# The Importance of Confidential Microdata for Economic Research *

Abhishek Nagaraj
UC Berkeley & NBER

Fernando Stipanicic
University of Oslo

Matteo Tranchero
The Wharton School

August 14, 2024

**Abstract**

This chapter presents a data-driven analysis of the use and impact of U.S. confidential administrative data in economics research. Despite the high quality and growing use of these datasets, access remains heavily restricted due to privacy and security concerns. Our findings indicate that while confidential data significantly enhances research quality, as evidenced by higher publication rates in top journals and increased citations, its adoption is predominantly limited to established researchers from prestigious institutions. Our results contribute to the ongoing discussion on how to balance privacy protection with broader accessibility to confidential data.

## 1 Introduction

In recent decades, economics has become increasingly empirical, driven by the availability of high-quality, large-scale, longitudinal microdata (Angrist et al., 2020; Hamermesh, 2013). This shift has enabled significant empirical and theoretical advancements. For example, plant-level records from the U.S. Census have revealed crucial facts about exporters, influencing models like Melitz (2003) in international trade. Similar data have been invaluable in studying employment (Haltiwanger et al., 2013), wages (Abowd et al., 2018), racial disparities (Chetty et al., 2020), innovation (Jaravel et al., 2018), firm productivity (Olley and Pakes, 1996), healthcare (Finkelstein et al., 2021), mortgage markets (Beraja et al., 2019), and many other topics.

Statistical surveys and administrative data from government agencies have been prime sources of such high-quality data (Cole et al., 2020). Nevertheless, there is a shortage of work documenting their diffusion and the associated benefits. On the one hand, despite the general consensus that microdata are precious for research (Abraham et al., 2022; Chetty, 2012), evidence on their impact is still largely anecdotal (Atrostic, 2007; CES, 2017; Davis and Holly, 2006). On the other hand, the granularity and sensitivity of government microdata pose significant privacy and security risks. In many jurisdictions, these concerns have led to stringent data access controls (Foster et al., 2009), often judged too limiting by academic researchers (Card et al., 2010). A systematic investigation of confidential data use and impact would be needed to support these claims and inform data-access policies.

*This chapter presents a synopsis of results from the research project "The value of administrative data access for economics research and policy," funded by the Alfred P. Sloan Foundation (Grant Number: G-2021-16965). The interested reader is referred to Nagaraj and Tranchero (2024) and Nagaraj et al. (2024) for greater details on the results discussed.

Several questions remain unanswered: What is the diffusion of confidential data from administrative sources in the United States? How does it impact research quality and policy-relevance? What types of researchers are using these data, and what are they studying? How can we best design a data infrastructure that fosters scientific progress? Current access restrictions might penalize specific groups of researchers or reduce scientific progress on certain topics. While the optimal level of data access is ultimately a function of social preferences (Abowd and Schmutte, 2019; Fobia et al., 2020), evidence of the scientific benefits of confidential data access can help policymakers navigate the trade-off between access and confidentiality.

This chapter addresses these questions summarizing our investigation of the use and impact of confidential administrative data distributed by the U.S. Census Bureau (hereafter "Census data" for brevity) on economic research. The U.S. Census Bureau collects both statistical data through surveys (e.g., the Economic Census) and administrative records from other government agencies. These datasets are highly restricted, accessible only through physical enclaves after rigorous approval processes, and results undergo strict disclosure review (Foster et al., 2009). As part of our project, we assemble and analyze a comprehensive dataset of economic publications, including over 90,000 articles from 158 peer-reviewed journals, focusing on U.S.-based researchers (Nagaraj and Tranchero, 2024). We track the use of Census data at the paper level by using a combination of natural language processing, FOIA requests, and publicly available information. We complement the dataset with information on authors' characteristics, scientific citations, and references in policy documents.

Our findings indicate that the use of Census data has steadily increased over time. These papers are influential, with higher chances of being published in top journals and receiving more citations. Significantly, this result is not driven by selection into usage (Nagaraj et al., 2024): even when restricting the comparison to papers of the same author, papers written in an FSRDC are 50% more likely to be published in the top five journals and 28% more citations than other papers of the same author. However, we find that usage is predominantly limited to established researchers from prestigious institutions. Echoing the results from Nagaraj and Tranchero (2024), these patterns suggest that while these datasets are very valuable for research, their adoption is still limited due to obstacles that are particularly binding for researchers with fewer resources. Our findings have potentially important implications for the design of the federal data infrastructure, which we discuss in the conclusion.

This chapter is structured as follows. First, we provide an overview of Census data and their access regime. Next, we describe our novel database of economics research. We then present key findings and discuss their implications. The final section concludes with an overview of the broader research agenda and future research needed to craft better data access policies that benefit the research community.

## 2 Balancing Research Access with Data Confidentiality

Economic scholarship has changed from being primarily theory-driven into a more data-intensive discipline (Angrist et al., 2020; Hamermesh, 2013). The increased availability of government administrative data, intended as records arising as a by-product of some non-research activity, has played an essential role in driving these trends (Cole et al., 2020; Groves, 2011). Indeed, this is not surprising when taking into account the tremendous advantages that administrative data and statistical surveys offer (Cole et al., 2020). These data are usually individual-level and display a longitudinal structure that allows tracking of the same economic units over time and before and after specific interventions. Unlike custom-made surveys, administrative data afford very large samples and do not suffer from common non-response issues

(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, 2014).

The unique value of government microdata is also what might put the privacy of respondents at risk (Foster et al., 2009). The level of detail and the sensitivity of the information collected about individuals and companies might put their privacy and security at risk. Even beyond direct economic harms, moral and legal frameworks 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.[1] Data providers thus face a trade-off between their duty to protect the confidentiality of the information entrusted to them and granting broader access to their administrative records for research purposes (Lane, 2021).

In the context of the United States, the main Federal Agency collecting and distributing confidential data is the U.S. Census Bureau (CES, 2017). Faced with the tension between disseminating its data while protecting individual privacy, the U.S. Census Bureau has experimented with several solutions over the years. Some of them include the development of synthetic data or releasing anonymized Public Use Microdata Samples (Abowd and Lane, 2004; Kinney et al., 2011). Yet, these data are an imperfect substitute for the possibility of working with the universe of respondent-level micro-data. Therefore, the Census Bureau established the Center for Economic Studies (CES) in 1982 (Atrostic, 2007). The objective of the CES was to develop longitudinal databases and to host qualified academic researchers who could analyze confidential data directly onsite, thus enabling non-employee data access to microdata (Foster et al., 2009).

To minimize the risk of privacy breaches, the U.S. Census Bureau has allowed only in-situ analysis in a data enclave. Since traveling to the Census Headquarters is not practical for most researchers, the Bureau has opened multiple secure facilities across the country as part of a program known as the Federal Statistical Research Data Centers (FSRDCs).[2] FSRDCs are operated by Census staff in partnership with local universities or research institutions. Each branch meets the same physical security standards as the CES. Researchers must seek pre-approval to carry out their projects by writing a detailed proposal showing how the proposed research benefits the Census Bureau, as well as justifying the project's feasibility and proving that the project does not pose a risk of unauthorized disclosure. Approved research must obtain a "special sworn status" (SSS) with the U.S. Census Bureau, which involves passing a background check and swearing to protect data confidentiality. Finally, anyone accessing a data enclave is closely monitored, ensuring that 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 for academic research purposes. It must be noted that the Census Bureau is bound by regulation to respect its mandate to protect data confidentiality. Nevertheless, this access regime has been criticized by researchers for being too restrictive and potentially shifting the research focus away from U.S.-relevant topics towards regions with more accessible data (Card et al., 2010; Hoelzemann et al., 2024). While the continuous expansion of the FSRDC network was indeed successful in easing access constraints (Nagaraj and Tranchero, 2024), an empirical examination of the research benefits stemming from increased data access could help justify regulatory changes aimed at

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

[2] Other Federal Agencies that make their data available to researchers through the FSRDCs include 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. As already mentioned, we refer to all confidential data distributed by the U.S. Census Bureau as "Census data" for brevity.

increasing access.

# 3 Data

As part of our research project, we started by collecting qualitative evidence regarding the research impact of confidential Census data. We conducted over fifteen semi-structured interviews with users and providers of confidential data. This body of qualitative evidence allowed us to learn the history and institutional details of how the Census Bureau makes its confidential data available to researchers (see Nagaraj and Tranchero 2024 for excerpts). The interviews have been crucial to inform and guide our quantitative exercises.

Next, we compiled a new database of economic scholarship using article-level data from several sources Nagaraj and Tranchero (2024). The starting point is constituted by *EconLit*, a comprehensive bibliographic database curated by the American Economic Association. Unlike other publication databases, EconLit provides extensive coverage of economics journals and includes JEL codes for classifying research fields (Angrist et al., 2020). Our analysis includes articles published in peer-reviewed journals between 1991 and 2019 by U.S.-based researchers. These data do not contain unique identifiers for authors or institutions. We employed a disambiguation process using an algorithm by Önder and Schweitzer (2017) to identify outcomes for 17,820 researchers affiliated with 344 North American institutions. We supplemented the paper-level data with information on authors' characteristics and citations from the Web of Science database, and policy document citations from Overton. Information on the ranking of economics departments was obtained from Kalaitzidakis et al. (2003).

To identify the use of confidential data from the Census Bureau and the FSRDC network, we used several approaches. First, we searched for acknowledgments of U.S. Census confidential microdata in published papers. We employed natural language processing techniques to detect common acknowledgment phrases across databases such as Web of Science, Scopus, JSTOR, Google Scholar, IDEAS RePEc, and the NBER website. However, not all publications included these acknowledgments. To supplement our search, we obtained information on approved projects directly from the Census Bureau through FOIA requests.[3] We then manually matched these projects to their corresponding publications in EconLit.

Table 1: Descriptive statistics of our sample of Economics publications.

| Census data use | N. Articles | N. Authors | N. Institutions | Avg. yearly citations | % Articles in Top 5 journal | Avg. yearly cit. Top 5 journal | Avg. yearly policy cit. |
|---|---|---|---|---|---|---|---|
| No | 90,981 | 17,776 | 342 | 4.00 | 7.65 | 9.99 | 0.58 |
| Yes | 589 | 525 | 122 | 7.26 | 14.60 | 17.25 | 2.36 |
| Total | 91,570 | 17,820 | 342 | 4.02 | 7.69 | 10.08 | 0.59 |

*Note*: the table is reproduced with permission from Nagaraj et al. (2024).

Descriptive statistics of our sample are presented in Table 1. We identified 589 papers using confidential Census data authored by 525 U.S.-based researchers between 1991 and 2019. These papers, which tend to come from a limited number of institutions, are more likely to appear in prominent economics journals and much more impactful on average.

[3]Data obtained via FOIA request No. DOC-CEN-2020-001640. This information is updated regularly by Census officials at: `https://www.census.gov/about/adrm/fsrdc/about/ongoing-projects.html`.

# 4 The Diffusion and Impact of Confidential Microdata

Our interviews confirmed the revolutionary impact that confidential microdata had on economics research, even suggesting that they "opened up fields of research that weren't possible before" (interview T14). Yet, most of our interviewees remarked how even relatively small geographic barriers are a major hindrance to using data with tight confidentiality regimes. One administrator told us that a one-hour commute is enough to discourage a researcher from ever applying to the data (interview S94). This qualitative evidence confirms anecdotal evidence suggesting that the Census data provided research benefits thanks to the granularity of confidential data (CES, 2017; Davis and Holly, 2006; Einav and Levin, 2014), while posing challenges due to the confidentiality regime (Card et al., 2010).

Next, we leveraged our database of economics research to quantitatively document the facts emerging in our interviews. Our analysis revealed that the use of U.S. confidential microdata in economic research has grown steadily. Figure 1 shows that the proportion of published papers using these data increased from 0.21% in 1991 to 1.27% in 2019. These papers are often published in some of the most prestigious scientific outlets: around 15% of them appeared in a top five journal throughout our sample period. 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 five journal. Comparing papers of the same author, papers using confidential microdata are 50% more likely to appear in a top five journals relative to their other publications. The higher amount of citations received 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.

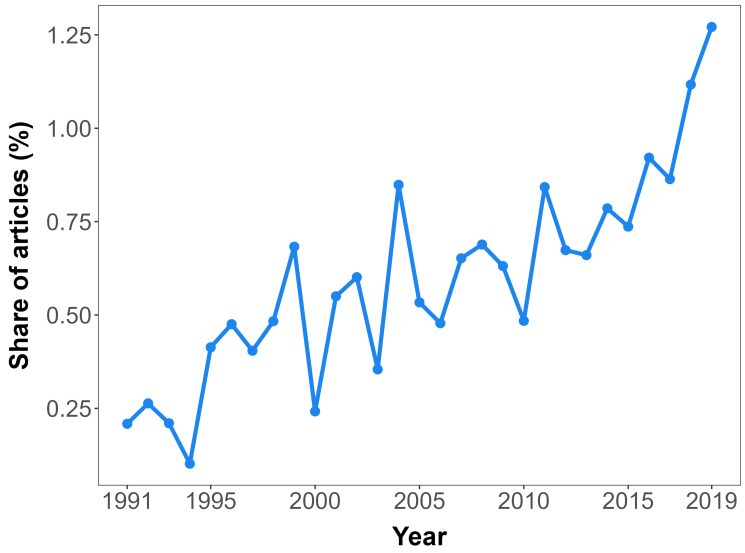


Figure 1: Use of confidential U.S. Census Bureau data in economics research

*Note*: the figure is reproduced with permission from Nagaraj et al. (2024).

Furthermore, articles utilizing Census data are more frequently cited in policy documents. While 44% of articles in our sample have at least one policy citation, the share is 75% for papers using confidential Census data. Unreported regressions confirm that the latter articles receive 80% more policy citations (Nagaraj et al., 2024). This evidence confirms what we learned in our interviews, where our respondents felt that Census data yielded findings with an outsized policy impact. Our respondents said that this was due to

two characteristics of Census data: representativeness and granularity. Without this kind of data, "special populations get lost in the noise of aggregated data sets," as Bill Maurer remarked during the opening of the UC Irvine FSRDC.[4] In practice, confidential microdata allow the study of heterogeneous effects for sub-populations, which is what policymakers need (interview P39).

We learned in our interviews that the research benefits of Census confidential data should be larger among applied researchers in environmental, international, labor, and public economics (Nagaraj and Tranchero, 2024). Indeed, 30% of articles using such data can be classified as labor economics and an additional 28% as applied microeconomics. The incidence of these two fields in the set of Census papers is more than double 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). 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).

Finally, we found that articles using Census data tend to include authors that are more established: 59% of articles have authors who previously published in a top-five economics journal, compared to 38% of those who do not use such data. Regression analysis confirms that this difference persists after accounting for factors such as number of co-authors, research field, publication year, average seniority, and seniority difference between co-authors. Additionally, 42% of these papers include at least one author from a top-ten economics department, versus 35% for other papers. This suggests that the current access model may disadvantage early-career researchers and those from less prestigious institutions, potentially hindering scientific advancement and diversity by limiting access to essential data resources (Card et al., 2010).

# 5 Discussion and the Research Agenda Ahead

This chapter summarizes our recent work to empirically document the role and impact of confidential microdata in economic research. Our findings indicate that papers utilizing U.S. Census Bureau's microdata are more likely to be published in leading economics journals, receive a higher number of scientific citations, and contribute to evidence-based policymaking. However, the use of such data is more common among established researchers from prestigious institutions, suggesting potential barriers to access for a broader range of scholars.

These preliminary results are part of a broader research agenda to understand better the role played by confidential data in scientific research. Data providers face a trade-off between enabling research access and protecting individual confidentiality (Foster et al., 2009). This situation 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 (both monetary and in terms of privacy risks) are borne by the data provider (Hoelzemann et al., 2024; Ritchie and Welpton, 2011). Therefore, it is reasonable to expect that from a welfare perspective, access to confidential data could be under-provided. Our contribution is a first step to better document the research value provided by data access so that data providers can make informed comparisons with the potential privacy loss (Abowd and Schmutte, 2019). Countries differ significantly in their policies regarding data access, project approvals, and the public disclosure of research results. More comparative work is needed to fully grasp the implication of alternative access regimes (Cole et al., 2020), and develop best practices to inform the creation of secure data access infrastructure.

[4] `https://news.uci.edu/2015/01/27/new-uci-based-center-gives-researchers-direct-link-to-us-census-data/`

In the United States, access to confidential administrative data is strictly regulated under Title 13 and 26 of the U.S. Code. The U.S. Census Bureau facilitates data access through secure physical enclaves with stringent security protocols. Recent research by Nagaraj and Tranchero (2024) suggests that expanding these facilities has increased data use by researchers located near the newly opened centers. Nevertheless, their result does not speak to how the current regulatory environment supports or stifles academic progress: besides an empirical exploration of the scientific benefits of access to Census data, we need to better understand the distributional consequences of confidentiality restrictions for researchers lacking FSRDC access. Quantifying such dimensions could help the design of alternative policies for the provision of administrative data (Card et al., 2010; Lane, 2021).

Several other research questions still require investigation. While our work to date investigates how access to data shapes the rate and quality of research, we still need to study its impact on the direction of economics research. Empirical researchers cannot study what they cannot measure (Griliches, 1994). However, data availability might affect scientific progress in non-linear ways. Perhaps counterintuitively, access to high-quality administrative data may increase the quality of economic research while creating crowding-in and limiting research to the topics answerable by these particular datasets, crowding out custom data collection projects that might have unlocked unique insights (Hoelzemann et al., 2024).

Relatedly, more work is needed to understand how access to restricted data differentially affects economics student and faculty productivity and their career outcomes. Restricted data can be a source of rents and career inequality in science, no differently than unequal access to funding (Bol et al., 2018). Policymakers might be particularly interested in the heterogeneous impact of access to administrative data across researchers' demographics and the status of the institutions where they are affiliated. In particular, minority-serving institutions are usually operating with fewer resources than high prestige research universities, which means that data access constraints can be particularly binding for them (Nagaraj and Tranchero, 2024). Changes in the mode of access could thus be an equalizing force that eliminates rent-seeking behaviors and favors scientists endowed with better ideas and talent.

Finally, more research should consider how access to data shapes the balance between theoretical and empirical approaches in economics. Access to data might facilitate the rate of progress in applied research, but it could crowd out theoretical contributions. Economists have traditionally warned against research that constitutes "measurement without theory" (Koopmans, 1947), wary that the empiricization of scientific progress could lead to policy failures such as the use of the Phillips curve to guide monetary policy (Lucas Jr, 1976). On the other hand, access to empirical data might inspire theorists to create new insights that were previously not possible, potentially in collaboration with empiricists. How precisely data shapes the idealized view of the scientific method remains a big unanswered question.

## Disclosure Statement

The authors have no conflicts of interest to declare.

## Acknowledgments

We thank Randol Yao for excellent research assistance and participants at the NBER conference "Data Privacy Protection and the Conduct of Applied Research: Methods, Approaches and their Consequences" in May 2024 for their feedback. We gratefully acknowledge the financial support of the Alfred P. Sloan Foundation (Grant Number: G-2021-16965). Any opinions and conclusions expressed herein are those of

the authors only, and any errors are our own.

## Contributions

All authors contributed equally.

# 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.

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.

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

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.

BOL, T., M. DE VAAN, AND A. VAN DE RIJT (2018): "The Matthew effect in science funding," *Proceedings of the National Academy of Sciences*, 115, 4887–4890.

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*.

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.

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

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.

GRILICHES, Z. (1994): "Productivity, R&D, and the data constraint," *The American Economic Review*, 84, 1–23.

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.

HOELZEMANN, J., G. MANSO, A. NAGARAJ, AND M. TRANCHERO (2024): "The streetlight effect in data-driven exploration," *NBER Working Paper 32401*.

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.

KOOPMANS, T. C. (1947): "Measurement without theory," *The Review of Economics and Statistics*, 29, 161–172.

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

LUCAS JR, R. E. (1976): "Econometric policy evaluation: A critique," in *Carnegie-Rochester conference series on public policy*, North-Holland, vol. 1, 19–46.

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

NAGARAJ, A., F. STIPANICIC, AND M. TRANCHERO (2024): "U.S. Census Data in Economic Research: Adoption, Diffusion, and Impact," *Working Paper*.

NAGARAJ, A. AND M. TRANCHERO (2024): "How Does Data Access Shape Science? The Impact of Federal Statistical 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*.