Using Paradata to Understand Panel Effects in the Current Population Survey

John Dixon

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

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

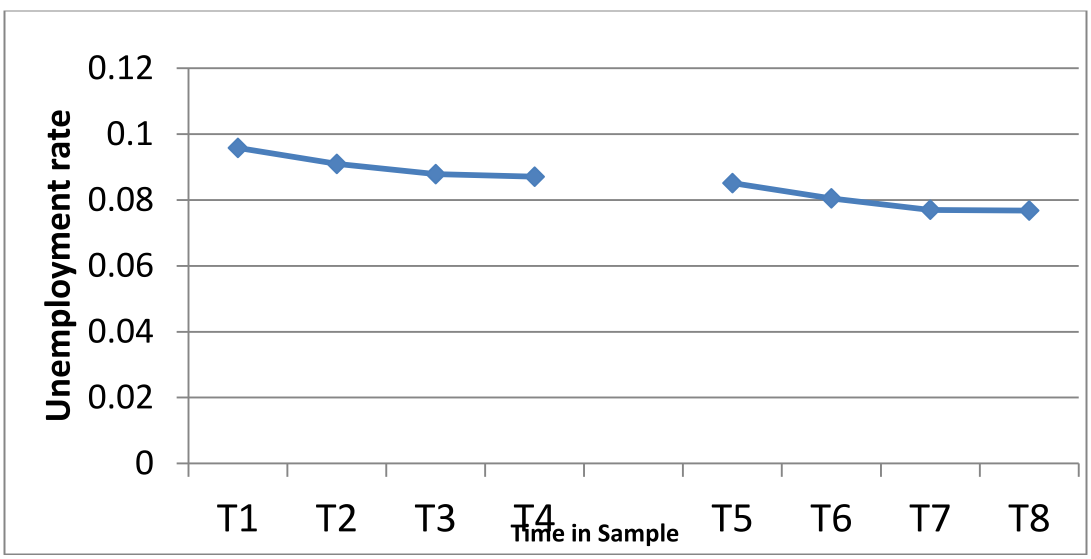
This study examines panel effects (rotation group bias in unemployment estimates) using paradata from the Current Population Survey (CPS). The CPS is administered to a household for four consecutive months, followed by a break of eight months, and then interviewed another four consecutive months. These eight interviews (panels) form the basis of this analysis. There are multiple theories that might explain an observed decline in unemployment rates over the waves, which this study explores. Paradata, including respondent contact history recorded by interviewers, may help understand those effects.

Key Words: Nonresponse, longitudinal survey, Contact History Instrument, panel effects, rotation group bias

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). I wanted to explore the panel effect of rotation group bias in the Current Population Survey. The most studied bias is the decline of the unemployment rate over the time in sample (Bailar, 1975; Solon, 1986; Erkens, 2012; Mansur and Cheng, 2012; Cheng, Larsen, and Wakim, 2013).

Figure 1: Unemployment rate by time in sample for 8 panels.



One of the concerns with repeated interviews is when the estimates vary in a pattern that suggests bias. The decline in unemployment (Figure 1) is nearly 2 percentage points from the first to the last time in the sample.

There is no economic reason for the decline in unemployment over the time in sample to happen. It has been hypothesized that it could be due to a number of possible reasons. The four hypotheses considered in this article are:

1. Attrition: unemployed leaving the sample at a higher rate (nonresponse bias propensity models are used to study this effect).
2. Moving: Those who move could similarly affect the unemployment rate if those who move out were unemployed and those who moved in were employed.
3. Measurement error: over the course of eight waves, respondents may differentially understand, or learn, the concepts of unemployment and "not in the labor force". Thus, they change their status in later interviews, based on this understanding.
4. Proxy reporting: There may also be a Hawthorne effect (Mayo,1993), where the act of interviewing changes the behavior of the respondents.

2. Data Sources

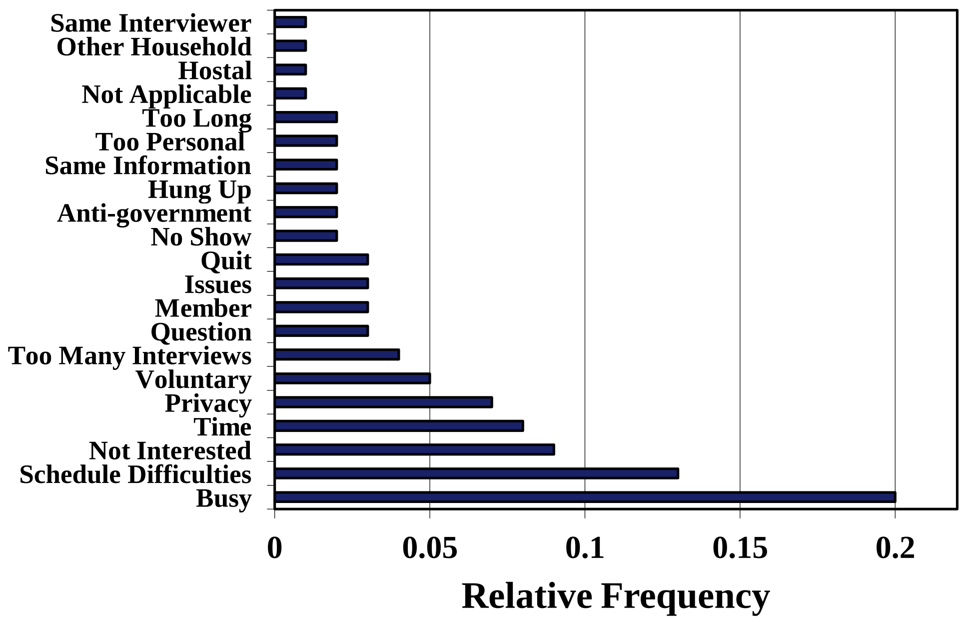
The CHI was designed to collect information about each contact attempt made by a field representative (FR); including information about why respondents refuse (Dyer, 2004). The Current Population Survey (CPS) is a source of estimates for the unemployment rate. Details about the CPS can be found in Technical 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.

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. However, this study only used the answers recorded by interviewers in response to a question about reasons for not responding reported by reluctant and difficult to contact households. Data from 2009 through 2011 was matched longitudinally to provide the changes in concerns and employment status studied here, with 159,860 households.

3. Methods

Understanding concerns that respondents have about participation and reasons behind their reluctance can help in estimating nonresponse bias. Logistic models were used to predict refusal and noncontact using the CHI data as predictors. The predicted values from those models serve as the propensity to respond or be contacted. 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 (refusal and noncontact are analysed separately to study the effect of attrition). The CHI also has information on household moving. 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 CPS also has information on proxy status, with about half the cases collected by proxy.

Figure 2: Relative frequency of concerns in Wave 1.



4. Results

In the CHI, the most common concern expressed by respondents during the first interview was "busy" (Figure 2), 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 analysis. However, it did not prove to be a problem in creating the propensity scores.

The largest change in unemployment between interviews is between times 1 and 2 (Figure 3). If the first 4 interviews were used, the measurement error hypothesis that the decline in employment was due to change in understanding of the employment questions might make sense. The idea is that the respondent has time to think about the concepts after the interview, and may respond differently at the second interview. The decline in changes from time 1 to time 4 would support the idea. The changes in the last 4 interviews do not fit this hypothesis. Aggregating the later differences T3-T8 would lead to too small a difference to explain the rotation group bias anyway.

Figure 3: Unemployment rate differences by time in sample.

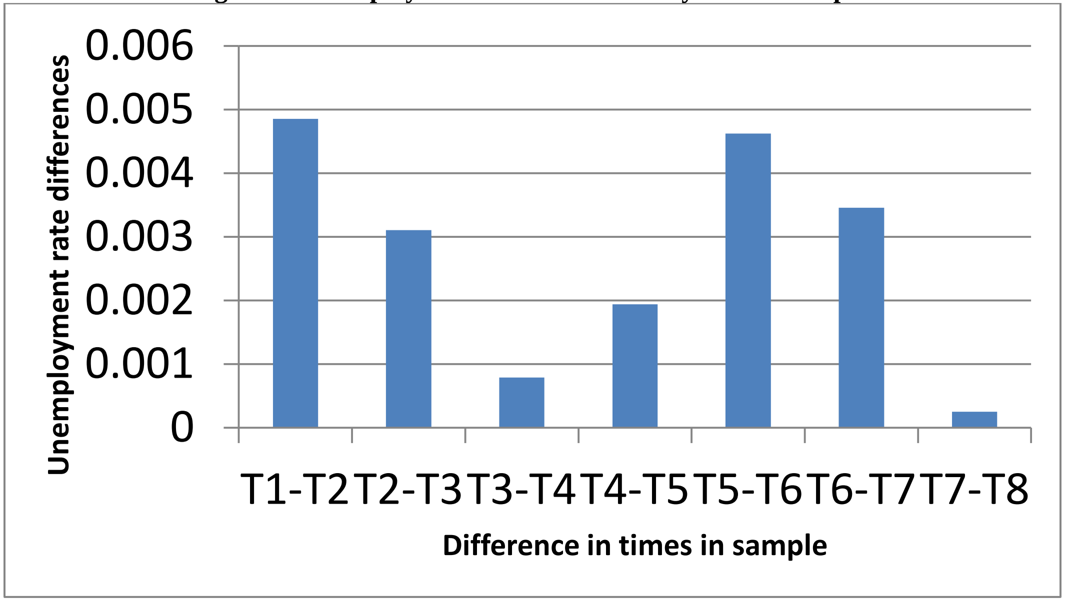
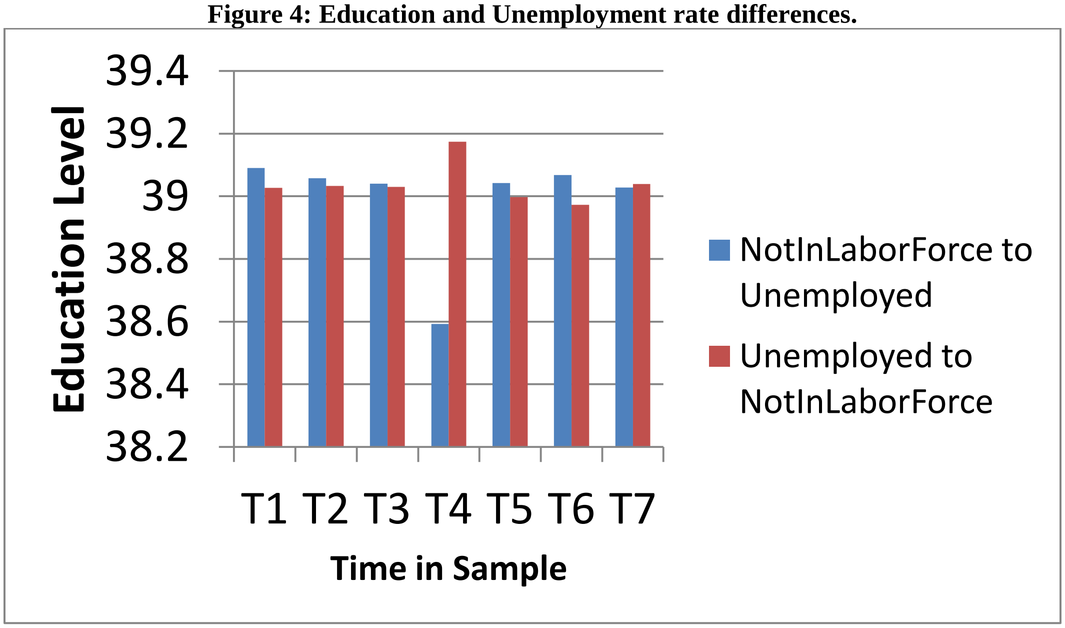


Figure 4: Education and Unemployment rate differences.



Note: The scale for education is grouped from no high school through graduate degrees; the mean represented here is between "High school graduate (39)" and "some college (40)".

One of the strongest correlates of measurement error in other studies is education level. This graph (Figure 4) shows the difference in education between those who change from "not in the labor force" to unemployed (blue bar), to those who change from unemployed to "not in labor force". These are the two concepts that are most likely to be confused (those who are employed are clearer on their status). The differences in education levels are very slight for most interview times, and do not conform to the idea that the changes would happen earlier in the interview process.

Figure 5: Moving by time in sample.

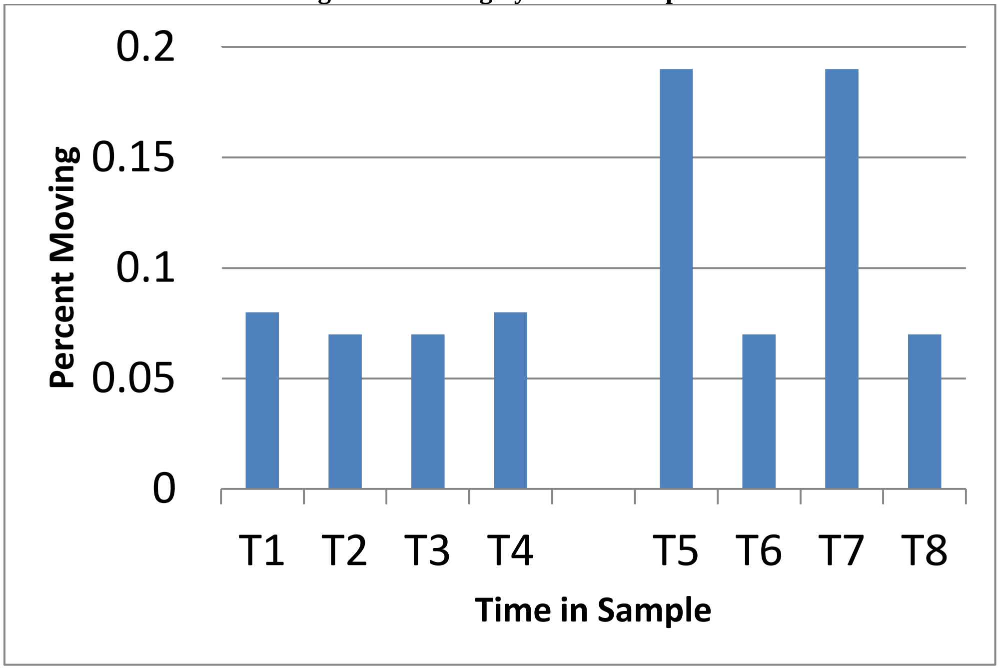


Figure 5 shows the proportion of households that moved. The rate is fairly constant (there are 8 months between 4 and 5, which would account for that increase), except for interview 7. The number of movers is not sufficient to account for much of the changes in employment.

Figure 6: Moving by time in sample.

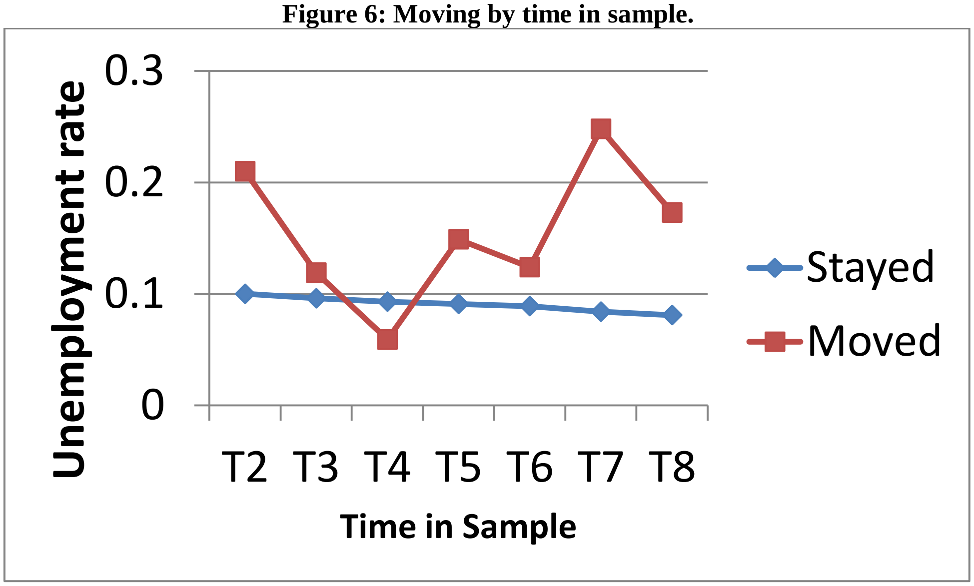


Figure 6 shows that the unemployed are more likely to have moved the next time they are contacted. There is a large increase in unemployment for the 5th month (more move by the next time in sample, 8 months later), but the other large effects (1st and 7th) are not explainable by the hypotheses put forward here.

Because there are so few movers, the impact on the unemployment rate overall is negligible (in the hundredth of a percentage point, rather than the 2 point decline in unemployment observed in the total sample. The greater variability of those who moved is possibly related to the small number of movers.

Figure 7; Proxy reporting and unemployment.

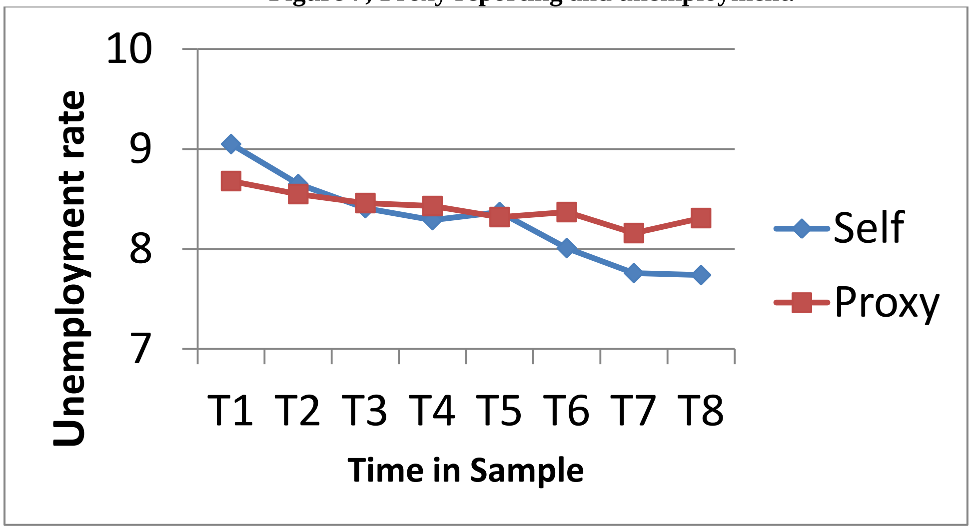


Figure 7 shows the difference between proxy and self-reporters. "Self-reporters" are those reporting employment status for themselves, proxy reports are those reporting for others. The interviewers are supposed to ask each eligible person their labor force questions, but if they are not available a proxy report can be accepted.

The proxy reports start lower, but do not change as much as for self-reports. Could it be that asking about employment is motivating the person responding to search for employment (a type of Hawthorne effect)? The 1 point difference overall would represent about a half point decline in overall estimates, the largest effect so far.

Figure 8: Predicting nonresponse – logistic model coefficients.

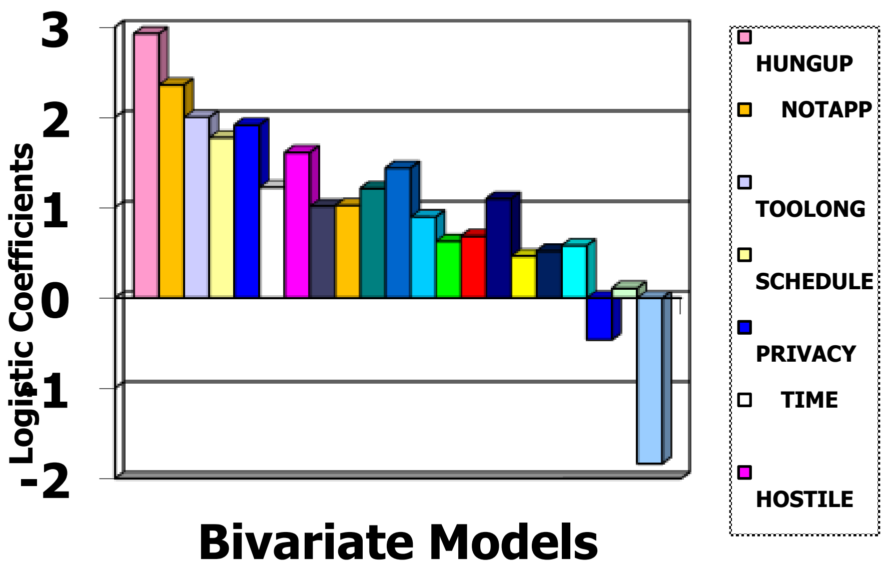


Figure 8 shows the coefficients from the bivariate logistic models for the CPS using the concerns expressed on the CHI to predict refusal. They showed positive relationships between most of the concerns and refusal during some of the interviews. “Family issues” (issues, which was not significantly related to refusal) and “intends to quit” are the two related to not refusing. Additional information on the models and their propensity scores are described in Dixon (2005).

Figure 9: Nonresponse bias.

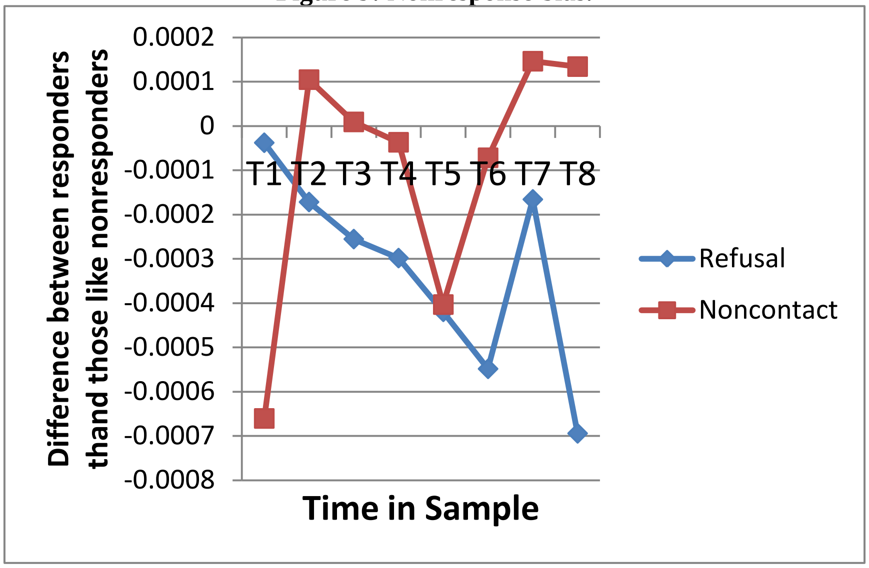


Figure 9 shows that those who are like the nonrespondents based on the CHI propensity scores usually have a slightly higher unemployment rate, leading to a very small negative bias in the estimates. The size of the bias and the pattern would not explain the rotation group bias. I was looking to explain almost 2 percentage points, and this would explain only .05 of one point.

5. Discussion

The CHI data were useful in modeling the relationship between concerns expressed by respondents and refusal/noncontact. The resulting propensity models indicated very slight nonresponse bias. The effect was too small to support the attrition hypothesis. The 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.

Moving had a very small effect on bias. The limitations are that the number of movers were small over the time period and does not capture any relationship between nonresponse and moving. The small number of movers combined with the small number of nonrespondents make any confounding unlikely to contribute to the rotation group bias with a large effect.

Measurement error had an even smaller effect on bias, and was not supported by the patterns of change in unemployment. The limitation is the lack of a better model of measurement error. The largest effect was for proxy/self-reporting. The relationship seems to point to something like a Hawthorne effect, but the proxy reports also show some decline. If that is also related to the household being interviewed, then the effect could account for much of the rotation group bias. It would be interesting to design an experiment to test the hypothesis. If administrative records for unemployment benefits for those interviewed in the CPS showed a similar rotation group bias, then an observational study could use those applying for benefits as a natural comparison group, with little added burden on the respondents.

Future research could include combining the hypothesized sources of error into one model so their relative contribution could be determined. The changes in labor force status would be modeled by the two nonresponse propensity scores, a measurement error indicator, moving status, and proxy reporting indicator all in one model. The relative contribution would be clearer when they are all dealing with the same data in combination, rather than separately as in this study.

References

BLS Handbook of Methods, http://www.bls.gov/opub/hom/

Bailar, B., "The Effects of Rotation Group Bias on Estimates from Panel Surveys", Journal of the American Statistical Association, Volume 70, Issue 349, 1975.

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.

Cheng, Y., Larsen, M., and Wakim, A., "Comparisons of CPS Unemployment Estimates by Rotation Panel", Presented at the Joint Statistical Meetings, 2013.

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.

Dyer, W. (2004). Contact History Instrument (CHI). **http://www.blaiseusers.org/2004/papers/03.pdf**

Erkens, G., "Changes in Panel Bias in the U.S. Current Population Survey and its Effects on Labor Force Estimates", Presented at the Joint Statistical Meetings, 2012.

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.

Mansur, K, and Cheng, Y., "The impact of Different Rotation Patterns on the Composite Estimator", Presented at the Joint Statistical Meetings, 2012.

Mayo E. The human problems of an industrial civilization. 2. Vol. 3. New York, MacMillan; 1993. pp. 53–73.

MPLUS (Version 6.11). [Computer Software]. Los Angeles, CA: Muthén & Muthén.

Solon, G., "Effects of Rotation Group Bias on Estimation of Unemployment", Journal of Business and Eonconmic Statistics, Janucayr 1986, V. 4.1, pp. 105-109.

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