Measurement Error in Contact Para data and its Relationship to Response Propensity.

John Dixon

Bureau of Labor Statistics
2 Massachusetts Ave, N.E.
Washington, DC 20212

Abstract

Propensity scores are proving very useful in studies of nonresponse error (e.g., Schouten and Leufkens, 2010 ITSEW). The Contact History Instrument (CHI), used in U.S. Census Bureau demographic surveys, provides interviewer-assigned outcomes resulting from each contact attempt, including a long list of reasons for noncontact, reluctance or refusals, which can be used to generate propensity scores. The interviewer selects the most appropriate reasons for nonresponse or noncontact, but many of these reasons for noncontact or refusal are related (e.g., "Too busy" and "Scheduling difficulties" are both related to time issues). Many reasons given for nonresponse from paradata could be reduced to a smaller set of reasons for noncontact and refusal. Dixon (nonresponse workshop 2010) found four factors related to refusal, similar to Maitland et al. (2008 Joint Statistical Meetings). The factors were named "cooperation", "time concerns (busy)", "privacy", and "gatekeeper issues".
Many of the reasons given for refusal may have measurement error (Bates et al., 2010 Joint Statistical Meetings). For example, some potential respondents might find "too busy" to be a more socially acceptable reason rather than "privacy concerns" or "anti-government sentiment". This idea of measurement error is different from other concepts of measurement error which refer to incorrect data. The measures are collected accurately; they represent different things. The present study will attempt to tease out the different subgroups of respondents where the meaning of the concerns could relate to other reasons. A bi-clustering procedure (in the R library) will show different subgroups of respondents which respond differently to a common clustering of reasons. These reasons could predict different propensity scores for the subgroups of respondents where their meaning differs, which hopefully will better predict nonresponse than the simpler model which assumes the measures are consistent for everyone.

Key Words: Nonresponse, Contact History Instrument, Measurement error.

1. Introduction

The present study uses reluctance concerns from the Contact History Instrument (CHI) and other paradata to explore the experience of the respondents with the Current

Population Survey (CPS). Nonresponse propensity models are used to study nonresponse bias in surveys. Paradata is one of the best predictors of nonresponse, and respondents' reasons for not responding is a very relevant form of paradata.

This paradata 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.

This study hopes the patterns of concerns might relate to different expressions of concern. For example; "Cooperation" and "Privacy" might mean something very different from "Cooperation" and "Busy". Cooperation might be more of a style of expression, rather than relating to a willingness to respond.


2. Data Sources

The Current Population Survey (CPS) is a source of estimates for the unemployment rate. Details about the CPS can be found in Technical Paper 66 (Census, 2006). The CPS is the primary source of information on the labor force characteristics of the U.S. population. The CPS uses a multistage probability sample based on the population counts from the decennial Census. The proportion of sample households not interviewed in the CPS due to non-contact or refusals typically varies between eight and ten percent. Data may be collected either in person or by telephone, although the first and fifth interviews are supposed to be in person. This study does not consider households where data are collected by telephone centers (CATI, about 11%), but does consider those where the field interviewer chooses to collect data by telephone. The CPS has information on education, which is used as an indicator of potential measurement error. Other studies have found education level to be one of the better predictors of measurement error, although a more sophisticated model could be used in a future study.

The CHI was designed to collect information about each contact attempt made by a field representative (FR); including information about why respondents refuse (Dyer, 2004). CHI was added to the CPS in 2009 to collect detailed contact history data (Bates, 2004). The interviewer records times and outcomes of attempted contacts, problems or concerns reported by reluctant households, and strategies used to gain contact or overcome reluctance. This provides a very rich source for studying the interview process since the data is available for both responding and non-responding households. However, this study only used the answers recorded by interviewers in response to a question about reasons for not responding reported by reluctant households (those who expressed concerns). Data from 2009 through 2013 were combined to provide concerns and employment status studied here, with 482535 individuals. Only the first interview was used.

Geographic characteristics may influence how a respondent may express concerns, depending on the culture of the area. The areas identified by the anthropologist Colin Woodard will be used to represent those differences in culture. While he identified different areas based on history and politics, I hope those cultural differences might influence how concerns about the survey are expressed.

Geographic Area characteristics
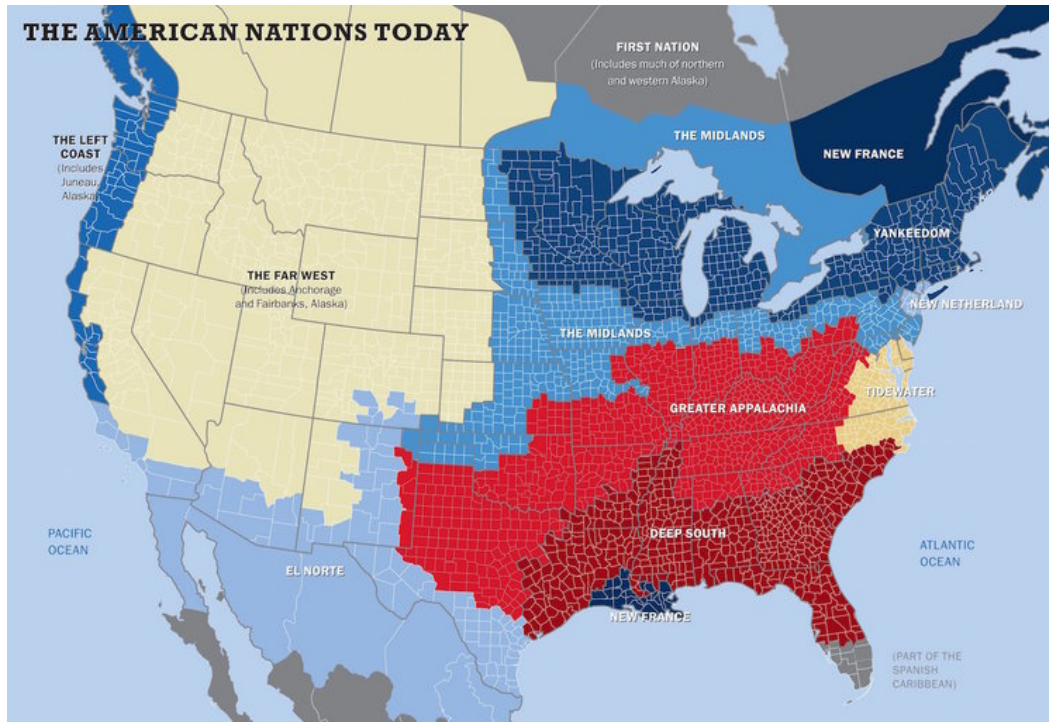
The idea is that cultural differences may contribute to differences in expressing concerns. The American Nations study is an anthropological typology of the United States based on historical and political differences. The county assignments to the different groups were merged with the current data, so each household was assigned to a group.

3. Methods

Logistic models were used to predict refusal using the CHI concern data as predictors. The predicted values from those models serve as the propensity to respond. Three logistic models will be used. The first only uses the concerns from the CHI. The second adds the clustering from the biclust procedure. The third includes the CHI concerns, the clusters, and adds the geographic characteristics. For each of the models, propensity scores will be created for the entire sample (both respondents and nonrespondents). Only respondents will be used to evaluate bias (since they are the only ones with measures from the survey). Those respondents who were most like the nonrespondents were used as substitutes for the nonrespondents, and the comparison of the two groups give a measure of potential nonresponse bias. The groups were assigned, so the proportion representing nonrespondents is the same as in the original sample.

Wikipedia defines biclustering as; "**Biclustering**, **block clustering** , **co-clustering**, or **two-mode** **clustering** is a data mining technique which allows simultaneous clustering of the rows and columns of a matrix."

In this study, the concerns recorded by the interviewers are the columns, and the respondents are the rows. Where patterns are consistent, there would be a single cluster. Where the patterns are different, separate clusters with similar patterns would be formed. It is hoped that those different patterns might relate to different meanings of the concerns expressed. By including the cluster groups in the logistic model, the different meanings would be controlled in producing propensity scores. These different meanings might relate differently to the propensity to respond, and so could be considered a source of error if not included in the model. The improvement in the model fit would be used to evaluate the effect of the clustering.

4. Results

I will describe the CHI descriptives, the biclust results, the geographic characteristics, and then the logistic models which put them all together.

In the CHI, the most common concern expressed by respondents during the first interview was "busy", followed by "schedule difficulties", and "not interested". Other notable concerns were "time the interview takes" and "privacy concerns". Since many of the categories are low frequency, this is a challenge for previous factor analysis. However, it did not prove to be a problem in creating the propensity scores.

The Biclust procedure produced three column clusters of concerns; Cooperation, Busy, and Privacy. The characteristics of households in the row clusters were;
One large cluster, with few if any concerns expressed.
Pure clusters (which wouldn't provide any indications of measurement differences);
   Cooperation, which had more females and persons who were not-in-labor-force).
   Privacy, which had more persons who were not-in-labor-force.
Mixed clusters (where the meanings may differ from the pure clusters);
   Cooperation-Busy, which had more Hispanics and unemployed.
   Privacy-Cooperation, which had more females and Hispanics)
   Cooperation-Privacy-Busy, which had more employed.

Geographic Characteristics (American Nations) provided a cultural reference for the households. The refusal rates were highest in the DC area, Left Coast, and New Netherland. The rates were lowest in Deep South, New France, Spanish Caribbean (Miami), and Tidewater. The most common concern was Privacy (Deep South, Far West, and Tidewater). Busy also combined with Privacy (El Norte, Greater Appalachia, Midlands, New France, Spanish Caribbean (Miami), and Yankeedom. Cooperation concerns were added in DC, Left Coast, and New Netherland. There weren't any areas which had only Cooperation or Busy concerns without Privacy, and the areas which had Cooperation responses also had higher rates of both Privacy and Busy concerns.

I speculate that the cultures in the Deep South, Far West, and Tidewater may prefer to respond to survey requests with privacy concerns, rather than Busy, which may be considered rude.

Three logistic models were run to try to investigate the effect of concerns, different groups based on patterns of concerns, and area differences relating different cultures within the US.
1. Refuse = CHI factors
    a. R-square: 0.2528
    b. bias 0.0521877-0.0525140=-0.0003263 or (-.6%)
2. Refuse = CHI factors and clusters
    a. R-square: 0.3148
    b. bias 0.0521877-0.0524981=-0.0003104 or (-.6%)
    c. No clusters proved significant.
3. Refuse = CHI factors and clusters and areas
    a. R-square: 0.3179
    b. bias 0.0521877-0.0524122=-0.0002245 or (-.4%)
    c. DC, Left Coast, Midland, and New Netherland > refusal

Bias was estimated by dividing the sample to produce a group most like the refusers based on the concerns expressed in the Contact history instrument. The propensity scores from each of the three models were used to produce a sample of similar proportions to the nonresponse rate (for this sample, around 5%). The difference in unemployment between the overall population and the same sample with the "refusers" removed gave an estimate of bias. Each of the models gave estimates of small bias, with the more complicated models giving smaller bias estimates.

   The model with the cluster groups improved the model fit (from an R-square of 0.2528 to an R-square of 0.3148). None of the clusters showed a significant effect relative to the large overall cluster.

   The model with the American Nations areas slightly improved the model fit (R-square of 0.3179). With "Yankeedom" as the reference group; DC, Left Coast, Midland and New Netherland had a greater likelihood of refusal. Deep South and Miami were associated with lower likelihood of refusal.

5. Discussion

   The CHI data were useful in modeling the relationship between concerns expressed by respondents and refusal (based on the R-squares). The resulting propensity models indicated very slight nonresponse bias.

   Measurement differences didn't relate as well to nonresponse as expected. While the clusters improved the models, it didn't improve as much as I hoped. Since the clusters were small relative to the entire sample, this may mask some of the effects. With surveys which have higher nonresponse (and more households expressing concerns) this method may be more useful. I expected differences in education level between clusters (which might produce differences understanding or differences in expressing concerns). There weren't any education differences.

Cultural differences based on geographic areas showed small effects. While it's interesting to speculate on what might be going on in the mind of respondents, the cultural typology was designed on historical and political differences. There may be interactions with other factors, such as how mobile the population is, the population density, and other sociological variables. The small improvement in the models when geographic area indicators are added makes the slightly smaller estimate in bias difficult to accept as a generalizable effect.

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, and may not have shown up in the different patterns of concerns. Another limitation on the conclusions are the assumptions behind the propensity models. In other words, the CHI relates to nonresponse, and those who had similar concerns could represent nonrespondents' employment status.

For future studies different clustering methods may produce different clusters; further work might produce clusters better related to the measurement issues of interest. The logistic models might be improved by allowing random effects. Other research (Dixon, 2013) had found "Cooperation" concerns were related to subsequent response, rather than nonresponse. Random effects would allow the relationship to vary between clusters.

References

BLS Handbook of Methods, http://www.bls.gov/opub/hom/
Bates, N. (2004). Contact Histories: A Tool for Understanding Attrition in Panel Surveys., A
      paper presented at the 59th Annual AAPOR Conference, May 11-13
Bates, N., Dahlhamer, J., and Singer, E. (2008). "Privacy Concerns, Too
      Busy, or Just Not Interested? Exploring the Relationship Between Doorstep Concerns
      and Nonresponse Bias", *Journal of Official Statistics*, Vol. 24.4, December.
Dahlhamer, J., Simile, C., and Taylor, B. (2006). "Exploring the Impact of
      Participant Reluctance on Data Quality in the National Health Interview Survey (NHIS)",
      paper presented at the Statistics Canada Symposium; Methodological Issues in Measuring
      Health.
Dixon, J. (2005). "Comparing different models of nonresponse bias using
      Census/Survey match data" presented at the International Workshop on Household
      Survey Nonresponse in Stockholm, Sweden.
Dixon, J. (2010). "Assessing Nonresponse Bias and Measurement Error Using
      Statistical Matching", presented at the Joint Statistical Meetings.
Dixon, J. (2013), "Using Paradata to Understand Panel Effects in the Current Population Survey",
      Presented at the Joint Statistical Meetings.
Dyer, W. (2004). Contact History Instrument (CHI).
      http://www.blaiseusers.org/2004/papers/03.pdf
Henley, M., and Bates, N. (2006). "Using Call Records to Understand Response
      in Longitudinal Surveys", paper presented at the American Association for Public
      Opinion Research.
Maitland, A., Casas-Cordero, C. and Kreuter, F. (2009). An evaluation of
      Nonresponse bias using paradata from a health survey. Pp. 2250-2255 in Proceedings of
      the Section on Survey Research Methods. Alexandria, VA: American Statistical
      Association.
U.S. Census Bureau. 2006. *Current Population Survey: Design and Methodology*.

Technical Paper 66. Washington, DC: Government Printing Office.
http://www.census.gov/prod/2006pubs/tp-66.pdf
Woodard, Colin, "American Nations", (2011) Penguin Books.