Reviewing Selective Editing Thresholds at ONS

A Retail Sales Pilot Study

Charlotte Gaughan

Overview

- Introduction
- Selective Editing Method
- Review
- Sample Design
- Pilot
- Results
- Conclusions and future work

Introduction

- The Retail Sales Inquiry (RSI) measures the value and volume of retail sales in Great Britain on a monthly basis
- In 2010 RSI moved to selective editing for two variables; Turnover and Employment.
- 5000 businesses sampled every month
- Savings of around 50% were possible without affecting the estimates

Selective Editing Method

- Selective editing score s_i calculated as

$$s_i = 100 \times \frac{w_i \left| z_i - \hat{y}_i \right|}{\hat{T}_{d,t-1}}$$

z_i is the unedited value for unit i

$\hat{y}_i$ is a predicted value for unit i, usually previous value

$\hat{T}_{d,t-1}$ is the previous period's estimate for the domain

- Scores calculated separately for turnover and employment and combined
- Thresholds set based on keeping domain bias within 1%

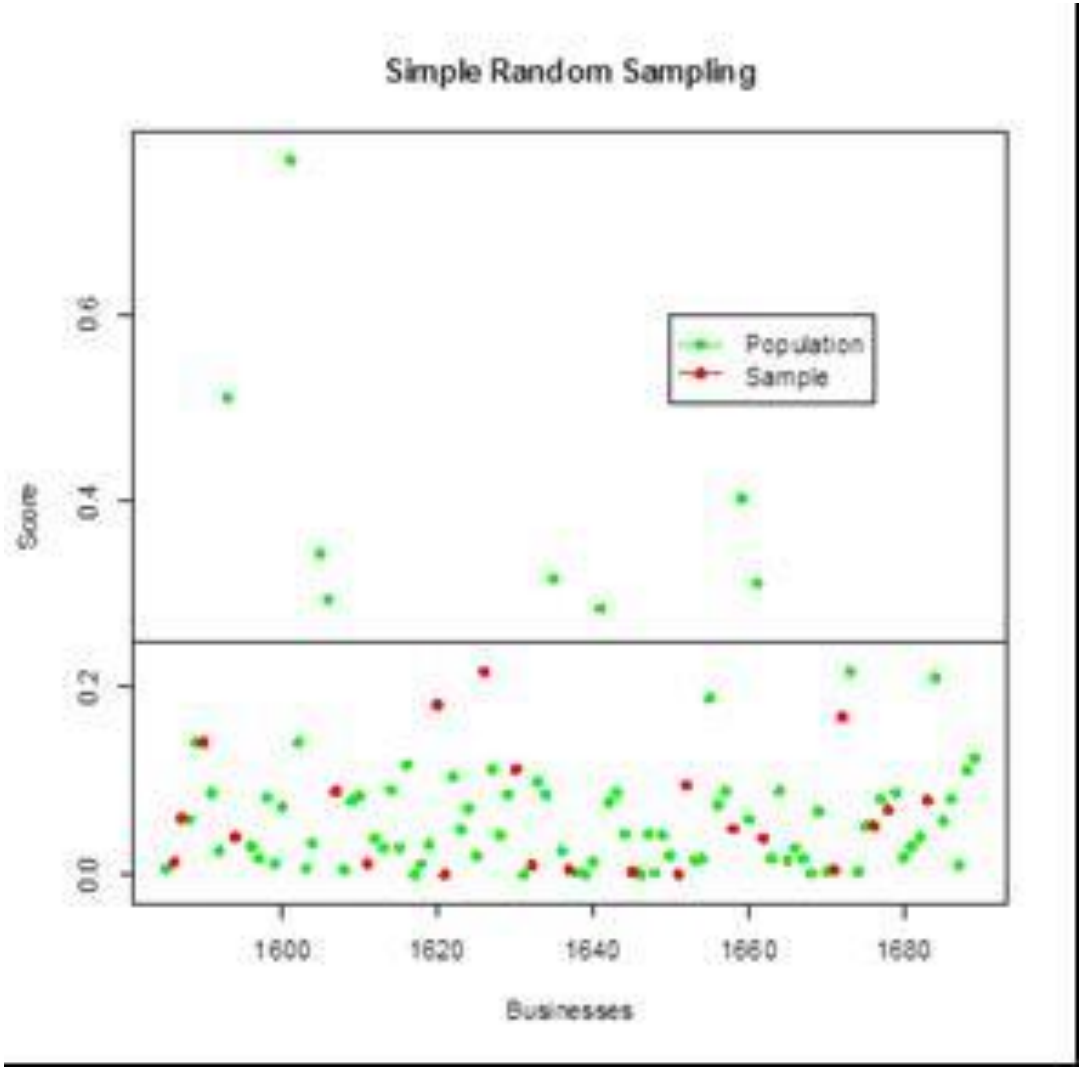
Review

- Thresholds set using past data which need to be updated periodically
- No fully edited dataset available going forward so take sub-sample of businesses that pass selective editing
- Re-contact those businesses to confirm their data
- Use results to check bias is still acceptable and re-set thresholds if required

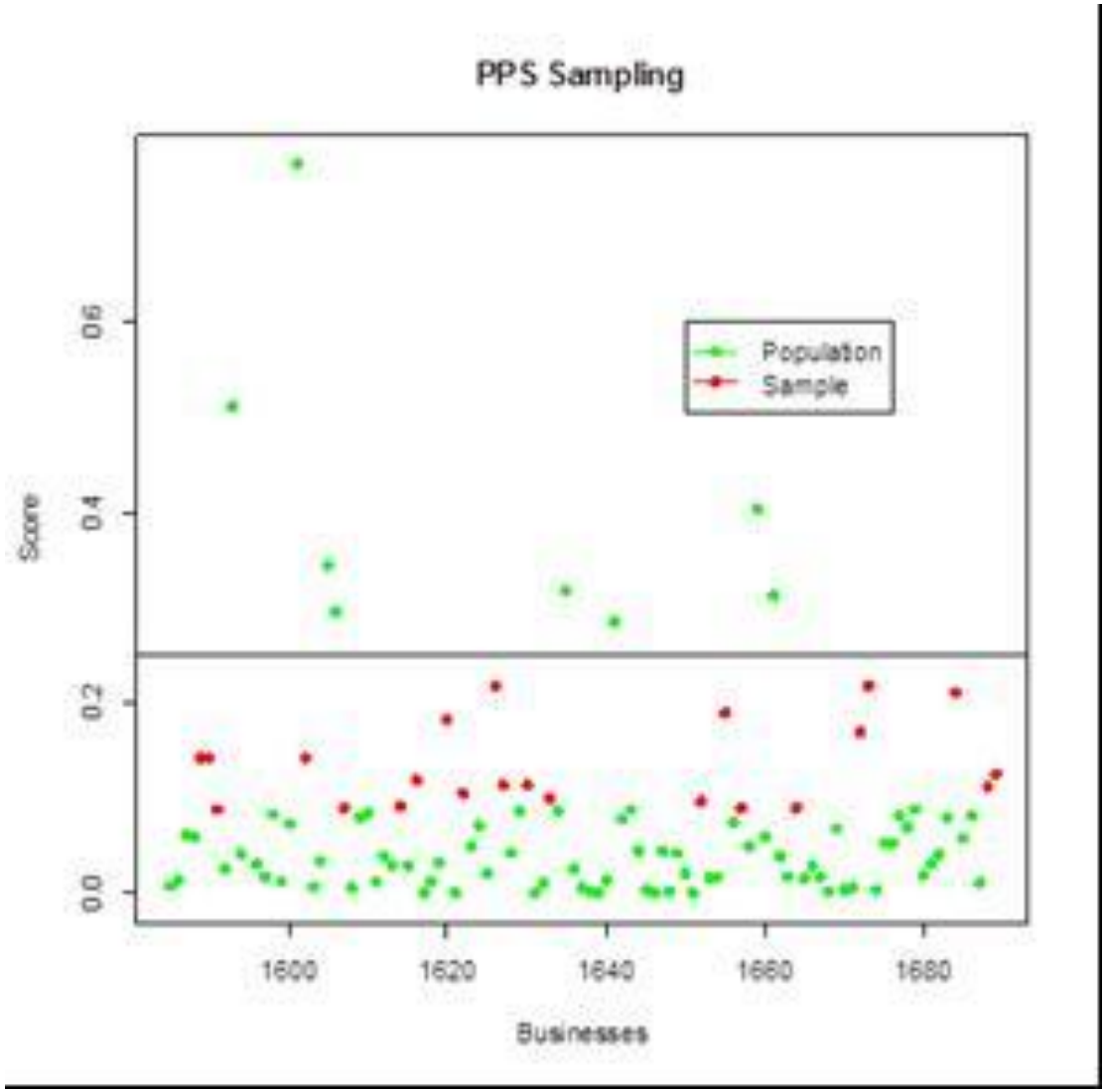
Sample design

- Agreed sample size of 600 businesses
- Tested variety of sample designs, two main options:
  - Stratified random sampling
  - Stratified probability proportional to size (PPS)
- Strata based on 25 industry groups, relating to publication domains
- Ensure minimum of 20 businesses in each stratum

Stratified random sample



Stratified PPS sample



Choice of sample design

- PPS is better at identifying businesses that are just below selective editing threshold
- However PPS SAS procedure did not allow for sampling from every stratum
- Compromised using stratified random sampling and splitting into 4 further strata:

Secondary Stratum	Scores of the Population	Percentage of n_h selected
1	75th percentile < Score <= 100th percentile	40%
2	50th percentile < Score <= 75th percentile	30%
3	25th percentile < Score <= 50th percentile	20%
4	Scores <= 25th percentile	10%

Pilot

- Selected sample of businesses which passed selective editing in December 2014
- Estimated bias from sample data
  - Accuracy of bias estimate tested using data from before introduction of selective editing
- Sample of 600 businesses taken from a month where employment is asked.

Results

- Thresholds for most domains are still fit for purpose – bias is within 1%
- Beneficial to slightly lower thresholds in a handful of domains, to ensure bias in control
- This can be offset by slightly increasing domains in some domains with very low bias

Conclusion and future work

- Taking sub-sample of businesses passing selective editing is an effective way to fine tune thresholds
- This method will be applied to all ONS surveys using this selective editing method
- Rolling programme of regular threshold reviews
  - Every 3 years for monthly surveys, every 5 years for annual surveys