DIW BERLIN

1334

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2013

Entrepreneurship versus Joblessness: Explaining the Rise in Self-Employment

Paolo Falco and Luke Haywood

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

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

Paolo Falco† and Luke Haywood‡

November 5, 2013

Abstract

The self-employed constitute a large proportion of the workforce in developing countries and the sector is growing. Different accounts exist as to the causes of this development, with *pull* factors such as high returns to capital contrasted with *push* factors such as barriers to more desirable salaried jobs. Using data from Ghana, we investigate the changing structure of earnings in self-employment relative to salaried work. We decompose earnings in a two-sector labour market allowing for flexible patterns of sorting on unobservables by means of a correlated random coefficient model estimated by IV-GMM. A unique panel dataset provides us with suitable instruments to tackle the endogeneity of sector choice and capital accumulation. We show that returns to productive characteristics in SE have increased significantly over the period 2004-11 and the sector has attracted workers with higher skills. We conclude that pull factors have significantly strengthened, pointing against the grim view of self-employment as an occupation of last resort.

JEL: O15, J24, J42, C14
Keywords: self-employment, semiparametric models, comparative advantage, segmentation, African labour markets.

*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 Jörg Breitung, Andrew Clark, Marc Gurgand, Clément Imbert, David Margolis, Jean-Marc Robin, Francis Teal and seminar and conference participants in Aix-en-Provence, Düsseldorf, Göteborg, Marseille, Oxford and Paris. The usual disclaimer applies.

†University of Oxford, Manor Road, OX13UQ, Oxford, paolo.falco@economics.ox.ac.uk
‡DIW Berlin, Public Economics Department, Mohrenstr.58, 10117 Berlin, LHaywood@diw.de

1 Introduction

The labour markets of most developing countries are dominated by self-employment in its various forms: smallholder farmers, retail business owners, basic manufacturers and petty traders are generally self-employed. West-African labour markets are a particularly stark example, with small trading activities occupying a large fraction of the active population throughout the region. Moreover, recent large-scale household data have confirmed an increase in self-employment both in rural and (more prominently) in 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, albeit heterogeneous (Banerjee and Duflo (2005)). More individuals becoming self-employed may thus be the result of increased wealth relaxing capital constraints and creating profitable investment opportunities (Evans and Jovanovic (1989), Magnac and Robin (1996) and Blanchflower and Oswald (1998)). Self-employment under this perspective is often interpreted as entrepreneurship (i.e. the establishment of a business transforming capital and labour into output) and it may be a road to higher employment, reduced inequality and better working conditions. This interpretation explains the strong recent focus in policy circles on microcredit and entrepreneurship initiatives. On the other hand, a more *pessimistic* perspective is that for many individuals desirable salaried work is simply unavailable and in the absence of social protection or family transfers, self-employment becomes the only means of survival. In this light self-employment is more closely acquainted to unemployment and does not constitute a viable engine of development. In fact, some of the urban self-employed may be viewed as forming a queue for a restricted pool of wage jobs in a Harris-Todaro-type model with labour-market segmentation (Harris and Todaro (1970)). This more pessimistic view notes that many self-employed in developing countries operate with little to no capital, whilst wage jobs, often formal and in the public sector, are hard to access. In other words, self-employment may not constitute the preferred option in the presence of barriers to entry for wage jobs.

This article contributes to the debate by addressing three interrelated questions using new panel data from Ghana. First, what are the relative returns to observable *and unobservable* skills in self- and wage-employment? Second, how much of the observed earning differentials can be attributed to differences in workforce composition between sectors? Third, upon controlling for observable and unobservable determinants of productivity, what is the nature of the remaining unexplained differential, and is it consistent with the segmentation hypothesis? By investigating how the answers to these questions have evolved over the past decade (2004-2011), we aim to shed new light on the recent rise in self-employment.

A robust answer to these questions hinges upon being able to account for unobservable worker skills and concomitantly tackle the endogeneity of sector movement and input choices. This article takes unobservables seriously, focusing on endogenous sorting and selection on unobservables as key features of the modeling strategy. By means of a random-coefficient model estimated by Instrumental Variable GMM, we are able to improve on standard fixed-effect estimation, which is unable to control for individual unobservables if their effect varies across sectors. Furthermore, we do not wish to make the assumption - standard in classic selection models - that certain variables monotonously increase the likelihood of selection into one of the sectors, the so-called "single-crossing property". For instance, since many wage-jobs are in the public sector, market forces may in fact compete with other, more complex, selection mechanisms (e.g. based on social capital). Similarly, the costs of entry into self-employment may vary in ways related to unobservables, for example by way of differences in access to credit, which requires a more flexible model of selection. Imposing minimal structure on the sorting process, our model allows for rich patters of endogenous selection. Moreover, a uniquely long panel dataset provides us with suitable instruments to tackle the endogeneity of sector-choice and capital accumulation. Existing studies are often unable to deal with this issue and need to rely on stronger assumptions about the exogeneity of sector allocation and input factors.

A burgeoning literature investigates similar questions on the nature of self-employment world-wide. Among the most recent studies, Poschke (2010) uses *Global Entrepreneurship Monitor* data about declared reasons for self-employment to investigate the phenomenon. In his dataset individuals state whether their self-employment is voluntary or not (with the latter named "necessity self-employment"). He is able to identify a number of prevailing characteristics among the "self-employed out of necessity": they typically own smaller businesses, they are less educated and more likely female than the ”self-employed out of choice”. 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” versus “last resort” category of informal employment, and conclude that the latter comprises around 45% of the entire informal sector.

If the increase in self-employment in Ghana is largely due to motivations of “last resort”, we should expect levels of capital, schooling and unobserved ability to be declining over time among the self-employed (as the best select out of the sector); and, in parallel, we might expect to find a significant (and possibly increasing) wage-employment premium. On the other hand, more optimistic *pull* factors leading to an increase in self-employment would include growing returns to human and physical capital and a decreasing wage premium. These are the empirical findings we concentrate upon.

Our results show that returns to both observable and unobservable skills in self-employment have increased vis-à-vis wage-employment over the period analysed. Moreover, self-employed workers possess increasing levels of physical capital and human capital. Taken together, the two findings lend support to the hypothesis that the sector is increasingly attracting (i.e. pulling) productive resources. Furthermore, we document a decreasing (unexplained) self-employment premium, which we interpret as evidence of decreasing barriers to enter the sector. This result is highly consistent with the strong growth rates experienced by Ghana over the period, where poverty was significantly reduced and capital accumulation may have contributed to the progressive relaxation of credit constraints. We conclude that the returns to being self-employed in Ghana have increased significantly over the past decade. While the sector remains highly heterogeneous, our evidence points against the grim view of self-employment as an occupation of last resort.

The paper proceeds as follows. Section (2) describes the context of the Ghanaian labour market. In section (3) we present our earnings model with a specific focus on selection mechanisms. Section (4) considers identification and presents our empirical strategy based on a correlated random coefficients model estimated by means of IV-GMM. Section (5) presents the data. Section (6) discusses the results. Section (7) considers factors beyond earnings that may influence occupational choices and section (8) concludes. The appendix contains numerous robustness checks of the key results.

2 The 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. Nsowah-Nuamah et al. (2012) find that poverty was halved over the period from 1991 to 2005. They also find that this occured 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, bauxit), 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 here >>

Reporting evidence from the Ghana Statistical Service, Baah-Boateng (2004) shows 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 spread of self-employment may be 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-employment in the economy, with the possibility of a trend reversal.[1] The panel dataset at our disposal (GHUPS) is not well-suited to settle this debate, since it follows the same group of workers over time (census data would be much better placed to resolve the controversy). However, by carefully recording the movement of workers between wage- and

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

self-employment, our data is ideally suited to identify changes in the structure of earnings and study the relative performance of wage- and self-employment over time.

3 The Model

In this section we describe a simple model capturing the key features of the earning structure in our two sectors of interest: 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-invaryng characteristics (e.g. human capital h_i) and unobservable productivity θ_i^j for $j \in \{SE, w\}$ varying across two 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, over the relatively short sample period, and do not consider the interactions between sector choice and capital accumulation.[2] Given these endowments, worker i optimally chooses between two alternative *employment strategies*: working as a self-employed or searching for a wage-job.

Earnings in wage (w) and self-employment (SE) can be described in general terms using the following log-linearised specification:

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

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

where $K_{i,t}$ and h_i indicate physical and human capital; δ_t^{SE} designates macroeconomic

[2] One could endogenize physical capital accumulation in line with Magnac and Robin (1996) and human capital in line with Keane and Wolpin (1997) in a structural framework.

effects on the self employed (viz. $\delta_t^w - \delta_t^{SE}$ the time-changing wage employment premium) and $u_{i,t}$ subsumes individual idiosyncratic factors.

We can think of β^w as the market rate of return on savings (e.g. interest rate on bank deposits), while β^{SE} captures returns to capital investment in self-employment. 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. Using measures of capital for workers in *both* sectors, we will be able to estimate relative returns to assets across employment catoegories.

Based on this simple earnings model, workers are allowed to sort into self-employment or salaried work. In what follows, we will allow maximal flexibility in the sorting mechanism and rather concentrate on its implications for the structure of earnings across 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 we now show how we explicitly 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 opportunities for salaried work on one side (provided by larger, more likely formal, private firms and by the public secor) and self-employment on the other side. 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 existing work. 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 e.g. the Harris-Todaro framework suggests), we need to take into account not only self-selection by workers, but also 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, such as personal connections and political affiliation, may also play a role in firms' selection choices, e.g. for access to the public sector (a large employer in many developing countries).[3] Rationing could also be the result of efficiency wage setting, institutional constraints or imperfect information.[4] For these reasons, we believe one should best remain agnostic about the exact nature and direction of sorting. For our purposes, 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 operate in their chosen sector. Magnac (1991) models barriers to entry as costs $c(.)$ resulting from queueing with a probability in every period of τ of remaining unemployed. This implies that we can expect $c(.)$ 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 to become self-employed (d^{SE}) as follows:[5]

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

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

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

[4]Modeling 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.

[5]Given 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.

[6]This 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

includes observable characteristics (e.g. h, K in equations (1),(2)), we can expect sorting on observables based on relative returns net of differential entry costs. Third, to the extent that x_i includes unobservable worker characteristics (e.g. θ_i in (1),(2)), we must flexibly allow for their role in the sorting process.

Given that individual unobservables are a key driver of sector choice, a random effects framework will not be suitable to estimate the model. A naïve fixed effects framework will also be inappropriate since our model allows individual abilities to be rewarded differently across sectors. Furthermore, it may be the case that institutions select individuals with comparative advantage in self-employment to be wage employees (e.g. if wealthier individuals, who could invest their resources to start a business, also have easier access to public-sector jobs). In such cases, the classic assumptions of selection models would be violated. Indeed, standard two-stage models that instrument the sector choice or rely on a specific parametrization to control for endogenous selection are commonly based on the assumption that certain factors *monotonously* increase the likelihood of finding a wage job (Heckman (1976)). We believe this assumption is too strong and we could not identify valid instruments driving selection, but not earnings in our data. Our plan is, instead, to derive a consistent estimate of $\Delta R(x_{i,t})$ allowing for unspecified sorting operating via a general entry costs $c(.)$.

4 Empirical Strategy

4.1 The Correlated Random Coefficients Model

In this section we outline our strategy to estimate the earnings model and identify sector differentials under the assumption that job queues may exist and selection is not necessarily based on a single threshold (as in standard self-selection models). To achieve this goal we devise a panel-based procedure that enables us to identify the differential cross-sector effect of competing earning determinants, whilst allowing for unspecified patterns of selection on unobservables. The estimation procedure rests on careful observation of the movers between

those who find employment.

the two employment states considered here, and on few functional form assumptions.[7]

Combining (1) and (2), we construct the following switching model of earnings in the two sectors (using d_i^{SE} as a dummy indicating self-employment):

$$\begin{aligned} R_{i,t} = \delta_t^w + d_{i,t}^{SE}\left(\delta_t^{SE} - \delta_t^w\right) \qquad (5) \\ + \alpha^w \ \ln h_i + d_{i,t}^{SE} \ \ln h_i \left(\alpha^{SE} - \alpha^w\right) \\ + \beta^w \ \ln K_{i,t} + d_{i,t}^{SE} \ \ln K_{i,t} \left(\beta^{SE} - \beta^w\right) \\ + \theta_i^w + d_{i,t}^{SE} \ \left(\theta_i^{SE} - \theta_i^w\right) + \varepsilon_{i,t} \qquad (6) \end{aligned}$$

where

$$\varepsilon_{i,t} \equiv \ u_{i,t}^w + d_{i,t}^{SE}\left(u_{i,t}^{SE} - u_{i,t}^w\right). \qquad (7)$$

Moreover, we further break down the unobservables $(\theta_i^{SE}, \theta_i^w)$ into an individual effect that is remunerated equally in both sectors and plays no role in sector choice (τ_i), and an unobservable effect θ_i that is remunerated differentially in the two sectors and hence generates a comparative advantage. We also introduce a parameter ψ indicating the average returns to unobservables in the SE vis-à-vis the wage sector.[8] Combining these additional elements,

[7]A key assumptoin is the multiplicativity of ψ. The procedure follows Lemieux (1998), who considers the union wage premium and has recently been applied by Suri (2011) to technological change.

[8]As 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.

we obtain:[9]

$$\theta_i^{SE} = \psi\ \theta_i + \tau_i \tag{10}$$

$$\theta_i^{w} = \theta_i + \tau_i \tag{11}$$

Earnings can then be written as follows:

$$\begin{aligned}
R_{i,t} = \delta_t^w + d_{i,t}^{SE} \left(\delta_t^{SE} - \delta_t^w\right) \\
+ \alpha^w\ \ln h_i + d_{i,t}^{SE}\ \ln h_i \left(\alpha^{SE} - \alpha^w\right) \\
+ \beta^w\ \ln K_{i,t} + d_{i,t}^{SE}\ \ln K_{i,t} \left(\beta^{SE} - \beta^w\right) \\
+ \theta_i + d_{i,t}^{SE}\ (\psi - 1)\ \theta_i + \tau_i + \varepsilon_{i,t}
\end{aligned} \tag{12}$$

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 determines incentives for sorting on unobservables. Our preferred choice is to remain agnostic about the actual selection mechanism when estimating the relative returns (ψ).[10] In summary, if we wish to

[9] Using 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 (8) and (9) are the same by subtracting (9) from (8) and noting that $b_{SE} + b_w = 1$ by construction.

$$\theta_i^{SE} = b_{SE}\ \left(\theta_i^{SE} - \theta_i^w\right) + \tau_i \tag{8}$$

$$\theta_i^{w} = b_{w}\ \left(\theta_i^{SE} - \theta_i^w\right) + \tau_i \tag{9}$$

$$\theta_i \equiv b_{w}\ \left(\theta_i^{SE} - \theta_i^w\right)$$

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 (10) and (11) 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 (10) and (11) directly follow.

[10] Note 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).

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.

4.2 Identification

Model identification relies on the classic panel data restriction that the idiosyncratic error terms $u_{i,t}$ be uncorrelated with the covariates in all time periods. In other words, movement across sectors depends on expected earnings and not on realisations of the wage shock $u_{i,t}$. Note that this allows for rich patterns of selection on the unobservables θ_i. Clearly, large income shocks may influence sector choice, though they are unlikely to explain the bulk of sector transitions. However, we note that if sector choice can operate in advance of such shocks (e.g. due to formalised recruitment cycles in the wage sector or informal safety nets cushioning the self-employed), sector allocation may be considered a pre-determined variable, leaving us with a smaller but sufficient set of moment restrictions (i.e. using longer lags). The same reasoning applies to capital. Though shocks may deplete individuals' capital levels (violating the assumption of strict exogeneity), it may still be reasonable to assume pre-determinedness of capital holdings if the effect occurs with a lag. We discuss this in more detail in section (4.3), where we outline in detail how the strict exogeneity assumption can be relaxed.

Identification rests on the following intuition. The sector-specific coefficients on the time-varying covariates (β^w, β^{SE}) are identified through variation in returns across different levels of those covariates *within* each sector.[11] The mechanism that identifies ψ, instead, is best explained by noting that standard fixed-effects models are overidentifying when movers across sectors are observed. Differential fixed effects across sectors are then identified by time-demeaning in both sectors - as long as we observe movers.

As in standard fixed-effects frameworks, we rely on a model transformation to suppress the unobservable fixed effects from the estimation. However, standard differencing or time de-meaning will not remove the individual effect (θ_i) in our setup. Instead, following

[11]For time-invarying characteristics, instead, we can only identify the effect of the quasi-difference, as outlined below.

Chamberlain (1982) and Lemieux (1998) we solve explicitly for θ_i in (12) and replace the following analytic expression in the general earnings equation.

$$\theta_i = \frac{R_{i,t} - \left[\delta_t^w + D_{i,t}\ (\delta_t^{SE} - \delta_t^w) + D_{i,t}\ \beta^w\ K_{i,t} + D_{i,t}(\beta^{SE} - \beta^w)\ K_{i,t} + \alpha^w\ h_i + D_{i,t}\ (\alpha^{SE} - \alpha^w)\ h_i + \tau_i + \varepsilon_{i,t}\right]}{1 + D_{i,t}\ (\psi - 1)} \tag{13}$$

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

For convenience of exposition, we define:

$$\begin{aligned} G_{i,t}(K,h,D) &\equiv \beta^w\ K_{i,t} + D_{i,t}(\beta^{SE} - \beta^w)K_{i,t} + \alpha_w\ h_i + D_{i,t-1}\ (\alpha^{SE} - \alpha^w)\ h_i \\ and\ \ \delta_t &\equiv \delta_t^w + D_{i,t}\ (\delta_t^{SE} - \delta_t^w) \end{aligned}$$

Then, combining (13) with the equivalent expression at $t-1$, we obtain:

$$\frac{R_{i,t} - G_{i,t}(K_{i,t}, h_i, D_{i,t}) - \delta_t - \tau_i - \varepsilon_{i,t}}{1 + D_{i,t}\ (\psi - 1)} = \frac{R_{i,t-1} - G_{i,t-1}(K_{i,t-1}, h_i, D_{i,t-1}) - \delta_{t-1} - \tau_i - \varepsilon_{i,t-1}}{1 + D_{i,t-1}\ (\psi - 1)} \tag{14}$$

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

$$R_{i,t} = G_{i,t} + \delta_t + \varepsilon_{i,t} + \tau_i + C_{i,t}\ \left(R_{i,t-1} - G_{i,t-1} - \delta_{t-1} - \varepsilon_{i,t-1} - \tau_i\right) \tag{15}$$

We have thus removed the influence of the unobservables θ_i and, recognizing that τ_i is a random effect (conditional on θ_i), we can now consistently estimate (15).[12]

The transformed model yields the following quasi-differenced errors, which can be used to construct a set of moment conditions:

$$\begin{aligned} erm_{i,t}\ &\equiv \left(R_{i,t} - G_{i,t} - \delta_t\right) - C_{i,t}\ \left(R_{i,t-1} - G_{i,t-1} - \delta_{t-1}\right) \Leftrightarrow \\ erm_{i,t}\ &\equiv \left(\varepsilon_{i,t} + \tau_i\right) - C_{i,t}\ (\varepsilon_{i,t-1} + \tau_i) \end{aligned} \tag{17}$$

[12]Given our quasi-differenced framework, the final required step is a simple normalisation of the individual effects(analogous to standard differenced models where time constants are not identified unless the fixed effects are normalized). Our chosen normalization is the following:

$$\frac{1}{T\ N}\sum_{t=1}^{T}\sum_{i=1}^{N}\theta_{i(t)} = 0 \tag{16}$$

Another way of expressing these quasi-differenced errors considers the different sector histories (i.e. the combinations of all possible employment histories over two consecuitve periods):[13]

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

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 sectors.[14]

Next, note that by assumption, $D_{i,t}$ is not correlated with $\left(\varepsilon_{i,t} + \tau_i\right)$.[15] However, since $R_{i,t-1}$ is by construction correlated with $\varepsilon_{i,t-1}$ and hence also with $erm_{i,t}$, we will need to instrument for $R_{i,t-1}$. The requirement is that an instrument be correlated with $R_{i,t-1}$ but not with $\varepsilon_{i,t}$. The assumption of strict exogeneity (the standard assumption in panel data models which we will relax), combined with the random effects character of τ_i (see (10) and (11)) implies that any combination (past and present) of covariates will satisfy this requirement. When the assumption of strict exogeneity is relaxed (see the next section for a detailed discussion) only a sub-set of those instruments will satisfy the requirement (i.e. those generated by 'sufficiently long lags' of the endogenous variables). Once a set of valid instruments is identified, the model is estimated by minimising the empirical sum of moment

[13]That 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 $s \in \{t, t-1\}$.

[14]More 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 $\alpha^{SE} - \frac{1}{\psi}\ \alpha^{w}$

[15]Whilst $D_{i,t}$ does appear in the expression for ε, this is unproblematic as long as $D_{i,t}$ is orthogonal to $u_{i,t}^{SE} - u_{i,t}^{w}$, see equation (7).

conditions:

$$E(erm_{i,t}\ Z_{i,t}) = 0 \tag{19}$$

where $Z_{i,t}$ is a vector of valid instruments.

4.3 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 the assumption of strict exogeneity. In other words, capital may 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 thus exploit its longitudinal dimension. Following Anderson and Hsiao (1982), Holtz-Eakin et al. (1988) and Arellano and Bond (1991), we can 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.

$$E[K_{is}\varepsilon_{i,t}] = 0 \quad \forall\ \ t \geq s \tag{20}$$

$$E[K_{is}D_{i,t}] = 0 \quad \forall\ \ t \geq s \tag{21}$$

$$E[K_{is}\tau_i] = 0. \tag{22}$$

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} (i.e. the minimum lag length for a valid instrument is equal to 2). In this case, we have fewer exclusion restrictions:

$$E[K_{is}\varepsilon_{i,t}] = 0 \quad \forall\ \ t > s \tag{23}$$

$$E[K_{is}D_{i,t}] = 0 \quad \forall\ \ t > s \tag{24}$$

Allowing for sequential (as opposed to strict) exogeneity increases data requirements: we now need three waves of consecutive information on each individual, as we aim to instrument period $t-1$ using information from period $t-2$. To conserve space, we document the details of how we construct our instrument matrix whilst allowing for the potential endogeneity of capital in appendix (A).

5 Data

We apply our model to 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 carried out in 2004 and now spans 8 years, an unusual length for panel data-sets in developing countries. The GHUPS covers the four largest cities in the country: 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.[16]

Figure (1) shows the split of employed workers in the sample into self-employed and salaried workers. Note that the increase in self-employment is not as marked here as in other datasets. The panel design, with its focus on reducing attrition, may have led to an understatement of changes in the split of self- and wage-employment. However, such changes would be expected to have a strong influence also on the wage structure, something the data is very well suited to research.

<< Figure 1 here >>

Table (2) shows changes in the workforce composition for the sample of GHUPS respondents who are employed at the time of the interview (pooling the first four waves (2004-7) and the second four waves (2008-11), as in our subsequent analysis).

<< Table 2 here >>

[16] 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.

What constitutes self-employment in our predominantly urban sample? Figure (2) shows that the largest self-employed category 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)).

<< Figure 2 here >>

<< Figure 3 here >>

<< Figure 4 here >>

<< Figure 5 here >>

5.1 Trends in Earnings

Table (2) shows that whilst average earnings in wage employment increased by roughly 50% from the first to the second half of the panel, they more than doubled in self-employment, leading to a reversal of the gross earning differential. However, in the most recent period the difference in *median* earnings between the two sectors is not large (figure 6) - and not statistically significant (not shown). Moreover, figure (7) shows an increse in the variance of earnings, especially for self-employed workers, alongside considerable real earnings growth. Strong earnings growth among the self-employed is also shown in figure (8), which reveals that real earnings have increased considerably faster among those workers than for employees in the wage-sector. This is prima facie evidence against the common negative view of self-employment as an occupational category inferior to salaried employment and populated by workers with lack of alternatives in the wage sector (the *push* interpretation of the rise in self-employment).

<< Figure 6 here >>

<< Figure 7 here >>

<< Figure 8 here >>

5.2 Transitions between Wage and Self-employment

Our empirical identification relies crucially on job-movers. Table (3) summarises respondents' transitions between the two sectors over consecutive years of the panel, pooling 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 only use data on the respondents who are employed and interviewed in all eight years (and for whom none of the necessary variables is missing), we are left with 42 observations. Decreasing the number of waves jointly used in the estimation, as we will do below, substantially increases sample size (see Table 4).

<< Table 3 here >>

<< Table 4 here >>

Upon investigating their characteristics the movers appear to be younger (figure (9)), while there is no difference in their relative levels of education (not shown).

<< Figure 9 here >>

5.3 Measures of Capital

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 levels and variance between wage and self-employed (see figure (10) and table (2)), especially in the later period. 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.

<< **Figure 10 here** >>

We are interested in the income stream or usage value generated by the assets respodents 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 (though we acknowledge that heterogeneous returns may matter)[17]. Figure (11) plots the (real) value of tools/equipment against earnings, including imputed capital usage values in the latter.

<< **Figure 11 here** >>

Capital holdings may be derived from past labour market outcomes. This would mean that earning shocks in the past may be correlated with current capital holdings. This type of reverse causality motivates our assumption that capital levels are *predetermined* - as explained in detail in section (4.3).

<< **Figure 12 here** >>

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 is 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

[17] The exact assumption we are making is that assets yield a 2% annual usage value, which we divide by 12 and add to the monthly earnings figure. 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.

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.

Investigation of the first-stage results from our IV-model shows that power is higher in samples with larger N. Thus, we focus on two three-year samples in the analysis, choosing the years at the two extremes of the calendar period in order to focus on the evolution of the labour market over time. For these reasons, our preferred approach is to estimate the model over two three-wave panels: from 2004 to 2006 and from 2007 to 2011. Standard errors are bootstrapped, but the relatively small number of movers in the data constrains our ability to perform certain additional robustness checks (out-of-sample prediction, comparing estimates for voluntary and involuntary movers etc.).

6 Results

6.1 IV-GMM

In this section we present the main results from estimating our earnings model using the Instrumental Variable Generalized Method of Moments (IV-GMM) and allowing for saving and dissaving of capital stocks ("endogenous capital"). We contrast these to estimates from simpler OLS and Fixed Effects models, obtained under the assumption that ψ is equal to 1. Appendix C includes various robustness checks and presents alternative results assuming that capital levels are exogenous.[18]

<< Table 5 here >>

Our key results are presented in table (5). We find increasing returns to productive characteristics in self-employment and a decreasing self-employment premium between the first and the second half of the sample period. More specifically, four aspects stand out.

[18]Relaxing the assumption of capital exogeneity increases the information needed because longer lags need to be used to instrument for past earnings. We focus on the results allowing for saving and dissaving of capital from earnings shocks since we believe this to be an important element of sector transition dynamics. Appendix (C.6) finds the same qualitative pattern in the results assuming strict exogeneity - all trends in returns to factors are the same, though the estimated coefficients vary significantly.

First, and perhaps unsurprisingly, *returns to capital* are significantly higher in self-employment than in the wage-sector, for which we have assumed $\beta^{SE} = 0.2$.[19] Moreover, returns to capital in self-employment appear to increase significantly between the first and the second half of the period. Our specification assumes a Cobb-Douglas-type production function (with constant elasticity) such that returns to capital vary over capital levels and earnings. Figure (13) plots the 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 returns to capital vary significantly, median annualised values are 70.21% in the first period and 80.35% in the second. These rates (compounded from monthly rates implied by $\hat{\beta}$) may appear high. 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 practiced in informal saving schemes are often negative in real terms, e.g. Susu collectors), they provide a strong incentive for holders of capital to start a self-employed business. The comparative advantage of being self-employed for individuals with enough capital has increased over time in all our specifications. The strong increase in β^{SE} is robust to changes in all calibrated parameters (as shown in appendix (C)) and it is also present in the OLS, though not in the FE, results (see table (6)). This result is an important contribution to the existing literature, which has often been unable to explicitly control for endogenous sector selection when estimating returns to capital in self-employment.

<< Figure 13 here >>

Second, *returns to human capital* (RORE) have significantly increased in self-employment over the period analysed. Since only the quasi-difference across sectors ($\alpha^{SE} - \psi\ \alpha^{w}$) of

[19]This assumption might be challenged. Most evidence points to very low (and sometimes even negative) real returns on liquid assets in Ghana. For instance, the Ghana Household Urban Panel survey reveals that 20% of respondents save some money with a Susu collector. Susus are a typical saving scheme whereby people deposit their money with a local informal financial institution (i.e. the Susu) and collect it back periodically (usually once a month). This saving device, which helps them keeping money away from immediate consumption, costs them a fee and pays no interest. Another (simpler) piece of evidence is that despite the urban focus of our survey, the majority of our respondents don't have a Bank account, and therefore have no access to formal interest-earning bank accounts. Appendix (C) considers changing the value of β^{w} by orders of magnitude but finds that results remain very similar.

time-invarying characteristics is identified, we set the value of returns to formal education to 8% p.a. in wage-employment (in line with the average found by Fasih et al. (2012)). In section section (C) we show that our results are not sensitive to increasing this figure up to 20% (the highest RORE found in studies of (higher) education in Ghana). We find that during the $2004 - 2008$ period, returns to education in self-employment are lower than in wage employment (see table (5)). The estimates for the more recent period, however, imply a strong growth in RORE among the self-employed between the first and the second half of the panel. This result demonstrates the increasing benefits of formal education, even in a country with high levels of informal self-employment. The trend is consistent across estimators and various robustness checks (see section (C.3) in the appendix).

Third, although *returns to unobservable skills* appear to be significantly lower for the self-employed than in the wage sector in both periods, the difference is shrinking. The estimated value of ψ in table (5) changes from below 0.5 for the period 2004-2007 to being larger than 0.8 in subsequent years, indicating that self-employment is *increasingly* becoming a desirable opportunity. However, when we consider changes in the composition of the workforce by analysing the distribution of estimated θ_i values across sectors, we do not detect an increase in the proportion of individuals with high unobserved skills (high θ_i) in self-employment over the period, in line with the continued incentives for able individuals to join the wage sector arising from $\psi < 1$.

Fourth, we find a decreasing self-employment premium ($\delta^{SE} - \delta^{w}$), suggesting that the barriers to enter this sector have weakened over time. This interpretation is highly consistent with the strong growth rates experienced by Ghana over the period, when poverty was significantly reduced and capital accumulation may have facilitated access to entrepreneurial opportunities. In analysing the coefficients, one must remember that they cannot be interpreted on their own, since no individuals have zero levels of all factors of production. We focus the discussion on the trend of this gap over time, which indicates a significant change.

How can our four key results be reconciled and what do they tell us about recent developments in the Ghanaian labour market? The following section relates our estimates to compositional changes over the period of interest and proposes an interpretation.

6.2 Understanding the Rise in Self-employment

We now employ our results to study changes in the composition of self-employment, focussing on the sectoral distribution of *observed and unobserved* factors of production. One concern driving this exercise is that in the presence of diminishing returns to input factors, such as education and other unobservable skills,[20] the estimated increase in returns may be a result of decreased deployment of those factors in a given sector. In other words, with less educated individuals in self-employment, marginal returns to education may be on the rise. However, when we attempt the decomposition of observed skills, we find *increasing* levels of factor inputs in self-employment for all observable factors of production, as already shown in table (2). By contrast, levels of *estimated unobserved skills (ability)* are higher in the wage sector and we do not detect a significant change between the two periods (in line with the fact that, although significantly increasing, ψ remains below 1 throughout).

Capital holdings increase substantially among the self-employed from an average of 203.24 USD in the first half to 473.11 USD in the second half of the period. By contrast, they fall among the wage-employed (from a mean (real) value of 336.11 US dollars to a value of 313.14). For schooling the differential trends are less marked (but statistically significant), with the level of formal schooling among the self-employed rising from 7.45 to 7.94 on average, whereas in the wage sector the same figure is essentially constant at 9.81 and 9.76 respectively.

In order to compare the levels of *unobserved skills* in the two sectors across time, we use the analytic expression for θ_i derived above (eq. 13), averaging over different years to obtain a consistent estimate. Unfortunately, only few observations underlie these calculations (a maximum of 4 per person if we use the 4-wave panel 2004-2007), thus our estimtaes suffer from small-sample bias. However, since our estimator is consistent, it is hoped that the

[20] Note that with respect to capital we are in a constant-elasticity Cobb-Douglas specification: our log-log formulation assumes decreasing returns to capital.

distribution of θ will be approximately correct. Graphs (14) and (15) show that the variance of skills in self-employment is very high, especially in the earlier period. In the first period our estimates show that average unobserved productivity is higher in the wage sector than in self-employment. This pattern does not change markedly in the later period, though the variance in self-employment decreases significantly.

<< Figure 14 here >>

<< Figure 15 here >>

In summary, it appears that changes in returns are driving changes in the composition of the workforce (and associated input factors), not vice-versa. The exception to this rule is individuals with high unobserved ability, who still receive higher returns in wage-employment (since $\psi < 1$) and, despite the increase in ψ we documented in the previous section, are not sufficiently attracted by the prospects of self-employment. This could change if returns to ability in self-employment increased further in the future ($\psi > 1$).

Upon decomposing the earning gap into *price effects* and *selection effects* (table 12 in the appendix), we find that the difference in the level of unobservable skills between sectors outweighs the differential effects of all other factors. In other words, despite a substantial increase in the returns to self-employment and a significant improvement in the composition of the self-employed workforce, unobserved worker characteristics remain a salient determinant of earnings. We discuss the decomposition in more detail in section (D) in the appendix.

6.3 Comparison with Other Estimators

Table (6) shows benchmark results from estimating our model with OLS (columns 1-2) and a standard Fixed Effect estimator (columns 3-4). The latter is nested in our previous model, and corresponds to the case where $\psi = 1$. We can therefore test the implications of this restriction, which is common in the existing literature.

<< Table 6 here >>

In the OLS and FE results (reported in table (6)) we detect a significant decrease in the self-employment premium ($\delta^{SE} - \delta^{w}$) from the first to the second half of the period. The result, however, is not significant once we control for fixed effects. The reduction in sector premia between the OLS and FE results is generally interpreted as showing that unobserved heterogeneity (e.g. ability) drives sector selection. How can we explain that by allowing $\psi \neq 1$ (i.e. in the benchmark IV-GMM results) we obtain significant sector premia that are absent in FE-estimates? Selection appears to be driven by differential remuneration of skills across sectors over and above the common effect of unobservables in all sectors. Once we allow for unobservables to be remunerated differentially in wage- and self-employment, it appears that, ceteris paribus, it is more attractive to be in the latter category. Both in OLS and in the IV-GMM results, the SE premium declines over time (and it is in fact reversed in OLS in the second half of the period).

Cross-sector differentials in the gender premia and returns to schooling are estimated more precisely in OLS than FE (as expected). Once we control for fixed effects, we no longer detect a difference in gender premia and the difference in returns to schooling is only significant in the first half of the panel. The change in the coefficient on education, however, suggests that returns to schooling in self-employment have increased over time relative to wage-employment, consistently with the IV-GMM results. Returns to capital appear to increase over time in OLS, while they remain stable in the fixed effect model. As for the raw sector differentials discussed above, it appears that allowing for $\psi \neq 1$ in the IV-GMM model was crucial to to uncover growth in β^{SE}.

7 Beyond Average Earnings

The framework presented so far assumes that workers maximise expected material gains. Alternative determinants of sector choice may however be invoked.

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)). Falco et al. (2012) find similare evidence in urban Ghana among entrepreneurs who employ others. This may indicate that working conditions, managerial independence, flexibility (or any other characteristics of self-employment) may be valued in addition to material compensation. A cursory glance at average levels of job-satisfaction[21] in table (2) shows a reduction in job-satisfaction among salaried workers over the period, alongside an increase among the self-employed. 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 Huneeus et al. (2012) do for Chile). Alternately, one could consider job-satisfaction 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 future work.

Second, not only the amount, but also the variance of earnings may be a criterion for occupational choice. Differences in risk aversion may explain different choices between self-employment and wage work for given levels of capital. 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 raising the attractiveness of self-employment (and part of the self-employment premia we estimate may in fact amount to risk-premia).[22]

8 Conclusions

Informal self-employment is the most common form of occupation throughout the developing world. The share of self-employed workers in the labour force is often very large and has been on the rise in recent decades. This article empirically investigates changes in the structure of earnings in self-employment and wage-employment in Ghana, over the period 2004-2011. Our results may help determine whether the rise in self-employment is the result

[21] 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).

[22] Given that some experimental data on risk aversion is available for the population we are studying (see Falco (2013)), the quantitative implications of taking into account a positive level of risk aversion may be considered.

of improved opportunities for successful entrepreneurship (*pull factors*), or the reflection of limited opportunities in wage-employment (*push factors*).

We construct a two-sector earnings model for wage and self-employment allowing for rich patterns of selection on observable and unobservable worker characteristics and for differential returns to unobservable factors, going beyond previous panel data analyses of developing countries' labour markets. We estimate the model using a unique panel dataset from urban Ghana, covering a representative sample of workers over 8 years in the four largest cities of the country and crucially including information on individuals moving from self- to wage-employment. The model is estimated using the Instrumental Variable Generalized Method of Moments (IV-GMM), with valid instruments to tackle the endogeneity of sector selection and imput choice. Contrasting our results with ordinary least squares and fixed effects estimators demonstrates the importance of allowing for differential returns to unobservable characteristics and endogenous sector allocation. The results are consistent across numerous robustness checks. We highlight three main findings.

First, we find evidence of increasing returns to productive factors in self-employment over the period analysed. Returns to capital, to schooling and to unobserved skills are all found to have increased significantly from 2004 to 2011. Returns to capital have increased substantially, from levels already above those in wage employment at the beginning of the panel. Returns to schooling have also increased, but we find conflicting evidence on their strength relative to the RORE in wage-employment. Returns to individual unobservables in self-employment have also trended upwards according to our model, but they remained higher in wage-employment throughout the period (a result that is robust across several specifications). Overall, our results indicate that the incentives for capital-rich, educated and otherwise able individuals to sort into self-employment have increased considerably over a decade.

Second, we find that the incentives generated by higher returns have started to work. Individuals with better productive characteristics are increasingly found in self-employment (where average levels of capital and education have increased significantly over the period, relative to wage-employment). This is evidence of an increasingly strong *pull* effect. However, despite the upward trend leading 'talents' into self-employment, we estimate that unobserved individual determinants of productivity are still higher (and more highly rewarded) in wage-employment.

Third, once our model controls for both observable and unobservable worker-characteristics, we find a sizeable self-employment premium which has decreased significantly over time. We interpret this decrease as evidence of increased sector mobility and lower barriers to self-employment. This interpretation is highly consistent with a period of strong growth and significant poverty reduction, when capital accumulation may have helped relaxing credit constratins.

We conclude that the returns to being self-employend in Ghana have increased significantly over the past decade and this appears to have attracted (i.e. pulled) productive resources and talents. While the sector remains highly heterogeneous, our evidence points against the grim view of self-employment as an occupation of last resort.

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*.

Banerjee, A. V. and E. Duflo (2005): "Growth Theory through the Lens of Development Economics," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1 of *Handbook of Economic Growth*, chap. 7, 473–552.

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.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Evans, D. and B. Jovanovic (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. (2013): "Does risk matter for occupational choices? Experimental evidence from an African labour market," *CSAE Working Paper Series*.

Falco, P., W. F. Maloney, B. Rijkers, and M. Sarrias (2012): "Heterogeneity in subjective wellbeing : an application to occupational allocation in Africa," .

Fasih, T., G. Kingdon, H. A. Patrinos, C. Sakellariou, and M. Soderbom (2012): "Heterogeneous Returns to Educatino in the Labor Market," *Policy Research Working Paper*, 6170.

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.

Greene, W. (2003): *Econometric Analysis*, Harvard University Press.

Harris, J. and M. Todaro (1970): "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 60, 126142.

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.

HUNEEUS, F., O. LANDERRETCHE, AND E. PUENTES (2012): "Multidimensional Measure of Job Quality Persistence and Heterogeneity in a Developing Country," *Documentos de Trabajo, Universidad de Chile*, 357.

IMBERT, C. (2013): "Decomposing the Labor Market Earnings Inequality: The Public and Private Sectors in Vietnam, 19932006," *World Bank Economic Review*, 27, 55–79.

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*.

NOPO, H., N. DAZA, AND J. RAMOS (2012): "Gender earning gaps around the world: a study of 64 countries," *International Journal of Manpower*, 33, 464–513.

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.

Roodman, D. (2009): "How to do xtabond2: An introduction to difference and system GMM in Stata," *Stata Journal*, 9, 86–136.

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.

Figures

Figure 1: Shares of Wage and Self-employed workers

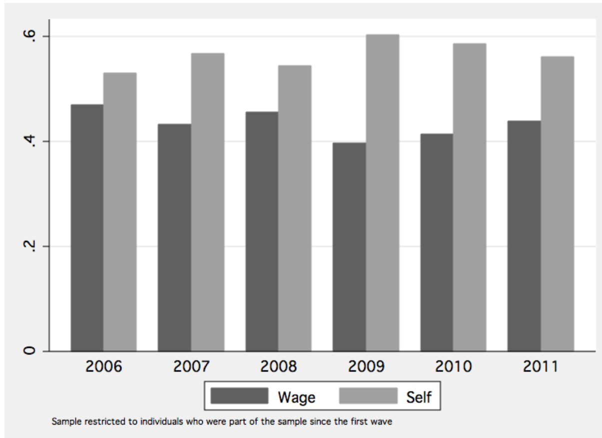


Figure 2: Different types of self-employed workers (pooling all waves)

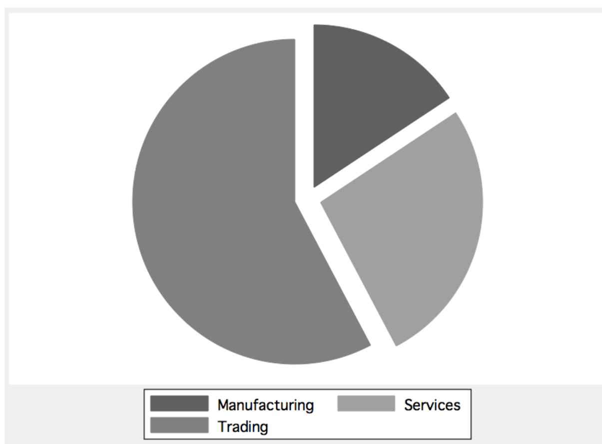


Figure 3: Age of Wage- and Self-Employed Workers

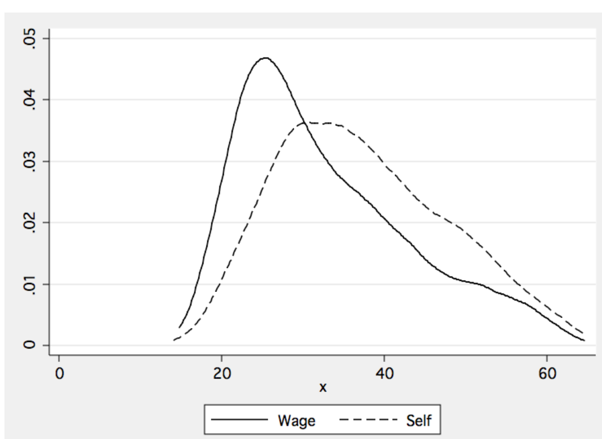


Figure 4: Education Levels of Wage- and Self-employed workers

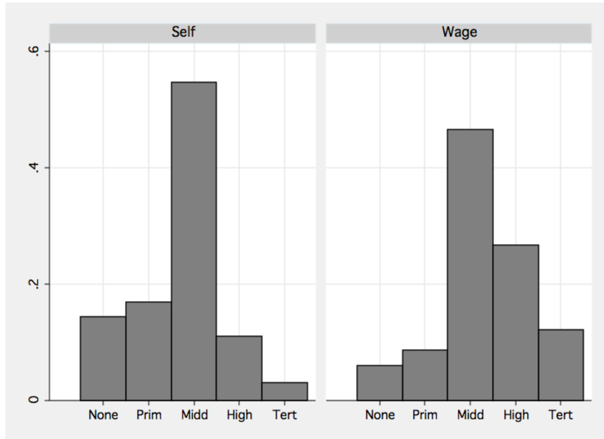


Figure 5: Changes in Gender Composition of Wage- and Self-Employed Workers

Figure 6: Real Monthly Median Earnings by Sector (in US dollars)

Figure 7: Earnings Distributions by Sector

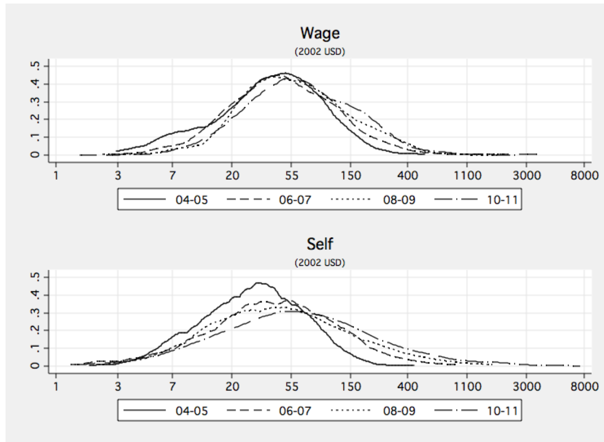


Figure 8: Median Growth Rates of Real Earnings by Sector

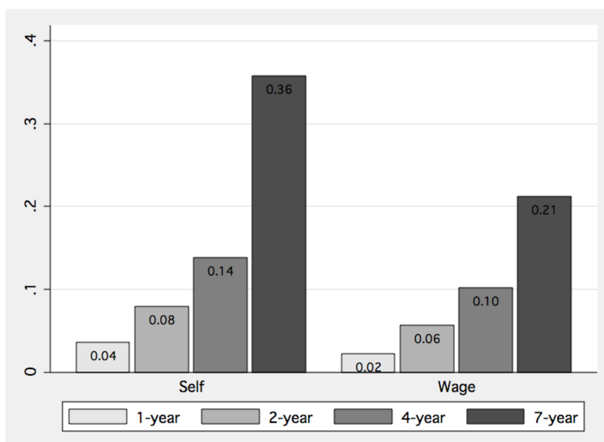


Figure 9: Distribution of Age by Transition Status
(Transition between Wage- and Self-Employment)

Figure 10: Distribution of Real Capital for Wage- and Self-Employed Individuals (log scale)

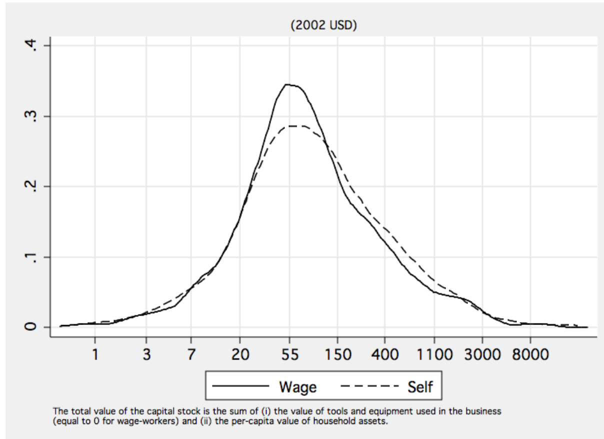


Figure 11: Capital and Earnings

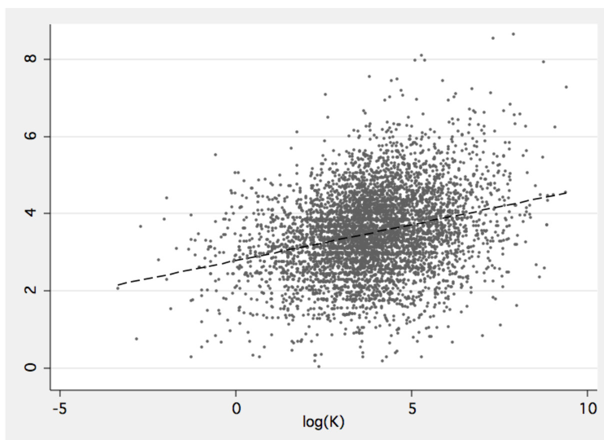


Figure 12: Distribution of (log) Capital by Sector

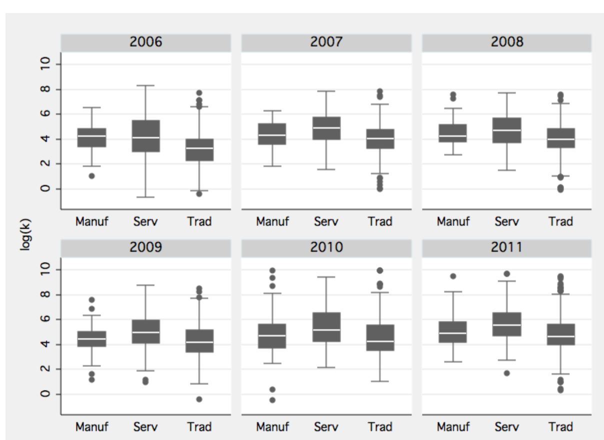


Figure 13: Distribution of Implied Annualised Interest Rates using $\hat{\beta}$

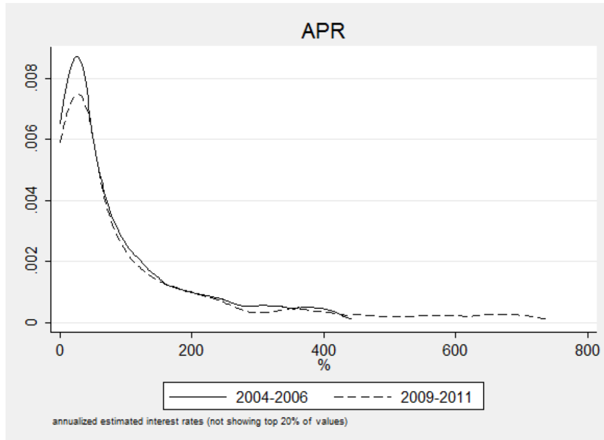


Figure 14: Distribution of $\hat{\theta}$ (Wage)

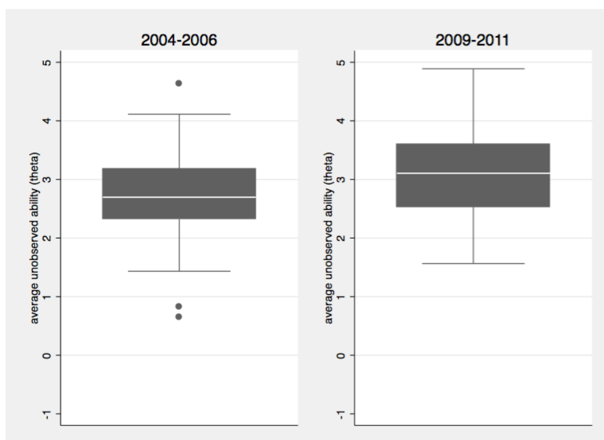
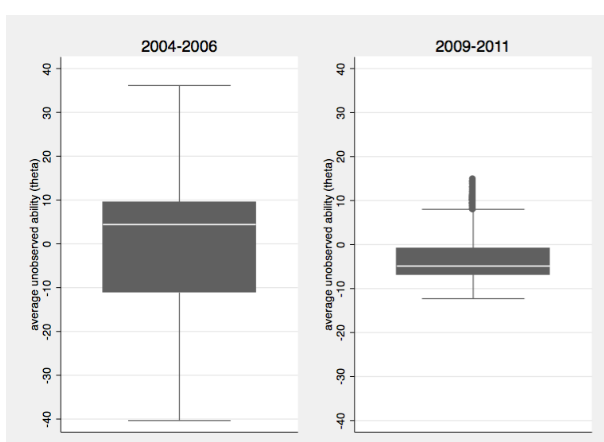


Figure 15: Distribution of $\hat{\theta}$ (Self)



Tables

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)

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

Table 3: Transitions between Wage- and Self-Employment

	$Salaried\ Wage\ Emp_{\ t}$	$Self-Emp_{\ t}$	Total
$Salaried\ Wage\ Emp_{\ t-1}$	1,369 (85.30)	236 (14.70)	1,605 (100)
$Self-Emp_{\ t-1}$	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

Table 4: Panel Attrition - Samples Sizes for Different Panels

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

Table 5: Estimation Results

(3-year sub-panels; bootstrapped 95% confidence interval with 500 resamples)

	IV-GMM 2004-2006	IV-GMM 2009-2011
ψ	0.4785	0.8162
	(0.4423 - 0.5147)	(0.7798 - 0.8522)
β^{SE}	0.2117	0.4134
	(0.1881 - 0.2354)	(0.3892 - 0.4377)
α^{SE}_{Educ}	-1.9472	0.4049
	(-1.7312 - -2.1633)	(0.3315 - 0.4785)
$\delta^{SE} - \delta^{w}$	19.0036	2.9974
	(17.3224 - 20.6855)	(2.4824 - 3.5093)
Number of Observed Individuals (N)	311	354

Table Footnotes: Calibrated Values: $\alpha^{w}_{educ} = 0.08$, $\beta^{w} = 0.02$

Table 6: OLS and Fixed Effects estimators ($\psi = 1$)

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

A Construction of the Instrument Matrix

To illustrate how we constructed our instrument matrix, let us focus on the quasi-differenced error term for the years 2008 to 2011:

$$\mathbf{erm} = \begin{pmatrix} erm_{1,09} & erm_{1,10} & erm_{1,11} \\ \vdots & \vdots & \vdots \\ erm_{N,09} & erm_{N,10} & erm_{N,11} \end{pmatrix}$$

where: $erm_{i,t} \equiv \left(\varepsilon_{i,t} + \tau_i\right) - C_{i,t}\ (\varepsilon_{i,t-1} + \tau_i)$.

In our estimations we use four instruments - past earnings, capital stocks, occupational choice and an interaction between capital levels and occupational choice. For illustration purposes, let us first concentrate on capital (leaving aside the other vectors of the instrument matrix).

$$\mathbf{Z} = \begin{pmatrix} k_{1,08} \\ \vdots \\ k_{N,08} \end{pmatrix}$$

If we assume *pre-determinedness* of capital (i.e. $E(k_{it} erm_{is}) = 0 \quad \forall \quad s \geq t$), an obvious candidate for an instrument would be to use the information on capital levels for 2008, which were determined prior to the realisation of any of the error terms in the sample period. This strategy, however, discards additional valid instruments from other lags of capital. Given pre-determinedness, we could in fact be using $k_{i,09}$ and $k_{i,10}$ as additional valid instruments in years 10 and 11 respectively, when they are uncorrelated with $erm_{i,10}$ and $erm_{i,11}$, and would therefore generate valid moment conditions (i.e. $E[k_{i,09} erm_{i,10}] = 0$ and $E[k_{i,10} erm_{i,11}] = 0$)

Taking this into account, we stack our error terms as follows:

$$\mathbf{erm} = \begin{pmatrix} erm_{1,09} & 0 & 0 \\ \vdots & \vdots & \vdots \\ erm_{N,09} & 0 & 0 \\ 0 & erm_{1,10} & 0 \\ \vdots & \vdots & \vdots \\ 0 & erm_{N,10} & 0 \\ 0 & 0 & erm_{1,11} \\ \vdots & \vdots & \vdots \\ 0 & 0 & erm_{N,11} \end{pmatrix}$$

which can also be collapsed to the following (generally with some minimal loss of efficiency, as explained by Roodman (2009), but in our current setup it should not make a difference):

$$\mathbf{erm} = \begin{pmatrix} erm_{1,09} \\ \vdots \\ erm_{N,09} \\ erm_{1,10} \\ \vdots \\ erm_{N,10} \\ erm_{1,11} \\ \vdots \\ erm_{N,11} \end{pmatrix}$$

such that the instrument matrix is then:

$$\mathbf{Z} = \begin{pmatrix} k_{1,08} & 0 & 0 \\ \vdots & \vdots & \vdots \\ k_{N,08} & 0 & 0 \\ k_{1,09} & k_{1,08} & 0 \\ \vdots & \vdots & \vdots \\ k_{N,09} & k_{N,08} & 0 \\ k_{1,10} & k_{1,09} & k_{1,08} \\ \vdots & \vdots & \vdots \\ k_{N,10} & k_{N,009} & k_{N,08} \end{pmatrix}$$

This structure allows the usual matrix multiplication **Z' erm** to include the larger set of moment conditions outlined above in the calculation of the objective function. In our case, assuming pre-determinedness of k, E[**Z' erm**] $= 0$ gives rise to a set of **6 valid moment conditions**, whose empirical analog can be minimised in usual GMM fashion (i.e. they can be used to form an objective function which we then minimize). Why 6? Intuitively, because we are generating one moment condition from each of the following capital-error combinations: $k_{08} * erm_{09}$, $k_{08} * erm_{10}$, $k_{08} * erm_{11}$, $k_{09} * erm_{10}$, $k_{09} * erm_{11}$, $k_{10} * erm_{11}$.

Let us now extend the analysis to include the additional instruments we can derive from assumptions about occupational status, D. For the case where we assume k and D are pre determined[23], the following matrix contains all the valid instruments we can use.

$$\mathbf{Z} = \begin{pmatrix} k_{1,08} & D_{1,08} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k_{N,08} & D_{N,08} & 0 & 0 & 0 & 0 \\ k_{1,09} & D_{1,09} & k_{1,08} & D_{1,08} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k_{N,09} & D_{N,09} & k_{N,08} & D_{N,08} & 0 & 0 \\ k_{1,10} & D_{1,10} & k_{1,09} & D_{1,09} & k_{1,08} & D_{1,08} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k_{N,10} & D_{N,10} & k_{N,009} & D_{N,009} & k_{N,08} & D_{N,08} \end{pmatrix}$$

By the same reasoning as above, E[**Z' erm**] = 0 gives now rise to **12 valid moment conditions**.

B Weighting Moments - 2-Step Feasible GMM

In our estimations we use four instruments - past earnings, capital stocks, occupational choice and an interaction between capital levels and occupational choice. In this section, we explain how our estimation strategy weights the different moments derived from our overidentified model and how we deal with the autocorrelation and heteroskedasticity inherent in panel data with individual effects. We estimate a 2-Step Feasible GMM estimator, which uses the variance (and covariance) of moment in a first-step estimation, to weigh the moments in the second step and obtain a more efficient estimator.

GMM estimators are derived from the minimization of the following objective function:

$$\beta_{GMM} = argmin_{\beta} \quad N[(\frac{1}{N}\mathbf{Z'erm})'\mathbf{A}(\frac{1}{N}\mathbf{Z'erm})] \tag{25}$$

where **A** is a quadratic weighting matrix. For efficiency, **A** must weigh moments in inverse proportion to their variances and covariances (i.e. de-emphasise the high-variance instruments and those which are highly correlated with others and do not convey much

[23] Assuming predeterminedness of D is a sensible choice if we have already assumed K to be pre-determined, since we want to allow capital to influence occupational choice in the future.

additional information). Hence,

$$A_{EGMM} = Var[\mathbf{z}erm]^{-1} \tag{26}$$

A feasible estimator of A_{EGMM} is:

$$Var[\mathbf{z}erm] = plim_{N \to \inf} \frac{1}{N} \mathbf{Z'}\Omega\mathbf{Z} \tag{27}$$

where Ω is the variance-covariance matrix of the residuals.

Hence,

$$\beta_{GMM} = argmin_{\beta} \quad N[(\frac{1}{N}\mathbf{Z'erm})'(\frac{1}{N}\mathbf{Z'}\Omega\mathbf{Z})^{-1}(\frac{1}{N}\mathbf{Z'erm})] \tag{28}$$

In the simplest case, where the errors are assumed to be *homoskedastic*,

$$\Omega = \sigma^2 \mathbf{I} \tag{29}$$

and

$$A_{EGMM} = \frac{\sigma^2}{N} \mathbf{Z'Z} \tag{30}$$

This gives us the **One-Step Estimator** (estimated as robustness check in table (9) in this paper).

$$\beta_{GMM} = argmin_{\beta} \quad N[(\frac{1}{N}\mathbf{Z'erm})'(\frac{\sigma^2}{N}\mathbf{Z'Z})^{-1}(\frac{1}{N}\mathbf{Z'erm})] \tag{31}$$

Note that since σ^2 is a constant, including it in our objective function does not in practice change the results of the maximisation. The main problem in our situation is that the error terms are unlikely to be *i.i.d.* - for example, several observations of one individual will be correlated as a result of the individual effect. In this case, a more general specification for Ω is required, that may handle arbitrary patterns of correlation across errors. We use the Newey-West estimator (see Greene (2003), p.546):

$$\hat{\Omega}_i = e\hat{r}m_i e\hat{r}m_i' = \begin{pmatrix} e\hat{r}m_{i,09}^2 & e\hat{r}m_{i,09}^2 e\hat{r}m_{i,10}^2 & e\hat{r}m_{i,09}^2 e\hat{r}m_{i,11}^2 \\ e\hat{r}m_{i,10}^2 e\hat{r}m_{i,09}^2 & e\hat{r}m_{i,10}^2 & e\hat{r}m_{i,10}^2 e\hat{r}m_{i,11}^2 \\ e\hat{r}m_{i,11}^2 e\hat{r}m_{i,09}^2 & e\hat{r}m_{i,11}^2 e\hat{r}m_{i,10}^2 & e\hat{r}m_{i,11}^2 \end{pmatrix}$$

where $\mathbf{e\hat{r}m}_i$ is the vector of quasi-differenced residuals for individual i.

In line with the common approach in the literature, we are thus assuming that errors are correlated within individuals across time, but not across individuals. In order to implement this strategy, we need an estimate of $e\hat{r}m$ that can be obtained only with an initial estimate of β. In order to obtain such an initial estimate, we need to make an assumption about the weighting matrix in the first step. Any full-rank choice of $\mathbf{A}$ for the initial GMM estimate will suffice. It is common to choose $A = (\frac{1}{N}Z'HZ)^{-1}$, where $\mathbf{H}$ is an estimate of Ω based on homoskedasticity.

C Robustness Checks

To reduce computational burden and because coefficients of time-invarying characteristics are only identified in quasi-differences we have fixed the values of various parameters in the estimation. In this section we systematically vary these parameters to the upper and lower bound of plausible values. We also consider the results of imposing strict exogeneity (in line with classical panel data analysis) and show results without optimal weighting of the different instruments considered here.

C.1 Returns to Capital in the Wage Sector

We currently assume that $\beta^w = 0.02$. Given that, in contrast to the parameters of the production function, returns to liquid assets are typically considered in terms of interest rates, figure (16) presents the distribution of interest rates implied by the benchmark value of β. Since very high returns to capital in self-employment (as estimated in our analysis) appear to coexist with very low (and sometimes negative) real returns on informal and bank savings in Ghana, there may be disagreement about the appropriate level of this parameter. We thus choose particularly extreme values of β^w, varying it by an order of magnitude from 0.2 to 0.002 in estimations (I) and (II) in tables (8) and (9). The former implies very high interest rates as figure (17) illustrates, whereas the latter essentially implies that savers have no real gains from saving capital (inferior to 3% annually, as figure (18) reveals). The key results appear to be fairly insensitive to these alternative hypotheses.

Figure 16: Distribution of implied interest rates: benchmark β^w

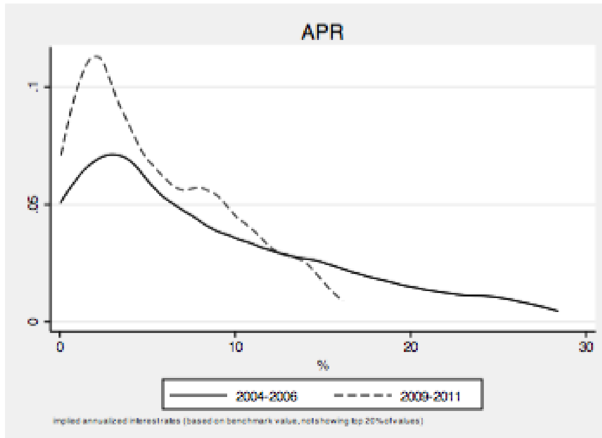


Figure 17: Distribution of implied interest rates: high level of β^w

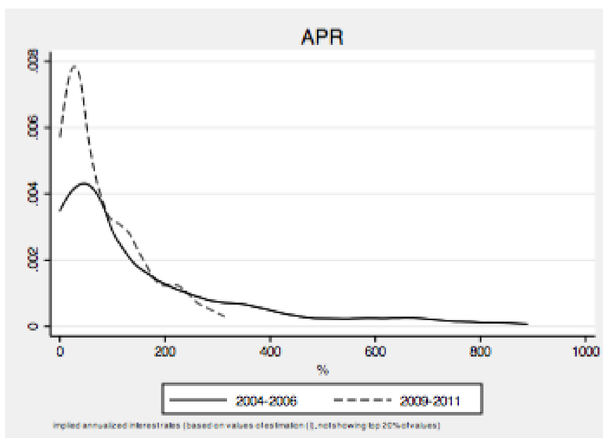
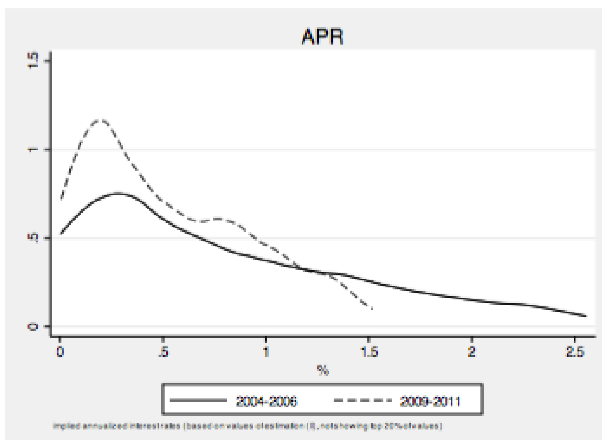


Figure 18: Distribution of implied interest rates: low level of β^w



C.2 Female Wage Discrimination

Differences in earnings across genders are taken into account in all our estimations. Based on previous findings (see in particular Nopo et al. (2012)) we set the benchmark value such that women earn 25% lower wages, ceteris paribus, i.e. $\alpha^w_{male} = 0.25$. However, as the sources upon which we base this assumption use different estimation strategies we increase and decrease the value of α^w_{male} from 15% to 35% in estimations (III) and (IV) reported in tables (9) and (10), a plausible range of wage discrimination. The results show that our main results are little affected by these changes.

C.3 Returns to Education in Wage Sector

There is fortunately no lack of studies on returns to education in Ghana and a hopefully consensual mean rate of returns to one year of formal schooling may be the 8% used as our benchmark. However, this presumably masks heterogeneity in returns across different forms of education, a subtlety that we could not easily incorporate. We thus test the impact of a very low level of under 1% of returns and a very high level of 20% of returns to schooling. Although point estimates are somewhat different as estimations (V) and (VI) in tables (9) and (10) report, all signs as well as the trends in coefficient values between the two periods remain unaffected.

C.4 Results without Optimal Weighting of Moments

If we use the identity matrix to weight our different instruments rather than taking into account the standard errors associated with them, estimation (VII) shows that point estimates are somewhat altered, but that coefficients carry the same sign and that the evolution of coefficients across time remains unchanged. It should be noted that ψ is now statistically insignificant from unity in the more recent period - the trend increase in ψ we pick up in our benchmark estimation is even stronger here and leads to complete equalization of the returns to unobservable individual skills across sectors.

C.5 Results with Reduced Instrument Vector

Our strategy for estimation requires us to find instruments for earnings in period $t-1$ not influenced by contemporaneous earnings shocks in t. This article uses four variables as instruments, as detailed in section (A): preceding earnings up until period $t-2$, past levels of capital, past occupational choices and an interaction between occupational choices and capital levels. In estimation (VIII) in table (11) we consider the impact of dropping the interaction term as an instrument. The results show considerably different point estimates but only in the case of ψ do we find a reversal of the change over time: the values of ψ has not increased over time but remained approximately constant (point estimates have even decreased). All other parameters follow similar patterns.

C.6 Results under the Assumption of 'Exogenous Capital'

In order to allow for the possibility that earning shocks may affect capital accumulation, we have restricted the instrument vector to include only information from periods up to $t-1$ in line with our assumption of predeterminedness - and to use information from past earnings only up until period $t-2$ (since we wish to instrument for earnings in $t-1$). In this section we do not change the use of the instrument "past earnings", however, we assume strict exogeneity for the other instruments and thus use information from all periods to instrument for past earnings. The results are presented in table (10) in estimation (IX).

We confirm the main patterns of increasing returns to all factors of production in self-employment (human capital, observable and unobservable, as well as physical capital), associated with a reduction in the self-employment premium. The point estimates and the degree of change is somewhat different, however our key findings do not appear to be sensitive to this less restrictive assumption on what determines sector mobility.

Table 7: IV GMM Estimation Robustness checks - 95% confidence interval

(normal naïve bootstrap, 500 samples)

	Benchmark 2004-06	Benchmark 2009-11	I(a) 2004-06	I(b) 2009-11
ψ	0.4785	0.8162	0.4585	0.8050
	(0.4423 - 0.5147)	(0.7798 - 0.8522)		
β^{SE}	0.2117	0.4134	0.1883	0.3793
	(0.1881 - 0.2354)	(0.3892 - 0.4377)		
β^{w}	0.02	0.02	**0.20**	**0.20**
	(set)	(set)	**(set)**	**(set)**
α^{SE}_{Educ}	-1.9472	0.4049	-1.6324	0.5269
	(-1.7312- -2.1633)	(0.3315 - 0.4785)		
α^{w}_{Educ}	0.08	0.08	0.08	0.08
	(set)	(set)	(set)	(set)
α^{SE}_{Male}	-3.8964	-12.8949	-2.6752	-10.0685
	(-3.740 - -4.4190)	(-12.3057 - -13.4777)	(0.1234 - 0.1234)	(0.1234 - 0.1234)
α^{w}_{Male}	0.25	0.25		
	(set)	(set)		
$\delta^{SE} - \delta^{W}$	19.004	2.9974	16.1080	1.1096
	(17.3224 - 20.6855)	(2.4824 - 3.5093)		
Optimal Weighting Ω	yes	yes	yes	
Instruments	full	full	full	
Sample N (movers)	311 (41)	354 (73)	311 (41)	354 (73)

(**Bold** figures indicate deviations from benchmark values)

Table 8: IV GMM Estimation Robustness checks - 95% confidence interval

(normal naïve bootstrap, 500 resamples)

	II(a) 2004-06	II(b) 2009-11	III(a) 2004-06	III(b) 2009-11
ψ	0.4796	0.8135	0.4749	0.8206
β^{SE}	0.2151	0.4169	0.2173	0.41622
β^{w}	**0.002** **(set)**	**0.002** **(set)**	0.02 (set)	0.02 (set)
α^{SE}_{Educ}	-1.9790	0.3928	-1.9625	0.3991
α^{w}_{Educ}	0.08 (set)	0.08 (set)	0.08 (set)	0.08 (set)
α^{SE}_{Male}	-4.0152	-13.1255	-3.9703	-13.1528
α^{w}_{Male}	0.25 (set)	0.25 (set)	**0.15** **(set)**	**0.15** **(set)**
$\delta^{SE} - \delta^{W}$	19.2901	3.16659	19.1357	3.1297
Optimal weighting Ω	yes	yes	yes	yes
Instruments	full	full	full	full
Sample N (movers)	311 (41)	354 (73)	311 (41)	354 (73)

(**Bold** figures indicate deviations from benchmark values)

Table 9: IV GMM Estimation Robustness checks - 95% confidence interval

(normal naïve bootstrap, 500 resamples)

	IV(a) 2004-06	IV(b) 2009-11	V(a) 2004-06	V(b) 2009-11
ψ	0.4820	0.8124	0.4330	0.8720
β^{SE}	0.2061	0.4107	0.2878	0.4573
β^{w}	0.02	0.02	0.02	0.02
	(set)	(set)	(set)	(set)
α^{SE}_{Educ}	-1.9317	0.4106	-2.1454	0.2387
α^{w}_{Educ}	0.08	0.08	**0.008**	**0.008**
	(set)	(set)	**(set)**	**(set)**
α^{SE}_{Male}	-3.8214	-12.6472	-4.1572	-15.6198
α^{w}_{Male}	**0.35**	**0.35**	0.25	0.25
	(set)	**(set)**	(set)	(set)
$\delta^{SE} - \delta^{W}$	18.8697	2.8693	20.3861	5.2154
Optimal weighting Ω	yes	yes	yes	yes
Instruments	full	full	full	full
Sample N (movers)	311 (41)	354 (73)	311 (41)	354 (73)

(**Bold** figures indicate deviations from benchmark values)

Table 10: IV GMM Estimation Robustness checks

(changes vis-à-vis benchmark in bold)- 95% confidence interval (normal naïve bootstrap, 500 resamples)

	VI(a) 2004-06	VI(b) 2009-11	VII(a) 2004-06	VII(b) 2009-11
ψ	0.5510	0.6751	0.3262	0.9974
β^{SE}	0.1089	0.3375	0.2125	0.3153
β^{w}	0.02 (set)	0.02 (set)	0.02 (set)	0.02 (set)
α^{SE}_{Educ}	-1.4754	0.5929	-0.4481	0.5552
α^{w}_{Educ}	**0.20** **(set)**	**0.20** **(set)**	0.08 (set)	0.08 (set)
α^{SE}_{Male}	-3.1353	-8.3363	-3.9199	-15.3240
α^{w}_{Male}	0.25 (set)	0.25 (set)	0.25 (set)	0.25 (set)
$\delta^{SE} - \delta^{W}$	15.3445	0.0695	7.4962	3.1701
Optimal weighting Ω	yes	yes	**no**	**no**
Instruments	full	full	full	full
Sample N (movers)	311 (41)	354 (73)	311 (41)	354 (73)

(**Bold** figures indicate deviations from benchmark values)

Table 11: IV GMM Estimation Robustness checks - 95% confidence interval

(normal naïve bootstrap, 500 resamples)

	VIII(a) 2004-06	VIII(b) 2009-11	IX(a) 2004-06	IX(b) 2009-11
ψ	0.8270	0.7946	0.3716	1.2221
β^{SE}	0.0000	0.3623	0.0000	0.2401
β^{w}	0.02	0.02	0.02	0.02
	(set)	(set)	(set)	(set)
α^{SE}_{Educ}	-2.3502	0.3974	-0.3603	0.7807
α^{w}_{Educ}	0.08	0.08	0.08	0.08
	(set)	(set)	(set)	(set)
α^{SE}_{Male}	-7.2135	-12.8496	-5.4953	-21.5158
α^{w}_{Male}	0.25	0.25	0.25	0.25
	(set)	(set)	(set)	(set)
$\delta^{SE} - \delta^{W}$	24.4853	3.2628	8.2803	3.6581
Optimal weighting Ω	yes	yes	yes	yes
Instruments	**partial**	**partial**	**full & strict exog.**	**full & strict exog.**
Sample N (movers)	311 (41)	354 (73)	311 (41)	354 (73)

(**Bold** figures indicate deviations from benchmark values)

D Earning Gap Decomposition

The earning gap decomposition in table (12) reveals that despite the aforementioned increase in returns to observable and unobservable factors of production in self-employment, and despite the higher prevalence of those factors in the sector, a positive unadjusted earning gap in favour of wage-employment remains (self-employed earnings are 28.28% lower in the earlier and 15.51% lower in the later period). After controlling for worker characteristics, we document a positive self-employment premium, which tends to shrink over the period. Given that our analysis conditions on both observables and unobservables, this residual premium can be interpreted as an indication of imperfect mobility across sectors (potentially driven by significant, albeit weakening, credit constraints).[24]

Overall, our analysis shows that differential returns to the same factors of production, both observable and unobservable, increasingly favour self-employment (i.e. the "price effect" is positive and increasing). However, the differential deployment of factor inputs (the "selection/composition effect") still favours wage-jobs, with a dominant contribution from unobservable individual characteristics.

[24]The estimated downward trend in the self-employment premium can equivalently be interpreted as an upward trend in the wage-premium. This development may be partly 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 (e.g. 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 policies (see Imbert (2013) for an analysis of the effects on the wage structure of public sector pay reform in Vietnam), of which we are not aware, 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.

Table 12: Earnings gap decomposition

Effect of Differential Returns to Self-employment	**2004-2006**	**2009-2011**
Negative Wage premium $\delta^{SE} - \delta^{w}$	19.0039	2.9973
Effect of Observable Skills	-15.7649	1.2221
Effect of Capital $ln(\overline{K_{SE}})\ (\beta^{SE} - \beta^{w})$	0.7416	1.9279
Effect of Gender $\overline{h_{SE}}\ (\alpha^{SE} - \alpha^{w})$	-1.3181	-3.3071
Effect of Education $\overline{h_{SE}}\ (\alpha^{SE} - \alpha^{w})$	-15.1884	2.6013
Effect of Unobservable Skills $(\psi - 1)\ \overline{\theta_{SE}}$	0.7634	0.2748
Total Price Effect	**4.0025**	**4.4941**
Effect of Differential Input Factors		
Difference in Observed Factors	-0.2949	-0.224
Difference in Capital $\left(ln(\overline{K_{SE}}) - ln(\overline{K_{w}})\right)\ \beta^{w}$	0.0014	0.0083
Difference in Gender $\left(\overline{h_{SE}} - \overline{h^{w}}\right)\ \alpha^{w}$	-0.0969	-0.0921
Difference in Education $\left(\overline{h_{SE}} - \overline{h^{w}}\right)\ \alpha^{w}$	-0.1993	-0.1402
Difference in Unobserved Factors $\overline{\theta_{SE}} - \overline{\theta_{w}}$	-3.9904	-4.4252
Total Selection Effect	**-4.2852**	**-4.6492**