Private sector development and income dynamics: A panel study of the Tanzanian labour market

Submitted paper:
THIRD IZA / WORLD BANK CONFERENCE ON EMPLOYMENT AND DEVELOPMENT

Conference 'Key Topic':
'CAUSES AND CONSEQUENCES OF FORMALITY AND INFORMALITY'

SIMON QUINN* and FRANCIS TEAL†

This draft: January 19, 2008

Abstract

In this paper, we use a three-period panel of Tanzanian households to explore the determinants of earnings and earnings growth from 2004 to 2006. In doing so, we draw particular attention to the role of education and to the importance of heterogeneity between more and less formal occupations. Several important conclusions emerge. Education is found to have a significant convex effect upon earnings levels, but to have had no significant effect upon earnings growth (indeed, there is some suggestion that education may have had a *negative* impact). This suggests that recent Tanzanian growth may have reflected an 'unskill-biased technological change', providing relative reward to informal skills rather than to formal education. Further, there are interesting insights into the age-earnings relationship: the relationship is found significantly to be concave in levels, yet age is *not* found significantly to have affected earnings growth. This suggests that the concave levels relationship is driven by workers' participation decisions, rather than by a concave earnings trajectory at the level of the individual worker. Finally, we find significant evidence of variation between formal and informal enterprises, and between sizes of enterprises within these different employment sectors.

*Centre for the Study of African Economies and All Souls College, University of Oxford (D.Phil candidate): simon.quinn@economics.ox.ac.uk.
†Centre for the Study of African Economies, University of Oxford.

1 Introduction

The distinction between formal and informal employment is fundamental to understanding the Tanzanian labour market. This paper presents results from a recent panel survey of households in urban Tanzania (the Tanzania Household Worker Survey) that covers both wage-earners and the self-employed. We explore the determinants of earnings levels in a standard Mincerian framework. However, the longitudinal structure of the data allows us to go further: in this paper, we focus primarily upon explaining income *growth*, and we do so at the level of the individual worker. This allows new insights into the income process, and provides a new angle from which to understand the key features observed in the standard earnings level regressions.

Several insights emerge. We find that education has a significant relationship to earnings levels, and that the shape of the relationship is convex. Intuitively, one might therefore expect that education is also significant in explaining earnings growth. However, we find that this is *not* the case. This suggests that recent Tanzanian growth may have reflected an 'unskill-biased technological change', providing relative reward to informal skills rather than to formal education. In contrast, the age-earnings profile is found to be significantly concave. The levels regression can shed no light on whether this concavity represents a concave earnings trajectory at the level of the individual worker or whether it simply reflects systematic differences in workforce participation. However, the growth regressions can assist here: we find no significant negative age effect on growth, suggesting the importance of participation differences rather than a concave underlying trajectory.

Finally, we find substantial and significant differences between different kinds of occupations — in effect, between formal and informal employment. The Tanzania Household Worker Survey includes imputed earnings figures for the self-employed; we are therefore able to compare directly the experiences of self-employed workers and wage-earners. We find important differences between those two categories: in terms of earnings levels, in the relative return to education and in the return to firm size. Even within the category of wage-earners, we find substantial differences in the level and the structure of returns to different sectors (private enterprises, public enterprises and the civil service). In short, we find a labour market marked substantially by heterogeneity.

The paper proceeds as follows. **Section 2** summarises the data, with particular reference to formality and informality in the labour market. **Section 3** explores the determinants of earnings levels. It uses a Mincerian framework, and shows a significant convex relationship between education and earnings; this relationship is shown to be robust to concerns of survey attrition. **Section 4** extends the Mincerian framework to consider the determinants of earnings growth. Education is *not* found to explain significantly workers' earnings growth; to the extent that it does so, higher education appears to be associated with *lower* relative earnings growth. In this section, too, we deal formally with issues of survey selection, and again find that they do not substantially change the key results. **Section 5** concludes.

2 Summary of data

2.1 The Tanzania Household Worker Survey

The Tanzania Household Worker Survey is a panel survey of Tanzanian households conducted by the Centre for the Study of African Economies at the University of Oxford. The survey has been run in 2004, 2005 and 2006, in various urban Tanzanian locations (Arusha, Dar es Salaam, Iringa, Morogoro, Mwanza and Tanga). Early results from the first two survey rounds have already been published: see **Sandefur, Serneels and Teal (2007)**. This paper is the first to analyse all three survey rounds. The Survey records information about a wide variety of issues. In this paper, we confine our attention to issues of income, with particular focus upon the role of age and tenure, of attained education and of occupation characteristics. We confine attention in each survey round to individuals aged between 16 and 65 (inclusive) reporting an occupation and a strictly positive income; we leave issues of unemployment for further research. **Table 1** summarises the sample size across each period; it shows that the survey records a total of 1651 income-earning observations on 957 individuals, of whom 358 were observed twice and 168 observed three times.

Table 1: **Summary of sample**

	2004	2005	2006	Individuals
2004 only	114	0	0	114
2005 only	0	111	0	111
2006 only	0	0	206	206
2004 & 2005 only	230	230	0	230
2004 & 2006 only	57	0	57	57
2005 & 2006 only	0	71	71	71
2004, 2005 & 2006	168	168	168	168
Observations	**569**	**580**	**502**	
			Total observations:	**1651**
			Total individuals:	**957**

2.2 Education in Tanzania

The Survey records individuals' education as a categorical variable for highest educational attainment; a years-of-education variable can then be constructed from this measure. Since the end of World War II, attainment categories in Tanzania have been similar to those of the British system: the structure "was fully formalised at the end of the British period into four years of primary, four years of middle, and four or six years of secondary education (O and A level, respectively)" (Buchert 1994, 61). The structure was reformed again in the late 1960s:

> [a]fter 1968, formal education comprised seven years of primary, four years of 'ordinary' secondary and two years of 'advanced' secondary education. . . The examination and certification points introduced by the British administraation were maintained after standard VII (the Primary School Leaving Certificate), form IV (the Certificate of Secondary Education) and form VI (the Advanced Certificate of Secondary Education), with additional examination points for quality control at standard II and form II. . .
>
> (Buchert 1994, 107)

Table 2 summarises this, with reference to the Tanzania Household Worker Survey sample.

Table 2: **Educational attainment in the survey sample**

		Number of respondents				
Highest attainment	**Years**	**Total**	**All years**	**2004**	**2005**	**2006**
None	0	137	30	94	78	73
Pre-1968 system. . .						
Primary	4	55	8	35	35	28
Middle	8	52	7	39	24	20
O Level	12	29	2	23	20	10
A Level	14	6	3	5	5	4
Tertiary	17	4	1	3	4	1
Post-1968 system. . .						
Primary	7	462	82	245	279	256
O Level	11	178	32	100	111	95
A Level	13	31	6	24	21	14
Tertiary	16	3	1	1	3	1
Total observations. . .		**957**	**168**	**569**	**580**	**502**
Average years of education. . .		**7.08**	**7.09**	**7.02**	**7.25**	**6.97**

2.3 Income and employment in the Survey

All income-earners in the Survey were required to assign themselves to one of two mutually exclusive categories: wage-earners and the self-employed. Wage-earners were then further required to categorise themselves into one of three mutually exclusive sectors: the civil service (defined to include the public sector, NGOs, religious or charitable organisations, *etc*), state-owned enterprises and private businesses (whether Tanzanian- or foreign-owned). Wage-earners in the private sector and in state-owned enterprises were later asked about the number of their employees; this was a categorial variable from which a single measure was constructed.[1] The single measure was constructed in order to provide a basis of comparison to the number of employees reported to be working for the businesses of the self-employed; this was constructed by adding the number of family employees and the number of non-family employees (both entered as an integer response, rather than categorically). In subsequent regression analysis, the 'firm size' and 'number of employees' enter in logs; this variable was constructed using the $f(x) = \ln(x + 1)$ transformation.

Tables 3 and 4 summarise the occupational and sectoral categories in the Survey; **Table 3** shows the decomposition for each year, while **Table 4** summarises movement for individuals observed in more than one round. Several points may be made immediately. First, on the maintained assumption that the first round of the survey is a representative sample, it is clear that informality is a key characteristic of the Tanzanian labour market: approximately two-thirds of interviewed respondents in 2004 reported being self-employed. Within the remaining one third, private enterprise was clearly the dominant sector. Second, sectoral movement was reasonably substantial over the course of the panel; this was particularly the case between wage earnings and self-employment, with more respondents moving from the latter to the former than *vice versa*. Third, it bears noting that the sectoral decomposition of the 2006 round seems noticeably different to that of the earlier rounds; relative to those rounds, the 2006 round appears to over-sample wage-earners relative to the self-employed; within the wage-earning sector, it appears to over-sample the civil service and private enterprise relative to public enterprise. Though occupational transitions reflect some of this change (in particular, the transition from self-employment to wage income), the magnitude of those transitions does not seem to justify the extent of the year-by-year change in sectoral decomposition. In short, subsequent analysis will need to remain cognisant that the representativeness of the pooled sample appears to have slipped, particularly in the 2006 survey round.

Table 3: **Occupational categories, pooled**

Occupational category	2004	2005	2006	Total
Self-employed	**379**	**410**	**263**	**1052**
Wage-earners	**190**	**170**	**239**	**599**
Civil service	46	44	79	169
Public enterprise	36	41	21	98
Private enterprise	108	85	139	332
Total	**569**	**580**	**502**	**1651**

Table 5 summarises the interaction between education and occupation. It shows that civil servants and employees of public firms have, on average, the most education among the pooled observations, followed by employees of private firms. Importantly, it also shows that the number of respondents having completed a tertiary qualification is very few; this will require care when interpreting the relative effect of tertiary education upon income and income growth.

We turn, then, to consider income — including the relationship between income and occupation. Income in the Survey is calculated in one of two ways. For wage earners, income is taken to be the respondent's reported wage earnings. For the self-employed, income is calculated as imputed revenue less inputs, labour and indirect costs. This approach allows a direct comparison between the income of those in formal and informal employment. Income is measured thousands of Tanzanian shillings per week (deflated), and is expressed in natural log terms. **Table 6** summarises the resultant income measure across different employment categories. The table shows that, on average, wage-earners earn more than the self-employed — and that, among wage-earners, civil servants and employees of public enterprises have similar earnings profiles (with civil servants generally appearing to earn more), which are noticeably higher than those of employees of private enterprises. **Figures 1 and 2** use kernel density estimates to show these features across the entire sample distribution.

[1] Categories were: (i) fewer than five employees (constructed as two employees); (ii) between six and 10 employees (constructed as eight employees); (iii) between 11 and 20 employees (constructed as 15 employees); (iv) between 21 and 50 employees (constructed as 35 employees); (v) between 51 and 100 employees (constructed as 75 employees); and over 100 employees (constructed as 150 employees).

Table 4: **Occupational and sectoral transitions**

			2004 → 2005	2004 → 2006	2005 → 2006
Wage-earner	→	**Wage-earner**	**115**	**72**	**79**
Civil	→	Civil	33	26	26
Civil	→	Public	0	0	0
Civil	→	Private	0	0	1
Public	→	Civil	0	0	1
Public	→	Public	29	15	17
Public	→	Private	0	1	0
Private	→	Civil	1	0	0
Private	→	Public	1	0	0
Private	→	Private	51	30	34
Wage-earner	→	**Self-employed**	**10**	**10**	**5**
Civil	→	Self-employed	0	0	0
Public	→	Self-employed	1	0	1
Private	→	Self-employed	9	10	4
Self-employed	→	**Wage-earner**	**5**	**28**	**20**
Self-employed	→	Civil	1	5	4
Self-employed	→	Public	1	3	1
Self-employed	→	Private	3	20	15
Self-employed	→	**Self-employed**	**268**	**115**	**135**
Total:			**398**	**225**	**239**

Table 5: **Education and occupation**

	Number of respondents															
	Self-Employed				**Civil**				**Public**				**Private**			
Highest attainment	’04	’05	’06	Pool	’04	’05	’06	Pool	’04	’05	’06	Pool	’04	’05	’06	Pool
No education	76	63	41	**180**	1	1	2	**4**	4	2	1	**7**	13	12	29	**54**
Primary	209	253	179	**641**	11	12	27	**50**	7	9	3	**19**	53	40	75	**168**
Middle	20	17	11	**48**	3	2	6	**11**	5	3	2	**10**	11	2	1	**14**
O Level	67	69	29	**165**	17	18	35	**70**	15	19	11	**45**	24	25	30	**79**
A Level	6	6	2	**14**	14	11	9	**34**	2	5	3	**10**	7	4	4	**15**
Tertiary	1	2	1	**4**	0	0	0	**0**	3	3	1	**7**	0	2	0	**2**
Average years	**6.3**	**6.6**	**6.3**	**6.4**	**10.2**	**10.0**	**9.4**	**9.8**	**9.2**	**10.1**	**10.3**	**9.8**	**7.4**	**7.6**	**6.5**	**7.1**
Total observations	**379**	**410**	**263**	**1052**	**46**	**44**	**79**	**169**	**36**	**41**	**21**	**98**	**108**	**85**	**139**	**332**

Table 6: **Log earnings by occupation (’000 Tsh per month, deflated)**

	Obs.	**Median**	**Mean**	**S.Dev.**	**Min.**	**Max.**
Pooled	1651	10.881	10.914	0.908	7.480	14.590
Self-employed	1052	10.698	10.725	0.884	7.480	14.141
Wage-earners	599	11.175	11.247	0.853	8.255	14.590
Civil	169	11.638	11.646	0.692	9.611	13.917
Public	98	11.676	11.670	0.779	8.774	13.609
Private	332	10.922	10.919	0.808	8.255	14.590

Figure 1: **Income kernel densities: Wage-earners and self-employed**

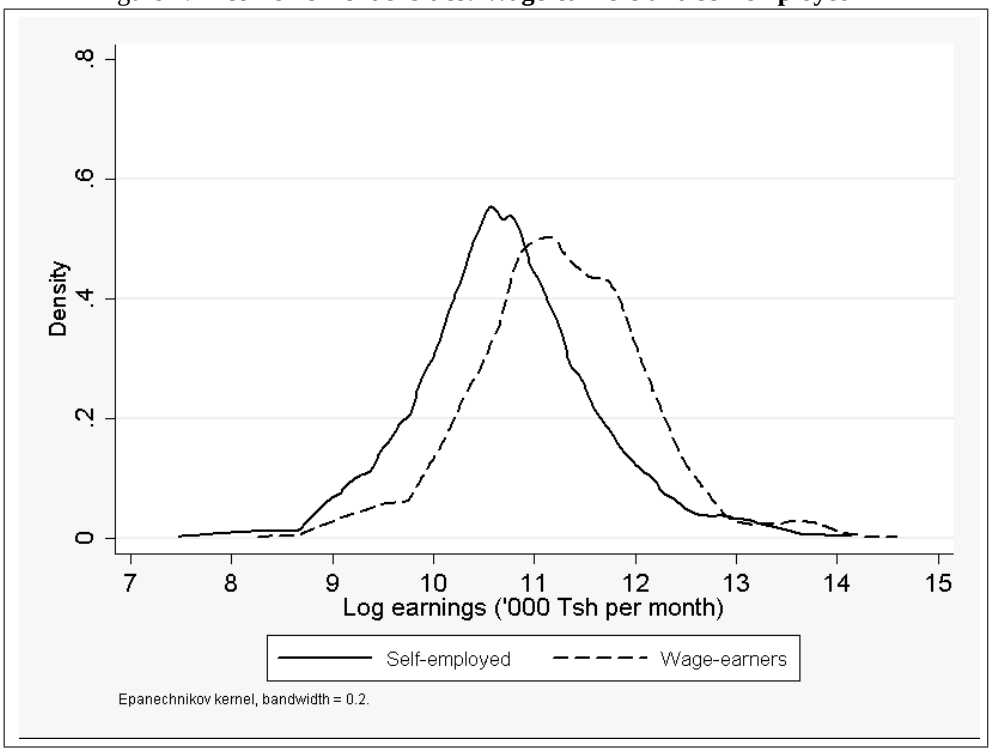
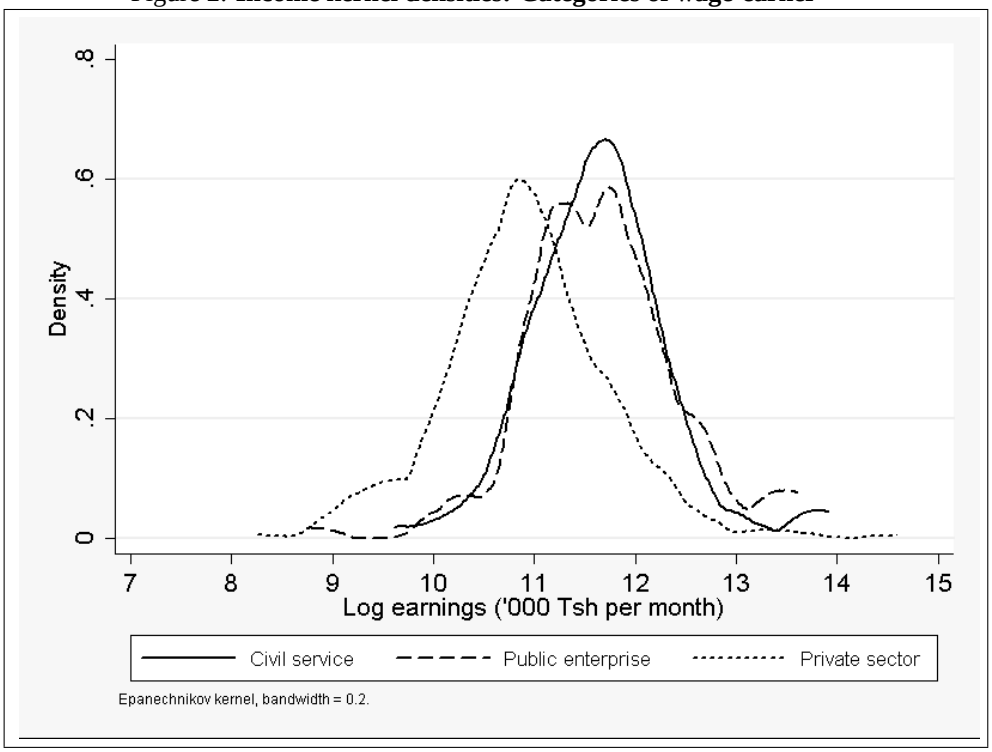


Figure 2: **Income kernel densities: Categories of wage-earner**



2.4 Other data in the Survey

Several other variables deserve explanation. **Region dummies** record in which of the six enumeration regions the respondent was found. The Survey is not a tracking survey, and respondents were not followed into new regions; thus, each respondent's region variable remains constant across all observations. **Table 7** summarises.

Table 7: **Regional disaggregation**

Region	2004	2005	2006	Observations	Respondents
Arusha	80	83	104	267	120
Dar es Salaam	232	241	161	634	356
Iringa	67	72	25	164	108
Morogoro	60	43	74	177	137
Mwanza	77	73	54	204	116
Tanga	53	68	84	205	120

Age has the natural meaning and is measured in years; by construction, every respondent aged exactly one year between survey rounds. **Tenure** refers to the number of years that each respondent has worked in his or her present occupation. By construction, tenure is not allowed to exceed the respondent's age less ten years.

3 Earnings levels

Quantitative research on the determinants of earnings in Tanzania has a long and intriguing history. One of the earliest contributions was that of **Boissiere, Knight and Sabot (1985)** (revised as Knight and Sabot (1990, Chapter 3)).[2] Boissiere et al. used a Mincerian framework to study a cross-sectional sample of 179 Tanzanian income-earners surveyed in 1980 in Dar es Salaam. The authors found an estimated return to completing secondary education (relative to not doing so) of approximately 25%, which was highly significant. However, when proxy measures of cognitive skill and reasoning ability were added to the regression (including the result from a Raven's Coloured Progressive Matrices test), the coefficient fell to approximately 10% and lost its significance. Boissiere et al. showed, in the Tanzanian context, that more educated employees are likely to earn more — but that this may be attributable to underlying cognitive abilities rather than schooling itself.

Pissarides (2002) similarly reported Mincerian estimates of returns to education in Tanzania, from a household survey of 1046 households across the country and from a survey of 546 small enterprises in five urban areas (both conducted in 1991). Pissarides estimated a return of over 10% per year of schooling from the household survey, and of approximately 4% from the enterprise survey. These results suggest that the returns to education differed significantly across different sectors of the Tanzanian labour market, and that there was a lower return to education in the informal sector; this paper will shortly consider both issues in the current context.

More recently, **Söderbom, Teal, Wambugu and Kahyarara (2006)** considered results from repeated cross-section surveys in the Kenyan and Tanzanian manufacturing sectors. The Tanzanian surveys occurred in 1993, 1994, 1999 and 2000/2001, and related to a total of 2738 workers. For Tanzania, the authors found an average marginal return to a year of education of between approximately 6% and approximately 13% — with evidence that this return had increased between 1993 and 2000/2001 (particularly among employees younger than 30 years old). Importantly, Söderbom et al. analysed not only the *slope* of the return function, but also its *shape*; this concern was particularly motivated by policy considerations (page 262):

> The shape of the earnings function is a key factor for understanding how policies of education expansion will impact on incomes. If innovations in educational policy impact primarily on those with high education costs, and the earnings function is concave, then returns to such reforms will be relatively high. . . If in fact the earnings function is convex, so that the marginal returns to education are lowest for the individuals with the least education, giving priority to investment in primary education may have little impact on poverty unless the individuals affected by the reforms proceed to higher levels of education.

Traditional views suggest that the earnings function is concave in education: see Psacharopoulos (1994) and Psacharopoulos and Patrinos (2004) (cited in Söderbom et al. (2006)). However, Söderbom et al. find that marginal returns *increase* with increased education: both in Tanzania and Kenya, the earnings function is found to be convex, and this result is robust to endogeneity (by an IV approach).

[2] Psacharopoulos (1981) summarises earlier studies of the issue across various countries; no research on Tanzania is mentioned.

The literature on earnings levels in Tanzania thus suggests several points. First, education is generally found to have a significant effect upon log earnings. Second, this relationship appears to be convex. Finally, these conclusions are susceptible to concerns of endogeneity, which may be addressed by instrumental variables and/or by ability proxies. These are important issues to bring to the present data. We do so now.

3.1 Levels identification: A Mincerian framework

We identify the basic levels relationship with a Mincerian semi-log earnings equation: see Mincer (1974). For any individual i, we allow log earnings (y_{it}) to be explained linearly by a vector of time-variant characteristics ($\mathbf{x}_{1it}$) and a vector of time-invariant characteristics ($\mathbf{x}_{2i}$),

$$y_{it} = \alpha_0 + \alpha_1 \mathbf{x}_{1it} + \alpha_2 \mathbf{x}_{2i} + u_i + \epsilon_{it}, \tag{1}$$

where the unobservables u_i and ϵ_{it} represent a mean-zero fixed effect and mean-zero time-variant shock respectively, and α_1 and α_2 are conformable coefficient vectors.

Four categories of explanatory variables are used; as reported in subsequent regressions, they are as follows. First, the **age-earnings profile**: this comprises (time-varying) measures of age and of tenure, each entering with a polynomial specification. Second, and of central interest, are measures of individuals' **educational attainment**. As explained, the primary measure is categorical, with the 'Education (years)' measure imputed as explained in **Table 2**. By construction, an individual's education measure is time invariant. Third are **occupation characteristics**: measures of enterprise size (as explained), and sector dummy variables. Both enterprise size and employment sector are allowed to vary over time. Finally, there are **other variables** — time dummies, region dummies and the gender dummy. By construction, only the time dummies are time-variant.

3.2 Levels estimation: OLS

Without specifying a distribution for ϵ_{it}, assume initially that the explanatory variables are linearly independent of the unobservable characteristics:

$$\mathbb{E}(\mathbf{x}_{1it} u_i) = \mathbb{E}(\mathbf{x}_{1it} \epsilon_{it}) = \mathbf{0};$$
$$\mathbb{E}(\mathbf{x}_{2i} u_i) = \mathbb{E}(\mathbf{x}_{2i} \epsilon_{it}) = \mathbf{0}.$$

The model, then, is semi-parametric and the appropriate estimator is OLS, run on all observations pooled. **Tables 8a and 8b** report the results. Regressions (1)–(3) use a single measure of education; regressions (4)–(6) include the squared term; and regressions (7)–(9) use instead the underlying categorical variables. In each set of three regressions, the first regression pools all observations, while the second and third regressions bifurcate the sample into self-employed and wage-earners respectively; self-employment thus becomes the omitted reference category for the sectoral dummy variables.

Tables 9a and 9b further decompose the initial results, using the categorical education variables. Regressions (1)–(3) repeat the earlier regressions (7)–(9); regressions (4)–(6) separate the different wage sectors comprising regression (3).

The results are interesting for several reasons. First, consider the **age-earnings profile** and **tenure-earnings profile**. All basic regressions show a significant and concave age-earnings profile; though the tenure coefficients also suggest a concave tenure-earnings profile, tenure is only found to be significant in the linear term (and is not significant once the regression is decomposed by sector). The concavity of the age-earnings relationship is an interesting result, though the levels regression alone does not allow it a clear interpretation. On the one hand, the relationship could reflect an age-earnings trajectory *at the level of the individual worker*. Alternatively, the relationship may result from the labour-force participation decision varying with workers' earning ability; if, for example, higher earners are likely to leave the labour force earlier, a levels regression may show a concave age-earnings profile *even if* no individual worker follows a concave earnings trajectory. A consideration of growth dynamics, however, will shortly shed some light on this distinction.

Given the close relationship between the variables discussed earlier, it is not necessarily surprising that significant concave relationships cannot be identified for *both* age and tenure. However, it is relevant that, as between the two, it is age that is significant — this may suggest a system of remuneration across the Tanzanian labour market that rewards seniority more prominently than it rewards the development of workplace-specific skills. The important exception is the public enterprise sector (regression (5) of Tables 9a and 9b). In this sector, *neither* age nor tenure is significant; the importance of this will be considered shortly.

Second, consider the **education-earnings relationship**. As expected, this is significant and positive. In the simplest linear specification (regressions (1)–(3) of Tables 9a and 9b), we estimate a general return of approximately 5.5%

per additional year of education; approximately 4% for the self-employed and approximately 7% for wage-earners. When the variable is allowed to enter under a quadratic specification, the quadratic term is significant and *positive*: education is found to have a convex effect upon earnings. The importance of this result in the context of the existing literature and potential policy issues has already been discussed. Specifying education according to its underlying categorical variable (regressions (7)–(9) of Tables 9a and 9b) confirms the same trend, with some nuance — specifically, the coefficients on middle school certificates and tertiary degrees are not significantly different from zero for the self-employed, and the coefficients on both A-Levels and tertiary degrees are smaller than the coefficient on O-Levels. It is not clear how to interpret this. On the one hand, this may be an artifact of a small number of relevant identifying observations; as **Table 5** showed, there are only 14 and 4 observations on self-employed workers with A-Level and tertiary qualifications respectively. Thus, it may be that the true relationship is indeed convex in that region, but not identifiable as such with the present data. Alternatively, it may also be the case that concavity in the education-earnings function arises at a much lower level of education among the self-employed than among wage-earners. Far from obscuring a true *convex* relationship, the small number of relevant identifying observations may be hiding a *concave* return, peaking at O-Level standard. This would certainly match an intuitive sense that delaying the completion of formal education may impede, rather than assist, a student moving into self-employment. The data are simply not able to resolve the issue.

At any rate, however, the data *do* suggest a markedly different education-earnings relationship between different sectors of the economy. The returns for the self-employed are an example of this. So too are the returns to education disaggregated across wage-earning sectors: regressions (4)–(6) of Tables 9a and 9b. It was noted earlier that neither age nor tenure is estimated to have a significant effect upon the earnings of those in public enterprises. Concomitantly, it is not surprising that significant returns to education are estimated for those employees; the sector appears strongly to reward education rather than other demographic characteristics (interestingly, gender does not appear to have a significant separate effect either). In contrast, the civil service — where age returns are highest — appears not to have any significant positive education-earnings relationship (save for the exception that, at the 90% level, the 50 civil service respondents with primary education appear to earn less than the four civil service respondents with no formal education). Regressions (4)–(6) are consistent with the intuition that the civil service strongly rewards seniority rather than education; that public firms seem to place the emphasis so strongly in the other direction — even compared to private firms — may be a matter of more surprise.

Third, consider **occupation characteristics**. Two points deserve noting. First, as the kernel density graphs showed (**Figures 1 and 2**), employees of the civil service earn more than employees of public enterprises, who earn more than those in the private sector; there is no significant difference between the return to employees of the private sector and the self-employed. Second, there is a large and significant coefficient on enterprise size (whether the number of employees, for the self-employed, or the 'firm size', for wage employees). This suggests immediately the importance of unobservable firm-specific attributes in determining worker income (whether causative both of higher firm size and higher earnings — *e.g.* management quality — or whether factors allowing firm size to drive earnings directly — *e.g.* production synergies, *etc*). It is notable that — across the three specifications allowing direct comparison (regressions (1), (4) and (7) of Tables 8a and 8b) the coefficient on number of employees is so much larger than the coefficient on firm size. It is unclear — and beyond the scope of the present work — the extent to which this dichotomy is driven by the attenuation involved in recording firm size by a categorical variable; at any rate, it shows that unobservable enterprise-specific characteristics are highly important both for wage-earners and for the self-employed. Thus, the data suggest not only the importance of heterogeneity *between* sectors, but also significant heterogeneity effects among different enterprises *within* each sector. Finally, note the estimates on the **region dummies**. They show — with the exception of Morogoro — that, after controlling for the other factors, workers outside Dar es Salaam (the capital) earn significantly less than those within the capital.

In short, several important conclusions emerge from the levels relationship. First, we estimate a concave age-earnings and tenure-earnings profiles, though the concavity is significant only for the former. Second, we estimate a significant effect of education on earnings, and find the effect to be significantly convex. Third, we find substantial and significant heterogeneity, both between sectors and within sectors. In regressions that are omitted here for brevity, we instrument for the potential endogeneity of education (using as instruments a worker's parents' education and occupation) and, separately, include as a separate regressor a Raven's Progressive Matrices score (to proxy for unobserved ability); neither approach changes these key conclusions.

Table 8a: **Determinants of log earnings, 2004-2006**

	Pooled 1	Self-Emp. 1	Wage 1	Pooled 2	Self-Emp. 2	Wage 2	Pooled 3	Self-Emp. 3	Wage 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable: ***Log earnings***									
Age and tenure:									
Age (years)	0.049 (0.013)***	0.053 (0.016)***	0.043 (0.019)**	0.051 (0.013)***	0.054 (0.016)***	0.05 (0.019)***	0.048 (0.013)***	0.048 (0.016)***	0.052 (0.019)***
$Age^2/100$	-.051 (0.015)***	-.060 (0.02)***	-.037 (0.025)	-.056 (0.015)***	-.062 (0.02)***	-.048 (0.024)**	-.052 (0.015)***	-.052 (0.019)***	-.053 (0.025)**
Tenure (years)	0.013 (0.007)*	0.009 (0.009)	0.017 (0.01)*	0.014 (0.007)**	0.011 (0.009)	0.016 (0.01)	0.014 (0.007)**	0.01 (0.009)	0.017 (0.01)
$Tenure^2/100$	-.022 (0.019)	-.020 (0.025)	-.025 (0.026)	-.022 (0.018)	-.023 (0.025)	-.014 (0.026)	-.022 (0.019)	-.022 (0.025)	-.020 (0.028)
Education:									
Education (years)	0.055 (0.006)***	0.041 (0.008)***	0.069 (0.008)***	-.002 (0.015)	0.008 (0.021)	-.018 (0.024)			
$Education^2/100$				0.465 (0.11)***	0.301 (0.17)*	0.631 (0.154)***			
Highest is primary certificate (0/1)							0.215 (0.061)***	0.223 (0.075)***	0.144 (0.107)
Highest is middle school certificate (0/1)							0.233 (0.122)*	0.013 (0.152)	0.431 (0.192)**
Highest is O Level (0/1)							0.6 (0.071)***	0.616 (0.099)***	0.532 (0.106)***
Highest is A Level (0/1)							0.804 (0.095)***	0.401 (0.197)**	0.891 (0.118)***
Highest is a tertiary degree (0/1)							1.127 (0.226)***	0.384 (0.407)	1.466 (0.225)***

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Continued...

Table 8b: **Determinants of log earnings, 2004-2006** (cont.)

	Pooled 1	Self-Emp. 1	Wage 1	Pooled 2	Self-Emp. 2	Wage 2	Pooled 3	Self-Emp. 3	Wage 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Occupation characteristics:									
Log (no. employees) (=0 if wage)	0.53 (0.071)***	0.543 (0.071)***		0.511 (0.072)***	0.527 (0.072)***		0.505 (0.071)***	0.519 (0.071)***	
Log (firm size) (=0 for civil service)	0.118 (0.02)***		0.106 (0.022)***	0.117 (0.02)***		0.106 (0.022)***	0.116 (0.021)***		0.106 (0.022)***
Civil service dummy (public sector, NGO, etc)	0.687 (0.066)***		0.651 (0.088)***	0.637 (0.066)***		0.638 (0.087)***	0.628 (0.068)***		0.641 (0.089)***
Public enterprise dummy (state-owned firm)	0.33 (0.109)***		0.294 (0.086)***	0.261 (0.107)**		0.245 (0.084)***	0.274 (0.108)**		0.261 (0.083)***
Private sector dummy	-.081 (0.069)			-.097 (0.068)			-.100 (0.069)		
Other:									
Time dummy, 2005	0.089 (0.045)**	0.069 (0.057)	0.138 (0.07)**	0.091 (0.045)**	0.072 (0.057)	0.132 (0.068)*	0.089 (0.044)**	0.072 (0.056)	0.14 (0.069)**
Time dummy, 2006	0.291 (0.047)***	0.294 (0.064)***	0.292 (0.068)***	0.306 (0.047)***	0.307 (0.065)***	0.294 (0.068)***	0.304 (0.048)***	0.31 (0.064)***	0.306 (0.068)***
Region dummy (reference: Dar): Arusha	-.162 (0.055)***	-.212 (0.072)***	-.052 (0.083)	-.179 (0.055)***	-.227 (0.073)***	-.054 (0.082)	-.182 (0.055)***	-.199 (0.071)***	-.071 (0.083)
Region dummy (reference: Dar): Iringa	-.271 (0.071)***	-.219 (0.088)**	-.336 (0.119)***	-.251 (0.071)***	-.212 (0.088)**	-.279 (0.119)**	-.261 (0.072)***	-.227 (0.089)**	-.291 (0.119)**
Region dummy (reference: Dar): Morogoro	-.094 (0.067)	-.097 (0.091)	-.069 (0.095)	-.071 (0.066)	-.093 (0.09)	-.015 (0.094)	-.070 (0.066)	-.086 (0.09)	-.020 (0.093)
Region dummy (reference: Dar): Mwanza	-.154 (0.058)***	-.154 (0.08)*	-.126 (0.076)*	-.167 (0.057)***	-.158 (0.079)**	-.146 (0.076)*	-.159 (0.058)***	-.140 (0.079)*	-.169 (0.08)**
Region dummy (reference: Dar): Tanga	-.125 (0.069)*	-.094 (0.089)	-.122 (0.112)	-.116 (0.069)*	-.095 (0.09)	-.091 (0.111)	-.117 (0.069)*	-.084 (0.09)	-.120 (0.111)
Gender dummy (1 = male)	0.288 (0.038)***	0.357 (0.05)***	0.177 (0.06)***	0.283 (0.038)***	0.353 (0.05)***	0.155 (0.059)***	0.283 (0.038)***	0.363 (0.05)***	0.158 (0.061)***
Const.	8.971 (0.234)***	9.045 (0.307)***	8.836 (0.354)***	9.065 (0.233)***	9.095 (0.304)***	8.958 (0.356)***	9.117 (0.233)***	9.136 (0.301)***	8.951 (0.366)***
Obs.	1646	1049	597	1646	1049	597	1646	1049	597
R^2	0.341	0.224	0.43	0.349	0.227	0.448	0.349	0.24	0.445

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Table 9a: **Determinants of log earnings, 2004-2006**

	Pooled	Self-Employed	Wage	Civil	Public	Private
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: ***Log earnings***						
Age and tenure:						
Age (years)	0.048 (0.013)***	0.048 (0.016)***	0.052 (0.019)***	0.135 (0.044)***	0.028 (0.059)	0.077 (0.024)***
$Age^2/100$	-.052 (0.015)***	-.052 (0.019)***	-.053 (0.025)**	-.144 (0.049)***	-.003 (0.076)	-.093 (0.032)***
Tenure (years)	0.014 (0.007)**	0.01 (0.009)	0.017 (0.01)	0.008 (0.019)	-.002 (0.022)	0.015 (0.015)
$Tenure^2/100$	-.022 (0.019)	-.022 (0.025)	-.020 (0.028)	0.011 (0.047)	-.004 (0.06)	-.009 (0.047)
Education:						
Highest is primary certificate (0/1)	0.215 (0.061)***	0.223 (0.075)***	0.144 (0.107)	-.702 (0.422)*	0.448 (0.372)	0.2 (0.112)*
Highest is middle school certificate (0/1)	0.233 (0.122)*	0.013 (0.152)	0.431 (0.192)**	-.178 (0.49)	0.466 (0.211)**	0.672 (0.381)*
Highest is O Level (0/1)	0.6 (0.071)***	0.616 (0.099)***	0.532 (0.106)***	-.097 (0.429)	0.857 (0.216)***	0.489 (0.119)***
Highest is A Level (0/1)	0.804 (0.095)***	0.401 (0.197)**	0.891 (0.118)***	-.038 (0.423)	1.238 (0.347)***	1.160 (0.197)***
Highest is a tertiary degree (0/1)	1.127 (0.226)***	0.384 (0.407)	1.466 (0.225)***		1.462 (0.367)***	1.200 (0.197)***

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Continued...

Table 9b: **Determinants of log earnings, 2004-2006** (cont.)

	Pooled	Self-Employed	Wage	Civil	Public	Private
	(1)	(2)	(3)	(4)	(5)	(6)
Occupation characteristics:						
Log (no. employees) (=0 if wage)	0.505 (0.071)***	0.519 (0.071)***				
Log (firm size) (=0 for civil service)	0.116 (0.021)***		0.106 (0.022)***		0.092 (0.051)*	0.116 (0.028)***
Civil service dummy (public sector, NGO, etc)	0.628 (0.068)***		0.641 (0.089)***			
Public enterprise dummy (state-owned firm)	0.274 (0.108)**		0.261 (0.083)***			
Private sector dummy	-.100 (0.069)					
Other:						
Time dummy, 2005	0.089 (0.044)**	0.072 (0.056)	0.14 (0.069)**	0.085 (0.138)	0.126 (0.149)	0.174 (0.098)*
Time dummy, 2006	0.304 (0.048)***	0.31 (0.064)***	0.306 (0.068)***	0.131 (0.113)	0.228 (0.249)	0.375 (0.095)***
Region dummy (reference: Dar): Arusha	-.182 (0.055)***	-.199 (0.071)***	-.071 (0.083)	-.172 (0.227)	-.294 (0.272)	0.016 (0.104)
Region dummy (reference: Dar): Iringa	-.261 (0.072)***	-.227 (0.089)**	-.291 (0.119)**	-.128 (0.29)	-.228 (0.252)	-.284 (0.153)*
Region dummy (reference: Dar): Morogoro	-.070 (0.066)	-.086 (0.09)	-.020 (0.093)	-.245 (0.125)*	-.243 (0.166)	0.252 (0.153)*
Region dummy (reference: Dar): Mwanza	-.159 (0.058)***	-.140 (0.079)*	-.169 (0.08)**	-.251 (0.14)*	-.155 (0.234)	-.095 (0.129)
Region dummy (reference: Dar): Tanga	-.117 (0.069)*	-.084 (0.09)	-.120 (0.111)	-.347 (0.138)**	-.179 (0.227)	0.032 (0.161)
Gender dummy (1 = male)	0.283 (0.038)***	0.363 (0.05)***	0.158 (0.061)***	0.139 (0.111)	0.102 (0.155)	0.17 (0.092)*
Const.	9.117 (0.233)***	9.136 (0.301)***	8.951 (0.366)***	8.792 (0.963)***	9.409 (1.053)***	8.451 (0.455)***
Obs.	1646	1049	597	168	98	331
R^2	0.349	0.24	0.445	0.34	0.518	0.342

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

3.3 Selection correction

The emphasis of this paper lies in its exploration of the determinants of growth. In that context, our primary concern for endogeneity must lie in the endogeneity of selection: the fear that our results on the determinants of earnings growth (to be presented shortly) are driven by the changing composition of the survey, rather than by underlying growth dynamics. We will shortly seek, therefore, to address formally the issue of endogenous attrition and its effects upon our growth results. It is an important parallel to that forthcoming analysis that we understand what effects, if any, endogenous selection plays in driving the levels results just presented. Thus, though we omitted for brevity robustness checks using instrumental variable and proxy variable approaches, we pause to analyse formally the issue of endogenous attrition.

We have already noted potential selection problems — the sectoral decomposition of the 2006 survey appeared quite different to that of the previous rounds; moreover, it is clear from **Table 7** that different regions were sampled in noticeably different proportions in the different rounds. The standard method of correcting for such selection concerns is that of **Heckman (1979)**. This method could be implemented in this case straightforwardly. However, when we turn shortly to consider the determinants of growth, we will need to treat the data as a panel (rather than, as in the present case, simply pooling the observations). Dealing with attrition issues in a linear unobserved-effects panel model is complicated when implementing Heckman's methodology: **Wooldridge (2002, 585)**. We seek, therefore, a methodology that will prove relatively straightforward and intuitive in the panel case; for consistency, we now implement the same methodology in the present pooled case.

The methodology used is the inverse probability weighting ('IPW') method of **Moffitt, Fitzgerald and Gottschalk (1999)**. We assume that the first period (2004) is a representative cross-section of the underlying population. We allow that an individual is observed in subsequent period t if $s_{it} = 1$, and we make the strong assumption that, conditional upon some *initial* observed vector $\mathbf{z}_{i1}$, s_{it} is independent of log earnings y_{it} and the time-variant and time-invariant explanatory variables ($\mathbf{x}_{1it}$ and $\mathbf{x}_{2i}$ respectively, using the previous notation). That is, we assume 'selection on observables':

$$\Pr(s_{it} = 1|y_{it}, \mathbf{x}_{1it}, \mathbf{x}_{2i}, \mathbf{z}_{i1}) = \Pr(s_{it} = 1|\mathbf{z}_{i1}). \tag{2}$$

Moreover, the methodology assumes that "attrition is an absorbing state" (Wooldridge 2002, 585). This requires, then, that we confine attention to those individuals observed in the 2004 round, and discard individuals observed only in 2004 and 2006; our concern, then, will be whether correcting for endogeneity substantially changes the basic OLS estimates for *those* individuals. Further, it requires the choice of some conditioning vector $\mathbf{z}_{i1}$. For present purposes, we choose simply the current dependent and explanatory variables in the first period: y_{i1}, $\mathbf{x}_{1i1}$ and $\mathbf{x}_{2i}$; these seem the most important determinants of individual outcomes and opportunity, and hence survey inclusion. (Importantly, they also include the region dummies, so capture regional differences in survey efficacy.)

The Moffitt et al. (1999) methodology — like that of Heckman — requires the estimation of an *a priori* probability of inclusion for each individual for each period after the original period: $\hat{p}_{it}, t \neq 1$. This can conveniently be done by a probit estimation. **Tables 10a and 10b** report the results (including the marginal effects), for the probability of observation in 2005 and 2006 respectively. (Given the assumptions of the IPW methodology — specifically, that attrition is an absorbing state — the probit for 2006 retention is actually a probit on the probability of retention in 2005 *and* 2006.)

Even before proceeding to the second stage, these regressions are highly informative of attrition issues in the sample. Relative to the 2004 sample, the 2005 survey is biased towards formality (as indicated by the significant coefficients on firm size and the public enterprise dummy), and significantly under-samples the Morogoro region. The 2006 survey apparently suffered more serious attrition problems; aside from a formality bias (significant coefficients on the civil service and public enterprise dummies), the survey appears to under-sample poorer and younger respondents, and shows clear regional variation in retention. Importantly, educational attainment does not significantly explain attrition in either the 2005 or 2006 round. While this should hardly eliminate concerns of selection bias on the education coefficients, it suggests that the concern is more pressing regarding regional variation and issues concerning formality and informality.

Following the Moffitt et al. (1999) methodology, we proceed to weight observations in the subsequent OLS regression by the inverse of the estimated probability: thus observation i at time t is weighted by $1/\hat{p}_{it}$. By implication of equation (2), $\hat{p}_{i1} = 1$ for all individuals. **Tables 11a and 11b** perform this second step. Regressions (1)–(3) repeat the basic OLS specification on the subsample of observations for individuals observed in the first period; regressions (4)–(6) then show the corresponding estimates from the weighted OLS procedure.

Several points of caution emerge for our interpretation of the OLS estimates. First, the estimates on age — while remaining very close to the OLS results — largely lose their significance. Further, the estimated linear effect of

tenure increases noticeably for the self-employed, producing a significant coefficient on the pooled linear term. Thus, it appears that endogeneity of selection may have contributed to the strong result on age, and to the apparent relative importance of age over tenure. Second, the selection correction changes marginally the significance on occupational dummies in the pooled regression. Finally, the correction removes the significance on the negative coefficient on the Iringa dummy; unsurprising, perhaps, given the high attrition in that region in 2006.

Much more important, though, are the estimates that do *not* substantially change: the point estimates and the significance of education (with the exception of the dummy for attaining a primary certificate) and of most of the occupation variables. With the exception of the estimate on primary education, the significance of the education dummies is not substantially affected by the correction; nor are the estimates nor the significance of the occupation characteristics; nor are the estimates on the occupation dummies (notwithstanding the changes in significance noted earlier). Despite changing the magnitude and the significance of the estimated return to primary education, **Tables ?? and ??** show that the selection correction does *not* change the earlier conclusion that the education-earnings relationship is significantly convex. In short — at least for those individuals observed in 2004 — the OLS estimates appear to be robust to endogenous attrition. Importantly, the important conclusions reached earlier for the pooled OLS regressions also seem to hold generally for the same subsample. We conclude that the earlier results are robust not only to corrections for endogeneity of regressors (addressed by the IV and the ability proxy), but also to corrections for endogeneity of selection.

4 Earnings growth

> Everything we do stresses book learning, and underestimates the value to our society of traditional knowledge and the wisdom which is often acquired by intelligent men and women as they experience life, even without their being able to read at all.
>
> (Nyerere 1968)

So warned Tanzania's founding president in a famous polemic on education policy in 1967. Whatever may have been the wisdom of the sentiment then, it is certainly a valuable caution to labour economists today. In this section, we seek to discover the key determinants of income *growth* — as distinct from income *levels* — by exploiting the panel structure of the data. We are particularly concerned to track the key findings of the levels regressions into the growth context. Thus, we are concerned to understand the relationship between age and income growth; this will shed light on whether the concave age-earnings profile identified earlier represents an individual-level earnings trajectory or simply an artifact of systematic differences in participation. We are concerned to understand whether the endogeneity in earnings levels — both between sectors and within sectors — is similarly evident in experiences of earnings growth. Finally — and perhaps most importantly — we are interested in the extent to which education determines earnings growth. This is a question having obvious policy relevance. However, more generally, it is also a question that suggests insights into the nature of the recent Tanzanian growth experience: insights into how that growth has been distributed; into whether recent growth has followed earnings levels in emphasising education or whether, as Nyere suggested, there has been comparable value in informal skills.

The ability to track the dynamics of growth at an employee level is an exciting development that has accompanied the increase in panel data sets for developing nations. The reason — and a fundamental justification for using such data — is well explained by **Cichello, Fields and Leibbrandt (2005, 145–146)**:

> This use of a series of cross-sectional surveys has added to our understanding of the evolving nature of the labour market over the 1990s. However, there are inherent difficulties associated with using a series of cross-sections to explore labour market dynamics. If the data sets tell similar stories over time, as is the case with the unemployment studies, there is no way of knowing whether this is because the labour market has operated in a stable fashion between the surveys or whether there have been changes in earnings and employment for certain individuals and groups but these changes have netted out to similar aggregate snapshots. Generally, repeated cross-sections cannot deal with the movement of people between labour market segments, or between jobs within sectors or with related real earnings changes over time. This is a particular concern if policy makers are really interested in knowing which specific individuals or groups are experiencing movement in the labour market and, in particular, who are the winners and losers from the current operation of the labour market.

Intuitively, one might expect the 'winners' from the dynamic operation of the labour market to match closely those identified as 'winners' by the levels relationship, so that wealthier and more educated workers enjoy both higher income *levels* and more rapid income *growth*: success begets success. Certainly, this would accord with the intuitive sense of many that liberalisation and growth effects a divergence of income and opportunity between the wealthier and the poorer.

Table 10a: **First-stage selection regression: Probit for sample retention against 2004 variables**

	2005 (coefficients)	2005 (marginal effects)	2005 and 2006 (coefficients)	2005 and 2006 (marginal effects)
	(1)	(2)	(3)	(4)
Dependent variable: ***Sample retention***				
Log earnings	0.053	0.018	-.145	-.045
	(0.078)	(0.027)	(0.085)*	(0.026)*
Age and tenure:				
Age (years)	0.035	0.012	0.082	0.025
	(0.039)	(0.014)	(0.044)*	(0.014)*
$Age^2/100$	-.045	-.016	-.111	-.034
	(0.047)	(0.016)	(0.055)**	(0.017)**
Tenure (years)	0.008	0.003	0.025	0.008
	(0.021)	(0.007)	(0.024)	(0.008)
$Tenure^2/100$	-.037	-.013	-.090	-.028
	(0.055)	(0.019)	(0.067)	(0.021)
Education:				
Highest is primary certificate (0/1)	0.113	0.039	0.299	0.092
	(0.175)	(0.06)	(0.189)	(0.058)
Highest is middle school certificate (0/1)	-.309	-.113	0.17	0.055
	(0.285)	(0.109)	(0.337)	(0.114)
Highest is O Level (0/1)	-.024	-.008	-.035	-.011
	(0.207)	(0.072)	(0.231)	(0.07)
Highest is A Level (0/1)	0.06	0.02	-.112	-.033
	(0.33)	(0.11)	(0.339)	(0.097)
Highest is a tertiary degree (0/1)			0.457	0.16
			(0.733)	(0.282)

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Continued...

Table 10b: **First-stage selection regression: Probit for sample retention against 2004 variables** (cont.)

	2005 (coefficients)	2005 (marginal effects)	2005 and 2006 (coefficients)	2005 and 2006 (marginal effects)
	(1)	(2)	(3)	(4)
Occupation characteristics:				
Log (no. employees) (=0 if wage)	0.049	0.017	-.023	-.007
	(0.18)	(0.062)	(0.2)	(0.062)
Log (firm size) (=0 for civil service)	-.142	-.049	-.066	-.020
	(0.078)*	(0.027)*	(0.083)	(0.026)
Civil service dummy (public sector, NGO, etc)	0.032	0.011	0.623	0.221
	(0.255)	(0.086)	(0.265)**	(0.102)**
Public enterprise dummy (state-owned firm)	1.118	0.26	0.874	0.32
	(0.433)***	(0.055)***	(0.421)**	(0.164)*
Private sector dummy	0.004	0.001	0.115	0.036
	(0.26)	(0.089)	(0.268)	(0.087)
Other:				
Region dummy (reference: Dar): Arusha	0.26	0.084	0.916	0.329
	(0.187)	(0.057)	(0.179)***	(0.069)***
Region dummy (reference: Dar): Iringa	0.288	0.092	-1.500	-.274
	(0.203)	(0.06)	(0.351)***	(0.028)***
Region dummy (reference: Dar): Morogoro	-.554	-.207	-1.225	-.243
	(0.19)***	(0.074)***	(0.3)***	(0.032)***
Region dummy (reference: Dar): Mwanza	0.162	0.054	-.103	-.031
	(0.19)	(0.061)	(0.19)	(0.055)
Region dummy (reference: Dar): Tanga	0.288	0.092	0.435	0.149
	(0.217)	(0.064)	(0.208)**	(0.077)*
Gender dummy (1 = male)	-.063	-.022	0.212	0.065
	(0.121)	(0.042)	(0.129)	(0.04)
Const.	-.697		-.800	
	(1.040)		(1.148)	
Obs.	565		569	
R^2	0.069		0.194	

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Table 11a: **Second-stage selection (Inverse Probability Weighting): Log earnings, 2004-2006**

	Pooled	Self-Emp.	Wage	Pooled (IPW)	Self-Emp. (IPW)	Wage (IPW)
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: ***Log earnings***						
Age and tenure:						
Age (years)	0.034	0.042	0.016	0.031	0.031	0.039
	(0.016)**	(0.02)**	(0.026)	(0.023)	(0.027)	(0.038)
$Age^2/100$	-.035	-.044	-.007	-.042	-.043	-.040
	(0.019)*	(0.023)*	(0.033)	(0.027)	(0.03)	(0.049)
Tenure (years)	0.007	0.0002	0.016	0.029	0.024	0.023
	(0.009)	(0.011)	(0.014)	(0.011)***	(0.015)	(0.016)
$Tenure^2/100$	-.006	-.002	-.020	-.033	-.035	-.018
	(0.023)	(0.031)	(0.037)	(0.029)	(0.041)	(0.045)
Education:						
Highest is primary certificate (0/1)	0.194	0.184	0.157	0.083	0.072	0.147
	(0.064)***	(0.081)**	(0.099)	(0.089)	(0.106)	(0.099)
Highest is middle school certificate (0/1)	0.148	-.160	0.404	0.29	-.005	0.562
	(0.144)	(0.175)	(0.223)*	(0.189)	(0.223)	(0.288)*
Highest is O Level (0/1)	0.612	0.616	0.58	0.574	0.596	0.577
	(0.078)***	(0.111)***	(0.098)***	(0.101)***	(0.128)***	(0.1)***
Highest is A Level (0/1)	0.778	0.264	0.907	0.732	0.274	0.894
	(0.106)***	(0.2)	(0.118)***	(0.133)***	(0.164)*	(0.168)***
Highest is a tertiary degree (0/1)	1.226	0.601	1.545	1.176	0.442	1.520
	(0.249)***	(0.448)	(0.283)***	(0.27)***	(0.468)	(0.34)***

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Continued...

Table 11b: **Second-stage selection (Inverse Probability Weighting): Log earnings, 2004-2006** (cont.)

	Pooled	Self-Emp.	Wage	Pooled (IPW)	Self-Emp. (IPW)	Wage (IPW)
	(1)	(2)	(3)	(4)	(5)	(6)
Occupation characteristics:						
Log (no. employees) (=0 if wage)	0.489 (0.078)***	0.492 (0.078)***		0.48 (0.091)***	0.481 (0.091)***	
Log (firm size) (=0 for civil service)	0.135 (0.025)***		0.129 (0.027)***	0.111 (0.026)***		0.116 (0.027)***
Civil service dummy (public sector, NGO, etc)	0.692 (0.086)***		0.754 (0.114)***	0.504 (0.11)***		0.6 (0.152)***
Public enterprise dummy (state-owned firm)	0.2 (0.126)		0.206 (0.102)**	0.237 (0.128)*		0.165 (0.133)
Private sector dummy	-.155 (0.083)*			-.076 (0.116)		
Other:						
Time dummy, 2005	0.14 (0.05)***	0.132 (0.063)**	0.164 (0.078)**	0.14 (0.051)***	0.137 (0.065)**	0.153 (0.078)*
Time dummy, 2006	0.289 (0.061)***	0.283 (0.083)***	0.31 (0.09)***	0.371 (0.07)***	0.389 (0.088)***	0.345 (0.106)***
Region dummy (reference: Dar): Arusha	-.224 (0.066)***	-.275 (0.083)***	-.039 (0.103)	-.161 (0.071)**	-.166 (0.089)*	-.047 (0.109)
Region dummy (reference: Dar): Iringa	-.238 (0.083)***	-.238 (0.098)**	-.178 (0.16)	-.138 (0.122)	-.022 (0.135)	-.355 (0.222)
Region dummy (reference: Dar): Morogoro	-.121 (0.08)	-.123 (0.109)	-.111 (0.101)	-.014 (0.11)	0.057 (0.136)	-.195 (0.162)
Region dummy (reference: Dar): Mwanza	-.260 (0.065)***	-.290 (0.089)***	-.162 (0.082)**	-.170 (0.087)**	-.134 (0.117)	-.168 (0.11)
Region dummy (reference: Dar): Tanga	-.055 (0.095)	-.008 (0.125)	-.057 (0.155)	-.013 (0.095)	0.015 (0.124)	-.009 (0.145)
Gender dummy (1 = male)	0.315 (0.046)***	0.425 (0.057)***	0.13 (0.074)*	0.296 (0.059)***	0.439 (0.07)***	0.059 (0.091)
Const.	9.443 (0.298)***	9.339 (0.383)***	9.544 (0.46)***	9.527 (0.434)***	9.511 (0.543)***	9.201 (0.655)***
Obs.	1189	780	409	1189	780	409
R^2	0.363	0.257	0.47	0.369	0.277	0.458
Log-likelihood	-1332.076	-910.994	-387.050	-1296.445	-884.885	-376.776

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Recent panel-data research on income dynamics indeed provides some support for these notions. **Carter and May (2001)**, for example, used poverty transition matrices to explore income movements between 1993 and 1998 in data from the KwaZulu-Natal Income Dynamics Study ('KIDS'); they estimated that a significant number of respondents were caught in a 'structural poverty trap' from which escape would be difficult. **Dreze, Lanjouw and Stern (1992)** draw similar conclusions using a four-period panel over 26 years in Palanpur, in the Indian state of Uttar Pradesh.

However, not all of the research has pointed in this direction. Indeed, the general trend in the literature — at least when considering income earners — has been to find *progressive* gains; that is, lower gains among those with higher initial wealth, with the traditional determinants of earnings levels having little explanatory value for earnings growth. **Gunning, Hoddinott, Kinsey and Owens (2000)** used a panel of households resettled on formerly white-owned farms following Zimbabwean independence — the respondents were first interviewed in 1983-84, were re-interviewed in 1987 and were then interviewed annually from 1992 to 1998. Using a regression approach, the authors found income growth to be a progressive process, shared across all households: "the largest percentage increases in predicted incomes [were] recorded by households that had the lowest predicted incomes at the beginning of the survey" (p.151). Similarly, **Fields, Cichello, Freije, Menéndez and Newhouse (2003*a*)** compared income dynamics using panel data from Indonesia, Spain, Venezuela and South Africa (the KIDS data, again); they found that all four countries experienced progressive growth, and that the finding was robust to "reasonable amounts of measurement error" in South Africa and Venezuela. Subsequent work by the same authors on the data found, in all four countries, that "initial income and job changes of the head are consistently the most important variables in accounting for household per capita income changes" **(Fields, Cichello, Freije, Menéndez and Newhouse 2003*b*, 31)**. Most recently, **Cichello et al. (2005)** found significantly progressive gains in the KIDS data — to the extent that, despite large general gains among labour force participants, the highest quintile of 1993 earners and those originally in the formal sector were found to have experienced zero or negative growth. Specifically, the authors highlighted the importance of sectoral movement and initial income, rather than other demographic variables, as key determinants of growth:

> We found that sector change is the most important variable and initial earnings is a close second. Together, these two variables account for nearly all of the explained variation in earnings changes. The remaining variables — most importantly, the worker's education and gender, but also other demographic and industry variables — explain virtually nothing about earnings change.
>
> (Cichello et al. 2005, 182)

The causes for such progressive growth remain matters of conjecture — certainly, neither the existing literature nor the present data can support any sweeping claims about the general nature of income growth in developing economies. However, at least one plausible explanation does emerge: that progressive income growth is the result of strong growth in the informal sector — 'unskill-biased technological change', as it were — particularly in the context of a diminishing role for the public sector. Thus, in the KIDS data, Cichello et al. (2005, 143) find that '[t]he dynamism of the informal sector over this period is shown to be an important contributor to the progressive growth in earnings'. **Calvès and Schoumaker (2004, 1343)** make a similar point in documenting substantial trends towards informality in a survey conducted in the year 2000 in Burkina Faso: "[i]n a context where recruitment in the public sector, a traditionally preferred employment location for new graduates, has considerably slowed or completely stopped, diplomas are no longer an automatic passport to secure jobs nor a protection against unemployment".

Intuitively, one might expect Tanzania to be experiencing a similar dynamic. Indeed, the former president's polemic itself reflected on the traditional link between education and formal employment:

> . . . a few people go to university. . . their idea of service is related to status and the salary which a university education is expected to confer upon its recipient. The salary and the status have become a right automatically conferred by the degree.
>
> (Nyerere 1968)

To the extent that recent Tanzanian experience reflects a weakening of the role of the public sector as employer, one might expect similarly progressive dynamics. This point has been made by Söderbom et al. (2006, 285) in summarising their findings from repeated cross-sections in Tanzania and Kenya:

> Knight and Sabot (1990). . . argue that the high returns in Kenya relative to Tanzania reflected a willingness to allow market processes to work in Kenya relative to Tanzania. Over the 1990s Tanzanian policies have become much more similar to those of Kenya, and we have shown that by the end of the 1990s the earnings profiles were quite similar in the two countries.

Thus, the literature poses several important questions about the recent growth process in Tanzania. What have been the main determinants of individual-level income dynamics? Has Tanzanian income growth — like that of several other countries — been progressive in its distribution? What role have sectoral differences played? These are the questions that we consider now.

4.1 Tanzanian growth: A first look

We begin our growth analysis with some descriptive statistics. **Table 12** shows the basic summary statistics for the growth process; income growth is defined, as throughout this paper, as referring to the change in log earnings: $\Delta(\text{log earnings})$. The table — along with **Figures 3 and 4**, showing the distribution of income and income growth — suggest that the survey reflects a wide variety of income growth experiences.

Table 12: **Growth in log earnings by year ('000 Tsh per month, deflated)**

	Obs.	Median	Mean	S.Dev.	Min.	Max.
One-year growth…						
Pooled	637	0.113	0.171	0.753	-2.502	2.994
2004 → 2005	398	0.012	0.133	0.792	-2.344	2.994
2005 → 2006	239	0.157	0.235	0.681	-2.502	2.188
Two-year growth…						
2004 → 2006	225	0.403	0.439	0.852	-2.241	2.957

Figure 3: **Income over time**

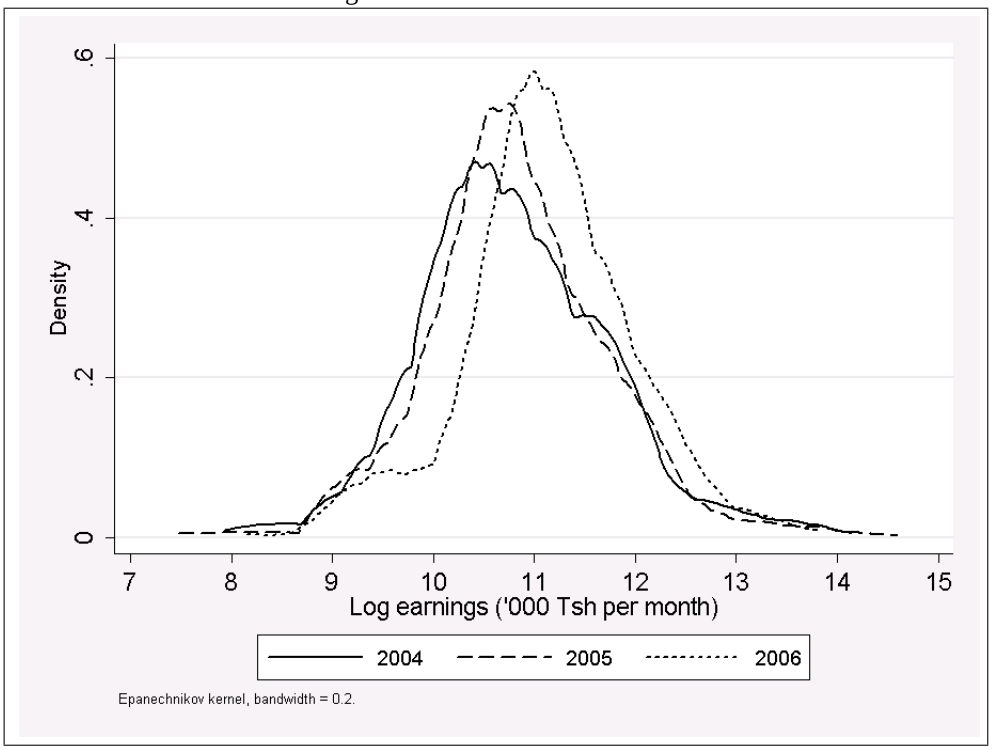
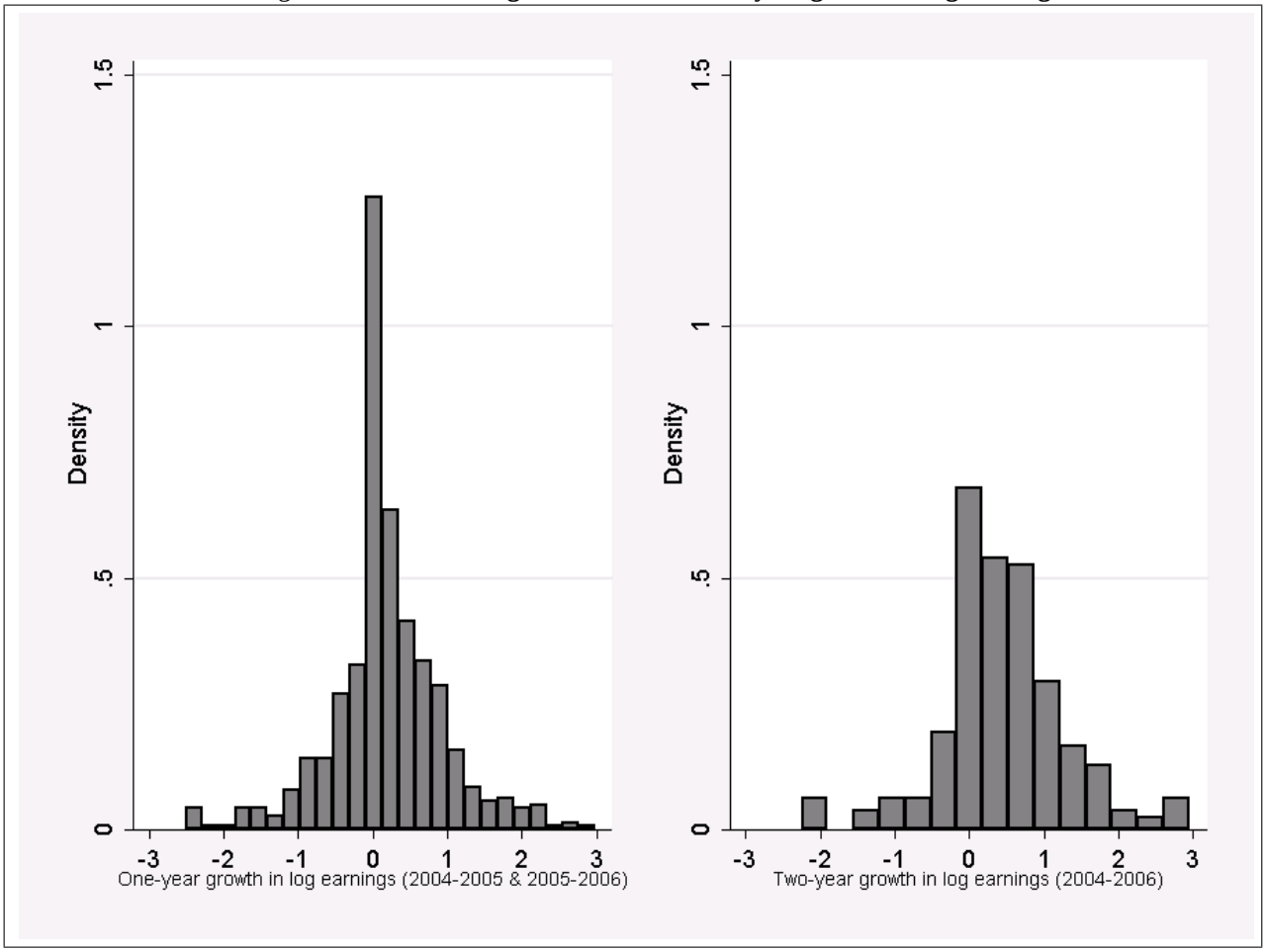


Figure 4: **Growth histograms: One- and two-year growth in log earnings**



These descriptive statistics and graphs suggest more than merely a wide variety of growth experiences; they also suggest that income growth has been higher from 2005 to 2006 than from 2004 to 2005. **Figure 5** shows the sample cumulative densities for each year of the survey; 2006 indeed lies clearly to the right of the earlier years. Further, it appears that the growth has been most pronounced (certainly, most pronounced in relative terms) for workers with relatively low initial incomes. Both these suggestions deserve further and more formal attention shortly.[3]

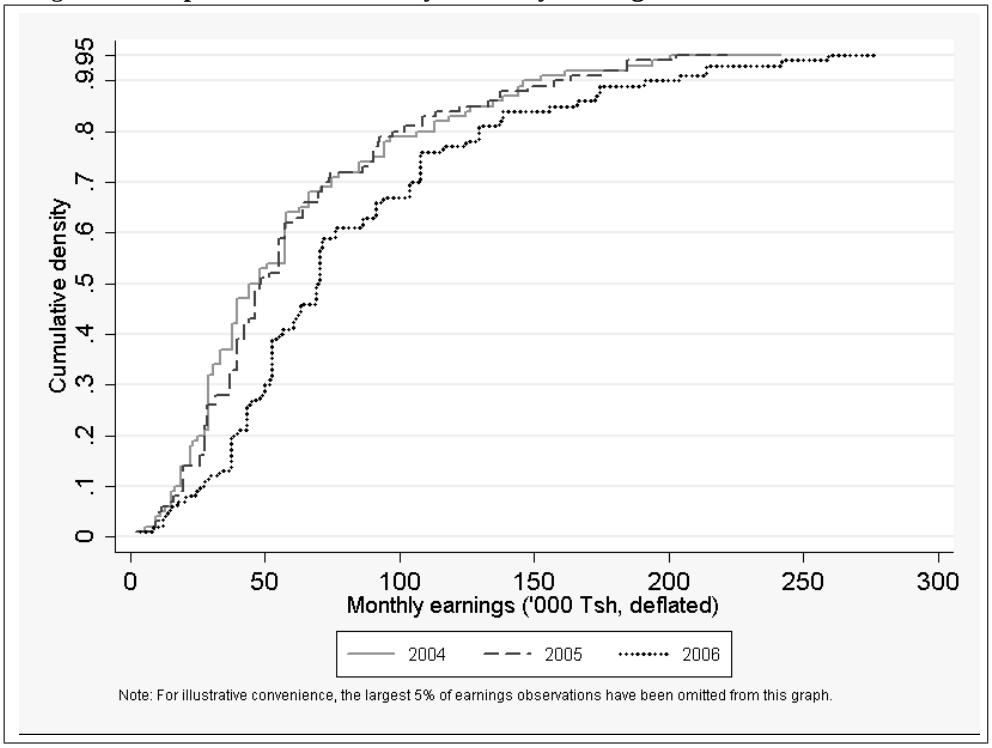
This prompts immediately a consideration of heterogeneity in earnings growth experiences; in particular, concern about differences in growth between employment sectors. **Table 13** suggests a substantial difference between self-employed workers and wage-earners (the former appear to have experienced higher earnings growth on average, but with a much greater variability), whereas there appear to have been minimal differences appear between the sub-categories of wage-earners. **Figures 6 and 7** illustrate.

Table 13: **One-year growth in log earnings by occupation ('000 Tsh per month, deflated)**

	Obs.	**Median**	**Mean**	**S.Dev.**	**Min.**	**Max.**
Pooled	637	0.113	0.171	0.753	-2.502	2.994
Self-employed	428	0.181	0.207	0.808	-2.344	2.589
Wage-earners	209	0.014	0.098	0.620	-2.502	2.995
Civil	60	0.015	0.017	0.831	-2.502	2.995
Public	49	-0.018	-0.026	0.420	-2.344	1.032
Private	100	0.074	0.208	0.535	-1.140	2.627

[3] Among other issues, we must be concerned whether these phenomena are merely reflections of the attrition pattern identified earlier. We do not pursue that issue directly here; rather, we correct for selection formally in a regression framework shortly.

Figure 5: **Sample cumulative density: Monthly earnings in 2004, 2005 and 2006**



What of other demographic factors? **Figure 8** compares mean earnings growth — both one-year and two-year — against the quartile of 2004 earnings; **Figure 9** goes further, to show literally the relationship between earnings growth and initial earnings. Intriguingly, both graphs suggest that *higher* percentage income growth accrued to those with *lower* initial earnings. Of course, one would expect a negative bias in dynamic models of this kind (see **Nickell (1981)**); we cannot, therefore, interpret either graph as providing persuasive evidence of progressive growth.[4] Nonetheless, the graphs are suggestive of a progressive relationship; at the least, they do not rule it out.

Finally, **Figure 10** shows mean earnings growth against highest educational attainment. In many respects, this is the most interesting graph of all. Intuitively, one would expect — as discussed earlier — that individuals with higher education enjoy both higher earnings levels *and* higher earnings *growth*. Figure 10 shows that — at least in terms of mean growth — this is not the case. Indeed, there is some evidence that individuals with higher education have enjoyed *lower* or even *negative* earnings growth (depending, of course, upon the weight that one intuitively attaches to the small number of individuals with tertiary degrees).

The suggestion from the descriptives, then, is that the growth experience has been markedly heterogeneous — between wage-earners and the self-employed, between individuals with different initial incomes, and between workers with different educational backgrounds. The extent to which this is so — and the extent to which such demographic factors can be taken to have explained earnings growth — is fundamental to understanding Tanzania's recent growth experience. To explore these issues further, we need a formal identification framework — and one that builds coherently upon the levels identification used earlier.

[4] We *could*, of course, explore such issues using a variant of the instrumenting methodology of Anderson and Hsiao (1981). We decline to pursue this in the present paper, primarily because of the relatively small number of relevant observations (*i.e.* only 168 individuals were observed in all three periods).

Figure 6: **Growth kernel densities: Wage-earners and self-employed**

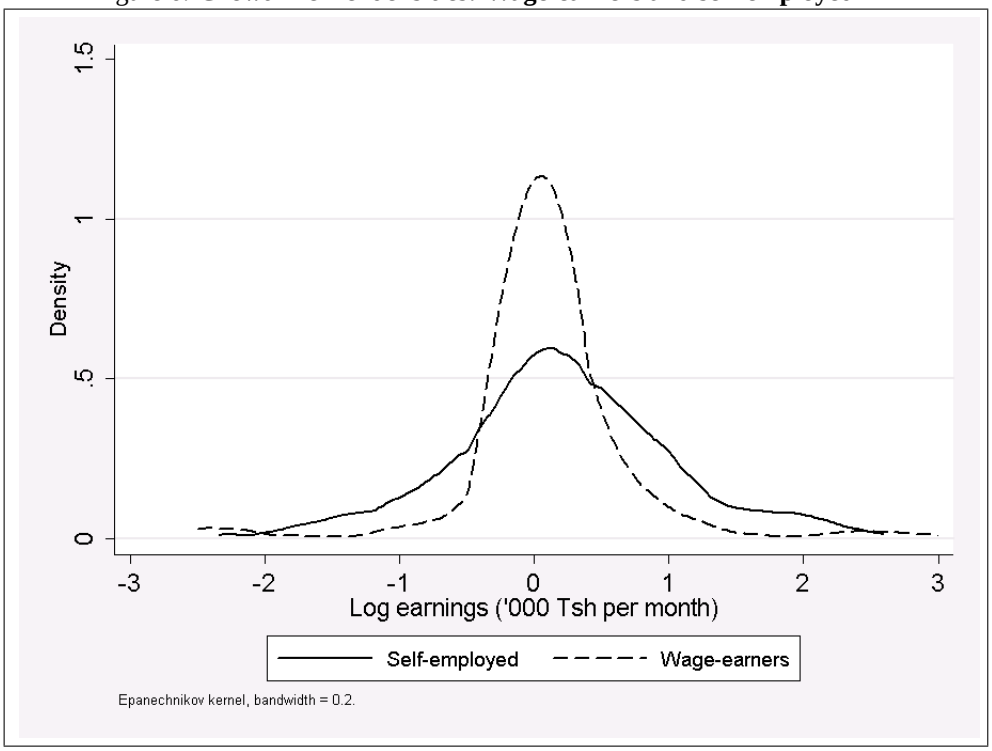


Figure 7: **Growth kernel densities: Categories of wage-earner**

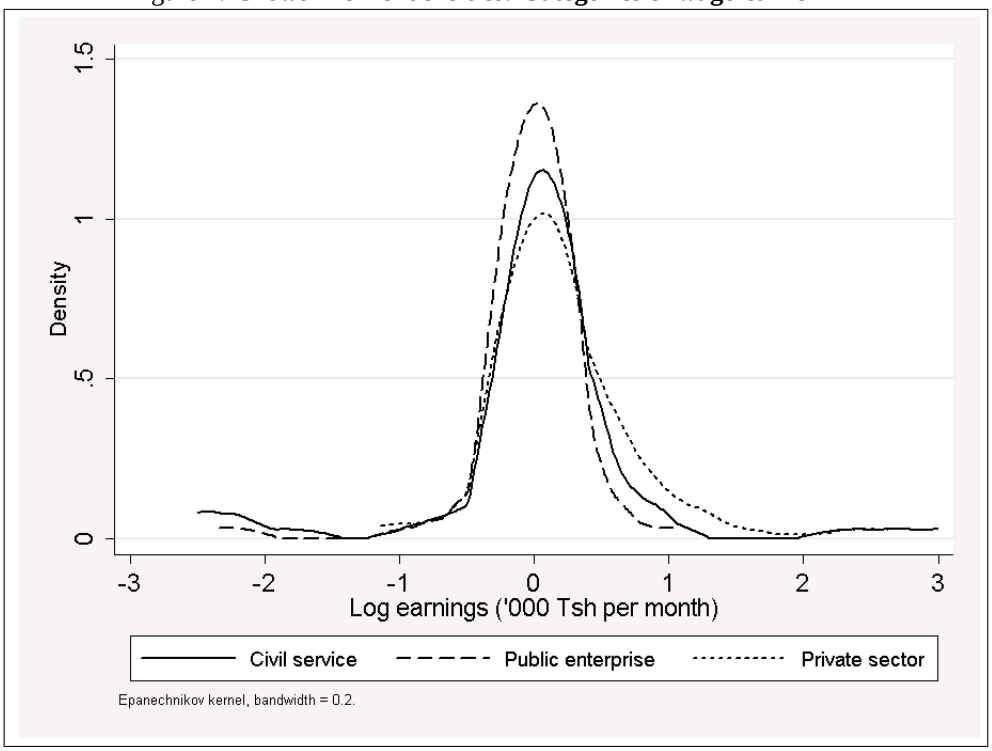


Figure 8: **Earnings growth by initial earnings**

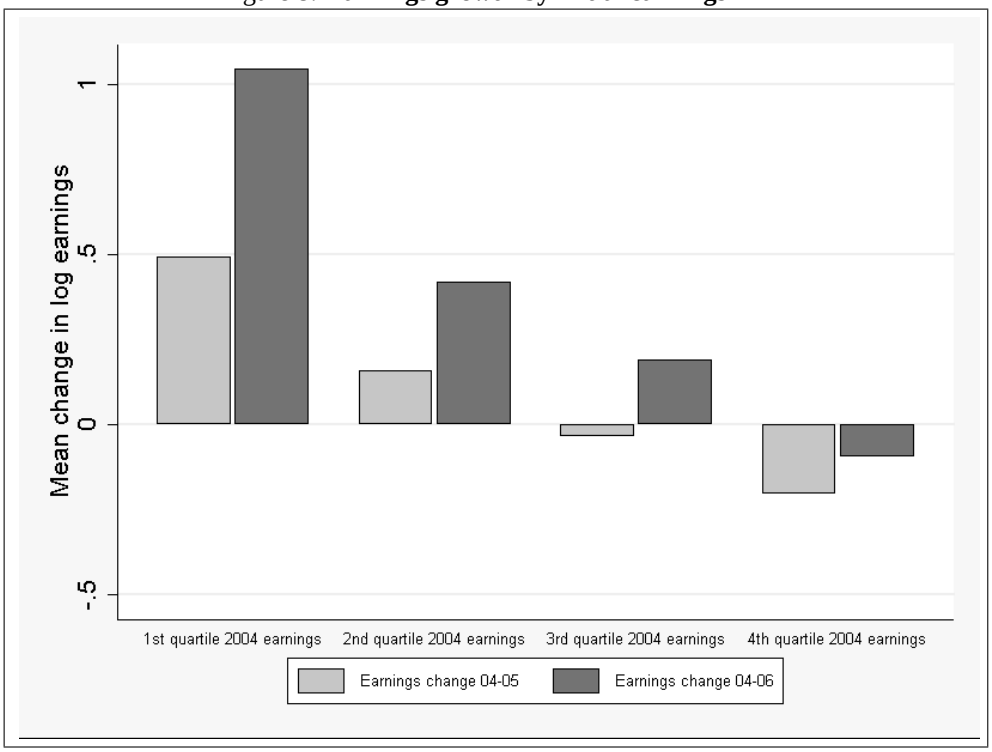


Figure 9: **Earnings growth by initial earnings**

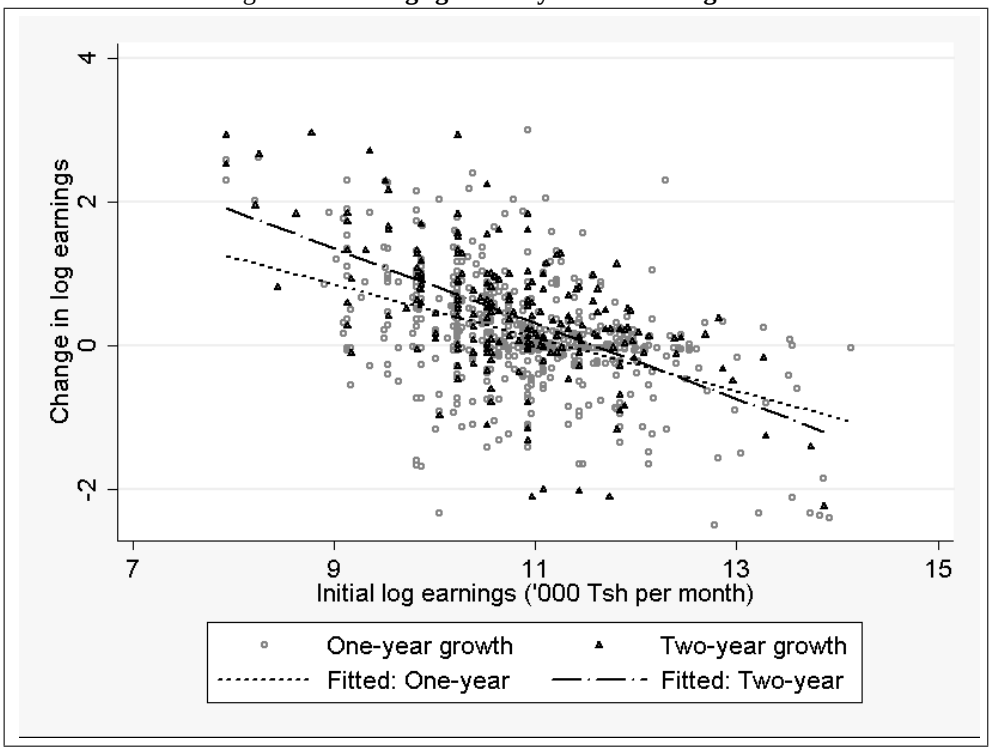
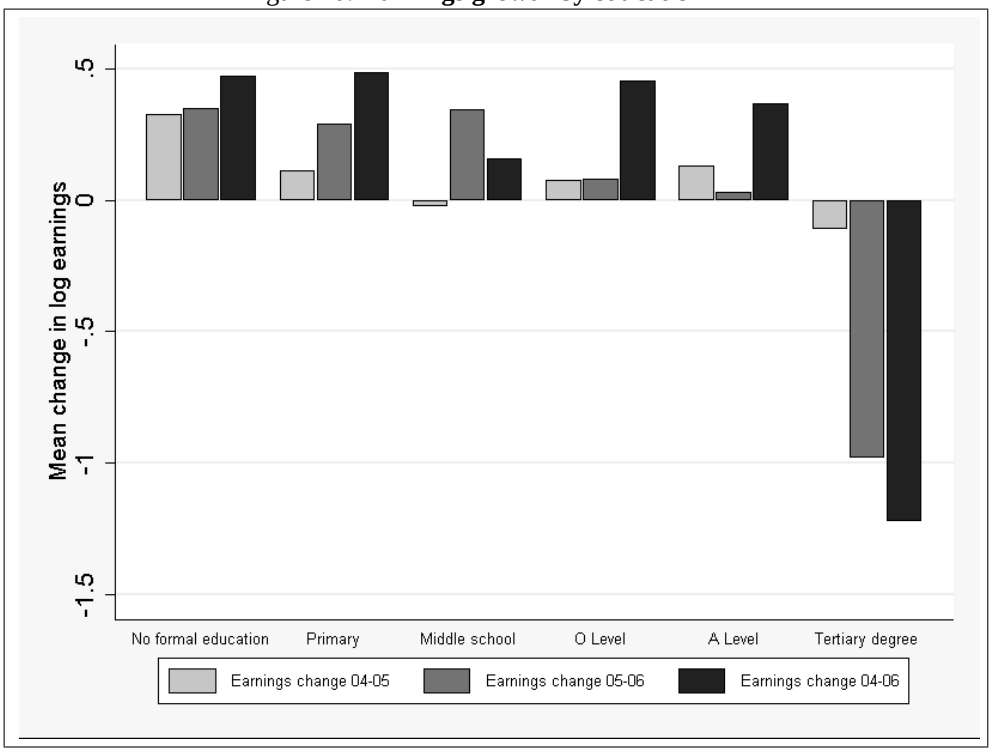


Figure 10: **Earnings growth by education**



4.2 Growth identification: Extending the Mincerian framework

We earlier identified the determinants of earnings levels by a standard Mincerian semi-log relationship. We now seek to extend that framework to identify the determinants of earnings growth. Importantly, we must do so in a manner that is consistent with the levels identification — so that a time-specific 'snapshot' of the earnings process produces a levels equation of the same structure used earlier.

Levels identification was achieved by the following relationship:

$$y_{it} = \alpha_0 + \alpha_1 \mathbf{x}_{1it} + \alpha_2 \mathbf{x}_{2i} + u_i + \epsilon_{it} \tag{1}$$

For simplicity and generality, we now specify an analogous relationship for the rate of growth in log earnings (that is, for the percentage growth in earnings), where — without loss of generality — the time-variant and time-invariant determinants are again denoted by the vectors $\mathbf{x}_{1it}$ and $\mathbf{x}_{2i}$:

$$\frac{\partial y_{it}}{\partial t} = \beta_0 + \beta_1 \mathbf{x}_{1it} + \beta_2 \mathbf{x}_{2i} + v_i + \eta_{it}. \tag{3}$$

These two equations may be combined in a consistent framework by integrating equation (3) by time and treating equation (3) as determinative of an employee's 'entry income' (whether that be 'entry' to an occupation or to the workforce as a whole): E_i. Denoting s_{it} as the duration of an employee's stay in an occupation (or, alternatively, in the workforce as a whole), we have:

$$\begin{aligned}
\frac{\partial y_{it}}{\partial t} &= \beta_0 + \beta_1 \mathbf{x}_{1it} + \beta_2 \mathbf{x}_{2i} + v_i + \eta_{it} && (3)\\
\therefore y_{it} &= E_i + \int_0^{s_{it}} \frac{\partial y_{it}}{\partial t} \; dt \\
&= E_i + (\beta_0 + \beta_2 \mathbf{x}_{2i} + v_i) \times s_{it} + \int_0^{s_{it}} (\beta_1 \mathbf{x}_{1it} + \eta_{it}) \; dt \\
&= \alpha_0 + \alpha_1 \mathbf{x}_{1it} + \alpha_2 \mathbf{x}_{2i} + u_i + \epsilon_{it} + (\beta_0 + \beta_2 \mathbf{x}_{2i} + v_i) \times s_{it} + \int_0^{s_{it}} (\beta_1 \mathbf{x}_{1it} + \eta_{it}) \; dt \\
&= \alpha_0 + \beta_0 s_{it} + \alpha_1 \mathbf{x}_{1it} + (\alpha_2 + \beta_2 s_{it}) \mathbf{x}_{2i} + u_i + \epsilon_{it} + v_i s_{it} + \int_0^{s_{it}} (\beta_1 \mathbf{x}_{1it} + \eta_{it}) \; dt. && (4)
\end{aligned}$$

At this level of generality, this relationship cannot be used for identification — identification of the remaining integral term depends upon identification of the history both of $\mathbf{x}_{it}$ and η_{it}. This makes intuitive sense: if an individual's

earnings *growth* is allowed to vary over time, the consequent earnings *level* will depend upon the shape of the growth path. This demands too much of the available data — which covers only three periods — and, as subsequent analysis will show, is not necessary to explore the determinants of growth. We therefore simplify the identification by imposing that an indivdiual's earnings growth is constant over time: $\beta_1 = \mathbf{0}$, $\eta_{it} = 0$.

Equation (4) then becomes:

$$y_{it} = \alpha_0 + \beta_0 s_{it} + \alpha_1 \mathbf{x}_{1it} + (\alpha_2 + \beta_2 s_{it})\mathbf{x}_{2i} + u_i + \epsilon_{it} + v_i s_{it}. \tag{5}$$

This equation, then, shows how the Mincerian levels equation may be combined with a growth relationship. Importantly, it shows that incorporating income growth does not change the underlying structure of the Mincerian levels relationship. In effect, it shows that we can reinterpret the constant term and the coefficients on time-invariant regressors as being the sum of a levels effect and an integrated growth effect. We *could* use this equation for identification, if some measure of s_{it} were defined — for example, we could use a worker's tenure, or experience. However, this produces unnecessary ambiguity — after all, the variable s_{it} was introduced as a mathematical concept rather than a specific concept — and raises unnecessary questions about the role of the fixed effect u_i.

More appropriate is the first difference of **Equation (5)**, where it is assumed that $\Delta s_{it} = 1$, and we denote throughout that $\Delta X_{it} \equiv X_{i,t+1} - X_{it}$:

$$\Delta y_{it} = \beta_0 + \alpha_1 \Delta \mathbf{x}_{1it} + \beta_2 \mathbf{x}_{2i} + \Delta\epsilon_{it} + v_i. \tag{6}$$

This equation, then, is used to identify the determinants of earnings growth. Our exercise in combining the levels equation with a growth equation shows an important point: the growth regression must include the first difference of the time-variant determinants of the earning level (*i.e.* $\Delta\mathbf{x}_{1it}$); otherwise, if $\mathbb{E}(\mathbf{x}_{2i}\Delta\mathbf{x}_{1it}) \neq 0$, the estimate of β_2 will be biased and inconsistent. Of course, for growth from period t to $t+1$, there is nothing to prevent us from augmenting the vector $\mathbf{x}_{2i}$ with time-variant measures taken at period t; this still remains essentially consistent with the identification framework laid out. Indeed, we do this shortly when considering including occupation characteristics (rather than merely the *change* in such characteristics) and age in our growth regression.

The estimation of **Equation 6** in the present semi-parametric framework requires the further imposition of moment conditions. The previous levels estimation depended for its validity — both in the OLS and the IV case — upon the assumption that the explanatory variables are linearly independent of the time-specific unobservable:

$$\mathbb{E}(\mathbf{x}_{1it}\epsilon_{it}) = \mathbf{0};$$
$$\mathbb{E}(\mathbf{x}_{2i}\epsilon_{it}) = \mathbf{0}.$$

From that maintained assumption, it follows trivially that $\mathbb{E}(\mathbf{x}_{2i}\Delta\epsilon_{it}) = \mathbf{0}$. However, the identification of **Equation 6** requires the imposition of several further assumptions. First, we must assume that the time-variant explanatory variables in period $t-1$ are linearly independent of the time-variant unobservable in period $t+1$, and *vice versa*. In effect, we need to assume strict exogeneity of the time-variable unobservable and the time-variant explanatory variables:

$$\mathbb{E}(\mathbf{x}_{1it}\epsilon_{i,t-1}) = \mathbb{E}(\mathbf{x}_{1i,t-1}\epsilon_{it}) = \mathbf{0}.$$

Finally, we must assume that both the level of the time-invariant explanatory variables and the change in the time-variant explanatory variables are linearly independent of the growth fixed-effect:

$$\mathbb{E}(\Delta\mathbf{x}_{1it}v_i) = \mathbf{0};$$
$$\mathbb{E}(\mathbf{x}_{2i}v_i) = \mathbf{0}.$$

The former restriction is nothing more than an implicit assumption made every time one differences a levels equation to eliminate an additive fixed effect. The latter restriction is stricter; it amounts, for example, to a requirement that an individual's choice of education is not endogenous to an unobservable fixed effect upon the individual's *growth* of earnings.

From these assumptions, it follows that:

$$\mathbb{E}(\Delta\mathbf{x}_{1it}(\Delta\epsilon_{it} + v_i)) = \mathbf{0} \text{ and}$$
$$\mathbb{E}(\mathbf{x}_{2i}(\epsilon_{it} + v_i)) = \mathbf{0},$$

so that OLS is the appropriate estimator. We proceed, then, to estimate the determinants of growth, using the same set of explanatory variables used earlier.[5] Specifically, we allow education, occupation and region (as well as age) to have growth effects, and include changes in occupation characteristics as the vector $\Delta\mathbf{x}_{1it}$, as just explained.

[5] Two subtle differences must be noted, both relating to the age-tenure relationship. First, we do not attempt, in the growth relationship, to identify separately the age and tenure effects; they are separately identified only by workers changing jobs, which — as **Table 4** suggested — does not often occur. Thus, we drop the tenure measure. Second, since the age measure enters with a quadratic form in the level regression, it is appropriate that the quadratic term is dropped in the growth equation, *i.e.* $\frac{\partial}{\partial t}(A \cdot \text{age}_{it} + B \cdot \text{age}_{it}^2) = A + 2B \cdot \text{age}_{it}$.

4.3 Growth estimation: OLS

> "Is there any point to which you would wish to draw my attention?"
> "To the curious incident of the dog in the night-time."
> "The dog did nothing in the night-time."
> "That was the curious incident," remarked Sherlock Holmes.
>
> *The Memoirs of Sherlock Holmes*, Doyle (2000 [1894])

Tables 14a and 14b report the OLS regression identified by equation (6), with the one-year change in log earnings as the dependent variable. As earlier, regression (1) is run on the pooled sample, with regressions (2) and (3) performed on the self-employed and wage-earners respectively; the set of wage-earners is then further subdivided into the civil service (regression (4)), public enterprises (regression (5)) and the private sector (regression (6)).[6]

Several significant coefficients deserve consideration. First, there appears to be significant **regional variation** in growth rates, expressed particularly through the self-employed; after correcting for the other regressors, Arusha and Mwanza appear to have enjoyed significantly higher income growth than Dar es Salaam, whose growth was itself significantly higher than that of Iringa. Similarly, **change in occupation characteristics** was significant in some respects: growth in the number of employees was substantial and significant for explaining income growth among the self-employed; this reflects the significant levels effect between the variables shown earlier. Interestingly, though, growth in firm size did not significantly explain income growth among wage-earners, despite also having a significant levels effect; even the point estimate was an order of magnitude less than in the levels regression.

This raises as a possibility that that the significant levels effect of firm size may be driven by levels fixed effects, but not the significant effect of the number of employees. That is, it suggests that the significant size effect among wage-earners may be driven by labour-market 'sorting' (whereby workers with higher ability are matched to larger enterprises), but that the same is not true of the significant size effect among the self-employed. Though necessarily a tenuous suggestion, the comparison between the levels results and the growth results across wage-earners and the self-employed suggests the possibility of quite fundamental differences in the behaviour of the two sectors; moreover, the significant size coefficient for the self-employed again reinforces the importance of heterogeneity *within* sectors, as well as between them. This clearly raises intriguing questions for further research. Finally, we note that some sectoral movements are also significant: workers shifting from self-employment to wage-earning received significantly higher incomes (identified by 25 observations), while the opposite was true for public servants shifting into self-employment (identified by two observations).

At least as interesting, though, are the dogs that do not bark: the coefficients that, despite common intuition, do *not* show a significant positive growth effect. This is true of all of the regressors shown in Table 14a; each set deserves consideration in turn. First, it is notable that **age** does not significantly explain growth. This is an important insight into understanding the levels results on age and tenure. Earlier, we found a significant linear tenure-earnings effect, and a significant *concave* age-earnings effect. This is notable in its own right, as discussed earlier. However, as explained earlier, the levels result leaves open the question of whether the concave age-earnings profile observed is a result of (i) *given* individuals following a concave earnings path as they age, and/or (ii) *different* individuals entering and leaving the labour force in such a way that a concave relationship is observed. Examining the age-growth relationship sheds light on this distinction: for if *given* individuals follow a concave path, one would expect age to have a significant negative effect upon earnings. This is not the case in the regression results. Of course — as with all of the non-results considered here — this could be a consequence of the regression having inadequate power to reject; it is relevant that age *does* have a negative point estimate for wage-earners (regressions (3)–(6)). However, the failure to find a significant negative effect nonetheless suggests a preference for the latter explanation: that the concave age-earnings relationship is a consequence of participation dynamics, rather than a concave individual-earnings path.

Second, it is important that **education** does not significantly explain earnings growth (except to the extent that self-employed tertiary-qualified employees apparently experienced *lower* earnings growth — though this is identified by six observations). As discussed, one might intuitively expect higher education to have a significant positive effect upon earnings levels *and* earnings growth; however, the data simply do not support this. One might wonder whether this non-result is simply an artifact of cutting the education measure so finely; when we reduce the education variable simply to a measure of whether an individual has or has not completed A Level (omitted here for brevity), we still find no significant positive education effect (indeed, the point estimate for the effect of having completed A Level is negative for growth regressions across the whole sample, across the self-employed and across wage-earners).

This is an important result for understanding the relationship between education and earnings in Tanzania. It is surprising given the common intuition that education has both levels and growth effects on income. It is all the more

[6] For growth from period t to period $t+1$, these are the employment categories occupied at period t.

perplexing given the earlier finding that education has a convex levels effect — a positive education-growth relationship is surely the readiest explanation for that shape. Two primary possibilities emerge.[7] First, it could be that education *has* previously had a positive growth effect in Tanzania — thus explaining the convex levels relationship observed — but that the recent experience observed has been far more progressive. Thus, the data may speak to an 'unskill-biased technological change' in Tanzania over recent years, with income growth higher in less formal enterprises with less educated employees. This would place Tanzania alongside other developing countries — discussed earlier — for which growth has been found to be progressive and biased towards informal employment. It is intriguing, in light of this possibility, that education does not have a significant positive relationship to earnings in *any* of the sectors considered, and that the occupational dummies themselves do show any significant sectoral growth differences. Nonetheless, the possibility deserves consideration. Second, however, it may be that — as we just suggested for the case of the age-earnings relationship — the shape of the education-earnings relationship is driven by participation decisions, rather than a convex earnings path at the individual level. Thus, it may be that higher education does not cause higher growth — *and has never done so* — but that more highly-educated individuals have a higher reservation wage; education could therefore be observed to have a convex 'effect' on earnings levels, but no observable growth effect. Importantly, this explanation would be consistent with different sectors having broadly similar experiences, as discussed. For now, we are content merely to note these possibilities, without further exploration. Nonetheless, it bears noting that the Tanzania Household Worker Survey does include (i) information on labour force non-participants, and (ii) a self-reported reservation wage. Future research might, at the least, draw profitably upon both of these sources.

Similar issues arise when we try to understand why neither firm size (for wage-earners) nor the number of employees (for the self-employed) significantly explain earnings growth, despite each having a significant positive levels effect. At the least, this may — again — be explained by (i) inadequate power of the test, (ii) firm characteristics not having, and never having had, significant growth effects, (iii) recent experience differing markedly from the past, and/or (iv) the employment participation decision. As explained earlier, the relationship between occupation characteristics and income is not a focus of this paper; we again therefore merely note the intriguing results and leave further exploration for future research.

Finally, note that — despite the suggestion in the descriptives that growth from 2005 to 2006 was significantly higher than that from 2004 to 2005, this does not appear to be the case once other regressors are controlled for. The suggestion, then, is that the result in the descriptives may have been driven by selection issues (in particular, over-sampling from more prosperous or more rapidly-growing regions), and/or by concomitant growth in other determinants (for example, the number of employees). Regressions using two-year growth as the dependent variable (omitted here for brevity) show the same broad patterns (though the number of observations is necessarily lower: the total pooled two-year growth relationship relates only to 169 individuals).

4.4 Growth estimation: Selection correction

Finally, we address one potential endogeneity issue in the context of the growth relationship: that of panel attrition. The IPW methodology used has already been explained, and it was noted that the methodology is sufficiently general to cover the present case. The first-stage regression remains that reported in **Tables 10a and 10b**. **Tables 15a and 15b** report the second-stage. Regressions (1)–(3) report the basic OLS regression, using the subset of individuals observed in 2004 (that is, the subset of individuals for whom the IPW methodology may be applied); regressions (4)–(6) repeat the same specification on the same subset, with inverse probability weighting.[8]

The weighting changes little. The weighted regression produces a significant age coefficient in the pooled regression; however, the coefficient is *positive*: if anything, this strengthens our earlier sense that the concavity of the age-earnings profile is driven by the participation choice. Under weighting, we no longer have a significant relationship between growth in the number of employees and growth in income; further, the point estimate on employee growth is substantially reduced. The suggestion, then, is that the significant levels effects for *both* number of employees and firm size may relate to the worker fixed effect, rather than a causal relationship; in effect, the levels relationship may be explicable wholly by workers 'sorting' into firms on the basis of their ability. This is an important distinct issue (for example, see Fafchamps, Söderbom and Benhassine (2006)), and one that deserves further attention in respect of the present data.

Importantly, the weighting does *not* change the earlier result that education has no positive impact upon income growth; the issues and suggestions arising from this finding cannot, therefore, be discarded merely as artifacts of sample attrition.

[7] Aside, of course, from the low-power explanation, already acknowledged.

[8] For growth between period t and period $t+1$, the probability used is the predicted probability of retention for the *later* period: $\hat{p}_{i,t+1}$.

Table 14a: **Determinants of one-year growth in log earnings, 2004-2005 & 2005-2006**

	Pooled	Self-Employed	Wage	Civil	Public	Private
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: ***One-year change in log earnings***						
Age and tenure:						
Age (years)	0.001	0.0009	-.003	-.016	-.009	-.005
	(0.003)	(0.004)	(0.005)	(0.015)	(0.009)	(0.006)
Education:						
Highest is primary certificate (0/1)	-.073	-.133	0.096	0.205	-.099	0.087
	(0.089)	(0.107)	(0.164)	(0.961)	(0.317)	(0.159)
Highest is middle school certificate (0/1)	-.262	-.285	-.116	-.573	0.238	-.150
	(0.165)	(0.214)	(0.259)	(1.095)	(0.308)	(0.379)
Highest is O Level (0/1)	-.140	-.224	0.054	0.006	-.045	0.035
	(0.106)	(0.137)	(0.167)	(0.948)	(0.241)	(0.185)
Highest is A Level (0/1)	-.177	-.164	0.031	0.01	-.334	0.188
	(0.162)	(0.292)	(0.203)	(0.989)	(0.36)	(0.278)
Highest is a tertiary degree (0/1)	-.642	-1.269	-.306		-.407	
	(0.319)**	(0.556)**	(0.378)		(0.321)	
Occupation characteristics:						
Log (no. employees) (=0 if wage)	0.022	0.066				
	(0.113)	(0.123)				
Log (firm size) (=0 for civil service)	0.002		-.006		0.088	-.028
	(0.045)		(0.04)		(0.061)	(0.04)
Civil service dummy (public sector, NGO, etc)	-.175		-.142			
	(0.113)		(0.162)			
Public enterprise dummy (state-owned firm)	-.131		-.083			
	(0.211)		(0.137)			
Private sector dummy	-.055					
	(0.148)					

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Continued...

Table 14b: **Determinants of one-year growth in log earnings, 2004-2005 & 2005-2006 (cont.)**

	Pooled	Self-Employed	Wage	Civil	Public	Private
	(1)	(2)	(3)	(4)	(5)	(6)
Change in occupation characteristics:						
Δ Log(employees) ($\equiv 0$ for sector movement)	0.322 (0.1)***	0.346 (0.107)***				
Δ Log(firm size) ($\equiv 0$ for sector movement)	0.013 (0.042)		0.017 (0.038)		0.046 (0.041)	0.057 (0.045)
Within-wage sector movement	0.174 (0.371)		0.023 (0.328)	0.258 (0.979)	-.188 (0.357)	0.201 (0.412)
Wage-earner → self-employed	0.053 (0.206)		-.048 (0.184)		-1.332 (0.304)***	0.171 (0.179)
Self-employed → wage-earner	0.205 (0.154)	0.7 (0.301)**				
Other:						
Region dummy (reference: Dar): Arusha	0.245 (0.082)***	0.41 (0.109)***	0.01 (0.121)	0.349 (0.385)	-.310 (0.185)*	0.066 (0.157)
Region dummy (reference: Dar): Iringa	-.383 (0.112)***	-.358 (0.134)***	-.295 (0.213)	-.349 (0.54)	0.072 (0.297)	-.515 (0.299)*
Region dummy (reference: Dar): Morogoro	0.018 (0.135)	0.091 (0.171)	-.183 (0.212)	0.435 (0.513)	-.381 (0.216)*	-.167 (0.396)
Region dummy (reference: Dar): Mwanza	0.343 (0.095)***	0.535 (0.122)***	-.006 (0.144)	0.052 (0.418)	-.176 (0.25)	0.034 (0.183)
Region dummy (reference: Dar): Tanga	0.179 (0.096)*	0.131 (0.125)	0.26 (0.144)*	0.019 (0.432)	-.365 (0.255)	0.42 (0.162)***
Time dummy, 2005-2006 (reference: 2004-2005)	-.027 (0.066)	0.002 (0.085)	-.068 (0.096)	-.322 (0.252)	0.232 (0.157)	0.036 (0.114)
Gender dummy (1 = male)	0.054 (0.061)	0.071 (0.078)	0.059 (0.095)	0.236 (0.263)	0.026 (0.115)	0.007 (0.123)
Const.	0.141 (0.158)	0.12 (0.2)	0.22 (0.258)	0.552 (1.065)	0.218 (0.518)	0.265 (0.287)
Obs.	637	428	209	60	49	100
R^2	0.111	0.158	0.085	0.107	0.603	0.199

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Table 15a: **Second-stage selection (Inverse Probability Weighting): One-year growth in log earnings, 2004-2005 & 2005-2006**

	Pooled	Self-Emp.	Wage	Pooled (IPV)	Self-Emp. (IPV)	Wage (IPV)
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: ***One-year change in log earnings***						
Age and tenure:						
Age (years)	0.002	0.002	-.001	0.007	0.005	0.002
	(0.003)	(0.004)	(0.006)	(0.004)*	(0.005)	(0.005)
Education:						
Highest is primary certificate (0/1)	-.059	-.107	0.045	0.104	0.074	0.05
	(0.094)	(0.113)	(0.171)	(0.108)	(0.127)	(0.112)
Highest is middle school certificate (0/1)	-.227	-.181	-.159	0.149	0.255	-.096
	(0.177)	(0.238)	(0.265)	(0.214)	(0.321)	(0.213)
Highest is O Level (0/1)	-.132	-.237	0.069	0.073	-.001	0.031
	(0.112)	(0.145)	(0.174)	(0.124)	(0.151)	(0.117)
Highest is A Level (0/1)	-.162	-.147	0.013	0.085	0.124	0.059
	(0.168)	(0.299)	(0.209)	(0.189)	(0.253)	(0.144)
Highest is a tertiary degree (0/1)	-.641	-1.213	-.395	-.524	-1.114	-.440
	(0.325)**	(0.567)**	(0.384)	(0.287)*	(0.69)	(0.202)**
Occupation characteristics:						
Log (no. employees) (=0 if wage)	-.041	0.019		-.187	-.103	
	(0.118)	(0.128)		(0.134)	(0.126)	
Log (firm size) (=0 for civil service)	-.005		-.014	-.016		-.021
	(0.047)		(0.042)	(0.036)		(0.04)
Civil service dummy (public sector, NGO, etc)	-.175		-.169	-.411		-.247
	(0.123)		(0.173)	(0.219)*		(0.192)
Public enterprise dummy (state-owned firm)	-.108		-.065	-.174		-.084
	(0.219)		(0.144)	(0.182)		(0.133)
Private sector dummy	-.041			-.056		
	(0.158)			(0.11)		

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.

Continued...

Table 15b: **Second-stage selection (Inverse Probability Weighting): One-year growth in log earnings, 2004-2005 & 2005-2006 (cont.)**

	Pooled	Self-Emp.	Wage	Pooled (IPV)	Self-Emp. (IPV)	Wage (IPV)
	(1)	(2)	(3)	(4)	(5)	(6)
Change in occupation characteristics:						
Δ Log(employees) ($\equiv 0$ for sector movement)	0.216 (0.11)**	0.245 (0.117)**		0.056 (0.141)	0.09 (0.131)	
Δ Log(firm size) ($\equiv 0$ for sector movement)	0.009 (0.047)		0.016 (0.041)	0.02 (0.028)		0.014 (0.031)
Within-wage sector movement	0.304 (0.436)		0.09 (0.383)	0.37 (0.184)**		0.341 (0.298)
Wage-earner $\rightarrow$ self-employed	0.042 (0.22)		-.043 (0.192)	0.066 (0.255)		0.085 (0.267)
Self-employed $\rightarrow$ wage-earner	0.233 (0.179)	0.833 (0.331)**		0.243 (0.172)	0.681 (0.177)***	
Other:						
Region dummy (reference: Dar): Arusha	0.245 (0.089)***	0.424 (0.119)***	-.038 (0.131)	0.191 (0.089)**	0.368 (0.118)***	-.051 (0.136)
Region dummy (reference: Dar): Iringa	-.394 (0.114)***	-.366 (0.137)***	-.290 (0.223)	-.217 (0.298)	-.0003 (0.272)	-.995 (0.606)
Region dummy (reference: Dar): Morogoro	0.012 (0.141)	0.118 (0.177)	-.214 (0.222)	-.059 (0.132)	-.094 (0.158)	-.020 (0.218)
Region dummy (reference: Dar): Mwanza	0.306 (0.1)***	0.487 (0.13)***	-.025 (0.147)	0.323 (0.105)***	0.47 (0.147)***	-.007 (0.088)
Region dummy (reference: Dar): Tanga	0.189 (0.109)*	0.125 (0.144)	0.296 (0.157)*	0.111 (0.129)	0.068 (0.187)	0.144 (0.157)
Time dummy, 2005-2006 (reference: 2004-2005)	-.054 (0.074)	-.040 (0.097)	-.061 (0.105)	0.037 (0.101)	0.079 (0.121)	-.053 (0.119)
Gender dummy (1 = male)	0.063 (0.066)	0.062 (0.085)	0.103 (0.102)	0.035 (0.085)	0.069 (0.096)	0.02 (0.11)
Const.	0.08 (0.17)	0.081 (0.216)	0.189 (0.271)	-.215 (0.198)	-.225 (0.244)	0.168 (0.187)
Obs.	566	378	188	566	378	188
R^2	0.099	0.148	0.095	0.099	0.134	0.18
Log-likelihood	-614.054	-428.010	-166.182	-594.362	-408.055	-161.199

Confidence: ***** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%**.

5 Conclusion

This paper has provided a first look at the three-period Tanzania Household Worker Survey. It has used the panel structure of that data to consider the determinants of earnings growth, and to relate these results to findings on the determinants of earnings levels. The key results that emerge relate to the effect of education: education is found to have a significant convex effect upon earnings levels, but to have had no significant effect upon earnings growth (indeed, there is some suggestion that education may have had a *negative* impact). This suggests that recent Tanzanian growth may have reflected an 'unskill-biased technological change', providing relative reward to informal skills rather than to formal education. Further, there are interesting insights into the age-earnings relationship: the relationship is found to be significantly concave in levels, yet age is *not* found significantly to have affected earnings growth. This suggests that the concave levels relationship is driven by workers' participation decisions, rather than by a concave earnings trajectory at the level of the individual worker. Finally, we find significant evidence of variation between formal and informal enterprises. This is evidenced in significant differences between wage-earners and the self-employed, *and* between different sectors within wage-earning employment, *and* between different sizes of enterprise; such differences are particularly evident in the levels relationship.

The literature on the determinants of earnings — and, particularly, on the relationship between education and earnings — has been plagued by endogeneity issues, driven particularly by concerns of unobserved ability and of non-random survey attrition. The panel structure of the Tanzania Household Worker Survey has allowed us to deal with some of those concerns, and to suggest several new insights. It has allowed a consideration of the determinants of earnings growth at an individual level. In doing so, it has shed light on several important issues necessarily left ambiguous by a study of earnings levels. In short, the ability to track individuals over time represents an ability to understand more comprehensively the nature of the income process *as a process*. In doing so, it prompts any number of avenues for future research.

References

Anderson, T. and Hsiao, C. (1981), 'Estimation of Dynamic Models with Error Components', *Journal of the American Statistical Association* **76**(375), 598–606.

Boissiere, M., Knight, J. and Sabot, R. (1985), 'Earnings, Schooling, Ability, and Cognitive Skills', *The American Economic Review* **75**(5), 1016–1030.

Buchert, L. (1994), *Education in the Development of Tanzania, 1919-1990*, Ohio University Press.

Calvès, A. and Schoumaker, B. (2004), 'Deteriorating Economic Context and Changing Patterns of Youth Employment in Urban Burkina Faso: 1980–2000', *World Development* **32**(8), 1341–1354.

Carter, M. and May, J. (2001), 'One Kind of Freedom: Poverty Dynamics in Post-apartheid South Africa', *World Development* **29**(12), 1987–2006.

Cichello, P., Fields, G. and Leibbrandt, M. (2005), 'Earnings and Employment Dynamics for Africans in Post-Apartheid KwaZulu-Natal: A Panel Study of KwaZulu-Natal', *Journal of African Economies* **14**(2), 143–190.

Doyle, A. (2000 [1894]), *The Memoirs of Sherlock Holmes*, Oxford University Press, USA.

Dreze, J., Lanjouw, P. and Stern, N. (1992), 'Economic Mobility and Agricultural Labour in Rural India: A Case Study', *Indian Economic Review* **Special Number 27**, 25–54.

Fafchamps, M., Söderbom, M. and Benhassine, N. (2006), Job Sorting in African Labor Markets. The Centre for the Study of African Economies Working Paper Series, Paper 246.

Fields, G., Cichello, P., Freije, S., Menéndez, M. and Newhouse, D. (2003*a*), 'For Richer or for Poorer? Evidence from Indonesia, South Africa, Spain, and Venezuela', *Journal of Economic Inequality* **1**(1), 67–99.

Fields, G., Cichello, P., Freije, S., Menéndez, M. and Newhouse, D. (2003*b*), 'Household income dynamics: a four-country story', *Journal of Development Studies* **40**(2), 30–54.

Gunning, J. W., Hoddinott, J., Kinsey, B. and Owens, T. (2000), 'Revisiting Forever Gained: Income Dynamics in the Resettlement Areas of Zimbabwe, 1983-1996', *Journal of Development Studies* **36**(6), 131–154.

Heckman, J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica* **47**(1), 153–161.

Knight, J. B. and Sabot, R. H. (1990), *Education, Productivity and Inequality: The East African Natural Experiment*, Oxford University Press.

Mincer, J. (1974), *Schooling, Experience and Earnings*, National Bureau of Economic Research.

Moffitt, R., Fitzgerald, J. and Gottschalk, P. (1999), 'Sample Attrition in Panel Data" The Role of Selection on Observables', *Annales d'Economie et de Statistique* **55**(56), 129–152.

Nickell, S. (1981), 'Biases in Dynamic Models with Fixed Effects', *Econometrica* **49**(6), 1417–1426.

Nyerere, J. (1968), *Freedom and Socialism: A Selection from Writings & Speeches, 1965–1967*, Oxford University Press.

Pissarides, C. A. (2002), Human Capital and Growth: A Synthesis Report, Technical Report 168, OECD Development Centre.

Psacharopoulos, G. (1981), 'Returns to Education: An Updated International Comparison', *Comparative Education* **17**(3), 321–341.

Psacharopoulos, G. (1994), 'Returns to Investment in Education: A Global Update', *World Development* **22**, 1325–1343.

Psacharopoulos, G. and Patrinos, H. A. (2004), 'Returns to investment in education: a further update', *Education Economics* **12**(2), 111–134.

Sandefur, J., Serneels, P. and Teal, F. (2007), *Employment and Shared Growth: Rethinking the Role of Labor Mobility for Development*, World Bank Publications, chapter 5 ('Poverty and Earnings Mobility in Three African Countries').

Söderbom, M., Teal, F., Wambugu, A. and Kahyarara, G. (2006), 'The Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing', *Oxford Bulletin of Economics and Statistics* **68**(3), 261–288.

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.