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

Luke Haywood† and Paolo Falco‡

1st March 2012

Work in Progress

Abstract

The self-employed constitute a large proportion of the workforce in developing countries and the sector has been found to be growing further. This article assess potential causes for the increase in self employment observed in Ghana by considering individuals moving between sectors taking into account the fact that selection into sectors may not be voluntary. Models of segmented labour markets typically consider sorting on unobservables to be important, typically by positing a sector choice model. If there are barriers to entry in one of the sectors, however, selection on unobservables may not go in the direction of self-selection. Assumptions about the direction of sorting on unobservables rest on shaky foundations.

We present a simple model of a two-sector labour market and estimate earnings using a correlated random coefficients model that allows for multiple patterns of sorting ansd selection on unobservables using instrumental variables GMM.

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

†Deutsches Institut für Wirtschaftsforschung & Paris School of Economics, *E-mail :* haywood@pse.ens.fr

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

1 Introduction

The self-employed constitute a large proportion of the workforce in developing countries and recent large-scale household data have confirmed an increase in self-employment both in rural but especially urban contexts (see Kingdon et al. (2006)). In econonmics, self-employment is interpreted as entrepreneurship, i.e. the establishment of a business transforming capital and labour into output. However, in developing countries many self-employed operate with little to no capital, whilst wage jobs, often formal and in the public sector, are viewed as hard to access. Self-employment may not be chosen but rather result from barriers to entry for wage jobs.
An optimistic interpretation of this development notes that returns to capital in countries with many low-paid workers are typically high. More individuals becoming self-employed may then be a result of increased wealth and opportunities since it has been argued that capital constraints on the establishment of self-employed businesses are often binding (Evans and Jovanovic (1989), Magnac and Robin (1996) and Blanchflower and Oswald (1998)). Self-employment may be a road to higher employment, reduced inequality in earnings and better working conditions.

A more pessimistic perspective is that for many individuals, wage work is not available and in absence of social protection or family transfers, self-employment may be the only means of survival. In this sense, part of the self-employment can be interpreted as the equivalent of unemployment in countries without social welfare systems. In fact, some of the urban self-employed may be viewed as forming a queue for salaried employment in Harris-Todaro-type models. This would imply that there exists a wage premium even controlling for selection on unobservables. We will test this hypothesis. We will also investigate additional factors influencing earnings in the two sectors.

Recently, Poschke (2010) uses Global Entrepreneurship Monitor data about declared reasons for self-employment. In particular, individuals state whether the reason for their self-employment is voluntary or not. Relevant findings are:

- there is more necessity self-employment in non-OECD countries
- necessity self-employment typically concerns small businesses
- the necessity self-employed are typically less educated and more often female

- in some non-OECD countries the proportion of one-person self-employed firms reporting necessity self-employment reaches 50%

The strategy proposed in this article takes unobservables seriously, and in particular, sorting or selection on unobservables. Panel fixed effects are insufficient to control for individual unobservables, if their effect varies across sectors. Furthermore, we do not wish to make the assumption - standard in selection models - that individuals self-select into the sector which is for them more advantageous. This feature of selection models (the so-called single-crossing property) implies labour markets that efficiently match individuals to sectors based on unobservable characteristics. However, search models imply that sorting into the correct sector may be limited due to information problems. More importantly in our context, in developing countries many wage jobs are in the public sector where competitive market forces may compete with other selection mechanisms (e.g. social capital). Similarly, entry costs in the self-employed sector may vary in ways related to unobservables, for example by way of differences in access to credit.

The paper proceeds as follows. In section (2) we present the basic model with a special focus on selection issues. Section (3) considers identification and presents an estimation strategy using a correlated random coefficients model. Section (4) presents the data. We then discuss results for sector earnings in section (5) and show how these can be used to assess the extent of choice and necessity self-employment. This section is currently work in progress.

2 The Model

In this section we present a simple occupational choice framework. 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 occurr, but the single-crossing property may not hold.

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

Let worker i be endowed with time-varying characteristics (e.g. physical capital K), time-invarying characteristics (e.g. human capital h) and unobservable sector-specific productivity θ^w, θ^{SE}.

2.1 Sector Earnings

Our definition of physical capital encompasses both liquid savings and assets (under the assumption that the latter can easily be sold and re-invested). We take physical and human capital as given and do not make explicit the impact that sector choice may have on capital accumulation[1].

Given these endowments, worker i optimally chooses between two alternative *employment strategies*: working as a self-employed or searching for a wage-job.

Using a log-linearised wage equation for self-employment, the worker will earn:

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

where $K_{i,t}, h_i$ indicate physical and human capital; δ_t^{SE} designates macroeconomic effects on the self employed sector and $u_{i,t}$ subsumes individual idiosyncratic factrs common across across sectors.
In wage-employment, workers earn:

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

We can think of β^w as the market rate of return on savings (e.g. interest rate on bank deposits). Empirical studies consistently find that access to capital is an important determinant of self-employment, indicating that many "latent entrepreneurs" (REF) are credit constrained. We have measures of capital and estimate returns to both sectors[2]

Our model depicts a dual labour market, with employers of skilled labour on the one side (large private firms and the public secor, i.e. 'formal employers') and the self-employed on the other side (micro/small private enterprises). 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. Such rationing could be the result of efficiency wage setting, institutional constraints or result from informational frictions[3].

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

[2] The issue of endogeneity of capital with respect to past earnings maybe particularly problematic in the wage sector, an issue which we do not currently address.

[3] 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.

We now provide a motivation for considering non-standard sorting by unobservables.

2.2 Selection and Barriers to Entry

The classic Roy model is based on specialisation by comparative advantage. However, we would like to avoid the assumption of free sector choice according to earnings underlying most of the literature. Indeed, if there are job queues for entry into wage employment, we need to take into account not only self-selection by workers, but selection of workers by firms. It is however unclear how firms choose individuals - for example, to what extent firms can observe individual sector-specific performance ($\theta_i^w, \theta_i^{SE}$). Non-productive factors may also play a role in firms' selection choices, e.g. in the public sector (the dominant wage job employer in many developing countries).

A convenient way of specifying sector assignment is as a cost of entry: individuals self-select subject to entry costs (implicitly set by firms) into sectors which may depend on individuals' characteristics. One such exposition is given by Magnac (1991) moels entry costs as the result of job-queues whereby individuals in every period have a probability τ of not finding a job. The waiting time can be formulated as a cost $c(.)$. We can expect $c(.)$ to be a function of the determinants of wages:

$$\begin{aligned}
\ln \pi^w(x_i, \theta_i)^* =& (1-\tau)\ \ln \pi^w(x_i, \theta_i) + \tau\ 0 \\
\ln \pi^w(x_i, \theta_i)^* =& \ln \pi^w(x_i, \theta_i) - \tau\ \pi^w(x_i, \theta_i) \\
\ln \pi^w(x_i, \theta_i)^* =& \ln \pi^w(x_i, \theta_i) - c(x_i, \theta_i) \qquad (3)
\end{aligned}$$

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 not only with earnings differentials but (potentially individual-specific) entry costs[4]. We can then think about the constrained choice

[4] 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.

of workers as follows:

$$\begin{aligned} Pr(d_i^{SE}) &= Pr\left(E\left(R^{SE}(x_{i,t},\theta_i^{SE})\right) - c^{SE}(x_{i,t},\theta_i^{SE}) + v_{i,t}^{SE} > E\left(R^{w}(x_{i,t},\theta_i^{w})\right) - c^{w}(x_{i,t},\theta_i^{w}) + v_{i,t}^{w}\right) \\ &= Pr\left(v_{i,t} < \Delta\hat{R}(x_{i,t},\theta_i) - c(x_{i,t},\theta_i)\right) \\ &= \Phi_v\left(\Delta\hat{R}(x_{i,t},\theta_i) - c(x_{i,t},\theta_i)\right) \end{aligned} \tag{4}$$

where $v_{i,t} \equiv v_{i,t}^{w} - v_{i,t}^{SE}$, $c(x_{i,t},\theta_i) = c^{SE}(x_{i,t},\theta_i^{SE}) - c^{w}(x_{i,t},\theta_i^{w})$ and $\Delta R(x_{i,t},\theta_i) = R^{SE}(x_{i,t},\theta_i^{SE}) - R^{SE}(x_{i,t},\theta_{i,t}^{w})$ and for some cumulative density function Φ_v of the differenced idiosyncratic error, $v_{i,t}$.

In a standard wage wage regression, coefficients on x would give net returns taking into account the effect of x on earnings via $\Delta\ R$ and c. In a two-stage model using instruments for sector choice or relying on parametric assumptions one could then take into account sector choice (Heckman (1976)). However, these strategies require the unappealing assumption that workers who are more productive in the wage employed sector find a job there and vice-versa.

Our plan is to derive consistent estimate of $\Delta R(x_{i,t},\theta_i^{SE},\theta_i^{w})$ allowing for sorting according to $c(.)$. Given a consistent estimate of expected sector earnings differentials we can then use information about sector choice according to expected earnings. We are particularly interested in the evolution over time of the sector wage premia δ and of potential changes in the returns to other observables (β, α).

2.3 Inactive and Unemployed Workers

Our current model rules out the possibility of unemployment/inactivity. It should not be forgotten, however, that results presented here are conditional on participation in the labour market, and evolutions in the labour market over time may also affect the composition of the labour force.

3 Identification and Estimation in the Correlated Random Coefficients Model

In this section we present a panel-based procedure that allows us to estimate the counterfactual earnings implied by our model allow for unspecified patterns of selection on unobservable characteristics. Thus we can estimate fractions of choice and necessity self-employed under the assumption that job queues may exist and selection is not necessarily based on a single threshold (as in self-selection models). In fact, we can estimate the importance and direction of selection effects. The

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

Following (1) and (2)), let earnings in wage and self employment be given by

$$R_{i,t}^{w} = \ln \pi^{w}(h_{i,t}, K_{i,t}) + \delta_{t}^{w} + \theta_{i}^{w} + u_{i,t}^{SE} \tag{5}$$
$$R_{i,t}^{SE} = \ln \pi^{SE}(h_{i,t}, K_{i,t}) + \delta_{t}^{SE} + \theta_{i}^{SE} + u_{i,t}^{w} \tag{6}$$

We can then state earnings as function of the sector dummy d_{i}^{SE} indicating whether an individual is self-employed:

$$\begin{aligned} R_{i,t} =& \delta_{t}^{w} + d_{i,t}^{SE}\left(\delta_{t}^{SE} - \delta_{t}^{w}\right) \\ &+ \alpha^{w}\ \ln \underline{\mathrm{h}}_{i,t} + d_{i,t}^{SE}\ \ln \underline{\mathrm{h}}_{i,t}\left(\alpha^{SE} - \alpha^{w}\right) \\ &+ \beta^{w}\ \ln K_{i,t} + d_{i,t}^{SE}\ \ln K_{i,t}\left(\beta^{SE} - \beta^{w}\right) \\ &+ \theta_{i}^{w} + d_{i,t}^{SE}\ \left(\theta_{i}^{SE} - \theta_{i}^{w}\right) + \varepsilon_{i,t} \end{aligned} \tag{7}$$

where $\varepsilon \equiv\ d_{i,t}^{SE}\left(u_{i,t}^{SE} - u_{i,t}^{w}\right)$. Using a simple projection, we can separate an absolute advantage component, τ_{i}, from a comparative advantage component of individual unobserved heterogeneity, θ_{i}[5].

$$\theta_{i}^{SE} = b_{SE}\ \left(\theta_{i}^{SE} - \theta_{i}^{w}\right) + \tau_{i} \tag{8}$$
$$\theta_{i}^{w} = b_{w}\ \left(\theta_{i}^{SE} - \theta_{i}^{w}\right) + \tau_{i} \tag{9}$$
$$\theta_{i} \equiv b_{w}\ \left(\theta_{i}^{SE} - \theta_{i}^{w}\right) \tag{10}$$

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 (11) and (12) 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.

$$\theta_{i}^{SE} = \psi\ \theta_{i} + \tau_{i} \tag{11}$$
$$\theta_{i}^{w} = \theta_{i} + \tau_{i} \tag{12}$$

[5] 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.

Earnings can then be written as follows:

$$\begin{aligned} R_{i,t} =& \delta_t^w + d_{i,t}^{SE}\left(\delta_t^{SE} - \delta_t^w\right) \\ &+ \alpha^w \ \ln \underline{\mathrm{h}}_{i,t} + d_{i,t}^{SE} \ \ln \underline{\mathrm{h}}_{i,t}\left(\alpha^{SE} - \alpha^w\right) \\ &+ \beta^w \ \ln K_{i,t} + d_{i,t}^{SE} \ \ln K_{i,t}\left(\beta^{SE} - \beta^w\right) \\ &+ \theta_i + d_{i,t}^{SE} \ (\psi - 1) \ \theta_i + \tau_i + \varepsilon_{i,t} \end{aligned} \tag{13}$$

The newly introduced parameter ψ provides information about sorting: if $\psi = 1$, unobservable individual characteristics are not remunerated differently across sectors, i.e. comparative sector advantages, where they exist, are not remunerated accordingly. 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 whose comparative advantage is in self-employment if they work in this sector, whereas if $\psi < 1$ there is a premium for these workers in the wage sector. How might a situation arise in which $\psi < 1$? We sketch one situation below, and stress that if we wish to make no assumptions about sector allocation it is important not to restrict the potential direction of selection bias.

3.1 Identification

Model identification relies on the classic panel data restriction that the idiosyncratic error terms $u_{i,t}$ are uncorrelated with the covariates in all time periods. This allows for rich patterns of selection on the unobservables θ_i but does not allow for sector choice to be a function of idiosyncratic errors $u_{i,t}$. Large income shocks may be thought to influence on sector choice. We can note that if such shocks operate by depleting individuals' level of capital this is an observable we can condition on. We can thus allow for selection due to this kind of shock.

Coefficients for the time-varying covariants β are identified by variation in returns across different levels of the covariates. The intuition for identification of θ_i and ψ can best be seen when we consider the error structure for individuals with different histories of sector transition. For illustration, let us focus on the case where $T = 2$. We observe four transition types: self-employed stayers (S, S), wage-employed stayers (w, w), wage-to-self-employment (w, S) and self-employment-to-wage-employment (S, w).

$$E\left[R_{t=1|S,S}\right] = \pi^{SE} + \delta_{t=1}^{SE} + \psi\ \theta_{S,S} \qquad E\left[R_{t=2|S,S}\right] = \pi^{SE} + \delta_{t=2}^{SE} + \psi\ \theta_{S,S} \tag{14}$$

$$E\left[R_{t=1|w,w}\right] = \pi^{w} + \delta_{t=1}^{w} + \theta_{w,w} \qquad E\left[R_{t=2|w,w}\right] = \pi^{w} + \delta_{t=2}^{w} + \theta_{w,w} \tag{15}$$

$$E\left[R_{t=1|w,S}\right] = \pi^{SE} + \delta_{t=1}^{SE} + \psi\ \theta_{w,S} \qquad E\left[R_{t=2|w,S}\right] = \pi^{SE} + \delta_{t=2}^{SE} + \psi\ \theta_{w,S} \tag{16}$$

$$E\left[R_{t=1|S,w}\right] = \pi^{w} + \delta_{t=1}^{w} + \theta_{S,w} \qquad E\left[R_{t=2|S,w}\right] = \pi^{w} + \delta_{t=2}^{w} + \theta_{S,w} \tag{17}$$

After a normalization of theta using the restriction (18), we see that with eight observations on the mean earnings in the eight groups in (14)-(17) we can identify the five structural parameters: $\delta_{t=1}^{SE}, \delta_{t=2}^{SE}, \delta_{t=1}^{w}, \delta_{t=2}^{SE}, \psi$ as well as three points of the distribution of theta $(\theta_{w,w}, \theta_{w,S}, \theta_{S,w})$[6].

$$\frac{1}{T\,N}\sum_{t=1}^{T}\sum_{i=1}^{N}\theta_{i(t)} = 0 \tag{18}$$

We can note that a fixed effects framework would assume that θ does not depend on sector mobility, i.e. would impose that $E(\theta_{j,k}) = 0 \;\forall\; j,k \in [S,w]$. Selection-corrected models in the spirit of Heckman (1976) impose sign restrictions on the expected mean values of θ depending on the types of transition, positing for example that $E\left[\theta_{S,S}\right] \geq E\left[\theta_{w,w}\right]$.

Flexibility with respect to the distribution of θ is important in our context: If there are job queues for wage sector employment it is unclear how the sorting process will work. Sorting on unobservables may conceivably go either way:

- Assume first that firms do not observe workers' comparative advantage θ_i and choose randomly from the queue of workers applying for a job. Assume furthermore that waiting costs are easier to shoulder for individuals with more friends and family find it easier to support them. These same characteristics may make individuals (conditional on observables) relatively more succesful in self-employment (acquaintances being potential customers): then we might expect a positive association between wage employment and comparative advantage for self-employment.
- Second, assume workers and firms observe workers' comparative advantage (i.e. their match quality) and waiting costs are randomly distributed. Other things being equal this will lead to an assignment of individuals with relatively low values of θ to wage employment.

Since we are agnostic about the actual selection mechanism, we want to allow for the fact that either of the two scenarios may be possible.

Note thatm 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

[6]From this we can easily see that our method of estimating θ is consistent only for large T, as in standard panel data models.

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[7].

We can note that the sign of ψ also has direct welfare consequences: 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[8].

3.2 Estimation

Due to the assumption of sector-specific rewards to individual effects, standard panel data methods using differencing or fixed effects analyses will not remove the unobservable individual effect θ_i. However, following Chamberlain (1982) and Lemieux (1998) we can solve explicitly for θ_i in (13) and replace this expression in the general earnings equation.

$$\theta_i = R_{i,t-1} - \left[\frac{\delta^w_{t-1} + d^{SE}_{i,t-1}\, \delta_{t-1} + \gamma^w\ \ln x_{i,t-1} + d^{SE}_{i,t-1}\ \ln x_{i,t-1}\, \gamma + \tau_i + \varepsilon_{i,t-1}}{1 + d^{SE}_{i,t-1}\ (\psi - 1)} \right] \tag{19}$$

where $x = (h, K)$, $\gamma = (\alpha, \beta)$ and $\gamma \equiv \gamma^{SE} - \gamma^w$.

$$R_{i,t} = G_t\left(x_{i,t}, d^{SE}_{i,t}\right) + \frac{1 + d_{i,t}\,\psi}{1 + d_{i,t-1}\,\psi}\ \left[R_{i,t-1} - G_{t-1}\left(x_{i,t-1}, d^{SE}_{i,t-1}\right)\right] + \varepsilon_{i,t} \tag{20}$$

where $G_t\left(x_{i,t}, d^{SE}_{i,t}\right) \equiv \delta^w_t + d^{SE}_{i,t}\ \delta_t + \gamma^w\ \ln x_{i,t} + d^{SE}_{i,t}\ \ln x_{i,t}\ \gamma$ and $\varepsilon_{i,t} \equiv u_{i,t} + \tau_i - \frac{1+d_{i,t}\ \psi}{1+d_{i,t-1}\ \psi}\ (u_{i,t} + \tau_i)$.

4 Data

We apply our model using data from the Ghana Household Urban Panel Survey ('GHUPS'), conducted by the Centre for the Study of African Economies (CSAE) at the University of Oxford. The survey was first conducted in 2004 and it now spans 7 years, an unusual length for panel data-sets in developing countries. The

[7] 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 $\beta^{SE} - \frac{1}{\psi}\,\beta^w$

[8] For a nice parametric example in the spirit of Roy (1951) where specialisation reduces inequality overall, see Heckman and Honore (1990).

GHUPS covers four cities: Accra, Kumasi, Takoradi and Cape Coast. Respondents were drawn by stratified random sampling of urban households from the Population and Housing Census of 2000. The survey was designed to cover all household members of working age at the time of the interview. After the first wave, the sample expanded by incorporating new members of the original households, as well as new households formed by individuals who had left their original household and were tracked to their new locations. Table (1) and (2) contain some key summary statistics for the sample of GHUPS respondents who are currently employed, split by sector and pooling all waves of data. Figure 1 shows the distribution of (log) real monthly earnings among self-employed and wage-employed workers. Most importantly for our analysis, the dataset includes information on assets and business capital for self-employed individuals, as well as information on transitions between labour market states over the years. Table (3) summarises transitions between sectors, pooled across all panel waves. It is confined to workers who are employed both at t and $t-1$ and shows that 14 % of all workers who are in wage-emlpoyment in any given period move to self-employment in the next period, while 9.5% of self-employed workers move to wage-employment. Though sizable in percentage terms, the number of observed transitions will pose a challenge to the precision of our more complex estimators.

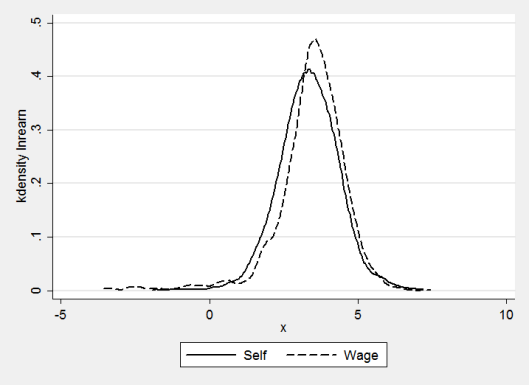
Table 1: Summary Statistics - Self

Variable	Mean	Std. Dev.	Min.	Max.	N
Male	0.273	0.445	0	1	1141
Age	37.493	10.742	15.208	69.659	1141
Educ	7.461	3.98	0	17	1141
Real Val Tools/Equip	76.658	825.081	0	25889.287	1141
Real Asset Val	224.828	737.194	0	11610.914	613

Table 2: Summary Statistics - Wage

Variable	Mean	Std. Dev.	Min.	Max.	N
Male	0.617	0.486	0	1	1304
Age	33.517	10.882	16.042	68.207	1428
Educ	9.231	3.663	0	18	1344
Real Val Tools/Equip					0
Real Asset Val	206.321	475.748	0	6076.791	446

Figure 1: Earnings by Sector



4.1 Measures of capital

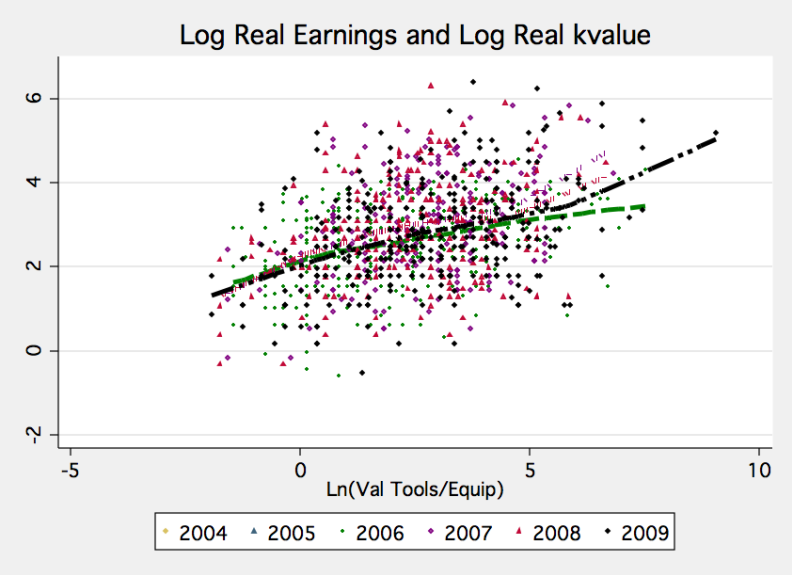
Given its traditionally important role in analysing self-employment decisions, our measure of capital deserves some discussion. The descriptives will serve the double role of helping us check data consistency and providing useful insights.

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 problem of 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. As figure (2) shows, there appears to be a clear, stable and strong relationship between the (real) value of tools/equipment and earnings (once we discard some very small capital values from 2006).

Studying the distribution of capital by sector, figure (3) shows that service providers are the ones who need the most capital, followed by manufacturers and traders. To put this into context, we provide a summary of the relative size of the different sectors in which the self-employed work: Figure (4) shows that the largest category of self-employed are traders, followed by service providers and manufacturers.

In order to test our model, we need data on the income stream or the usage value generated by the assets respodents own, in any given period. 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

Figure 2: Capital and earnings



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 reposndents (clearly an over-simplification, since heterogeneous returns might play an important role).[9]

Capital holdings may be derived from past labour market outcomes - thus related to not only schooling but earnings. This would mean that earnings shocks in the past may be correlated with capital holdings. This type of reverse causality has been the motivation for models using functions of predetermined (as opposed to current) levels of capital as instruments for capital holdings. We leave integrating this type of (Arellano and Bond (1991), Blundell and Bond (2000)) strategy into the current framework to future work.

[9]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. This is likely an overestimate for the past years covered by our survey, where inflation has been on average higher. We will attempt a more detailed construction of the usage value interest rate in future drafts of the paper.

Figure 3: Distribution of lnrkvalue by sector

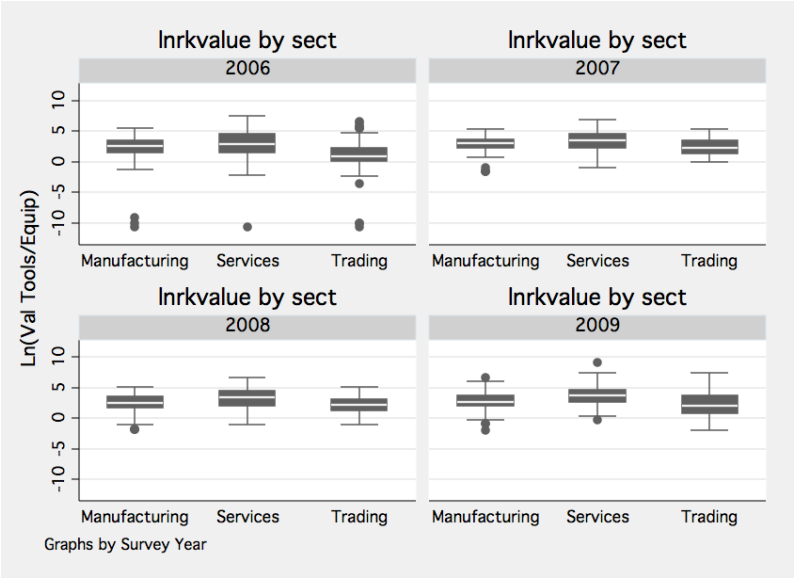


Figure 4: Sector shares

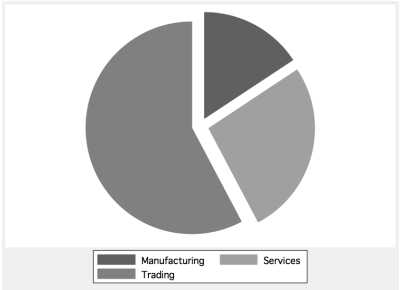


Table 3: Sector Movement

	$Salaried\,Wage\,Emp_{\,t-1}$	$Self-Emp_{\,t}$	Tot
$Salaried\,Wage\,Emp_{\,t-1}$	1,008 (85.86)	166 (14.16)	1,174 (100)
$Self-Emp_{\,t-1}$	171 (9.45)	1,638 (90.55)	1,809 (100)
Tot	1,179 (39.52)	1,804 (60.48)	2,983 (100)

One-period transitions pooled across wavews; Percentages reported in parentheses

5 Results - *Work in progress*

In this section we present a set of *preliminary* results from the estimation of our quasi-differenced earnings model using the Generalized Method of Moments (GMM). We begin by estimating the model over *pairs* of consecutive years, which helps us curtailing the problem caused by missing observations over multiple panel waves, which significantly reduces the size of the available sample. Moreover, the computation of the standard errors to be attached to our GMM estimates is currently work in progress and we are not able to draw precise statistical inference yet. However, presenting our preliminary results constitutes prima facie evidence of whether our model yields intuitively meaningful insights. As a check on the limitations of our GMM estimation, we parallel our results with estimates from Non-Linear Least Squares (NLS) and standard Fixed-Effects models. The results are reported in table (4).

First, the GMM estimates consistently show a value of ψ larger than 1, evidence that workers with a comparative advantage in self-employment tend to be rewarded if they do choose to be self-employed, i.e. evidence of positive sorting on unobservables into the self-employed sector (consistent, for example, with self-selection). In our first models we only control for physical capital, leaving aside human capital and other observables that we will include at a later stage. In this model, the time and sector-specific constant for the self-employed sector is larger than the one for wage employment: i.e. with the exception of 2009-2010, there is a positive earnings premium for wage employees over the self employed once we control for unobserved heterogeneity and capital (($\delta^{SE} - \delta^{\omega}$>0).

Our dataset contains additional information about the reasons for a sectoral change. Namely, we can separate *voluntary* from *involuntary* separations. Using this data we will test whether self-selection into different occupations is indeed the driving force behind our results. As the data requires further cleaning, we are unable to provide these results in the current draft.

5.1 Workers' comparative advantage

In this section we use the results of our estimation to compute the individual levels of comparative advantage (θ_i). Recovering those individual parameters will inform us of the degree of positive selection of workers in the labour market. Namely, we will analyse the distribution of θ among the wage-employees and the self-employed. Following our model, estimates of θ_i can be obtained as:

$$\hat{\theta}_i = \frac{1}{T}\sum_{1}^{T}\left(\frac{R_{i,t} - \hat{R}_{i,t}}{1 + d_{i,t}^{SE}\ \left(\hat{\psi} - 1\right)}\right) \tag{21}$$

$$\hat{\theta}_i = \frac{1}{T}\sum_{1}^{T}\left(\frac{R_{i,t} - (\hat{\delta}_t^w + d_{i,t}^{SE}\ \hat{\delta}_t + \hat{\gamma}^w\ \ln x_{i,t} + d_{i,t}^{SE}\ \ln x_{i,t}\ \hat{\gamma})}{1 + d_{i,t}^{SE}\ \left(\hat{\psi} - 1\right)}\right) \tag{22}$$

Upon plotting the density of θ_i by sector of employment in Figure (5) - (7), the first interesting finding is that there is very large overlap in the unobserved heterogeneity of workers across the two sectors. However, we find that the distribution of θ for wage workers (first order) statistically dominates that of the self-employed. This means that despite a wage structure in the wage employed sector that does not favor individuals who have good characteristics for self-employment, many of these individuals are actually found in the wage employed sector.

Finally, the estimated model parameters can be employed to answer the central question to this paper: taking into account the unobserved heterogeneity, what are the causes of the increase in self-employment over time? We can analyse differences in returns to observables, unobservables as well as the sector dummy δ which - given we are conditioning on observables and unobservables and in across sectors - can be interpreted as an indication of a lack of perfect mobility across sectors. Other candidate explanations for changes in sector composition of the labour force include changes in risk or hedonic characteristics.

How many new self-employed are choice and how many necessity self-employed[10]?

One statistic which generalises the question of choice or necessity self-employment is "For how many individuals would it have been beneficial to move into self-meployment"?, i.e. we are interested in $E\left(R_t^j - R_{t-1}^k\right)$ for $j, k \in \{SE, w\}$ and $j \neq k$. Given our non-linear panel model we can only identify $E\left(R_t^w - \frac{1}{\psi}R_{t-1}^{SE}\right)$ and $E\left(R_t^{SE} - \psi\ R_{t-1}^w\right)$ however.

On the one hand, developing countries are still characterised by the lack of widely accessible formal banking infrastructure. On the other hand, informal saving mechanisms (e.g. susu collectors in Ghana) are characterised by astonishingly

[10]Note that without additional assumptions it is not possible to answer this question in levels, i.e. to estimate the share of choice vs. necessity employed in the population. This would require information of the returns to human capital in both sectors and not only of the quasi-difference.

Table 4: First Results

	GMM (04-05)	GMM (05-06)	GMM (09-10)	NLS (04-08)	FE (04-08)
$\delta^{SE} - \delta^{w}$	0.091	0.056	-0.031	0.343 (.326)	0.218 (0.198)
$\beta^{W}_{Male} - \Psi\beta^{S}_{Male}$				-0.287** (0.177)	-0.287** (0.127)
$\beta^{W}_{Educ} - \Psi\beta^{S}_{Educ}$				0.064** (0.025)	0.057*** (0.017)
$\beta^{SE}_{\ln K}$	0.078	0.023	0.069	0.033 (0.031)	0.029 (.026)
$\beta^{W}_{\ln K}$	0.002	0.138	-0.003	0.039** (0.017)	.076*** (0.016)
Ψ	1.122	1.136	1.551 (0.99)	1.033	
Obs.	384	506	543	1440	2126
R^2				.899	.909

low interest rates (sometimes negative in real terms). Whilst these forms of savings may have other justifications, the observed returns to capital in these widely used schemes almost certainly imply (**??**).[11]

6 Beyond average earnings

The framework presented so far assumes that workers' objective function consists of maximising expected material gains. Alternative factors determining the choice of self-employment may of course then be invoked to challenge the classification of individuals who are predicted to earn less on transiting to self-employment. The literature has considered two factors in particular (excluding capital constraints, which we assume to be complete).

First, empirical evidence suggests that job-satisfaction is higher among self-

[11] Clearly, the existence of a uniform and exogenous return to saving in the economy that applies to all agents is an over-simplification. Considerably more realistic will be, in a more complex version of the model, to allow for heterogeneity in workers' access to credit markets.

Figure 5: Estimated comparative advantage effect by sector (2004-2005)

Figure 6: Estimated comparative advantage effect by sector (2005-2006)

Figure 7: Estimated comparative advantage effect by sector (2009-2010)

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

With respect to risk aversion, empirical evidence from urban Ghana supports the intuition that self-employed individuals have lower levels of risk aversion (REF). Hence, given that our model assumes risk-neutrality, we are implicitly raising the attractiveness of self-employment. Our estimated share of necessity self-employed movers should thus be viewed as a lower bound on the actual figure, that would emerge if attitudes to risk were taken into account [12].

Similarly, differences in job satisfaction that derive from non-pecuniary job-attributes or unobservable individual characteristics, may also be related to occupational choice. If non-material benefits increase the attractiveness of the self-employed option, our estimated share of necessity self-employment will be an

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

over-estimate of the truth. To deal with this problem, we can consider

the evolution of job satisfaction analogously to that of wages. Alternately, we can consider job satisfaction (for which a subjective indicator is available in the data) as one of the factors influencing wages in the two sectors and test whether differences in working conditions compensate for part of the earnings differential.

7 Conclusion

This article has argued that despite a multiplicity of self employment experiences, two distinct conditions (with strong welfare implications) exist and can be empirically identified: on the one hand involuntary or *necessity self employed* workers would prefer a wage-job at the going wage rate and, on the other hand, the voluntary or *choice* self employed, who benefit from the relatively high returns to capital available in many developing countries. We propose a model that allows us an empirical translation of this distinction and analyze the causes of the increase in self-employment under this perspective.

Identification of the fraction of necessity self employed movers relies on information about savings and hypothetical earnings in the two sectors, which we calculated using information on transitions, earnings, personal characteristics and data on capital holdings - all of which are available in the Ghana Household Urban Panel Survey. Finally, we allow for an unobserved relative performance effect capturing the possibility of differential unobserved heterogeneity in wage employment and self employment.

We believe the distinction between choice and necessity of self-employment is meaningful because there is strong evidence of a lack of labour market clearing as well as very restricted financial markets. The distinction is relevant for policy discussions, since controversy exists about the desirability of the strong increase in self-employment that has been observed in many African countries over the last ten to twenty years. The focus of development work - whether capital market or labour market imperfections ought to be targetted, for example - often presupposes a view of whether self-employment is chosen or suffered. Surprisingly, there has been very little direct evidence on this. In particular, it ought to be noted that studies which focus on whether individuals are engaged in job search - the indicator with which standard labour market models test for the degree of market imperfection in wealthy countries - may be much less appropriate in a situation where the formal wage labour market is so small that a large proportion of job-less workers may be discouraged. They nevertheless suffer the consequences of self-employment.

The extent of necessity self-employment among the newly self-employed thus appears to be an important question and one to which we this paper finds a quantitative answer using a novel methodology with very few precedents in developing countries. Our results show, first, that the structure of earnings in the two sectors should encourage sorting on unobservables: i.e. it is the case that individuals who have a comparative advantage in self-employment are rewarded by higher earnings - ceteris paribus. Second, our model gives us some indication about what explains the wage gap in favour of the wage employees? Using the GMM model allowing for correlated random coefficients our estimate of the pure sector effect (δ) is actually negative at the beggining of the period, indicating a self-employment wage premium. The effect may however be insignificant (standard errors are work in progress) as in the non-linear least squares and fixed effects estimations.

References

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.

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.

BLUNDELL, R. AND S. BOND (2000): "GMM Estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19, 321–340.

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

EVANS, D. AND B. JOVANOVIC (1989): "An estimated model of entrepreneurial choice under liquidity constraints," *The Journal of Political Economy*, 97, 808–827.

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.

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.

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.

POSCHKE, M. (2010): "Entrepreneurs out of Necessity A Snapshot," *IZA Discussion Paper*, 4893.

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.