Working paper

International Growth Centre

The Returns to Formality and Informality in Urban Africa

Paolo Falco
Andrew Kerr
Neil Rankin
Justin Sandefur
Francis Teal

November 2009

When citing this paper, please use the title and the following reference number:
F-2003-CCP-1

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid
from the British people

CSAE WPS/2010-03

THE RETURNS TO FORMALITY AND INFORMALITY IN URBAN AFRICA

Paolo Falco‡, Andrew Kerr‡, Neil Rankin†
Justin Sandefur‡ and Francis Teal‡.

November 2009

Abstract

This paper addresses the question as to why we observe such large differentials in earnings in urban African labour markets after controlling for observable human capital. We first use a three year panel across Ghana and Tanzania and find common patterns for both countries assuming that movement between occupations is exogenous. Unobserved individual market ability is by far the most important factor explaining the variance of earnings. Sector differences do matter even with controls for ability and the sectoral gap between private wage employment and civil servants is about 50 per cent, once we control for unobserved time-invariant factors. Wage earners earn the same as the self-employed in both Ghana and Tanzania. An additional important aspect of formality is enterprise size. At most half of the OLS effect of size on earnings can be explained by unobservable ability. Workers in largest firms are the high earners with wage rates which exceed those of civil servants. We then use an extension of the Ghana panel to five years to assess the extent of possible biases from the assumption of exogenous movement. We find evidence that this is important and that OLS may be understating the extent of both the size effect and the private sector wage (negative) premium. The implications of our results for understanding the nature of formal and informal employment in Africa are discussed.

The data used in this paper were collected by the Centre for the Study of African Economies, Oxford, in collaboration with the Ghana Statistical Office (GSO) and the Tanzania National Bureau of Statistics (NBS). The research, and the surveys on which it is based, has been funded by the Department for International Development (DfID) and the Economic and Social Research Council (ESRC) of the UK and by the IDRC in Canada. We are greatly indebted to numerous collaborators for enabling this data to be collected, particularly Emilian Karugendo and Trudy Owens in Tanzania, and Moses Awoonor-Williams, Geeta Kingdon and Andrew Zeitlin in Ghana. All errors are ours.
†School of Economic & Business Sciences, University of the Witwatersrand, South Africa
‡Centre for the Study of African Economies, Oxford University

1 Introduction

This paper investigates the role of formality and informality in determining earnings in urban Africa using individual panel data. While the importance of activities which take place in small scale, unregulated and untaxed enterprises has been widely recognised, how these activities link to self-employment and how both link to larger scale enterprises which are regulated and taxed has not been investigated with data that allows for the potential importance of skills, both observable and unobservable, in determining these outcomes. There has been an extensive discussion of how the informal sector should be defined and of its implications for poverty, initiated by the seminal ILO (1972) study of Kenya. Since then, informality has been a generic term given to the diverse range of economic activities that takes place among the small scale enterprises and the self-employed which dominate urban labour markets in Africa. The complexities in measuring incomes, outputs and employment when labour markets are not organised in the form of a formal wage contract are central to the analysis of duality in labour markets synthesised and extended by Sen (1975). The existence of the informal sector has been hypothesised to be the result of the regulations in the formal sector, see Schneider (2005) for this view and a cross-country analysis of the size of the informal sector.

The existence in tandem of an informal and a formal sector has been viewed as the result of labour market segmentation in which the formal sector sets wages above the market clearing rate, leaving either a large pool of unemployed or an informal sector oversupplied with labour at markedly lower wages rates; the classic model of this interpretation being Harris and Todaro (1970). This model was extended by Fields (1975) who introduced the notion of there being a "murky" sector in which all those who wish to work can and postulating a divide within the urban sector between this "murky" sector and the formal sector. In contrast, it is possible to view the informal sector as an efficient outcome of a labour market in which a technology intensive in unskilled labour exists alongside a formal sector using relatively skilled workers at much higher wages. Fajnzylber, Maloney and Rojas (2006) provide evidence that small scale self-employment is a preferred outcome and not the result of an inability to find a rationed formal sector job. Given the scarcity of skills in low income countries and the much wider range (a significant part of the working population has no education) the greater diversity of

economic activities than are observed in higher income economies is to be expected. The fact, which earnings functions clearly show, that workers in large firms or in the formal sector earn more, even with controls for observable skills in the form of education and work experience, simply shows the importance of unobservable skills in determining these labour market outcomes.[1]

Our objective in this paper is to advance understanding of which of these interpretations is most consistent with data for earnings across these sectors for urban labour markets in Ghana and Tanzania. One reason why explanations have proliferated much faster than testing is the problem posed by measuring incomes in the informal sector. If we seek to address the possibility that the outcomes we observe are due to unobservable skills by means of panel data we need to measure changes too. That essentially is the contribution of this paper.

This paper presents new survey data specifically designed to address this question. Like many existing data sets, the survey collects a wide range of socioeconomic data from workers in urban areas of two African countries: Ghana and Tanzania. The survey is novel in collecting income data on the self-employed in the urban informal sector. This data is intended to be comparable to wage data for urban employees. In both Ghana and Tanzania, the self-employed outnumber wage employees by at least 2-to-1 in the urban labor force. Understanding earnings determination in this sector, and how workers sort between sectors is thus central to any understanding of labor markets and income distribution in these countries.

The following section presents our theoretical framework, section 3 our data and section 4 our empirical results. Our empirical strategy for disentangling the effects of human capital and labor market segmentation acknowledges that workers may endogenously sort between occupations based on observable and unobservable skill differences. Our approach exploits potential instruments for occupation choice as well as

[1] Söderbom and Teal (2004) show that wages in manufacturing firms are clearly linked to size, Söderbom, Teal and Wambugu (2005) show for both Kenya and Ghana that the firm size effect for wages does not appear to be driven by unobservables, Kahyarara and Teal (2007) show for Tanzania the importance of size in determining the return to both academic and vocational education. As these studies are confined to wage employment they are uninformative of the determinants of size in explaining gaps between self and wage employment.

the panel dimension of our data set to investigate the extent of biases from both OLS and fixed effects approaches in section 5. A final section concludes.

2 Observable and Unobservable Skills

In the literature for developed countries there has been an extensive discussion initiated by Abowd, Kramarz and Margolis (1999) of the relative importance of unobserved individual characteristics on earnings relative to the factors associated with the firm in which the individual is employed. Their basic equation which we will adapt in this paper is:

$$y_{it} = x_{it}\beta + \theta_i + \psi_{J(i,t)} + \delta \cdot time + \varepsilon_{it} \quad (1)$$

where y_{it} is the natural log of earnings, x_{it} is a vector of time-varying observables, θ_i is the time-invariant individual characteristics, $\psi_{j,t}$ will be the sector j in which individual i works at time t. A time dummy captures all effects that are common at a point in time and ε_{it} is an error term. In Abowd, Kramarz and Margolis (1999) the term $\psi_{j,t}$captured the firm in which the worker is employed. In the case of urban labour markets in Africa most employment is not wage employment so we seek to identify movement across sectors where our data allows us to know if employment is in the civil service, a state or private enterprise, all these being forms of wage employment, and a final category for the self-employed. These four sectors capture some aspects of what has been meant by formal and informal enterprises. Indeed some models simply define wage employment as formal and see self-employment as a sector into which those who cannot find wage jobs end up working. We believe such an approach is to miss the fact that much wage employment is in small enterprises which can readily be characterised as "informal" and some form of self-employment activities, for example lawyers and accountants, are well paid. As we cannot usefully model the firm in which the individual works, we have chosen, in addition to modeling sector, to measure the size of the enterprise. We do this for both wage employees and for the self-employed by creating variables which measure the number of employees in the enterprise in which they work and thus form part of the time varying observable vector x_{it}. By controlling for size we are implicitly identifying the individual and the sector effect, conditioning on the size of the enterprise. Not only

does this enable us to model informality as a heterogeneous phenomenon related to enterprise size, it also allows for a set of controls which might be biasing what we can think of as the "pure" sector effect.

The identification strategy used by Abowd, Kramarz and Margolis (1999) is based on the assumption of exogenous movement. For the first part of this paper we will emulate their approach using our panel to identify the effects of time varying observables, which in our case include enterprise size, and following their procedure we will assess the relative importance of θ_i and $\psi_{J(i,t)}$ for the determination of earnings. We can write:

(2) $$\theta_i = \alpha_i + u_i\eta$$

where u_i is a vector of time invariant observables (in our model these will be controls for education and gender) and α_i captures the time invariant unobservables. The conclusion of Abowd, Kramarz and Margolis (1999) is that these α_i terms are by far the most important factor determining earnings. While more recent work has qualified the relative importance of such measures of market ability as a determinant of earnings, their importance is a common factor across all labour market datasets. Their potential correlation with the observables factors, such as size and sector, ensure that the estimates for these effects may be subject to serious bias. We derive an index of such ability defined as:

(3) $$Index_{ability} = \mathbf{D} \cdot \alpha$$

where **D** is an *N* x *N* matrix of indicators of individual $i = 1,2 \ldots N$.

We will estimate equation (1) by panel techniques, so in using fixed person effects it will not be possible to identify separately α_i and η. We use our estimates of the individual fixed effects as our dependent variable in (2) and identify α from the residuals of that equation. By imposing $\alpha = 0$ we scale our ability index in the same manner as Abowd, Kramarz and Margolis (1999). To identify θ_i we need to observe individuals over time.

On the other hand, to identify $\psi_{j,t}$ we need to observe workers move across sectors. Given that, we can use fixed effects results to calculate an index which summarises how important are sectors in determining earnings. We define such index as:

(4) $$Index_{\text{sec}\,tor} = \mathbf{F} \cdot \psi$$

where **F** is the *NT* x *J* matrix of indicators for the firm effect at which *i* works at date *t*.

Our assumption is that sector simply acts as a shifter of the earnings-profile while tenure effects are common across sectors and, therefore, do not feature in the sector index. The Abowd, Kramarz and Margolis (1999) approach is slightly more general, as seniority effects are included in ψ_{ji} They have :

(5) $$\psi_{ji} = \phi_j + \gamma_j s_{it}$$

(6) $$y_{it} = x_{it}\beta + \theta_i + \phi_j + \gamma_j s_{it} + \delta \cdot time + \varepsilon_{it}$$

When we tested for the impact of sector-specific seniority on earnings, we did not find significant differences. Hence, we assume that tenure has a homogeneous effect across sectors.

We still need to distinguish between general work experience which is in x_{it} and the within sector tenure (seniority) effect s_{it}. Our sector index, therefore, is defined as follows:

(7) $$\mathbf{F}\psi = \mathbf{F_0}\phi$$

where $\mathbf{F}_0$is the *NT* x *J* design matrix associated with the vector of sector specific intercepts ϕ is a *J* x *1* vector of sector-specific intercepts is the So:

(8) $$Index_{\text{sec}\,tor} = \mathbf{F}\psi = \mathbf{F_0}\phi$$

The answer to the question as to how important are time varying observables relative to person and sector effects can be found from a variance decomposition of the following equation:

(9) $$\mathbf{Y} = \beta\mathbf{X} + \eta\mathbf{U} + Index_{ability} + Index_{\text{sec}\,tor} + \delta \cdot time + \mathbf{E}$$

Equation (9) is our initial focus in this paper. The debate on informality in Africa has been essentially a debate as to the importance of the sector index relative to the ability index. To simplify the views of the alternative models: the Harris-Todaro (1970) model sees sector differences driving employment and unemployment outcomes while a human capital model of the world sees a sector differential as being an artifact of the failure to control for ability.

The above analysis is implicitly for earning rates, ie wage rates for wage employees or their hourly equivalent for the self-employed. Wage contracts often specify both earnings and hours. What we measure is earnings per period of time but we also have data on hours spent working. In our data the majority of work is self-employment and hours worked are both important to determine the earnings rate which can be

compared with the wage rate available from wage employment and likely to be measured with considerable error. Rather than compute hourly rates of earnings, which will then incorporate the errors from the hours data, we report the results with controls for hours. We interpret these regressions with controls for (log of) hours as the equivalent to the earnings functions that underlie the Abowd, Kramarz and Margolis (1999) analysis.

The next issue that needs to be addressed is selection into employment. A substantial part of the urban population aged 15 to 65 is classified both in our data, and in household surveys in Africa more generally, as out of the labour force. One interpretation of this finding is that these are individuals for whom either wage or self-employment jobs offer such low returns that they choose not to take them. Whatever the reason there is clear evidence that those in employment may be a selected sample. We will address this problem by assumption; namely that our ability to control for unobservables with the panel also controls for selection. This assumption does however change the interpretation of $\alpha_i$which is now not simply ability in the job but includes the (time-invariant and unobservable) ability to get a job.

Finally we need to recognize the key role that the assumption of exogenous movement has played in the model up this point. To interpret any of our results as causal, even with controls for unobservables, requires that individuals do not sort across occupations in the light of unobserved sector specific ability or that there are time varying unobservables which are correlated with the regressors of interest. In particular as Heckman, Lochner and Todd (2009) have forcibly argued if there is more than one ability "type" then it is possible that IV estimates will exceed any OLS estimate if the IV procedure succeeds in identifying the increase in earnings for an individual in a sector, relative to what would have been earned in the alternative sector *for that individual*. We return to this point in section 5 where we seek to allow for endogenous movement. To implement such a model needs panel data and the creation of the panel is described in the next section.

3 The Data

Our data is taken from a longitudinal labor market survey conducted by the Centre for the Study of African Economies (CSAE) at Oxford University, under the direction of the

authors and in collaboration with the Ghana Statistical Office (GSO) and the Tanzania National Bureau of Statistics (NBS). We refer to the data set as the Ghana and Tanzania Urban Panel Survey (UPS). The survey collects information on incomes, education and labor market experience, household characteristics and various other modules for labor force participants (ages 15 to 60) in urban areas. For Ghana these areas span the four largest urban centers in the country: Accra (and neighboring Tema), Kumasi, Takoradi and Cape Coast. In Tanzania, the sample covers several of the largest urban areas including Arusha, Dar es Salaam, Iringa, Morogoro, Mwanza, and Tanga.

The samples were based on a stratified random sample of urban households from the 2000 census in Ghana and the 2000 Household Budget Survey (HBS) in Tanzania.[2] While the initial sample was household based, interviews were conducted on an individual basis, and the unit of analysis in most of what follows will be at the individual level. A total of 830 and 543 individuals were interviewed in the first round of the survey in Ghana and Tanzania respectively, which was conducted between October 2003 and June 2004.

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

Collecting income data on the self-employed in low-income countries is a controversial endeavor. The problem is that self-employed business people in the informal sector rarely keep written accounts and their self-reported income data may be too noisy to be of use. For household based enterprises, the distinction between business and personal expenditures may be completely alien to respondents. We acknowledge the validity of these concerns.

[2] We should note that the analysis in this paper does *not* incorporate data from the Ghana Manufacturing Enterprise Survey (GMES). The UPS and the GMES are conducted in parallel with a common survey instrument. However, we restrict ourselves in this paper to the population based sample of the UPS, excluding the firm-based sample of manufacturing employees interviewed through the GMES.

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

A second unique feature of the UPS data set is its panel dimension. The individuals have been resurveyed for a period of three years in Tanzania and for five in Ghana. Thus the UPS constitutes one of very few household panel data sets in sub-Saharan Africa, and the only panel of income dynamics for the self-employed that we are aware of. In Table 1a and 2 we present the transition matrices for the two Ghana panels we will be using, that for the three years 2004-06 and the five years 2004-08. Table 1b presents similar data for the three year 2004-06 Tanzania panel. These matrices introduce the sectors that will be the focus of the analysis in the next two sections. Individuals are classified into five mutually exclusive groups, self-employed, private wage, civil servants, public enterprise and not working. The first four of these have earnings and it is these who will be used in the regressions reported below. Possibly surprisingly there is less movement in the five year than the three year Ghana panel. For this panel the largest change is within public enterprise which is also the category with the smallest numbers. The self-employed in both Ghana and Tanzania have the lowest mobility. The summary statistics on which the analysis will be based are given in the Appendix.

Table 3 presents the OLS results for each wave of the survey and for a pooled regression for both Ghana and Tanzania. The results confirm what numerous cross-section based studies have established namely that sectoral differences do appear to be an important determinant of earnings. Civil servants earn some 2.5 times what private wage

employees earn, *with a full set of controls for human capital*. Strictly this comparison is for wage employees in firms of size 1 as the equation also controls for firm size which has a large impact on earnings for both countries. In moving from a firm of size 1 to one of size 100 the regression predicts a 2-2.5 times rise in earnings, again this comparison is with the full set of controls for observable human capital. Two other features of the data stand out. The first is the strongly convex relationship between earnings and education, something widely found in developing country data sets and secondly that wage employees earn less than the self employed in firms of size unity showing the importance of not assuming that the self-employment sector is a free entry, low income, sector. As has previously been noted in Sandefur, Serneels and Teal (2006) a striking feature of this data is how great is the overlap in the distributions within these categories. Figure 1shows the distribution for earnings not only by wage and self-employment but by the size dimensions within each. This confirms that controls for observable human capital do not alter the ranking in terms of income where working in large firms and in self-employment with employees is the high return activity. The data for education confirms for both countries a monotonic increase in education from self-employment through small to large firms to the public sector. Also true for both countries is that younger workers are concentrated in small firms.

4 Measuring the Role of Unobservables

In the previous section we have presented a standard earning function and shown that it has the characteristics one would expect. We now proceed to ask how far any of the findings with respect to sectoral effects or size survive once we allow for the role of unobservables.

Table 4 presents a series of results which do control for time invariant unobservables. These are the first key results of the paper. All columns control for experience. In columns (1) and (2) we focus simply on the sectoral effects. As can be seen by a comparison of columns (3) and (4) with columns (1) and (2) controls for size are important. While these effects are much reduced from the OLS they are still large and significant in both countries. In both countries the dummy on civil servants is significant in columns (3) and (4) but the differential between these and private wage employees is

now only 38 per cent, down from a differential of 2.5 times. Before considering the interpretation of these results we complete the decomposition set out above. In Table 5 we decompose the fixed effect θ_i, obtained from the First Stage Regression in Table 3, into true unobserved heterogeity α_i and a part that can be associated with time invariant observables such as gender and education. Not surprisingly this re-creates the convex return on education we observed in the OLS earnings functions. Given the results from Table 5 we can extract from the data our Ability Index and we present the results in Figure 2. Both distributions are close to normal. We also create the Sectoral Index set out above and in Table 6 we bring together the ingredients of our analysis to address the relative importance of formality and informality in the determination of earnings.

Table 6 presents the second set of key results of the paper. The Table is set up so that the relative importance of time varying determinants of earnings in Columns (1) and (2) can be compared with the inclusion of time invariant gender and human capital measures in Columns (3) and (4). These results can in turn be compared to Columns (5) and (6) which include the Sectoral index and finally Columns (7) and (8) add the Ability Index. It is apparent that sectoral effects are far less important than individual effects in explaining the variance of earnings, although they are highly significant. The R^2 rises from 0.27 to 0.83 for Ghana and from 0.33 to 0.83 for Tanzania when we add the Ability Index to the equation. This result is very similar to that in Abowd, Kramarz and Margolis (1999) who find that these individual effects dominate earnings variance. It also is consistent with the argument of Mortensen (2005) that observable human capital explains relatively little of the variance of earnings.

For both countries we find that the size effects both within firms and among the self-employed remain large and significant even after controlling for sector and ability. We can see from Table 4 that in both Ghana and Tanzania the wage employees in small enterprises (strictly enterprises of size 1) do not earn significantly different amounts from the self-employed. The implication of the size coefficient is that those working in large firms (of more than 100) will earn on average more than the civil servants who have the highest sector differentials. In the next section we discuss how the results might be affected by movement and sorting across the occupations.

5 Endogenous Movement

The results in our second stage regression appear to confirm the importance of sector and size effects as determinants of earnings. The two most clear-cut findings are a gap between the private and the public sector, larger for Tanzania than Ghana but highly significant in both, and a highly significant firm size effect for both countries. In this section we address the question as to how robust are these results. Identification of the size and sectoral effects has been obtained by following the procedure of Abowd, Kramarz and Margolis (1999) and assuming the exogeneity of movement. We now relax this assumption and investigate possible endogeneity biases. Before presenting the results of these estimators we return to the potential sources of bias in our results raised in section 2 above. The exogeneity assumption we have used to identify the coefficients of interest does not allow endogenous sorting of workers across occupations. As pointed out by Heckman, Lochner and Todd (2009), in the context of interpreting the expected bias in the standard Mincerian earnings function, it is far from obvious that OLS estimates will be biased upwards. In this context the usual intuition would be that high ability individuals sort into wage employment assuming that wage employment is the preferred outcome and that ability is uni-dimensional. It assumes in the language of treatment effects that:

$$E(Y_0|W = 1) > E(Y_0|W = 0)$$

In words, the income for those treated if they had not been treated would have been higher than the actual outcome for the untreated. Heckman, Lochner and Todd (2009) argue that sorting may well reverse this inequality in which case OLS estimates will be biased down. In this context a high wage premium may be due to the fact that those good at being wage employees sort into that occupation and that the OLS estimate will be biased down as a result of such sorting.

Our results cannot be directly compared to the standard treatment literature as we are seeking to control for unobserved dimensions of the wage contract by means of a firm-size variable. The interpretations of the size effect discussed by Oi and Idson (1999) mainly focus on the role of unobserved skill in creating a size effect with the implication that fuller controls for skills will diminish any "size" effect. Our results are certainly consistent with that view as the firm size effect approximately halves when we use fixed

effects. However if size is proxying the unobserved "ability at being a wage worker across firms of different size" then the selection bias points in the opposite direction. Another problem is measurement error which will in general bias down the estimates in the fixed effects regressions. We need instruments to address these potential sources of endogeneity.

We proceed by exploiting the panel dimension of the data to use lags as instruments, as initially proposed by Arellano and Bond (1991). In their estimator the lags of the variables are used as instruments for the first differences (the DIFF-GMM estimator programmed in STATA). Our approach is to treat all the time-varying job-characteristics as *endogenous*, since they may be linked to unobservables determinants of movement. [3] This implies that in order to instrument the first-difference of these variables by their lags, one needs to go back at least 2 periods to find valid instruments. In addition to instrumenting the first differences of the regressors by means of their lags, the System GMM estimator as proposed by Blundell and Bond (1998), introduces a set of moment conditions based on the assumption that lagged differences of the regressors are independent from current unobservables. We present results for both estimators. As is well known the results of these estimators can be sensitive to the number of instruments so in order to assess how robust are the results to these different instrument procedures we present results making different assumptions about the instrument set.

The data requirements posed by GMM techniques in terms of lags is high. In order to ensure that our results are as robust as possible we use only the Ghana data where we have a five, rather than three, year panel. Our results up to this point have enabled us to make a direct comparison of Ghana and Tanzania. By confining ourselves to the longer panel data set for Ghana we can no longer make comparative statements but we can assess the potential importance of the biases that arise from the failure to allow for the biases due to movement.

Tables 7 and 8 present the results first for the differenced then for the system GMM estimator. As we are now using a five year panel we present in the first two

[3] The only exceptions are *Total Labour Market Experience* (which, by construction, is just a function of time) and *Common Time Effects* (captured by the time-dummies). These variables are treated as exogenous with respect to individual unobservables at time *t* and therefore treated as "classic instruments". One could view these variables as "instrumenting themselves"in the system GMM regressions.

columns the OLS and FE results which enable a comparison to be made for the Ghana results of the three year panel presented in Tables 3 and 4. We also present the results for the first differenced estimator (FD) to assess if any of our result are sensitive to whether we use this or the FE. It can be seen that there are some differences between the FE and FD but the broad pattern is very similar.

Our primary interest now is how our instrument sets affects our results. For both Tables 7 and 8 in Column (4) we use a set of instruments which includes all lags between periods *t-2* and *t-4*. In Column (5) we collapse the lags in the manner suggested by Roodman (2009). In Column (6) we confine the instruments to those for periods *t-2* and *t-3* only. Finally in Column (7) we restrict the instrument set still further and use only lags for *t-2*. For reasons which have been extensively discussed in the literature where variables are highly persistent the system GMM estimator should provide more robust results assuming that the data are stationary. Thus our preferred results are those in Table 8. The rather striking result for this Table is that for all the instrument sets with which we have experimented the IV estimate for the size effect is substantially higher than the OLS. It is also the case that the point estimate on the private dummy which has a magnitude of -0.354 in the OLS has a maximum (negative) value of -0.969 in our second instrument set. These instrumental variable estimates suggest that the OLS is understating the sector and size effect, not overstating them. While the differenced GMM results presented in Table 7 are less robust there is no evidence there either that OLS is upward biased.

5 Conclusions

We have used panel data from Ghana and Tanzania to address a question that has been central to how their urban labour markets work since the seminal model of Harris and Todaro (1970) on the links between formal wages and employment. A development of this framework by Fields (1975) postulated the importance, within the urban economy, of a distinction between a low income "murky" and a high income formal sector. By exploiting two sets of individual based panel data we have sought to adapt the framework provided by Abowd, Kramarz and Margolis (1999) to show the importance of sectoral effects relative to the unobserved aspects of ability in determining these differences in

incomes. We have then extended their procedure by instrumenting the variables using the estimators developed by Arellano and Bond (1991) and Blundell and Bond (1998).

Our results using the FE estimator and the IV point in opposite directions as to the signs of the bias in the OLS. Confining attention to our largest sample which is the Ghana five year panel the use of either FE or FD reduces, in some cases substantially, the point estimates for both the private wage (negative) premium and the size effect. In contrast the IV estimates increase both in absolute size. This latter result is consistent with the argument advanced by Heckman, Lochner and Todd (2009) that endogenous sorting is a key feature of these markets. If this is correct then the focus of Harris and Todaro (1970) and Fields (1975) on sectoral differences in income misses the quantitative importance of sorting by type within these markets.

A result which is common across both the FE and the IV estimators is that the size effect cannot be explained by time invariant skill unobservables. If the IV estimates have captured a key element of how dimension of firm size do impact on earnings then these effects are very large and suggest new insights into what drives the heterogeneity in labour incomes in these economies.

References

Abowd, J.M., Kramarz, F. And D.N. Margolis (1999) 'High wage workers and high wage firms', Econometrica, Vol. 67, No. 2, 251-333.

Arellano, M., and S. Bond (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", Review of Economic Studies, 58: 277-297

Arellano, M. and O. Bover. (1995), "Another look at instrumental variables estimation of error-components models", Journal of Econometrics, 68:29-51

Bardhan, P. and C. Udry (1999), "Development Microeconomics", Oxford University Press.

Bazzi, S. and M. Clemens (2009) "Blunt Instruments: On establishing the causes of economic growth" Working Paper Number 171 (May) Centre for Global Development.

Becker, G. S. (1964), "Human capital: A theoretical and empirical analysis, with special reference to education", First edition. Columbia University Press, New York.

Blundell, R. and S. Bond (1998). "Initial conditions and moment restrictions in dynamic panel data models", Journal of Econometrics, 87: 115-143

Fajnzylber, P., Maloney, W., & Rojas, G. (2006). Microenterprise dynamics in developing countries: How similar are they to those in the industrialized world? Evidence from Mexico. *World Bank Economic Review*, 20, 389-419.

Fields, G. (1975) "Rural-urban migration, urban unemployment and underemployment and job search activity in LDCs" *Journal of Development Economics*, 2, 165-188.

Fox, L. and M. Gaal (2008), "Working out of Poverty: Job creation and the quality of growth in Africa", The World Bank, Washington

Harris, J. and M. Todaro (1970) "Migration, unemployment and development: A two sector analysis", *American Economic Review*, 60 (1).

Heckman, J.J., L.J. Lochner and P.E. Todd, (2009) "Earnings functions, rates of return, and treatment effects: the Mincer Equation and Beyond", Chapter 7 in *Handbook of the Economics of Education*, vol. 1, 2009.ILO (1972) Employment, Income and Equity: A Strategy for Increasing Productive Employment in Kenya, 1972

Imbens, G. and J. Wooldridge (2009) "Recent developments in the econometrics of program evaluation", *Journal of Economic Literature*, 47;1, 5-86.

Jamal, V. and J. Weeks (1993) *Africa Misunderstood, or, Whatever happened to the rural-urban gap?* Macmillan Press, London.

Jolliffe, D. (1998): "Skills, Schooling, and Household Income in Ghana," *World Bank Economic Review*, 12(1), 81–104.

Kingdon, G., J. Sandefur, and F. Teal (2006): "Labour market flexibility, Wages and Incomes in Sub-Saharan Africa in the 1990s," *African Development Review*, Vol. 18, No. 3, December 2006.

Lewis, W.A (1954) "Economic development with unlimited supplies of labour", *The Manchester School*, 22, 141-45.

Mortensen, D.T. (2005) *Wage Dispersion: Why are similar workers paid differently?*, The MIT Press, Cambridge, Massachusetts.

Oi, W.Y. and Idson, T.L. (1999) "Firm size and wages". Chapter 33 in Handbook of Labor Economics, vol. 3, Sections 3-5.

Owens, T and F. Teal (2005) "Policies towards poverty: Ghana and Tanzania in the 1990s" at: http://www.gprg.org/themes/t1-pov-house-well/pol-pov-ghan-tanz/pol-pov-ghan-tanz-indepth-05.htm.

Psacharopoulos, G. (1994). 'Returns to investment in education: a global update,' *World Development*, Vol. 22, pp. 1325-1343.

Psacharopoulos, G. and Patrinos, H. A. (2002). 'Returns to investment in education: a further update,' World Bank Research Working Paper 2881, The World Bank.

Roodman, D. (2009) "A Note on the Theme of Too Many Instruments" *Oxford Bulletin of Economics and Statistics*, 71, 1, pp. 135-158.

Sandefur, J., P. Serneels and F. Teal (2006) "African poverty through the lens of labor economics: Earnings & mobility in three countries" Chapter 5 in P. Paci and P. Serneels (eds.) *Employment and shared Growth: rethinking the role of labor mobility for development*, The World bank, Washington, DC. at http://www.gprg.org/pubs/workingpapers/pdfs/gprg-wps-060.pdf.

Schneider, F. (2005) "Size and Measurement of the Informal Economy in 110 Countries around the World" Working Paper 2005-13, Center for Research in Economics, Management and the Arts, Johannes Kepler University of Linz.

Sen, A.K. (1966) "Peasants and dualism with or without surplus labour", *Journal of Political Economy*, 74.

Sen, A.K. (1975) *Employment, Technology and Development*, Oxford University Press.

Söderbom, M. and F. Teal (2004) "Size and efficiency in African manufacturing firms: Evidence from firm-level panel data", *Journal of Development Economics*, 73, pp. 369-394.

Söderbom, M., Teal, F. and A. Wambugu (2005) "Unobserved heterogeneity and the relation between earnings and firm size: Evidence from two developing countries" *Economic Letters* , 87, 153-159.

Söderbom, M., Teal, F., Wambugu, A. and G. Kahyarara (2006*)* "The dynamics of returns to education in Kenyan and Tanzanian manufacturing", *Oxford Bulletin of Economics and Statistics,* June, Vol. 68, No. 1, pp.261-396.

Table 1a: 1-Year Sector Transitions - Ghana 3 Ys Panel

	Self	Priv Wage	Civil	Pub Ent	Not Working	
Self	672	66	0	2	91	831
%	80.87	7.94	0	0.24	10.95	100
Priv Wage	47	262	8	5	46	368
%	12.77	71.2	2.17	1.36	12.5	100
Civil	2	5	28	3	7	45
%	4.44	11.11	62.22	6.67	15.56	100
Pub Ent	3	15	13	20	8	59
%	5.08	25.42	22.03	33.9	13.56	100
Not Working	123	112	10	3	395	643
%	19.13	17.42	1.56	0.47	61.43	100
Total	847	460	59	33	547	1,946
	43.53	23.64	3.03	1.7	28.11	100

Table 1b: 1-Year Sector Transitions - Tanzania 3 Ys Panel

	Self	Priv Wage	Civil	Pub Ent	Total
Self	403	19	4	2	428
%	94.16	4.44	0.93	0.47	100
Priv Wage	13	87	26	11	137
%	9.49	63.5	18.98	8.03	100
Civil	0	7	27	1	35
%	0	20	77.14	2.86	100
Pub Ent	2	9	8	18	37
%	5.41	24.32	21.62	48.65	100
Total	418	122	65	32	637
%	65.62	19.15	10.2	5.02	100

Table 2: 1-Year Sector Transitions - Ghana 5 Year Panel

	Self	Priv Wage	Civil	Pub Ent	Not Working	
Self	1,214	104	3	2	150	1,473
%	82.42	7.06	0.2	0.14	10.18	100
Priv Wage	93	607	16	9	102	827
%	11.25	73.4	1.93	1.09	12.33	100
Civil	3	11	68	4	10	96
%	3.13	11.46	70.83	4.17	10.42	100
Pub Ent	4	16	16	29	8	73
%	5.48	21.92	21.92	39.73	10.96	100
Not Working	184	170	17	6	920	1,297
%	14.19	13.11	1.31	0.46	70.93	100
Total	1,498	908	120	50	1,190	3,766
	39.78	24.11	3.19	1.33	31.6	100

Table 3: OLS Regressions by wave and by Country

	GH04	TZ04	GH05	TZ05	GH06	TZ06	GHPool	TZPool
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total Exp	.021	.024	.039	.014	.052	.041	.042	.027
	(.012)*	(.012)**	(.010)***	(.010)	(.010)***	(.013)***	(.007)***	(.008)***
$exp^2/100$	-.044	-.030	-.046	-.015	-.082	-.056	-.061	-.036
	(.025)*	(.021)	(.019)**	(.019)	(.021)***	(.022)**	(.014)***	(.013)***
Ln(Hours)	.396	.025	.355	.298	.095	.102	.130	.028
	(.126)***	(.117)	(.104)***	(.136)**	(.078)	(.121)	(.062)**	(.073)
Ln(firsiz)	.170	.180	.220	.168	.166	.127	.182	.149
	(.030)***	(.033)***	(.035)***	(.048)***	(.028)***	(.036)***	(.020)***	(.022)***
Ln(emps)	.431	.615	.170	.492	.279	.418	.267	.555
	(.112)***	(.123)***	(.065)***	(.135)***	(.112)**	(.130)***	(.057)***	(.076)***
Tenure	.030	-.006	.017	.017	.007	.007	.008	.006
	(.013)**	(.013)	(.010)*	(.012)	(.010)	(.008)	(.007)	(.007)
$Tenu^2/100$	-.061	.011	-.029	-.040	.021	-.014	.0007	-.012
	(.047)	(.033)	(.034)	(.037)	(.033)	(.022)	(.022)	(.020)
Male	.240	.217	.243	.299	.319	.356	.271	.313
	(.071)***	(.065)***	(.060)***	(.065)***	(.055)***	(.068)***	(.042)***	(.044)***
Educ	-.060	.0004	-.019	-.008	-.060	.037	-.042	.010
	(.023)***	(.024)	(.021)	(.026)	(.022)***	(.035)	(.015)***	(.019)
$Educ^2/100$	.601	.634	.518	.496	.804	.239	.626	.436
	(.157)***	(.183)***	(.149)***	(.172)***	(.150)***	(.239)	(.105)***	(.131)***
Pub Ent	.055	-.085	-.348	.193	-.462	.101	-.258	.112
	(.201)	(.215)	(.197)*	(.236)	(.198)**	(.186)	(.127)**	(.135)
Civ Ser	.687	.985	.646	.908	.520	.589	.635	.826
	(.109)***	(.170)***	(.120)***	(.125)***	(.108)***	(.100)***	(.077)***	(.073)***
Priv Wag	-.200	-.241	-.457	-.177	-.359	-.157	-.317	-.121
	(.112)*	(.124)*	(.111)***	(.116)	(.093)***	(.118)	(.066)***	(.074)
Const.	.208	9.626	-.087	8.735	1.181	9.293	.980	9.634
	(.484)	(.483)***	(.428)	(.569)***	(.321)***	(.511)***	(.251)***	(.316)***
Obs.	617	568	817	579	995	496	2429	1643
R^2	.231	.354	.29	.35	.242	.291	.232	.327

Dep.Var: Log of Real Monthly Earnings; Robust Standard Errors in parentheses; (Sig.Levels: * .10, **.05, ***.001)

Table 4: First Stage Fixed Effect Estimation

	GH1	TZ1	GH2	TZ2	GH3	TZ3	P-GH	P-TZ
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$exp^2/100$	.010	.009	.021	.015	.002	.015	.012	.020
	(.103)	(.112)	(.104)	(.112)	(.105)	(.116)	(.104)	(.114)
Ln(Hours)	.077	.114	.074	.085	.080	.087	.080	.090
	(.082)	(.091)	(.083)	(.090)	(.083)	(.090)	(.082)	(.090)
Tenure							-.012	.011
							(.013)	(.014)
$Tenu^2/100$							.054	-.031
							(.046)	(.037)
ten-self					-.023	-.006		
					(.026)	(.024)		
ten-privwage					.003	.020		
					(.028)	(.018)		
ten-civil					-.002	.138		
					(.034)	(.088)		
ten-pubent					-.016	.007		
					(.023)	(.074)		
tensq-self					.098	.019		
					(.088)	(.063)		
tensq-privwage					-.014	-.038		
					(.124)	(.045)		
tensq-civil					.041	-.634		
					(.141)	(.357)*		
tensq-pubent					.047	-.078		
					(.061)	(.253)		
Ln(firsiz)			.069	.087	.066	.083	.069	.086
			(.030)**	(.030)***	(.030)**	(.031)***	(.030)**	(.031)***
Ln(emps)			.146	.307	.145	.308	.145	.305
			(.074)*	(.085)***	(.075)*	(.085)***	(.075)*	(.085)***
Priv Wag	.040	.138	-.063	.022	-.093	-.045	-.064	.044
	(.098)	(.118)	(.112)	(.159)	(.146)	(.211)	(.114)	(.169)
Civ Ser	.175	.168	.285	.342	.231	.323	.275	.396
	(.148)	(.165)	(.154)*	(.153)**	(.189)	(.197)	(.158)*	(.188)**
Pub Ent	.238	.244	.066	.038	.090	.112	.055	.064
	(.144)*	(.163)	(.161)	(.226)	(.227)	(.212)	(.163)	(.243)
2005	.128	.118	.112	.121	.109	.136	.105	.121
	(.058)**	(.073)	(.058)*	(.072)*	(.059)*	(.076)*	(.059)*	(.073)*
2006	.361	.376	.346	.328	.352	.350	.344	.330
	(.093)***	(.126)***	(.093)***	(.124)***	(.094)***	(.130)***	(.094)***	(.126)***
Const.	1.920	10.185	1.840	10.175	1.956	10.189	1.887	10.062
	(.595)***	(.894)***	(.603)***	(.890)***	(.617)***	(.952)***	(.613)***	(.921)***
Obs.	2432	1653	2432	1652	2429	1643	2429	1643
e(N-g)	1321	961	1321	961	1320	954	1320	954
R^2	.076	.097	.083	.126	.085	.136	.084	.127

Dep.Var: Log of Real Monthly Earnings; Robust Standard Errors in parentheses; (Sig.Levels: * .10, **.05, ***.001)

Table 5: Regression of Total Individual Fixed Effect

	GH1	TZ1	GH2	TZ2	PREDICT-GH	PREDICT-TZ
	(1)	(2)	(3)	(4)	(5)	(6)
Educ	.006	.017	-.105	-.094	-.122	-.105
	(.002)**	(.003)***	(.010)***	(.011)***	(.011)***	(.012)***
$Educ^2/100$			1.021	1.151	1.063	1.194
			(.092)***	(.114)***	(.091)***	(.115)***
Male					.247	.109
					(.043)***	(.045)**
Obs.	1320	954	1320	954	1320	954
R^2	.004	.03	.09	.125	.112	.13

Dep.Var: $\hat{\theta}_i$ from First Stage (obtained using the "PREDICT" specification); (Sig.Levels: * .10, **.05, ***.001)

Table 6: Second Stage Regression with Indices

	GH1	TZ1	GH2	TZ2	GH3	TZ3	GH4	TZ4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total Exp	.041	.037	.045	.038	.042	.028	-.00002	-7.81e-11
	(.008)***	(.009)***	(.007)***	(.008)***	(.007)***	(.008)***	(.0008)	(.0009)
exp^2/100	-.070	-.064	-.066	-.047	-.062	-.036	.012	.020
	(.015)***	(.015)***	(.015)***	(.014)***	(.014)***	(.014)***	(.001)***	(.001)***
Ln(Hours)	.136	.074	.141	.053	.167	.073	.080	.090
	(.064)**	(.082)	(.060)**	(.076)	(.061)***	(.074)	(.032)**	(.028)***
Ln(firsiz)	.158	.155	.088	.096	.134	.109	.069	.086
	(.013)***	(.015)***	(.013)***	(.014)***	(.014)***	(.014)***	(.004)***	(.004)***
Ln(emps)	.337	.538	.260	.430	.286	.540	.145	.305
	(.055)***	(.077)***	(.055)***	(.078)***	(.055)***	(.078)***	(.027)***	(.028)***
Tenure	.020	.001	.018	.004	.019	.011	-.012	.011
	(.007)***	(.008)	(.007)***	(.007)	(.007)***	(.007)	(.002)***	(.002)***
Tenu^2/100	-.015	.017	-.021	-.002	-.027	-.024	.054	-.031
	(.024)	(.023)	(.023)	(.021)	(.022)	(.020)	(.005)***	(.004)***
Male			.265	.275	.257	.281	.249	.109
			(.042)***	(.046)***	(.042)***	(.044)***	(.005)***	(.004)***
Educ			-.055	.009	-.044	.010	-.122	-.105
			(.015)***	(.020)	(.015)***	(.019)	(.0005)***	(.0006)***
Educ^2/100			.802	.637	.637	.488	1.063	1.194
			(.102)***	(.142)***	(.106)***	(.136)***	(.006)***	(.007)***
Ab Ind							1.000	1.000
							(.002)***	(.002)***
Sec Ind					2.483	1.851	1.000	1.000
					(.244)***	(.193)***	(.069)***	(.078)***
2005	.048	.119	.061	.102	.065	.097	.105	.121
	(.040)	(.038)***	(.038)	(.038)***	(.038)*	(.037)***	(.028)***	(.029)***
2006	.281	.356	.301	.363	.310	.267	.344	.330
	(.042)***	(.049)***	(.041)***	(.046)***	(.040)***	(.046)***	(.029)***	(.028)***
Const.	1.061	9.810	.673	9.208	.643	9.250	1.887	10.062
	(.257)***	(.339)***	(.245)***	(.331)***	(.247)***	(.322)***	(.127)***	(.112)***
Obs.	2429	1643	2429	1643	2429	1643	2429	1643
R^2	.132	.142	.223	.297	.249	.335	.783	.851

Dep.Var: Log of Real Monthly Earnings; Robust Standard Errors in parentheses; (Sig.Levels: * .10, **.05, ***.001)

Table 7: Arellano and Bond - DIFF-GMM

	OLS	FE	FD	BB1a	BB1b	BB1c	BB1d
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	.299						
	(.041)***						
Educ	-.029						
	(.015)*						
Educ^2/100	.563						
	(.106)***						
Total Work Experience	.045						
	(.007)***						
exp^2/100	-.062	-.006	.011	.016	.099	.127	.090
	(.014)***	(.072)	(.088)	(.108)	(.138)	(.157)	(1.112)
Ln(Hours)	.265	.193	.085	-.349	-.338	.026	-.765
	(.052)***	(.057)***	(.039)**	(.265)	(.577)	(.509)	(2.118)
Tenure	.007	-.014	-.024	.035	.049	.088	.014
	(.007)	(.013)	(.013)*	(.056)	(.096)	(.127)	(1.825)
Tenu^2/100	-.005	.052	.082	-.153	-.152	-.347	.526
	(.023)	(.049)	(.052)	(.193)	(.332)	(.473)	(7.837)
Ln(firm size)	.183	.113	.078	.138	.433	.130	.286
	(.019)***	(.027)***	(.029)***	(.202)	(.222)*	(.392)	(3.549)
Ln(no. employees)	.271	.120	.113	.071	-.055	-.041	-1.321
	(.055)***	(.076)	(.053)**	(.363)	(.761)	(.948)	(12.523)
Private wage dummy	-.354	-.285	-.218	-.945	-1.185	-.167	-.359
	(.059)***	(.097)***	(.088)**	(.645)	(.767)	(.988)	(1.337)
civil	.558	.235	.130	.853	.955	1.168	2.925
	(.078)***	(.141)*	(.165)	(.655)	(1.497)	(1.635)	(27.383)
pubent	-.231	-.163	-.156	-.461	-1.631	-1.673	-.013
	(.119)*	(.159)	(.178)	(.898)	(.927)*	(1.855)	(7.934)
TimeDummies							
Const.	.245	1.616	.151				
	(.213)	(.427)***	(.056)***				
Obs.	3529	3529	1806	1806	1806	1806	1806
e(N-g)		1485		944	944	944	944
e(hansenp)				.317	.741	.668	.
e(j)				53	29	21	13
e(ar1p)				2.88e-10	2.32e-09	7.65e-07	.243
e(ar2p)				.584	.529	.428	.957

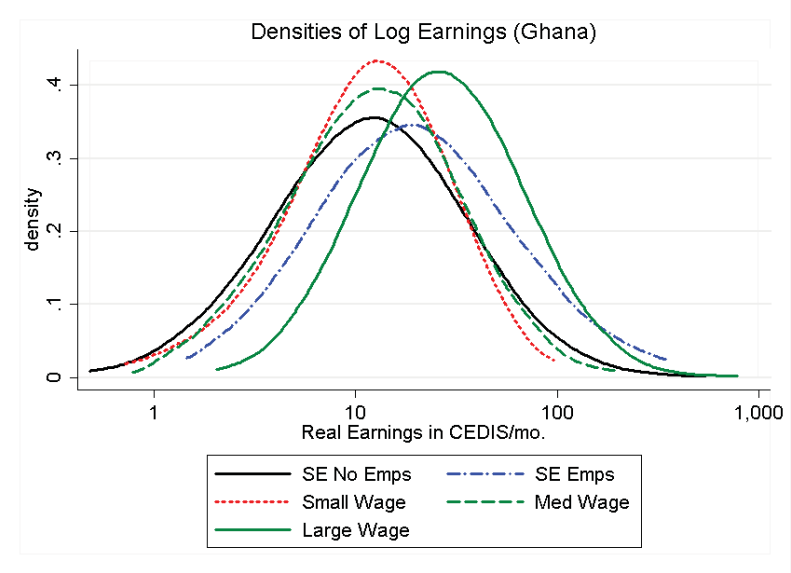
Dep.Var: Log of Real Monthly Earnings; All estimates are obtained via a Two-Step Procedure, where an optimal weighting matrix is estimated in the first stage and the Standard Errors are corrected using the Windmeijer methodology; N-g: Number of Included Individuals; hansenp: p-Value of Hansen Stat; j: Number of Instruments; ar$i$p: p-value of AR(i) statistic; (Sig.Levels: * .10, **.05, ***.001).

Table 8: Blundell and Bond - SYS-GMM

	OLS	FE	FD	BB1a	BB1b	BB1c	BB1d
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	.299						
	(.041)***						
Educ	-.029						
	(.015)*						
Educ^2/100	.563						
	(.106)***						
Total Work Experience	.045						
	(.007)***						
exp^2/100	-.062	-.006	.011	.018	.017	.019	-.013
	(.014)***	(.072)	(.088)	(.007)**	(.009)*	(.014)	(.034)
Ln(Hours)	.265	.193	.085	.196	.476	.348	-.991
	(.052)***	(.057)***	(.039)**	(.187)	(.218)**	(.485)	(1.011)
Tenure	.007	-.014	-.024	-.012	-.026	-.047	-.045
	(.007)	(.013)	(.013)*	(.024)	(.027)	(.050)	(.075)
Tenu^2/100	-.005	.052	.082	.043	.082	.157	.281
	(.023)	(.049)	(.052)	(.087)	(.105)	(.174)	(.240)
Ln(firm size)	.183	.113	.078	.460	.443	.325	.455
	(.019)***	(.027)***	(.029)***	(.104)***	(.119)***	(.185)*	(.264)*
Ln(no. employees)	.271	.120	.113	.378	.500	-.491	.098
	(.055)***	(.076)	(.053)**	(.156)**	(.189)***	(.685)	(1.110)
Private wage dummy	-.354	-.285	-.218	-.971	-.969	-.821	-.759
	(.059)***	(.097)***	(.088)**	(.308)***	(.339)***	(.593)	(.760)
civil	.558	.235	.130	.706	.699	.513	1.943
	(.078)***	(.141)*	(.165)	(.223)***	(.194)***	(.577)	(2.286)
pubent	-.231	-.163	-.156	-1.624	-1.311	-2.500	-2.068
	(.119)*	(.159)	(.178)	(.520)***	(.579)**	(1.604)	(2.363)
TimeDummies							
Const.	.245	1.616	.151	1.442	.448	1.197	5.968
	(.213)	(.427)***	(.056)***	(.721)**	(.866)	(1.799)	(3.639)
Obs.	3529	3529	1806	3529	3529	3529	3529
e(N-g)		1485		1485	1485	1485	1485
e(hansenp)				.369	.249	.865	.998
e(j)				78	38	30	22
e(ar1p)				1.75e-12	1.25e-10	3.66e-07	.012
e(ar2p)				.389	.358	.578	.816

Dep.Var: Log of Real Monthly Earnings; All estimates are obtained via a Two-Step Procedure, where an optimal weighting matrix is estimated in the first stage and the Standard Errors are corrected using the Windmeijer methodology; N-g: Number of Included Individuals; hansenp: p-Value of Hansen Stat; j: Number of Instruments; ar$i$p: p-value of AR(i) statistic; (Sig.Levels: * .10, **.05, ***.001).

Figure 1: Densities of Log-Earnings for Ghana and Tanzania



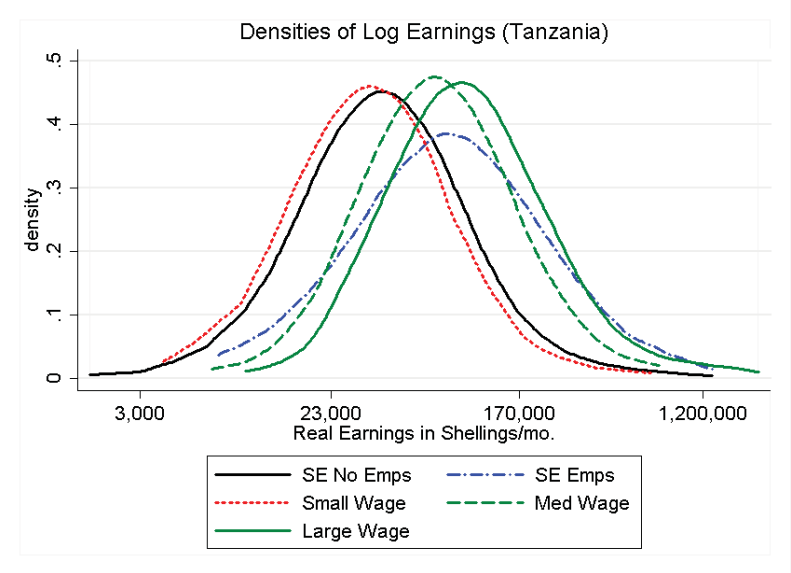
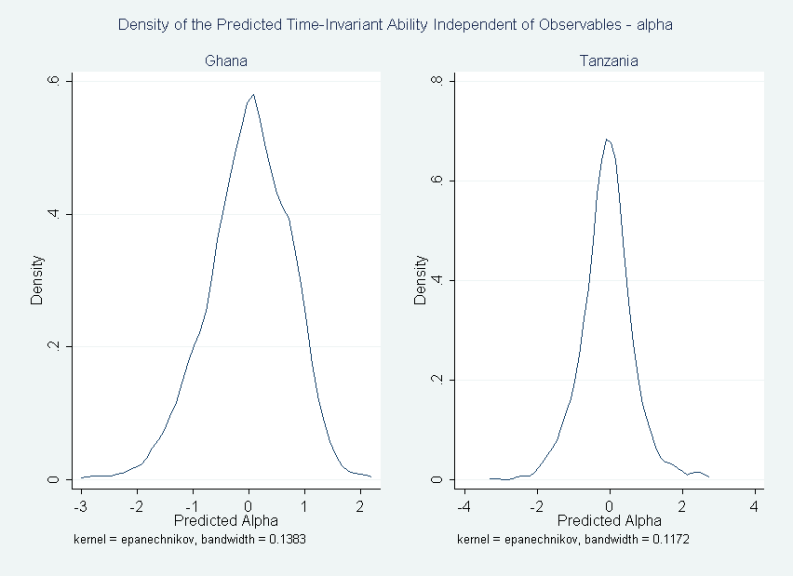


Figure 2: Predicted Time-Invariant Ability (independent of observables) - α_i



APPENDIX

Table 1: Summary Stats for entire 3wave sample GHANA

Variable	Mean	Std. Dev.	Min.	Max.
Male	0.447	0.497	0	1
Age (years)	34.727	10.481	16.083	64.233
Age$ˆ2/100$	13.157	7.891	2.587	41.259
Total Exp	20.284	11.431	0	57.417
exp$ˆ2/100$	5.42	5.689	0	32.967
Educ	8.444	3.896	0	19.5
Educ$ˆ2/100$	0.865	0.555	0	3.803
Real Monthly Earnings	17.848	17.591	0.481	207.9
Log Earnings	2.488	0.930	-0.731	5.337
hours	47.707	12.497	2	80
Ln(Hours)	3.819	0.349	0.693	4.382
Tenure	8.752	9.212	0	45.58
Tenu$ˆ2/100$	1.614	2.708	0	20.775
employees	1.2	0.854	1	12
Ln(emps)	0.092	0.334	0	2.485
firmsize	13.424	34.711	1	150
Ln(firsiz)	0.887	1.515	0	5.011
self	0.608	0.488	0	1
Priv Wag	0.315	0.464	0	1
Public wage	0.077	0.267	0	1
2004	0.254	0.435	0	1
2005	0.336	0.473	0	1
2006	0.41	0.492	0	1
Civ Ser	0.044	0.204	0	1
Pub Ent	0.033	0.18	0	1
Country	1	0	1	1
N		2429		

Table 2: Summary Stats for entire 3wave sample TANZANIA

Variable	Mean	Std. Dev.	Min.	Max.
Male	0.502	0.5	0	1
Age (years)	38.263	11.184	16	65
Age$ˆ2/100$	15.89	9.192	2.56	42.25
Total Exp	25.2	12.186	2	59
exp$ˆ2/100$	7.834	7.149	0.04	34.81
Educ	7.063	3.748	0	17
Educ$ˆ2/100$	0.639	0.504	0	2.89
Real Monthly Earnings	87146.599	127990.06	1771.48	2170061.75
Log Earnings	10.913	0.915	7.48	14.59
hours	52.537	14.91	12	120
Ln(Hours)	3.919	0.3	2.485	4.787
Tenure	9.920	9.263	0	46
Tenu$ˆ2/100$	1.842	3.267	0	21.16
employees	1.219	0.776	1	11
Ln(emps)	0.114	0.337	0	2.398
firmsize	14.901	38.038	1	150
Ln(firsiz)	0.809	1.565	0	5.011
self	0.637	0.481	0	1
Priv Wag	0.237	0.425	0	1
Public wage	0.126	0.332	0	1
2004	0.346	0.476	0	1
2005	0.352	0.478	0	1
2006	0.302	0.459	0	1
Civ Ser	0.083	0.277	0	1
Pub Ent	0.043	0.202	0	1
Country	2	0	2	2
N		1643		

Table 3 - Median Monthly Incomes in Nominal and Real Domestic Currency and in USD

		Self no Employees	Self with employees	Public	Small firm <20	Large firm >20
Nominal	GH04 cedis	32	60	83	30	71
	TZ04 TSH	40000	80000	112500	40000	98000
	GH06 cedis	63	83	150	45	88
	TZ06 TSH	62500	91666	150000	65000	80000
Real	GH04 cedis	9	17	24	9	20
	TZ04 TSH	36934	73868	103878	36934	90489
	GH06 cedis	14	19	35	10	20
	TZ06 TSH	51824	76008	124378	53897	66334
USD	GH 04	36	66	91	33	78
	TZ 04	37	73	103	36	90
	GH 06	68	90	162	49	95
	TZ 06	50	73	120	52	64

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

IGC
International
Growth Centre

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid
from the British people

Designed by soapbox.co.uk