The Effect of the U.S. Census Bureau Realignment on the National Crime Victimization Survey and the Consumer Expenditure Quarterly Interview Survey

Lindsay Longsine and Danielle Castelo
U.S. Census Bureau, 4700 Silver Hill Road, Washington DC 20233

Abstract

This paper is a description of the processes used to monitor the change in the management structure of the U.S. Census Bureau's Regional Offices and a recommendation for future monitoring projects for the National Crime Victimization Survey and the Consumer Expenditures Quarterly Interview Survey.

Key Words: Census, Realignment, NCVS, CEQ

Introduction

The U.S. Census Bureau conducts a variety of national-level demographic surveys; among them are the National Crime Victimization Survey (NCVS) and the Consumer Expenditures Quarterly Interview Survey (CEQ).

The NCVS data are used to estimate crime victimization rates. An annual, overall (entire U.S.) victimization rate is estimated, as well as the rates of certain subtypes of crimes (e.g. robbery, assault, etc.). Rates can also be estimated for different demographic groups (e.g. age, race, etc.). The data can also be used to determine how often crimes go unreported to the police and how often victims know their offenders. The Bureau of Justice Statistics uses NCVS data in their published annual crime reports (Garland 2009).

The data collected for CEQ are used to estimate how much money American households spend on a wide range of items (from big-ticket items like cars, to smaller, everyday items, like groceries). Other uses for this data include market research and determining spending patterns/habits of the population. This survey also provides input to the Consumer Pricing Index and the Supplemental Poverty Measure (BLS 2009).

Background on Restructure

Due to foreseen budget cuts, in 2012 the Census opted to reduce the number of their Regional Offices (ROs) from 12 to 6. This was a gradual process completed over the course of the year (2012) in seven waves (Christy 2011).

Table 1 summarizes this transition.

Table 1: The Seven Waves for the RO transition

Wave	Months
1	January – March
2	April – May
3	June – July
4	August
5	September
6	October
7	November – December

During this downsizing of ROs, the management structure of the remaining ROs changed. Each household sampled for our surveys is linked to one of these ROs based on geographic location. Our field interviewers report to a specific RO and are in charge of interviewing the sampled households covered by that particular RO. Throughout this paper, the terms 'Census realignment' and 'new management structure' are used interchangeably.

In each wave, a percentage of NCVS and CEQ cases were transitioned into the new management system. What this basically means is that the interviewer in charge of that particular case had gone through the required training and started to report to his or her new chain of command. At the end of 2012, all the cases were under the new system.

Figure 1 shows the transition for NCVS, CEQ, and all U.S. counties.

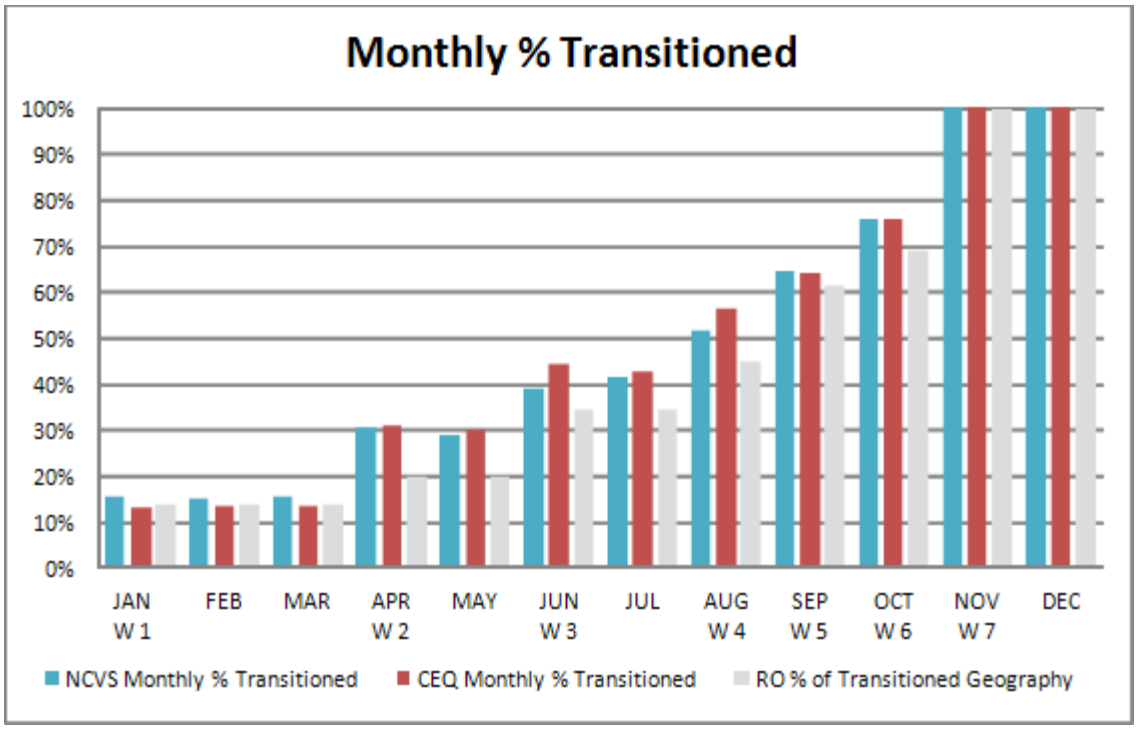


Figure 1: Percent of cases transitioned for NCVS and CEQ and National % of Transitioned Geography for 2012

In Figure 1 we see that for the month of January, roughly 15% of the cases on the NCVS data file were under the new management structure, about 12% of the CEQ cases were under the new management structure, and about 13% of all U.S. counties had been transitioned.

We believed the changes in the number of ROs and the management structure had the potential to affect data quality and, in turn, the final estimates for the Census demographic surveys (including NCVS and CEQ). As a result, we set up a monitoring system which we could check daily for any evidence of a management effect.

This was the first time we have developed a daily monitoring system for the NCVS or CEQ surveys and we have documented our experiences in this report and provide possible recommendations for any projects of this kind.

Responsibilities/Scope

The purpose of this project was to build a daily monitoring system (using statistical models) to determine whether the restructure had an effect on our demographic surveys. The authors of this paper were involved specifically with NCVS and CEQ. For both surveys we selected key variables from our data to represent our estimates, as well as covariates that are believed to have an association with our final estimates. Using these variables, we developed models to monitor the restructuring effects on the survey estimates. These models were run and checked on a daily basis to address any problems that arose. The Data Monitoring charter provides more information on the scope of this project (U.S. Census Bureau 2011).

Data

For this project, raw, unedited data was provided on a daily basis from the interviewers' instruments. If there were missing values for the numeric variables, values were imputed using PROC STANDARD in SAS. This procedure replaces missing values with the mean value for the given variable. For categorical variables, we created a missing category called 'M'. Missing values were not imputed for the key variables.

Because our final estimates for NCVS and CEQ are calculated using a number of factors and a year's worth of data, we selected similarly-related variables from the daily data to serve as our estimates. These variables were referred to as proxy response variables or key variables. The key variables for NCVS were binary (0,1) and the key variables for CEQ were numeric, continuous variables.

Tables 1 and 2 outline these key variables.

Table 1: Key Variables for the NCVS Data-Monitoring Models

Key Variable	Description	Values
Total Crime	All cases where a crime was reported	0 = no crime reported 1 = crime reported
Property Crime	All cases where a household-level crime was reported	0 = no crime reported 1 = crime reported
Person Crime	All cases where a person-level crime was reported	0 = no crime reported 1 = crime reported
Violent Crime	Person-level crimes minus pick-pocketing and purse theft	0 = no crime reported 1 = crime reported

Table 2: Key Variables for the CEQ Data-Monitoring Models

Key Variable	Description	Values
Rent	Quarterly rent value	Continuous ≥ 0
Rent Equivalent	Monthly rent equivalent value	Continuous ≥ 0
Property Value	Property value	Continuous ≥ 0
Gas Expenditures	Monthly gas expenditures	Continuous ≥ 0
Health Expenditures	Quarterly health and medical expenditures	Continuous ≥ 0
Food Expenditures	Weekly food expenditures	Continuous ≥ 0
Contributions	Quarterly contributions	Continuous ≥ 0

The main explanatory variable of interest was the management variable. This variable would measure the change in the management structure. The management variable (called z-var) was binary (0,1) where

Table 3 defines these two restructure variables

Table 3: Restructure variables for the Data-Monitoring Models

Restructure Variable	**Description**	**Values**
Current Wave	The wave of the Census realignment on the day the model was run.	1 - 7
Wave RO Change	The wave each case was assigned to transition to the new management structure.	1 - 7

Thus a case assigned to change in the sixth wave would have the following values for the restructure variables throughout 2012

	Current Wave	*Wave RO Change*	*z-var*
January - March	1	6	0
April – May	2	6	0
June - July	3	6	0
August	4	6	0
September	5	6	0
October	6	6	1
November – December	7	6	1

Models

We created a model for each key variable (4 models for NCVS and 7 models for CEQ). For the NCVS models, because crime is a rare occurrence, we first tried to fit zero-inflated models to the data. Though we would get output, these models would not converge, and the validity of these models was questionable. Ultimately, we opted for a simpler logistic regression approach for the NCVS data (Killion 2011). For CEQ we used linear regression models. The key variables were transformed using the natural log in order to meet normality assumptions (Killion 2011).

The models were run on a daily basis and the beta values (or estimated coefficients) for the z-var variable from the model output were saved and graphed. Included in these graphs were 95% confidence bounds around zero. These graphs were checked daily to look for instances where the beta value fell outside these bounds (when z-var was deemed statistically significant at α=0.05).

Results

Figure 2 shows the z-var beta values for the NCVS models for the year of 2012.

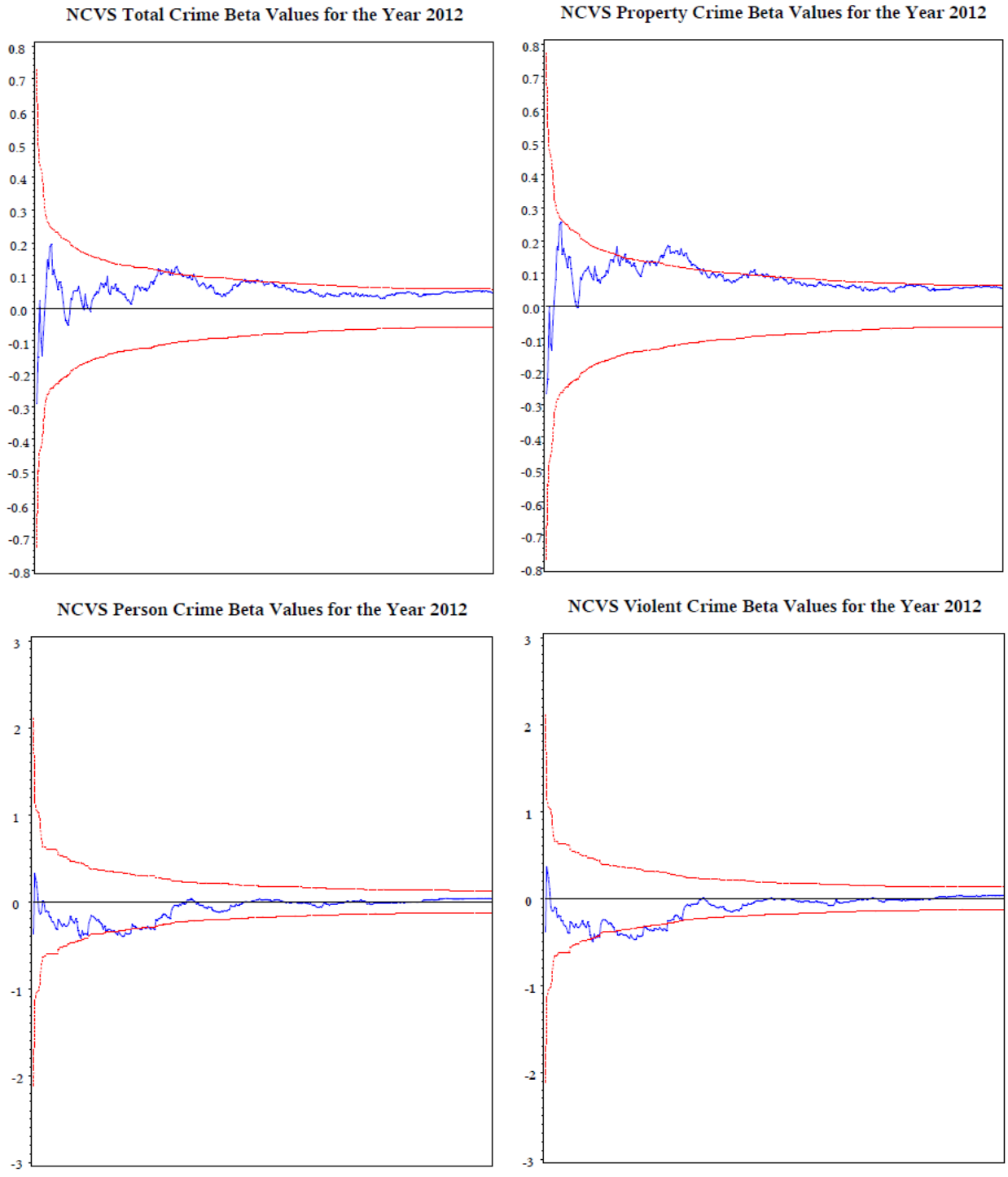


Figure 2: NCVS Beta Values for the entire year of 2012

The management variable was out of bounds 5.3% of the time for the total crime model, 15.0% of the time for the property crime model, 6.7% of the time for the person crime model, and 12.3% of the time for the violent crime model. When using a 95% confidence interval, z-var is expected to fall outside the bounds 5% of the time just by chance.

The out of bounds percentages for NCVS are higher than expected given the null hypothesis of no management effect. All instances where the management variable went out of bounds occurred in the first three waves of 2012.

At first, we thought there was confounding with another NCVS training that overlapped with the new management training. In late 2011 and early 2012, NCVS interviewers went through Refresher Training, with roughly half of the interviewers being trained in August 2011 and the other half in February 2012. The reason behind this training was a concern that crime incidents were being missed by interviewers who were not reading the survey questions completely or as worded (Schafer 2013).

The NCVS interview process consists of a screener interview and an incident report (with the latter happening only if a crime was reported in the screener). In the design of the NCVS screener questionnaire, questions were worded in a way to trigger the respondent's memory. If simply asked, "Were you a victim of a crime in the last six months?" respondents may only remember more traumatic crimes or crimes that they would report to police. The survey questions are developed to uncover crimes that the respondent may not have remembered or perhaps did not think was a crime. If these questions are not read completely and as worded, the potential to miss crime increases.

The length of time required to read the screener correctly typically falls between 3.5 and 4 minutes. When we looked at the screener times for our interviewers before the Refresher Training, the average time was approximately 90 seconds. Concerned that we were underreporting the crime rate, we decided to schedule a Refresher Training. In addition to this training, the interviewers were informed that their screener times were going to become part of the performance review process. These two changes (NCVS Refresher Training and performance monitoring) had the potential to greatly affect the NCVS data. We theorized that NCVS crime rates would spike after these changes, and an experiment was conducted to analyze these effects.

For confidentiality and availability reasons, indicator variables for the Refresher Training/performance monitoring were not available for the Realignment project. Thus we were unable to control for these changes.

Results from the Refresher Training/performance monitoring project show that, although the screener times increased, the 2012 crime rates were not affected. These models also had an indicator variable for the Census Realignment, which was not significant (Schafer 2013).

With this information, we think that there was no real Realignment effect to the NCVS data and that it is possible that the crime rates for the geographic areas in each wave could be different from the others.

To test the idea that the waves differ, we decided to treat each wave as its own independent test. We created wave-specific data files that were not cumulative. We ran the NCVS Total Crime model for each wave. The results mirror what we found for the entire year. For waves 1, 4, 5, 6, and 7, z-var stayed within bounds, but waves 2 and 3 showed significant effects.

Figure 3 shows the output for waves 2 and 3.

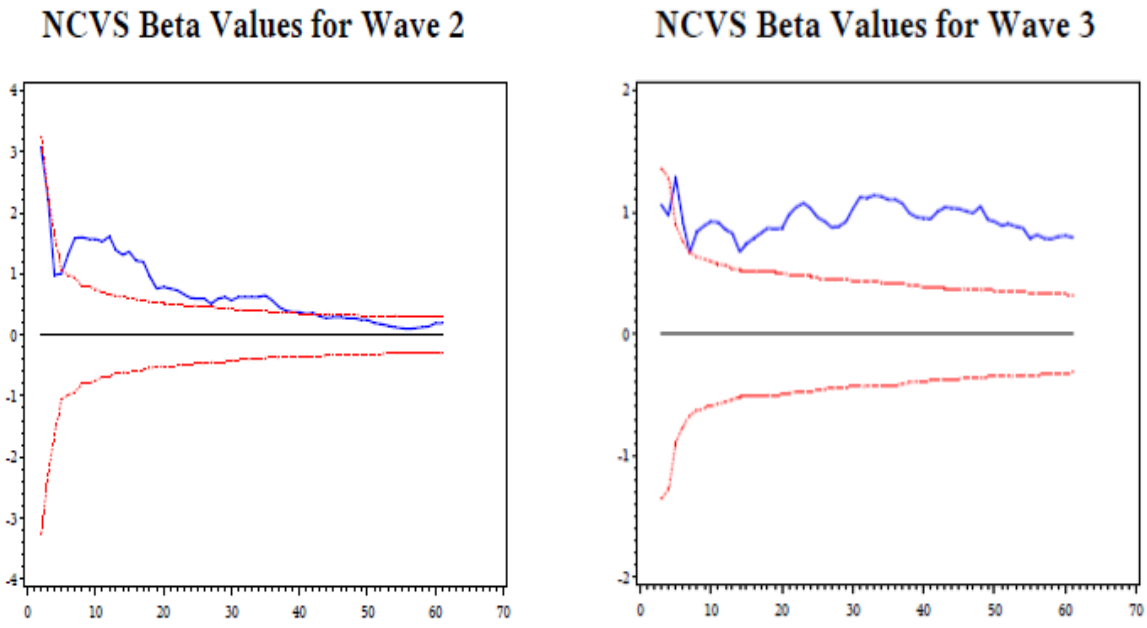


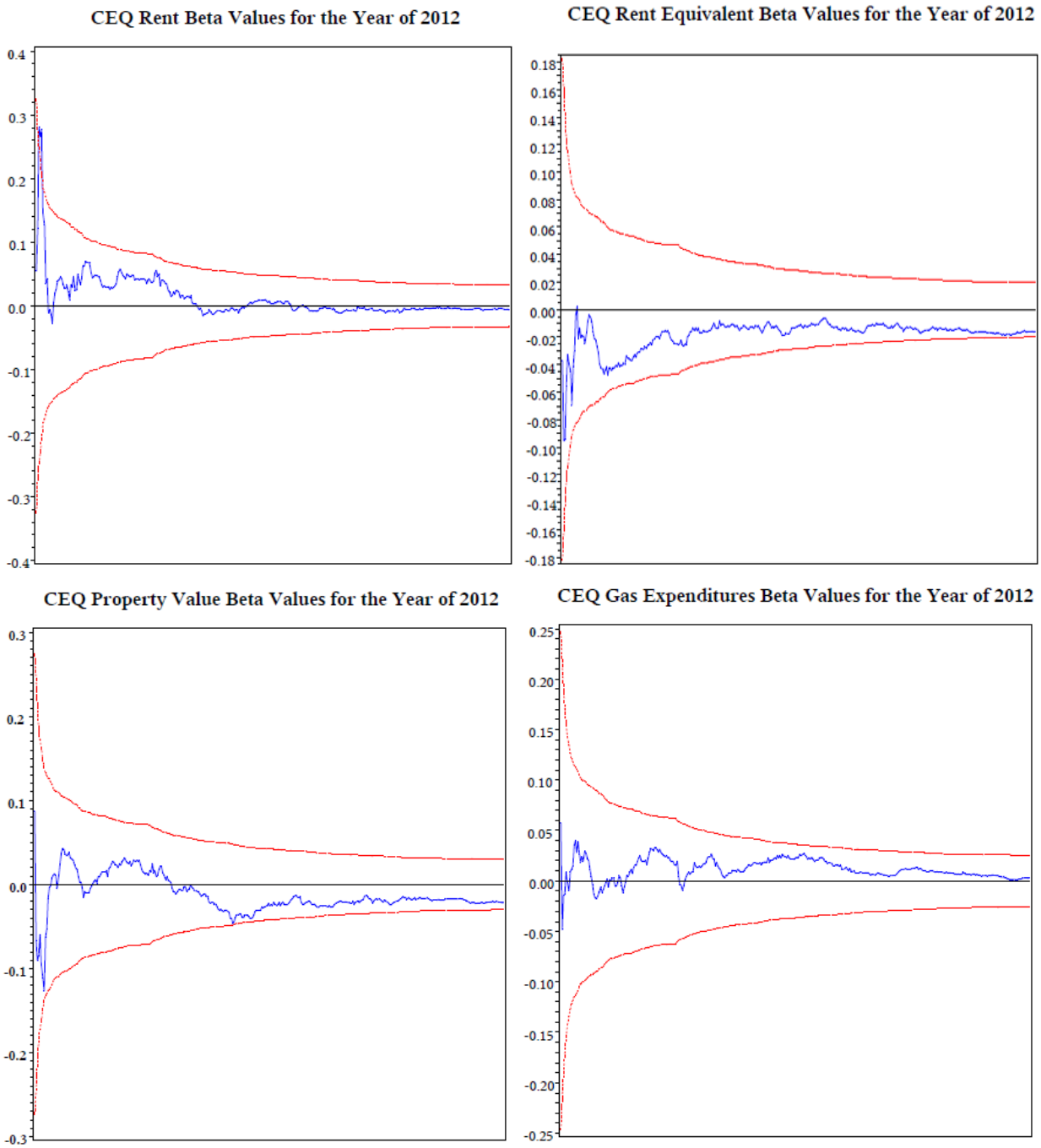
Figure 3: NCVS Beta Values for Waves 2 and 3 (Total Crime)

The graphs for waves 2 and 3 suggest the management change could have affected the crime data reported for the months of April through July. There could also be confounding factors specific to these waves and/or the crime rates for the geography in these two waves could be different from the others.

We believe that although there could be a management effect for the months of April through July, the final, yearly estimates for NCVS were not affected by the Census realignment.

Note: NCVS interviews are collected monthly. This means our interviewers have a whole month to collect interviews. The first week of the month, generally very few cases are reported. Because of this, the standard errors at the beginning of each wave are unusually large and unreliable due to the small number of cases used in the model fitting. More information flows in towards the end of the month. Thus our beta estimates at the end of the month are more reliable than at the beginning. For readability reasons, the few first days of each wave (the extreme standard errors) were removed from the graphs in Figure 3.

Figure 4 shows the z-var beta values for the CEQ models for the year of 2012.



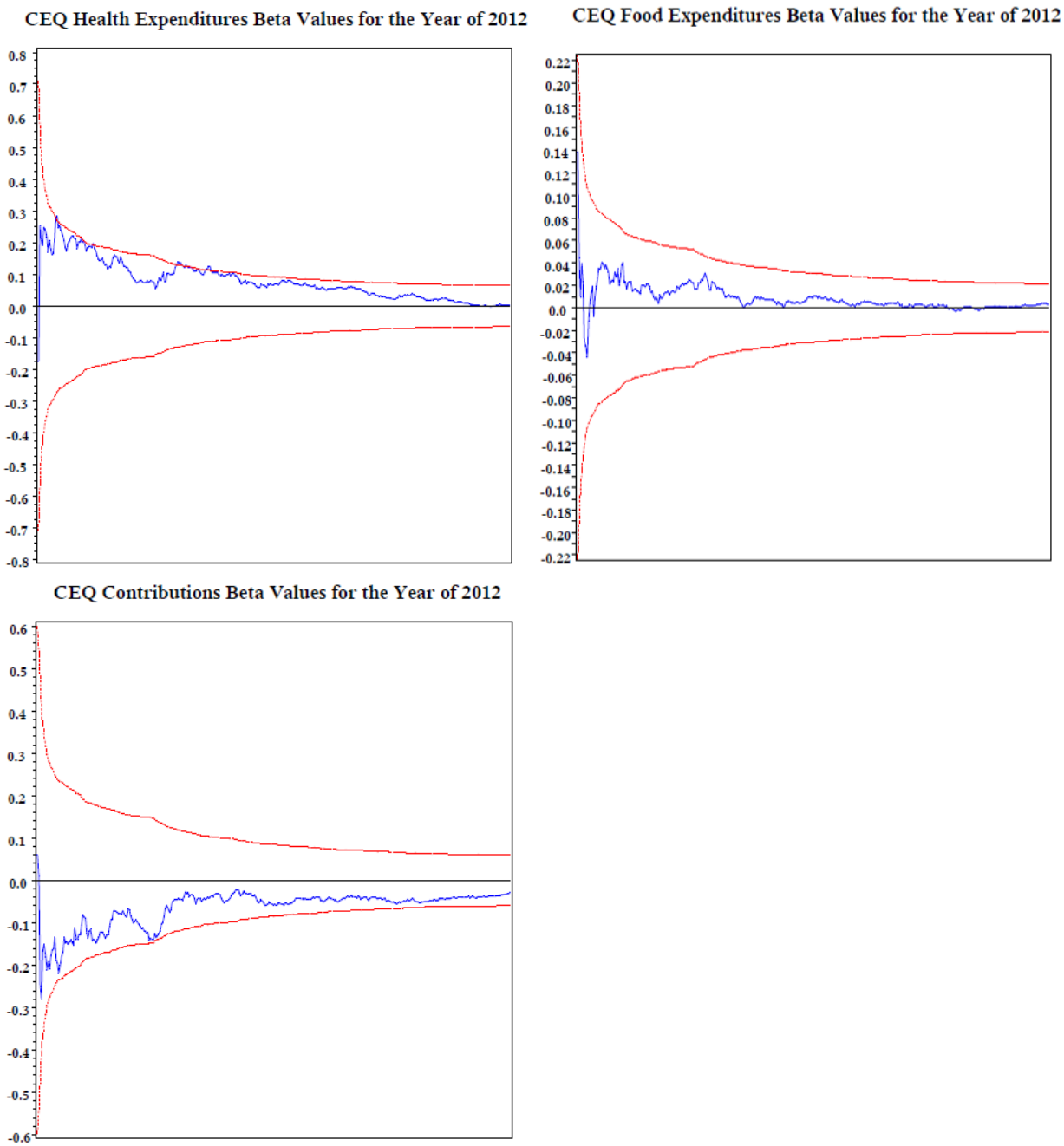


Figure 4: CEQ Beta Values for the entire year of 2012

The management variable for five of the seven CEQ models (Rent Equivalent, Property Value, Gas Expenditures, Food Expenditures, and Contributions) stayed within bounds for the entire year. For the remaining two models, z-var fell out of bounds 0.8% of the time for Rent and 4.2% of the time for Health Expenditures. These numbers fall within the expected 5% and thus we feel there is not evidence to support a new management effect on the CEQ estimates.

Conclusions

Overall, based on our findings for this project, we do not see sufficient evidence that the Census realignment affected the final estimates for NCVS and CEQ in 2012.

Limitations

Our biggest limitation was time. We were given roughly four months to develop and test our models and create a monitoring system. As a result, we decided to go with simpler, easy to interpret models and did not include such things as weights or random effects.

A limitation of the models used is the assumption of independence. There is most likely dependence among the cases collected by the same interviewer in the same month. This may cause the estimated standard errors to be too small.

For future projects of this nature, we recommend a mixed model analysis similar to that used in Schafer (2013).

References

Bureau of Labor Statistics. *Handbook of Methods*. Washington, D.C. 2009. Bureau of Labor Statistics. www.bls.gov/opub/hom. 24 July 2013.

Garland, Kathlene. (2009). U.S. Census Bureau Memorandum "General Description of the 2000 Sample Design on the National Crime Victimization Survey," from Kathlene Garland to Stephen Ash, dated October 22, 2009.

Killion, Ruth Ann. (2011). U.S. Census Bureau Memorandum "Description of Models Used to Monitor Key Variables for the Consumer Expenditure Quarterly Interview Survey," for documentation.

Killion, Ruth Ann. (2011). U.S. Census Bureau Memorandum "Description of Models Used to Monitor Key Variables for the National Crime Victimization Survey," for documentation.

Christy, James. "Responding to a Changing Future: Restructuring the Census Bureau Regional Offices." PowerPoint presentation. U.S. Census Bureau, Washington D.C. July, 2011.

U.S. Census Bureau Memorandum "Charter for the Demographic Data Monitoring Modeling Team," dated August 14, 2011

Schafer, Joseph L. (2013). "Modeling the Effects of Field Interventions in the 2012 National Crime Victimization Survey." Research Report (Statistics), Center for Statistical Research and Methodology, U.S. Census Bureau.