IFPRI

Discussion Paper BRIEFS

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

Discussion Paper 95

ATTRITION IN THE KWAZULU NATAL INCOME DYNAMICS STUDY 1993-1998

John Maluccio

This paper focuses on sample attrition, and the possible ensuing selectivity, with special reference to a recently collected South African panel survey of African and Indian households, the KwaZulu-Natal Income Dynamics Study (KIDS).

Using Panel Data

The analysis of panel (or longitudinal data, where the same individuals or households are interviewed multiple times, contributes substantially to the understanding of a variety of phenomena. For example, while two cross-sectional surveys of different households at two points in time might reveal a constant poverty rate, they are silent as to whether this reflects chronic poverty, i.e., the same households in poverty in each period, or transitory poverty with an equal proportion of households exiting and entering poverty between surveys. If appropriate policy action depends on the chronic or transitory nature of poverty, it is critical to be able to distinguish between the two, something that panel data allow. Thus, panel data often permit an understanding of the dynamic behavior of individual households not possible with cross-sectional or time-series information alone.

A second advantage of panel data is that they enable us to resolve, or at least reduce concern about, a key econometric problem: omitted variable (or unobserved heterogeneity) bias. For example, rarely do surveys observe or measure a family's preferences and priorities for educating its children. It is quite likely that families that put a high priority on education will perform additional work to obtain income needed to pay school fees. If we use cross-sectional data alone to determine the effect of family income on education, we risk making incorrect inferences, i.e., families with the highest income may also be those that prioritize education the most. In other words, omitted preferences for education are correlated with included income measures. Estimates derived from such data will tend to overstate the impact that an income transfer would have on educational decisions of families that give only an average priority to education. In contrast, with panel data, econometric methods can be used to control for these sorts of time-invariant preferences and family characteristics, allowing unbiased estimates of the effect of income on education.

Panel data are not a panacea, however. In practice, one must balance the potentially substantial benefits against the many real difficulties encountered in survey work that lead to, in particular, errors of measurement and sample attrition. (One should also keep in mind that unless refresher samples are added in later rounds, the current period representativeness of the panel sample deteriorates over time, and this may occur more quickly in rapidly changing societies. Thus many analyses appropriate for a representative cross-sectional survey are not appropriate for individual rounds of a panel survey.) Either of these can introduce different sources of bias, inhibiting anew the capacity to make correct inferences from the data.

The Study

This analysis examines attrition in KIDS to (1) document the procedures and outcomes of the survey as a resource both for those using this publicly available data and for those embarking on their own survey work, (2) describe the characteristics of households that attrited in the second round sample and explore their correlates in a multivariate framework, and (3) propose a simple methodology to assess and correct for attrition bias, using information reflecting the quality of the fieldwork in the first round as identifying instruments. Because it is a comprehensive

"Since it is typically difficult to determine the bias for a particular analysis a priori, it behooves researchers using panel data not to avoid using panel data when there is attrition, but to always evaluate the effect of such bias on the analysis at hand."

survey and can be used for a variety of analyses, however, it is not possible to make global statements about attrition bias. Rather, the results presented here should be treated as methods to be replicated by other analysts using the data.

Results

The evidence presented here indicates the following. A large percentage (84 percent) of the original sample was successfully reinterviewed after nearly five years, and the ability to follow those who had moved contributed a substantial portion of the overall success rate. However, attrition in the KIDS survey is nonrandom and varies with, among other things, household size and household- and community-level resources. Furthermore, attrition is closely linked to migration. The characteristics of the households that were not reinterviewed but left no trace differ from the other movers, which suggests that the processes underlying their attrition may have been different. Indicators of quality of the interview in 1993 were identified that significantly influence the likelihood of being in the no trace group and might be used to correct for sample selection based on unobservables.

While observable differences between attritors and non-attritors (as well as within the former group) indicate that attrition is nonrandom, this does not necessarily imply that estimated relationships based on the non-attriting sample suffer from attrition bias. To more directly explore attrition bias, which is by its nature model-specific, I estimate household-level expenditure functions correcting for attrition bias using standard Heckman selection procedures and quality of 1993 interview variables as identifying instruments. The results suggest that, at least for this simple case, attrition does appear to be biasing the "behavioral" coefficients.

In a related paper focused on attrition on unobservables, Alderman et al. (2000) [Food Consumption and Nutrition Division Discussion Paper 96] use some of the above techniques to explore attrition bias for three developing country data sets, including the KIDS data examined in this brief. They also document that a variety of family background characteristics are significant predictors of attrition, indicating it is indeed nonrandom. Nevertheless, for a majority of the outcome variables considered across the different countries, coefficient estimates for the influence of those same family background characteristics are not significantly affected by attrition. In particular, for the KIDS sample, estimates of a variety of child anthropometric outcomes indicate attrition bias in only a few of them.

These examples demonstrate that attrition bias for models estimated on panel data is indeed model specific. Large levels of attrition do not always lead to attrition bias; however, sometimes they do. Since it is typically difficult to determine the bias for a particular analysis a priori, it behooves researchers using panel data not to avoid using panel data when there is attrition, but to always evaluate the effect of such bias on the analysis at hand.

Keywords: South Africa, panel data, longitudinal data, attrition, attrition bias

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.**

"Since it is typically difficult to determine the bias for a particular analysis a priori, it behooves researchers using panel data not to avoid panel data when there is attrition, but to always evaluate the effect of such bias on the analysis at hand." DP95