Exploring Noise Infusion for Disclosure Avoidance at BEA

Dan Yorgason and John Bockrath, June 9th 2023

Federal Economic Statistics Advisory Committee Meeting

Overview

Proposal: Replace cell suppression with simple **"EZS" noise infusion** to protect confidentiality for a key—trade in services—survey

Benefits: 1) Publish values for **all cells**; 2) Introduce flexibility to **publish greater detail** and/or across additional domains; 3) Simplify application of disclosure avoidance

Cost: Distortions in non-vulnerable cells

Feedback sought: 1) **Usability** of statistics with noise vs. suppression; 2) Choice of method; 3) Importance of transparency of methods

Noise infusion proposed to replace cell suppression

- U.S. international trade in **services**
  - Quarterly surveys of U.S. trade in various service types (IP, telecommunications, insurance etc.);
  - Surveys account for roughly 55% of U.S. exports and 40% of U.S. imports in services
  - Included in international transactions accounts (ITAs); also monthly trade in goods and services statistics, annual detailed services statistics, and annual ICT and ICT-enabled services statistics

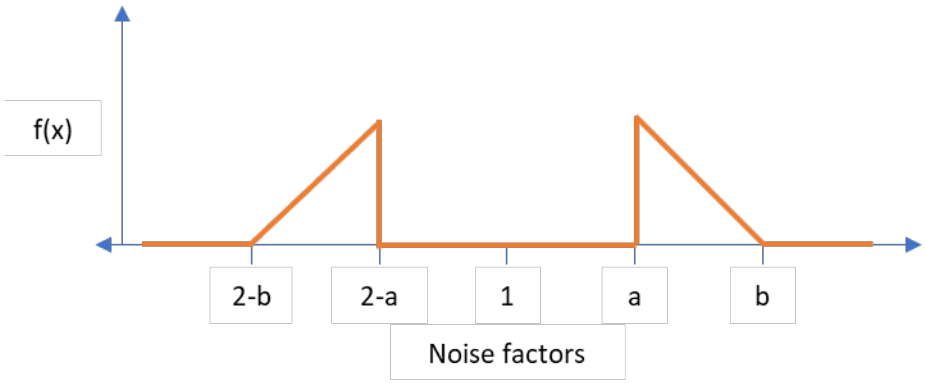
Noise not (yet) proposed; cell suppression to continue

- **Direct investment** transactions and positions
  - Transactions included in ITAs
- Activities of **multinational enterprises**
  - Not included in ITAs

Proposed (EZS-based) Disclosure Avoidance Approach

Randomly perturb each observation: produces noise-infused microdata used to construct published tables

- Draw multiplicative noise factors from a symmetric distribution—for example, see prototypical EZS noise factor distribution at right
- No suppression used in published tables
- Circumscribed perturbations → circumscribed cell distortion
- **Vulnerable cells are distorted more** on average
- Perturbations from individual contributions **offset** to significant degree for larger aggregates
- Like suppression, no formal privacy guarantees
- No privacy budget constraint

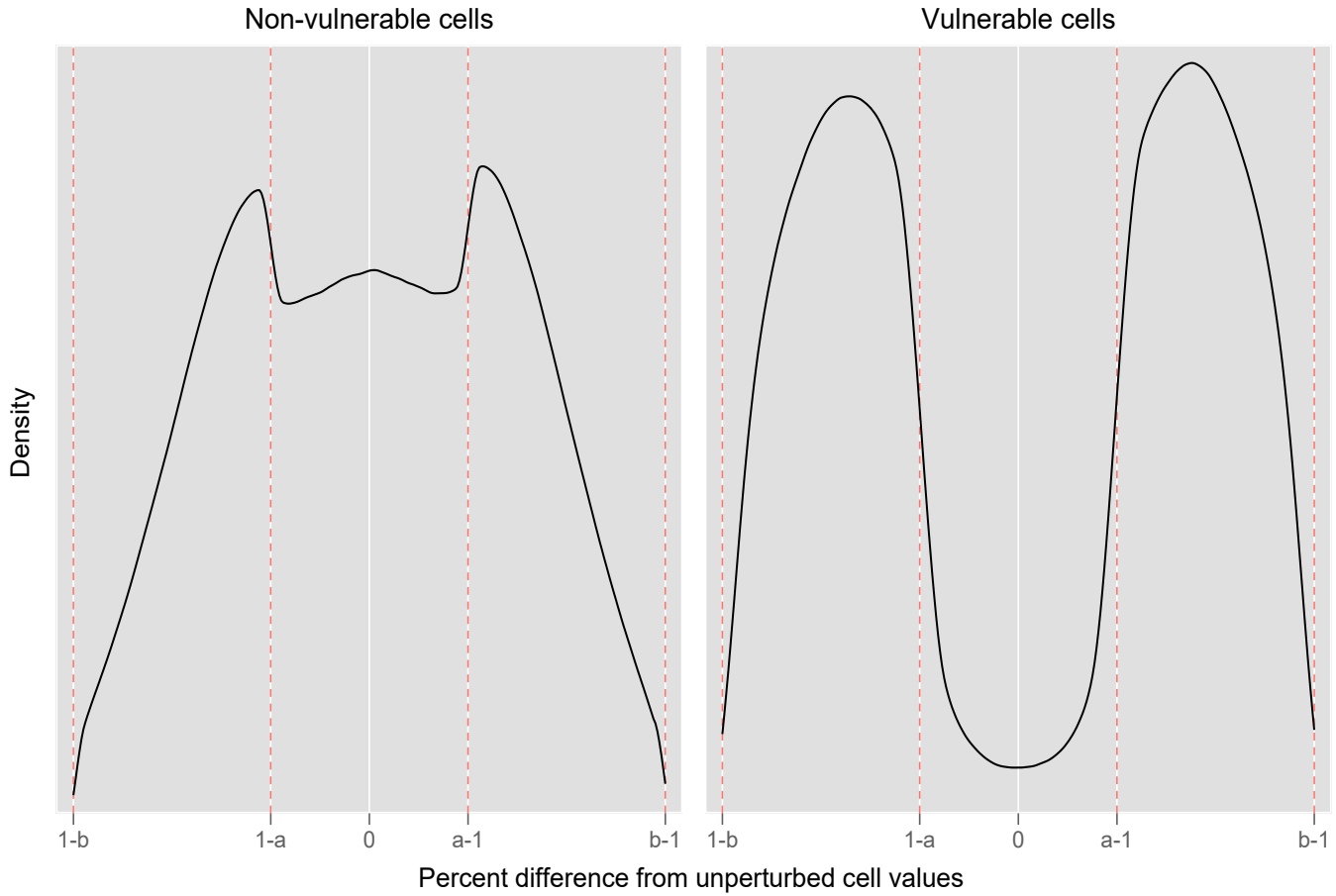


Add-ons

- **To protect reporters**, multiple noise factors for a given reporter are either all above one or all below one
- **To limit time series distortion**, noise factors for a given contribution are either always above or always below one

Distortion distributions

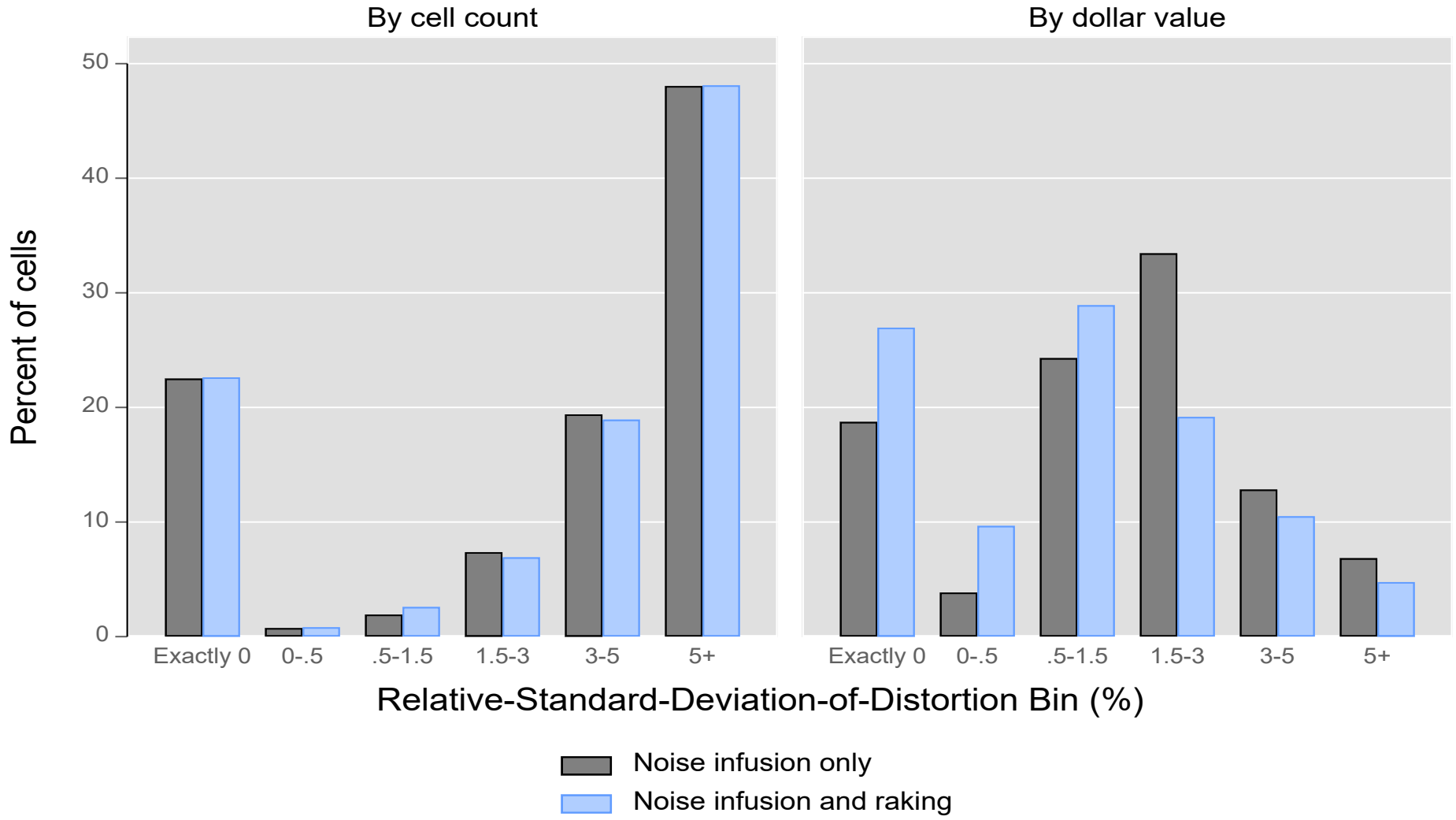
- Method produces:
  - Relatively **small distortions**, on average, for most cells
  - Larger distortions, on average, for vulnerable cells
- Results will **vary** with:
  - Distributions used for noise infusion (including whether bounded or not)
  - Distribution parameters
  - Underlying concentration of data within cells
- Protection is an on-average notion, not a yes/no notion like cell suppression



Additional considerations

- **Raking** to preserve undistorted aggregates, especially those that **contribute to GDP**
  - Increases variability of most unraked cells
  - May diminish average distortion of vulnerable cells, so considering excluding distorted vulnerable cells from raking procedure
- Communicating with users about switch in disclosure avoidance methods
- Informing users about magnitude of distortions
  - Cannot provide full information on individual distortions
  - Cannot provide full information on distributions used to infuse microdata with noise
  - Could provide, e.g., **average standard deviations** of distortions for cell groups or **range flags of standard deviations** for individual cells

Flagging distortions for users



Questions for Committee

- Is noise infusion an **improvement** over cell suppression?
- What level of distortion is **too much**?
- How much insight should be **provided to users** into the noise infusion process?
- How much information should be provided to users on **variability** of cell distortions?
  - Flags?
  - Something else?
- Any other questions users might like to see answered?