Examining Relationships between Doorstep Concerns and Edit Check Frequency in the CE Interview Survey

Taylor J. Wilson
Economist
Bureau of Labor Statistics
American Association for Public Opinion Research (AAPOR)
18 May 2019

Roadmap

- Introduction
- Data
- Methods
- Results
- Conclusion

Roadmap

- Introduction
- Data
- Methods
- Results
- Conclusion

Statement of Purpose

- New machine readable error audit trails made accessible to internal researchers.
- Enables researchers to explore new questions that previously could not be supported with data.
- This research explores whether interviewer doorstep concerns are related to the frequency of edit checks over the course of the interview. This is just one of many potential questions that can be investigated!

Introduction

- Consumer Expenditure Interview Survey
  - Administered in person for at least the first interview
  - A household is selected to be in the survey four times
  - Entirely voluntary with approximately a 60% response rate
  - Designed to capture large, recurring expenses
    - Diary Survey designed to capture smaller expenses

Roadmap

- Introduction
- Data
- Methods
- Results
- Conclusion

Data: Error Audit Trails

- A new dataset that tracks errors triggered in the survey instrument by the interviewer.
- Errors are classified by type
  - Hard Edit Checks (You can't do that!)
  - Soft Edit Checks (Are you sure this is right?)
- Interviewer response to edit checks
  - Suppress, Go to, and Escape

Data: Hard and Soft Edit Checks

- Hard Edit Check Example:

  - You must enter all ten digits of the contact person's phone number. No spaces.

- Soft Edit Check Example:

  - WARNING! You are about to DELETE the ENTIRE current household and create a NEW REPLACEMENT case.

  - THE VALUE ENTERED IS UNUSUALLY HIGH OR LOW PLEASE VERIFY

Data: Contact History Attempt

- Contact history attempt data are created by the interviewer.
- Track important information related to the interview
  - Outcome of the contact attempt? (e.g., completed interview, demolished household, partial interview, etc.)
  - Were there doorstep concerns? (e.g., hostile, too busy, etc.)
- For this analysis we are interested in **completed interviews** with **doorstep concerns** for the most recent releasable quarters of data (2018 Q1 and 2017 Q4).

Roadmap

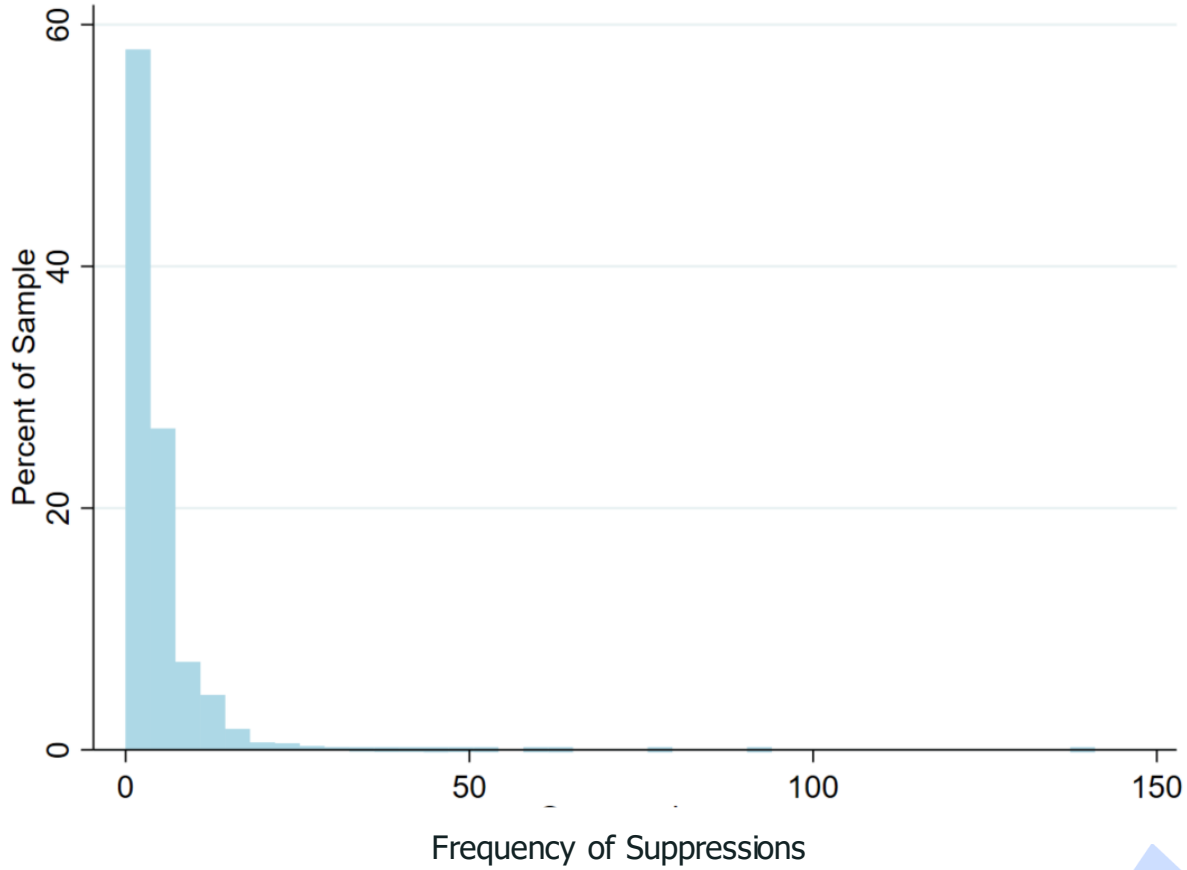
- Introduction
- Data
- Methods
- Results
- Conclusion

Methods: Hypotheses

- In the ideal case, any given door step concern **would not impact the number of edit checks.**
- In reality, interviewers are human so some doorstep concerns may influence the frequency of edit checks. A rushed respondent may try and speed through the interview resulting in more typos.
- Reduction in edit checks is also a plausible outcome. A respondent who was skeptical but convinced by the interviewer may be easier to work with than an apathetic respondent.

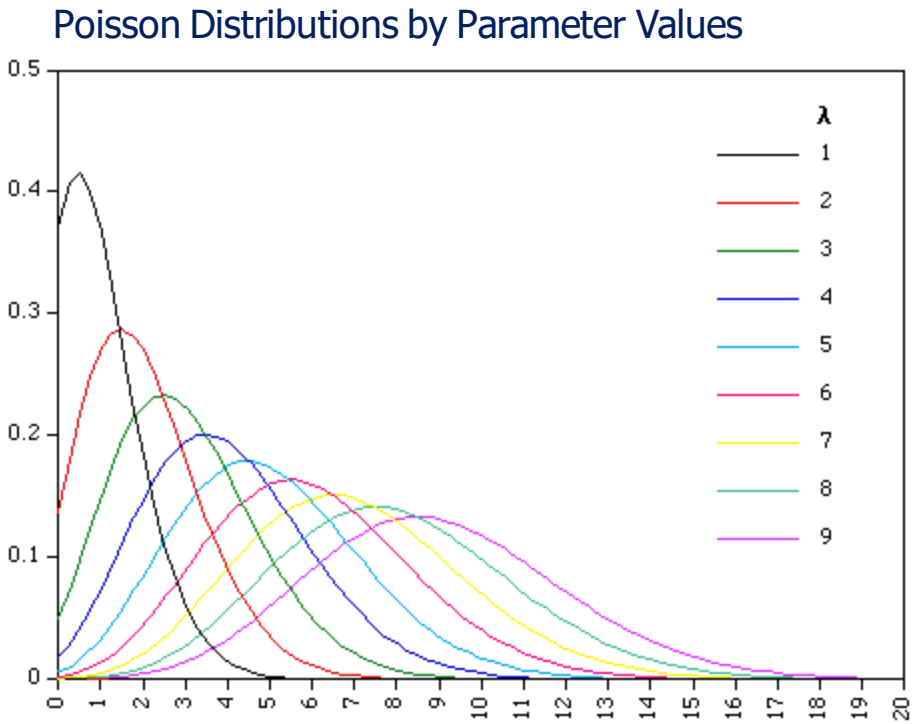
Methods: Define Response Variables

- Total number of edit checks
  - Hard
  - Soft
- Total number of resolutions
  - Suppress
  - Go to
  - Escape

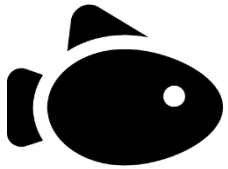


Methods: Poisson Regression

- Count data lend themselves to a Poisson regression
  - Assumption that the response is Poisson distributed
    - Mean and variance are equal
- Potentially overdispersed
  - Greater variability than is expected by the model

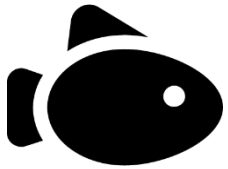


Methods: Poisson Regression



- Overdispersion is present in the response variables due to the excess number of zeros.
- Therefore, the analysis will be modified by using a **zero-inflated Poisson regression**.
  - Deals with the excess zeros by estimating them separately from the Poisson process.
  - Zero-inflated Negative Binomial Model used as a robustness check.
- Household demographics used as controls (e.g., race, region, income, etc.)

Methods: Poisson Regression



- Produce incidence rate ratios (IRR) to determine the multiplicative likelihood compared to the baseline.
  - 1.5 IRR, for example, indicates that the result is 1.5 times as likely compared to the control group.
- Detectable differences vs. Meaningful differences
  - Large samples often lead to over rejection of hypotheses
  - Consider both magnitude and significance of result
    - Detectable IRR of 0.98 doesn't matter much in a practical sense

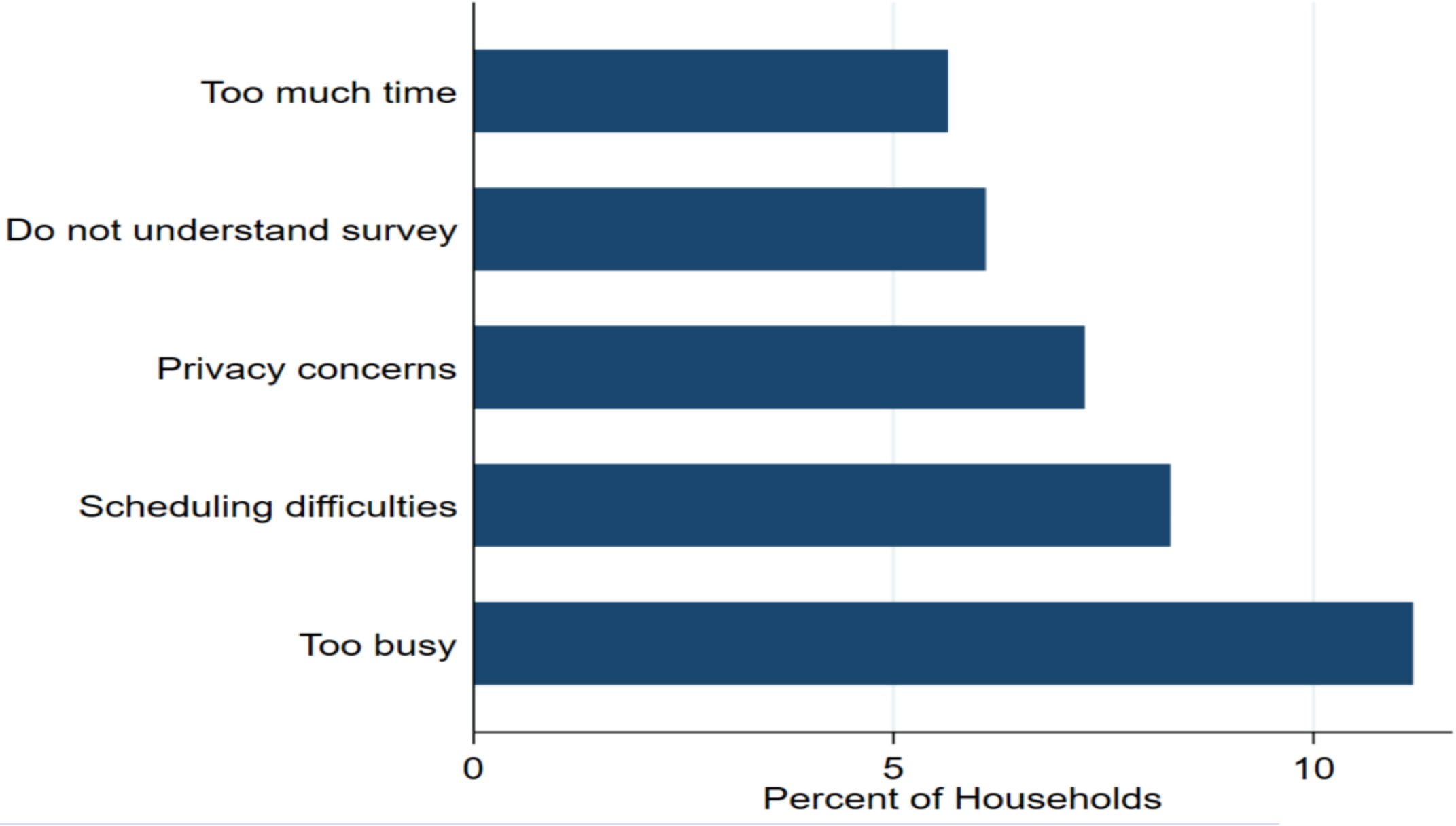
Roadmap

- Introduction
- Data
- Methods
- Results
- Conclusion

Results: Descriptive Statistics

- 75% of interviews reported “No Concerns”
  - There is a small, but statistically detectable reduction in edit checks for households reporting no concerns.
- The most common concern was “too busy” at 11% of the sample.
- The least common concern was “Hostile or threatens FR” at 0.1% of the sample. (A positive thing)
- The average number of edit checks is about 5.2 per interview with a standard deviation of 6.4.

Results: Most Common Doorstep Concerns



Results: IRR for Errors/Edit Check

- Very few doorstep concerns resulted in any changes from the control group.
  - Generally, homogenous interviewer behavior is preferred.
- Only ‘intends to quit the survey’ resulted in a statistically significant increase in frequency of edit checks controlling for demographics. (IRR = 1.4) : Potentially satisficing on the part of the interviewer?
- ‘Not interested’ and ‘Too many interviews’ both had meaningful* IRR < 1

edit_check	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
edit_check						
rspdnt01	.836045	.0598583	-2.50	0.012	.7265846	.9619956
rspdnt02	.983287	.0454023	-0.37	0.715	.8982079	1.076425
rspdnt03	1.061421	.0658361	0.96	0.337	.9399193	1.198628
rspdnt05	1.030921	.0538246	0.58	0.560	.9306448	1.142002
rspdnt06	.8279138	.0578452	-2.70	0.007	.7219595	.9494179
rspdnt07	.8789207	.0519975	-2.18	0.029	.7826941	.9869777
rspdnt08	1.132407	.1085197	1.30	0.194	.9384932	1.366388
rspdnt09	.9175287	.0533586	-1.48	0.139	.8186877	1.028303
rspdnt11	1.177803	.29054	0.66	0.507	.7262709	1.910059
rspdnt15	.9801296	.088912	-0.22	0.825	.8204783	1.170846
rspdnt17	.9415239	.0979018	-0.58	0.562	.7679297	1.15436
rspdnt18	1.133754	.1720325	0.83	0.408	.8420934	1.526433
rspdnt19	.7233275	.0483631	-4.84	0.000	.6344858	.8246089
rspdnt21	1.302019	.1646288	2.09	0.037	1.016226	1.668185
rspdnt22	.8951468	.0508786	-1.95	0.051	.8007804	1.000634
rspdnt23	1.120636	.1006163	1.27	0.205	.939809	1.336256

Reduced frequency of errors	Increased frequency of errors	No deviation from control
Not interested **Too many interviews** **Survey is voluntary** Privacy Concerns	**Intends to quit the survey**	Too busy Interview takes too much time Scheduling difficulties Anti-government concerns Do not understand survey Hang-up/Slams door Family issues Gave information last time Too many personal questions Other

Results: Incidence Rate Ratios for Suppressions

- Since ‘Escape’ is an incredibly sparse solution for dealing with errors (only 9% of all error resolutions), Suppressions were modeled.
- Again, almost all doorstep concerns did not result in any changes from the control.
- ‘Survey is voluntary’ and ‘Too many interviews’ both correlate with a reduced frequency of suppressions at meaningful levels.

suppress	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
suppress						
rspdnt01	.8540425	.066532	-2.03	0.043	.7331095	.9949246
rspdnt02	.9113208	.0376723	-2.25	0.025	.8403964	.9882309
rspdnt03	1.112184	.0664233	1.78	0.075	.9893273	1.250297
rspdnt05	1.049207	.048496	1.04	0.299	.9583348	1.148696
rspdnt06	.8409663	.0616638	-2.36	0.018	.7283905	.9709411
rspdnt07	.8472386	.0494261	-2.84	0.004	.7556983	.9498676
rspdnt08	1.029309	.1005634	0.30	0.767	.8499302	1.246545
rspdnt09	.9663523	.0537255	-0.62	0.538	.8665866	1.077604
rspdnt11	1.062332	.2584358	0.25	0.804	.6594568	1.71133
rspdnt15	.9755133	.0837198	-0.29	0.773	.8244836	1.154209
rspdnt17	.9775344	.1090187	-0.20	0.839	.7856019	1.216358
rspdnt18	1.20376	.1897577	1.18	0.239	.8838096	1.639537
rspdnt19	.7873321	.053801	-3.50	0.000	.6886405	.9001676
rspdnt21	1.172224	.1498204	1.24	0.214	.912472	1.505919
rspdnt22	.9097631	.0456429	-1.89	0.059	.8245622	1.003768
rspdnt23	1.136075	.093978	1.54	0.123	.966038	1.336041

Reduced frequency of Suppressions	Increased frequency of Suppressions	No deviation from control or sample size too small
Survey is voluntary **Too many interviews** Not interested Too busy Privacy concerns		Interview takes too much time Scheduling difficulties Survey is voluntary Anti-government concerns Do not understand survey Hang-up/Slams door Family issues Gave information last time Too many personal questions Intends to quit the survey Other

Roadmap

- Introduction
- Data
- Methods
- Results
- Conclusion

Conclusions

- The majority of door step concerns **are not associated with the number of edit checks or suppressions.**
- Of the top 5 most common doorstep concerns, only “privacy concerns” was significantly associated with edit checks.
- Observing a reduction in errors or suppressions could be a result of interviewer effects (i.e., interviewers who get completed interviews from challenging households may be characteristically different)

Conclusions: Limitations

- Interviewer characteristic data are not available to researchers so they cannot be included in the model.
  - Some interviewers could be better than others at getting respondents to complete the interview.
  - Could be a function of experience, general safety of the area, etc.

- Data analyzed are the most recently available.

Questions?

Contact Information

Taylor J. Wilson
Economist
Bureau of Labor Statistics
www.bls.gov/cex
202-691-6550
wilson.taylor@bls.gov