Volume 0 Issue 0
DOI: 00.000
ISSN: 2644-2353

Disclosing Economists' Privacy Perspectives: A Survey of American Economic Association Members on Differential Privacy and Data Fitness for Use Standards

Aaron R. Williams[†], Joshua Snoke[‡], Claire McKay Bowen[†,*], Andrés F. Barrientos[◇],
[†] Urban Institute
[‡] RAND Corporation
[◇] Florida State University

Abstract. Formal privacy provides great opportunities to how researchers conduct their work. Unlike traditional statistical data privacy methods, formal privacy can quantify privacy loss and enable policymakers to defend releases of information against growing threats to privacy from expanding computational power, swelling sources of auxiliary information, and novel techniques for attacking data. Yet, formal privacy disrupts how researchers conduct statistical analyses. Most formally private methods add large amounts of noise to statistics, and there are few feasible formally private methods that apply to statistics common in research, such as providing full inference on regression coefficients. Finally, although the design of formally private methods and systems should account for users' expectations and needs, little is known about the expectations and needs of data users — much less their knowledge and perceptions of formal privacy.

We demonstrate a framework to evaluate and identify the expectations and needs of potential users in the context of building a formally private validation server for calculating statistics on administrative tax data. We expect economists to be major users of such a validation server, so we seek to understand economists' knowledge of differential privacy (DP) and attitudes toward it. We conduct a convenience sample survey of members of the American Economic Association, receiving over 1,000 responses and a response rate in excess of 10%. Our questions identify baseline knowledge about DP, attitudes toward the differentially private framework, types of statistical methods that are most useful to economists, and how the injection of noise under DP would affect the value of the queries to the user. We provide suggestions for how the adoption of differentially private methods and tools can align with users' expectations, possible next steps to create better privacy-preserving tools for the economics community, and recommendations to improve the development of formally private tools.

Keywords: administrative data, data confidentiality, data privacy, differential privacy, social science, tax policy analysis

[*] This draft is preliminary. Please do not quote or share without permission. Please send comments to cbowen@urban.org

1. The Opportunity and Disruption of Formal Privacy

Many public policy decisions rely on official statistics and administrative data, such as education funding, emergency evacuation routes, infrastructure updates, unemployment benefits, and voter redistricting lines. The underlying data often contain sensitive information and must undergo statistical privacy modifications to protect the data. For more than half a century, privacy researchers implemented statistical disclosure control methods (we also refer to these as traditional statistical data privacy methods) based on ad hoc measures of disclosure risk that tried to predict how much information a bad actor had or how that actor might attack the data.

Now, the field of statistical data privacy and confidentiality is experiencing a paradigm shift. As computational power and availability of auxiliary information continue to increase rapidly, the demand for more robust data privacy protections has increased as well (Snoke & Bowen, 2020). When Dwork et al., 2006 introduced differential privacy (DP), this new concept of data privacy provide a completely different way for how privacy experts think about and implement a statistical privacy framework and promised to address the challenges that the traditional statistical disclosure control methods faced. At a high level, DP links the potential for privacy loss to how much the answer of a statistic is changed given the presence or absence of any possible record from any possible data set and provides a relative measure of disclosure risk. We provide the formal definition in Appendix C.

DP inspired a new research area of developing mathematical frameworks for providing a provable and quantifiable amount of privacy protection, called formal privacy. Most notably, the U.S. Census Bureau updated their 2020 Disclosure Avoidance System for the 2020 Decennial Census with a formal privacy framework (United States Census Bureau, 2021). Recently, the Internal Revenue Service has started considering formal privacy to protect tax data (Barrientos et al., 2021a, 2021b; Taylor et al., 2021). Yet, formal privacy presents both these promising opportunities and sudden disruptions to how researchers and data practitioners conduct their work.

1.1. **Formal Privacy Opportunities for Research.** This shift to formalize privacy protections through DP and other formal privacy definitions offers the exciting possibility of releasing new sources of administrative data. For instance, many United States federal statistical agencies allow researchers and data practitioners, a group we will call data users, to access confidential data for analysis through special programs that require analysis results to first undergo a thorough review before being publicly released. This process is often slow because it relies on subjective human review. It also cannot precisely account for how much disclosure risk is created by multiple individual releases within or across agencies. Further, many federal statistical agencies require U.S. citizenship to access some administrative data.

DP and other formal privacy definitions could automate the review process, removing the human element. Also, these definitions provide an accounting framework for the accumulative disclosure risk of multiple statistics released from the same data set, where past statistical disclosure control methods could not. Barrientos et al., 2019 created an example of a formally private system for academic research on administrative data. Their system, a verification server, allowed data users to calculate statistics on synthetic data and then confirm the statistics with the confidential data. For example, verification servers might report whether the inferences about the sign and statistical significance derived from the synthetic data are consistent with the confidential data.

In contrast, Barrientos et al., 2021b and Taylor et al., 2021 explored the idea of creating an automated *validation server* that releases formally private statistics, such as regression estimates and their associated standard errors, that meet a pre-defined privacy guarantee. They focused on summary statistics and regression methods for cross-sectional administrative tax data. Their research found that summary statistics like means and percentiles performed well, whereas obtaining full inference on regression coefficients performed poorly. However, determining what formally private outputs are accurate enough for public policy decisions is still an open question.

1.2. **Formal Privacy Disruptions for Research.** Although formally private frameworks show great promise, they will disrupt and affect results derived from official statistics, impact the availability of data, and demand new skills and knowledge for data analysis. In other words, methods that satisfy formal privacy definitions often add large amounts of noise to statistics. This situation forces data users to account for this noise in their analyses, but many do not have the tools or knowledge to make these adjustments.

In an ideal scenario, formally private approaches should provide full inference for a wide variety of statistical problems without requiring extensive customization. However, current formal privacy methods are far from this ideal. For example, Barrientos et al., 2021b demonstrated that most formally private methods do not provide confidence intervals for regression analyses. The few methods that do provide formally private confidence intervals require specific tailoring and may lack accuracy in scenarios where moderate to high levels of privacy are desired. A potential solution to reduce extensive customization is to work with less stringent notions of privacy and make assumptions about the data structure. This may result in a more broadly applicable formal privacy model with less customization required. For instance, Chetty et al., 2022 implemented a formal privacy method for tabular statistics and a method based on local sensitivities for more complex analyses. However, this alternative solution has other limitations. Agencies and other institutions must accept weaker notions of privacy and assume a specific data structure can be restrictive. These challenges highlight the ongoing need for developing practical and effective formal privacy methods for a diverse range of statistical applications. They also emphasize the need for further research to develop more versatile and practical solutions.

Even in the ideal scenario where there are plenty of formally private methods that allow full inference for a wide variety of statistical models, there would still be an additional problem. Data users cannot directly access data for implementing exploratory data analysis or interrogating model assumptions. This challenge potentially presents the most significant obstacle, as many data users do not know what analyses they want to run until they see the data. Although formal privacy could prevent bad exploratory data analysis practices, such as p-hacking, formally private frameworks still force data users to change how they conduct their research. For instance, the restriction of a privacy loss budget would limit how much exploratory data analysis a researcher could conduct. In other words, how do data users make robustness checks or determine model specifications without using their entire budget? What happens when a journal reviewer asks for alternative model specifications and the privacy budget is already spent?

These are all unanswered questions within the field that need to be further explored. We do not seek to answer every question in this paper, but we argue that the field will not progress until we start answering these questions and that these are *not entirely* theoretical or methodological questions. These questions should be answered, in part, by directly querying data users' needs and balancing these against privacy requirements.

2. User-Centered Formal Privacy Development

We now understand that formally private systems show promise but have the potential to disrupt how data users conduct statistical analyses. As a result, privacy researchers developing formally private methods and data stewards implementing formally private frameworks should actively seek input from data users to gain a better understanding of their needs and expectations. These privacy researchers could adopt user-centered design ideas from Abras et al., 2004, where the design process actively has "end-users influence how a design takes shape." Privacy researchers should engage potential users through interviews, focus groups, questionnaires, and extensive user testing to identify needs and expectations. Ideally, the design process should be iterative and involve regular benchmarking against these needs and expectations throughout. By incorporating user-centered design methodologies, privacy experts can ensure that the tools meet the complex needs of users, include useful statistical methods, have a clear user interface, and release standard errors that are protective and fit for use.

We highlight several areas of inquiry that are essential for developing and implementing systems that use statistical privacy methods and that can be explored through a user-centered design process. These areas are more important than ever with formal privacy, particularly in the context of public policy decisions.

2.1. **Understanding Users' Needs and Expectations.** The first area of inquiry focuses on the knowledge and perceptions of potential data users about formal privacy. Will privacy researchers need to persuade skeptical users, or is there demand for the change? How much training will data users need to use a system and to clearly communicate and understand the methodological consequences of injecting noise into results?

In the first area of inquiry, how much do data users know about DP or formal privacy? If so, what are their perceptions of DP? Abowd et al., 2019 say the economics profession must actively participate in the privacy protection debate, but, to our knowledge, no work has quantified the knowledge of formal privacy among data users.

In the second area of inquiry, what types of methods will potential users need to access? Is cross-sectional data sufficient? Are data users more interested in statistical inference or prediction? For a given method like linear regression, do users need all diagnostic information including residuals or only a subset of information like estimated coefficients?

For the third area of inquiry, how much error are users willing to tolerate? This question may prove difficult to answer because, all else equal, data users will likely want unlimited accuracy. Therefore, questions need to create a salient trade-off for researchers to weigh the accuracy of results against some adverse outcome.

2.2. **Benchmark Formally Private Results Based on User Input.** This last area of inquiry is important because it allows privacy researchers to benchmark methods against users' expectations, instead of just benchmarking against competing methods. This is necessary to determine fitness for use, an important dimension that cannot be ascertained simply by comparing to the performance of other existing methods.

Various frameworks have been proposed for benchmarking DP algorithms, such as Garrido et al., 2021; Hay et al., 2016; Tao et al., 2021, but none of these frameworks set a target based on users' input. Similarly, papers that introduce or evaluate formally private methods compare them against each other on example data or with simulation studies (Barrientos et al., 2021b; Bowen & Liu, 2020; Bowen & Snoke, 2021; Couch et al., 2019; Gillenwater et al., 2021).

None of these works answer the question, "Are formally private methods actually fit for use?" Privacy experts have worked with subject matter experts to determine thresholds for fitness for use for specific data privacy applications (Bowen et al., 2022; Li et al., 2022). This is a necessary step, but more may be learned from expanding beyond a small pool of subject matter experts.

In our prior work, Barrientos et al., 2021b, we explored creating a formally private query system for a target a specific application. Our goal was to allow for interactive queries that included statistic-specific uncertainty estimates for hypothesis testing, such as confidence intervals to allow tax researchers to estimate simple statistics and linear regressions on confidential IRS data. Throughout the process, we've asked ourselves: how would data users use a validation server? What types of methods would data users need for a validation server to be useful? How would data users use their privacy loss budget? Because users have never been directly asked these questions, we had no meaningful way of answering these questions.

3. Case Study: A Survey of Economists

We demonstrate a user-centered process that informs the development of a formally private validation server for tax economists. Economists are a large group of empirical researchers who regularly use government data, such as census data products and Internal Revenue Service tax data. We want to understand economists' knowledge and attitudes toward DP, as well as their research needs and tolerance for additional errors added to the data to protect privacy. This process informs us what queries to implement and how much error will be acceptable in developing a potential validation server for tax policy research.

3.1. Questionnaire Design and Data Collection.

We conducted a survey of members of the American Economic Association (AEA) to gather this information. We ran the survey using the AEA's existing mechanism, which invites all members via email who opted in to receiving surveys. This mechanism constitutes a convenience sample, and we cannot guarantee its representativeness. Regardless, we received a large response of over 1,000 individuals, which was a response rate in excess of 10%.[1] The contents of the the questionnaire can be found in Appendix B.

The questionnaire begins asking about demographic and professional characteristics. This allows us to determine if the respondents to our questionnaire resemble our population of interest — potential data users of a formally private validation server, or more simply, research economists. The questionnaire next asks about the types of methods research economists use with cross-sectional data. The questionnaire includes responses in randomized order and open responses for responses not included in the list. The third section evaluates research economists' knowledge and perceptions of formal privacy and DP. The final section of the questionnaire includes vignettes to explore research economists' tolerance for errors introduced by implementing DP and their preferences for using DP. Vignettes can approximate real-world behavior by presenting respondents with competing choices (Hainmueller et al., 2015).

All else equal, a researcher will generally want as little error introduced into the data as possible. We therefore try to evaluate two different trade-offs between data access and utility across four different error metrics. For each metric, we ask how much error the survey respondent would

[1] The questionnaire, which was implemented online with Qualtrics and contains four sections.

tolerate before sacrificing access to the administrative data and how much error they would tolerate before adversely responding to a journal submission as a referee.

To evaluate the clarity of our questionnaire, we performed twelve cognitive interviews with economists, sociologists, and criminologists. During the cognitive interviews, respondents spoke their thoughts aloud while working through the questionnaire (Redmiles et al., 2017). After the test respondents completed the questionnaire, we asked probing questions to further improve the questionnaire. We repeated this process with different respondents until no more clarifications were needed. We also received feedback from two expert survey methodologists at the Urban Institute. Through this process, we refined our questions and adopted best practices, such as including "Other" and "I don't know" responses to many questions, randomizing the order of responses, and reducing the length of the questionnaire (Redmiles et al., 2017).

We sent our questionnaire to members of the AEA, a professional organization of about 23,000 professionals or graduate-level students dedicated to economics research and teaching.[2] When individuals join the AEA, the membership form includes an option that reads, "I would like to receive academic surveys regarding economics or the economics profession." We sent the questionnaire to members who elected to receive these surveys from the AEA. The questionnaire was emailed to 8,850 economists on Monday, April 25, 2022 and a follow-up email was sent on Monday, May 9, 2022. We stopped accepting responses on Friday, May 13, 2022. We did not offer any incentives for completing the survey. The recruitment email is available in Appendix A.

3.2. **Response Quality and Limitations.** We received 1,028 responses to the questionnaire, over a 10% response rate, which is very good for a questionnaire without incentives. Overall, the responses capture representation of research economists and potential users of a formally private validation server, though the results may have selection bias because it is a convenience sample.

The survey includes demographics shown in Table 1. We collect these covariates to better understand the differing knowledge and needs of economists with different backgrounds and research interests. Table 1 also includes some summary statistics for our respondents.

We intended to obtain population data to create survey weights to improve the representation of the sample. Unfortunately, accurate population statistics do not exist for the population of economists, and we were unsuccessful obtaining any statistics from AEA on the population of their membership. Because of this, our results reflect a convenience sample, and we cannot assume that our results reflect an unbiased sample of AEA members or economists. This is one of our primary limitations, and we recommend that future surveys work with sampling frames or organizations where population information is collected.

Unsurprisingly, our respondents were primarily U.S. residents with doctoral degrees that work at academic institutions. We also received the strongest response among early career (fewer than 10 years' experience) respondents and individuals who are active in the peer-review process. We grouped JEL codes into three categories using a simple correlation-based variable clustering in **R** given that individuals could choose up to three codes. Generally, we find that these groupings make sense based on the content of the sub-fields. A complete list of the number of individuals who selected each JEL code in shown in Figure 1.

The quality of responses is high according to standard survey metrics. Qualtrics returns information about the IP addresses and longitudes and latitudes of respondents. Qualtrics also reports

[2] For more information about the American Economic Association, see their website at https://www.aeaweb.org/about-aea

Table 1. Selected Summary Statistics for Demographic Covariates Collected in the Survey of AEA Members

Covariate	**Survey Estimate**
Percent Residing in U.S.	70.2%
Percent Student	10.8%
Highest Degree Completed:	
Masters	15.5%
PhD or DBA	80.0%
Year of Highest Degree Completion:	
Early-Career (2013 – 2022)	49.2%
Mid-Career (2003 – 2012)	16.6%
Late-Career (Before 2003)	32.2%
Primary Employer:	
University or College	68.1%
Federal Government	10.0%
State or Local Government	2.72%
Not-For-Profit	6.61%
For-Profit Business	11.4%
Percent Peer-Reviewer in the Past Five Years	69.7%
Journal of Economic Literature Codes[3] (Select up to three):	
JEL Codes Group 1: A, B, C, K, M, N, P, Q, Y, or Z	50%
JEL Codes Group 2: D, E, F, I, J, or R	79.2%
JEL Codes Group 3: G, H, L, or O	58.0%

that 95% of IP addresses are unique, and 70% of non-missing longitude/latitude pairs are unique. For both measures, the non-unique values are low frequency. Additionally, Qualtrics includes a tool called "ExpertReview" that checks the overall quality of the collected data. Further, 100% of observations passed quality checks for completing the questionnaire within 24 hours, finishing the questionnaire abnormally fast, and questionnaire completion. The response times were reasonable and highly concentrated after the emails were sent, indicating that the recipients of the email list were the primary drivers of the responses. This also implies that the link did not circulate to unintended recipients or scammers.

Our analysis has two potential limitations. First, we see significant sample attrition over the course of the questionnaire, which corresponded to the four sections of the survey. Approximately 99% of individuals completed the demographics section, while between 80% to 85% of individuals completed the sections on knowledge and opinions of DP and the methods used with cross-sectional data. Roughly 45% – 50% of individuals completed the last section on error tolerance in analyses. The large drop-off in the last section is likely due to the unfamiliarity of the topic for many respondents.

Second, reported JEL codes do not align with (Wohlrabe & Rath, 2016). As shown in Figure 1, the respondents over-represent codes like "J. Labor and Demographic Economics" and "H. Public Economics." This is likely because the recruitment email mentioned "administrative data," which could be more popular for the overrepresented JEL codes than the underrepresented JEL codes like "G. Financial Economics." While the survey may contain selection bias due to the interest

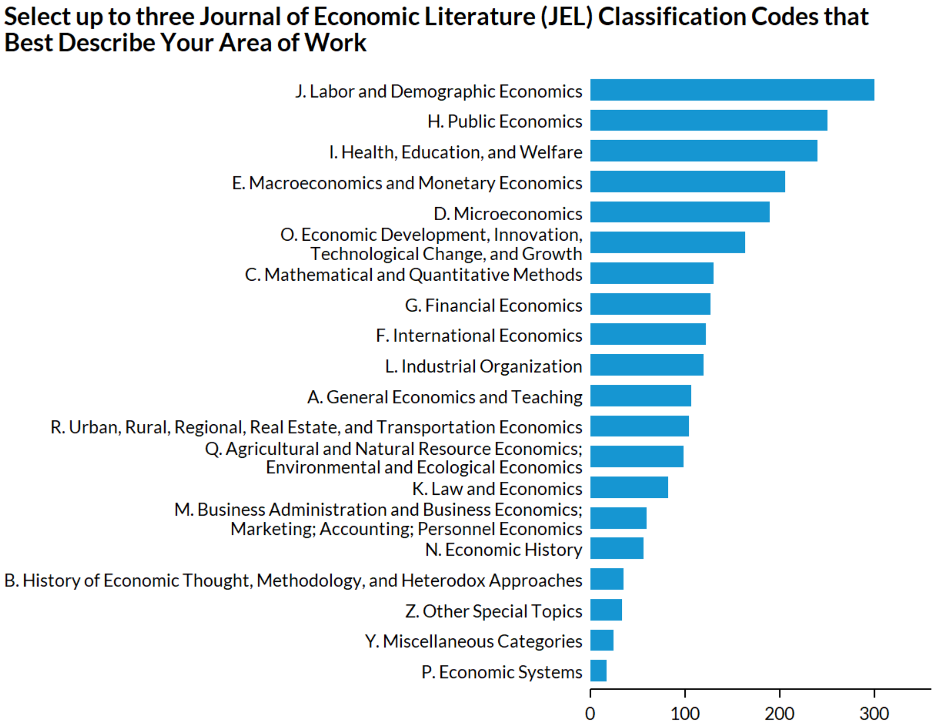


Figure 1. Respondents' JEL Sub-Fields.

of respondents in administrative tax data, the respondents may better reflect the population of potential validation server users than a representative survey of all economists.

3.3. **Summary of Results.** Our goal is to understand how economists might interact with a validation server to access administrative data. We describe the results from our survey and how they help to identify economists' baseline knowledge about DP and formal privacy, attitudes toward those frameworks, types of statistical methods that are most useful to them, and how the injection of noise under DP and formal privacy would affect the value of the queries to the user. In each section, we present results first for all respondents and then disaggregate the results by demographic covariates. We provide uncertainty estimates in the form of confidence intervals, but we do not conduct any formal hypothesis testing. The results should be viewed as informative but possibly biased. The survey serves as a proof-of-concept first and foremost. Finally, we only present selected results from the survey. We provide complete results on our GitHub Repo.[4]

3.3.1. *Economists' Knowledge and Opinions of Differential Privacy.* The first set of questions concerns economists' knowledge and opinions of differential privacy. Table 2 shows the distribution of responses to four of our questions. We found that 53.6% of respondents have "Never heard of the concept" and another 24.3% of respondents selected "Have heard the term but am not familiar with any of the details." Survey design research suggests using prominent events for question framing,

[4] GitHub repo website, https://github.com/UrbanInstitute/formal-privacy-aea-questionnaire

Table 2. Selected Survey Results for Economists' Knowledge and Opinions of Differential Privacy

Question	**Survey Estimate**
Please rate your familiarity with the concept of differential privacy (N = 847):	
Have never heard of the concept	53.6% [50.2%, 56.9%]
Have heard the term but am not familiar with any details	24.3% [21.5%, 27.3%]
Have read a blog, newspaper, or non-academic report, or discussion on the topic	12.4% [10.3%, 14.8%]
Have read an academic paper on the topic	6.97% [5.43%, 8.89%]
Feel confident implementing these methods on my own	2.72% [1.81%, 4.06%]
What share of economists you know in your professional circles have discussed the U.S. Census Bureau's adoption of differential privacy/formal privacy for the 2020 Census? (N = 843):	
None	66.5% [63.3%, 69.7%]
A minority	26.6% [23.7%, 29.7%]
A majority	6.17% [4.73%, 8.01%]
All	0.712% [0.320%, 1.58%]
Differential privacy/formal privacy is an approach to protecting privacy that distorts the results more than necessary given the actual risks (N = 816):	
Agree	18.4% [15.9%, 21.2%]
Disagree	6.74% [5.21%, 8.68%]
I don't know	74.9% [71.8%, 77.7%]
Differential privacy/formal privacy is a needed approach to preserve privacy in the face of expanding disclosure risks that current methods are unable to address (N = 813):	
Agree	13.3% [11.1%, 15.8%]
Disagree	13.5% [11.3%, 16.1%]
I don't know	73.2% [70.0%, 76.1%]

such as "before the COVID-19 pandemic" or "after 9/11." Our equivalent question is whether respondents know of anyone in their professional circles who have discussed the U.S. Census Bureau's adoption of DP/formal privacy for the 2020 Decennial Census. This change in the Census Bureau's disclosure avoidance system affected a major source of data for empirical research and spawned widespread debate (Ruggles & Van Riper, 2022) and popular news coverage (Bahrampur & Lang, 2021; Wang, 2021; Wines, 2022). Despite being asked about the highly debated adoption of DP by the U.S. Census Bureau, a significant proportion of respondents (66.5%) reported that they did not know of anyone in their professional circles who discussed it.

The results indicate that a significant majority of AEA economists lack meaningful knowledge of DP or formal privacy. This has important policy implications as it highlights the need to increase awareness and understanding of these methods and their potential impact on individuals' work through better communication strategies. Additionally, the findings suggest that most researchers

Table 3. Differences Among Respondents in the Impact and Need of Differential Privacy

Question	**Survey Estimate**
Among those who *agreed* that DP distorts the data more than necessary (N = 150):	
Agreed that it is a needed approach	20.7% [14.9%, 27.9%]
Disagreed that it is a needed approach	68.0% [60.1%, 75.0%]
Don't know that it is a needed approach	11.3% [7.15%, 17.5%]
Among those who *agreed* that DP is a needed approach (N = 107):	
Agreed that it distorts the data more than necessary	29.0% [21.1%, 38.3%]
Disagreed that distorts the data more than necessary	41.1% [32.2%, 50.7%]
Don't know that distorts the data more than necessary	29.9% [22.0%, 39.3%]

do not have firmly entrenched views on the proper interaction between privacy-preserving methods and statistical analyses

We then more closely examine those who did express an opinion about DP. We find that while many more agreed than disagreed that DP distorted the data more than necessary, an equal number of respondents agreed and disagreed that it was a needed approach that older methods cannot address. We discover that individuals who thought DP distorts the data more than necessary generally also thought that the approach was not needed to protect privacy. Conversely, among those who thought it was a needed approach, the results were mixed concerning whether it distorted the data too much.

Table 3 displays the results. One may be concerned that because a binary Agree/Disagree scale was used rather than a likert scale, some individuals felt uncomfortable answering and opted to choose "I Don't Know" instead. Given the similar rates (and strong correlation) of those who expressed no familiarity with DP and did not provide an opinion, we do not believe this played a significant role in our results.

Disaggregating results on the familiarity with DP by demographics, we learn a few differences. For simplicity we collapse the question about familiarity to three levels. We group those who have never heard the term or heard the term but are not familiar with any details and label them as "No Familiarity," and we group those who have read an academic paper or are comfortable implementing the methods themselves as "Research Familiarity."

Table 4 shows the familiarity results by demographic characteristics. Our analysis reveals that U.S. residents, economists in late stages of their careers, and those working in government are more likely to have read about or worked on the topic of DP. While not unexpected, it is noteworthy that respondents in the later stages of their careers were the most familiar with the topic. This is primarily due to a higher percentage of late-career economists having read blogs or discussions on DP, while the familiarity rates for research on DP remain consistent across experience levels.

We see the largest differences when separating respondents by their primary employer. Our analysis shows that economists working in government are more likely to have some level of familiarity with DP, which makes sense given the close connection between official statistics and privacy-preserving methods. Among those in government, almost half of respondents had some familiarity with DP.

Table 4. Familiarity with Differential Privacy by Selected Demographic Characteristics

Please rate your familiarity with the concept of differential privacy (N = 847):	**No Familiarity**	**Read Blog or Discussion**	**Research Familiarity**
U.S. Resident	73.8% [70.2, 77.2]	14.6% [12.0, 17.6]	11.6% [9.31, 14.4]
Not U.S. Resident	88.6% [83.8, 92.0]	6.78% [4.19, 10.8]	4.66% [2.59, 8.24]
Early-Career	81.0% [76.9, 84.5]	8.54% [6.19, 11.7]	10.5% [7.86, 13.9]
Mid-Career	79.3% [72.1, 85.1]	12.0% [7.68, 18.3]	8.67% [5.09, 14.4]
Late-Career	72.7% [67.3, 77.6]	18.2% [14.1, 23.1]	9.09% [6.26, 13.0]
Federal, State, or Local Gov.	60.7% [51.3, 69.4]	19.6% [13.3, 28.1]	19.6% [13.3, 28.1]
Industry (For or Not-for Profit)	75.0% [67.3, 81.4]	13.2% [8.56, 19.8]	11.8% [7.45, 18.2]
Academic Institution	81.9% [78.6, 84.8]	10.8% [8.56, 13.6]	7.28% [5.44, 9.67]
JEL Codes Group 1	81.9% [78.0, 85.3]	9.39% [6.96, 12.6]	8.69% [6.35, 11.8]
JEL Codes Group 2	76.0% [72.6, 79.0]	13.2% [10.8, 15.9]	10.9% [8.74, 13.4]
JEL Codes Group 3	80.0% [76.3, 83.3]	11.7% [9.15, 14.8]	8.32% [6.20, 11.1]

Table 5 displays the results for the two questions on respondents' opinions of DP. The first question asks whether respondents agree that DP distorts the data more than necessary, and the second question asks whether DP is a necessary technique to preserve privacy risks that the current methods cannot address. As stated previously, these questions did not always return opposing responses.

We again see notable differences between U.S. and non-U.S. residents, different career stages, and government versus industry and academia. For the first notable difference, we see that non-U.S. residents are (by a surprisingly large magnitude) less likely to agree that DP distorts the data more than necessary and are somewhat more likely to agree that it is necessary to preserve privacy. Given the larger adoption of DP in the U.S., we did not expect this result. A possible explanation may be that our sample of non-U.S. economists is biased because they are also AEA members. Without AEA membership statistics, we cannot dig deeper to understand this further. However, these results likely reflect non-U.S. economists who have reasonably strong ties to the U.S. research community.

Another notable difference, and perhaps less surprisingly, is that late-career respondents have a more negative opinion of DP. This may reflect that these researchers are more accustomed to other existing privacy-preserving methods. Further, we see that late-career respondents are more likely to have an opinion about DP, whether positive or negative, which is consistent with the familiarity results.

Our analysis next shows that respondents working in federal, state, or local government positions are more likely to agree that DP distorts the data more than necessary *and* believe that it is needed to preserve privacy. This results reflects the current conundrum that many in official statistics face: although stronger privacy methods are needed, formally private tools do not currently exist that allow for the type of statistical analyses that researchers have come to expect.

Last, as with the familiarity question, we note that there are only minor differences between economists in different JEL subfields.

Table 5. Opinions of Differential Privacy by Selected Demographic Characteristics

Differential privacy distorts the results more than necessary given the actual risks (N = 816):	**Agree**	**Disagree**	**Don't Know**
U.S. Resident	21.1% [18.1, 24.7]	6.40% [4.69, 8.68]	72.4% [68.6, 75.8]
Not U.S. Resident	10.8% [7.34, 15.6]	7.66% [4.80, 12.0]	81.5% [75.9, 86.1]
Early-Career	13.1% [10.1, 16.8]	8.04% [5.74, 11.2]	78.9% [74.6, 82.6]
Mid-Career	18.2% [12.8, 25.3]	3.38% [1.41, 7.89]	78.4% [71.0, 84.3]
Late-Career	26.5% [21.5, 32.1]	6.72% [4.27, 10.4]	66.8% [60.9, 72.2]
Federal, State, or Local Gov.	21.5% [14.7, 30.3]	10.3% [5.76, 17.7]	68.2% [58.8, 76.4]
Industry (For or Not-for Profit)	16.8% [11.4, 24.0]	8.76% [5.03, 14.8]	74.5% [66.5, 81.1]
Academic Institution	18.2% [15.2, 21.6]	5.59% [3.98, 7.81]	76.2% [72.6, 79.5]
JEL Codes Group 1	14.5% [11.4, 18.3]	6.90% [4.80, 9.82]	78.6% [74.3, 82.3]
JEL Codes Group 2	20.1% [17.2, 23.3]	6.59% [4.93, 8.74]	73.4% [69.9, 76.6]
JEL Codes Group 3	18.9% [15.7, 22.7]	5.91% [4.13, 8.38]	75.2% [71.1, 78.8]

Differential privacy is a needed approach to preserve privacy that current methods are unable to address (N = 813):	**Agree**	**Disagree**	**Don't Know**
U.S. Resident	11.3% [9.02, 14.2]	16.4% [13.6, 19.6]	72.3% [68.5, 75.7]
Not U.S. Resident	18.5% [13.9, 24.1]	5.86% [3.42, 9.84]	75.7% [69.6, 80.9]
Early-Career	15.4% [12.2, 19.3]	8.59% [6.19, 11.8]	76.0% [71.5, 80.0]
Mid-Career	10.1% [6.19, 16.2]	13.5% [8.87, 20.1]	76.4% [68.8, 82.5]
Late-Career	11.9% [8.56, 16.4]	20.9% [16.4, 26.2]	67.2% [61.3, 72.5]
Federal, State, or Local Gov.	21.7% [14.8, 30.6]	14.2% [8.69, 22.2]	64.2% [54.5, 72.7]
Industry (For or Not-for Profit)	15.9% [10.7, 23.1]	10.9% [6.64, 17.3]	73.2% [65.1, 79.9]
Academic Institution	11.1% [8.74, 13.9]	14.1% [11.4, 17.2]	74.9% [71.1, 78.3]
JEL Codes Group 1	12.8% [9.91, 16.5]	9.88% [7.32, 13.2]	77.3% [72.9, 81.1]
JEL Codes Group 2	13.5% [11.1, 16.3]	14.8% [12.3, 17.8]	71.7% [68.1, 75.0]
JEL Codes Group 3	12.7% [10.0, 15.9]	13.7% [10.9, 17.1]	73.6% [69.5, 77.3]

3.3.2. *Economists' Privacy Error Tolerance.* We next present information on questions related to economists' tolerance for the errors introduced by formal privacy metrics. As mentioned earlier, we observe a lower response rate in this section of questions compared to others, with approximately only 45% to 50% of respondents completing it. Despite conducting cognitive tests with economists of varying backgrounds and experience levels, we suspect this lower response rate is due to the unfamiliarity of the topic. We acknowledge this as an area for further investigation in Section 4.

We asked about four different ways that error might impact the results of a statistical query. We asked respondents to provide their tolerance for (1) the proportion of significance mismatch of the confidential and noisy statistics, (2) sign mismatch of the confidential and noisy statistics, (3) absolute relative bias in the point estimate, and (4) the confidence interval ratio between the confidential and noisy CIs. We observe a strong correlation between responses to the first three

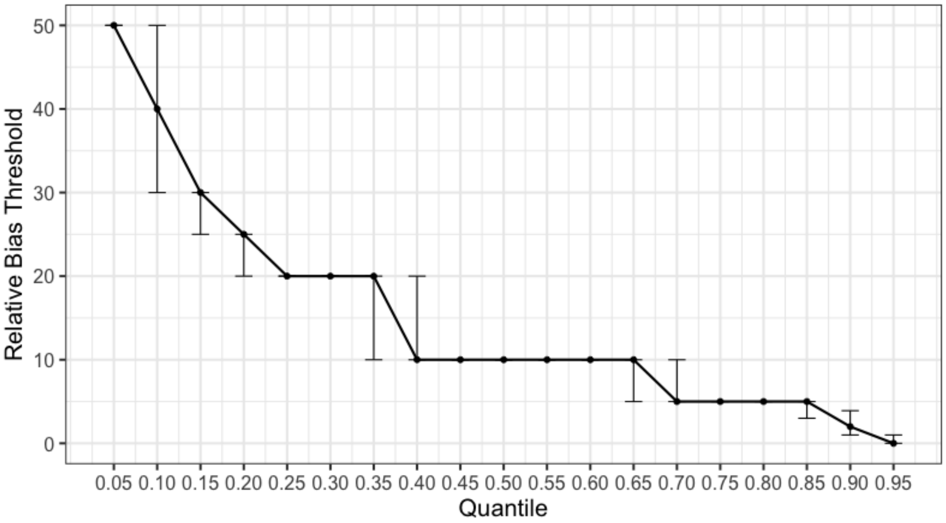


Figure 2. Respondents' absolute relative bias thresholds by quantiles. Bootstrap 90% confidence intervals shown.

questions but not the fourth question. We expect the concept of a confidence interval ratio is unfamiliar to many respondents, which may explain the inconsistency between this question and the other three[5].

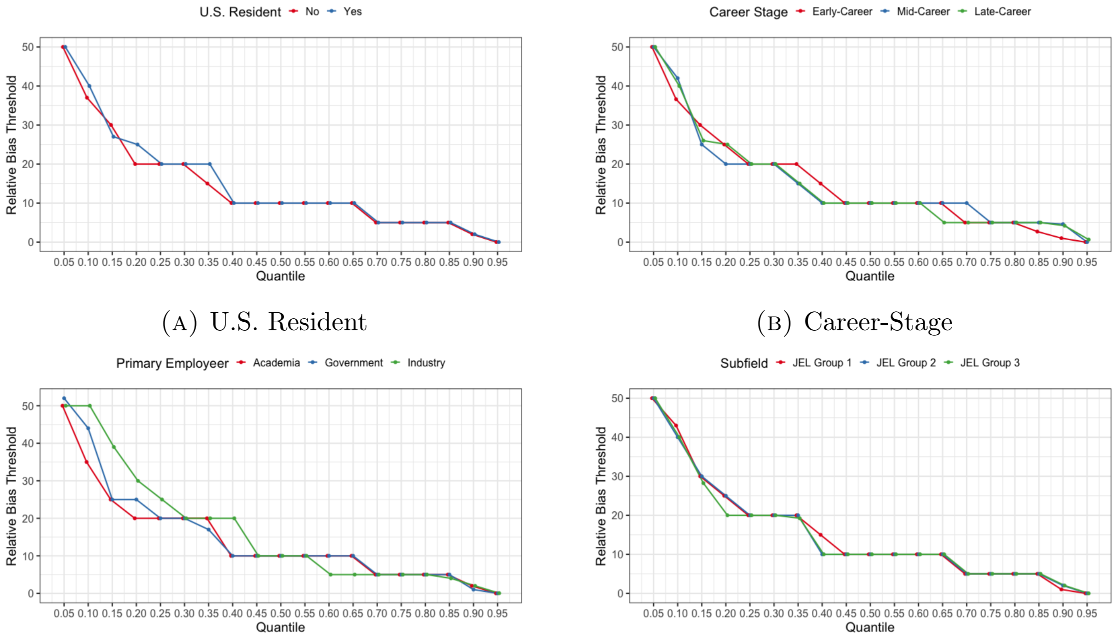
For each error metric, we ask respondents about two scenarios. At what point would they be willing to "sacrifice access to the administrative data" or "adversely respond to a journal submission" as a referee due to the amount of noise added? However, we observed no significant difference between the two scenarios across all four metrics. We hoped that respondents would provide more lenient tolerances when the alternative was sacrificing access entirely, but either respondents did not fully grasp the question or do not have separate preferences.

We only present the results for absolute relative bias using the question about sacrificing access to administrative data. Those interested in results for the other metrics can find them on our GitHub repository.[6] These results give an example of how one might use survey responses to help develop practical privacy implementations. Given the trade-off between privacy and accuracy, such as relative bias, the maintainers of a validation server might use users error tolerances to understand where they would need to set a privacy budget to meet users' needs.

Absolute relative bias is the amount of noise introduced into an estimate divided by the estimate without noise. For example, if the estimate without noise is 10 and the difference between the estimate with noise and the estimate without noise is 5, then the absolute relative bias is 50. Responses range from 0 to 150 with most responses concentrated below 50. The median response is 10, and as mentioned earlier most respondents provided the same value whether they were "sacrificing access to the administrative data" and "adversely responding to a journal submission" as a referee.

[5] We also removed one individual who responded "10,000" as their confidence interval ratio tolerance, which we assumed was not an accurate answer.

[6] GitHub repo website, https://github.com/UrbanInstitute/formal-privacy-aea-questionnaire



(A) U.S. Resident

(B) Career-Stage

(C) Primary Employeer

(D) JEL Subfields

Figure 3. Differences in relative bias thresholds for demographic groups. Confidence intervals not shown but no significant differences between groups detected.

We present the results for respondents' relative bias tolerance by quantile in Figure 2. This visual allows us to consider what percentage of potential users we would satisfy given different levels of added noise. For example, only 5% of respondents would accept an absolute relative bias of 50, while 25% of respondents would use the server if the expected relative bias was only 20. We bootstrap the estimates to get confidence intervals to assess the variability in the results.

We further separate the results by demographic groups, shown in Figure 3. We do not find any large differences between groups, which suggests that users are fairly consistent on their error tolerances regardless of background. We observe the largest difference between those who worked in industry (either for- or not-for profit) and those in government and academia. A larger proportion of industry economists report higher error tolerances, but overall the median is the same. For simplicity, we do not depict the confidence intervals, but given the uncertainty in the estimates none of the groups differed by more than the estimated intervals.

While these results do not reveal much difference between groups, this type of survey and analysis could easily inform a practical implementation that may have specific target users. We see these results as a proof-of-concept that useful metrics can be gathered through directly surveying potential data users.

Finally, we report respondents answers about how they would use a finite privacy budget. Of those who responded, the pattern in the responses is very clear. We find 62.9% of respondents prefer "One regression specification with moderate noise and five robustness checks with more noise," whereas 26.8% respondents prefer "One regression specification with less noise." Only 10.3% respondents want "Ten regression specifications with more noise." This suggests that respondents want to spend extra budget on at least one estimated regression model with relative precision. For

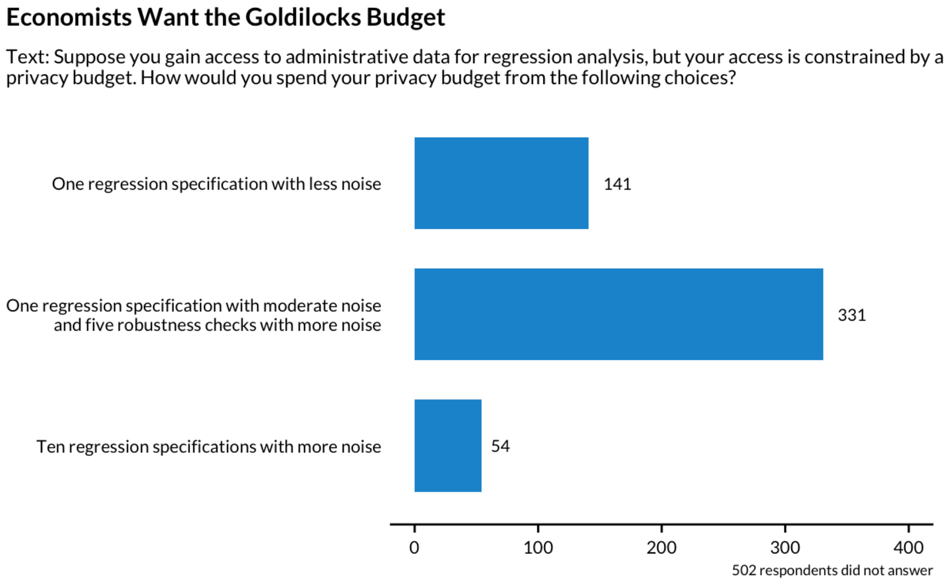


Figure 4. AEA economists' preferred means of using a privacy budget given three simple scenarios.

some respondents, this means using all budget on one estimated model, while for most respondents this means using extra budget on one model and saving some budget for robustness checks.

3.3.3. *Methodological Directions for Privacy Research.* Lastly, we show results from the the questions about what output potential users of a validation server would use. We first asked the types of methods users frequently employ for their data analysis, with the results shown in Figure 5.

We asked respondents to rank the frequency of their use of various methods over the past year. We examine how often each method was rated "Always" or "Frequently." We found almost two-thirds of respondents selected "Multiple linear regression" and "Merging multiple data source." Meanwhile, more advanced methods for cross-sectional data like "Regression discontinuity or kink design" (0.17) and "Tobit" (0.08) are much less popular methods. Because respondents could select more than one option, we find that most selected "multiple linear regression" along with at least one more complex method.

In addition to the structure options, a follow-up open-ended question allowed respondents to suggest methods not in the original list. There were 288 responses with a median response length of 32 characters. We converted all responses to lowercase, removed extraneous white spaces, and dropped blatantly incorrect responses. Next, we used a dictionary to reconcile minor differences between responses with the same meaning (e.g., "difference in difference" and "difference in differences").

Figure 6 shows the frequency of individual words, also called unigrams, after stemming the words and dropping stop words from the "onix," "SMART," and "snowball" lexicons. Although many of the open-ended responses were brief, they consisted of multiple words. Therefore, it is important to consider the collocation of words. Additionally, Figure 6 shows a bigram graph for words that occur more than five times. "Difference-in-difference," a panel method, and "panel data" were very popular responses along with "Time series." This suggests that economists want access to data that allows for more sophisticated research methods than cross-sectional data.

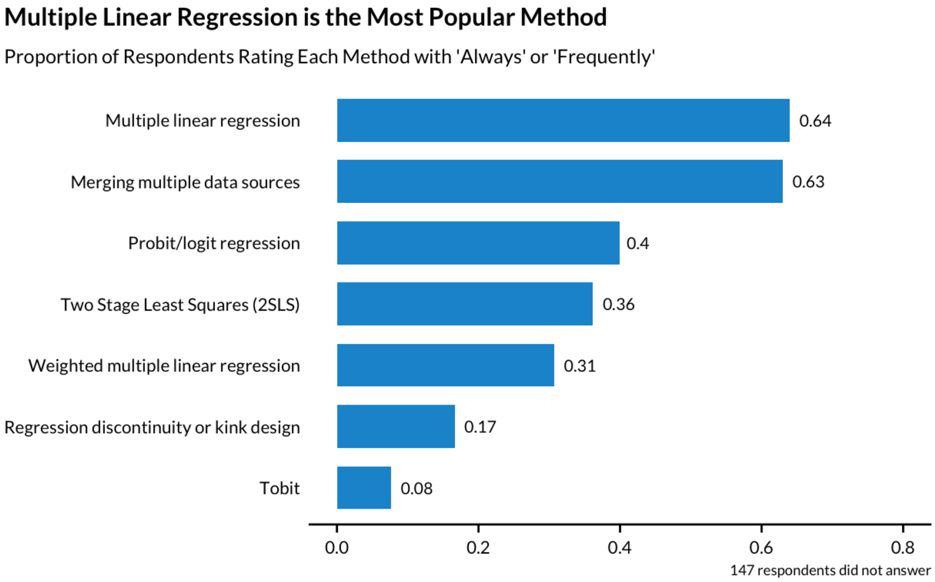


Figure 5. Please rate the following methods based on how frequently you have used the method on cross-sectional data in the past year.

Table 6. Unigram frequency from open-ended responses to methods used.

word	frequency
difference	101
regression	37
method	33
moments	30
analysis	27
data	27
generalized	25
models	25
panel	24
matching	22

Anticipating that multiple linear regression would be the most popular method, we asked about the importance of including specific information returned after fitting a linear regression model. This is crucial because some differentially private methods can only provide certain information and not everything one might expect from standard statistical packages. For example, methods relying on objective-perturbation-based approaches (Chaudhuri et al., 2011; Fang et al., 2019; Gong et al., 2019; Zhang et al., 2012) return estimated coefficients but do not return estimated standard errors and cannot be used for full statistical inference.

We compare how frequently respondents rated each method as "Very important" or "Important." We find 75% of respondents selected "Estimated coefficients," and 72% selected "Estimated standard errors, T-statistics for coefficients, or P-values for coefficients." Roughly half of respondents selected

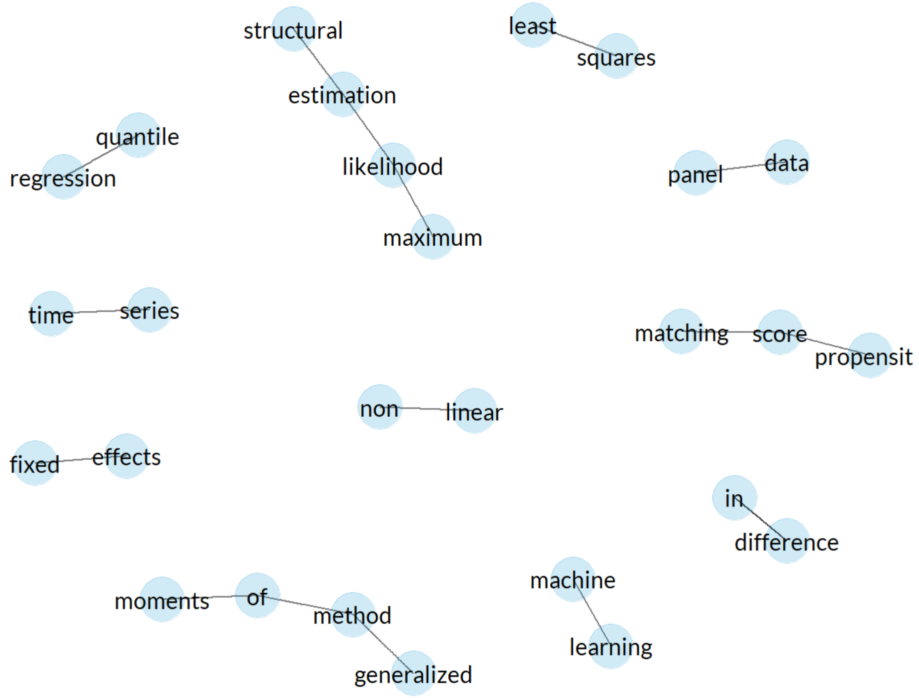


Figure 6. Bigram graph of the most common responses from the open response question.

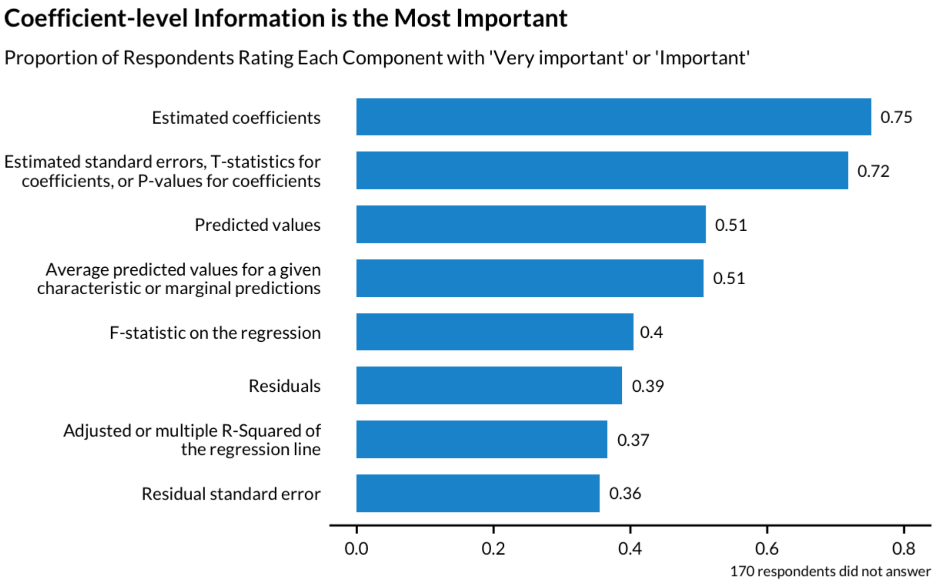


Figure 7. Please rate the following information obtained from multiple linear regression based on its importance for your work.

"Predicted values" and "Average predicted values for a given characteristics or marginal predictions." Predicted values have been a focus of formal privacy experts Dwork and Roth, 2014; however, the latter is an approach that is likely more popular with economists than statisticians or computer scientists. Lastly we find that a smaller but non-negligible amount of economists desire residual and model performance information.

The results of this section can help inform future work in the area by highlighting the types of methods that economists use. Unsurprisingly, multiple linear regression forms the backbone for the largest percentage of potential users, and more complex methods are often built on top of regression. Providing uncertainty estimates such as standard errors is more important for economists than predicted values or model diagnostic tools. We choose not to evaluate differences in methods by demographic characteristics, but a similar analysis could be performed on a target group of interest.

4. Discussion

We conducted a convenience sample survey of AEA members to understand their knowledge of formal privacy, their attitudes toward formally private frameworks, the types of statistical methods most useful to them, and their tolerance for the errors introduced by formal privacy metrics. We provide concluding thoughts on implementing a formally private validation server for administrative tax data and lessons learned for integrating user-centered design for statistical disclosure limitation.

4.1. A Validation Server for Administrative Tax Data.

Based on our survey results, we learn that economists have a limited understanding of DP and formal privacy. As a result, implementing a formally private validation server would require substantial training in its usage and reporting of results with injected noise.

Our survey also identifies economists' mixed skepticism about formal privacy. This suggests that we need to thoroughly motivate a formally private validation server. We can accomplish this by highlighting real confidentiality risks releasing statistics from administrative data and making the case that privacy-preserving technologies could expand access to administrative data that would otherwise be unavailable.

Another important result is that privacy researchers should benchmark their methods against users' error tolerances and expectations instead of solely benchmarking against other formally private methods. Generally speaking, we find that respondents expected modest errors even when facing a trade-off as extreme as totally sacrificing access to administrative data. Precise estimates of error tolerances under different scenarios need to be replicated with additional surveys before they can lead to specific policy implementations.

Finally, although multiple linear regression on cross-sectional data may suffice for some data users, there is a growing demand for panel data and difference-in-differences techniques, which allow for more sophisticated research designs. However, methods for the former are insufficient while methods for the latter are almost non-existent.

4.2. User-Centered Design for Statistical Data Privacy.

Despite rapid growth in the field of formal privacy, the modern field of statistical data privacy is still in an early stage. Many questions remain about the practical implementation of formal privacy, and public policymakers have not yet made many choices about the trade-off between privacy loss and data accuracy.

Our experience in developing a validation server and conducting this questionnaire leads us to recommend that privacy experts include prospective data users in the developments of formally private methods and tools to ensure that data users' expectations are met. Based on our case study, we identify the following key takeaways and lessons learned for collecting high-quality data user feedback.

We recommend including a question about whether a potential user would consider using a method or tool (in our case, a formally private validation server). This is important for understanding the size of a potential user base and for stratifying results. We also recommend ensuring that statistics are collected on the population of interest, so that results can be appropriately calibrated to the population of interest to account for selection bias.

Care should be taken to reduce survey attrition. We asked questions about DP and formal privacy. Given low levels of familiarity about both, we recommend only asking questions about DP, which is more familiar. Similarly, we tried to distinguish between losing access to the data and adversely responding to a journal review when asking about error tolerance. We didn't see much difference. One possible reason for this result could be that we presented each scenario in succession for each error metric. Future work could decouple these scenarios to elicit more varied responses or drop one of the scenarios to reduce respondent attrition.

Estimating population error tolerance is a key motivation and novel contribution of this work. Respondents appeared to consider the trade-off instead of stating an arbitrarily high standard for accuracy. However, we believe there is room for improvement. Future work should attempt to develop simpler vignettes. We use vignettes that are very general and don't mention specific statistics so our results generalize to many types of benchmarking. A drawback is this may have confused or exhausted respondents. Future vignettes should consider using specific statistics and examples to ease respondent burden and improve response quality.

Adopting user-centered design is an important mindset for building tools. While sending a questionnaire to potential users is an important and necessary step for understanding the needs and users of any tool, it is not sufficient. Privacy researchers and data stewards should consider a range of additional techniques including interviews and focus groups, user testing of interfaces, benchmarking example studies and simulation studies against user-derived error tolerances, and experimental studies for users, decision making.

Disclosure Statement

The authors have no conflicts of interest to declare.

Acknowledgments

Funding. This research was funded by the Alfred P. Sloan Foundation [G-2020-14024] and National Science Foundation National Center for Science and Engineering Statistics [49100420C0002 and 49100422C0008].

Collaborators. We would like to thank our collaborators at the Statistics of Income Division, especially Barry Johnson and Victoria Bryant, for their amazing support for this project. We also thank our stellar validation server project team, consisting of Leonard Burman, John Czajka, Surachai Khitatrakun, Graham MacDonald, Rob McClelland, Sybil Mendonca, Josh Miller, Livia Mucciolo, Madeline Pickens, Silke Taylor, Deena Tamaroff, Erika Tyagi, and Doug Wissoker. Thank you to Tim Triplett for reviewing our survey design and methodology. Finally, we thank our advisory board for their invaluable advice and support. The members are John Abowd, Jim Cilke, Jason DeBacker, Nada Eissa, Rick Evans, Dan Feenberg, Max Ghenis, Nick Hart, Matt Jensen, Barry

Johnson, Laura Kawano, Ithai Lurie, Ashwin Machanavajjhala, Shelly Martinez, Robert Moffitt, Amy O'Hara, Jerry Reiter, Emmanuel Saez, Wade Shen, Aleksandra Slavković, Salil Vadhan, and Lars Vilhuber.

Contributions. We used the CRediT taxonomy[7] to indicate author contributions to this paper.
ARW: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing
JS: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing
CMB: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft, and Writing – review & editing
AFB:

Appendix A. Recruitment Email

This appendix contains the recruitment email we sent to the American Economic Association (AEA) members who indicated they would like to receive academic surveys regarding economics or the economics profession.

Dear [AEA member's name],

We are researchers at the Urban Institute who are developing tools that could expand access to administrative data that are not publicly available. We are interested in your responses to a brief questionnaire. Responding to this questionnaire will benefit you and economists by shaping future access to administrative data. Interested respondents can also opt in to see aggregated results after the questionnaire has been completed.

Accessing administrative data currently requires a lengthy approval process and eligibility requirements, such as being a US citizen. Our tools would allow researchers to access these administrative data through a validation server, where researchers submit queries and receive results that satisfy privacy control criteria. The data included could be covered by Title 13, CIPSEA, and/or Title 26.

We are interested in identifying the types of methods that would be most useful to economists and how the impacts of disclosure control techniques (or methods of data privacy and confidentiality) would affect the value of the data. The questionnaire should take about 10-15 minutes.

Please use the link below to get started.

[Questionnaire link]

You received this email because you indicated on your AEA membership that you would like to receive academic surveys regarding economics or the economics profession.

Thanks again for your participation.

Sincerely,

Claire McKay Bowen, PhD, and Aaron R. Williams

Appendix B. Questionnaire

This appendix contains the content of the Qualtrics survey we sent to the American Economic Association members who indicated they would like to receive academic surveys regarding economics or the economics profession.

[7] https://credit.niso.org

We are researchers at the Urban Institute who are developing tools that could expand access to administrative data that are not publicly available. **We are interested in your responses to a brief questionnaire. Responding to this questionnaire will benefit you and economists by shaping future access to administrative data.** Interested respondents can also opt in to see aggregated results after the questionnaire has been completed.

Accessing administrative data currently requires a lengthy approval process and eligibility requirements, such as being a US citizen. Our tools would allow researchers to access these administrative data through a validation server, where researchers submit queries and receive results that satisfy privacy control criteria. The data included could be covered by Title 13, CIPSEA, and/or Title 26.

We are interested in identifying the methods that would be most useful to economists and how the impacts of disclosure control techniques (or methods of data privacy and confidentiality) would affect the value of the data. The questionnaire should take about 10-15 minutes.

1.1: Do you reside in the United States?

- Yes
- No

1.2: Are you currently a student?

- Yes
- No

1.3: What is the highest degree or level of school you have completed?

- High school or General Education Development (GED)
- Bachelor's degree
- Master's degree
- JD, MD, or other terminal degree
- PhD or DBA

1.4: In what year did you finish your highest degree or level of school?

1.5: Which of the following best describes your most recent primary employer?

- University or college
- For-profit business
- Not-for-profit
- Federal government
- State or local government

1.6: Have you refereed at least one peer-reviewed journal article in the past five years?

- Yes
- No

1.7: Select up to three Journal of Economic Literature (JEL) Classification Codes that best describe your area of work:

- A. General Economics and Teaching
- B. History of Economic Thought, Methodology, and Heterodox Approaches
- C. Mathematical and Quantitative Methods
- D. Microeconomics
- E. Macroeconomics and Monetary Policy
- F. International Economics

- G. Financial Economics
- H. Public Economics
- I. Health, Education, and Welfare
- J. Labor and Demographic Economics
- K. Law and Economics
- L. Industrial Organization
- M. Business Administration and Business Economics; Marketing; Accounting; Personal Economics
- N. Economic History
- O. Economics Development, Innovation, Technological Change, and Growth
- P. Economic Systems
- Q. Agricultural and Natural Resource Economics; Environment and Ecoloogival Economics
- Y. Miscellaneous Categories
- Z. Other Special Topics

2.1: Please rate the following methods based on how frequently you have used the method on cross-sectional data in the past year:

Multiple linear regression

- Never
- Infrequently
- Frequently
- Always

Two Stage Least Squares (2SLS)

- Never
- Infrequently
- Frequently
- Always

Tobit

- Never
- Infrequently
- Frequently
- Always

Probit/logit regression

- Never
- Infrequently
- Frequently
- Always

Merging multiple data sources

- Never
- Infrequently
- Frequently
- Always

Weighted multiple linear regression

- Never
- Infrequently
- Frequently
- Always

Regression discontinuity or kink design

- Never
- Infrequently
- Frequently
- Always

2.2: Please add any methods not listed above that you use for a typical research project on cross-sectional administrative data:

2.3: Please rate the following information obtained from multiple linear regression based on its importance for your work:

Adjusted or multiple R-Squared of the regression line
Residuals
Residual standard error
Predicted values
Estimated coefficients
Average predicted values for a given characteristics or marginal predictions
F-statistic on the regression
Estimated standard errors, T-statistics for coefficients, or P-values for coefficients

2.5: Please add any linear regression information that is not listed above:

3.1: Please rate your familiarity with the concept of *differential privacy.*

- Have never heard of the concept
- Have heard the term but am not familiar with any details
- Have read a blog, newspaper, or non-academic report or discussion on the topic
- Have read an academic paper on the topic
- Feel confident implementing these methods on my own

3.2: Please rate your familiarity with the concept of *formal privacy*

- Have heard of the concept
- Have heard the term in relation to Differential Privacy, but I do not know the difference
- Am familiar ith the concept and the distinction from differential privacy

3.3: What share of economists you know in your professional circles who have discussed the US Census Bureau's adoption of *differential privacy/formal privacy* for the 2020 Census?

- None
- A minority
- A majority
- All

3.4: What share of economists you know in your professional circles work with differential privacy/formal privacy methods?

- None
- A minority
- A majority

- All

The following statements are about *differential privacy/formal privacy.* Please choose the option that best reflects your view of the statement.

3.5 I do not have an opinion on *Differential Privacy/formal privacy*

- Agree
- Disagree

Differential privacy/formal privacy is an approach to protecting privacy that distorts the results more than necessary given the actual risks.

- Agree
- Disagree
- I don't know

Differential privacy/formal privacy is a new approach to protecting privacy that seems to be mostly popular among computer scientists.

- Agree
- Disagree
- I don't know

Differential privacy/formal privacy is a needed approach to preserve privacy in the face of expanding disclosure risks that current methods are unable to address.

- Agree
- Disagree
- I don't know

We are investigating a system to allow researchers or data practitioners to submit a statistical analysis through a web-based interface (a validation server) that then applies the analysis on cross-sectional administrative data (e.g., health records, tax data, or unemployment insurance data) using differential privacy or related privacy-preserving methods. The researcher or data practitioner would then receive a result that has some random error added to preserve privacy. The following questions are based on this framework and assume your **only** access to this administrative data would be the noisy estimates. **The trade-off you should consider is between noisy estimates and no access**.

For questions 4.1 and 4.2, suppose we submitted a regression analysis 1,000 times to the validation server and the outputs received are a coefficient estimate and associated standard error (1,000 of them), both with some random error added. Consider the following accuracy measures for the 1,000 coefficient estimates and associated standard errors.

4.1: Significance mismatch is the relative frequency with which a noisy estimate has a different statistical significance (assume 0.05 level) than the estimate without noise. For example, 0% means the coefficients with and without noise always have the same statistical significance, 40% means the noisy coefficient has a different statistical significance than the coefficient without noise for 400 of 1,000 estimates, and 100% means the coefficients with and without noise always have different statistical significance. For the 1,000 coefficient estimates:

- What is the largest percentage of **significance mismatch** you would accept before sacrificing access to the administrative data?
- What is the largest percentage of **significance mismatch** you would accept as a referee before adversely responding to a journal submission?

4.2: Sign mismatch is the relative frequency with which a noisy estimate is expected to have a different sign (positive or negative) than an estimate without noise. For example, 0% means the coefficients with and without noise always have the same sign, 40% means the noisy coefficient has a different sign than the coefficient without noise for 400 of 1,000 estimates, and 100% means the coefficients with and without noise always have different signs. For the 1,000 coefficient estimates:

- What is the largest percentage of **sign mismatch** you would accept before sacrificing access to the administrative data?
- What is the largest percentage of **sign mismatch** you would accept as a referee before adversely responding to a peer-review journal submission?

4.3: Absolute relative bias is the amount of noise introduced into an estimate divided by the estimate without noise. For example, if the estimate without noise is 10 and the difference between the estimate with noise and the estimate without noise is 5, then the absolute relative bias is 50

- What is the highest amount of **absolute relative bias** you would accept in an estimate before sacrificing access to the administrative data?
- What is the highest amount of **absolute relative bias** you would accept in an estimate as a referee before adversely responding to a journal submission?

4.4: Confidence interval ratio is the width of the noisy confidence interval divided by the width of the confidence interval without noise. For example, 1 means the confidence intervals have identical widths and values, and >1 means the noisy confidence interval is wider than the confidence interval without noise.

- What is the largest **confidence interval ratio** you would accept in an estimate before sacrificing access to the administrative data?
- What is the largest **confidence interval ratio** you would accept in an estimate as a referee before adversely responding to a journal submission?

5.1: Suppose you gain access to administrative data for regression analysis, but your access is constrained by a privacy budget. How would you spend your privacy budget from the following choices?

- One regression specification with less noise
- One regression specification with moderate noise and five robustness checks with more noise
- Ten regression specifications with more noise

Appendix C. Background on Differential Privacy

This work is part of an effort to develop a validation server for querying United States taxpayer data from the Internal Revenue Service, which allows individuals to query noisy statistics without accessing the confidential data (Barrientos et al., 2021b). In collaboration with the Internal Revenue Service Statistics of Income Division, we are developing tools that could expand access to administrative tax data that are not publicly available using differentially private methodologies.

We introduce formal privacy, differential privacy, validation servers, and the evaluation of differentially private methods in this section.

Formally private methods are privacy methods that can mathematically prove the privacy loss from the publication of data. *Differential privacy* is a formally private method for releasing statistics (Dwork et al., 2006).

Definition 1. ***Differential Privacy*** *(Dwork et al., 2006): A sanitization algorithm, $\mathcal{M}$, satisfies ϵ-DP if for all subsets $S \subseteq Range(\mathcal{M})$ and for all X, X' such that $d(X, X') = 1$,*

$$\frac{\Pr(\mathcal{M}(X) \in S)}{\Pr(\mathcal{M}(X') \in S)} \leq \exp(\epsilon) \tag{C.1}$$

where $\epsilon > 0$ is the privacy loss budget and $d(X, X') = 1$ represents the possible ways that X' differs from X by one record.

Differential privacy places a bound, called ϵ, on the amount of information released by individual statistics. Differential privacy has appealing properties including sequential composition.

Theorem 1. ***Sequential Composition*** *(McSherry, 2009): Suppose a mechanism, $\mathcal{M}_j$, provides ϵ_j-DP. The sequence of $\mathcal{M}_j(X)$ applied on the same X provides $\sum_{j=1}^{J} \epsilon_j$-DP.*

Sequential composition is important because it means that the amount of information released by multiple statistics can be tracked. In other words, the privacy loss of many statistics, ϵ_j, can be summed to a global privacy loss budget, ϵ.

References

Abowd, J. M., Schmutte, I. M., Sexton, W. N., & Vilhuber, L. (2019). Why the economics profession must actively participate in the privacy protection debate. *AEA Papers and Proceedings*, *109*, 397–402.

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Encyclopedia of human-computer interaction* (pp. 445–456). Sage Publications.

Bahrampur, T., & Lang, M. J. (2021). New system to protect census data may compromise accuracy, some experts say. *The Washington Post*.

Barrientos, A. F., Reiter, J. P., Machanavajjhala, A., & Chen, Y. (2019). Differentially private significance tests for regression coefficients. *Journal of Computational and Graphical Statistics*, *28*(2), 440–453.

Barrientos, A. F., Williams, A. R., Snoke, J., & Bowen, C. M. (2021a). Differentially private methods for validation servers.

Barrientos, A. F., Williams, A. R., Snoke, J., & Bowen, C. M. (2021b). A feasibility study of differentially private summary statistics and regression analyses for administrative tax data. *arXiv preprint arXiv:2110.12055. Under Review.*

Bowen, C. M., Bryant, V. L., Burman, L., Khitatrakun, S., McClelland, R., Mucciolo, L., Pickens, M., & Williams, A. R. (2022). Synthetic individual income tax data: Promises and challenges. *National Tax Journal*, *75*(4), 767–790.

Bowen, C. M., & Liu, F. (2020). Comparative study of differentially private data synthesis methods.

Bowen, C. M., & Snoke, J. (2021). Comparative study of differentially private synthetic data algorithms from the nist pscr differential privacy synthetic data challenge. *Journal of Privacy and Confidentiality*, *11*(1).

Chaudhuri, K., Monteleoni, C., & Sarwate, A. D. (2011). Differentially private empirical risk minimization. *Journal of Machine Learning Research*, *12*(3).

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., et al. (2022). Social capital i: Measurement and associations with economic mobility. *Nature*, *608*(7921), 108–121.

Couch, S., Kazan, Z., Shi, K., Bray, A., & Groce, A. (2019). Differentially private nonparametric hypothesis testing. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 737–751.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Theory of cryptography* (pp. 265–84). Springer.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3–4), 211–407.

Fang, X., Yu, F., Yang, G., & Qu, Y. (2019). Regression analysis with differential privacy preserving. *IEEE access*, *7*, 129353–129361.

Garrido, G. M., Near, J., Muhammad, A., He, W., Matzutt, R., & Matthes, F. (2021). Do i get the privacy i need? benchmarking utility in differential privacy libraries. *arXiv preprint arXiv:2109.10789*.

Gillenwater, J., Joseph, M., & Kulesza, A. (2021). Differentially private quantiles. *International Conference on Machine Learning*, 3713–3722.

Gong, M., Pan, K., & Xie, Y. (2019). Differential privacy preservation in regression analysis based on relevance. *Knowledge-Based Systems*, *173*, 140–149.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–400.

Hay, M., Machanavajjhala, A., Miklau, G., Chen, Y., & Zhang, D. (2016). Principled evaluation of differentially private algorithms using dpbench. *Proceedings of the 2016 International Conference on Management of Data*, 139–154.

Li, Y., Coull, B. A., Krieger, N., Peterson, E., Waller, L. A., Chen, J. T., & Nethery, R. C. (2022). Impacts of census differential privacy for small-area disease mapping to monitor health inequities. *arXiv preprint arXiv:2209.04316*.

McSherry, F. D. (2009). Privacy integrated queries: An extensible platform for privacy-preserving data analysis. *Proceedings of the 2009 Association for Computing Machinery's Special Interest Group on Management of Data International Conference on Management of Data*, 19–30.

Redmiles, E. M., Acar, Y., Fahl, S., & Mazurek, M. L. (2017). A summary of survey methodology best practices for security and privacy researchers.

Ruggles, S., & Van Riper, D. (2022). The role of chance in the census bureau database reconstruction experiment. *Population Research and Policy Review*, *41*, 781–788.

Snoke, J., & Bowen, C. M. (2020). How statisticians should grapple with privacy in a changing data landscape. *Chance*, *33*(4), 6–13.

Tao, Y., McKenna, R., Hay, M., Machanavajjhala, A., & Miklau, G. (2021). Benchmarking differentially private synthetic data generation algorithms. *arXiv preprint arXiv:2112.09238*.

Taylor, S., MacDonald, G., Ueyama, K., & Bowen, C. (2021). A privacy-preserving validation server prototype.

United States Census Bureau. (2021). *Disclosure avoidance for the 2020 census: An introduction* (tech. rep.). https://www.census.gov/library/publications/2021/decennial/2020-census-disclosure-avoidance-handbook.html

Wang, H. L. (2021). For the u.s. census, keeping your data anonymous and useful is a tricky balance. *NPR*.

Wines, M. (2022). The 2020 census suggests that people live underwater. there's a reason. *The New York Times*.

Wohlrabe, K., & Rath, K. (2016). Trends in economics publications represented by jel categories between 2007 and 2013. *Applied Economic Letters*, *23*(9), 660–663.

Zhang, J., Zhang, Z., Xiao, X., Yang, Y., & Winslett, M. (2012). Functional mechanism: Regression analysis under differential privacy. *arXiv preprint arXiv:1208.0219*.