The Ghana and Tanzania Urban Household Panel Surveys: 2004-2006[1]

Francis Teal
CSAE

February 2008

Overview of the Surveys

The Centre for the Study of African Economies (CSAE) at Oxford University in collaboration with the Ghana Statistical Office (GSO) and the Tanzania National Bureau of Statistics (NBS) has been conducting a labour market panel survey of urban sectors in Ghana and Tanzania since 2004. There are now three waves of this survey covering the period 2004 to 2006. No survey was conducted in 2007. It is planned to repeat the surveys during the course of 2008.

The survey collects information on incomes, education and labor market experience, household characteristics and various other modules for labor force participants (ages 15 to 60) in urban areas. For Ghana these areas span the four largest urban centers in the country: Accra (and neighboring Tema), Kumasi, Takoradi and Cape Coast. In Tanzania, the sample covers several of the largest urban areas including Arusha, Dar es Salaam, Iringa, Morogoro, Mwanza, and Tanga.

The samples were based on a stratified random sample of urban households from the 2000 census in Ghana and the 2000 Household Budget Survey (HBS) in Tanzania. While the initial sample was household based, interviews were conducted on an individual basis, and the unit of analysis can be at the individual level. A total of 830 and 543 individuals were interviewed in the first round of the survey in Ghana and Tanzania respectively, which was conducted between October 2003 and June 2004. In Ghana only there was a follow-up survey of workers in Ghana's manufacturing firms who had been surveyed from 1995. Thus the Ghana data contains those sampled on the basis of households and those drawn from firms. In using the data it is important to allow for the different basis of the two components of the sample.

Two unique features of the UPS data set are important. First, the UPS provides comparable information, including income data, on both wage employees and the self employed. All labor force participants in the selected households were to be interviewed. Thus the sample of workers spans the formal and informal sectors, public and private employees, the self-employed, unemployed and so on.

[1] These surveys have been made possible by funding from the Economic and Social Research Council (ESRC) and the Department for International Development (DFID) of the UK government. Neither the ESRC nor DFID are responsible for any uses made of this data or of views expressed based on the data. Numerous individuals have assisted in designing the questionnaires and carrying out the surveys. Neil Rankin was responsible for setting them up in the form they currently run and Justin Sandefur was the principal person responsible for both the Ghana and Tanzania survey work in 2005 and 2006.

Collecting income data on the self-employed in low-income countries is a controversial endeavor. Field guides for the World Bank's Living Standards Measurement Surveys (LSMS), which serve as the international standard for household surveys in development economics, recommend survey managers not collect this information. The stated rationale is that self-employed business people in the informal sector rarely keep written accounts and their self-reported income data may be too noisy to be of use. For household based enterprises, the distinction between business and personal expenditures may be completely alien to respondents. We acknowledge the validity of these concerns.

However, because the non-agricultural self-employed constitute a majority of the urban working population in both Ghana and Tanzania, we feel the incomes in this sector are too important to ignore. Our income measure for the self-employed is based on self-reported profits. Profits are net of routine operating expenses and gross of fixed capital expenditure, if any. The concepts of "revenue", "business costs", and "profits" are explained to respondents by enumerators with experience in conducting firm and household surveys. As the surveys are entered directly onto handheld computers, a simple mechanical check forces enumerators to go over the numbers again if revenue, cost and profit figures are inconsistent. While enumerators have reported few conceptual difficulties with this portion of the questionnaire, we feel a better test of the validity of self-reported income data is our ability to explain its variation with personal, household and business characteristics, as we attempt to do below.

A second unique feature of the UPS data set is its panel dimension. During the course of July-August 2005 the initial UPS sample was resurveyed and questions were asked which enabled us to link their activities and earnings in 2005 with the same variables in 2003/04, creating a panel of individual workers. During the period August-October 2006 a further survey in both countries was conducted. Thus the UPS constitutes one of very few household panel data sets in sub-Saharan Africa.

Methods of Data Collection

The first survey was carried out in Ghana over the period from October 2003 to early 2004. This survey was initially conducted with a paper questionnaire which was a development of the labour market questionnaires which had been used as part of the manufacturing firm enterprise surveys in which the CSAE has participated since 1992. During the course of this first round survey the paper questionnaire was replaced by one based on a hand held computer. Thus for the first round of the Ghana survey there is a mixture of paper based data and hand held computer based answers. This regrettably makes the data rather complex to get into a usable form. However after this initial experiment with hand held computers was successful all the other surveys were based on this technology. It is easiest to get a sense of how the initial questionnaire was structured from the paper

based form and we have made this available as an excel file [Ghana Survey 2003_04wkrhsehld with codes]. It is important to remember that most of the data was *not* collected by this means.

The hand held computers used a Pocket Survey based design. An html version of these surveys is available on the web for both Ghana and Tanzania.

In the first round of the survey recall questions were asked of the individuals to enable us to create a job history of their activities since they left school. This was done in the paper questionnaire by allowing for separate jobs, and in the hand held computer by allowing for as many jobs as the individual wished to identify. The time when jobs were begun and when they ended were also identified.

The Ghana household and firm worker dataset, collected during 2003/2004.

This dataset is made up of three component datasets:

1. Firm workers collected on paper questionnaires;
2. Household workers collected on paper questionnaires; and
3. Household workers collected using handheld computers.

As already mentioned the firm workers were not a random sample, they were workers who had been interviewed previously as part of work surveying the manufacturing sector. There are 1860 valid observations and the breakdown between firm and household, and paper and handheld is:

	Household	Firm	Total
Handheld	627	0	627
Paper	538	695	1,233
Total	1,165	695	1,860

The origin of the data is indicated by the dummy variables: ‘paper’, ‘handheld’, ‘household’ and ‘firm’.

The paper and handheld questionnaires differ slightly. We have attempted to set up the dataset so that it is similar for both types of questionnaire. In the dataset a prefix of *pw* (as in *pw3a5a1*) indicates that the variable comes from the paper questionnaire. A prefix of *hh* (as in *hh3a5a1*) indicates that the variable comes from the handheld questionnaire. In this case although these variables refer to the same question the paper version is numeric and the handheld version a string. If using this variable you will have to create a new variable that is numeric regardless of the origin of the data.

Those variables which are associated with activities have a suffix _? where ? is the number of the activity. *Activities are sequentially numbered: 1 for the current activity, 2 for the immediately previous activity, etc.* The suffix allows the data to be reshaped and the suffix will then become the activity_number (or whatever you choose to name it when reshaping). Using this is only necessary if you wish to use the recall dimension of the data.

Cleaning of the data.

A substantial amount of cleaning has been carried out on the data but it is essential that any user of the data check the data further.

The Tanzanian household worker dataset, collected during 2003/2004.

A virtually identical questionnaire was used in Tanzania as for Ghana. There was though no attempt to link to previous firm-level surveys.

The Ghana and Tanzania Household Worker Surveys 2005 Data

In the first round of the surveys a household roster was not compiled. In the second round it was, and an attempt was made to expand the sample by interviewing all members of the household. The identifier *wid* is a unique identifier across the rounds of the survey.

The Ghana and Tanzania Household Worker Surveys 2006 Data

In this round of the surveys there were three questionnaires, one individual based, one household based, and a fertility questionnaire addressed to women. At present only the first of these is available. The questionnaire for this round of the survey was adjusted to be in line with the objectives of the DFID funded Research Programme Consortia "Outcomes of education for pro-poor development" (RECOUP). This programme draws on researchers across three UK institutions: the Centre for Commonwealth Education in the Faculty of Education at Cambridge University; the University of Edinburgh; and the CSAE. There are a further four consortium partner institutions in South Asia and Africa: India, Pakistan, Kenya and Ghana.

The questionnaire builds on questions from a large number of existing surveys but is focused very much on the purpose of the RPC, namely to facilitate investigation of the outcomes of education. The objective is to collect detailed information about education, health, fertility, labour market outcomes, savings, assets, shocks, social capital, opinions, attitudes, and perceived well-being.

Section 1: General Household Information

This section asks households questions about location (to allow easy revisit for a possible panel), language, religion, sect and caste/biradari's/zaats/kinship structures.

Section 2: Household Roster

This section asks standard questions about all individuals in a household, irrespective of age. The 'household' has been tightly defined. The roster questions ensure we have information on household size, demographic composition, relation to head, gender, migrant status, education status, etc. of all members in the household.

Section 3: Household Services, Consumption, Asset Ownership, Borrowing and Saving

The section contains questions at the household level on standard of living, asset ownership (including land and animals), consumption expenditure and debt/savings. Section 3.5 includes questions to measure any exogenous shocks to the household in well-defined time periods.

Section 4: Education – Current and Past

This section asks detailed questions on current and past education of all individuals aged 12 and above in the household. The 4 sub-sections focus respectively on primary education, secondary education, post-secondary education and vocational schooling. Several new questions are asked. Just to take two examples:

1. Distance to (and walking time to) the nearest primary/secondary school when of school-going age are asked of *all* individuals (regardless of enrolment status when of school-going age) and will serve as instruments for endogenous schooling in earnings, health and fertility equations.
2. Whether an individual took board examinations 'privately' or as a 'regular' candidate, and whether the individual took any private tuition when studying for given exam levels are pertinent questions to ask in the South Asian context because a number of individuals (particularly girls) never enrol in secondary school but still complete various levels of education 'privately'.

Section 5: Household Livelihood and Non-Employment Income

Aims to reveal sources of household income and the main livelihood for the household.

Section 6: Economic/Other Activities – Current and Past

This section has standard questions on labour force participation, unemployment, self- and wage-employment, occupation, industry and earnings. A special feature is that it captures information on both current and most recent previous employment. This allows us to construct a pseudo-panel dataset which will enable us to see not only how education affects level of earnings at a point in time but also how it affects *change* in economic outcomes over time, e.g. change in occupation/industry/firm-size/sector and growth in earnings.

Section 7: Health, Disability and Fertility

Information on illness, injuries and accidents is collected here by allowing individuals to reveal information about the three most important episodes of ill-health/accidents within the past year. Detailed disability information is also asked, as is fertility information from all married women aged 12 and above. 'Health Outcomes' are measured both subjectively and objectively, the latter via anthropometric measures (height and weight to construct BMI), functional ability measures and questions on health behaviours such as smoking, drinking etc.

Section 8: Time Allocation

There are hardly any surveys of time allocation in South Asia or Africa, and none with links to detailed questions on education. Information collected in this section will thus enable us to break new ground and say something about the association between education and time use and whether/how that association varies by gender, poverty-status, etc.

Section 9: Subjective Well-being and Social/Other Attitudes

This section asks questions about subjective well-being, opinions/attitudes, network membership and social capital.

Section 10: Empowerment, Autonomy and Exposure to Media

This section asks questions on media-exposure, empowerment/autonomy and dowry and bride/groom prices. One particular feature is that most surveys collect no information about parental education and family background of married women. This section asks questions on father's, mother's and maternal grandfather's education, number of siblings and whether the individual's parents were/are alive during adolescence.