UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert meeting on Statistical Data Confidentiality

26–28 September 2023, Wiesbaden

The case of bounds in noisy protection methods: Selected risk and utility perspectives from official population statistics

Fabian Bach (European Commission—Eurostat, Luxembourg)[*]

fabian.bach@ec.europa.eu

Abstract

Noise-based approaches to protecting statistical confidentiality have become increasingly popular over the past decade, including for official population statistics. Many different concepts and practical methods exist meanwhile and have been studied at length. There are some generic risk/utility aspects shared by many of them, for instance the particular effects of bounding the maximum noise magnitude by a fixed value (or not). We focus on such effects of noise bounds in tabular population statistics outputs, showing on the one hand that the additional disclosure risks related to bounding noise can be controlled and on the other hand that there are important specific utility benefits of bounding noise in such outputs.

[*] The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

1 Introduction

While traditional SDC methods tend to focus on the protection of small counts only,[1] methods based on noise injection have been increasingly studied over the past two decades. There is meanwhile a rich choice of well-tested noise methods and protection setups available, with all kinds of different characteristics and often tailored to specific risk or utility priorities. Despite this rich variety, one of the rather generic qualitative features of many noise methods is whether the noise is bounded, i.e. its magnitude is limited by a fixed finite parameter, or not.

This paper addresses noise-based approaches to statistical confidentiality in official population statistics with a focus on specific risk and utility implications stemming from the presence—or absence—of a noise bound. A further focus on census-like statistics is chosen because of the global relevance (2020/2021 census round), and because unweighted counts simplify technical discussions without major loss of generality in the key issues.

There are two classes of such methods, each representing one of the two scenarios of bounded vs. unbounded noise, that have indeed been used in census production for the global 2020 round: On the one hand, the U.S. Census Bureau has adopted a strictly differentially private[2] noise method for the 2020 U.S. census (Abowd, 2018; Abowd et al., 2022), which received mixed reactions down to grave utility concerns (Ruggles et al., 2019; Santos-Lozada et al., 2020) and ensuing debates (Muralidhar and Domingo-Ferrer, 2023). It is a characteristic feature of *strictly* differentially private methods that the underlying noise must be unbounded (Dwork et al., 2006). On the other hand, the European Statistical System[3] has developed recommendations for a harmonised protection of 2021 EU census outputs (Antal et al., 2017; De Wolf et al., 2019a,b) based on the cell key method (Fraser and Wooton, 2005; Marley and Leaver, 2011; Thompson et al., 2013). This method provides a dedicated parameter to control the noise bound explicitly. For the quantitative analysis of specific risk and utility aspects related to noise bounds, these two methods are employed in this paper as generic representatives of their respective classes (with or without noise bound).

2 Key concepts and terms used

Differential privacy (DP) was initially proposed by Dwork et al. (2006) as a rigorous privacy or risk measure. The concept is appealing from a risk-aware view because it gives a DP guarantee to each individual contributor of a given statistic; cf. annex A.1. This can come as a strict ε-DP guarantee with a single privacy budget measure ε, and as a relaxed (ε, δ)-DP guarantee with a second measure δ quantifying the potential leakage from a strict guarantee. Various noise protection methods were proposed specifically to implement a certain (strict or relaxed) DP guarantee by construction; see e.g. Rinott et al. (2018).

Noise distributions are probability distributions over the range of the statistical outputs of interest, e.g. non-negative integers in population counts. The noise distribution is used to draw a dedicated random noise term x to be added to each statistical value in the output. Typically, risk and utility considerations influence the detailed design shape of the distribution, but many broad aspects can be studied rather generically based on just two parameters: noise variance V[4] and bound E (see next paragraph). Examples in annex A.2 include manifestly ε-DP distributions and those used by the cell key method (Marley and Leaver, 2011).

Bounded noise comes from a noise distribution with a parameter $E > 0$ such that $\Pr(|x| > E) \equiv 0$, i.e. limiting the magnitude of any noise term x. Note importantly that strict ε-DP, in contrast to (ε, δ)-DP, does not allow

[1] E.g. suppression or rounding of small counts, topcoding or general recoding of rare attributes.

[2] See section 2 and annex A.1 for a short outline of differential privacy.

[3] The joint body of Eurostat and the national statistical institutes of all EU countries and Iceland, Liechtenstein, Norway and Switzerland. It is responsible for the development and quality assurance of official European statistics.

[4] With a conservative assumption that the distribution is reasonably centred, as is the case with the Laplace, Gaussian and derived discrete distributions used in this paper.

$E < \infty$ (see annex A.2). A key goal of this paper is to quantify specific utility flaws of *unbounded* noise (section 3), but also additional disclosure risks of *bounded* noise (section 4).

3 Specific utility flaws of *unbounded* noise

This section concentrates on generic tail effects of unbounded noise distributions, using the vanilla ε-DP two-tailed geometric distribution of Eq. (11) (annex A.2) as a generic toy method. There are already many studies assessing utility aspects of DP methods or testing them in statistical applications—e.g. Machanavajjhala et al. (2008); Dwork and Smith (2010); Ghosh et al. (2012); Hsu et al. (2014); Wang et al. (2015); Petti and Flaxman (2019). In particular, Rinott et al. (2018) is a key reference for population statistics, but all DP noise distributions there were truncated (i.e. bounded and thus 'just' (ε, δ)-DP), so results do not cover tail effects from unbounded noise. On the other hand, the U.S. Census Bureau used unbounded ε-DP noise for its 2020 census (Abowd, 2018), which triggered severe utility concerns (Ruggles et al., 2019; Santos-Lozada et al., 2020). Petti and Flaxman (2019) assessed some utility implications of published test setups, but explicitly left the issue of tail effects open.

3.1 Parameter setup

Aiming for a realistic setup in a census context, we try to guess the incremental ε budget spent on a single output table in the hypothetical U.S. census DP scenario described in Petti and Flaxman (2019). There, discrete ε-DP noise is drawn from the two-tailed geometric distribution with a *global* privacy budget $\varepsilon_{\text{global}} \in \{0.25, 0.5, 1.0, 2.0, 4.0, 8.0\}$ (Garfinkel, 2019; Petti and Flaxman, 2019). This global budget is then distributed across six hierarchical geographies (Garfinkel, 2019). Certain optimisations may shift the relative shares away from an even split, but we assume 1/6 for practical purposes as Petti and Flaxman (2019) do. Further intricacies include that noisy total population counts are generated for each geographic level[5] and all further breakdowns are optimised to sum to those totals. The reference also suggests that at each geographic level, 67.5 % of the budget are spent on the more important person aggregate tables. In summary, we assume

$$\varepsilon_{\text{table}} = 67.5\,\% \times 1/6 \times \varepsilon_{\text{global}} \simeq 10\,\% \times \varepsilon_{\text{global}}, \tag{1}$$

so $\varepsilon_{\text{table}} \in \{0.025, 0.05, 0.1, 0.2, 0.4, 0.8\}$ for tabular (count-level) ε-DP noise. This corresponds to noise sizes, in terms of noise variance V, at single count level of

$$V \in \{3200, 800, 200, 50, 12.5, 3.125\}, \qquad \sqrt{V} \in \{56.6, 28.3, 14.1, 7.1, 3.5, 1.8\}.$$

For comparison, the CK variances tested for the 2021 EU census round are in the range $V \in [1, 5]$ (Antal et al., 2017), so barely touching the above DP range at its risky end ($\varepsilon_{\text{table}} \gtrsim 0.4$). Moreover, no tails effects $> E$ are present by definition.

3.2 Demographics at high geographic detail

Accurate demographics at a high geographic detail is one of the key unique census features in many world regions. For instance, the 2021 EU census round will cover ca. 110 000 local administrative units (LAUs) with a total population of roughly 4.5×10^8 people across the whole EU.[6] Coincidentally this matches well with U.S. census outputs at tract level, covering ca. 75 000 geographic units (Garfinkel, 2019) with a total population of 3.3×10^8 people. However, the following analysis is intended solely to discuss effects of a generic unbounded noise scenario on key EU census outputs. Whether any of the conclusions may apply to tract-level U.S. census

[5] Except at State level, where the U.S. Constitution requires the U.S. Census Bureau to publish unperturbed totals (Petti and Flaxman, 2019).

[6] The LAU data used for this section are 2011 census outputs from all EU Member States as available at ec.europa.eu/CensusHub2.

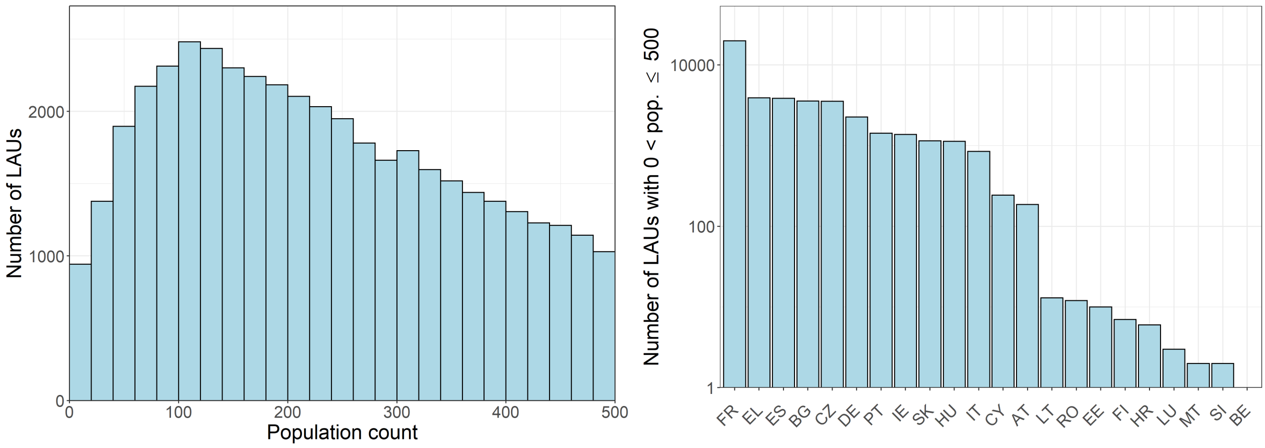

FIGURE 1. Distribution of populated LAUs with ≤ 500 residents across the total population count (left) and across EU countries (right).

outputs depends critically on the correctness of parameter assumptions, Eq. (1), and also on the comparability of population distributions across EU LAUs vs. U.S. tracts.

The statistics of LAUs There is an extreme variety of total population by LAU, with populated units ranging from $O(1)$ residents (450 LAUs with < 10 people) to 3.3×10^6 residents (Berlin; in total 14 LAUs with $> 10^6$ people). Now the key point is that statistics *across* LAUs is only part of the purpose of these census results; they are also the only source to obtain accurate demographic information on *individual* LAUs. For this purpose, even very unlikely but very large noise outliers can have severe, maybe unacceptable, consequences. Furthermore, if the method of adjusting inner tables to their geographic totals after drawing noise is applied (Petti and Flaxman, 2019), a single large noise outlier on a given small LAU total would systematically and heavily distort all statistics published for that LAU. Therefore, the subsequent focus is on LAUs with counts < 500 illustrated in Fig. 1.

The demographics of LAUs To add a demographic element, we include a sex breakdown into females, males and a total, i.e. SEX $= \{F, M, T\}$ as in section 4. This is the spine of all LAU-level person tables in table groups 3 and 8 of the 2021 EU census programme[7]. It also reflects a possible notion of picking more important ‘aggregate tables’ to which all further breakdowns would then be adjusted (Petti and Flaxman, 2019). To cover both large distortions of totals as well as of sex balances, the counts of F, M and T are treated independently. In total, there are ~ 167 000 LAU counts of F, M or $T < 500$ in the 2011 data.

Estimating distortions In the ε range of Eq. (1), the discrete two-tailed geometric distribution used already converges well to the continuous $\text{Lap}(1/\varepsilon)$. So the cumulative inverse distribution function of $\text{Lap}(1/\varepsilon)$ can be used to estimate the probability for the noise magnitude $|x|$ to exceed a certain threshold E:

$$\Pr(|x| > E|\varepsilon) = \exp\left(-\varepsilon E\right) . \tag{2}$$

This probability is plotted in the lower-right of Fig. 2 as a function of ε inside the relevant range, and for $E \in \{20, 50, 100\}$. Now Eq. (2) can be convoluted with the distribution of LAU counts (left plot in Fig. 1) to estimate how many LAU counts in each bin will end up with noise exceeding a given absolute relative error RE = 20, 50 or 100 %. These binned estimates can be tested by sampling some noise on the LAU data, and counting occurrences of RE magnitudes above a given threshold. Fig. 2 (left column) overlays the estimates with counts found in the noise-sampled data. Clearly the analytic estimates describe very well the sampled noise data.

[7] Commission Regulation (EU) 2017/712 of 20 April 2017 establishing the reference year and the programme of the statistical data and metadata for population and housing censuses provided for by Regulation (EC) No 763/2008 of the European Parliament and of the Council (OJ L 105, 21.4.2017, p. 1).

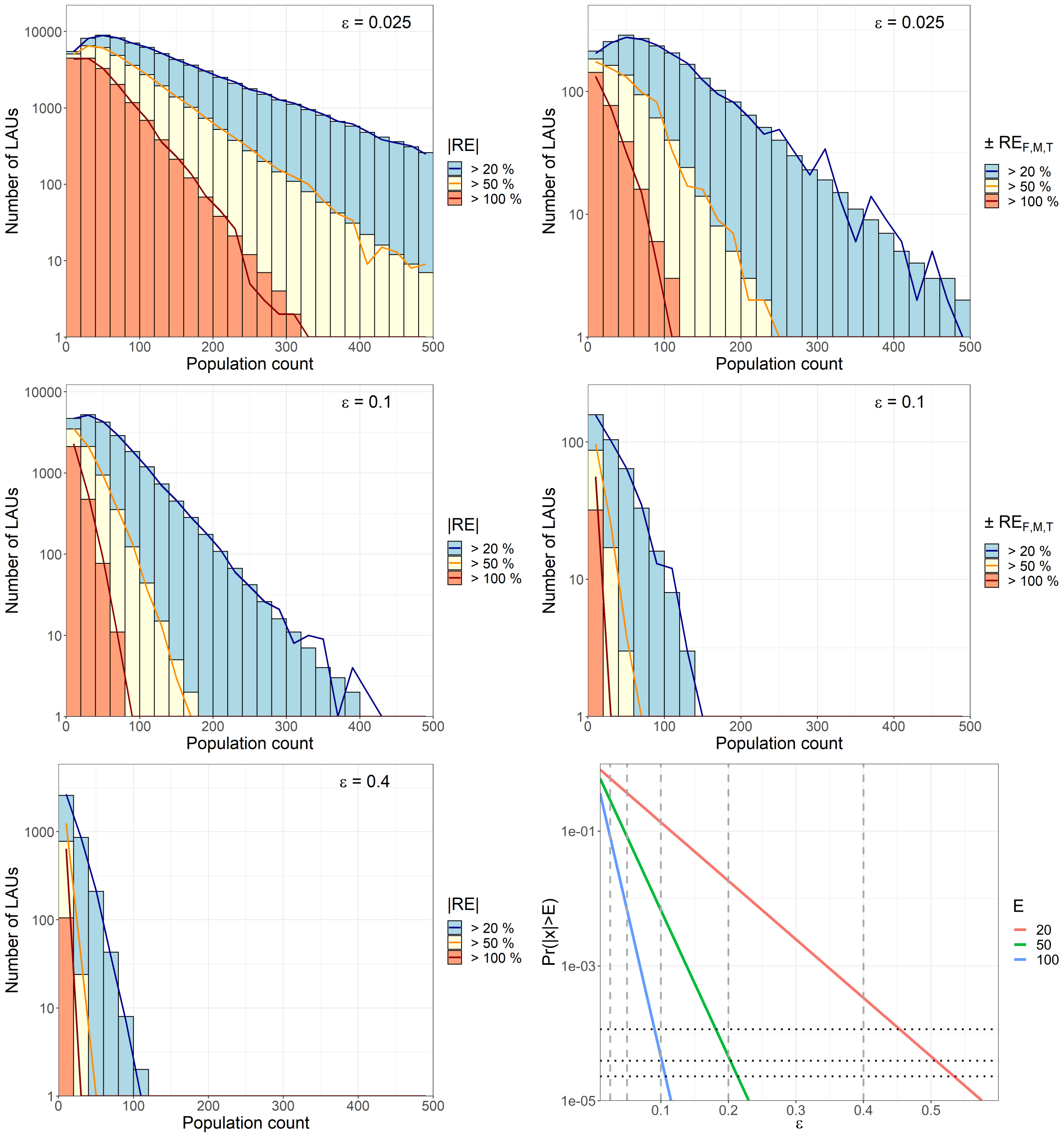

FIGURE 2. Log-linear estimates for frequencies of relative error (RE) magnitudes exceeding 20 % (blue), 50 % (yellow) and 100 % (orange) occurring in LAU counts, by total count: bins show the analytic estimate obtained from Eq. (2), while lines show the actual distortion frequencies found in the data with noise sampled.

The rows vary $\varepsilon = 0.025$ (top) to 0.1 (middle) to 0.4 (bottom). The left column counts single observations (F, M or T) exceeding a given RE, while the right column counts LAUs where F, M and T all exceed RE in the same direction.

The lower right histogram (F, M and T distorted in the same direction for $\varepsilon = 0.4$) is almost empty and thus replaced by a plot illustrating Eq. (2): log-linear $\Pr(|x| > E)$ as a function of ε with $E = 20$ (orange), 50 (yellow) and 100 (blue). Vertical dashed lines indicate ε choices from Eq. (1), while horizontal dotted lines show 1 over the number of LAUs with $T \le E = 20$, 50 or 100.

Distortions of single counts Looking now at the actual distortions in the left column of Fig. 2, one finds a sizeable dependence on ε, which is not surprising due to the exponential scaling in Eq. (2). In fact, noise distortions of single counts in these LAU statistics may be said to become manageable from $\varepsilon > 0.4$ (and we do not show the upper end of the ε range, $\varepsilon = 0.8$, for this reason). However, for $\varepsilon \lesssim 0.1$ there are many LAU counts expected with $|\mathrm{RE}| > 50\,\%$ or even $> 100\,\%$.

For instance, with $\varepsilon = 0.025$ ($\sqrt{V_{\text{table}}} = 56.6$) there are 1 648 observations above 100 affected by ±100 % or more, and still 87 observations above 200 with RE ± 100 % or more. Recall that every third of these observations describes a total count, and every 6th a total count with RE $< -100\,\%$, thus wiping out the whole population of that LAU. The largest LAU where this happens is Aragnouet, France, with originally 239 residents (now -7[8]). The situation does improve with $\varepsilon = 0.1$ ($\sqrt{V_{\text{table}}} = 14.1$), but we still find 122 observations above 40 and 11 observations above 60 with RE ± 100 % or more. The largest depopulated LAU is again in France, Mélagues with originally 63 residents (now -9).

Distortions of entire LAUs The findings on single counts are disconcerting in their own right, but there is an added danger: if the total count is distorted so severely and inner table cells are adjusted to the new total[9], entire LAU populations may disappear from the census output. If inner cells are not adjusted, constraints like $F + M = T$ can be exploited to improve knowledge a bit; e.g. $\widehat{T} = (F + M + T)/2$.

However, such ad hoc 'repair' estimates exploiting 3-tuple constraints will not always help. This is the case when F, M and T are all distorted in the *same* direction ("broadband distortions"), so the distorted 3-tuple is internally consistent and no ad hoc estimate can improve the user's knowledge. To quantify this, one can count all LAUs affected by such broadband distortions; results are shown in the right column of Fig. 2. For $\varepsilon = 0.025$ there are 28 LAUs above 40 residents and 4 LAUs above 80 with a broadband distortion $-100\,\%$ or more. The largest such LAU is Landremont, France with $F = 61 \longrightarrow -8$, $M = 74 \longrightarrow -26$ and $T = 135 \longrightarrow -83$. For $\varepsilon = 0.1$, most broadband distortions of ±100 % only occur in the lowest count bin $(0, 20]$, but there is one above: Cidamón, Spain with $F = 15 \longrightarrow -9$, $M = 20 \longrightarrow -1$ and $T = 30 \longrightarrow -17$. Broadband distortions ±20 % still occur for 61 LAUs with 100 or more residents. The largest LAU where this happens is Ellend, Hungary with $F = 112 \longrightarrow 74$, $M = 94 \longrightarrow 65$ and $T = 206 \longrightarrow 158$. Even distortions around ±20 % may have significant policy impacts at local level.

3.3 Population shares at high geographic detail

Going beyond simple population counts provides further insights into unbounded noise effects. For example, we consider one of the simplest derived indicators within the setting of this section: the share of females $r := F/T$ in any given geographic unit (LAU here).[10] For the unbounded noise, we choose the ε-DP setup with tabular $\varepsilon = 0.8$ of section 3.1, and for the bounded noise a CK setup with $V = 3.125$ (corresponding to $\varepsilon = 0.8$, cf. section 3.1) and $E = 6$—a conservative choice for the given V within the EU census scenario, according to Fig. 4.

A first question is how large the typical noise-induced r variations are. This is given by the standard deviation of r derived from the propagation of $\mathrm{Var}(F) = \mathrm{Var}(T) = V$ for both noise setups:

$$\mathrm{sd}_r\,(V) = \frac{1}{T}\sqrt{V\left(1 + r^2\right)}. \tag{3}$$

[8] Negative output counts are a typical consequence of standard DP noise. These may be lifted to 0, as proposed e.g. by Ghosh et al. (2012). However, this generally introduces a (normally small) overall bias to the output and may have other negative impacts on output utility, pointed out by Rinott et al. (2018). In any case, the discussion is not relevant here: all negative counts mentioned in this section can be replaced by 0 without changing any conclusion.

[9] I.e. in this example, noise on T would be fixed but noise on F and M would be post-processed to minimise the violation of the 3-tuple constraint $F + M = T$.

[10] All following findings on r transcend to any share indicator, and even to more complex ones like the index of dissimilarity, with the sole complication that other shares, such as minority shares, are typically much less centred around 50 %. This is relevant for the 2021 EU census outputs, which will provide migrant background variables by sex at LAU level (table group 8 in footnote 7).

The left-hand side of Fig. 3 shows this propagation model for sd_r as a function of T, with $V = 3.125$ as introduced and $r = 0.5$ fixed. The latter is a reasonable approximation as the factual share of females in the population is strongly centred around 50 % across all LAUs. In the plot, the model is overlaid with mean standard deviations computed from the noise samples in bins of width 20, showing that it is indeed a very good approximation of the properties of both noise setups.

Now note the $\mathrm{sd}_r \sim 1/T$ dependence for fixed V: while the typical noise variation of r drops below 1 % point above $T = 200$, it is in the 10 % points region for T around 20. Since these are absolute percentage points, this can mean *relative* variations of e.g. minority shares ($r \ll 0.5$) of $\mathrm{sd}_r/r = 100\,\%$ and more for small populations. Moreover, a sizeable share of LAUs will have an r variation $> \mathrm{sd}_r$; somewhat depending on the exact noise distribution, we find here roughly 28 % (DP setup) to 35 % (CK setup) of LAUs independent of T. So for instance, about every third LAU with $T \simeq 100$ will have an absolute r deviation > 2 % points (for $V = 3.125$). In fact, these are very generic insights irrespective of any noise: the total variance of any statistic is generally composed of intrinsic contributions (e.g. measurement or statistical uncertainties) and extrinsic ones such as noise injection, where the described effects scale with the total resulting variance V. The tangible utility argument to accept noisy protection is that its *added* variation is limited or negligible compared to the intrinsic components.

Now moving to tail effects, we start by approximating the r variation to leading order in the noise terms: using $i = i_0 + x_i$ with $i \in \{F, T\}$ and noise terms x_i, one finds

$$r - r_0 = r\,(\xi_F - \xi_T) + O\left(\xi^2\right) \quad \text{with} \quad \xi_i \equiv x_i/i \ll 1\,, \tag{4}$$

where $r_0 = F_0/T_0$. In the CK setup, the maximum absolute variation is bounded by $|x_i| \leq E$ (= 6 here) and thus from Eq. (4)

$$\max|r - r_0| \overset{CK}{\simeq} \frac{E}{T}\,(1 + r)\,, \tag{5}$$

whereas the unbounded noise from the DP setup does not respect such an upper bound. The right-hand side of Fig. 3 shows the CK limit model of Eq. (5) as a function of T, again for $V = 3.125$ and $r = 0.5$. Overlaid are the largest $|r - r_0|$ values found for the DP and CK noise setups in each bin of width 20. This shows first that Eq. (5) indeed describes a tight upper bound on the CK noise variations (the blue bin centres are always below the grey line); and second, that the largest variations from unbounded noise—for same V—are typically significantly larger (the black line is always above the blue one, and often above the grey one). These differences can be sizeable (note the log scales): e.g. in the $T = [220, 240)$ bin, we find $\max|r - r_0| = 5.3$ % points in the DP noise, but only 2.9 % points in the CK noise, with Eq. (5) setting a tight CK limit $\lesssim 4.1$ % points. Again, these are variations in absolute percentage points, so differences in relative variations between bounded and unbounded noise setups can be huge for small shares $r < 0.1$, e.g. minority groups.

3.4 Discussion

The two simple analyses above have shown that the tails of unbounded noise distributions, such as strictly ε-DP ones, may have grave effects at small geographies. For absolute population counts (section 3.2), this starts to kick in severely around count-level $\varepsilon_{\text{table}} < 0.4$ ($V > 12.5$) for most countries ($> O(10^3)$ LAUs), whereas effects on population shares (section 3.3) such as minority groups can be sizeable already at $\varepsilon_{\text{table}} = 0.8$ ($V = 3.125$). These results point at similar conclusions as in Santos-Lozada et al. (2020): with unbounded noise it is very difficult to maintain a certain minimum utility per individual small area unit, for *every* small area unit in the output.

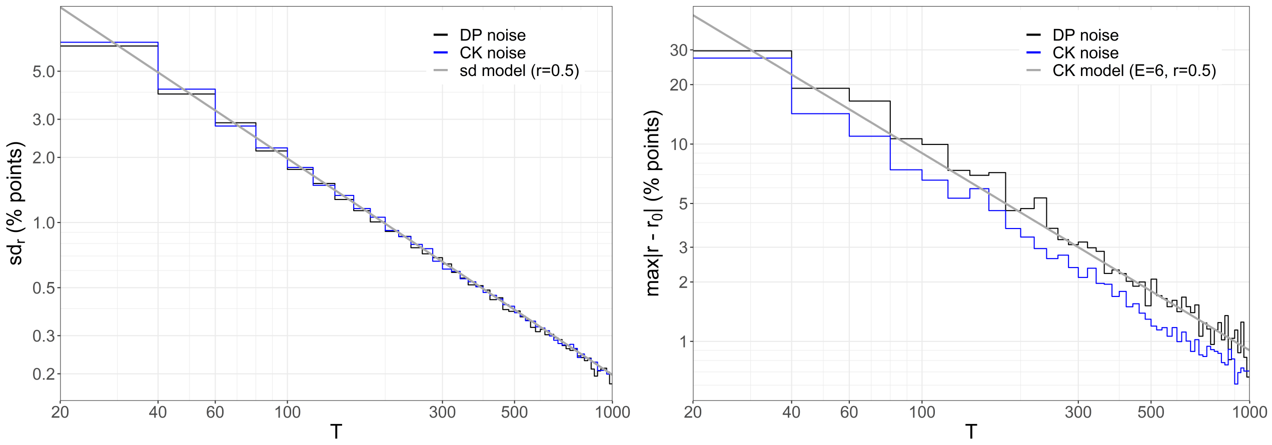

FIGURE 3. The left-hand side shows a log-log plot of the standard deviation sd_r of the share r over the total population T. The binned distributions represent the mean sd_r from the DP (black) and CK (blue) noise samples, while the grey line shows the sd_r model of Eq. (3). The right-hand side shows (same colour code) a log-log plot of the largest absolute r variations found in each T bin, and overlaid the CK limit model from Eq. (5).

4 Specific additional disclosure risks of *bounded* noise

While unbounded noise has specific utility flaws, as shown in section 3, bounded noise is known to come with certain additional disclosure risks stemming directly and rather generically from the noise bound; see e.g. Gießing (2016). Asghar and Kaafar (2020) have elaborated a typical generic attack along this line that exploits a fixed noise bound. This attack exploits extreme noise patterns in constrained n-tuples with generic bounded noise protection. Normally the noise bound E should be non-public, so the first step is to find the value of E.

4.1 Revealing the bound

The attack of Asghar and Kaafar (2020) relies on m output 3-tuples of noisy observations with independent noise but respecting a linear constraint. The type of 3-tuples is not important, e.g. a sex breakdown including total count $\{F, M, T\}$ with expectation $\mathrm{E}(F + M - T) = 0$ so that $F + M - T$ values are sampling the noise distribution. This gives an estimator for the noise bound

$$\widehat{E} = \left\lceil \left| \frac{F + M - T}{3} \right| \right\rceil, \tag{6}$$

where the probability of revealing E correctly from a single 3-tuple is fixed by the noise distribution as $p_1 := \Pr[|F + M - T| > 3(E - 1)]$.[11] Given p_1, the number of independent 3-tuples needed to infer E at confidence level α is

$$m = \left\lceil \frac{\log(1 - \alpha)}{\log(1 - p_1)} \right\rceil \simeq \left\lceil \frac{1}{p_1} \right\rceil \quad \text{for} \quad \alpha = 68\,\% \text{ and } p_1 \ll 1. \tag{7}$$

Results of Asghar and Kaafar (2020) are for uniform noise only, but in general m will depend heavily on p_1 and thus on the particular noise distribution. For instance, in CK-like methods p_1 is fixed by the p-table (Thompson et al., 2013) and thus by V and E parameters, which allows to control the required complexity m. Fig. 4

[11] E.g. uniform noise $\in \{-E, E\}$ gives $p_1 = 20/(2E + 1)^3$ by simple combinatorics (Asghar and Kaafar, 2020).

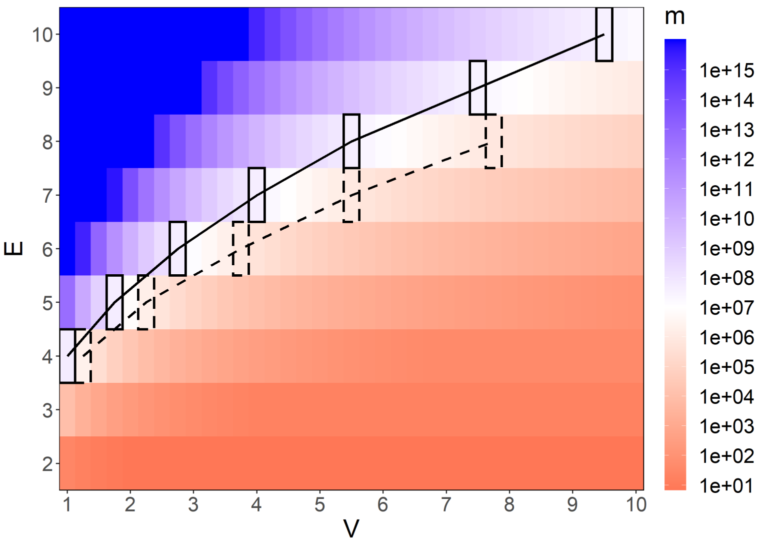

FIGURE 4. Heat map showing the number m of 3-tuples required to infer E at confidence level $\alpha = 68\ \%$ over the V–E parameter space of CK-like methods (and $p_1 \simeq 1/m$, cf. Eq. (7)). Black boxes highlight the parameter settings where m exceeds the number of independent 3-tuples (i.e. sex breakdowns) available in the 2021 EU census output of Germany (solid) and Malta (dashed).

illustrates m over the typical V–E space in a generic CK setup using the p-table tool recommended for the 2021 EU census (De Wolf et al., 2019b).[12]
Note that m converges to the uniform limit for increasing $V > E$ (because the p-table converges to the uniform distribution with maximum variance $V = E(E+1)/3$), but diverges quickly for decreasing $V < E$ (because large noise magnitudes become increasingly unlikely). This suggests that CK setups with moderately large $E \lesssim 10$ and considerably smaller V (e.g. $E = 5$ to 10 and $V = 2$) perform as "quasi-unbounded" noise on attempts to disclose E. In conclusion, Asghar and Kaafar (2020) have argued that E cannot be sufficiently protected, but it was shown above that this depends critically on the noise distribution and relative choice of V and E: while uniform noise seems E-disclosive, generic bounded noise distributions can be set up to protect E effectively while keeping strong utility guarantees (moderate variance and hard noise bound).

4.2 Exploiting margins

Nevertheless, assume now E is known to complete the discussion. Then one can search the whole output for constraint n-tuples with extreme noise combinations, which can only be obtained by a single noise pattern applying $\pm E$ to each count. In such a case, all true counts of the n-tuple are disclosed: e.g. find $F = 3$, $M = 2$ and $T = 11$ with $E = 2$ known, which discloses true $F = 5$, $M = 4$ and $T = 9$.[13] However, the abundance of such noise combinations in the output depends again on p_1 described above (or its generalisation for > 2 categories), and thus becomes increasingly unlikely in the "quasi-unbounded" regime ($V \ll E \lesssim 10$): Bailie and Chien (2019) estimate such risks in a typical scenario as $O(10^{-3} - 10^{-16})$, but assuming $E = 2$ fixed, while similar risk scaling with $E = 1$, 2 or 5 (for fixed V) is observed by Enderle et al. (2018). Fig. 4 also shows

[12] The setup is 'generic' because we use the implemented generic p-table generating algorithm that maximises entropy under the sole constraints of fixed V and E (cf. Gießing (2016)). If p-tables are further tailored to specific needs, e.g. adding more constraints, this may affect p_1 and thus m.

[13] There are more disclosive patterns when true 0s are not perturbed (Enderle et al., 2020), but these require very specific true count patterns combined with a specific noise pattern drawn; such patterns become very unlikely for moderately large E, as suggested by Enderle et al. (2020).

$p_1 \simeq 1/m$ for a two-category variable (most disclosive) as a function of E and V. For instance, with $E = 5$ and $V = 2$ each output 3-tuple has just an individual chance $p_1 \simeq O(10^{-7})$ of being divulged, and for $E = 10$ and $V \leq 4$ it is practically zero.[14] Note finally that such an attack cannot be "aimed" at specific statistics of interest; it is limited to wherever extreme noise patterns happen to occur.

4.3 Heuristic parameter constraints

To generalise this, a heuristic risk constraint can be inferred on the V–E parameter space: to avoid E-disclosure, choose V and E for fixed m—i.e. a static output scenario[15]—such that the E disclosure risk according to Eq. (7) is below 68 %, even when all available 3-tuples are used. The respective contours are added to Fig. 4 for Germany (most independent sex 3-tuples) and for Malta (fewest independent sex 3-tuples). In such a static output scenario, such a limit can always be set by requiring the noise distribution to satisfy $p_1 \lesssim 1/m$. Note that this constraint is very conservative: even if E is disclosed correctly, there are then only $mp_1 = O(1)$ 3-tuples in the whole output where the noise can be removed.

5 Conclusions

While traditional SDC methods in population statistics mainly focused on small counts at high risk of direct re-identification (e.g. suppressing those, etc.), often some generic features of such noise methods can be captured well with just two parameters, namely the variance V and noise magnitude bound E. Whether the noise is bounded ($E < \infty$) or unbounded ($E = \infty$) is a key question, as it has specific consequences for both utility and risk of the respective noise method.

On the one hand, the two simple analyses of section 3 have shown that the tails of unbounded noise may have grave effects on population statistics at small geographies, such as census results at the level of municipalities. Large tail effects may be rare, but not rare enough in typical noise setups to guarantee that the resulting statistics are still useful for *all* small areas. In particular, share indicators for minority groups may become severely affected already at very modest $V \sim O(1)$. On the other hand, section 4 has illustrated that specific additional disclosure risks of noise bounds can typically be controlled systematically by tuning V and E. In conclusion, unbounded noise seems not fit for population statistics where a certain minimum accuracy for *every* output count is a design requirement.

Differential privacy (DP) is a useful concept to *quantify risk* irrespective of a particular protection scenario, and hence to compare risk levels consistently between various SDC approaches. However, the flexibility of strictly ε-DP *protection methods* is heavily limited with just a single parameter (the privacy budget ε). It is the presence of a second parameter—generically the noise bound E, or δ in relaxed (ε, δ)-DP methods—that adds the flexibility needed to arbitrate risk vs. utility efficiently.

[14] E.g. the number of independent sex breakdowns in the 2021 EU census output depends on the geographic breakdowns and hence on country size, giving the largest $m = 2.8 \times 10^7$ for Germany.

[15] Meaning here that the total amount of output statistics is fixed, e.g. as a set of contingency tables. This typically gives more control over the remaining disclosure risks than a flexible or dynamic output scenario, where the user may query a less limited amount of statistics.

Appendix A Extended discussion of concepts and terms used

A.1 Differential privacy: a risk measure

Differential privacy was first proposed by Dwork et al. (2006), in the wake of the database reconstruction theorem. In plain words, its paradigm is that every query result (output) should be robust against addition to, or removal from, the input database of any single record, e.g. picking one record and removing it from the database should not significantly change any outputs (hence *differential* privacy).

There are various mathematical definitions of differential privacy, so we repeat here the most generic one, introducing both strict as well as relaxed (or approximate) differential privacy in one go as Dwork et al. (2006): given two neighbouring input databases d and d' that differ exactly in one record, any mechanism $\mathcal{M}(\cdot)$ acting on the universe of input databases to generate outputs must fulfil

$$\Pr(\mathcal{M}(d) \in S) \leq e^{\varepsilon} \Pr(\mathcal{M}(d') \in S) + \delta \tag{8}$$

for all subsets $S \subseteq \mathrm{Range}(\mathcal{M})$ to be *δ-approximately ε-differentially private* or short (ε, δ)-DP, where ε and δ are parameters establishing the differential privacy level. For $\delta \to 0$, Eq. (8) reduces to a definition of *strictly ε-differentially private* or short ε-DP mechanisms.

The definition implies that, for any single output $s \in \mathrm{Range}(\mathcal{M})$—singleton S in Eq. (8)—with nonzero probability on d, the probability to obtain s from d' should also be nonzero for the mechanism to be possibly ε-DP. This suggests some kind of noise injection applied by $\mathcal{M}$ as an option to comply with Eq. (8). It is important to note here that ε-DP or (ε, δ)-DP are attributes or qualifiers of any given $\mathcal{M}$, thus measuring the individual information leakage from any thinkable output. Therefore, ε and δ are handy risk measures to compare different output scenarios and noise mechanisms, as done e.g. by Rinott et al. (2018).

A.2 Noise distributions: bounded or unbounded?

Recall that the discussion is confined to outputs representing unweighted person counts, or sets of such counts (e.g. contingency tables). Then the most generic output mechanism $\mathcal{M}(\cdot)$, in the sense of annex A.1 , returns an ordered k-tuple of frequencies representing the answers to k individual counting queries passed to $\mathcal{M}$. Further let $\widetilde{\mathcal{M}}(\cdot)$ denote an *exact* output mechanism without any noise injected, so that $\mathrm{Range}(\widetilde{\mathcal{M}}) = \mathbb{N}_0^k$. Then by noise distribution we mean the probability distribution underlying the process of drawing an additive (pseudo) random noise term $x \equiv (\mathcal{M} - \widetilde{\mathcal{M}})(d)$ for $k = 1$ and any given d. Among the popular options are e.g. Laplace, Gaussian, or entropy-maximising distributions, which may come in various flavours and with auxiliary constraints, but many properties can be captured by just two generic attributes: the noise variance $\mathrm{Var}(x)$ and its magnitude bound $|x| \leq E$ $(\leq \infty)$. Here we just give a crude classification based on the DP categories introduced in annex A.1.

ε-DP noise distributions manifestly comply with Eq. (8) for any possible singleton S (single output count) with $\delta = 0$. It is easy to show (Dwork, 2011) that e.g. the Laplace distribution

$$\mathrm{Lap}\,(\Delta/\varepsilon) : \; x \sim \frac{\varepsilon}{2\Delta} \exp\left(-\frac{\varepsilon|x|}{\Delta}\right) \tag{9}$$

with $\mathrm{Var}(x) \equiv V = 2(\Delta/\varepsilon)^2$ fulfils this requirement, where Δ is the *global sensitivity* of $\widetilde{\mathcal{M}}$ defined as

$$\Delta := \max_{d,d'} \sum_{i=1}^{k} \left| \widetilde{\mathcal{M}}(d)_i - \widetilde{\mathcal{M}}(d')_i \right| \tag{10}$$

with i running through output k-tuple indices. Clearly for $k = 1$ and unweighted person counts, $\Delta = 1$ and $x \sim \mathrm{Lap}(1/\varepsilon)$ in this case.[16] Now this distribution is over $\mathbb{R}$, so that $\mathrm{Range}(\mathcal{M}) = \mathbb{R}^k$ which may return

[16] Apart from unweighted counts, the issue with Δ is that it is generally hard to obtain, and arbitrarily difficult for some queries on weighted or magnitude data: e.g. Bambauer et al. (2014) argue that in an average income query the global sensitivity is theoretically

non-integer person counts. This can be lifted this by using the discrete two-tailed geometric distribution (Ghosh et al., 2012)

$$x \sim \frac{1-\exp(-\varepsilon)}{1+\exp(-\varepsilon)} \exp(-\varepsilon|x|), \tag{11}$$

which gives $\mathrm{Range}(\mathcal{M}) = \mathbb{Z}^k$ and approximates to $\mathrm{Lap}(1/\varepsilon)$ for $\varepsilon \ll 1$.
Note finally that noise distributions, continuous or discrete, must be unbounded to be ε-DP. To see this, assume bounded noise with $\Pr(x > E) = 0$. Then in Eq. (8), choose without loss of generality $d > d'$ and $s = \widetilde{\mathcal{M}}(d) + E$ (i.e. $k = 1$). Thus, $\Pr(\mathcal{M}(d') = s) = 0$ as $s - \widetilde{\mathcal{M}}(d') = E + 1$ and the inequality requires $\delta \geq \Pr(\mathcal{M}(d) = s) > 0$ to hold, which contradicts $\delta = 0$.

(ε, δ)-DP and other noise distributions In a sloppy manner, most noise distributions that are not ε-DP are (ε, δ)-DP: If a distribution fails the strict ε-DP requirement, Eq. (8) with $\delta = 0$, a $\delta > 0$ can usually be found to establish (ε, δ)-DP. In particular, unbounded noise distributions can usually be truncated to give (ε, δ)-DP, where δ depends on the resulting probability distribution close to its discontinuity (Rinott et al., 2018). Also for cell key noise of Thompson et al. (2013), taking variance V and noise bound $|x| \leq E$ as input parameters, an (ε, δ)-DP level can be inferred (Bailie and Chien, 2019). However, the issue is not about finding a δ but about dealing with its value: clearly it should be $\delta \ll 1$ but how small exactly? For instance, $\delta < 1/n$ is stated in Dwork and Roth (2014), but higher values are also discussed in Rinott et al. (2018). It is also argued there that often the choices of δ (and ε) are policy decisions, not statistical decisions.

Bounded vs. unbounded noise Why select an unbounded noise distribution? As shown above, if a strictly ε-DP output mechanism is ultimately desired, the underlying noise distribution must be unbounded. Moreover, Asghar and Kaafar (2020) recently claimed that a tight noise bound poses additional disclosure risks. However, sections 3 and 4 argue that unbounded noise may come at too high a price on utility, while the additional risks of bounded noise can be controlled.

References

Abowd, J., R. Ashmead, R. Cumings-Menon, S. Garfinkel, M. Heineck, C. Heiss, R. Johns, D. Kifer, P. Leclerc, A. Machanavajjhala, B. Moran, W. Sexton, M. Spence, and P. Zhuravlev (2022). The 2020 Census Disclosure Avoidance System TopDown Algorithm. *Harvard Data Science Review* (Special Issue 2).

Abowd, J. M. (2018). The U.S. Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, New York, NY, USA, pp. 2867. Association for Computing Machinery.

Antal, L., M.-L. Buron, A. Cabrera, T. Enderle, S. Gießing, J. Lukan, E. Schulte Nordholt, and A. Smukavec (2017). Harmonised protection of census data. https://ec.europa.eu/eurostat/cros/content/harmonised-protection-census-data_en. Accessed on 13 Jul 2023.

Asghar, H. J. and D. Kaafar (2020). Averaging attacks on bounded noise-based disclosure control algorithms. *Proceedings on Privacy Enhancing Technologies 2020*(2), 358 – 378.

Bailie, J. and C.-H. Chien (2019). ABS perturbation methodology through the lens of differential privacy. In *Joint UNECE/Eurostat work session on statistical data confidentiality*.

Bambauer, J., K. Muralidhar, and R. Sarathy (2014). Fool's gold! An illustrated critique of differential privacy. *Vanderbilt J. Entertain. Technol. Law 16*, 701–755.

De Wolf, P.-P., T. Enderle, A. Kowarik, and B. Meindl (2019a). Perturbative confidentiality methods. https://ec.europa.eu/eurostat/cros/content/perturbative-confidentiality-methods_en. Accessed on 13 Jul 2023.

driven by the highest-income person in the world, because the query result must be robust also against addition of that person to the database. Naturally such a Δ drives the noise through the roof and renders all outputs useless. On the other hand, capping Δ arbitrarily dilutes the individual privacy guarantee.

De Wolf, P.-P., T. Enderle, A. Kowarik, and B. Meindl (2019b). SDC Tools - user support and sources of tools for statistical disclosure control. https://github.com/sdcTools. Accessed on 13 Jul 2023.

Dwork, C. (2011). A firm foundation for private data analysis. *Commun. ACM 54*(1), 86–95.

Dwork, C., K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor (2006). Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 486–503. Springer.

Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin (Eds.), *Theory of Cryptography*, Berlin, Heidelberg, pp. 265–284. Springer Berlin Heidelberg.

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

Dwork, C. and A. Smith (2010). Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality 1*(2).

Enderle, T., S. Gießing, and R. Tent (2018). Designing confidentiality on the fly methodology - three aspects. In J. Domingo-Ferrer and F. Montes (Eds.), *Privacy in Statistical Databases - UNESCO Chair in Data Privacy, International Conference, PSD 2018, Valencia, Spain, September 26-28, 2018, Proceedings*, Volume 11126 of *Lecture Notes in Computer Science*, pp. 28–42. Springer.

Enderle, T., S. Gießing, and R. Tent (2020). Calculation of risk probabilities for the cell key method. In J. Domingo-Ferrer and K. Muralidhar (Eds.), *Privacy in Statistical Databases - UNESCO Chair in Data Privacy, International Conference, PSD 2020, Tarragona, Spain, September 23-25, 2020, Proceedings*, Volume 12276 of *Lecture Notes in Computer Science*, pp. 151–165. Springer.

Fraser, B. and J. Wooton (2005). A proposed method for confidentialising tabular output to protect against differencing. In *Monographs of Official Statistics: Work Session on Statistical Data Confidentiality*, pp. 299–302.

Garfinkel, S. L. (2019). Deploying differential privacy for the 2020 census of population and housing. In *JSM 2019 Session: Formal Privacy - Making an Impact at Large Organizations*.

Ghosh, A., T. Roughgarden, and M. Sundararajan (2012). Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing 41*(6), 1673–1693.

Gießing, S. (2016). Computational issues in the design of transition probabilities and disclosure risk estimation for additive noise. In J. Domingo-Ferrer and M. Pejic-Bach (Eds.), *Privacy in Statistical Databases - UNESCO Chair in Data Privacy, International Conference, PSD 2016, Dubrovnik, Croatia, September 14-16, 2016, Proceedings*, Volume 9867 of *Lecture Notes in Computer Science*, pp. 237–251. Springer.

Hsu, J., M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth (2014). Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, Volume 2014-January, pp. 398–410.

Machanavajjhala, A., D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber (2008). Privacy: Theory meets practice on the map. In *IEEE 24th International Conference on Data Engineering (ICDE)*, pp. 277–286.

Marley, J. K. and V. L. Leaver (2011). A method for confidentialising user-defined tables: Statistical properties and a risk-utility analysis. In *Int. Statistical Inst.: Proc. 58th World Statistical Congress (Session IPS060)*, pp. 1072–1081.

Muralidhar, K. and J. Domingo-Ferrer (2023). Legacy statistical disclosure limitation techniques for protecting 2020 decennial us census: Still a viable option. *Journal of Official Statistics Forthcoming*.

Petti, S. and A. Flaxman (2019). Differential privacy in the 2020 US census: what will it do? Quantifying the accuracy/privacy tradeoff. *Gates Open Research 3*.

Rinott, Y., C. O'Keefe, N. Shlomo, and C. Skinner (2018). Confidentiality and differential privacy in the dissemination of frequency tables. *Statistical Science 33*, 358–385.

Ruggles, S., C. Fitch, D. Magnuson, and J. Schroeder (2019). Differential privacy and census data: Implications for social and economic research. *AEA Papers and Proceedings 109*, 403–08.

Santos-Lozada, A. R., J. T. Howard, and A. M. Verdery (2020). How differential privacy will affect our understanding of health disparities in the United States. *Proceedings of the National Academy of Sciences 117*(24), 13405–13412.

Thompson, G., S. Broadfoot, and D. Elazar (2013). Methodology for the automatic confidentialisation of statistical outputs from remote servers at the Australian Bureau of Statistics. In *Joint UNECE/Eurostat work session on statistical data confidentiality*.

Wang, Y., J. Lee, and D. Kifer (2015). Revisiting differentially private hypothesis tests for categorical data. arXiv: 1511.03376 [cs.CR].