

# Statistical Examination of Rounding Tendencies in the Consumer Expenditure Interview Survey

**Taylor J. Wilson**

Bureau of Labor Statistics

Joint Statistical Meetings (JSM)

Baltimore, Maryland

30 July 2017



# Research Question

Does the use of records help to reduce rounding in the Consumer Expenditure Interview Survey?



# Motivation

- Understanding the affect of respondent rounding on data quality.
  - ▶ Reducing total survey error
- Encouraging the use of records in the CE.
  - ▶ Reducing rounding
  - ▶ Improving data accuracy
- Data on record use are available.
- “Heaped” values prevalent in the data.

# Literature Review

- Response Heaping In Interviewer-Administered Surveys: A Cognitive Shortcut to Satisfice? (Holbrook et al, 2009)
- Encouraging Record Use for Financial Asset Questions in a Web Survey (Couper et al, 2013)
- Heaping at Round Numbers on Financial Questions: The Role of Satisficing (Gideon et al, 2017)

# Identifying Rounded Values

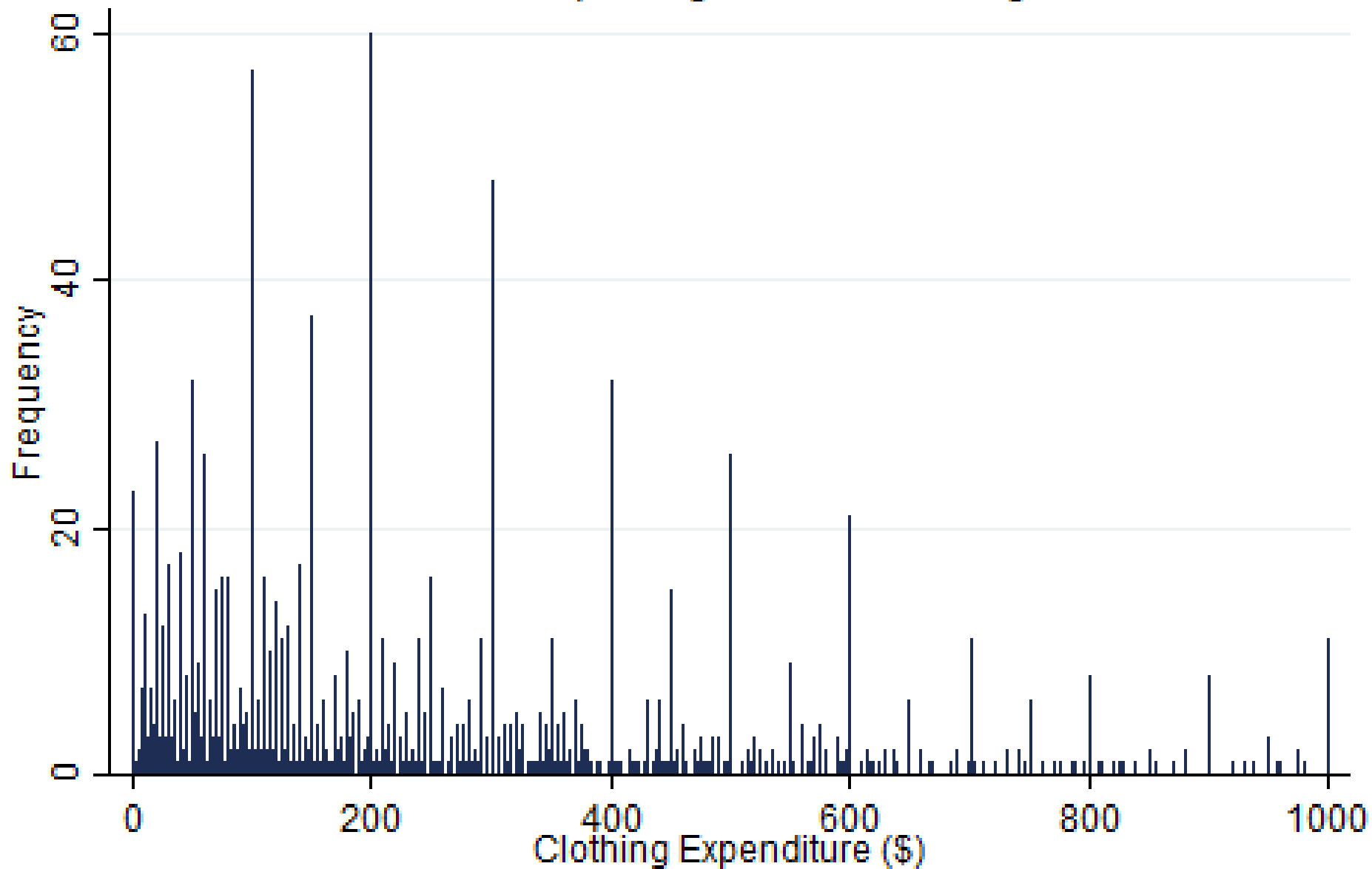
- Guess (educated)
- Cut off points from psychological studies
  - ▶ Works with ratios
  - ▶ Extrapolated to small expenditures
- Heaped values
  - ▶ Highest probability of being rounded
  - ▶ Need systematic identification

# What is a heaped value?

- “Heaping is a common type of measurement error emerging when data are collected with various degrees of coarseness. We say that a dataset is ‘heaped’ when it contains a mixture of exact and rounded-off values.”

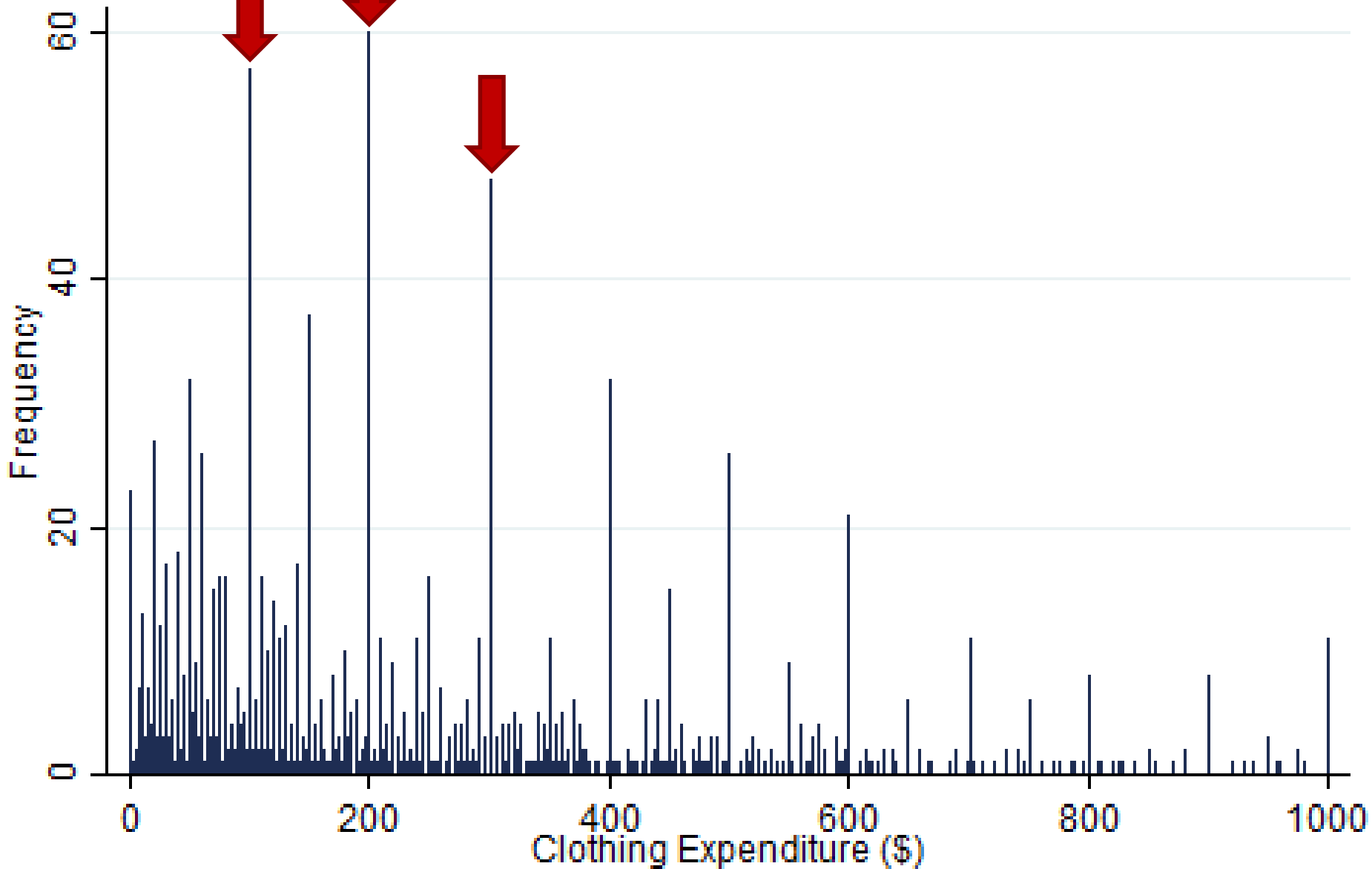
Wang, Hao, "Statistical methods for heaped data" (2009). *Dissertations available from ProQuest*. AAI3395726.

Frequency Histogram of Clothing Expenditures  
observations spanning two orders of magnitude



# Frequency Histogram of Clothing Expenditures

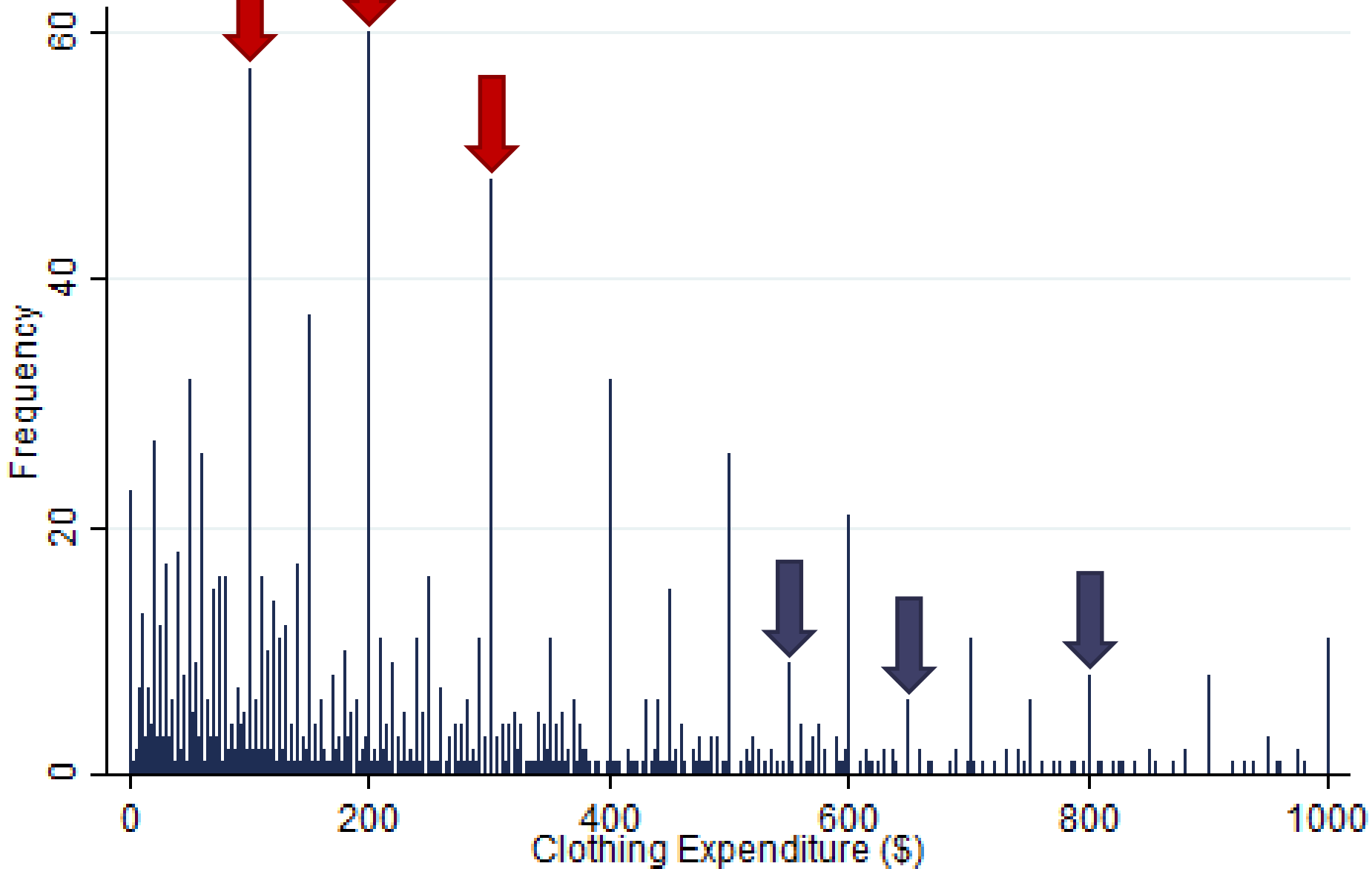
observations spanning two orders of magnitude





# Frequency Histogram of Clothing Expenditures

observations spanning two orders of magnitude



# Properties of Heaping

- As a function of rounding
  - ▶ Order of magnitude effects and relative value
  - ▶ Use heaped data to identify rounded records
- As a function of natural prices
  - ▶ Common and uncommon price points for items
- More common in discrete data
  - ▶ CE data is already rounded to the dollar

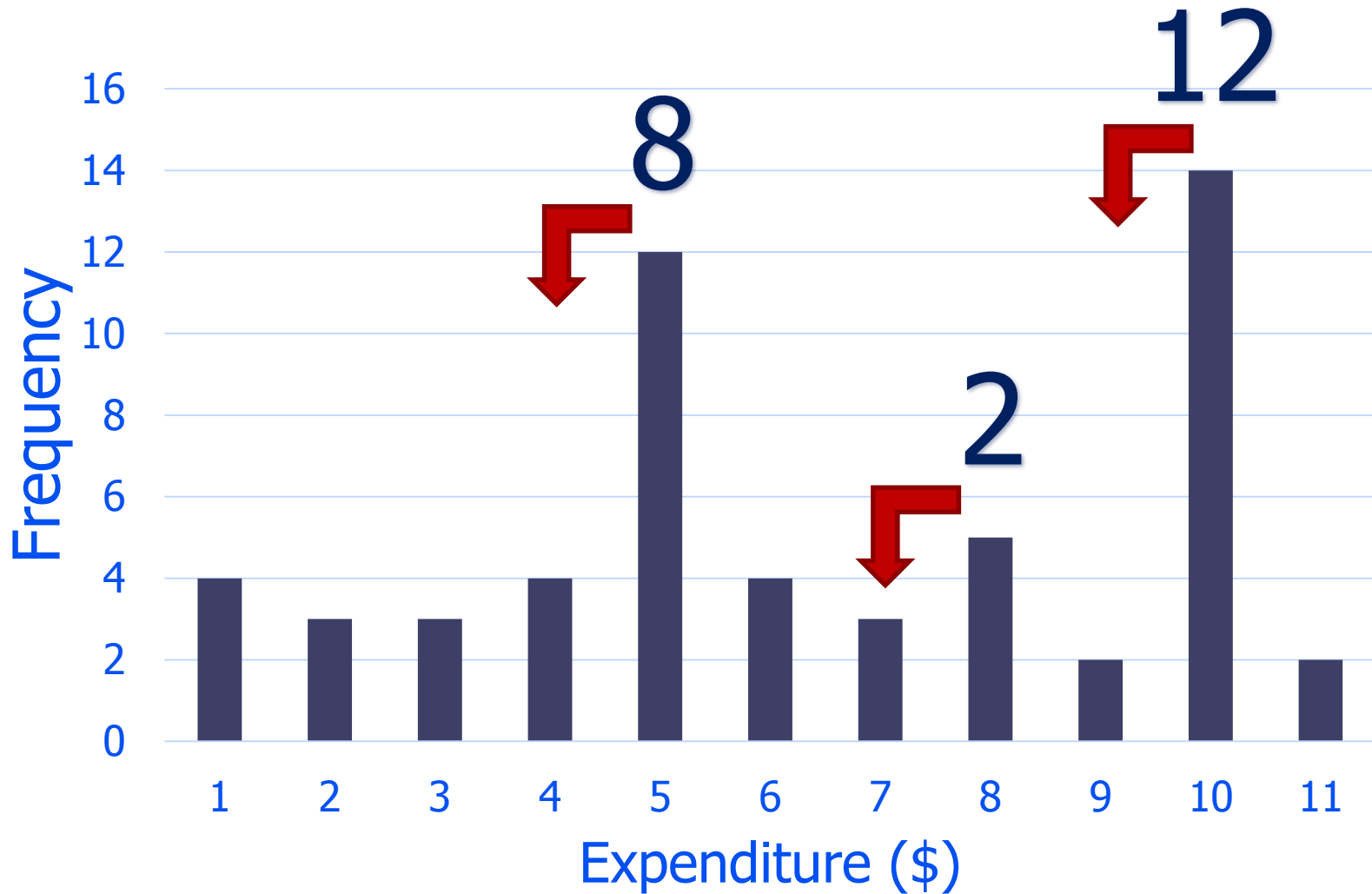
# How to identify heaping?

- Unique problems with order of magnitude effects related to expenditure data.
- The order of magnitude problem requires a more nuanced approach than the simple cut-off points that are typically suggested.
- Relative value and rounded dollar amounts.

# How to identify heaping?

- Dynamic record evaluation
  - ▶ Programmatic solution
  - ▶ Relativistic and non-arbitrary solutions
- How to determine sufficient concentration?
  - ▶ Deviation from typical heaped value
  - ▶ Relative to order of magnitude

# Average "Fall" Approach



# Dynamic Evaluation

- Requires understanding your data
  - ▶ How many orders of magnitude does your expenditure cover?
- Compute the conditional mean
  - ▶ Grouped on order of magnitude
- Determine the standard deviation for non-zero distances to the left and right of a spike within the groups.
- Loop through each observation and assign a t statistic to determine if it is significantly different from the norm.



# Testing Significance

- The distance distributions constructed from the falls are generally non-normal right skewed.
- 68-95-99 Rule fails to provide accurate measure of significance.
- Chebyshev's Inequality as a benchmark for significance.

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

# Record Use

- A household is said to have “used records” when the field representative notes that the household used records greater than 50% of the time.
- Natural Error and Variance in response.
- Hypothesis is that more record use implies less heaping.



# Record Use

- About half of households have data for record use with a small variance depending on the selected time frame for analysis.
- Of those who had data collected, about a fourth of these households used records over half the time (defined as a record user)
- Independent Variable of Interest

# Mann-Whitney U Test

- Non-parametric test for record use because of the non-normal underlying distributions.
- Allows us to test the hypothesis that the probabilities of randomly selecting a value from two independent non-normal distributions are equal.
- Rank-Sum procedure on two expenditure types on a fixed order of magnitude value domain.



# Mann-Whitney U Test

- Record use appears to be generally useful for smaller, large price-variance goods and services that aren't purchased on a repeatable basis.
- The following expenditure types were selected to exemplify the general behavior and to present this juxtaposition.
  - ▶ Clothing and Accessories
  - ▶ Subscriptions

## Clothing and Accessories on Value Domain [0,99]

Two-sample Mann-Whitney U test

records	obs	rank sum	expected
0	291	60178	58345.5
1	109	20022	21854.5
combined	400	80200	80200

$H_0: \Pr(\text{rounded value}(\text{records}=0)) = \Pr(\text{rounded value}(\text{records}=1))$

$z = 2.512$

**P value = 0.0120**

The use of records **dramatically decreased** incidence of identified rounded values. Natural prices do not typically fall on highly divisible values in this expenditure category so the **effect of records is expected to be greater.**

## Subscriptions on Value Domain [0,99]

Two-sample Mann-Whitney U test

records	obs	rank sum	expected
0	542	195798.5	195662
1	179	64482.5	64619
combined	721	260281	260281

$H_0: \Pr(\text{rounded value}(\text{records}=0)) = \Pr(\text{rounded value}(\text{records}=1))$

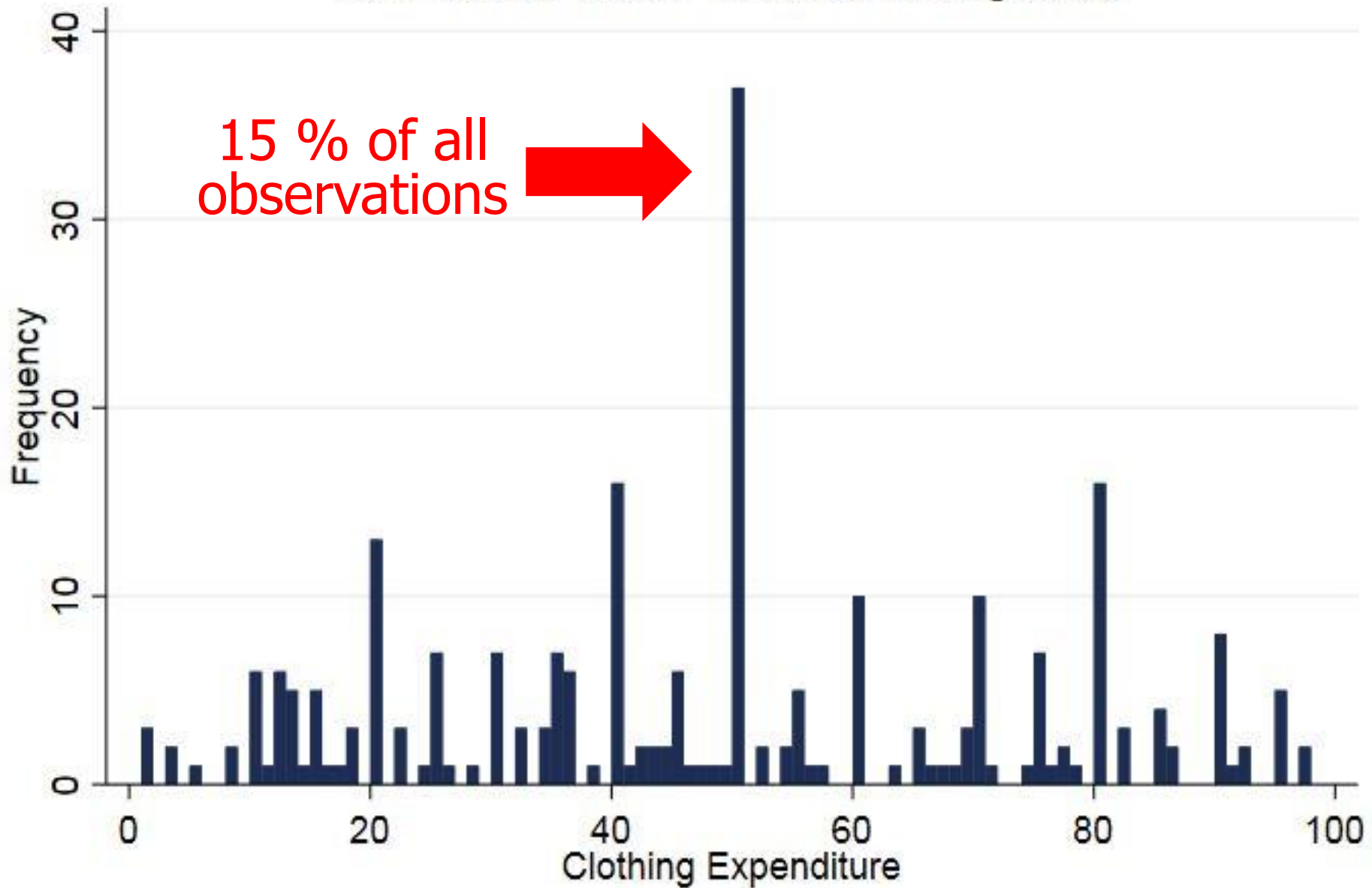
$z = 0.110$

**P value = 0.9123**

The use of records **did not decrease** incidence of identified rounded values. Natural prices typically fall on highly divisible values in this expenditure category so the **use of records is expected to be ineffectual.**

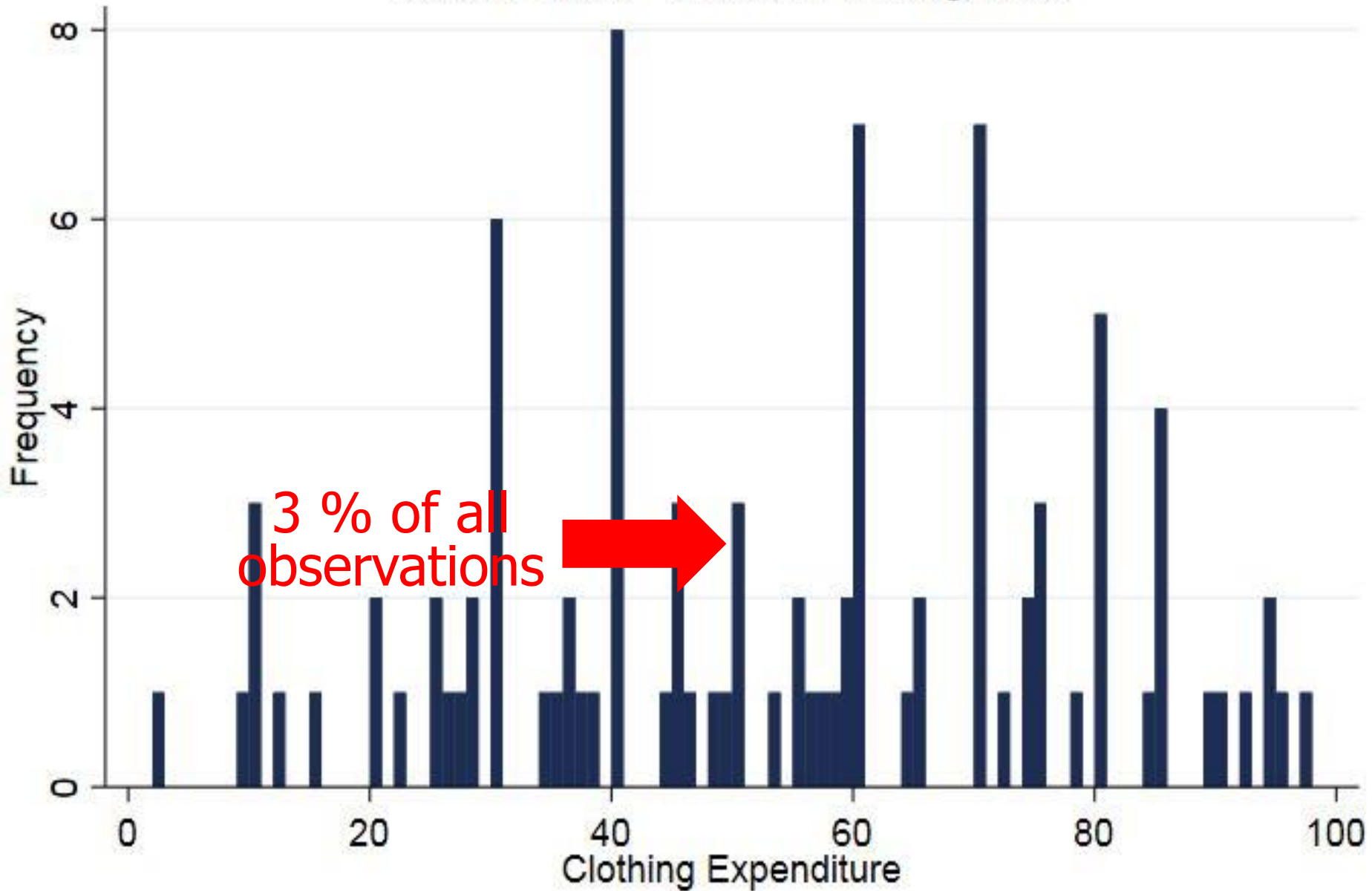
# Frequency Histogram of Clothing Expenditures

## Non-Record Users - 2 Orders of Magnitude



# Frequency Histogram of Clothing Expenditures

## Record Users - 2 Orders of Magnitude



# Conclusions

- Record use is helpful in improving data quality by reducing the coarseness of data for certain expenditure types.
- Recommendation is to repeat the analysis for every record type on a regular basis to evaluate any changes in consumer preferences or natural prices that may influence the effectiveness of records in a particular expenditure group.
- Spend resources targeting expenditure categories with record use incentives where you know record use makes a difference.



# Contact Information

**Taylor J. Wilson**  
Economist

Division of Consumer Expenditure Surveys

[www.bls.gov/cex](http://www.bls.gov/cex)

202-691-6550

[wilson.taylor@bls.gov](mailto:wilson.taylor@bls.gov)

