

Haywood, Luke; Falco, Paolo

Conference Paper

Entrepreneurship versus Joblessness - Explaining the Rise in Self-Employment

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Occupational Choice and Self-Employment, No. F11-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Haywood, Luke; Falco, Paolo (2013) : Entrepreneurship versus Joblessness - Explaining the Rise in Self-Employment, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Occupational Choice and Self-Employment, No. F11-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/80016>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Entrepreneurship versus Joblessness: Explaining the rise in Self-Employment*

Luke Haywood[†] and Paolo Falco[‡]

10th December 2012

Work in Progress

Abstract

The self-employed constitute a large proportion of the workforce in developing countries and the sector has been found to be growing further. Different accounts exist as to the cause of this development, with pull factors such as high returns to capital and increased wealth contrasted with push factors such as barriers to entry into the wage sectors following traditional segmented labour market models. This article considers changes in the structure of earnings for the self-employed in Ghana and compares them with the wage employed. Models of segmented labour markets typically consider sorting on unobservables to be important, and often posit a sector choice model. If there are barriers to entry into one of the sectors, however, selection on unobservables there may be no clear selection rule. We apply a simple model of a two-sector labour market and estimate earnings using a correlated random coefficients model that allows for multiple patterns of sorting and selection on unobservables using instrumental variables GMM. We find evidence of increasing return to productive characteristics for the self-employed, but also a large wage premium.

*This paper uses data from the six rounds of the Ghana Urban Household Panel Survey, conducted by the Centre for the Study of African Economies (CSAE). The dataset forms part of ongoing CSAE research into urban African labour markets funded by the ESRC, RECOUP, IDRC, DFID and the Gates Foundation. We are greatly indebted to Moses Awoonor-Williams and members of the Ghana Statistical Office, who assisted in the data collection. The paper has benefited from comments and discussion by Andrew Clark, Marc Gurgand, Clément Imbert, David Margolis, Jean-Marc Robin, Francis Teal and seminar and conference participants in Paris and Marseille. The usual disclaimer applies.

[†]DIW Berlin & Paris School of Economics, *E-mail* : haywood@pse.ens.fr

[‡]University of Oxford, *E-mail* : paolo.falco@economics.ox.ac.uk

1 Introduction

West-African labour markets are no exception in being dominated by self-employment in various forms: smallholder farming, retail businesses, mechanics and petty traders are most often self-employed. Furthermore, recent large-scale household data have confirmed an increase in self-employment both in rural but especially urban contexts (see [Kingdon et al. \(2006\)](#)).

An optimistic interpretation of this development notes that returns to capital in countries with many low-paid workers are typically high. More individuals becoming self-employed may then be a result of increased wealth and opportunities in combination with binding capital constraints on the establishment of self-employed businesses ([Evans and Jovanovic \(1989\)](#), [Magnac and Robin \(1996\)](#) and [Blanchflower and Oswald \(1998\)](#)). Self-employment may be a road to higher employment, reduced inequality in earnings and better working conditions. Self-employment under this perspective is often interpreted as entrepreneurship, i.e. the establishment of a business transforming capital and labour into output.

A more pessimistic perspective is that for many individuals, wage work is not available and in absence of social protection or family transfers, self-employment may be the only means of survival. In this sense, part of the self-employment can be interpreted as the equivalent of unemployment in countries without social welfare systems. In fact, some of the urban self-employed may be viewed as forming a queue for wage employment in a Harris-Todaro-type setting ([Harris and Todaro \(1970\)](#)). This would imply that even if we were able to control for both observable and unobservable characteristics of workers in different sectors, we should observe a wage premium. The pessimistic view notes that in developing countries many self-employed operate with little to no capital, whilst wage jobs, often formal and in the public sector, are viewed as hard to access. Self-employment may not constitute the preferred option in the presence of barriers to entry for wage jobs.

We investigate the existence of unexplained pay differentials across sectors as well as more generally the evolution of the structure of earnings in Ghana as an element in understanding the causes of the increase in self-employment. We find evidence of a wage premium but also of higher and increasing returns to observable and unobservable characteristics in the self-employed sector. Our results suggest both barriers to entry into wage employment (push) and increasing returns (pull) as causes of the increase in self-employment.

Similar questions concerning the nature of self-employment have been asked

in the literature. Recently, [Poschke \(2010\)](#) uses Global Entrepreneurship Monitor data about declared reasons for self-employment. In particular, individuals state whether the reason for their self-employment is voluntary or not (the latter named “necessity self-employment”). He finds that the characteristics of the necessity self-employed workers are specific - they typically own small businesses, are less educated and more likely to be female. Necessity self-employment is especially common in non-OECD countries, rising to up to 50% in some areas.

[Launov and Guenther \(2012\)](#) estimate a model with latent probabilities of being in the “opportunity” or “last resort” sector of informal employment. If the reason for the increase in self-employment in Ghana is largely due to motivations of “last resort”, we might expect levels of capital, schooling and unobserved ability to be declining for the self-employed. Accompanying this emphasis on a *push* into self-employment we might then expect to find a significant (possibly increasing) premium for wage employment. Alternative and more optimistic *pull* factors for increasing self-employment would include increasing returns to human or physical capital in the self-employed sector compared to the wage sector or a decreasing wage premium.

Explaining which factors can account for the rise in self-employment in developing countries appears crucial to assess the implications of this trend and to make informed policy decisions.

To test these competing views this article takes unobservables seriously, and in particular, sorting or selection on unobservables. Panel fixed effects are insufficient to control for individual unobservables if their effect varies across sectors. Furthermore, we do not wish to make the assumption - standard in selection models - that there are variables which monotonously increase the likelihood of selection into one of the sectors, the so-called “single-crossing property”. Many wage jobs are in the public sector where market forces may compete with other selection mechanisms (e.g. based on social capital). Similarly, entry costs in the self-employed sector may vary in ways related to unobservables, for example by way of differences in access to credit.

The paper proceeds as follows. Section (2) describes the context of the Ghanaian labour market. In section (3) we present the basic model with a special focus on selection issues. Section (4) considers identification and presents an estimation strategy using the correlated random coefficients model. Section (5) presents the data. We then discuss results showing an increase in returns to productive characteristics in the self-employed sector, a corresponding change in composition of the

workforce as well as a rising wage premium in section (6).

2 Labour Market in Ghana

Ghana is one of the most stable countries in West Africa and has shown fairly strong economic growth rates over the last decade. In fact, [Nsowah-Nuamah et al. \(2012\)](#) find that poverty was halved over the period from 1991 to 2005. They also find that this occurred whilst employment in the public sector fell and employment in small firms increased. Indeed, next to the traditionally important role of the agricultural sector, [Robson et al. \(2009\)](#) cite evidence that in Ghana “[t]he economic structure is polarised between a small number of large corporations and large volume of micro and small enterprises”. Large corporations are active, amongst others, in mining (gold, bauxite), oil exploration (significant oilfields were discovered in 2007) and timber and derived products. Whilst agriculture contributes to the largest proportion of GDP (40.4%) in 2005, services account for 32.4% and industry for 27.7%. Self-employed businesses are active in all of these fields.

Table 1: Changes in the composition of the Ghanaian Labour market

Type of Employment	1991/92	1998/99	% change
Public Wage Employment	9.1	6.2	- 2.9
Private Formal Wage Employment	2.3	1.4	- 0.9
Informal Wage Employment	2.3	1.9	- 0.4
Export Farmer	7.8	6.9	- 0.9
Food Crop Farmer	57.3	58.1	+ 0.8
Non-farm self-employed	20.5	24.5	+ 4.0
Others and Non-working	0.7	1.1	+ 0.4

Source: Table (2) in Baah-Boateng (2004)

Reporting evidence from the Ghana Statistical Service, [Baah-Boateng \(2004\)](#) finds evidence of rising rates of self-employment for the period until 1999 (see table (1)), especially non-farm self-employment, which is what our urban sample will cover. [Palmer \(2007\)](#) notes that “[m]ost new jobs are created in the informal economy, with formal sector employment growth largely stagnant.” [Gollin \(1995\)](#) shows evidence that one of the reasons for the importance of the self-employed sector may be a result of differential tax rules by firm size - most self-employed businesses being rather small.

There is conflicting evidence on the more recent development of the share of self-employed in the economy, with the possibility of a trend reversal¹. This article focusses on the relative earnings performance of two of the most important types of economic activity in Ghana and their determinants. Census data are better suited to investigate the relative proportions of these two sectors. We here exploit the comparative advantage of our panel dataset in tracing trends in the structure of earnings.

3 The Model

In this section we describe the simple model with which we attempt to capture some of the features of the wage structure in two important sectors of the Ghanaian labour market - self-employment and wage employment. In line with occupational and sector choice models since Roy (1951), we take into account the fact that unobservable factors may importantly determine sectoral preferences. The situation is one in which selection on unobservables will occur, but the single-crossing property may not hold.

3.1 Sector Earnings

Let worker i be endowed with time-varying characteristics (in particular, physical capital $K_{i,t}$), time-invariant characteristics (e.g. human capital h_i) and unobservable productivity θ_i^j for $j \in \{SE, w\}$ varying across sectors self-employment (SE) and wage employment (w). Our definition of physical capital encompasses both liquid savings and assets (under the assumption that the latter can easily be sold and re-invested). We take physical and human capital as given and do not consider the interactions between sector choice and capital accumulation². Given these endowments, worker i optimally chooses between two alternative *employment strategies*: working as a self-employed or searching for a wage-job.

For employed workers we observe wages and worker and job characteristics. Earnings in wage (w) and self-employment (SE) can be given in fairly general

¹ See table (1) in Nsowah-Nuamah et al. (2012)

² One could endogenize physical capital accumulation in line with Magnac and Robin (1996) and human capital in line with Keane and Wolpin (1997), whereby putting a structure on individuals' expected ease of entry into the wage sector will be crucial in our set-up of entrance barriers, since the selection process of wage employment is not observed.

terms using the following log-linearised earnings equations:

$$R_{i,t}^{SE} = \alpha^{SE} h_i + \beta^{SE} K_{i,t} + \delta_t^{SE} + \theta_i^{SE} + u_{i,t}^{SE} \quad (1)$$

$$R_{i,t}^w = \alpha^w h_i + \beta^w K_{i,t} + \delta_t^w + \theta_i^w + u_{i,t}^w \quad (2)$$

where $K_{i,t}$ and h_i indicate physical and human capital; δ_t^{SE} designates macroeconomic effects on the self employed sector (viz. $\delta_t^w - \delta_t^{SE}$ the time-changing wage employment premium) and $u_{i,t}$ subsumes individual idiosyncratic factors common across across sectors. We can think of β^w as the market rate of return on savings (e.g. interest rate on bank deposits). Empirical studies consistently find that access to capital is an important determinant of self-employment, indicating that many “latent entrepreneurs” (in the words of [Blanchflower et al. \(2001\)](#)) are credit constrained. We have measures of capital and estimate returns to both sectors.

What if the employment strategy of searching for a wage job does not lead to individuals finding a job? We do not directly consider unemployment, but we take into account one way in which it might influence the structure of wages.

3.2 Selection and Barriers to Entry

Our model depicts a dual labour market, with on the one side large private firms and the public sector, i.e. ‘formal employers’ and the self-employed on the other side (typically micro/small private enterprises). This general situation has been analysed in the literature using variants of the classic Roy model, based on specialisation by comparative advantage. However, we would like to avoid the assumption of free sector choice underlying most of the literature. Access to formal jobs may be rationed in the sense that the number of workers who are willing to work for the equilibrium wage (at a given skill level) exceeds the number of available jobs. Indeed, if there are job queues for entry into wage employment (as eg. the Harris-Todaro framework suggests), we need to take into account not only self-selection by workers, but selection of workers by firms. It is however unclear how firms choose individuals - for example, to what extent firms can observe individual unobserved sector-specific performance ($\theta_i^w, \theta_i^{SE}$). Non-productive factors may also play a role in firms’ selection choices, e.g. for allocation in the public sector (an important wage job employer in many developing countries³). Rationing could be the result of efficiency wage setting, institutional constraints or result from infor-

³The public sector accounts for around 20% of wage employment in our data, see table (2).

mational frictions⁴.

It is convenient to consider the constrained sector assignment as a form of waiting costs resulting from barriers to entry: individuals face waiting periods before being able to exercise in their chosen sector. Magnac (1991) models barriers to entry as costs $c(\cdot)$ resulting from queueing with a success chance in every period of τ of remaining unemployed. This shows that we can expect $c(\cdot)$ to be a function of all the determinants of wages (x_i):

$$\begin{aligned}\pi^w(x_i) &= (1 - \tau) R^w(x_i) + \tau 0 \\ &= R^w(x_i) - \tau R^w(x_i) \\ &= R^w(x_i) - c(x_i)\end{aligned}\tag{3}$$

We thus contend that individuals self-select subject to two constraints: They must choose employment strategies based on expected earnings in the two sectors (i.e. with knowledge of the determinants of earnings, but subject to stochastic variation) and they are faced with (potentially individual-specific) entry costs. We can then think of workers' choice of becoming self-employed (d^{SE}) as follows⁵:

$$\begin{aligned}Pr(d_i^{SE}) &= Pr(E(R^{SE}) > E(R^w) - c) \\ &= Pr(E(R^{SE} - R^w) - c > 0)\end{aligned}\tag{4}$$

In this simple framework we may expect, first, a difference in mean earnings across sectors (a wage premium) corresponding to the mean value of c ⁶. Second, to the extent that x_i includes observable characteristics (e.g. h, K in equations (1),(2)), sorting by observable characteristics into sectors according to the relative returns net of differential entry costs. Third, to the extent that x_i includes unobservables (e.g. θ_i in (1),(2)), it will be important to allow for different forms of

⁴Modeling wage-setting in the formal sector is beyond the scope of this article and we choose to remain suggestive on the causes of the imbalance. What we are interested in is allowing for workers seeking employment in the formal sector to be unsuccessful.

⁵Given the binary choice framework used here only relative entry costs will determine choice. If entry costs are the same in the two sectors there is no role for entrance costs to determine sector choice, although labour market participation and hours of work - margins we do not consider here - may be affected by the overall level of such costs. Most obviously, capital constraints have been argued to create important entry costs for the self-employed sector. To the extent that we observe capital, this factor can be controlled for. Conditional on capital holdings, we would however expect relatively low barriers to entry in the self-employed sector.

⁶This is true conditional on observing employed individuals. Following the Harris-Todaro argument, costs and benefits of *intending* to enter a sector are equalized in expectation. Differences in rewards then exist for those who find employment.

sorting on unobservables.

Given that individual effects determine sector choice, a random effects framework for analysing unobserved heterogeneity will not be suitable. Fixed effects will also be inappropriate to the extent that individual effects may be rewarded differently across sectors.

Furthermore, if the diverging influence of unobservables on entry costs is strong (e.g. due to selection into public sector jobs not based on productive characteristics), it may be the case that institutions select individuals with comparative advantage in self-employment to be wage employees. In this case, the assumptions made in selection models may also be violated. A two-stage model using instruments for sector choice or relying on parametric assumptions to take into account selection effects requires an assumption that certain factors monotonously increase the likelihood of workers finding a wage job (Heckman (1976)). We could not identify credible instruments affecting sector selection but not earnings in our data.

Our plan is to derive a consistent estimate of $\Delta R(x_{i,t})$ allowing for unspecified sorting operating via some entry costs $c(\cdot)$. As a way of understanding the increase in self-employment in Ghana, we are particularly interested in the evolution over time of the sector wage premia δ and of potential changes in the returns to other characteristics (β, α, ψ) : If the increase in self-employment is accompanied by an unexpected wage differential, we might view the rise in a different light than if we find that changes in productive characteristics or selection on unobservables is responsible for differences in earnings across sectors.

4 The Correlated Random Coefficients Model

In this section we show how we estimate earnings functions and identify the wage premium under the assumption that job queues may exist and selection is not necessarily based on a single threshold (as in self-selection models). Our reduced form panel-based procedure that allows us to estimate the determinants of earnings allowing for unspecified patterns of selection on unobservable individual characteristics. The estimation procedure relies on careful observation of the movers between the two employment states considered here, and with few functional form assumptions⁷. Thus we can consistently estimate the impact of covariates in the two sectors for unspecified selection as well as evolution over time of the sector

⁷A key assumption is the multiplicativity of ψ .

wage differential⁸.

Following (1) and (2), express earnings using the self-employed dummy d_i^{SE} :

$$\begin{aligned} R_{i,t} = & \delta_t^w + d_{i,t}^{SE} (\delta_t^{SE} - \delta_t^w) \\ & + \alpha^w \ln h_i + d_{i,t}^{SE} \ln h_i (\alpha^{SE} - \alpha^w) \\ & + \beta^w \ln K_{i,t} + d_{i,t}^{SE} \ln K_{i,t} (\beta^{SE} - \beta^w) \\ & + \theta_i^w + d_{i,t}^{SE} (\theta_i^{SE} - \theta_i^w) + \varepsilon_{i,t} \end{aligned} \quad (5)$$

where

$$\varepsilon_{i,t} \equiv u_{i,t}^w + d_{i,t}^{SE} (u_{i,t}^{SE} - u_{i,t}^w). \quad (6)$$

We differentiate an individual effect that is remunerated equally in both sectors (τ_i) and will play no role in sector choice from an unobservable effect θ_i that is remunerated differentially in the two sectors and will generate a comparative advantage in a sector, *ceteris paribus*. We introduce a parameter ψ indicating the average returns to unobservables in the SE vis-à-vis the wage sector⁹. This formulation can be derived from the above θ^{SE} and θ^w ¹⁰.

$$\theta_i^{SE} = \psi \theta_i + \tau_i \quad (9)$$

$$\theta_i^w = \theta_i + \tau_i \quad (10)$$

⁸The procedure follows Lemieux (1998), who considers the union wage premium and has recently been applied by Suri (2011) to technological change.

⁹As Evdokimov (Evdokimov (2010)) notes, the introduction of only one parameter to capture differential returns to unobservables constitutes a semiparametric restriction that can be relaxed. However, with the sample sizes available here, non-parametric techniques appear all but infeasible.

¹⁰Using a simple projection, we can separate an absolute advantage component, τ_i , from a comparative advantage component of individual unobserved heterogeneity we name θ_i . As Suri (2011) notes, one can easily see that the τ_i in equations (7) and (8) are the same by subtracting (8) from (7) and noting that $b_{SE} + b_w = 1$ by construction.

$$\theta_i^{SE} = b_{SE} (\theta_i^{SE} - \theta_i^w) + \tau_i \quad (7)$$

$$\theta_i^w = b_w (\theta_i^{SE} - \theta_i^w) + \tau_i \quad (8)$$

$$\theta_i \equiv b_w (\theta_i^{SE} - \theta_i^w)$$

Where the projection coefficients are $b_w \equiv \frac{\sigma_w^2 - \sigma_{w,SE}}{\sigma_{SE}^2 - \sigma_w^2 - 2\sigma_{w,SE}}$ and $b_{SE} \equiv \frac{\sigma_{w,SE} - \sigma_{SE}^2}{\sigma_w^2 - \sigma_{SE}^2 - 2\sigma_{w,SE}}$.

We can then see in equations (9) and (10) that the model implies that the comparative advantage effect, θ_i , is remunerated differentially in the two sectors unless $\psi \equiv \frac{b_w}{b_{SE}} = 1$, an equality we can test for. Equations (9) and (10) directly follow.

Earnings can then be written as follows:

$$\begin{aligned}
 R_{i,t} = & \delta_t^w + d_{i,t}^{SE} (\delta_t^{SE} - \delta_t^w) \\
 & + \alpha^w \ln h_i + d_{i,t}^{SE} \ln h_i (\alpha^{SE} - \alpha^w) \\
 & + \beta^w \ln K_{i,t} + d_{i,t}^{SE} \ln K_{i,t} (\beta^{SE} - \beta^w) \\
 & + \theta_i + d_{i,t}^{SE} (\psi - 1) \theta_i + \tau_i + \varepsilon_{i,t}
 \end{aligned} \tag{11}$$

Contrast this formulation with other techniques for analysing unobservables in sector choice: if $\psi = 1$, unobservable individual characteristics are not remunerated differently across sectors. With no sorting on unobservables there is then no selection bias by estimating a first-differenced or fixed-effects model. If $\psi > 1$ there exists a premium for workers with high levels of unobserved skills in the self-employed sector. By contrast, if $\psi < 1$ there is a premium for these workers in the wage sector. The sign of ψ thus provides incentives for sorting on unobservables. Thus we would like to remain agnostic about the actual selection mechanism when estimating the relative returns ψ ¹¹.

In summary, if we wish to make no assumptions about sector allocation it is important to allow for differential returns to unobservables in the two sectors and not to restrict the potential direction of selection bias.

4.1 Identification

Model identification relies on the classic panel data restriction that the idiosyncratic error terms $u_{i,t}$ are uncorrelated with the covariates in all time periods. In other words, movement across sectors depends on expected earnings as specified in our model. This allows for rich patterns of selection on the unobservables θ_i but does not allow for sector choice to be a function of idiosyncratic errors $u_{i,t}$. Large income shocks may be thought to influence sector choice. We can note that if sector choice may be thought to operate in advance of such shocks (e.g. due to formalised recruitment cycles in the wage sector), sector choice will be a predetermined variable. In the same line of reasoning we may think that shocks operate by depleting individuals' level of capital. This would violate the assumption of strict

¹¹Note that for θ_i , if $\psi < 1$ then the degree of inequality across sectors is lower than it would be if sector allocation was random whereas for $\psi > 1$ the opposite holds. Note that this does not imply that specialisation increases inequality overall. For a nice parametric example in the spirit of Roy (1951) where specialisation reduces inequality overall, see Heckman and Honore (1990).

exogeneity. We discuss this in more detail in section (4.2).

Coefficients for the time-varying covariants β are identified by variation in returns across different levels of the covariates within each sector¹². The intuition for identification of ψ can best be seen when we consider that fixed individual effects models are overidentifying if we observe movers between sectors. By the same means that fixed effects are identified within a sector (i.e. by time-demeaning), differential fixed effects by sector are then time-identified by time-demeaning in both sectors - as long as we have movers.

Standard panel data methods using differencing or fixed effects analyses will not remove the unobservable individual effect θ_i . However, following Chamberlain (1982) and Lemieux (1998) we can solve explicitly for θ_i in (11) and replace this expression in the general earnings equation.

$$\theta_i = \frac{R_{i,t} - [\delta_t^w + D_{i,t} (\delta_t^{SE} - \delta_t^w) + D_{i,t} \beta^w K_{i,t} + D_{i,t} (\beta^{SE} - \beta^w) K_{i,t} + \alpha^w h_i + D_{i,t} (\alpha^{SE} - \alpha^w) h_i + \tau_i + \varepsilon_{i,t}]}{1 + D_{i,t} (\psi - 1)}, \quad (12)$$

where $k_{i,t} = \ln(K_{i,t})$.

For convenience, define $G_{i,t}(K, h, D) \equiv \beta^w K_{i,t} + D_{i,t} (\beta^{SE} - \beta^w) K_{i,t} + \alpha^w h_i + D_{i,t-1} (\alpha^{SE} - \alpha^w) h_i$ and $\delta_t = \delta_t^w + D_{i,t} (\delta_t^{SE} - \delta_t^w)$. Combining (12) with the equivalent expression in $t - 1$:

$$\frac{R_{i,t} - G_{i,t}(K_{i,t}, h_i, D_{i,t}) - \delta_t - \tau_i - \varepsilon_{i,t}}{1 + D_{i,t} (\psi - 1)} = \frac{R_{i,t-1} - G_{i,t-1}(K_{i,t-1}, h_i, D_{i,t-1}) - \delta_{t-1} - \tau_i - \varepsilon_{i,t-1}}{1 + D_{i,t-1} (\psi - 1)} \quad (13)$$

Or, defining the transfer term $C_{i,t} \equiv \frac{1 + D_{i,t} (\psi - 1)}{1 + D_{i,t-1} (\psi - 1)}$:

$$R_{i,t} = G_{i,t} + \delta_t + \varepsilon_{i,t} + \tau_i + C_{i,t} (R_{i,t-1} - G_{i,t-1} - \delta_{t-1} - \varepsilon_{i,t-1} - \tau_i) \quad (14)$$

We have now removed the inobservable θ and, recognizing that τ is - conditional on θ - a random effect, we can estimate (14). Given the quasi-differenced framework, some normalisation on the individual effects or the time constants is required (analogously to the situation in other differenced models where time constants are not identified). We proceed by enforcing the following restriction:

$$\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \theta_{i(t)} = 0 \quad (15)$$

¹²This is only correct for time-varying observables. For time-invarying characteristics only the quasi-difference can be estimated, see below.

In particular, we can now try to minimize the quasi-differenced errors. Define

$$erm_{i,t} \equiv (\varepsilon_{i,t} + \tau_i) - C_{i,t} (\varepsilon_{i,t-1} + \tau_i) \quad (16)$$

Note that by assumption, $D_{i,t}$ is not correlated with $(\varepsilon_{i,t} + \tau_i)$ ¹³. However, since $R_{i,t-1}$ is by construction correlated with $\varepsilon_{i,t-1}$ and hence also with $erm_{i,t}$, we need to instrument for $R_{i,t-1}$. The requirement is that an instrument be correlated with $R_{i,t-1}$ but not with $\varepsilon_{i,t}$. Given our identification assumption (the standard strict exogeneity assumption of panel data models), and combined with the random effects character of τ_i (see (9) and (10)), any combination (past and present) of covariates will do¹⁴.

$$erm_{i,t} = (R_{i,t} - G_{i,t} - \delta_t) - C_{i,t} (R_{i,t-1} - G_{i,t-1} - \delta_{t-1}) \quad (17)$$

Another way of expressing this relationship considers the different sector histories (combinations of employment histories)¹⁵.

$$\begin{aligned} erm_{i,t} = (R_{i,t} - G_{i,t}) - C_{i,t} (R_{i,t-1} - G_{i,t-1}) - \\ D_{i,t} D_{i,t-1} (\delta_t^{SE} - \delta_{t-1}^{SE}) \\ (1 - D_{i,t}) D_{i,t-1} \left(\delta_t^w - \frac{1}{\psi} \delta_{t-1}^{SE} \right) \\ D_{i,t} (1 - D_{i,t-1}) (\delta_t^{SE} - \psi \delta_{t-1}^w) \\ (1 - D_{i,t}) (1 - D_{i,t-1}) (\delta_t^w - \delta_{t-1}^w) \end{aligned} \quad (18) \quad (19)$$

We now see that identification of δ relies on movers across sectors and that these cross-sector differences are only identified up to a function of ψ . Note that measures of human capital such as education, experience or age typically do not vary (conditional on a linear time trend) and are thus not identified within sectors. As in other panel frameworks, we can identify the difference in remuneration of these factors only by looking at differences in remuneration across sectors¹⁶.

¹³Whilst $D_{i,t}$ does appear in the expression for ε , this is unproblematic as long as $D_{i,t}$ is orthogonal to $w_{i,t}^{SE} - u_{i,t}^w$, see equation (6).

¹⁴We later consider the restriction of predeterminedness vis-à-vis capital to avoid reverse causality issues.

¹⁵That the following expression is equivalent to (17) can easily be seen by going through the possible combinations of values that $D_{i,s}$ can take for $s \in \{t, t-1\}$.

¹⁶More precisely, in the current context we are not performing linear differencing, but non-linear quasi-differencing. This means that we identify not the difference, but the quasi-difference $\alpha^{SE} - \frac{1}{\psi} \alpha^w$.

With the assumption of strict exogeneity, we can then use the following moment conditions:

$$E(erm_{i,t} Z_i) = 0 \quad (20)$$

Where Z_i is a vector of interactions of strictly exogenous variables (including values of variables from other time periods, in line with the Chamberlinian tradition).

4.2 Endogenising Capital

Capital accumulation may respond to unobservable shocks that affect the entrepreneur's optimal input choices and as a result also influence sector choice, violating our identifying assumption. Our capital variable, therefore, may be thought to be endogenous with respect to past and, possibly, current time-varying unobservables. For this reason, the literature has commonly attempted to relax the assumption of strict exogeneity and instrumented capital accordingly. Our dataset does not currently provide us with satisfactory external instruments (i.e. instruments that do not belong to the set of variables included in our model), and we would like to exploit its longitudinal dimension.

Following [Anderson and Hsiao \(1982\)](#), [Holtz-Eakin et al. \(1988\)](#) and [Arellano and Bond \(1991\)](#), we could use K_{t-1} as a valid instrument if capital is assumed to be *pre-determined*, in the sense that it is affected by past (but not contemporaneous and future) shocks and sector of employment, i.e.

$$E[K_{is}\varepsilon_{i,t}] = 0 \quad \forall t \geq s \quad (21)$$

$$E[K_{is}D_{i,t}] = 0 \quad \forall t \geq s \quad (22)$$

$$E[K_{is}\tau_i] = 0. \quad (23)$$

If on the other hand capital is affected by both past and contemporaneous (but not future) shocks and sector of employment, the first valid instrument is K_{t-2} (minimum lag length for a valid instrument is equal to 2). In this case, the first two assumptions can be made less stringent:

$$E[K_{is}\varepsilon_{i,t}] = 0 \quad \forall t > s \quad (24)$$

$$E[K_{is}D_{i,t}] = 0 \quad \forall t > s \quad (25)$$

However, allowing for sequential exogeneity (as opposed to strict) exogeneity increases data requirements: we now need three waves of consecutive information

on individuals because we need to instrument period $t - 1$ with period $t - 2$ information. We plan to implement this procedure to compare our results. In the current draft of the paper, we assume that capital is not affected by past shocks and sector of employment.

5 Data

We apply our model using data from the Ghana Household Urban Panel Survey (GHUPS), conducted by the Centre for the Study of African Economies (CSAE) at the University of Oxford. The survey was first conducted in 2004 and now spans 8 years, an unusual length for panel data-sets in developing countries. The GHUPS covers four cities: Accra, Kumasi, Takoradi and Cape Coast. Respondents were drawn by stratified random sampling of urban households from the Population and Housing Census of 2000. The survey was designed to cover all household members of working age at the time of the interview. After the first wave, the sample expanded by incorporating new members of the original households, as well as new households formed by individuals who had left their original household and were tracked to their new locations.

Figure (1) shows the split of the sample in self-employment and wage employment observed in our sample. Note that the increase in self-employment is not as marked here as in other samples. This is one of the reasons we focus on the structure of earnings in this sample. The panel design with its focus on reducing attrition may have led to an understatement of changes in the split of self-employment and wages. However, such changes would be expected to have a strong influence also on the wage structure, something that the data is very well suited to research.

Table (2) shows the change in composition of the workforce in our two sectors for the sample of GHUPS respondents who are currently employed over the sample period, pooling waves 2004-2007 and waves 2008-2011.

What does self-employment consist of in our predominantly urban sample? Figure (2) shows that the largest category of self-employed are traders, followed by service providers and manufacturers. Contrasting the wage employed to the self-employed we find that the latter are older and less educated on average - see figures (3) and (4). They are also more likely to be female (see figure (5)).

Table 2: Demographic Composition by sector 2004-2007 vs. 2008-2011

Wage employment	Mean		Std. Dev.	
N=508 / N=782	04-07	08-11	04-07	08-11
Male	0.70	0.63	0.46	0.48
Age	34.72	35.84	10.38	10.39
Educ	9.81	9.76	3.26	3.86
Real Capital (2002 USD)	336.11	313.14	1155.96	804.21
Real Monthly Earnings (2002 USD)	72.11	106.32	75.79	206.74
Civil Service / Public Sector	0.2	0.23	0.4	0.42
Job satisfaction	3.27	3.22	0.84	1.06

Self-employment	Mean		Std. Dev.	
N=825 / N=1231	04-07	08-11	04-07	08-11
Male	0.3	0.27	0.46	0.44
Age	37.35	39.36	10.02	9.76
Educ	7.45	7.94	4.03	3.78
Real Capital (2002 USD)	203.24	473.11	419.23	1512.45
Real Monthly Earnings (2002 USD)	64	134.02	85.79	332.71
Job satisfaction	3.24	3.37	0.85	1.08

Figure 1: Shares of Wage and Self-employed workers

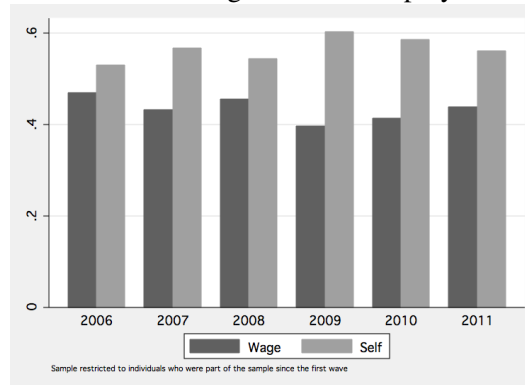


Figure 2: Different types of self-employed workers (pooling all waves)

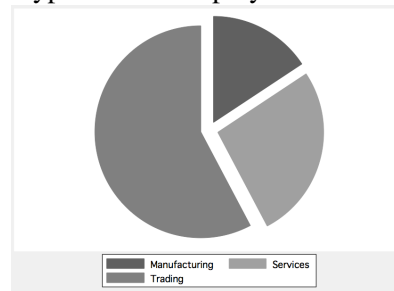


Figure 3: Age of Wage and Self-employed workers

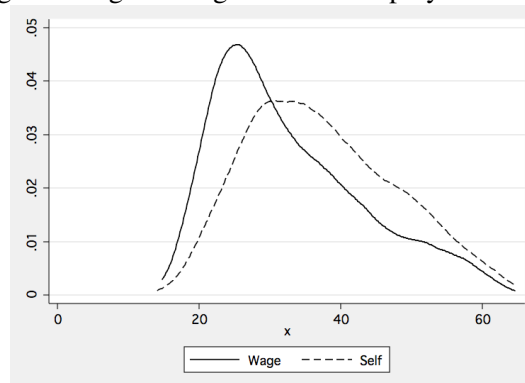


Figure 4: Education levels of Wage and Self-employed workers

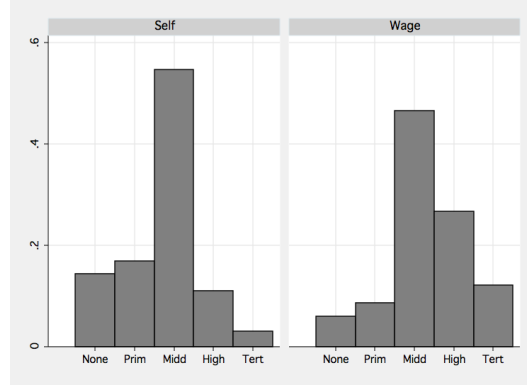


Figure 5: Changes in gender composition of Wage and Self-employed workers



5.1 Trends in Earnings

Table (2) shows that whilst earnings in wage employment increased by roughly 50% they more than doubled in self-employment, leading to a reversal of the gross wage differential. However, figure (6) shows that in the more recent period the difference in *median* earnings across the two sectors is not large - and not statistically significant (not shown). However, figure (7) shows an increase in the variance of earnings especially for self-employed workers - alongside considerable real earnings growth over the sample period. Strong earnings growth among the self-employed is also shown in figure (8), which reveals that real earnings have increased considerably faster among the self-employed than for employees in the wage-sector over the survey period. This is *prima facie* evidence against the common negative view of self-employment as an occupational category inferior to

wage-employment and populated by workers with lack of alternatives in the wage sector (the *push* interpretation for the rise of self-employment).

Figure 6: Real monthly median earnings by sector (in US dollars)



5.2 Transiting between Wage and Self employment

Identification relies crucially on the movers. Table (4) summarises transitions between sectors, pooled across all panel waves. It is confined to workers who are employed both at t and $t - 1$ and shows that 14.7 % of all workers who are in wage-employment in any given period move to self-employment in the next period, while 11.2% of self-employed workers move to wage-employment. Though sizeable in percentage terms, the number of observed transitions will pose a challenge to the precision of our panel estimators.

Specifically, if we use information only from individuals for who we have all the information required for all eight years, we are left with only 42 observations. Decreasing the number of waves jointly estimated naturally increases the sample size that can be exploited (see 3).

The movers are maybe predictably found to be younger (see figure (9)) whereas there appears to be no difference in relative levels of education (not shown).

Table 3: Panel Attrition - Samples sizes for different panels

years covered by different panels	panel combinations	observations (movers)
2004-11	one 8-wave panel	42 (15)
2004-07, 08-11	two 4-wave panel	122 (33), 188 (52)
2004-06, 05-07, 08-10, 09-11	four 3-wave panel	311 (41), 222 (51), 226 (52), 354 (73)
2004-05, 05-06, 06-07, 08-09, 09-10, 10-11	six 2-wave panel	408 (40), 588 (58), 337 (58), 303 (55), 448 (65), 792 (118)

Table 4: Transitions between Wage and Self-employment

	<i>Salaried Wage Emp_t</i>	<i>Self – Emp_t</i>	Tot
<i>Salaried Wage Emp_{t-1}</i>	1,369 (85.30)	236 (14.70)	1,605 (100)
<i>Self – Emp_{t-1}</i>	260 (11.22)	2,058 (88.78)	2,318 (100)
Total	1,629 (41.52)	2,294 (58.48)	3,923 (100)

Consecutive period transitions pooled across waves; Percentages reported in parentheses

Figure 7: Earnings distributions by sector

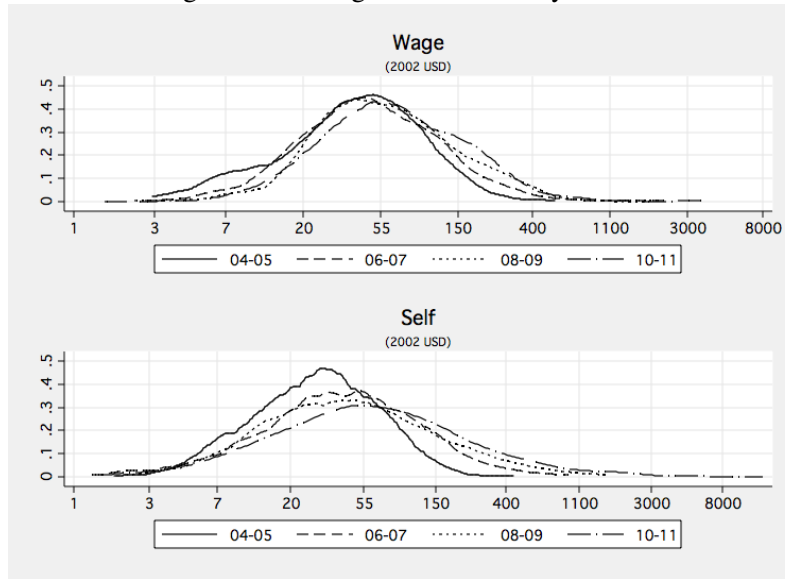


Figure 8: Median growth rates of real earnings by sector

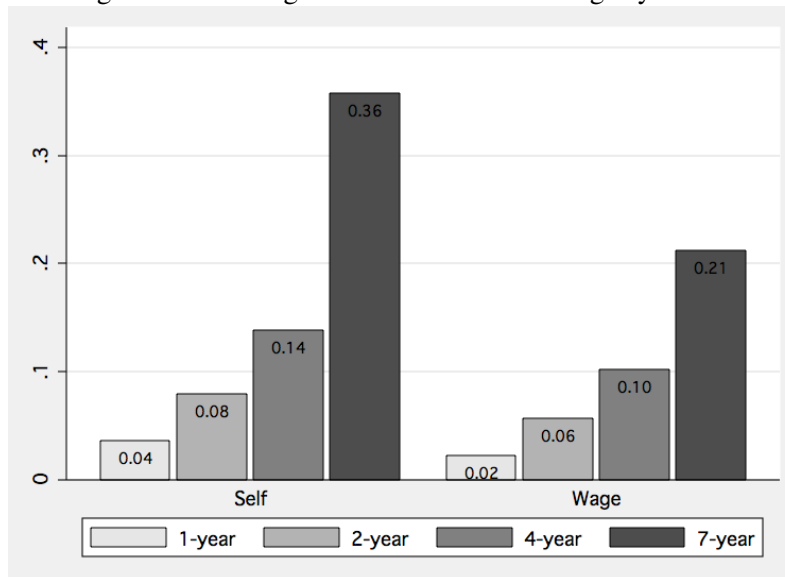
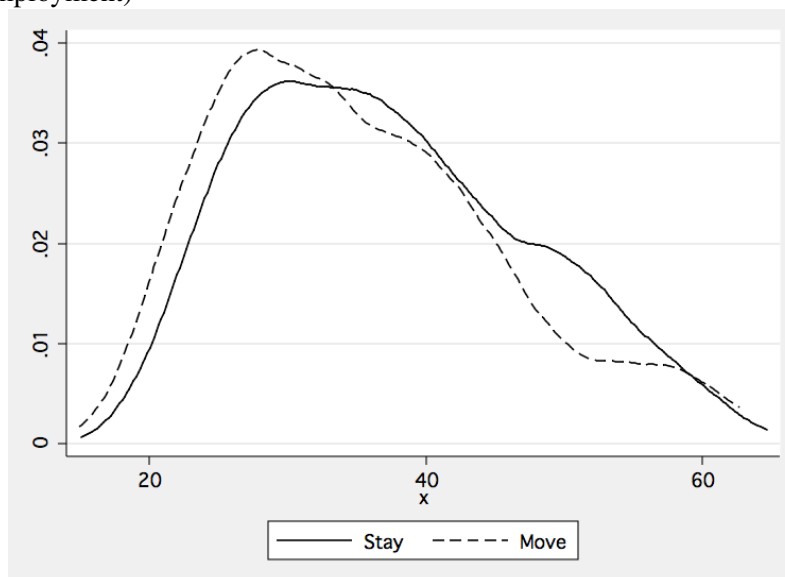


Figure 9: Distribution of age by transition status (transition between wage and self-employment)

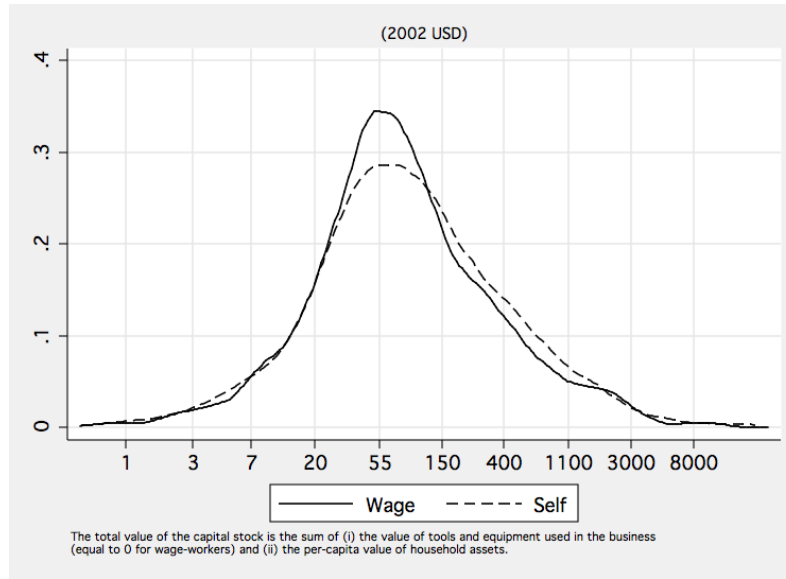


5.3 Measures of capital

Given its traditionally important role in analysing self-employment decisions, our measure and treatment of capital deserves some discussion. The dataset includes information on assets and business capital for self-employed individuals. Various types of capital are observed - agricultural land, real estate, tools and equipment - for each of which respondents are asked to report monetary valuations. Ownership of agricultural land is very rare in urban Ghana, while the value of real estate is measured very noisily and suffers from the problems of clearly identifying ownership, especially in those areas where urban development has been largely unregulated and official titling is absent. For these reasons we choose to focus on the value of household and business assets. In line with previous findings, there is a considerable difference in capital holdings in favour of the self-employed (table (2) finds this only for the later period however), however the sometimes proposed idea of a very large variance of capital in self-employed business (maybe even bi-modal distribution of necessity self-employed on the lower end and self-employed business owners at the upper end) is not evidenced when we consider the density of (log) capital values in figure (10). Studying the distribution of capital by sector, figure (12) shows that many self-employed operate at low levels of capital intensity - in trading especially, but also in manufacturing and service provision. Credit

constraints might cause this if they operate not by impeding the establishment of a business, but maybe by reducing returns.

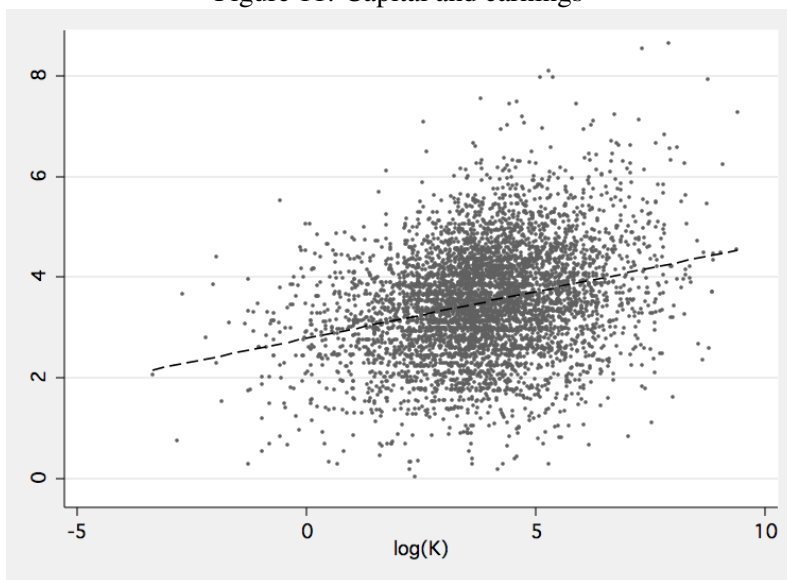
Figure 10: Distribution of real capital for wage and self-employed individuals (log scale)



We are interested in the relationship between capital assets and incomes. In particular, we are interested in the income stream or usage value generated by the assets respondents own, in any given period and for both sectors. However, since the majority of our respondents lacks access to formal banking and does not own real estate that can be rented out, we do not expect to observe significant cash streams generated by asset ownership in our sample. Therefore, we choose to impute a usage value of assets for every period. In doing so, we assume that usage-returns to assets accrue at a constant rate to all respondents (heterogeneous returns may matter)¹⁷. Figure (11) plots (real) value of tools/equipment and earnings, including the imputed capital usage values.

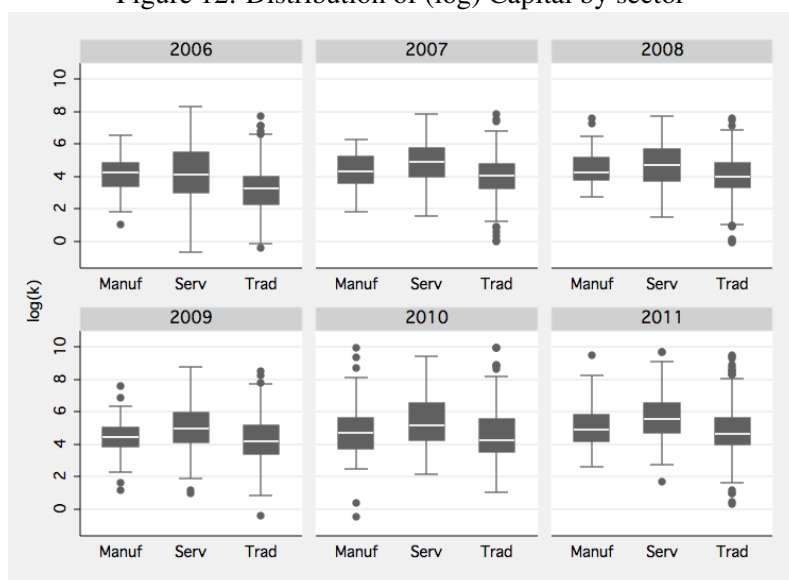
¹⁷ At the time of writing, Feb 2011, the Bank of Ghana Official interest rate is 13.5 % p.a, while the latest inflation figures report that inflation was 10 % over the past year. The real official interest rate is therefore 3.5 %. We believe this is likely to be an upper bound on what our respondents can gain on the value of their assets, since they mostly lack access to official banking and many of their assets are not sufficiently liquid. Therefore, we choose to apply a 2 % AER interest rate, which amounts to 1.16 % per month.

Figure 11: Capital and earnings



Capital holdings may be derived from past labour market outcomes. This would mean that earnings shocks in the past may be correlated with current capital holdings. This type of reverse causality has been the motivation for models using functions of predetermined (as opposed to current) levels of capital as instruments for capital holdings (Arellano and Bond (1991), Blundell and Bond (2000)). Allowing for this type of dependence implies using fewer observations in estimation but will be implemented in the future. Our current estimation assumes exogenous capital.

Figure 12: Distribution of (log) Capital by sector



5.4 Attrition and Data Aggregation

As discussed in a previous section, although the panel dataset is long and large by the standards of developing countries, attrition is an issue. If we use only observations for which we have coverage in all time periods, the sample size becomes very small. The alternative is to estimate the model separately using shorter panels. Indeed, [Muris \(2010\)](#) shows that by optimally weighting subsamples of different time periods for which data are available, a consistent and efficient GMM estimator can be constructed. We might then be tempted to focus on estimates from pairs of years. The disadvantage, however, is that fewer instruments are available and Monte Carlo tests confirmed better performance of our estimator with at least three time periods. A second issue concerns information for the year 2007, which is derived from recall data collected in 2008. We find that the number of movers across sectors between these two years is so low that the transition from 2007 to 2008 cannot be analysed. For these reasons, our preferred approach is to estimate the model over two four-wave panels: from 2004 to 2007 and from 2008 to 2011. Standard errors are bootstrapped (in this first version over 100 resamples), whereby the relatively small number of movers constrains our ability to perform certain additional robustness checks (out-of-sample prediction, comparing estimates for voluntary and involuntary movers etc.).

6 Results

In this section we present a set of results from the estimation of our quasi-differenced earnings model (some results are preliminary and bootstrap resample sizes are currently small), using instrumental variables Generalized Method of Moments (IV-GMM). We contrast these to estimates from simpler OLS and Fixed Effects models, estimated under the assumption that ψ is equal to 1¹⁸.

The results are presented in table (5). These indicate increasing returns to productive characteristics in the self-employed sector and an increasing wage premium in the wage sector. More specifically, four aspects stand out.

First, returns to capital. Unsurprisingly, returns to capital are significantly higher in the self-employed compared to the wage sector, for which we have assumed an annualised interest rate of 2%. Returns to capital in self-employment

¹⁸Note that despite the fact that $\psi \neq 1$ implies that the fixed effects estimator is inconsistent it is not necessarily strongly biased. Bias depends on the average level of unobserved skill levels across sectors, see figures (14) and (15) for evidence on this.

appear to increase significantly between the first and the second half of the panel. Our specification assumes a Cobb-Douglas-type production function (with constant elasticity) such that returns to capital vary over capital and earnings. However, figure (13) plots the implied annualised returns that our value of β^{SE} implies for individuals with different K and earnings. We find that the distribution for the later period first-order stochastically dominates returns in the previous period. Calculated interest rates vary significantly, median annualised values are 18.15% in the first period and 94, 49% in the second. These rates (compounded from monthly rates implied by $\hat{\beta}$) may appear high, and our treatment here falls short of the standards in the literature estimating production functions. However, our results are consistent with other estimates of returns to capital in Ghana: Udry and Anagol (2006) estimate a lower bound of 60% annualised returns to capital and report rates up to 250 – 300% for farmers of certain crops. If these high rates can be believed and given that there is no evidence that investment opportunities offer similar returns (to the contrary, interest rates have often been thought to be negative in real terms), they provide a strong incentive for holders of capital to starting a self-employed business. Thus they provide additional credence to the argument that self-employment may partly be related to a lack of investment opportunities. The comparative advantage of being in the self-employed sector *for individuals with enough capital* has increased over time in all our specifications. This increase in returns to capital is present on a similar order of magnitude in the ordinary least squares (OLS) estimates, although these estimates do not take any precautions regarding the endogeneity of capital (or unobservables) (see table (7)). We do not find the result for the fixed effects estimates (FE).

Second, returns to human capital have also increased in self-employment, and are in the more recent period found to actually be higher than in wage employment¹⁹, whereas during the 2004 – 2008 period, returns to education in self-employment are found to be lower than in wage employment. Given that we only have access to quasi-differenced estimates for this time-invariant characteristic, levels of returns to education cannot be given - however the estimates for the more recent period imply rates of return to education (RORE) of at least 10% higher in self-employment than wage employment, an important finding supporting investment in schooling in countries such as Ghana with fairly high levels of self-employment. The trend is consistent across estimators, the difference in levels is not however: When we consider the results in the four 3-year subpanels (see

¹⁹Note that only the quasi-difference across sectors $\alpha^{SE} - \psi \alpha^w$ is identified. However, given the positive point estimates of the quasi-differential and the fact that $\hat{\psi} > 1$, we know that $\alpha^{SE} - \alpha^w > 0$.

Figure 13: Distribution of implied annualised interest rates using $\hat{\beta}$

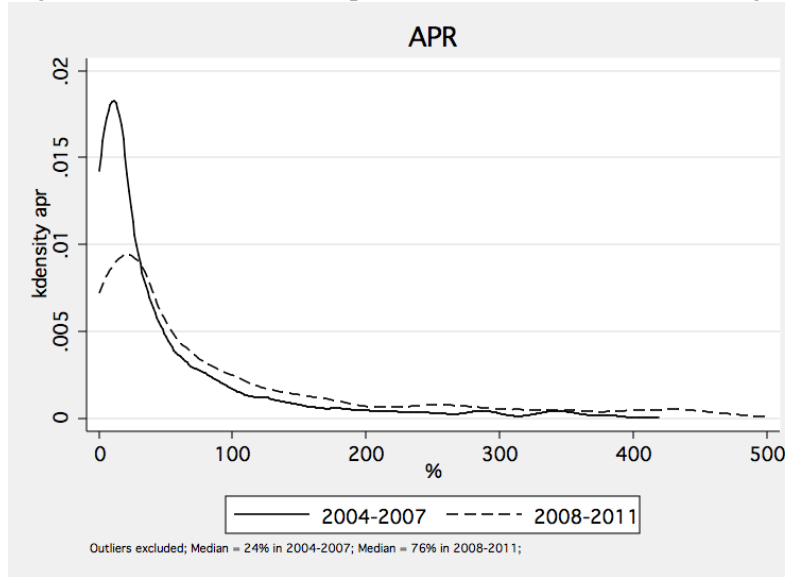


table (6)), only one of the four subperiod estimates shows higher RORE in the self-employed compared with the wage employed sector. The differential trend in RORE is also visible - albeit to a lesser degree - in the OLS and FE estimates, the self-employed are not, in those estimates, found to have higher RORE in the more recent period. Thus we conservatively conclude that there appears to be a significant trend increase in RORE for the self-employed.

Third, returns to unobservable skills appear to be significantly higher in the self-employed than the wage sector only in the more recent period: The estimated value of ψ in table (5) changes from below unity for the period 2004-2007 to being larger than one in subsequent years, indicating that self-employment has *increasingly become a desirable opportunity* - potentially because of lower entry costs - for those with large unobservable skills. In fact, when we consider changes in the composition of the workforce using estimated values of theta, we can show that it is the case that in the more recent period we find more individuals with high levels of unobserved skills (this is discussed in detail in the following section).

Fourth, the unexplained constant sector effect (the wage premium) has increased over time, from a negative 97% lower monthly earnings in the wage sector to a 250% higher level of base earnings. Whilst these figures (and the swing) may

Table 5: Determinants of log monthly earnings - 4-year sub-panels; 95% confidence interval (normal naïve bootstrap, 100 resamples - *to be extended*)

	IV-GMM 2004-2008	IV-GMM 2008-2011
$\delta^{SE} - \delta^w$ (average over year pairs)	0.9723 (0.8555 - 1.0892)	-2.4967 (-2.7223 - -2.2711)
β^{SE}	0.0302 (0.0092 - 0.7970)	0.1242 (0.1038 - 0.1445)
$\alpha_{Male}^{SE} - \psi \alpha_{Male}^W$	0.0203 (-0.079 - 0.1203)	-0.3960 (-0.5256 - -0.2664)
$\alpha_{Educ}^{SE} - \psi \alpha_{Educ}^W$	-0.0708 (-0.0789 - -0.0626)	0.1083 (0.0968 - 0.1198)
ψ	0.8396 (0.7970 - 0.8823)	1.3456 (1.2921 - 1.3990)
Sample N (movers)	122 (33)	188 (52)

appear very high, it should be remembered that a large part of SE earnings derive from their use of capital. Looking at the four 3-year subpanels we find a consistent pattern of rising wage premia throughout the period.

How can these four key results²⁰ be reconciled and what might they tell us about the recent developments in the labour market in Ghana? The following section relates these developments in returns to compositional changes and proposes an interpretation.

6.1 Understanding the Rise in Self-employment

We now employ our estimated model parameters to consider the evolution of self-employment: its trend increase and the increase in the variance of earnings over time. We can analyse differences in returns to observables, unobservables as well

²⁰For completeness: In other results, differences in the gender gap across sectors are not precisely estimated. For the 2004-2008 period we find positive, negative and insignificant results depending on the instrument set. For the 2008-2011 period we find a negative quasi-difference, however vis-a-vis the previous period the increased ψ -value could be the cause here. More specifically, a negative quasi-difference in combination with $\hat{\psi} > 1$ implies that we cannot make any statement about the relative size of the gender gap in the two sectors.

Table 6: Determinants of log monthly earnings - 3-year sub-panels; 95% confidence interval (normal naïve bootstrap, 100 resamples - *to be extended*)

	IV-GMM 2004-06	IV-GMM 2005-07	IV-GMM 2008-10	IV-GMM 2009-11
$\delta^{SE} - \delta^W$	1.0169 (0.8561-1.1777)	0.2273 (0.0416-0.4129)	-0.2888 (-0.4531- -0.1235)	-1.9888 (-2.195- -1.7813)
β^{SE}	0.0001 (-0.0120-0.0120)	0.0690 (0.0569-0.0811)	0.0640 (0.0443-0.0836)	0.0450 (0.0226-0.0675)
$\alpha_{Male}^{SE} - \psi^W \alpha_{Male}^W$	0.2840 (0.2149-0.3531)	-0.2366 (-0.3173- -0.1559)	-0.0505 (-0.1420- -0.0408)	-0.6617 (-0.7608- -0.5626)
$\alpha_{Educ}^{SE} - \psi^W \alpha_{Educ}^W$	-0.0871 (-0.0999- -0.0743)	-0.0496 (-0.0582- -0.0409)	0.0716 (0.0600-0.0831)	-0.0356 (-0.0496- -0.0216)
ψ	0.8467 (0.8029-0.8905)	1.0232 (0.9604-1.0860)	0.9004 (0.8545-0.9464)	1.8001 (1.7403-1.8600)
Sample N (movers)	311 (41)	222 (51)	226 (52)	354 (73)

as the sector dummy δ and distinguish the effect of changes in returns to productive characteristics from changes in the composition of characteristics across sectors.

A particular concern may be that if we have diminishing returns to input factors such as education or unobservable skills²¹, the increase in returns may be a result of decreased deployment of these factors in this sector. With less educated individuals in the self-employed sector, marginal returns to education may be on the rise. However, we find the opposite. For all factors of production (physical and human capital as well as unobserved ability) we find *increasing* levels of factor inputs in the self-employed sector, as table (2) shows.

For capital we find lower levels (falling from a mean value of 336.11 US dollars to a value of 313.14 in real terms) in the wage sector in the more recent period, despite the considerable increase in earnings in the period. By contrast, average levels of capital in the self-employed sector more than doubled from 203.24 to 473.11 US dollars in the more recent period.

For schooling the differential trends are less marked (but statistically significant), with levels in the self-employed sector rising from 7.45 to 7.94 on average whereas in the wage sector levels are essentially unchanged at 9.81 and 9.76 respectively.

In order to compare the levels of unobserved skills θ in the two sectors across time, we use the expression derived above (see equation (12))²², averaging over different years of observation. We have very few observations underlying our estimates of θ_i (a maximum of 4 observations if we use the 4-wave panel 2004-2007), thus our estimates of θ will suffer from small-sample bias. However, since our estimates are consistent, it is hoped that sufficient cross-sectional observations will not bias the distribution of θ_i . Graphs (14) and (15) show that whilst the variance of skills in self-employment appears higher in the early period, the difference in mean values between the two sectors appeared small²³. In the later period, by contrast, unobservable skills in the self-employed skills are clearly higher, on average.

²¹Note that with respect to capital we are in a constant-elasticity Cobb-Douglas world: Our log-log formulation assumes decreasing returns to capital.

²²There is a snag, as we do not observe the levels of our time-invariant characteristics α , only the quasi-difference. However, we find little sensitivity to varying different values of α . Additional figures in section (9) provide equivalent kernel density plots to those presented here under differing assumptions on levels of α .

²³This is even more so for some of the other values of α we tested, see the additional figures in section (9)

Figure 14: Distribution of $\hat{\theta}$ by sector

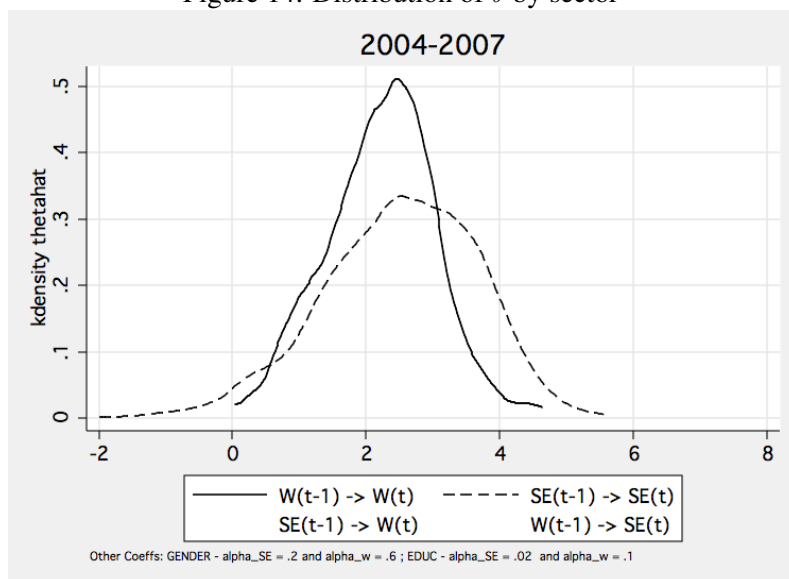
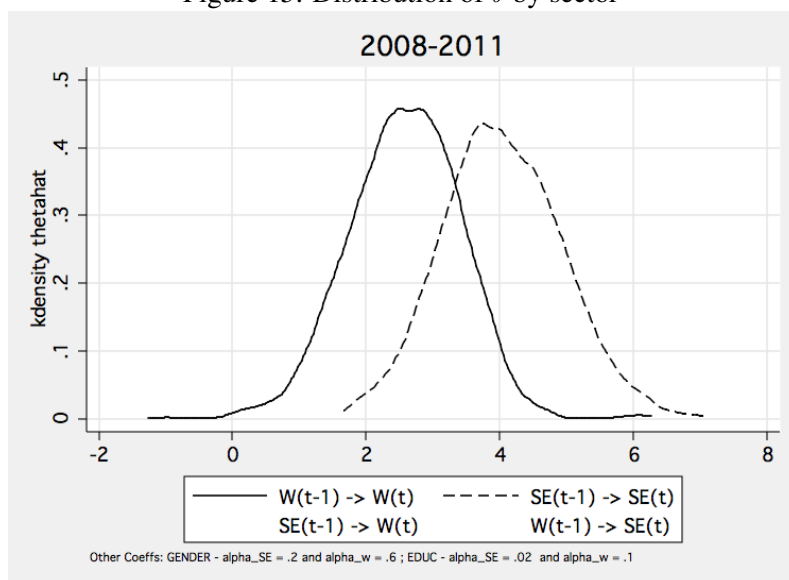


Figure 15: Distribution of $\hat{\theta}$ by sector



There appears to have been a move of high-skilled individuals to the self-employed sector (or recent high-skilled individuals joined the self-employed sector in larger

numbers).

Thus we interpret this as a sign that the changes in returns are driving changes in the composition of the workforce (and associated input factors), not vice-versa.

Given rising returns to the observable and unobservable productive characteristics over time in self-employment and an increase in the prevalence of these factors of production amongst the self-employed, earnings may be expected to be increasing. However, countervailing this is an increasing wage premium. Thus after controlling for all the (unobservable and observable) characteristics, we find starkly increasing wage premia. Given that we are conditioning on observables and unobservables this wage premium can be interpreted as an indication of a lack of perfect mobility across sectors, e.g. as a result of barriers to entry into a sector.

Note that the lower returns to productive characteristics may be in part related to the relative importance of the public sector amongst the wage employed. Public sector “equalising” effects with higher basic pay (here expressed by higher values of δ^w) and lower returns to productive traits have been found in other developing countries (this wage compression feature of public sectors has been found in [Gosling and Lemieux \(2004\)](#)). However, unless there was a change in public sector pay policies²⁴, this cannot explain the increase in the wage premium since the proportion of public sector workers only increased from 20% to 23% amongst the wage employed workers.

6.2 Comparison with other estimators

Table (7) shows benchmark results from estimating our model with OLS (columns 1-2) and a Fixed Effect estimator (columns 3-4). These simpler models are nested in our previous estimation framework, and corresponds to the case where $\psi = 1$. With this comparison, therefore, we can test the implications of this restriction, which is common in the existing literature.

Similar to our IV-GMM results, we detect a significant reversal in the wage premium ($\delta^w - \delta^{SE}$), which changes from being significantly negative in the first years of the panel, to being positive later on. The result, however, is not significant once we control for Fixed Effects. Allowing for selection on unobservables thus appears necessary to uncover the significant wage premia workers in the wage

²⁴See [Imbert \(2012\)](#) for an analysis of the effects on the wage structure of public sector pay reform in Vietnam.

Table 7: OLS and Fixed Effects estimators ($\psi = 1$)

	OLS		FE	
	2004-2007	2008-2011	2004-2007	2008-2011
$\delta^{SE} - \delta^w$	0.283 (0.111)**	-0.320 (0.114)***	-0.001 (0.220)	-0.328 (0.209)
β^{SE}	0.125 (0.015)***	0.180 (0.016)***	0.094 (0.024)***	0.089 (0.026)***
α_{Male}^w	0.375 (0.057)***	0.341 (0.053)***		
$\alpha_{Male}^{SE} - \alpha_{Male}^w$	-0.062 (0.075)	0.302 (0.075)***	-0.046 (0.147)	0.162 (0.146)
α_{Educ}^w	0.095 (0.008)***	0.078 (0.007)***		
$\alpha_{Educ}^{SE} - \alpha_{Educ}^w$	-0.087 (0.010)***	-0.060 (0.009)***	-0.056 (0.021)***	-0.003 (0.019)
δ^w	2.328 (0.089)***	2.906 (0.078)***		
Obs.	2,926	3,988	2,926	3,988
R^2	0.124	0.13	0.015	0.01

sector receive according to the present estimates. The idea that selection on unobservables might help to remove often large wage premia found in OLS regressions (reported in table (7)) is not verified here. How might it be the case that by allowing $\psi \neq 1$ we get significant sector premia absent in FE-estimates? Once we allow for unobservables to be remunerated differentially across sectors, it no longer appears to be the case that the more “able” individuals are to be found in wage employment. Rather, as evidenced in graph (15), the more able individuals are now found in self-employment, thus giving rise to sizeable wage premia.

Gender differentials and returns to capital are estimated more precisely in OLS (and their level effects are *only* identified in OLS, since both these characteristics are time-invariant). Once we control for fixed effects, we find no gender differential between self and wage employment. The education differential is only significant (and negative) in the first half of the panel, while it becomes insignificant in the second half.

Alternative explanations for the increase in self-employment that we cannot take into account include changes in risk or hedonic characteristics and will be discussed in the following section.

7 Beyond average earnings

The framework presented so far assumes that workers' objective function maximises expected material gains. Alternative factors determining sector choice may however be invoked to challenge the interpretation of earnings differentials as unjustified sectoral earnings premia. The literature has considered two factors in particular (additional to credit constraints).

First, empirical evidence from the developed world suggests that job-satisfaction is higher among self-employed workers than among wage-employees, after controlling for other workers' characteristics (Blanchflower (2004), Benz and Frey (2008), Benz and Frey (2003)). This may indicate that working conditions, managerial independence, flexibility (or any other characteristics of self-employment) may be valued in addition to material compensation. Second, not only the amount, but also the variance of earnings may be an argument of the objective function. Differences in risk aversion may explain different choice between self-employment and wage work for given levels of capital (rather than frictions in the labour market, as assumed here).

Differences in job satisfaction that derive from non-pecuniary job-attributes or unobservable individual characteristics, may also be related to occupational choice. In fact, a cursory glance at average levels of job satisfaction²⁵ in table (2) does appear to show a reduction in job satisfaction in the wage sector alongside an increase in job satisfaction in the self-employed sector. To deal with this issue, one could evaluate the evolution of job satisfaction analogously to that of wages (or generate a job quality index as Federico Huneeus and Puentes (2012) do for Chile). Alternately, one could consider job satisfaction (for which a subjective indicator is available in the data) as one of the factors influencing wages in the two sectors and test whether differences in working conditions compensate for part of the earnings differential. We leave integrating job satisfaction into the current framework to fu-

²⁵ Answers in table (2) refer to the subjectively evaluated question asked about job satisfaction in the GHUPS: "All things considered, how satisfied are you with your current work?" Potential answers were: "Very Dissatisfied" (coded 1); "Dissatisfied" (2); "Neither Satisfied Nor Dissatisfied"; "Satisfied" (4); "Very Satisfied" (5).

ture work.

With respect to risk aversion, empirical evidence from urban Ghana supports the intuition that self-employed individuals have lower levels of risk aversion. Hence, given that our model assumes risk-neutrality, we are implicitly raising the attractiveness of self-employment, thus going against our finding of wage premia in recent years²⁶.

8 Conclusion

Informal self-employment is the most common form of occupation throughout the developing world and the share of self-employed workers in the labour force is high and has been on the rise in recent decades in Ghana. This article has empirically assessed changes in the structure of earnings in the self-employment and wage sectors which can help determine whether the rise in self-employment is the result of improved opportunities for successful entrepreneurship (*pull*), or the reflection of limited opportunities in wage-employment (*push*).

We construct a two-sector earnings model with unobserved sector-specific heterogeneity (i.e. allowing workers to have comparative advantages). Our approach allows for rich patterns of selection on observable and unobservable worker characteristics and for differential returns to unobservable factors.

We estimate the model using a unique panel dataset from urban Ghana, covering a representative sample of workers over 8 years (2004-2011) in the four largest cities of the country and including information on individuals moving from self-employment to wage employment. Panel datasets of this kind are very rare in developing countries. The model is estimated using instrumental variables Generalized Method of Moments (IV-GMM). Contrasting our results with ordinary least squares and fixed effects estimators demonstrates the importance of allowing for differential returns to unobservable characteristics. We have three main findings.

First, we find evidence of increasing returns to productive factors in self-employment. Returns to capital, to schooling and to unobserved skills were all found to have increased significantly over the period of our sample (2004-2011). Returns to capital

²⁶Given that some experimental data on risk aversion is available for the population we are studying (see Falco (2012)), the quantitative implications of taking into account a positive level of risk aversion may be considered.

were found to have increased from levels already above those in wage employment at the beginning of the panel. We find conflicting evidence about relative levels of returns to education in the two sectors. Higher returns to unobserved skills in the self-employed sector are found to be robust across different specifications. These developments indicate that incentives for capital-rich, educated and otherwise able individuals to sort into self-employment have increased considerably - although levels of returns to education may still be higher in the wage sector.

Second, we find that the incentives of higher returns have worked. Individuals with better productive characteristics are now found in self-employment (apparently there has been *pull*): Levels of capital, education and unobserved skills in the self-employment sector are found to be higher in the more recent period. Thus, composition effects in the workforce would tend to work against these findings (supposing diminishing returns to factors of production), strengthening our argument for differential trends in returns.

Third, we find large and rising wage premia over time. Given that our estimations are controlling for observed *and unobserved* workers' characteristics, we interpret this finding as evidence for the existence of significant barriers to entry into different sectors, which may prevent market forces to equilibrate wages across them.

The results indicate that both push and pull factors are at work: we find increased incentives for self-employment but also increased wage premia that may indicate barriers to entry into the wage sector. The results thus provide a rationale for taking seriously reports of large returns to capital and are consistent with large labour-market distortions resulting from missing capital markets. Alternately or additionally, non-competitive pay may be the cause of unexplained sector differentials.

To put our results into context, it should be noted that our interpretation of barriers to entry is based on the assumption that sector choice is determined by expected monetary gains, leaving no room for risk aversion or non-monetary characteristics. Furthermore, all results reported here are based on comparatively small samples (a common feature of developing country datasets) but appear robust to varying the length of the panel to include more individuals at the cost of efficiency in estimation. Our preferred specification compares a four-year panel from 2004-2007 with a four-year panel from 2008-2011. Finally, our current model rules out the possibility of unemployment/inactivity. We do not incorporate capital accumulation. Evolutions in labour market participation over time may also affect the composition of the labour force.

9 Additional figures

Figure 16: Distribution of $\hat{\theta}$ by sector - alternative assumption on α

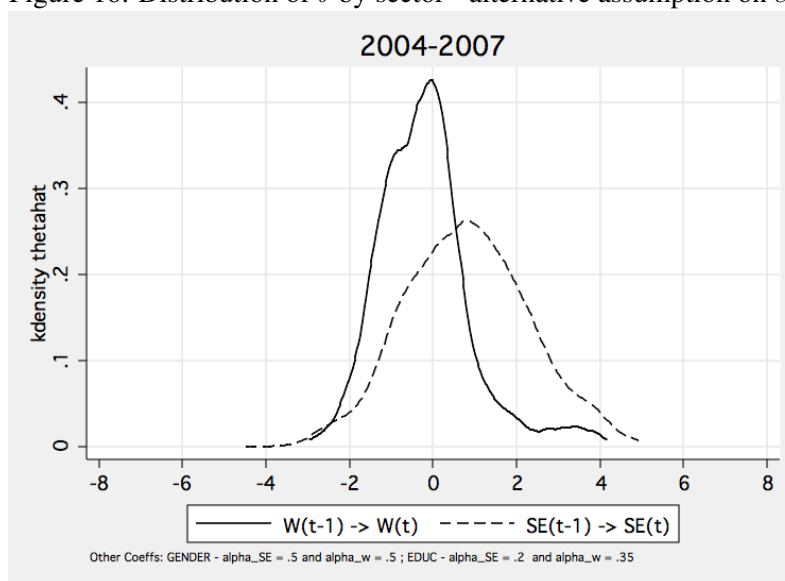


Figure 17: Distribution of $\hat{\theta}$ by sector - alternative assumption on α

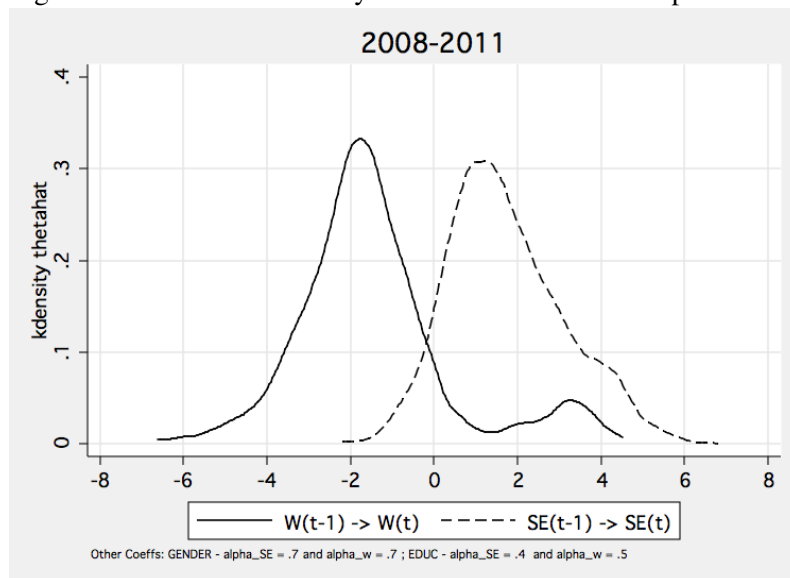


Figure 18: Distribution of $\hat{\theta}$ by sector - alternative assumption II on α

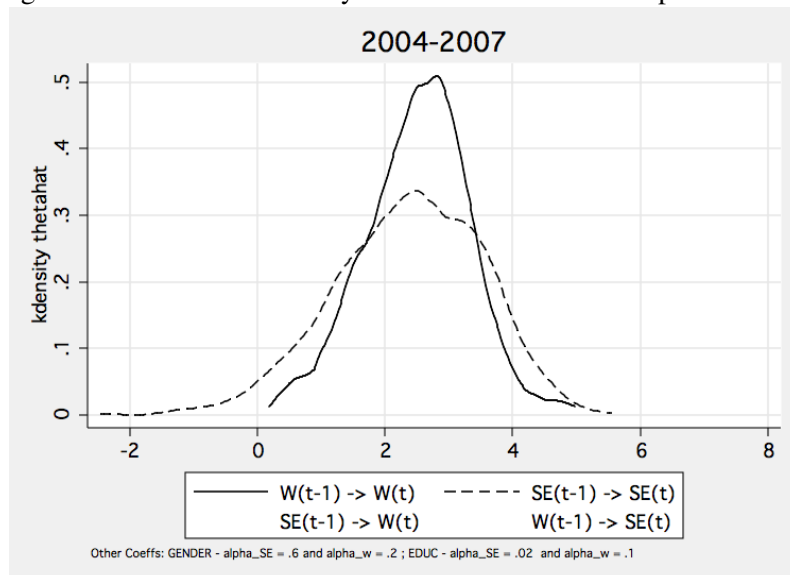
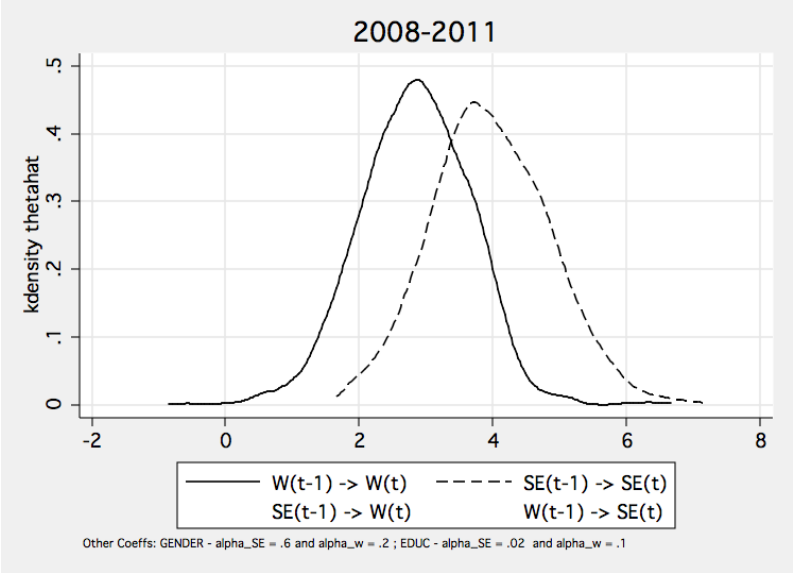


Figure 19: Distribution of $\hat{\theta}$ by sector - alternative assumption II on α



References

- ANDERSON, T. W. AND C. HSIAO (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47–82.
- 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, 277–97.
- BAAH-BOATENG, W. (2004): "Employment policies for sustainable development: the experience of Ghana," *University of Ghana*.
- BENZ, M. AND B. FREY (2003): "The Value of Autonomy: Evidence from the Self Employed in 23 countries," *University of Zurich Working Paper*.
- (2008): "Being Independent is a Great Thing: Subjective Evaluations of Self-Employment and Hierarchy," *Economica*, 75, 362–383.
- BLANCHFLOWER, D. (2004): "Self-Employment: More may not be better," *Swedish Economic Policy Review*, 11.
- BLANCHFLOWER, D. AND A. OSWALD (1998): "What Makes an Entrepreneur?" *Journal of Labor Economics*, 16, 26–60.
- BLANCHFLOWER, D., A. OSWALD, AND A. STUTZER (2001): "Latent entrepreneurship across nations," *European Economic Review*, 45, 680–691.
- BLUNDELL, R. AND S. BOND (2000): "GMM Estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19, 321–340.
- CHAMBERLAIN, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.
- EVANS, D. AND B. JOVANOVIĆ (1989): "An estimated model of entrepreneurial choice under liquidity constraints," *The Journal of Political Economy*, 97, 808–827.
- EVDOKIMOV, K. (2010): "Nonparametric Identification of a Nonlinear Panel Model with Application to Duration Analysis with Multiple Spells," *mimeo*.
- FALCO, P. (2012): "Risk-aversion and occupational choices: evidence from matched field experiments and survey data in urban Ghana," *mimeo*.

- FEDERICO HUNEEUS, O. L. AND E. PUENTES (2012): "Multidimensional Measure of Job Quality Persistence and Heterogeneity in a Developing Country," *Documentos de Trabajo, Universidad de Chile*, 357.
- GOLLIN, D. (1995): *Do taxes on large firms impede growth? Evidence from Ghana*, Economic Development Center, University of Minnesota.
- GOSLING, A. AND T. LEMIEUX (2004): "Labor market reforms and changes in wage inequality in the United Kingdom and the United States," in *Seeking a Premier Economy: The Economic Effects of British Economic Reforms, 1980-2000*, ed. by D. Card, R. Blundell, and R. B. Freeman, University of Chicago Press, 275–312.
- HARRIS, J. AND M. TODARO (1970): "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 60, 126–142.
- HECKMAN, J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables," *NBER Annals of Economic and Social Measurement*, 5, 120–137.
- HECKMAN, J. AND B. HONORE (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149.
- HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–95.
- IMBERT, C. (2012): "Decomposing wage inequality: public and private sectors in Vietnam 1993-2006," *World Bank Economic Review*, forthcoming.
- KEANE, M. AND K. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.
- KINGDON, G., J. SANDEFUR, AND F. TEAL (2006): "Labour market flexibility, wages and incomes in sub-Saharan Africa in the 1990s," *African Development Review*, 18, 392–427.
- LAUNOV, A. AND I. GUENTHER (2012): "Informal employment in developing countries: Opportunity or last resort?" *Journal of Development Economics*, 97, 88–98.
- LEMIEUX, T. (1998): "Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection," *Journal of Labor Economics*, 16, 261–291.

- MAGNAC, T. (1991): “Segmented or Competitive Labor Markets,” *Econometrica*, 59, 165–187.
- MAGNAC, T. AND J.-M. ROBIN (1996): “Occupational choice and liquidity constraints,” *Ricerche Economiche*, 50, 105–133.
- MURIS, C. (2010): “Efficient GMM estimation with a general missing data pattern,” *mimeo*.
- NSOWAH-NUAMAH, N., F. TEAL, AND M. AWOONOR-WILLIAMS (2012): “Jobs, Skills and Incomes in Ghana: How was poverty halved?” *Comparative Education*, 48, 231–248.
- PALMER, R. (2007): “Education, Training and Labour Market Outcomes in Ghana: A review of the evidence,” Tech. rep., RECOUP Working paper.
- POSCHKE, M. (2010): “Entrepreneurs out of Necessity A Snapshot,” *IZA Discussion Paper*, 4893.
- ROBSON, P., H. HAUGH, AND B. OBENG (2009): “Entrepreneurship and innovation in Ghana: enterprising Africa,” *Small Business Economics*, 32, 331–350.
- ROY, A. (1951): “Some thoughts on the distribution of earnings,” *Oxford Economic Papers*, 3, 135–146.
- SURI, T. (2011): “Selection and Comparative Advantage in Technology Adoption,” *Econometrica*, 79, 159–209.
- UDRY, C. AND S. ANAGOL (2006): “The Return to Capital in Ghana,” *Yale University Economic Growth Center Discussion Paper*, 932.