Conference 2010

Conference Paper

Country Case Study for Nigeria on Skills Development for Informal Sector Workers

Alberto Behar
The World Bank

27 - 29 October 2010
Indaba Hotel and Conference Centre
Johannesburg
South Africa

Hosted by

DPRU
Development Policy Research Unit

TIPS

Country Case Study for Nigeria on Skills Development for Informal Sector Workers

Alberto Behar*; 8 AUGUST 2010

1. INTRODUCTION

Since first recognized in the early 1970s, the informal sector of Sub-Saharan Africa has become a growing source of employment. Initially viewed as a safety net for those unable to find employment in the modern sector, as in the classic treatment by Harris and Todaro (1970), it is now an open question whether the informal sector is an inferior alternative to formal work. It may be a preferred destination offering opportunities for those wanting to become entrepreneurs or for others with certain characteristics (Fajnzylber and others, 2006).

In Sub-Saharan Africa, informal sector activities claim a significant and at times dominant share of their respective markets. The sector was estimated to account on average for 42 percent of GDP in 23 African countries in 2000 and is forcing governments to acknowledge its existence and importance to the national economy and the welfare of those employed therein (Schneider, 2005). In countries throughout the Africa region, where job growth in the formal wage sector has stagnated, self-employment has opened opportunities for youths with some skills (Fox and Gaal 2008) and there is evidence that skills are rewarded with higher earnings in this sector (World Bank, 2008).

A multi-year program of economic and sector work is being undertaken by PREM and HD to provide African policymakers, Bank staff, and other development agencies with knowledge of the informal sector and how skills can contribute to growth and poverty reduction. As part of this work, this paper examines the role of various skills in the informal sector in the case of Nigeria.

Nigeria's much improved economic performance seems to have done little to reduce unemployment, especially among the young. Wage employment has declined, as retrenchment in the civil service and state-owned enterprises has not been compensated by job creation in new industries. Fast growing sectors have been to a significant extent made up of informal workers (Treichel, 2010).

This paper draws in large part on the framework devised by Adams (2008) to answer a number of questions. How do the demographic and skills profiles of those in the informal sector compare to the formal sector? How do skills influence selection into these sectors and do some forms of skill acquisition complement others? Are skills associated with higher earnings in the informal sector and how does this compare to the formal sector? Are earnings higher in the formal or informal sector and to what extent is this attributable to skills or other observables?

*World Bank, abehar@worldbank.org

Section 2 discusses the data set used and some of the approaches used to construct informative variables. Section 3 profiles the employment status of Nigeria and argues that the choice of informal definition does not materially affect the relative sizes of the sectors. 57-63% of males in the non-agricultural sector can be classified as informal while the dominant majority of females (78-81%) in the non-agricultural sector are informal. Section 4 provides descriptive evidence that those in the formal sector have more education or other courses while those in the informal sector are more likely to have done apprenticeships. It finds that formal sector earnings are about 70% higher than informal sector earnings. For men, part of this is due to an educational composition effect. Section 5 discusses patterns of complementarity in skill acquisition; for example apprentices tend to have less education but are more literate/numerate.

Section 6 estimates earnings functions and participation equations. It finds that the returns to education are quite low in the informal sector and lower than in the formal sector, although this is in part due to the latter having low earnings at low education levels and not just high earnings at high education levels. Much of the formal sector gain happens at the post-secondary level. We find no evidence that other skill acquisition methods contribute to earnings. However, apprenticeships are positively associated with selection into the informal sector while other forms are positively associated with selection into the formal sector. The effect of education on selection into the informal sector is low and potentially negative beyond a certain level.

Section 7 synthesizes the findings and suggests that some or all of the observed formal sector premium can be accounted for by observable covariates, at least for men. It nonetheless remains an open question whether people with certain observed or unobserved characteristics would prefer a job in the formal sector.

2. DATA

The dataset is the 2003–04 Nigeria Living Standards Survey (NLSS), which is presented and analyzed in Haywood (2007) and also Haywood & Teal (2010). Variables of interest are employment status, earnings and human capital (formal education, apprenticeships, on-the-job training, literacy/numeracy and other courses). This study is aimed at understanding skills in the informal sector for people aged 15-65.

When we consider employment status, it refers to who the respondent says they work for during their primary activity. We do not report results from another question on self-reported employment status because it does not easily separate out agricultural workers. We did however compare answers to these

[Type text]

questions and drop observations where the answers were inconsistent. [1] Further, we concentrate on primary activity as secondary or subsequent activities reflect a small number of observations.

Respondents were asked the highest level of schooling completed as well as the highest certification attained. While the latter has the advantage of conveying achievement as opposed to attendance, the former allows us to distinguish between those who have only some primary education as opposed to those who finished primary school. We use the former (schooling completed in years) for much of the analysis but also use discrete measures of educational attainment. Further, we removed observations where answers to these two questions were clearly inconsistent. [2]

To assign years, we aggregate the 24 potential education completion categories into seven, namely those listed in Table 1, which assigns years of education to each category. This is in line with the 6-3-3-4 education system. We have assigned vocational training a value of 10 so that it lies between junior and senior secondary. Koranic education is assigned the same value as having some primary education.

Educcompleted	Years
None	0
some primary or Koranic	3
primary complete / some junior secondary	6
junior secondary complete / some senior secondary	9
Vocational	10
senior secondary complete	12
post-secondary	16

Table 1: Aggregate education categories (based on education completed) and years assigned.

We also need measures of earnings or income. As noted by Falco and others (2009), collecting income data on the self-employed in low-income countries is problematic because those in the informal sector, especially the self-employed, rarely keep written accounts and their data may be too noisy to be of use. With this important caveat, we discuss two sources of this information.

1. Primary earnings. Respondents were asked if they received remuneration for their primary activity and then how much in various time units, which we standardized to monthly amounts. All measures exclude those who reported they are unpaid, even if they report a number, and exclude those for whom income is zero (or missing). Furthermore, we provide measures which are censored at the top/bottom 1% or 5% of values.[3] Both censored samples also exclude

---

[1] For example, people may have answered they work for themselves in one question but then don't state self employed as their employment status. Another example is where they state unpaid family work but then report they received remuneration.

[2] For example, if people said they attained a higher level than they completed, we drop them. While achievement implies completion, the reverse is not true. Nonetheless, we did drop unlikely combinations; for example, people who completed post secondary but who attained less than primary school were also dropped.

[3] The 1% thresholds are 300 Naira (less than three US dollars and well below the poverty line) and 285,000 Naira per month. The 5% thresholds are 1,000 and 50,000.

[Type text]

people who are reported to have worked more than 320 hours a month (which corresponds to a 16 hour day, 5 days a week, 4 weeks a month) and fewer than 40 hours a month (which corresponds to a 2 hour day). These observations are excluded for two reasons. First, a 40 hour month corresponds to less than 25% of a standard 160 hour month, which is inconsistent with the activity being primary. Second, this excludes numbers that are likely to have been inaccurately given or reported.[4]

2. Screening income. In addition to the earnings reported from primary activity, we use a measure based on screening income. Recipients were asked how much they received over the entire year rather than in a given respondent-selected time unit. Furthermore, questions were asked about all sorts of sources of income, including asset sales, loans, and government transfers. However, for much of the analysis, we construct our (monthly) measure by summing the amounts from employment income and non-farm business income. We also exclude observations with obviously inconsistent earnings/income answers and condition on positive income. We also censor this measure.

We also analysed a measure that accounts for monetary estimates of benefits. This affected a relatively small proportion of respondents and is likely to be susceptible to error, so we exclude these from the analysis. It's also important to note that primary earnings question asked whether people had had taxes deducted but the follow up question did not say how much, so it is impossible to adjust for this in a reliable way. Indications are that those with taxes deducted had higher remuneration even after deduction.

Table 2 compares the primary earnings and screening income measures as well as the effects of censoring.

Source	Primary Earnings			Screening Income		
Censoring	None	1%	5%	None	1%	5%
weighted mean	50455	15916	11908	9004	8396	7203
unweighted mean	39730	15082	11474	8230	7760	6754
median	9700	9800	9500	5000	5000	5000
USD weighted mean	388	122	92	69	65	55
USD median	75	75	73	38	38	38
std deviation	624221	31141	11290	12564	9967	7252
N	6183	4709	4405	6266	6135	5577

Table 2 Measures of earnings/income at various censoring thresholds. All measures conditional on a positive answer and exclude inconsistent answers.

If we omit the top and bottom 5% of primary earnings, the weighted average monthly wage is 92 US Dollars[5] or about 11 500 Naira.[6]   The overall median is about 75 USD. The standard error of the

---

[4] Respondents could give hours worked on the primary activity per day, week, fortnight, month, quarter or year and it is likely that the unit of measure was not always accurately captured or reported.
[5] In 2003, the exchange rate was just below 130 and in 2004 it was just above, so we use an exchange rate of 130.
[6] This compares with a mean of about 11 000 Naira in Haywood (2008), although he adds benefits and uses a different sample. Further, the nature of censoring, if any, is not clear.

[Type text]

estimated mean can be used to calculate the standard deviation, which is useful for calculating the standard deviation with population weights.[7] Censoring dramatically reduces the standard deviation.

We see the screening income measure is much lower than the primary earnings measure, regardless of the measure used. For example, the 1% censored mean or median income is approximately half of the primary earnings measure. The earnings measure is particularly susceptible to censoring of hours and incomes. Because the wage is from primary activity only while the income is potentially from all activity, the latter should if anything be higher. In this regard, two remarks are in order:

    i.   There are two potential sources of error when constructing the primary earnings. First, like the income measure, the actual amount received in a period may be inaccurate. Second, unlike the income measure, the unit reported may be incorrectly reported or recorded. When combined to yield a monthly amount, these errors can be dramatically multiplied.

   ii.   The higher primary wage can be the result of the implicit assumption of full employment in this activity. Thus, people reporting weekly wages from the primary activity are assumed to have worked this way for 4 weeks in the month. In practice, many are likely to be underemployed.

As a result, it is important to present results based on both measures of income/earnings.

3. PROFILE OF EMPLOYMENT STATUS (DEFINING INFORMALITY)

Table 3 provides a breakdown of employment status as given by the answer to the question on who respondents work for as their primary activity. The full sample of people describing a primary activity has almost two thirds engaged in family agriculture. The next biggest category is self-employed workers. To have a better description of contributions within the non-agricultural sector, we exclude family agriculture and also those classified as other.[8] More than half of those employed in the non-agricultural sector work for themselves. The sample proportion (61%) is slightly higher for women than for men (52%) while the ratio of self employed persons to paid employees is almost three times higher for women.

---

[7]Calculated by multiplying the estimated standard error of the estimated mean (with the *mean* command in *Stata*) by the square root of N to get the standard deviation. By specifying population weights, one can do the same with the weighted mean and weighted standard error.

[8] The 'other' category also includes people who gave inconsistent answers regarding employment status.

	All			Nonag		
	Male	Female	Total	Male	Female	Total
Family Agriculture	11,314	8,806	20,120			
	64.29	*61.31*	*62.95*			
Public	1,313	532	1,845	1,313	532	1,845
	7.46	*3.7*	*5.77*	*31.68*	*14.09*	*23.3*
Private	554	215	769	444	175	619
	3.15	*1.5*	*2.41*	*10.71*	*4.63*	*7.82*
Self Employed	2,259	2,367	4,626	2,148	2,287	4,435
	12.84	*16.48*	*14.47*	*51.83*	*60.57*	*56*
Unpaid Family	239	782	1,021	239	782	1,021
	1.36	*5.44*	*3.19*	*5.77*	*20.71*	*12.89*
Other	1,920	1,660	3,580			
	10.91	*11.56*	*11.2*			
Total	17,599	14,362	31,961	4,144	3,776	7,920
	100	*100*	*100*	*100*	*100*	*100*
Self employed ratio	1.21	3.17	1.77	1.22	3.23	1.80

Table 3: Employment Status. All refers to all respondents with a primary employment activity.Nonag excludes family agriculture and the "other category". Percentages in italics. Self emplyed ratio is ratio of self employed workers to public or private employees.

The survey asks more detailed questions about the job or occupation. Response rates were low and there is some obvious inconsistency between questions.[9] Nonetheless, almost a third of respondents in the informal sector said they were engaged in proprietary wholesaling, where women outnumber men by two to one. Various forms of service activity were also prominent.

We now classify the non-agricultural workers according to whether their primary activity is in the formal or informal sectors. As discussed in Adams (2008), there is no obvious way to do so despite recent moves towards consistency. We considered a number of alternative definitions, which we list from narrowest to broadest.

1. Informal if self-employed or does unpaid family work (narrow definition).
2. Informal if self-employed or does unpaid family work, or works for a co-operative.
3. Informal if self-employed or does unpaid family work, or works for a co-operative, or is employed by a private firm with less than 10 employees and gets no benefits.[10]
4. Informal if self-employed or does unpaid family work, or works for a co-operative, or is employed by a private firm with less than 10 employees (broad definition).

In practice, the difference between definitions 1 and 4 is small. In Figure 1, we show this by classifying those who are not informal under (4) as "formal", classifying those who are informal under (1) as "informal" and those who are informal in definitions 2-4 but not in definition 1 (ie work for a co-op or a

---

[9] For example, 31 people say they do hairdressing while 93 say they are hairdressers.
[10] Benefits are a catch-all phrase for either having a formal contract, having a pension, the firm being unionized, entitlements to paid holidays / sick leave / medical care or other benefits.

[Type text]

small private firm) as "intermediate". Figure 1 shows that this intermediate category reflects a small proportion of workers for both genders: only 5% of males are ambiguously defined while less than 3% of females are intermediate. As a result, we will use the broadest definition in most of the analytical work.
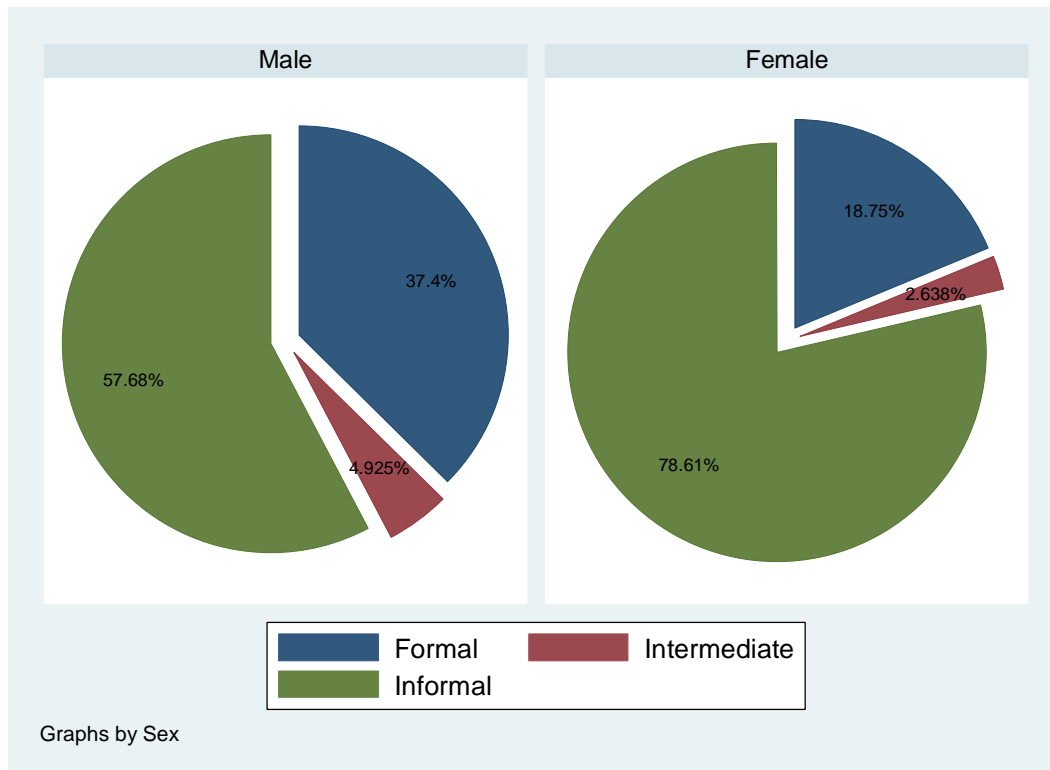

Figure 1: Formal/informal breakdown of primary activity. Informal workers are self-employed or unpaid. Formal workers are those who work for the government/parastatals and/or small private sector firms. Intermediate workers are those who work for small private firms (with or without benefits) or NGOs. Population weighted.

Further, the chart shows that 57-63% of males in the non-agricultural sector can be classified as informal while the dominant majority of females (78-81%) in the non-agricultural sector are informal.

Table 4 attaches actual numbers rather than percentages. First, it shows the total number of people reporting a secondary employment activity is much smaller than those reporting a primary activity. Second, it confirms that agricultural activity is the dominant primary activity and comprises about half of secondary activity. Third, it shows that informality is relatively more important in the secondary activity than in the primary activity. Even in the primary activity, the informal share of non-agricultural workers is more than two thirds. Forth, it illustrates the relatively minor consequences of using the narrow as opposed to broad definition of an informal worker.

[Type text]

	Primary narrow		Primary broad		Secondary narrow		Secondary broad	
	Number	%	Number	%	Number	%	Number	%
Agriculture	20,120	70.89	20,120	70.89	1,245	48.16	1,245	48.16
Formal	2,614	9.21	2,230	7.86	172	6.65	119	5
Informal	5,647	19.9	6,031	21.25	1,168	45.18	1,221	47.23
Total	28,381	100	28,381	100	2,585	100	2,585	100
Informal:Formal	2.2		2.7		6.8		10.3	
Informal share	68.4		73.0		87.2		91.1	

Table 4: Breakdowns according to definition of informal sector. Narrow measure of informality includes self employed and unpaid family workers. Primary and secondary refer to the activity. Broad definition includes employees in the private/NGO/Co-op sector in small firms. Informal:Formal is ratio of informal to formal sector wokers while informal share refers to the share of the non-agriculture workforce. Unweighted, so figures not directly comparable with pie chart.

4. COMPARING THE FORMAL AND INFORMAL SECTORS

This section compares demographic, human capital (education, training, literacy etc) and earnings features between the formal and informal sectors.

DEMOGRAPHY

Table 5 shows that the formal sector has a lower proportion of women and a slightly higher proportion of urban dwellers. Age is approximately the same. Table 6 concerns marital status. Those in the formal sector are even more likely to be monogamously married and less likely to be in a polygamous marriage.

	Narrow		Broad	
	Formal	Informal	Formal	Informal
Proportion of females	30%	54%	30%	53%
Proportion of urban	74%	68%	74%	68%
Mean Age	40.8	39.6	41.4	39.4

Table 5: Gender, sector and age by formality. Population weighted proportions.

[Type text]

	Male	Female	Formal	Informal	Total
Monogamous	2,672	2,106	1,496	3,282	4,778
	64.48	*55.77*	*68.59*	*57.19*	*60.33*
Polygamous	497	833	245	1,085	1,330
	11.99	*22.06*	*11.23*	*18.91*	*16.79*
Informal	10	13	10	13	23
	0.24	*0.34*	*0.46*	*0.23*	*0.29*
Divorced	27	48	25	50	75
	0.65	*1.27*	*1.15*	*0.87*	*0.95*
Separated	104	165	77	192	269
	2.51	*4.37*	*3.53*	*3.35*	*3.4*
Widowed	59	325	60	324	384
	1.42	*8.61*	*2.75*	*5.65*	*4.85*
Never	775	286	268	793	1,061
	18.7	*7.57*	*12.29*	*13.82*	*13.4*
Total	4,144	3,776	2,181	5,739	7,920
	100	*100*	*100*	*100*	*100*

Table 6: Marital status by gender (non-agricultural) and formality (broad definition). Percentages in italics.

SKILLS

Tables 7 and 8 compare education levels. Males have more years of education the females, while people in the formal sector have more education than those in the informal sector. We would expect intermediate workers to have an intermediate level of education. This is consistent with what happens when we move from the broad to the narrow definition of informality, which raises both means.

	Nonag		Nonag narrow		Nonag broad	
	male	female	formal	informal	formal	informal
Mean	10.49	9.71	12.80	8.53	13.15	8.64
std error	0.09	0.11	0.10	0.08	0.10	0.08

Table 7: Years of education by gender (non-agricultural) and formality.

The main difference occurs at the post-secondary level, where over half of formal sector workers completed that level of education while only 6% of informal sector workers did (Table 8). Other than vocational work, bigger proportions of informal sector workers completed all lower levels, including the senor secondary level. Furthermore, we report that 2% of formal sector workers attended technical (part of post-secondary) or vocational (shown) school and 1.2% of informal workers did.[11]

---

[11] Our measure of educational attainment has a correlation of 0.96 with educational completion, in part because obviously inconsistent answers were excluded. Thus, results from those statistics re-enforce the pattern of Table 8.

[Type text]

Education completed	Narrow		Broad	
	Formal	Informal	Formal	Informal
none	6	40	2	44
	0.27	*1.18*	*0.1*	*1.21*
some primary / Koran	62	495	39	518
	2.83	*14.57*	*1.99*	*14.25*
primary & some sec	344	1,346	279	1,411
	15.68	*39.62*	*14.26*	*38.83*
junior sec & some senior	74	242	55	261
	3.37	*7.12*	*2.81*	*7.18*
vocational	19	21	17	23
	0.87	*0.62*	*0.87*	*0.63*
senior secondary	586	1,059	503	1,142
	26.71	*31.17*	*25.7*	*31.43*
post-secondary	1,103	194	1,062	235
	50.27	*5.71*	*54.27*	*6.47*
Total	2,194	3,397	1,957	3,634
	100	*100*	*100*	*100*

Table 8: Education completion by formality. Percentages in Italics.

Table 9 shows a greater proportion of men receive apprenticeships but they are not necessarily longer. One in five informal sector workers did an apprenticeship. Furthermore, the proportion of former apprentices in the informal sector is higher than in the formal sector, even if we use a narrow definition which excludes paid work for someone else. We also state that less than 10% of respondents paid for their apprenticeship in cash or in kind.

	male	female	Narrow formal	informal	Broad Formal	informal
Did Apprenticeship	21%	12%	8%	21%	6%	21%
Duration	34.7	29.5	31.9	33.2	30.4	33.3
Respondents	803	367	162	1008	110	1060

Table 9: Apprenticeships by gender and informality. Weighted. Duration in weeks conditional on doing an apprenticeship.

Table 10 compares on-the-job training (OJT) in the formal and informal sectors. In contrast to apprenticeships, OJT is rare or nonexistent in the informal sector. Almost all recipients of OJT are in the formal sector. Furthermore, we report that the employer paid in over half of the cases. Of the entire survey, only eight people said they paid in the form of lower earnings.

	Narrow Formal	Informal	Broad Formal	Informal
Did OJT	6.3%	0.0%	6.9%	0.7%
Duration	4.9	.	4.9	2.8
Respondents	144	0	138	6

Table 10: On-the-job training by formality. Duration in weeks conditional on getting OJT.

[Type text]

Table 11 presents information on literacy and numeracy. Formal sector workers tend to do better on these counts.

	Nonag narrow		Nonag broad	
	Formal	Informal	Formal	Informal
read	91%	61%	92%	62%
write	89%	57%	91%	58%
literacy course	3.4%	2.1%	3.8%	2.0%
calc	89%	62%	90%	63%

Table 11: Literacy and numeracy by sample, gender and formality. Read & write refers to English. Course is for a literacy course attended.

EARNINGS

Recall that the overall weighted mean monthly wage from primary earnings for the censored (5%) sample was 92 USD and that from the uncensored screening income measure was 69. In table 12, we compare the formal and informal sectors.

	Primary earnings (Censored5)				Screening Income (Paid)			
	Narrow		Broad		Narrow		Broad	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
weighted mean	121	74	127	74	95	54	99	55
weighted std error	2	1	2	1	2	1	3	1
unweighted mean	117	71	121	71	89	48	93	48
median	100	48	108	50	69	31	77	31
N	13	21	12	22	18	30	16	32

Table 12: Earnings by sector in US Dollars.

The key message is that unconditional income is higher in the formal than informal sector. For example, weighted average primary earnings are $74 for broadly defined informal sector workers while $127 for the formal sector workers. Given the low standard errors, these are statistically significantly different. Further, this does not depend on the income measure used nor does it depend on the informality definition chosen.[12] One average, across four comparisons, the premium is 73%.

Figure 2 compares the distributions of log wages in the formal and informal sectors. In both sectors, the densities reflect approximately lognormally distributed earnings. For the informal sector, lower (log) wages have higher densities with a mode of about 9. For the formal sector, the mode is at almost 10, which implies a mode at close to double the wage. The lines only cross once and there is a substantial degree of overlap: many people in the informal sector are earning more than many people in the formal

---

[12] In the completely uncensored primary earnings sample, the informal sector wage had a higher mean due to some extremely high informal sector wages in the top percentile of the distribution.

[Type text]

sector.[13] As discussed in Adams (2008) and Falco and others (2009), this mildly supports the view that the informal sector is not necessarily inferior to the formal sector.
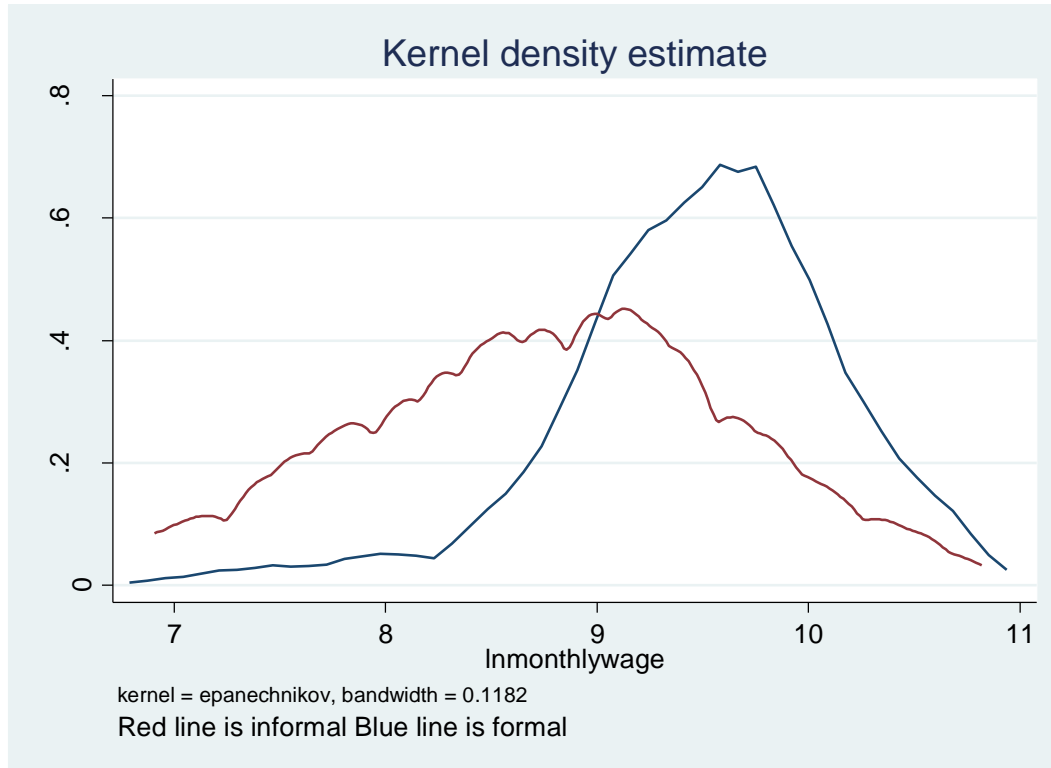

Kernel density estimate

kernel = epanechnikov, bandwidth = 0.1182

Red line is informal Blue line is formal

Figure 2: Distribution of log monthly wage (in Naira) by sector. Censored 5% primary earnings

Furthermore, it is possible that some of the formal sector earnings premium observed on average can be accounted for once one conditions on other covariates. To illustrate, we compare earnings in the formal and informal sectors but classify by level of education. This is done in Table 13.

For male primary earnings, on the left hand side of Table 13, informal primary earnings are higher than formal primary earnings up the junior secondary level, where they are approximately the same. Formal wages are higher for vocational qualifications, which are few in number, but it is mainly at the post-secondary level at which the formal sector premium appears.   Thus, on this evidence, it seems that the

---

[13] However, our analysis of the cumulative distribution functions (available on request) confirms that distribution for the informal sector is everywhere above that for the formal sector. Thus, there is always a greater proportion of informal sector workers than formal sector workers earning less than a certain amount, regardless of that amount. We can say that the distribution of formal sector workers stochastically dominates the informal sector. In other words, the worst paid formal sector worker earns more than the worst paid informal sector worker, the second worst paid earns more in the formal sector and so on, continuing throughout the distributions such that the best paid formal sector worker earns more than the best paid informal sector worker.

[Type text]

higher overall mean for the formal sector is due in large part to a composition effect because many people with a post-secondary education are in the formal sector and many people with lower levels are at lower levels (eg completed primary). Thus, according to this evidence, whether or not earnings are higher in the formal or informal sectors depends very much on the level of education.

For male screening income, the pattern is not as strong. Informal sector incomes are lower at almost all levels. However, informal incomes are only much lower than formal incomes at the high education levels, so a composition effect is again playing an important role.

For women, the table suggests formal sector earnings are higher than informal sector earnings at all education levels, regardless of the earnings measure. The differentials overall are much greater than they are for men.

| | | Men | | | | Women | | | |
| | | Primary Earnings | | Income | | Primary Earnings | | Income | |
		Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
none	Mean	12942	17000	6613	10000	4508	.	2314	.
	Number	12	1	14	1	13	0	19	0
some primary or	Mean	11778	9715	7197	7429	6003	13400	3076	9483
Koranic	Number	112	18	169	32	84	3	131	3
primary complete	Mean	12105	11496	7618	8253	6577	10556	4308	9320
	Number	412	168	595	219	412	27	544	36
junior secondary	Mean	11592	11682	7280	5401	6052	8173	4094	6857
	Number	74	27	107	35	68	12	103	17
vocational	Mean	7543	11139	7652	8907	5200	8550	6889	5565
	Number	7	6	11	10	5	5	9	7
senior secondary	Mean	12171	13660	8645	10136	6733	10593	5247	7989
	Number	428	270	553	350	305	97	417	122
post-secondary	Mean	16823	20188	11658	14287	9498	17820	7958	13561
	Number	99	486	129	642	52	260	72	333
Total	Mean	12454	16399	8232	11678	6664	15165	4661	11602
	Number	1144	976	1578	1289	939	404	1295	518

Table 13: Comparing formal and informal earnings by education level. Unconditional earnings measures censored at 5% (primary earnings) and 1% (screening income). Naira.

Furthermore, comparing post-secondary to some primary earnings in the informal sector for men, the rise is about 5000 Naira or less than 50%. However, for the formal sector, post-secondary earnings are more than double those of someone with only some primary. In other words, earnings rise more by education level in the formal sector, but this is as much a depiction of low formal sector earnings at low education levels as it is one of high formal sector earnings at high education levels. For screening income, earnings also appear to rise faster up the education ladder but do not start off lower than in the

informal sector. Because there are so few women with low education in the formal sector, it is hard to compare earnings across education levels in a reliable manner.

5. MORE ON SKILLS ACQUISITION

Table 14 goes into more detail on apprenticeships. For example, it describes the association between informal status apprenticeships; the middle left of the table indicates that having had an apprenticeship is slightly more likely among the self employed. Further, the higher propensity for former apprentices to be in the informal sector is not affected by education breakdown. It appears that there is no relationship between average education levels and having been an apprentice. There appears to be no relationship between on-the-job training (OJT) and apprenticeship either. This suggests that the relative scarcity of former apprentices in the formal sector is not due to substitution by OJT. Finally, in the informal sector, people who can read, write and/or calculate are more likely to have been apprentices.

Analogously, we now go into more detail on OJT, as shown in Table 15. OJT is more prevalent among public sector workers and those who work in large private organizations than those who are self employed or unpaid. People who attended a literacy course and work in the formal sector are more likely to have received OJT than those who did not attend or who are in the informal sector. Unlike the case of apprenticeships, there is evidence that people with more years of education were more likely to have received OJT. Further analysis revealed that the likelihood rises at senior secondary and beyond. In fact, we report that over 50% of those with OJT have a post-secondary education; those with senior secondary or more comprise 85%. In contrast, the biggest contribution to apprenticeships comes from those with only primary school, who make up almost half.

These descriptive statistics are consistent with probit analysis on the probability of being apprenticed or receiving OJT (results available on request). So, it seems that apprenticeships and education are neither complements nor substitutes, apprenticeships and basic literacy/numeracy are complements, and apprenticeships and OJT are substitutes. OJT and advanced formal education and formal literacy courses appear to be complements.

[Type text]

	Men		Women		Urban		Rural	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	95	708	15	352	70	639	40	421
	6.1	*27.38*	*2.41*	*11.16*	*6.02*	*24.47*	*3.93*	*13.46*
No	1,463	1,878	608	2,801	1,093	1,972	978	2,707
	93.9	72.62	97.59	88.84	93.98	75.53	96.07	86.54
Total	1,558	2,586	623	3,153	709	3,065	1,018	3,128
	100	100	100	100	18.79	81.21	100	100
	Monthly wage > 9700		Monthly wage <=9700		Monthly income >= 3333		Monthly income < 3333	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	61	410	38	570	78	629	32	431
	4.02	*26.11*	*7*	*22.37*	*5.1*	*23.31*	*4.91*	*14.17*
No	1,456	1,160	510	1,978	1,451	2,069	620	2,610
	95.98	73.89	93	77.63	94.9	76.69	95.09	85.83
Total	1,517	1,570	548	2,548	1,529	2,698	652	3,041
	100	100	100	100	100	100	100	100
	Worked for				Educ>=10 years		Educ<10 years	
	Public	Private	Self	Unpaid Family	Formal	Informal	Formal	Informal
Yes	62	100	991	17	60	328	40	533
	3.36	*16.16*	*22.34*	*1.67*	*3.87*	*24.1*	*11.24*	*26.6*
No	1,783	519	3,444	1,004	1,492	1,033	316	1,471
	96.64	83.84	77.66	98.33	96.13	75.9	88.76	73.4
Total	1,845	619	4,435	1,021	1,552	1,361	356	2,004
	100	100	100	100	100	100	100	100
	Received OJT		No OJT		Can Calculate		Can't Calculate	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	9	170	101	69	98	848	12	212
	6.25	*5.09*	*5*	*5.32*	*5.1*	*26.7*	*4.65*	*8.27*
No	135	3,171	1,942	1,229	1,825	2,328	246	2,351
	93.75	94.91	95.06	94.68	94.9	73.3	95.35	91.73
Total	144	3,341	2,043	1,298	1,923	3,176	258	2,563
	100	100	100	100	100	100	100	100
	Can Read		Can't Read		Can Write		Can't Write	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	102	845	8	215	99	759	11	301
	5.16	*27.18*	*3.94*	*8.17*	*5.06*	*26.42*	*4.91*	*10.5*
No	1,876	2,264	195	2,415	1,858	2,114	213	2,565
	94.84	72.82	96.06	91.83	94.94	73.58	95.09	89.5
Total	1,978	3,109	203	2,630	1,957	2,873	224	2,866
	100	100	100	100	100	100	100	100

Table 14: Apprenticeship received by informality and other possible covariates. Broad definition.

[Type text]

	Men		Women		Urban		Rural	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	117	4	21	2	82	5	56	1
	7.51	*0.91*	*3.37*	*0.23*	*7.05*	*1.98*	*5.5*	*0.1*
No	1,441	434	602	864	1,081	248	962	1,050
	92.49	99.09	96.63	99.77	92.95	98.02	94.5	99.9
Total	1,558	438	623	866	1163	253	1,018	1,051
	100	100	100	100	18.79	81.21	100	100
	Monthly wage > 9700		Monthly wage <=9700		Monthly income >= 3333		Monthly income < 3333	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	112	3	92	6	46	0	113	4
	7.38	*3.45*	*6*	*1.32*	*7.06*	*0*	*7.28*	*2.06*
No	1,405	84	1437	447	606	851	1439	190
	92.62	96.55	94	98.68	92.94	100	92.72	97.94
Total	1,517	87	1529	453	652	851	1552	194
	100	100	100	100	100	100	100	100
	Worked for				Educ>=10 years		Educ<10 years	
			Priv.	Unpaid				
	Public	Priv. Big	Small	Family	Formal	Informal	Formal	Informal
Yes	121	17	6	0	113	4	15	1
	6.56	*5.06*	*2.12*	*0*	*7.28*	*2.06*	*4.21*	*0.3*
No	1,724	319	277	1,021	1,439	190	341	331
	93.44	94.94	97.88	100	92.72	97.94	95.79	99.7
Total	1,845	336	283	1,021	1,552	194	356	332
	100	100	100	100	100	100	100	100
	Attended Literacy		Didn't attend		Can Calculate		Can't Calculate	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	16	0	122	6	127	6	11	0
	16.84	*0*	*6*	*0.47*	*6.6*	*1.29*	*4.26*	*0*
No	79	14	1,964	1,284	1,796	458	247	840
	83.16	100	94.15	99.53	93.4	98.71	95.74	100
Total	95	14	2,086	1,290	1,923	464	258	840
	100	100	100	100	100	100	100	100
	Can Read		Can't Read		Can Write		Can't Write	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Yes	129	6	9	0	129	6	9	0
	6.52	*1.33*	*4.43*	*0*	*6.59*	*1.42*	*4.02*	*0*
No	1,849	444	194	854	1,828	418	215	880
	93.48	98.67	95.57	100	93.41	98.58	95.98	100
Total	1,978	450	203	854	1,957	424	224	880
	100	100	100	100	100	100	100	100

Table 15: Receipt of OJT by formality and other possible covariates. Broad definition.

[Type text]

6. SKILLS, EARNINGS AND PARTICIPATION

This section presents a multivariate analysis of participation in the formal/informal sectors as well earnings in each of those sectors. Our analysis commences with specifications that are linear in the years of education but will proceed to present additional insights from alternative specification of education. In terms of other right hand side variables, a number of alternative specifications were attempted but the results are robust, so we present only one set. The regressions do not include on-the-job training because this restricted the sample size. However, those findings will be discussed where relevant. Three estimation procedures are presented.

I.   Linear OLS, where separate earnings regressions are run for the formal and informal sectors.
II.   Heckman 2-step correction approach, where we control for selection into the informal sector (as opposed to informal and agricultural) or we control for selection into the formal sector (as opposed to the other two) before estimating the earnings function in the second stage.
III. Multinomial logit selection model, which explicitly recognizes the fact that there are three choices. The first stage jointly estimates the probability of selection into the formal or informal sectors relative to agriculture using a multinomial logit model. Conditioning on this, the second earnings function stage is separately estimated for the formal and informal sectors.[i]

We have not attempted to instrument for education or the other forms of human capital acquired, which leaves the estimates vulnerable to endogeneity bias.[ii] We will focus on earnings first before moving onto participation. Table 16 presents results where the dependent variable is earnings from the primary activity and Table 17 presents results for the alternative income screening variable. In all cases, we use the broad definition of informality.

EARNINGS

Conditioning on other covariates, there is little robust statistical evidence that apprenticeships have an impact on earnings, regardless of sector. The coefficients are generally negative for the primary earnings and generally positive for the screening income measure. Similarly, the coefficients for the literacy course are insignificant on either side of zero. For primary earnings regressions, the short course variable is insignificant but most coefficients are positive. Screening income regressions suggest a positive statistically significant relationship in the formal sector but not the informal sector. OJT was also insignificant but excluded from the presented regressions because it substantially reduced the sample size (results on request).

In the informal sector, coefficient estimates for education completed range from less than 1% (primary earnings using MNL method) to 6.4% (screening income on MNL method). For the formal sector, the range is between 6% and 7% for four of the estimates, almost 9% for the Heckman-correction screening

income estimate and over 15% for the MNL screening income estimate.[14] As a point of comparison, recent Mincerians from seven African countries have an average of 7%.[15] These do not distinguish by formality/informality. The average marginal returns for the informal sector from 8 West African cities (taken from Kuepie and others, 2006*),* is calculated to be 5%. Thus, while not unique to Nigeria or this study, the earnings gains from education appear to be low in the informal sector.

Comparing the informal and formal estimates, we see that the coefficients are always lower for the informal sector. In many cases, the informal coefficient is less than half the formal coefficient. This is consistent with the gain to education being higher in the formal sector than the informal sector.[16] This is also consistent with what we observed in Table 13, where income/earnings rose more by education level in the formal sector than the informal sector and where we saw that formal sector primary earnings are low at low education levels.

Comparing across tables, it is clear that the primary earnings estimates are lower than the screening income estimates, especially for the MNL specifications.[17]   Regressions run on primary earnings censored at 1% yielded results similar to the other primary earnings regressions. In other words, the difference is not due to censoring.

With a view to some of the potential complementarities between education/training vehicles noted in the previous section before, we tried a number of interactions between education and the apprenticeship/course dummies. However, we found no reliable evidence of interaction effects in the earnings functions.

---

[14] The MNL result is out of line from the others but robust to a number of alternative specifications and exclusion variables.

[15] Taken from data in Behar (2009)

[16] However, it is possible that measurement error is more serious for the informal sector than the formal sector, which would lead to more attenuation bias in the former.

[17] *If* the primary earnings estimates have more measurement error than the screening income estimates, you would expect the primary coefficients to be lower, *ceteris paribus*. However, the measurement error could well be cancelling any upward endogeneity bias, so it would not necessarily follow that the primary estimates are too low.

[Type text]

		Primary Earnings OLS		Primary Earnings Heckman		Primary Earnings MNL	
		Informal	Formal	Informal	Formal	Informal	Formal
2nd stage	main						
	male	0.607***	0.115**	0.637***	0.130**	0.6187***	0.093
	urban	0.110*	0.122*	-0.199	0.118*	-0.0869	0.2753*
	education	**0.0250***	**0.0605***	**0.0379***	**0.0660***	**0.0054**	**0.0672***
	age	0.0563***	0.0636***	0.0442**	0.0650***	0.0393*	0.0723*
	agesq	-0.0005***	-0.0006**	-0.0004*	-0.0006**	-0.0003	-0.0007
	apprentice	**-0.0606**	**-0.0741**	**-0.271***	**-0.111**	-0.0897	-0.0734
	shortcourse	**0.281**	**0.0202**	**0.425**	**0.0432**	0.2302	0.0278
	literacycourse	**0.148**	**0.0279**	**0.199**	**0.0365**	-0.0664	-0.004
	literate/numerate	0.0337	0.0803	-0.0188	0.092	0.0806	0.1321
	selection_ag					0.1902	-0.0597
	selection_formal					-0.2389	0.0863
	selection_informal					-0.0896	0.4987
1st stage	male			-0.0902*	0.339***	-0.3287***	0.2994***
	urban			0.997***	-0.127	2.097***	1.730***
	education			**-0.0568***	**0.161***	**0.0297***	**0.313***
	age			0.0444***	0.125***	0.0585***	0.2858***
	agesq			-0.0006***	-0.0013***	-0.0008***	-0.0030***
	apprentice			**0.859***	**-0.701***	**1.487***	**-0.0574**
	shortcourse			**-0.438***	**0.821***	**0.053**	**1.313***
	literacycourse			**-0.270***	**0.193**	**-0.4111****	**-0.0392**
	literate/numerate			0.199***	0.179*	0.0764	0.3447***
	hh_kids			-0.0214**	-0.0001	-0.0384***	-0.0613***
	married			0.0489	-0.160**	-0.0837	-0.1453*
	Government grants					-0.00001**	-0.0000*
Ancillary statistics	N	2169	1391	12542	5014	12546	12546
	lambda			-0.393	0.0512		
	rho_ag					0.3116	-0.1077
	rho_formal					-0.3913	0.1556
	rho3_informal					-0.1468	0.8989

Table 16: Earnings functions; dependent variable is monthly wages from primary activity in Log Naira. 1st stage is either probit equation for participation in formal or informal sectors (Binary) or a multinomial logit simultaneously estimated for the formal and infrmal sectors relative to agriculture (coefficients, not marginal effects); lambda indicates correlation between errors (the inverse mills ratio) in probit and 2nd stage while rho_terms are analagous for the MNL; selection_ terms are control terms for participation in each of the three sectors. Clustered standard errors in linear and Heckman models; adjusted for two-stage procedure in Heckman; bootstrapped in MNL; population weights used in OLS and Heckman; *** 0.1% ** 1% * 5%; state dummies included but not reported.

[Type text]

		Screening Income OLS		Screening Income Heckman		Screening Income MNL	
		Informal	Formal	Informal	Formal	Informal	Formal
2nd stage	main						
	male	0.496***	0.0945	0.507***	0.159*	0.5800***	0.1399
	urban	0.364***	0.213*	0.356***	0.299***	0.2145	0.8072***
	education	**0.0453***	**0.0614***	**0.0452***	**0.0894***	**0.0643***	**0.1531**
	age	0.0533***	0.120***	0.0548***	0.127***	0.0732***	0.1581**
	agesq	-0.0005***	-0.0012***	-0.0005**	-0.0013***	-0.0007***	-0.0016**
	apprentice	**0.0706**	**0.204**	**0.0719**	**0.122**	-0.1304	0.1301
	shortcourse	**0.0847**	**0.296***	**-0.0468**	**0.396****	0.1017	0.4964*
	literacycourse	**-0.158**	**-0.153**	**-0.192**	**-0.0954**	-0.3592	-0.1222
	literate/numerate	-0.0313	-0.0366	-0.062	-0.0037	0.0259	0.1204
	selection_ag					0.263	-1.123**
	selection_formal					0.7418*	0.3431
	selection_informal					-0.0415	0.4744
1st stage	male			-0.108**	0.318***	-0.3287***	0.2994***
	urban			1.010***	0.454***	2.097***	1.731***
	education			**-0.0557***	**0.157***	**0.0297***	**0.3130***
	age			0.0411***	0.150***	0.0585***	0.2858***
	agesq			-0.0005***	-0.0016***	-0.0007***	-0.0030***
	apprentice			**0.877***	**-0.534***	**1.487***	**-0.0574**
	shortcourse			**-0.548**	**0.719***	**0.053**	**1.313***
	literacycourse			**-0.354**	**0.0396**	**-0.4111**	**-0.0392**
	literate/numerate			0.190***	0.153*	0.0762	0.3447***
	hh_kids			-0.0193**	-0.0365***	-0.0384***	-0.0613***
	married			0.055	-0.0913*	-0.0837	-0.1453*
	Government grants					-0.00001**	-0.0000*
Ancillary statistics	N	3036	1833	13332	13943	12546	12546
	lambda			0.00397	0.252***		
	rho_ag					0.2736	-0.9281***
	rho_formal					0.7717**	0.2834
	rho3_informal					-0.0431	0.392

Table 17: Earnings functions; dependent variable is monthly wages from screening income in Log Naira. 1st stage is either probit equation for participation in formal or informal sectors (Binary) or a multinomial logit simultaneously estimated for the formal and informal sectors relative to agriculture (coefficients, not marginal effects); lambda indicates correlation between errors (the inverse mills ratio) in probit and 2nd stage while rho_terms are analagous for the MNL; selection_ terms are control terms for participation in each of the three sectors. Clustered standard errors in linear and Heckman models; adjusted for two-stage procedure in Heckman; bootstrapped in MNL; population weights used in OLS and Heckman; *** 0.1% ** 1% * 5%; state dummies included but not reported.

SELECTION

For the primary earnings regressions, indications are that sample selection issues are not affecting the second stage estimates. This can be gleaned from the insignificance of the lambda term and the analogous multinomial selection_ terms. However, the screening income specifications indicate some importance, especially involving the formal sector. Besides helping produce more accurate earnings functions, the selection equations are of direct interest.

[Type text]

Returning to Tables 16 and 17, the Heckman first stage probits indicate a positive relationship between education and the probability of entering the formal sector as well as a negative relationship between education and entering the informal sector. These are coefficients, not marginal effects. Education is also significant in the MNL specifications[18] but the coefficient signs do not necessarily give the direction of the relationship.[19] For this reason, and to capture some of the non-linearity in predicted probability models, we present figure 3, which plots the predicted probabilities for each sector against the years of education. It indicates that predicted participation in agriculture falls with education while predicted participation in the formal sector rises with education. For the informal sector, the predicted probability is relatively flat, with evidence of a slight fall for higher levels of education.
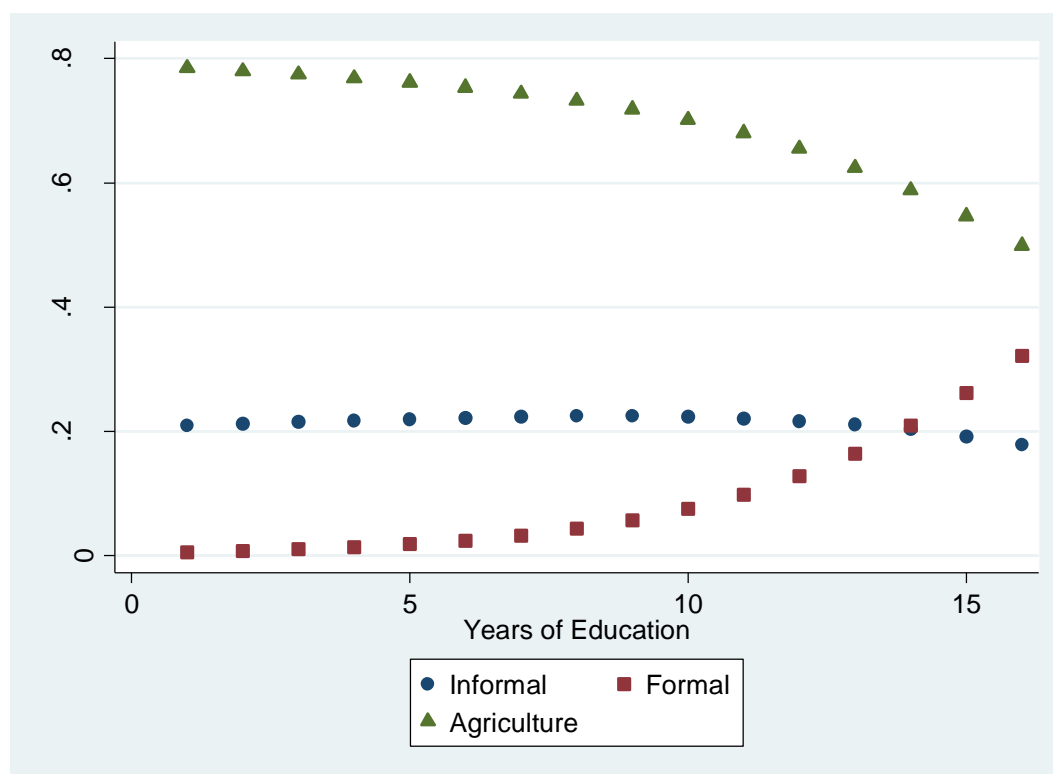

Figure 3: Predicted probabilities generated from multinomial logit; evaluated at means of other variables. Note: the three probabilities sum to 1.

As indicated in Tables 16 and 17, apprenticeship appears to be strongly positively associated with participation in the informal sector but negatively associated with participation in the formal sector. Table 18 shows significant marginal effects based on the multinomial logit. For example, getting an apprenticeship increases the predicted probability of informal sector participation by more than 30

---

[18] These are exactly the same in both tables.

[19] The reason for this is that, while education may have a slight positive effect on entering the informal sector relative to agriculture, it has a large positive effect on entering the formal sector relative to agriculture. Thus, overall, the probability of entering the informal sector relative to either the formal or agricultural sectors could feasibly fall.

[Type text]

percentage points, which is a lot. Part of this is at the expense of the formal sector, where the marginal effect is negative.

In contrast to apprenticeships, literacy or short courses appear to be means of getting into the formal sector instead of the informal or agricultural sectors. [20] The marginal effects for education correspond to the slopes of the lines in Figure 3 at just over 8 years of education.

	probability (formal)		probability (informal)	
variable	marginal effect	p value	marginal effect	p value
education	0.030	0	-0.010	0
apprentice*	-0.067	0	0.317	0
short course*	0.231	0.004	-0.147	0.093
literacy course*	0.013	0.575	-0.115	0.017

Table 18: Selected marginal effects taken from multinomial logit. Evaluated at means (eg about 8 years of education); * effect of discrete jump in dummy variable.

DISCRETE EDUCATION MEASURES

So far, we have only allowed for the years of education to enter linearly in the earnings and selection equations, but some forms of education may be more influential than others. As an alternative, and to aid future comparisons with other informal sector studies and work on Nigeria (Haywood & Teal on Nigeria, 2010 and Cojucaru on Tanzania, 2009), we present specifications where education attainment is measured in discrete jumps relative to no education. This measure is based on certifications rather than the level attended/completed.[21]

Tables 19 and 20 are like Tables 16 and 17 except for the alternative specification of education. Focusing on the formal sector coefficients first, we see that they get progressively higher as the level of attainment gets higher (except for vocational training). This means that returns are positive at almost all stages. Furthermore, the coefficient on post-secondary education is much higher than that on senior secondary or lower levels, which suggests some convexity in returns. This is apparent in the informal sector too, although not to the same extent as for the formal sector and not as consistently across

---

[20] We note that the size and significance of the marginal effects depends on the values of the explanatory variables chosen.

[21] Note that actual certification, as opposed to learning, may be additionally more important for the formal than informal sector.

specifications. [22] We also note that, although the post-secondary coefficient is consistently the highest, it is not always significant at 5%. [23]

		Primary OLS		Primary Binary		Primary MNL	
		Informal	Formal	Informal	Formal	Informal	Formal
2nd stage	Male	0.588***	0.115**	0.636***	0.122*	0.6167***	0.1467*
	Urban	0.114*	0.137**	-0.253	0.147*	-0.0701	0.2778**
	Primary	0.0367	0.106	0.0184	0.17	0.0212	0.227
	Junior Secondary	0.0461	0.247	-0.0171	0.307	-0.0154	0.4124
	Vocational	0.0572	-0.0624	-0.0507	0.237	-0.1118	0.3908
	Senior Secondary	0.107	0.211	0.0327	0.309	0.0321	0.4579
	Post-secondary	0.506***	0.643***	0.656***	0.752	0.4021	1.04*
	Age	0.0561***	0.0570***	0.0278	0.0660*	0.0417	0.0813**
	Agesq	-0.0005***	-0.0005**	-0.0002	-0.0006*	-0.0004	-0.0007**
	apprentice1	-0.0349	-0.055	-0.282*	-0.12	-0.0788	-0.1113
	shortcourse1	0.269	0.0424	0.283	0.0817	0.254	0.1618
	litcourse1	0.0855	-0.0243	-0.0537	-0.0396	-0.1214	-0.0446
	Litnum	0.0709	0.133	0.0853	0.144*	0.1068*	0.175*
	selection_ag					0.02	-0.4047
	selection_formal					-0.1875	0.064
	selection_informal					-0.1462	-0.0784
1st stage	Male			-0.121***	0.269***	-0.2536***	0.3573***
	Urban			0.980***	0.419***	2.104***	1.729***
	Primary			0.0299	0.460***	-0.0859	0.9841***
	Junior Secondary			0.0984	0.751***	0.1945	1.72***
	Vocational			-0.147	1.239***	0.1089	2.419***
	Senior Secondary			0.133*	1.049***	0.2955**	2.26***
	Post-secondary			-0.797***	2.197***	0.1128	4.174***
	Age			0.0618***	0.142***	0.0589***	0.2670***
	Agesq			-0.0008***	-0.0015***	-0.0008***	-0.0029***
	apprentice1			0.910***	-0.388***	1.499***	0.0773
	shortcourse1			-0.102	0.652***	0.225	1.340***
	litcourse1			-0.276**	-0.0378	-0.3908*	-0.1429
	Litnum			0.0811	0.214***	0.0883	0.4555***
	hh_kids			-0.0273***	-0.0185*	-0.0348***	-0.0585***
	Married			0.00074	-0.052	-0.089	-0.1936*
	Government grants					-0.00001*	-0.0000*
Ancillary statistics	N	2058	1360	11530	12367	11558	11558
	Lambda			-0.477	0.037		
	rho_ag					0.034	-0.7876
	rho_formal					-0.3186	0.1246
	rho3_informal					-0.2485*	-0.1525

Table 19: Earnings functions with discrete education attainment measures. Dependent variable is primary earnings.

---

[22] We also ran a specification with education completed and its square. This suggested strong convexity for the formal sector but, for the informal sector, both terms were close to zero and insignificant.
[23] One possible reason is difficulty in separately identifying all the attainment coefficients, which raises standard errors.

[Type text]

		Earnings OLS		Earnings Binary		Earnings MNL	
		Informal	Formal	Informal	Formal	Informal	Formal
2nd stage	Male	0.475***	0.115	0.490***	0.169*	0.5615***	0.201
	Urban	0.350***	0.235**	0.338***	0.304***	0.2196	0.7542***
	Primary	0.193*	-0.0817	0.287**	-0.0892	0.332***	0.2603
	Junior Secondary	0.283*	-0.0362	0.408**	-0.00781	0.4695***	0.321
	Vocational	0.706**	0.0619	0.740**	0.098	0.8067*	1.036
	Senior Secondary	0.435***	0.172	0.527***	0.263	0.5862***	0.9302*
	Post-secondary	0.642***	0.570**	0.700***	0.797***	1.017***	1.939***
	Age	0.0566***	0.110***	0.0587***	0.110***	0.07***	0.1456***
	Agesq	-0.0005***	-0.0011***	-0.0005**	-0.0011***	-0.0006**	-0.0015***
	apprentice1	0.0533	0.234	0.0562	0.191	-0.0923	0.1252
	shortcourse1	0.0488	0.314*	-0.0759	0.385**	0.0624	0.5826*
	litcourse1	-0.198	-0.225	-0.24	-0.179	-0.4057*	-0.1515
	Litnum	-0.0168	-0.0154	-0.071	0.0123	0.0197	0.1711
	selection_ag					0.0697	-1.363*
	selection_formal					0.5268	0.3002*
	selection_informal					-0.0884	-0.0781
1st stage	Male			-0.128***	0.322***	-0.2536***	0.3573***
	Urban			0.991***	0.452***	2.104***	1.729***
	Primary			0.0282	0.449***	-0.0859	0.9841***
	Junior Secondary			0.168	0.756***	0.1945	1.72***
	Vocational			0.256	1.059***	0.1089	2.42***
	Senior Secondary			0.041	1.030***	0.2955**	2.26***
	Post-secondary			-0.903***	2.077***	0.1128	4.174***
	Age			0.0525***	0.136***	0.0589***	0.269***
	Agesq			-0.0006***	-0.0014***	-0.0007***	-0.0029***
	apprentice1			0.813***	-0.470***	1.499***	0.0773
	shortcourse1			-0.454*	0.674***	0.225	1.340***
	litcourse1			-0.382***	-0.0122	-0.3908*	-0.1429
	Litnum			0.0791	0.215**	0.0883	0.4555***
	hh_kids			-0.0213**	-0.0348***	-0.0348***	-0.0585***
	Married			0.0666	-0.105*	-0.089	-0.1936*
	Government grants					-0.00001*	-0.0000*
Ancillary statistics	N	2872	1786	12273	12779	11558	11558
	Lambda			0.00237	0.203		
	rho_ag					0.07812	-1.072*
	rho_formal					0.5906	0.2361*
	rho3_informal					-0.0991	-0.0615

Table 20: Earnings functions with discrete education attainment measures. Dependent variable is screening income.

The coefficients for attainment are generally higher when the earnings measure is used (Table 20) rather than the primary activity measure (Table 21), as was the case for years of education. As before, it is also the case that the coefficients are generally higher for the formal sector and that, in general the implied returns are low.

[Type text]

The first stage regressions from the binary (Heckman) procedure indicate the main difference in the influence on participation occurs at the post-secondary level, where the coefficient is strongly negative for the informal sector and strongly positive for the formal sector.

To have a better understanding of participation predicted by the multinomial logit models, Table 21 presents marginal effects. Each marginal effect presents the impact on the probability of selecting the formal (or informal) sector after moving from no education to that level of attainment. Evaluating at the means, the indications are that all levels of education, even completed primary, reduce the probability of entering the informal sector; this is progressively the case except for vocational. For the formal sector, we see the impact on participation generally getting progressively larger, with the big effect from post-secondary education being suggestive of a convex effect on participation.

In a specification with a number of dummies, it can be difficult to interpret an effect evaluated at the means of all variables. As an alternative, we evaluate the effects where all dummies take on a value of zero (eg the person is female and lives in the rural sector) and we have assigned an age of 17 so that the person is young and potentially in formal education rather than of average age. The marginal effects are lower in absolute value in this case.

For the informal sector, the highest coefficients occur at the junior secondary and vocational level (9 or 10 years), which suggests the highest participation likelihood is at this level. Although this table allows for an additional source of non-linearity (the use of dummies rather than education in addition to the non-linearity of marginal effects), this is consistent with the evidence in Figure 3. In all cases, a post-secondary education is associated with lower participation in the informal sector. For the formal sector, the coefficients again suggest the biggest effect on the probability of entry comes from a post-secondary education. This is consistent with the figure even though we have allowed an additional source of convexity.

variable	Evaluated at means				Age 17; zero otherwise			
	Informal		Formal		Informal		Formal	
	ME	p value	ME	p value	ME	p value	ME	p value
Primary	-0.072	0.019	0.138	0.000	-0.002	0.013	0.006	0.018
Junior sec	-0.087	0.101	0.298	0.000	0.035	0.070	0.016	0.017
Vocational	-0.165	0.059	0.440	0.000	0.042	0.110	0.030	0.081
Senior sec	-0.122	0.002	0.351	0.000	0.023	0.049	0.027	0.007
Post-sec	-0.361	0.000	0.735	0.000	-0.010	0.016	0.140	0.002

Table 21: Marginal effects (ME) from Multinomial logit for selection into the formal or informal sectors.

[Type text]

7. SYNTHESIS: SKILLS AND EARNINGS IN THE FORMAL AND INFORMAL SECTORS

The large majority of non-agricultural workers are in the informal sector. This paper has produced evidence that, compared to the formal sector, informal sector workers tend to have less formal education, especially at the highest levels. Informal sector workers also tend to be more likely to have done an apprenticeship but less likely to have other formal courses or on-the-job training. Indications are that apprentices are overwhelmingly those with low education levels although there is not strong association between the years of education and apprenticeship. However, there is some evidence of complementarity between apprenticeship and basic numeracy/literacy as well as substitution between apprenticeship and with OJT.

On average, earnings are about 70% higher in the formal sector than the informal sector. The formal sector premium occurs throughout the distribution but there is some overlap, so some informal sector workers earn more than some formal sector workers. Furthermore, for men, there is evidence that formal earnings are close to or even below informal earnings at low levels of education. Given the higher propensity of educated people in the formal sector, much of the observed earnings premium can be attributed to an education composition effect for men.

While there is little evidence that other forms of skill acquisition affect earnings, the correlation between earnings and formal education is reproduced in a series of earnings functions. Estimated coefficients depend on the income/earnings measure used, but the coefficient is higher for the formal sector. This is consistent with the view that returns are higher in the formal sector but at least partly due to low formal sector earnings for those with low education levels. Furthermore, returns appear to be convex, especially in the formal sector, and much of the difference in returns between sectors occurs at post-secondary attainment.

Given the earnings functions and the composition effects, the formal sector premium seems to be at least partially accounted for by skills. Given the differences in skills profiles and the correlations between skills and earnings, and given that other demographic factors interact with earnings, sectoral choice and skill acquisition, we analyse the premium conditioning on these observable characteristics.

Table 22 provides regressions that are only for males and that pool the formal and informal sectors, but include a dummy for whether or not the sector is formal or informal, which is our main object of interest. We do not present results where no education is included at all, but these would reveal much higher dummies than those in the table. Regressing primary earnings by OLS, a comparison between columns 1 and 2 suggests that adding a higher order term for education lowers the predicted formal sector wage premium but that it remains.

For screening income, we see that, once one has some form of control for education, the premium is statistically insignificant (not even at 10%). When we use the Heckman approaches to control for selection into non-agricultural work, the premium for the primary earnings regression falls and, once we

[Type text]

add a quadratic for education, stops being significant at 10%. For income, the premium is about zero in both cases.

	OLS				Heckman			
	Primary Earnings		Income		Primary Earnings		Income	
Urban	0.117+	0.123+	0.292+	0.296+	0.102+	0.103+	0.246+	0.239+
Formal	**0.121+**	**0.102+**	**0.0897**	**0.076**	**0.0909+**	**0.0613**	**0.0072**	**-0.0262**
Education	0.0393+	-0.0809+	0.0553+	-0.0209	0.0390+	-0.0838+	0.0547+	-0.0267
Education^2		0.0058+		0.0037+		0.006+		0.004+
age	0.0792+	0.0709+	0.0818+	0.0772+	0.0776+	0.0687+	0.0772+	0.0712+
Age^2	-0.0008+	-0.0007+	-0.0008+	-0.0007+	-0.0008+	-0.0007+	-0.0007+	-0.0007+
apprentice	-0.0930+	-0.0821+	0.097	0.103	-0.106+	-0.0983+	0.0608	0.0592
shortcourse	0.1	0.0913	0.169	0.168	0.1	0.0913	0.171	0.171
litcourse	-0.0089	-0.016	-0.275+	-0.277+	-0.0069	-0.0135	-0.269+	-0.270+
litnum	0.0141	0.0866	-0.0791	-0.0378	0.0104	0.0832	-0.0896	-0.0481
_cons	6.965+	7.610+	5.633+	6.031+	7.053+	7.739+	5.878+	6.360+

Table 22: Pooled regressions across the formal and informal sectors but with a formal sector dummy; *males only*; non-agricultural workers only; Heckman estimates account for selection into non-agricultural work (only 2nd stage reported); + indicates significance at 10% or better.

The results indicate that accounting for education and other observables can reduce or remove the formal sector wage premium for males. These regressions are illustrative. After all, having just presented many regressions in which the returns to education in the formal and informal sectors differ, those in Table 22 assume they are the same.

Furthermore, we are not able to control for unobservable characteristics like ability or diligence,[24] so a significant formal sector dummy woulld not necessarily imply that the same person would earn more in the formal sector. As a corollary, because of the endogeneity of education to unobservable characteristics, a finding that the premium is removed once one controls for education does not necessarily merit the causal implication that, if you gave all current informal sector workers the education levels of formal sector workers, they would earn the same.

As indicated by the descriptive statistics and selection equations, it appears that a high (for example post-secondary) education is very helpful for acquiring formal sector employment. As noted, apprenticeships are associated with entry into the informal sector. One interpretation is that this is a means of establishing activity in the informal sector, but it is not clear whether these apprentices would prefer to stay in the formal sector after their employment if given the choice. Perhaps surveys should ask what the intention of the apprenticeship is.

More generally, while we generally see that the formal sector has higher earnings, that there are selection effects into formality for those with higher education levels and that observables can account

---

[24] See Falco and others (2009) for an attempt to analyse the effects of unobservables.

[Type text]

for much of the formal premium, we cannot with confidence ascertain the characteristics of informal sector workers who would prefer to be in the formal sector given the choice. Perhaps we should ask them.

REFERENCES

Adams (2008), A Framework for the Study of Skills Development in the Informal Sector of Sub-Saharan Africa

Aslam & Kingdon (2009), Public–private sector segmentation in the Pakistani labour market, Journal of Asian Economics

Asadullah (2009), Returns to Education in Bangladesh, Education Economics

Behar (2009), Directed technical change, the elasticity of substitution and wage inequality in developing countries, Department of Economics Discussion Paper

Bourguignon, Fournier & Gurgand (2007), SELECTION BIAS CORRECTIONS BASED ON THE MULTINOMIAL LOGIT MODEL: MONTE CARLO COMPARISONS, Journal of Economic Surveys

Card (1999), The causal effect of education on earnings, in Ashenfelter & Card (eds), Handbook of Labor Economics vol 3A, Elsevier

Cojucaru (2009), Tanzania: Skills and the Informal Sector

Durbin and McFadden (1984), An econometric analysis of residential electric appliance holdings and consumption, *Econometrica*

Kuepie, Nordam & Roubaud (2006), Education and Labour Market Outcomes in Sub-Saharan West Africa, DIAL Working Paper

Harris and Todaro (1970), Migration, unemployment and development: A two sector analysis, A*merican Economic Review*

Haywood (2007), The Performance of the Labor Market in Nigeria: A Report on the 2003–2004 Nigeria Living Standards Survey

Haywood (2008), The Performance of the Labor Market in Nigeria

 Haywood & Teal (2010), Employment, Unemployment, Joblessness, and Incomes in Nigeria, 1999–2006, in Treichel (ed), Putting Nigeria to Work, World Bank

Falco, Kerr, Rankin, Sandefur & Teal (2009), THE RETURNS TO FORMALITY AND INFORMALITY IN URBAN AFRICA, CSAE Working Paper

Fajnzylber, Maloney & Rojas (2006), Microenterprise dynamics in developing countries: How similar are they to those in the industrialized world? Evidence from Mexico, *World Bank Economic Review*

Fox & Gaal (2008), Working Out of Poverty

Heckman (1979), Sample selection bias as a specification error, *Econometrica*

Lee (1983), Generalized econometric models with selectivity, *Econometrica*

Schneider (2005), "Size and Measurement of the Informal Economy in 110 Countries around the World", Center for Research in Economics, Management and the Arts Working Paper

Treichel (2010), Employment and Growth in Nigeria, in Treichel (ed), Putting Nigeria to Work, World Bank

World Bank (2008), Ghana: Job Creation and Skills Development Draft Report, Vol. I, Main Document

[Type text]

ENDNOTES

---

[i] Earnings functions can be inconsistently estimated if one fails to account for selection because the characteristics which determine the decision to participate in either of the sectors may also affect earnings. This is a particularly important issue in developing countries, where we have seen a minority of people actually work in the formal or informal sectors. Thus, conditional on being in a certain sector, the expected value of the error term may not be zero but a function of the variables in the earnings function. This is the source of bias. A common approach, often used for the decision to participate in the labor market, is Heckman's 2-step approach, where the first stage estimates a probit model for participation and uses that predicted model to build a control variable (the Mills Ratio, denoted by Lambda in the tables) to account for the conditional expected value of the error term.

We estimate both stages jointly by maximum likelihood using Stata's *heckman* command unless the use of a two-step procedure is indicated. To aid identification, we specify two variables in the first stage that are not in the second. Following the literature (eg Aslam & Kingdom 2009; Asadullah, 2009), we used a dummy for whether or not the person is married (monogamous, polygamous or informal union) and constructed a measure for the number of children in the household. The intuition is that this affects the decision to seek non-agricultural employment but does not affect the earnings received once in it. To what extent this distinguished between the formal and informal sectors is not clear. While the validity of the exclusion restrictions is untestable, we draw comfort from the fact that neither was significant if included in the second stage. Further, we used the number of children under the age of 6 but this did not affect the results.

It may be more appropriate to explicitly recognise that there are three choices, namely formal, informal and agricultural. Lee (1983) developed a procedure analogous to that of Heckman (1979), but the first stage uses a multinomial logit to estimate the probability of each option relative to the others instead of a binary probit. Controls analogous to the Mills ratio are then included in the second stage, which is separately estimated for each option of interest, although the precise meaning of the controls depends on the nature of the method used. Similar methods have evolved since.

This method has received renewed attention since the work of Bourguignon, Fournier & Gurgand (2007), who compare alternative methods and who have developed *Stata* code for it. We implement this procedure using the *selmlog* add-in. Following the recommendations of their study, we use the *dmf(1)* option, which is their modification of the Durbin and McFadden (1984) approach. However, our results are robust to alternatives. When implementing the procedure, we added a third variable to the first stage, namely the receipt of government grants. Following Asadullah (2009), we examined a number of alternative sources of income (eg loans, asset sales), but this was the only income measure that was significant in the first stage. The use of the multinomial approach has a number of advantages, as discussed in Bourguignon et al, but we note that the standard errors, while bootstrapped (25 replications, which is probably not enough but doesn't make a difference in practice), do not explicitly allow for clustering. Furthermore, we do not have the facility to account for population weights.

[ii] As is well known, endogeneity of these variables can lead to inconsistent estimates and also prevent the attachment of causal interpretations. For education, this is typically thought to lead to overestimates, but the direction may be unclear for other forms. Furthermore, a number of studies attempting to instrument have lead to increased estimates (see for example Card, 1999). One plausible reason is measurement error, which has a downward effect on estimates when estimated by OLS. We do not attempt to control for this source of endogeneity because the dataset does not offer credible instruments for all of these variables. Nonetheless, potential endogeneity bias must be borne in mind.