Application of Calibration Estimators in the Diary Data of the Finish Time Use Survey Based on Household Sample

Paavo Väisänen
Statistics Finland, Research and Development
P.O.Box 3A, FIN-00022 Tilastokeskus
Helsinki, Finland
Paavo.Vaisanen@stat.fi

1. Introduction

The time use survey (TUS) were conducted in 1999 - 2000 in Finland. Data were collected on household and individual level by using interviews and diaries. The sampling design was a single stage cluster sampling where households served as clusters and individuals were elementary units. The sample was based on a sample of individuals drawn from the Population Register by using a master sample which was a large sample serving as sampling frame for the second phase sampling. The selection of the sample was based on individuals, and the household to which the sampled person belonged was included in the sample. In this design sampling probabilities depended on the number of household members belonging to the sampling frame.

The estimates of such auxiliary variables as age and sex were benchmarked to the population totals by using calibration estimators (Deville et al., 1993). The weights were calibrated on the individual level, under the constraint that all individuals in a household had equal weight, by using the method presented by Lemaitre & Dufour (1987). Nonresponse was adjusted on household level by using response probabilities. Calibration estimators adjusted nonresponse bias according to auxiliary variables so special nonresponse adjustment on individual level was not needed. Several auxiliary variables were available from administrative files. The first phase sample was used for an adjustment of household size. Other auxiliary variables consisted of regional demographic data available from the Population Register.

In the TUS, the estimation needs some special steps due to the diaries and household sample therefore the estimators of the time use variables may be rather complicated. For instance, the allocation of diary days affects on the weighting.

2. Sampling design and weighting procedures

Individuals which were 10 years or older formed the target population. Households made up another target population for variables studied on a household level. The sample size was 4800 households including 10 978 individuals.

Weights were composed on several steps which took into account sampling design, the household level information of population and master sample, household level nonresponse adjustment, calibration of demographic data, the allocation of the sample to diary days, and balancing the diary weights to correspond equal size groups.

The household sample was composed around the sampled person. A household may enter the sample by anyone of the members which were belonging to the sampling frame. When the sample size is n households, and households are drawn with replacement then in the master sample, the inclusion probability weight of the household is $w_i = M/nm_i$ where M is the total number of population and m_i is the number of household members belonging to the sampling frame. The final sample was drawn from the master sample by simple random sampling, when the weight is the inverse of sampling fraction. Response homogeneity group model was used to adjust the household level nonresponse. The sample was divided into homogenous subsets according to household size and regions. The inclusion probability that a household

belonged to the response group h was estimated by using the observed response proportions. The nonresponse adjusted inclusion probability weight was composed by multiplying these weights which were calibrated according to gender and five years age groups, unemployed jobseeker and a week. The classes of the variables were transformed to binary variables so that the value of the binary variable was one if the individual belonged to the class and zero otherwise. The classes were transformed to the 94 variables for which, the household level frequencies were calculated and used as calibration variables. The weekly variation of the response rates of households and individuals was adjusted by including a constant in the calibration vector. The constant was received by dividing the total number of the population by 52. Calculations were done by using Calmar SAS-macro written by Sautory (1993). Individuals kept the diaries two days a week day and a weekend day, when the allocation dependent weights of diaries were 2/7 for weekend and 5/7 for weekdays. The sample was allocated equally to weeks and weekdays. Some persons returned only one diary. Nonresponse was high, 48%, having seasonal variation and in some cases the respondents were allowed to postpone the diary keeping one or two weeks which affected the variation to weekly and daily responses. Diary weights were balanced on the day level to the constant corresponding the value received by dividing the total number of population by 364.

Finally, the diary weight was composed as product $w_i = w_{Ai} w_B w_{Rh} g_i w_W w_D$ where w_{Ai} is the Master sample weight, w_{Bi} the weight of the sample of the TUS, w_{Ri} the household level nonresponse adjustment weight, g_i the calibration weight, w_W the diary allocation weight and w_D is the balancing weight of a diary day. The diary weights followed approximately lognormal distribution.

Calibration benchmarked the point estimates of the auxiliary variables to the true values. In the official statistics, the results received from different sources should be consistent. These numbers should correspond those published by Statistics Finland why the calibration according to the demographic variables was necessary. In the time use surveys, the sample often covers the whole year, and when seasonal variation in response rates occurs, the calibration should include the balancing of time periods.

REFERENCE

Deville, J.-C., and C.-E. Särndal (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, Vol. 87, No. 418, 376?382

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). General Raking Procedures in Survey Sampling. *JASA, September 1993. Vol.88, No. 423, Theory and Methods*, American Statistical Association.

Lemaitre, G. and Dufour, J. (1987). An Integrated method for Weighting Persons and Families. *Survey Methodology,13, no. 2*, pp 199-207. Statistics Canada

Sautory O. (1993) *La macro Calmar. Redressement d' un echantillon par calage sur marges.* Document no. F9310. Institut National de la Statistique et des etudes economiques. Direction Generale. 25.11.1993

RÉSUMÉ

Dans les enquêtes sur l'emploi du temps des ménages il faut des pondérations pour les ménages, les individus et les journaux. Des estimateurs de calibrage ont été appliqués pour standardiser le nombre de réponses variable d'une semaine à l'autre, de sorte que leur nombre corresponde à l'ensemble de la population de base. Dans la standardisation des journaux on a utilisé, en plus, des estimateurs de ratio. Le poids du journal est alors la combinaison du poids ménage-individu calibré et redressé et du poids journalier standardisé au moyen d'estimateurs de ratio.