U.S. Census Data in Economic Research: Adoption, Diffusion, and Impact *

Abhishek Nagaraj
UC Berkeley-Haas

Fernando Stipanicic
University of Oslo

Matteo Tranchero
UC Berkeley-Haas

May 6, 2024

Abstract

Administrative data from government agencies is frequently used in modern economics research and yet, access to this data is restricted due to concerns about privacy and security. We propose a data-driven assessment of the use and impact of administrative data made accessible by the U.S. Census Bureau. Our findings show that while the data is of high quality and its use is growing, adoption is limited to established researchers from prestigious institutions. Our results and discussion inform the creation of policies that balance privacy protection with accessibility to confidential administrative data.

*We thank Randol Yao for excellent research assistance. Funding from the Alfred P. Sloan Foundation is gratefully acknowledged. Please direct any comments or feedback to nagaraj@berkeley.edu.

1 Introduction

In recent decades, economics has taken a sharp turn towards becoming a more empirical science (Angrist et al., 2020; Hamermesh, 2013). While the potential drivers for this shift are many, the increasing availability of high-quality, large-scale, longitudinal data such as administrative data from government agencies is a major factor (Currie et al., 2020; Einav and Levin, 2014a). Important empirical and theoretical breakthroughs have come from the availability of microdata that were originally collected for administrative needs. For example, plant-level records from the U.S. Census revealed numerous descriptive facts about exporters (Bernard and Jensen, 1999) that inspired the analysis of Melitz (2003), perhaps *the* workhorse model in international trade.[1] Similar examples can be found in a variety of other fields.[2]

Despite their scientific potential, administrative microdata pose serious privacy and security risks. In many jurisdictions, these confidentiality concerns have translated into a system that prioritizes strong guardrails (Cole et al., 2020). In other locations, such data are more freely available to academic researchers (Card et al., 2010). Notwithstanding the growing importance of administrative data for economics and enormous variation in access rules, the question of how one might balance privacy and scientific progress has ironically been data-free to date. A few papers have pointed to the growing diffusion of administrative data in general (Abraham et al., 2022; Chetty, 2012; Currie et al., 2020), but evidence on the impact and limitations of such data remains mostly anecdotal (Atrostic, 2007; CES, 2017; Davis and Holly, 2006).

Given the dearth of data-driven work, many questions remain unanswered. What is the extent and rate of adoption of administrative data, and how does its use influence research quality? What types of researchers are using the data and what questions are they exploring? Are current access restrictions shaping these patterns by excluding a subset of researchers and research topics? While balancing the trade-off between access and confidentiality is a function of social preferences (Abowd and Schmutte, 2019; Fobia et al., 2020), data-driven answers to these questions could help policy-makers strike a better balance between protecting privacy and enabling scientific innovation.

In this paper, we shed light on these questions by examining the use and impact of confidential, administrative data distributed by the U.S. Census Bureau ("Census data" for brevity) on

[1] At the moment of writing, Melitz (2003) appears as the 15th most cited article of all times in Ideas: `https://ideas.repec.org/top/top.item.nbcites.html`. This article has more than 19,000 citations in Google Scholar.

[2] Including in employment and wages (Haltiwanger et al., 2013) and (Abowd et al., 2018), racial disparities (Chetty et al., 2020), innovation (Jaravel et al., 2018), firm productivity (Hsieh and Klenow, 2009; Olley and Pakes, 1996), health care provision (Finkelstein et al., 2021) and mortgage markets (Beraja et al., 2019).

economics research.[3] Census data are perhaps the most pre-eminent sources of administrative data and have tight access restrictions, making them a prime setting for study. For example, data can only be accessed via physical enclaves, projects need to be approved in advance, and the disclosure of results is intensely scrutinized (Abowd and Schmutte, 2019; Cole et al., 2020; Foster et al., 2009; Nagaraj and Tranchero, 2023). This system has been criticized for being too restrictive and disadvantageous to U.S.-focused research, potentially shifting scientific attention toward other regions with more openly accessible data (Card et al., 2010). By bringing new evidence on the use of Census data, we hope to shed light on how the current access regime shapes their adoption and impact.

We document new facts on the use and impact of confidential Census data using a comprehensive database of publications in economics. The primary source of the data is *EconLit*, a proprietary database of economic scholarship curated by the American Economic Association. We consider in our analyses over 90,000 articles published in 158 peer-reviewed journals by researchers based in the United States. These data do not possess unique identifiers for researchers or institutions. We engage in a painstaking disambiguation effort, which allows us to measure outcomes for 17,820 researchers affiliated with 344 North American institutions between 1991 and 2019. Through various methods, including natural language processing and FOIA requests to the U.S. government, we are able to track the adoption of Census data at the paper-level. Finally, we augment these paper-level data with information on authors' characteristics, paper-to-paper citations and policy document-to-paper citations.

The analysis of our publication data shows that the incidence of articles using Census data has steadily increased over time, from 0.21% of all published papers in 1991 to 1.27% in 2019. Almost 6% of all papers in economics cite at least one paper using such confidential data. Scientific papers based on Census confidential data are highly impactful, being 50% more likely to be published in the "top five" economics journals, receiving 28% more citations from other papers and 80% more citations from policy documents. Census data are predominantly used in labor, applied microeconomics, industrial organization, and international economics research. However, papers using such data are more likely to include authors who have previously published in a top five

[3]Technically, the U.S. Census Bureau collects *statistical data* through surveys for the fulfillment of its mandate (e.g., the Economic Census). The protection of these data is regulated by the Title 13 of the U.S. Code. The Bureau also collects *administrative records* from other agencies that provide services to the public. Such data are often protected by different legal provisions (e.g., tax returns from the IRS fall under the Title 26 of the U.S. Code). For our purposes, this distinction is blurred since many datasets commonly used by researchers contain both type of data (e.g., the LEHD, see Abowd et al. 2009). Therefore, we adopt a broad definition of administrative data as any record not originally collected for research purposes which is instead collected as a by-product of economic or government activities (Cole et al., 2020; Groves, 2011). This includes all surveys and censuses that the Census Bureau carries out during its routine functioning.

journal or are affiliated with high-status U.S. universities. Taken together, our analysis suggests that while these data are precious in producing impactful research, their adoption remains relatively limited and restricted to established researchers from prestigious institutions.

We organize the article as follows. We first provide an overview of the confidential, administrative data distributed by the U.S. Census. We then describe the dataset we have built to assess the adoption and impact of Census data in the economics profession. We then overview key facts emerging from our publication data. We end with a discussion of the policy implications of our work.

2 Background

2.1 Administrative Data in Economics Research

Economic scholarship has changed from being primarily theory-driven into a more data-intensive discipline (Backhouse and Cherrier, 2017). The share of theoretical papers published in top economics journals has decreased from 50.7% in 1963 to 19.1% in 2011 (Hamermesh, 2013), while empirical work has surged both in incidence and impact (Angrist and Pischke, 2010; Angrist et al., 2020). The increased availability of government administrative data, intended as records arising as a by-product of some non-research activity, has played an important role (Cole et al., 2020; Groves, 2011). Around a fifth of the articles recently published in the five most prestigious economics journals employ data derived from administrative sources (Currie et al., 2020). This share increases up to 70% when focusing on studies about high-income countries. (Chetty, 2012)

Administrative data offer tremendous advantages for economic research (Cole et al., 2020). These data are usually granular, highly disaggregated, and display a longitudinal structure that naturally allows tracking of the same individuals, or companies, over time and before and after certain interventions. Unlike surveys, administrative data do not suffer from non-response issues and can include very large samples that allow precise estimation for subgroup heterogeneity analyses (Heckman, 2001). Administrative data can not only provide better answers to old questions, but they can open up new fields of inquiry based on new questions as well (Einav and Levin, 2014b). Creative combinations of administrative records from multiple sources have been instrumental during the COVID-19 pandemic in following real-time developments and targeted assistance programs (Bartlett and Morse, 2021; Vavra, 2021).

The unique value of administrative data for policy-relevant research derives from the level of

detail that it affords. However, this very characteristic might put the privacy of respondents at risk. Information on an individual's employment, earnings, tax identification numbers, etc., must be kept private to prevent identity fraud and harmful targeting. Leakage of granular sales and employment information at the business level could also have negative competitive implications for firms. Even beyond these economic harms, moral and legal frameworks in most contexts, and especially in the U.S., support privacy as a fundamental right that must be preserved. For example, Title 13 of the U.S. Code makes it illegal for the U.S. Census to disclose or publish any private information that identifies an individual or business, including names, addresses, Social Security Numbers, and telephone numbers.[4]

Data providers thus face a trade-off between granting access to their administrative records and their duty to protect the confidentiality of the information entrusted to them (Foster et al., 2009). How might a data provider handle this problem with seemingly countless moving parts? The literature points to the "five safes" framework (Desai et al., 2016) that provides a useful way to organize the different levers through which data access can be regulated (projects, people, data, settings, and outputs). The five safes framework prioritizes controls on a) which projects are approved, b) which individuals are provided access, c) how data is modified to preserve anonymity, d) locations or devices where data access is provided and monitored, and e) how results are disclosed. This framework has been used by the ICPSR (provider of IPUMS data), the OECD, statistics offices in the U.K., Australia, New Zealand, Luxembourg, Mexico, the NORC data enclave, and the U.S. National Research Council to develop procedures regulating administrative data access.

The trade-off between limiting access and value is further complicated because access to confidential data faces problems similar to public goods provision. The benefits from realized projects accrue to the whole society while costs are borne by the data provider (Ritchie and Welpton, 2011). Therefore, it is reasonable to expect that from a welfare perspective, access to confidential data could be under-provided. Speaking specifically about the U.S. Census Bureau (our focus), informed commentators have argued that current access policies prioritize security and privacy risks without fully considering its impacts on value creation (Lane, 2021).

2.2 Balancing Access and Security at the U.S. Census Bureau

Over the years, the U.S. Census Bureau has experimented with several solutions to disseminate its data while protecting individual privacy. Some of them include releasing anonymized Public Use Microdata Samples (PUMS) and the development of synthetic data (Abowd and Lane, 2004;

[4] https://www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html

Kinney et al., 2011). Yet, these data are only poor replacements for the possibility of working with the universe of respondent-level micro-data. This includes data collected by the Census itself as well as by partner agencies such as the Agency for Healthcare Research and Quality, the Bureau of Economic Analysis, the Bureau of Justice Statistics, the Bureau of Labor Statistics, the National Center for Health Statistics, and the National Center for Science and Engineering Statistics.

In 1982, the Census Bureau established the Center for Economic Studies (CES) to enable non-employee data access to administrative microdata (Atrostic, 2007; McGuckin et al., 1993). The objective of the CES was to develop longitudinal databases and to host qualified academic researchers that could analyze confidential data directly onsite (Foster et al., 2009). However, before researchers can access data, the Bureau has imposed limits on using each of the five types of controls highlighted by the five safes framework. First, the Bureau enforces tight restrictions on the types of projects and people who can work with Census data. Researchers must write a detailed proposal showing how the proposed research benefits the Census Bureau, justifies the feasibility of the project and requirement for non-public data, and proves that the project does not pose a risk of unauthorized disclosure. Researchers must also apply for "special sworn status" (SSS) with the U.S. Census Bureau, which involves passing a background check, among other legal requirements. Further, the proposal must specify which variables and datasets the researchers will use. Access is provided to only the data requested and approved.

Regarding how the data is accessed, the U.S. Census Bureau has adopted the approach of allowing only in-situ analysis in a data enclave to minimize the risk of privacy breaches. Since traveling to the Center for Economic Studies in Suitland, MD is not feasible for most researchers, the Bureau has opened over 30 multiple secure facilities across the country as part of a program known as the Federal Statistical Research Data Centers (FSRDCs). FSRDCs are operated by Census staff in partnership with local universities or research institutions. Each branch meets the same physical security standards as the Center for Economic Studies. Researchers are monitored closely when accessing data in these enclaves, and no data or outputs can leave the secure facilities without a detailed disclosure review from Census officials. Through these multiple steps, the Census ensures that the privacy and security risks of sharing sensitive data are minimized while simultaneously allowing access to academic research.

Despite the strict limits on access, anecdotal evidence suggests that the FSRDC program helped foster data diffusion and provided benefits for the Census Bureau (CES, 2017; Davis and Holly, 2006). Several policy-relevant findings were enabled by the granularity of confidential

data uniquely accessible in Research Data Centers (Card et al., 2010; Einav and Levin, 2014a). Nagaraj and Tranchero (2023) show that the expansion of the FSRDC network alleviated part of access constraints, leading to a large increase in the use of administrative data by researchers in nearby institutions and spillover effects in terms of increased productivity for empirical researchers.

Nevertheless, their result does not speak to how the current regulatory environment supports or stifles academic progress. Additional evidence is needed to understand how researchers respond to access restrictions and whether procedures can be streamlined and made more researcher-friendly without significantly compromising privacy or security. In this project, we provide a data-driven exploration of the potential benefits of access to Census data. These parameters could help inform the design of alternative policies to address the public goods problem inherent in the provision of administrative data.

3 Data

We assemble a database of economic scholarship using article-level data from *EconLit*, a proprietary database of economic scholarship curated by the American Economic Association. Compared to other popular databases of scientific publications, EconLit has a wider coverage of economics journals and includes JEL codes that allow classifying papers into economics fields (Angrist et al., 2020; Nagaraj and Tranchero, 2023). We consider in our analyses over 90,000 articles published in 158 peer-reviewed journals by researchers based in the United States. These data do not possess unique identifiers for researchers or institutions. We engage in a painstaking disambiguation effort using the disambiguation algorithm developed by Önder and Schweitzer (2017). This procedure allows us to measure outcomes for 17,820 researchers affiliated with 344 North American institutions between 1991 and 2019. We augment these paper-level data with information about authors' characteristics and paper-to-paper citations from the Web of Science database. We obtain policy document-to-paper citations from Overton. Information on the ranking of economics departments is taken from Kalaitzidakis et al. (2003).

We collected new information to measure the adoption of administrative data available only at the Census Bureau or through the FSRDC network. Since no official bibliographic record is available, we carefully sifted published records with several complementary strategies. We exploited the fact that projects using U.S. Census confidential microdata are expected to indicate it clearly in the acknowledgement of the published version of the paper. Using natural language processing, we searched for the most commonly used acknowledgement formulas in databases

such as Web of Science, Scopus, JSTOR, Google Scholar, IDEAS RePEc, and the NBER website. However, we noticed that several publications omit the disclaimer. As a solution, we gathered information on what projects had been approved for in-situ analysis by the U.S. Census Bureau through FOIA requests.[5] We then manually matched projects with their resulting output in EconLit.

Descriptive statistics of our article-level sample are presented in Table 1. In total, we have 589 papers that used confidential Census data written by 525 U.S.-based researchers in the period 1991-2019. The table suggests that Census data are within the purview of a restricted set of institutions. Papers using confidential data from the Census Bureau are more impactful on average and more likely to appear in high-profile economics outlets.

Table 1: Descriptive statistics data Economics academic articles

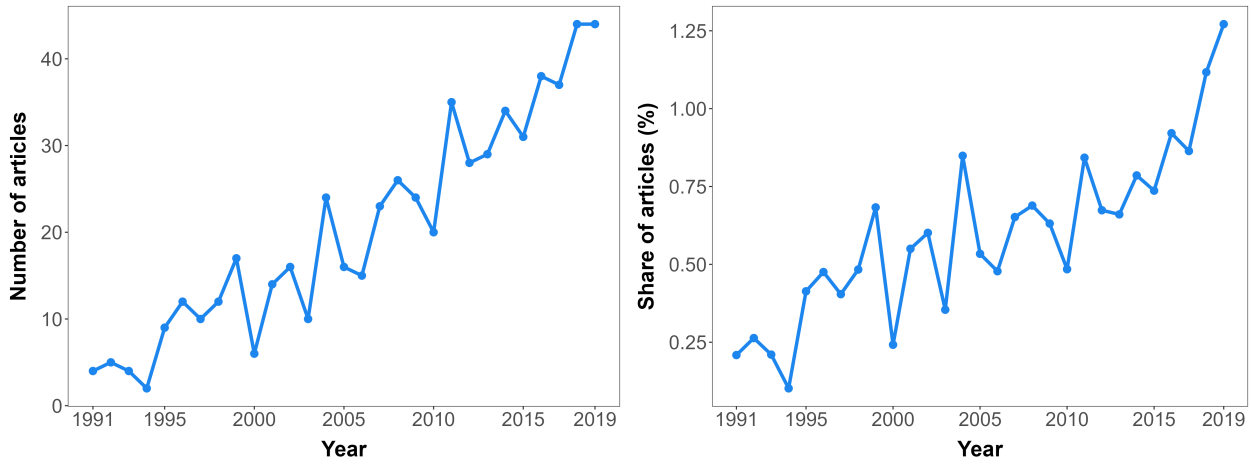
Census data use	N. Articles	N. Authors	N. Institutions	Avg. yearly citations	% Articles in Top 5 journal	Avg. yearly citations Top 5 journal
No	90,981	17,776	342	4.00	7.65	9.99
Yes	589	525	122	7.26	14.60	17.25
Total	91,570	17,820	342	4.02	7.69	10.08

4 Descriptive Statistics from Census dataset

We present preliminary findings emerging from our analysis of research based on Census confidential data. To the best of our knowledge, we are the first to provide an empirical assessment of use and impact of these data in the economic profession. Our results can be summarized in the following five facts. For this proposal, we discuss these facts as an example of the kind of insights our data can surface, and hope to dig deeper into additional aspects in our final paper.

Fact 1: The scientific impact of Census data is increasing over time. The number of peer-reviewed publications employing confidential, administrative data from the Census Bureau has steadily increased during our sample period. In relative terms, these papers went from 0.21% to 1.27% of all published scholarship in the last three decades. These papers are often published in the most prestigious economics outlets. The share of papers using Census data that each year appear in a top 5 journal is consistently around 15% throughout our sample period.

[5]The information we used were given to us as a response to FOIA request No. DOC-CEN-2020-001640. These records have also been published online for the benefit of everyone interested in tracking the use of Census Bureau's administrative data. Census officials are now regularly updating the list available at: `https://www.census.gov/about/adrm/fsrdc/about/ongoing-projects.html`.



(a) Articles using Census data

(b) Share of articles using Census data

Figure 1: Use of confidential U.S. Census Bureau data in economics research

Fact 2.a: Articles using confidential Census data are more impactful. Despite constituting a fraction of all published economic scholarship, papers using confidential Census data are disproportionally impactful. These papers represent 0.64% of all papers published in 1991-2019, but they received 1.16% of all citations and constitute 1.22% of the papers appearing in a top 5 journal. Results of regression analysis presented in Table 2 confirm these patterns. Even when restricting the comparison to papers of the same author, papers written in an FSRDC are 50% more likely to be published in top 5 journals. The effect is not driven by the prestige of the journal: these papers receive on average 28% more citations than other papers of the same author appearing in the same journal.[6] As a further measure of scientific quality, we collected data on papers that won best paper prizes (e.g., the *AEJ Best Paper Award* or the *Frisch Medal*). We found that the incidence of paper prizes is around 2.5 times higher in the sample of papers using administrative Census data.

Fact 2.b: Articles using confidential Census data receive more policy citations. While 44% of articles in our sample have at least one policy citation, the share is 75% for papers using confidential Census data. Restricting to articles published in top 5 journals, 65% of all articles receive policy citations, while the share is 89% for articles that use confidential Census data. The regression analysis presented in Table 2 shows that the difference persists after controlling for field-year, author and journal fixed effects: articles that use confidential US Census data are 24% more likely to receive policy citations and receive 80% more policy citations.

[6]Columns (4) through (7) of Table 2 use the inverse hyperbolic sine (*asinh*) of citations as the dependent variable. The *asinh* function closely follows the natural logarithm function. Given that the right hand side variable is a dummy, coefficients on columns (4) through (7) should be interpreted as being close to the semi-elasticity of citations to using Census data. The percentage increase in citations due to a paper using Census data is $\exp(\beta) - 1$, which is approximately 28% using the coefficient of column (7). Estimations through Poisson Pseudo Maximum Likelihood give a similar result (25% increase using the fixed effects of column (4) in Table 2).

Table 2: Quality of Economics research using Census data

Dependent Variables:	Top 5 publication (0/1)			Citations received			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Uses Census data	0.077***	0.078***	0.056***	0.641***	0.658***	0.398***	0.250***
	(0.015)	(0.013)	(0.014)	(0.052)	(0.052)	(0.043)	(0.038)
Fixed-effects							
Year-Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Author			Yes			Yes	Yes
Journal							Yes
Observation level	Art.	Art.-Author	Art.-Author	Art.	Art.-Author	Art.-Author	Art.-Author
Cluster s.e.	Year-Field	Author	Author	Year-Field	Author	Author	Author
Observations	91,563	132,259	132,259	91,563	132,259	132,259	132,259
Mean LHS	0.077	0.090	0.090	50 citations	55 citations	55 citations	55 citations

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Results of estimating by OLS: $y_{iajft} = \beta \times \text{Uses Census data}_i + FE_{...} + \epsilon_{iajft}$ for article i, author a, journal j, field f and year of publication t. In columns (1) to (3) the outcome variable y_{iajft} is whether the article was published in a top 5 journal (i.e. American Economic Review, Econometrica, Journal of Political Economy, Review of Economic Studies, The Quarterly Journal of Economics). In columns (4) to (7) the outcome variable y_{iajft} is the inverse hyperbolic sine (asinh) of the amount of citations that the article i has received. Columns (1) and (4) are estimated at the observation level of a article ($ijft$), while columns (2), (3), and (5) to (7) are at the article-author level ($iajft$). Each article is classified into only one 16 economics fields. In order to identify the author fixed effect it is required that the author publishes more than one article. We keep the sample of articles constant across regressions at the article-author level by dropping 3,377 single-article authors (18.9% of authors accounting for 3.6% of articles). The amount of observations reported is the effective sample size.

Table 3: Policy impact of Economics research using Census data

Dependent Variables:	Policy citation received 0/1			Policy citations received			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Uses Census data	0.254***	0.240***	0.116***	1.18***	1.20***	0.701***	0.587***
	(0.021)	(0.013)	(0.017)	(0.079)	(0.065)	(0.062)	(0.055)
Fixed-effects							
Year-Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Author			Yes			Yes	Yes
Journal							Yes
Observation level	Art.	Art.-Author	Art.-Author	Art.	Art.-Author	Art.-Author	Art.-Author
Standard-Errors	Year-Field	Author		Year-Field		Author	
Observations	91,563	132,259	132,259	91,563	132,259	132,259	132,259
Mean LHS	0.439	0.469	0.469	6 citations	7 citations	7 citations	7 citations

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Results of estimating by OLS: $y_{iajft} = \beta \times \text{Uses Census data}_i + FE_{...} + \epsilon_{iajft}$ for paper i, author a, journal j, field f and year of publication t. In columns (1) to (3) the outcome variable y_{iajft} is whether the paper received at least one citation from a policy document. In columns (4) to (7) the outcome variable y_{iajft} is the inverse hyperbolic sine of the amount of policy citations that the paper i has received. Columns (1) and (4) are estimated at the observation level of a paper ($ijft$), while columns (2), (3), and (5) to (7) are at the paper-author level ($iajft$). Each paper is classified into only one of 16 Economics fields. In order to identify the author fixed effect it is required that the author publishes more than one paper. We keep the sample of papers constant across regressions at the paper-author level by dropping 3,377 single-paper authors (18.9% of authors accounting for 3.6% of papers). The amount of observations reported is the effective sample size.

Fact 3: Census data are mostly used in labor economics and applied microeconomics. 30% of articles using Census data can be classified as labor economics and 28% as applied microeconomics. Notably, the incidence of these two fields in the set of Census papers is more than double of the incidence of these fields in the rest of the economic scholarship. Confirming our priors, articles based on confidential microdata are substantially less focused on macroeconomics or econometric methods (5% and 2% respectively) relative to non-Census articles (11% and 9% respectively). Within the body of papers using confidential Census data, we noticed a decrease of papers in industrial organization from 18.5% in the first decade to 5.6% in last decade of our sample. Meanwhile, there has been a constant growth of reduced-form applied micro papers, consistent with the broader trends documented by Hamermesh (2013) and Angrist et al. (2020).

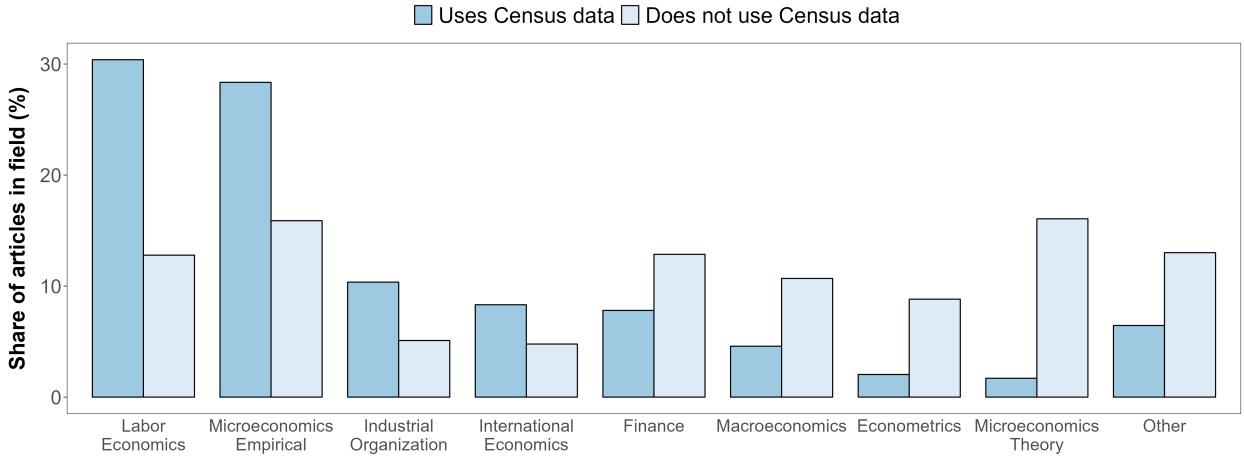


Figure 2: Share of articles using confidential U.S. Census Bureau data by research field

Fact 4: Articles using administrative microdata have larger and more age-diverse co-authorship teams. Census papers have on average 2.1 authors, while all other economics papers have on average 1.5 coauthors. Compared to the rest of the sample, the within-paper minimum seniority of coauthors is lower for Census papers, while the maximum seniority is higher.[7] The result is that papers using administrative Census data are more likely to combine experienced and inexperienced researchers. Figure 3 plots that within-paper difference in seniority between coauthors, that is a synthetic metric capturing the age composition of authorship teams. We notice that the distribution is shifted to the right for Census papers, suggesting that they tend to include senior authors working with junior colleagues. This fits well with oral accounts of the division of labor in those projects, whereby it is usually the junior team member that physically access the data enclave to carry out analyses.[8]

[7]Age is computed as the year of publication minus the first year in which the author published an article. Within-paper minimum seniority of coauthors is 6.8 years for Census papers vs. 9.7 years for non-Census papers; the average maximum seniority is 18.1 years for Census papers and 15.8 years for non-Census papers.

[8]Authorship teams using confidential data are also more likely to include female coauthors, but this effect is explained

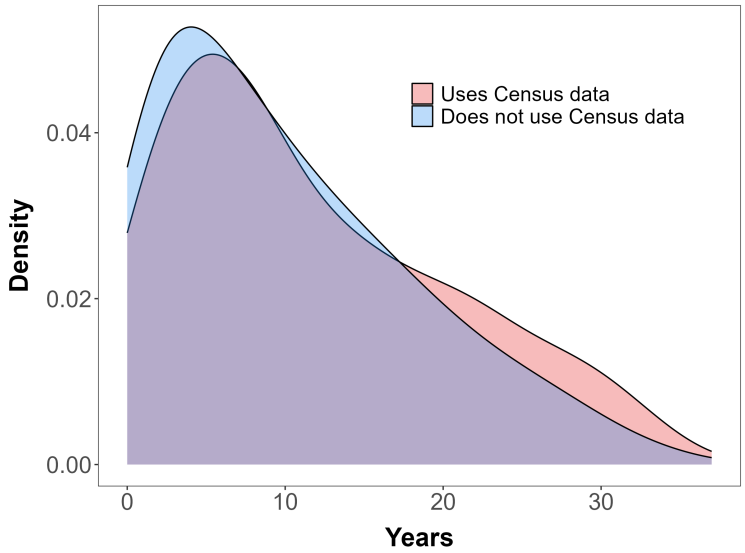


Figure 3: Maximum age difference between coauthors of articles using U.S. Census Bureau data

Fact 5: Articles using Census data are more likely to include established researchers affiliated to high-status institutions. Authors of research based on confidential Census data are more established in the discipline: 59% of Census papers have authors whom previously published on a top 5 journal, while this is true only for 38% of the papers in our sample. Regressions show that the difference remains after controlling for field, year, number of co-authors, average seniority, and maximum seniority difference between co-authors. Similarly, the share of papers that have at least one author in a top 10 institution is 42% for Census papers while it is 35% for non-Census papers.

5 Conclusion

Our analysis shows that the use of confidential US Census data is incredibly valuable for Economics research: articles using these data are 50% more likely to be published in top 5 journals, receive 28% more citations from other academic articles and 80% more citations from policy documents. However, our analysis also shows that the use of these data remains limited, and those who use it are more likely to be established researchers from prestigious institutions.

The provision of access to confidential administrative data shares similarities with the provision of a public good. While the benefit of using such data for research accrue to the society as a whole, the costs of privacy and security risks are borne by the data provider (Ritchie and Welpton, 2011). In the case of U.S. Census Bureau, the access to confidential administrative data is tightly circumscribed by Title 13 and 26 of the U.S. Code. Within these limits, the U.S. Census Bureau has

by the distribution across fields of Census papers.

provided access to researchers by setting up physical enclaves with strict security controls. Nagaraj and Tranchero (2023) show that the expansion and opening of new physical enclaves led to an increase in the use of administrative data by researchers in nearby institutions. In spite of this, informed commentators have argued that current access policies prioritize security and privacy risks without fully considering its impacts on value creation (Lane, 2021).

Recently the U.S. Census Bureau has taken steps towards providing remote access to confidential administrative data. This process is in very early stages and an evaluation of such program would be very valuable to know whether and how this policy change leads to a democratization of access to data.

Overall, through this article article we hope to have provided an overview of the role of administrative data in economics, a sense of its growing impact, a description of the challenges associated in making these data more accessible and the impacts of access restrictions on researchers' research trajectories. We hope this work will provide a solid foundation for policy discussions about whether and how administrative data should be disseminated for academic scholarship.

References

ABOWD, J. M. AND J. LANE (2004): "New approaches to confidentiality protection: Synthetic data, remote access and research data centers," in *International Workshop on Privacy in Statistical Databases*, Springer, 282–289.

ABOWD, J. M., K. L. MCKINNEY, AND N. L. ZHAO (2018): "Earnings inequality and mobility trends in the United States: Nationally representative estimates from longitudinally linked employer-employee data," *Journal of Labor Economics*, 36, S183–S300.

ABOWD, J. M. AND I. M. SCHMUTTE (2019): "An economic analysis of privacy protection and statistical accuracy as social choices," *American Economic Review*, 109, 171–202.

ABOWD, J. M., B. E. STEPHENS, L. VILHUBER, F. ANDERSSON, K. L. MCKINNEY, M. ROEMER, AND S. WOODCOCK (2009): "The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators," in *Producer dynamics: New evidence from micro data*, University of Chicago Press, 149–230.

ABRAHAM, K. G., R. S. JARMIN, B. MOYER, AND M. D. SHAPIRO (2022): *Big Data for 21st Century Economic Statistics*, NBER Book Series Studies in Income/Wealth.

ANGRIST, J., P. AZOULAY, G. ELLISON, R. HILL, AND S. F. LU (2020): "Inside job or deep impact? Extramural citations and the influence of economic scholarship," *Journal of Economic Literature*, 58, 3–52.

ANGRIST, J. D. AND J.-S. PISCHKE (2010): "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of Economic Perspectives*, 24, 3–30.

ATROSTIC, B. (2007): "The Center for Economic Studies 1982-2007: A brief history," *CES working paper*.

BACKHOUSE, R. E. AND B. CHERRIER (2017): "The age of the applied economist: the transformation of economics since the 1970s," *History of Political Economy*, 49, 1–33.

BARTLETT, R. P. AND A. MORSE (2021): "Small-Business survival capabilities and fiscal programs: evidence from Oakland," *Journal of Financial and Quantitative Analysis*, 56, 2500–2544.

BERAJA, M., A. FUSTER, E. HURST, AND J. VAVRA (2019): "Regional heterogeneity and the refinancing channel of monetary policy," *The Quarterly Journal of Economics*, 134, 109–183.

BERNARD, A. B. AND J. B. JENSEN (1999): "Exceptional exporter performance: cause, effect, or both?" *Journal of International Economics*, 47, 1–25.

CARD, D., R. CHETTY, M. S. FELDSTEIN, AND E. SAEZ (2010): "Expanding access to administrative data for research in the United States," *American economic association, ten years and beyond: Economists answer NSF's call for long-term research agendas*.

CES (2017): "Center for Economic Studies and Research Data Centers Research Report: 2016," *Available on the US Census Bureau's website*.

CHETTY, R. (2012): "Time Trends in the Use of Administrative Data for Empirical Research," *NBER Summer Institute presentation. Available at the author's website*.

CHETTY, R., N. HENDREN, M. R. JONES, AND S. R. PORTER (2020): "Race and economic opportunity in the United States: An intergenerational perspective," *The Quarterly Journal of Economics*, 135, 711–783.

COLE, S., I. DHALIWAL, A. SAUTMANN, L. VILHUBER, ET AL. (2020): "Handbook on Using Administrative Data for Research and Evidence-Based Policy," *https://admindatahandbook. mit. edu/book/v1. 0-rc5/index. html*.

CURRIE, J., H. KLEVEN, AND E. ZWIERS (2020): "Technology and big data are changing economics: Mining text to track methods," in *AEA Papers and Proceedings*, vol. 110, 42–48.

DAVIS, J. C. AND B. P. HOLLY (2006): "Regional analysis using Census Bureau microdata at the Center for Economic Studies," *International Regional Science Review*, 29, 278–296.

DESAI, T., F. RITCHIE, AND R. WELPTON (2016): "Five Safes: Designing data access for research," *Economics Working Paper Series 1601, University of the West of England*.

EINAV, L. AND J. LEVIN (2014a): "The data revolution and economic analysis," *Innovation Policy and the Economy*, 14, 1–24.

——— (2014b): "Economics in the age of big data," *Science*, 346, 1243089.

FINKELSTEIN, A., M. GENTZKOW, AND H. WILLIAMS (2021): "Place-based drivers of mortality: Evidence from migration," *American Economic Review*, 111, 2697–2735.

FOBIA, A. C., J. H. CHILDS, AND C. EGGLESTON (2020): "Attitudes toward Data Linkage: Privacy, Ethics, and the Potential for Harm," *Big Data Meets Survey Science: A Collection of Innovative Methods*, 683–712.

FOSTER, L., R. JARMIN, AND L. RIGGS (2009): "Resolving the tension between access and confidentiality: Past experience and future plans at the US Census Bureau," *Statistical Journal of the IAOS*, 26, 113–122.

GROVES, R. M. (2011): "Three eras of survey research," *Public Opinion Quarterly*, 75, 861–871.

HALTIWANGER, J., R. S. JARMIN, AND J. MIRANDA (2013): "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 95, 347–361.

HAMERMESH, D. S. (2013): "Six decades of top economics publishing: Who and how?" *Journal of Economic Literature*, 51, 162–72.

HECKMAN, J. J. (2001): "Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture," *Journal of Political Economy*, 109, 673–748.

HSIEH, C.-T. AND P. J. KLENOW (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124, 1403–1448.

JARAVEL, X., N. PETKOVA, AND A. BELL (2018): "Team-specific capital and innovation," *American Economic Review*, 108, 1034–73.

KALAITZIDAKIS, P., T. P. MAMUNEAS, AND T. STENGOS (2003): "Rankings of academic journals and institutions in economics," *Journal of the European Economic Association*, 1, 1346–1366.

KINNEY, S. K., J. P. REITER, A. P. REZNEK, J. MIRANDA, R. S. JARMIN, AND J. M. ABOWD (2011): "Towards unrestricted public use business microdata: The synthetic longitudinal business database," *International Statistical Review*, 79, 362–384.

LANE, J. (2021): *Democratizing our data: A manifesto*, MIT Press.

MCGUCKIN, R. H., R. H. MCGUKIN, AND A. P. REZNEK (1993): "The statistics corner: research with economic microdata: the Census Bureau's Center for Economic Studies," *Business Economics*, 52–58.

MELITZ, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71, 1695–1725.

NAGARAJ, A. AND M. TRANCHERO (2023): "How Does Data Access Shape Science? Evidence from the Impact of U.S. Census's Research Data Centers on Economics Research," *NBER Working Paper 31372*.

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

ÖNDER, A. S. AND S. SCHWEITZER (2017): "Catching up or falling behind? Promising changes and persistent patterns across cohorts of economics PhDs in German-speaking countries from 1991 to 2008," *Scientometrics*, 110, 1297–1331.

RITCHIE, F. AND R. WELPTON (2011): "Sharing risks, sharing benefits: Data as a public good," *Work Session on Statistical Data Confidentiality, Eurostat*.

VAVRA, J. (2021): "Tracking the pandemic in real time: Administrative micro data in business cycles enters the spotlight," *Journal of Economic Perspectives*, 35, 47–66.