0.273**	4.008
	(3.623)	(0.121)	(2.452)
No. Manuf. Establishments	0.001	0.000	-0.000
	(0.001)	(0.000)	(0.000)
Pop. Change (from t-1)	-0.004	0.000	0.009
	(0.019)	(0.000)	(0.009)
N	2832	2832	2832

Replication of Table 10 with additional control variables.

Table A.8
The Impact of Nuclear Power Plants on Employment and Wages
(Excluding Major Post-Accident Periods)

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Excluding Post-Fukushima (2011)			
Construction	11.136**	0.360***	7.832**
	(4.650)	(0.115)	(3.208)
Commercial Operation	2.811	0.190**	2.698
	(3.654)	(0.092)	(2.033)
N	2451	2451	2451

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Excluding Post-Chernobyl (1986)			
Construction	10.533***	0.393***	7.735***
	(3.790)	(0.111)	(2.547)
Commercial Operation	-0.039	0.029	-0.384
	(2.966)	(0.105)	(1.878)
N	1026	1026	1026

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Excluding Post-TMI (1979)			
Construction	9.766**	0.339***	6.610**
	(4.280)	(0.112)	(2.666)
N	627	627	627

Replication of Table 10 excluding particular time periods from the sample.

Table A.9
The Impact of Additional Nuclear Reactors on Employment and Wages

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction of 1st reactor	5.755**	0.268**	4.487*
	(2.815)	(0.124)	(2.410)
Construction of 2nd reactor	5.340	0.113	2.809
	(4.105)	(0.137)	(3.537)
Construction of 3rd+ reactors	16.188	0.344	9.769
	(13.601)	(0.303)	(7.502)
Comm. Op. of 1st reactor	-0.388	0.140	1.604
	(4.540)	(0.130)	(3.190)
Comm. Op. of 2nd reactor	-0.080	0.071	-1.210
	(5.309)	(0.127)	(3.891)
Comm. Op. of 3rd+ reactor	14.193*	0.267	10.516**
	(8.176)	(0.169)	(4.169)
N	2907	2907	2907
Mean of the outcome	48.376	16.844	16.844

Estimates are obtained from estimating equation (2.3) with dummies for the construction and operation of the first, second, and third reactor of the NPP.

Table A.10
A Falsification Test

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	-1.046	0.010	-0.325
	(0.820)	(0.026)	(0.350)
Commercial Operation	-0.152	0.035	0.929
	(1.797)	(0.059)	(1.172)
N	2907	2907	2907
Mean of the outcome	48.882	15.589	15.589

Results are obtained from the falsification test described in section 2.5.2.

Table A.11
The Impact of Nuclear Power Plants on Employment and Wages
(Excluding Specific Control Counties Where NPP Projects Were Canceled)

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
A. Excluding control counties where NPPs were canceled due to low energy demand forecasts			
Construction	10.787**	0.369***	8.199**
	(4.476)	(0.117)	(3.237)
Commercial Operation	3.496	0.211*	3.169
	(4.669)	(0.105)	(2.562)
N	1938	1938	1938

	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
B. Excluding control counties where NPPs were canceled due to unanticipated cost increases			
Construction	10.955**	0.377***	8.319**
	(4.490)	(0.118)	(3.245)
Commercial Operation	3.614	0.198*	3.055
	(4.658)	(0.106)	(2.559)
N	1938	1938	1938

	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
C. Excluding control counties that were excluded in Panel A or B			
Construction	10.286**	0.352***	7.996**
	(4.394)	(0.117)	(3.256)
Commercial Operation	3.734	0.197*	3.006
	(5.158)	(0.112)	(2.780)
N	1683	1683	1683

	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
D. Excluding all control counties			
Construction	7.732*	0.259**	6.990*
	(3.924)	(0.121)	(3.552)
Commercial Operation	5.979	0.272*	3.156
	(7.769)	(0.145)	(4.113)
N	1173	1173	1173

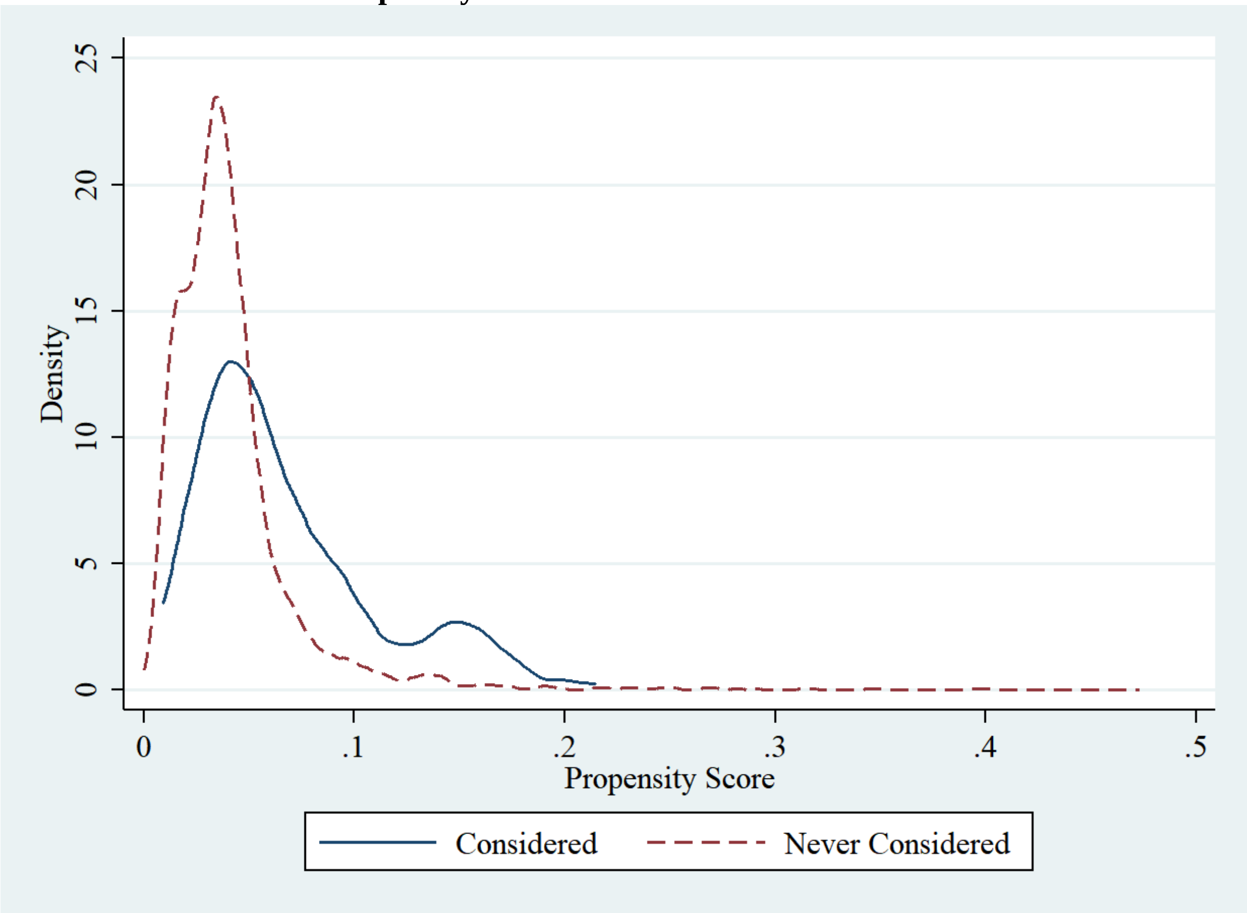
Replication of Table 10 excluding various groups of control counties.

Table A.12
The Impact of Nuclear Power Plants on Employment and Wages
(Commuting Zone Level)

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	1.181*	0.062**	1.057**
	(0.647)	(0.024)	(0.456)
Commercial Operation	1.306	0.049	1.194
	(1.082)	(0.038)	(0.820)
N	2244	2244	2244
Mean of the outcome	52.509	19.200	19.200

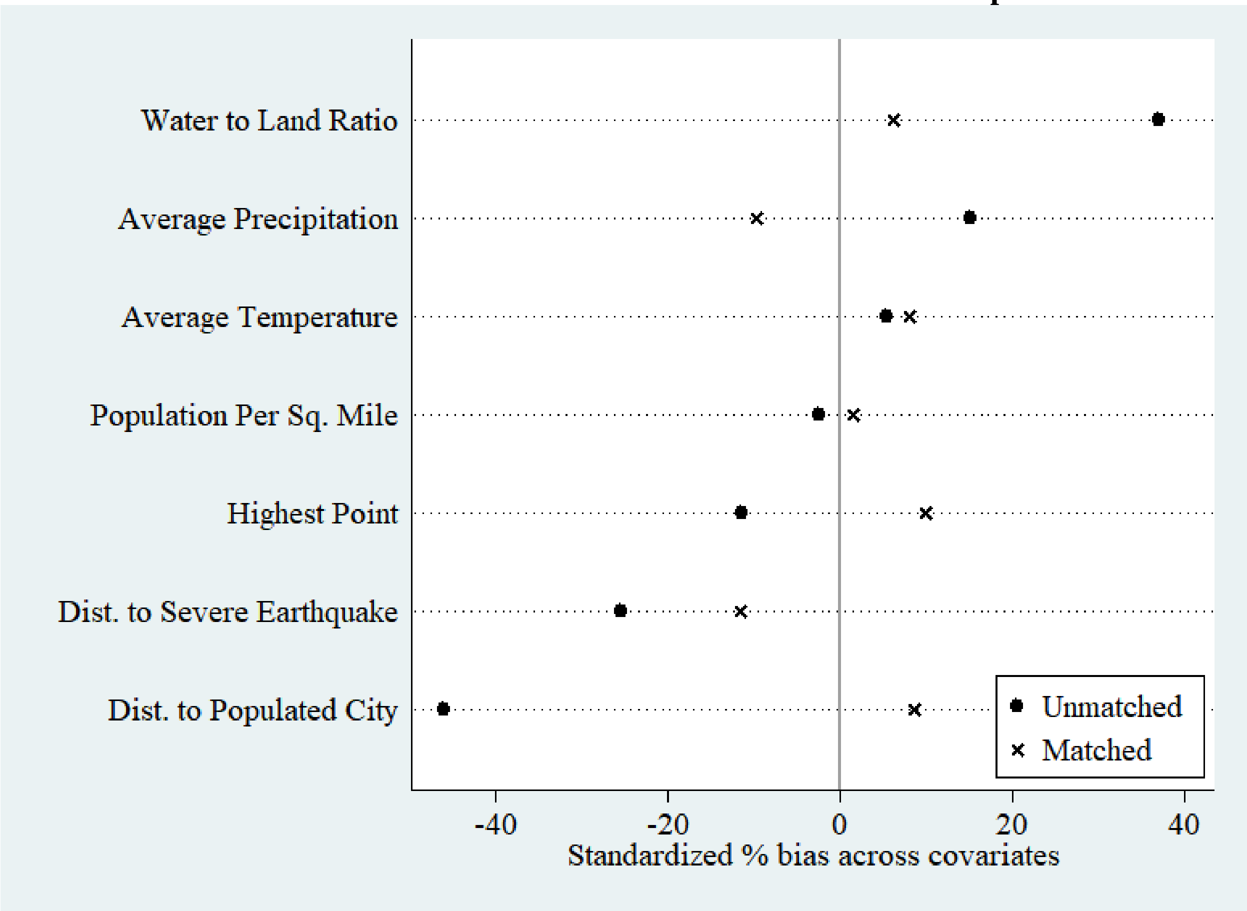
Results are obtained from Equation 2.3 that is estimated at the commuting zone level. See notes in Table 10.

Figure A.5
The Distribution of Propensity Scores in Considered vs Never Considered Counties



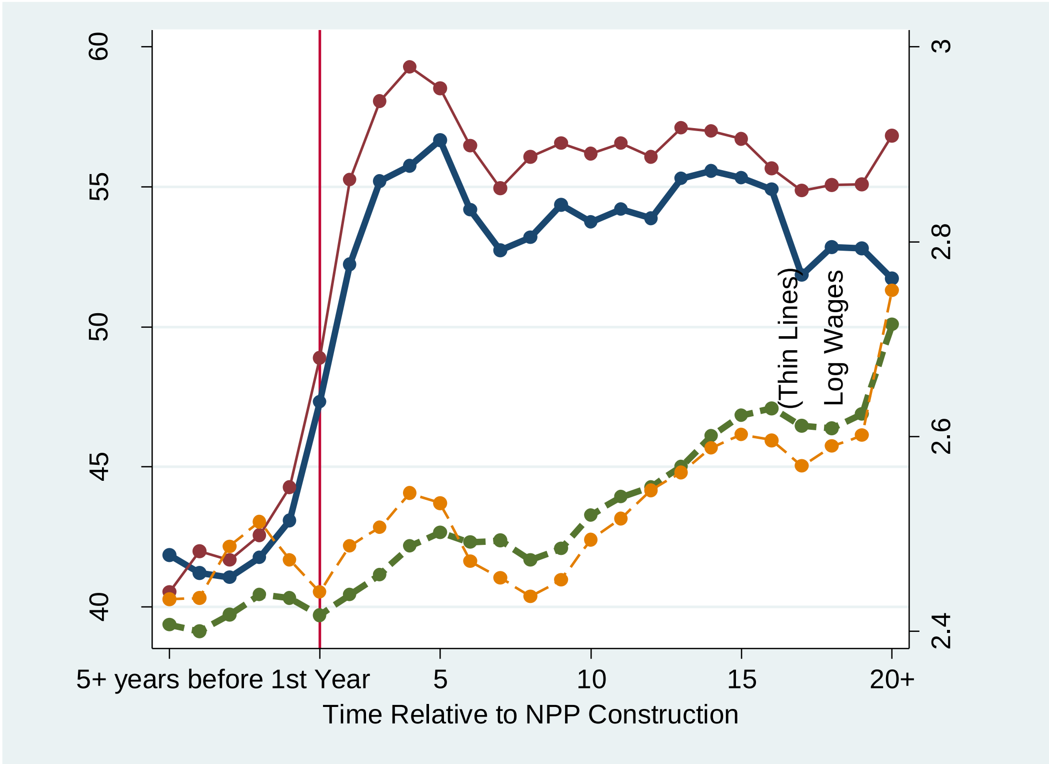
The figure presents the distribution of propensity scores obtained from Equation 2.1.

Figure A.6
Covariate Balance in Matched and Unmatched PSM Samples



The figure presents the differences (as a percentage of standard deviations) in the predictor variables between the considered and never considered counties in the matched and unmatched samples for the PSM estimation reported in Table 7.

Figure A.7
Trends in Raw Means of the Outcomes in the Treated and Control Counties



The figure depicts the averages of the employment rate (thick lines, measured on the left axis) and log of per capita wages (thin lines, measured on the right axis) over time in the treated (solid lines) and control (dashed lines) counties. For the treated counties, time is standardized at the year of the commencement of the first nuclear reactor (red vertical line) in the county. "5+ years before" is the average of all years more than four years before the construction's start, and "20+" is the average of all years after the nineteenth year after construction's start. For control counties, "1st Year" indicates 1975, and other years on the time axis are numbered accordingly (5 means 1980, 10 is 1985, etc.).

Figure A.8
Event Study Analysis Using the Chaisemartin and D'Haultfœuille (2023) Estimator (Event: Construction, Outcome: Employment-to-Population Ratio)

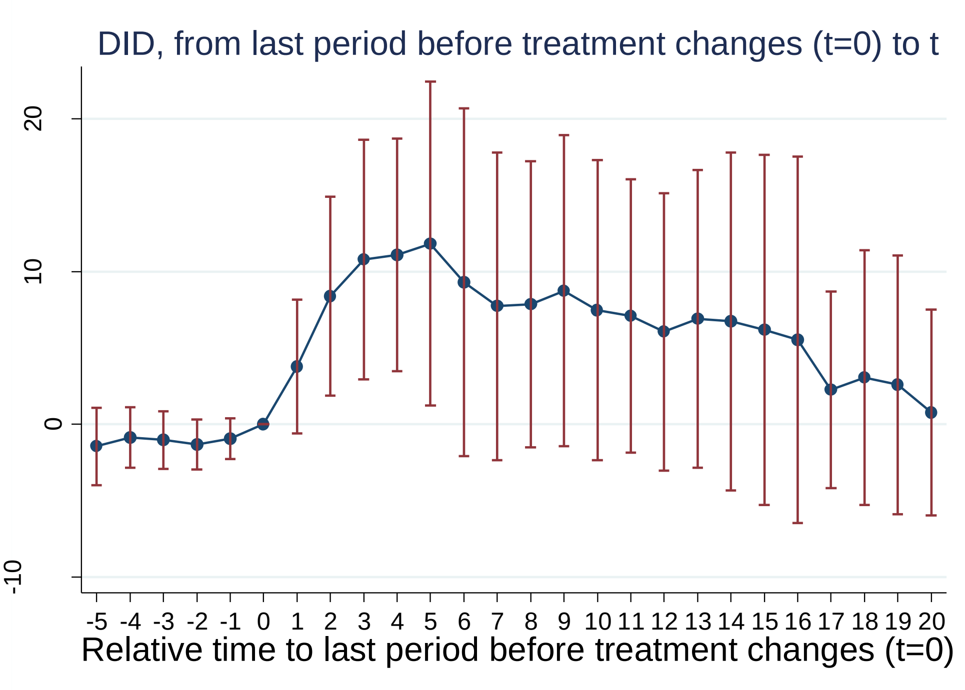


Figure A.9
Event Study Analysis Using the Chaisemartin and D'Haultfœuille (2023) Estimator (Event: Commercial Operation, Outcome: Employment-to-Population Ratio)

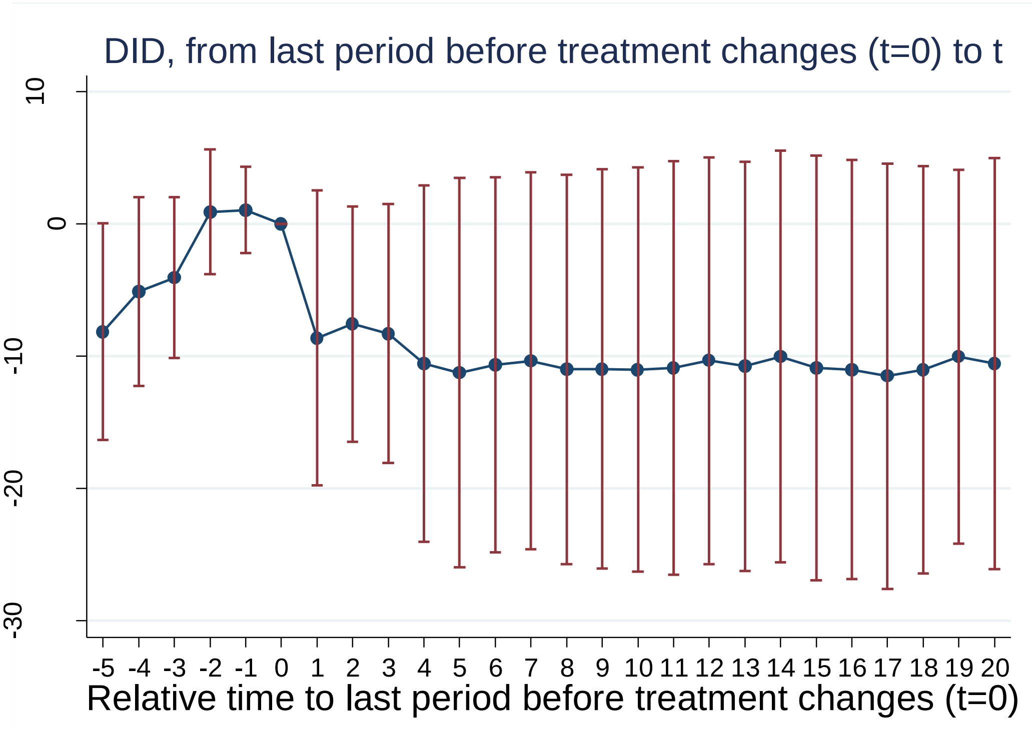


Figure A.10
Event Study Analysis Using the Chaisemartin and D'Haultfœuille (2023) Estimator
(Event: Construction, Outcome: Log Per Capita Wages and Salaries)

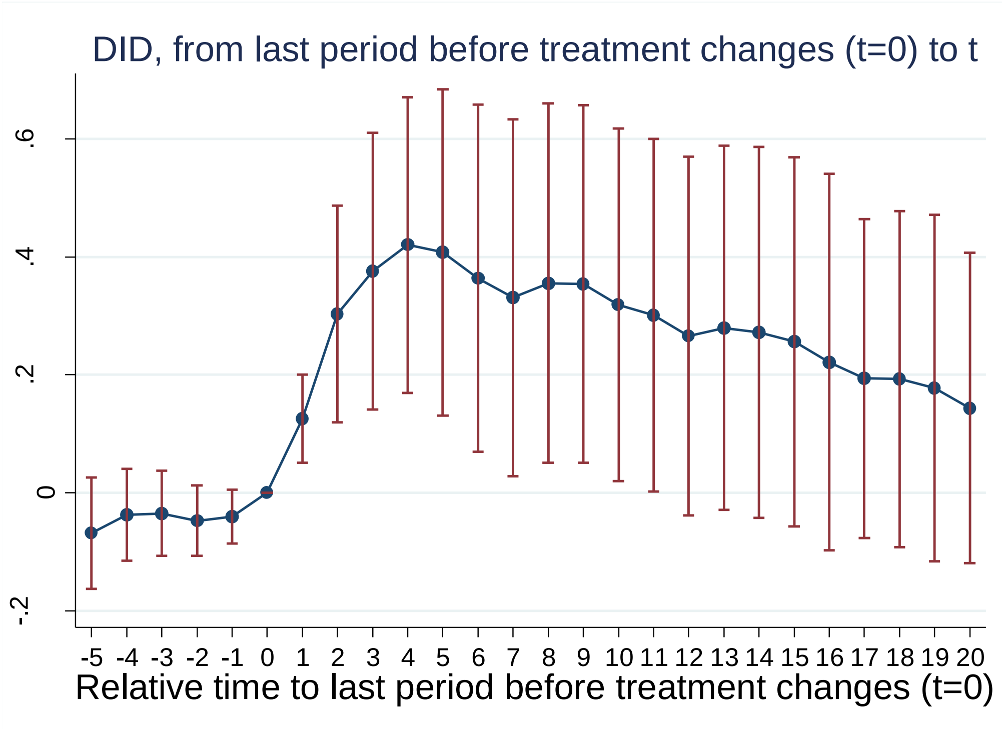
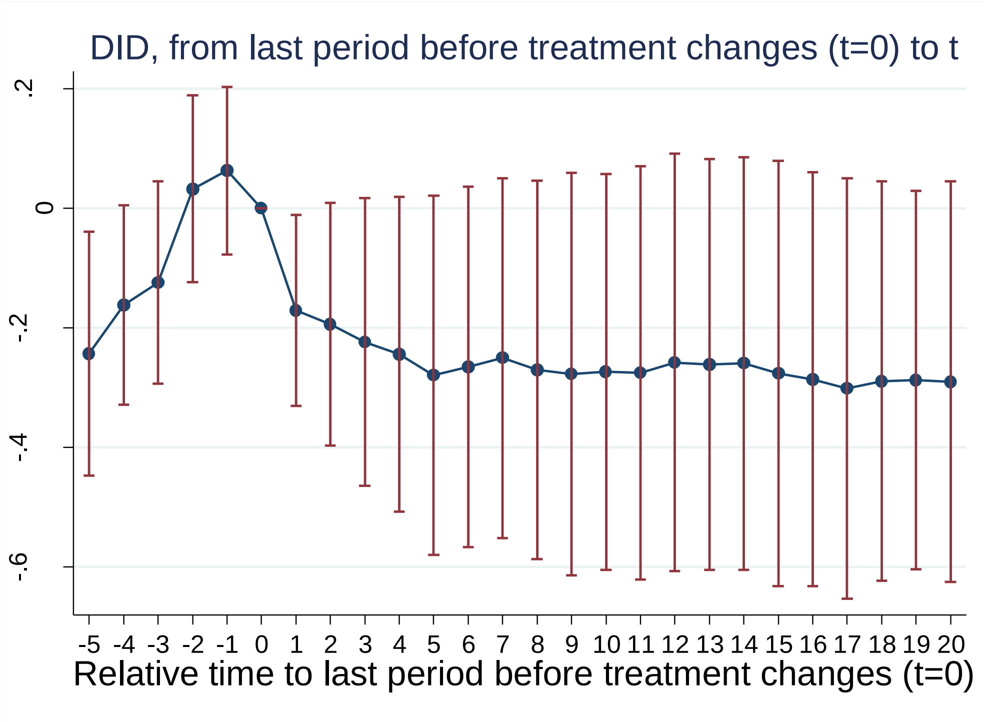


Figure A.11
Event Study Analysis Using the Chaisemartin and D'Haultfœuille (2023) Estimator
(Event: Commercial Operation, Outcome: Log Per Capita Wages and Salaries)



Appendix 3: Chapter 2: Synthetic Control Estimates

In this appendix section, we estimate the impact of NPPs on labor market outcomes using a synthetic control (SC) strategy. As described in Abadie (2021), the SC method may be suitable for this setting (in addition to the DD method) for a number of reasons. First, SC is appropriate for analyzing outcomes from aggregate entities, such as counties. Second, the abrupt alteration in construction and operations of NPPs triggered by TMI, together with the temporal variation in construction and operation phases, furnishes a rich treatment structure well-suited for the SC methodology. Also, in most cases, an NPP likely represents a "large intervention" relative to the size of the local economy, in which SC is more likely to detect an effect. Finally, because the treatment intensity, i.e., the magnitude of construction and operations, may vary greatly across counties, the best counterfactual counties may differ across treated counties. Thus, a case study-like comparison could be more appropriate.

One issue with using the SC method in this context is the small number of pre-intervention periods. SC requires a sizable number of pre-intervention periods to match the outcome path (Abadie, 2021). However, because the county-level employment and wage outcomes data are available on the BEA website since 1969, and our sample includes counties whose NPP construction started in 1974-1979, we only have a handful of pre-treatment periods. Nevertheless, we proceeded with the SC estimation, given that in several past SC applications, the number of pre-treatment periods ranged between 5 and 8 (Ferman et al. 2020).[109]

We implement the SC strategy using only the counties that were ever considered for an NPP. This is because the considered counties constitute a homogeneous group with similar attributes that are not correlated with the probability of cancelation (see Table 7 in the main text and its discussion in the Background section). Some of the considered counties ultimately experienced the construction of an NPP. They form the treatment group. The remaining considered counties (where an NPP's construction never started because the project was canceled) are in the donor pool. For each of the 23 treated counties whose NPP constructions commenced between 1974 and 1978, we construct a synthetic counterfactual county constructed as a weighted average of the 34

[109] Bartel et al. (2018) uses five pretreatment periods and included all outcome lags to construct the synthetic controls. Lindo and Packham (2017) and Dustmann et al. (2017) take into account 6 pre-treatment periods. Baccini et al. (2014) and Gobillon and Magnac (2016) use 7 and 8 pre-treatment periods, respectively, to create the synthetic controls. However, there are others, such as Abadie et al. (2010), who use a larger number of pre-intervention years.

donor counties with canceled NPP projects. The weights are determined using the algorithm described in Abadie et al. (2010) so that the pre-treatment outcomes of the counterfactual county closely track the pre-treatment outcomes of the treated counties. The treatment effect is then estimated using the treated county's deviations from the counterfactual county's trajectory in the post-treatment period.

Appendix Table B.1 presents the means of the yearly outcome variables in the pre-treatment period (1969-1973) for various groups of counties. Columns 1, 2, and 3 (which are identical to Table 3 in the main text) show the means for the group of treated counties (those in which an NPP was constructed), donor/control counties (those that were considered for an NPP but never got one), and the rest of the US. Column 4 contains the averages of the variables in the synthetic control counties. The stars indicate whether the average in a sample is statistically different from the one in column 1. We can infer from Appendix Table B.1 that the synthetic control counties are similar to the treated counties in terms of employment-to-population ratio and per capita wages before the treatment. This result indicates that the estimation of weights for the counterfactual counties was successful on average. We present the averages of the remaining variables used for the estimation of synthetic control weights in Appendix Table B.2. These county characteristics include the pre-treatment 1969-1973 average of the yearly employment-to-population ratio, log of per capita real wages, the percentage of Whites, the percentage of the population 65 years and older in the county, as well as the county's housing and labor market characteristics from the 1950, 1960, and 1970 Censuses.[110]

The results of the match of treated counties with their counterfactuals can be visualized in graphs such as Appendix Figures B.1 and B.2, where we depict the employment-to-population ratio and the per capita wages in Burke County, Georgia (solid black line) and its synthetic counterfactual (dashed black line). The NRC permitted the construction of two reactors in Burke County for the first time in its history in 1974. The construction started in the same year. This is represented by the vertical dotted line on the figures. The first and second reactors' constructions were completed in 1986 and 1988 (solid red lines) when they started commercial operations. The

[110] For two counties (Tishomingo, MS and Trousdale, TN), we were unable to find some of the socioeconomic characteristics, such as the percentage of foreign-born individuals in the county. Therefore, the estimation of these counties' synthetic controls does not include these variables. Also, there were two other counties (West Feliciana, LA and Claiborne, MS) that could not be analyzed because of technical errors in the calculation; for example, their Hessian matrices are unstable or asymmetric, or Stata cannot calculate their numerical derivatives.

county started building a third reactor in 2012 (green vertical line), which had not been completed as of 2019. Both Appendix Figures B.1 and B.2 show that the outcomes (employment-to-population ratio and per capita wages) in Burke and its counterfactual are identical or very close before 1974. However, during the construction of its nuclear reactors between 1974 and 1988, Burke County experienced a dramatic increase in both employment and wages. These gains appear to be reversed immediately after the completion of the projects. Wages and employment in Burke County have been increasing compared to its synthetic control since 2012 during the construction of its new reactor.

Appendix Figures B.3 and B.4 provide the employment-to-population ratios of two other counties where the NPP construction started but was never completed, namely Tishomingo, MS, (B.3) and Cherokee, SC (B.4). The project in Tishomingo (Cherokee) commenced in 1978 (1977), and the project was abandoned partially completed in 1984 (1983) due to economic and regulatory challenges. Both figures show that these counties' employment outcomes almost overlap with their synthetic controls before the treatment. During the construction periods, NPPs elevated employment in these counties. The magnitude of the increase was smaller in Cherokee. After the cancellation of the NPP projects, the outcomes returned to the baseline levels.[111]

Similar exercises are implemented 23 times for each of the treated counties. The complete list of 23 treated counties, along with the description of their synthetic controls, i.e., which donor county contributed to the construction of their counterfactual, are in Appendix Table B.3. To aggregate and average the estimates from all of the treated counties, we estimate a regression of the form below:

(B1) $$Y_{ct} = \beta_1 Construction_{ct} + \beta_2 Commercial\ Operation_{ct} + \gamma_{mc} + \tau_t + u_{ct},$$

[111] The construction and cancellation of the NPPs in these counties are covered by the local and national media. For example, see the news pieces at the following links: https://www.nytimes.com/1984/08/30/business/tva-cancels-4-reactors-knoxville-tenn-aug-29.html; https://www.youtube.com/watch?v=1kQv6jLIAaE. In these canceled reactor sites, other endeavors took advantage of the infrastructure built. For example, portions of Abyss, a 1989 James Cameron movie that won an Academy Award in 1990, were filmed in the reactor containment building of the NPP in Cherokee, SC (https://www.telegraph.co.uk/films/0/abyss-how-james-camerons-underwater-thriller-almost-killed). Similarly, in the late 1980s, NASA began an establishment to build rocket motors in close proximity to the canceled NPP site in Tishomingo, MS. This initiative was dead-ended as Congress stopped funding in 1993 (https://www.nytimes.com/1993/11/29/us/hit-twice-town-feels-misled-by-us.html). Our analyses show that neither of these projects impacted employment in these counties.

where the unit of observation is a county-year (ct). To compare treated counties to their synthetic counterfactual counties, we include a matched pair dummy, γ_{mc}, in the specification. For example, regressions include a dummy that equals one for Burke County and its synthetic control county, another dummy that takes the value of one for another treated county and that county's counterfactual county, and so on. Thus, this regression includes the yearly observations of 23 treated counties and 23 synthetic control counties between 1969 and 2019, i.e., 46 counties × 51 years = 2,346 observations. The right-hand side variables also contain year fixed effects, denoted by τ. Finally, u stands for the error term. From this analysis, we draw statistical inference using bootstrap standard errors clustered at the match-pair level.[112]

The estimates obtained from Equation B1 are presented in Appendix Table B.4. The coefficient of the *Construction* dummy in column 1 is 10.2, indicating that when a nuclear reactor is being built in a county, the employment-to-population ratio in that county goes up by about ten percentage points. This corresponds to a 20% increase from the mean employment-to-population ratio in our sample (50.1 percentage points). The estimate of *Commercial Operation* is not statistically different from zero. Column 2 shows that per capita wages in a county go up both during the construction of the NPP and its commercial operations. The effect of *Construction* (*Commercial Operation*) corresponds to a 44% (32%) increase in per capita wages.[113] Results in column 3, where the outcome is the level of per capita wages, are similar to those in column 3.[114]

We implement a series of "leave-one-out" tests, in which we re-estimate our SC models, omitting the donor counties from the analysis one by one. Specifically, we recalculated the weights for the synthetic control of each treated county, excluding one of the 34 donor counties, and estimated Equation B1. This results in 34 separate samples for each outcome. The coefficients of the *Construction* and *Commercial Operation* dummies from the 34 regressions that are obtained from these samples in Employment-to-Population Ratio and Wages per capita regressions are displayed in Appendix Figures B.5, B.6, B.7, and B.8. As shown in Appendix Figure B.5, in all of the regressions, the estimate of the *Construction* indicator for the employment-to-population ratio is stable at around 10. However, this is not the case for the *Commercial Operation* dummy, point

[112] The results are robust to clustering at the county level and wild-t bootstrap clustering.
[113] We computed these figures by $100 \times (\exp(\text{Estimate}) - 1)$.
[114] We use separate weights for each outcomes based on the matching of the corresponding outcomes. We compared the counties that constitute the synthetic counterfactual for each outcomes. In the majority of cases, these counties are identical but their weights are different. When we repeated our analysis using the weights we computed for the other outcome (e.g., employment weights for the wage outcome and vice-a-versa) our main inference did not change.

estimates of which display greater variation (Appendix Figure B.6). In addition, the *Commercial Operation* confidence intervals contain zero several times. A similar picture arises for the Wages per capita outcome (Appendix Figures B.7 and B.8). The point estimates of *Construction*, but not *Commercial Operation*, are not sensitive to the set of donor counties.

We also conduct a falsification test, in which we compare the outcomes of donor counties with those of other donor counties. Specifically, we start with the first donor county (i.e., no NPP construction took place in this county) and assume that an NPP construction started in 1974 in that county. We also assume the construction period lasted ten years, after which the commercial operations commenced. Then, we estimate this donor county's synthetic control using the remaining 33 donor counties. We repeat the same exercise for the second, third, and remaining donor counties, assuming that an NPP was constructed between 1974 and 1984 (ten years) and there was a commercially operational NPP after 1984. We then estimate Equation B1 using this sample. We repeated this analysis assuming a donor county was treated in 1975, 1976, 1977, and 1978. Thus, we have five separate regressions.

Since we are comparing untreated counties with other untreated counties, we expect the coefficients of *Construction* and *Commercial Operation* dummies to be zero. Appendix Figures B.9, B.10, B.11, and B.12 show that this is, indeed, the case. In all of the regressions, the point estimates of the *Construction* and *Commercial Operation* variables are small, and the 95% confidence intervals include zero. These results suggest that the paths of the employment and wage outcomes of the counties in the donor pool are alike, and our baseline results are not driven by the peculiarities in the counties in the donor pool.

Overall, the SC results in this appendix point to similar conclusions that are drawn from the DD results described in the main text. The construction of an NPP improves the employment and wages in a county. Baseline SC estimates indicate that commercial operations increase wages (but not employment). However, this is not a robust result.

Additional References for Appendix 3

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." Journal of Economic Literature 59 (2): 391–425.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." Journal of the American Statistical Association 105 (490): 493–505.

Baccini, Leonardo, Quan Li, and Irina Mirkina. 2014. "Corporate Tax Cuts and Foreign Direct Investment." Journal of Policy Analysis and Management 33 (4): 977–1006.

Bartel, Ann P, Maya Rossin-Slater, Christopher J Ruhm, Jenna Stearns, and Jane Waldfogel. 2018. "Paid Family Leave, Fathers' Leave-taking, and Leave-sharing in Dual-earner Households." Journal of Policy Analysis and Management 37 (1): 10–37.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler. 2017. "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment." The Quarterly Journal of Economics 132 (1): 435–83.

Ferman, Bruno, Cristine Pinto, and Vitor Possebom. 2020. "Cherry Picking with Synthetic Controls." Journal of Policy Analysis and Management 39 (2): 510–32.

Gobillon, Laurent, and Thierry Magnac. 2016. "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls." Review of Economics and Statistics 98 (3): 535–51.

Lindo, Jason M, and Analisa Packham. 2017. "How Much Can Expanding Access to Long-Acting Reversible Contraceptives Reduce Teen Birth Rates?" American Economic Journal: Economic Policy 9 (3): 348–76.

Appendix Table B.1
Summary Statistics of the Outcomes in the Pre-treatment Period (1969-1973)

	(1)	(2)	(3)	(4)
	Treated	Control	Rest of the US	Synthetic Controls
Employment-to-Population Ratio	40.785	39.595	40.977	40.269
Per capita wages and salaries	12.615	12.77	10.821***	12.627
Per capita personal income	24.582	25.765	23.447*	25.652
Total observations	115	170	14791	115

The first three columns are identical to Table 9 in the main text. See notes in Table 9 in the main text. Column 4 shows the means of the synthetic control counties.

Table B.2
Summary Statistics of the Variables Used in the Synthetic Control Method

	(1)	(2)	(3)	(4)
	Treated	Control	Rest of US	Synthetic Control
% of white (1969-1973)	89.810	90.127	89.824	87.839
% of 65 and older (1969-1973)	10.704	10.873	12.309	10.758
% of Non-White				
1950	11.765	11.321	10.708	14.517
1960	11.013	10.591	10.637	13.328
1970	10.460	9.862	10.190	12.146
% of Urban				
1950	35.121	38.025	27.679	39.701
1960	41.135	41.694	31.526	43.497
1970	45.270	46.291	34.165	
% of Foreign Born				
1950	2.701	4.239	2.767	3.189
1960	2.048	3.297	2.172	2.554
Schooling: % of Pop. 25+				
1950				
with completed HS	26.891	29.809	27.128	28.489
with 4 years of college	4.718	4.587	4.133	4.563
1960				
with 12+ years of schooling	34.309	37.365	34.570	36.997
1970				
with 12+ years of schooling	44.791	49.359	44.578	48.008
with 4+ years of schooling	7.804	8.126	7.227	8.272

Table B.2 Continued

	(1)	(2)	(3)	(4)
	Treated	Control	Rest of US	Synthetic Control
Log of Median Rent				
1950	3.463	3.521	3.440	3.450
1960	3.995	4.084	3.954	4.052
1970	4.446	4.500	4.343	4.501
% Dwellings Owner Occ.				
1950	59.713	60.674	61.051	60.552
1960	67.020	67.380	67.070	67.413
1970	70.352	70.121	70.825	70.110
Income less than				
$2000 (1950)	41.900	37.656	44.689	39.268
$3000 (1960)	32.913	27.276	35.976	28.698
$3000 (1970)	13.922	12.482	16.839	13.304
% of Low Income families				
1970	14.339	12.797	17.423	13.850
% in LF				
1950	36.552	36.872	35.973	37.182
1960	36.270	36.183	35.366	35.881
% of LF in Agriculture				
1950	26.426	21.424	32.711	21.479
1960	14.355	12.232	20.849*	12.606
% Unemployed				
1950	3.485	4.360	3.640	4.062
1960	4.217	5.518**	5.211**	4.961
1970	4.091	4.876	4.572	4.512

The table presents the means of the variables used in the construction of the synthetic control counties. See notes in Table B.1.

Table B.3
Treated Counties and Their Synthetic Controls

Treated County	Weights (Emp. To Pop. Ratio)	Weights (Log of Per Capita Wages & Salaries)
Jackson AL	Davie NC (0.035), Jasper IA (0.037), Roane TN (0.291), Dallas TX (0.156), Chilton AL (0.482)	Washington RH (0.007), Stanislaus CA (0.023), Roane TN (0.248), Chilton AL (0.311), Gilliam OR (0.411)
Maricopa AZ	Gilliam OR(0.019), Charles MD (0.038), Jefferson LA (0.087), Washington RI (0.183), Riverside CA (0.257), Orange FL (0.416)	Stanislaus CA (0.116), Orange FL (0.142), Jefferson LA (0.151), Washington RH (0.221), Los Angeles CA (0.328)
Burke GA	Austin TX (0.137), Newton TX (0.165), Dallas AL (0349, Chilton AL (0.349)	Austin TX (0.171), Dallas AL (0.337), Newton TX (0.492)
De Witt IL	Los Angeles CA (0.018), Lancaster PA (0.02), Chilton AL (0.05), Gilliam OR (0.071), Carroll IL (0.128), Sheboygan WI (0.196), Cayuga NY (0.237), Rogers OK (0.28)	Riverside CA (0.062), Elmore AL (0.091), Gilliam OR (0.183), Rogers OK (0.327), Sheboygan WI (0.336)
Ogle IL	Jefferson LA (0.088), Gilliam OR (0.156), Charles MD (0.184), Jasper IA (0.211), Sheboygan WI (0.362)	Washington RH (0.004), Sheboygan WI (0.129), Carroll IL (0.138), Jasper IA (0.209), Gilliam OR (0.236), Lancaster PA (0.243)
Will IL	Charles MD (0.002), Tuscola MI (0.07), Washington RH (0.263), St. Clair MI (0.665	Washington RH (0.084), Jefferson LA (0.165), St. Claire MI (0.301), Erie OH (0.45)
Jefferson IN	Orange FL (0.038), Jasper IA (0.054), Washington RH (0.086), Carroll IL (0.131), Roane TN (0.15) Waldo ME (0.17), Lancaster PA (0.372)	Rogers OK (0.01), Tuscola MI (0.012), Dunn WI (0.045), Lancaster PA (0.053), Carroll IL (0.081), Waldo ME (0.115), Jasper IA (0.189), Washington RH (0.215), Roane TN (0.28)
Porter IN	Tuscola MI (0.154), Lancaster PA (0.22), Jefferson LA (0.626)	Bernardino CA (0.044), Tuscola MI (0.059), Washington RH (0.082), Erie OH (0.405), St. Claire MI (0.41)

Table B.3 Continued

Treated County	Weights (Emp. To Pop. Ratio)	Weights (Log of Per Capita Wages & Salaries)
Coffey KS	Jasper IA (0.153), Waldo ME (0.172), Austin TX (0.675)	Waldo ME (0.152), Rogers OK (0.161), Charles MD (0.229), Newton TX (0.457)
St. Charles LA	Roane TN (0.01), Washington RH (0.086), Newton TX (0.144), Dallas AL (0.268), Jefferson LA (0.492)	Stanislaus CA (0.007), Jefferson LA (0.053), Newton TX (0.099), Dallas AL (0.116), Charles MD (0.231), Roane TN (0.493)
Tishomingo MS	Newton TX (0.101), Roane TN (0.144), Dunn WI (0.186), Stanislaus CA (0.569)	Jasper IA (0.028), Tuscola MI (0.034), Roane TN (0.105), Davie NC (0.184), Chilton AL (0.266), Dunn WI (0.382)
Callaway MO	Roane TN (0.018), New Hanover NC (0.023), Newton TX (0.057), Charles MD (0.07), Gilliam OR (0.122), Austin TX (0.148), Jasper IA (0.193), Waldo ME (0.369)	New Hanover NC (0.011), Tuscola MI (0.023), Washington RH (0.023), Roane TN (0.048), Chilton AL (0.054), Austin TX (0.111), Sheboygan WI (0.122), Charles MD (0.132), Lancaster PA (0.15), Waldo ME (0.326)
Rockingham NH	Gilliam OR (0.01), Washington RH (0.011), Erie OH (0.032), Waldo ME (0.038), Orange FL (0.078), St. Chair MI (0.09), Jefferson LA (0.2), Sonoma CA (0.541)	Washington RH (0.001), Erie OH (0.11), Sheboygan WI (0.021), Sonoma CA (0.044), Bernardino CA (0.071), Waldo ME (0.196), Gilliam OR (0.224), Skagit WA (0.431)
Wake NC	Dallas AL (0.004), Orange FL (0.996)	Dallas AL (0.102), Los Angeles CA (0.253), Orange FL (0.646)
Lake OH	Tuscola MI (0.031), Skagit WA (0.087), Jefferson LA (0.882)	St. Claire MI (0.218), Bernardino CA (0.331), Skagit WA (0.451)
Montgomery PA	Lancaster PA (0.096), Los Angeles CA (0.184), Orange FL (0.72)	Los Angeles CA (0.423), Erie OH (0.577)
Cherokee SC	Lancaster PA (0.09), Davie NC (0.43), Aiken SC (0.48)	Roane TN (0.147), Dallas AL (0.148), New Hanover NC (0.168), Davie NC (0.537)

Table B.3 Continued

Treated County	Weights (Emp. To Pop. Ratio)	Weights (Log of Per Capita Wages & Salaries)
York SC	Charles MD (0.217), New Hanover NC (0.22), Dallas AL (0.245), Lancaster PA (0.319)	Charles MD (0.039), Elmore Al (0.075), New Hanover NC (0.181), Dallas AL (0.192), Lancaster PA (0.222), Aiken SC (0.29)
Hawkins TN	Chilton AL (0.079), Victoria TX (0.189), Davie NC (0.249) Newton TX (0.483)	Newton TX (0.105), Chilton AL (0.163), Dunn WI (0.732)
Trousdale TN	Roane TN (0.072), Gilliam OR (0.928)	Austin TX (0.451), Roane TN (0.549)
Matagorda TX	Riverside CA (0.045), Waldo ME (0.064) Austin TX (0.218), Dallas AL (0.275), Jefferson LA (0.397)	Gilliam OR (0.049), Rogers OK (0.071), Newton TX (0.083), Dallas AL (0.101), Washington RH (0.141), Victoria TX (0.155), Charles MD (0.17), Austin TX (0.228)
Somervell TX	Elmore AL (0.111), Austin TX (0.435), Rogers OK (0.454)	Elmore AL (0.032), Newton TX (0.067), Chilton AL (0.16), Rogers OK (0.352), Austin TX (0.389)
Grays Harbor WA	Los Angeles CA (0.036), Roane TN (0.119), Sheboygan WI (0.17), Skagit WA (0.193), St. Clair MI (0.236), Mendocino CA (0.246)	Elmore AL (0.039), Skagit WA (0.142), Sheboygan WI (0.819)

The table displays each treated county (that ultimately had an NPP construction) and the composition of its synthetic control. The numbers in parentheses are the weights assigned to the donor counties. Others' weights are zero.

Appendix Table B.4
The Impact of Nuclear Power Plants on Employment and Wages
(Synthetic Control Strategy)

	(1)	(2)	(3)
	Emp. to Pop. Ratio	Log of Per Capita Wages & Salaries	Per Capita Wages & Salaries
Construction	10.222**	0.364***	8.084**
	(4.838)	(0.119)	(3.951)
Commercial Operation	1.292	0.274***	4.490***
	(1.967)	(0.103)	(1.648)
N	2346	2346	2346
Mean of the outcome	50.097	17.585	17.585

The table presents the estimates obtained from Equation B1. See notes in Table 10 in the main text.

Appendix Figure B.1
The Employment-to-Population Ratio in Burke County, GA, and Synthetic Control

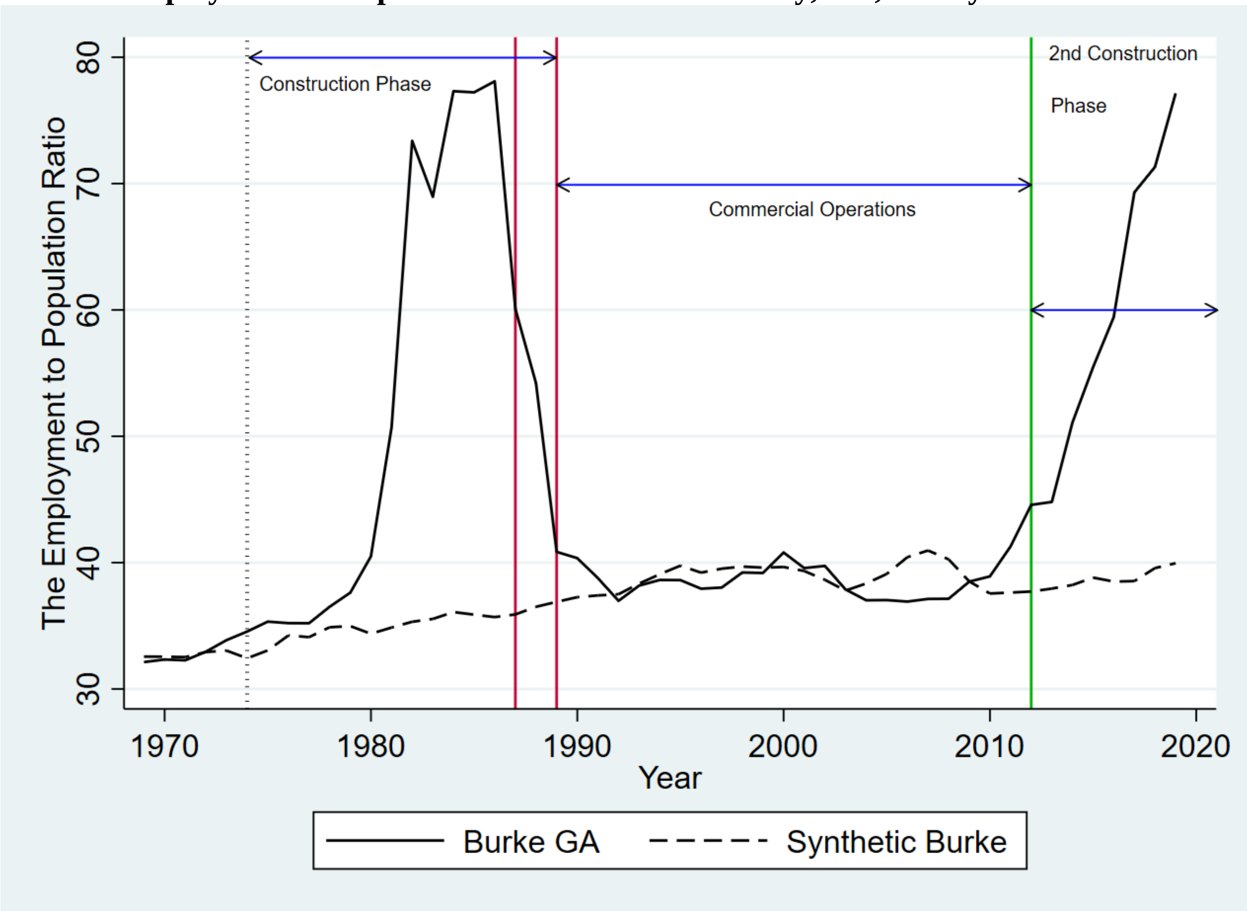
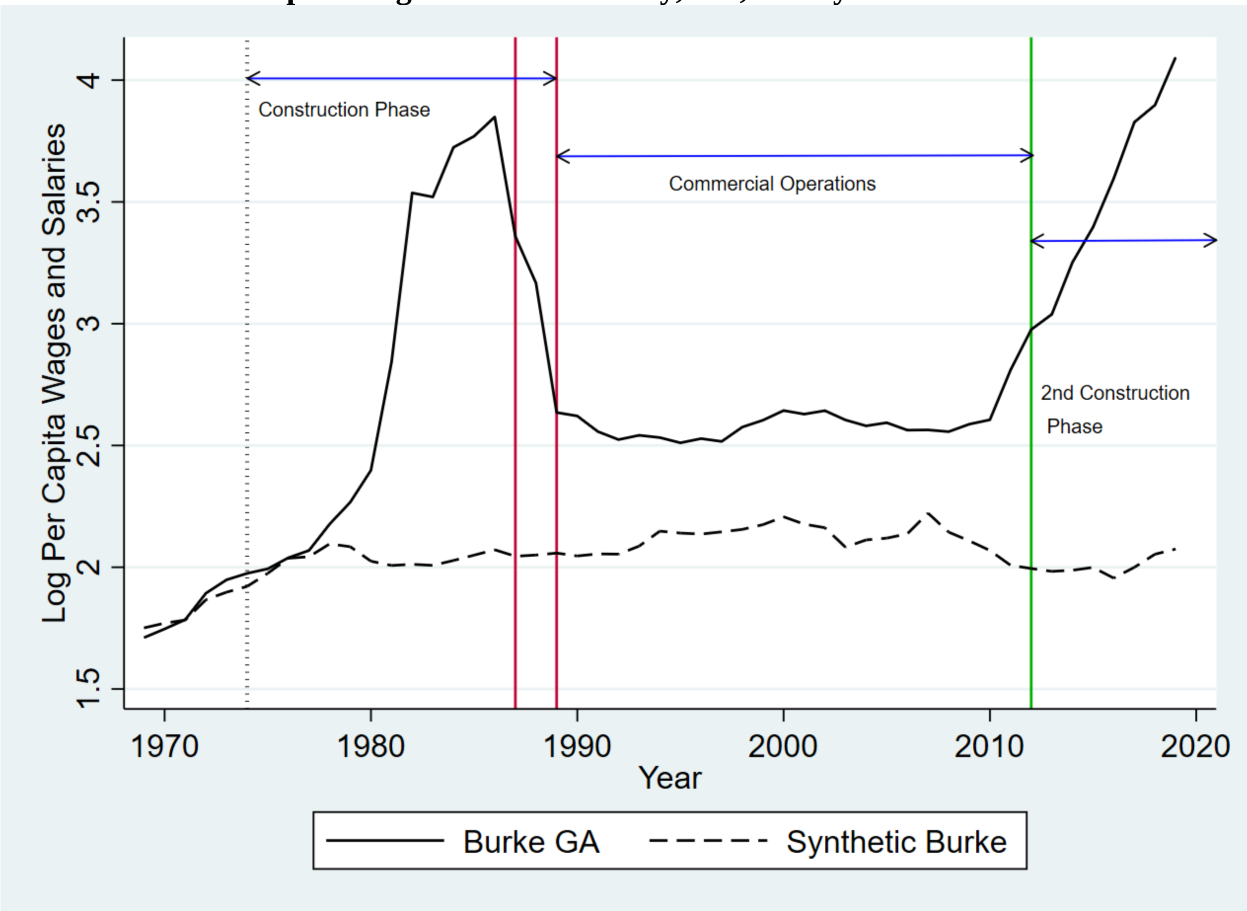


Illustration of the results of the synthetic control method for Burke County, GA. The outcome is the employment-to-population ratio. The construction of two reactors, both of which are operational now, started in 1974 (indicated by the vertical dotted line in the figure). One of them began commercial operations in 1987, and the other in 1989 (vertical solid lines in the corresponding years). The construction of another two units commenced in 2012 (vertical solid line in 2012).

Appendix Figure B.2
Per Capita Wages in Burke County, GA, and Synthetic Control



The outcome is per capita wages and salaries. See the notes in Appendix Figure B.1.

Appendix Figure B.3
Employment to Population Ratio in Tishomingo County, MS, and Synthetic Control

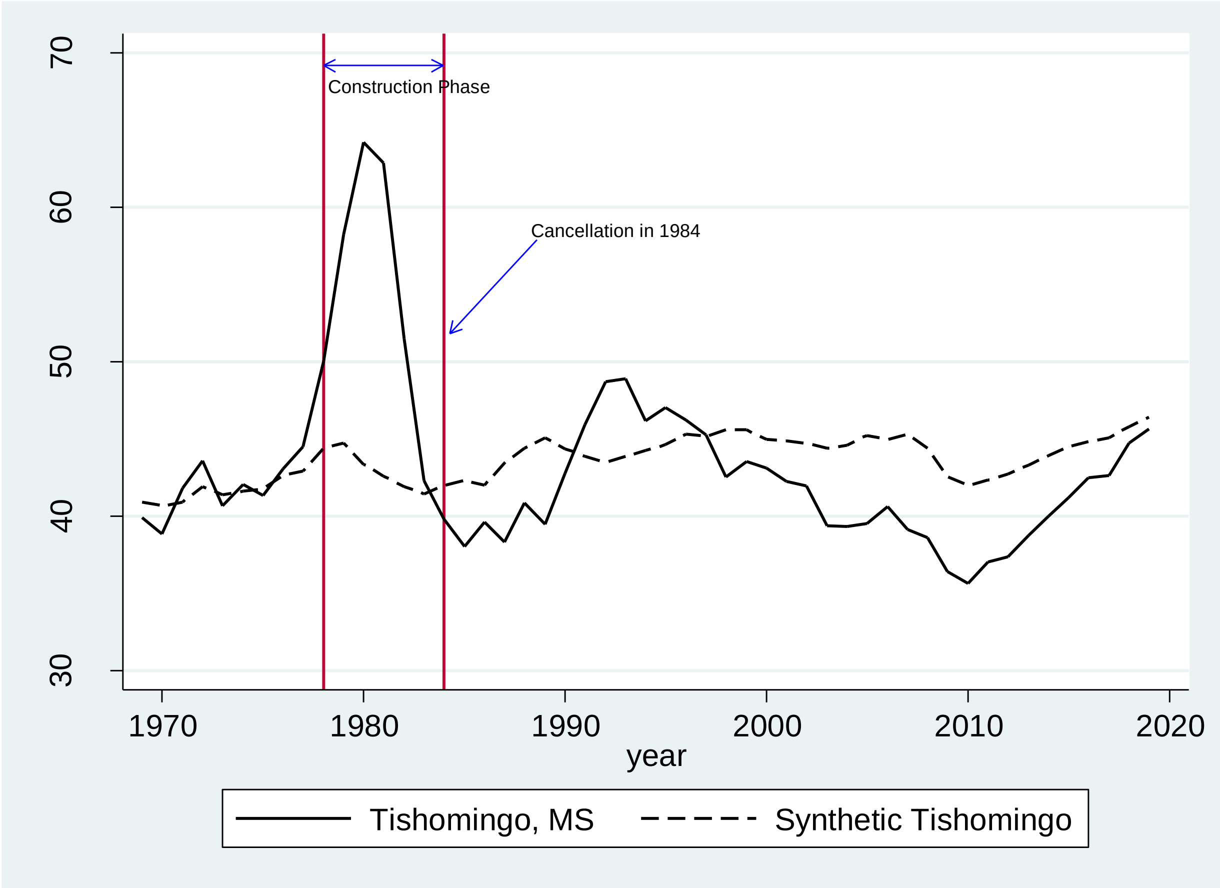


Illustration of the results of the synthetic control method for Tishomingo County, MS. The outcome is the employment-to-population ratio. The construction of a reactor started in 1978. The project was canceled before its completion in 1984. The vertical lines represent the construction period.

Appendix Figure B.4
Employment to Population Ratio in Cherokee County, SC, and Synthetic Control

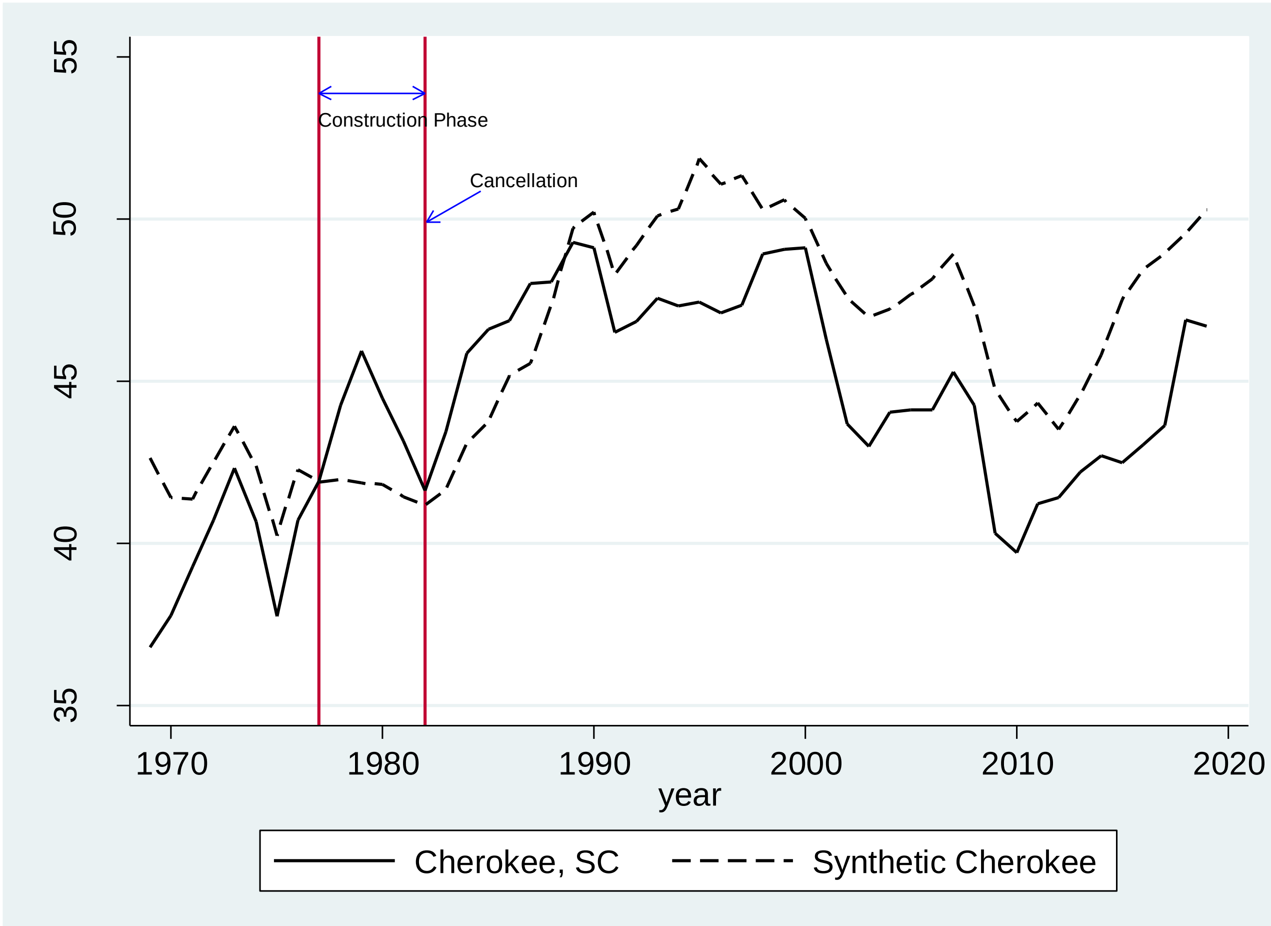
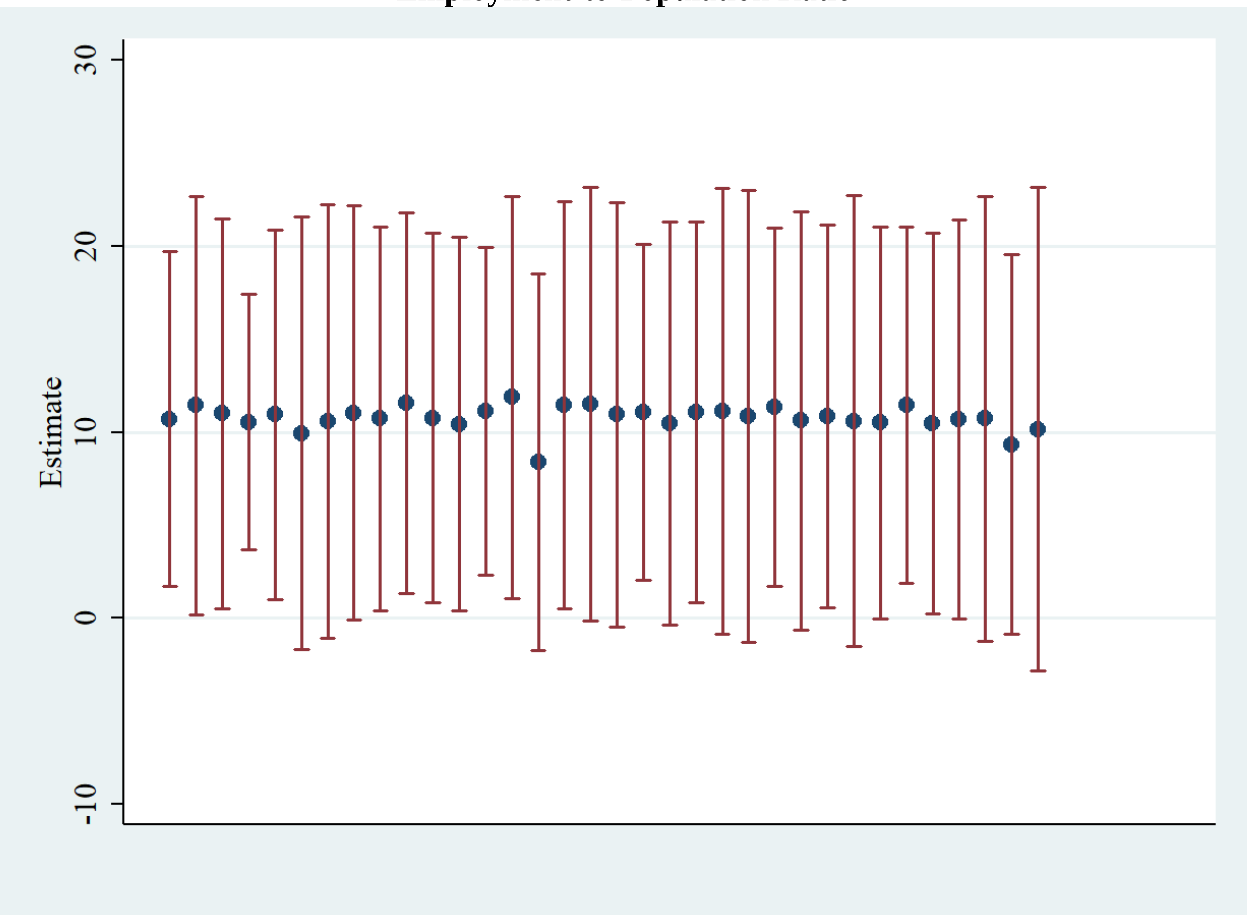


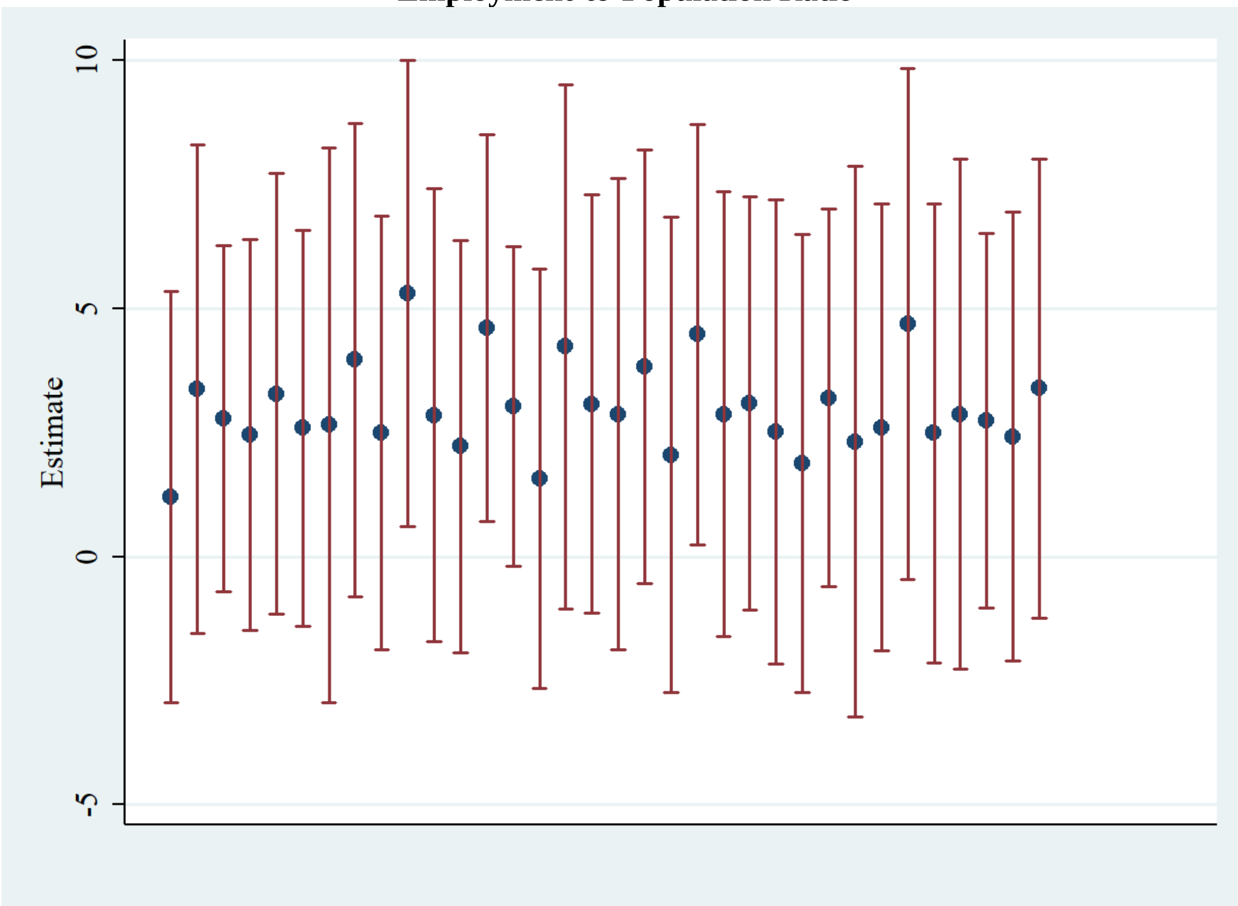
Illustration of the results of the synthetic control method for Cherokee County, SC. The outcome is the employment-to-population ratio. The construction of a reactor started in 1977. The project was canceled before its completion in 1983. The vertical lines represent the construction period.

Appendix Figure B.5
Estimates of Construction in the "Leave-One-Out" Analysis:
Employment-to-Population Ratio



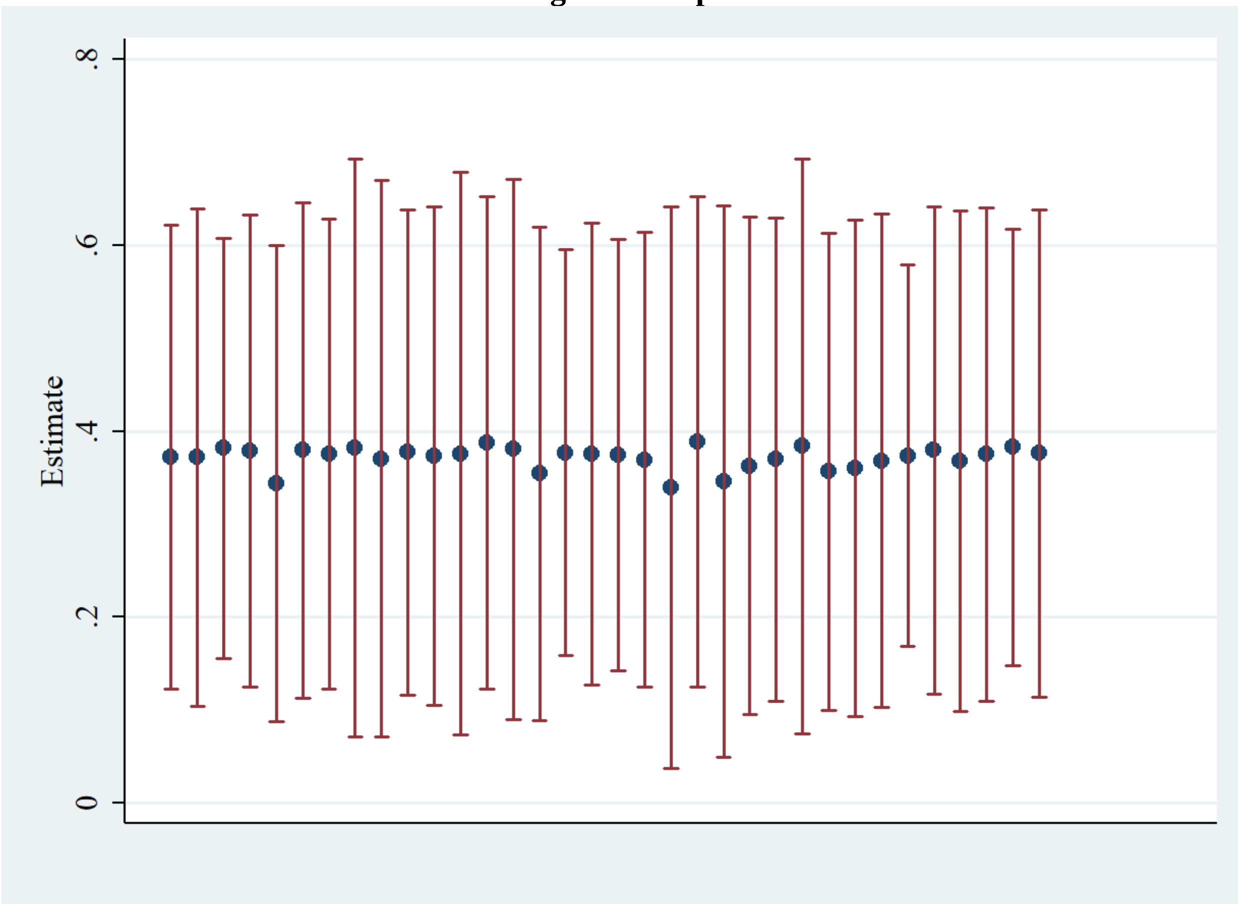
Point estimates and the 95% confidence bands of the Construction indicator in the employment-to-population ratio regressions obtained from a "Leave-one-out" analysis in which each of the 34 counties in the donor pool is excluded one by one when constructing the synthetic controls.

Figure B.6
Estimates of Commercial Operation in the "Leave-One-Out" Analysis: Employment-to-Population Ratio



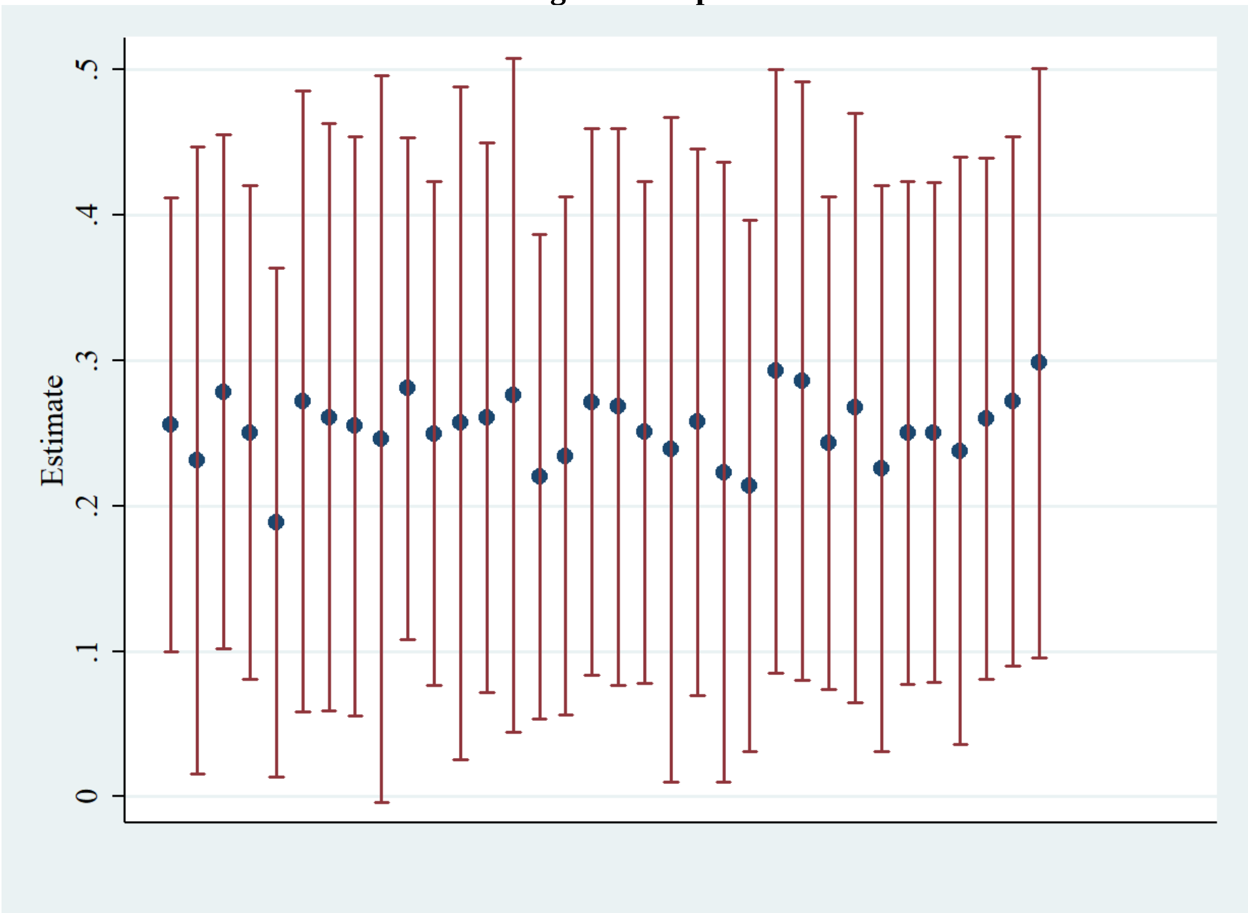
Point estimates and the 95% confidence bands of the Commercial Operation indicator in the employment-to-population ratio regressions obtained from the "Leave-one-out" analysis in which each of the 34 counties in the donor pool is excluded one by one when constructing the synthetic controls.

Figure B.7
Estimates of Construction in the "Leave-One-Out" Analysis:
Wages Per Capita



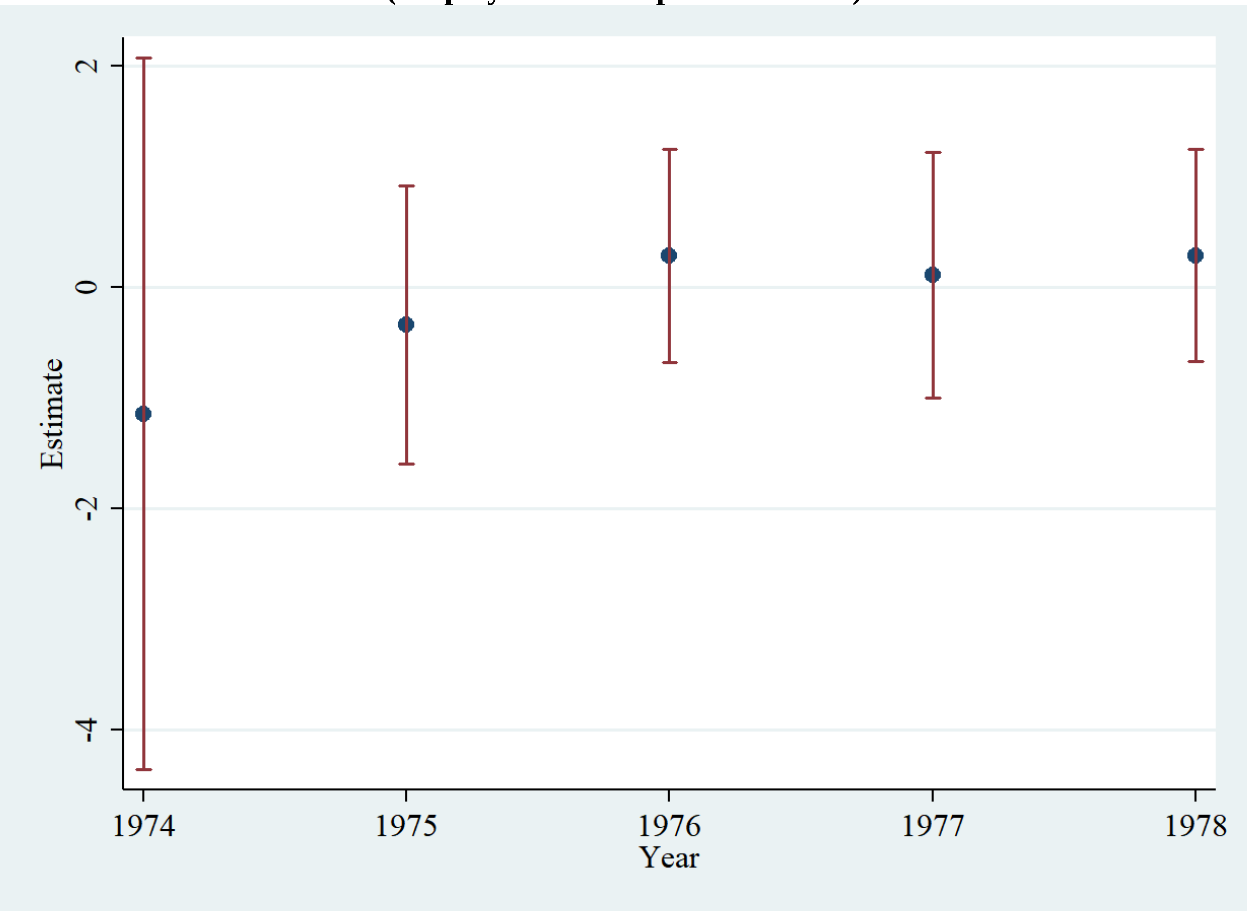
The outcome is per capita wages and salaries. See notes in Appendix Figure B.1.

Figure B.8
Estimates of Commercial Operation in the "Leave-One-Out" Analysis:
Wages Per Capita



The outcome is per capita wages and salaries. See notes in Appendix Figure B.2.

Figure B.9
Estimates of Construction in the Falsification Test in Appendix
(Employment-to-Population Ratio)

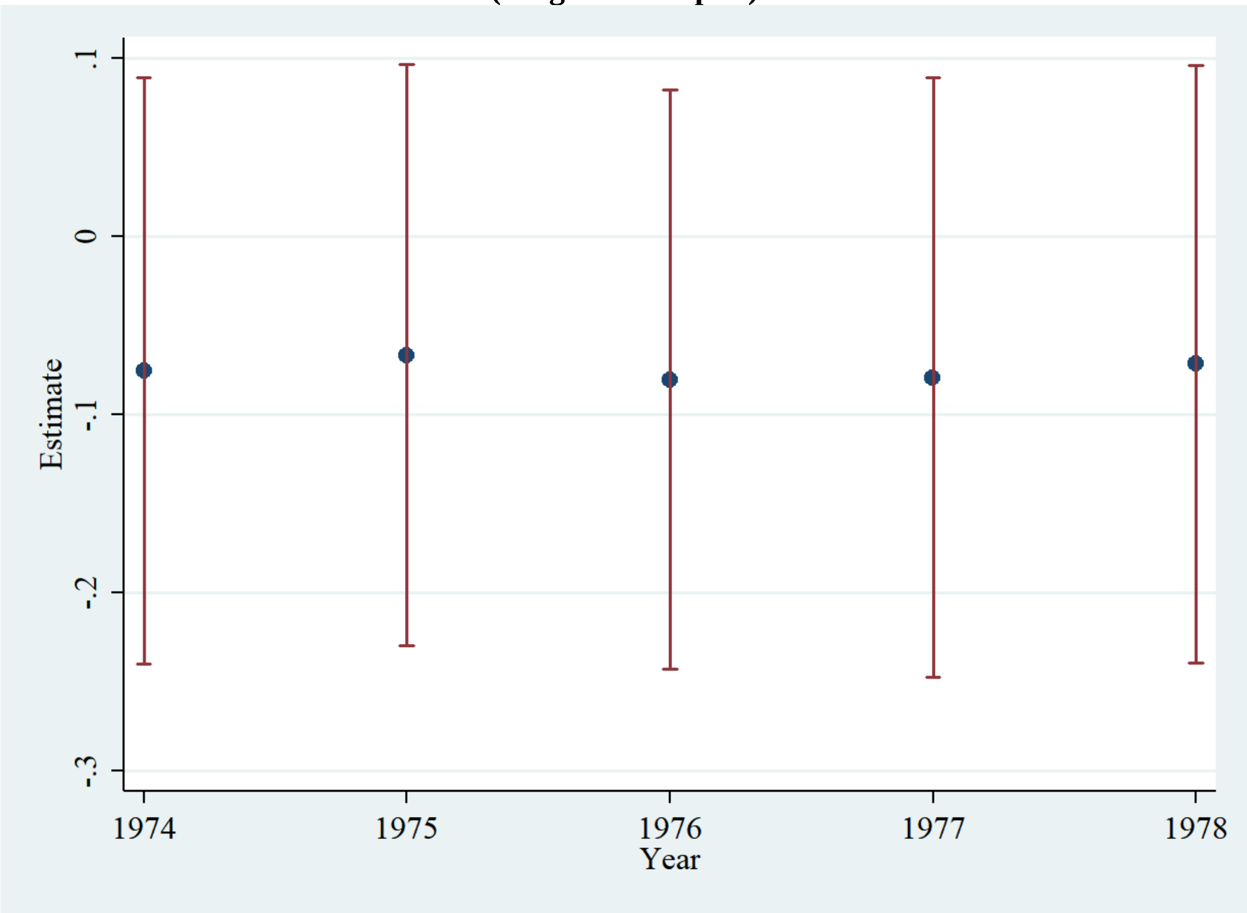


The point estimates and the 95% confidence intervals of *Construction* indicator in employment-to-population ratio regressions obtained from the falsification test described in the Appendix 3.

Figure B.10
Estimates of Commercial Operation in the Falsification Test in the Appendix
(Employment-to-Population Ratio)

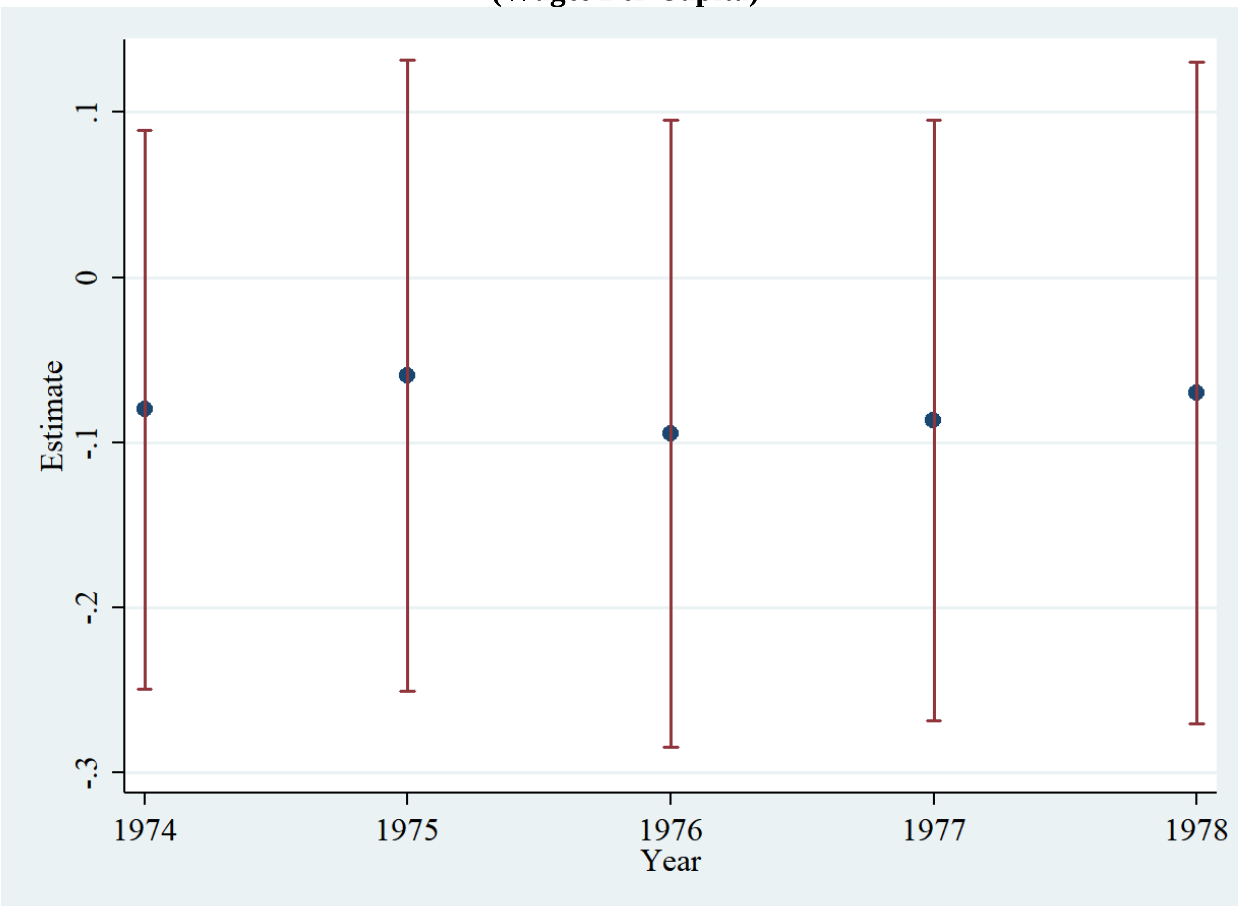
The point estimates and the 95% confidence intervals of *Commercial Operation* indicator in employment-to-population ratio regressions obtained from the falsification test described in the Appendix 3.

Figure B.11
Estimates of Construction in the Falsification Test in the Appendix
(Wages Per Capita)



The outcome is per capita wages and salaries. See notes in Appendix Figure B.5.

Figure B.12
Estimates of Commercial Operation in the Falsification Test in the Appendix
(Wages Per Capita)



The outcome is per capita wages and salaries. See notes in Appendix Figure B.6.