Transpn. Res.-A Vol. 21A, No. 4/5, pp. 249–263, 1987
Printed in Great Britain.

0191-2607/87 $3.00 + .00

THE ROLE OF PANEL STUDIES IN RESEARCH ON ECONOMIC BEHAVIOR

GREG J. DUNCAN, F. THOMAS JUSTER and JAMES N. MORGAN
Survey Research Center and Economics Department, University of Michigan, P.O. Box 1248, Ann Arbor, Michigan 48106, USA

Abstract—The analytic and monetary costs and benefits of panel surveys are assessed in light of experiences from the Panel Study of Income Dynamics, an 18-year panel survey on the economic status and behavior of the U.S. population. The analytic benefits of panel are formidable, ranging from description of gross change to various analytic advantages of continuous and discrete time modelling. Analytic costs such as the conditioning of responses in subsequent participation or nonresponse bias are possible in panel surveys, but their effects can be minimized with proper data collection procedures and analytic adjustments. Surprisingly, the monetary costs of panel surveys are less than the costs of comparable repeated cross-sectional surveys.

INTRODUCTION

Although numerous panel studies of specialized topics on specialized samples had been conducted prior to 1970, a major development in social science research in the United States in the 1970's and in several other Western nations in the 1980's is the collection and analysis of large-scale, nationally representative panel studies. Much has been learned from these studies because they provide more reliable information on changes in the objective and subjective circumstances of their samples. In addition, they have provided much more detailed point-in-time information than many existing cross-sectional studies and for much larger samples. As a result, they have also been viewed by analysts as better sources of cross-sectional information. But, because of their size and scope, they are seen as expensive ways of collecting data.

The focus of this paper is on the relative costs and benefits of panel studies, with special reference to mobility and transportation behavior. We first illustrate the analytic advantages of panel designs with several striking findings from the Panel Study of Income Dynamics, a large-scale, general purpose panel study of the economic status of the American population begun in 1968. These examples are followed by a more general discussion of the usefulnmess of panel data for testing economic behavioral theories. Next discussed are the possible analytic costs of panel designs, including bias due to nonresponse and panel conditioning. We conclude that these costs are likely to be small in well-run surveys of economic behavior.

Analytic costs and benefits must be balanced against monetary costs, and the monetary costs of a population sample survey in the United States are the subject of the third section of the paper. Not unexpectedly, we find that a two-wave panel design is somewhat more expensive than a single-wave cross-sectional study, even when the questionnaire of the cross-sectional study is lengthened to attempt to measure change through retrospective questions. Field costs of the panel are 20% more expensive if the waves are one year apart and 25% more expensive if the waves are five years apart. But a more appropriate comparison of costs is often between a panel design and a series of cross sections. Contrary to conventional wisdom, many panel designs, especially those that measure family-level variables, need not become less representative of younger families and individuals over time due to the "aging" of their sample. Often it is possible to represent in the sample new families and individuals "born" into the population and thus maintain a representative sample over time—a sample that is as well suited for making point-in-time estimates as would be possible from interviews from a fresh cross section of the population. In these cases, the field costs of a panel design are 40% less than a series of comparable cross-sectional surveys, although these cost differentials will be offset to some extent if panel respondents are paid for their cooperation. The reputation that panel studies have as more costly alternatives to cross-sectional studies seems

to have come from the size and specialized nature of the samples chosen for those studies, as well as from the fact that comparisons are often made implicitly between single cross sections and panels with multiple observations.

I. THE ANALYTIC BENEFITS OF PANEL DESIGN

The terminology surrounding survey designs across time is somewhat ambiguous, so it is best to begin with some definitions (Duncan and Kalton, 1985). A repeated survey makes similar measurements on samples from an equivalent population at different points in time, but without an attempt to ensure that any respondent is included in more than one round of data collection. A panel survey is one in which similar measurements are made on the same sample at different points in time. There are several different types of panel surveys. Panel surveys that are based on elements from population subgroups that have shared a similiar experience—i.e. birth during a given year, marriage during a given month—are called cohort studies. A panel survey in which sample elements are kept in the panel for only a portion of the duration of the survey is called a rotating panel survey. A split panel survey is a combination of a panel and a repeated or rotating panel survey. The term "longitudinal survey" is sometimes used interchangeably with "panel surveys."

It is important to bear in mind the distinction between panel (longitudinal) surveys and panel (longitudinal) data. The latter consist of continuous or periodic measurements of the variables of interest. In theory it is possible to obtain reliable panel data from a survey conducted at a single point in time. But measurement considerations usually argue for the use of a panel survey design rather than retrospective questioning to collect reliable panel data on individuals. Thus, in large part, the analytic advantages derived from panel survey designs are synonymous with those inherent in reliable panel data.

Reliable panel data provide several analytic advantages over cross-sectional data. First, they make possible reliable measurement of change in the objective and subjective situation of families and individuals and enable the analyst to estimate truly dynamic models. For example, most attitudes can be measured reliably only at the time of the survey. Models of attitude change require repeated measurements of the attitudes, obtainable only in panels. If accurate measurements of commuting behavior, utility expenditures or family income history are required for the analysis, such data cannot be obtained reliably from retrospective questions asked of a cross section, but must be obtained by repeated measurements from a panel survey. Although it is possible to obtain some retrospective information about circumstances surrounding events of interest [e.g. Mackett (1983) on the effects of travel alternatives on residential mobility], information about financial conditions or, most importantly, perceptions of alternatives available at the time choices were made are virtually impossible to obtain retrospectively. Thus, any modeling effort that involves dynamic processes can only be carried out effectively, with rare exceptions, when the data come from a panel survey with repeated measures from the same households or individuals.

Baanders and Slootman (1983) summarize a number of ways in which repeated measurement from a household panel would enhance our understanding of travel behavior. These include (1) describing and analyzing changes in travel behavior in response to changing prices or the availability of public transportation, (2) analyzing the sequencing of joint decisions about place of residence, place of work and home–work trips, (3) understanding changes in energy consumption in response to changes in energy prices, (4) forecasting car ownership and driving licenses and (5) estimating the price elasticity of public transportation by measuring behavior before and after price changes.

Second, a panel design reduces the amount of time between an event of interest and the date of an interview and therefore reduces the imprecision of reporting the event due to memory loss. Memory losses are severe enough in some cases to make panel designs the only feasible method for collecting some kinds of data. A classic example is the study of time use, where methodological work has shown that the most reliable method for collecting general time use data is a 24-hour recall of activities (Robinson,

1985; Juster, 1985).* A single wave of interviews that collects one such time "diary" per respondent is sufficient to give reliable aggregate information on time use, but precise time use information at the level of the individual or household requires several diaries from the same respondent, and thus a panel format. Balanced against this benefit of improved data quality is the possibility that responses in a given panel wave may be affected by past participation in that panel and, therefore, contain a panel bias effect. This possibility is discussed below.

Finally, there is a statistical advantage to estimating net change from panels rather than from repeated cross sections: sampling errors of the net change between two points in time are usually substantially lower when the measurements are obtained from a longitudinal panel than when they are obtained from repeated cross sections. One element that reduces the variance of net change—the covariance between the levels at times one and two—is zero with two independent samples but usually positive in a panel.

The role of panel data in describing and explaining change

The basic descriptive information on individual change that has come from panel studies challenges much of the conventional wisdom that has been based on cross-sectional data. A set of cross-sectional surveys provides a series of "snapshots" of the population at several points in time. The apparent stability of the population that is inferred from the similarity of the snapshot pictures is almost always incorrect, as can be seen most clearly in aspects of the distribution of family income.

• Successive cross sections from the U.S. Census Bureau show a steady decline in the extent of poverty during the 1960's but little change during the 1970's. Year-to-year changes in these estimates are typically quite small, and the characteristics of those found to be poor also changes little from one year to the next. These facts have created the stereotype that the same families are poor from one year to the next. In fact, the Panel Study of Income Dynamics has shown remarkable turnover in the poverty population. Only a little over half of the poor in a given year are poor in the next, and most individuals who come into contact with poverty do so for very brief periods of time (Hill, 1982).

• Administrative records on welfare programs show steady growth in the costs and numbers of recipients of welfare programs during the 1970s in the United States. When combined with evidence from case studies, the yearly figures are taken as evidence that the typical welfare case is on the rolls for a long time and that the welfare system traps generations of families into dependency. But Panel Study data show that, as with poverty, the typical episode with the welfare system is brief, assisting families in the early stages of recovery from an economic crisis caused by the death, departure, or disability of a husband, a process that often culminated in finding full-time employment, remarriage, or both. Furthermore, most of the children coming from welfare families were not themselves receiving welfare benefits after they had left home and formed their own households (Coe, 1982).

More important than their ability to describe change is the usefulness of longitudinal data for explaining change through the testing of dynamic models of economic and social processes. Longitudinal data often provide more statistical leverage in sorting out causal patterns because they enable the analyst to separate effects of persistent interpersonal differences from real inter-temporal relationships. Some of these advantages can be illustrated more formally.

Social scientists have spent a great deal of time and effort in testing their theories with cross-sectional data. Typical of the empirical specifications are variants of a simple

*This methodological work compared 24-hour recall methods and activity-specific recall methods to patterns of time use obtained from activities recorded at the times that a beeper programmed to go off at random times during the day sounded. The 24-hour recall patterns were much closer to the beeper patterns. Godard (1983) reports a similar inconsistency between "previous day" and "habitual behavior" questions regarding travel behavior, but did not validate the two reports against a third, more accurate source.

linear and additive model relating two independent variables, X, and Z, to Y, the dependent variable:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_{it} + \epsilon_{it} \tag{1}$$

where Y_{it} is the value of the dependent variable for the i^{th} individual at time t, and X_{it} and Z_{it} are values of the two independent variables for the i^{th} individual at time t. In most survey applications, t corresponds to the time of the interview. Although the model as specified by eqn (1) is static in the sense that all of the measures are taken at a single point in time, Schoenberg (1977) has shown that such models are generally consistent with dynamic processes in which the state at time t is a function of the initial conditions of the system. Such models are termed "nonergodic" and are contrasted with ergodic models in which the past can be forgotten in the sense that the current state does not depend upon the initial conditions.

Suppose that a panel survey design gives us measures of Y, X, and Z for the same individual at a subsequent point in time, say $t + 1$, and further suppose that β_1 and β_2 (the effects of X and Z and Y, respectively) do not change between t and $t + 1$. Can those panel data give us better estimates of the crucial parameters in eqn (1)?

The answer to this question depends upon the measurement properties of the X and Z variables and upon the nature of the error term in eqn (1). Under certain circumstances, a simple difference equation such as (2) below, in which changes in Y are regressed on changes in the X and Z variables, give better estimates of β_1 and β_2 than would come from estimation of eqn (1):

$$\Delta Y = \Delta\beta_0 + \beta_1 \Delta X + \beta_2 \Delta Z + \Delta\epsilon \tag{2}$$

where Δ represents simple change between t and $t + 1$ and the individual subscript i has been (now and hereafter) suppressed for notational convenience. Those circumstances include:

1. *Instances where the* Z *variables are unmeasured and unchanging.* Most social science theories reserve a place for the effects of unchanging "taste," background or personality factors. Typically these factors are not measured. If they are correlated with observed X variables and with the Y variables, then estimation of eqn (1) with cross-sectional data may well be biased by the omission of these Z-variables. (This is demonstrated more formally below.) A change equation like (2) solves this omitted variable problem because changes in these unchanging Z-variables are, by definition, zero and eqn (2) reduces to a regression of ΔY on ΔX. The parameter β_1 in this change equation corresponds to the same parameter in level eqn (1), but the estimation problems caused by the Z-variable have been eliminated. Note, however, that this method of obtaining cleaner parameter estimates of observed X variables will only work in instances where the values of X change for a substantial portion of individuals over time.* Less obvious but also crucial is the assumption that the process that generates changes in X is independent of the error term in (2).

2. *Instances where observed* X *variables are measured with errors that persist over time.* Measurement error in the X variable is likely to bias estimates of β_1 in (1) toward zero. To see this, suppose that the "true" value of X_t is related to the observed value of X_t by:

$$X_t^* = X_t + u_t \tag{3}$$

where X_t^* is the observed amount of X at time t, X_t is the true amount, $u_t \sim N(0, \sigma_u^2)$, and $\text{cov}(u_t, Y_t) = \text{cov}(u_t, \epsilon_t) = 0$.

For the cross-sectional eqn (1), it can be shown that as the sample size becomes

*Hausman and Taylor (1980) develop a method for estimating parameters of unchanging X-variables in first difference equations.

very large, the OLS estimate of β_1 will not, in general, collapse on the value β_1. Instead, this probability limit (*p*lim) will be:

$$p\lim(\hat{\beta}_1) = \frac{\beta_1 + \beta_2 b_{ZX}}{1 + [\mathrm{Var}(u_t)/\mathrm{Var}(X_t)]} \tag{4}$$

where b_{ZX} is the regression coefficient of X when Z is the dependent variable.

There are two sources of inconsistency affecting the estimate of β_1 from cross-sectional eqn (1), the first one due to measurement error and the other due to the correlation between the omitted Z-variable and X. In general, the larger the measurement error ($\mathrm{Var}(u_T)$) and the greater the correlation between X and Z, the greater the inconsistency of the OLS estimate of β_1 from eqn (1).

But now consider the change eqn (2). If measurement error in the X variable is correlated over time according to:

$$u_t = \rho u_{t-1} + v_t, \tag{5}$$

and if the Z variable is unchanging, then the probability limit of the OLS estimate of parameter β_1 from change eqn (2) is:

$$p\lim(\hat{\beta}_1) = \frac{\beta_1}{1 + [(1 - \rho)^2 \mathrm{Var}(u_t) + \mathrm{Var}(v_t)]/\mathrm{Var}(\Delta X)} \tag{6}$$

where $\mathrm{Var}(\Delta X) = \mathrm{Var}(X_t) + \mathrm{Var}(X_{t-1}) - 2\,\mathrm{Cov}(X_t, X_{t-1})$. If there is no autocorrelation in the measurement error in the X variable (i.e. $\rho = 0$) and no biasing effect of unmeasured Z variables, then the cross-sectional form is generally preferred to the change form since the variance of ΔX will generally be smaller than the variance of X. On the other hand, a remarkable result from (6) is that perfect autocorrelation between the measurement errors of X at the same two points in time (i.e. $\rho = 1$ and $v_t = 0$) will cause no estimation problems for the method of first differences; the OLS-estimate of β_1 from (2) is consistent.

Although perfect autocorrelation will be rare in survey data, substantial positive autocorrelation is likely to arise when respondents persistently over- or understate responses or persistently misinterpret questions (Duncan and Hill, 1985). In an analysis of compensating wage differentials, Duncan and Holmlund (1983) argue that errors in reporting on working conditions were likely to contain a great deal of serial correlation, since workers will consistently apply their own standards to what constitutes "unhealthy" or "unsafe" conditions. They find much more reasonable estimates of wage functions when using a first difference form.

3. *Instances where the panel data give more reliable measurement of changes in* X *between* t *and* t + 1 *than in the level of* X *at time* t. Suppose that Y is earned income and X is work experience. Cross-sectional surveys obtain measures of X retrospectively, a procedure filled with possible memory error. But a panel that provides annual measures of work hours over the period between t and $t + 1$ will provide relatively more reliable measurement of changes in X between t and $t + 1$ (Corcoran *et al.*, 1983).

In sum, the case for the superiority of change eqn (2) over level eqn (1) is obvious from a comparison of (6) and (4). The advantages of first difference equations over cross-sectional formulations are greatest when: (1) powerful, unmeasured Z variables bias the cross-sectional estimates; (2) errors in the X variable over time are highly autocorrelated; and (3) ΔX is measured more reliably than X. The advantages of first differences are weakened (and cross-sectional formulations may be preferred) when these conditions are not true and when the X variables are highly correlated over time.

Extensions of model (1) to include lags in both dependent and independent variables are common in transportation, energy and other areas of consumer research. Houthakker *et al.* (1974) develop a distributed lag model of gasoline demand and estimate it with

quarterly time series data at the state level. Such models can also be estimated with panel data on individuals or families, with the added advantage that response parameters can differ by subgroup.

Panel data are also instrumental in estimating models of discrete events and transitions between states.* Examples include changes in the states of commuting mode (Johnson and Hensher, 1982), states of employment, unemployment, and out of the labor force (Flinn and Heckman, 1982), states of poverty and nonpoverty (Hill, 1982), marital states (Hannan *et al.*, 1977), and occupations (Pullman, 1978). Individuals found in a particular state at a point in time typically have a much higher than average probability of being in that state at a subsequent period of time. A crucial question is whether the association between states comes about because there is something about the state itself that changes the likelihood of remaining in that state, or because some unmeasured and unchanging set of personal characteristics of individuals causes observed states to persist? To illustrate, is it the experience of unemployment or poverty itself that causes subsequent unemployment or poverty, or is it personal characteristics of the unemployed or poor that increases their chances of remaining in those states. These two possibilities, termed respectively "state dependence" and "heterogeneity," are among the many problems associated with the identification of models of the duration of events (Heckman and Singer, 1982).

Although there have been many event history analyses of states and transitions between states based on data recalled in a single interview, what little methodological work that has been done on episodic recall of such events suggests that respondents have substantial problems in recalling the timing of past events with an acceptable degree of accuracy. A more detailed review of the evidence is given below, but two recent examples can serve to illustrate the problem. First, Mathiowetz and Duncan (forthcoming) found that nearly two-thirds of all spells of unemployment experienced by a sample of workers during a 30-month period were not reported in an interview with those workers. Second, rates of incorrect reporting of occupational status were also quite high when the time between the interview and the data being recalled exceeded six months (Duncan and Mathiowetz, forthcoming). The extent of the errors in these data are likely to cause serious bias if they are used to estimate event history models.

Longitudinal surveys can provide more accurate information to test these models in two ways. First, they can provide a set of point-in-time observations on the state of individuals or families in the sample. Second, and perhaps more important, they increase the accuracy of the report of the entire history of the event, thus reducing measurement errors and providing greater statistical leverage in the estimation of continuous time models.

Panels as successive cross sections

A common but mistaken criticism of longitudinal designs is that their samples must necessarily "age" over time and become unrepresentative of the composition of the initial population. This is clearly true in cohort study designs that select a sample of individuals with a common birth or other event date and continue to follow only those same individuals over time. Individuals who join the population by birth, immigration, or by aging into the relevant age range are not represented in this design.

But consider the more general problem of using a panel design to maintain a representative sample of the entire population of families and individuals over time. Such panels must cope with the problems of the births, immigration or entry by other means, of individuals into the population. But they must also handle the "birth" of new *families* into the population when children leave their parental homes, when a married couple splits into two families through divorce or separation, or through immigration. If a panel study is to maintain a representative sample of both the individuals and the families in the population, then there must be a mechanism that allows families and individuals to enter the sample with known selection probabilities.

*Much of this discussion parallels that given in Tanur (1981).

The Panel Study of Income Dynamics contains such a mechanism. It began in its initial year with a sample of families and individuals living in those families, each of whom had a known selection probability. The replacement mechanism for births of individuals into the population is to include individuals born into sample families as part of the sample of individuals, assigning to them the selection probability and weight of the family into which they are born. (No attempt has been made to make the PSID representative with respect to immigration since 1968, although one could make periodic additions of recent immigrants to the sample.)

The "birth" and "death" of families is a more complicated process to account for in a study design. New families formed from old family members retain the selection probabilities of the old family only as long as the new family does not contain nonsample adults in it. If it does contain nonsample adults, then the new family has a double chance of falling into the sample, and some assumptions must be made about the selection probability of such families.* An analysis of the representativeness of the PSID sample by Duncan and Hill (1984) found that the sample in 1980 was almost as closely matched to the characteristics of the Current Population Survey sample as was the sample in the study's first year, 1968, when compared with the 1968 Current Population Survey sample. With a mild and testable set of such assumptions, this type of panel design will continue to provide information each year about a representative sample of nonimmigrant families and individuals. Differential nonresponse will detract from this representativeness only if the crucial dimensions of nonresponse cannot be accounted for with an adjustment to the sample weights.†

An alternative method for adding into the sample new elements in the population is to sample dwellings and to continue to interview households in those same dwellings even if those households change between surveys. With these procedures new households moving into these dwellings represent new households that are formed in the population. This procedure, followed in the U.S. Current Population Surveys, is less expensive than one that follows the migrating households, but has the disadvantage of providing no information on change surrounding the residential moves. This disadvantage is potentially quite serious for some types of behavioral modelling, as with Hill's (1986) analysis of post-OPEC commuting responses, which found very large responses for families that had changed residences and very small ones for families that had not moved.

Thus, panel studies with annual interviews can be designed to yield representative information for their samples in every year of their existence. In this sense, there is no tradeoff between panel versus cross-sectional designs because the panel provides the cross-sectional information as well. This will necessarily be the case, however, only if prior interviews do not adversely affect the subsequent information obtained in panels and only if enough resources are devoted to keeping the panel representative. These issues are explored in the next section.

II. PANEL DESIGNS AND DATA QUALITY

A panel design may add to or detract from the quality of the data. Repeated interviewing reduces the effects of memory loss and may motivate the respondent to perform the cognitive work required for an accurate response. But, on the other hand, panels typically have higher cumulative rates of nonresponse and run the risk that an

*In fact, the PSID assumes that individuals marry or live with individuals like themselves. This doubles the selection probabilities of such families and gives them a family weight that is half as large as the individual weight of the sample individual. A more complete explanation of this is given in Duncan (1982). Dynamic change analysis is best done with individuals as the unit of analysis, even if many measures are at the family level. There is no satisfactory definition of what is the same family over time.

†Nonresponse adjustments are made periodically in the PISD. The first step in the adjustment process is to estimate a predictive model of nonresponse subsequent to the initial wave on the basis of information gathered in the initial wave. Mutually exclusive groups that differ maximally in their response rates are formed on the basis of this information and then each response case is assigned to its appropriate group. The original weight attached to each case based on its selection probability is then multiplied by the inverse of the response rate for its group.

individual's subsequent responses are "contaminated" by having been interviewed in earlier waves of the panel. These issues are discussed in this section.

Effects of memory on response error

A great deal of methodological work has shown that the reliability of survey reporting depends heavily upon time. Using hospital records as a check on validity, Cannell and his associates have found that the accuracy of reports of visits to physicians (Cannell and Fowler, 1963) and of hospitalization (Cannell *et al.*, 1965) decline dramatically with time—nearly half of hospitalization episodes go unreported when they occurred more than one year ago.* Duncan and Hill (1984) found considerably higher error variance in reports of earnings, work hours and unemployment for two years prior to the interview than in the year immediately prior to the interview. Jenkins *et al.* (1979) used the same method to ascertain memory losses in the reporting of major life events and found that, on average, between one-third and one-half of the events were forgotten or mis-dated with especially heavy losses and items related to work and financial matters.† Juster (1985) found that the quality of time diaries obtained with recall methods showed substantial deterioration with lags of more than one day if the diary were obtained for a weekday, but surprisingly little or no deterioration with lags of up to one week if the diary were obtained for a weekend day—a difference presumably due to the incidence of salient events. In their study of the validity of voting behavior, Parry and Crossley (1950) generally found more valid responses for more recent elections. Even when events are remembered, there is a tendency for respondents to "telescope" the past event by remembering them as occurring more recently than they in fact did (Sudman and Bradburn, 1974). Repeated interviewing over time will reduce both of these sources of error by reducing the amount of time between the event and the interview and by providing an opportunity for procedures such as "bounded recall" that appear to improve the quality of recalled information (Neter and Waksberg, 1964).‡

The case for panel designs is especially strong when attitudes and other subjective measures are the focus of attention because it is very unlikely that respondents can recall these types of measures accurately.§ In the words of Moss and Goldstein (1979), "there are very strong reasons for rejecting the idea that subjective experiences, like emotions, can be recalled at all" (p. 111).

Respondents in panel studies have repeated contact with the studies' interviewers and questionnaires, raising the possibility that the quality of the data might be influenced by this contact. Repeated contact has an ambiguous effect on the quality of the data. Repeated contact increases the chances that respondents understand the purposes of the study and are more motivated to do the work necessary to give more accurate answers. Research by Cannell and his associates (summarized in Cannell *et al.*, 1981) has shown that survey respondents typically have little idea of the purposes of survey studies, of

*Although evidence for the deleterious effects of memory on response error is strong, it is not unanimous. Lansing *et al.* (1961) found in two separate methodological studies that the accuracy of reports of savings account balances were not significantly different between recall periods of one month, six months, and one year. This may have resulted from the fact that the balances in many savings accounts change little over the course of a year.

†Several other methodological studies of life events (e.g. Uhlenhuth *et al.*, 1977; Casey *et al.*, 1967; and Thurlow, 1971) have found that the number of events reported falls with the amount of time that has passed since they might have occurred. None of these studies was able to validate the reports of life events with checks of official records.

‡Other procedures can improve the accuracy of time-related reports as well. In general, reliability of reporting is a function of time, of the salience of the event to the respondent, of the social desirability of the event, and of the motivation of the respondent to do the work necessary for an accurate report (Cannell *et al.*, 1981). While there is some substitutability among the items on this list (e.g. an extremely salient event like an auto accident is not as subject to memory error as a less salient event), each of the characteristics can be thought of as an independent constraint on reliability. Repeated measurement in a panel design can improve data quality substantially only if none of the other constraints are binding.

§See Smith (1982) for a brief, recent review of this literature. Smith's own analysis and that of Duncan *et al.* (1973) found that respondent reports of some attitude change can be used to obtain a fairly accurate aggregate distribution of the atitude at a prior point in time, but this may result from a large amount of offsetting individual reporting error. Few would claim that the retrospective reports of attitude change are reliable enough for analysis of attitude change at the individual level.

the cognitive tasks required of them by the interview, and are usually not motivated to undertake these tasks. Panel studies can summarize results from prior waves for respondents and thus show them in a direct way how the information they are providing is used in useful and interesting ways. Even without such motivation, there is evidence that data quality may improve in later waves of a panel. Juster (1985) finds that the quality of time-use data, measured indirectly by quality proxies such as the number or variety of activities reported on a 24-hour diary, tends to improve in later waves of a panel, probably because of learning by respondents, interviewers, or both.

Nonresponse issues

Nonresponse problems associated with the initial wave of a panel survey are not analytically distinct from nonresponse problems associated with a cross-sectional survey. Very little is usually known about these nonrespondents and very little can be done to adjust for the possible effects of the omission of their responses from the data. Panel survey designs also face potential problems with nonresponse subsequent to the initial interviewing wave. In contrast to initial wave nonresponse, however, a great deal of information has typically been gathered about these subsequent nonrespondents—information that can be used to determine the characteristics of the nonrespondents and to model nonresponse as part of the more general behavioral modelling that is of primary interest.

The Panel Study of Income Dynamics devotes considerable resources to chasing after and persuading its respondents. Although most of the interviewing takes place between March and May each year, the field operation does not stop completely until early September. (Data processing is not held up by this, since the early interviews are processed in batches shortly after they arrive.) One full-time staff person spends those months attending to the "care and feeding" of respondents, often writing persuasion letters tailored to the source of the respondent's reluctance. Furthermore, we instruct interviewers to attempt to contact respondents many more times than is the case with most studies. A result of these efforts is that the reinterview response rate in the PSID has been 97% or more in each of the years subsequent to the second one. The response rate between the first and the second year, when pursuit procedures were still being developed, was 86%.

The random nature of these losses can be shown if response rates are calculated separately by the characteristics of individuals and the heads of the families in which they lived, measured in the initial year of interviewing. Few of the characteristics account for much of the variation in actual nonresponse, and those that do may be explained by death.† For example, response rates are quite uniform across age categories up until age 65. Response rates are lower than average among the retired and among those with a nonnuclear relationship to the head of the household. There is some evidence that response rates are lower among the less educated but there are few differences across income/needs deciles. For example, the response rates within ascending deciles of family income/needs were 73, 66, 70, 71, 72, 73, 72, 73, 75, and 74%, respectively.

A major investigation of nonresponse in the Panel Study of Income Dynamics was performed by Finis Welch and his colleagues and documented in Becketti *et al.* (1983). They investigated whether nonrespondents subsequent to the initial wave differed in any crucial behavioral sense by estimating a series of models of economic behavior using information from the first wave of data, forming dichotomous variables indicating whether each case had subsequently been lost due to nonresponse and then interacting these dummy variables with key behavioral parameters of interest. They found virtually no evidence of statistically significant interactions.

The generally benign nature of nonresponse in the PSID is also illustrated with an experiment, described below, in which hypothetical nonresponse is generated using information on the difficulty of obtaining the actual interviews. Less extensive pursuit

†Over a long period of time, additional uncounted nonresponse could result from children born to adults who were lost from the panel.

of PSID respondents does not appear to produce a bias in the remaining sample that cannot be corrected with adjustment of the sampling weights. Of course, these specific results on nonresponse may not generalize to other survey nor to behavioral models not included in the past PSID methodological work. But the robustness of the findings, coupled with a realization of the potential analytic power of the information already collected prior to the nonresponse in adjusting for its detrimental effects does argue that concern about low panel response rates as such has been overstated.

Effects of panel participation

Care must be taken in the feedback process to avoid possible contamination of subsequent responses. Tanur (1981) reports on evidence that intitial-wave responses in a panel study may differ from responses in subsequent waves and that in some panel surveys the initial interview is not used for comparative purposes. Panel projects may also have a cumulative effect on responses if membership in a panel has an effect on attitudes or behaviors. Bailor (1975) shows how responses from the initial waves of CPS rotation groups differ from subsequent waves but cannot tell from her data which responses are more accurate.

A crucial question is whether it is merely the reporting of behavior or the behavior itself that is affected by panel membership. A definitive answer to this question usually requires independent validation. One such validation study was conducted by Traugott and Katosh (1979). They found that while extended participation in an election survey improved the accuracy of data on voting behavior, it also appeared to increase the amount of voting itself—thus making the behavior of the panel unrepresentative of the behavior of the population. Differential nonresponse in the panel may explain some of this result, since less politically motivated individuals tended to drop out of the panel.

Although it obviously can cause specific problems, the possibility that panel participation has pervasive effects on behavior or attitudes seems remote. The illustration above, where voting behavior appears to have been affected by panel participation, falls into a class of cases where the behavior being studied (voting) is one where a change requires very little effort by the respondent, and where the heightened perception of political issues occasioned by participation in a survey may plausibly lead respondents to decide to vote rather than to stay home. It seems much less likely that participation in a panel study concerned with economic behavior, where the focus is on work effort, saving, energy usage, asset accumulation, durable goods purchase, commuting behavior, etc., will be a sufficiently salient experience so that answering questions about employment activities will cause employment behavior to change. In fact it is quite likely that repeated contacts improve the quality of the data as respondents get trained to their task, and cross-year data consistency checks are applied.

Analytic costs of "cheaper" panel studies

Optimal research designs are dependent upon the economic costs and benefits of the alternatives. These issues become very apparent in sample surveys when increments to the efficiency of parameter estimates, to the response rates, to the amount of data "cleaning," to the quality of the documentation, and to the quality of the analysis itself come about only when additional resources are devoted to them. The nature of the "production function" associated with the generation of scientific knowledge is not well known,† but it is useful to investigate how the quality of the data might be affected by various ways of reducing survey costs.

Telephone vs. personal interview. An obvious way to reducing survey costs is to interview by telephone rather than in person. This is particularly appealing for subsequent waves in panel studies, where the first wave, done in person, establishes rapport with respondents, does not eliminate those without telephones from the sample, and provides the opportunity for in-person observation by the interviewer. The case for personal interviews becomes less compelling over time since rapport has already been established

†The general role of measurement in the knowledge production function is discussed in Juster (1970, 1972).

and geographic mobility will increase the costs of personal interviews substantially.* Perhaps the ideal design for the subsequent waves is to use both interviewing modes—telephone interviewing whenever possible and personal interviews for those without access to telephones or with a reluctance to share their responses with those on the same party line.

What are the implications of switching to telephone interviewing for the quality of data? It is obvious that personal, on-site observation by the interviewer is impossible, as are data collection methods that require visual aids. However, some of these methods can be converted to a form that can be administered by telephone. Groves and Kahn (1979) conducted a careful comparison of telephone and personal interviewing and found surprisingly few differences between the two modes. Telephone respondents did tend to avoid placing themselves on the extremes of some attitudinal scales, did give somewhat more rounded financial figures, and did give shorter responses to open-ended questions. But, on the whole, the patterns of univariate and bivariate relationships observed for most of the 200 questions were not significantly different for the two modes. Analysis of the quality of the information from personal vs. telephone interviews in the Panel Study of Income Dynamics also shows few differences (Benus, 1975), and the quality of time-use data collected by personal interview or by telephone appears to be comparable (Juster, 1985). When combined with the fact that telephone interviewing costs are roughly half those of personal interviews, telephone interviewing appears to be a promising method for adapting to scarcer research resources.†

Less extensive pursuit of panel respondents. A prominent cost associated with longitudinal studies are those incurred in the final stages of the field work, when inaccessible or reluctant respondents are pursued or persuaded. The costs of these final interviews are relatively high, and it is tempting to cut costs by accepting lower response rates in the reinterviews. The potential costs of losing these respondents is especially high in panel studies, however, since respondents lost in early waves are typically not sought after in subsequent waves. Analytic costs will be incurred if these potential respondents differ systematically from the more accessible respondents, especially if the differences cannot be handled with adjustments to the weights for the differential nonresponse.

To see if the respondents obtained through persistent and relatively expensive efforts differed systematically from more accessible ones, we conducted an experiment that involved creating a "new" and smaller PSID sample characterized by lower costs and lower response rates. Specifically, we eliminated those families that would have been lost either if interviewing had stopped on July 1 of each year (as opposed to September 1), or if we had stopped trying to contact households after four personal attempts or seven telephone attempts. This experiment was repeated each year between 1973 and 1980, and thus generated a new and smaller sample.

Only 62% of the individuals living in families that were part of the actual 1980 sample would have been retained if these less intensive pursuit rules had been applied. (The weighted response would have been 68%.) In contrast to the largely random patterns of actual nonresponse, these additional losses can be distinguished by their initial characteristics, particularly race, income, and age.

Whether the differential nonresponse from less intensive field procedures distorts the sample to the point that parameter estimates from it become biased depends upon the extent to which adjustments for differential nonresponse can be incorporated into the weight variable. To determine this, we searched through a set of demographic variables, measured in 1972, to find the combination that maximally differentiated groups

*White and Huang (1982) conduct a careful study of the costs of following movers in the Income Survey Development Program. All interviews were done in person, and families that moved more than 50 miles from a PSU in which study interviews were located were not pursued. They found that following the rest of the movers over the 6 quarterly waves of the study added 7% to the total number of interviewer hours and 11.4% to the number of miles charged.

†An added advantage of telephone interviewing from a central location is the ease with which methodological experiments can be built into the design and thus be used to identify sources of response error. Interviewers can be randomly assigned so that interviewer contributions to variance can be identified separately from geographic cluster components.

according to whether they would have been lost with less intensive pursuit. We then adjusted the PSID weights for this differential nonresponse and compared the characteristics of the smaller, newly weighted sample of accessible respondents with the larger sample of all PSID respondents. Our focus in this comparison was on dynamic characteristics of the respondents in the period 1972 to 1980 (e.g. income fluctuation, poverty dynamics, unemployment experience) that may well have differed between accessible and inaccessible respondents. We are thus asking how those who would have been lost differed after they would have disappeared.

Surprisingly, we found only minor differences. Even without adjusting the weights for the additional (hypothetical) nonresponse, the two samples agreed on most univariate distributions, and with new weights they were close even on race. Although our list of search variables was fairly extensive, it is of course possible that the two samples differ along dimensions that we did not investigate or that interactions would be biased. Our tentative conclusion, however, is that while there are clear, understandable differences between inaccessible respondents and accessible ones, adjustments to the weights for differential nonresponse should assure the representativeness of the smaller sample for most purposes.

The simple implications of this are that the extra cost of chasing people for maximum response rate has little apparent effect on distributions estimated from the sample, even on situations after some individuals would have been lost, and that if anything the quality of the data might have been improved less than the increased sample size would imply. It remains possible, however, that multivariate estimates, particularly with models not perfectly specified, might suffer some biases, although we did estimate some simple earnings models with the full and artificially truncated samples and found no significant difference between the weighted estimates of these models. In any case, since we did not have this information at the time, the high response rates were probably crucial if the Panel Study was to have received continued support, since the growing unrepresentativeness of panels is often thought to be a major potential deficiency of such studies. More important than all these considerations is the actual sample size for some crucial subgroups, such as black splitoffs, which would suffer disproportionately from relaxation of efforts to maximize response rates.

III. RELATIVE MONETARY COSTS OF LONGITUDINAL STUDIES

To answer questions about the relative costs of longitudinal studies, it is important to state clearly the alternatives to them. One obvious comparison is between a single cross-sectional survey with memory questions and a two-wave panel. But if the longitudinal study is designed to keep its basic sample representative each year and if enough resources are devoted to that task, then the longitudinal study can serve as an annual source of representative cross-sectional information and ought to be compared with a series of cross-sectional studies rather than just a single one. Both comparisons are made in this section.

To estimate field costs of the single cross section versus two-wave panel, it was assumed that the cross-sectional interview was done in person, averaged 90 min in length (to include some retrospective questions on the changes that would be answered directly in the panel), achieved a response rate of 75%, and resulted in 2,000 completed interviews. The field costs of such a survey at the Survey Research Center were estimated to be roughly $250,000.† A "comparable" panel study was assumed to have two waves, separated by one year. The initial wave would be done in person and average 60 min in length. The second wave was assumed to be done in 30 min by telephone. The combined response rate was targeted at about 70%, obtained in the cheapest possible manner. It turned out that the optimal strategy was to obtain an 85% response rate in the first wave and an 80% response rate in the second wave.‡ The first-wave field costs

†All of the estimates of field costs given in this section should be considered as very rough, but unbiased.

‡The relatively low response rate in the second wave of the panel results from the loss of respondents without telephones.

of such a design also total $250,000 (the savings due to the shorter interview are almost exactly offset by the increased costs of attaining the higher response rate). The monetary costs of the second wave were estimated to be $50,000. The combined cost of the panel was $300,000, a figure that is 20% higher than the single cross section.† If the waves of a two-wave panel are separated by five rather than one year, the costs would perhaps increase by an additional $10,000, and the response rate in the second wave would fall to perhaps 70% from about 80%. Thus, the total monetary cost of the five-year, two-wave panel is $310,000, which is roughly 25% higher than the cross-sectional design.

If a panel design is used to provide representative cross-sectional information, then a proper cost comparison is between additional waves of the panel and fresh cross-sectional surveys of comparable size. A useful comparison here is between the actual field costs of additional waves of the Panel Study of Income Dynamics and a comparable cross-sectional survey. To price the latter, it is assumed that the same number of interviews are taken (6,700 in this case), that the response rate is 75%, that 4,700 of the interviews are done by telephone through random-digit dialing, and 2,000 personal interviews are taken from an area probability sample.‡ The field costs of such a study would be roughly $325,000 if the length of the interview is comparable to the PSID's and $425,000 if it is expanded by 30 minutes to include retrospective questions. This compares to the actual field costs of the PSID of $250,000. Thus, the field costs of each successive wave of the cross section are 30 to 70% higher than for additional waves of the panel, depending on the length of the interview.

Most of the added costs are due to the additional time needed by interviewers to contact and persuade respondents. The total amount of time taken by interviewers for each completed interview is roughly one hour more in a fresh cross-section than in additional waves of a panel.

Other costs vary with the survey design as well. Sampling costs of $15,000 are incurred with each fresh cross-section, but only once with a panel. On the other hand, there are additional data processing costs associated with panels, particularly those that attempt to keep family composition changes recorded accurately each wave and possible respondent payments in the panel. The processing costs might amount to one to two person-years of time plus some computer expenses, perhaps $30,000 to $50,000 in all. Panels also provide an opportunity for a great deal of checking for inconsistencies, analysis of nonresponse subsequent to the first wave and so forth. The costs of these activities are part of the total cost of panel designs, although it is important to realize that each of them provides information that cannot be obtained in cross-sectional surveys.

SUMMARY

In some areas of research, the case for longitudinal studies is exceedingly strong. Not only can panel data offer the analyst more reliable measurement and much greater leverage in estimating truly dynamic models of behavior, but the costs of panel waves are less than the costs of comparable cross sections. These benefits are most apparent when the focus of the research is a subjective phenomena, is objective but subject to reporting error through memory loss or inadequate motivation of respondents, or is measured at the family level rather than the individual level. The ability of panels to maintain a representative sample of families and individuals each year is often misunderstood but very important, since it means that panels can substitute for the cross-

†The possible analytic costs associated with the loss of an additional 5% of the respondents must be added to these monetary costs. In one important sense, second wave nonrespondents are "better" than initial wave nonrespondents since so much more is know about them.

‡This mixture of 4,700 telephone and 2,000 personal interviews appears to maximize the amount of information per dollar spent, taking into account the noncoverage problem of telephone samples and the design effects associated with area probability samples. The telephone interviewing is assumed to be conducted with a Computer Assisted Telephone Interviewing (CATI) system, thus saving a substantial amount of coding costs. In the Panel Study of Income Dynamics, approximately 400 interviews are conducted in person each year in cases where there is no telephone or when the respondent insists on a personal interview. Telephone interviewing is done by the field interviewers, not from the Ann Arbor facility and not with the CATI system.

sectional surveys. Thus for some research areas, panel studies have all of the analytic benefits of successive cross-section, have the additional benefits of better measurement, the capability of being used for estimating truly dynamic behavioral models, and do not cost anymore than the successive cross sections.

This exceedingly optimistic assessment of panel studies does not apply in all research areas, however, nor for situations when they must be conducted on a shoestring budget or by an inept survey organization. It is possible that the attitudes and behavior of respondents become contaminated over time through their participation in the study particularly if the interval between interviews is short. It is unlikely that all types of behavior or attitudes are affected by this, but the evidence suggests that it can happen, and concern for it ought to be part of an assessment of whether a panel design is an appropriate one.

More importantly if not enough resources are devoted to keeping the panel representative or if for some other reason the panel is not run well, then not only will the panel fail to provide cross-sectional information each year but the longitudinal results will be suspect as well. We find that reluctant respondents are different from less reluctant ones. They are younger, lower income, and are more likely to be black, and considerable resources must be devoted to keeping them in a panel sample. But we also find that adjustment to weights for this differential nonresponse restores most of the descriptive and analytic representativeness of the smaller sample.

Much of the case for panel studies rests on their unique ability to provide data for the estimation of dynamic models of behavior. And yet most social science research is still confined to the analysis of cross-sectional data. But the development of social science models goes hand-in hand with the availability of data to test them. If nothing else, the longitudinal studies of the 1970's have prompted a flood of work on methodological issues that surround the specification and estimation of truly dynamic models of behavior. The path-breaking articles are just beginning to filter down through the social sciences. What advances will come from the next generation of social scientists, schooled in the new methods, cannot be stated, but there are strong grounds for optimism.

Acknowledgement—The authors are Program Directors at the Institute for Social Research, University of Michigan. Juster and Morgan are also Professors in the Economics Department at the University of Michigan. Charles Cannell, Robert Groves, Steven Herringa, Graham Kalton, Ronald Kessler, Leslie Kish, Stanley Pressor, and Howard Schuman all gave valuable comments on earlier drafts of this paper, portions of which were presented at a Social Science Research Council conference entitled "Designing Research with Scarce Resources," November, 1982.

REFERENCES

Baanders B. and Slootman K. (1983) A panel for longitudinal research into travel behavior. In *Recent Advances in Travel Demand Analysis* (Edited by Carpenter S. and Jones P.) Hants, England, Gower.

Bailor B. (1975) The effect of rotation group bias on estimates from panel surveys. *J. Am. Stat. Assoc.* **70,** 23–31.

Becketti S., Gould W., Lillard L. and Welch F. (1983) Attrition from the PSID. Unicon Research Corp., Santa Monica, CA.

Benus J. (1975) Response rates and data quality. In *Five Thousand American Families—Patterns of Economic Progress* (Edited by Duncan G. D. and Morgan J. N.). Vol. III. Institute for Social Research, Ann Arbor.

Cannell C. F., Fisher G. and Bakker T. (1965) Reporting of hospitalization in the health interview study. *Vital and Health Statistics* **2(6),** 1–71.

Cannell C. F. and Fowler F. J. (1963) *A study of the reporting of visits to doctors in the national health survey.* Survey Research Center. (mimeo), Ann Arbor.

Cannell C. F., Miller P. V. and Oksenberg L. (1981) Research on interviewing techniques. In *Sociological Methodology.* Jossey-Bass, San Francisco.

Casey R. L., Masuda, M. and Holmes T. H. (1967) Quantitative study of recall of life events. *J. Psychosomatic Res.* **11,** 239–247.

Coe R. D. (1982) A preliminary empirical examination of the dynamics of welfare use. In *Five Thousand American Families—Patterns of Economic Progress,* Vol. IX, (Edited by Hill D. H., Hill M. S. and Morgan J. N.) Institute for Social Research, Ann Arbor.

Corcoran M.E., Duncan G. J. and Ponza M. (1983) A longitudinal analysis of white women's wages. *J. Hum. Resources* **18,** 497–520.

Duncan G. J. (1982) The implications of changing family composition for the dynamic analysis of family economic well-being. Prepared for Conference on Analysis of Panel Data on Income, London.

Duncan G. J. and Hill D. H. (1985) An investigation of the extent and consequences of measurement error in labor economic survey data. *J. Labor Econ.* **3(4),** 508–532.

Duncan G. J. and Hill D. H. (1984) How representative is the PSID? Survey Research Center, Ann Arbor (mimeo).
Duncan G. J. and Holmlund B. (1983) Was Adam Smith right, after all? another test of the theory of compensating wage differentials. *J. Labor Econ.* **11(4),** 366–379.
Duncan G. J. and Kalton G. (1985) Issue of design and analysis of surveys across time. Prepared for the Centenary Session of the International Statistical Institute. Amsterdam.
Duncan G. J. and Mathiowetz N. (1984) A validation study of economic survey data. Survey Research Center, University of Michigan (mimeo).
Duncan O. D., Schuman H. and Duncan B. (1973) *Social Change in a Metropolitan Community.* New York, Russell Sage.
Flinn C. and Heckman J. (1982) New methods for analyzing labor force dynamics. *J. Econometrics* **18,** 115–168.
Godard X. (1983) Presentation of some results of the Dijon survey: linkage between activities and mobility. In *Recent Advances in Travel Demand Analysis* (Edited by Carpenter S. and Jones P.) Hants, England, Gower.
Groves R. M. and Kahn R. L. (1979) *Surveys by Telephone: A National Comparison with Personal Interviews.* New York, Academic Press.
Hannan M. T., Tuma N. and Groeneveld L. P. (1977) Income and marital events: evidence from an income maintenance experiment. *Am. J. Sociology* **82,** 1186–1211.
Hausman J. A. and Taylor W. R. (1980) Panel data and unobservable individual effects. *Econometrica* **49,** 1377–1398.
Heckman J. J. and Singer B. (1982) The identification problem in econometric models for duration data. Discussion paper #82–6. Economic Research Center, National Opinion Research Center, Chicago (mimeo).
Hill D. H. (1986) Dynamics of household driving demand. *Review of Economics and Statistics.* **68(1),** 132–141.
Hill M. S. (1982) Some dynamic aspects of poverty. In *Five Thousand American Families—Patterns of Economic Progress,* Vol. IX, (Edited by Hill D. H., Hill M. S. and Morgan J. N.), pp. 93–120. Institute for Social Research, Ann Arbor.
Houthakker H. S., Verliger P. K. and Sheehan D. P. (1974) Dynamic demand analysis for gasoline and residential electricity. *Am. J. Agri. Econ.* **56,** 412–418.
Jenkins C. D., Hurst M. W., and Rose R. M. (1979) Life changes: do people really remember? *Arch. Gen. Psychiatry* **36,** 379–384.
Johnson L. and Hensher D. (1982) Application of a multinominal probit to a two-period panel data set. *Transp. Research* **16A(5–6),** 457–464.
Juster F. T. (1982) *Government support of economic research: an evaluation.* Paper presented to the American Economic Association. Institute for Social Research, The University of Michigan. Ann Arbor.
Juster F. T. (1972) Microdata requirements and public policy decisions. *Annals Econ. Social Measurement* **1(2),** 7–16.
Juster F. T. (1970) Microdata, economic research, and the production of economic knowledge *Am. Econ. Rev.* **40(2),** 138–148.
Juster F. T. (1985) The quality and validity of time use estimates obtained from recall diaries. In *Time, Goods, and Well-Being* (Edited by Juster F. T. and Stafford F.). Institute for Social Research. Ann Arbor.
Lansing, J. B., Ginburg G. P., and Bratten, K. (1961) *An Investigation of Response Error.* Studies in Consumer Saving, No. 2. University of Illinois, Urbana.
Mackett R. L. (1983) The inter-relationships betwen travel and locational decisions. In *Recent Advances in Travel Demand Analysis* (Edited by Carpenter S. and Jones P.) Hants, England, Gower.
Mathiowetz N. and Duncan G. J. (forthcoming) Out of work, out of mind: response error in retrospective reports of unemployment Survey Research Center, University of Michigan *J. Bus. and Econ Statis.*
Moss L. and Goldstein H. (1979) *The Recall Method in Social Surveys.* University of London Institute of Education, London.
Neter J. and Waksberg J. (1964) A study of response errors in expenditure data from household interviews. *J. Am. Stat. Assoc.* **59,** 18–55.
Parry H. J. and Crossley H. M. (1950) Validity of responses to survey questions. *Publ. Opinion Q.* **14(1),** 61–80.
Pullman T. (1978) Postscript to social mobility. In *International Encyclopedia of Statistics* (Edited by Kruskal W. H. and Tanur J. M.). Free Press, New York.
Robinson J. P. (1985) The validity and reliability of alternative time use measures. In *Time, Goods, and Well-being* (Edited by Juster F. T. and Stafford F. P.). Institute for Social Research, Ann Arbor.
Schoenberg R. (1977) Dynamic models and cross-sectional data: the consequences of dynamic misspecification, *Social Science Research* **7,** 133–144.
Smith T. W. (1982) Recalling attitudes: an analysis of retrospective questions on the 1982 general social survey. GSS Technical Report No. 35, National Opinion Research Center (mimeo).
Sudman S. and Bradburn N. N. (1974) *Response Effects in Surveys.* Aldine, Chicago.
Tanur J. M. (1981) Advances in methods for large-scale surveys and experiments. Paper commissioned by the Social Science Research Council for the Five-Year Outlook on Science and Technology.
Thurlow H. J. (1971) Illness in relation to life situations and sick-role tendency. *J. Psychsomatic Res.* **15,** 73–88.
Traugott M. W., and Katosh J. P. (1979) Response validity in surveys of voting behavior. *Publ. Opinion Q.* **42,** 359–377.
Uhlenhuth E. H., Haberman S. J., Balter M. D. and Lipman R. S. (1977) Remembering life events. In *The Origins and Course of Psychopathology* (Edited by Strauss J. S., Babigan H. M. and Roff M.). Plenum Press, New York,
White G. D. Jr. and Huang H. (1982) Mover follow-up costs for the income survey development program. *Proceedings,* Section on Survey Research Methods, American Statistical Association, 376–81.