WESTAT'S PROPOSED WEIGHTING PLAN FOR THE PDII RDD SURVEY

DRAFT
June 26, 2008

In order to obtain a sample of 2,000-2,500 respondents for the RDD survey, a self-weighting sample of 91,000 telephone numbers was selected from a national sampling frame consisting of all working banks with at least 1 listed residential number. The selected numbers were then screened to identify households with eligible respondents (i.e., adults 18 years of age or older in the labor force at the time of interview),and from these one respondent was randomly selected for the extended interview.

In general, the analysis of data from the RDD survey will require weights to compensate for variable selection probabilities, differential response rates and possible undercoverage of the sampling frame. Since the sampling process started with the selection of telephone numbers, an initial or "base weight" will first be attached to each telephone number. The weighting of subsequent sampling units (i.e., households and persons) will then follow. The weighting process will be carried out in the seven steps outlined below.

1.   Base Weight for Sample Telephone Numbers

The base weight for a sampled telephone number is the inverse of the selection probability. Each of the t sample telephone numbers was selected with a probability of t/T, where T represents the size of the sampling frame (number of telephone numbers in the working banks from which the t sampled numbers were drawn). For the PDII survey, the base weight for a sampled telephone number is:

$$w_i^{(0)} = \frac{T}{t} = \frac{288,865,000}{91,000} = 3174.34.$$

2.   Residential Status Adjustment

Only residential telephone numbers are of interest. Since it is not always possible to determine residential status, the weight of unknown cases (telephone numbers with unknown residential status) will be distributed so as to preserve the distribution of the cases for which residential status is known.

This adjustment will be carried out as a nonresponse adjustment within a number of adjustment cells defined by the Census region, metropolitan status, and minority status. All of these variables are exchange-level variables available in the RDD sampling frame. A cross-classification of these variables will yield a total of $L = 16$ adjustment cells.

- The 4 Census regions to be used are the following: (1) Northeast, (2) Midwest, (3) South, and (4) West;

- The 2 metropolitan-status categories are (1) MSA, and (2) nonMSA;

- The 2 minority-status categories are (1) low minority and (2) high minority.

Within each adjustment cell h, we will calculate the proportion of telephone numbers for which residential status was determined as follows:

$$\hat{q}_h = \frac{\sum\limits_{i \in S_h^{(r)}} w_i^{(0)}}{\sum\limits_{i \in S_h} w_i^{(0)}}$$

where $S_h^{(r)}$ is the set of telephone numbers in cell h for which residential status is known, and S_h is the set of all sample telephone numbers in cell h.

Within each adjustment cell, the base weight of a sampled telephone number will be adjusted as follows:

$$w_{hi}^{(r)} = \begin{cases} w_i^{(0)} / \hat{q}_h, & \text{if the residential status of the telephone number was determined,} \\ \\ 0, & \text{otherwise.} \end{cases}$$

3. Eligibility Status Adjustment

Some of the residential telephone numbers will have no eligible persons. These are telephone numbers associated with households that have no adults 18 years of age or older who are in the labor force at the time of interview. It is anticipated that for some residential telephone numbers, it will not be possible to ascertain eligibility status. Therefore, an eligibility status adjustment will be performed within the same adjustment cells specified in Step 2.

The household screening questionnaire determines how many eligible adults live in the household. There are three categories of interest. The first category includes households in which 1 or more residents are eligible for the extended interview. The second category is households in which no resident is eligible. The third category includes households in which eligibility is not known.

It is possible that some of the unknown cases are eligible. Therefore, the weight of the eligible cases will be adjusted upwards as follows:

- Within each of the adjustment cells defined above, the proportion of households with known eligibility status will be calculated as:

$$\hat{\theta}_h^{(e)} = \frac{\sum\limits_{i \in S_h^{(e)}} w_{hi}^{(r)}}{\sum\limits_{i \in S_h} w_{hi}^{(r)}}$$

where $S_h^{(e)}$ is the subsample of households with known eligibility status in adjustment cell h, and S_h is the subsample of identified households in adjustment cell h.

- The weight adjusted for eligibility status is given by:

$$
w_{hi}^{(re)} = \begin{cases} w_{hi}^{(R)} / \hat{\theta}_h^{(e)}, & \text{if eligibility status of the household was determined,} \\ \\ 0, & \text{otherwise.} \end{cases}
$$

4. Nonparticipation Adjustment for Eligible Households

Some eligible households will refuse to participate in the extended interview during the screening process. The adjustment for nonparticipation will be carried out within new adjustment cells defined by the Census region, MSA status, income level of exchange, and possibly other characteristics

For each adjustment cell l, the participation rate is estimated by:

$$
\hat{\phi}_l = \frac{\sum\limits_{j \square S_l^{(p)}} w_{lj}^{(re)}}{\sum\limits_{j \square S_l} w_{lj}^{(re)}}
$$

where $S_l^{(p)}$ is the set of sample households in cell l that agreed to participate in the study, and S_l, the set of all sample households in cell l.

The eligible household weight adjusted for nonparticipation is given by:

$$
w_{lj}^{(pre)} = w_{lj}^{(re)} / \hat{\phi}_l .
$$

5. Sample Person Weight

It should be noted that eligible persons in households having more than one residential telephone number will have a higher selection probability than eligible persons in households with only one residential telephone number. Therefore, the probability of selecting individual j is equal to the probability of selecting at least one residential telephone number from the household to which he or she belongs multiplied by the probability of selecting individual j within that household.

Each sample person j belongs to one household with a residential telephone number i. Let m_j be the number of eligible persons in the household in which sample person j resides, and let k_j be the number of telephone numbers asssociated with that household (in practice, we usually set $k_j = 1$ if there is only one residential telephone number, and $k_j = 2$ if there are 2 or more). The weight attached to sample person j in household i is then given by:

$$w_{ij}^{(prs)} = \frac{m_j}{k_j} w_{ij}^{(pre)} \quad .$$

6. Person-Level Nonresponse Adjustment

This adjustment compensates for nonresponse resulting from sample persons who agreed to participate in the study but for some reason did not complete the intake interview. The adjustment will be performed within cells defined by the following variables:

- Census region (1,2,3,4);

- MSA status (MSA, nonMSA);

- Median income level of exchange (4 or 5 categories);

- Employment Status

The final cells will be determined by a CHAID analysis. Note that the first three variables are exchange-level variables, whereas the remaining variables are person-level variables available from the screening interview. Also note that missing values of these variables will be imputed if necessary.

For each adjustment cell c defined by the CHAID analysis, we will estimate the response propensity by:

$$\hat{\psi}_c = \frac{\sum\limits_{j \square S_c^{(p)}} w_{cj}^{(prs)}}{\sum\limits_{j \square S_c} w_{cj}^{(prs)}}$$

where $S_c^{(p)}$ is the set of sample persons in cell c who completed the intake questionnaire, and S_c is the set of all sample persons in cell c who initially agreed to participate in the study.

The person-level weight adjusted for nonresponse is given by:

$$w_{2cj}^{(prs)} = w_{cj}^{(prs)} / \hat{\psi}_c \quad .$$

7. **Calibration Adjustments**

Finally, the nonresponse-adjusted weights will be adjusted (calibrated) to known population control totals using a raking algorithm. The nonresponse-adjusted weights will be calibrated to CPS population counts (marginal totals) defined by the following variables.

- Sex;

- Age group - suggested categories: *18-29, 30-39, 40-49, 50-59, 60+;*

- Educational attainment - suggested categories:

    - *Less than High School/GED*
    - *High School/GED*
    - *Some college/Tech School/Associate college degree*
    - *Bachelor's degree*
    - *Master's degree or higher;*

- Census region;

- MSA status;

- Race of sample person;

- Hispanic origin of sample person;

- Employment status

- Type of Employer