

Using Respondent Verbal Paradata to Predict Income Nonresponse:

How They Say It Can Predict What They'll Say

Matt Jans
US Census Bureau
Statistical Research Division

Acknowledgments

- **Funding:**
 - Charles Cannell Fund in Survey Methodology
 - US Census Bureau Dissertation Fellowship
- **Data:** Richard Curtin and Rebecca McBee at SCA
- **Dissertation Committee:** **Frederick G. Conrad**, Co-Chair, **James M. Lepkowski**, Co-Chair, Norbert Schwarz, Jose R. Benki, Frauke Kreuter, Floyd Jackson Fowler, Jr.
- **Coders:** Maya Burns, Alex Dopp, Zach Hartley, Clare Levijoki, Lian Liu, Melinda Mosher, Miyuki Nishimura, Katie Singer, Jenna Stein, Travis Pashak

Verbal Paradata

- Byproducts of data collection (Couper 1998)
- *How* an answer is delivered provides a lot of verbal paradata
 - Pauses
 - Rate of speech
 - Pitch
- Requests for clarification
- Potential predictors irrespective of mechanism

Income Nonresponse

- As high as 50% (Moore, Stinson, & Welniak, 2000)
- 25-30% common (Atrostic & Kalenkoski, 2002)
- Bias if response rate differs across income groups
- Analytic annoyance regardless

Why do Respondents not Provide Income?

- Income value may be *sensitive*...
 - Intrusiveness, threat of disclosure, social desirability

Why do Respondents not Provide Income?

- Income value may be *sensitive*...
 - Intrusiveness, threat of disclosure, social desirability
- Remembering and calculating income may be *cognitively complex*...
 - Comprehension, retrieval, judgment and mathematical calculation, report

Why do Respondents not Provide Income?

- Mechanisms
 - Affect (i.e., emotion)
 - Cognitive difficulty

Verbal Paradata as Measures of Affect

- Voice acoustics
 - Pitch and pitch variability (Bachorowski, 1999)
- Question answering
 - Refusals

Verbal Paradata as Measures of Cognitive Difficulty

- Indicators of cognitive difficulty
 - Speech
 - Pauses, fillers, repairs (Schober & Bloom, 2004; Conrad, Schober, & Dijkstra, 2008)
 - Question answering
 - Requests for repeat or clarification, reports, don't knows (Mathiowetz, 1999; Schober & Bloom, 2004)

Related Measures of Psychological Processes

- Direct ratings of affect and difficulty
 - Affect valence
 - Affect intensity
 - No, Medium & High cognitive difficulty

This Study

Utterance-level Verbal Paradata

- Utterance:

I: What was your income in 2009?

R1: Well, that's a tough one.

R2: What do you mean by income?

Utterance-level Verbal Paradata

- Analyzed at question-level
 - Proportion of utterances having a certain behavior
 - e.g., proportion of utterances with a request for clarification
 - Average per utterance
 - e.g., average cognitive difficulty rating at question
 - Presence at the question
 - e.g., presence of request for clarification

Types of Coded Variables

- Question Answering Behavior
 - e.g., answering with qualification
- Natural Speech
 - e.g., pauses and fillers (*ums* and *uhs*)
- Voice Pitch
- Coder Ratings
 - Affect
 - Cognitive difficulty

Income Nonrespondent Types

- Household income in past calendar year
- Income Nonrespondents
 - No income value or bracketed value given

Income Nonrespondent Types

- Household income in past calendar year
- Income Nonrespondents
 - No income value or bracketed value given
- Bracketed Respondents
 - Do not provide a dollar value, but provide a bracketed response

Income Nonrespondent Types

- Household income in past calendar year
- Income Nonrespondents
 - No income value or bracketed value given
- Bracketed Respondents
 - Do not provide a dollar value, but provide a bracketed response
- Dollar Amount Respondents
 - Provide a dollar amount

Respondents

- 185 completed interviews from the Reuters/UM Surveys of Consumers monthly household survey

<u>Income Nonrespondent Type</u>	<u>Number of Respondents</u>	<u>Percent of Sample</u>
Income Nonrespondent	60	32.4%
Bracketed Respondent	56	30.3%
Dollar Amount Respondent	69	37.3%
Total	185	100.00%

NOTICE!!!

- Data **before** the income question only

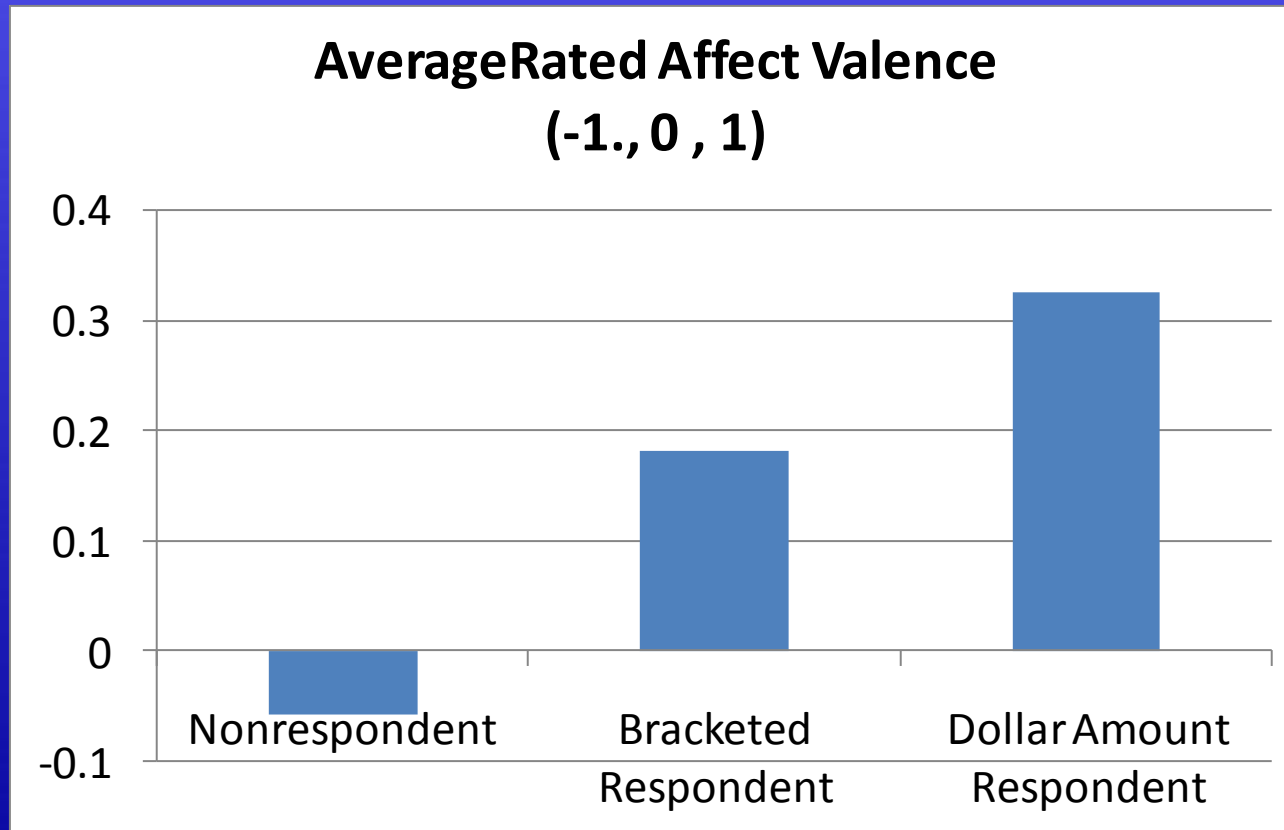
Nonrespondent Type Analysis

- One-way ANOVA
 - Differences between nonrespondent types
- “Nonrespondent type” is a fixed status (independent variable)

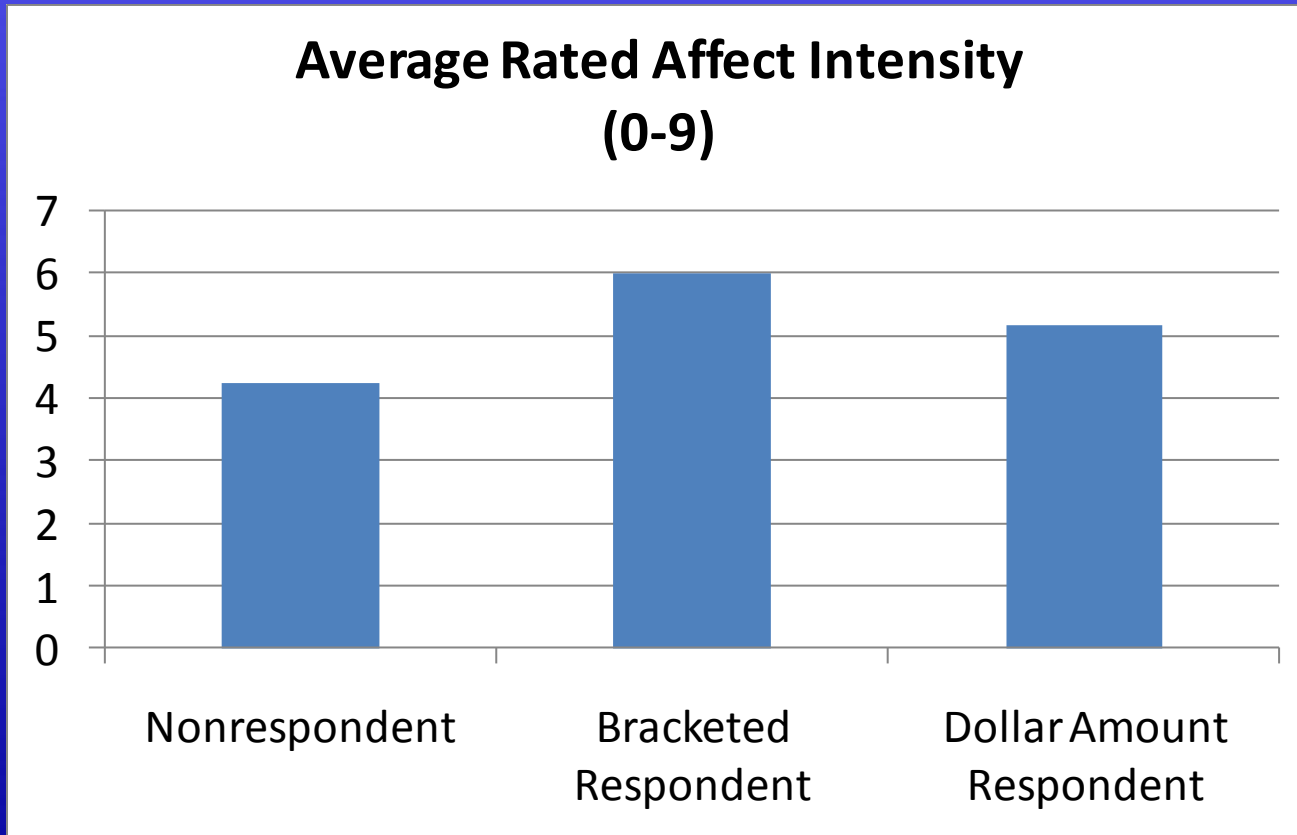
Differences between Nonrespondent Types

- Data before the income question show...
 - Nonrespondents distinguished by lower affect than bracketed respondents or dollar amount respondents
 - Negative comments
 - Affect intensity
 - Affect valence

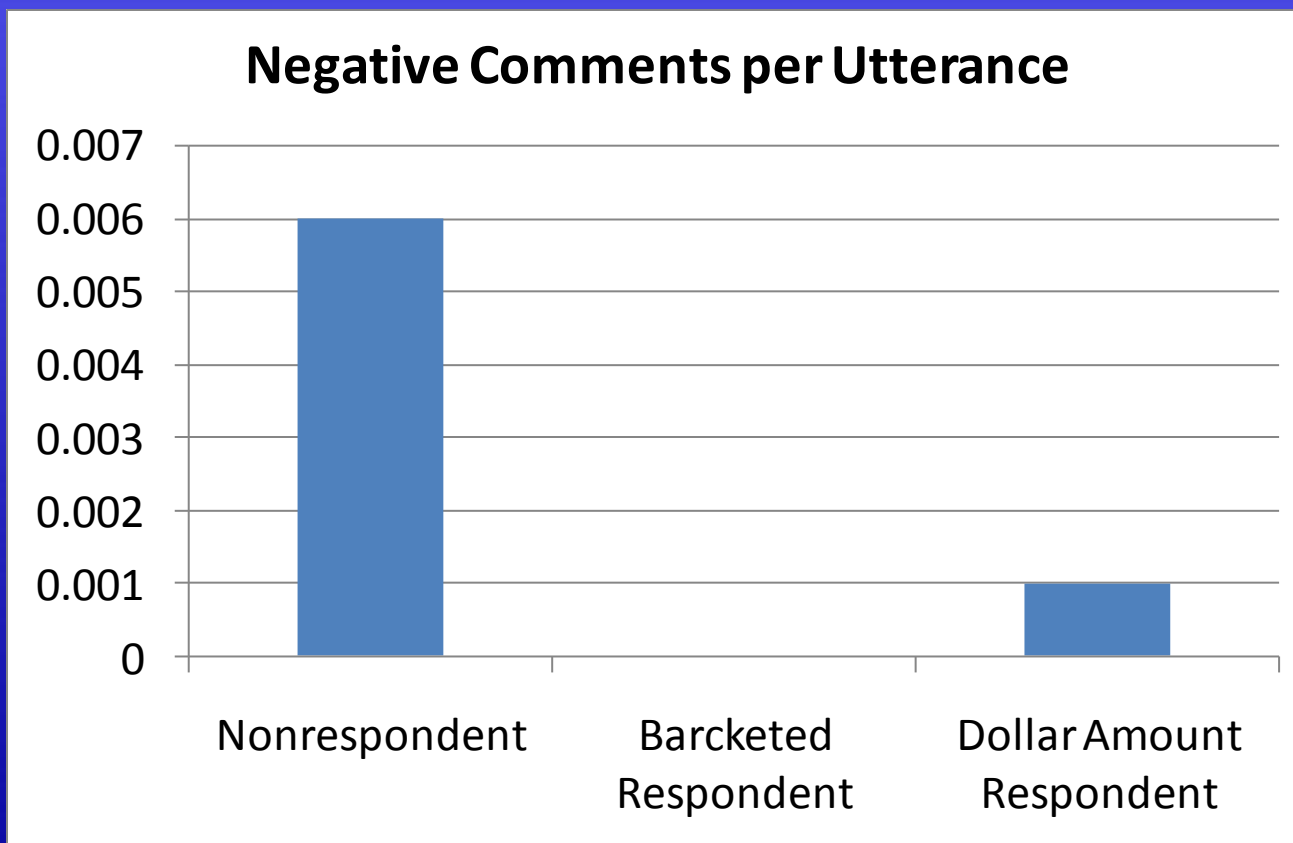
Affect Valence



Affect Intensity



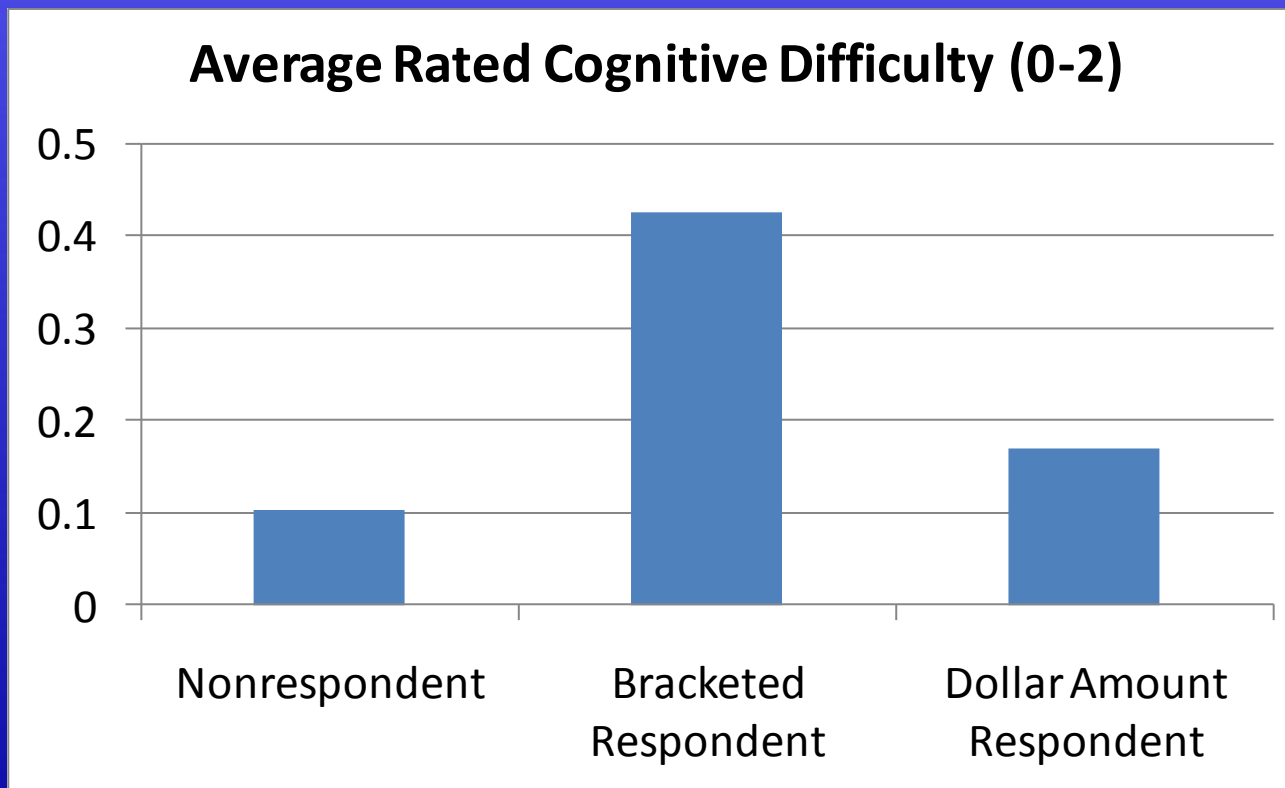
Negative Comments



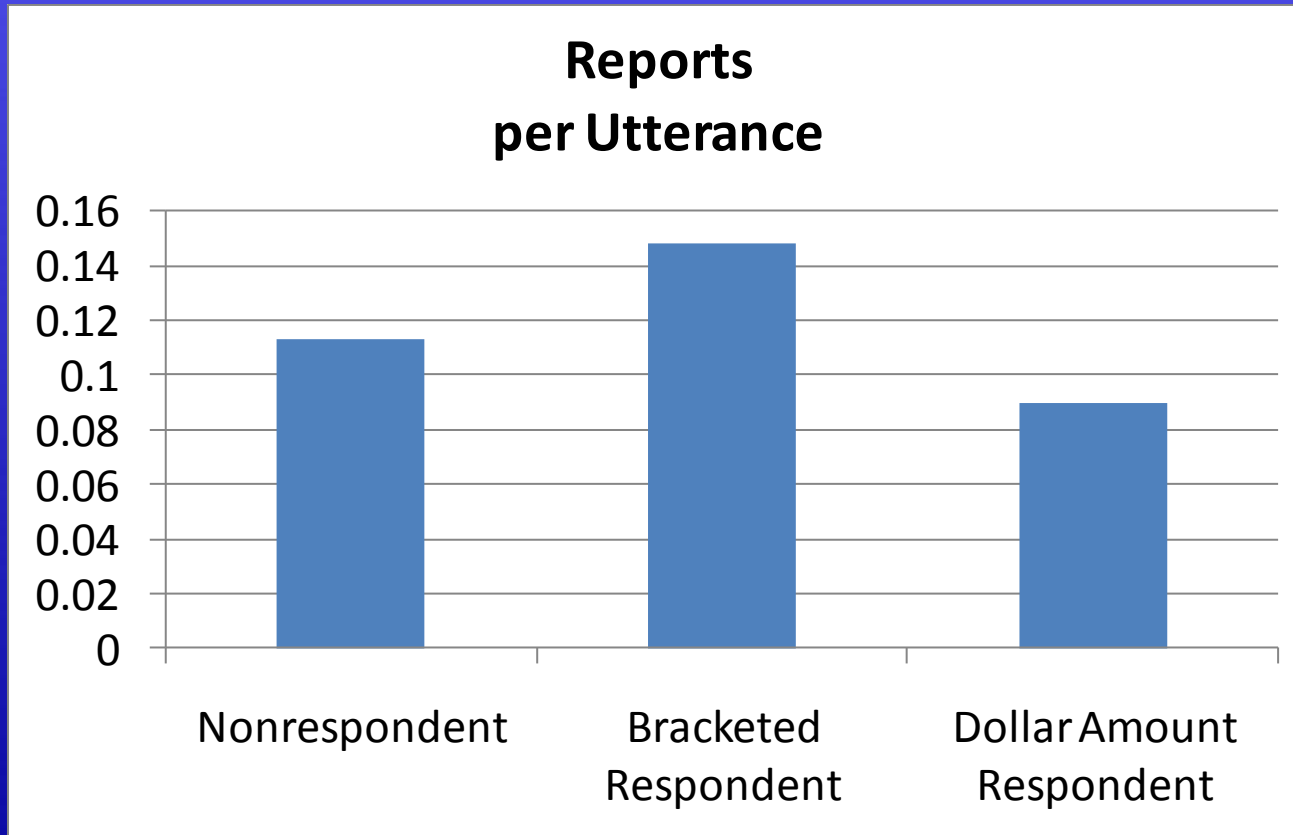
Differences between Nonrespondent Types

- Data before the income question show...
 - Bracketed respondents show more cognitive difficulty than nonrespondents or dollar amount respondents
 - Rated cognitive difficulty
 - Reports
 - Digressions

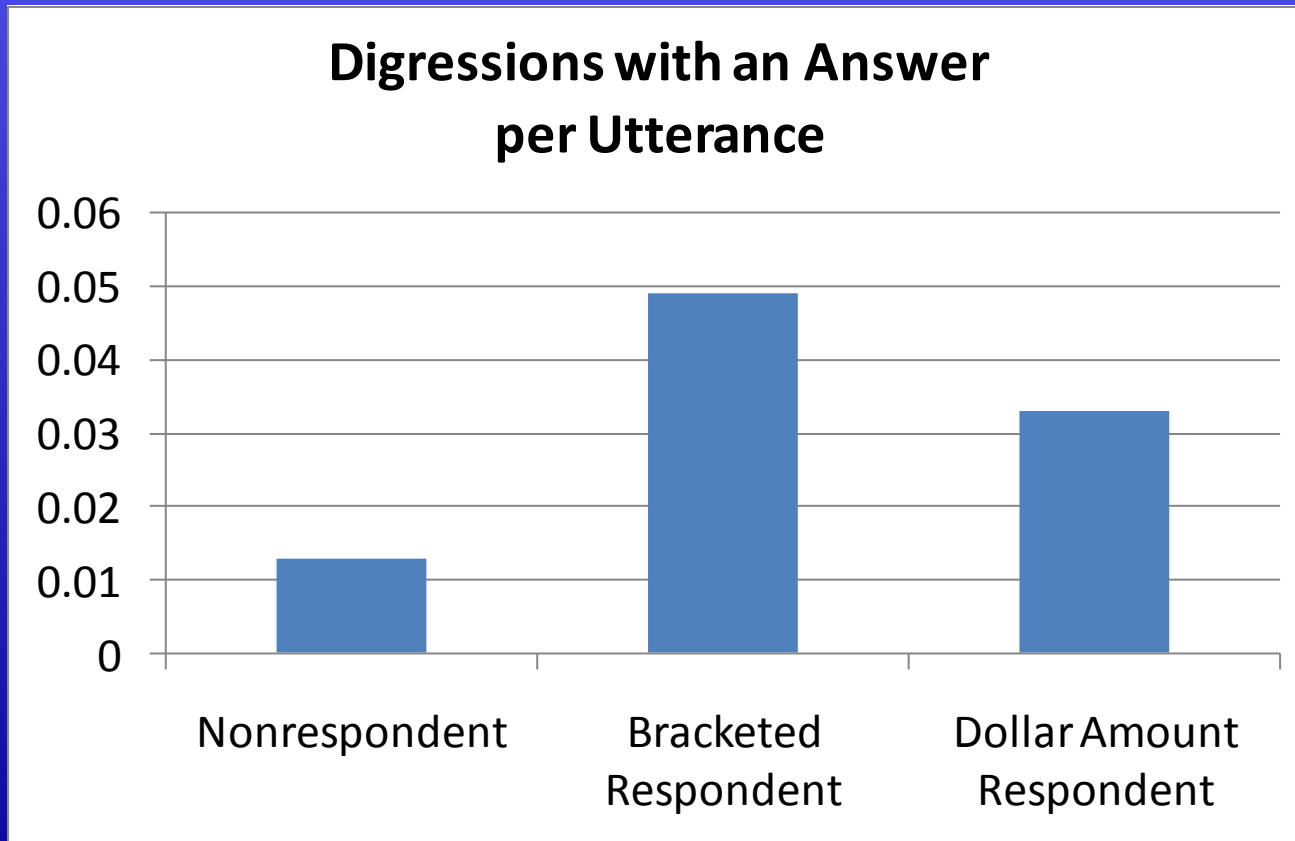
Rated Cognitive Difficulty



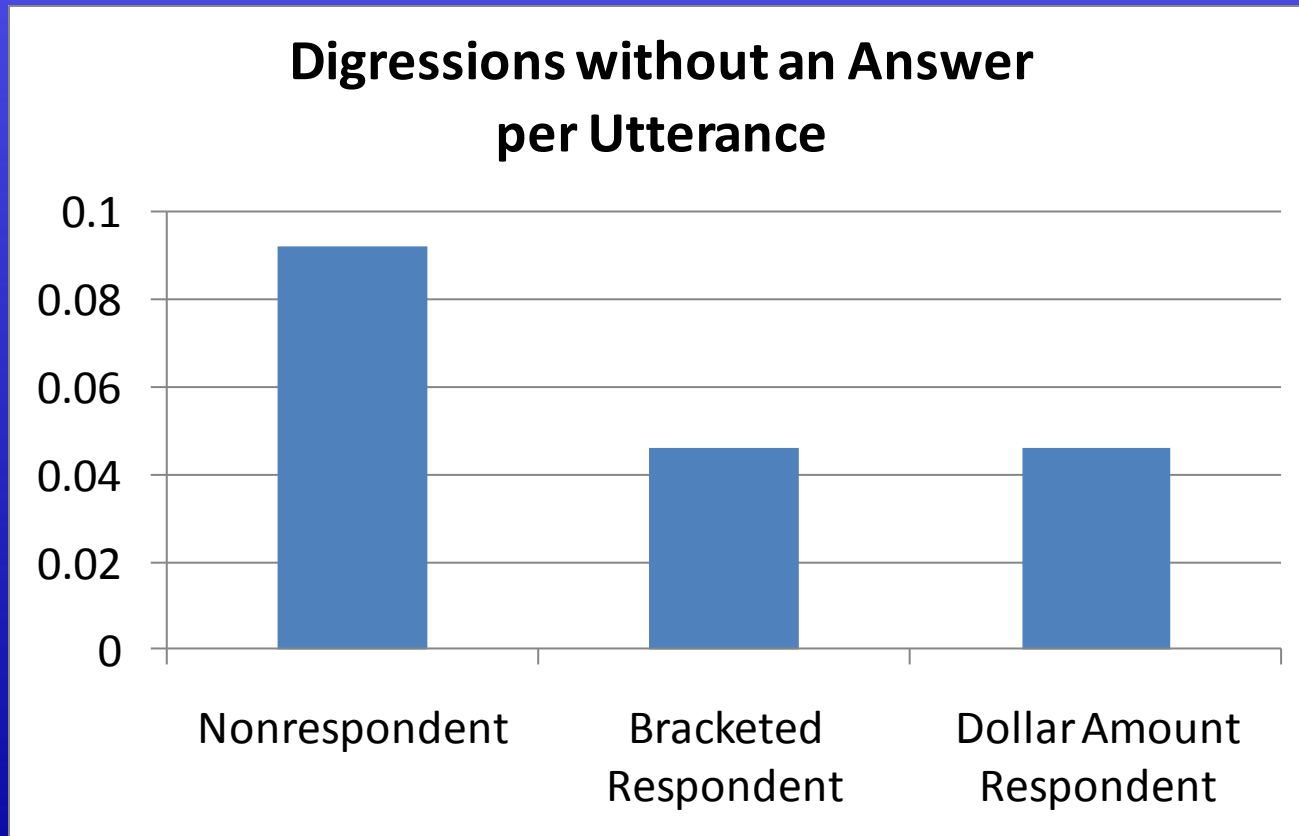
Reports



Digressions



Digressions



Predicting Nonresponse

- Multinomial logistic regression predicting nonresponse
- Reverses prediction (NR = indicators)
- 739 cases (one per Q per R)
- Standard errors corrected for nesting
- Several indicators recoded as presence at the question

- Similar results to ANOVA

Predicting Income Nonresponse

Indicator	Estimate	S.E.	p-value	Odds Ratio
Backchannels	-0.995	0.472	0.035	0.370
Affect Intensity	-0.409	0.070	< 0.0005	0.664
Affect Valence	-0.774	0.173	< 0.0005	0.461
Overspeech	0.098	0.032	0.002	1.103
Negative Comments	2.387	1.137	0.036	10.886
Request for Clar'n or Repeat	0.613	.308	.047	1.847
Reports	0.735	0.286	0.01	2.116
<u>Borderline Significant Predictors</u>				
No Difficulty	1.113	0.591	0.060	3.042
Don't Know	0.650	0.373	0.081	1.916

Predicting Bracketed Response

Indicator	Estimate	S.E.	p-value	Odds Ratio
Pitch at last 50 ms Voicing in First Utterance	0.002	0.001	0.035	1.002
No Difficulty	-1.173	0.482	0.015	0.309
Reports	0.657	0.251	0.025	1.930

Comparing Analytic Methods

ANOVA	Multinomial Model
Negative Comments & NR	Negative Comments & NR
Rated Affect & NR	NO
Cognitive Difficulty & Brackets	Cognitive Difficulty & Brackets
NO	Cognitive Difficulty & NR
Reports & Brackets	Reports & Brackets, NR

Predicting Income Nonresponse

- Mechanisms leading to nonresponse
 - Income nonresponse more affective
 - Bracketed response more cognitive
- Differences in affect and difficulty begin **before** the income question
- Nonresponse predicted before income, bracketed response at the income question

Future Work

- Model the multivariate factor structure that describes verbal paradata and predicts income nonresponse (SEM)
- Explore the order of verbal events within questions (i.e., what follows what?)
- Inclusion of interviewer data

Implications and Applications

- Interviewer interventions based on perception of respondent verbal paradata
 - Can interviewers perceive signs of upcoming income nonresponse and intervene to increase response
- Imputation models using verbal paradata
 - If paradata predict income nonresponse, they can be applied to imputation for missing income data (Mathiowetz, 1999)

Implications for Practice

- Can interviewers (like coders) listen for verbal paradata and intervene to produce better quality income data?

Thank you

- Questions?
- matthew.e.jans@census.gov