Scientific Inferences
From Privatized Census Data

Gary King\(^1\)

Institute for Quantitative Social Science
Harvard University

Panel Discussion on the Future Direction of the Census Bureau’s American Community Survey, for Federal Reserve Staff, 11/30/2021

\(^1\)GaryKing.org
Requirements for Scientific Measurement

1. Quantity of interest defined separately from any measure
   - E.g.: Forecasts, descriptions, causal effects

2. Measure with known statistical properties
   - E.g.: If we apply this rule to data we have lots of times, on average we'll get the right answer ("unbiasedness")
   - E.g. 2: The more data, the closer we'll likely get to the right answer ("consistency")

3. Accurate uncertainty estimates
   - E.g.: Margins of error (CIs), SEs, hypothesis tests, etc.
   - A scientific statement: not one that is necessarily correct, but one that comes with accurate uncertainty estimates
Requirements for Scientific Measurement

1. **Quantity of interest** defined separately from any measure
Requirements for Scientific Measurement

1. **Quantity of interest** defined separately from any measure
   - E.g.: Forecasts, descriptions, causal effects
Requirements for Scientific Measurement

1. **Quantity of interest** defined separately from any measure
   - E.g.: Forecasts, descriptions, causal effects
2. **Measure** with known statistical properties
Requirements for Scientific Measurement

1. **Quantity of interest** defined separately from any measure
   - E.g.: Forecasts, descriptions, causal effects
2. **Measure** with known statistical properties
   - E.g.: If we apply this rule to data we have lots of times, on average we’ll get the right answer (“unbiasedness”)
1. **Quantity of interest** defined separately from any measure
   - E.g.: Forecasts, descriptions, causal effects
2. **Measure** with known statistical properties
   - E.g.: If we apply this rule to data we have lots of times, on average we’ll get the right answer (“unbiasedness”)
   - E.g.2: The more data, the closer we’ll likely get to the right answer (“consistency”)

---

2/5
Requirements for Scientific Measurement

1. **Quantity of interest** defined separately from any measure
   - E.g.: Forecasts, descriptions, causal effects

2. **Measure** with known statistical properties
   - E.g.: If we apply this rule to data we have lots of times, on average we’ll get the right answer (“unbiasedness”)
   - E.g.2: The more data, the closer we’ll likely get to the right answer (“consistency”)

3. **Accurate uncertainty estimates**
1. **Quantity of interest** defined separately from any measure
   - E.g.: Forecasts, descriptions, causal effects
2. **Measure** with known statistical properties
   - E.g.: If we apply this rule to data we have lots of times, on average we’ll get the right answer (“unbiasedness”)
   - E.g.2: The more data, the closer we’ll likely get to the right answer (“consistency”)
3. **Accurate uncertainty estimates**
   - E.g.: Margins of error (CIs), SEs, hypothesis tests, etc.
Requirements for Scientific Measurement

1. **Quantity of interest** defined separately from any measure
   - *E.g.*: Forecasts, descriptions, causal effects

2. **Measure** with known statistical properties
   - *E.g.*: If we apply this rule to data we have lots of times, on average we’ll get the right answer (“unbiasedness”)
   - *E.g.2*: The more data, the closer we’ll likely get to the right answer (“consistency”)

3. **Accurate uncertainty estimates**
   - *E.g.*: Margins of error (CIs), SEs, hypothesis tests, etc.
   - *A scientific statement*: not one that is necessarily correct, but one that comes with accurate uncertainty estimates
### The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$+Privacy$</th>
<th>$\vdots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret</td>
<td>3</td>
<td>127</td>
<td>97</td>
</tr>
<tr>
<td>Robert</td>
<td>3</td>
<td>76</td>
<td>103</td>
</tr>
<tr>
<td>Bhashkar</td>
<td>3</td>
<td>145</td>
<td>75</td>
</tr>
<tr>
<td>Gary</td>
<td>3</td>
<td>96</td>
<td>113</td>
</tr>
<tr>
<td>Jerome</td>
<td>3</td>
<td>86</td>
<td>125</td>
</tr>
<tr>
<td>Sandra</td>
<td>3</td>
<td>127</td>
<td>97</td>
</tr>
<tr>
<td>Amanda</td>
<td>3</td>
<td>72</td>
<td>101</td>
</tr>
<tr>
<td>Ari</td>
<td>3</td>
<td>132</td>
<td>128</td>
</tr>
<tr>
<td>Hannah</td>
<td>3</td>
<td>95</td>
<td>83</td>
</tr>
<tr>
<td>Indraneel</td>
<td>3</td>
<td>134</td>
<td>201</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quantity of Interest</th>
<th>Usually no direct relevance</th>
<th>No direct relevance</th>
</tr>
</thead>
</table>

- **Mean income:**
  - [Classical Inference](#)
  - [Query-Response](#)
The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Mean income:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret</td>
<td>$48</td>
</tr>
<tr>
<td>Robert</td>
<td>$122</td>
</tr>
<tr>
<td>Bhashkar</td>
<td>$145</td>
</tr>
<tr>
<td>Gary</td>
<td>$76</td>
</tr>
<tr>
<td>Jerome</td>
<td>$145</td>
</tr>
<tr>
<td>Sandra</td>
<td>$96</td>
</tr>
<tr>
<td>Amanda</td>
<td>$127</td>
</tr>
<tr>
<td>Ari</td>
<td>$72</td>
</tr>
<tr>
<td>Hannah</td>
<td>$132</td>
</tr>
<tr>
<td>Indraneel</td>
<td>$108</td>
</tr>
</tbody>
</table>

*Quantity of Interest*

*Mean income:*

*Statistically Valid Inferences from Privacy Protected Data*
## The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret</td>
<td>✓</td>
</tr>
<tr>
<td>Robert</td>
<td>✓</td>
</tr>
<tr>
<td>Bhashkar</td>
<td>✓</td>
</tr>
<tr>
<td>Gary</td>
<td>✓</td>
</tr>
<tr>
<td>Jerome</td>
<td>✓</td>
</tr>
<tr>
<td>Sandra</td>
<td>✓</td>
</tr>
<tr>
<td>Amanda</td>
<td>✓</td>
</tr>
<tr>
<td>Ari</td>
<td>✓</td>
</tr>
<tr>
<td>Hannah</td>
<td>✓</td>
</tr>
<tr>
<td>Indraneel</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Mean income:

$48

### Quantity of Interest
## The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>:</td>
<td>✕</td>
<td>?</td>
</tr>
<tr>
<td>Margaret</td>
<td>✓</td>
<td>122</td>
</tr>
<tr>
<td>Robert</td>
<td>✓</td>
<td>76</td>
</tr>
<tr>
<td>Bhashkar</td>
<td>✓</td>
<td>145</td>
</tr>
<tr>
<td>Gary</td>
<td>✓</td>
<td>96</td>
</tr>
<tr>
<td>Jerome</td>
<td>✓</td>
<td>86</td>
</tr>
<tr>
<td>Sandra</td>
<td>✓</td>
<td>127</td>
</tr>
<tr>
<td>Amanda</td>
<td>✓</td>
<td>72</td>
</tr>
<tr>
<td>Ari</td>
<td>✓</td>
<td>132</td>
</tr>
<tr>
<td>Hannah</td>
<td>✓</td>
<td>95</td>
</tr>
<tr>
<td>Indraneel</td>
<td>✓</td>
<td>134</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean income:</th>
<th>Classical Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$48</td>
<td>$108</td>
</tr>
</tbody>
</table>

### Quantity of Interest
- Usually no direct relevance

---
# The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret</td>
<td>✔️</td>
<td>122</td>
</tr>
<tr>
<td>Robert</td>
<td>✔️</td>
<td>76</td>
</tr>
<tr>
<td>Bhashkar</td>
<td>✔️</td>
<td>145</td>
</tr>
<tr>
<td>Gary</td>
<td>✔️</td>
<td>96</td>
</tr>
<tr>
<td>Jerome</td>
<td>✔️</td>
<td>86</td>
</tr>
<tr>
<td>Sandra</td>
<td>✔️</td>
<td>127</td>
</tr>
<tr>
<td>Amanda</td>
<td>✔️</td>
<td>72</td>
</tr>
<tr>
<td>Ari</td>
<td>✔️</td>
<td>132</td>
</tr>
<tr>
<td>Hannah</td>
<td>✔️</td>
<td>95</td>
</tr>
<tr>
<td>Indraneel</td>
<td>✔️</td>
<td>134</td>
</tr>
</tbody>
</table>

| Mean income: | $48 | Classical Inference | $108 |

| Quantity of Interest | Usually no direct relevance |

Population + Privacy = dp

...
### The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
<th>+Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret</td>
<td>✓</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>Robert</td>
<td>✓</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Bhashkar</td>
<td>✓</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>Gary</td>
<td>✓</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Jerome</td>
<td>✓</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Sandra</td>
<td>✓</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>Amanda</td>
<td>✓</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Ari</td>
<td>✓</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>Hannah</td>
<td>✓</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Indraneel</td>
<td>✓</td>
<td>134</td>
<td></td>
</tr>
</tbody>
</table>

Mean income: $48 \rightarrow \text{Classical Inference} \rightarrow $108

Quantities of Interest: Usually no direct relevance
### The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>Quantity of Interest</th>
<th>$ +Privacy</th>
<th>=dp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret</td>
<td>✓</td>
<td>?</td>
<td>122</td>
<td>85</td>
</tr>
<tr>
<td>Robert</td>
<td>✓</td>
<td>?</td>
<td>76</td>
<td>103</td>
</tr>
<tr>
<td>Bhashkar</td>
<td>✓</td>
<td>?</td>
<td>145</td>
<td>75</td>
</tr>
<tr>
<td>Gary</td>
<td>✓</td>
<td>?</td>
<td>96</td>
<td>113</td>
</tr>
<tr>
<td>Jerome</td>
<td>✓</td>
<td>?</td>
<td>86</td>
<td>125</td>
</tr>
<tr>
<td>Sandra</td>
<td>✓</td>
<td>?</td>
<td>127</td>
<td>97</td>
</tr>
<tr>
<td>Amanda</td>
<td>✓</td>
<td>?</td>
<td>72</td>
<td>101</td>
</tr>
<tr>
<td>Ari</td>
<td>✓</td>
<td>?</td>
<td>132</td>
<td>128</td>
</tr>
<tr>
<td>Hannah</td>
<td>✓</td>
<td>?</td>
<td>95</td>
<td>83</td>
</tr>
<tr>
<td>Indraneel</td>
<td>✓</td>
<td>?</td>
<td>134</td>
<td>201</td>
</tr>
</tbody>
</table>

### Mean income:

- Classical Inference: $48
- Query-Response: $108
- $111

### Noise & Censoring

- Usually no direct relevance
- No direct relevance
# The Role of Differential Privacy: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
<th>+Privacy</th>
<th>=dp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>:</td>
<td>☑️</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margaret</td>
<td>☑️</td>
<td>122</td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>Robert</td>
<td>☑️</td>
<td>76</td>
<td></td>
<td>103</td>
</tr>
<tr>
<td>Bhashkar</td>
<td>☑️</td>
<td>145</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>Gary</td>
<td>☑️</td>
<td>96</td>
<td></td>
<td>113</td>
</tr>
<tr>
<td>Jerome</td>
<td>☑️</td>
<td>86</td>
<td></td>
<td>125</td>
</tr>
<tr>
<td>Sandra</td>
<td>☑️</td>
<td>127</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>Amanda</td>
<td>☑️</td>
<td>72</td>
<td></td>
<td>101</td>
</tr>
<tr>
<td>Ari</td>
<td>☑️</td>
<td>132</td>
<td></td>
<td>128</td>
</tr>
<tr>
<td>Hannah</td>
<td>☑️</td>
<td>95</td>
<td></td>
<td>83</td>
</tr>
<tr>
<td>Indraneel</td>
<td>☑️</td>
<td>134</td>
<td></td>
<td>201</td>
</tr>
</tbody>
</table>

**Mean income:**

Classical Inference: $48 → $108

Query-Response: $108 → $111

Statistically Valid Inferences from Privacy Protected Data
Analyzing Differentially Private Data (Data + Noise)

• Statistical methods: must change!

• Consequence of ignoring DP noise

• Bias: any direction, any magnitude

• Proper analysis of DP data (with corrected methods)

• Estimates with known statistical properties (as with raw data)

• Accurate uncertainty estimates (as with raw data)

• The only change with DP: larger CIs

• The only valid objections to DP

• Added privacy protections: not necessary

• The larger CIs: too large for my QOI
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods:** must change!
- Consequence of ignoring DP noise
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods:** must change!
- **Consequence of ignoring DP noise**
  - **Bias:** any direction, any magnitude
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- **Consequence of ignoring DP noise**
  - **Bias**: any direction, any magnitude
- **Proper analysis of DP data (with corrected methods)**

- **The (only) valid objections to DP**
  - **Added privacy protections**: not necessary
  - **The larger CIs**: too large for my QOI
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- **Consequence of ignoring DP noise**
  - **Bias**: any direction, any magnitude
- **Proper analysis of DP data (with corrected methods)**
  - estimates with *known statistical properties* (as with raw data)

The (only) valid objections to DP

- Added privacy protections: not necessary
- The larger CIs: too large for my QOI
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods:** must change!
- Consequence of ignoring DP noise
  - **Bias:** any direction, any magnitude
- Proper analysis of DP data (with corrected methods)
  - estimates with **known statistical properties** (as with raw data)
  - accurate uncertainty estimates (as with raw data)
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- Consequence of ignoring DP noise
  - **Bias**: any direction, any magnitude
- Proper analysis of DP data (with corrected methods)
  - estimates with known statistical properties (as with raw data)
  - accurate uncertainty estimates (as with raw data)
  - the only change with DP:
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- Consequence of ignoring DP noise
  - **Bias**: any direction, any magnitude
- Proper analysis of DP data (with corrected methods)
  - estimates with known statistical properties (as with raw data)
  - accurate uncertainty estimates (as with raw data)
  - the only change with DP: larger CIs
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- **Consequence of ignoring DP noise**
  - **Bias**: any direction, any magnitude
- **Proper analysis of DP data (with corrected methods)**
  - estimates with **known statistical properties** (as with raw data)
  - accurate uncertainty estimates (as with raw data)
  - the only change with DP: larger CIs
- **The (only) valid objections to DP**
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- **Consequence of ignoring DP noise**
  - **Bias**: any direction, any magnitude
- **Proper analysis of DP data (with corrected methods)**
  - estimates with **known statistical properties** (as with raw data)
  - accurate uncertainty estimates (as with raw data)
  - the only change with DP: larger CIs
- **The (only) valid objections to DP**
  - I don’t wanna learn new statistical methods!
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods:** must change!
- **Consequence of ignoring DP noise**
  - **Bias:** any direction, any magnitude
- **Proper analysis of DP data (with corrected methods)**
  - estimates with known statistical properties (as with raw data)
  - accurate uncertainty estimates (as with raw data)
  - the only change with DP: larger CIs
- **The (only) valid objections to DP**
  - I don’t wanna learn new statistical methods!
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- Consequence of ignoring DP noise
  - **Bias**: any direction, any magnitude
- Proper analysis of DP data (with corrected methods)
  - estimates with **known statistical properties** (as with raw data)
  - accurate uncertainty estimates (as with raw data)
  - the only change with DP: larger CIs
- The (only) valid objections to DP
  - I don’t wanna learn new statistical methods!
  - Added privacy protections: not necessary
Analyzing Differentially Private Data (Data + Noise)

- **Statistical methods**: must change!
- Consequence of ignoring DP noise
  - **Bias**: any direction, any magnitude
- Proper analysis of DP data (with corrected methods)
  - estimates with **known statistical properties** (as with raw data)
  - accurate uncertainty estimates (as with raw data)
  - the only change with DP: larger CIs
- The (only) valid objections to DP
  - I don’t wanna learn new statistical methods!
  - Added privacy protections: not necessary
  - The larger CIs: too large for my QOI
US Census Privatization Strategies

1990-2010: Secret (failed) Privatization
- Methods: Swapping, top coding, cell suppression (no details)
- Privatization fails: most people can be reidentified
- Valid inferences: impossible

2020: Public Privatization
- Method: Add DP noise to census block counts (public DGP)
- Privatized "Noisy Measurements File"
- Valid inferences: easy, but data not (yet) released!
- Post-Processed data released: "TopDown Algorithm"
- Motivation: CB's legacy code, users' statistical confusion
- Valid inferences: (most are) extremely difficult
- Proper statistical methods: not developed yet

What can the Fed do?
- Push Census Bureau to release the Noisy Measurements File
US Census Privatization Strategies

1990-2010: Secret (failed) Privatization

Methods:
- Swapping, top coding, cell suppression

Privatization fails:
- Most people can be reidentified

Valid inferences:
- Impossible

2020: Public Privatization

Method:
- Add DP noise to census block counts (public DGP)

Privatized "Noisy Measurements File"

Valid inferences:
- Easy, but data not (yet) released!

Post-Processed data released:
- "TopDown Algorithm"

Motivation:
- CB's legacy code, users' statistical confusion

Valid inferences:
- (Most are) extremely difficult

Proper statistical methods:
- Not developed yet

What can the Fed do?
- Push Census Bureau to release the Noisy Measurements File
US Census Privatization Strategies

1990-2010: Secret (failed) Privatization

- Methods:
  - Swapping, top coding, cell suppression (no details)

- Privatization fails:
  - Most people can be reidentified

- Valid inferences:
  - Impossible

2020: Public Privatization

- Method:
  - Add DP noise to census block counts (public DGP)

- Privatized “Noisy Measurements File”

- Valid inferences:
  - Easy, but data not (yet) released!

- Post-processed data released:
  - “TopDown Algorithm”

- Motivation:
  - CB’s legacy code, users’ statistical confusion

- Valid inferences:
  - (Most are) extremely difficult

- Proper statistical methods:
  - Not developed yet

What can the Fed do?

- Push Census Bureau to release the Noisy Measurements File
# US Census Privatization Strategies

## 1990-2010: Secret (failed) Privatization

- **Methods:** Swapping, top coding, cell suppression (no details)
- **Privatization fails:** most people can be reidentified
- **Valid inferences:** impossible

## 2020: Public Privatization

- **Method:** Add DP noise to census block counts (public DGP)
- **Privatized “Noisy Measurements File”**
- **Valid inferences:** easy, but data not (yet) released!
  - **Post-Processed data released:** “TopDown Algorithm”
- **Motivation:** CB’s legacy code, users’ statistical confusion
- **Valid inferences:** (most are) extremely difficult
- **Proper statistical methods:** not developed yet

## What can the Fed do?

- Push Census Bureau to release the Noisy Measurements File
US Census Privatization Strategies

1990-2010: Secret (failed) Privatization

- **Methods**: Swapping, top coding, cell suppression (no details)

2020: Public Privatization

What can the Fed do?
<table>
<thead>
<tr>
<th>US Census Privatization Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1990-2010: Secret (failed) Privatization</strong></td>
</tr>
<tr>
<td>• <strong>Methods</strong>: Swapping, top coding, cell suppression (no details)</td>
</tr>
<tr>
<td>• <strong>Privatization fails</strong>: most people can be reidentified</td>
</tr>
<tr>
<td><strong>2020: Public Privatization</strong></td>
</tr>
<tr>
<td><strong>What can the Fed do?</strong></td>
</tr>
</tbody>
</table>
US Census Privatization Strategies

1990-2010: **Secret** (failed) Privatization

- **Methods**: Swapping, top coding, cell suppression (no details)
- **Privatization fails**: most people can be reidentified
- **Valid inferences**: impossible

2020: **Public** Privatization

- **Methods**:
  - Add DP noise to census block counts (public DGP)
  - Privatized "Noisy Measurements File"
- **Valid inferences**: easy, but data not (yet) released!
- **Post-Processed data released**:
  - "TopDown Algorithm"
- **Motivation**:
  - CB's legacy code, users' statistical confusion
- **Valid inferences**:
  - (most are) extremely difficult
- **Proper statistical methods**:
  - not developed yet

What can the Fed do?

- Push Census Bureau to release the Noisy Measurements File
US Census Privatization Strategies

1990-2010: *Secret* (failed) Privatization

- **Methods**: Swapping, top coding, cell suppression (no details)
- **Privatization fails**: most people can be reidentified
- **Valid inferences**: impossible

2020: *Public* Privatization

- **Method**: Add DP noise to census block counts (public DGP)

What can the Fed do?
US Census Privatization Strategies

1990-2010: Secret (failed) Privatization
- Methods: Swapping, top coding, cell suppression (no details)
- Privatization fails: most people can be reidentified
- Valid inferences: impossible

2020: Public Privatization
- Method: Add DP noise to census block counts (public DGP)
- Privatized “Noisy Measurements File”

What can the Fed do?
## US Census Privatization Strategies

### 1990-2010: Secret (failed) Privatization

- **Methods:** Swapping, top coding, cell suppression (no details)
- **Privatization fails:** most people can be reidentified
- **Valid inferences:** impossible

### 2020: Public Privatization

- **Method:** Add DP noise to census block counts (public DGP)
- **Privatized “Noisy Measurements File”**
  - **Valid inferences:** easy, but *data not (yet) released!*

### What can the Fed do?

Push Census Bureau to release the Noisy Measurements File
US Census Privatization Strategies

1990-2010: *Secret* (failed) Privatization

- **Methods:** Swapping, top coding, cell suppression (no details)
- **Privatization fails:** most people can be reidentified
- **Valid inferences:** impossible

2020: *Public* Privatization

- **Method:** Add DP noise to census block counts (public DGP)
- **Privatized “Noisy Measurements File”**
  - **Valid inferences:** easy, but *data not (yet) released!*
- **Post-Processed data released:** “TopDown Algorithm”

What can the Fed do?
# US Census Privatization Strategies

## 1990-2010: Secret (failed) Privatization
- **Methods:** Swapping, top coding, cell suppression (no details)
- **Privatization fails:** most people can be reidentified
- **Valid inferences:** impossible

## 2020: Public Privatization
- **Method:** Add DP noise to census block counts (public DGP)
- **Privatized “Noisy Measurements File”**
  - **Valid inferences:** easy, but *data not (yet) released!*
- **Post-Processed data released:** “TopDown Algorithm”
  - **Motivation:** CB’s legacy code, users’ statistical confusion

## What can the Fed do?
- Push Census Bureau to release the Noisy Measurements File
### US Census Privatization Strategies

#### 1990-2010: Secret (failed) Privatization
- **Methods:** Swapping, top coding, cell suppression (no details)
- **Privatization fails:** most people can be reidentified
- **Valid inferences:** impossible

#### 2020: Public Privatization
- **Method:** Add DP noise to census block counts (public DGP)
- **Privatized “Noisy Measurements File”**
  - **Valid inferences:** easy, but *data not (yet) released!*
- **Post-Processed data released:** “TopDown Algorithm”
  - **Motivation:** CB’s legacy code, users’ statistical confusion
  - **Valid inferences:** (most are) extremely difficult

### What can the Fed do?

Push Census Bureau to release the Noisy Measurements File
## US Census Privatization Strategies

### 1990-2010: Secret (failed) Privatization
- **Methods**: Swapping, top coding, cell suppression (no details)
- **Privatization fails**: most people can be reidentified
- **Valid inferences**: impossible

### 2020: Public Privatization
- **Method**: Add DP noise to census block counts (public DGP)
- **Privatized “Noisy Measurements File”**
  - **Valid inferences**: easy, but *data not (yet) released!*
- **Post-Processed data released**: “TopDown Algorithm”
  - **Motivation**: CB’s legacy code, users’ statistical confusion
  - **Valid inferences**: (most are) extremely difficult
  - **Proper statistical methods**: not developed yet

### What can the Fed do?
- Push Census Bureau to release the Noisy Measurements File
US Census Privatization Strategies

1990-2010: **Secret** (failed) Privatization

- **Methods**: Swapping, top coding, cell suppression (no details)
- **Privatization fails**: most people can be reidentified
- **Valid inferences**: impossible

2020: **Public** Privatization

- **Method**: Add DP noise to census block counts (public DGP)
- **Privatized “Noisy Measurements File”**
  - **Valid inferences**: easy, but *data not (yet) released!*
- **Post-Processed data released**: “TopDown Algorithm”
  - **Motivation**: CB’s legacy code, users’ statistical confusion
  - **Valid inferences**: (most are) extremely difficult
  - **Proper statistical methods**: not developed yet

What can the Fed do?

Push Census Bureau to release the Noisy Measurements File