CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Research Papers in Economics

IFPRI

Discussion Paper BRIEFS

Food Consumption and Nutrition division of the International Food Policy Research Institute

Discussion Paper 96

Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing Country Samples

Harold Alderman, Jere R. Behrman, Hans-Peter Kohler, John A. Maluccio, Susan Cotts Watkins

Use of longitudinal, or panel, household data can have considerable advantages over more widely used cross-sectional data for social science analysis. Longitudinal data permit (1) tracing the dynamics of behaviors, (2) identifying the influence of past behaviors on current behaviors, and (3) controlling for unobserved fixed characteristics in the investigation of the effect of time-varying exogenous variables on endogenous behaviors. The collection of longitudinal data, however, may be difficult and expensive. Some question whether the gains from collecting such data in developing countries are likely to be worth the costs.

One issue that has concerned analysts is sample attrition, i.e., not reinterviewing households and individuals in all rounds of a longitudinal survey. Many analysts share the intuition that attrition is likely to be selective on characteristics such as schooling and that high attrition is likely to bias estimates based on longitudinal data. Such attrition may be particularly severe in areas of the developing world where there is considerable mobility due to migration between rural and urban areas.

Methodology

This paper considers the extent and implications of attrition for three different developing country longitudinal household surveys including (1) a Bolivian household survey designed to evaluate an early childhood development intervention in poor urban areas, with survey rounds in 1995/1996 and 1998; (2) a Kenyan rural household survey designed to investigate the nature of social networks in the dissemination of contraceptive use and behaviors related to HIV/AIDS, with survey rounds in 1994/1995 and 1996/1997; and (3) a South African rural and urban household survey designed for more general purposes, with survey rounds in 1993 and 1998.

Although attrition is a pervasive problem in longitudinal surveys, its magnitude varies substantially from one survey to the next. For example, the attrition rates for the three samples considered here are considerable—35 percent for the Bolivian sample, 28 percent for women and 41 percent for couples in the Kenyan sample, and 16 percent for households and 22 percent for preschool children in the South African sample. These translate to attrition rates ranging from 3.2 to 20.5 percent per year between survey rounds.

Using the methodology outlined in Fitzgerald, Gottschalk, and Moffitt (1998) [*Journal of Human Resources* 33 (2): 251-299], this paper conducts a series of tests on these three datasets to evaluate the extent to which attrition biases estimates of some "standard" behavioral relations:

1. Comparison of Means for Major Outcome and Control Variables. The first test directly compares the means for major outcome and control variables measured in the first rounds of the respective data sets for (eventual) attritors versus nonattritors.
2. Probits for Probability of Attrition. The second test estimates the probability of attrition using combinations of the outcome and control variables as independent variables to examine whether they are significantly associated with attrition in a multivariate framework (as opposed to the univariate comparisons from the first test).
3. Estimates of Behavioral Relationships. Do attritors have different coefficient estimates than nonattritors in behavioral equations? The final set of tests estimates some "standard" behavioral relationships for outcome variables in the initial wave of the survey and tests whether the coefficients of the predetermined variables and the constant differ for those observations that eventually attrite versus those that do not. The aim is to determine whether those who subsequently leave the sample differ in their initial behavioral relationships; if they do, then special care must be taken when interpreting estimates based only on the nonattriting sample.

Experience from three longitudinal surveys shows that sample dropout, or attrition, is often <u>not</u> a source of bias in multivariate analysis.

Results

1. The means for a number of critical child development outcome and family background variables do

differ significantly between attritors and nonattritors. For the Bolivian PIDI data, there is a tendency for attritors to have worse child development outcomes and family background than do nonattritors. In the poor urban communities in which PIDI concentrates, it appears that it is the worse-off households that are most mobile and most difficult to follow over time. This contrasts with the Kenyan rural data and the South African rural and urban data for which it is households and individuals with better backgrounds that are most mobile and thus more likely to attrite.

2. A number of the Bolivia family background variables—but not child outcome variables—are significant predictors of attrition. The result for the child outcome variables is similar to that for the outcome variables in the Kenyan case. But the significance of a number of background variables in predicting attrition in the Bolivian data again contrasts with the limited significance of such background variables in predicting attrition in the Kenyan and South African data. For South Africa, the overall probit relation does not significantly predict attrition, even though some individual variables appear to predict greater attrition of children.

3. The coefficients estimates for "standard" family background variables in regressions and probit equations for the majority of the Bolivian child development outcome variables are not affected significantly by attrition.

The coefficients on "standard" variables in equations with the major outcome and family planning social network variables in the Kenyan data also are unaffected by attrition. For six of the seven child anthropometric measures in the South African data, moreover, there are no significant effects of attrition on the coefficient estimates of the "standard" variables. Therefore, attrition apparently is not a problem for obtaining consistent estimates of the coefficients of interest for most of the child development outcomes. In contrast to concerns often expressed about attrition, and despite suggestions of systematic attrition from univariate comparisons, multivariate estimates of many of the behavioral relations of interest do not appear to be biased due to attrition.

For the Bolivian child development outcomes related to child weight and for South African child moderate stunting and morbidity, however, the results differ strikingly and suggest that attrition bias is likely to be a problem in multivariate estimates of related behavioral relations. Attrition selection bias appears to be model specific: changing outcome variables may change the diagnosis even within the same data set. Thus, as a general observation, analysts should assess the problem for the particular model and the particular data they are using.

Keywords: Bolivia, Kenya, South Africa, household survey, attrition

Recent FCND Discussion Papers

Attrition in Kwazulu Natal Income Dynamics Study 1993-1998, John Maluccio, October 2000, ***DP95***

Targeting Urban Malnutrition: A Multicity Analysis of the Spatial Distribution of Childhood Nutritional Status, Saul Sutkover Morris, September 2000, ***DP94***

Mother-Father Resource Control, Marriage Payments, and Girl-Boy Health in Rural Bangladesh, Kelly K. Hallman, August 2000, ***DP93***

The full text of this document and other FCND Discussion Papers are available on our Website (*www.ifpri.org/divs/fcnd/dp.htm*) or via B.McClafferty@cgiar.org

FCND BRIEFS

International Food Policy Research Institute

2033 K Street, N.W.
Washington, D.C. 20006 U.S.A.

***Experience from three longitudinal surveys shows that sample dropout, or attrition, is often not a source of bias in multivariate analysis.*--DP96**