

# Evaluating Noise Infusion for Disclosure Protection for Two Time Periods

Jared Martin  
U.S. Census Bureau

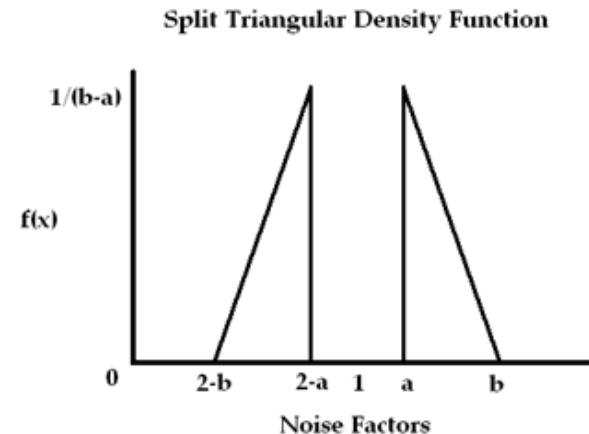
Note: Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

# The Economic Census of Island Areas (IA)

- Quinquennial census
- Provides comprehensive data on economic activity and structure for Island Areas of the U.S.
- Covers American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, Puerto Rico and the U.S. Virgin Islands

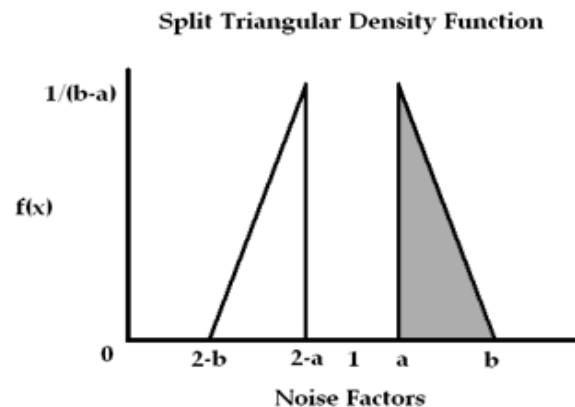
# Noise Infusion for Disclosure Avoidance

- Perturbs each individual contributor to a tabulated result by a small amount
- For the Economic Census of Island Areas, noise infusion applied using multiplicative factor from split triangular distribution
- Guarantees minimal protection by choosing appropriate noise factor distribution



# Noise Infusion as Applied to IA

- Noise factors are ‘balanced’ using reported sales data at the industry category level
- If a company reports with more than one establishment, we assign each establishment’s noise factor in the same direction as the company’s noise factor.



# Concerns About Noise In Consecutive Periods

- Tabulated results with few contributing companies or with a few large contributors have large variation due to noise
- Large variation in a tabulated value due to noise may drive change in tabulated value from cycle to cycle and interfere with trend over time

# Example of Noise Effects on Totals

| Company      | Cycle 1        |              |                | Cycle 2        |              |               |
|--------------|----------------|--------------|----------------|----------------|--------------|---------------|
|              | Actual Sales   | Noise Factor | Noisy Sales    | Actual Sales   | Noise Factor | Noisy Sales   |
| A            | \$ 10.0        | 1.14         | \$ 11.4        | \$ 10.0        | 0.85         | \$ 8.5        |
| B            | \$ 0.5         | 0.9          | \$ 0.45        | \$ 0.5         | 1.10         | \$ 0.55       |
| C            | \$ 0.5         | 0.9          | \$ 0.45        | \$ 0.5         | 1.10         | \$ 0.55       |
| <b>Total</b> | <b>\$ 11.0</b> |              | <b>\$ 12.3</b> | <b>\$ 11.0</b> |              | <b>\$ 9.6</b> |

# New Method Suggestions

- Preserve noise factor direction for contributors appearing from cycle to cycle
- Some combination of preserving direction and magnitude of factor from cycle to cycle

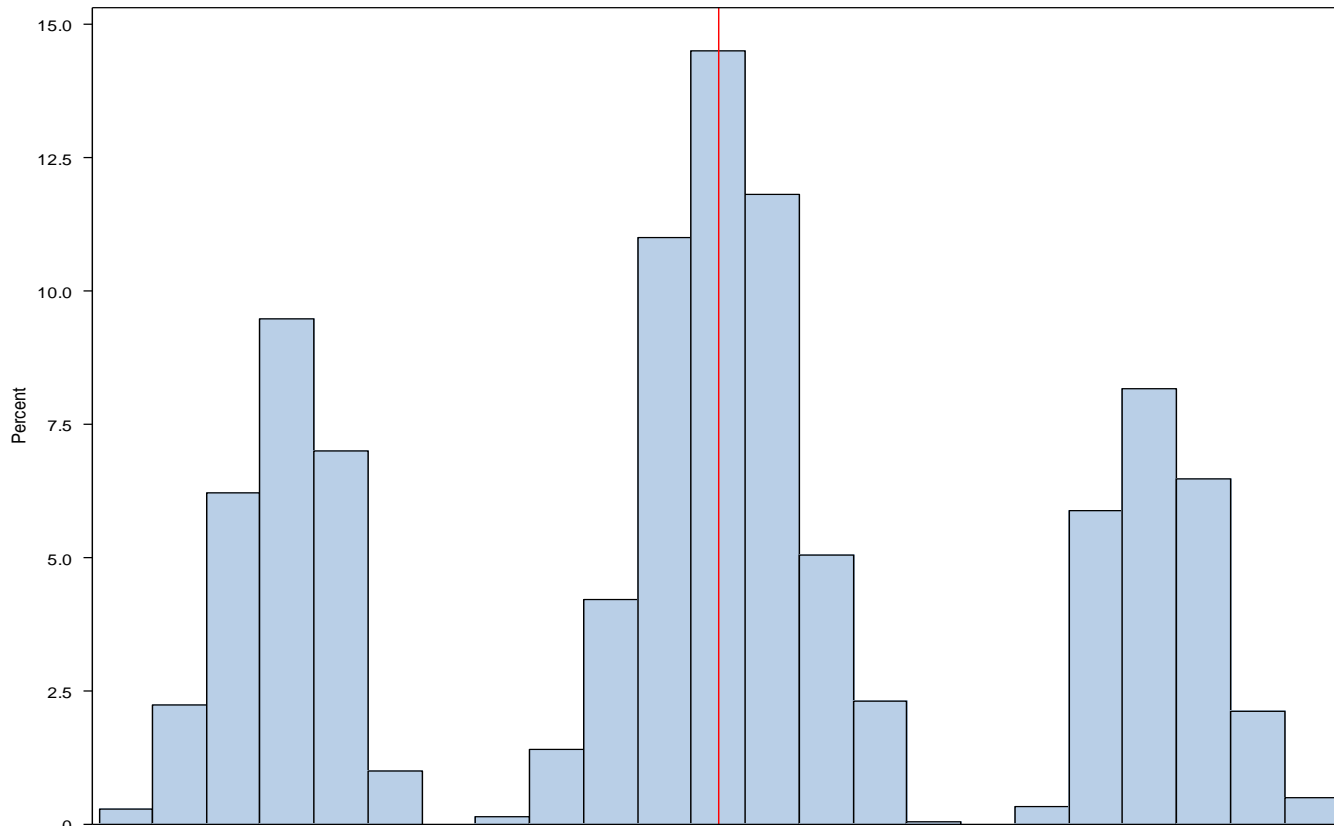
# Methods 1–4 (out of 10)

## Methods Evaluated Using Simulations

- Method 1 – Current methodology with direction and magnitude of noise not fixed
- Method 2 – Fixing the direction and magnitude of noise factors for companies appearing from cycle to cycle
- Method 3 – Fixing only the direction of noise factors appearing from cycle to cycle
- Method 4 – Fixing direction and magnitude of noise factors for companies appearing from cycle to cycle and contributing more than 50% of an estimate or are in top two contributors that together contribute at least 75%



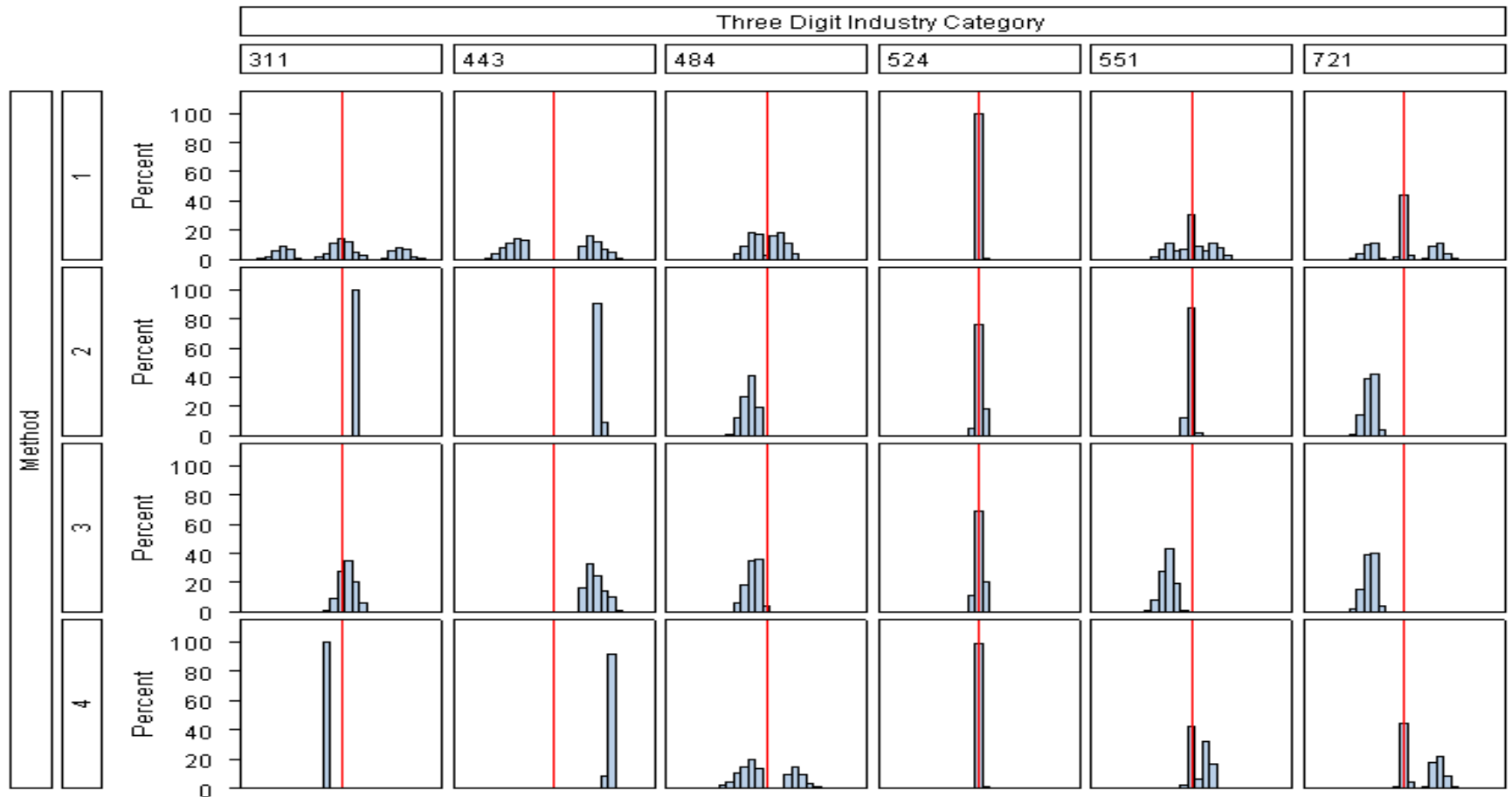
# Simulation of Variation Due to Noise



Non-Noisy Tabulation Value

# Results of Simulation

## Distribution of Distortion as a Percent of Non-Noisy Total Sales for American Samoa



Method1 – Reassigned noise direction and magnitude

Method2 – Fixed noise direction and magnitude

Method3 – Fixed noise direction

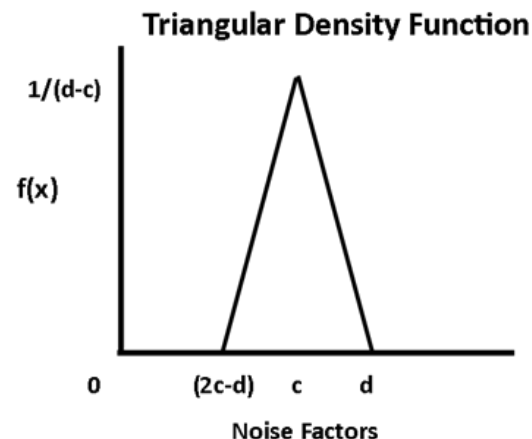
Method4 – Fixed noise direction and magnitude for cases > 50% of an estimate or in the top two contributors that together are > 75% of an estimate

# Results

- Tabulated values using Method 1 had the largest variation due to noise
- In Method 3, fixing direction of noise factors reduced variation of tabulated values due to noise
- In Method 2, fixing direction and magnitude of noise factors further reduced variation of tabulated values due to noise, but raises concerns of not enough variation to provide adequate disclosure avoidance
- In Method 4, fixing direction and magnitude of noise factors for “large” companies provided some reduction in variation due to noise, but “large” is subjective

# Methods 5 and 6

- Method 5 – Fixing direction and randomly picking magnitude of noise factors from triangular density ( $c =$  old factor magnitude,  $d = c + 0.0125$ ) for companies appearing from cycle to cycle and contributing more than 50% of an estimate
- Method 6 – Same as Method 5, except  $d = c + 0.005$

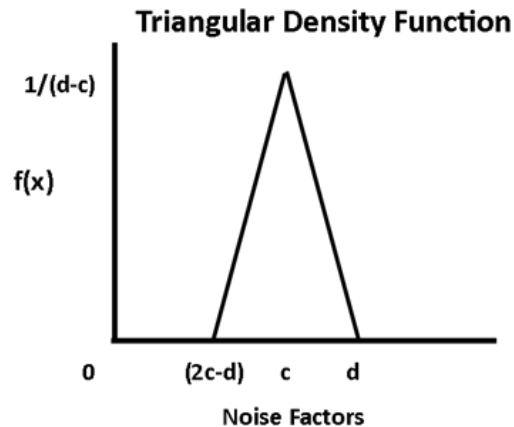


# Methods 7 and 8

- Method 7 – Fixing direction of noise factors for companies appearing from cycle to cycle and contributing more than 50% of a tabulated result
- Method 8 – Fixing direction for establishments appearing from cycle to cycle and contributing more than 75% of a tabulated result

# Methods 9 and 10

- Method 9 – Same as Method 5, except applied to establishments appearing from cycle to cycle and make up more than 50% of an estimate or are a company with more than one establishment
- Method 10 – Same as Method 4, except only direction fixed, allowing magnitude to vary

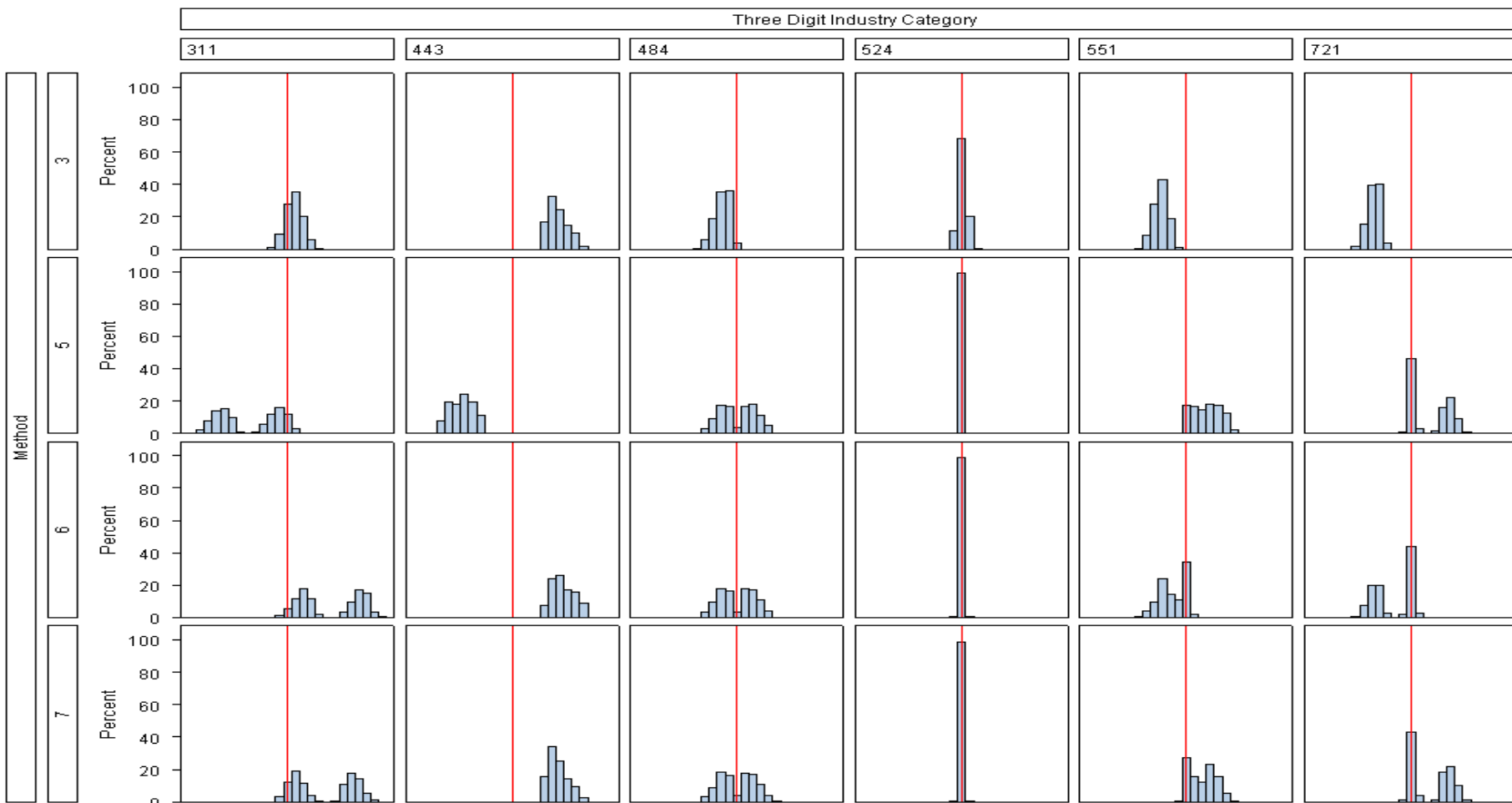


# Methods Summary

| Methods | Fixed Direction | Fixed Magnitude | Small Magnitude Variation | > 75% of Value | > 50% of Value | Top 2 >75% of Value | Companies With Multiple Establishments | All Companies |
|---------|-----------------|-----------------|---------------------------|----------------|----------------|---------------------|--|---------------|
| 1       |                 |                 |                           |                |                |                     |  | X             |
| 2       | X               | X               |                           |                |                |                     |  | X             |
| 3       | X               |                 |                           |                |                |                     |  | X             |
| 4       | X               | X               |                           |                | X              | X                   |  |               |
| 5       | X               |                 | X                         |                | X              |                     |  |               |
| 6       | X               |                 | X                         |                | X              |                     |  |               |
| 7       | X               |                 |                           |                | X              |                     |  |               |
| 8       | X               |                 |                           | X              |                |                     |  |               |
| 9       | X               |                 | X                         |                | X              |                     | X                                      |               |
| 10      | X               |                 |                           |                | X              | X                   |  |               |

# Results

## Distribution of Distortion as Percent of Non-Noisy Total Sales for American Samoa



Method 3 – Fixed noise direction

Method 5 – Fixed noise direction and magnitude varies by 0.0125 for cases > 50% of an estimate

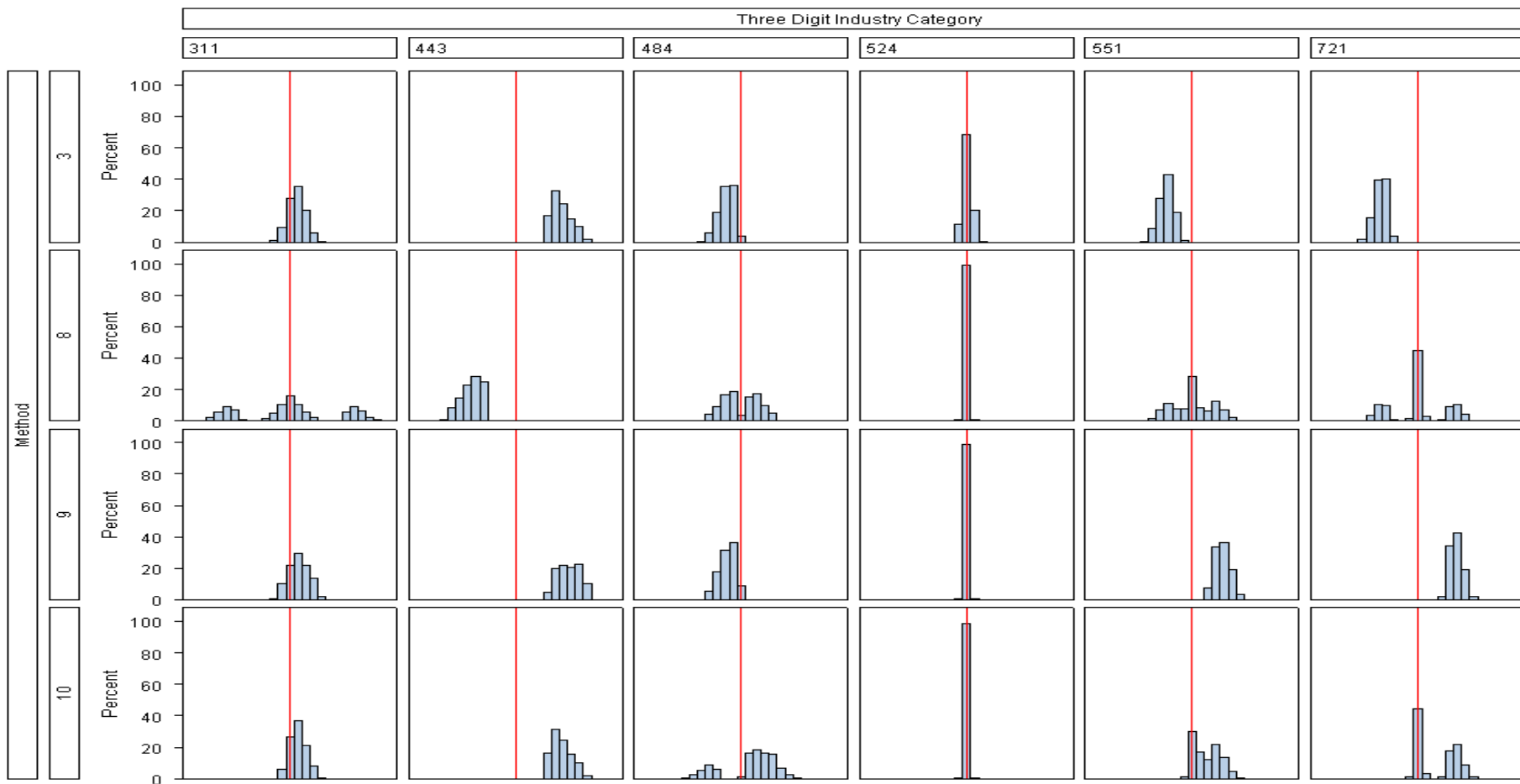
Method 6 – Fixed noise direction and magnitude varies by 0.005 for cases >50% of an estimate

Method 7 – Fixed noise direction for companies > 50% of an estimate



# Results

## Distribution of Distortion as Percent of Non-Noisy Total Sales for American Samoa



Method 3 – Fixed noise direction > 50% of an estimate

Method 8 – Fixed noise direction for companies that are > 75% of an estimate

Method 9 – Fixed noise direction and magnitude varies by 0.0125 for cases > 50% of an estimate or are a company with multiple establishments.

Method 10 – Fixed noise direction for cases > 50% of an estimate or in top two contributors that together are > 75% of an estimate

# Conclusion

- Rejected methods 5-7 because of exceptions where the distribution of overall distortion was biased or had too much variation.
- Rejected method 8 because too few establishments had fixed noise and the result nearly identical to current method.
- Methods 3, 9 and 10 produced similar results, but method 3 recommended because it most consistently reduced variance and bias in the distribution of overall percent noise of an estimate and it is the simplest of methods 3, 9 and 10

# Contact Information

## **Jared Martin**

Mathematical Statistician  
Company Statistics Division  
U.S. Census Bureau  
Office: 301-763-6974  
Email: [Jared.Martin@census.gov](mailto:Jared.Martin@census.gov)