

The Papua New Guinea Household Survey

John Gibson[†]
Department of Economics
University of Waikato

Acknowledgements

Appreciation is expressed for the financial support from the governments of Australia (TF-032753), Japan (TF-029460), and New Zealand (TF-033936) and for the helpful advice of Malcolm Levett, Scott Rozelle and Christopher Scott. All views in this paper are those of the author and should not be attributed to the World Bank.

[†] Department of Economics, University of Waikato, Private Bag 3105, Hamilton, New Zealand.
Fax: (64-7) 838-4331. E-mail: jkgibson@waikato.ac.nz.

1. Introduction

The Papua New Guinea Household Survey (PNGHS) is a multi-topic survey of living standards in Papua New Guinea. The survey was conducted for the World Bank and is modelled on the Living Standards Measurement Study (LSMS) surveys (Gross and Glewwe, 1996). This cross-sectional survey was the first national study of consumption, living standards and poverty in Papua New Guinea and these data are likely to be of interest to development and other applied microeconomists. The similarity with the LSMS surveys may aid cross-country studies based on unit record data (Lancaster, *et. al.*, 1998), especially because there are few other publicly available living standards surveys for the Asia-Pacific region. The data may hold particular interest for Australian researchers because Papua New Guinea is the largest recipient of Australian overseas development aid.

2. Content of the Survey

In addition to estimating the value of household consumption, the survey also collected data on demographics, education levels and costs, income sources, use of health facilities, diet, housing conditions, agricultural production, durable assets, cash transfers, and participants' perception of their quality of life. Anthropometric measurements (height and weight) on young children and their parents were recorded. A community questionnaire was used to collect data on access to education, health, transport and communication services and also included a consumer price survey. Details of the information collected under each of these themes is reported in Table 1.

(Table 1 about here)

The scope of the survey allows a wider range of topics to be investigated than the standard demand, inequality and poverty analyses that are typically carried out with household

consumption data. The survey is well-suited to multivariate analysis, with several potential instruments available to deal with the endogeneity problems that typically arise when studying the human capital investments of households. Amongst the potential topics that can be studied with these data are the determinants of educational attainment, malnutrition and poverty; gender bias and the intra-household allocation of resources; and the effect of informal mechanisms for sharing risk and redistributing income.

3. Sample Design and Data Collection Methods

The survey used a stratified, two-stage random sample of 1200 households, residing in 120 Census Units that were selected from the 1990 Census sample frame. The stratifying variables were sector (urban and rural), environmental conditions (three elevation and two rainfall classes), and level of agricultural development (three classes). All provinces were covered except for Bougainville, where civil war prevented field operations. Sampling weights were generated from the variation between the Census estimates of the size of each cluster and the actual size found in 1996, and from the deviation of the actual number of households surveyed in each cluster from the target number. A total of 1144 households have complete information and there is no evidence of sample selection bias (Gibson and Rozelle, 1998). Although the sample is small, the relatively large number of clusters and detailed stratification should aid statistical efficiency (Howes and Lanjouw, 1998).

Household interviews for the survey took place between January 1996 and March 1997. The interviewers were either graduates or senior students of the University of Papua New Guinea, while the supervisors were all senior academics. Interviewers worked in mixed-gender teams of four, spending approximately two weeks in each selected Census Unit. Random checks were carried out at approximately one-quarter of the interview locations by the current author

and one other senior supervisor. An additional quality control check was to rotate interviewer teams and then send them back to 20 of the Census Units to re-identify the respondents.

A closed interval recall method was used, with each household interviewed twice so that the start of the consumption recall period was signalled by the first interview. Consumption data were collected on all food (36 categories) and other frequent expenses (20 categories) during the recall period (which averaged two weeks) and then extrapolated to annual totals. In an economy such as Papua New Guinea, a large share of consumption is from self-produced items and gifts so the survey paid particular attention to these. Food quantities were based on conversions from volume measures (households were given empty sacks with marked graduations for recording garden produce). Imputed values of self-produced and gifted items were based on respondent reports, but replacing these with either cluster medians of the unit values or cluster averages of market prices has no effect on mean consumption (Gibson and Rozelle, 1998). An annual recall covered 31 categories of infrequent expenses. An inventory of durable assets was used to estimate the value of the flow of services from these assets and hedonic rent equations were used to estimate the value of services provided by dwellings.

Anthropometric measurements (weight and height) were made on all children in the surveyed households who were age five years and under ($n=969$), and also on the parents of these children. Both children and their parents were weighed and measured twice, once during the first visit to the household and again during the consumption recall interview. This duplication allowed the average of the two measures to be used, which should reduce the effects of measurement error. Documentary evidence on children's date of birth was requested from the parents (e.g., birth certificates, health books) and visits were also made to health centres and hospitals to check birth records, and to churches to check baptismal records.

Market prices were collected twice during the survey in each Census Unit, with up to six outlets being surveyed on each occasion. Reports on prices were also obtained at household level to deal with the issue of items that are missing from local markets. These household-level price reports were for goods of defined specification to avoid the biases that result from using unit values (expenditure divided by quantity) as a proxy for price (Deaton, 1988).

An unusual feature of the survey was that one-sixth of the survey clusters were chosen as a “longitudinal sub-sample” and households in these clusters were revisited approximately seven months after the first interviews. All parts of the survey, including the expenditure recall, were gathered again during these revisits. One use of these longitudinal data is to correlate the two sets of observations on the same household as an estimate of the reliability ratio of survey variables (i.e. the fraction of observed variance that is due to variance in the true variable). Another use is for studying consumption smoothing and decomposing poverty estimates into transient and chronic components. However, this longitudinal sub-sample can complicate the drawing of a cross-sectional sample because, with the exception of the month of the interview, the two observations on the same household have the same identifying variables.

4. Obtaining the Data

All data from the survey are freely available to researchers and may be downloaded from the internet at www.worldbank.org/lsms/country/png/pnghome.html. The data are available in three formats: ASCII, portable SAS, and STATA. The data dictionaries, all questionnaires and manuals used in the survey, and reports on the processing of the data are also available for downloading from this site. There are also links to papers published from this and other LSMS surveys so that researchers can see what analyses have already been done with these data and what other types of analyses are possible with living standards survey data.

In addition to the raw data, a number of constructed data files are available for downloading. These constructed data files include the consumption estimates for each commodity, the consumption aggregate, and a set of temporal and spatial price deflators. The spatial deflators are the most important because the rugged terrain and poor transport infrastructure cause considerable regional price variation. One of the deflators is based on “cost-of-basic-needs” poverty lines (Ravallion and Bidani, 1994), while the other deflators are based on Fisher, Laspeyre’s, Paasche, and Törnqvist indexes. The constructed data files also include sampling weights for community, household and individual level effects, and variables describing the stratification and clustering in the sample.

Although not provided in the constructed data files, some researchers may wish to use equivalence scales. Estimates with the PNGHS data suggest that children in the 0-6 year age group are equivalent to 0.5 adults, other children are equivalent to adults, and economies of household size appear absent (Gibson and Rozelle, 1998). However, there is controversy about the identifying assumptions of these scales (Deaton, 1997) so researchers may prefer to use simple per capita measures.

With free access to the raw data files, researchers should find it easy to replicate, extend or modify the variables in the constructed data files while also contributing new empirical results. Given the increasing use of living standards survey data and the paucity of published microeconomic research on Papua New Guinea, it is to be hoped that many economists will find the PNGHS data to be a valuable addition.

References

- Deaton, A. 1988, 'Quality, quantity, and spatial variation of price', *American Economic Review*, vol. 78, pp. 418-430.
- Deaton, A. 1997, *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Johns Hopkins, Baltimore.
- Gibson, J., and Rozelle, S. 1998, 'Results of the household survey component of the 1996 Poverty Assessment for Papua New Guinea', *mimeo*, Poverty and Human Resources Division, The World Bank.
- Grosh, M., and Glewwe, P. 1996, 'Household survey data from developing countries: progress and prospects', *American Economic Review*, vol. 86, pp. 15-19.
- Howes, S. and Lanjouw, J. 1998, 'Does sample design matter for poverty rate comparisons?', *Review of Income and Wealth*, vol. 44, pp. 99-109.
- Lancaster, G., Ray, R. and Valenzuela, M.R. 1998, 'A cross-country study of household poverty and inequality on unit record household budget data', *Economic Development and Cultural Change*, vol. 48, pp. 177-208.
- Ravallion, M., and Bidani, B. 1994, 'How robust is a poverty profile?', *World Bank Economic Review*, vol. 8, pp. 75-102.

Table 1 Selected Contents from PNGHS

| <i>Theme</i> | <i>Selected contents</i> |
|--------------------------------|---|
| Demographics | Age, sex, relationship to head, birth province, marital status, daily attendance during survey period for each individual |
| Education | Qualifications, literacy and enrolments for each individual. School costs (five categories), travelling time and dropping out for current students. |
| Income sources | Activity types (24 categories) for each adult, major income source. |
| Health | Number of visits (four provider types) for each individual, birth facilities, attendants and ante-natal visits for young children. |
| Diet | 24-hour dietary recall, current and previous food shortages. |
| Housing | Dwelling materials, cost and area (measured); electricity and toilet facilities; water source, cost and fetching time for each individual. |
| Agriculture | Animal numbers, cash crops grown, agricultural input use, agricultural assets, receipt of extension advice. |
| Anthropometrics | Height, weight and age of young children (<5 years) and parents. |
| Food and recurrent consumption | Value and quantity of food produced, purchased, gifted, sold, and stocked during survey period (36 categories), value of purchases, gifts and exchanges of other recurrent consumption (20 categories). |
| Other consumption | Value of purchases and gifts given and received of non-recurrent consumption (31 categories), capital value and flow of services from durable household goods (16 categories). |
| Cash transfers | Value and intended purpose of large cash transfers between households, details on transfer partner (relationship, location, occupation, education). |
| Quality of life | Qualitative perceptions on quality of life, adequacy of services and assessment of needs (for individual male and female adults). |
| Public services | Inventory of community assets, travelling time to nearest schools, medical facilities, and transport and communication services. |
| Prices | Market prices at community level (34 categories), price reports at household level. |