Using Paradata to Model Total Survey Error in the Current Population Survey

John Dixon

U.S. Bureau of Labor Statistics
June 2-4, 2013

The opinions expressed are those of the author, not the BLS
Nonresponse Bias and Measurement Error

- Nonresponse bias and measurement error both refer to difficult to measure errors in surveys. Nonresponse bias refers to unmeasured persons, and measurement error refers to an unmeasured construct (which is mis-measured by the survey). This will be examined with propensity models based on contact history variables.

- Contact history has the potential to describe the concerns of potential respondents as well as contactability. Those concerns have been found to relate to nonresponse, but little relationship to measurement error. Measurement error will be assessed in the relationship between unemployment and education level (as a surrogate for covariates of measurement error).

- There may also be a relationship between movers and unemployment.
Current Population Survey (CPS)

- Sponsored by Bureau of Labor Statistics
- Data collection by Bureau of the Census
- Primary household labor force survey for the U. S.
- Multi-stage clustered sample of 50,000 households per month
- 4 - 8 - 4 rotation pattern
- 5.5% refusal, 3.5% noncontact for 1st month in sample
CONCERN / BEHAVIOR / RELUCTANCE

Select the categories that describe respondent concerns, behaviors, or reluctance during this contact attempt. Enter all that apply, separate with commas.

- Not interested / Does not want to be bothered
- Too busy
- Interview takes too much time
- Breaks appointment (puts off FR indefinitely)
- Scheduling difficulties
- Survey is voluntary
- Privacy concerns
- Anti-government concerns
- Does not understand survey / Asks questions about the survey
- Survey content does not apply (retired, healthy, no crimes to report)
- Hang-up / slams door on FR
- Hostile or threatens FR
- Other household members tell respondent not to participate
- Talk only to specific household member
- Family issues
- Respondent requests same FR as last time
- Gave that information last time
- Asked too many personal questions last time
- Too many interviews
- Last interview took too long
- Intends to quit survey
- No concerns
- Other - specify
Mean Rates of CHI concerns

Mean Concerns (vertical axis) vs. Concerns (horizontal axis):
- busy
- schedule
- notint
- time
- privacy
- volunteer
- toomany
- question
- member
- issues
- quit
- noshow
- antigov
- hungup
- sameinf
- toopers
- too long
- notapp
- hostile
- othhh
- samefr
Employment by Time in Sample

Unemployment rate by time in sample

T1 T2 T3 T4 T5 T6 T7 T8
Employment by Time in Sample

Differences in Unemployment rate by time in sample

- T1-T2
- T2-T3
- T3-T4
- T4-T5
- T5-T6
- T6-T7
- T7-T8

The diagram shows the differences in unemployment rate by time in the sample. The highest difference is observed between T1-T2, followed by T5-T6 and T6-T7. The differences are relatively small between T3-T4 and T4-T5, and the smallest difference is between T7-T8.
Unemployment change and education

<table>
<thead>
<tr>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.3</td>
<td>38.4</td>
<td>38.5</td>
<td>38.6</td>
<td>38.7</td>
<td>38.8</td>
<td>38.9</td>
</tr>
</tbody>
</table>

- **NotInLaborForce to Unemployed**
- **Unemployed to NotInLaborForce**
Moving by time in sample

Moved

T1 T2 T3 T4 T5 T6 T7 T8
Movers and unemployment

Stayed

Moved

T2 T3 T4 T5 T6 T7 T8
Predicting Nonresponse
Logistic model Coefficients

Bivariate Models

- HUNGUP
- NOTAPP
- TOOLONG
- SCHEDULE
- PRIVACY
- TIME
- HOSTILE
- NOTINT
- MEMBER
- QUESTION
- SAMEFR
- BUSY
- OTHHH
- SAMEINF
- VOLUNTAR
- NOSHOW
- ANTI
- ISSUES
- TOOPERS
- ISSUES
- TOO_MANY
- QUIT
Predicting Nonresponse Logistic Model Coefficients
Factor Pattern for Contact History Concerns

![Factor Pattern Chart]

- Hostile
- Time
- Busy
- Privacy

Legend:
- Hostile
- NotInt
- Hangup
- Samefr
- Time
- SameInf
- TooPers
- TooMany
- TooLong
- Quit
- Busy
- NoShow
- Schedule
- Privacy
- AntiGov
- Question
- NotApp
- OthHH
Nonresponse bias

Refusal
Noncontact
Nonresponse:

- The CHI data was useful in modeling the relationship between concerns expressed by respondents and refusal/noncontact.
- The CHI data showed factor patterns which could describe broad areas of concern. They related well in predicting nonresponse.
- The propensity models indicated very slight nonresponse bias.

Moving had a very small effect on bias.

Measurement error had an even smaller effect on bias, and wasn’t supported by the patterns of change in unemployment.
Limitations and Future Research

- The CHI data is limited in that it only reflects the concerns expressed by respondents. Some of the most common concerns may mask the real reasons, for example, “busy” may hide concerns about privacy, which weren’t expressed to the interviewer.
- Other models for measurement error may be needed, such as “stayer/mover” Markov Chain Monte Carlo latent class models.
- Replicating the models with another survey may help make the model more general, and give different insights which would help with the CPS.
- Put all the pieces into a structural equation model to better estimate the relative and combined effects on rotation group bias.
Contact Information

Using Paradata to Model Total Survey Error in the Current Population Survey

John Dixon
202-691-7516
dixon.john@bls.gov