Differentially Private Methods for Validation Servers
A Feasibility Study on Administrative Tax Data

Andrés F. Barrientos
FLORIDA STATE UNIVERSITY

Aaron R. Williams
URBAN INSTITUTE

Joshua Snoke
RAND CORPORATION

Claire McKay Bowen
URBAN INSTITUTE

November 2021
ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Executive Summary</td>
<td>v</td>
</tr>
<tr>
<td>Differentially Private Methods for Validation Servers</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Background on Accessing Confidential Data</td>
<td>1</td>
</tr>
<tr>
<td>Background on Differential Privacy</td>
<td>3</td>
</tr>
<tr>
<td>Tax Data Use Cases</td>
<td>7</td>
</tr>
<tr>
<td>Tested Differentially Private Algorithms</td>
<td>8</td>
</tr>
<tr>
<td>Tabular Statistics</td>
<td>9</td>
</tr>
<tr>
<td>Quantile Statistics</td>
<td>10</td>
</tr>
<tr>
<td>Means and Confidence Intervals</td>
<td>11</td>
</tr>
<tr>
<td>Regression Analyses</td>
<td>12</td>
</tr>
<tr>
<td>Conclusion</td>
<td>16</td>
</tr>
<tr>
<td>Challenges for Accuracy and More Complex Analyses</td>
<td>16</td>
</tr>
<tr>
<td>Avenues for Future Work</td>
<td>17</td>
</tr>
<tr>
<td>Notes</td>
<td>20</td>
</tr>
<tr>
<td>References</td>
<td>21</td>
</tr>
<tr>
<td>About the Authors</td>
<td>24</td>
</tr>
<tr>
<td>Statement of Independence</td>
<td>26</td>
</tr>
</tbody>
</table>
Acknowledgments

This report was funded by the Alfred P. Sloan Foundation and National Science Foundation National Center for Science and Engineering Statistics. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

We would like to thank our collaborators at the Internal Revenue Service 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, Silke Taylor, Kyle Ueyama, Doug Wissoker, and Noah Zwiefel. Thank you to Gabriel Morrison for reviewing our code.

Finally, we thank our advisory board, who have provided invaluable advice throughout the course of this project. The members are John Abowd, Jim Cilke, Jason DeBacker, Nada Eissa, Rick Evans, Dan Feenberg, Max Ghenis, Nick Hart, Matt Jensen, Barry Johnson, Ithai Lurie, Shelly Martinez, Robert Moffitt, Amy O'Hara, Jerry Reiter, Emmanuel Saez, Wade Shen, Aleksandra Slavković, Salil Vadhan, and Lars Vilhuber.
Executive Summary

Federal tax data, derived from individuals’ and businesses’ tax and information returns, are invaluable resources for research on a range of topics. That research improves our understanding of individuals’ and firms’ responses to economic incentives. However, full access to these data is available only to select government agencies, to a very limited number of researchers working in collaboration with analysts in those agencies, or through highly selective programs within the Internal Revenue Service Statistics of Income Division. In addition, the existing process of manually vetting each statistical release for disclosure risks is labor intensive and imperfect because it relies on subjective human review.

As part of larger project to implement an automated validation server, we conduct an extensive feasibility study on several differentially private methods for releasing tabular statistics, mean and quantile statistics, and regression analyses with cross-sectional data. We provide a discussion on which methods we tested and which methods could not be implemented in practice. We then evaluate the selected differentially private methods based on their impact on tax public policy decisions and several other utility metrics. From our findings, we outline the outstanding challenges and future work.
Federal tax data, derived from individuals’ and businesses’ tax and information returns, are invaluable resources for research on a range of topics. That research improves our understanding of individuals’ and firms’ responses to economic incentives. Researchers can also use these data to study areas far removed from taxation. For example, Chetty, Friedman, and Rockoff (2014) have used tax data to study economic mobility across generations and how elementary school teacher quality affects economic outcomes later in life (Chetty, Friedman, Hilger, and colleagues 2011).

However, full access to these data is available only to select government agencies, to a very limited number of researchers working in collaboration with analysts in those agencies, or through highly selective programs within the Internal Revenue Service (IRS) Statistics of Income (SOI) Division. In addition, the existing process of manually vetting each statistical release for disclosure risks is labor intensive and imperfect because it relies on subjective human review. The tremendous demand to participate in such projects, which is limited by SOI resource constraints, indicates that much more high-quality research could be conducted if a safe and less resource-intensive method were developed to expand access.

Background on Accessing Confidential Data

At the IRS, the current process to release analytic results on confidential datasets requires the researcher to undergo an extensive background check (IRS clearance) to access the data, and then, an IRS staff member must review any results the researcher wants to release. This process reflects the norm for researchers wishing to access federal confidential data. Researchers either gain access from a public use file that is an altered version of the confidential data or have direct access to the confidential data.

As a potential middle ground between the two extremes, the US Census Bureau provides research access to two experimental synthetic databases via the Synthetic Data Server at Cornell University: the Synthetic Longitudinal Business Database and the Survey of Income and Program Participation's Synthetic Beta Data Product (Benedetto, Stinson, and Abowd 2013; Drechsler and Vilhuber 2014). The
Synthetic Data Server provides a validation server that allows researchers to submit their statistical programs to run on the underlying administrative data after testing it on the publicly available synthetic data.

However, the Synthetic Data Server has two disadvantages. First, because it is not automated, the process consumes limited staff time, which demand often exceeds. This situation causes long delays for approval. Second, reviews may be inconsistent because they are manually evaluated by humans and do not adhere to formal notions of privacy that constrain the allowable output.

To address these problems for a validation server, some privacy researchers have proposed using a newer privacy loss definition, differential privacy (DP), as a means to automate the process (Dwork, McSherry, and colleagues 2006). DP emerged from the computer science community as a rigorous definition for privacy loss associated with data publishing. Since then, many data privacy experts regard DP as the gold standard for privacy protection. It is part of a larger class of approaches often called formally private methods because statisticians can mathematically prove the privacy loss that would result from a data publication that uses differentially private methods.

DP differs from prior statistical disclosure control or limitation methods because it does not require a simulated attacker or the same strong assumptions concerning how much information an intruder may have or what kind of disclosure is likely to occur. This does not imply that DP protects from all attacks, but, for a defined type of privacy loss, it offers provable amounts of protection.

At a high level, DP links the potential for privacy loss to how much the answer of a query (such as a statistic) is changed given the absence or presence of the most extreme possible person or observation in the data population. DP requires that the level of protection is set proportionally to this maximum potential change, thereby providing formal privacy protections scaled to the worst-case scenario. For further details, Dwork and Roth (2014) provide a rigorous mathematical review of DP. Bowen and Garfinkel (2021) cover the basics of DP and its challenges for adoption geared toward a general, mathematical audience, whereas Nissim and colleagues (2017) and Snoke and Bowen (2019) describe DP for a nontechnical, general audience.

In this report, we examine the feasibility of differentially private methods for our target analyses and more complex ones. Specifically, we highlight the general findings from our extensive feasibility study on several differentially private methods for releasing tabular statistics, mean and quantile statistics, and regression analyses with cross-sectional data (Barrientos and colleagues 2021). Based on informal interviews and our tax expert collaborators, we prioritized these analyses for the first stage of the validation server. There are several other analyses, such as model selection, that have been...
identified as important but will be explored for later development stages of the validation server. Further, we outline the methodological and practical challenges for implementing differential privacy on more complex analyses, such as regression discontinuity design.

Background on Differential Privacy

Differential privacy (DP) offers a provable and quantifiable amount of privacy protection, colloquially referred to as the privacy loss budget. Those in the data privacy and confidentiality community should note that DP provides a statement about the algorithm (or mechanism), not the data—a common misconception. In other words, DP requires that the mechanism or algorithm produces an output that provably meets the definition. We refer to these methods as differentially private algorithms or mechanisms.

In this section, we reproduce the pertinent definitions and theorems of DP with the following notation: $X \in \mathbb{R}$ is the original dataset with dimension $n \times r$ and $X^*$ is the private version of $X$ with dimension $n^* \times r$. We also define a statistical query as a function $u : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^k$, where the function maps the possible datasets of $X$ to $k$ real numbers.

Definitions and Theorems

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

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

(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.

Definition 1 provides what is known as $\epsilon$-DP. There are varying understandings of what it means to differ by one record. One interpretation is the presence or absence of a record, and the other has the difference as a change, where $X$ and $X'$ have the same dimensions. Li and colleagues (2016) refers to these interpretations as *unbounded DP* for addition or removal of a record and *bounded DP* for the change of a record. They prove that unbounded DP satisfies an important composition theorem we will discuss later in this section (see Theorem 1), whereas bounded DP does not. Because many DP methods rely on Theorem 1, we assume unbounded DP in this paper.

Several relaxations of $\epsilon$-DP have been developed in order to inject less noise into the output, such as approximate DP (Dwork, Kenthapadi, and colleagues 2006), probabilistic DP (Machanavajjhala and
colleagues 2008), concentrated DP (Dwork and Rothblum 2016), Rényi differential privacy (Mironov 2017), and zero-concentrated DP (Bun and Steinke 2016). Though these definitions are still formally private, they offer slightly weaker privacy guarantees. In return, they typically lessen the amount of noise required. We will cover approximate DP, also known as $(\epsilon, \delta)$-DP, and zero-concentrated DP in depth, because most of the methods we test in our study use one of these two definitions.

**Definition 2.** $(\epsilon, \delta)$-Differential Privacy (Dwork, Kenthapadi, and colleagues 2006): A sanitization algorithm, $M$, satisfies $(\epsilon, \delta)$-DP if for all $X, X'$ that are $d(X, X') = 1$,

$$
\Pr(M(X) \in S) \leq \exp(\epsilon) \Pr(M(X') \in S) + \delta
$$

where $\delta \in [0, 1]$. $\epsilon$-DP is a special case of $(\epsilon, \delta)$-DP when $\delta = 0$.

Definition 2 provides a simple relaxation of Definition 1 by adding the parameter $\delta$. This allows, with small probability, that the strict bound given does not hold, which can be useful when dealing with extreme yet very unlikely cases.

Dwork and Rothblum 2016 proposed concentrated DP, which aimed to reduce the privacy loss over multiple computations (more on composition of multiple queries soon when discussing Theorem 1).

This definition of privacy was later on improved by Bun and Steinke 2016 who introduced zero-concentrated DP ($\rho$-zCDP), given in Definition 3. They also show in their Proposition 1.3 that if $M$ satisfies $\rho$-zCDP, then $M$ is $(\rho + 2\sqrt{\rho \log(1/\delta)}, \delta)$-DP for any $\delta > 0$. For the other direction, their Proposition 1.4 states that if $M$ satisfies $\epsilon$-DP, then $M$ satisfies $(1/2\epsilon^2)$-zCDP, which allows us to relate $\rho$-zCDP algorithms to an $\epsilon$-DP equivalent.

**Definition 3.** Zero-Concentrated Differential Privacy (Bun and Steinke 2016): A sanitization algorithm, $M$, satisfies $(\xi, \rho)$-zero-concentrated differential privacy if for all $X, X'$ that are $d(X, X') = 1$ and $\alpha \in (1, \infty)$,

$$
D_\alpha(M(X)||M(X')) \leq \xi + \rho \alpha,
$$

where $D_\alpha(M(X)||M(X'))$ is the $\alpha$-Rényi divergence between the distribution of $M(X)$ and the distribution of $M(X')$, $\xi$ and $\rho$ are positive constants, and $\alpha \in (1, \infty)$.

As mentioned before, many DP algorithms require repeated responses from a query system, such as a validation server. Each time a statistic or output is released, data information “leaks” and must be protected. DP protects the information by splitting the amount of $\epsilon$ used for each output, and the composition theorems formalize this concept.

**Theorem 1.** Composition Theorems (McSherry 2009; Dwork and Rothblum 2016; Bun and Steinke 2016):
Suppose a mechanism, $M_j$, provides $(\epsilon_j, \delta_j)$-DP or $(\xi_j, \rho_j)$-zCDP for $j = 1, \ldots, J$.

a) **Sequential Composition:** The sequence of $M_j(X)$ applied on the same $X$ provides
$$(\sum_{j=1}^J \epsilon_j, \sum_{j=1}^J \delta_j)$$-DP or $$(\sum_{j=1}^J \xi_j, \sum_{j=1}^J \rho_j)$$-zCDP.

b) **Parallel Composition:** Let $D_j$ be disjoint subsets of the input domain $D$. The sequence of $M_j(X \cap D_j)$ provides
$$(\max_{j \in \{1, \ldots, J\}} \epsilon_j, \max_{j \in \{1, \ldots, J\}} \delta_j)$$-DP or
$$(\max_{j \in \{1, \ldots, J\}} \xi_j, \max_{j \in \{1, \ldots, J\}} \rho_j)$$-zCDP.

More simply, suppose there are $J$ many statistical queries on $X$. The composition theorems state that we may allocate a portion of the overall desired level of $\epsilon$ to each statistic by sequential composition. A typical appropriation is dividing $\epsilon$ equally by $J$. For example, a data practitioner might want to query the mean and standard deviation of a variable. These two queries will require using the sequential composition, allocating an equal amount of privacy budget to each query. Conversely, parallel composition does not require splitting the budget because the noise is applied to disjoint subsets of the input domain. Privacy experts will often leverage parallel composition, for instance, to sanitize histogram counts, where the bins are disjoint subsets of the data. In this example, noise can be added to each bin independently without needing to split $\epsilon$.

The post-processing theorem is another important theorem, which states that any function applied to a DP output also satisfies DP. Many DP methods use the post-processing theorem to correct any inconsistencies or values that are not possible and to compute additional summaries required to perform statistical inference.

**Theorem 2. Post-Processing Theorem** (Dwork, McSherry, and colleagues 2006; Nissim, Raskhodnikova, and Smith 2007; Bun and Steinke 2016): If $M$ be a mechanism that satisfies $(\epsilon, \delta)$-DP or $(\xi, \rho)$-zCDP, and $g$ be any function, then $g(M(X))$ also satisfies $(\epsilon, \delta)$-DP or $(\xi, \rho)$-zCDP.

**Differentially Private Mechanisms**

In this section, we present two of the fundamental mechanisms we consider that satisfy $\epsilon$-DP and $(\epsilon, \delta)$-DP. We employ additional mechanisms that will be discussed in Section ???. For a given value of $\epsilon$ and $\delta$, an algorithm that satisfies DP or approximate DP will adjust the amount of noise added to the output based on the maximum possible change between two databases that differ by one row. This value is commonly referred to as the global sensitivity (GS), given in Definition 4.

**Definition 4. $l_1$-Global Sensitivity** (Dwork, McSherry, and colleagues 2006): For all $X, X'$ such that $d(X, X') = 1$, the global sensitivity of a function $M$ is

$$
\Delta_1(M) = \sup_{d(X, X') = 1} ||M(X) - M(X')||_1
$$

\[4\]
We can calculate sensitivity under different norms. For instance, $\Delta_2(M)$ represents the $l_2$ norm $GS$, $l_2$-GS, of the function $M$. Another way of thinking about the GS is that it measures the statistical query’s robustness to outliers. Though the definition is straightforward, calculating the GS can often be difficult. For instance, we cannot directly calculate an upper bound for the GS of one of the most common statistical analyses, regression, where the coefficients are unbounded. To address this issue, privacy researchers had to be creative in figuring out how to add noise to regression analyses.

The most basic mechanism satisfying $\epsilon$-DP is the Laplace mechanism, given in Definition 5, and was first introduced by Dwork, McSherry, and colleagues 2006. Another popular mechanism is the Gaussian mechanism that satisfies $(\epsilon, \delta)$-DP, given in Definition 6, which uses the $l_2$-GS of the statistical query.

**Definition 5. Laplace mechanism** (Dwork, McSherry, and colleagues 2006): The Laplace mechanism satisfies $\epsilon$-DP by adding noise $(\eta_1, \ldots, \eta_k)$ to $M$ that are independently drawn from a Laplace distribution with the location parameter at 0 and scale parameter of $\Delta_1(M)\epsilon^{-1}$ such that

$$M(X) = M(X) + (\eta_1, \ldots, \eta_k).$$  \hfill (5)

**Definition 6. Gaussian mechanism** (Dwork and Roth 2014): The Gaussian mechanism satisfies $(\epsilon, \delta)$-DP by adding Gaussian noise with zero mean and variance, $\sigma^2$, such that

$$M(X) = M(X) + (\eta_1, \ldots, \eta_k)$$  \hfill (6)

where $\eta_1, \ldots, \eta_k$ are independently drawn and $\sigma = \Delta_2(M)\epsilon^{-1}\sqrt{2 \log(1.25/\delta)}$.

Although Dwork and Roth 2014 proposed Gaussian mechanism for $(\epsilon, \delta)$-DP, the Gaussian mechanism satisfies $(\Delta_2(M)^2/2\sigma^2)$-zCDP, per Proposition 1.6 from Bun and Steinke 2016. Additionally, if the $l_2$-GS is 1 (which is true for all counting queries), then the Gaussian mechanism satisfies $(\alpha, \sigma^2)$-Rényi DP, per Corollary 3 from Mironov 2017. We use this relationship for multiple counting queries to reduce the amount of noise being added from the Gaussian mechanism.

Both the Laplace and Gaussian mechanisms are simple and quick to implement, but they apply only to numerical values (without additional post-processing, Theorem 2). A more general $\epsilon$-DP mechanism is the Exponential mechanism, given in Definition 7, which allows for the sampling of values from a noisy distribution rather than adding noise directly.

**Definition 7. Exponential mechanism** (McSherry and Talwar 2007): The Exponential mechanism releases values with a probability proportional to

$$\exp \left( \frac{\epsilon u(X, \theta)}{2\Delta_1(u)} \right)$$  \hfill (7)
and satisfies $\epsilon$-DP, where $u(X, \theta)$ is the score or quality function that determines the values for each possible output, $\theta$, on $X$.

Determining the Privacy Loss Parameters

The mechanisms reviewed provide levels of privacy protection determined by the value of privacy parameters, such as $\epsilon$, $\delta$, or $\rho$. For a validation server, the relative merits and interpretations for chosen values of these parameters would need to be understood and communicated to policymakers. It is also important to understand that when considering the total privacy loss budget, choosing the value of $\epsilon$ for any single query is sensitive to other factors, such as the sample size, the total number of desired queries, and the size of the population from which data are sampled (if sampled). Those questions concern the overall framework that would need to be created for a fully implemented validation server, and we do not seek to answer those questions at this stage.

We explore a wide range of $\epsilon$ and $\delta$ values for this study, ranging from ones seen in theoretical work to practical applications. By doing so, we will gain a better sense of the privacy-utility trade-off. Specifically, we will examine the affects of $\epsilon$ and $\delta$ on accuracy within the context of our study, which we hope will contribute to the conversation about how to set the privacy-loss budget.

Tax Data Use Cases

When deciding the scope of the statistical analyses the validation server will cover, we interviewed six economists who specialize in taxation, econometrics and causal inference, and other fields that use tax data. They emphasized the importance of custom tabulations for their work, which one economist referred to as “low-hanging fruit.” Many recent journal articles using tax data relied on aggregated data. Examples include the ongoing analyses of income inequality (Auten and Splinter 2018) and wealth inequality (Smith, Zidar, and Zwick 2019), the growth of gig work in the economy (Collins and colleagues 2019), and decisions on retirement savings (Brady and Bass 2020).

Other researchers have used aggregations of data in linear regressions on topics such as the effects of attitudes toward government and tax evasion (Cullen, Turner, and Washington 2021), the effect of tax credits on the labor market (Tong and Zhou 2014) and optimal taxation (Piketty, Saez, and Stantcheva 2014). Researchers have also used aggregated data to infer behavioral responses by examining bunching at kink points (Chetty, Friedman, Olsen, and colleagues 2011) and notches (Kleven and Waseem 2013) in the tax code. Mortenson and Whitten (2020), for example, investigate bunching at kink points formed from the formula for computing earned income tax credits.
With these analyses in mind, we tested the differentially private methods on the IRS SOI public use file (PUF) (Barrientos and colleagues 2021). This is a database of sampled individual income tax returns with privacy protections applied that is developed and released annually by IRS SOI. Several organizations, such as the American Enterprise Institute, the Urban-Brookings Tax Policy Center, and the National Bureau of Economic Research, develop PUF-based microsimulation models that help inform the public on potential impacts of policy proposals. But access to this public file is limited to certain institutions, and we cannot provide the full data. For this reason, we also test on the 1994 to 1996 Current Population Survey Annual Social and Economic Supplements, accessed through IPUMS USA Ruggles and colleagues 2021, a publicly available dataset. Crucially, it has similar variables as the PUF, and the case study results on the Current Population Survey will be similar to those run on the IRS SOI.

A tax expert would likely query the following as exploratory data analysis before statistical modeling, which we based our feasibility study on:

- **Single counting query**: How many tax returns have salary and wage income in excess of $100,000?
- **Mean statistic query**: What are the means for the total population and subsets of the total population?
- **Quantile statistic query**: What is the income threshold for the top 10 percent of earners?
- **Regression query**: What effects do marginal tax rates have on capital investment?

### Tested Differentially Private Algorithms

In this section, we survey differentially private methods we explored for the feasibility study and highlight the general results. We provide full details in Barrientos and colleagues (2021).

When testing our methods, we explored various values of $\epsilon$ and $\delta$. For the tabular, we set $\epsilon = \{0.01, 0.05, 0.1, 0.5, 1, 5, 10, 15, 20\}$ and $\delta = \{10^{-3}, 10^{-7}, 10^{-10}\}$. For the mean and quantile statistics, we set $\epsilon = \{0.1, 0.5, 1, 5, 10, 15, 20\}$ and $\delta = \{10^{-3}, 10^{-7}, 10^{-10}\}$. For the regression analyses, we set $\epsilon = \{0.5, 1, 5, 10, 15, 20\}$, $\delta = \{10^{-3}, 10^{-7}, 10^{-10}\}$. We based these values from what is seen in other practical applications and from our initial tests. We replicated all differentially private methods 100 times to assess variability.
Tabular Statistics

Researchers have fully developed differentially private tabular statistics. The literature already shows that adding Laplace noise produces the most accurate estimates for single tabular queries. For example, Rinott and colleagues (2018) found for disseminating frequency tables, the Laplace mechanism performed the best overall. Other DP research reports similar results because the Laplace mechanism is still "hard to beat" when the data have a large number of observations, or there are a lot of parameters and attributes to consider (Bowen, Liu, and Su 2021; Shlomo 2018; Liu 2018).

Although the Gaussian mechanism produces wider tails, theory suggests adding Gaussian noise will perform better than adding Laplace noise for multiple counting queries because of how the tails compose (Wang, Balle, and Kasiviswanathan 2019). We can improve this composition property by leveraging the Rényi differential privacy, which reduces the amount of noise added to each count. A drawback to using the Gaussian mechanism is that it satisfies $(\epsilon, \delta)$-DP, requiring the data user to balance two privacy parameters instead of one.

There are other differentially private algorithms that tackle more complex tabular statistics. However, the tax analyses we selected for the feasibility study are simple counting queries. This means we expect the Laplace and Gaussian mechanisms will meet our accuracy goals. Therefore, we do not choose to explore the more extensive differentially private tabular algorithms or apply any sophisticated postprocessing.

This brings up the question of why we tested the Laplace and Gaussian mechanisms for our study. Even though there is extensive literature, we still test the Laplace and Gaussian mechanisms for two reasons. First, we must verify these methods work well on our tax data. Second, the feasibility study can provide additional context for the privacy-utility trade-off in calculating an appropriate privacy loss budget.

In Barrientos and colleagues (2021), we confirmed that the Laplace mechanism outperformed the Gaussian mechanism, even for larger values of $\delta$. We note that both methods are unbiased, and for $\epsilon \geq 5$ they both perform well enough that there is little difference between them. Overall, these results suggest the Laplace mechanism would perform well enough to enable researchers to query accurate private histograms.
Quantile Statistics

The leading methods for generating differentially private quantiles use either the Laplace mechanism or the Exponential mechanism. For instance, Smith (2011) proposed an algorithm, \textit{IndExp}, for selecting individual quantiles using the Exponential mechanism. \textit{IndExp} has since been implemented in the SmartNoise\textsuperscript{1} and IBM DP libraries.\textsuperscript{2} This method was recently extended for querying multiple quantiles by Gillenwater, Joseph, and Kulesza (2021) to two other algorithms, \textit{AppIndExp} and \textit{JointExp}. The former is the same as \textit{IndExp} but uses the composition theorem from Dong, Durfee, and Rogers (2020) to choose an optimal $\epsilon$ for multiple queries given a total $(\epsilon, \delta)$, and the latter samples multiple quantiles jointly. \textit{JointExp} uses similar techniques to \textit{IndExp} but outputs multiple quantiles simultaneously. Using the Laplace mechanism, Nissim, Raskhodnikova, and Smith (2007) developed an approach for sampling median values using smooth sensitivity that can be extended to query any other quantile. We will hereafter refer to this method as \textit{Smooth}. A few other approaches or adaptations of the approaches listed here have been proposed, but we did not test them because they require fine-tuning based on distribution assumptions that a researcher might not realistically have in our validation server setting.

In Barrientos and colleagues (2021), we found overall that both \textit{AppIndExp} and \textit{JointExp} offer high-quality quantiles, and they were generally preferable to \textit{Smooth} for all but the highest values of $\epsilon$. \textit{JointExp} performs better at estimating the zero-valued quantiles, though \textit{AppIndExp} offers high-quality performance at most levels of $\epsilon$ and $\delta$. Similarly, \textit{JointExp} has slightly lower relative bias on average for the nonzero quantiles, but the difference between the two algorithms is very small.

For a practical validation server implementation, \textit{AppIndExp} and \textit{JointExp} both appear sufficient to return accurate quantile estimates. The choice would likely depend on whether the system deploys pure DP or approximate DP. We do not recommend that a system use \textit{Smooth} unless queries are made with very high levels of $\epsilon$. Additionally, we find that \textit{AppIndExp} returns more equally biased results for each quantile because they are drawn independently. On the other hand, \textit{JointExp} can return biased results, which are more biased for some quantiles than others because they are drawn from a noisy joint distribution.

In our application, the quantiles follow an exponential trend. \textit{JointExp} returns more accurate results for the lowest and highest quantiles but returns less accurate results for those in-between. It may be preferable to choose one or the other algorithm, depending on the application.
Means and Confidence Intervals

For mean estimates, we specifically reviewed differentially private methods that released means with their associated confidence intervals (CI). When we explored the general literature, we found common approaches use the Laplace mechanism, Gaussian mechanism, or Exponential mechanisms for releasing some means with CIs.

As with the quantiles, some methods require more information than is realistic for our application. For instance, Karwa and Vadhan (2017), Bowen and Liu (2020), D’Orazio, Honaker, and King (2015), and Biswas and colleagues (2020) require the researcher to set bounds for the standard deviation to calculate the global sensitivity to add Laplace noise. In a validation server setting, a researcher might not have a good sense of the bounds. Given this limitation, we did not test them for our study.

Du and colleagues (2020) conducted a comprehensive research study that aimed to move the theory to practice for releasing differentially private CIs. The authors developed five new methods, two based on directly applying Laplace noise named NOISYVAR and NOISYMAD and three based on querying quantiles from the Exponential mechanism to estimate the standard deviation named CENQ, SYMQ, and MOD. Du and colleagues (2020) also compared their methods against Karwa and Vadhan (2017); D’Orazio, Honaker, and King (2015); and Brawner and Honaker (2018). They found the two best performing for simulated Gaussian data were NOISYMAD and SYMQ. NOISYMAD adds Laplace noise to the mean absolute deviation and the associated standard deviation, whereas SYMQ computes two different quantiles that are equal distance away from the median and adds noise from the Exponential mechanism. A potential limitation to the approach using quantiles is that it strongly assumes the data are approximately Gaussian, since the confidence intervals are centered on the median rather than the mean. We chose not to test these methods on skewed data because preliminary tests showed extremely poor performance.

In Barrientos and colleagues (2021), the results show that the non-quantile-based methods are approximately unbiased. NOISYMAD and NOISYVAR perform similarly and both provide highly accurate statistics. BHM performs comparably to the other two methods only when $\delta = 0.01$ and when $\epsilon < 1$. Otherwise, the other two clearly outperform BHM.

For the associated confidence intervals, we find NOISYVAR offers the best performance at all but the lowest level of $\epsilon$. NOISYMAD performs well for very small values of $\epsilon$, but, as $\epsilon$ gets larger, the width of the CIs produced by NOISYMAD shrinks and are consistently more narrow than the confidential CI. These results would produce overly confident inference for researchers making this query. Overall, NOISYVAR is preferable, and it is capable of providing sufficiently accurate confidence intervals for all
Regression Analyses

In this section, we begin with an overview of the currently available differentially private approaches for regression analysis. We then explain the criteria for including methods in the feasibility study, where we discuss the selected methods and any adaptations made before inclusion in more detail.

Traditional Differentially Private Approaches for Regression Analyses

Differentially private methods for regression can be classified according to the outputs they produce: (i) point estimates only, (ii) point estimates and uncertainty estimates, and (iii) other outputs. Because we are particularly interested in methods that provide more comprehensive inferences, we focus our report on methods from category (ii). For interested readers, Barrientos and colleagues (2021) has a brief description of the other two types of methods.

For providing both point and uncertainty estimates, Sheffet (2017) developed \((\epsilon,\delta)\)-differentially private algorithms that, with certain probability, output summary statistics useful for either traditional linear regression or ridge regression. When the outputs are summaries for ridge regression, the penalization parameter is a function of the algorithm's inputs instead of being predefined by the user. Sheffet (2017) derives CIs and t-statistics that account for the noise added to the confidential summaries and shows how such statistics relate to the underlying truth. In addition, the author discusses a different algorithm that adds Gaussian noise directly to the sufficient statistics and shows that users can possibly obtain CIs for the regression coefficients under certain conditions (e.g., the norm of the true regression coefficients is upper bounded). Despite this method's promise, we do not include it in our study because of implementation barriers.

In the category of methods that add noise to the sufficient statistic for normal linear models, Sheffet (2019) proposed a \((\epsilon,\delta)\)-differentially private mechanism that draws random noise from the Wishart distribution. The mechanism defines a noisy statistic that preserves the property of being positive-definite, but only works for \(\epsilon < 1\). Wang, Kifer, and Lee (2019) also propose \((\epsilon,\delta)\)- and \(\epsilon\)-differentially private methods that release noisy versions of the sufficient statistic while preserving positive-definiteness. These methods add noise using either a normal distribution (for \((\epsilon,\delta)\)-differentially privacy) or the spherical analogue of the Laplace distribution (for \(\epsilon\)-differential privacy). The positive-definiteness is achieved by using eigenvalue decomposition and censoring the eigenvalues falling below a given threshold. Although Sheffet (2019) and Wang, Kifer, and Lee (2019) do not derive CIs or t-values under the proposed mechanisms for the normal linear model, these
contributions pave the way to develop differentially private methods that allow full inference for regression coefficients.

Ferrando, Wang, and Sheldon (2020) developed a general approach to produce point and interval estimates in different scenarios, including linear regression. The authors outline two differentially private strategies, both employing a noisy version of the sufficient statistic. The first one applies the noisy statistics in classic OLS point and interval estimators. The accuracy and coverage of this strategy is ensured by large-sample arguments. The second strategy also uses plug-in estimators, but computes CIs by means of parametric bootstrap. The method accounts for the injected noise and the underlying sampling distribution. A major drawback of these two approaches is that it is not clear how to compute points and interval estimates of the coefficients when the inverse of the noisy matrix is not positive-definite. But as the most generally applicable method for producing confidence intervals, we employ this method as part of the feasibility study.

The approach described in Bernstein and Sheldon (2019) also relies on a noisy version of the sufficient statistics. The procedure uses a Bayesian framework and employs a large-sample distributional characterization of the sufficient statistics. Using a Bayesian approach allows the authors to draw from the posterior distribution of the regression coefficient and, thus, provide point and interval estimates. Another method that draws from the posterior distribution of the regression coefficients is Wang (2018). However, under this approach, users need to spend part of the privacy budget for each draw. This particular aspect limits the applicability of Wang (2018)'s algorithm since any accurate Monte Carlo approximations would require dividing the privacy budget into too many small values. Dividing the privacy parameter too much dramatically decreases the method's statistical usefulness.

When narrowing down methods for the feasibility study, we only consider frequentist approaches. Therefore, we do not include Bernstein and Sheldon (2019) and Wang (2018)'s approaches in this iteration of the validation server. Nonetheless, we plan to consider Bayesian approaches in future versions of the validation server.

Criteria for Method Inclusion

When selecting methods to include in the study, we sought to maximize the statistical usefulness of the outputs and the feasibility of implementation. We select methods that are frequentist, can be used for linear regression models with normal errors, and can handle multiple predictors. We restrict our selection to methods that allow for full inference. For example, we exclude methods that only provide point estimates or only t-values for regression coefficients. Finally, we exclude procedures that meet the criteria above yet are hard to implement. Thus, we omits instances when:
• a manuscript lacks the information needed to implement the method it describes,
• a pseudocode is absent and implementation requires a deep understanding of the theory supporting the proposed method,
• the method has difficult-to-fix errors,
• the authors failed to reply to inquiries about their method or its implementation, or
• the method achieves differential privacy under nontestable assumptions.

Selected Methods and Adaptations

Only one method, that of Ferrando, Wang, and Sheldon (2020), met all the inclusion criteria as originally proposed. We also included Brawner and Honaker (2018)’s method, because, even though it was not originally designed to for linear regression, a small adaptation would make it eligible. We modified both Ferrando, Wang, and Sheldon (2020)’s and Brawner and Honaker (2018)’s method and obtained new versions of both methods to compare in the feasibility study.

Additionally, we used elements of the algorithm from Ferrando, Wang, and Sheldon (2020) to perform full inference with other mechanisms. Specifically, Ferrando, Wang, and Sheldon (2020)’s approach employs a parametric bootstrap to approximate the distribution of the coefficients’ estimator while accounting for both the underlying data-generating distribution and the differentially private mechanism (see algorithm 3 for the method’s implementation). Although the original method uses the Laplace mechanism to achieve differential privacy, it can also employ other mechanisms after simple adaptations. For that reason, we adapted algorithm 3 to compare the method’s performance using the Analytic Gaussian mechanism from Balle and Wang (2018), the mechanism in algorithm 2 from Sheffet (2019), and the mechanisms in algorithm 2 from Wang, Kifer, and Lee (2019) (hereafter, the Normal and Spherical Laplace mechanisms). We refer to the two methods of obtaining confidence intervals as bootstrap and asymptotic. A full discussion on each of the methods and their adaptations can be found in Barrientos and colleagues (2021).

Results

We summarized our findings from the regression experiments, focusing on the methods that may offer feasible results for a validation server. Unlike the previous analyses, the complexity of regression results makes it more difficult to assign the best performance to one method. For simplicity, we describe only results for the three best methods that have similar overall performance at $\epsilon = 5$ and $\delta = 1e - 06$. Barrientos and colleagues (2021) provide results for all values of $\epsilon$ and $\delta$.

We first examined the bias and relative bias of the point estimates. Overall, some of the replicates produced satisfactory results, but other replicates produced estimates that are very far off. This means
there is a reasonably high probability that the differentially private methods could return improper estimates. Even as $\epsilon$ increased as high as 20, the noise distribution for some coefficients remained quite wide.

We also compare the noisy confidence intervals to the original confidence intervals because we are particularly interested in performing full inference with the regression output. As mentioned earlier, we utilize two different approaches for estimating confidence intervals; bootstrap and asymptotic approaches. The former we expect to have wider confidence intervals that account for the variability introduced by the noise mechanism. The asymptotic approaches, on the other hand, produce intervals with lengths more similar to the length of the confidential estimates because they assume large sample sizes that minimize the noise from the privacy mechanisms. Because both of these uncertainty estimates can be output from the algorithms without spending additional budget, a validation server could return both "for free." We thus evaluate the different inference one might make using the different uncertainty estimates. We find that the two approaches are quite divergent. When using the bootstrap confidence intervals, the width can be as much as 50 times wider than the original CIs, suggesting the noise mechanism is introducing a significant amount of uncertainty. On the other hand, the asymptotic CIs can be badly biased and frequently do not overlap with the original CI at all.

Lastly, we look at regression inferences more simply. We consider only whether the noisy results match the sign and significance of the confidential estimates. In this case, we assume researchers are primarily interested in understanding whether a predictor has a positive or negative relationship and whether that relationship is statistically significant, and we assume they are less concerned about the exact magnitude of the estimate.

Looking at these results again for both the bootstrap and asymptotic uncertainty estimates, we see perhaps a more hopeful picture of the results for regression models, but they are still with significant limitations. Both approaches achieve relatively high sign match, which suggests that, though biased, the results will more often point in the right direction. The significance match varies quite a lot by the uncertainty method used. All regression predictors were significant in the confidential models, and we see that most of the time, the bootstrap estimates are not significant because of the very wide intervals. On the other hand, the asymptotic approaches all match significance at high levels, and they match both sign and significance for the majority of the results. The drawback for the asymptotic approach is that they have a high percentage of results which match the significance but not the sign, which could be seen as the most undesirable result. In summary, the bootstrap approach is likely to produce only null results, and the asymptotic can be seen as a high-risk, high-reward approach which will produce either correct or incorrect significant results.
Conclusion

We believe this feasibility study is the first comprehensive evaluation of various differentially private methods for practical applications within a validation server framework for real-world data (Barrientos and colleagues 2021). We found that differentially private algorithms for summary statistics performed moderately to well if the privacy loss budget is larger than 1, whereas methods for regression analyses still need improvement when it comes to performing full inference. Practical applications would likely require either larger sample sizes or allocating more $\epsilon$ on every query in order to return estimates with satisfactory levels of uncertainty. Based on conducting the study and reviewing the results, we identified a few challenges and avenues for future work.

Challenges for Accuracy and More Complex Analyses

We found that existing differentially private regression analyses were limited in applicability and often added more noise than necessary to protect privacy. For example, suppose there are four coefficients in the regression model. The tested differentially private methods add noise to the design matrix, which has 10 sufficient statistics. This means the noise is multiplicative rather than additive, resulting in estimates that were often so noisy that they would be of little practical use. Additionally, we did not test many other differentially private regression analyses because they did not meet our inclusion criteria, such as not reporting standard errors. Also, similarly to some of the mean and quantile methods, more research is needed to develop methods that are robust to other data types than Gaussian.

Besides the methodological issues, we encountered practical challenges with coding the various differentially private methods. Some manuscripts did not include enough detail or pseudocode to produce the algorithm. We sometimes discovered errors in pseudocode from a manuscript. Other times, we found bugs from code we collected from GitHub or from the author(s). Overall, these are minor issues, and we informed the author(s) of bugs.

However, code for particular applications were more problematic. Though we do not expect privacy researchers to provide production-ready code, we often discovered the research code to be messy, hard to read, and difficult to alter for our use cases. This situation is still preferable to not having any code or hand-coding the method based on the paper. We encountered this problem a few times, which prevented us from implementing some approaches. We reached out to author(s) when no code was available. If the author(s) did not respond, we attempted to code the methods ourselves. But, in some cases, the manuscript did not provide enough information for us to implement the method and was thus excluded from the feasibility study.
These issues emphasize the importance of considering practical applications when developing theory and releasing open-source code to facilitate wider use and acceptance of differentially private algorithms in practice. Recognizing this need, OpenDP from Harvard University and SmartNoise from Microsoft are developing a suite of open-source software tools to implement differentially private methods. These platforms are still under development (particularly for regression), which is why we did not use their code for the feasibility study.

Avenues for Future Work

We identified a few areas for future work based on the feasibility study results and consulting with colleagues in the field. One area for improvement is testing differentially private algorithms on data that are not Gaussian. Many of the methods we tested performed well in their respective papers, because the authors tested on well-behaved or normally distributed data. Real-world data are often skewed, such as the 2012 SOI PUF and Current Population Survey Annual Social and Economic Supplements data, which resulted in these same methods performing poorly.

Another area for improvement is developing differentially private algorithms with a focus on statistical inferences. Many data privacy experts create differentially private regression analyses to output accurate predictions, such as classification. But, for our use cases, these methods performed poorly on reporting sufficiently accurate statistical inferences or lacked reporting the standard error. This appears to be a significant gap in the data privacy and confidentiality literature that must be addressed.

With these potential research areas in mind, we will focus first on improving formal privacy-preserving methodology for ordinary least squares regression coefficient and uncertainty estimates and then expand that research to address other important economic use cases. The next steps after ordinary least squares are to develop new privacy-preserving methodologies driven by tax policy use cases and to determine how to handle survey weights. Tax experts indicated that regression discontinuity and regression kink designs are essential for much of their research using annual tax data. We are not aware of any research on formal methods to protect privacy in regression discontinuity and regression kink designs.

When developing new differentially private algorithms, we will examine both relaxed differential privacy definitions and alternative privacy definitions. For relaxed definitions, we will focus our efforts on the differentially private methods that use the $(\epsilon, \delta)$-DP and zCDP definitions, because methods that satisfied those definitions had the highest accuracy in the feasibility study. For possible alternative
privacy definitions, a promising area for further study is computing the sensitivity of an estimate based on the observed dataset instead of the possible population data. This sensitivity, known as the local sensitivity, is defined as the maximum change in a statistic from adding or removing a single record within the observed data, rather than the superset of possible datasets or population data (Nissim, Raskhodnikova, and Smith 2007). Within this framework, we can calculate the bounds of the outputs of the statistical analyses more easily than the other differentially private relaxations that rely on calculating the global sensitivity.

However, the drawback for this alternative definition is that methods under this framework are no longer formally private because we use local sensitivity instead of global sensitivity. In other words, because the definition depends on the observed data, we cannot quantify the privacy loss under the same assumptions as DP—the absolute worst-case scenario. Also, we lose the ability to automate the system more easily to prevent certain privacy attacks. For instance, validation server users might try to "game the system." This is difficult to predict. An automated system using a differentially private framework can already handle this scenario, because DP must consider all the ways a user might leak private information.

If we use an alternative definition, we would need to define a threat model to provide context for the potential privacy loss and possible ways people would try to game the system. This approach could be considered reverting back to traditional statistical disclosure control privacy definitions that tend to be ad hoc. But, many in the data user community have criticized DP (and other DP relaxations) as being too conservative in assuming the absolute worst-case scenario. Exploring these alternative definitions with a threat model could be a way to bridge traditional statistical disclosure control privacy definitions and DP.

Another area for future work is determining an appropriate privacy loss budget and how to split it among users and queries for the validation server. Deciding the total privacy loss budget and how to allocate it involves many factors that require careful consideration and research, such as determining the total number of queries. Within this research area, we could explore whether we should advise users on how to optimize their privacy loss budget, such as via the global sensitivity or the associated standard error.

We must also consider other social factors beyond limiting the total number of queries. For example, suppose user A queries a statistic, and a couple months later, user B queries the same statistic. How should the validation server handle this scenario? There are two general options: treat both queries A and B as separate queries and charge both users their respective privacy loss budgets, or use the privacy preserving result from A for B. Both have pros and cons.
The former avoids the conflict of notifying user B that user A has already conducted the analyses. Sometimes researchers do not want others to know what kind of analyses they are conducting for various reasons, such as to avoid being scooped. But, submitting the queries separately would result in user A and user B having slightly different answers to the same statistic. In addition to other educational materials, we would need to clearly and carefully inform validation server users that both answers are valid.

For the latter scenario, the users could benefit from reducing the privacy loss cost. User A and user B could split the cost evenly, preserving more of their individual privacy loss budget. Additionally, user A and user B would have the same statistical result, removing the necessity to carefully explain why the same analyses produces different results when submitted by two different users.

Additionally, we should explore the impact and outcome of a user publishing their privacy-preserving results in a peer-reviewed journal. Some journals require users to provide open-source code and data when possible. To meet these requirements, the validation server would assign an identification number to every query submitted to the system. The journal could then contact the validation server managers to verify whether the user conducted the particular analysis. We plan to partner with more tax experts to see if privacy-preserving results from our validation server would be accepted in a peer-reviewed journal.

In future stages of the validation server project, we will need to develop a privacy-preserving model selection for the regression analyses. A good model selector would ideally prevent the user submitting several queries to identify the "best" model, preserving more of a user's privacy loss budget. Another issue we must tackle soon is how to handle error messages or privacy violations that could lead to a server crash. Specifically, a user could submit queries on a subset that has too few observations or is empty. The validation server would then send an error message that would indirectly inform the user additional information, such as there are few or no observations in the subset. Given this possibility, we would also need to limit multiple queries on subsampling the data for various analyses to avoid a user trying to game the system and discover more information. A potential solution is to limit queries that only work on the synthetic data, which will likely reduce a significant number of possible error messages or server crashes. More research is needed to determine how much the synthetic data could avoid these problems on the validation server.

We hope this study provides the data privacy and confidentiality community with a better understanding the capabilities, limitations, and challenges of current differentially private methods for summary statistics and regression analyses.
Notes

1 "SmartNoise Differential Privacy Library," Microsoft, accessed July 20, 2021,


3 To learn more about OpenDP and SmartNoise, see https://opendp.org/ and https://smartnoise.org/

4 Scooped refers to an instance where user B uses the ideas of user A but publishes before user A.
References


Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek. 2021. "IPUMS USA: Version 11.0 [dataset]."


About the Authors

Andrés F. Barrientos is an assistant professor of statistics at Florida State University. His research interests focus on developing and studying methodological procedures in statistical science that incorporate flexible Bayesian modeling and approaches for data privacy. His methodological research involves single and regression density estimation, hypothesis testing, massive data, probabilistic approaches for complex data, differentially private procedures, and synthetic data generators.

Barrientos obtained his PhD in statistics at the Pontificia Universidad Catolica de Chile, in Chile. Previously, he was a postdoctoral associate in the department of statistical science at Duke University, under the mentoring of Dr. Jerry Reiter.

Aaron R. Williams is a senior data scientist in the Income and Benefits Policy Center and Program on Retirement Policy at the Urban Institute, where he works on retirement policy, microsimulation models, data privacy, and data imputation methods. He has worked on Urban's Dynamic Simulation of Income microsimulation model, the Social Security Administration's Modeling Income in the Near Term microsimulation model, and the Tax Policy Center's synthesis of individual tax records.

Williams is an adjunct professor in the McCourt School of Public Policy at Georgetown University. He holds a BS in economics from Virginia Commonwealth University, a BA in music from Virginia Commonwealth University, and an MS in mathematics and statistics from Georgetown University.

Joshua Snoke is a statistician at RAND. His primary research focuses on novel estimation procedures when data are unavailable because of privacy restrictions. These methods seek to increase researchers' ability to answer policy questions using hard to obtain data due to privacy concerns. He has expertise broadly in the field and has published papers on topics such as the release of synthetic datasets, the enabling of model estimation across multiple databases, and the creation of differentially private algorithms.

Snoke received his BS in mathematics and economics from Wheaton College. He received his PhD in statistics with a graduate minor in social data analytics from the Pennsylvania State University.

Claire McKay Bowen is the lead data scientist of privacy and data security at the Urban Institute. Her research focuses on assessing the quality of differentially private data synthesis methods and science communication. In 2021, the Committee of Presidents of Statistical Societies identified her as an emerging leader in statistics for her technical contributions and leadership to statistics and the field of
data privacy and confidentiality.

Bowen holds a BS in mathematics and physics from Idaho State University and an MS and PhD in statistics from the University of Notre Dame. After completing her PhD, she worked at Los Alamos National Laboratory, where she investigated cosmic ray effects on supercomputers.
STATEMENT OF INDEPENDENCE

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.