

Nonresponse Bias Analysis for the Passport Demand Survey

The art and science of nonresponse bias analysis, which aims to quantify the potential bias that could result due to differential nonresponse, provide an empirical foundation that serves two distinct objectives. First, a data-driven strategy will be developed to guide the process that will be used to weight the survey data. Specifically, prior to the final step of post-stratification or raking, design weights will be adjusted for nonresponse. As such, our nonresponse bias analysis will help identify effective subgroups within which this interim adjustment should take place to significantly reduce bias due to differential nonresponse.

Secondly, nonresponse bias analysis will identify domains for which the resulting survey estimates may be subject to non-ignorable bias. This is particularly important for surveys that experience low rates of response, for which some measure of bias may remain even after adjustments for differential nonresponse and calibration of weights to target totals. While specific steps that will be taken to implement nonresponse bias analysis for this survey will be established after availability of survey data, some of the main steps that will be taken are outlined below.

Typically, surveys that secure low to modest rates of response are expected to conduct a nonresponse bias analysis to assess the potential magnitude of nonresponse bias before the data or any analysis using the data are released. For this purpose, estimates of survey characteristics for nonrespondents and respondents are required to assess the potential nonresponse bias. Theoretically, the bias in an estimated mean, say, can be expressed as follows:

$$B(\bar{y}_R) = \bar{y}_R - \mu$$

In the above μ represents the unknown population parameter of interest that would result if a complete census of the entire population is conducted and all units respond, and \bar{y}_R represents its estimate based on data from respondents only. However, for variables that are available from the sampling frame for both respondents and nonrespondents, μ can be estimated by $\hat{\mu}$ and the bias in \bar{y}_R can then be approximated by:

$$\hat{B}(\bar{y}_R) = \bar{y}_R - \hat{\mu}$$

Moreover, with $\hat{\eta}$ representing the weighted unit nonresponse rate, $\hat{\mu}$ can be decomposed into two components by:

$$\hat{\mu} = (1 - \hat{\eta}) \bar{y}_R + \hat{\eta} \bar{y}_{NR}$$

Consequently, the bias in \bar{y}_R can be estimated by:

$$\begin{aligned} \hat{B}(\bar{y}_R) &= \bar{y}_R - \hat{\mu} \\ &= \bar{y}_R - [(1 - \hat{\eta}) \bar{y}_R + \hat{\eta} \bar{y}_{NR}] \\ &= \hat{\eta} (\bar{y}_R - \bar{y}_{NR}) \end{aligned}$$

That is, the estimate of the nonresponse bias is the difference between the mean for respondents and nonrespondents multiplied by the weighted nonresponse rate, using the design weight prior to nonresponse adjustment. Here, a respondent will be defined as any sample member who is determined to be eligible for the study and has valid data for the selected set of key analytical variables.

Once the potential bias for variables available for both respondents and nonrespondents have been estimated, they will be tested (after adjusting for multiple comparisons) to determine if the observed bias is significant at the 5 percent level significance. In the second step, nonresponse adjustment procedure

will be designed to reduce nonresponse bias based on the information obtained during the first step. In the third step and after the nonresponse-adjusted weights have been computed, for key variables statistical tests will be performed to determine the significance of any remaining nonresponse bias.

While the exact steps that will be taken to implement the above will be specified once survey data become available, we anticipate relying on relevant variables that are available from the sampling frame as well as other ancillary data that may be obtained from external sources. For instance, the sampling frame provided by MSG will append – among others – the following data items on each address:

1. Delivery Point Type Code
2. Drop, Seasonal, and Vacant Codes
3. County FIPS Code
4. Tract, Block Group, and Block
5. Latitude
6. Longitude
7. Dwelling Type
8. Walk Sequence

Moreover, the following data items will be available for various subsets of addresses. Presence or absence of these data items, which will be known for both respondents and nonrespondents, will provide a rich source of investigation for nonresponse bias as described earlier.

1. Phone (Landline)
2. Head of Household:
 - a. Surname
 - b. Age
 - c. Gender
 - d. Race
 - e. Ethnicity
 - f. Marital Status
 - g. Education
3. Number of Adults
4. Number of Children
5. Presence of 18 to 24 Year-Olds
6. Presence of 25 to 34 Year-Olds
7. Presence of 35 to 64 Year-Olds
8. Presence of Over 65 Year-Olds
9. Household Income
10. Own Rent Status

Also, from sources such as the American Community Survey we can obtain estimates for passport-relevant attributes, including:

1. Language
2. Ancestry
3. Foreign Born
4. International Migration
5. Place of Birth
6. Year of Entry

Obviously, the above estimates will be available only at higher levels of aggregation and not for individual households; nonetheless, they can provide further insights about key outcome measures of this study – passport demand.