

Falco, Paolo; Haywood, Luke

Article — Accepted Manuscript (Postprint)

Entrepreneurship versus joblessness: Explaining the rise in self-employment

Journal of Development Economics

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Falco, Paolo; Haywood, Luke (2016) : Entrepreneurship versus joblessness: Explaining the rise in self-employment, Journal of Development Economics, ISSN 0304-3878, Elsevier, Amsterdam, Vol. 118, pp. 245-265, <https://doi.org/10.1016/j.jdeveco.2015.07.010>

This Version is available at:

<https://hdl.handle.net/10419/202468>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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



<http://creativecommons.org/licenses/by-nc-nd/4.0>

This is the postprint of an article published in *Labour Economics* 38 (2016), pp. 1-11,
available online at: <http://dx.doi.org/10.1016/j.labeco.2015.10.002>

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

Paolo Falco[†] and Luke Haywood[‡]

July 27, 2015

Abstract

A large share of the workforce throughout the developing world is self-employed, and this proportion has increased in recent decades. Assessments of this development vary, with pull factors such as high returns to capital contrasted with push factors such as barriers to more desirable salaried jobs. Using a long panel dataset from Ghana, we empirically investigate the changing structure of earnings in self-employment relative to salaried work. Our analysis allows workers to sort into self-employment for different reasons, including expected earnings differentials and endogenous capital accumulation whilst highlighting constraints workers may face. We estimate a correlated random coefficient model using IV-GMM, taking into account the endogeneity of sector choice and capital accumulation. We find that returns to productive characteristics in self-employment have increased significantly between 2004 and 2011 whilst self-employment has attracted increasingly skilled workers. This points against the negative view of self-employment as an occupation of last resort.

JEL: O15, J24, J42, C14

Keywords: self-employment, African labour markets, semiparametric models, comparative advantage, segmentation.

© 2016. This manuscript version is made available under the CC-BY-NC-ND 4.0 license



*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, Federico Cingano, Andrew Clark, Marc Gurgand, Clément Imbert, David Margolis, Jean-Marc Robin, Francis Teal, Christopher Udry, two anonymous referees and seminar and conference participants in Aix-en-Provence, Düsseldorf, Göteborg, Marseille, Oxford and Paris. The usual disclaimer applies.

[†]OECD & University of Oxford, 2, rue André Pascal, 75016, Paris, paolo.falco@oecd.org

[‡]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 of various types - smallholder farmers, retail business owners, basic manufacturers and petty traders. Recent large-scale household data show an increase in self-employment especially 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 self-employed workers may thus indicate that increased wealth and better-functioning credit markets are relaxing capital constraints and creating investment opportunities (Evans and Jovanovic (1989), Magnac and Robin (1996) and Blanchflower and Oswald (1998)). This interpretation underlies the support in policy circles for microcredit and entrepreneurship. A more *pessimistic* perspective notes that desirable salaried work is unavailable for many individuals 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 related to unemployment or job queues à la 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.

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 do differences in returns and workforce composition combine to explain sector earnings differentials? 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? In answering these questions we focus especially on changes over the past decade (2004-2011) in order to shed new light on the recent rise in self-employment.

We concentrate on the choice between self-employment and wage employment in urban areas. This focus is informed by increasing urbanisation, the role of informality and policy relevance. Whilst 85% of the Ghanaian population lived in rural areas in the 1950s, that figure has come down to 50% today and is projected to decrease to 30% by 2050. This trend

has been accompanied by a rapid rise in self-employment, as recent arrivals in search for better opportunities in urban areas are likely to start in self-employment (Fields (1990)). By contrast, in rural areas the most important occupational decision is typically whether to stay (and most likely be employed in low-productivity agriculture), or migrate to urban centres - as analyzed since Harris and Todaro (1970). In 2005, wage employment accounted for only 7.3% of employment in rural areas and 33.8% in urban areas (Ghana Living Standard Survey (GLSS)). Conversely, the share of non-agricultural self-employment was very low in rural areas (14.9%), but over 50% in urban areas. Only recently have non-farm alternatives in rural contexts been investigated - and the fall-back option remains agriculture (Chawanote and Barrett (2014)). By contrast, the fall-back option in urban areas is typically marginal self-employment. Distinguishing between “opportunity” versus “last resort” self-employed (Launov and Guenther (2012)) or “choice” versus “necessity” self-employment (Poschke (2010)) is thus a particular challenge of urban areas¹.

Allocation of workers to different parts of the labour market has been viewed as a problem of occupational *choice*. However, the hierarchy between necessity- and last-resort-self-employment suggests significant constraints to sector choice. The standard “single-crossing property” of selection models (certain variables monotonously increase the likelihood of selection into a sector) is then likely to be violated. Nguimkeu (2014) builds a model of occupational choices to explain how workers sort themselves between subsistence self-employment and entrepreneurship taking into account such constraints. His model stresses the importance of individual unobservable skills in determining sector allocation. Without panel data, however, identification of individual effects relies on strong assumptions since it cannot be informed by observing movers between different labour market states.

Robust empirical modelling of segmented labour markets hinges upon being able to account for unobservable worker skills and concomitantly tackle the endogeneity of sector movement and input choices. Models with sorting on unobservable comparative advantage

¹The segmentation of the Ghanaian labour market into wage or self-employment is closely linked to the segmentation between formal and informal employment since the vast majority of self-employed workers in Ghana work informally, while formality is largely concentrated among wage-workers. Jutting et al. (2009) estimate that over 70% of informal workers were self-employed in the 1990s. Our analysis of the choice between self- and wage-employment, therefore, also sheds light on the relative merits of formality and informality.

date back to Roy (1951). Since the costs of entry into self-employment may vary in ways related to unobservables (e.g. as a result of differential credit constraints), we require a model that is robust to different forms of selection. We allow for a semi-parametric relationship between returns to self-employment and individual earnings ability - more beneficial for some individuals and less so for others, in line with the large effect heterogeneity that the literature has highlighted. This gives rise to a correlated random-coefficient (CRC) model which we estimate using Instrumental Variable GMM. Lemieux (1998) contrasts earnings of union and non-union members in this framework. Suri (2011) uses a CRC model to shed new light on farmers' decision to adopt hybrid seeds in Kenya. Imbert (2012) studies wage-premia in the public sector of Vietnam. These studies do not allow for determinants of sector choice to be related to past realizations of earnings, however. Since the literature shows that savings are an important factor explaining transition into self-employment, we want to allow past earnings shocks to determine capital accumulation and sector choice. In line with parallel work by Nyshadham (2014) we thus relax the assumption of strict exogeneity which inhibits any feedback from past earnings to future labour market behaviour. Nyshadham (2014) focuses on the individual dimension of occupational choice as a function of learning about earnings and does not estimate returns to capital in self-employment. By contrast, we allow greater flexibility in the sorting mechanisms (by imposing less structure on it) and estimate changes in returns to self-employment, focusing on both observable and unobservable inputs. In line with this focus on macro-level changes, we split our unusually long panel dataset into two periods and study the evolution of returns to self-employment over time. Contrary to Suri (2011), Imbert (2012) and Nyshadham (2014) who estimate their CRC models using minimum distance based on projections following Chamberlain (1982), we use IV-GMM. Since the two strategies are asymptotically equivalent in our formulation, our preference is largely practical. We also believe there is value added in presenting an additional tool to analyze this class of models.

If the increase in self-employment in Ghana is largely due to motivations of “last resort”, we should expect declining levels of capital, schooling and unobserved ability among the self-employed (as the best select out of the sector); and, in parallel, we might expect a significant (and possibly increasing) wage-employment premium. If, instead, increases in

self-employment are largely based on opportunity motives, growing returns to human and physical capital in self-employment and a decreasing wage premium may be more likely. 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 and human capital. Taken together, the two findings lend support to the hypothesis that the sector is increasingly attracting productive resources. Furthermore, we document a decreasing (unexplained) self-employment premium, which we tentatively interpret as evidence of decreasing barriers to enter the sector (e.g. credit constraints). This result is consistent with the strong growth rates experienced by Ghana over the period, which allowed for poverty reduction and capital accumulation. While the self-employed 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 estimation strategy. Section (5) presents the data. Section (6) discusses the results and assesses the importance of estimated changes in returns in explaining recent labour market trends in Ghana. Section (7) concludes.

2 The Labour Market in Ghana

Ghana is one of the most stable countries in West Africa and has shown strong economic growth rates over the last decade². Nsowah-Nuamah et al. (2012) find that the poverty rate halved between 1991 and 2005. They also find that this occurred whilst employment in the public sector fell and employment in small firms increased. Indeed, next to the traditionally important role of the agricultural sector, Robson et al. (2009) cite evidence that in Ghana “[t]he economic structure is polarised between a small number of large corporations and large volume of micro and small enterprises”.

<< **Table 1 here** >>

²According to the World Bank’s World Development Indicator database, GDP growth rates over the period this paper considers were 5.6% (2004), 5.9% (2005), 6.4% (2006), 4.0% (2007), 8.4% (2008), 4.0% (2009), 8.0% (2010) and 15.0% (2011).

Reporting evidence from the Ghana Statistical Service, Baah-Boateng (2004) shows rising rates of self-employment in the 1990s (see table (1)), especially non-farm informal self-employment, which is what our urban sample covers. Palmer (2007) notes that “[m]ost new jobs are created in the informal economy, with formal sector employment growth largely stagnant.”. 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.³ The panel dataset at our disposal, the Ghana Urban Household Panel Survey (GUHPS), is not well-suited to settle this question, since it follows the same group of workers over time (census data would be more appropriate). However, by carefully recording the movement of workers between wage- and 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 earnings 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 comparative advantages may determine earnings and choice of sector. However, we impose minimal structure on the selection process. We believe that selection on unobservables will occur, but we do not impose a specific structure on the mechanism, nor assume that we can identify variables which can serve as instruments for occupational choice (by satisfying the single-crossing property, as in Heckman (1976)).

3.1 Sector Earnings

Let worker i be endowed with time-varying characteristics (in particular, physical capital $K_{i,t}$), time-invariant characteristics (e.g. human capital h_i) and unobservable productivity θ_i^j for $j \in \{SE, w\}$ varying across two sectors: self-employment (SE) and wage employment (w). Our definition of physical capital encompasses both liquid savings and assets⁴. We take human capital as given. We then assume that log earnings R for worker i in wage- (w) and

³See table (1) in Nsowah-Nuamah et al. (2012).

⁴Under the assumption that the latter can easily be sold and re-invested.

self-employment (SE) are

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

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

where $K_{i,t}$ and h_i indicate physical and human capital; δ_t^{SE} designates macroeconomic effects on the self employed (viz. $\delta_t^w - \delta_t^{SE}$ the 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. Using measures of capital for workers in *both* sectors, we are able to estimate relative returns to assets across employment categories and shed further light on the link between capital and sector choice.

Based on this simple earnings model, workers may sort into self-employment or salaried work. However, we will not assume that workers sort into the sector with highest returns. Instead, we allow for flexibility in the sorting mechanism and rather concentrate on the resulting structure of earnings across sectors⁵.

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 firms and the public sector) and diverse self-employment on the other side. Similar settings have been analysed in the literature using variants of the classic Roy model based on specialisation by comparative advantage. We relax the assumption of free sector choice underlying most of these models. First, access to formal jobs may be rationed, in the sense that the number of workers who are willing to work for the current wage (at a given skill level) exceeds the number of available jobs.⁶ If that is the case, workers may need to queue for their preferred job (as in the classic Harris-Todaro framework) and the allocation mechanism may not necessarily be in line with sorting by comparative advantage. For instance, if firms cannot perfectly observe

⁵We do not directly consider unemployment, but at the end of section (4.1) we show how search unemployment might be related to earnings differentials across sectors.

⁶This may occur as a result of efficiency wages, legal minimum wages or union wage bargaining. Modeling the exact process of 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.

workers' skills, they may be unable to choose the best suited candidates for their jobs. In addition, non-productive factors, such as personal connections and networks, may play a significant role in the selection process (e.g. Heath (2011), Cingano and Rosolia (2012)). Second, workers themselves may not know their comparative advantage. By not specifying the precise nature of the sorting mechanism, we do not assume perfect knowledge of θ^{SE} and θ^w . This allows workers to learn their comparative advantage over time and move to the most suited sector⁷. For these reasons, we believe one should best remain agnostic about the exact nature (and direction) of the selection process. We return to the empirical implications of barriers to sector choice after presenting our empirical framework.

4 Empirical Strategy

4.1 The Correlated Random Coefficients Model

In this section we outline our empirical strategy to identify sector earnings differentials under the assumption that job queues may exist and selection does not necessarily follow simple patterns. Our panel-based approach enables us to identify the effects of self-employment on different determinants of earnings, whilst allowing for unspecified patterns of selection on unobservables.

Combining (1) and (2) gives the following switching model of earnings in the two sectors

⁷We think this is one factor in the choice of wage- versus self-employment but certainly not the only one. It may be more important if we considered larger household businesses which are the focus of Nyshadham (2014), who presents a model of entrepreneurial choice based on learning. If learning were the predominant determinant of transitions, we should observe a strong decrease in transitions across wage- and self-employment over workers' lifetimes and that the difference in predicted earnings across sectors increases. We tested this using the model's estimated parameters and found that the probability of being in the "wrong" sector only declines slowly from a peak of around 40% to around 30% between ages 30 and 55 (available on request). Raw transition rates decline somewhat with age also.

(using $D_{i,t}^{SE}$ as a dummy indicating self-employment of worker i at time t):

$$R_{i,t} = \delta_t^w + D_{i,t}^{SE} \left(\delta_t^{SE} - \delta_t^w \right) \quad (3)$$

$$\begin{aligned} & + \alpha^w \ln h_i + D_{i,t}^{SE} \ln h_i \left(\alpha^{SE} - \alpha^w \right) \\ & + \beta^w \ln K_{i,t} + D_{i,t}^{SE} \ln K_{i,t} \left(\beta^{SE} - \beta^w \right) \\ & + \theta_i^w + D_{i,t}^{SE} \left(\theta_i^{SE} - \theta_i^w \right) + \varepsilon_{i,t} \end{aligned} \quad (4)$$

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

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 SE vis-à-vis the wage sector.⁸ Combining these additional elements, we obtain:⁹

$$\theta_i^{SE} = \psi \theta_i + \tau_i \quad (7)$$

$$\theta_i^w = \theta_i + \tau_i \quad (8)$$

⁸Whilst we make few functional form assumptions, a key assumption is the multiplicativity of ψ . As 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 at our disposal, non-parametric techniques appear infeasible.

⁹We can justify separating the absolute advantage component, τ_i , from a comparative advantage component of individual unobserved heterogeneity, θ_i , by using a projection. As Suri (2011) notes, one can easily see that the τ_i in equations (5) and (6) are the same by subtracting (6) from (5) and noting that $b_{SE} + b_w = 1$ by construction.

$$\theta_i^{SE} = b_{SE} \left(\theta_i^{SE} - \theta_i^w \right) + \tau_i \quad (5)$$

$$\theta_i^w = b_w \left(\theta_i^{SE} - \theta_i^w \right) + \tau_i \quad (6)$$

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

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 (7) and (8) 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 (7) and (8) directly follow.

Earnings can then be written as

$$\begin{aligned}
 R_{i,t} = & \delta_t^w + d_{i,t}^{SE} \left(\delta_t^{SE} - \delta_t^w \right) \\
 & + \alpha^w \ln h_i + d_{i,t}^{SE} \ln h_i \left(\alpha^{SE} - \alpha^w \right) \\
 & + \beta^w \ln K_{i,t} + d_{i,t}^{SE} \ln K_{i,t} \left(\beta^{SE} - \beta^w \right) \\
 & + \theta_i + d_{i,t}^{SE} (\psi - 1) \theta_i + \tau_i + \varepsilon_{i,t}.
 \end{aligned} \tag{9}$$

The coefficient ψ is central to our analysis, as it captures differential returns to comparative advantages and it helps us shed light on the sorting mechanism. If $\psi = 1$ unobservable individual characteristics are not remunerated differently across sectors and we would expect 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 self-employment. By contrast, if $\psi < 1$ there is a premium for these workers in wage employment. The sign of $\psi - 1$ thus determines earnings incentives for sorting on unobservables¹⁰. In summary, if we wish to make no assumptions about the nature of 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.

Note that our flexibility with respect to the possibility of choosing the “wrong” sector contrasts with standard two-stage models that instrument for sector choice or rely on a specific parametrization to control for endogenous selection. These models are typically based on the assumption that certain factors monotonously increase the likelihood of finding a wage job - the so-called single-crossing property (Heckman (1976)). We believe this assumption is too strong and we know of no instruments in our data that determine selection but do not influence earnings independently. We thus follow Lemieux (1998) and Suri (2011) in estimating a model that allows for unspecified selection on individual unobservables.

Alternately, we could think of workers choosing their sector subject to entry costs (e.g. costs of access to credit or job search costs) that may vary across individuals. Magnac (1991)

¹⁰Note that with $\psi < 1$ the degree of inequality across sectors is lower than it would be if sector allocation was random, whereas for $\psi > 1$ the opposite holds. Specialisation may nevertheless decrease inequality overall, see Heckman and Honore (1990).

models this scenario in a way that will prove useful for interpreting our results: A worker's choice to become self-employed (D^{SE}) is given by

$$Pr(D^{SE}) = Pr\left(E\left(R^{SE}(x_i)\right) > E\left(R^w(x_i)\right) - c(x_i)\right),$$

where the existence of relative entry costs $c(x_i)$ may justify the persistence of sector differentials in equilibrium. That is, we may expect a difference in mean earnings across sectors (a wage premium) corresponding to the mean value of c conditional on observable and unobservable dimensions of worker productivity¹¹. In the empirical analysis we will test for the existence of such a premium.

4.2 Identification

All determinants of earnings in our model may have sector-specific returns - including the unobservable individual effect. Since identification of the individual effects relies on using earnings from several periods jointly (to assess earnings persistence), panel data models typically place conditions on feedback from errors in one period on covariates in other time periods. In common with the identifying assumptions of other correlated random effects models (Lemieux (1998), Suri (2011), Imbert (2012), Nyshadham (2014)) we posit that the unexpected part of earnings (the “shocks” $\varepsilon_{i,t}$) must not influence individuals' choices in certain other periods.

One option is to assume strict exogeneity of choices with respect to the entire history of transitory shocks - as in Chamberlain (1982), Lemieux (1998) and the vast literature using linear panel data models. Suri (2011) justifies this assumption in her context of agricultural technology choice on the grounds that input choices, in any given period, are made before the most important transitory shocks hit (in her case, rainfall). The same logic could plausibly apply in our context, given that sector allocation and capital investments are most likely the result of careful decision processes and unlikely to be affected by contemporaneous shocks. However, strict exogeneity also implies that shocks occurring in a given period t do not affect input choices in *subsequent* periods. Crucially, strict exogeneity implies that

¹¹This is true conditional on observing employed individuals. Following Harris and Todaro (1970), costs and benefits of *intending* to enter a sector are equalized in expectation. Differences in rewards then exist for those who find employment.

conditional on the model's current observable and unobservable determinants of earnings, mobility across sectors is random. We feel much less comfortable with this assumption since it will be violated if high earnings in one period are saved and influence future sector choices. Thus, in line with parallel work by Nyshadham (2014), we relax the assumption of strict exogeneity and let input and sector choices depend on *past* shocks. Shocks are allowed to have a *lagged* effect on entrepreneurial choices, whilst current shocks must not affect household choices in the same period. If our earnings function includes the important determinants of earnings, shocks cannot easily be foreseen. This implies that there must be no serial correlation in the error term $\varepsilon_{i,t}$. Our assumption, then, is that sector choice is random conditional on predicted earnings, i.e. individuals base their mobility decisions on expected earnings.

We now consider the intuition for the identification of the different constituents of our earnings function. Parameters giving sector-specific returns (β^w, β^{SE}) are identified through variation in earnings across different levels of inputs (e.g. capital) *within* each sector.¹² Identification of δ relies on movers across sectors and these cross-sector differences are only identified up to a function of ψ . Identification of our comparative advantage effect ψ relies on observing how earnings of individuals with different unobservable skill levels change as they move between wage and self-employment¹³.

4.3 Estimation

As in most panel models, we treat θ_i as a nuisance parameter we need to remove for consistent estimation. However, standard differencing or de-meaning will not remove the individual effect (θ_i), since it is interacted with self-employment status (D_i^{SE}). There are then two alternative routes to estimation, both exploit strict or sequential exogeneity.

¹²Our measures of human capital such as education, experience or age do not change over time (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 comparing earnings differentials across sectors. Since we are not performing linear differencing, but quasi-differencing we identify the quasi-difference $\alpha^{SE} - \frac{1}{\psi} \alpha^w$.

¹³Intuitively, the standard fixed effects model is overidentified as it allows for an individual fixed effect and a sector fixed effect, but constrains the interaction of both to be zero. The sector-specific returns to unobserved skill (ψ) can be interpreted as a coefficient on the interaction effect of unobserved skill and the sector dummy. Analogously to other interaction effects, ψ will be identified as long as there is variation in the interaction that is different from its constituent terms.

One strategy is to replace the unobserved individual effect by its projection on covariates in all (or certain) periods. Strict (or sequential) exogeneity ensures the resulting projection can be inserted into the original outcome equation without causing endogeneity. One can then decompose the raw coefficients to obtain the parameters of interest¹⁴. Alternatively, one can remove the individual effect using the earnings function, by differencing across both time and sector (quasi-differencing). One can then exploit strict or sequential exogeneity to generate moment conditions in a GMM framework. We follow Lemieux (1998) in taking this route and outline details below.

Importantly, the two strategies are asymptotically equivalent if the instruments used for estimation in the IV-GMM context are the same that are used to project the individual unobserved effects on to¹⁵. We have no strong preference for either of the two approaches, but two considerations encourage us to present the first estimation of a correlated random coefficient model with partially endogenous variables estimated by IV-GMM.

First, instruments are easily understood. This allows for a more straightforward discussion of identifying assumptions than using the projection approach. Second, the selection of covariates in the projections of θ_i may be called into question. Related to this, whilst adding more waves and covariates is trivial with IV-GMM, projections become increasingly “cumbersome” (Suri (2011)). Given the number of waves in the GUHPS panel, we thus prefer the flexibility of IV-GMM. By contrast, GMM estimators have sometimes been found to be unstable in small samples and the choice of the weighting matrix may have important implications for the results when models are over-identified (like in our case). We thus add numerous robustness checks, including a specification using the identity matrix as an alternative to optimal weighting. On the important issue of instrument validity we feel confident that we are making less restrictive assumptions than are standard in the literature by adopting sequential exogeneity rather than strict exogeneity.

¹⁴For details, see Chamberlain (1982) or Suri (2011).

¹⁵Angrist and Newey (1991) present a proof for the linear case with strictly exogenous covariates. The equivalence also holds in our context. The additional restriction of sequential exogeneity will be reflected in the choice of variables to project on to (in the projection approach) and in the choice of instruments (in the IV-GMM setting). A proof is available upon request.

Following Lemieux (1998) we now use our earnings equation (9) to solve for θ_i .

$$\theta_i = \frac{R_{i,t} - \left[\delta_t^w + D_{i,t} (\delta_t^{SE} - \delta_t^w) + D_{i,t} \beta^w k_{i,t} + D_{i,t} (\beta^{SE} - \beta^w) k_{i,t} + \alpha^w h_i + D_{i,t} (\alpha^{SE} - \alpha^w) h_i + \tau_i + \varepsilon_{i,t} \right]}{1 + D_{i,t} (\psi - 1)} \quad (10)$$

where $k_{i,t} = \ln(K_{i,t})$ and for convenience we define $\delta_t \equiv \delta_t^w + D_{i,t} (\delta_t^{SE} - \delta_t^w)$ and

$$G_{i,t}(k, h, D) \equiv \beta^w k_{i,t} + D_{i,t} (\beta^{SE} - \beta^w) k_{i,t} + \alpha^w h_i + D_{i,t-1} (\alpha^{SE} - \alpha^w) h_i.$$

Combining our expression (10) for θ from t with the equivalent expression at $t - 1$:

$$\frac{R_{i,t} - G_{i,t}(k_{i,t}, h_i, D_{i,t}) - \delta_t - \tau_i - \varepsilon_{i,t}}{1 + D_{i,t} (\psi - 1)} = \frac{R_{i,t-1} - G_{i,t-1}(k_{i,t-1}, h_i, D_{i,t-1}) - \delta_{t-1} - \tau_i - \varepsilon_{i,t-1}}{1 + D_{i,t-1} (\psi - 1)} \quad (11)$$

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

$$R_{i,t} = G_{i,t} + \delta_t + \varepsilon_{i,t} + \tau_i + C_{i,t} (R_{i,t-1} - G_{i,t-1} - \delta_{t-1} - \varepsilon_{i,t-1} - \tau_i) \quad (12)$$

The transformed model has now removed the influence of the unobservables θ_i and we can now use the following quasi-differenced errors to construct a set of moment conditions:

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

where

$$erm_{i,t} \equiv (\varepsilon_{i,t} + \tau_i) - C_{i,t} (\varepsilon_{i,t-1} + \tau_i). \quad (13)$$

Another way of expressing these quasi-differenced errors considers the different sector histories (i.e. the combinations of all possible employment histories over two consecutive

periods):¹⁶

$$\begin{aligned}
erm_{i,t} = & (R_{i,t} - G_{i,t}) - C_{i,t} (R_{i,t-1} - G_{i,t-1}) - D_{i,t} D_{i,t-1} (\delta_t^{SE} - \delta_{t-1}^{SE}) \\
& - (1 - D_{i,t}) D_{i,t-1} \left(\delta_t^w - \frac{1}{\psi} \delta_{t-1}^{SE} \right) \\
& - D_{i,t} (1 - D_{i,t-1}) (\delta_t^{SE} - \psi \delta_{t-1}^w) \\
& - (1 - D_{i,t}) (1 - D_{i,t-1}) (\delta_t^w - \delta_{t-1}^w) \quad (14)
\end{aligned}$$

The model can then be estimated by minimizing the empirical analogue of (16) upon identifying a suitable set of instruments ($Z_{i,t}$), recognizing that τ_i is a random effect (conditional on θ_i)¹⁷.

$$E(erm_{i,t} Z_{i,t}) = 0 \quad (16)$$

Under strict exogeneity, we would simply use all covariates in all time periods and minimize $\sum_t E(X_i' erm_{i,t})$ where $X_i = [K_{i,1}, D_{i,1}, K_{i,2}, D_{i,2}, \dots]$. Based on the assumption of sequential exogeneity, our valid instrument set, Z , contains suitably chosen lags of the endogenous variables. Importantly, we allow input choices to be affected by past (but *not contemporaneous or future*) shocks, i.e. they are predetermined¹⁸. The most recent valid instruments are then the first lags of the endogenous variables, Z_{t-1} , since $erm_{i,t}$ contains

¹⁶That the following expression is equivalent to equation (13) can easily be seen by going through the possible combinations of values that $D_{i,s}$ can take for $s \in \{t, t-1\}$.

¹⁷Given our quasi-differenced framework, the final 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:

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

¹⁸An even less restrictive alternative would be to allow for sector and input choices to be affected by past *and contemporaneous* shocks. Our first valid instrument is Z_{t-2} and we have fewer exclusion restrictions:

$$E[Z_{i,s} \varepsilon_{i,t}] = 0 \quad \forall t > s \quad (17)$$

$$E[Z_{i,s} \tau_i] = 0. \quad (18)$$

This strategy is also more demanding of the data, however, as we now need at least three waves of consecutive information on each individual. We found that results were not robust. We thus base our estimation on the assumption of predeterminedness.

both $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$. Our identifying assumptions can thus be summarised as follows,

$$E[Z_{i,s}\varepsilon_{i,t}] = 0 \quad \forall t \geq s \quad (19)$$

$$E[Z_{i,s}\tau_i] = 0. \quad (20)$$

We document the details of how we construct our instrument matrix in appendix (B).

5 Data

We apply our model to data from the Ghana Urban Household Panel Survey (GUHPS), conducted yearly since 2004 by the Centre for the Study of African Economies (CSAE) at the University of Oxford¹⁹. The GUHPS 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.²⁰

Figure (1) splits 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 between self- and wage-employment.

<< **Figure 1 here** >>

Table (2) shows changes in the workforce composition for the sample of GUHPS respondents who are employed either in our earlier or later estimation sample (2004-6 or 2009-11).

<< **Table 2 here** >>

Contrasting the wage employed to the self-employed, we find that the latter are older on average - see figure (2). They are also less educated and more likely to be female (not shown).

¹⁹At the time of writing eight waves have been released. The survey was not conducted in 2007. Data from that year is obtained from recall questions in the 2008 questionnaire. The latest wave was conducted 1.5 years after the previous one and it took place in early 2012. For simplicity, we label it 2011 in order to maintain the yearly structure.

²⁰After the first wave, the sample expanded by incorporating new members of the original households and tracking original interviewees to their new households.

<< **Figure 2 here** >>

Table (2) shows that earnings grew much faster in self-employment, compared to wage employment, leading to a reversal of the initial advantage for wage workers in mean earnings. It also shows a strong catch-up of median earnings in self-employment (see figure (3)). The difference between average and median earnings is driven by a skewed distribution with relatively few highly rewarded jobs, typical of many developing countries. The large (and increasing) variance of self-employed earnings can be seen in figure (4).

<< **Figure 3 here** >>

<< **Figure 4 here** >>

Our empirical identification relies crucially on movers between wage- and self-employment. Table (3) summarises respondents' transitions between the two sectors over consecutive years of the panel. We find that between any two waves, 14.7% of all wage workers move to self-employment, while 11.2% of the self-employed move to wage-employment. The fraction of movers increases, naturally, when we follow individuals longer through the panel, reaching one quarter of all individuals as figure (5) shows.

<< **Figure 5 here** >>

A longer panel thus increases the number of movers. At the same time, however, as we increase the length of the panel we are confronted with larger rates of attrition. The issue is severe if we consider attrition from the initial sample over the full eight waves (see figure (15) in appendix (D)) - full information for all eight years is in fact only available for 42 individuals. The alternative is to estimate the model separately using shorter panels²¹. We could even focus on estimates from pairs of years. The disadvantage, however, is that longer panels allow better estimation of the individual effects (and of their sector-specific returns ψ). In the analysis we focus on two three-year periods, choosing the years at the two extremes of the calendar period in order to focus on the evolution of the labour market over time: from 2004 to 2006 and from 2009 to 2011²². Levels of attrition are then acceptable:

²¹Muris (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.

²²The choice of these specific years is informed by another fact: for the year 2007 we only have recall data collected in 2008. The number of movers between these two years is so low that transitions from 2007 to 2008 cannot be analysed, raising questions about the quality of data for 2007.

we have full information for over 300 workers in both periods, see figure (14) in appendix (D). Among these we still observe considerable (and increasing) mobility across sectors - over one quarter of workers move at least once during the three years, with over 10% moving more than once (see appendix figures (11) and (12)). Our analysis of attrition in appendix (D) also shows that sample attrition is largely uncorrelated with worker characteristics and earnings.

Finally, our choice of capital measures deserves a brief discussion. We have information on the value of agricultural land, real estate and household assets, for both wage- and self-employed workers²³. It also includes the value of business capital for the self-employed. 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 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 exclude these two categories and focus on the value of household and business assets. We are interested in the income stream or usage value generated by the assets respondents own, in any given period and for both sectors. However, since the majority of our respondents lacks access to formal banking and does not own real estate that can be rented out, we do not expect cash streams generated by asset ownership except for the self-employed. We could restrict ourselves to the reported earnings for both wage and self-employed. This would increase the opportunity cost of moving to wage-employment for the self-employed, since business capital would yield no monetary returns in wage employment. We believe, however, that assets such as cars, tools or machinery have a private usage value that should be taken into account. In an effort to provide a fair comparison across sectors, we thus impute a usage value of assets at a constant rate for all respondents²⁴.

²³The data does not contain information on financial assets, except for the final three waves. Moreover, these variables are likely to be subject to severe misreporting. For these reasons, we do not include them in the analysis.

²⁴The exact assumption we are making is that assets yield a modest 2% annual usage value which we add to monetary earnings. Our estimation results are not very sensitive to removing this assumption, but we were averted from doing so, so as not to bias the comparison between wage- and self-employment in favour of the latter.

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 estimate $\psi, \beta^{SE}, \alpha_{Educ}^{SE}, \alpha_{Male}^{SE}$, and the average difference across sectors $\delta^{SE} - \delta^w$. By contrast, we assume specific values for β^w, α_{Male}^w and α_{Educ}^w .

We set β^w since only a minority of individuals in our data have access to formal banking and may be faced with low or negative real interest rates. Distinguishing changes in asset holdings as a result of depreciation is thus difficult. Rather, we set β^w equal to 0.02 to target the average real interest rate available in Ghana over the period 2005-2009 (around 5%)²⁵. Given that we find large rates of return to capital in self-employment, we feel that using fairly high values in the benchmark estimation for wage employment is a conservative approach²⁶. We can only identify relative returns to time-invariant characteristics²⁷. 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)). We take the estimated value of α_{male}^w from Nopo et al. (2011) who find a male wage premium of 0.25% in Ghana. In appendix (F) we show that our results are not sensitive to changing any of these set values.

Table (4) presents our key results. Four aspects stand out. First, *returns to capital* are significantly higher in self-employment than in wage employment. Moreover, returns to capital in self-employment increase significantly over 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 (6) plots the annualised returns that our value of β^{SE} implies for different individuals. Calculated median annualised returns to capital are 81.13% in the first period and 84.08% in the second. These rates may appear

²⁵Our Cobb-Douglas formulation implies constant returns to log capital, generating a distribution of returns over different levels of capital.

²⁶As a robustness test, appendix (F) varies the value of β^w by an order of magnitude in either direction.

²⁷In fact, in our non-linear panel model only the quasi-difference across sectors ($\alpha^{SE} - \psi \alpha^w$) of time-invariant characteristics is identified.

unusually 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 of up to 250 – 300% for farmers of certain crops. The comparative advantage of being self-employed for individuals with enough capital has increased over time in all our specifications. The strong increase in $\hat{\beta}^{SE}$ is robust to changes in all calibrated parameters (as shown in appendix (F)) and is also evident in simpler OLS and Fixed Effects specifications (see table (6) in appendix (E)). We view this result as a relevant contribution to a literature which has often not been able to control for selection on unobservables when estimating returns to capital in self-employment.

<< **Table 4 here** >>

Second, *returns to human capital* (RORE) have significantly increased in self-employment over the period analysed. We find that during the period 2004-2006 returns to education in self-employment are lower than in wage employment (see table (4)), but they grow rapidly from the first to 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 (F.3) in the appendix).

Third, relative *returns to unobservable individual ability* have increased significantly over time, as shown by the increase in the estimated value of ψ in table (4). Self-employment is *increasingly* becoming a desirable opportunity for highly able individuals. However, when we estimate the distribution of $\hat{\theta}_i$ across sectors, we do not detect an increase in the proportion of highly skilled individuals in self-employment over the period, in line with the continued incentives for able individuals to join the wage sector as indicated by the fact that the estimated value of ψ remains lower than unity.

<< **Figure 6 here** >>

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, we focus the discussion on the trend of this gap over time, which indicates a significant change.

We contrast our IV-GMM results to estimates from simpler OLS and Fixed Effects models reported in appendix (E). The results show that failing to account for sector-specific returns to individual effects may lead us to misleading conclusions. Appendix (F) presents additional robustness checks, including an alternative estimation under the assumption that capital levels are exogenous.

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 the growth of self-employment and proposes an interpretation.

6.2 Understanding the Rise in Self-employment

Our estimated earnings functions can be used to determine counterfactual earnings in both sectors for all individuals. Hypothesizing how earnings differentials may influence individuals' choice of becoming self-employed, we now analyze to what extent changes in returns to self-employment can be related to the recent increase in self-employment.

One way of presenting the influence of the changes in returns and worker characteristics is to ask how the relative attractiveness of the two sectors has changed over time. In figure (9) we present earnings differences between self- and wage-employment which we predict given individuals' characteristics and our estimated parameters. Consistent with our findings, the distribution of earnings differentials based on more recent estimated returns statistically dominates the earlier distribution, i.e. self-employment has become more attractive.

We now ask whether our micro-econometric results can be used to interpret macro trends. Levels of self-employment have risen strongly over the last decades, and we have found increasing returns to factors of production in self-employment over the 2000s in this

study. Whilst our estimations are not based on a model of sector choice, we have noted that expected differences in earnings are one likely cause of transitions across sectors. In order to determine how much of the increase in self-employment might be explained by the change in returns we have documented, we now posit that a fraction of workers choose their sector as a function of earnings differentials. We estimate a reduced-form relationship between the size of predicted earnings differentials and the observed frequency of transiting into (and out of) self-employment. We estimate this relationship on the 2004-2006 panel. We can then estimate by how much the transition probabilities change if we replace the earnings differentials based on the estimates from the first half of the panel with the higher earnings differentials based on returns from the later period. As expected, the predicted inflow probability into self-employment increases, while the outflow probability falls. More specifically, the higher (second-period) returns predict an inflow probability of 16% from wage- to self-employment and an outflow probability of 9.2% every year²⁸. With these transition rates, we would predict self-employment to have grown from 59.9% in 2009 to 61.6% in 2011. We can thus compare this predicted growth to the actual growth that occurred in the share of self-employment: from 59.9% to 62.4%. This means that over two thirds of the increase in self-employment over this period can be explained by changes in returns alone²⁹.

Our results point to increasing returns to different factors of production in self-employment over the period analysed. May this be a result of increasing scarcity of these factors³⁰? To answer this question, we consider the sectoral distribution of *observed and unobserved* factors of production. For self-employment we find *increasing* levels of factor inputs for all observable factors of production, as noted in table (2). Mean capital holdings among the self-employed increase from 242.97 USD in 2004-06 to 563.81 USD in 2009-11

²⁸At lower (first-period) returns to self-employment these predictions are 13.1% and 9.4% respectively

²⁹In these calculations we only vary returns to factors, not the factor composition. We could also take into account the changed composition of the workforce. We would then predict self-employment to reach 69% in 2011 (calculations available on request), overpredicting the actual rate. However, since we argue that the changed composition of factors is itself a result of the returns, the adjustment implies utility costs, thus changing the nature of the relationship between earnings differentials pre- and post adjustment. In sum, only taking into account price changes underpredicts the true influence of changed earnings differentials, fully including behavioural reactions would however lead to overprediction, as in the case of Paasche vs. Laspeyres indices.

³⁰For example, the standard assumption of diminishing returns to input factors suggests that at lower levels of education or unobservable skills in a certain sector, marginal returns increase.

and fall among the wage-employed (from 410.56 to 322.51 USD). Trends are less marked for schooling, with education levels amongst the self-employed rising from 7.54 to 8.00 years and falling slightly in the wage sector from 10.23 to 9.82 years. In order to compare the levels of *unobserved ability* in the two sectors across time, we use expression (10) for θ_i derived above, averaging over different years³¹. Graphs (7) and (8) show that despite a large variance of skills (especially in self-employment) average unobserved productivity is higher in the wage sector than in self-employment. This pattern does not change markedly in the later period, in line with the fact that, although significantly increasing, ψ remains smaller than unity throughout.

In summary, we find higher levels of factors of production where returns have increased. This is consistent with the idea that changes in returns are driving changes in the composition of the workforce (and associated input factors). There is no evidence that the changes in returns are driven by diminishing marginal returns.

<< **Figure 7 here** >>

<< **Figure 8 here** >>

Decomposition techniques provide a systematic way of comparing the influence of differences in returns and factor composition on earnings across different sectors or groups. To this end, we first use common returns across sectors but let input factor levels vary, then fix input factors and apply the relevant sector-specific returns. Table (12) in appendix (G) explains the results of this procedure in more detail. In particular, we find a very strong effect of the one factor of production that is rewarded more highly in wage-employment: unobserved individual skills (since $\psi < 1$), which moderates the earnings differentials.

We have discussed the nature, size and potential effects arising from changes in returns to wage- and self-employment, but what has caused them? We can only provide suggestive evidence. We believe that credit constraints exist in Ghana. Lacking direct information we take as evidence a strong association between assets and transitions into self-employment.

³¹These calculations are based on few observations, thus our estimates suffer from small-sample bias. However, it is hoped that the comparisons between wage- and self-employment sectors are not affected too much.

We also find that transitions into self-employment increase with age - and that the increase in self-employment over our sample period was mainly driven by older individuals³². Credit constraints may slowly be weakening as a result of policy support (e.g. microcredit) and economic growth. If that is the case, we would expect factors of productions to flow more easily to the sectors where they are most productive, increasing overall efficiency and decreasing gross (unexplained) sector premia.

Finally, mean earnings differentials are not the only determinant of the decision to become self-employed: the associated risk may matter also (see Falco (2014)). However, given that the variance in self-employed earnings has increased, this would not appear to explain the increase in urban self-employment. Non-monetary aspects have also been found to favour self-employment (e.g. independence, flexibility), see e.g. Benz and Frey (2008). In our sample, reported levels of job-satisfaction in table (2) do fall among salaried workers, arguably also an indication of the increased relative attractiveness of self-employment³³.

7 Conclusions

Informal self-employment is the most common type of employment throughout the developing world and the share of self-employed workers has been on the rise in recent decades. This article tracks changes in the structure of earnings between 2004 and 2011 with a particular focus on differences between wage- and self-employment. Our results may help determine whether the rise in self-employment is the result of improved opportunities for successful entrepreneurship (*pull factors*), or the reflection of limited opportunities in wage-employment (*push factors*).

To study earnings across wage and self-employment, we construct a two-sector correlated random coefficient model that allows for rich patterns of selection on observable and unobservable worker characteristics, differential returns to unobservable factors and endogenous capital accumulation. The model is estimated on two periods of a long panel dataset from urban Ghana using Instrumental Variable Generalized Method of Moments (IV-GMM).

³²All results available on request.

³³Respondents replied on a 5-item Likert scale to the following question “All things considered, how satisfied are you with your current work?”.

Contrasting our results with ordinary least squares and fixed effects estimators demonstrates the importance of allowing for differential returns to unobservable characteristics. 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 in self-employment increase substantially. Returns to schooling also increase, but we find conflicting evidence on relative returns in self-employment relative to wage-employment. Returns to individual unobservable ability in self-employment rise but remain lower than in wage-employment throughout the period. Overall, our results indicate that the incentives for capital-rich, educated and otherwise able individuals to sort into self-employment have increased.

Second, we find some evidence that the incentives generated by higher returns have indeed started to attract more productive resources into self-employment. 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 increasing *pull* effect. Despite the trend, estimated levels of unobserved individual ability are still higher (and more highly rewarded) in wage-employment.

Third, once we control for both observable and unobservable worker characteristics, we find that unexplained sector earnings differentials have decreased over time. We interpret this as evidence of increased sector mobility and lower barriers to self-employment. This interpretation is consistent with a period of strong economic growth, when capital accumulation and improved credit markets may have helped relaxing credit constraints.

We conclude that while the sector remains highly heterogeneous, our evidence points against the grim view of self-employment as an occupation of last resort.

References

- ANGRIST, J. AND W. NEWEY (1991): “Over-Identification Tests in Earnings Functions with Fixed Effects,” *Journal of Business & Economic Statistics*, 9, 317–323.
- 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 (2008): “The Value of doing what you like: Evidence from the self-employed in 23 countries,” *Journal of Economic Behaviour and Organization*, 68, 445–455.
- BLANCHFLOWER, D. AND A. OSWALD (1998): “What Makes an Entrepreneur?” *Journal of Labor Economics*, 16, 26–60.
- CHAMBERLAIN, G. (1982): “Multivariate Regression Models for Panel Data,” *Journal of Econometrics*, 18, 5–46.
- CHAWANOTE, C. AND C. BARRETT (2014): “Farm and Non-farm Occupational and Earnings Dynamics in Rural Thailand,” *mimeo*.
- CINGANO, F. AND A. ROSOLIA (2012): “People I Know: Job Search and Social Networks,” *Journal of Labor Economics*, 30, 291 – 332.
- EVANS, D. AND B. JOVANOVIĆ (1989): “An estimated model of entrepreneurial choice under liquidity constraints,” *The Journal of Political Economy*, 97, 808–827.
- EVDOKIMOV, K. (2010): “Nonparametric Identification of a Nonlinear Panel Model with Application to Duration Analysis with Multiple Spells,” *mimeo*.
- FALCO, P. (2014): “Does risk matter for occupational choices? Experimental evidence from an African labour market,” *Labour Economics*, 28, 96–109.

- FASIH, T., G. KINGDON, H. A. PATRINOS, C. SAKELLARIOU, AND M. SODERBOM (2012): “Heterogeneous Returns to Education in the Labor Market,” *Policy Research Working Paper*, 6170.
- FIELDS, G. (1990): “Labour market modelling and the urban informal sector: Theory and evidence,” in *The informal sector revisited*, ed. by D. Turnham, B. Salomé, and A. Schwarz, OECD Paris.
- 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.
- HEATH, R. (2011): “Why Do Firms Hire Using Referrals?: Evidence from Bangladeshi Garment Factories,” Ph.D. thesis, Yale University.
- 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.
- IMBERT, C. (2012): “Decomposing wage inequality: public and private sectors in Vietnam 1993-2006,” *World Bank Economic Review*, 27, 55–79.
- (2013): “Decomposing the Labor Market Earnings Inequality: The Public and Private Sectors in Vietnam, 1993-2006,” *World Bank Economic Review*, 27, 55–79.
- JUTTING, J., J. R. D. LAIGLESIA, ET AL. (2009): *Is informal normal?: towards more and better jobs in developing countries*, OECD.

- 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*.
- NGUIMKEU, P. (2014): “A structural econometric analysis of the informal sector heterogeneity,” *Journal of Development Economics*, 107, 175–191.
- NOPO, H., N. DAZA, AND J. RAMOS (2011): “Gender Earnings gaps in the World,” *World Development Report 2012*.
- (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.
- NYSHADHAM, A. (2014): “Learning about Comparative Advantage in Entrepreneurship: Evidence from Thailand,” *mimeo*.
- 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

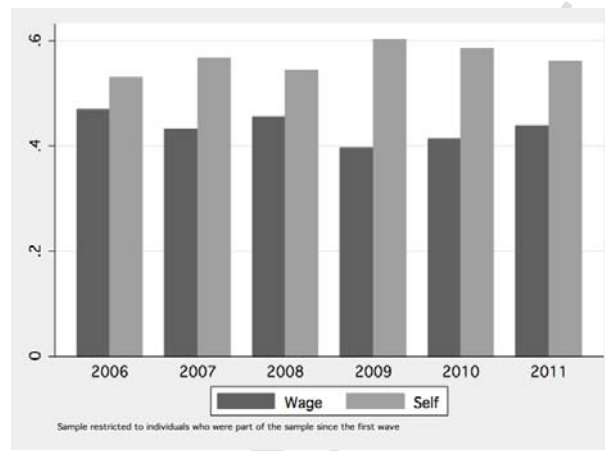


Figure 2: Age of Wage- and Self-Employed Workers

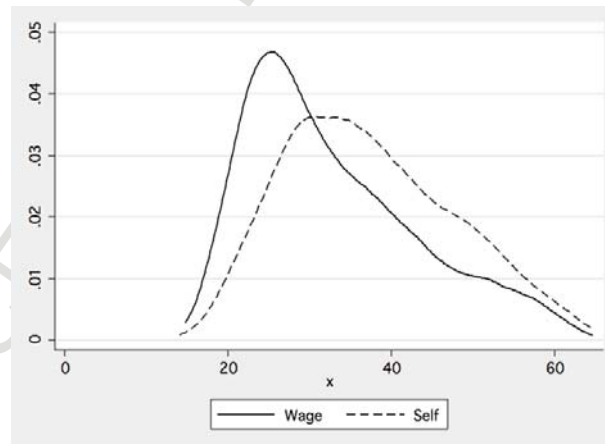


Figure 3: Real Monthly Median Earnings by Sector (in US dollars)

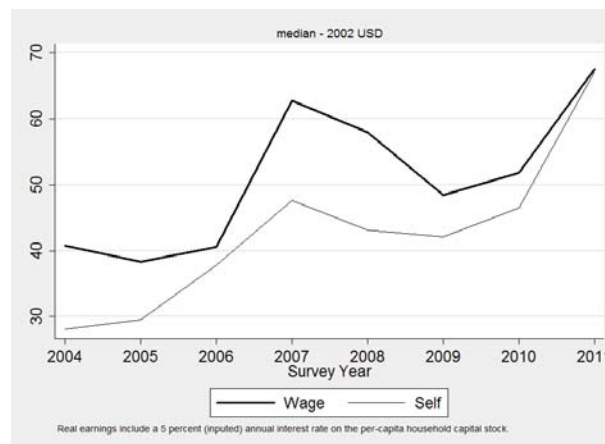


Figure 4: Earnings Distributions by Sector

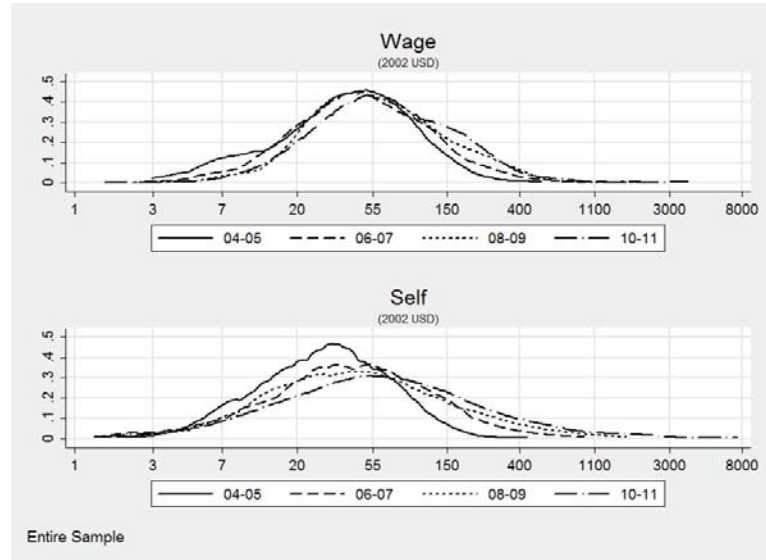


Figure 5: Probability of sector transition by panel length

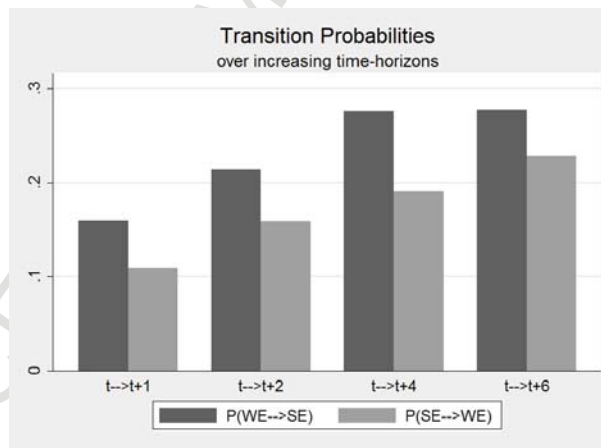


Figure 6: Distribution of Implied Annualised Interest Rates using $\hat{\beta}^{SE}$

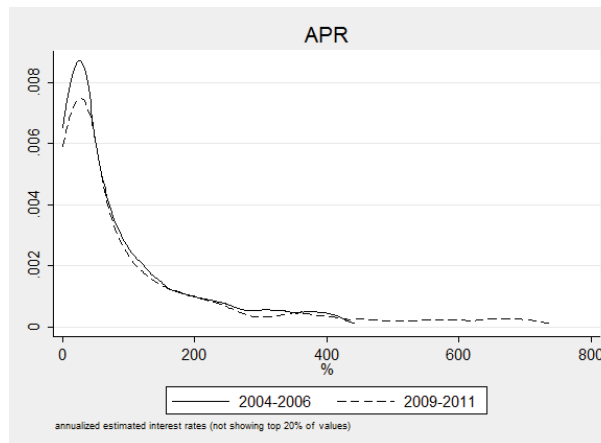


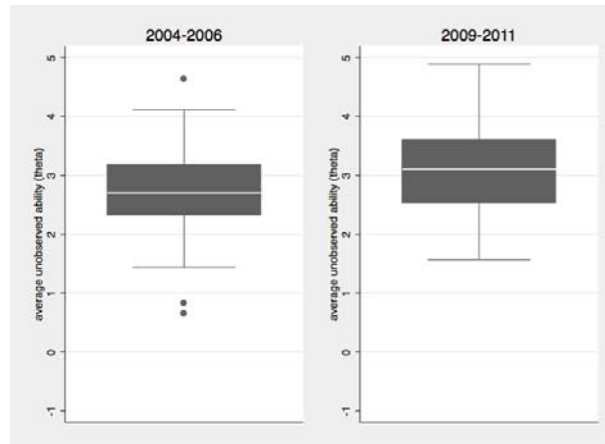
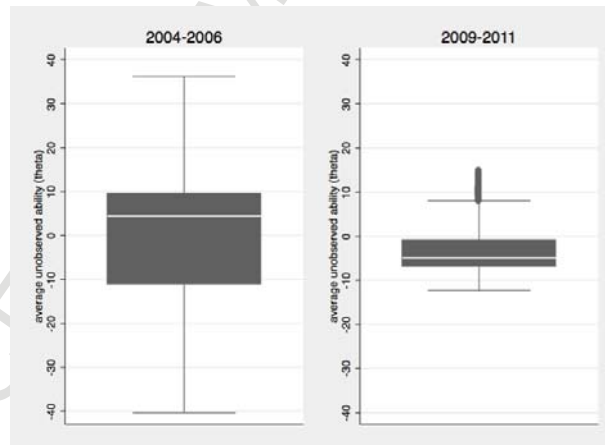
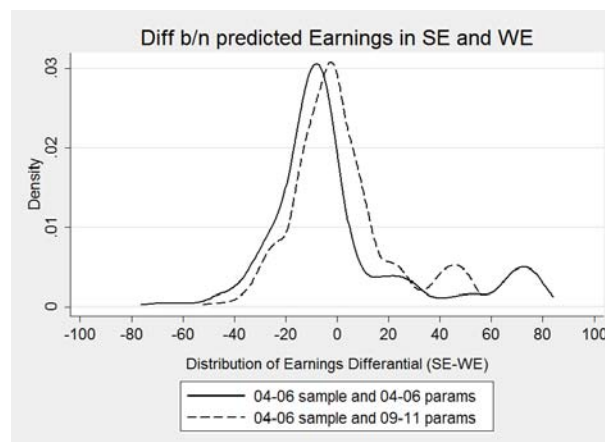
Figure 7: Distribution of $\hat{\theta}$ (Wage)Figure 8: Distribution of $\hat{\theta}$ (Self)

Figure 9: Evolution of cross-sector earnings differentials for given input factors



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: Descriptive Statistics by Sector 2004-2006 vs. 2009-2011

Wage Employment	Mean		Std. Dev.	
	04-06	09-11	04-06	09-11
Male	0.76	0.63	0.43	0.48
Age	35.66	36.36	9.97	10.33
Educ	10.23	9.82	2.71	4.07
Real Capital (2002 USD)	410.56	322.51	1381.82	744.54
Real Monthly Earnings (2002 USD)	65.59	95.33	53.54	91.76
Log Real Monthly Earnings (2002 USD)	3.66	3.96	0.79	0.86
Civil Service / Public Sector	0.28	0.28	0.45	0.45
Job satisfaction	3.45	3.26	0.96	1.12
Self-Employment	Mean		Std. Dev.	
	04-06	09-11	04-06	09-11
Male	0.33	0.25	0.47	0.44
Age	36.20	39.44	9.59	9.41
Educ	7.54	8.00	4.03	3.85
Real Capital (2002 USD)	242.97	563.81	612.89	1836.12
Real Monthly Earnings (2002 USD)	46.28	130.48	43.77	372.37
Log Real Monthly Earnings (2002 USD)	3.22	3.72	0.91	1.27
Job satisfaction	3.36	3.37	0.89	1.12

Only individuals used in the estimation included in sample. N=311 in period 2004-2006, N=357 in period 2009-2011, total observations N*T=2004.

Table 3: Transitions between Wage- and Self-Employment

	Wage Emp _t	Self-Emp _t	Total
Wage Emp _{t-1}	943 (80.46)	229 (19.54)	1,172 (100)
Self-Emp _{t-1}	217 (13.66)	1,371 (86.34)	1,588 (100)
Total	1,160 (42.03)	1,600 (57.97)	2,760 (100)

Only individuals in the estimation samples included; Consecutive period transitions pooled across all waves (including waves not used for estimation); Percentages reported in parentheses

Table 4: Estimation Results

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

	IV-GMM 2004-2006	IV-GMM 2009-2011
ψ	0.4444 (0.3781-0.5205)	0.8027 (0.7107 - 0.8734)
β^{SE}	0.2371 (0.1657-0.2869)	0.4124 (0.3454 - 0.4660)
α_{Educ}^{SE}	-1.8718 (-1.3737 - -2.2808)	0.405 (0.2593 - 0.5783)
$\delta^{SE} - \delta^w$	18.1939 (18.5523 - 18.1939)	2.9314 (1.5835 - 3.6463)
Number of Observed Individuals (N)	311	357

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

A Additional figures and tables

Figure 10: Transitions between wage- & self-employment, 2004-2006 vs. 2009-2011

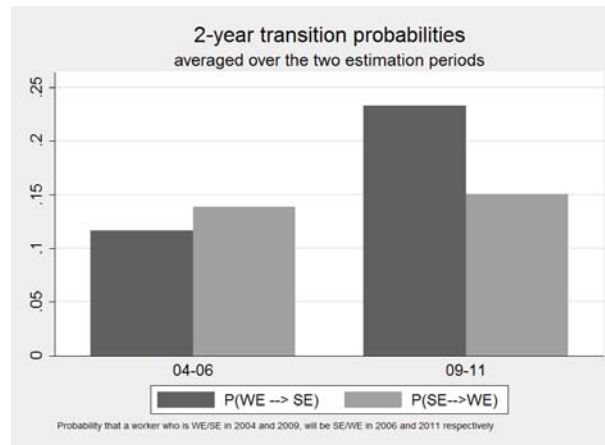


Figure 11: Transitions between wage- and self-employment, 2004-2006

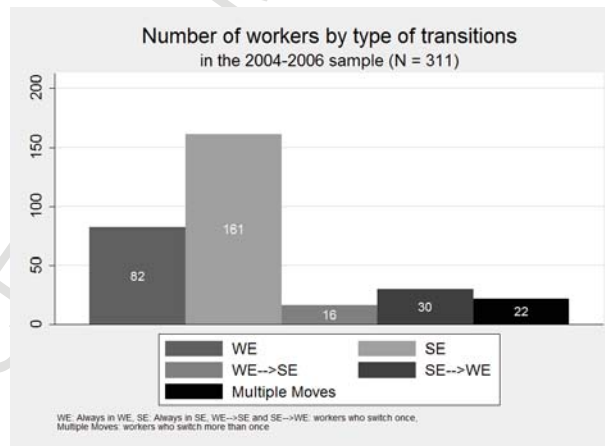


Figure 12: Transitions between wage- and self-employment, 2009-2011

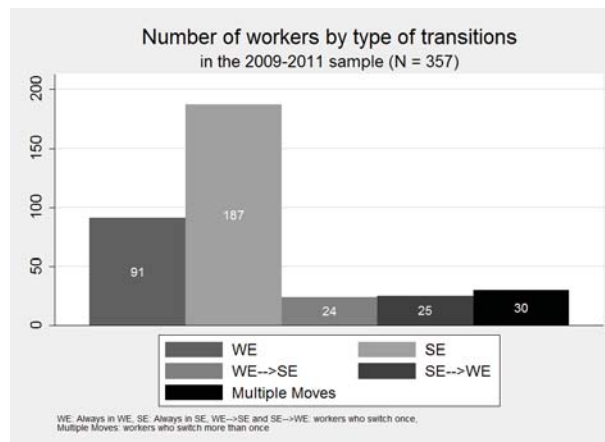
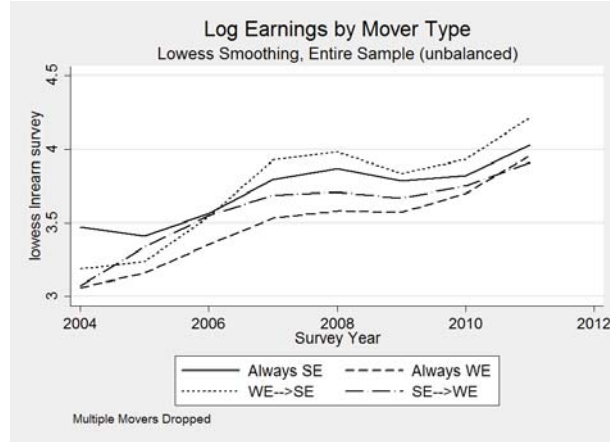


Figure 13: Mean earnings by mover type



B Construction of the Instrument Matrix

To illustrate how we construct 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: $erm_{i,t} \equiv (\varepsilon_{i,t} + \tau_i) - C_{i,t} (\varepsilon_{i,t-1} + \tau_i)$.

In our estimations we use three instruments - 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. $E(k_{it}erm_{is}) = 0 \quad \forall \quad s \geq t$), capital levels

for 2008 are an obvious instrument, since they 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 can in fact use $k_{i,09}$ and $k_{i,10}$ as additional valid instruments in years 10 and 11 respectively, since they are uncorrelated with $erm_{i,10}$ and $erm_{i,11}$, and 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 $\mathbf{Z}' \mathbf{erm}$ 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 , $E[\mathbf{Z}' \mathbf{erm}] = 0$ 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 occupational status, D , which we also assume to be pre-determined, since we want to allow capital to influence occupational choice in the future. 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, $E[\mathbf{Z}' \mathbf{erm}] = 0$ gives now rise to **12 valid moment conditions**.

C Weighting Moments - 2-Step Feasible GMM

In our estimations we use three instruments - past capital stocks, sector choice and an interaction between capital levels and sector choice. Using the quasi-difference over three time periods then gives six moments which are used to identify five parameters $(\alpha_{educ}^w, \alpha_{male}^w, \beta^{SE}, \psi, \bar{\delta})$. In this section, we explain how our estimation strategy weights the moments derived in this 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 moments in a first-step estimation, to weight the moments in the second step and obtain a more efficient estimator. GMM estimators minimize the following objective function:

$$\beta_{GMM} = \underset{\beta}{argmin} N[(\frac{1}{N}\mathbf{Z}'\mathbf{erm})'\mathbf{A}(\frac{1}{N}\mathbf{Z}'\mathbf{erm})] \quad (21)$$

where \mathbf{A} is a quadratic weighting matrix. For efficiency, \mathbf{A} must weight moments in inverse proportion to their variances and covariances (i.e. de-emphasize the high-variance instruments and those which are highly correlated with others and do not convey much additional information). Hence,

$$A_{EGMM} = Var[\mathbf{z}'erm]^{-1} \quad (22)$$

A feasible estimator of A_{EGMM} is:

$$Var[\mathbf{z}'erm] = plim_{N \rightarrow \infty} \frac{1}{N}\mathbf{Z}'\Omega\mathbf{Z} \quad (23)$$

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

$$\beta_{GMM} = \underset{\beta}{argmin} N[(\frac{1}{N}\mathbf{Z}'\mathbf{erm})'(\frac{1}{N}\mathbf{Z}'\Omega\mathbf{Z})^{-1}(\frac{1}{N}\mathbf{Z}'\mathbf{erm})] \quad (24)$$

In the simplest case, where the errors are assumed to be *homoskedastic*, $\Omega = \sigma^2 \mathbf{I}$, and

$$A_{EGMM} = \frac{\sigma^2}{N} \mathbf{Z}' \mathbf{Z} \quad (25)$$

This gives us the **One-Step Estimator** (estimated as one of our robustness checks in table (9)).

$$\beta_{GMM} = \underset{\beta}{\operatorname{argmin}} N \left[\left(\frac{1}{N} \mathbf{Z}' \mathbf{erm} \right)' \left(\frac{\sigma^2}{N} \mathbf{Z}' \mathbf{Z} \right)^{-1} \left(\frac{1}{N} \mathbf{Z}' \mathbf{erm} \right) \right] \quad (26)$$

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):

$$\hat{\Omega}_i = \mathbf{erm}_i \mathbf{erm}_i' = \begin{pmatrix} \mathbf{erm}_{i,09}^2 & \mathbf{erm}_{i,09} \mathbf{erm}_{i,10} & \mathbf{erm}_{i,09} \mathbf{erm}_{i,11} \\ \mathbf{erm}_{i,10} \mathbf{erm}_{i,09} & \mathbf{erm}_{i,10}^2 & \mathbf{erm}_{i,10} \mathbf{erm}_{i,11} \\ \mathbf{erm}_{i,11} \mathbf{erm}_{i,09} & \mathbf{erm}_{i,11} \mathbf{erm}_{i,10} & \mathbf{erm}_{i,11}^2 \end{pmatrix}$$

where \mathbf{erm}_i is the vector of quasi-differenced residuals for individual i .

In line with the most 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 \mathbf{erm} 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 $A = \left(\frac{1}{N} \mathbf{Z}' \mathbf{H} \mathbf{Z} \right)^{-1}$, where \mathbf{H} is an estimate of Ω based on homoskedasticity.

D Attrition

Despite best efforts to track individuals, attrition from the initial sample is large, as figure (15) shows. However, using two three-year panels gives us more acceptable levels of attrition, as figure (14) indicates.

Figure 14: Attrition from our two estimation samples, 2004-2006 vs. 2009-2011

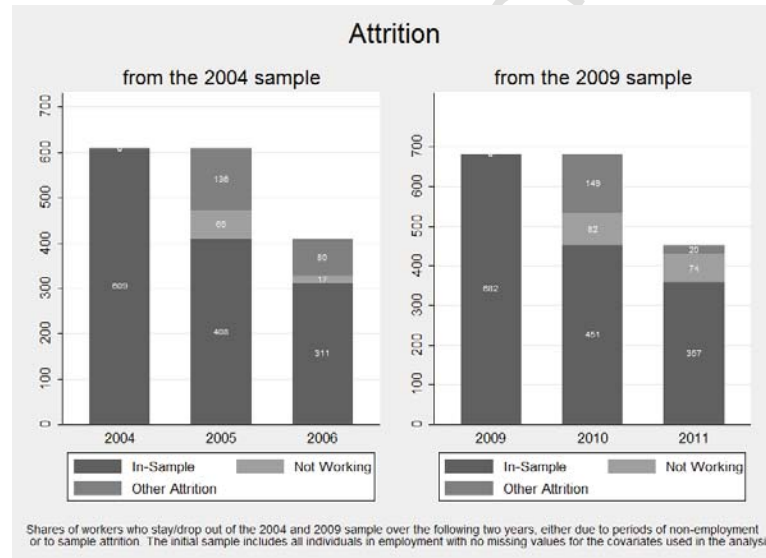
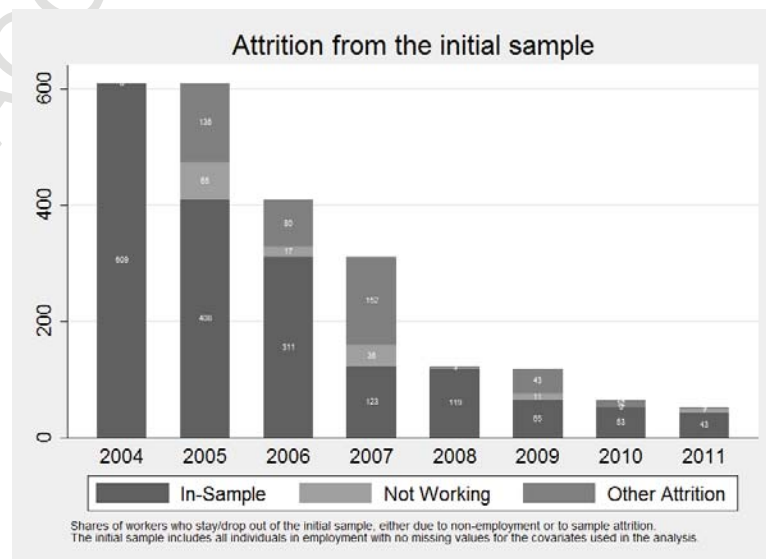


Figure 15: Attrition from the initial sample



Given that we focus on the evolution of earnings, we have two causes of attrition in our estimation sample. Some individuals' responses are missing as a result of classical

attrition (individuals were not interviewed) or because individuals are not employed. In the latter case we cannot use the observations to study earnings dynamics, naturally. We may be concerned if a large number of individuals exit employment into unemployment, suggesting that an exclusive focus on the dynamics between self- and wage-employment may be misplaced. However, in our data under 5% of workers transit from employment to unemployment³⁴. In comparison to high-income countries with more established social security systems, unemployment dynamics are thus less important here³⁵.

We now compare the observable characteristics of the estimation sample with the attrition sample. In the first estimation sample, table (5) shows that individuals exiting the panel are significantly more educated - however the difference is only just significant at the 5% level. By contrast, in the second sample we find significantly more men attrite - we have no explanation for this large difference, which may make us cautious about interpreting gender differences in earnings, for example. By contrast, all other characteristics are not significantly different across the two groups at the 5% level.

Table 5: Characteristics of attrition sample

	2004-2006				2009-2011			
	Sample	Attrition	Diff.	p-val	Sample	Attrition	Diff.	p-val
Male	0.45	0.444	0.006	0.888	0.353	0.503	-0.15	0.001
Age	33.975	34.452	-0.477	0.557	36.803	35.182	1.621	0.091
Education	8.639	9.282	-0.644	0.045	8.593	8.893	-0.301	0.406
Log Earnings	3.2	3.159	0.041	0.581	3.577	3.725	-0.149	0.12
Log Capital	4.304	4.35	-0.046	0.751	4.428	4.227	0.201	0.115

E OLS & Fixed Effects Estimators

Table (6) shows benchmark results from estimating our model with OLS (columns 1-2) and a standard Fixed Effects estimator (columns 3-4). This corresponds to the case where

³⁴Defined as having actively sought work for at least four weeks.

³⁵The unemployment rate in Ghana is 1-3% amongst 25-64 year olds, and relatively high only for 15-24 year olds (at 8.4%), according to the Ghana Living Standards Survey 6, 2012.

Table 6: OLS and Fixed Effects estimators

	OLS		FE	
	2004-2006	2009-2011	2004-2006	2009-2011
$\delta^w - \delta^{SE}$	-2.475*** (.045)	-3.100*** (.057)	-2.214*** (.093)	-3.636*** (.109)*
β^{SE}	0.072*** (.022)	0.117*** (.030)	0.019 (.033)	0.096** (.043)
$\alpha_{male}^{SE} - \alpha_{male}^w$	-0.171 (.073)	-0.1190*** (.098)	-0.04 (.253)	-0.247 (.276)
$\alpha_{educ}^{SE} - \alpha_{educ}^w$	-0.150*** (.008)	-0.142*** (.011)	-0.184*** (.039)	-0.123 (.037)
Obs.	933	1071	933	1071
R ²	.078	.05	.019	.011

Table Footnotes: $\alpha_{educ}^w = 0.08$, $\beta^w = 0.02$; using the same sample as in IV-GMM estimation.

$\psi = 1^{36}$. We can therefore test the implications of this restriction, which is common in the existing literature.

As in the IV-GMM results, the OLS and FE results also show a significant decrease of the self-employment premium ($\delta^{SE} - \delta^w$) from the first to 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 suggests that returns to schooling in self-employment have increased over time relative to wage-employment, consistently with the IV-GMM results.

Estimated returns to capital increase over time in OLS and FE as well. In the earlier period, however, returns to capital are not significant in the FE model. It appears that allowing for $\psi \neq 1$ in the IV-GMM model was crucial to uncover growth in β^{SE} .

³⁶The model is actually not a nested version of our IV-GMM model, since we are estimating the within-transformed model rather than a first-differenced model.

F 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 ranges. We also consider the results of imposing strict exogeneity (in line with classical panel data analyses) and show results without optimal weighting of the different instruments considered here.

F.1 Returns to Capital in the Wage Sector

We have so far assumed 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

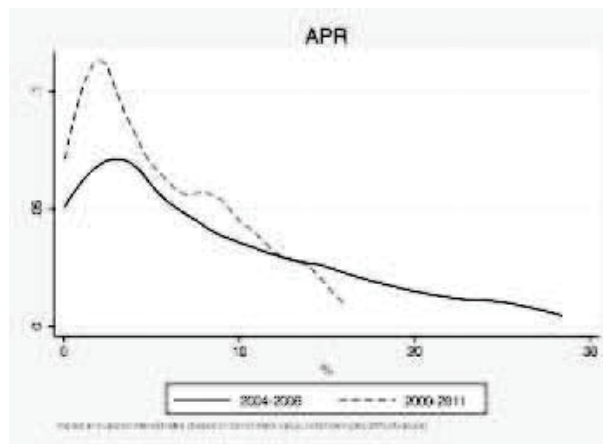
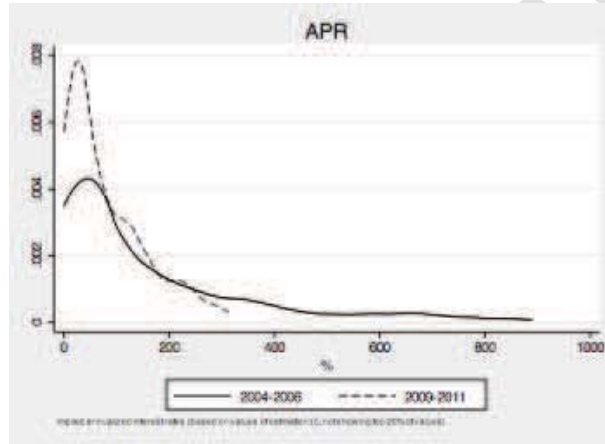
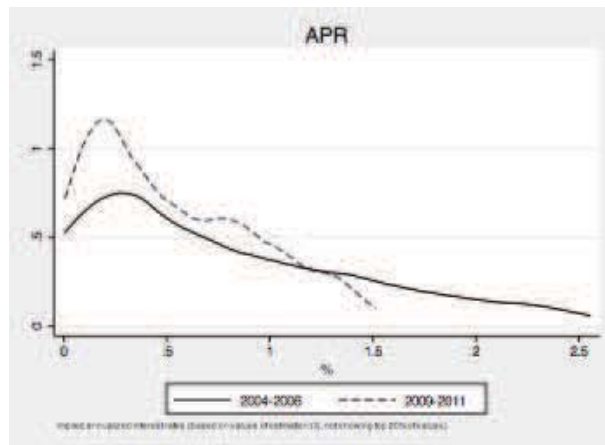


Figure 17: Distribution of implied interest rates: high level of β^w Figure 18: Distribution of implied interest rates: low level of β^w 

F.2 Female Wage Gap

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. $\alpha_{male}^w = 0.25$. However, as the sources upon which we base this assumption use different estimation strategies (and obtain different results) we increase and decrease the value of α_{male}^w from 15% to 35% in estimations (III) and (IV) reported in tables (9) and (10), a plausible range for the gender wage gap. The results show that our main findings are little affected by these changes.

F.3 Returns to Education in Wage Employment

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.

F.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 over 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.

F.5 Results under the Assumption of 'Exogenous Capital'

In order to allow for the possibility that earnings 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 of the other instruments and thus use information from all periods to instrument for past earnings. The results are presented in table (11) in specification (VIII).

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 are somewhat different, but 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	0.4444 (0.3781-0.5205)	0.8027 (0.7107 - 0.8734)	0.4067	0.7064
β^{SE}	0.2371 (0.1657-0.2869)	0.4124 (0.3454 - 0.4660)	0.2110	0.3793
α_{Educ}^{SE}	-1.8718 (-1.3737 - -2.2808)	0.405 (0.2593 - 0.5783)	-1.5327	0.5024
α_{Educ}^w	0.08 (set)	0.08 (set)	0.08 (set)	0.08 (set)
α_{Male}^{SE}	-3.6213 (-4.3495 - -2.4846)	-12.7072 (-10.7328 - -13.4324)	-2.3140	-8.9479
$\delta^{SE} - \delta^w$	18.1939 (18.5523 - 18.1939)	2.9314 (1.5835 - 3.6463)	15.0930	0.9111
Optimal Weighting Ω	yes	yes	yes	yes
Instruments	full	full	full	full
Sample N (movers)	311 (41)	357 (73)	311 (41)	357 (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	0.4476	0.8126	0.4424	0.8069
β^{SE}	0.2401	0.4154	0.2414	0.4150
β^w	0.002 (set)	0.002 (set)	0.02 (set)	0.02 (set)
α_{Educ}^{SE}	-1.9074	0.3927	-1.8892	0.3997
α_{Educ}^w	0.08 (set)	0.08 (set)	0.08 (set)	0.08 (set)
α_{Male}^{SE}	-3.7538	-13.1208	-3.7009	-12.9598
α_{Male}^w	0.25 (set)	0.25 (set)	0.15 (set)	0.15 (set)
$\delta^{SE} - \delta^w$	18.5163	3.1723	18.3487	3.0627
Optimal weighting Ω	yes	yes	yes	yes
Instruments	full	full	full	full
Sample N (movers)	311 (41)	357 (73)	311 (41)	357 (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	0.4464	0.7987	0.4182	0.8677
β^{SE}	0.2327	0.4097	0.2946	0.4547
β^w	0.02 (set)	0.02 (set)	0.02 (set)	0.02 (set)
α_{Educ}^{SE}	-1.8542	0.4108	-2.1051	0.2398
α_{Educ}^w	0.08 (set)	0.08 (set)	0.008 (set)	0.008 (set)
α_{Mate}^{SE}	-3.5403	-12.4595	-4.0258	-15.5678
α_{Mate}^w	0.35 (set)	0.35 (set)	0.25 (set)	0.25 (set)
$\delta^{SE} - \delta^w$	18.0358	2.8049	19.9850	5.1983
Optimal weighting Ω	yes	yes	yes	yes
Instruments	full	full	full	full
Sample N (movers)	311 (41)	357 (73)	311 (41)	357 (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	0.4814	0.6683	0.3317	1.0008
β^{SE}	0.1525	0.3385	0.2140	0.3153
β^w	0.02 (set)	0.02 (set)	0.02 (set)	0.02 (set)
α_{Educ}^{SE}	-1.3923	0.5897	-0.4563	0.5551
α_{Educ}^w	0.20 (set)	0.20 (set)	0.08 (set)	0.08 (set)
α_{Male}^{SE}	-3.1353	-8.2539	-3.9894	-15.3679
α_{Male}^w	0.25 (set)	0.25 (set)	0.25 (set)	0.25 (set)
$\delta^{SE} - \delta^w$	14.3533	0.0638	7.5826	3.1881
Optimal weighting Ω	yes	yes	no	no
Instruments	full	full	full	full
Sample N (movers)	311 (41)	357 (73)	311 (41)	357 (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
ψ	1.5068	1.3266
β^{SE}	0.0790	0.3170
β^w	0.02 (set)	0.02 (set)
α_{Educ}^{SE}	-0.9488	0.8527
α_{Educ}^w	0.08 (set)	0.08 (set)
α_{Male}^{SE}	-23.2016	-20.8091
α_{Male}^w	0.25 (set)	0.25 (set)
$\delta^{SE} - \delta^w$	19.8595	2.6844
Optimal weighting Ω	yes	yes
Instruments	full + strict exogeneity	full + strict exogeneity
Sample N (movers)	311 (41)	357 (73)

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

G Earnings Gap Decomposition

In analyzing differences in earnings across wage- and self-employment we here seek to differentiate the impact of differences in returns to factors from compositional effects. The upper panel of table (12) fixes input factors (at the mean value amongst the self-employed) and assesses the influence of sector-specific returns. For example, what is the earnings differential that would occur if all individuals had the same level of capital given differential returns? For capital, we find a more than doubling of the earnings advantage of the self-employed between the first and second period. The complementary lower panel of table (12) fixes returns (at the self-employed level) and shows the effect of changing the input mix on mean earnings across sectors. The self-employed have higher capital stocks and this differential increases over time.

Despite the 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 earnings gap in favour of wage-employment remains (self-employed earnings are 41.75% lower in the earlier and 20.52% 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 (and its fall over time is potentially driven by weakening credit constraints)³⁷.

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, despite substantial growth in the productive assets of the self-employed (especially capital), total factor endowments still favour wage-jobs (the “selection/composition effect”), with a dominant contribution from unobservable individual characteristics.

³⁷The 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 (Gosling and Lemieux (2004)). However, we are not aware of a change in public sector pay policies (Imbert (2013) analyzes a public sector pay reform in Vietnam). The proportion of public sector workers is constant at 28% amongst the wage employed workers.

Table 12: Earnings gap decomposition

Effect of Differential Returns to Self-employment	2004-2006	2009-2011
Negative Wage premium $\delta^{SE} - \delta^w$	18.1939	2.9314
Effect of Observable Skills	-15.3109	0.864
Effect of Capital $\ln(\overline{K_{SE}}) (\beta^{SE} - \beta^w)$	0.8514	1.9163
Effect of Gender $\overline{h_{SE}} (\alpha^{SE} - \alpha^w)$	-1.2228	-3.5976
Effect of Education $\overline{h_{SE}} (\alpha^{SE} - \alpha^w)$	-14.9396	2.5453
Effect of Returns to Unobservable Skills $(\psi - 1) \overline{\theta_{SE}}$	0.6658	0.211
Total Price Effect	3.5488	4.0064
Effect of Differential Input Factors		
Difference in Observed Factors	-0.2324	-0.2004
Difference in Capital $(\ln(\overline{K_{SE}}) - \ln(\overline{K_w})) \beta^w$	0.0039	0.0098
Difference in Gender $(\overline{h_{SE}} - \overline{h^w}) \alpha^w$	-0.0767	-0.08
Difference in Education $(\overline{h_{SE}} - \overline{h^w}) \alpha^w$	-0.1596	-0.1302
Difference in Unobserved Factors $\overline{\theta_{SE}} - \overline{\theta_w}$	-3.7339	-4.0112
Total Selection Effect	-3.9663	-4.2116