Nonresponse Bias Patterns in the Current Population Survey

John Dixon
Bureau of Labor Statistics, Room 1950
2 Massachusetts Ave., NE
Washington, DC 20212-0001

Abstract

Nonresponse rates have been used as a proxy for survey quality since they indicate the relative potential for nonresponse bias. The Patterns of nonresponse rates(e.g.: seasonal, time in sample) can provide insight into those rates. This study uses different measures of nonresponse bias for the Current Population Survey (CPS) using match data from the Decennial Census and longitudinal CPS data to explore patterns of bias.

Introduction

Studying nonresponse to household surveys is difficult because of a lack of information about nonrespondents. For panel surveys information can be borrowed from other panels. Survey households may also be matched with other sources, usually administrative data (registers) or censuses (Dixon, 2004).

For a single administration of a survey, information can be modeled based on characteristics of those interviewed early and late in the interview process. The lateness of response (for example, the last 5 percent) can be used, since if the effort to collect the data had ended earlier, they would have been nonrespondents (Bates and Crieghton, 2000; Chiu, Riddick, and Hardy, 2001). The nonresponse to items can also be used as a surrogate for nonresponse (Dixon, 2002; Loosveldt, Pickery, and Billiet, 2002).

Data Sources

A key source of data in this study resulted from matching Census long-form data to Current Population Survey (CPS) cases. Therefore, information obtained from the Census could be used to describe nonresponse cases in the CPS. Data from the CPS was selected for February through May, 2000 to cover the response time frame for the 2000 Census long form[1] . For longitudinal patterns, CPS data from several years was used.

Details about the CPS can be found in Technical Paper 63. The CPS is the primary source of information on the labor force characteristics of the U.S. population. Similar estimates can be generated from the Census. However, many methodological differences may contribute to differences between the CPS and Census;

- Reference period (CPS: asks about the week containing the 2nd Tuesday of the month, Census: asks about last week, but over a several month period).
- The CPS consists of 8 separate interviews spread out over a 16 month period using a complex sample rotation design. The Census long form was done once.
- Collection mode (CPS: personal visit on 1st and 5th interviews, other interviews done predominately by telephone; Census: self-administered done mostly by mail; 72 percent, drop off form, 18.8 percent; and the rest mostly by personal interview).

---

[1] Census Day was April 1, 2000.

- Interviewers (CPS interviewers are much more experienced).
- Instrument (Census paper form, CPS computer-assisted interview).
- Questions (CPS asks about active search for work, self employment, owning a business, multiple jobs, retirees); Census is more general and asks fewer questions about labor-force status).
- Collection period: CPS for 10 days, Census for over a month (as long as 7 months).

Methods

The matching process failed to match about 10 percent of the CPS household members using the Census long form. The match was less successful for those who refused the CPS interview (no match for 25 percent of refusers).

The variables used to model nonresponse were adapted from Groves and Couper (1998), and Dixon (2001). A model with 17 predictors and 72 interactions was examined and reduced to a model with 8 predictors and 5 interactions. The adjusted pseudo r-square went from .23 to .20. While the goodness of fit statistics indicated there were other terms which should be added to the model, this model represented a trade-off between complexity and fit.

Two methods for modeling nonresponse based on the current survey respondents were used. The last 5 percent of the respondents was used to represent potential nonrespondents (some noncontact, some refusal), and item refusal was used as a surrogate for unit refusal.

Two sources of information on nonrespondents are also used in the models: panel information from nonrespondents who had responded in previous months, and information from the Census long form matched to the CPS. The models used household demographics from previous months for the panel estimates, and the last 5% of respondents for the late responder estimates. The data from the Census long form was used to develop models based on the household demographics. The labor force estimates were based on the long form for the geograpic analysis and the month-in-sample analysis. The estimates from the CPS were used for the seasonal patterns, since the long form only covered April and nearby months. The models from the Census/CPS match were used to score the data from the CPS for 2006 for the seasonal analysis.

Results

Figure 1 shows the pattern of nonresponse over several years. The refusal rate has drifted up, accounting for most of the change in the overall nonresponse rate.

Figure 2 shows the nonresponse rate by time in sample. The first and fifth time in sample indicates higher nonresponse for the first contact and the recontact after an 8 month interval.

Figure 3 shows the types of nonresponse by time in sample. The first time shows higher "no one home" and "temporarily absent" indicating difficulty in contacting and scheduling interviews in the short 10 day window available to CPS interviewers. "Refuse" is also higher, possibly indicating the short time avaialbe to convert refusals. The fifth time is also higher for the non-contact types, since the interviewers have lost contact with the households and their schedules may have changed over the 8 months since the last interview. Refusals are also higher, although there are many conversions

since the fourth interview, there are even more refusals which produces a net increase in refusals.

Figure 4 shows the refusal rates for many of the sampled counties. The highest rates are in the Northeast and the Southwest. The lowest rates are in the midwest and the South. It is difficult the disentangle the cultural effects from regional office practices.

Figures 5 and 6 show the seasonal patterns of nonresponse. The highest rates are around March and December. The noncontact types have increases in the summer (vacations?) and December (vacations, shopping?). Refusals increase mostly around March, associated with the annual supplement, which is considered burdensome. It is interesting to note that there is an increase even for the first time in sample, so it isn't entirely an experience effect, the attitude of the interviewer may also have an impact. There is also a slight increase in December, possibly because of households being so busy during the holiday season.

Figure 7 shows the refusal rates for counties in 2000, which is similar to the map for the earlier decade. Figures 8 through 10 show the bias in employment, unemployment, and not-in-labor-force (NILF) estimates (the absolute value of the bias), which is more scattered than the rates. This lack of relationship between rates and bias is shown in Figure 11, which is a plot of the log of the refusal rates and the log of the absolute bias.

Figure 12 shows the noncontact rates. It is interesting to note that the Northeast tends to a higher proportion of noncontact, while the Southwest tends to a higher proportion of refusals. They have similar nonresponse rates, but from different sources. Figures 13 through 15

shows the noncontact bias for the different employment estimates, which were similarly unrelated to the noncontact rate as was the nonresponse bias.

Figures 16 through 18 show the refusal bias in employment estimates (not absolute bias, but in the original scale) by month in sample (also called time in sample). Employment and NILF are mirror images, which is expected since they are complementary measures. Both show a reduction of bias as the sample progresses. Unemployment varies more by which bias estimating technique is used. The "last 5%" shows high bias in the first interviews and the 5th interview. This is in keeping with the higher rates, although the 6th interview is odd. The "panel" based estimates of unemployment bias show a higher bias in the 2nd and 7th interviews, which don't correspond to the rates. Overall, the bias is very small.

The noncontact bias is studied in Figures 19 through 21. They show a very similar pattern to refusal bias for employment and NILF estimates. This would be expected for the "last 5%" since it is sure to contain some noncontact bias from both the refusal and noncontact models. Those who are difficult to contact are likely to be either reluctant (screening calls and not answering the door), or not often at home. The unemployment bias estimates are relatively flat for the "last 5%" method, but the panel method shos an increase in the 7th interview.

The monthly refusal rates for 2006 show a similar pattern to that seen in the previous decade (Figure 22), with a high March rate and increasing toward December. Refusal bias shows a large difference (mostly decreasing) around July for both employment and NILF

estimates (Figures 23-25). The unemployment estimates increase near June. The change in composition of responders during the summer months may shed light on the problem. Since many people are on vacation or take temporary jobs during the summer this might be affecting the estimates. The two methods of estimating bias were very similar. This could be due to the similarity of those who are difficult to contact and those who refuse. It could also be an artifact of the methods used.

Noncontact bias is studied in Figures 26 through 29. The rates show difficulty contacting respondents around March (probably due the difficult supplement), then again in the early summer, and again during the holiday season. The bias estimates match the refusal estimates.

Discussion

The study of bias from nonresponse has many difficulties. The assumptions that those not matched wouldn't dramatically affect the results is a strong assuption. The ability of the models to capture enough of the characteristics of nonresponders to estimate the effects is also questionable. The study here should be considered a graphical exploration, leading to further research. The geographic analysis shows a large urban effect for both rates and bias, but this varied considerably by region. Los Angeles and New York showed different patters of rates, but very similar patterns of bias. The "hot spots" scattered throughout the maps might be interesting, in particular those spots with high bias relative to their nonreponse rate.

The month-in-sample bias patterns followed the rate patterns for the most part, suggesting effort could be spent on the first and fifth interview.

The seasonal patterns were unexpected, since they didn't follow the rates. The peaks in June and July suggest further study of responses in those months. It could be due to different employment patterns in the summer months, or it could be that the models (based on April), are giving a false reading. If the indices of bias could be trusted, it would be very encouraging to see that there isn't bias from the March or holiday increases in nonresponse. Understanding the summer increase would be helpful in building confidence in those measures.

References:

- Bates, N. and Crieghton, K., "The Last Five Percent: What Can We Learn from Difficult Interviews?", In Proceedings of the Annual Meetings of the American Statistical Association, August 13-17, 2000. http://www.fcsm.gov/committees/ihsng/asa2000proceedings.pdf
- Beerten, R., and Freeth, S., "Exploring Survey Nonresponse in the UK: The Census-Survey Nonresponse Link Study", Paper presented at the 2004 American Association for Public Opinion Research, Phoenix, Arizona.
- Biemer, P., and Bushery, J., "Estimating the Error in the Labor Force Data Using Markov Latent Class Analysis", Proceedings of the FCSM Conference, 1999. http://www.fcsm.gov/99papers/biemer.pdf
- Chiu, P., Riddick, H. and Hardy, A., "A Comparison of Characteristics Between Late/Difficult and Non-late/Difficult Interviews in the National Health Interview Survey", In Proceedings of the Annual Meetings of the American Statistical Association, August 5-9, 2001. http://www.fcsm.gov/committees/ihsng/chiupasa2000.pdf
- Dixon, J. and Tucker, C., Modeling Household and Interviewer Nonresponse Rates from Household and Regional Characteristics. A paper presented at the International Workshop on Household Survey Nonresponse, Budapest, September, 2000. http://www.fcsm.gov/committees/ihsng/buda5.pdf
- Dixon, J., "Relationship Between Household Nonresponse, Demographics and the Unemployment Rate in the Current Population Survey", In Proceedings of the Annual Meetings of the American Statistical Association, August 5-9, 2001. http://www.fcsm.gov/committees/ihsng/jdixon2001.pdf
- Dixon, J., "The Effect of Item and Unit Nonresponse on Estimates of Labor Force Participation", In Proceeding of the Annual Meetings of the ASA, August 11-15, 2002. http://www.fcsm.gov/committees/ihsng/asa02a.pdf
- Dixon, J., "Using Census Match Data to Evaluate Models of Survey Nonresponse", In Proceeding of the Annual Meetings of the ASA, August 8-12, 2004. http://www.fcsm.gov/committees/ihsng/040812asa.pdf
- Fraboni, R., Rosina, A., Orsini, S., and Baldazzi, B., "A Multilevel Approach to the Analysis of Refusal Risk in the Italian Multipurpose Survey", paper presented at the International Conference on Improving Surveys, August 2002, Copenhagen, Denmark.
- Groves, R. and Couper, M., Nonresponse in Household Interview Surveys, 1998, New York: Wiley.
- Loosveldt, G., Pickery, J., and Billiet, J., "Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey", Journal of Official Statistics, v.18.4, 2002.
- Palumbo, T. and Siegel, P., "Accuracy of Data for Employment Status as Measured by the CPS-Census 2000 Match", Final Report for the Census 2000 Evaluation B.7, May 4, 2004.
- U.S. Department of Labor, Bureau of Labor Statistics. *Current Population Survey Technical Paper 63 - Design and Methodology*. Washington, D.C. 2000.
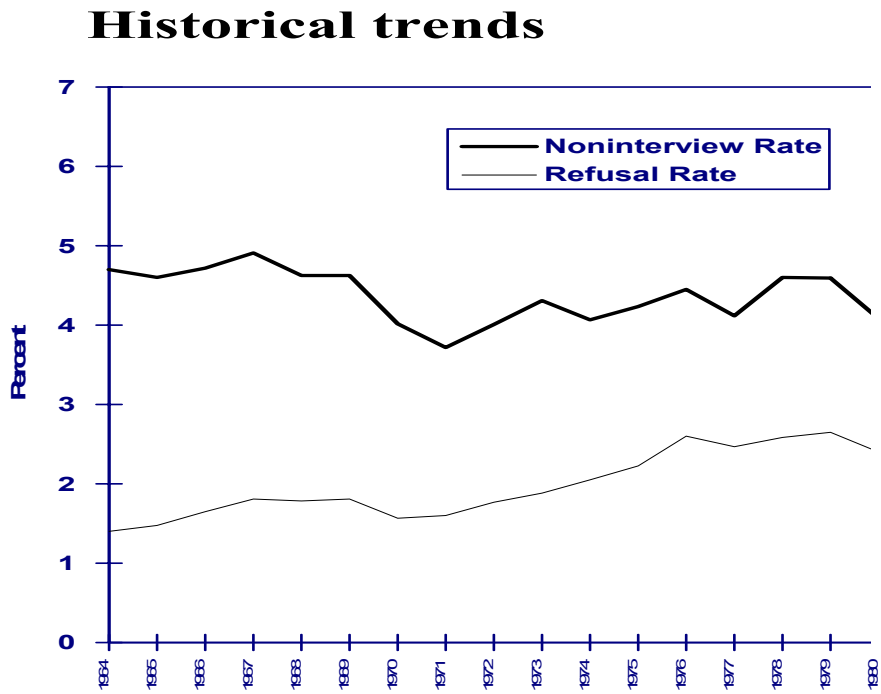
Figure 1: Historical trends

Historical trends

Figure 2: Month in Sample
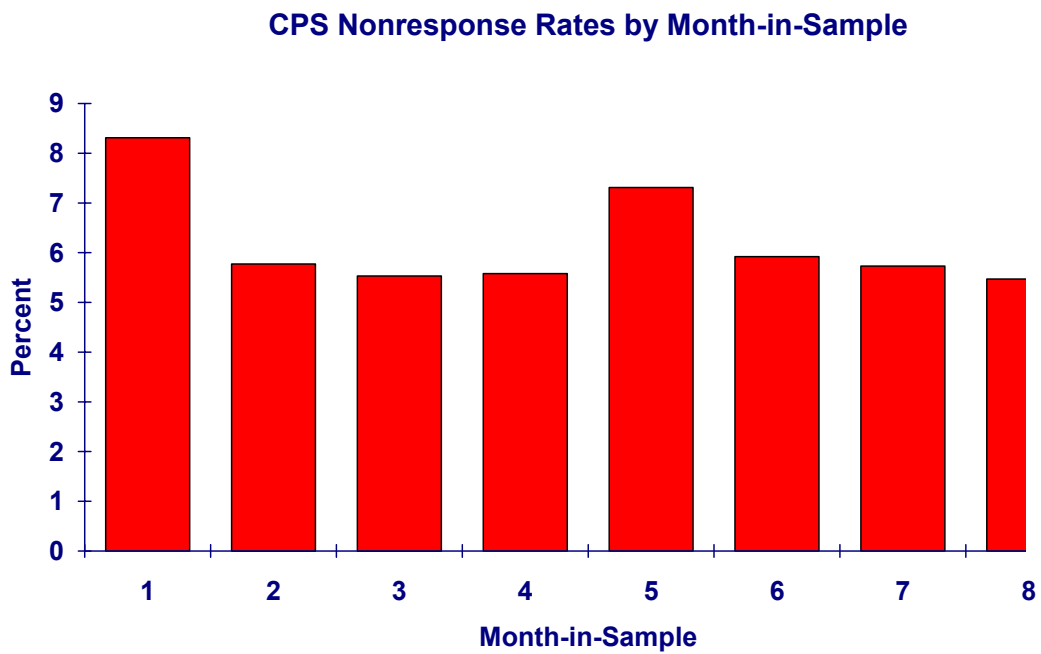
CPS Nonresponse Rates by Month-in-Sample

Figure 3: Type of Nonresponse by Month in Sample
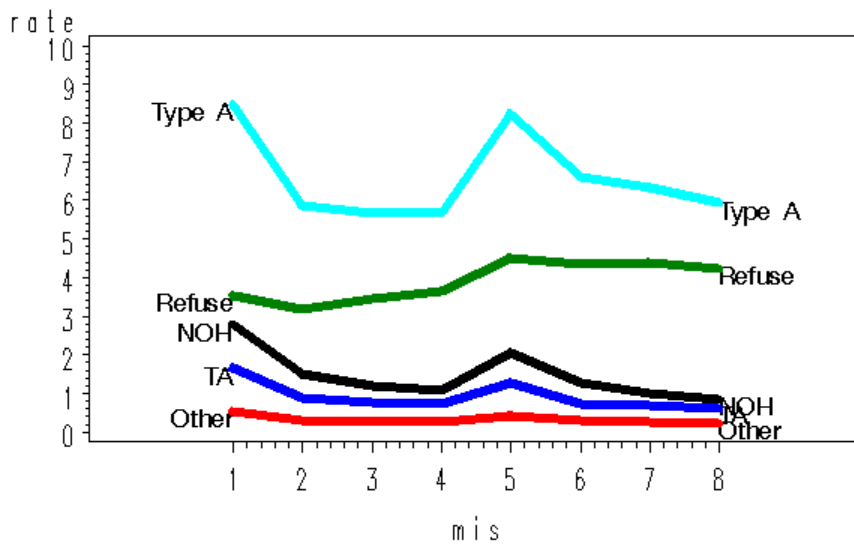
Figure 4: Geographic Variation of Refusal Rates
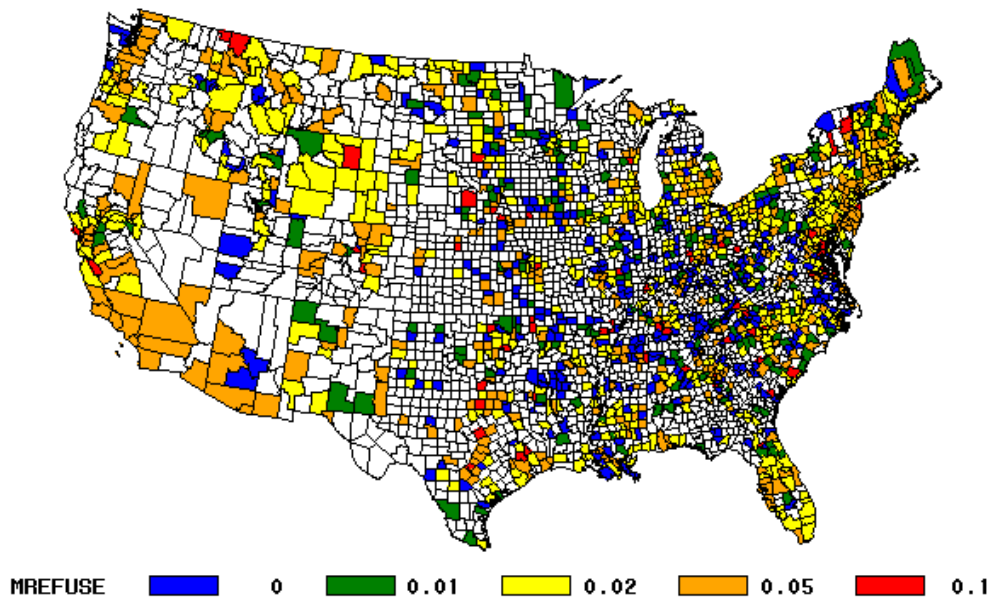

Map of county average refusal rates
1994-1999 MIS 1

MREFUSE ▮ 0 ▮ 0.01 ▮ 0.02 ▮ 0.05 ▮ 0.1

Figure 5: Seasonal patterns

Seasonal patterns
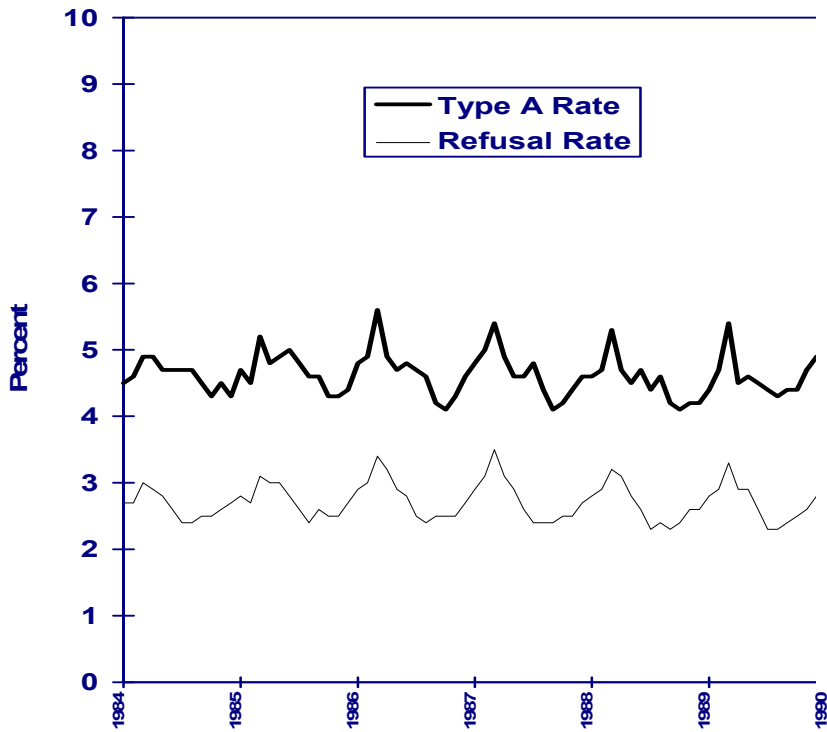
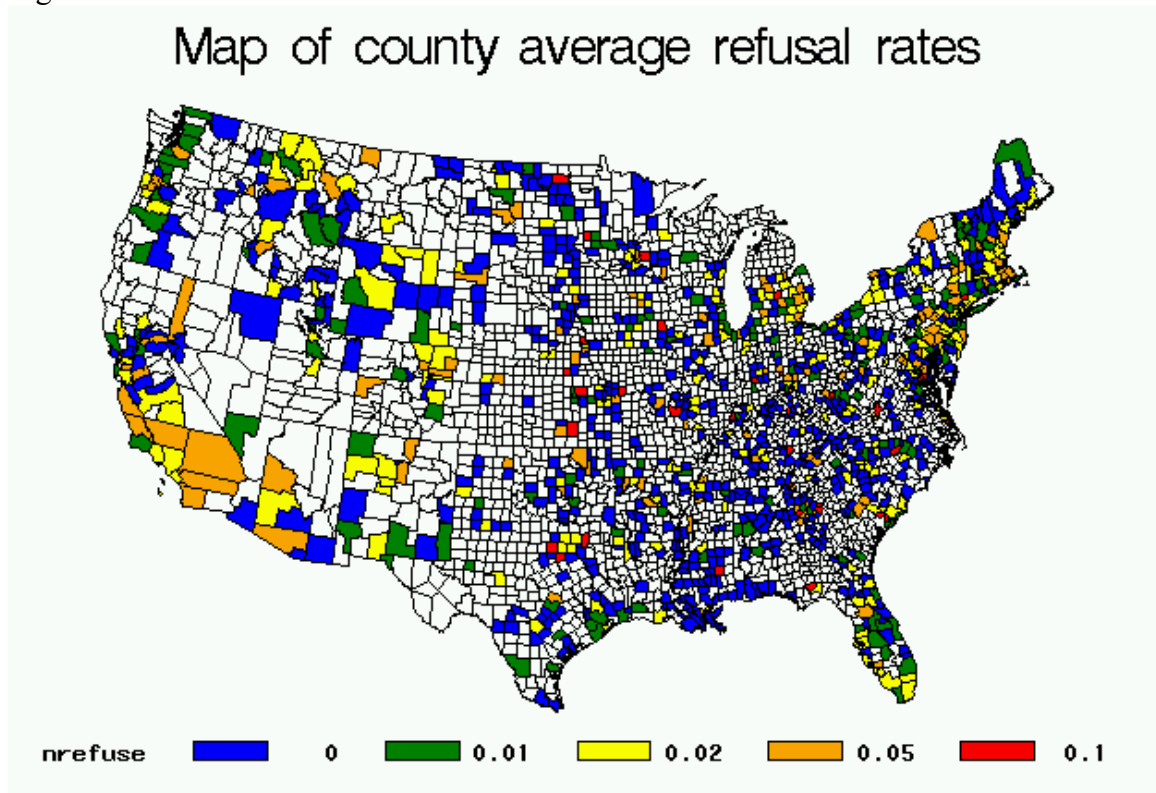Figure 6: Type of Nonresponse by Month
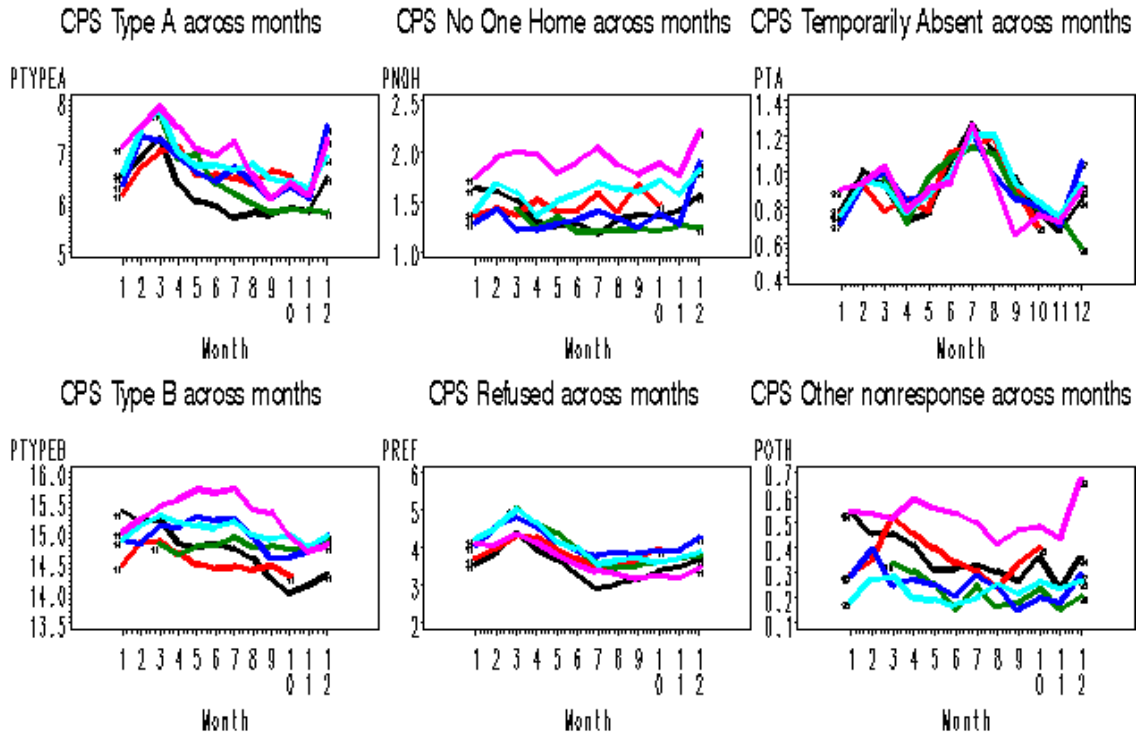
Figure7: 2000 Refusal rates

Map of county average refusal rates

Figure 8: Refusal Employment Bias

Map of county average employment absolute refusal bias
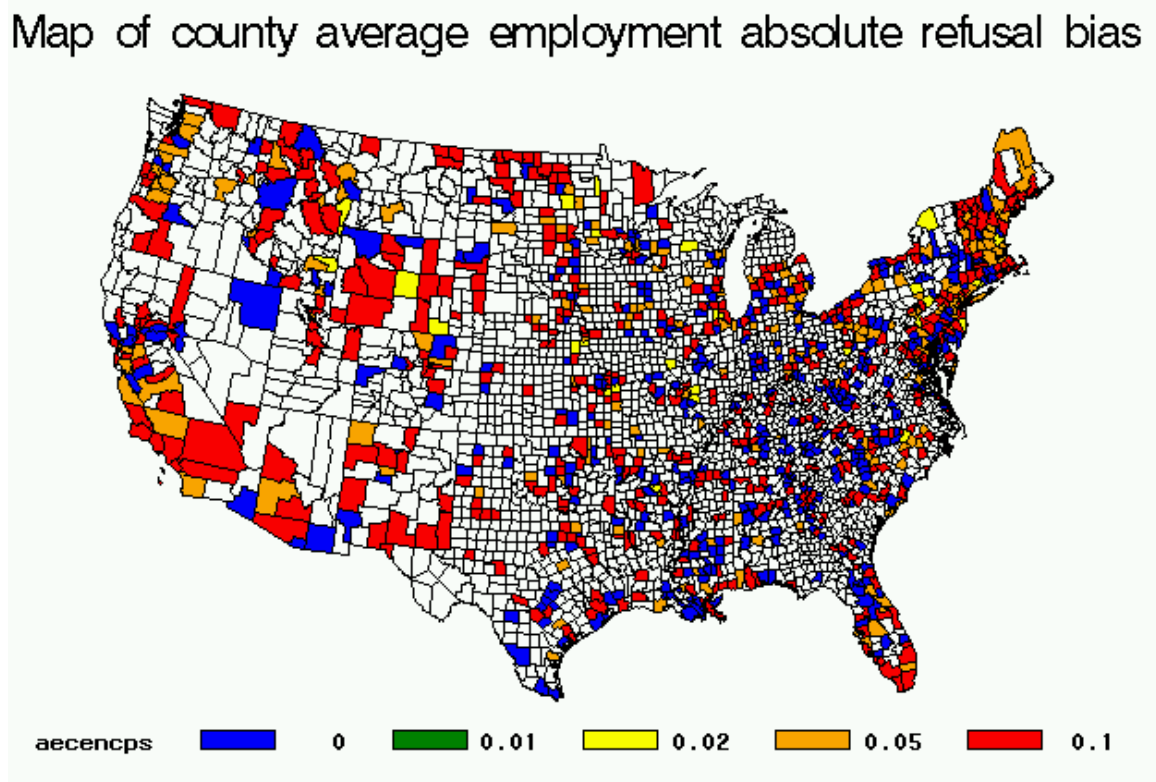
Figure 9: Refusal Unemployment Bias

Map of county average unemployment absolute refusal bias

aucencps   ■ 0   ■ 0.01   ■ 0.02   ■ 0.05   ■ 0.1
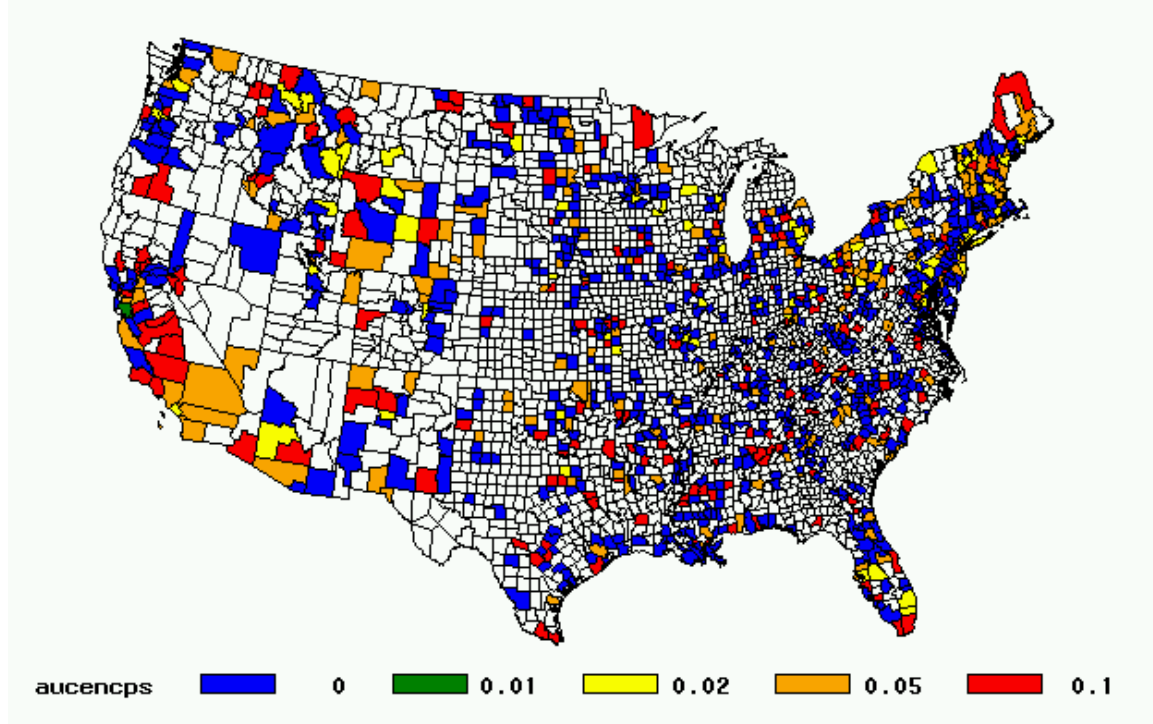
Figure 10: Refusal NILF bias

Map of county average not−in−labor−force absolute refusal bias

ancencps   ■ 0   ■ 0.01   ■ 0.02   ■ 0.05   ■ 0.1
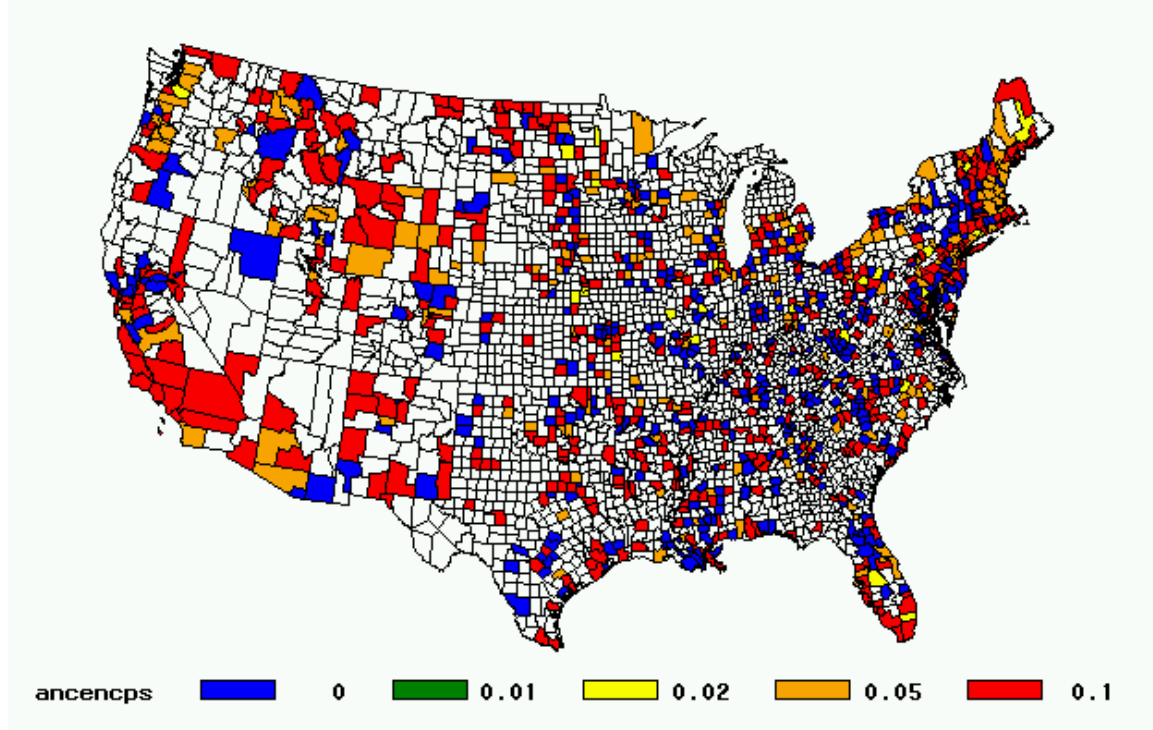
Figure 11: Relationship between refusal rate and absolute bias in employment
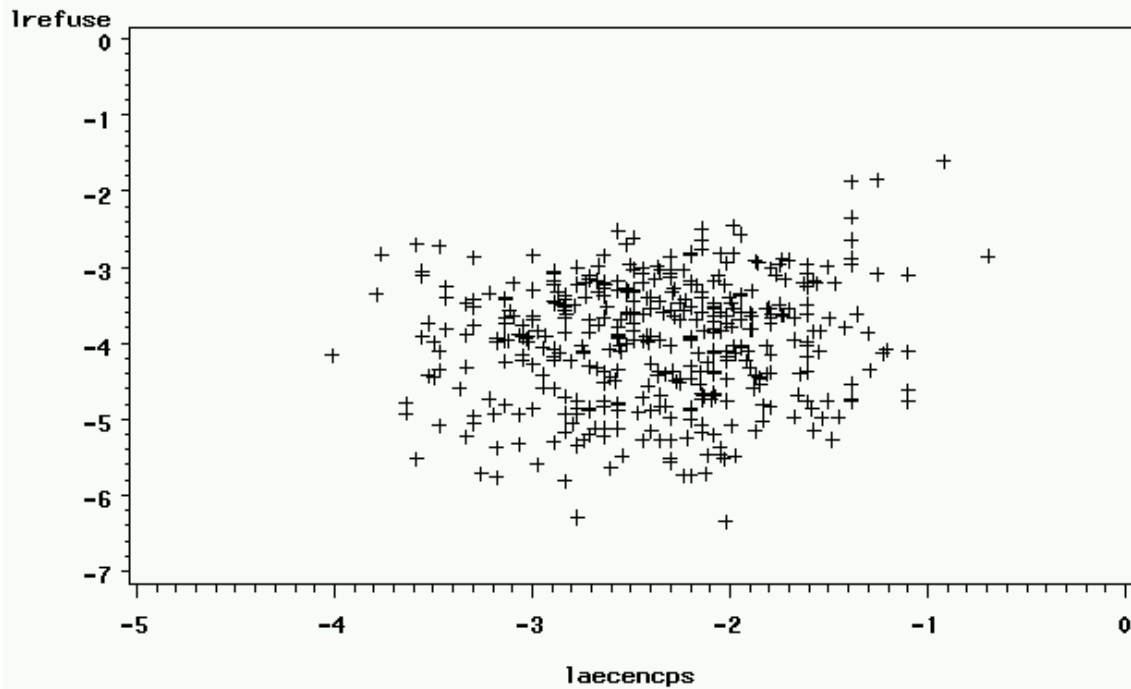
Figure 12: 2000 Noncontact rates

Map of county average noncontact rates
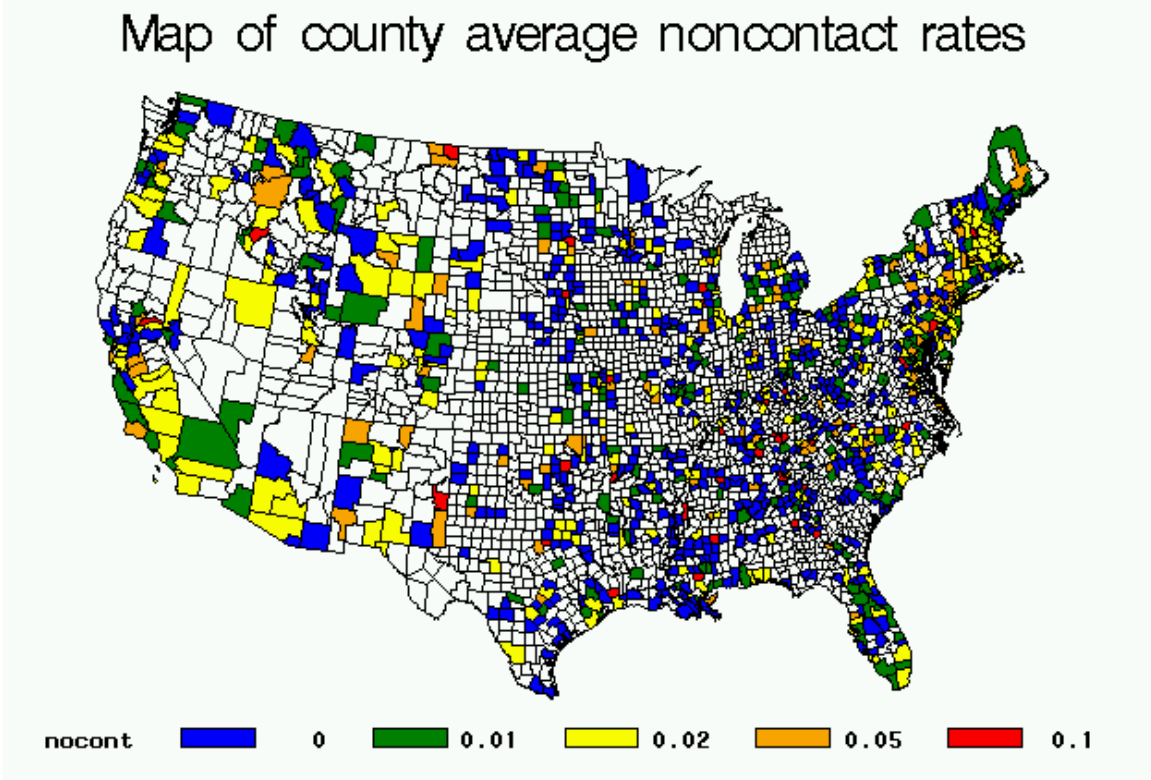
Figure 13: Noncontact bias in employment

Map of county average employment noncontact absolute bias

aecencps | 0 | 0.01 | 0.02 | 0.05 | 0.1
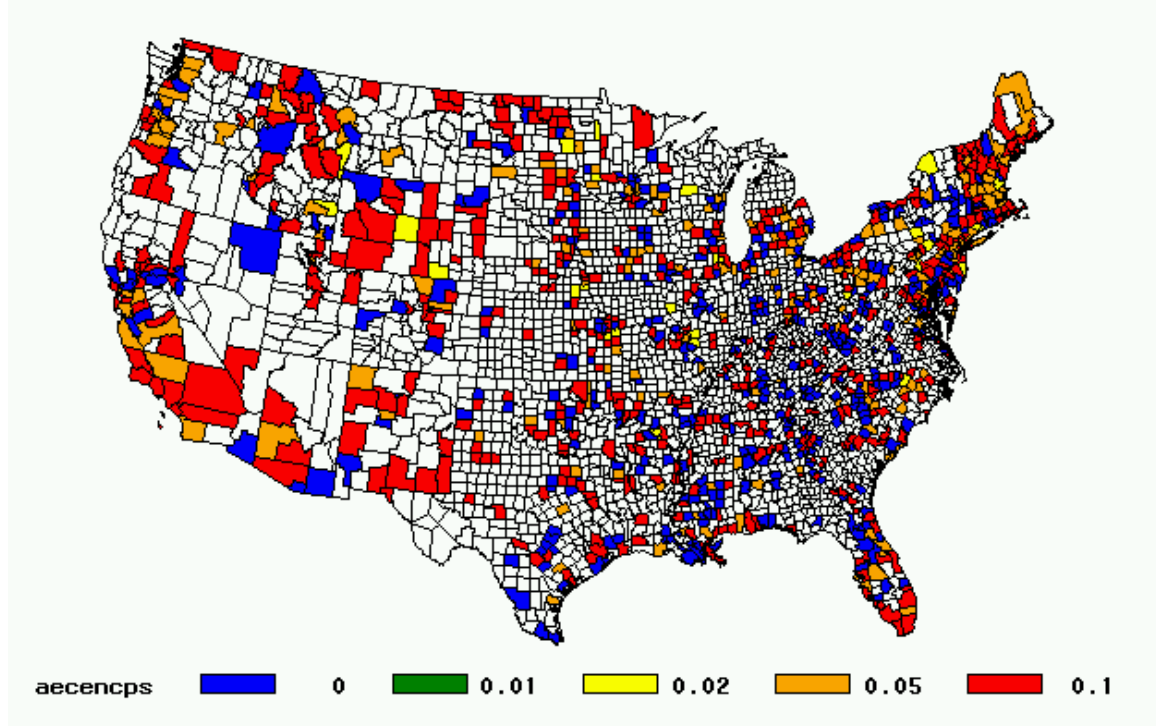
Figure 14: Noncontact bias in unemployment

Map of county average unemployment noncontact absolute bias

aucencps | 0 | 0.01 | 0.02 | 0.05 | 0.1
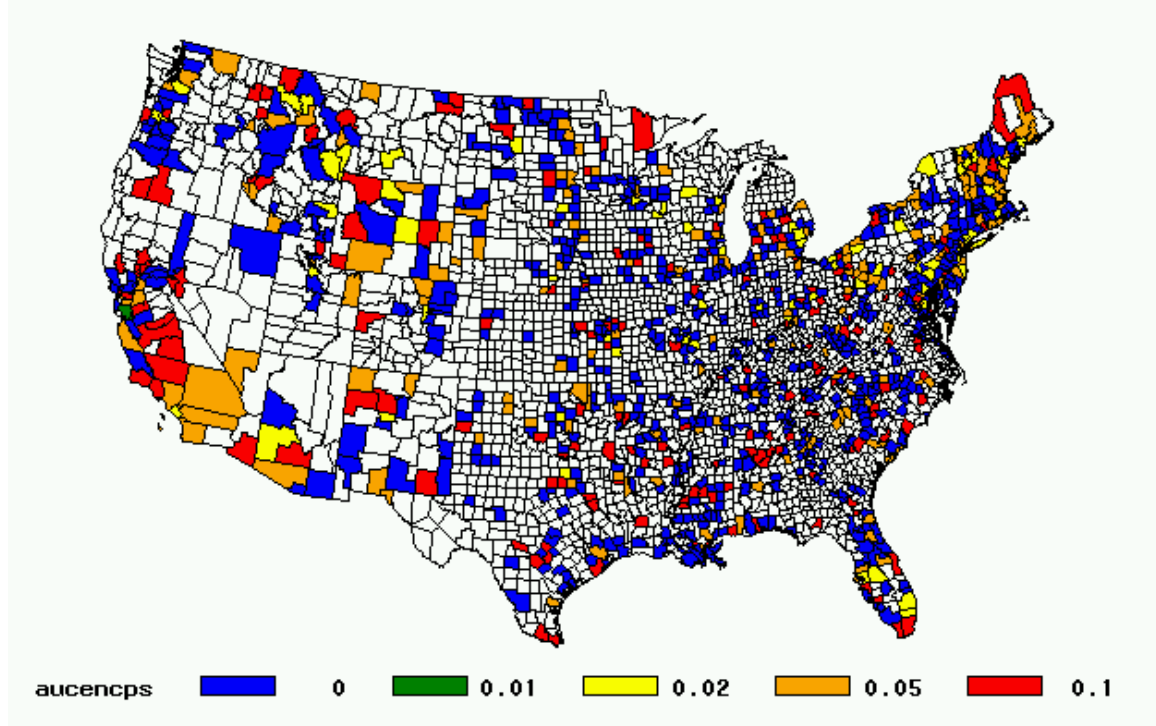
Figure 15: Noncontact bias in not-in-labor-force

Map of county average not-in-labor-force noncontact absolute bias

ancencps   █ 0   █ 0.01   █ 0.02   █ 0.05   █ 0.1
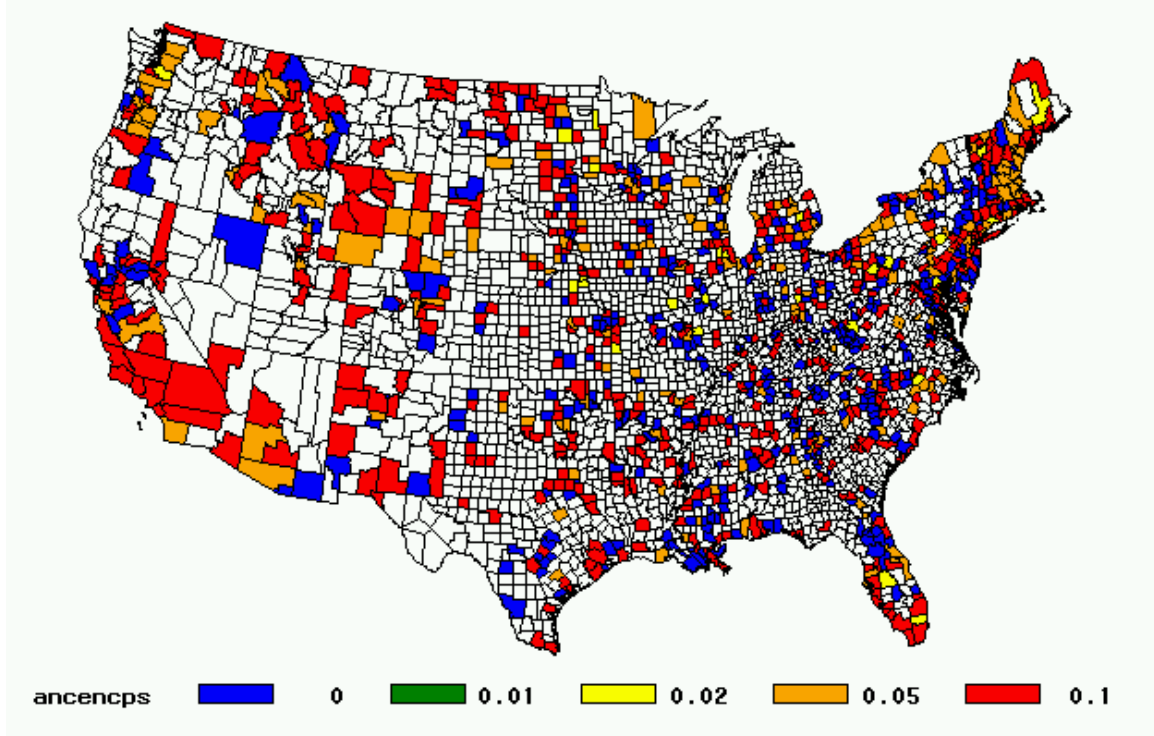
Figure 16: Refusal bias in unemployment by Month-in-Sample

Refusal Bias in Unemployment by Month in Sample
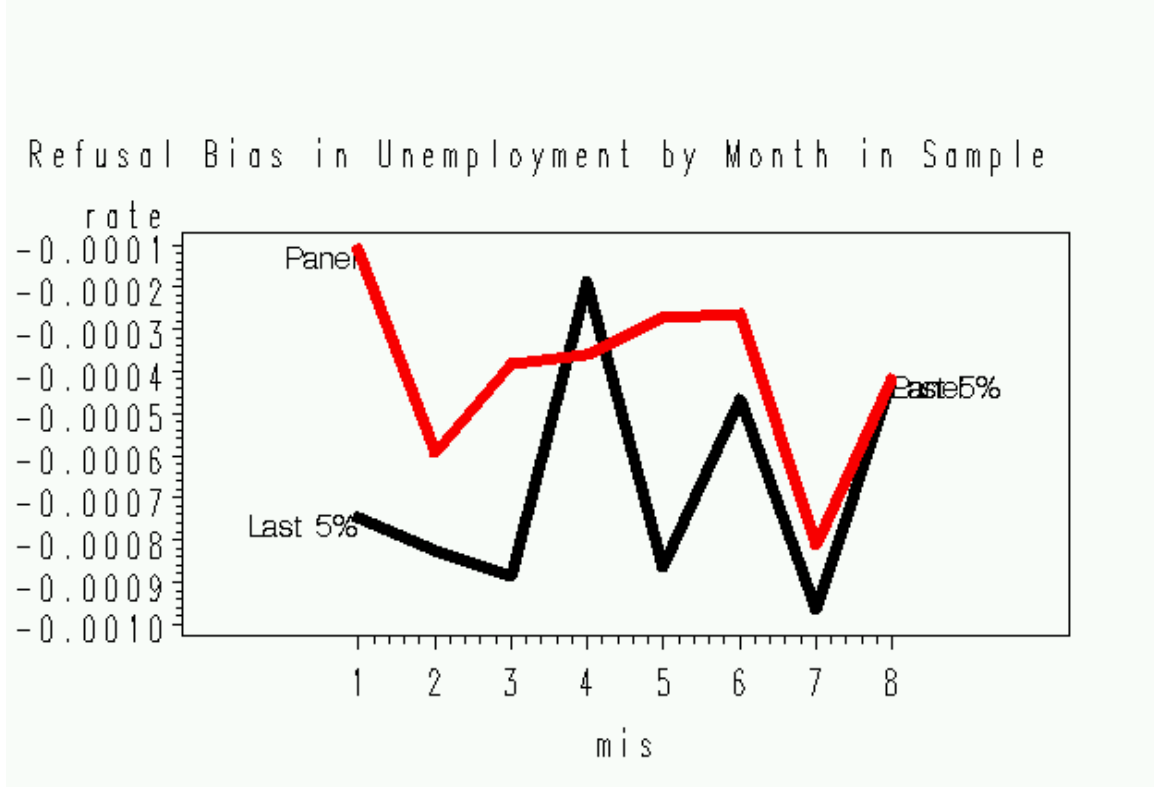
Figure 17: Refusal bias in employment by Month-in-Sample

Refusal Bias in Employment by Month in Sample
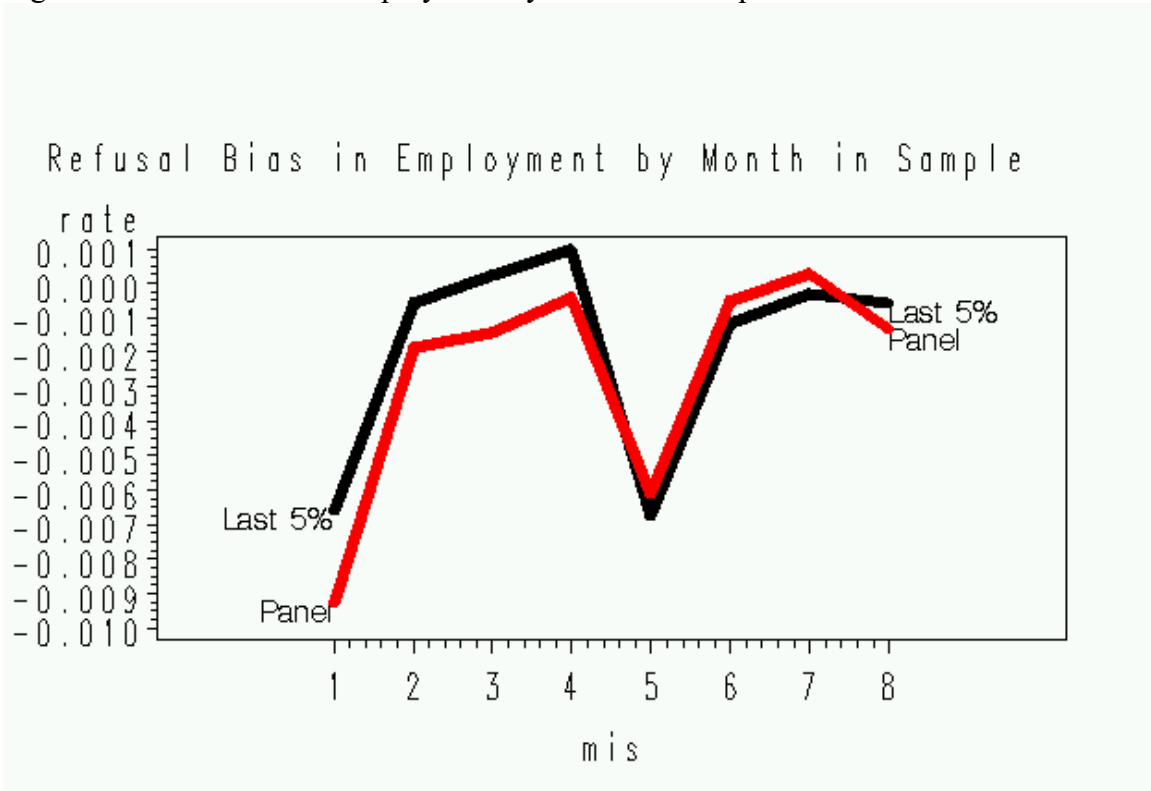
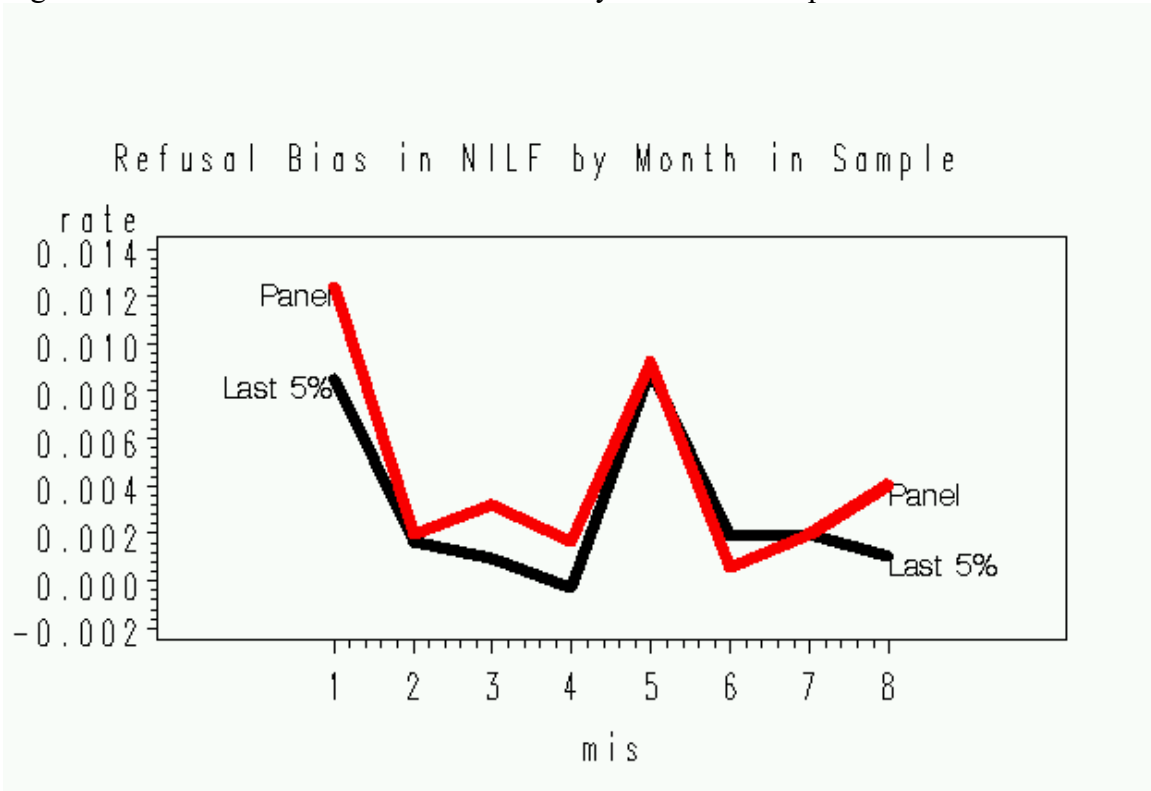Figure 18: Refusal bias in not-in-labor-force by Month-in-Sample

Refusal Bias in NILF by Month in Sample

Figure 19: Noncontact bias in unemployment by Month-in-Sample

Noncontact Bias in unemployment by Month in Sample
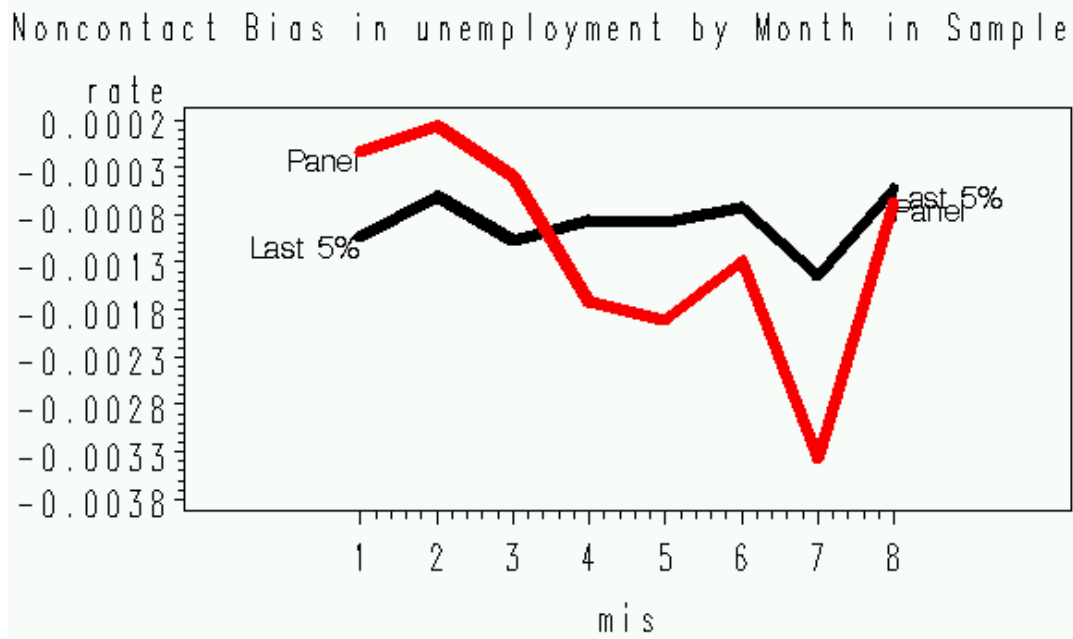
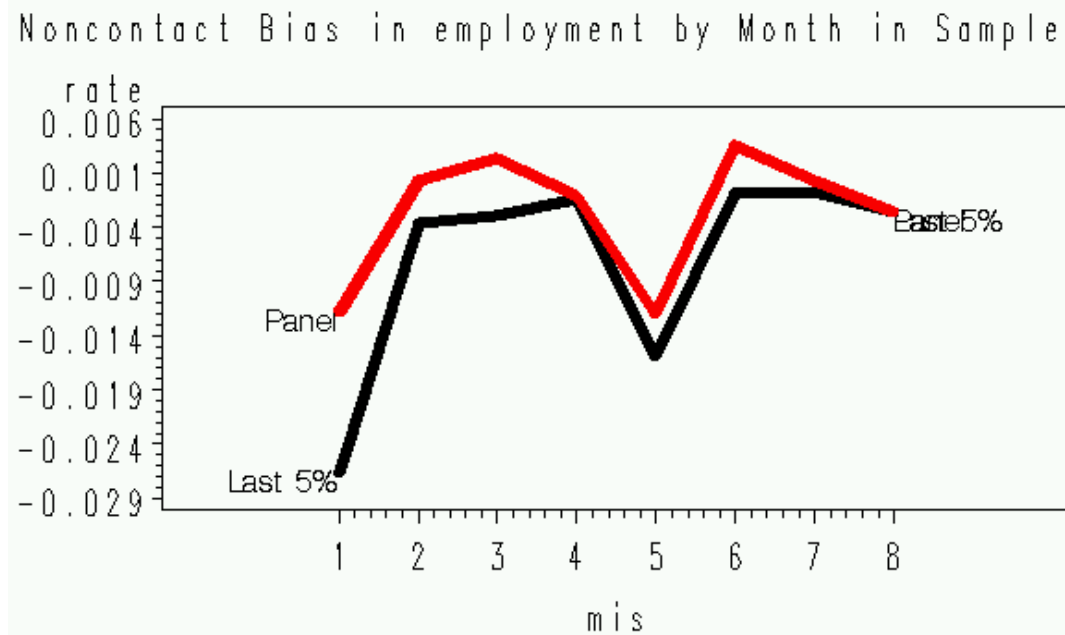Figure 20: Noncontact bias in employment by Month-in-Sample

Noncontact Bias in employment by Month in Sample

Draft- Opinions expressed are of the author and not BLS

Figure 21: Noncontact bias in not-in-labor-force by Month-in-Sample

Noncontact Bias in NILF by Month in Sample

rate

Last 5%

Panel

Panel 5%
Last 5%

mis
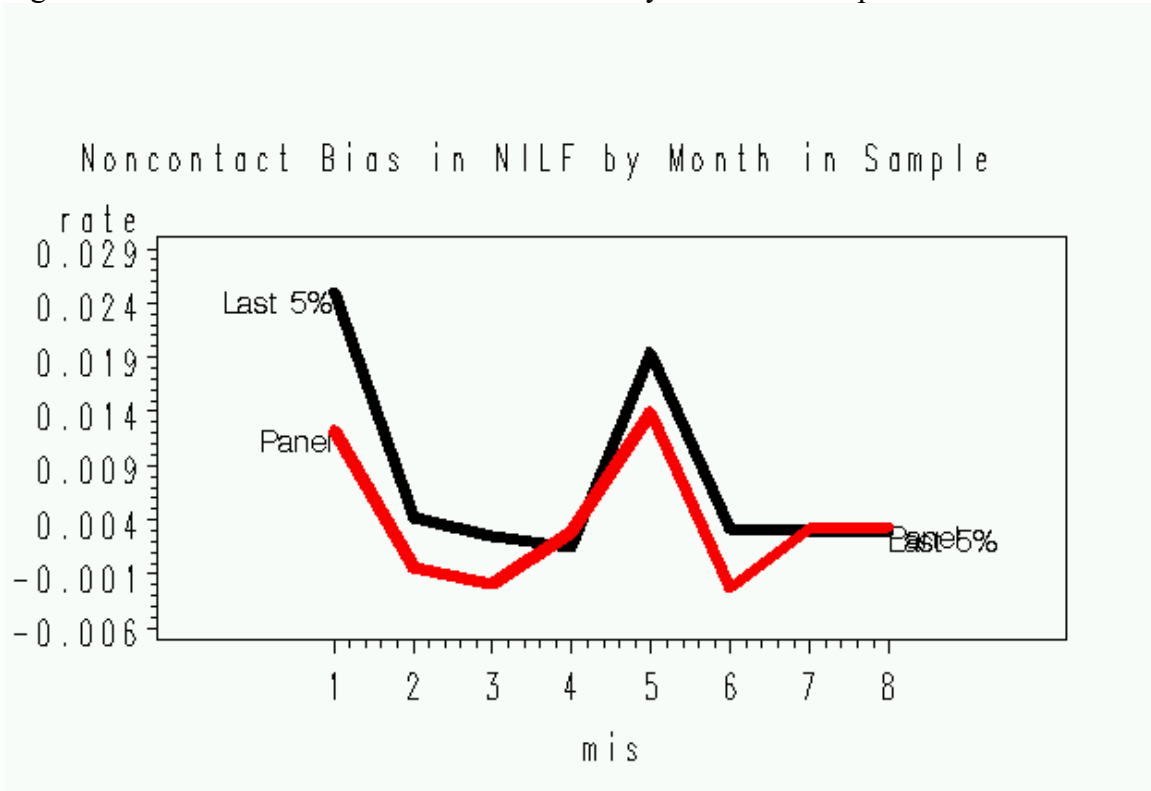
Figure 22: Refusal Rates by Month (2006)

Refusal rates by Month

nrefuse

hrmonth
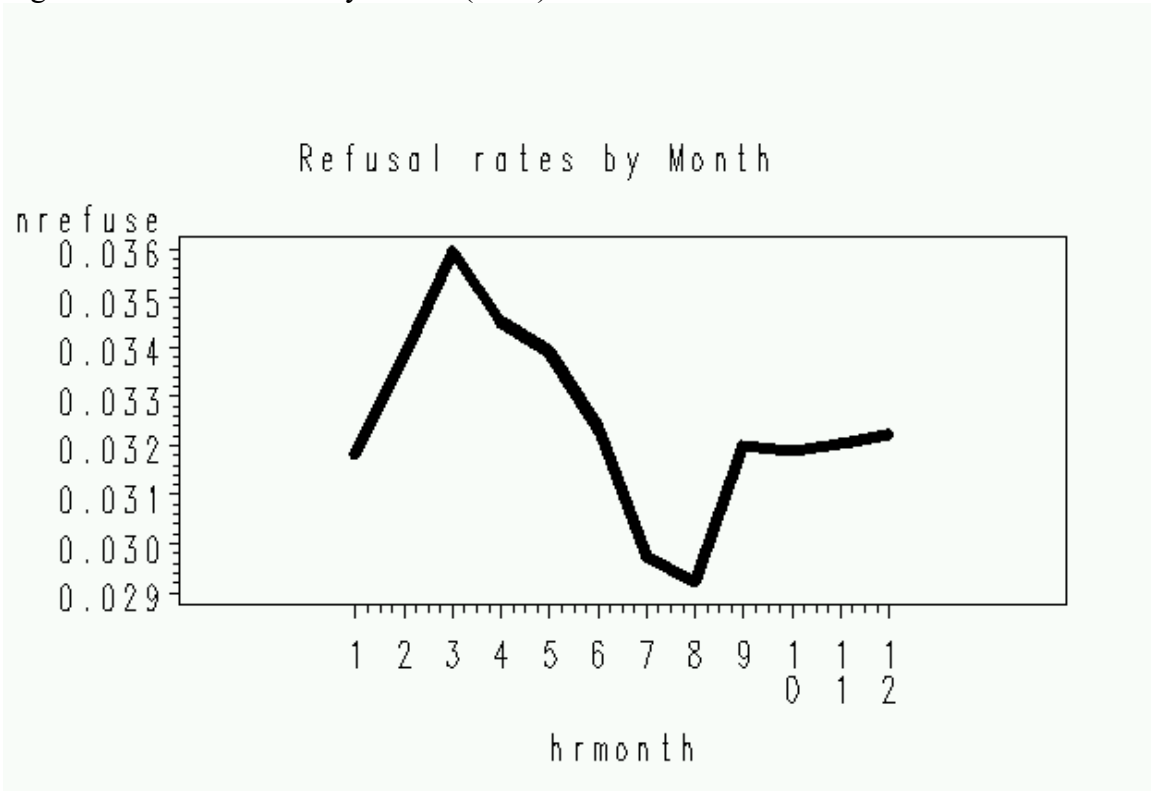
Figure 23: Refusal bias in employment by Month (2006)

Refusal Employment Bias by Month
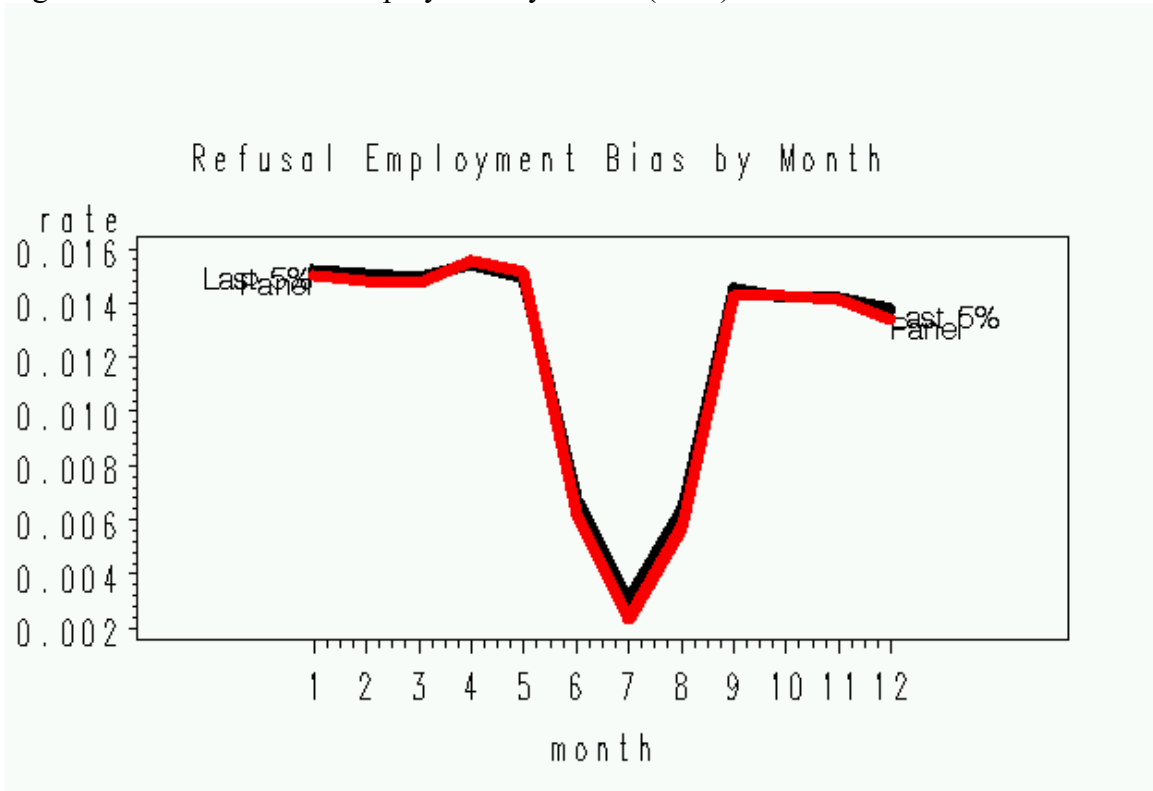
Figure 24: Refusal bias in unemployment by Month (2006)
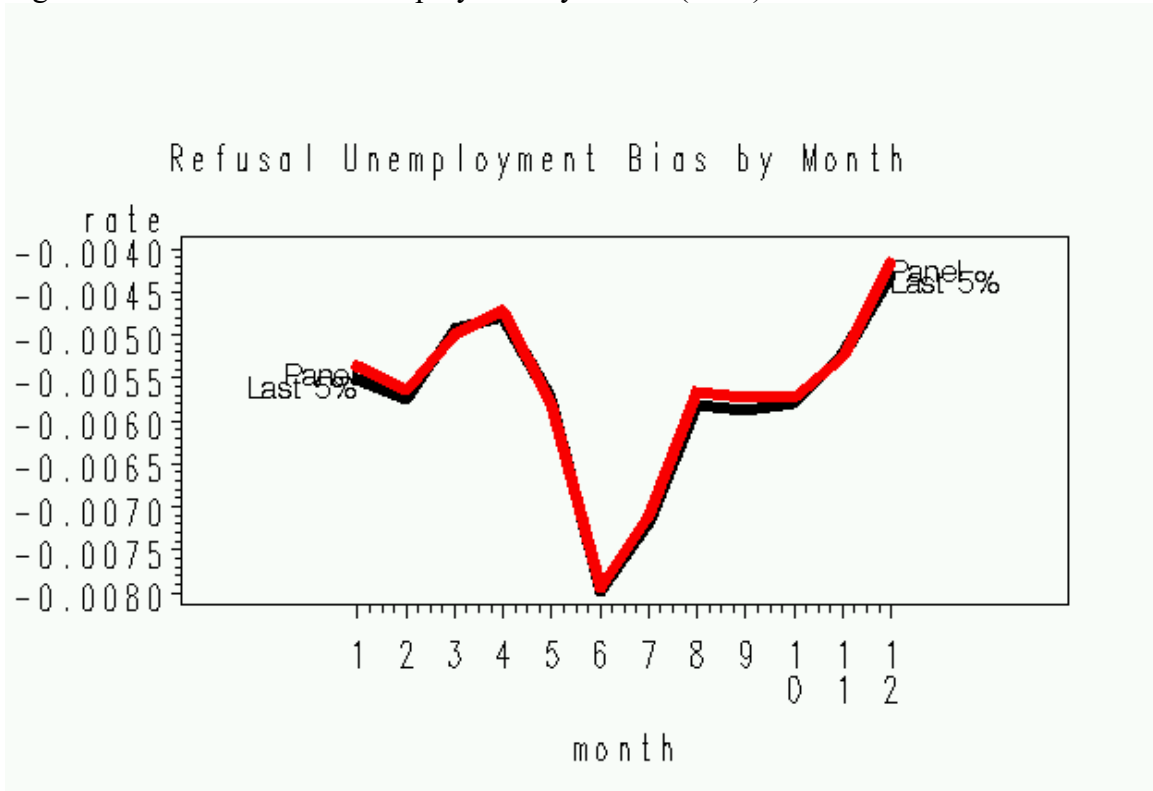

Refusal Unemployment Bias by Month
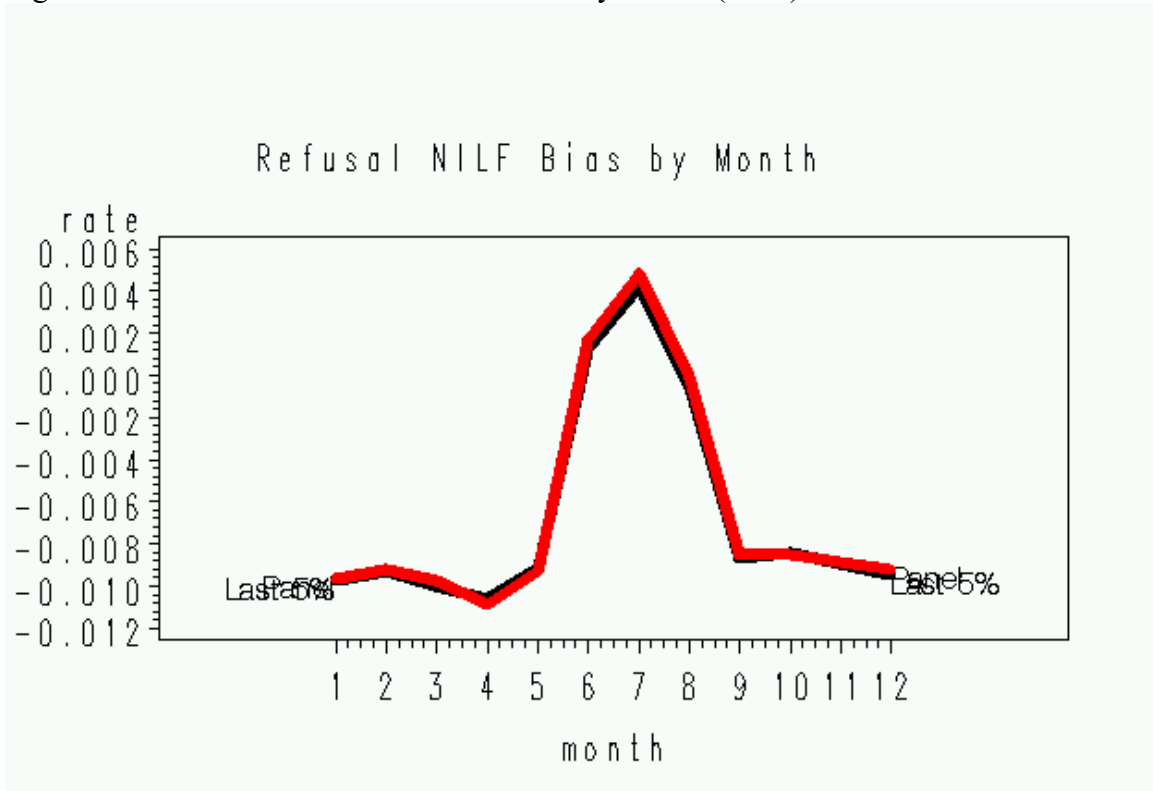
Figure 25: Refusal bias in not-in-labor-force by Month (2006)
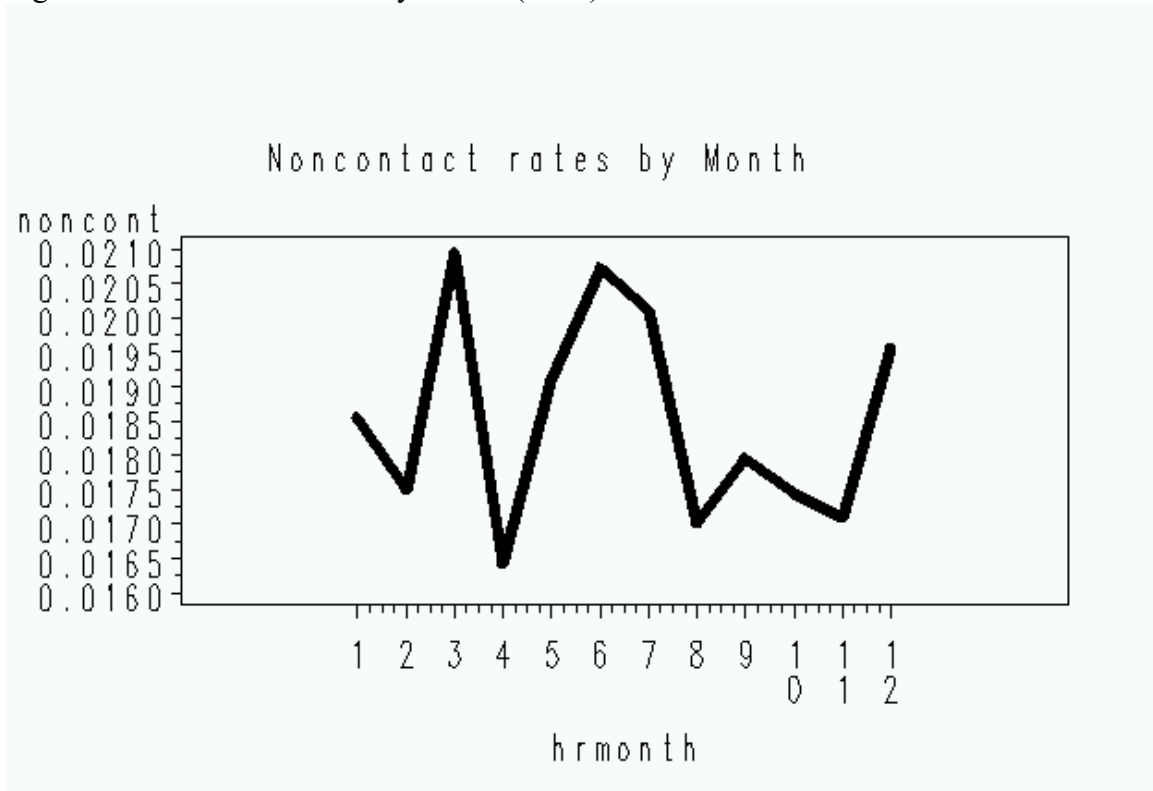

Figure 26: Noncontact Rates by Month (2006)

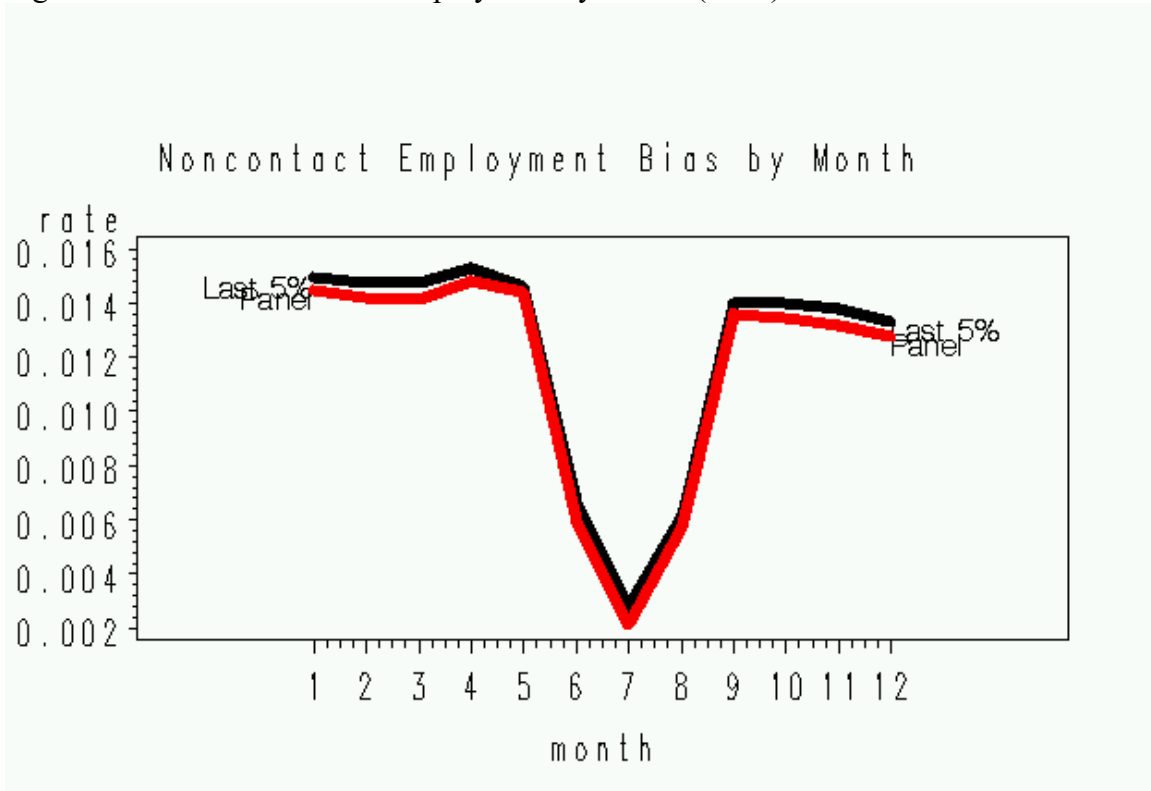Figure 27: Noncontact bias in employment by Month (2006)

Noncontact Employment Bias by Month

rate

0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002

Last 5%
Panel

Last 5%
Panel

1 2 3 4 5 6 7 8 9 10 11 12

month

Figure 28: Noncontact bias in unemployment by Month (2006)

Noncontact Unemployment Bias by Month

rate

-0.0039
-0.0044
-0.0049
-0.0054
-0.0059
-0.0064
-0.0069
-0.0074
-0.0079
-0.0084

Last 5%
Panel

Last 5%
Panel

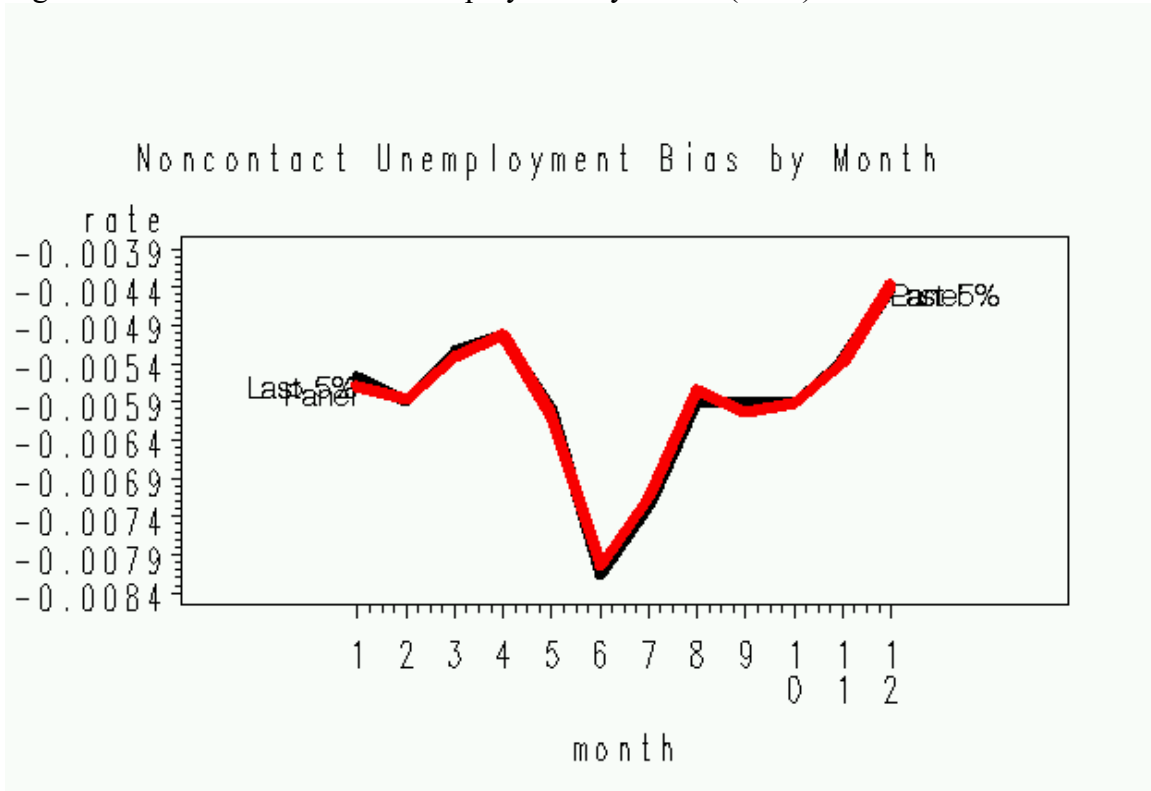1 2 3 4 5 6 7 8 9 1 1 1
                    0 1 2

month

Figure 29: Noncontact bias in not-in-labor-force by Month (2006)

Noncontact NILF Bias by Month