The transition from school to the labor market in Uganda

Preliminary outline - Do not cite or quote

Lars Vilhuber[1]

December 16, 2005

[1]Lars Vilhuber: Cornell University, CISER. lars.vilhuber@cornell.edu

Abstract

This document is a draft outline, and should not be considered in any way final. It should not be distributed.

Contents

List of Tables

List of Figures

1 Introduction

It is widely acknowledged that schooling quality is critical to the overall welfare of a developing country (Lee and Barro, 2001XXX; Fields (2004)). And developing countries, and in particular Uganda, have made tremendous investments into their educational systems. Uganda's Universal Primary Education (UPE) program, which reduced the cost of education at the primary level dramatically for most families, has had a strong impact on primary school enrollment, which increased from 3 million in 1997, year of its introduction, to 7.6 million in 2002 (World Bank; 2005; Deininger; 2003).

But increasing the average education level, and improving the quality of that education, is by itself no panacea. If job opportunities are absent for the educated, then schooling will only have delayed entry into poverty, while creating disincentives for others to continue to participate in the educational system. Countries have acknowledged this (World Bank; 2003), and have actively implemented policies to alleviate credit constraints, improve infrastructure, and generally created a better business environment, with the double goal of facilitating wealth creation by existing businesses, and removing barriers to the creation of new enterprises, in particular of small businesses and self-employment. An improved investment climate goes hand in hand with an improved educational system. Whereas the educational system provides the skills to feed the improved investment climate, a successful economy provides the incentives for individuals to participate in the educational system in the first place.

Prima facie, the incentive mechanism households and families in Uganda are faced with has the right structure. The education-earnings profile is quite steep, whether one considers individual labor earnings (Table 1 and Figure 1, based on 1999 UHS data), or total household income. Even when controlling for generational or cohort effects by restricting the sample to only workers between 20 and 30 years old, these results hold (see Table 2 and Figure 2).

However, only a small fraction of the work force actually derives their income from the formal

sector. Table 3 breaks out the fraction of workers with positive labor income, showing that in particular the agricultural sector has a large fraction of their workforce outside of the formal sector. Nevertheless, even when considering total household income, which varies much less by sector, the education-earnings profile for the household head is still very steep (Table 4 and Figure 3).

Finally, it should be noted that other policies might negatively impact or constrain education policy. In particular, since families may be faced with credit constraints despite the abolition of school fees (World Bank; 2003, pg. 163), foregone earnings of children in schools may constitute a significant barrier for more widespread education. (World Bank; 2003, pg. 163, Table 4.5) cites "engaging in income-generating activities" as the top reason for boys and one of the top five reasons for girls not attending school, despite UPE. Thus, as the investment climate improves, the potential earnings of these out-of-school children also increase, and with it the likelihood that at least some children continue to stay away from school.

The goal of this paper is twofold. In a first section, we will describe how household participation of young men and women in the educational system has varied over time. In particular the UPE should have a noticeable impact in school enrollment in the primary sector, but particular attention will be paid to the simultaneity of schooling and work within households. Furthermore, some authors (e.g. Deininger (2003)) have cautioned that without an equivalent facilitation of access to secondary, problems may arise (but see Court (1999) for a positive evaluation of the change to a fee-based structure at Uganda's Makerere University).

The second, analytical section will use both the first section in this paper as well as the separate background paper to correlate employment choices made with policy changes over time as well as historical education changes. This analysis will depend critically on the availability of the 1992/1999 panel data to be able to do a proper conditional analysis, but will also use multiple cross-sections to construct synthetic cohorts (Deaton; 1985) and how they differ from each other. Focussing on short time periods, we will be able to do comparisons between school participation in 1992, 1995, 1999 with labor market participation of the equivalent cohort in 1995, 1999, and

2002. A long-term perspective of the 1992 cohort when captured ten years later will also be a part of this analysis.

2 Theoretical Background

Consider the choice faced by families with children in school Becker (1975). The child's earnings can be noted $I(i, h, s, p)$, which is a function of the child's identity i, the fraction of total hours h spent earning such income, the level of schooling s achieved, and the policy environment p in effect at the time the work is performed. Thus, if the child were not going to school, earnings would be $I(i, 1, s, p)$, whereas if the child went to school for one additional year, earnings would be $I(i, 0, s, \cdot)$ during the school year, and $I(i, 1, s + 1, p)$ a year later. When deciding whether to send the child to school, parents also need to cover direct schooling costs $k(p)$, which may be book fees, food, and possibly lodging. Some of its components are directly related to policy, such as school fees. Uganda's UPE reduced some of those costs, but others persist. Thus, parents will send the child to school if

$$I(i, 0, s, \cdot) + \delta I(i, 1, s + 1, p) - k(p) > (1 + \delta) I(i, 1, s, p)$$

where δ is a discount factor. It can be rewritten as

$$\delta \left(I(i, 1, s + 1, p) - I(i, 1, s, p) \right) > \left(I(i, 1, s, p) - I(i, 0, s, \cdot) + k(p) \right)$$

where the left term is the (discounted) benefit from investing in an additional year of schooling, and the right term includes the direct and indirect costs of that additional year of schooling.

Although simple, this specification highlights two points. First, it is true at all levels of education, including after the end of free primary education. As long as the condition holds true, the child will go to school; when the condition no longer holds, the child leaves school. Second, the parameter p highlights the fact that policy decisions affect both sides of the equation. While UPE clearly has the effect of reducing k ($\frac{\partial k}{\partial p}$), the improvement of the economic climate can lead to a

(presumably short-term) reduction in school enrollment at certain levels of schooling if

$$\delta\left(\frac{\partial I(i,1,s+1,p)}{\partial p}-\frac{\partial I(i,1,s,p)}{\partial p}\right)<\left(\frac{\partial I(i,1,s,p)}{\partial p}+\frac{\partial k}{\partial p}\right)$$

i.e., if the reduction in schooling costs is not sufficient to offset the increase in alternate employment opportunities in the short-term. It is unlikely to occur in the early years, given the substantial reduction in schooling costs through UPE, but remember that $k > 0$ despite the UPE.

It is also worthwhile noting that other economic changes, such as an increase in the price of essential food, can be integrated into such a framework through changes in the discount factor, or through explicit modelling of credit and budget constraints.

3 Results

For this analysis, we have used cross-sections of the Uganda National Household Survey (UNHS) for 1992, 1996, 1999, and 2002. For reasons described elsewhere (see Appendix B), labor market status is not available for 2002, and weighted data is not available for 1996. Where appropriate, we present results in unweighted form if the inclusion of 1996 is useful, but otherwise present weighted results for the remaining years only.

For part of the analysis, we construct synthetic 5-year cohorts of young people, as a function of their age in 1992. Thus, the youngest cohort is between 5 and less than 10 years of age in 1992. Since the data is not in true panel form, we cannot follow these individuals over time. However, we can identify other survey respondents in later survey years, e.g in 2002, who whould have been in that same age group in 1992. Their outcomes will proxy for the unobserved outcomes of the actual 1992 survey respondents. The maintained assumption is that these two mostly distinct groups have the same characteristics, and experience on average similar outcomes.

We focus in general on two indicators of education: some degree of primary education, and

some degree of secondary education. Formally,

$$EDUC_{it}(1) = I(EDUCATION_{it} \in \{P1 - P7) \quad (1)$$

$$EDUC_{it}(2) = I(EDUCATION_{it} > P7) \quad (2)$$

3.1 Education

As other authors have pointed out (Deininger (2003); World Bank (2005)), enrollment and consequently education levels have increased. Table 5 presents results for the percentage of young people who have completed some degree at the primary level (P1-P7) and those who have completed any type of degree after the primary level, for different age groups[1].

[Table 5 about here]

In line with previous authors, and in line with the expectations of UPE, the number of children aged 5-10 and 10-15, of both genders, who have completed some level of primary education (Panel (a)), has risen substantially since 1992. In fact, by 2002, girls have surpassed boys in this category. However, in the age group of the 15 to 20-year olds, the number of young people who have no more than a primary education has dropped: Men show a drop of nearly 12 percent, and women of nearly 6 percent. This can be directly attributed to the results shown in the Panel (b), where the fraction of young people in each age group that have achieved at least one post-primary degree has risen dramatically for all of the older age groups.

Table 6 shows the same analysis as before, but for synthetic cohorts of young people, defined by their age in 1992. Thus, by design, the first column in Table 6 corresponds exactly to the first column in Table 5. Consider the first cohort, aged 5-10 in 1992. By 1997, when the UPE was implemented, these children were 10-15 - and might still have benefited from the UPE. And in

[1]Note that 1996 data cannot be tabulated, since weights were not available. Table 25 compares the unweighted results to the weighted results, for a subset of age groups and variables.

fact, we observe that their schooling rate, as a cohort, has risen by 1999. As they age another three years, that rate drops again (in Panel (a)), as they acquire additional education. Consider next the third cohort, aged 15-20 in 1992. By the time the UPE was implemented, these young people were in their early 20s, having exited the educational system: 70 percent had obtained some primary education, and another 20 percent had completed some level of secondary education (Panel (b)). By 1999, an additional 14 percent had gone on to some secondary education, but because the total number of people with either primary or secondary education had not increased, we can deduce that no signficant increase in the level of primary education occurred for this cohort. The second cohort, aged 10-15 in 1992, is intermediate to these two groups. Some members of this cohort might still have profited from the benefits of UPE, and the overall number of people with at least some primary education had grown by 3 percentage points by 1999, and continues to grow in the next 3 years to 94 percent.

[Table 6 about here]

Finally, consider how the cohort composition changes over time. Each cohort, when observed in 2002, can be compared to another cohort, 10 years older, as that cohort is observed in 1992. Thus, when we consider the cohort of 5-10 year olds, who by 2002 are 15-20 years old, the appropriate comparision is to the cohort of 15-20 year olds in 1992. Relative to that group, we observe a significantly higher educational level of the younger cohorts at the same stage in their development, relative to the older cohorts. Whereas among the younger cohort, over 95 percent had at least a primary education, only 90 percent of the older cohort had the same educational achievement at the same age, for men. For women, this differential is twice as large for the same cohort, and the differential is increasing in age.

Beyond the age of 30, the educational level of the cohorts is more or less stagnant or even decreasing for men (which might indicate some selection bias in the composition of the cohorts). The educational level of women, however, increases with age across all cohorts.

One measure of improvement of general educational level is to consider heads of household only. The fraction of young (20-30) male heads of households with at least a primary education has increased from 86% in 1992 to 93% in 2002 over the time period (Row 1 of Table 7), reflecting the move of the younger, higher-educated cohorts into a position of household heads. Interestingly, when compared to all household members (Table 5), heads of households are no more and no less likely to have a higher education nowadays.

[Table 7 about here]

When considering a disaggregation by region, all regions have seen a large increase in the educational attainment of household heads, though differences persist. Male household heads in the Central Region still have a higher educational level than those in other regions. However, in the Western Region, where only half as many household heads as in the Central Region had some level of secondary education, educational levels have caught up with the rest of the country. All non-Central Regions have nowadays achieved the 1992 level of the Central Region. Results by rural/urban status of the residence (Table 8, Panel b) also show strong upward trends in overall education levels, in particular for rural women (an increase of nearly 23 percentage points) and men (10 percentage points).

3.2 Employment patterns

3.3 Changes in the incentive structure

We next take a look at changes in the economy's incentive structure, as observed (a) by changes in the return to education and (b) by changes in the observed pattern of job obtention. We concentrate on heads of households, since a large portion of earnings are non-labor earnings, which cannot be consistently attributed to individuals. We choose to attribute them to the head of the household.[2]

[2]Furthermore, for reasons explained in Appendix B, we restrict our attention to 1992 and 1999 only.

The basic reduced form regression we estimate is

$$E_{ht} = \beta_0 + \beta_1 f_1(\text{demographics}) + \beta_2 f_2(\text{age} + \beta_3 f_3(\text{education}) \tag{3}$$

where we have generally used linear effects for demographics (gender, marital status, region), a quadratic in age, and dichotomous variables for education levels as defined in (1) and (2). E_{ht} can be either the household head's labor income, or the household's entire earnings. Depending on the left-hand side variable, Equation 3 can be interpreted as a Mincerian wage equation, or a reduced-form household production function. We focus in particular on the education coefficients.

3.3.1 Changes in the return to education

Table 10 reports select results from wage regressions by gender and survey year. For both men and women, the age profile has become steeper but also more concave between 1992 and 1999. However, the changes to the returns to education are striking. For men, returns to a primary education have decreased substantially, although for women, they have increased. For both genders, the returns to a secondary education, however, have significantly increased.

[Table 10 about here]

In Table 11, the regressions are done separately by region, for men only[3]. Results for the Eastern and Northern region differ quite markedly from the Central and Western regions. Thus, in the latter two, which experienced a much stronger increase in educational achievement (Table XX), the return to a primary education is essentially nil in 1999, whereas the returns to a secondary education are substantially increased when compared to 1992. In the Eastern and Northern regions, however, both primary and secondary education now provide very high returns, both when compared to the other regions and to the year 1992 in the own region.

[3]The sample sizes for women are quite small.

[Table 11 about here]

Table 12 presents results by age group of male household heads. In 1992, both of the younger age groups have positive returns to a primary education in the formal labor market. By 1999, only the group of 30-40 year olds still shows such returns. Returns for a secondary education also only changed significantly for the same group. Note however that the gradient in the returns to education (the difference between the returns to secondary and primary education) has increased for both of the younger groups: for the 20-30 year olds, through a decrease in the return to a primary education, for the 30-40 year olds through an increase in the return to a secondary education. This steepening of the wage-education profile did not occur for the oldest group.

[Table 12 about here]

A different way of looking at Table 12 is by cohort. Table 13 highlights results for two cohorts, those between 20 and 30 years old in 1992, and those between 30 and 40, and reports the return to education faced by these two groups in 1992 and 1999. These two groups were chosen because they were already most likely to have entered the labor market by 1992, and still be in the labor market in 1999. While the previous tables suggest that there is a secular change in the returns to education, Table 13 provides some evidence that the changes may be driven by the entry of younger cohorts into the labor market. To wit, the returns to education faced by the younger cohort in Table 13 reflect the patterns shown in previous tables, while the older cohort sees little change in the returns to secondary education as it ages.

[Table 13 about here]

However, Tables 12 and 13 also show that the number of household heads active in the formal labor market decreases with age. This may be related to a switch of the household head's income source from the formal employment economy to a status of self-employed or employer. To address this issue, we use an alternate measure of income is cumulative household income, which includes

labor and non-labor income for all household members. Tables 14 through 17 estimate the model outlined in Equation (3), using cumulative household income as dependent variable, controlling for the household head's education level only. Tables 18 through 20 expand on the same specification, adding controls for the average education of other household members, or more formally, the average presence of household members with a certain level of education, i.e.

$$EDUC_{ht}(k) = \frac{\sum_{i \in h, i \neq HEAD} I(EDUC_{it} = k)}{N_h - 1} \tag{4}$$

where h in the subscript indicates household h instead of individual i, and with some abuse of notation, individuals are a member of household h if $i \in h$, and N_h is the number of household members. By design, both sets of education variables are scaled to lie on the unit interval, making the coefficients from the regression analysis comparable. In the data, the variables are signficantly correlated, but the correlation for primary education is weak (about 10 percent), whereas the correlation for secondary education is stronger (about 39 percent).

The results show that the other household members' education contributes significantly to the explanatory power of the regression. However, for both sets of variables, the steepening of the education gradient is present for both the household head's education as well as for the other household members' education, albeit to a somewhat lesser degree. As the other tables document, the regional patterns (Table 19) and the age group patterns (Tables 20) tell essentially the same story as before.

3.3.2 Changes in the patterns of job obtention

The previous section suggests that the returns to education have grown. In this section, we consider whether the way individuals earn their income has changed. In particular, we document whether changes in the type of activity – a salaried job, or a self-employed job – constitutes the prime absorber of available workers.

Table 21 shows the fraction of each age group that is either employed as a wage- or salary-earning worker ("Salaried"), or self-employed, with or without other people as employees ("Employer"). Here, it is striking that all age groups of workers have experienced an increase in overall employment. Except for the youngest men (15-20), the increase in employment is driven primarily by increases in the "Employer" category. In general, the increase is small. Young men (15-30) have increases around 1 percentage point. Markedly, young women have seen a much larger increase in overall employment, in particular for women between 15 and 25.

The cohort analysis in Table 22 provides some elucidation. In each panel, the data in the 1999 column is roughly comparable to the 1992 column of the next-higher cohort. The cohort of 20-25 year old men in 1992 improved their employment rate from 94.20% to 95.54% by 1999. The next higher cohort of 25-30 year-old men only had an employment rate of 94.46% in 1992. Thus, the younger cohort had a higher employment rate when measured at a comparable stage of their career. This is in fact true for all male cohorts, but the pattern is not stable for the female cohorts. Note however that since there has been an across-the-board increase in employment rates, i.e., each cohort has increased their own employment rate, this result seems to be more due to an economy-wide upswing in employment opportunities.

The cross-sectional analysis in Table 21 does not show a broad pattern to how this increase in employment is achieved. For men, some age groups have seen a strong increase in salaried employment accounting for most or all of the increase in employment, whereas other age groups have seen stagnant or declining salaried employment, with employment gains stemming from increases in (self-)employer status.

Again, analysis by cohort, rather than cross-sectional age group, sheds some light on this phenomenon. The younger male cohorts have seen strong increases in salaried employment, while older cohorts have seen a decline in salaried employment. Female cohorts, on the other hand, have generally seen declines in salaried employment coupled with strong increases in employer status. For women, all increases in employment rates are driven by increased (self)-employer status.

These results also hold when contrasting rural with urban areas (Table 23). Most of the increase in female employment is driven by urban women, from a low base. Overall, employment rates were higher in rural areas in 1992, whereas they have been almost equalized by 1999.[4]

TODO: do a probit of being self-employed/employer, as a function of region, education, gender, rural/urban. Table 24.

4 Concluding remarks

[4]It is possible that this is due to data idiosynchrasies, since the questionnaires were not identical in the two years, and in particular the questions about labor market activity have changed. See Appendix C.2 for how we have chosen to standardize this variable.

References

Becker, G. S. (1975). *Human capital; a theoretical and empirical analysis, with special reference to education*, National Bureau of Economic Research.

Court, D. (1999). Financing higher education in Africa: Makerere, the quiet revolution, *Tertiary Education Thematic Group series 22883*, World Bank.

Deaton, A. (1985). Panel data from time series of cross-sections, *Journal of Econometrics* **30**(1-2): 109–126.

Deininger, K. (2003). Does cost of schooling affect enrollment by the poor? universal primary education in uganda, *Economics of Education Review* **22**(3): 291–305.
URL: *http://www.sciencedirect.com/science/article/B6VB9-4834GB0-3/2/4e5de18321c0f8e09c3ccddcf82e82fc*

Deininger, K. and Okidi, J. (2003). Growth and poverty reduction in Uganda, 1999-2000: Panel data evidence, *Development Policy Review* **21**(4): 481–509.

Fields, G. (2004). A guide to multisector labor market models, *Labor market conference*, World Bank.

Republic of Uganda (1992). Integrated Household Survey 1992 - Main questionnaire, *Technical report*, Ministry of Planning and Economic Development, Statistics Department.

Republic of Uganda (1996). Uganda National Household Survey 1996/1997 - Labour force pilot and socio-economic survey questionnaire, *Technical report*, Ministry of Planning and Economic Development, Statistics Department.

Republic of Uganda (1999a). Multipliers for Uganda National Household survey 1999/2000, *Technical report*, Ministry of Finance, Planning and Economic Development, Uganda Bureau of Statistics.

Republic of Uganda (1999b). Uganda National Household survey 1999/2000 - Socio-economic survey questionnaire, *Technical report*, Ministry of Finance, Planning and Economic Development, Uganda Bureau of Statistics.

Republic of Uganda (2002a). Uganda National Household survey 2002/2003 - Manual of instructions, *Technical report*, Uganda Bureau of Statistics.

Republic of Uganda (2002b). Uganda National Household survey 2002/2003 - Socio-economic survey questionnaire, *Technical report*, Uganda Bureau of Statistics.

World Bank (2003). Poverty reduction strategy paper annual progress report, *Report 26567-UG*, World Bank.

World Bank (2005). Uganda: From periphery to center, a strategic country gender assessment, *Report 30136-UG*, World Bank, Office of the Sector Director, Poverty Reduction and Economic Management, Africa Region.

5 Tables and Graphs

Table 1: Labor income by educational levels

Education level	1999 UGX
no level	220,701
pre-school	254,943
primary, not completed	405,444
primary completed, no secondary	716,667
secondary, not completed	1,019,301
secondary completed	1,475,608
tertiary	2,178,587

Author's calculations, Uganda NHS 1999.

Table 2: Labor income by educational levels, age 20-30

Education level	1999 UGX
no level	235,599
pre-school	242,151
primary, not completed	406,416
primary completed, no secondary	707,318
secondary, not completed	824,726
secondary completed	1,105,709
tertiary	1,654,026

Author's calculations, Uganda NHS 1999.

Table 3: Labor income by Industry (%)

	Positive labor income		
Industry	No	Yes	Total
agriculture	83	17	100
manufacturing	68	32	100
mining and quarrying	83	17	100
construction	20	80	100
utilities	11	89	100
commerce	81	19	100
banking/financial services	12	88	100
professional	10	90	100
public administration	4	96	100
transport	27	73	100
other	37	63	100
Total	72	28	100

Source: Uganda NHS 1999, author's calculations

Table 4: Household income by educational levels

Education level	Agricultural	Non-agri.
no level	934	876
pre-school	1,110	1,224
primary, not completed	1,254	1,610
primary completed, no secondary	1,374	2,350
secondary, not completed	1,946	2,654
secondary completed	2,939	3,528
tertiary	3,410	4,939

Author's calculations, Uganda NHS 1999. All numbers in thousands of 1999 UGX.

Figure 1: Labor income by educational levels

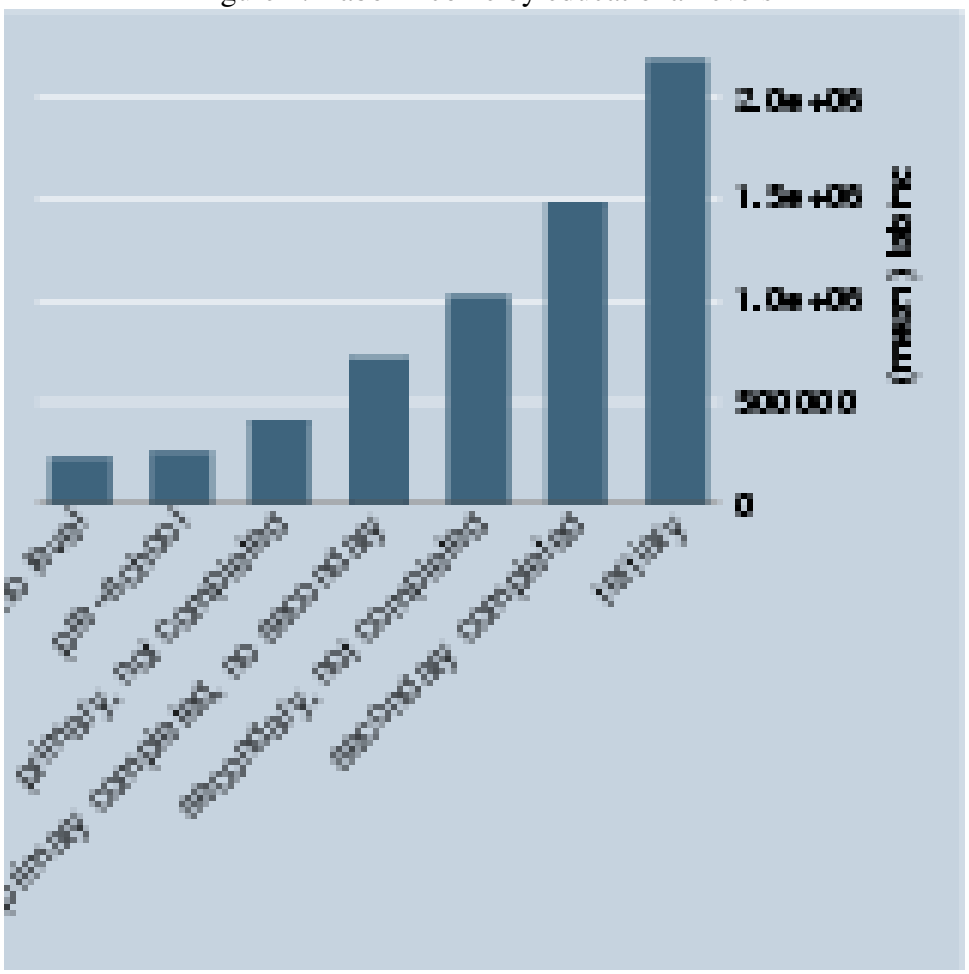

Figure 2: Labor income by educational levels, age 20-30

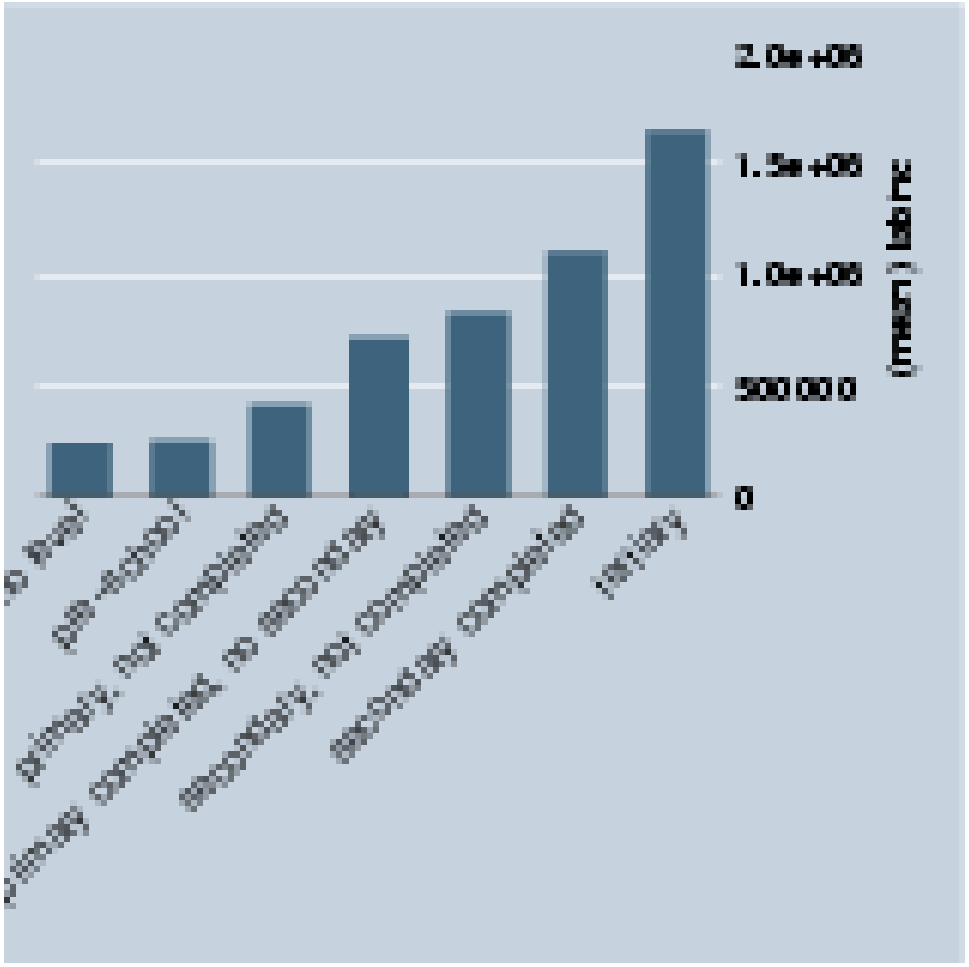

Figure 3: Household income by educational levels

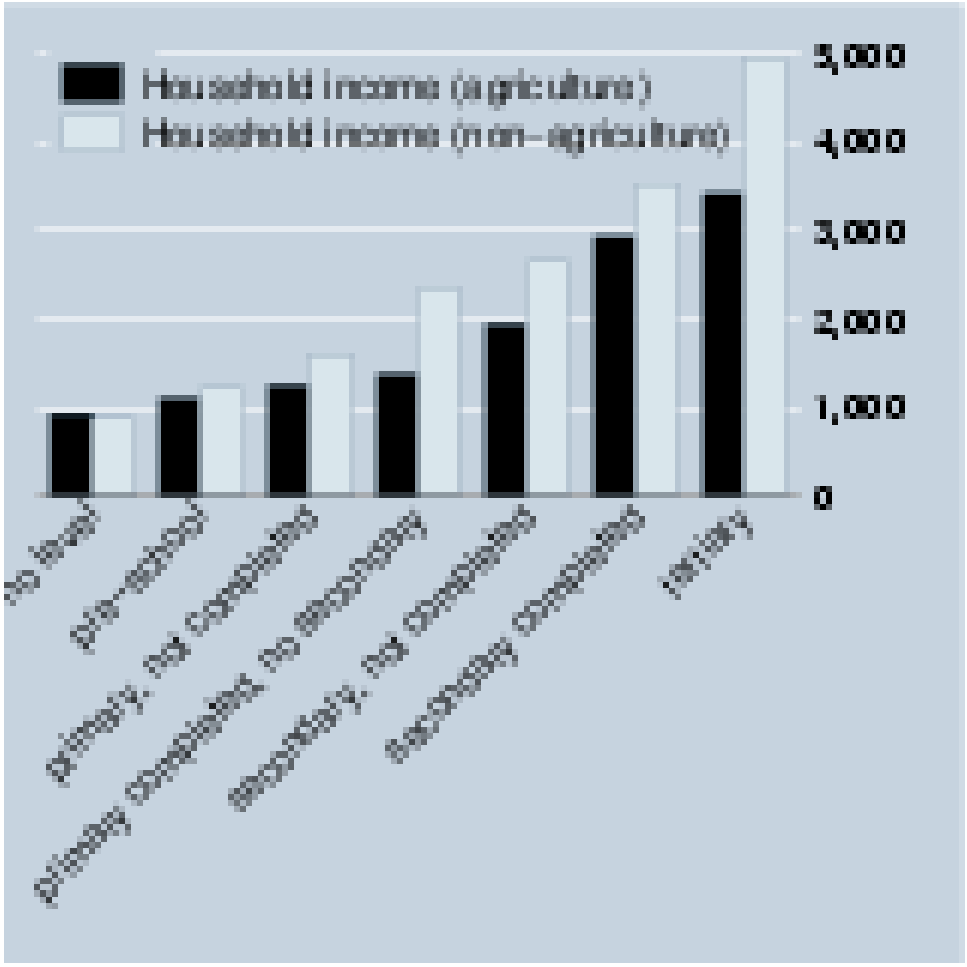

Table 5: Education by age, primary and higher

	1992	1996	1999	2002
(a) With some primary education				
Male				
5-10	39.87	.	54.58	53.40
10-15	85.25	.	89.27	90.59
15-20	70.05	.	59.95	58.09
20-25	59.57	.	52.37	52.15
Female				
5-10	39.25	.	54.56	54.95
10-15	77.67	.	87.66	90.09
15-20	63.55	.	55.69	57.13
20-25	55.91	.	53.73	55.20
(b) With primary or secondary education				
Male				
5-10	39.87	.	54.73	53.40
10-15	87.00	.	92.23	95.45
15-20	90.49	.	91.22	95.54
20-25	87.98	.	90.80	93.67
25-30	86.69	.	86.72	91.37
30-35	85.84	.	86.35	91.52
35-40	84.60	.	84.87	88.34
40-45	80.35	.	85.72	83.63
Female				
5-10	39.25	.	54.60	54.97
10-15	79.87	.	91.93	95.18
15-20	79.31	.	84.07	90.77
20-25	72.14	.	77.29	85.58
25-30	64.51	.	69.51	79.55
30-35	55.36	.	65.03	69.77
35-40	50.08	.	58.44	64.96
40-45	42.33	.	51.99	63.46

Author's tabulations, UNHS, various years.
All results weighted.

Table 6: Education by cohort

	1992	1996	1999	2002
(a) With some primary education				
Male				
5-10	39.87	.	82.73	58.09
10-15	85.25	.	50.85	52.15
15-20	70.05	.	56.18	56.36
20-25	59.57	.	57.78	58.42
Female				
5-10	39.25	.	77.84	57.13
10-15	77.67	.	52.09	55.20
15-20	63.55	.	54.35	57.77
20-25	55.91	.	52.57	51.19
(b) With primary or secondary education				
Male				
5-10	39.87	.	92.98	95.54
10-15	87.00	.	90.48	93.67
15-20	90.49	.	89.13	91.37
20-25	87.98	.	85.52	91.52
25-30	86.69	.	87.80	88.34
30-35	85.84	.	84.84	83.63
35-40	84.60	.	83.48	82.82
40-45	80.35	.	80.79	84.59
Female				
5-10	39.25	.	91.18	90.77
10-15	79.87	.	79.34	85.58
15-20	79.31	.	76.33	79.55
20-25	72.14	.	69.24	69.77
25-30	64.51	.	62.13	64.96
30-35	55.36	.	54.65	63.46
35-40	50.08	.	49.80	57.30
40-45	42.33	.	47.12	49.13

Author's tabulations, UNHS, various years.
All results weighted.

Table 7: Educational attainment of household heads

	Some primary	Some secondary	At least primary
(a)All ages			
Male			
1992	56.80	20.11	76.91
1999	56.74	25.15	81.89
2002	57.96	29.32	87.28
Female			
1992	33.44	10.67	44.11
1999	36.87	13.08	49.95
2002	45.87	19.56	65.44
(b)Young (20-30) heads			
Male			
1992	63.20	22.99	86.19
1999	64.73	25.01	89.74
2002	60.92	31.86	92.78
Female			
1992	46.27	23.21	69.47
1999	47.99	26.21	74.20
2002	54.14	31.64	85.77

Author's tabulations, UNHS, various years.

Table 8: Education by rural/urban residence, primary and higher

	1992	1996	1999	2002
(a) With some primary education				
Male				
Rural	59.99	.	63.41	63.24
Urban	51.85	.	50.97	50.69
Female				
Rural	48.32	.	56.00	60.32
Urban	56.71	.	54.92	52.24
(b) With primary or secondary education				
Male				
Rural	70.75	.	76.81	80.49
Urban	86.77	.	89.16	90.52
Female				
Rural	53.19	.	63.43	70.81
Urban	81.29	.	85.36	87.64

Author's tabulations, UNHS, various years.
All results weighted.

Table 9: Education by rural/urban residence, household heads only

	1992	1996	1999	2002
(a) With some primary education				
Male				
Rural	58.96	.	60.40	61.06
Urban	43.49	.	35.85	40.09
Female				
Rural	30.83	.	34.89	45.99
Urban	47.10	.	45.88	43.64
(b) With primary or secondary education				
Male				
Rural	74.37	.	80.21	85.83
Urban	92.66	.	91.50	95.37
Female				
Rural	37.21	.	42.56	58.44
Urban	79.57	.	83.23	86.69

Author's tabulations, UNHS, various years.
All results weighted.

Table 10: Wage regressions, by gender

Gender	Year	R^2	Married	Age	Age^2	Education:	
		(*N*)				Primary	Secondary
Male	1992	0.269	0.392 ***	0.023 **	-0.039 ***	0.317 ***	0.834 ***
		(1886)	(0.052)	(0.010)	(0.011)	(0.064)	(0.067)
Male	1999	0.365	0.489 ***	0.048 ***	-0.061 ***	0.295 ***	1.385 ***
		(1830)	(0.060)	(0.010)	(0.011)	(0.076)	(0.076)
Female	1992	0.263	0.113	0.054 **	-0.078 ***	0.277	0.879 ***
		(370)	(0.112)	(0.023)	(0.027)	(0.179)	(0.158)
Female	1999	0.384	0.240	0.163 ***	-0.202 ***	0.489 ***	1.539 ***
		(318)	(0.150)	(0.033)	(0.041)	(0.184)	(0.180)

Author's tabulations, UNHS, various years. All regressions use weights. Regressions also include region dummies.

Table 11: Wage regressions, by region

Region	Year	R^2 (*N*)	Married	Age	Age^2	Education: Primary	Education: Secondary
Central	1992	0.296	0.473 ***	0.057 ***	-0.081 ***	0.446 ***	1.084 ***
		(743)	(0.080)	(0.015)	(0.018)	(0.111)	(0.113)
Central	1999	0.356	0.491 ***	0.072 ***	-0.096 ***	0.135	1.254 ***
		(624)	(0.094)	(0.018)	(0.021)	(0.142)	(0.142)
Eastern	1992	0.133	-0.109	0.027	-0.053 **	0.147	0.497 ***
		(502)	(0.107)	(0.019)	(0.023)	(0.118)	(0.117)
Eastern	1999	0.355	0.517 ***	0.054 ***	-0.063 ***	0.674 ***	1.609 ***
		(486)	(0.130)	(0.018)	(0.019)	(0.148)	(0.145)
Northern	1992	0.270	0.057	-0.035	0.030	-0.539 *	0.425
		(117)	(0.307)	(0.048)	(0.055)	(0.301)	(0.315)
Northern	1999	0.381	0.576 **	0.008	-0.008	0.422 **	1.561 ***
		(250)	(0.222)	(0.021)	(0.021)	(0.180)	(0.167)
Western	1992	0.158	0.431 ***	-0.016	0.015	0.416 ***	0.755 ***
		(539)	(0.089)	(0.017)	(0.019)	(0.099)	(0.113)
Western	1999	0.307	0.271 **	0.064 ***	-0.072 ***	0.146	1.258 ***
		(485)	(0.128)	(0.023)	(0.027)	(0.138)	(0.143)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 12: Wage regressions, by age group and year

Age group	Year	R^2 (*N*)	Married	Age	Age^2	Education: Primary	Education: Secondary
20-30	1992	0.077	0.098	-0.095	0.212	0.554 ***	0.789 ***
		(641)	(0.082)	(0.251)	(0.484)	(0.120)	(0.119)
20-30	1999	0.162	0.032	0.083	-0.143	0.071	0.876 ***
		(546)	(0.095)	(0.325)	(0.618)	(0.146)	(0.147)
30-40	1992	0.154	0.279 **	0.074	-0.094	0.394 ***	0.997 ***
		(602)	(0.113)	(0.338)	(0.477)	(0.146)	(0.145)
30-40	1999	0.290	0.276 **	0.721 *	-1.001 *	0.396 **	1.528 ***
		(642)	(0.126)	(0.382)	(0.536)	(0.158)	(0.152)
40-50	1992	0.197	0.289 **	0.943	-1.083	0.174	0.920 ***
		(315)	(0.145)	(0.660)	(0.722)	(0.159)	(0.158)
40-50	1999	0.301	0.711 ***	0.268	-0.302	0.073	1.196 ***
		(341)	(0.192)	(0.700)	(0.765)	(0.182)	(0.183)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 13: Wage regressions, by cohort

Age group	Year	R^2	Married	Age	Age^2	Education:	
		(*N*)				Primary	Secondary
20-30	1992	0.077	0.098	-0.095	0.212	0.554 ***	0.789 ***
		(641)	(0.082)	(0.251)	(0.484)	(0.120)	(0.119)
30-40	1999	0.290	0.276 **	0.721 *	-1.001 *	0.396 **	1.528 ***
		(642)	(0.126)	(0.382)	(0.536)	(0.158)	(0.152)
30-40	1992	0.154	0.279 **	0.074	-0.094	0.394 ***	0.997 ***
		(602)	(0.113)	(0.338)	(0.477)	(0.146)	(0.145)
40-50	1999	0.301	0.711 ***	0.268	-0.302	0.073	1.196 ***
		(341)	(0.192)	(0.700)	(0.765)	(0.182)	(0.183)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 14: Regression results for household income, by gender

Gender	Year	R^2	Married	Age	Age^2	Education:	
		(*N*)				Primary	Secondary
Male	1992	0.210	0.583 ***	0.043 ***	-0.047 ***	0.267 ***	0.732 ***
		(6510)	(0.030)	(0.004)	(0.004)	(0.027)	(0.034)
Male	1999	0.384	0.473 ***	0.041 ***	-0.050 ***	0.249 ***	1.510 ***
		(2590)	(0.056)	(0.008)	(0.009)	(0.064)	(0.066)
Female	1992	0.191	0.124 ***	0.075 ***	-0.084 ***	0.255 ***	0.922 ***
		(2284)	(0.043)	(0.008)	(0.009)	(0.047)	(0.070)
Female	1999	0.449	-0.016	0.065 ***	-0.075 ***	0.755 ***	2.061 ***
		(459)	(0.132)	(0.021)	(0.023)	(0.146)	(0.147)

Author's tabulations, UNHS, various years. All regressions use weights. Regressions also include region dummies.

Table 15: Regression results for household income, by region

Region	Year	R^2 (N)	Married	Age	Age^2	Education: Primary	Education: Secondary
Central	1992	0.259	0.664 ***	0.035 ***	-0.041 ***	0.286 ***	0.939 ***
		(1937)	(0.046)	(0.007)	(0.007)	(0.057)	(0.064)
Central	1999	0.334	0.506 ***	0.033 **	-0.049 ***	0.186	1.423 ***
		(767)	(0.090)	(0.016)	(0.018)	(0.134)	(0.136)
Eastern	1992	0.161	0.569 ***	0.053 ***	-0.059 ***	0.274 ***	0.676 ***
		(1818)	(0.061)	(0.007)	(0.008)	(0.050)	(0.062)
Eastern	1999	0.351	0.447 ***	0.059 ***	-0.069 ***	0.439 ***	1.706 ***
		(723)	(0.115)	(0.015)	(0.016)	(0.120)	(0.125)
Northern	1992	0.122	0.214 **	0.061 ***	-0.064 ***	0.366 ***	0.704 ***
		(977)	(0.094)	(0.011)	(0.012)	(0.064)	(0.080)
Northern	1999	0.316	0.662 ***	0.017	-0.017	0.009	1.301 ***
		(418)	(0.172)	(0.018)	(0.019)	(0.150)	(0.156)
Western	1992	0.093	0.572 ***	0.033 ***	-0.033 ***	0.218 ***	0.500 ***
		(1793)	(0.062)	(0.007)	(0.008)	(0.048)	(0.068)
Western	1999	0.300	0.230 *	0.059 ***	-0.059 ***	0.273 **	1.547 ***
		(697)	(0.130)	(0.017)	(0.020)	(0.108)	(0.119)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 16: Regression results for household income, by age group and year

Age group	Year	R^2 (N)	Married	Age	Age^2	Education: Primary	Education: Secondary
20-30	1992	0.058	0.137 **	0.167	-0.272	0.279 ***	0.594 ***
		(2075)	(0.055)	(0.134)	(0.258)	(0.054)	(0.062)
20-30	1999	0.197	-0.195 **	0.291	-0.483	0.076	1.106 ***
		(781)	(0.097)	(0.288)	(0.549)	(0.134)	(0.140)
30-40	1992	0.141	0.343 ***	0.035	-0.038	0.328 ***	0.912 ***
		(1839)	(0.068)	(0.176)	(0.248)	(0.057)	(0.063)
30-40	1999	0.352	0.307 ***	0.767 **	-1.067 **	0.362 ***	1.844 ***
		(874)	(0.119)	(0.357)	(0.499)	(0.130)	(0.129)
40-50	1992	0.192	0.715 ***	-0.021	0.009	0.233 ***	0.917 ***
		(1025)	(0.079)	(0.341)	(0.373)	(0.070)	(0.084)
40-50	1999	0.288	0.653 ***	-0.354	0.376	0.102	1.384 ***
		(472)	(0.172)	(0.665)	(0.725)	(0.152)	(0.159)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 17: Regression results for household income, by cohort

Age group	Year	R^2	Married	Age	Age^2	Education:	
		(*N*)				Primary	Secondary
20-30	1992	0.058	0.137 **	0.167	-0.272	0.279 ***	0.594 ***
		(2075)	(0.055)	(0.134)	(0.258)	(0.054)	(0.062)
30-40	1999	0.352	0.307 ***	0.767 **	-1.067 **	0.362 ***	1.844 ***
		(874)	(0.119)	(0.357)	(0.499)	(0.130)	(0.129)
30-40	1992	0.141	0.343 ***	0.035	-0.038	0.328 ***	0.912 ***
		(1839)	(0.068)	(0.176)	(0.248)	(0.057)	(0.063)
40-50	1999	0.288	0.653 ***	-0.354	0.376	0.102	1.384 ***
		(472)	(0.172)	(0.665)	(0.725)	(0.152)	(0.159)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 18: Regression results for household income, incl. other's education, by gender

Gender	Year	R^2	Heads's education:		Other education:	
		(*N*)	Primary	Secondary	Primary	Secondary
Male	1992	0.238	0.181 ***	0.502 ***	0.325 ***	0.935 ***
		(5763)	(0.028)	(0.037)	(0.031)	(0.061)
Male	1999	0.415	0.171 **	1.179 ***	0.462 ***	1.254 ***
		(2198)	(0.071)	(0.077)	(0.072)	(0.104)
Female	1992	0.183	0.126 **	0.701 ***	0.284 ***	0.728 ***
		(1813)	(0.052)	(0.079)	(0.063)	(0.111)
Female	1999	0.492	0.492 ***	1.579 ***	0.358 *	1.597 ***
		(339)	(0.174)	(0.195)	(0.203)	(0.296)

Author's tabulations, UNHS, various years. All regressions use weights. Regressions also include region dummies.

Table 19: Regression results for household income, incl. other's education, by region

Region	Year	R^2	Heads's education:		Other education:	
		(*N*)	Primary	Secondary	Primary	Secondary
Central	1992	0.247	0.220 ***	0.715 ***	0.354 ***	0.860 ***
		(1580)	(0.063)	(0.074)	(0.065)	(0.093)
Central	1999	0.322	0.004	0.910 ***	0.611 ***	1.210 ***
		(576)	(0.177)	(0.186)	(0.158)	(0.187)
Eastern	1992	0.179	0.132 **	0.379 ***	0.457 ***	1.226 ***
		(1632)	(0.052)	(0.068)	(0.057)	(0.130)
Eastern	1999	0.374	0.450 ***	1.492 ***	0.506 ***	1.565 ***
		(625)	(0.130)	(0.139)	(0.143)	(0.236)
Northern	1992	0.137	0.294 ***	0.554 ***	0.266 ***	0.895 ***
		(946)	(0.067)	(0.087)	(0.070)	(0.248)
Northern	1999	0.322	-0.042	1.086 ***	0.190	1.288 ***
		(381)	(0.159)	(0.172)	(0.147)	(0.310)
Western	1992	0.082	0.146 ***	0.317 ***	0.217 ***	0.792 ***
		(1626)	(0.050)	(0.075)	(0.057)	(0.130)
Western	1999	0.364	0.134	1.101 ***	0.401 ***	1.734 ***
		(637)	(0.112)	(0.136)	(0.129)	(0.225)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 20: Regression results for household income, incl. other's education, by age group and year

Age group	Year	R^2	Heads's education:		Other education:	
		(*N*)	Primary	Secondary	Primary	Secondary
20-30	1992	0.101	0.170 ***	0.301 ***	0.345 ***	0.821 ***
		(1846)	(0.056)	(0.068)	(0.049)	(0.088)
20-30	1999	0.275	0.015	0.703 ***	0.509 ***	1.321 ***
		(647)	(0.149)	(0.163)	(0.123)	(0.166)
30-40	1992	0.210	0.184 ***	0.532 ***	0.516 ***	1.345 ***
		(1711)	(0.058)	(0.070)	(0.057)	(0.118)
30-40	1999	0.433	0.256 *	1.377 ***	0.551 ***	1.782 ***
		(794)	(0.135)	(0.142)	(0.126)	(0.180)
40-50	1992	0.255	0.067	0.536 ***	0.586 ***	1.926 ***
		(917)	(0.072)	(0.094)	(0.093)	(0.199)
40-50	1999	0.365	-0.098	0.837 ***	0.803 ***	2.628 ***
		(419)	(0.162)	(0.177)	(0.200)	(0.323)

Author's tabulations, UNHS, various years. All regressions are for male heads of household, and use weights.

Table 21: Employment status by age, weighted results

	1992			1999		
Age in 1992	Salaried	Employer	Overall	Salaried	Employer	Overall
Male						
15-20	22.51	68.19	90.70	25.03	67.78	92.81
20-25	21.05	73.15	94.20	20.90	75.40	96.29
25-30	24.20	70.27	94.46	24.34	71.60	95.94
30-35	24.11	71.64	95.75	25.32	70.90	96.22
35-40	23.85	72.06	95.91	22.93	73.52	96.45
40-45	25.71	67.62	93.33	16.20	79.30	95.51
Female						
15-20	22.69	47.94	70.62	13.11	69.92	83.03
20-25	17.77	57.78	75.56	18.38	60.78	79.16
25-30	16.97	65.82	82.79	14.22	71.50	85.72
30-35	9.80	73.25	83.05	11.63	71.09	82.72
35-40	13.13	70.77	83.90	7.75	78.12	85.86
40-45	7.06	76.16	83.22	10.56	74.83	85.39

Author's tabulations, UNHS, various years. Heads of household only. All means are weighted.

Table 22: Employment status by cohort, weighted results

	1992			1999		
Age in 1992	Salaried	Employer	Overall	Salaried	Employer	Overall
Male						
15-20	22.51	68.19	90.70	24.78	71.65	96.43
20-25	21.05	73.15	94.20	23.45	72.10	95.54
25-30	24.20	70.27	94.46	25.83	71.21	97.05
30-35	24.11	71.64	95.75	20.14	76.03	96.17
35-40	23.85	72.06	95.91	15.85	79.72	95.57
40-45	25.71	67.62	93.33	19.58	74.83	94.42
Female						
15-20	22.69	47.94	70.62	18.25	56.21	74.46
20-25	17.77	57.78	75.56	14.67	74.20	88.87
25-30	16.97	65.82	82.79	9.07	72.88	81.95
30-35	9.80	73.25	83.05	8.81	80.09	88.90
35-40	13.13	70.77	83.90	7.99	71.94	79.93
40-45	7.06	76.16	83.22	3.22	85.39	88.60

Author's tabulations, UNHS, various years. Heads of household only. All means are weighted.

Table 23: Employment status by urban/rural, weighted results

	1992			1999		
Age in 1992	Salaried	Employer	Overall	Salaried	Employer	Overall
Male						
Rural	17.90	78.37	96.27	15.41	80.85	96.27
Urban	51.66	34.53	86.19	55.24	39.45	94.69
Female						
Rural	8.15	75.24	83.39	7.12	77.66	84.77
Urban	30.41	44.94	75.35	26.95	55.26	82.21

Author's tabulations, UNHS, various years. Heads of household only. All means are weighted.

Table 24: Probit of choice of self-employed vs. salaried, conditional on employment

PROC PROBIT is modeling the probabilities of levels of wage_employed having LOWER Ordered Values in the response profile table (probability that wage_employed=0, not salaried)

---------- Sex, 1=Male Year of survey=1992 ----------

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.7186	0.0216	0.6762	0.7610	1103.45	<.0001
cat_married	1	0.5299	0.0033	0.5235	0.5363	26402.3	<.0001
demo_age	1	-0.0108	0.0014	-0.0135	-0.0080	58.82	<.0001
demo_age2	1	0.0055	0.0022	0.0012	0.0097	6.42	0.0113
cat_region2	1	0.2583	0.0029	0.2525	0.2641	7680.99	<.0001
cat_region3	1	0.4985	0.0037	0.4911	0.5058	17675.1	<.0001
cat_region4	1	0.0905	0.0029	0.0847	0.0962	946.42	<.0001
urban	1	-0.8272	0.0030	-0.8331	-0.8214	76529.6	<.0001
cat_primary	1	-0.0144	0.0033	-0.0209	-0.0078	18.44	<.0001
cat_secondary	1	-0.7771	0.0037	-0.7844	-0.7698	43351.7	<.0001

---------- Sex, 1=Male Year of survey=1999 ----------

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	2.2341	0.0234	2.1884	2.2799	9146.62	<.0001
cat_married	1	0.4985	0.0034	0.4919	0.5051	21891.0	<.0001
demo_age	1	-0.1096	0.0015	-0.1125	-0.1067	5447.94	<.0001
demo_age2	1	0.1751	0.0023	0.1707	0.1796	5974.30	<.0001
cat_region2	1	0.1100	0.0029	0.1043	0.1157	1418.64	<.0001
cat_region3	1	0.1904	0.0035	0.1835	0.1974	2884.78	<.0001
cat_region4	1	0.0414	0.0031	0.0354	0.0475	179.08	<.0001
urban	1	-0.9521	0.0028	-0.9576	-0.9466	116331	<.0001
cat_primary	1	0.0708	0.0036	0.0638	0.0778	396.27	<.0001
cat_secondary	1	-0.6149	0.0038	-0.6223	-0.6076	26851.4	<.0001

A Additional tables

Table 25: Education by age, comparing weighted and unweighted

	1992	1996	1999	2002
(a) With some primary education, weighted				
Male				
5-10	42.73	39.97	55.70	55.09
10-15	85.78	87.36	91.52	90.62
15-20	66.29	58.06	59.97	54.15
Female				
5-10	43.43	40.79	58.00	56.92
10-15	79.15	83.00	89.51	89.15
15-20	60.68	55.46	54.56	53.18
(b) With some primary education, unweighted				
Male				
5-10	39.87	.	54.58	53.40
10-15	85.25	.	89.27	90.59
15-20	70.05	.	59.95	58.09
20-25	59.57	.	52.37	52.15
Female				
5-10	39.25	.	54.56	54.95
10-15	77.67	.	87.66	90.09
15-20	63.55	.	55.69	57.13
20-25	55.91	.	53.73	55.20

Author's tabulations, UNHS, various years.

B Data appendix

We initially attempted to use data from 1992, 1996, 1999, and 2002 Uganda National Household Surveys (Republic of Uganda (1992, 1996, 1999b, 2002b)). According to the Republic of Uganda (1999a), a panel of households already surveyed in 1992 were re-surveyed in 1999, allowing for construction of a true household panel. However, several issues stymied the use of the full panel or pseudo-panel.

B.1 Time inconsistency in household identifiers

This problem is pertinent for the merging of 1999 with 1992 households in order to construct the 1992-1999 household panel. In 1999, the variable `hhcode` is a 13-char string, whereas in 1992, it is a numeric variable with a maximum of 9 significant digits.

It would appear that in 1992

```
hhcode92 = stratum || ea || hh
```

where all numbers are zero padded, treated as characters, and || designates the concatenation operator. In principle, it would seem possible to create

```
hhcode99 = stratum || ea || hhnumber
```

However, the effective length of ea in the 1999 survey is 5 digits, against 4 digits in 1992. Thus, even a reconstructed and hopefully consistently household identifier turns out to be time-inconsistent. Unfortunately, neither the documentation made available to us nor subsequent queries resulted in resolution of these problems, rendering the household panel inaccessible to the author.[5]

B.2 Errors in household identifiers

Household identifiers in 1996 are not unique across households in the person level files. The problem arises because the last two digits of the household identifier, which represent the sample identifier (household within enumeration area), seem to be compromised. To wit, in the household-level file "SEC1" (representing the first section of the household survey), there are only 1440 unique household identifiers, instead of each one of the 6655 records having a unique household identifier. Adding the variable `sample` to the household identifier yields a unique variable for each one of the 6655 records, but this variable is unfortunately not available in other person-level files. It is thus not possible to reliably merge different sections of the household survey with each other, making it impossible to correlate, say, economic activity (in Section 3) with educational achievement (in Section 2). It is also not possible to obtain weighted tabulations of any variable, or to tabulate any variable by region, since these data are on the household-level Section 1 file.

[5]Note that at least one pair of authors Deininger and Okidi (2003) report results from analysis of a panel. It would thus seem that such an analysis is feasible in future analyis, conditional on being able to construct such a panel.

B.3 Activity status not available in 2002

In 2002, the labor market activity was asked in a separate questionnaire. This questionnaire and the accompanying data, although described in Republic of Uganda (2002a), could not be identified clearly.

C Programs used

C.1 Standardize education

```
/* Time-stamp: <2005-09-22 14:20:47 vilhuber> */
```

attempt to standardize education

Utility macro

```
%macro educ1992(invar,outvar);
     if &invar. = 7 then &outvar.=.;
else if &invar. = 0 then &outvar.=1;
else if &invar. = 1 then &outvar.=2;
else if &invar. = 2 then &outvar.=2;
else if &invar. = 3 then &outvar.=3;
else if &invar. = 4 then &outvar.=.;
else if &invar. <18 then &outvar.=4;
else if &invar. <37 then &outvar.=6;
else if &invar. <42 then &outvar.=5;
else if &invar. <52 then &outvar.=7;
else if &invar. <92 then &outvar.=8;
else &outvar.=.;
%mend;
```

1992

```
data base1992a;
set in1992.sec4b (keep=hh tid pid s4bq2);
year=1992;
        %hhid_code(1992);
%educ1992(s4bq2,ed_current);
run;
proc sort;
by hhid_code pid tid;
run;
proc contents;
run;



data base1992b;
set in1992.sec3 (keep=hh tid pid s3q7);
year=1992;
        %hhid_code(1992);
%educ1992(s3q7,ed_highest);
run;
proc sort;
by hhid_code pid tid;
run;
proc contents;
run;

data base1992;
     merge base1992a(in=a)
           base1992b(in=b)
   ;
     by hhid_code pid tid;
     merge=a+2*b;
run;
proc freq;
table merge;
run;

proc print data=base1992(where=(merge=2) obs=10);
run;

data base1992(drop=merge);
     /* current=base1992a=merge=1 is asked of age<=30 only */
     /* highest=base1992b=merge=2 is asked of all */
     set base1992(where=(merge=3 or merge=2));
run;
```

```
/*=====================================================================
                              1996
=====================================================================*/
data base1996;
set in1996.sec2(keep=hh tid pid s2q9 s2q12);
year=1996;
        %hhid_code(1996);
     if s2q9 = 7 then ed_current=.;
else if s2q9 = -9 then ed_current=.;
else if s2q9 = 0 then ed_current=1;
else if s2q9 = 1 then ed_current=2;
else if s2q9 = 2 then ed_current=2;
else if s2q9 = 3 then ed_current=2;
else if s2q9 = 4 then ed_current=2;
else if s2q9 = 5 then ed_current=3;
else if s2q9 = 6 then ed_current=10;
else if s2q9 <17 then ed_current=4;
else if s2q9 <36 then ed_current=6;
else if s2q9 =40 then ed_current=5;
else if s2q9 =50 then ed_current=7;
else if s2q9 <90 then ed_current=8;
else ed_current=.;

%educ1992(s2q12,ed_highest);
run;
proc contents;
run;
```

```
/*=====================================================================
                                1999
=====================================================================*/
data base1999;
set in1999.sec3(keep=hhcode tid pid s3q3 s3q6);
year=1999;
        %hhid_code(1999);
     if s3q3 = -9 then ed_current=.;
else if s3q3 = 0 then ed_current=1;
else if s3q3 = 1 then ed_current=2;
else if s3q3 = 2 then ed_current=2;
else if s3q3 = 3 then ed_current=2;
else if s3q3 = 4 then ed_current=3;
else if s3q3 =11 then ed_current=10;
else if s3q3 = 5 then ed_current=4;
else if s3q3 = 7 then ed_current=6;
else if s3q3 = 6 then ed_current=5;
else if s3q3 = 8 then ed_current=7;
else if s3q3 = 9 then ed_current=8;
else if s3q3 =10 then ed_current=8;
else ed_current=.;
%educ1992(s3q6,ed_highest);
run;
proc contents;
run;


/*=====================================================================
                                2002
=====================================================================*/
data base2002;
set in2002.sec4_clean(keep=hh sid tid pid s4q2 s4q5);
year=2002;
        %hhid_code(2002);
ed_current=s4q2;
        if s4q2 = 9 then ed_current=8;
        if s4q5=1 then s4q5=0;
if s4q5=10 then s4q5=1;
%educ1992(s4q5,ed_highest);
run;

proc contents;
run;
```

now put all together

```
data lars.ed_current;
set
  base1992 (keep=hhid_code pid tid ed_current ed_highest year)
  base1996 (keep=hhid_code pid tid ed_current ed_highest year)
  base1999 (keep=hhid_code pid tid ed_current ed_highest year)
  base2002 (keep=hhid_code pid tid ed_current ed_highest year)
;
%pid_tid_code;
run;

proc sort data=lars.ed_current;
by hhid_code pid_code tid_code;
run;


data weight;
     set bryce.uganda_household(keep=hhid_code year weight
rename=(year=year_b));
     run;
proc sort data=weight;
by hhid_code;
run;

data ed_current;
     merge lars.ed_current (in=a )
           weight (in=b);
     by hhid_code;
     merge=a+b*2;
run;
```

Diagnostics

```
proc freq;
table merge*year_b merge*year /missing;
run;


proc freq data=ed_current(where=(merge=3));
table ed_current*year ed_highest*year/missing;
format ed_current ed_highest edcur.;
weight weight;
run;

proc freq data=ed_current(where=(merge=3));
title "Education weighted";
table ed_current*year ed_highest*year;
format ed_current ed_highest edcur.;
weight weight;
run;

proc freq data=lars.ed_current;
title "Education unweighted";
table ed_current*year ed_highest*year;
format ed_current ed_highest edcur.;
run;
```

C.2 Standardize activity

```
/* Time-stamp: <2005-09-22 14:24:04 vilhuber> */
```

standardize activity

Utility macro

```
%macro status1992(invar,outvar);
     if &invar. = 0 then &outvar.=0;
else if &invar. = 1 then &outvar.=1;
else if &invar. = 2 then &outvar.=2;
else if &invar. = 3 then &outvar.=3;
else if &invar. = 4 then &outvar.=4;
else if &invar. < 9 then &outvar.=5;
else if &invar. = 9 then &outvar.=6;
else if &invar. =10 then &outvar.=7;
else if &invar. <15 then &outvar.=8;
else if &invar. =15 then &outvar.=9;
else if &invar. =16 then &outvar.=8;
else if &invar. =99 then &outvar.=99;
else &outvar.=.;
%mend;
```

```
/*=====================================================================
1992
=====================================================================*/
data base1992a;
set in1992.sec6a (keep=hh tid pid s6aq2a s6aq2b s6aq2c);
year=1992;
        %hhid_code(1992);

%status1992(s6aq2a,act_status);
run;
proc contents;
run;

data base1992b;
     set in1992.sec6c (keep=hh tid pid s6cq2a s6cq2b s6cq2c);
year=1992;
        %hhid_code(1992);
%status1992(s6cq2a,act_current_status);
run;
proc contents;
run;


proc sort data=base1992a;
by hhid_code pid tid;
run;
proc sort data=base1992b;
by hhid_code pid tid;
run;


data base1992;
     merge base1992a(in=a)
           base1992b(in=b)
   ;
     by hhid_code pid tid;
     merge=a+2*b;
run;

proc freq;
table act_current_status;
run;
proc freq data=base1992;
table merge;
run;

data base1992(drop=merge);
     set base1992(where=(merge=3));
run;
```

```
/*=====================================================================
                    1996
=====================================================================*/

data base1996a;
set in1996.sec3(keep=hh tid pid s3q2 s3q9 s3q3 s3q4 s3q5);
year=1996;
        %hhid_code(1996);
        /* inactive */
        if s3q2 = 3 then do;
     if s3q9 = 1 then act_status=0;
 else if s3q9 = 2 then act_status=1;
 else if s3q9 = 3 then act_status=2;
 else if s3q9 = 4 then act_status=8;
 else if s3q9 = 5 then act_status=8;
 else if s3q9 = 6 then act_status=8;
 else if s3q9 = 7 then act_status=8;
 else if s3q9 = 8 then act_status=9;
 else if s3q9 = 9 then act_status=99;
end;
        else if s3q2 = 1 then do;
      if s3q3 =1 then act_status=4;
 else if s3q3 =2 then act_status=3;
 else if s3q3 =3 then act_status=7;
 else if s3q3 <7 then act_status=5;
 else if s3q3<10 then act_status=6;
end;
else act_status=.;

run;
proc sort data=base1996a;
by hhid_code pid tid;
run;


data base1996b;
set in1996.sec41(keep=hh tid pid s41q2 s41q3  );
year=1996;
        %hhid_code(1996);
run;
data base1996c;
set in1996.sec43(keep=hh tid pid s43q2  s43q5 s43q6 );
year=1996;
        %hhid_code(1996);
run;
proc sort data=base1996b;
by hhid_code pid tid;
run;
proc sort data=base1996c;
by hhid_code pid tid;
run;
```

```
data base1996d;
     merge base1996b(in=a)
           base1996c(in=c)
   ;
     by hhid_code pid tid;
     merge=a+2*c;

        /* inactive */
        if s41q2 = 3 then do;
           if  s43q2= 3 then do;
                if s43q6 = 1 then act_current_status=0;
            else if s43q6 = 2 then act_current_status=1;
            else if s43q6 = 3 then act_current_status=2;
            else if s43q6 = 4 then act_current_status=8;
            else if s43q6 = 5 then act_current_status=8;
            else if s43q6 = 6 then act_current_status=8;
            else if s43q6 = 7 then act_current_status=8;
            else if s43q6 = 8 then act_current_status=9;
            else if s43q6 = 9 then act_current_status=99;
   end;
           else if s43q2 = 1 or s43q2=2 then act_current_status=99;
        end; /* end of s41q2=3 */
        else if s41q2 = 1 then do;
      if s41q3 =1 then act_current_status=4;
 else if s41q3 =2 then act_current_status=3;
 else if s41q3 =3 then act_current_status=7;
 else if s41q3 <7 then act_current_status=5;
 else if s41q3<10 then act_current_status=6;
end;
else act_current_status=.;
run;
proc contents;
run;
proc freq;
table act_current_status;
run;
proc freq data=base1996d;
table merge;
run;
```

```
/* now put d and a together */
data base1996;
     merge base1996a(in=a)
           base1996d(in=d drop=merge)
   ;
     by hhid_code pid tid;
     merge=a+2*d;
run;

proc freq data=base1996;
table merge;
run;

data base1996(drop=merge);
     set base1996(where=(merge=3));
run;



/*=====================================================================
                              1999
=====================================================================*/

%macro status1999(invar,outvar);

     if &invar. <=3 then &outvar.=&invar.-1;
else if &invar. = 4 then &outvar.=4;
else if &invar. = 5 then &outvar.=3;
else if &invar. = 6 then &outvar.=7;
else if &invar. = 7 then &outvar.=5;
else if &invar. = 8 then &outvar.=6;
else if &invar. = 9 then &outvar.=8;
else if &invar. =10 then &outvar.=8;
else if &invar. =11 then &outvar.=8;
else if &invar. =19 then &outvar.=99;
else &outvar.=.;
%mend;


data base1999;
set in1999.sec2(keep=hhcode tid pid s2q9 s2q15);
year=1999;
        %hhid_code(1999);
%status1999(s2q9,act_status);
%status1999(s2q15,act_current_status);
run;
proc freq;
table act_current_status;
run;
proc contents;
run;
```

```
/*=================================================================
                              2002
=================================================================*/

/* no activity can be found in 2002 */
/*
data base2002;
set in2002.sec4_clean(keep=hh sid tid pid s4q2);
year=2002;
        %hhid_code(2002);
act_status=s4q2;
        if s4q2 = 9 then act_status=8;
run;

proc contents;
run;
*/
```

now put all together

```
data lars.act_status;
set
  base1992 (keep=hhid_code pid tid act_status act_current_status year)
  base1996 (keep=hhid_code pid tid act_status act_current_status year)
  base1999 (keep=hhid_code pid tid act_status act_current_status year)
/*   base2002 (keep=hhid_code pid tid act_status act_current_status year)
*/
;
%pid_tid_code;
run;



proc freq data=lars.act_status;
table act_status*year /missing;
format act_status status.;
run;
```

C.3 Putting it together

```
/* put the files together, for now without weights */

proc sort data=lars.act_status out=act_status;
by hhid_code pid_code tid_code;
run;

proc sort data=lars.ed_current out=ed_current;
by hhid_code pid_code tid_code;
run;

proc sort data=bryce.uganda_demo out=demo;
by hhid_code pid_code tid_code;
run;

proc sort data=bryce.uganda_household out=hhold;
by hhid_code;
run;

data uganda_demo_all;
merge
act_status (in=a)
ed_current (in=b)
demo       (in=c)
;
       by hhid_code pid_code tid_code;
merge=100*a+10*b+c;
        demo_cohort = demo_age-(year-1992);
        label demo_cohort = "demo_age-(year-1992)";
run;

proc freq data=uganda_demo_all;
table merge*year;
run;

data lars.uganda_demo_all;
     merge uganda_demo_all(in=a rename=(merge=pmerge))
           hhold(in=b);
     by hhid_code;
     hmerge=10*a+b;
run;

proc freq data=lars.uganda_demo_all;
table hmerge*year;
run;

proc contents data=lars.uganda_demo_all;
run;
```

C.4 Utilities

```
/* Time-stamp: <2005-09-22 11:53:36 vilhuber> */

%macro hhid_code(year);

/* requires Bryce's macros */
%hhid_code&year.;

%mend;

proc format ;
value edcur
1 = "Never attended"
2 = "Left school"
3 = "Nursery"
4 = "Primary"
5 = "Post-primary"
6 = "Secondary"
7 = "Post-secondary"
8 = "Higher ed (university, diploma)"
10= "Apprenticeship"
;
value $region
                '1' ='Central'
                '2' ='Eastern'
                '3' ='Northern'
'4' ='Western'
;
value status
0 = "Too young, too old"
1 = "Disabled"
2 = "Student"
3 = "Own account worker"
4 = "Employer"
5 = "Gov't employee"
6 = "Private employee"
7 = "Unpaid family worker"
8 = "Not working/unemployed"
9 = "Not interested in work"
99= "Other"
;
run;



*macros to create HHID_CODE;
*Use in data step;
```

```
%macro hhid_code1992;
   length hhid_code $ 17.;
   hhid_code='0000000000000';
   if hh le 99999999 then do;
       substr(hhid_code,2,8)=put(hh,8.);
   end;
   else do;
       substr(hhid_code,1,9)=put(hh,9.);
   end;
   substr(hhid_code,14,4)=1992;
   label hhid_code = "Household ID standardized";
%mend hhid_code1992;


%macro hhid_code1996;
   length hhid_code $ 17.;
   hhid_code='0000000000000';
   if hh le 99999999 then do;
       substr(hhid_code,2,8)=put(hh,8.);
       if newhh lt 10 then do;
   substr(hhid_code,13,1)=put(newhh,1.);
       end;
       else if newhh ge 10 then do;
   substr(hhid_code,12,2)=put(newhh,2.);
       end;
   end;
   else do;
       substr(hhid_code,1,9)=put(hh,9.);
       if newhh lt 10 then do;
   substr(hhid_code,13,1)=put(newhh,1.);
       end;
       else if newhh ge 10 then do;
   substr(hhid_code,12,2)=put(newhh,2.);
       end;
   end;
   substr(hhid_code,14,4)=1996;
   label hhid_code = "Household ID standardized";
%mend hhid_code1996;


%macro hhid_code1999;
    length hhid_code $ 17.;
    substr(hhid_code,1,13)=hhcode;
    substr(hhid_code,14,4)=1999;
    label hhid_code = "Household ID standardized";
%mend hhid_code1999;
```

```
%macro hhid_code2002;
    length hhid_code $ 17.;
    hhid_code='0000000000000';
    substr(hhid_code,1,11)=put(hh,11.);
    substr(hhid_code,14,4)=2002;
    label hhid_code = "Household ID standardized";
%mend hhid_code2002;


%macro pid_tid_code;
    length pid_code $ 2. tid_code $ 1.;

    pid_code='';
    tid_code='';

    if pid lt 10 then do;
substr(pid_code,2,1)=put(pid,1.);
    end;
    else if pid ge 10 then do;
substr(pid_code,1,2)=put(pid,2.);
    end;
    if tid ge 0 then do;
tid_code=put(tid,1.);
    end;
    label pid_code = "Person within HHID_CODE identifier";
    label tid_code = "Type of person within HHID_CODE identifier";
%mend pid_tid_code;
```

This version: $Id: uganda-report.tex 50 2005-12-06 04:14:29Z vilhuber $