Household Surveys

Duncan Thomas

CCPR-010-07

January 2007

California Center for Population Research
On-Line Working Paper Series

Household Surveys

Duncan Thomas

UCLA

January 2007

Abstract

Household surveys play a pivotal role in empirical economics. Cross-section and longitudinal surveys are regularly conducted throughout the globe. A description of survey design and sampling methods provides the foundation for discussing survey errors. These include errors associated with sampling, survey coverage and non response (which includes attrition from panel surveys) as well as errors of observation or measurement. In recent years, surveys have tended to become more complex and broader in scope with many reaching beyond measuring economic choices, constraints and outcomes. This trend will likely continue and exciting technological innovations in survey methods and implementation promise to revolutionize the field.

Household surveys provide one of the pillars upon which some of the most important innovations in economics during the last half of the 20th century have been built. Enumeration of households dates back at least to the collection of budget data in the late 18th century. Eden (1797) compiled information on the diet, dress, fuel, and habitation spending as well as earnings of households from 86 households in England, while Davies (1795) reported detailed budgets of 127 households engaged in agriculture. Both studies sought to describe the lot of the poorest in England and so the budgets are not representative of the English population at the time. Ducpétiaux (1855) published the budgets of 199 Belgian households. Those data provided the empirical foundation for Engel's Law (Engel, 1857) which posits an inverse relationship between income and the share of the budget spent on food.

Statistical foundations

The development of practical methods of probability sampling and a theory to support estimation and inference based on those samples had a major impact on the design and implementation of household surveys. Work by Neyman (1934) on stratified designs and work on randomization in agricultural experiments by Fisher (1935) were especially influential, and their work, in combination with contributions by inter alia Bowley (1926), Deming (1950), Kaier (1895) and Yates (1935) provided a theoretical foundation for survey design.

The importance of scientific surveys was underscored by some spectacular failures. For example, in 1936 the *Literary Digest* mailed out ten million questionnaires in a poll about the election of the next US president. About two million respondents mailed back their questionnaires, and the *Digest* predicted a victory for the Republican candidate, Alfred Landon. The election was won by a landslide – not by Landon but by his opponent, Franklin Roosevelt. There were also very influential survey successes. For example, Mahalanobis (1940) highlighted the advantages of surveys in terms of cost and timeliness of results. Using a sample survey of jute producers in Bengal, he estimated the area under jute within three per cent of the official estimate based on a complete census. The cost of the sample survey was only about eight per cent of the cost of the census..

These advances laid the foundation for an explosion in the quantity and quality of household surveys during the second half of the 20th century. Many of the surveys have been designed and implemented by national statistical agencies. At a substantive level, there are at least three important classes of household surveys, each of which has specific goals.

First, household budget surveys collect detailed information on the spending patterns of households. They are used to calculate price indices and poverty lines and to estimate the incidence of poverty. These include the Indian National Sample Survey, the Family Expenditure Survey (FES) in the UK and the Consumer Expenditure Survey (CEX) in the United States. Nowadays, virtually every country in the world conducts household budget surveys periodically. In some cases, respondents are asked to maintain a diary of spending over a pre-specified time period. In other surveys respondents are interviewed and asked to recall spending on items, often with varying recall periods depending on the item. The diary method typically covers a relatively short time period, which complicates modelling low frequency purchases and interpreting reported spending as indicative of longer-run resource availability. The interview method is potentially affected by recall error. This includes forgetting (which increases with the recall period) and telescoping, which may be positive (if spending before the recall time frame is telescoped into the recall period) or negative (if spending during the recall time frame is telescoped out of the period). Whether the interview or diary method yields less measurement error remains an open question.

Second, labour force and income surveys are collected routinely to monitor inter alia labour force participation, unemployment and earnings. Labour force surveys tend to be administered frequently and samples are large enough to detect small changes in the labour market. In the United States, for example, the Current Population Survey (CPS) is a monthly survey of over 50,000 households that has been conducted for over 50 years. Some surveys focus on income and wealth. The Survey of Consumer Finances measures the financial health of the US population and includes a special over-sample of the most wealthy households.

The third class of surveys measure non-economic domains of well-being. Fertility surveys provide information on marriage and living arrangements, reproductive health including pregnancies and births, and use of health services. These are important for documenting the dramatic changes in family formation, composition and size that has occurred over the 20th century. Health surveys monitor the health of the population. In some cases, such as the National Health and Nutrition Examination Survey (NHANES), an extensive physical examination is

performed by trained medical personnel in conjunction with a detailed questionnaire about health status and health-related behaviours. Several surveys integrate demographic with health information including the Demographic and Health Surveys (DHS), which grew out of the World Fertility Surveys and have been collected in over 75 countries. Surveys of attitudes, like the General Social Survey, are routinely collected across the globe.

In practice, the distinction between these classes of surveys is not clear-cut since many of the economic surveys record demographic, health or attitudinal information, and vice versa. To be sure, these topic-specific surveys are extremely important for monitoring the prevalence of indicators of interest to researchers and policymakers. However, the surveys are often inadequate for testing hypotheses about behaviours of individuals and their families.

In the late 1960s, surveys were designed to address this limitation, explicitly drawing on the theoretical models of household behaviour suggested by Becker, T.W. Schultz, and their collaborators and students. One class of surveys explicitly recognized the dual role of households in agricultural economies as both producers and consumers of food. See, for example, Evenson (1978) for a discussion of a series of innovative household surveys conducted by nutritionists and economists in Laguna Province, Philippines. These surveys collect detailed information on farm inputs and output, non-farm activities, consumption, health and demographic behaviour.

Another class of surveys relied on the economic model of household production to guide the collection of information on individual choices and constraints people face. For example, the RAND Malaysian Family Life Survey (MFLS) was designed to capture multiple domains of the lives of each individual respondent, their family and community to better understand the determinants of fertility and investment in children during early life (Butz and DaVanzo, 1975). As a result of the scope of the questions, MFLS has been used to address a far broader array of questions in economics and demography than those for which it was originally conceived. The International Crops Research Institute for Semi-Arid Tropics (ICRISAT) village-level studies (VLS) followed a similar approach. The best known of these was conducted in six villages in three regions of semi-arid India and collected very detailed data on a very broad array of topics from 240 farm households surveyed annually for ten years (Walker and Ryan, 1990).

The Living Standard Measurement Surveys (LSMS) conducted by the World Bank drew heavily on the experiences of the Laguna, MFLS and ICRISAT studies among others. Conceived as broad-purpose surveys to monitor poverty and material well-being in developing countries and

also contribute to the design of social policy, the surveys collect a wide array of indicators of well-being and behaviours of households along with extensive community data. Initiated in the mid-1980s, a hallmark of the LSMS program is a framework that is broadly consistent across many countries. Having been implemented in many low-income and transition countries around the world, LSMS and DHS stand out as leaders in the development of comparable survey data collected from a wide spectrum of social and economic contexts.

Survey design

A typical household survey selects a sample of households from a frame which is the population of interest for the research. In many cases, the frame is a census and the sample is representative of a geographic area, although this need not be the case. The simplest sampling strategy randomly selects households from the frame. In practice, most household surveys follow a two-stage (or multi-stage) sampling design in which clusters are selected and then households are selected from those clusters.

There are several advantages associated with geographically-defined clusters. Administration costs are lower for surveys that involve face-to-face interviews. Clusters may facilitate incorporating neighbourhood- or community-level data in the survey or, alternatively, models might highlight variation within communities and control community-level heterogeneity with a fixed effect, for example.

Clustering also carries disadvantages since two sampled units within a cluster tend to be more similar than two randomly selected units. The loss of independence across sampled units results in lower precision and thus larger standard errors of estimates. The magnitude of this effect for a particular indicator is often summarized by the design effect which is the ratio of the variance, with the cluster design taken into account, to the variance if households were randomly selected. An alternative summary statistic is provided by the intra-cluster correlation coefficient. The greater the covariance within clusters relative to differences across clusters, and the larger the number of households within a cluster, the greater is the design effect and the greater is the loss of precision due to clustering. It is standard practice to estimate standard errors by taking account of the clustering following the method of Huber (1967) or a re-sampling approach such as the jackknife or bootstrap (Efron, 1982). In short, clustering buys more information per unit cost but less information per sampled unit.

Many surveys are designed to oversample specific sub-populations, in which case estimates are typically adjusted for the probability of a household being selected into the sample. An important principle underlying population-based sampling is that because the probability of selection of every eligible unit is known and greater than zero, with appropriate weights, it is possible to reconstruct the population, although in some instances the complexity of survey designs becomes overwhelming.

Survey errors

There are at least two classes of error in any survey. 'Non-observational' errors occur when part of the target population is not measured. 'Observational' errors are the result of incorrect measurement.

Sampling error, the most familiar survey error, is a form of non-observational error. It reflects the fact that any sample is a subset of the underlying population and so an estimate based on the sample will not be identical to the population value.

Coverage error, another source of non-observational error, arises when the sampling frame excludes part of the target population. Many sample frames are based on a list of household dwellings; those samples exclude homeless people and so are not representative of the entire population. If a household listing is based on an old census, more mobile people are at risk of being under-represented. A sampling frame based on telephone numbers (or e-mail addresses) will exclude those who do not have a telephone (or e-mail address) and oversample those with multiple numbers (or addresses).

A third source of non-observation error arises from non-response, of which there are two categories. First, survey non-response occurs when a target respondent cannot be located. It will also arise if the respondent refuses to participate in the survey (or fails to answer the telephone, respond to an e-mail or return a mailed-out survey). Second, item non-response occurs when a respondent fails to answer one or more questions in the survey either because he or she refuses to answer or does not know the answer to the question(s). The incidence of the latter is reduced by probing, and unfolding brackets have proved to be particularly useful for economic quantities (Hurd et al., 1998).

Broadly speaking, non-response rates tend to rise with the value of time of the respondent, and there has been a secular trend of increased non-response in many developed countries. Survey

non-response in developing country household surveys is typically substantially lower than in higher income countries.

If, conditional on observed characteristics, coverage and non-response error are random, appropriate weights can be computed so that survey statistics are representative of the underlying population. Complications arise when these errors are selected on unobserved characteristics. Several procedures have been suggested to deal with non-response error including hot deck or matching procedures (Rosenbaum and Rubin, 1983) and modelling the selection process with a control function (Heckman, 1978).

The most familiar source of observational error is respondent failure to answer a question correctly. This may be intentional (in order to misrepresent reality) or unintentional. Interviewers may make errors in the administration of the survey, and there may be interviewer-specific effects in the ways questions are asked. Survey instruments are also prone to error. In general, the extent of observational error likely depends on interactions among the sources of error and also on the mode of the survey. Respondents in telephone surveys tend to provide shorter answers than those in face-to-face interviews, and web-based surveys are more likely to be ended prematurely.

While the distinction between observational and non-observational error is conceptually useful, in practice the distinction is often blurred. For example, survey non-response is typically related to interviewer characteristics. Both item non-response and respondent error have been shown to be related to questionnaire design and interviewer characteristics. Groves (1989) provides an excellent discussion of these and related issues.

Typology of surveys

Cross-section surveys provide a snap-shot of a target population at a point in time. They are the bread and butter of research based on household surveys. Many cross-section surveys are repeated regularly, with independent samples drawn from the same target population, so that it is possible to track the evolution over time of indicators such as unemployment, poverty or inequality as well as map changes for population sub-groups. Synthetic panels of individuals created using repeated cross-section follow the same population subgroup over time, such as a birth cohort. They are straightforward to interpret if there are no entrants into or exits from the target population via, for example, immigration, emigration or death. Synthetic panels of households are more complicated. Household composition changes due, for example, to marriage or divorce result in

changes over time in the unit being followed. It is difficult to distinguish composition changes from true change. Similar issues arise with synthetic panels of communities.

Longitudinal or panel surveys follow the same respondent over time, which provides opportunities for exploration not feasible with cross-section surveys. First, tracing the dynamic evolution of choices and outcomes over the individual's life provides insights into, for example, early life experiences and later life outcomes, resilience and recovery from adversity as well as the characteristics of those who cycle in and out of some state (such as poverty, unemployment, public assistance or poor health).

Second, panel data provide expanded options for treating unobserved heterogeneity in models like

$$y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$$

where μ_i is an unobserved individual-specific characteristic. If μ_i is correlated with x_{it}, OLS estimates of β are biased. With repeated observations on the same individual in a panel, μ_i can be estimated (or the model cast in first-differences) to consistently estimate β. The 'fixed effect' μ_i absorbs all time-invariant individual characteristics that enter the model in a linear and additive way.

The advantage of a longitudinal survey is that the same sampling unit is followed over time. This is also its Achilles heel. Attrition from longitudinal surveys is a particular form of non-response error. The nature and magnitude of attrition varies with the study design. For example, in face-to-face interviews in the home, individuals who move are followed to their new location and interviewed there. Those who move the furthest are often the hardest and most expensive to find. Attrition tends to be selected on traits associated with migration – younger, better-educated adults being the most like to move. The selectivity of the sample is exacerbated in panel surveys that do not follow people who leave the location in which they were interviewed at baseline. Attrition in telephone and web-based surveys have less to do with tracking people to new geographical locations and more to do with retaining the cooperation of respondents – an issue that also confronts face-to-face interviews. In multi-wave panel surveys, it is important to attempt to re-contact respondents who have been skipped in prior waves so that attrition does not cumulate. There are many examples of well-designed panel surveys that have kept attrition low across multiple rounds.

Statistical adjustments for attrition are the same as other forms of non-response error. Re-weighting will be effective when attrition is selected on observed characteristics. When selection is on unobserved characteristics, a control function approach is more likely to be successful. In analytical models, the importance of adjusting for attrition will vary with the research question. The stronger the association is between attrition and observed or unobserved characteristics in the model, the more important the adjustment is for attrition.

An alternative approach to treating attrition is to replace a respondent who attrits from the survey with a new, similar respondent – frequently people living in the same housing structure as the respondents in the previous wave (or the person who is assigned the telephone number, e-mail address, and so forth). There are several problems with this approach. First, it assures the study population appears stable since no primary sampling units will lose population; the reality may be quite different. Second, housing structures can change, be torn down or difficult to relocate, resulting in a different type of attrition. Third, even if populations are stable in aggregate and housing structures do not change, it is assumed that the replacement and original respondents are 'exchangeable' or effectively identical. It is not clear that this will be true as in the case of a respondent who died. Fourth, the key advantage of a longitudinal survey – following the same person through the life course – is lost.

It follows that a panel survey of households has little conceptual appeal. Although a household survey is often the baseline for a panel of individuals, households change over time and it is individuals who will be followed – possibly all the original household members. These respondents will often be interviewed along with the people in their new household and so the panel is a series of household surveys embedded in which is a longitudinal survey of individuals. A small number of longitudinal surveys has sought to follow family members over time.

The Panel Survey of Income Dynamics (PSID) is a long-running panel and one of the most widely used surveys in economics. Initiated in 1968, with a nationally representative sample of 5,000 households, interviews spanning 40 years with household members, and children born to them, has provided unique insights into the dynamics of income, human capital, health and living arrangements over the life course, across cohorts and across generations (Duncan, Hofferth and Stafford, 2004).

A cohort survey is a special type of longitudinal survey which follows a specific cohort of respondents, often a birth cohort. The advantages of the design are that, because of shared

environments, cohort members are less heterogeneous than the entire population and there are power benefits to comparing people making similar life course transitions at the same time. A disadvantage is that age and period effects cannot be disentangled. To address this, cohort studies often draw new cohorts. The British Cohort Studies, for example, have mounted four large-scale population-representative birth cohort studies since the 1930s. The Health and Retirement Survey (HRS) is an innovative cohort study that focuses on the health and economic well-being of older Americans. The HRS has been replicated in several countries across the globe.

Statistical innovation presaged the explosion in household surveys since the 1950s. Technological innovation is likely to provide the foundation for the next revolution in survey design. For example, electronic communication devices, geographical information systems and innovations in health measurement along with sophisticated analytical tools have already begun to profoundly affect the scope and quality of household surveys.

Bibliography

Bowley, A. L. 1926. Measurement of precision attained in sampling, *Bulletin de l'Institut International de Statistique,* 22 Supplement to Livre 1, 6-62.

Butz, W.P. and DaVanzo, J. 1975. *The Malaysian Family Life Survey: Summary Report*. Santa Monica: The Rand Corporation

Davies, D. 1795. *The State of Labourers in Husbandry Stated and Considered.* Bath.

Deming, W. E. 1950. *Some theory of sampling*. New York: Wiley.

Ducpetiaux, E. 1855. *Budgets économiques des classes ouvrières en Belgique.* Bruxelles.

Duncan, G., Hofferth, S. and Stafford, F. 2004. Evolution and change in family income and wealth: the Panel Survey of Income Dynamics 1968–2000 and beyond. In *A Telescope on Society: Survey Research and Social Science at the University of Michigan and Beyond*, ed. J. House, T. Juster and R. Kahn. Ann Arbor: University of Michigan Press.

Eden, F.M. 1797. *The State of the Poor*. London: J. Davis.

Efron, B. 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: SIAM.

Engel, E. 1857. Die Productions- und Consumptions verhältnisse des Königreichs Sachsen. *Zeitschrift des Statistischen Büreaus des Königlich Sächsischen Ministerium des Innern* 8(9), 22 November. Repr. in *Bulletin de I'Institut International de la Statistique* 9 (1985), 1–54.

Evenson, R. E. 1978. Time allocation in rural Philippine households. *American Journal of Agricultural Economics* 60, 322–30.

Fisher, R.A. 1935. *The Design of Experiments*. London: Oliver & Boyd.

Groves, R. M. 1989. *Survey Errors and Survey Costs.* New York: Wiley.

Heckman, J. J. 1978. Dummy endogenous variables in a simultaneous equation

system. *Econometrica* 46, 931–59.

Huber, P. J. 1967. The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 4, 221–33.

Hurd, M.D. McFadden, D. Chand, H. Gan, L. Merrill, A. and Roberts, M**.** 1998. "Consumption and Saving Balances of the Elderly: Experimental Evidence on Survey Response Bias" in D. Wise (ed.) *Frontiers in the Economics of Aging,* Chicago: University of Chicago Press, pp 353-387.

Kaier, A. N. 1895. Observations et expériences concernant des dénombrements représentatifs. *Bulletin of the International Statistical Institute,* 9, Livre 2, 176-183.

Mahalanobis, P. C. 1940. Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, *A* 109, 329–78.

Neyman, J. 1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97, 558–625.

Rosenbaum P.R. and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55

Walker, T.S. and Ryan, J.G**.** 1990. *Village and household economies in India 's Semi-Arid Tropics*. Baltimore : Johns Hopkins University Press

Yates, F. 1935. Complex experiments. *Journal of the Royal Statistical Society*, Supplement 2, 181-247.