D.Phil. Thesis

Occupational choices and their outcomes in African labour markets

PAOLO FALCO

Keble College, University of Oxford

Thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics at the University of Oxford, Michaelmas Term 2011.

Occupational choices and their outcomes in African labour markets

PAOLO FALCO, UNIVERSITY OF OXFORD

Abstract

This thesis is an investigation into the microeconomic mechanisms that govern some of the occupational choices faced by workers in Sub-Saharan Africa, and into the monetary and non-monetary returns to their decisions.

Chapter 1 begins by exploring the decision process that leads workers to allocate themselves to different occupations within the economy. In particular, I investigate the role of risk-aversion in the allocation of workers between formal and informal jobs in Ghana, hence attempting to explain a fundamental dimension of duality through an investigation into workers' preferences. In my model of sectoral allocation risk-averse workers can opt between entering the free-entry informal sector and queuing for formal occupations. Conditional on identifying the riskier option, the model yields testable implications on the relationship between risk-aversion and workers' allocation. My testing strategy proceeds in two steps. First, using the first three waves of the Ghana Household Urban Panel Survey (GHUPS) dataset, I estimate expected income uncertainty and find it considerably higher in the informal sector than in formal employment. Second, using experimental data to elicit risk-attitudes I estimate the effect of risk-aversion on occupational choices and I find that, in line with the first result, more risk-averse workers are more likely to queue for formal jobs and less likely to be in the informal sector. The conclusion of the first chapter is that attitudes to risk should feature more prominently in models of sector allocation and in the design of labour market policies, in particular when those policies aim to impact workers' vulnerability to risk and uncertainty.

Chapter 2 focuses on the largest occupational category in the Developing world, self-employed workers with small productive activities, and it tries to estimate the returns to different productive assets, namely physical capital, labour and human capital. These are the workers that form most of the informal sector analysed in chapter 1, which allows me to draw a direct link with the analysis so far. The chapter begins by specifying a model for the income-generating process grounded in the literature on firms' production and hence abridging the gap between the analysis of individual earnings and the study of firms' value added. Identification in the empirics is achieved by means of panel estimators that are suitable to address the endogeneity of input choices, which derives from both time-varying and time-invariant unobservable heterogeneity. The use of these estimators is made feasible by the length of the Ghanaian Household Urban Panel Survey dataset at CSAE. I also explore issues of endogeneity in the selection of different technologies, defined by their relative capital and labour intensity. Finally, I analyse the shape of returns to capital, with the aim to detect potential non-convexities in technology. The results show that capital and work-experience play the strongest role in income-generation, while the shares of value added attributed to labour and to formal schooling are low. Marginal returns to investment are high at low capital levels and they decrease very rapidly, pointing against the existence of non-convexities due to minimum scale requirements, but implying that real income gains resulting form micro-investment are modest.

Chapter 3 returns to the issue of earnings uncertainty and risk-aversion explored in Chapter 1, but it now takes the allocation choice as given and explores the direct welfare implications of income uncertainty for worker's well-being. Namely, the chapter explores the relationship between income and welfare, with a particular attention on the link between income vulnerability and happiness. Using unique longitudinal data on life-satisfaction and labour market outcomes, I estimate an individual measure of vulnerability (defined as the probability of falling below a low-income threshold) and investigate its effect on well-being. After controlling for unobservable individual fixed effects, work-satisfaction, relative income and other relevant worker characteristics, I find a sizable impact of vulnerability, over and above the income effect. When I explore the mechanisms behind my results, I find that aspiration adaptation to current income may result in a transitory income effect. Moreover, using my direct measure of attitudes to risk from field-experiments (already used in chapter 1), I can test directly the hypothesis that more risk-averse agents suffer more heavily from a given increase in income vulnerability. Overall, my findings support policy interventions that aim to reduce vulnerability, as I expect such policies to have a 'direct' impact on agents' happiness given the prevailing attitudes to risk and uncertainty in the population. Finally, from the point of view of overall social welfare, my results suggest that non-Rawlsian growth models, whereby 'someone may be left behind', may fail to enhance general welfare, for high enough levels of risk-aversion in the population, if the risk of falling behind is sufficiently widespread. *[Word count: approx. 56,500]*

Table of Contents

Chapter 3: Income Vulnerability and Happiness

Acknowledgements

It is difficult to thank all the people who have helped me achieve the results presented in this thesis and attempting to name them will not do justice to those I am likely to forget. A special acknowledgment goes to my Supervisor, Dr. Francis Teal, who has offered me his unconditional support during my entire time at Oxford and to all the researchers, the students and the staff at the Centre for the Study of African Economies, who have made my doctoral experience an unforgettable time of personal growth and professional enrichment. My gratitude goes in particular to Prof. Abigail Barr, Moses Awoonor-Williams, Prof. Stefan Dercon, Dr. Kim Lehrer, Babak Somekh, Dr. Mauro Caselli, Christian Westerlind Wigstrom, Andrew Kerr and Simon Quinn, who have lent me their help and contributed to my work with their stimulating ideas, and to Stefano A. Caria who has worked with me on the last chapter of this thesis. I would also like to thank the researchers at the World Bank, especially Dr. Bob Rijkers and Dr. Pierella Paci, and at the Paris School of Economics, in particular Prof. Francois Bourguignon, Dr. Milo Bianchi and Luke Haywood, who welcomed me in their research communities. My gratitude goes also to Prof. Glenn Harrison, who repeatedly dedicated his attention to my work, to Prof. Gary Fields and Prof. Steve Bond, for their precious comments, and to my College Supervisor Dr. Howard Smith for his continuous support.

I am grateful to my family, Andrea, Teresa and Camillo, to Marie and to my friends, who have put up with my moments of despair and have rejoiced in my achievements.

This thesis is dedicated to the people of Ghana, who have made me feel at home during long periods of field-work, in a sincere hope that the results of my research will be a small contribution to the progress of a beautiful Country.

Preface

Labour markets in developing countries have truly fascinating connotations. On the one hand, since labour is often the only productive asset available to workers, the decisions that determine its allocation and its returns in different sectors of the economy are crucially important to understand and foster the development process. On the other hand, a wealth of constraints to the smooth functioning of the market, such as limited access to information, poorly developed financial systems, limited rule of law and issues of political economy often give rise to outcomes that can only be rationalised through careful study of the microeconomic interactions that characterise workers' behaviour. For instance, how can we explain the paradoxical coexistence of the often very high returns to micro-investment with the elusive growth of most micro-enterprises? Or the existence of highly dual labour markets, with a restricted elite of workers who enjoy high protection while large fractions of the population live close to perennial unemployment? Moreover, it is my firm opinion that the attention of a careful researcher cannot overlook cultural and behavioural aspects of workers' decision-making that may emerge in specific contexts and may diverge substantially from general models of behaviour, advocating the use of carefully tailored empirical strategies to capture the more nuanced aspects of behaviour. Finally, from a methodological point of view, our understanding of workers' behaviour is complicated further by the informal nature of most economic contracts and transactions, which makes them difficult to document.

Despite all the complexity and methodological challenges, however, these labour markets are populated by economic agents whose decisions are amenable to rigorous economic analysis. This thesis is an investigation into the microeconomic mechanisms that govern some of the occupational choices workers make in their life and into the consequences of such choices in Sub-Saharan Africa. It is the result of several years of data collection conducted in Ghana by the author, in collaboration with the Centre for the Study of African economy, which have coupled standard survey data with behavioural evidence from field experiments designed to capture aspects of behaviour that have traditionally been difficult to investigate empirically.

In chapter 1 the thesis begins by exploring the decision process that leads workers to allocate themselves between different types of occupations. In particular, I investigate the role of risk-aversion in the allocation between *formal and informal jobs*, with the aim to explain this fundamental duality through an investigation of workers' preferences. The chapter introduces a model of sector choice that yields testable implications on the relationship between risk-aversion and workers' allocation. The main result is that risk-averse workers are more likely to queue for formal jobs and less likely to be in the riskier informal sector. The conclusion I draw from the analysis, which extends well beyond the context of developing countries, is that attitudes to risk should feature more prominently in models of sector allocation and in the design of labour market policies, in particular when such policies aim to impact workers' vulnerability to risk and uncertainty. Moreover, this chapter argues that social protection policies (e.g. unemployment benefits) play an instrumental role in assisting workers in their career choice, by affecting the riskiness of different career paths. As such, they should not be viewed as sheer transfers payments, but as a powerful tool for development.

Chapter 2 focuses on the largest occupational category in the developing world, *self-employed workers with small productive activities*, and it tries to estimate their returns to different productive assets, namely physical capital, labour and human capital. As these workers constitute the large majority of the informal sector analysed in chapter 1, this analysis delves directly into the outcomes of the allocation process described above. The chapter begins by specifying a model for the income-generating process, grounded in the literature on firms' production technology and hence abridging the gap between the analysis of individual earnings and the study of firms' value added. Identification in the empirics is achieved by means of panel estimators that are suitable to address the endogeneity of input choices. The results show that capital and work experience play the strongest role in income-generation, while the shares of value added attributed to labour and to formal schooling are low. Marginal returns to investment are high at low capital levels, but they decrease very rapidly, pointing against the existence of non-convexities due to minimum scale requirements and implying that real income gains resulting form micro-investment are modest. Most importantly, I conclude that informal self-employment continues to pose an important puzzle. How can we reconcile the high rates of returns on micro-investment I estimate with the very low growth rates of the micro-enterprises in the sample? The results in chapter 1 point to the possibility that high margins on micro-investment constitute a compensating differential for higher risk in self-employment. Poor saving devices and myopic preferences are competing explanations that we will explore in future research.

Finally, chapter 3 returns to the issue of earnings uncertainty and risk-aversion explored in chapter 1, but it now takes the allocation choice as given and explores the direct welfare implications of income uncertainty for workers' well-being. Namely, I explore the relationship between income and welfare, with a particular attention on the link between *income vulnerability and happiness*. Using unique longitudinal data on life-satisfaction, I find a sizable impact of vulnerability, over and above the income effect, indicating, as in chapter 1, that earning security has a direct impact on workers' well-being, which may offset the positive effect of higher income on average. Rather interestingly, I also find evidence that income aspirations adapt quickly, resulting in income effects that are only transitory. In other words, earning more also 'raises the bar' for the income level that is henceforth necessary to achieve the same level of happiness. This result may explain why societies that have grown considerably richer over the decades have failed to grow significantly happier. Like for the previous chapters, these findings extend beyond the context of developing countries and they constitute, to the best of my knowledge, the first attempt to carefully quantify the impact of income insecurity on subjective well-being. They also carry important policy implications. In line with chapter 1, the results of this analysis support policy interventions that aim to reduce vulnerability, as we can expect such policies to have a 'direct' impact on agents' happiness given the dominant preferences for risk and uncertainty in the population. Moreover, from the point of view of overall social welfare, non-Rawlsian growth models, whereby 'someone is left behind', may fail to enhance general welfare, for high enough levels of risk-aversion in the population, if the risk of falling behind is sufficiently widespread.

Each chapter of this thesis is self-contained and, after introducing the specific topic, will take the reader through a detailed explanation of the methodology, the data sources and the results. An overall summary will conclude, outlining some of the most important policy implications and leads for future research that have emerged from this challenging and captivating work.

Chapter 1:

Risk-aversion and occupational choices: evidence from matched field experiments and survey data in urban Ghana

1 Introduction

Modern Development Economics has been increasingly interested in the workings of labour markets. As explained by Fields (2005), “the poor are poor because they earn little from the work they do” and it is now widely recognised that fighting poverty entails improving labour-market opportunities for the most disadvantaged. A crucial step in this process will be to gain a clearer understanding of how workers allocate themselves between different occupations and of the constraints they face in their occupational choices. In particular, as the divide between the formal and the informal sector of the economy deepens, identifying what drives workers into such different segments will enhance our understanding of the development process and will help us design more effective labour market policies. Banerjee and Newman (1993) lend strong theoretical support to these claims by showing how economic development can be modelled as a dynamic process of institutional change that ultimately depends on workers’ occupational choices, given an initial wealth distribution.

Once we have a clear motivation for studying workers allocation, we are faced with the question of how to model occupational choices in labour markets without state funded insurance mechanisms to support the unemployed. My approach in this chapter will be to model the choices of individual workers between alternative job-options as a function of their *attitudes to risk*. In doing so, I will attempt to abridge a gap in the existing literature, which, in my opinion, has not dedicated enough attention to the effects that risk in occupational outcomes may have on the choices of risk-averse workers. The main innovation from this study will be the use of experimental data to elicit risk-aversion. During the summer of 2007 the author carried out three weeks of behavioural experiments over a sample of Ghanaian workers who had previously taken part in a three-year household survey by the Centre for the Study of African Economies (CSAE). The experiment consisted of a series of dichotomous decisions over 21 pairs of lotteries with real money prizes. Applying Maximum Likelihood techniques for the estimation of utility functions to this data, I have obtained a measure of individual risk-aversion that I can use to research the relationship between attitudes to risk and occupational choices.

To guide the investigation, I formulate a theoretical model of sectoral choice with *risk-averse* workers, motivated by the empirical observation that occupational decisions are "risky" in two important dimensions. First, the process of job-search is inherently uncertain, since workers who seek employment are generally only successful with some limited probability and face the risk of remaining unemployed if their search fails. Second, conditional on finding a job, earnings volatility over “good” and “bad” states of the world can be substantial in some occupations. These two forms of uncertainty should both influence the expected utility that risk-averse workers receive from prospective jobs and they should ultimately affect workers' allocation. Based on this observation, I propose an extension to the classical Harris and Todaro (HT, 1970) model of intersectoral linkages and to its further development by Fields (1975). In my model a worker can either seek employment in the formal sector or in the informal sector. If she chooses the former option/strategy, she will obtain a job with probability p, while she will remain unemployed with probability $(1-p)$. If, on the other hand, she chooses the informal sector, she will be employed with certainty (since the informal sector is presumed to be free entry), but her earnings will be subject to volatility over good and bad periods. These assumptions reflect the observed choice set available to workers in many developing countries, where earnings from informal jobs are easily accessible but highly volatile, while access to formal occupations, which pay more regular income streams, is rationed. In this setting, workers' occupational decisions are driven by their risk-aversion and by the relative magnitude of earnings uncertainty under different search strategies. Hence, conditional on being able to determine which strategy is riskier, the model yields testable implications on the relationship between workers' attitudes to risk and their sectoral allocation.

My empirical strategy will test those implications in the context of the Ghanaian labour market. First, I will estimate the marginal effect of risk-aversion on sectoral allocation using a multinomial logit model. The main finding will be that risk-aversion significantly increases the likelihood of queuing for formal occupations, while it decreases the likelihood of working in the informal sector. These effects are large and statistically significant. Second, using panel data on earnings, I will measure income uncertainty and I will conclude that in urban Ghana taking an informal job is on average a riskier option than queuing for a formal one in terms of expected income uncertainty. These two findings are consistent with each other and in line with the predictions of the theoretical model. Moreover, they survive an extensive set of robustness checks against a number of potential limitations in the empirical setup, including issues of workers' misclassification into formal and informal employment categories, of imprecision in the estimation of risk-aversion and of the potential endogeneity of risk-preferences.

My conclusion is that risk-preferences, like education and age, play an important role in determining occupational choices and, therefore, they should be taken more carefully into account when designing labour market policies for developing countries, especially if such policies are going to affect the degree of income uncertainty (and ultimately of risk) in different sectors of the economy. Employment schemes that produce highly volatile income streams, for instance, may fail to enhance welfare and they may result in low take-up if workers are sufficiently risk-averse, potentially causing a waste of development resources. Conversely, my results highlight the value of social protection and insurance, as instruments that can effectively assist workers in their occupational choices, enhancing their future employment prospects. As such, these tools should not be viewed as ‘mere’ transfer payments, but as powerful development tools. Given the novelty of the analytical framework, the evidence presented in this paper extends beyond the context of developing countries, and certainly beyond the African context. It is entirely consistent, for instance, with the results of the work by Bonin et al. (2007), who find that, among German workers, those with low willingness to take risks are more likely to work in occupations with low earnings risk.

The chapter is structured as follows. In Section 2 I will introduce my model of occupational choices with risk-averse workers and income uncertainty, and I will present its implications for empirical testing. In Section 3 I will describe the survey data and the experimental dataset collected in the summer of 2007. Section 4 will outline the Maximum Likelihood Estimation method used to estimate workers’ risk-aversion. In Section 5 I will introduce my empirical strategy, present my results and perform a number of robustness checks on them. Section 6 will conclude.

2 A model of occupational choices with risk-averse workers and earnings uncertainty

The model I propose builds upon the classical Harris-Todaro (HT, 1970) model of the dual economy (and more specifically on its later development by Fields, 1975), extending its basic intuition in two interrelated dimensions. First, my theoretical framework will not be solely concerned with differences in *expected earnings* across sectors, but it will also take into account differences in *earnings* *volatility* and more generally in *earnings uncertainty* as determinants of occupational preferences. Second, I will relax the assumption of workers' risk neutrality to allow risk-aversion to play a role in the occupational decisions.

Inspired by the Kenyan experience of the 1960s, Harris and Todaro (1970) set forth a compelling theory to explain rural-urban migration. Their model postulates that workers compare the *expected incomes* in the urban sector with agricultural wage rates and migrate if the former exceeds the latter. Urban sector wages are assumed to be set institutionally above market clearing and urban unemployment results. On the other hand, it is assumed that there is always full-employment in agriculture. Rural-urban migration acts as an equilibrating force, until the expected urban wage eventually equates the rural wage. As noted by Fields (1975), migration could be more generally regarded as an adjustment mechanism by which workers allocate themselves between different segments of the labour market, while attempting to maximise their expected incomes. In this light, the original HT model can be used to analyse the movement of workers across the formal and the informal sector of the economy (Fields, 1975). In the formal sector, generally defined as the sum of public sector workers and employees or owners of large private businesses (a precise definition will be provided in the next section), wages above market clearing cause unemployment (possibly due to institutional constraints, such as union bargaining and minimum wage laws, as well as efficiency wage setting). On the other hand, there is always full-employment in the freely accessible informal sector, commonly defined as the sum of salaried workers and entrepreneurs in small/micro enterprises. To choose their preferred job-options, workers compare the *expected wage* in the formal sector with the wage in the informal sector. An equilibrium is reached once the flows of workers equalise the expected wages in the two sectors.

One major shortcoming of the HT model is the assumption of risk-neutral workers, who are presumed to compare the expected income from different occupational prospects *without taking into account earnings volatility/uncertainty in the different occupations*. This is in stark contrast with most economic theories, where workers are generally assumed to display some degree of risk-aversion and to care about the degree of uncertainty associated with their occupational choices. It is even more at odds with the reality of Developing Countries, where it is not uncommon to observe high earnings volatility, which tends to be higher in certain sectors compared to others. For instance, a salaried worker in a large firm in the formal economy is generally less likely to experience a negative (or positive) shock to his income than a petty trader in the market, whose earnings are generally characterised by heavy idiosyncratic variations and are not protected by a formal contractual agreement. Therefore, it seems rather unrealistic to assume (like in the standard HT model) that an individual can base her occupational decisions on expected earnings only, without taking into account the degree of volatility of those earnings. More plausibly, we should model workers as choosing among different *income lotteries*, corresponding to the different occupational sectors, some of which are characterised by higher risk than others. To be more precise, I will refer to earnings distributions, rather than income lotteries. The nature of these distributions, as defined by their first two moments, determines the riskiness of an occupation. In what follows I will formalise this intuition in a model of occupational choices that will deliver clear-cut testable implications.

As in Fields (1975) and in Rankin, Sandefur and Teal (2008), my model extends HT (1970) to allow for the co-existence of an informal and a formal sector in urban labour markets. The former is free-entry, in the sense that whoever wants an informal job can get one. This is quite plausible, especially in urban contexts, where it only requires some minimal investment to produce some good or service that can be sold in a casual marketplace (Fields, 2005) and where low institutional barriers as well as a weak regulatory environment are unlikely to place significant constraints on the possibility of starting small-scale informal economic activities. The low values of fixed capital reported by the majority of the self-employed in our sample are supportive of this view. The left-hand side of Figure 1 in the empirical appendix shows that a large proportion of the self-employed (about 36%) reported having no fixed capital in 2006 (defined as the value of tools, equipment and structures).[1] Moreover, those who had some fixed capital reported very low values of such capital. The right-hand panel in Figure 1 shows that 60% of those workers reported a replacement value of their capital that was equal to only a month of current profit or less, and up to 80% of those workers could replace their entire capital stock with 3 months worth of profits

[1] They are predominantly traders who do not process their merchandise and do not own a fixed location. Their main cost is the variable cost of merchandise, which presumably constitutes a low barrier to entry as it is unlikely to be characterised by indivisibilities and minimum scale requirements.

or less).[2] On the other hand, the formal sector comprises the most desirable, stable and well-paid occupations that are only accessible to a minority of the labour force. Access to those occupations tends to be regulated by formal contracts (e.g. civil servants and employees in large private enterprises) and by a more 'structured' recruitment process (e.g. where workers are asked to provide a formal proof of their qualifications), and it is therefore likely to be a lengthier process, which motivates the idea of a 'queue'. Institutional wage setting above market clearing (especially in the public sector), in the tradition of dual economy models à la HT, tends to corroborate the idea of rationing and queuing. Although the assumption of a two-sector economy might seem like an over-simplification of reality, it enables us to capture the main insights from the Dual-Economy approach, while keeping the mechanics simple.[3]

The model operates over two adjacent periods, which I will call t and $t+1$. In period t, worker i has a choice between searching/queuing for formal employment (Q) and searching for an informal sector job (or, equivalently, setting up her own informal business), (I). The two strategies are mutually exclusive, an assumption that carries some obvious limitations but seems to be find support in my dataset from Ghana. Figure 2 in the empirical appendix shows that 84 percent of all the formal wage-employed in the GHUPS sample

---

[2] This should not mislead the reader to underestimate the importance of capital market failures in an economy like Ghana. What I am arguing is that the nature of the informal sector is such that a "petty" business can easily be created without facing large setup costs. Clearly, credit market failures might themselves be among the reasons for the existence of such a casual and uncertain informal sector.

[3] This point is forcefully explained in Basu (1997): *"It is unlikely that any of the initiators of the dual economy model would deny that the labour market may in reality be fragmented into more than two sectors. The assumption of duality is merely for analytical convenience. If fragmentation – irrespective of the number of parts – in itself causes some problems and we wish to examine these, then the simplest assumption to make is that of market dualism."*

were 'unemployed' before finding their current occupation, while only 16 percent came to their current job directly from self-employment[4] At time $t+1$, if the worker has chosen to go in the informal sector, she will now have a job with certainty (reflecting the free-entry nature of informal jobs), but she will face a bad or a good period with probability θ_i^G and θ_i^B respectively[5]. In a good period she will earn y_{iI}^G, while in a bad year she will get y_{iI}^B (where $y_{iI}^G > y_{iI}^B$). On the other hand, if she has chosen to search for a formal job, she will have gotten one with probability p_i, in which case she will earn a sure wage y_{iF} in period $t+1$. Yet, with probability $(1-p_i)$, her search will have been unsuccessful; she will now be unemployed and she will fall back on some minimal income level from external sources, $\underline{y_i}$ (e.g. financial support from her family).[6] The most important feature of the model is the introduction of *uncertainty in the outcomes of both strategies*. In the informal sector, the source

[4] An intuitive explanation behind this evidence is at least three-fold. First, the available channels of job-search are generally very different between the formal and the informal labour market. Hence, workers who seek informal employment are unlikely to come across formal job-opportunities during their job-search (and vice-versa). Second, if seeking employment in the informal sector carries some degree of negative stigma (because informal workers are perceived to be inferior in terms of skills and productivity), people who aspire to formal employment will keep clear of the informal labour market. Similarly, if aspirations play a role in sector allocation, workers who aspire to formal employment may refuse to take up informal employment in the meantime. Third, workers who choose to set up their own informal business will have little or no time (nor, probably, motivation) left to seek formal employment in the meantime.

[5]For simplicity, throughout the analysis and also in the empirical section I will assume that bad and good years are equally likely, i.e. $\theta_i^G = \theta_i^B = 1/2$

[6] It should be remarked that the job-prospects of a worker in different occupations are allowed to depend on the worker's personal characteristics (this is the reason for indexing all the above quantities and probabilities by i). For instance, a worker who is more educated or more experienced may have higher y_{iF} and a higher p_i, than a relatively unskilled one (because he is more productive and, perhaps, more effective at job-search). This is in line with the classical Mincerian approach to earnings functions; in the empirical analysis, I will control for these observables when I attempt to disentangle the effect of risk-aversion from the other determinants of a worker's job-choice.

of uncertainty is the volatility of earnings between good and bad states. In the formal sector, instead, the wage is constant, but the search for such jobs is only successful with some probability p_i. This is clearly an oversimplification of reality, but it aims to capture a dichotomy that is commonly observed in the labour market. In the formal sector, as it is generally described in the literature (see Fields 2005), working conditions and salaries are more stable than in the "murky" sectors of the economy, where the absence of formal contractual arrangements and the unsteady economic environment are among the causes of highly uncertain earnings streams. However, it is generally recognised that access to the best jobs is rationed and that the process of search is inherently uncertain. Moreover, the supporters of Segmented Labour Market theories, who have strongly advocated the presence of institutional wage-setting as a source of segmentation in the labour market, would support the view that different segments also differ in terms of earnings stability (Leontaridi, 1998).

The model is solved by backward induction, with the worker choosing in period t the occupation that yields the higher expected utility in period $t+1$. All workers have the same utility function, but they differ in their degree of risk-aversion. We assume the following CRRA utility specification.

$$U_i(y_{ij}) = \frac{{y_{ij}}^{1-\gamma_i}}{1-\gamma_i}$$

where

y_{ij} is the level of earnings accruing to individual i from occupation j

γ_i is the individual coefficient of relative risk aversion

(If $\gamma_i < 0, i$ is risk-prone/loving, if $\gamma_i = 0, i$ is risk-neutral, if $\gamma_i > 0, i$ is risk-averse).

$j = I$ "Informal"; F "Formal"

Utility only depends upon consumption and therefore ultimately upon earnings. As Fields (2005) states, this is a reasonable assumption in the context of developing countries where a large number of people value additional goods greatly compared to additional leisure and it would not be unreasonable even to replace utility maximisation with income-maximisation. Risk-attitudes are captured by the parameter γ_i, which can take both negative (risk-lovingness) and positive (risk-aversion) values. As already mentioned, while in the classic Harris and Todaro framework a worker compares the *expected income* from two alternative occupations, in this model she compares the *expected utilities* from the two search strategies. In fact, she will choose to seek formal employment if and only if:

$$E[U_i(Queue\ for\ formal\ job)] > E[U_i(Seek\ Informal)]$$

$$\Leftrightarrow$$

$$p_i \frac{{y_{iF}}^{1-\gamma_i}}{1-\gamma_i} + (1-p_i)\frac{{\underline{y}_i}^{1-\gamma_i}}{1-\gamma_i} > \theta_i^G \frac{{y_{iI}^G}^{1-\gamma_i}}{1-\gamma_i} + \left(1-\theta_i^G\right)\frac{{y_{iI}^B}^{1-\gamma_i}}{1-\gamma_i}$$

In order to determine how risk-aversion affects a worker's choice, it is necessary to determine how the expected utilities from the two strategies compare as γ_i increases and as the relative riskiness of the two job-options changes. To answer this question I have simulated expected utilities from the two strategies at increasing levels of γ_i and under different assumptions on the relative magnitude of the first two moments of the income distributions in the two sectors. From the results of these numerical exercises (contained in the

Empirical Appendix), I have distinguished the following four representative cases, with four corresponding analytical results, which will be tested in the Section 5. I will only describe the first two in detail, since case 3 and 4 are exactly symmetric and produce the exact opposite results. Nevertheless, all four of them, with the corresponding results, are summarised in Table 1.

<u>Case 1</u>

"Expected Earnings from seeking formal employment (Q) are lower than expected earnings from seeking informal employment (I), $\mathrm{E}[\mathrm{y}_{\mathrm{iQ}}] < E[\mathrm{y}_{\mathrm{iI}}]$; the variance of earnings from Q is lower than the variance of earnings from I, $\mathrm{Var}[\mathrm{y}_{\mathrm{iQ}}] < Var[\mathrm{y}_{\mathrm{iI}}]$". In this case, Q is the safer option, and a worker who is sufficiently risk-averse will be willing to accept a lower expected income from formal-job-search in exchange for a lower variance of earnings (i.e. lower uncertainty). This is the scenario represented in Simulation 2, 3 and 4.

Result (1): There exists a positive level of individual risk-aversion, above which a worker who faces the employment prospects described by Case 1 will prefer Q over I.

<u>Case 2</u>*:*

"Expected Earnings from Q are higher than expected earnings from I, $\mathrm{E}[\mathrm{y}_{\mathrm{iQ}}] > E[\mathrm{y}_{\mathrm{iI}}]$; the variance of earnings from Q is lower than the variance of earnings from I, $\mathrm{Var}[\mathrm{y}_{\mathrm{iQ}}] < Var[\mathrm{y}_{\mathrm{iI}}]$". In this case, Q pays higher expected earnings with lower variance. Therefore, a worker will prefer Q over I at any level of positive risk-aversion. However, there is a low enough negative level of risk-aversion (a high enough level of risk-lovingness) below which a worker will prefer the risky option (I), even though it pays less on average. This is the scenario depicted in simulation 5 and 6.

Result (2): There exists a negative level of individual risk-aversion above which a worker who faces the employment prospects described by Case 2 will prefer Q over I.

Taking Result (1) and (2) together, we reach the following *Conclusion* (and vice-versa for Result 3 and 4), which will be central in the empirical analysis:

"If $Var[y_{iQ}] < Var[y_{iI}]$ *, there will be a (positive or negative) level of risk-aversion above which a worker will choose Q over I. Hence, the higher is a worker's risk-aversion, the higher the likelihood that he will choose to search/queue for a formal job (Q) over seeking an informal sector job (I)"*

This is the proposition that I will attempt to test empirically in Chapter 6, conditional on being able to determine which one of the above four hypothetical cases best describes the options available to workers.[7]

[7] Further simulations have shown that these four results would also be valid if we substituted the binary distribution of informal sector earnings assumed in this stylised model (i.e. good/bad periods) with a continuous distribution. That would be a more realistic way to model earnings volatility in the informal economy (since reality is better approximated by a continuum of different states of the world) and it is the approach that I will implicitly follow in the empirical analysis in Section 5.

Table 1: Summary of the four possible Cases and Results from the model

	$\mathbf{E[y_{iQ}]} < \boldsymbol{E}[\mathbf{y_{iI}}]$	$\mathbf{E[y_{iQ}]} > \boldsymbol{E}[\mathbf{y_{iI}}]$
$\mathbf{Var[y_{iQ}]} < \boldsymbol{Var}[\mathbf{y_{iI}}]$	Case 1 – Result 1 *There exists a positive level of γ_i above which a worker prefers Q over I*	Case 2 – Result 2 *There exists a negative level of γ_i above which a worker prefers Q over I*
$\mathbf{Var[y_{iQ}]} > \boldsymbol{Var}[\mathbf{y_{iI}}]$	Case 4 – Result 4 *There exists a negative level of γ_i above which a worker prefers I over Q*	Case 3 – Result 3 *There exists a positive level of γ_i above which a worker prefers I over Q*

3 Data

This paper makes use of two *matched* sources of data, the Ghana Household Urban Panel Survey (alternatively referred to as GHUPS or UPS henceforth) and an experimental dataset containing the choices made by a sub-set of UPS respondents in a series of lottery games designed to elicit their attitudes towards risk. The UPS has been conducted since 2004 by the Centre for the Study of African Economies (CSAE) at Oxford University, in collaboration with the Ghana Statistical Office (GSO), and at the time of the risk-aversion experiment it spanned a period of three years (2004-2006). The sample was based on a stratified random sample of urban households from the 2000 census and it has been held to be representative of the Ghanaian urban population (RST, 2008). The peculiar feature of the dataset is that it contains comparable information, including earnings data, on both wage-employees and self-employed workers in the formal and the informal sector, spanning over a period of three consecutive years. The distinction between formal and informal jobs is based on international standards. The International Labour Organisation and the Economic Commission for Latin America and the Caribbean have defined the informal sector as the sum of non-professional self-employed, domestic workers, unpaid workers, and workers in small/micro enterprises.[8] I adopt the same definition in this paper and split those currently employed in my sample into informal sector workers, comprising the self-employed and the salaried workers in small/micro firms (max 10 employees) and formal sector workers, comprising employees of the public sectors and of large private firms (more than 10 employees in total). The resulting breakdown by survey wave is provided in the Table 1-3 in the Appendix. As acknowledged in Rankin, Sandefur and Teal (2008), surveying the earnings of the Self-Employed is a challenging practice, mainly because these workers tend to keep poor records of their activities. The UPS attempts to overcome this problem in two ways. First, respondents are explained concepts of “revenue”, “business costs” and “profits” to enhance the precision of their answers. Second, handheld computers are employed during the data collection to check on the field for possible inconsistencies across related figures. Throughout this study, although I will generally use the terms “earnings” and “wages”, I will actually be referring to *real monthly earnings*, using 1995 as the base year to calculate the inflation figures. To detect unemployment, the UPS uses the phrasing suggested by the

[8] ILO (2002). Women and Men in the Informal Economy: A Statistical Picture (Geneva: ILO).

International Labour Organisation. The standard ILO definition of an unemployed person in low-income countries is a person who did no work for pay in the preceding week, not even for one hour.[9]

The Experimental Dataset contains the decisions made by a sub-sample of survey respondents in a series lottery games designed by Dr Abigail Barr at the University of Oxford and conducted by the author, in collaboration with the GSO, during the summer of 2007.[10] A representative subset of 288 respondents from the UPS was invited to participate in a number of workshops, where they were presented with a series of 21 lottery games.[11] In each game (two examples of which are presented in the Empirical Appendix) they were asked to choose between two alternative lotteries. Each lottery took the shape of an opaque bag containing a number of coloured marbles. To each colour we attached a money prize and each respondent was asked to choose from which of the two bags he would prefer to pick a random marble. At the beginning of the task each participant was told that, once they had made their 21 choices, one would be picked at random and acted out, i.e., we would let them pick a marble from the bag they chose in that case and pay them the corresponding winnings. The winnings of the gamble ranged from 10,000 to 110,000 CEDIs, with average

[9] ILO (2003). Global Employment Trends (Geneva: ILO).

[10] A detailed description of the experimental setup is contained in Barr, A. (2007). "Attitudes to Risk in Ghana: Field Manual." Unpublished.

[11] To check whether this group was indeed representative of the whole UPS sample, I have compared the mean values of a number of workers' characteristics in the UPS and in the Experimental Dataset for every different survey-wave. The results are contained in the Empirical Appendix. The two sets of means are generally very close, with one noticeable exception: the ethnic composition of the two samples. This is explained by the fact that, given the time available, I was only able to conduct the behavioural experiments in two out of the four areas where the UPS is administered. Given the ethnic differences across different regions of Ghana, this should explain the difference with UPS. Overall, I conclude that the experimental sample is sufficiently representative of the two urban that I visited (Accra and Kumasi).

winnings calibrated just above 30,000 CEDIs. This sum was estimated to be larger than the average daily earnings of a worker in our sample[12]. It is therefore reasonable to argue that the prizes at stake were high enough to elicit participants' true risk-preferences. By applying a maximum likelihood routine developed by Glenn Harrison (2007) to this dataset, one can derive an estimate of each of the participant's individual Coefficients of Relative Risk Aversion.

4 Estimation of individual CRRA using Experimental Data

The experiment conducted in Ghana over the summer of 2007 was designed to provide estimated coefficients of CRRA for each of the participating individuals. By applying a statistical methodology designed by Glenn Harrison (2007), I can generate a distinct estimate of CRRA for each individual. This estimate is based on the assumption that all the participants in the experiment have a simple utility function of the form $U_i(x) = \frac{x^{1-\gamma_i}}{1-\gamma_i}$, where x is the lottery prize and $\gamma \neq 1$ is the CRRA parameter that we wish to estimate. So, $\gamma = 0$ corresponds to risk-neutrality, $\gamma < 0$ to risk loving and γ >0 to risk aversion[13]. This is the same utility function as the one

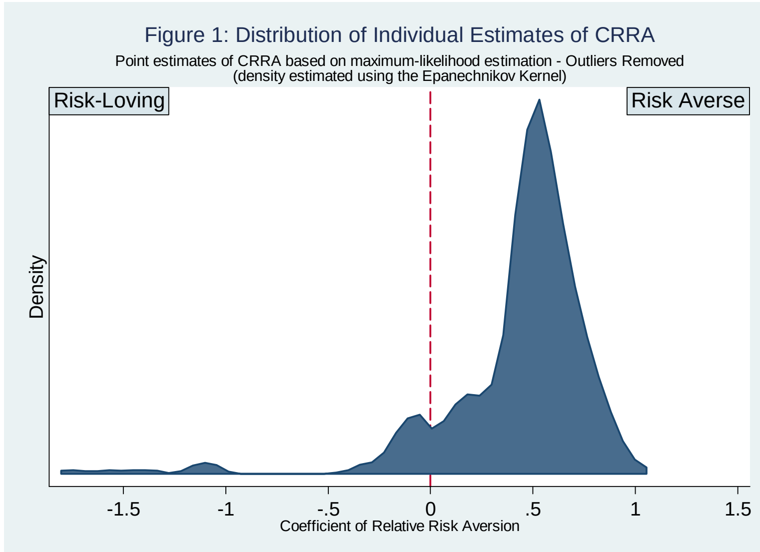
[12] To be precise, 30,000 CEDIs is higher than the average nominal daily earnings of workers in our sample in the last available wave of the UPS survey, 2006. Given the economic outlook of the regions we visited, it is reasonable to expect some (low) nominal earnings growth between 2006 and 2007 and therefore, it is possible that the average stakes of our games fell below average daily earnings, though not considerably. Even if that was the case, I still believe that the money prizes we offered were a high enough incentive for our participants to reveal their risk-preferences truthfully.

[13] If needed, I will assume $U_i(x) = \ln(x)$ when $\gamma_i = 1$.

which underlies the occupational choice model and for this reason the coefficients of risk aversion estimated from the lottery games can be directly transferred to our study of job-decisions.

The estimation method consists of a separate maximum likelihood estimation (using a customised likelihood functions by Harrison (2007)) for each individual, designed to identify the value of γ_i that best explains all of their 21 choices. The results of the estimation, carried out in collaboration with Dr Abigail Barr, are the following. The maximum likelihood estimator converged for 274 of the 288 subjects (90%). Of the 274 ML estimations undertaken, 259 returned estimates of r within the acceptable range and with acceptably small standard errors. So, after rejecting outliers, we obtain estimated γ_i for nearly 90% of the individuals who took part in the experiment. The mean estimated γ_i for these 259 participants is around 0.43, indicating moderate risk aversion. The distribution of these estimates is presented in Figure 1. The graph indicates that, according to the point estimates, only 10% of the individuals are risk-loving.[14]

[14] Implicit in my estimation strategy is the assumption that workers do not integrate their game endowments with their liquid wealth when they take part in the games. Recent evidence from a similar experiment by Andersen et al. (2011) has shown that only partial integration occurs. The authors report that their respondents “...behave as if some small fraction of personal wealth is combined with experimental prizes in a utility function”. Moreover, their evidence comes from a developed country (Denmark), where average endowments of liquid wealth are likely to be larger relative to game winnings than in our experimental setting in Ghana. Ideally, I would have liked to include liquid wealth in the utility model above and re-estimate the parameters of risk-aversion allowing for integration, but information on respondents’ wealth was not collected at the time of the experiment, and the previous waves of GHUPS do not contain a satisfactory proxy.



5 Empirical Results

The goal of this section will be to test the implications of my theory and to explore the empirical findings on the estimated relation between risk-aversion and occupational choices. First, using a Multinomial Logit with CRRA among the explanatory variables, I will model the likelihood of being in each of the following three occupational categories (which together comprise all the workers in my sample): informal sector, formal sector and unemployment[15]. To test the implications of the model, the latter category will proxy the pool of job-seekers who have remained unemployed as a consequence of choosing to search for a formal occupation (merits and limitations of this approximation will be discussed below). I will refer to this category as the "queue of unemployed

[15] Recall that according to our definition the informal sector includes the self-employed and the salaried workers in small/micro firms (max 10 employees in total). The formal sector includes public sector workers and those employed in large private firms (more than 10 employees in total).

formal-job-seekers", though at times, I will simply call it the unemployed pool. In the second part of the chapter I will discuss some fundamental estimation issues and perform a number of robustness checks on my results. Finally, I will gauge the degree of earnings uncertainty from different job-options, exploiting both the cross-sectional and the panel dimension of my dataset. I will then be able to interpret the empirical findings in the light of my occupational choice model.

Multinomial Logit of Occupational Choices

Occupation sorting will be modelled using a classic latent variable approach, similar to the one used by Rankin, Sandefur and Teal (2008). The main novelty, in this case, will be the introduction of CRRA among the right hand side variables in the specification. This approach is in line with my theoretical model whereby workers choose the sector with the highest expected utility. The latent variable ω^*_{ijt}, which can be considered as the propensity of worker i to sort into sector j and time t, is indeed a function of the same variables which are believed to drive expected earnings (and hence expected utility) from each sector. I model it as follows:

$$\omega^*_{ijt} = X_{it}\theta_j + \eta_{ijt}$$

where

j indexes different job categories: Informal (I), Formal (F) and Queue of Unemployed Job Seekers (Q)

$t = (2004, 2005, 2006)$, indexes the three waves of data.

X_{it} is a matrix of individual characteristics, including the estimated coefficients of relative risk-aversion

θ_j is the vector of multinomial logit coefficients

η_{ijt} is a sector-and-period-specific error term

The condition for choosing category k over the others is that:

$$\omega_{ikt}^* > \max_{j\neq k}\left(\omega_{ijt}^*\right)$$

If we define

$$\varepsilon_{ikt} = \max_{j\neq k}\left(\omega_{ijt}^* - \omega_{ikt}^*\right) = \max_{j\neq k}\left(X_{it}\theta_j + \eta_{ijt} - X_{it}\theta_k - \eta_{ikt}\right),$$

the condition for choosing sector k in period t becomes $\varepsilon_{ikt} < 0$. Following McFadden (1973), I assume that the η_{jt}'s are independent and identically Gumbel distributed (the well-known IIA assumption). It follows that their cumulative and density functions are respectively: $G\left(\eta_{ijt}\right) = exp(-e^{-\eta_{ijt}})$ and $g\left(\eta_{ijt}\right) = exp\left(-\eta_{ijt} - e^{-\eta_{ijt}}\right)$. From this model, it is possible to derive the following result, which constitutes the core of the Multinomial Logit:

$$P(Worker\ i\ choses\ sector\ k\ in\ period\ t) \equiv P(\varepsilon_{ikt} < 0|X_{it})$$

$$= \frac{exp\left(X_{it}\theta_j\right)}{\sum_j exp\left(X_{it}\theta_j\right)}$$

Using this expression, I can obtain consistent maximum likelihood estimates of θ_j, and from those estimates I can compute the marginal effects of each regressor on the probability of employment in any one sector. This method hinges upon the validity of the assumption of Independence of Irrelevant

Alternatives (IIA). This would be least likely to hold in the presence of close substitutability between different occupational options. Given the clear distinction which seems to exist between different sectors of the Ghanaian labour market, it can be deemed unlikely that the same worker could be indifferent between different employment strategies. In their paper on occupational choices in Ghana, Rankin, Sandefur and Teal (2008) set up a similar Multinomial Logit and they indeed accept the IIA assumption to be valid.

It should be remarked that in my model a worker only has a choice between joining the informal sector and searching for formal employment. The latter option might also be interpreted as joining a queue of “unemployed formal job-seekers” until a job opportunity becomes available. In other words, since access to the formal sector is rationed, workers cannot opt for a formal job directly (as it is the case in the informal sector), but they can only choose to search/queue for one, remaining unemployed in the meantime. It follows that the hypothesis we should test is whether risk-aversion has an impact on the likelihood that workers will sort into the two categories *that are actually open to them*: the informal sector and the queue of unemployed formal job seekers.

The full set of results from the estimation of the multinomial logit is reported in Table 4 in the Empirical Appendix. Below I report the marginal effects (with corresponding standard errors) of all the regressors on the likelihood of being in the job-queue (Table 2) and on the likelihood of being in the informal sector

(Table 3)[16]. As explained, these are the two categories between which workers can be presumed to actually have a choice, and therefore, only the marginal effect of CRRA on the likelihood of choosing these two sectors can possibly have a causal interpretation (this is why I do not compute the same marginal effects for the third category, the formal employees). My main finding is a positive and statistically significant marginal effect of risk-aversion on the likelihood of being in the queue of formal job seekers and a negative one on the probability of being in the informal sector. These results hold throughout seven different specifications and they are statistically significant at least at the 10% level. From the first specification (Column 1), it emerges that increasing relative risk-aversion by 1 increases the likelihood of queuing by 20.7 percentage points and reduces the likelihood of being in the informal sector by 21.8 percentage points when we do not control for other variables.

In column 2 to 4, I introduce a number of controls for workers' characteristics which constitute the most likely determinants of occupational choices other than risk-aversion. From the perspective of the theoretical model, these variables should be viewed as determinants of a worker's expected earnings (in the formal and the informal sector), expected earnings volatility (in the informal sector) and probability of finding a formal job (p_i)[17]. These are the elements

[16] This is more convenient than reporting the estimated coefficients, θ_j's, which in a multinomial logit model do not have a direct interpretation. The marginal effects are non-linear functions of all the regressors and, following common practice, I choose to evaluate them at the mean of X.

[17] It is well documented that Gender, Age and Years of Formal Education are important determinants of a worker's earnings and, as discussed above, they might also drive her earnings volatility if they affect her ability to deal with unexpected shocks. Moreover, Rankin, Sandefur and Teal (2008) show that these variables, and in particular formal education, are important

which, in the model, were shown to determine the expected utility from different occupational strategies and as such, they constitute the main drivers of occupational choices together with CRRA. It follows that controlling for the observed individual characteristics which are likely to determine such job prospects is necessary if we aim to disentangle the causal effect of increasing risk-aversion on the propensity of choosing each job-option. I also include dummy variables to control for period-specific effects (e.g. periods of recession causing higher rates of unemployment). With these controls in place I observe a reduction in the marginal effect of CRRA (in absolute terms). In particular, including linear (in 2) and non-linear (in 4) terms for age and education, and gender, reduces the marginal effect on the probability of queuing to 0.172 (in 2) and 0.148 (in 4) and it also reduces (in absolute terms) the marginal effect on the likelihood of informal employment to -0.177 (in 2) and -0.155 (in 4). These are still large effects and they remain statistically significant at the 10% and 5% level respectively. The year dummies, in column 3 and 4 do not seem to have any significant impact on the likelihood of queuing and they do not change the marginal effect of CRRA once they are introduced. To gauge the strength of the CRRA marginal effect, it helps to realize that moving from the 20th percentile point in the distribution of CRRA to the 80th percentile point (a difference of 0.4134 in CRRA), increases the likelihood of queuing by an average 6.1 percentage points (Spec 5). In comparison, increasing education by 3 years (the duration of Junior Secondary School, Senior Secondary School or a University

determinants of the probability of employment in the formal sector (i.e. people who are more educated are considerably more likely to have a formal job).

Bachelor in Ghana) increases the likelihood of queuing by an average of 8 points. Risk-aversion, together with education and age, is clearly playing an important role in determining workers' willingness to queue: younger, more educated and more risk-averse people are significantly more likely to seek formal employment than to work in the informal sector.

Specification 6 introduces controls for the marital status of the worker, for whether he/she is the head of the household and for the number of his/her children. The reason for having these variables in the model is that, within my dataset, they constitute some potential determinants of both risk-aversion and likelihood of employment, and I want to investigate whether omitting them from the previous specifications had caused any bias in the estimated effect of CRRA. For instance, the head of the household could be presumed to be more risk-averse than other members of the same family, since she is in charge not only of her own well-being, but also of the well-being of the other family members. At the same time, she might be expected to be more likely employed than the rest of her family, precisely because household headship tends to be assigned to the principal income earner. Marital status and children can also affect occupational outcomes through a number of channels (e.g. women who are married and have many children are more likely to be unemployed), while they could also have an impact on a person attitudes to risk. Nevertheless, when I include these variables, I observe only a small change in the estimated marginal effect of the CRRA on both the likelihood of queuing and of being employed in the informal sector, rejecting the concern of significant omitted variable biases in the previous specifications.

Finally, in column 5 and 7, family income per capita is introduced as an additional control. This is the closest approximation I have available for $\underline{y}_i$ (i.e. the external income upon which a job seeker can rely until her search is successful) which, as discussed in section 3, represents another important determinant of occupational choices, since it affects the expected utility from choosing to queue. Specifically, being able to rely on high enough non-labour income, makes the prospect of unemployment less dramatic and it reduces the risk of searching for a job in the formal sector. An additional reason to control for family income per capita is that in this way I can address a potentially important issue of simultaneity between CRRA and Employment. This would be a concern if, for instance, workers who are currently unemployed and *cannot rely on financial support from external sources* tend to reveal higher risk-aversion in the lottery games (since they may prefer some low but safe winnings to help sustain their living). Interestingly, when I control for family income per capita, I do not find evidence of any such simultaneity, as the marginal effect of the CRRA does not change dramatically in Table 2 (where it drops from 0.15 to 0.12) and it only slightly increases (in absolute terms) in Table 3 (where it goes from -0.155 to -0.182). In the last specification, where I include all the regressors proposed, the marginal effect of CRRA becomes just insignificant in the first table, whereas it keeps its significance at the 10% level in Table 3. The restricted sample size (due to the unavailability of family income per capita in 2006), represents the most likely determinant of such a drop in precision.

Figure 2 and 3 translate the results from the Multinomial Logit into a graphical representation. There, I plot the predicted likelihood of being in each of the three possible categories against the whole range of estimated coefficients of relative risk-aversion, while holding the other regressors at their mean. In Figure 2, I plot the results from specification 4 above, while Figure 3 summarises the results in specification 5. Since the marginal effects in the Multinomial Logit are non-linear functions of the regressors, these graphs are more informative than the Table 2 and 3. I also plot the 95% confidence intervals around the predicted probabilities, to investigate whether they are statistically different from one another. The results are clear and they largely confirm my conclusions so far. There is a positive relationship between risk-aversion and the likelihood of queuing for formal jobs, while the opposite relationship exists between CRRA and the likelihood of informal employment (throughout the whole range of CRRA values). For completeness, I have also plotted the predicted probability of being in the formal sector, although, as argued above, we do not expect risk-aversion to have a direct causal impact upon it. This probability remains very low throughout the graph and indeed, it does not seem to be significantly affected by changes in CRRA. In Figure 3, although the marginal effect is reduced and the precision of my estimates drops (due to a significantly smaller sample size), the effects described so far are still identifiable. The conclusion is that a positive relationship between risk-aversion and the likelihood of queuing (and a negative one between risk-aversion and the likelihood of informal employment) exists and remains statistically significant. It should also be remarked that in Section 4 we estimated the large majority of our sample to have a coefficient of relative risk-aversion between 0 and 1. This turns out to be the domain over which the marginal effects of risk-aversion are steepest, and the predicted probabilities most clearly follow different trends from each other.

Table 2: Marginal Effects on the Probability of Queuing for a Formal Sector Job

	(1)	**(2)**	**(3)**	**(4)**	**(5)** ††	**(6)**	**(7)** ††
CRRA	.207*** (.082)	.172** (.087)	.172** (.087)	.148* (.080)	.147* (.086)	.136* (.081)	.121 (.081)
Male†		-.038 (.053)	-.038 (.05332)	-.091* (.053)	-.049 (.056)	-.052 (.058)	-.023 (.058)
Age		-.012*** (.003)	-.012*** (.003)	-.090*** (.017)	-.076*** (.015)	-.075*** (.017)	-.065*** (.017)
Age^2				.001*** (.0003)	.001*** (.0002)	.001*** (.0002)	.001*** (.0002)
Years of Education		.035*** (.008)	.035*** (.008)	.051** (.026)	.030 (.030)	.066** (.027)	.043 (.030)
Years of Education^2				-.001 (.002)	.001 (.002)	-.002 (.002)	.0003 (.002)
Year=2005†			.014 (.043)	.001 (.046)	-.054 (.044)	-.011 (.049)	.068 (.045)
Year=2006†			-.011 (.047)	-.033 (.048)		-.032 (.056)	
Family Income per Capita					-8.97e-06*** (.00000)		-8.65e-06*** (.00000)
Married†						-.212*** (.067)	-.140** (.070)
Number of Children						.058*** (.018)	.052** (.023)
Head of Household†						-.112* (.066)	-.110* (.060)
Obs.	585	585	585	585	367	551	364

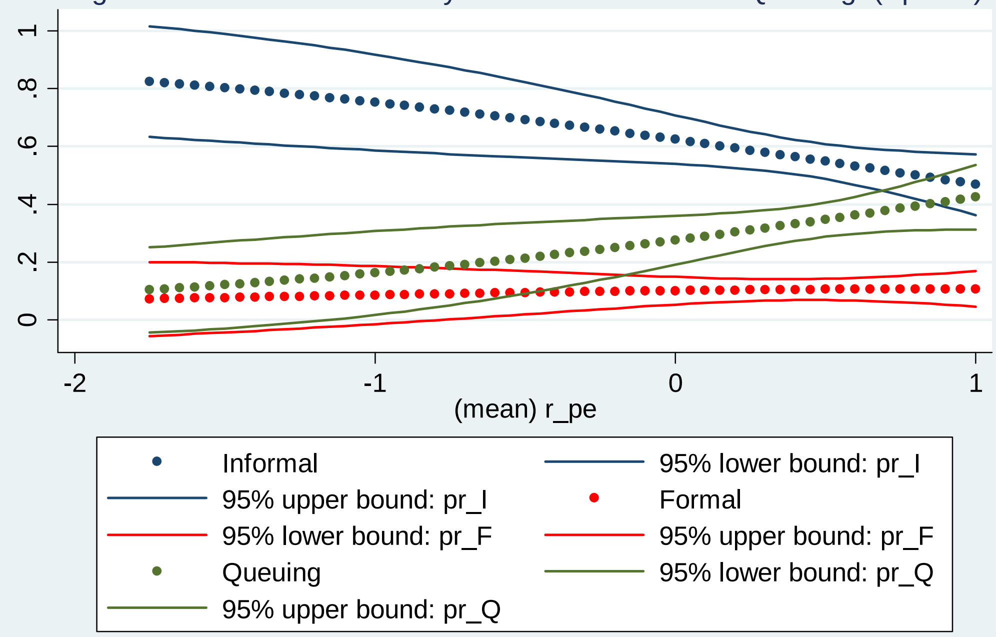
Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01
†Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Table 3: Marginal Effects on the Probability of Working in the Informal Sector

	(1)	(2)	(3)	(4)	(5) ††	(6)	(7) ††
CRRA	-.218*** (.077)	-.177** (.083)	-.177** (.083)	-.155** (.077)	-.178* (.094)	-.149* (.077)	-.183* (.094)
Male†		-.066 (.060)	-.066 (.060)	-.029 (.062)	-.106 (.077)	-.056 (.068)	-.138* (.083)
Age		.010*** (.003)	.010*** (.003)	.092*** (.019)	.076*** (.019)	.078*** (.019)	.062*** (.019)
Age^2				-.001*** (.0003)	-.001*** (.0003)	-.001*** (.0003)	-.0008*** (.0003)
Years of Education		-.058*** (.008)	-.058*** (.008)	-.062** (.026)	-.035 (.034)	-.074*** (.028)	-.043 (.034)
Years of Education^2				.0002 (.002)	-.002 (.002)	.0008 (.002)	-.002 (.002)
Year=2005†			-.016 (.048)	-.003 (.051)	.031 (.051)	.006 (.054)	-.038 (.054)
Year=2006†			-.002 (.051)	.018 (.054)		.029 (.061)	
Family Income per Capita					7.42e-06*** (1.23e-06)		7.22e-06*** (1.21e-06)
Married†						.199*** (.073)	.154* (.084)
Number of Children						-.047** (.021)	-.043* (.025)
Head of Household†						.081 (.070)	.094 (.076)
Obs.	585	585	585	585	367	551	364

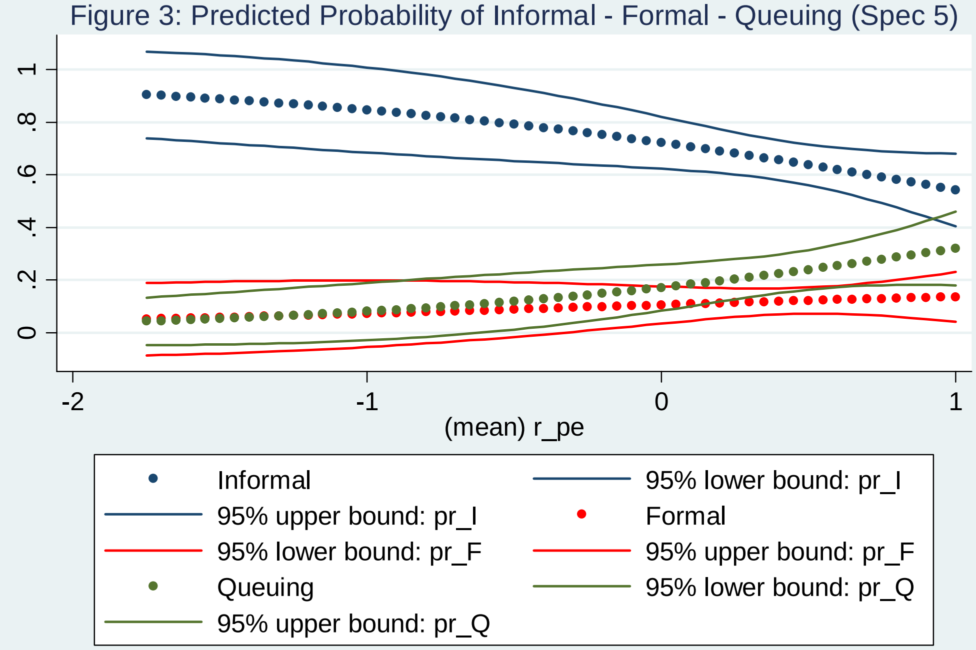
Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01
†Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Figure 2: Predicted Probability of Informal - Formal - Queuing (Spec 4)



The area between the solid lines shows the 95% CI of the Predicted Probability

Figure 3: Predicted Probability of Informal - Formal - Queuing (Spec 5)



The area between the solid lines shows the 95% CI of the Predicted Probability

Estimation Issues and Robustness Checks

In this section I will attempt to address the main concerns surrounding my estimation strategy and I will carry out a number of robustness checks on the results shown above.

First, since my dataset does not allow me to distinguish the formal job-seekers from the rest of the unemployed, I have treated the entire pool of current unemployed workers as a proxy for the queue of formal job-seekers. Though I acknowledge its potential drawbacks, there are reasons to believe in the validity of this approach. First, given that our sample only contains individuals between 15 and 65 years of age, the large majority of people who are currently not working are significantly more likely to be seeking employment than to be out of the labour force. In addition, the analysis of data from the latest waves of the UPS dataset, coupled with casual observation during survey work suggest that the vast majority of the currently-not-working population is searching for employment or, at least, would be willing to work if an opportunity appeared. Whether these people are specifically searching for formal employment is less clear, but given the extremely easy access to the informal labour market, it does not seem unreasonable to think that if these workers are staying unemployed, it could be because they are waiting for better opportunities in the formal sector. At least two critiques can be advanced against this view. First, it might well be possible that a queue for self-employment also exists, and that some of the unemployed in the sample may in fact be preparing to take on some form of (most likely informal) self-employment in the future, presumably by accumulating capital. The evidence I will present in the next chapter, however, shows that capital requirements in self-employment are very low (with entrepreneurs generating high margins from very small capital bases) which, coupled with evidence of widespread lack of savings (which may, of course, be endogenously determined by lack of saving mechanisms), points against the existence of a queue for self-employment. Second, my pool of queuers may mistakenly contain workers with high reservation wages and low incentives to search, who are effectively out of the labour force. My model tries to deal with this issue via the inclusion of unearned income sources, proxied by family income per capita in the estimation (of which reservation wages are likely to be a function). However, to the extent that I cannot control for unobserved sources of income, or for other sources of heterogeneity driving reservation wages, this concern can only be partly addressed in the empirics. Intuitively, though, it would seem unreasonable to believe that high reservation wages may keep a sizable portion of the sample out of the labour force in a context of widespread poverty and lack of social security. As a robustness check on my results I have repeated the analysis confining it to only the Men in the sample. This strategy is based on the idea that female labour market attachment is generally lower than it is for male workers, and confining the analysis to men should therefore reduce the proportion of workers who are out of the labour force but get misclassified as job-seekers. In other words, restricting the sample to the Men should make the unemployed pool a more accurate approximation of the formal job-seekers, shedding light on the extent of the potential bias that may arise from misclassification. The results (which are included in the Empirical Appendix) largely confirm the findings I obtained using the entire sample. The direction of the marginal effect of CRRA remains the same, its size increases (indicating an even stronger effect for Men) and its statistical significance holds (See Table 5 and 6 in the Empirical Appendix), potentially lending support to the idea that reducing mis-categorisation strengthens the results.[18]

Second, it should be remarked that the estimated marginal effect of Risk-Aversion can only be given a causal interpretation under the assumption that we are able control for all factors affecting occupational choices and attitudes to risk simultaneously (i.e. that we are able to control for the potential endogeneity of risk-aversion with respect to other determinants of occupational choices). For instance, if risk-aversion turned out to be a function of age and/or education, which are two important determinants of sectoral allocation, its estimated marginal effect would be highly misleading. As a check on my results, I have used the last wave of the panel (2006) to estimate a reduced form equation with CRRA as a function of the controls used in the multinomial logit (the results of this estimation are also contained in the Empirical Appendix). Quite strikingly, none of these variables, except being the head of the household, show a significant effect on risk-preferences, which points to the conclusion that risk-aversion might indeed be exogenous with respect to most personal characteristics and in particular to the other most important determinants of occupational outcomes. These results are in line with the fact that including an extensive set of controls in the multinomial logit did not significantly change the estimated marginal effect of CRRA. In fact, even controlling for the status of head of the household, which shows a statistically significant negative impact on Risk-

[18] For completeness I should report that the results also hold when I confine the same analysis to the Women in the sample.

Aversion in the reduced form, did not affect the result of the multinomial logit dramatically. Of course, I fully acknowledge the possibility that omitted sources of endogeneity might still be at play, but recent evidence from twin studies in behavioural economics lends support to the idea that attitudes to risk are largely heritable (Zyphur et al. (2009)) and therefore exogenous with respect to labour market outcomes that occur during a person's lifetime and mostly in adulthood (when the period of upbringing during which preferences and attitudes are most malleable is over).

Third, given that my empirical strategy is entirely driven by cross-sectional variation in attitudes to risk across different sectors of the labour market, the most natural way to use my panel dataset has been to pool observations from different survey waves into a unique cross-section. This approach is motivated by the design of my theoretical model, where workers' decisions are entirely based on current expectations of employment outcomes (i.e. expected earnings and expected variance of earnings) from different strategies. In other words, a worker chooses whether to queue for a formal job or to join the informal sector on the basis of her current expectations on the utility she will get from the two alternatives. Given that a worker's job-prospects can change over time and so can change her expectations (while their risk-aversion is assumed to be time-invariant), we should expect occupational choices to change over the years and we should consider the sectoral choices of the same worker in two different years as separate (since they are based on potentially different expectations)[19]. In fact, a non-negligible fraction of workers

[19] To be precise, according to the theoretical model, the sectoral choice of a worker in period t depends on her expectations (at time t) of earnings and earnings uncertainty in period $t+1$.

is observed to change occupational status across survey waves. Including time-variant individual characteristics (e.g. age, children, marital status) in the Multinomial Logit and period-specific dummies was an attempt to allow for the possibility that employment conditions may evolve systematically with a worker' characteristics, causing change in her occupational choices over the years. At the same time, I am aware of the imprecision of this pooling approach, arising from the fact that some unobserved determinants of the employment decisions will be correlated between observations on the same individual in subsequent years. To overcome this problem, I should remark that in the previous estimations I relaxed the assumption of independence of the error components and I allowed for correlation between data on the same worker, correcting the standard errors accordingly. As a further check on the robustness of the pooling approach, I have also estimated the same multinomial logit separately on each wave of the panel (results for the latest wave are included in the Empirical Appendix). The results I have obtained are largely consistent with the ones above. Although, as a likely consequence of the reduced sample size, precision decreases, the direction of the estimated marginal effects of CRRA remains the same and it is consistent across panel waves. This is indicative of the fact that with a larger cross-section, I should expect to improve upon the efficiency of my estimates, and that pooling across waves should not have distorted the nature of my results.

Presuming that period t (the choice period) is relatively small compared to period $t+1$ (the working period), I can assume that both of them are contained within the span of one survey year. In this sense, it is reasonable to treat the choice of a worker at the time of one survey wave as separate from her decision at the time of the next (and the previous) wave.

In order to tackle both the issue of endogeneity in risk-aversion and the potential problems of identification in the cross-section more thoroughly, it is my plan to re-visit the sample of respondents that participated in the experimental game in 2007 and re-elicit their preferences for risk. Coupled with information on their labour market outcomes collected over the past years, this data will provide an effective tool to study the evolution of risk-aversion in conjunction with the evolution in employment histories.[20]

Finally, the measure of risk-aversion I have used throughout this analysis is itself an estimate obtained from fitting a choice model on my experimental dataset using maximum likelihood. As such, it carries an estimation error; and as the last check on the robustness of my results, I have explicitly accounted for this source of imprecision by re-weighting the sample according to the error attached to each value of CRRA. The results are reported in Table 10 and 11 in the empirical appendix and they show that, despite a slight drop in the magnitude of the estimated marginal effect of risk-aversion, the estimated patterns remain consistent with the discussion so far.

<u>Estimation of Sectoral Risk and Interpretation of the Results</u>

Having established the existence of a positive and statistically significant relationship between risk-aversion and the likelihood of queuing for a formal job

[20] Of course, given the availability of UPS waves of data for the years after 2007, the reader may wonder why we have not already attempted to study labour market outcomes that followed the measurement of risk-aversion, a strategy that would have partly resolved concerns of endogeneity of RA. Unfortunately, such strategy has encountered a set-back due to the attrition rates in the panel that have made it difficult to trace a sufficiently high number of respondents from the small original experimental sample, hence leading us to necessitate the design of new behavioural tasks to be carried out in the future with a larger number of respondents.

(and of a negative and statistically significant relationship between risk-aversion and the likelihood of working in the informal sector), I will now attempt to ascertain whether these findings are in line with the theoretical framework proposed in Section 2. The *Conclusion* of my theoretical model was that, under Case 1 and 2, the likelihood of preferring formal job-search over informal employment should increase with a worker's risk-aversion. Therefore, for my empirical results to be consistent with the model, Case 1 or Case 2 should prove to be a plausible description of reality. Specifically, I will ask whether searching for a formal job is indeed the low-risk strategy, in the sense of a lower variance of earnings in period $t+1$, compared to working in the informal sector.[21] To tackle this question, I will focus on a representative worker (I will define her "a") with average observed characteristics. The reason for using this approach is that it pairs with the fact that the marginal effects from the multinomial logit were computed holding all the regressors at their average and therefore, the estimates in Table 2 and Table 3 were precisely indicative of the behaviour of an average individual in my sample.

First of all, I will estimate the income uncertainty at time $t+1$ that this average worker can expect if she chooses to queue for formal employment (i.e.$Var[y_{aQ}]$). This variance is a function of her expected earnings in the formal sector (y_{aF}), of the outside income that she can fall back on if her search is not successful ($\underline{y}_a$) and of the probability of successful search (p_a). To match the workings of the theoretical model as closely as possible, I maintain the assumption that, conditional on seizing a formal job, worker a can expect to earn y_{aF} with certainty at time $t+1$. This wage

[21] I will need to show that $Var[y_F] < Var[y_I]$ holds.

will depend on her observed (average) characteristics and I will predict it (for every different year) using a Mincerian earnings regression of log-wages in the formal sector[22]. The minimal income level $\underline{y}_a$, will be proxied by the average family income per capita among the unemployed. The third component of $Var[y_{aQ}]$ is the probability p_a that the average worker manages to get a formal job. The highest possible estimate of p_a is equal, in any one year, to the ratio of the total number of formal jobs in the economy divided by the sum of formal jobs and unemployed people. The reason I consider this estimate to be an upper bound is that it describes a situation where all the formal jobs are suddenly vacated and become available to the sum of the previous and the new unemployed. If I assume, more realistically, that only a fraction of the formal jobs available is vacated at any given period, p_a decreases. However, it can easily be shown that decreasing p_a from its upper-bound value decreases $Var[y_{aQ}]$ monotonically.[23] It follows that, once I can prove that $Var[y_{aQ}] < Var[y_{aI}]$ holds at the upper-bound, the same inequality will automatically hold at any lower (and more plausible) value of p_a. Having defined y_{aF}, $\underline{y}_a$ and p_a, as well as the assumptions of the model, the estimated variance of earnings for the average worker who chooses to queue for formal employment can be calculated as follows:

[22] $y_{itF} = \beta' X_{it} + u_{it}$, where y_{itF} is a vector of individual earnings in the formal sector in period t; X_{it} is a matrix of (possibly time-variant) individual characteristics, including gender, age (both linear and quadratic) and years of formal education (both linear and quadratic); u_{it} is a random error component.

[23] This is true as long as this upper bound for $p_a$is lower than ½ (which is highly plausible, on both empirical grounds and on the basis of the assumptions generally made in this literature).

$$\widehat{Var[y_{aQ}]} = \hat{p}_a * \left[\hat{y}_{aF} - \left(\hat{p}_a * \hat{y}_{aF} + (1-\hat{p}_a) * \underline{\hat{y}_a}\right)\right]^2 + (1-\hat{p}_a) * \left[\underline{\hat{y}_a} - \left(\hat{p}_a * \hat{y}_{aF} + (1-\hat{p}_a) * \underline{\hat{y}_a}\right)\right]^2$$

Next, I will outline the estimation procedure for $Var[y_{aI}]$, the variance of earnings in the case that the average worker opts for the informal sector. In those circumstances, she will face an uncertain level of earnings y_{aI} in period $t+1$; these earnings will be, in general terms, high in "good" years and low in "bad" years. To estimate such volatility, I construct a sub-sample of workers in the informal sector with similar skills and experience to the average worker.[24] Then, I compute the variance of their earnings in every given year and I treat it as a proxy for $Var[y_{aI}]$. The idea is that, since I cannot actually observe the earnings of the same individual in different states of the world, I assume that similar individuals should face similar earnings in the same job category and that the cross-sectional variation I observe among them is a good approximation of the uncertainty resulting from random shocks.

Using this procedure, I obtain the following estimates of the model parameters and of the variance of (real monthly) earnings in the two sectors and in the two survey waves, for the average worker in my sample.[25]

<u>In 2005</u>:

$\hat{y}_{aF} = 86{,}766.24$; $\underline{\hat{y}_a} =$ 18,953.8 ; $\hat{p}_a = .27$

[24] Specifically, I include in this sample all the workers in the informal sector who fall within the middle four deciles of the distribution of Years of Formal Education and Age. [In 2005: (7<years of education<=10) and age (27.54<age<40); In 2006: (7<years of education<=10) and age (26<age<38.91)]. I end up with a sample of 143 individuals in the 2005 wave of data and 157 in 2006.

[25] It was not possible to perform the same computation for the year 2004, since household income per capita was not available for that particular year.

$\widehat{Var[y_{aQ}]} = 8.99e+08 < 4.22e+09 = \widehat{Var[y_{aI}]}$

In 2006:

$\hat{y}_{aF} = 113{,}296.2$; $\underline{\hat{y}}_{a} = 27{,}576.78$; $\hat{p}_a = .31$

$\widehat{Var[y_{aQ}]} = 1.57e+09 < 6.37e+09 = \widehat{Var[y_{aI}]}$

In both years, the estimated variance of earnings in the informal sector is considerably (4 to 5 times) higher than the variance of earnings if the worker chooses to seek formal employment, indicating a much higher degree of income uncertainty to be expected from choosing strategy I (Informal sector) over strategy Q (Queuing for a formal job). This result corroborates the hypothesis that Case 1 and 2 from the theoretical model, which occur if $Var[y_{iQ}] < Var[y_{iI}]$, may indeed capture the reality of the labour market; and it suggests that the positive estimated relationship between risk-aversion and likelihood of queuing for a formal job is consistent with my theory of occupational choices. Working in the informal sector is the riskier job-option and the likelihood of choosing it over seeking formal employment should decrease in workers' risk-aversion.

An alternative and potentially more robust method to estimate the volatility of earnings under strategy I (i.e. $Var[y_I]$) would be to exploit the time-series dimension of my data to calculate the variance of the average worker's earnings in the informal sector over the three survey waves. This approach would be superior to the cross-sectional estimation used above, as a way to quantify earnings volatility over different states of the world. It would also constitute an improvement on previous work by Bonin et al. (2007), who rely on cross-sectional earnings risk in their analysis of occupational sorting and risk-attitudes in Germany. In fact, the cross-sectional variance I calculated previously might have been driven by unobserved differences among workers within the sub-sample "around the average", in which case it would constitute a poor estimate of volatility. The UPS dataset is ideally suited to solve this problem, since its panel structure allows me, in principle, to control for individual fixed effects and to isolate the variation in earnings that derives from idiosyncratic shocks in different periods[26]. Unfortunately, given the short time-horizon covered in the data, this method cannot be applied effectively. Three survey waves (and therefore, three observations per individual) are clearly too few to obtain reliable estimates of individual earnings variances. Nevertheless, I will produce some evidence using this second approach, keeping in mind that it is only a tentative estimation of the object of interest. To do that, I will first compute the variance of individual earnings over the three years of data for workers in the informal sector. Then, I will average these estimates across the same sub-sample of "average workers" (i.e. workers with age and skills close to the average) that I used before. With enough data, this procedure should result in a robust estimate of the earnings volatility that an average worker can expect in any particular period if she works in the informal sector[27]. The results I obtain are the following:

[26] To be precise, the variance of individual earnings over time would also contain sheer (real) earnings growth, but in our sample such growth was estimated to be low between 2004 and 2006 and it cannot account for a large share of the variation.

[27] A more elaborate way to carry out the same estimation would be to design a reduced form equation for the time-variance of individual earnings, as a function of observable workers' characteristics, and to estimate its coefficients via regression techniques (e.g. OLS). I could then use these estimates to predict the variance of earnings for an average worker. The predictions I obtained when I tried this alternative approach confirmed the result that risk is higher in the informal sector. However, my

In 2005:

$$\widehat{Var[y_{aI}]} = 2.39e + 09$$

In 2006:

$$\widehat{Var[y_{aI}]} = 2.40e + 09$$

These figures are still higher than the previously estimated variances of earnings from strategy Q, in both 2005 and 2006. My conclusion, therefore, remains the same: the informal sector is the riskier of the two options available to workers.[28]Finally, I fully acknowledge the possibility that measurement error may lead to overestimate the variance and that such bias is likely to be more significant for informal businesses (that generally lack precise records). However, both casual observation and formal empirical evidence point towards the conclusion that the informal sector is indeed more uncertain. Moreover, given the large difference in the magnitudes of the estimated variances above, I would be hesitant to attribute the entire gap to measurement error. Third, a recent methodological study by Fafchamps et al. (2011) in urban Ghana investigates the effectiveness of consistency checks, performed with handheld computers during the data collection, in reducing the degree of measurement error in the earnings of informal worker. The study is conducted in enumeration areas that are very similar to those surveyed by the GHUPS used in this analysis. Their main finding is that “the vast majority of large

estimates were rather imprecise and the explanatory power of the morel very low (this is why I have preferred the approach described in the main text). With additional earnings from the latest UPS wave I will be able to use this approach more effectively.

[28] This result also holds using data from the most recent waves of the GHUPS.

changes in enterprise sales and profits are confirmed by firm owners as genuine, highlighting the volatility of income in this sector".[29]

Finally, it should be remarked that the measure of risk I have used throughout the analysis focuses on short-term income uncertainty (and on how this differs across the available employment strategies). An alternative approach would have been to isolate persistent shocks (from short-term fluctuations), which might constitute a more relevant dimension of risk for career choices if transitory fluctuations can be smoothed out through precautionary savings. I would challenge this view on the grounds that in an economy that lacks effective insurance and saving devices, short-term volatility is unlikely to be smoothed out entirely and may therefore impact consumption significantly, constituting a very relevant dimension of risk.[30] Evidence on the role of different types of shocks on employment dynamics in Latin America has been recently presented by Fiess et al. (2010), and it would of course be an interesting extension to this analysis trying to separate out short term fluctuations form permanent shocks, using, for instance, the approach proposed by Meghir and Pistaferri (2004), but the limitations of our short panel foreclose this possibility until additional data becomes available.

[29] Moreover, Fafchamps et al. (2011) find that while consistency checks did succeed in reducing the coefficient of variation and increase the autocorrelation of earnings data across survey waves, the overall effects on the distribution of earnings were fairly modest, since few of the observations were identified as errors and consequently changed, and the changes were not order of magnitude changes..

[30] Preliminary evidence from recent waves of GHUPS supports this view, showing low levels of savings among respondents.

6 Conclusions

This chapter has investigated how attitudes to risk determine the allocation of workers into formal and informal occupations. A two-period theoretical model has been proposed, where workers choose between two alternatives: working in the free entry informal sector (defined as the sum of the self-employed and the salaried workers in small/micro enterprises) and searching for a formal occupation (in the public sector or in larger private firms). Both options are risky, although two different sources of risk are taken in consideration. On one hand, the informal labour market is free entry, but it is risky in the sense that earnings are volatile over good and bad states of the world. On the other hand, earnings in the formal sector are assumed to be constant once a job is found (i.e. not subject to idiosyncratic fluctuations), but access to those occupations is rationed and job-seekers run the risk of finding themselves unemployed. The structure of the model suggests how we should weigh these two sources of uncertainty against each other, allowing me to estimate the overall expected variance of earnings from either strategy and hence determine which one is “riskier”. In accordance to our priors, the evidence shows that uncertainty is higher in the informal sector and, therefore, that we should observe more risk-averse workers as less likely to sort into it. Such prediction is strongly confirmed in my empirical analysis. Using innovative econometric technologies for the estimation of utility functions, coupled with a new experimental dataset I gathered in collaboration with CSAE, it has been possible to estimate an exogenous measure of individual risk-aversion for a representative sample of Ghanaian workers, whose labour market histories had been previously followed for a period of 3 years. Including this measure among the explanatory variables in a Multinomial Logit model of occupational choices, it emerges that workers who showed to be more risk-averse in the experimental setting are generally more likely to prefer the safer occupational option: searching/queuing for a formal job. This effect is statistically significant and robust to concerns regarding the potential endogeneity of the risk-aversion coefficients.

The results of this paper should naturally contribute to the debate on the importance of attitudes to risk as a determinant of occupational outcomes. Developing a better understanding of how risk-aversion affects workers' decisions will be important to improve anti-poverty programmes that operate through the workings of the labour market. For instance, this paper suggests that a government scheme that aims to increase the number of jobs in the formal sector might fail to benefit the poor if their risk-aversion is so low that they would prefer to "try their luck" at the riskier income lottery that constitutes the informal sector. A crucially related question would then be whether poor workers are systematically more or less risk-averse than wealthier people and if so, why.[31] This is only one of the research routes left open by this paper. The more general question of *how risk-attitudes originate* would also deserve further investigation, given its potential relevance for both theoretical modelling and economic policy. Specifically, being able to precisely identify the direction of causality between risk-attitudes and economic outcomes would be crucially

[31] In the well-known paper "The Two Poverties" (2000), Banerjee proposes that the poor might either be more desperate (and therefore less risk-averse) or more vulnerable (and therefore more risk averse) than the non-poor. Discerning empirically between these two theoretically plausible alternatives would be extremely interesting.

important, as it would help resolve the debate on the endogeneity of risk-preferences, which was discussed at length in this paper.

Finally, my experimental design should contribute to a methodological discussion on the benefits and drawbacks of using experiments to estimate exogenous measures of individual risk-aversion that are amenable to be used as explanatory variables not only in the study of occupational choices, but also of other economic decisions of interest (e.g. saving/investment decisions, borrowing and microfinance).

References

ANDERSEN, S., J.C. COX, G.W. HARRISON, M. LAU, E.E.RUTSTROEM and V. SADIRAJ (2011): “Asset Integration and Attitudes to Risk: Theory and Evidence”, *Department of Economics, Durham University*, WP: 2011–10.

AZAM, J. P., B. BIAIS, et al. (2002): “Risk Aversion and Entrepreneurship in the Ivory Coast: A Structural Econometric Approach”, *University of Toulouse working paper*.

BANERJEE, A. V. (2000): “The Two Poverties.” *Nordic Journal of Political Economy,* vol.26, pages 129-141.

BANERJEE, A. V. and A. F. NEWMAN (1993): “Occupational Choice and the Process of Development.” *Journal of Political Economy*, University of Chicago Press, vol. 101(2), pages 274-98, April.

BARR, A. (2007): “Attitudes to Risk in Ghana: Field Manual.” *Unpublished*.

BASU, K. (1997): “Analytical development economics”, MIT Press.

BECKER, G. S. (1962) “Investment in Human Capital: A Theoretical Analysis.” *The Journal of Political Economy* 70(5): 9-49.

BECKER, G. S. (1964): “Human capital, National Bureau of Economic Research”; distributed by Columbia University Press New York.

BONIN, H., T. DOHMEN, A. FALK, D. HUFFMAN and U. SUNDE (2006): “Cross-sectional earnings risk and occupational sorting: The role of risk attitudes”, *Labour Economics,* Volume 14, Issue 6, December 2007, Pages 926-937.

BRUNNERMEIER, M. K. (2001) “Asset Pricing Under Asymmetric Information: Bubbles, Crashes, Technical Analysis, and Herding”, Oxford University Press.

DERCON, S. (1996): “Wealth, risk and activity choices: cattle in Western Tanzania.” The Centre for the Study of African Economies *WPS: 44*.

FAFCHAMPS, M., D. McKENZIE, S.QUINN, and C.WOODRUFF (2011): “Using PDA consistency checks to increase the precision of profits and sales measurement in panels,” *Journal of Development Economics*, 2011 (forthcoming).

FIELDS, G. S. (1975): “Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs.” *Journal of Development Economics* 2(2): 165-87.

FIELDS, G. S. (1980a). “Education and income distribution in developing countries: a review of the literature.” *World Bank Staff Working Paper*.

FIELDS, G. (2005): “A Guide to Multisector Labor Market Models.” *World Bank, Social Protection Working Paper*.

FIESS, N.M., M. FUGAZZA and W.F. MALONEY (2020): “Informal self-employment and macroeconomic fluctuations.” *Journal of Development Economics,* Elsevier, vol. 91(2), pages 211-226, March.

HARRIS, J. R. and M. P. TODARO (1970): “Migration, Unemployment and Development: A Two-Sector Analysis.” *The American Economic Review* 60(1): 126-142.

HARRISON, G. W. (2007): “Maximum Likelihood Estimation of Utility Functions Using Stata”, Working Paper 06.

HUMPHREY, S. J. and A. VERSCHOOR (2004): “The probability weighting function: experimental evidence from Uganda, India and Ethiopia.” *Economics Letters* 84(3): 419-425.

ILO (2002) “Women and Men in the Informal Economy: A Statistical Picture”. (Geneva:ILO).

ILO (2003): “Global Employment Trends”. (Geneva:ILO).

KANHEMAN, D. and A. TVERSKY (1979): "Prospect theory: An analysis of decision under risk*." Econometrica* 47(2): 263-291.

KIHLSTROM R.E. and LAFFONT J.J. (1979): "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion", *The Journal of Political Economy*, Vol. 87, No. 4 (Aug., 1979), 719-748.

KING, A. G. (1974): "Occupational Choice, Risk Aversion, and Wealth." *Industrial and Labor Relations Review* 27(4): 586-596.

LEONTARIDI, M. (1998): "Segmented Labour Markets: Theory and Evidence." *Journal of Economic Surveys* 12(1): 103-109.

LEWIS, W. A. (1954): "Economic Development with Unlimited Supply of Labour." *The Manchester School* 22(2): 139-191.

MEGHIR, C. and L. PISTAFERRI (2004): "Income Variance Dynamics and Heterogeneity." *Econometrica* 72(1): 1-32, 01.

RANKIN, N., J. SANDEFUR, et al. (2008): "Learning & Earnings in Africa: Evidence from Ghana", *CSAE Working Paper*, University of Oxford.

ZYPHUR, M.J., NARAYANAN, J., ARVEY, R.D. and GORDON, J.A. (2009): "The genetics of economic risk preferences", *Journal of Behavioral Decision Making,* Vol. 22(4): 367-377

APPENDIX

Table 1: Summary Statistics for workers' in the UPS and in the Experimental Dataset, in 2004

Variable	Mean (UPS)	Mean (Exper)	Obs (UPS)	Obs (Exper)
Male	.440	.411	1144	192
Age	31.250	31.538	1143	192
Years of Education	8.93	8.091	1144	192
Married	.480	.474	1143	192
Household Head	.318	.330	1087	176
Household Size	3.882	4.157	468	83
Children	1.658	1.81	1142	192
Maths Test Score	.362	.365	1029	188
English Test Score†	N/A	N/A	N/A	N/A
Ethnicity 1 **	.591	.354	1145	192
Ethnicity 2 **	.164	.318	1145	192
Ethnicity 3 **	.091	.094	1145	192
Ethnicity 4 **	.010	.182	1145	192
Ethnicity 5 **	.116	.052	1145	192
Informal Job	.533	.547	1145	192
Formal Job	.131	.120	1145	192
Unemployed	.336	.333	1145	192
Real Monthly Earnings††	52663.72	44342.07	1052	177
Height (cm)	163.781	163.584	739	130
Number of Days when ill in the last year†	N/A	N/A	N/A	N/A

†Variable not available in 2004; ††Old Ghana Cedi (prior to the 2007 reform that converted 10,000 Old Ghana Cedis into 1 New Ghana Cedi). ** Ethnicity 1 = Akan; Ethnicity 2 = Ga-Dangme; Ethnicity 3 = Ewe; Ethnicity 4 = Mole-Dagbani and Hausa ; Ethnicity 5 = Other.

Table 2: Summary Statistics for workers' in the UPS and in the Experimental Dataset, in 2005

Variable	Mean (UPS)	Mean (Exper)	Obs (UPS)	Obs (Exper)
Male	.431	.388	1380	255
Age	31.875	32.258	1377	255
Years of Education	8.594	8.16	1376	255
Married	.407	.424	1376	255
Household Head	.385	.368	1343	250
Household Size	4.800	5.127	1229	220
Children	1.644	1.9	1311	250
Maths Test Score	.355	.366	1373	255
English Test Score[†]	N/A	N/A	N/A	N/A
Ethnicity 1 [**]	.590	.345	1394	255
Ethnicity 2 [**]	.148	.302	1394	255
Ethnicity 3 [**]	.074	.075	1394	255
Ethnicity 4 [**]	.011	.200	1394	255
Ethnicity 5 [**]	.119	.078	1394	255
Informal Job	.522	.516	1349	250
Formal Job	.128	.116	1349	250
Unemployed	.351	.368	1349	250
Real Monthly Earnings[††]	56445.28	45298.89	1303	247
Height (cm) [†]	N/A	N/A	N/A	N/A
Number of Days when ill in the last year	3.169014	3.416	1349	250

[†]Variable not available in 2005; [††]Old Ghana Cedi (prior to the 2007 reform that converted 10,000 Old Ghana Cedis into 1 New Ghana Cedi). [**] Ethnicity 1 = Akan; Ethnicity 2 = Ga-Dangme; Ethnicity 3 = Ewe; Ethnicity 4 = Mole-Dagbani and Hausa ; Ethnicity 5 = Other.

Table 3: Summary Statistics for workers' in the UPS and in the Experimental Dataset, in 2006

Variable	Mean (UPS)	Mean (Exper)	Obs (UPS)	Obs (Exper)
Male	.443	.371	1692	240
Age	30.411	32.444	1689	239
Years of Education	8.335	8.244	1695	240
Married	.344	.394	1635	236
Household Head	.351	.374	1555	214
Household Size	4.908	5.216	1540	213
Children	.973	1.238	1502	210
Maths Test Score	.373	.361	1686	239
English Test Score	11.704	11.101	1687	238
Ethnicity 1 **	.420	.325	1696	240
Ethnicity 2 **	.110	.275	1696	240
Ethnicity 3 **	.042	.067	1696	240
Ethnicity 4 **	.008	.175	1696	240
Ethnicity 5 **	.090	.158	1696	240
Informal Job	.495	.55	1696	240
Formal Job	.156	.142	1696	240
Unemployed	.350	.308	1696	240
Real Monthly Earnings††	66841.55	66636.82	1630	232
Height (cm)	164.7158	165.829	1495	200
Number of Days when ill in the last year	2.816627	1.595833	1696	240

††Old Ghana Cedi (prior to the 2007 reform that converted 10,000 Old Ghana Cedis into 1 New Ghana Cedi).
** Ethnicity 1 = Akan; Ethnicity 2 = Ga-Dangme; Ethnicity 3 = Ewe; Ethnicity 4 = Mole-Dagbani and Hausa ; Ethnicity 5 = Other

Simulation 1:

$E(y_{iQ}) = 100$; $E(y_{iI}) = 100$; $Var(y_{iQ}) = 653$; $Var(y_{iI}) = 2500$;

γ_i	Expected Utility from Q	Expected Utility from I	Difference in Exp Utility (Q – I)	Strategy Chosen
-2	401400.7	583333.3	-181933	I
-1.8	166349.4	231449.7	-65100.3	I
-1.6	69436.55	92492.04	-23055.5	I
-1.4	29219.52	37274.36	-8054.85	I
-1.2	12409.69	15172.36	-2762.68	I
-1	5326.74	6250	-923.26	I
-0.8	2315.073	2611.934	-296.861	I
-0.6	1021.234	1110.918	-89.6843	I
-0.4	458.7825	482.9185	-24.1359	I
-0.2	210.9309	215.8111	-4.88025	I
0	100	100	0	=
0.2	49.51378	48.7062	0.807583	Q
0.4	26.21651	25.55876	0.657745	Q
0.6	15.65584	15.2529	0.402936	Q
0.8	12.49694	12.27699	0.219953	Q
1.2	-2.00531	-2.06101	0.055693	Q
1.4	-0.40307	-0.42986	0.026788	Q
1.6	-0.10827	-0.12092	0.012651	Q
1.8	-0.03279	-0.03868	0.005894	Q
2	-0.01061	-0.01333	0.002718	Q
2.2	-0.00359	-0.00483	0.001244	Q
2.4	-0.00125	-0.00181	0.000566	Q
2.6	-0.00044	-0.0007	0.000256	Q
2.8	-0.00016	-0.00028	0.000115	Q
3	-5.9E-05	-0.00011	5.19E-05	Q
3.2	-2.2E-05	-4.5E-05	2.33E-05	Q
3.4	-8.27E-06	-1.9E-05	1.04E-05	Q
3.6	-3.13E-06	-7.78E-06	4.65E-06	Q
3.8	-1.19E-06	-3.27E-06	2.07E-06	Q
4	-4.58E-07	-1.38E-06	9.25E-07	Q

Calibration: $p_i = .38$; $y_{iF} = 133$; $\underline{y} = 80$; $\theta_i^G = .5$; $y_{iI}^G = 150$; $y_{iI}^B = 50$;

Simulation 2:

$E(y_{iF}) = 100$; $E(y_{iI}) = 125$; $Var(y_{iQ}) = 662$; $Var(y_{iI}) = 5625$;

γ_i	Expected Utility from Q	Expected Utility from I	Difference in Exp Utility (Q – I)	Strategy Chosen
-2	403814	1354167	-950353	I
-1.8	167257.1	505311.2	-338054	I
-1.6	69777.98	189815.5	-120038	I
-1.4	29347.95	71869.94	-42522	I
-1.2	12457.99	27473.37	-15015.4	I
-1	5344.91	10625	-5280.09	I
-0.8	2321.907	4168.377	-1846.47	I
-0.6	1023.804	1664.787	-640.983	I
-0.4	459.7495	680.0695	-220.32	I
-0.2	211.2946	286.0067	-74.7122	I
0	100	125	-24.86	I
0.2	49.56524	57.61234	-8.0471	I
0.4	26.23586	28.7324	-2.49654	I
0.6	15.66311	16.38412	-0.721	I
0.8	12.49968	12.68031	-0.18063	I
1.2	-2.00493	-2.00969	0.004767	Q
1.4	-0.40293	-0.41155	0.008624	Q
1.6	-0.10822	-0.11439	0.00617	Q
1.8	-0.03277	-0.03635	0.003582	Q
2	-0.01061	-0.0125	0.001893	Q
2.2	-0.00358	-0.00453	0.000949	Q
2.4	-0.00125	-0.00171	0.00046	Q
2.6	-0.00044	-0.00066	0.000218	Q
2.8	-0.00016	-0.00026	0.000102	Q
3	-5.9E-05	-0.00011	4.71E-05	Q
3.2	-2.2E-05	-4.4E-05	2.15E-05	Q
3.4	-8.26E-06	-1.8E-05	9.79E-06	Q
3.6	-3.13E-06	-7.56E-06	4.43E-06	Q
3.8	-1.19E-06	-3.19E-06	2.00E-06	Q
4	-4.57E-07	-1.35E-06	8.97E-07	Q

Calibration: $p_i = .38$; $y_{iF} = 133$; $\underline{y} = 80$; $\theta_i^G = .5$; $y_{iI}^G = 200$; $y_{iI}^B = 50$;

Simulation 3:

$E(y_{iF}) = 100.14$; $E(y_{iI}) = \ 525$; $Var(y_{iQ}) = 661.8004$; $Var(y_{iI}) = \ 225{,}625$;

γ_i	Expected Utility from Q	Expected Utility from I	Difference in Exp Utility (Q – I)	Strategy Chosen
-2	403814	1.67E+08	-1.66E+08	I
-1.8	167257.1	4.49E+07	-4.47E+07	I
-1.6	69777.98	1.21E+07	-1.21E+07	I
-1.4	29347.95	3304351	-3275003	I
-1.2	12457.99	906031.5	-893574	I
-1	5344.91	250625	-245280	I
-0.8	2321.907	70092.2	-67770.3	I
-0.6	1023.804	19880.8	-18857	I
-0.4	459.7495	5745.722	-5285.97	I
-0.2	211.2946	1704.337	-1493.04	I
0	100.14	525	-424.86	I
0.2	49.56524	171.2837	-121.719	I
0.4	26.23586	61.29344	-35.0576	I
0.6	15.66311	25.78837	-10.1253	I
0.8	12.49968	15.41949	-2.91981	I
1.2	-2.00493	-1.77123	-0.23369	I
1.4	-0.40293	-0.34028	-0.06265	I
1.6	-0.10822	-0.0929	-0.01531	I
1.8	-0.03277	-0.02982	-0.00295	I
2	-0.01061	-0.0105	-0.00011	I
2.2	-0.00358	-0.00392	0.000332	Q
2.4	-0.00125	-0.00152	0.000269	Q
2.6	-0.00044	-0.0006	0.000158	Q
2.8	-0.00016	-0.00024	0.000083	Q
3	-5.9E-05	-0.0001	4.11E-05	Q
3.2	-2.2E-05	-4.2E-05	1.96E-05	Q
3.4	-8.26E-06	-1.7E-05	9.18E-06	Q
3.6	-3.13E-06	-7.36E-06	4.23E-06	Q
3.8	-1.19E-06	-3.12E-06	1.93E-06	Q
4	-4.57E-07	-1.33E-06	8.76E-07	Q

Calibration: $p_i = .38$; $y_{iF} = 133$; $\underline{y} = 80$; $\theta_i^G = \ .5$; $y_{iI}^G = 1{,}000$; $y_{iI}^B = 50$;

Simulation 4:

$E(y_{iF}) = 80$; $E(y_{iI}) = 525$; $Var(y_{iQ}) = 0$; $Var(y_{iI}) = 225{,}625$;

γ_i	Expected Utility from Q	Expected Utility from I	Difference in Exp Utility (Q – I)	Strategy Chosen
-2	170666.7	1.67E+08	-1.67E+08	I
-1.8	76119.15	4.49E+07	-4.48E+07	I
-1.6	34124.05	1.21E+07	-1.21E+07	I
-1.4	15388.8	3304351	-3288963	I
-1.2	6988.36	906031.5	-899043	I
-1	3200	250625	-247425	I
-0.8	1480.095	70092.2	-68612.1	I
-0.6	693.1448	19880.8	-19187.7	I
-0.4	329.76	5745.722	-5415.96	I
-0.2	160.1499	1704.337	-1544.19	I
0	80	525	-445	I
0.2	41.62766	171.2837	-129.656	I
0.4	23.10483	61.29344	-38.1886	I
0.6	14.427	25.78837	-11.3614	I
0.8	12.01124	15.41949	-3.40825	I
1.2	-2.08138	-1.77123	-0.31015	I
1.4	-0.43322	-0.34028	-0.09294	I
1.6	-0.12023	-0.0929	-0.02732	I
1.8	-0.03754	-0.02982	-0.00771	I
2	-0.0125	-0.0105	-0.002	I
2.2	-0.00434	-0.00392	-0.00042	I
2.4	-0.00155	-0.00152	-3.1E-05	I
2.6	-0.00056	-0.0006	3.91E-05	Q
2.8	-0.00021	-0.00024	3.55E-05	Q
3	-7.8E-05	-0.0001	2.21E-05	Q
3.2	-3E-05	-4.2E-05	1.21E-05	Q
3.4	-1.1E-05	-1.7E-05	6.16E-06	Q
3.6	-4.34E-06	-7.36E-06	3.02E-06	Q
3.8	-1.68E-06	-3.12E-06	1.45E-06	Q
4	-6.51E-07	-1.33E-06	6.82E-07	Q

Calibration: $p_i = 1$; $y_{iF} = 80$; $\underline{y} = 80$; $\theta_i^G = .5$; $y_{iI}^G = 1{,}000$; $y_{iI}^B = 50$;

Simulation 5:

$E(y_{iF}) = 106.6$; $E(y_{iI}) = 100$; $Var(y_{iQ}) = 1{,}154$; $Var(y_{iI}) = 2{,}500$;

γ_i	Expected Utility from Q	Expected Utility from I	Difference in Exp Utility (Q – I)	Strategy Chosen
-2	533313.3	583333.3	-50020	I
-1.8	215337.8	231449.7	-16111.9	I
-1.6	87630.26	92492.04	-4861.78	I
-1.4	35976.79	37274.36	-1297.57	I
-1.2	14919.51	15172.36	-252.851	I
-1	6259	6250	9	Q
-0.8	2661.373	2611.934	49.43896	Q
-0.6	1149.878	1110.918	38.95984	Q
-0.4	506.5739	482.9185	23.6554	Q
-0.2	228.6863	215.8111	12.87515	Q
0	106.6	100	6.599998	Q
0.2	51.96487	48.7062	3.258667	Q
0.4	27.12727	25.55876	1.568508	Q
0.6	15.99427	15.2529	0.741369	Q
0.8	12.6227	12.27699	0.345719	Q
1.2	-1.98794	-2.06101	0.073064	Q
1.4	-0.39662	-0.42986	0.033245	Q
1.6	-0.10587	-0.12092	0.015051	Q
1.8	-0.0319	-0.03868	0.006786	Q
2	-0.01028	-0.01333	0.00305	Q
2.2	-0.00346	-0.00483	0.001367	Q
2.4	-0.0012	-0.00181	0.000612	Q
2.6	-0.00043	-0.0007	0.000273	Q
2.8	-0.00015	-0.00028	0.000122	Q
3	-5.7E-05	-0.00011	5.42E-05	Q
3.2	-2.1E-05	-4.5E-05	2.41E-05	Q
3.4	-7.94E-06	-1.9E-05	1.07E-05	Q
3.6	-3.01E-06	-7.78E-06	4.77E-06	Q
3.8	-1.15E-06	-3.27E-06	2.12E-06	Q
4	-4.41E-07	-1.38E-06	9.42E-07	Q

Calibration: $p_i = .38$; $y_{iF} = 150$; $\underline{y} = 80$; $\theta_i^G = .5$; $y_{iI}^G = 150$; $y_{iI}^B = 50$;

Simulation 6:

$E(y_{iF}) = 149.38$; $E(y_{iI}) = 100$; $Var(y_{iQ}) = .2356$; $Var(y_{iI}) = 2500$;

γ_i	Expected Utility from Q	Expected Utility from I	Difference in Exp Utility (Q – I)	Strategy Chosen
-2	1111143	583333.3	527809.4	Q
-1.8	437393.6	231449.7	205943.9	Q
-1.6	173060	92492.04	80567.91	Q
-1.4	68880.83	37274.36	31606.46	Q
-1.2	27607.48	15172.36	12435.12	Q
-1	11157.31	6250	4907.31	Q
-0.8	4554.673	2611.934	1942.739	Q
-0.6	1882.566	1110.918	771.6478	Q
-0.4	790.4651	482.9185	307.5466	Q
-0.2	338.821	215.8111	123.0098	Q
0	149.38	100	49.38	Q
0.2	68.60308	48.7062	19.89688	Q
0.4	33.60653	25.55876	8.04777	Q
0.6	18.52066	15.2529	3.267759	Q
0.8	13.60906	12.27699	1.332075	Q
1.2	-1.83701	-2.06101	0.223995	Q
1.4	-0.33746	-0.42986	0.092399	Q
1.6	-0.08266	-0.12092	0.038265	Q
1.8	-0.02278	-0.03868	0.015908	Q
2	-0.00669	-0.01333	0.006639	Q
2.2	-0.00205	-0.00483	0.002781	Q
2.4	-0.00065	-0.00181	0.001169	Q
2.6	-0.00021	-0.0007	0.000493	Q
2.8	-6.8E-05	-0.00028	0.000209	Q
3	-2.2E-05	-0.00011	8.87E-05	Q
3.2	-7.48E-06	-4.5E-05	3.78E-05	Q
3.4	-2.52E-06	-1.9E-05	1.62E-05	Q
3.6	-8.55E-07	-7.78E-06	6.92E-06	Q
3.8	-2.92E-07	-3.27E-06	2.98E-06	Q
4	-1.00E-07	-1.38E-06	1.28E-06	Q

Calibration: $p_i = .38$; $y_{iF} = 150$; $\underline{y} = 149$; $\theta_i^G = .5$; $y_{iI}^G = 150$; $y_{iI}^B = 50$;

Table 4: Multinomial Logit of Occupational Categories (Spec 1 - 4)

	(1)		(2)		(3)		(4)	
	Formal	Unemployed	Formal	Unemployed	Formal	Unemployed	Formal	Unemployed
CRRA	.500 (.362)	1.028*** (.379)	.375 (.459)	.814** (.393)	.370 (.458)	.815** (.392)	.349 (.445)	.719** (.365)
Male†			1.082*** (.367)	.008 (.257)	1.088*** (.367)	.007 (.257)	1.048*** (.375)	-.231 (.265)
Age			-.003 (.014)	-.052*** (.015)	-.003 (.014)	-.053*** (.015)	-.179 (.116)	-.432*** (.081)
Age^2							.002 (.002)	.005*** (.001)
Years of Education			.347*** (.061)	.207*** (.036)	.348*** (.061)	.206*** (.036)	.223 (.161)	.262** (.120)
Years of Education^2							.006 (.008)	-.003 (.007)
Year=2005†					.047 (.297)	.070 (.203)	.018 (.302)	.008 (.220)
Year=2006†					.137 (.326)	-.027 (.218)	.106 (.337)	-.130 (.233)
Family Income per Capita								
Married†								
Number of Children								
Head of Household†								
Cons	-1.545*** (.241)	-.875*** (.205)	-5.126*** (.788)	-.874* (.519)	-5.203*** (.789)	-.887 (.542)	-1.72 (1.965)	4.929*** (1.361)
Obs.	586		585		585		585	

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Signif. Levels: * for p<.10, ** for p<.05, and *** for p<.01; Base Outcome: Informal
†Binary Variable (0,1);

Table 4 (Cont'ed): Multinomial Logit of Occupational Categories (Spec 5 to 7)

	(5)		(6)		(7) [††]	
	Formal	Unemployed	Formal	Unemployed	Formal	Unemployed
CRRA	.538 (.639)	.912* (.489)	.393 (.465)	.679* (.373)	.791 (.599)	.853* (.497)
Male[†]	1.264*** (.451)	-.055 (.353)	1.040** (.418)	-.066 (.297)	1.353*** (.472)	.101 (.396)
Age	-.116 (.106)	-.448*** (.091)	-.162 (.116)	-.370*** (.085)	-.074 (.102)	-.403*** (.109)
Age^2	.001 (.001)	.006*** (.001)	.002 (.002)	.005*** (.001)	.0010 (.001)	.005*** (.001)
Years of Education	.101 (.194)	.183 (.178)	.210 (.159)	.331** (.131)	.061 (.181)	.270 (.199)
Years of Education^2	.013 (.011)	.008 (.011)	.007 (.008)	-.006 (.008)	.015 (.011)	.004 (.011)
Year=2005[†]	.149 (.323)	.283 (.256)	.038 (.331)	-.044 (.238)	-.195 (.379)	.382 (.278)
Year=2006[†]			-.015 (.389)	-.152 (.275)		
Family Income per Capita	1.41e-06 (4.69e-06)	-.00005*** (8.83e-06)			1.27e-06 (4.64e-06)	-.951** (.467)
Married[†]			-.208 (.485)	-1.039*** (.346)	-.351 (.554)	.315** (.135)
Number of Children			-.028 (.152)	.263*** (.086)	-.018 (.196)	-.708* (.420)
Head of Household[†]			.157 (.390)	-.505 (.333)	-.003 (.481)	-.00005*** (9.30e-06)
Cons	-2.268 (1.991)	5.686*** (1.606)	-1.857 (1.967)	3.824*** (1.398)	-2.847 (1.862)	4.617** (1.807391)
Obs.	367		551		364	

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Signif. Levels: * for p<.10, ** for p<.05, and *** for p<.01; Base Outcome: Informal
[†]Binary Variable (0,1); [††]Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Table 5: Marginal Effects on the Probability of Queuing for a Formal Sector Job **[Men Only]**

	(1)	(2)	(3)	(4)	(5) ††	(6)	(7) ††
CRRA	.324* (.184)	.430** (.200)	.430** (.199)	.366* (.203)	.340** (.167)	.346 (.217)	.343** (.172)
Age		-.018*** (.005)	-.018*** (.005)	-.077*** (.022)	-.072*** (.019)	-.072*** (.025)	-.078*** (.027)
Age^2				.0009*** (.0003)	.0010*** (.0003)	.0009** (.0003)	.001*** (.0003)
Years of Education		.032* (.017)	.032* (.017)	.112* (.061)	.087** (.038)	.092* (.056)	.082** (.040)
Years of Education^2				-.004 (.003)	-.002 (.003)	-.004 (.003)	-.002 (.003)
Year=2005†			-.0009 (.066)	-.016 (.070)	-.005 (.072)	-.076 (.075)	.001 (.079)
Year=2006†			-.032 (.071)	-.071 (.070)		-.140** (.071)	
Family Income per Capita					-9.01e-06*** (2.08e-06)		-8.86e-06*** (2.14e-06)
Married†						-.268** (.121)	-.067 (.116)
Number of Children						.019 (.029)	-.003 (.035)
Head of Household†						.071 (.131)	.067 (.133)
Obs.	223	223	223	223	137	208	136

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01
†Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006)

Table 6: Marginal Effects on the Probability of Working in the Informal Sector **[Men Only]**

	(1)	**(2)**	**(3)**	**(4)**	**(5)** ††	**(6)**	**(7)** ††
CRRA	-.230 (.181)	-.410** (.180)	-.411** (.179)	-.325* (.185)	-.379* (.209)	-.321* (.186)	-.439** (.210)
Age		.012*** (.004)	.012*** (.005)	.076*** (.027)	.073** (.031)	.088*** (.033)	.095** (.044)
Age^2				-.0009** (.0004)	-.0010** (.0004)	-.001** (.0005)	-.001** (.0006)
Years of Education		-.070*** (.017)	-.070*** (.017)	-.074 (.053)	-.040 (.056)	-.067 (.055)	-.019 (.055)
Years of Education^2				.0002 (.003)	-.002 (.004)	-.0001 (.003)	-.004 (.004)
Year=2005†			-.014 (.069)	-.0008 (.074)	-.041 (.090)	.002 (.080)	-.065 (.099)
Year=2006†			-.009 (.080)	.018 (.087)		.036 (.093)	
Family Income per Capita					4.59e-06** (1.92e-06)		4.47e-06** (1.95e-06)
Married†						.035 (.160)	-.190 (.196)
Number of Children						.003 (.040)	.038 (.048)
Head of Household†						-.065 (.103)	-.065 (.143)
Obs.	223	223	223	223	137	208	137

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01
†Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Table 7: Marginal Effects on the Probability of Queuing for a Formal Sector Job **[2006 Only]**

	(1)	**(2)**	**(3)**	**(4)** ††	**(5)**	**(6)** ††
CRRA	.310*** (.111)	.285** (.117)	.237** (.100)	.211* (.111)	.191 (.122)	.144 (.111)
Male†		-.026 (.071)	-.099 (.071)	-.086 (.072)	-.084 (.079)	-.074 (.074)
Age		-.008** (.004)	-.080*** (.021)	-.076*** (.019)	-.075*** (.025)	-.066*** (.021)
Age^2			.001*** (.0003)	.0010*** (.0003)	.0010*** (.0003)	.0009*** (.0003)
Years of Education		.025** (.0101)	.052 (.039)	.020 (.042)	.055 (.042)	.028 (.040)
Years of Education^2			-.002 (.003)	.001 (.0030)	-.002 (.003)	.0008 (.003)
Family Income per Capita				-7.57e-06*** (1.50e-06)		-7.42e-06*** (1.49e-06)
Married†					-.108 (.098)	-.105 (.087)
Number of Children					.049* (.029)	.028 (.027)
Head of Household†					-.100 (.100)	-.111 (.076)
Obs.	207	207	207	187	184	184

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01
†Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Table 8: Marginal Effects on the Probability of Working in the Informal Sector **[2006 Only]**

	(1)	**(2)**	**(3)**	**(4)** ††	**(5)**	**(6)** ††
CRRA	-.302*** (.109)	-.270** (.114)	-.220** (.101)	.211* (.111)	-.212 (.129)	-.177 (.124)
Male†		-.093 (.081)	-.053 (.086)	-.086 (.072)	-.066 (.095)	-.098 (.102)
Age		.007** (.004)	.076*** (.023)	-.076*** (.019)	.078*** (.027)	.068*** (.024)
Age^2			-.0010*** (.0003)	.0010*** (.0003)	-.0010*** (.0004)	-.0009*** (.0003)
Years of Education		-.049*** (.011)	-.042 (.038)	.020 (.042)	-.047 (.043)	-.019 (.047)
Years of Education^2			-.0002 (.003)	.001 (.003)	-.0002 (.003)	-.003 (.003)
Family Income per Capita				-7.57e-06*** (1.50e-06)		6.42e-06*** (1.43e-06)
Married†					.099 (.106)	.091 (.100)
Number of Children					-.048 (.031)	-.034 (.029)
Head of Household†					.041 (.106)	.043 (.099)
Obs.	207	207	207	187	184	184

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01
†Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Table 9: Reduced Form Equation of CRRA

	Dependent Variable: CRRA						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Age	-0.018		-0.019	-0.020	-0.020	-0.020	-0.020
	(0.013)		(0.013)	(0.014)	(0.014)	(0.014)	(0.014)
Age^2	0.0002		0.0002	0.0002	0.0003	0.0003	0.0003
	(0.0002)		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Years of Education		-0.017	-0.026	-0.025	-0.013	-0.013	-0.012
		(0.020)	(0.021)	(0.021)	(0.021)	(0.0210)	(0.021)
Years of Education^2		0.002	0.002	0.002	0.002	0.002	0.002
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male			0.007	0.007	0.002	0.002	0.002
			(0.052)	(0.053)	(0.053)	(0.053)	(0.053)
Married				0.015	-0.013	-0.013	-0.012
				(0.059)	(0.062)	(0.062)	(0.063)
Number of Children					0.020	0.020	0.021
					(0.0184)	(0.018)	(0.019)
Head of Household					-0.125	-0.125	-0.125
					(0.060)**	(0.060)**	(0.061)**
Family Income per capita							0.00000
							(0.00000)
Constant	0.764	0.450	0.825	0.834	0.769	0.769	0.771
	(0.211)***	(0.074)***	(0.243)***	(0.255)***	(0.252)***	(0.252)***	(0.253)***
Observations	207	207	207	204	184	184	184
R-squared	0.01	0.01	0.03	0.03	0.07	0.07	0.07

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 10: Marginal Effects on the Probability of Queuing for a Formal Sector Job [**Weighting by CRRA s.e.**]

	(1)	**(2)**	**(3)**	**(4)**	**(5)** ††	**(6)**	**(7)** ††
CRRA	.132* (.069)	.118* (.063)	.128** (.060)	.140** (.064)	.116** (.053)	.122* (.063)	.109** (.044)
Male†		.018 (.076)	.012 (.074)	-.027 (.069)	-.024 (.050)	.023 (.076)	.002 (.050)
Age		-.003 (.004)	-.003 (.004)	-.065*** (.021)	-.028* (.016)	-.054** (.026)	-.021 (.019)
Age^2				.001*** (.000)	.0004* (.0002)	.001** (.0002)	.0002 (.0002)
Years of Education		.028*** (.010)	.030*** (.009)	.010 (.031)	-.003 (.027)	.019 (.032)	.001 (.029)
Years of Education^2				.002 (.002)	.003 (.002)	.001 (.002)	.003 (.002)
Year=2005†			-.088 (.085)	-.091 (.088)	.027 (.041)	-.124 (.094)	-.006 (.043)
Year=2006†			-.156** (.068)	-.171** (.068)		-.182*** (.069)	
Family Income per Capita					-7.75e-06*** (.00000)		-7.00e-06*** (.00000)
Married†						-.168** (.086)	-.107** (.053)
Number of Children						.039** (.019)	.043** (.023)
Head of Household†						-.087 (.082)	-.033 (.062)
Obs.	585	585	585	585	367	551	364

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01 †Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Table 11: Marginal Effects on the Probability of Working in the Informal Sector [**Weighting by CRRA s.e.**]

	(1)	(2)	(3)	(4)	(5) ††	(6)	(7) ††
CRRA	-.193*** (.068)	-.154** (.066)	-.164*** (.062)	-.181*** (.064)	-.159* (.057)	-.157** (.065)	-.158*** (.046)
Male†		-.068 (.091)	-.062 (.090)	-.026 (.087)	-.012 (.070)	-.096 (.092)	-.066 (.071)
Age		.002 (.004)	.002 (.004)	.059*** (.022)	.027 (.017)	.049* (.030)	.020 (.019)
Age^2				-.001** (.0003)	-.0003 (.0002)	-.001* (.0004)	-.0002 (.0002)
Years of Education		-.044*** (.011)	-.047*** (.010)	-.025 (.031)	-.017 (.036)	-.025 (.035)	-.008 (.030)
Years of Education^2				-.002 (.002)	-.003 (.003)	-.002 (.002)	-.003 (.002)
Year=2005†			.117 (.101)	.012 (.103)	-.034 (.047)	.153 (.107)	-.013 (.047)
Year=2006†			.169** (.079)	.182** (.078)		.218*** (.081)	
Family Income per Capita					7.14e-06*** (.00000)		6.41e-06*** (.00000)
Married†						.138 (.094)	.092 (.065)
Number of Children						-.022 (.029)	-.041* (.024)
Head of Household†						.101 (.094)	.047 (.073)
Obs.	585	585	585	585	367	551	364

Standard Errors in Parentheses; Correction for clustering at the level of single workers; Significance Levels: * for p<.10, ** for p<.05, and *** for p<.01 †Binary Variable (0,1); ††Family income is not reported in 2004 and we drop all the observations from the first wave in this specification (omitted year dummy = 2006).

Figure 1 – Capital in the Informal Sector

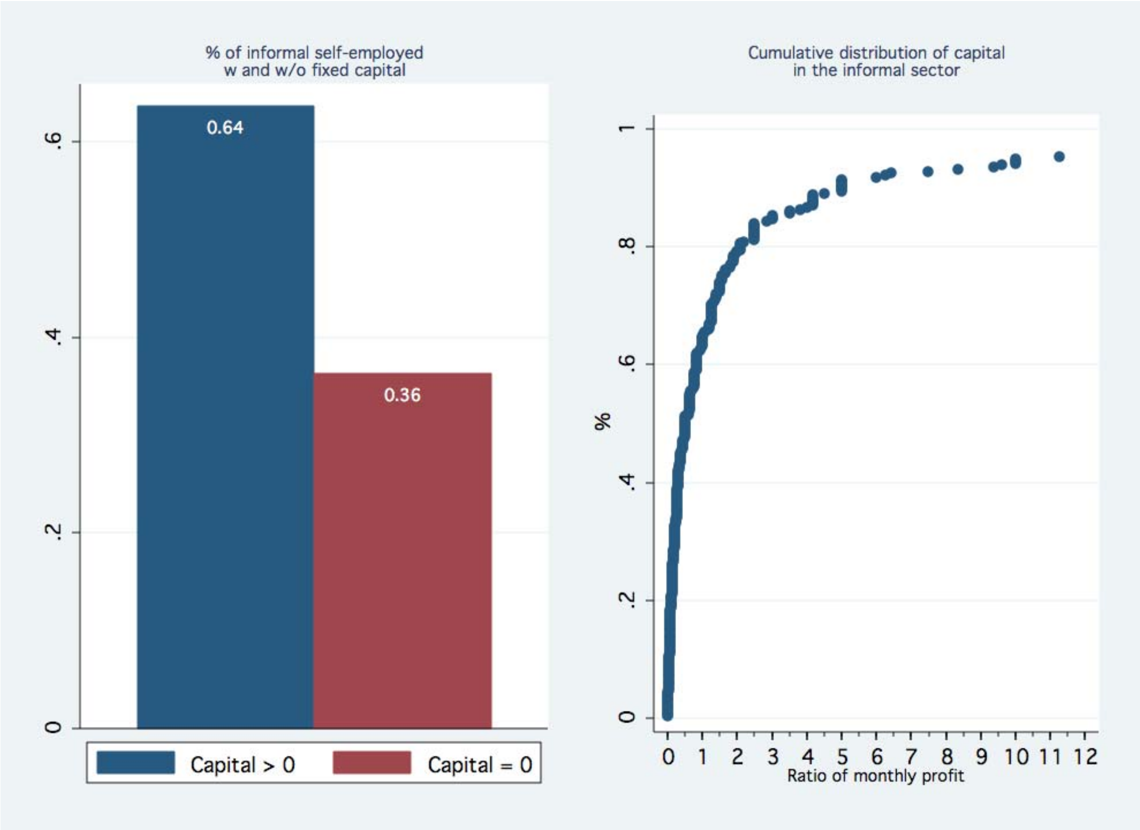


Figure 2 – Previous employment status of formal salaried workers

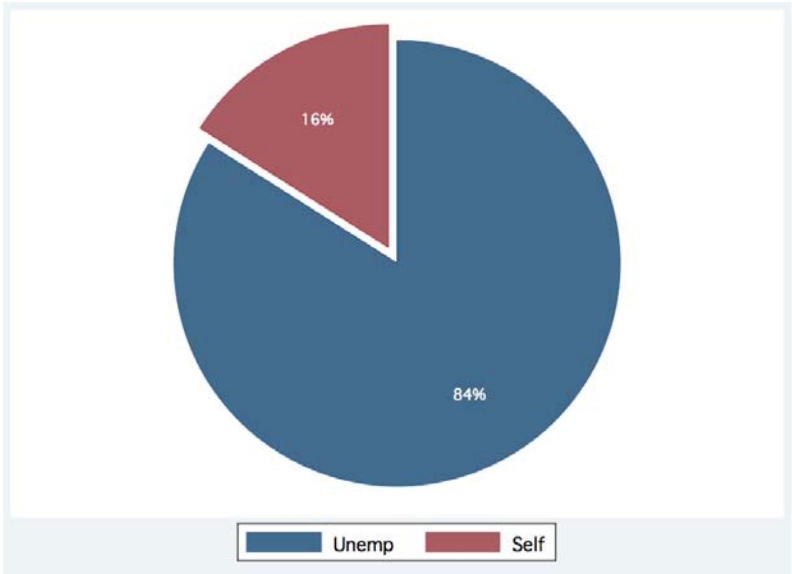
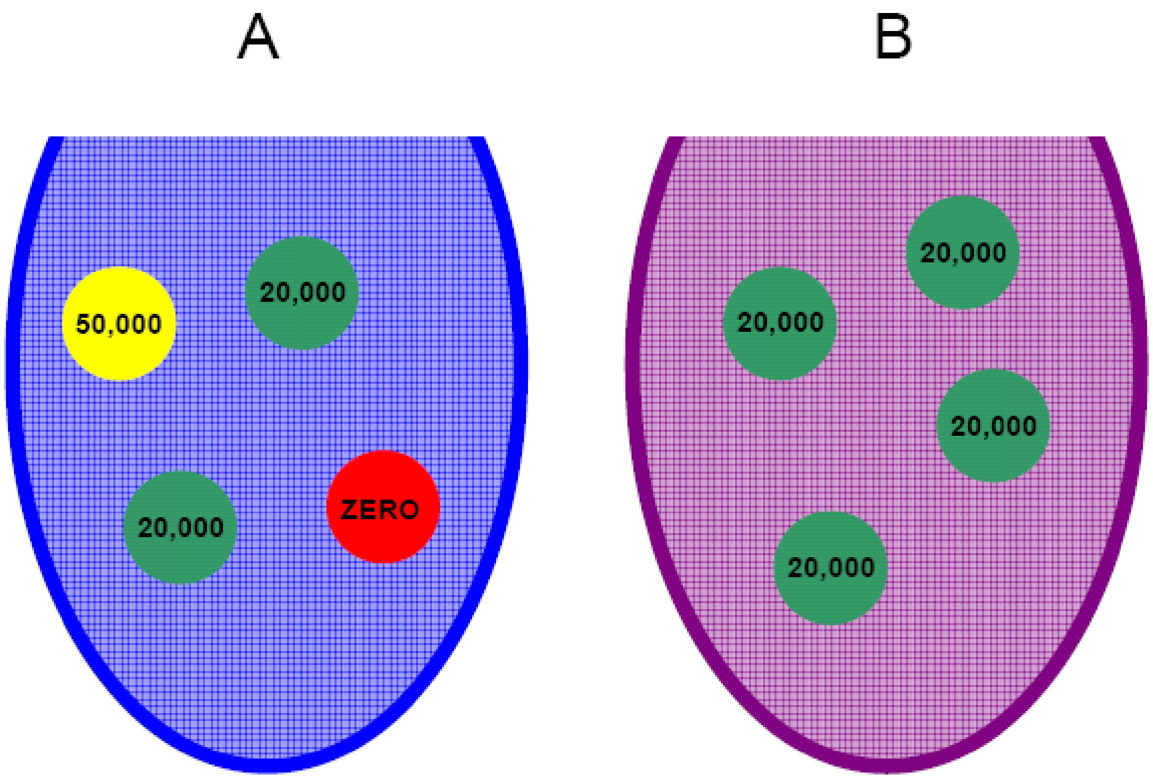


Figure 3 - Sample Game from the Set of Behavioural Experiments (1)

Task ID = MX1G

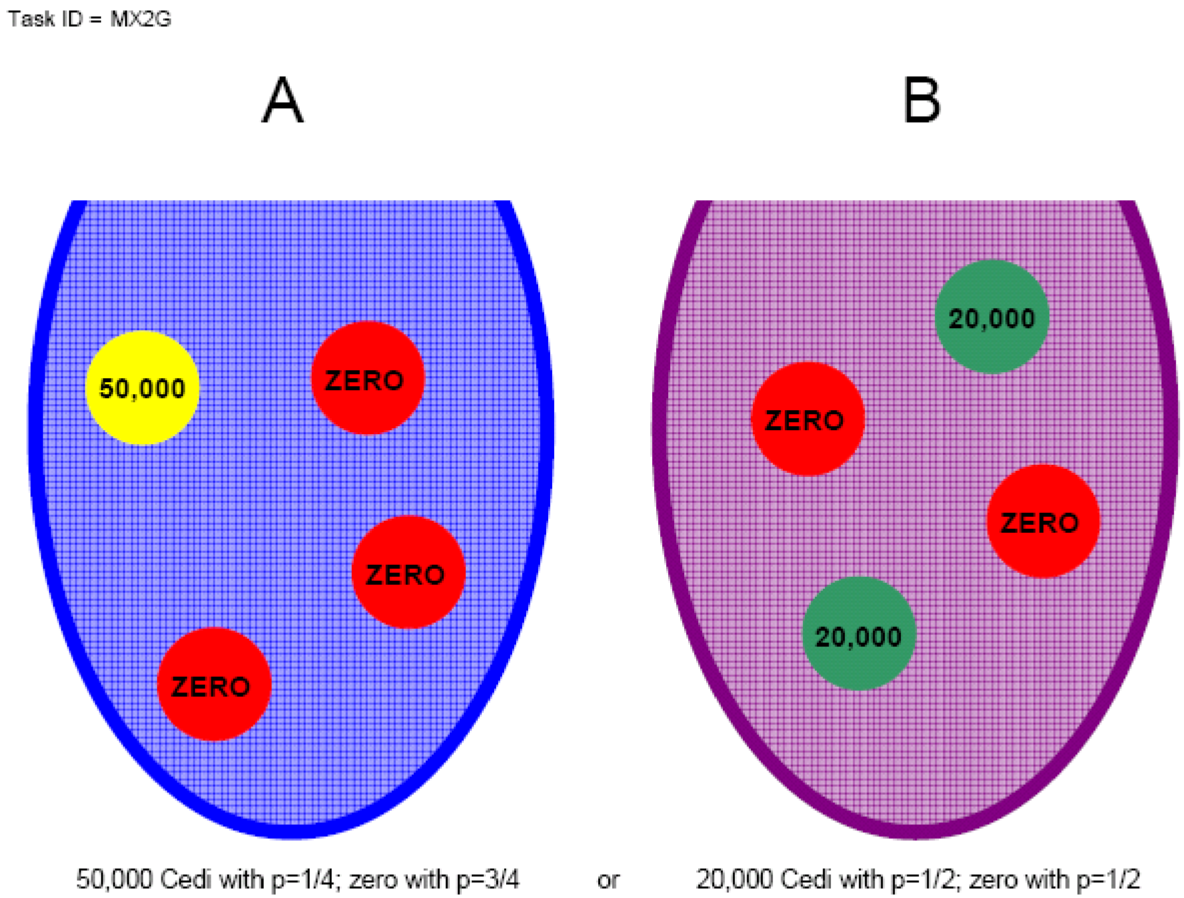


50,000 Cedi with p=1/4; zero with p=1/4; 20,000 Cedi with p=1/2

or

20,000 Cedi with p=1

Figure 4 - Sample Game from the Set of Behavioural Experiments (2)



Chapter 2:

Determinants of income in informal self-employment: new evidence from a long African panel

1 Introduction

Self-employment remains the predominant type of occupation in most parts of the developing world, including Sub-Saharan Africa, where the number of self-employed workers, mainly in the informal economy, has been rising in recent decades (Kingdon, Sandefur, and Teal (2006)). A positive view of this phenomenon would say that the progressive relaxation of credit constraints has allowed an increasing number of workers to reap the benefits from profitable investment opportunities. A negative view, on the other hand, would argue that a growing informal economy has resulted from the failure to create a sufficiently large industrial sector that could provide workers with desirable wage-opportunities. The central empirical issue in assessing these alternative views is the consistent estimation of the returns to workers' *productive assets*: *physical capital, labour and human capital* in self-employment. Can we successfully model the income generating process in informal self-employment and can we successfully measure the outcomes of interest in a context of widespread lack of numeracy and literacy skills? Are returns to productive assets high enough to support the optimists' argument? And how do these returns compare to the returns to the same assets in alternative occupations (e.g.

wage-employment)? These are some of the questions I will attempt to answer in this chapter.

The first challenge I face will be to specify an estimable *model* of the income generating process in informal self employment. In doing so, I will attempt to abridge the gap between the analysis of workers' earnings and the study of firms' production through a specification that accommodates both human capital and physical assets. Second, I will propose an *identification strategy* that will allow for consistent estimation of returns to physical capital, human capital and labour using a newly collected 'long' panel dataset from urban Ghana, gathered by the Centre for the Study of African Economies under my own supervision. The survey was conducted between 2004 and 2009 at yearly intervals and is now sufficiently long to allow the use of complex panel estimators, such as the Anderson and Hsiao (1982) instrumental variable estimator and the Arellano and Bond (1991) Difference-GMM estimator, which should enable me to purge my estimates from both time-invariant and time-varying sources of endogeneity in the choice of factors of production.

After estimating returns to capital and assessing the magnitude of the returns to different factors of production, I will dig deeper into the mechanics of the production process and I will attempt to explicitly control for the endogenous choice of capital and labour intensive production technologies by means of two-stage estimators, modeled after the the work of Heckman (1979) and Dubin and McFadden (1984). As a final step, I will attempt an analysis of the *shape* of the returns to capital over the range of capital stocks observed, with the aim to assess whether there exist any *non-convexities* in the production technology that may justify the existence of poverty traps at low capital levels. In the course of the analysis I will provide a precise definition of the type of non-convexities that may potentially give rise to such traps.

Given the computational intensity of my empirical approach and the scarcity of long panels in the African context, the results in this paper constitute an important contribution to the discussion on returns to productive assets in developing countries, well beyond the African context. As a growing literature has shown, in fact, one fundamental feature of the urban labour markets of Sub-Saharan Africa is closely replicated in latin America (e.g. McKenzie and Woodruff (2008) on Mexico), parts of Asia (e.g. de Mel, McKenzie, and Woodruff (2008) on Sri-Lanka) and, to some extent, parts of rural Africa (e.g. Udry and Anagol (2006) on Ghana), namely that a growing population of informal businesses operate at a very small production scale, generally without hired labour, and reap very high (and heterogeneous) returns to minimal capital investments (Banerjee and Duflo (2005)). This paper aims to complement this literature, also from a methodological standpoint. By means of survey data collected over a longer time-horizon than it is usually available, this paper will offer *observational* estimates of the returns to capital (including human capital) and labour that can be compared to estimates obtained through (increasingly common) Randomized Control Trials (RCTs) (see Fafchamps, McKenzie, Quinn, and Woodruff (2011) for a prominent example of a randomized capital drop intervention in Ghana). Such experimental evidence often faces criticisms of limited external validity and the results in this paper, obtained for a representative sample of urban self-employed workers, may constitute a useful benchmark.

The main finding in this chapter is that physical capital and labour market experience play the strongest role in the income generating process for the self-employed. On the other hand, the share of value-added attributed to labour is considerably smaller and imprecisely estimated. In addition, formal education does not appear to play a role in enhancing the productivity of the self-employed in the informal economy, leading me to conclude that learning on the job is a more important dimension of human capital than formal schooling. When I control for the endogenous choice of capital intensive production technologies using a first stage selection model, I find that my core results do not change significantly, although the limitations of my instruments for selection curtail the robustness of this conclusion. Finally, when I explore the shape of the production function over the range of capital observed, I find a *highly concave* technology. Marginal returns to investment are high at very low capital levels (it is not uncommon to find businesses that operate with capital value equal to 10 (real) USD or less), and they decrease rapidly.

The implications of these findings are at least two-fold. On the one hand, coupled with evidence of low entry costs among our survey respondents, my results overall point against the existence of non-convexities in the production technology driven by minimum-scale requirements. On the other hand, the real income gains that result from high marginal returns are modest, as they are the product of very low capital stocks. Hence, whether high marginal returns to investment will eventually translate into firm-growth (as firms re-invest their profits and attempt to *bootstrap* themselves out of poverty) remains an open question, the answer to which will partly depend on being able to study workers' inter-temporal preferences over saving and consumption, in relation to the degree of asset integration between businesses and households. Finally, the results presented in this chapter link directly to my analysis in chapter 1, where the determinants of the income generating process that drives workers' sectoral choice were intentionally left unexplained to concentrate on the allocation process. Clearly, the two mechanisms cannot be uncoupled and my aim in this chapter is to draw the link via a thorough investigation into the earnings of the self-employed, who constitute the majority of 'informal sector workers' as defined earlier in the thesis.

The chapter is structured as follows. Section 2 outlines the model of the income generating process. Section 3 describes the dataset and discusses my choice of specific measures of the capital stock, which will be central to the analysis. In section 4, I outline the results and discuss their potential interpretations. In Section 5 I test the robustness of the results against the possibility of endogeneity in the choice of the production technology. Section 6 explores the shape of the production function in greater detail, searching for potential evidence of non-convexities in the production set. Section 7 concludes.

2 Identification of the income-generating technology

Let the income of a self-employed worker be governed by the following process, based on a standard Cobb-Douglas production function. My choice of the model comes from the view that despite the small size of the enterprises in the sample (often reducing to a single worker), earnings in self-employment ought to be investigated using the analytical tools generally deployed to study firms' output (production functions), rather than individual earnings (earnings regressions). Like larger formal firms, one-worker enterprises generate 'value-added', transforming raw-materials into final products via a multi-factor technology. Crucially, however, in addition to capital and labour, this technology will be augmented by the human capital of the entrepreneur (education and labour market experience), whose effects are important to draw conclusions on the returns to a worker from choosing self-employment (presumably as an alternative to potential wage-opportunities).

Hence, let

$$Y_{it} = A(H_{it}, X_{it}, u_{it}) K_{it}^{\alpha} L_{it}^{\beta} \tag{1}$$

where Y_{it} denotes the output of firm i at time t, measured as 'value added'[1], K_i is the stock of physical capital, L_i denotes units of labour (measured as total hours of work, including the entrepreneur's), A_i captures firm's productivity, which I assume is a function of the entrepreneur's stock of human capital (H_{it}) (proxied by the number of years spent in formal education), labour market experience (proxied by age) and other individual characteristics such as gender (included in X_{it}). u_{it} is an unobserved component of productivity, which can be further decomposed into

[1] This choice follows the most common approaches in the literature (see Basu and Fernald (1995, 1997) and Eberhardt and Helmers (2010)) for a review

$$u_{it} = \gamma_0 + \delta_t + \eta_i + \omega_{it} \tag{2}$$

where γ_0 denotes average productivity across firms, δ_t captures period specific effects that are common across firms, η_i is a time-invariant firm-specific fixed effect and ω_{it} contains shocks to productivity that are period and firm-specific. Log-linearisation transforms the above production technology into the following empirical analog:[2]

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma H_i + \theta X_{it} + (\gamma_0 + \delta_t + \eta_i + \omega_{it}) \tag{4}$$

where lower case letters denote log-values.

Consistent estimation of (4) poses a number of challenges. First, the optimal choice of capital and labour by the firm is likely to depend on the unobservable components of productivity. In fact, it could easily be shown that the marginal product of capital and labour are a function of such unobservables. Hence, depending on the speed at which inputs can be adjusted, we can expect that they will be either (i) a function of only the time-invariant heterogeneity (η_i) or (ii) a function of both time-varying and time-invariant heterogeneity (δ_t, η_i, ω_{it}). As it is well-known, under either of these circumstances, OLS estimates

[2] This specification implicitly assumes:

$$A(H_{it}, X_{it,} u_{it}) = e^{\gamma H_i + \theta X_{it} + (\gamma_0 + \delta_t + \eta_i + \omega_{it})} \tag{3}$$

will be biased, as either of the following assumptions may not hold:

$$E[k_{it}u_{it}] = 0; \quad E[l_{it}u_{it}] = 0 \tag{OLS}$$

My identification strategy will first control for individual fixed effects by means of within group transformations (WG) and differencing (DIFF), which are both feasible given the panel structure of the data. However, even the following less restrictive identifying assumptions, necessary for WG and DIFF estimation to be unbiased, may fail to hold if time-varying heterogeneity plays a role in the choice of inputs:

$$E[k_{it}(\delta_t + \omega_{it})] = 0; \quad E[l_{it}(\delta_t + \omega_{it})] = 0 \tag{WG/DIFF}$$

As one can reasonably hypothesise sufficient flexibility in input choice over time, we believe this is a legitimate concern.

In what follows time-dependent shocks that are common across firms (δ_t) will be controlled for by means of time-dummies. The only remaining source of time-varying unobserved variation will therefore be ω_{it}, which will take center stage in the remainder of the identification strategy.

Before delving into such strategy, however, it should be remarked that the second challenge posed by the estimation of my production function comes from the fact that the optimal level of human capital accumulation chosen by the individual may also depend on his/her unobserved productivity. For instance, more productive (able) individuals may also have lower costs of school attendance and therefore acquire higher levels of human capital. This would bias the OLS coefficients upwards. Since human capital is time-invariant in our dataset (workers accumulate formal education in their youth and once they enter the labour market that capital stock remains fixed; and we do not allow for depreciation in human capital), panel techniques such as Differencing and WG transformations are not suitable in their simplest form to deal with this problem. In order to remedy this limitation, I could employ the Blundell and Bond (1998) System-GMM estimator, as well as complementary Instrumental Variable techniques that use instruments external to the income model (such as distance from schooling during childhood) to ascertain the true returns to schooling. Due to data limitations the results I obtain when applying System-GMM are still unstable, while the lack of reliable external instruments in our dataset makes the second option unavailable to me. However, it should soon be possible to repeat this analysis with additional data gathered during the most recent waves of the GHUPS, which might strengthen the estimation.

Going back to time-varying unobservables (ω_{it}), dealing with them constitutes the most challenging part of the identification strategy. Exploiting the length of the panel, I will base my procedure on a series of estimators that have been extensively used in the literature on the empirical estimation of production functions: these are the Anderson and Hsiao (1982) instrumental variable estimator, the Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991) (Difference GMM) estimator. A detailed discussion of these techniques is provided in the appendix. In the absence of reliable *external* instruments for input choices, these estimators provide a framework to use lags of the endogenous variables as instruments, after applying the first-difference transformation that controls for time-invariant heterogeneity. Making different identifying assumptions allows us to use different lag-lengths as instruments. Namely, one option is to assume that inputs are *pre-determined*, in the sense that input choice is affected by past, but not current productivity shocks.

$$E[k_{is}\omega_{it}] = 0; \quad E[l_{is}\omega_{it}] = 0 \quad \forall \ t \geq s \tag{GMM1}$$

Alternatively, one can assume that input choices are *endogenous*, in the sense that they are affected by both past and current productivity shocks.

$$E[k_{is}\omega_{it}] = 0; \quad E[l_{is}\omega_{it}] = 0 \quad \forall \ t > s \tag{GMM2}$$

In the following analysis I will first assume pre-determinedness of both capital and labour, and then relax the latter assumption, allowing labour inputs, which should be flexibly adjusted in the absence of formal contracts, to become endogenous. On the other hand, due to the likely presence of credit-constraints, capital stocks should be less flexible, which justifies the pre-determinedness assumption in the case of capital inputs.

The empirical methodologies I have chosen constitute one of at least two main subsets of estimators that rely on observational data to estimate production functions. The most prominent alternative would have been to employ the estimators proposed by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2000) (LP), but they have both posed challenges that I have been unable to overcome with the available data. Most notably, the GHUPS dataset lacks a reliable investment series for the years I have analysed, which impairs the use of the OP estimator. Data on variable inputs is similarly limited and imprecise, and using such data to proxy unobserved productivity shocks as suggested by LP would partly clash with the definition of capital I discuss in the next section. Moreover, the LP estimator suffers from the well-known critique by Ackerberg, Caves, and Frazer (2006) and Bond and Soderbom (2005) on the collinearity between variable inputs and unobserved shocks, which impairs identification when input prices are not observed and need to be assumed constant across firms (as in our case).[3]

3 Data

The production model will be estimated using data from the Ghana Household Urban Panel Survey ('GHUPS'), conducted by the Centre for the Study of African Economies (CSAE) at the University of Oxford under my own supervision. The survey was launched in 2004 and at the time of writing it spans 6 years, an unusual length for panel data-sets in developing countries. The GUHPS covers four cities: Accra, Kumasi, Takoradi and Cape Coast. Respondents were drawn by stratified random sampling of urban households from the Population and Housing Census of 2000. The survey was designed to cover all household members of working age at the time of the interview. After the first wave the sample expanded by incorporating new members of the original households, as well as new households formed by individuals who had left their original household and were tracked to their new locations.

[3] Improvements in the precision of the data series on investment, and the availability of firm-specific price data for variable inputs might enable me to attempt a structural estimation of the model in the future.

The GUHPS contains a wide range of workers' characteristics and, most importantly, a wide range of work-related variables, such as business size, location and, crucially, capital data. As it was already pointed out in the previous chapter, it is important to underline that the GUHPS overcomes important measurement issues, which have often raised skepticism about the possibility of measuring the earnings and, more generally, the business characteristics of informal self-employed workers with any degree of precision. These concerns are not unreasonable, given that informal businesses often lack written book-keeping and are run by workers with poor literacy and numeracy, who may find it hard to produce the figures they are asked to provide. Thanks to intensive enumerator training and to the use of portable computers (PDAs) in the data collection, it was possible to perform a number of *live* consistency checks during the interviews, which increased precision. However, in line with recent evidence on the relative merits of different methods to measure informal enterprises' earnings (de Mel, McKenzie, and Woodruff (2009) and Fafchamps, McKenzie, Quinn, and Woodruff (2010)), my preferred earnings measure comes from a simple question that asked respondents to report their overall profit. Table 1 reports key summary statistics for income (in the form of both profits and value added) and capital (in the form of capital stock (K) and working capital (R)) for all the respondents in the sample, pooling across survey waves, which should help the reader get a sense of the average scale of production in the data.

Table 1: Summary Statistics - Income per month and Value of Capital - 1997USD

Variable	Mean	Median	N
Value Added	135.21	71.36	1304
Profit	129.79	66.58	1304
K	212.98	27.09	1304
R	809.33	102.67	1304
$K > 0$	0.76		1304

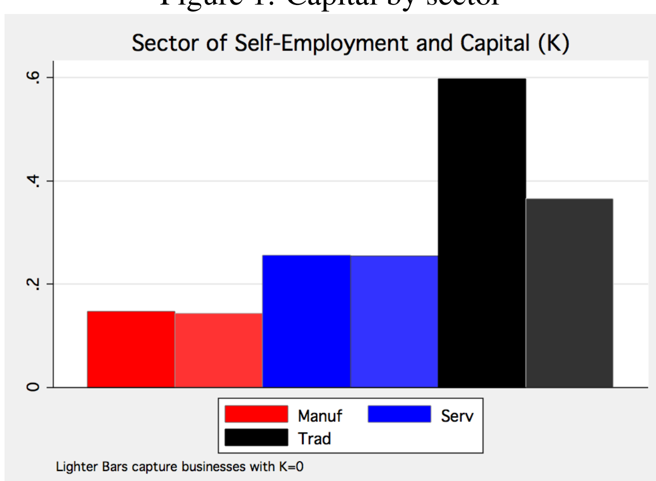
3.1 Labour, Human Capital and Physical Capital

Given the central role that workers' productive assets play in the analysis, I shall briefly discuss how these are measured (the appendix provides further details on how other key variables are constructed). Labour enters the production function in the form of *total hours* of work employed in the business. This includes both the hours worked by the entrepreneur/business owner and the hours worked by any hired labourers. The latter, however, is not observed in the data. To overcome this limitation, I generate total hours of hired labour as the product of the total number of hired labourers reported by the entrepreneur times 40 hours per week (an approximation based on evidence from wage-workers in the GHUPS dataset) and sum it to the hours worked by the entrepreneur to obtain L. Human capital is measured as the workers' number of *years in formal education*, which is directly observed in the data, and by his/her *labour market experience* (proxied by age).[4] The most reliable measure of physical capital available is the *total value of tools and equipment* employed in the business. Interestingly, this is reported to be 0 for 24% of the sample (see Table 1)

[4] As average education levels in the sample of self-employed are rather low, this approach should not diverge substantially from a commonly used alternative: calculating labour market experience as *(Age - Education - 5)*.

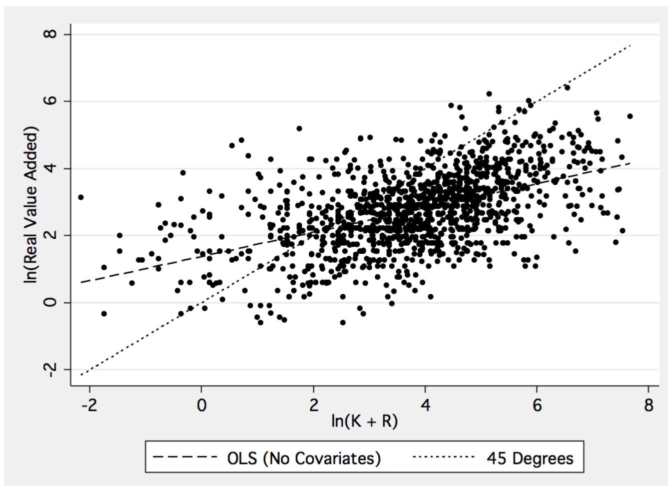
and, almost exclusively, by traders (see figure 1). Reflecting upon the nature of micro trading businesses, such as the unprocessed-food sellers who are one of the most common categories in our sample, this feature of the data does not seem implausible. Such workers are in many cases unlikely to require any equipment or tools for their income-generating technology, other than the *merchandise* they buy and re-sell. It results, therefore, that limiting the analysis to one of capital stocks in the strict sense of tools/equipment/machinery used in production, would overlook an important part of the picture. My approach, therefore, is to construct the capital measure as the sum of total value of tools and equipment (K) and of working capital (R) - the amount of money invested in business merchandise and raw-materials, which is also recorded in the data. This approach is further supported by the empirical observation that respondents who borrow for their businesses (e.g. from microfinance institutions) largely do so in order to finance the purchase of merchandise.

Figure 1: Capital by sector



As figure 2 shows, there appears to be a clear and stable relationship between the (real) value of total capital (K+ R) and earnings in the pooled dataset. In addition, the 3-dimensional graph in figure 4 suggests that the relationship holds when we add labour into the picture. In estimating the production model outlined above, we will test the strength of this relationship in a multivariate setting that attempts to control for endogeneity in input choice.

Figure 2: Capital and Value Added



4 Results

The results from estimating the production function (4) are reported in Table 2. First, my estimates show a strong and statistically significant effect of physical capital on value added. In line with my priors, a simple OLS regression delivers an upward-biased coefficient, presumably the result of unobserved ability or productivity shocks driving the choice of capital by the entrepreneur. Once individual fixed effects are controlled for (WG), the bias is significantly reduced (the coefficient drops from .27 to .196), but not entirely eliminated. Indeed, after first differencing and instrumenting the first-differences by means of lagged values of K and L (col. 6-8), we find that the coefficient drops further, but it stabilises in a relatively small range between the Anderson-Hiao (AH) estimate (.147) and the Arellano-Bond Diff-GMM estimate (.18). The Arellano-Bond GMM results (in col. 8) are robust to concerns of serial correlation in the error terms (the AR(2) test results show that we can reject the null of serial correlation) and pass the Hansen Test of over-identifying restrictions. It is well-known that the validity of this test has been subject to severe criticism in contexts where the use of lags leads to proliferation of instruments (Roodman (2009a,b)). Unfortunately, the literature does not offer a clear-cut rule to judge whether an instrument set is too large, except for the intuitive rule of thumb that when the instrument set approaches N, the model is invalid (Roodman (2009a). As reported at the bottom of table 2, the instrument set I use comprises 22 instruments. Therefore, I believe that with a dataset of over 400 respondents, my instrument set is 'safely' small.

The results on the role of labour in the production technology are less clear-cut. The OLS coefficient is .20, while the WG coefficient is .11. When I use lags as instruments, I am unable to pin down an estimate with sufficient precision. A reminder of how the labour variable is constructed may help clarify this result. In my estimation, labour is the sum of the hours worked by the entrepreneur and by his/her employees, where the latter is obtained by multiplying the number of employees by a standard number of weekly hours of work (set at 40). It follows that identification of the coefficient on L is achieved through variation in the number of hours worked by the entrepreneur and by the number of employees working in the business. Given that the large majority of workers in our sample does not employ any additional labour (other than themselves), the degree of variation in the data may therefore be insufficient for precise identification. This concern is particularly strong given that my panel estimators crucially achieve identification through within-firm variation over time, which is unlikely to be significant. Despite lack of precision in the GMM regressions, the OLS and WG results appear to suggest that returns to labour in micro-enterprises are considerably lower than returns to capital.

The weak identification of returns to labour may also be due to the fact that labour inputs rapidly adjust to capital. This suggests that if we aim to separately identify the effects of capital and labour in a GMM context we should consider relaxing the assumption of pre-determinedness of labour and let it be endogenous, while retaining the assumption that capital is pre-determined. When I attempt to do that, however, I do not attain any increase in precision (see table 10 in the Appendix). In fact, possibly due to the reduction in sample size that results from using deeper lags and due to the fact that deeper lags may constitute weaker instruments for the endogenous labour inputs, precision drops. An alternative strategy would have been to include additional instruments for labour. For instance, one could think of instruments that are motivated by the theory of household models with incomplete or imperfect markets, which shows that the shadow cost of labour varies depending on labour supply conditions in the household - such as changes in family size. However, when I tried to use the measures of time-varying household characteristics available in the dataset as additional instruments, the estimation results remained highly imprecise, a result most likely due to limitations in the data series on household characteristics (which contained many missing observations and caused a considerable reduction in sample size).[5]

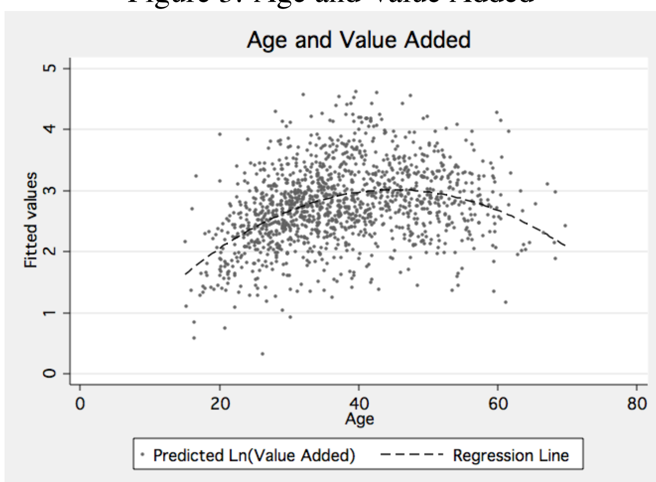
Given the nature of the businesses that prevail in the Ghanaian urban economy, the conclusion that capital (including 'working capital'), is a far more valuable factor of production than labour would seem plausible. Small trading businesses (e.g. food and clothes sellers) are indeed unlikely to display high returns to labour, since, with the exception of transportation (which may be sporadic or outsourced to the suppliers), there appear to be few 'processing tasks' on a sufficiently large scale that substantial labour inputs could be useful for. In fact, most frequently, the task of selling the goods is effectively fulfilled by the firm owner on his/her own (especially if the business is in a fixed location, like a market stall, where the entrepreneur can easily supervise its operations); and we could hypothesise that until a certain scale is reached (e.g. a formal shop, which would be observed rarely in our sample), the marginal product of additional labour may indeed be very small. Yet, inefficiently high levels of labour may be hired as a result of the availability of household workers (whose prospect would alternatively be unemployment) inducing the entrepreneur to over-employ despite very low marginal products.

Turning to Human Capital, the results show a clear and strong effect of labour market experience (proxied by age). The OLS regressions show a highly con-

[5] The results are not reported for the sake of conciseness and since they would not add much contents to the discussion. The available instruments were *family size* and *number of children*, which are not recorded in all waves, causing a drop in sample size and making it especially difficult to exploit their variation over time for identification.

cave age-earnings profile. After transforming the data to account for fixed-effects, I am no longer able to identify the linear effect of age separately from the average time-effect common across people (since age is assumed to change by exactly 1 between two waves for all the individuals in the sample), but we are still able to capture the concavity of the effect. Figure 3 plots the age-earning profile implied by the OLS regression.

Figure 3: Age and Value Added



Perhaps more interestingly, the OLS results show no significant relationship between formal education and the earnings of the self-employed (neither a linear nor a quadratic one). This finding is not robust to the potential endogeneity in human capital accumulation, but the striking weakness of the coefficient strongly suggests that formal schooling may indeed be playing only a marginal role in the production technology. A vast literature has attempted to estimate the causal impact of education on earnings (see Card (1999) for an extensive survey), and the most common point of departure has been the hypothesis that returns to education are biased *upwards* in OLS by the omission of unobserved ability. Under such hypothesis, which many previous studies have corroborated, the results I obtain are clearly compelling, as they do not detect any effect of education even when ability is not controlled for. One potential explanation may be the absence of 'sheepskin' effects among the self-employed in the informal sector. This result will seek confirmation when new waves of data become available,[6] but taken at face value it tells us that in an economy where the informal sector is quickly expanding and absorbs an increasing share of the population, formal schooling has not provided workers with the skills they require to increase their productivity. Such finding would support the hypothesis that education acts primarily as a signal in the Ghanaian labour market, allowing people to access desirable employment opportunities in the formal economy (e.g. public sector), while it does not add much to their productivity. An alternative explanation may be that formal education provides the *wrong* set of skills, which are not applicable in informal self-employment. In line with this argument, the result produces an interesting contrast with the existing evidence on the high value of apprenticeships for informal sector workers in Ghana (Monk, Sandefur, and Teal (2008)).

When I interact productive assets with gender, I find that a larger proportion of value added is attributed to capital among women than among men, while the opposite is true for labour. Labour market experience and education, on the other hand, do not appear to have significantly different coefficients among

[6] When I shall be able to employ the Blundell-Bond (1998) System-GMM estimator more effectively.

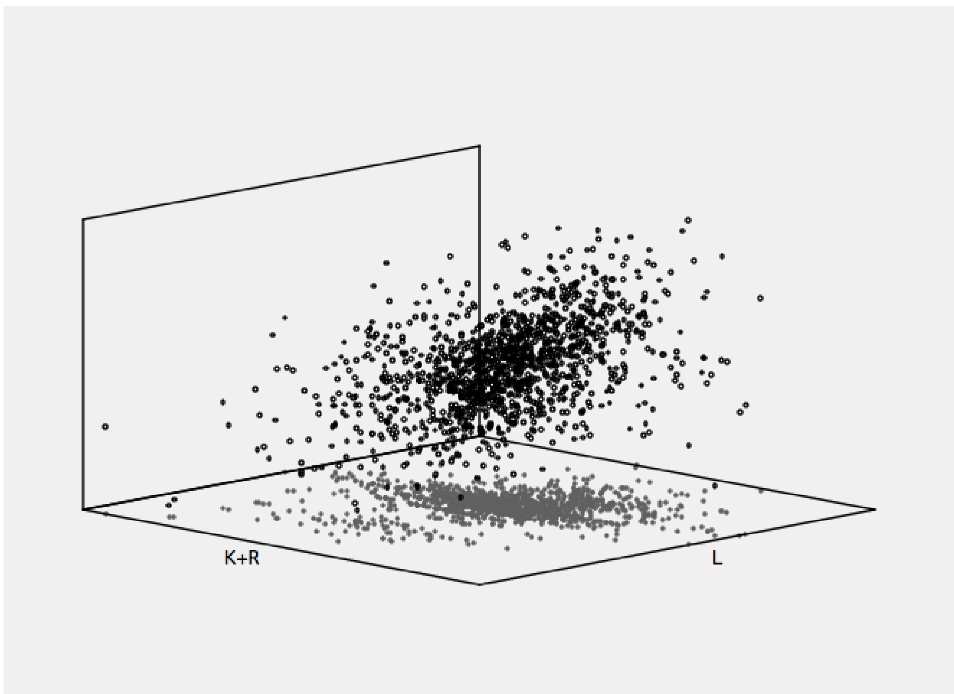
men and women. These results are suggestive of a number of hypotheses on the differences in occupational outcomes of male and female workers. One of them is that gender may be driving selection into different industries and sectors (e.g. female workers are more likely to be traders and less likely to work in manufacturing businesses) and consequently it may be associated with different returns to capital. Similarly, differential returns to capital may simply be driven by differences in labour market attachment of male and female workers. Though I acknowledge the shortcomings of the approach I have proposed, in particular those that may derive from unaccounted gender-driven selection, due to data limitations I am currently unable to split the analysis by gender. Similarly, it would be interesting to refine the estimation splitting it up by sector, to be able to infer potential differences in returns to capital across different industries, but the results I obtained upon pursuing that strategy suffered from a dramatic loss in precision due to the significant reduction in sample size. As new data becomes available, it will be interesting to attempt both these extensions.

Finally, despite the imprecision in the estimated labour coefficient, my results appear to indicate overall *decreasing returns to scale* with respect to capital and labour, a finding that deserves some discussion. The direct implication of decreasing returns to scale is that if the business attempted to expand its capital and labour inputs, the resulting increase in value added would be less than proportional. One possible explanation could be the existence of additional factors of production that are either unobserved or cannot be included in our regressions due to poor data quality, but are implicitly held constant when performing simple comparative statics.[7] Given the nature of the businesses that prevail in the Ghanaian economy (small trading enterprises), important omitted factors may be *location* and *information*; especially if these are the factors that crucially allow traders to gain from arbitrage of unprocessed goods across markets (e.g. whole-sale to retail, countryside to city).[8] Further refinements in the dataset and the availability of additional variables to proxy location, business knowledge and informational advantages may enable us to test some of these hypotheses in the future.

[7] This hypothesis would seem to be corroborated by the fact that, as in most empirical studies on production technologies, a large share of variation in the outcome variable is unexplained by our model (R2 in OLS is about .3). However, to the extent that the use of such potentially omitted factors is complementary to and hence positively correlated with capital and labour inputs, our estimated coefficients should be biased upwards, while actually, both in the case of capital and especially in the case of labour my concern is of a potentially negative bias.

[8] Factors of this kind, also including more general business practices and 'know-how', would commonly be considered part of overall TFP as captured by the estimation residual. Trying to capture them more explicitly and including them among the factors of production would be an interesting extension. The empirical challenge would be finding good proxies. Attempting to measure business practices will be among the goals of the next waves of GHUPS.

Figure 4: Income, Capital and Labour



NOTE: The chart is a 3-D plot of log-value-added (on the vertical axis) against the log value $K + R$ and log number of hours (L) (on the horizontal axes).

Table 2: Determinants of value-added in informal self-employment

	OLS	OLS2	OLSINT	WG	FD	AH	HNR	DIFF-2S
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K+R	.272	.272	.341	.196	.183	.147	.171	.180
	(.017)***	(.017)***	(.021)***	(.026)***	(.028)***	(.061)**	(.051)***	(.063)***
L	.197	.197	.127	.108	.059	.083	.077	.004
	(.038)***	(.038)***	(.044)***	(.051)**	(.054)	(.117)	(.095)	(.093)
Educ	.002	.010	.005					
	(.007)	(.021)	(.008)					
Educ2		-.0007			-.011			
		(.002)			(.011)			
Age	.074	.074	.077		-.008			
	(.016)***	(.016)***	(.019)***		(.012)			
Age2	-.0008	-.0008	-.0008	-.002	.002	-.001	-.001	-.0009
	(.0002)***	(.0002)***	(.0002)***	(.001)**	(.005)	(.002)	(.002)	(.001)
Male	.504	.506	.831		.148			
	(.060)***	(.061)***	(.705)		(.097)			
(K+R)*Male			-.111					
			(.035)***					
L*Male			.138					
			(.084)*					
Educ*Male			-.024					
			(.016)					
Age*Male			-.004					
			(.033)					
Age2*Male			-.0002					
			(.0004)					
2007	.222	.222	.179	.464	.339	.466	.448	.419
	(.076)***	(.076)***	(.075)**	(.109)***	(.093)***	(.153)***	(.150)***	(.142)***
2008	.130	.129	.108	.627	.201	.497	.480	.438
	(.073)*	(.073)*	(.072)	(.188)***	(.073)***	(.300)*	(.299)	(.242)*
2009	-.090	-.091	-.138	.717		.463	.434	.386
	(.070)	(.070)	(.069)**	(.283)**		(.468)	(.465)	(.388)
Const.	-.799	-.812	-1.092	4.980	.262			
	(.330)**	(.332)**	(.395)***	(1.545)***	(.197)			
Obs.	1304	1304	1298	1304	459	459	459	459
R^2	.313	.314	.346	.165	.171	.	.	
AR(2) test (p-val)								.022
Hansen test (p-val)								.759
Num. of Ins.								22

Confidence: *** $\leftrightarrow$ **99%**, ** $\leftrightarrow$ **95%**, * $\leftrightarrow$ **90%**.; DIFF-2S uses 2-step difference GMM with optimal weighting allowing for arbitrary patterns of heteroskedasticity and Widnmeijer (2005) small sample correction for se;

4.1 Returns to Capital

Using the results of the estimation and given the production model above, I will next compute marginal rates of return on capital (at current levels of output and of the capital stock) for all the firms in our sample using the following expression:

$$\frac{\partial Y}{\partial (K+R)} = \alpha A(K+R)^{(\alpha-1)}L^{\beta} = \alpha\frac{Y}{(K+R)} \tag{5}$$

Table 3 summarises the distribution of estimated returns *per month*, together with the distribution of the output/capital ratio (Y/K). First, it shows great heterogeneity across firms, the extent of which may cast doubts of potential measurement error, but is in fact not uncommon in the existing literature (Banerjee and Duflo (2005)). The median of the distribution is .05, a rather sensible value when compared to the rates of interest charged by credit institutions in this market.[9] Second, the distribution shows returns to capital that are very high for large portions of the sample, a finding that is also rather common across the available evidence. A number of studies have now documented that investment in small-scale and productive activities in the developing world often results in very high rates of return, as one would expect given that capital is still relatively scarce. McKenzie and Woodruff, for example, document returns of 15% per month among urban micro-enterprises in Mexico, while de Mel, McKenzie, and Woodruff (2008) find that exogenous increases to the capital stock of small Sri Lankan firms resulted in yearly average real returns of 55%

[9] For instance, one microcredit institution that we analysed as part of a companion study in Ghana charged interest rates of 4% per month.

to 63%.[10]

Figure 6 plots individual marginal returns to capital ($\partial Y/\partial(K+R)$) against the value of capital $(K+R)$. The plot shows the strong concavity of the production technology. Marginal returns are very high at micro-investment levels, but they decrease very rapidly over the range of capital we observe. The implication of this finding are at least two-fold. On the one hand, high marginal returns to micro-investments suggest that saving and re-investing business profits may be a viable growth opportunity, allowing small entrepreneurs to *bootstrap* their way out of poverty (McKenzie and Woodruff (2006)). This conclusion would seem to be reinforced by the empirical observation that entry-level capital stocks and start-up costs are minimal. On the other hand, when we convert the high marginal returns into *real income gains* (obtained by multiplying the marginal rate of return by the value of the capital stocks), we are left wondering whether absolute returns of the resulting magnitude can be effectively saved and re-invested. Figure 5 highlights this point by showing the distribution of marginal real income gains per month, which has a median of 15 (1997) USD.[11] Moreover, when I plot the same income gains against capital (Figure 7), I find that although marginal returns to capital were decreasing over the range of observed capital stocks, real gains increase steadily over the same range (after excluding extreme values, see part (b) of the graph). Most interestingly, real income gains appear to be rather modest in absolute value for the majority of entrepreneurs, who own very small capital stocks. Can profits of

[10] See McKenzie and Woodruff (2008) for similar evidence on Mexican microenterprises.

[11] The base year of the deflator series available for Ghana at the time of writing is 1997, and to facilitate interpretation of the results I convert the figures in 1997 USD.

such magnitude be effectively saved and re-invested (rather than consumed) in an economy where a substantial segment of the population lives below the poverty line and efficient saving mechanisms are lacking? And therefore, can we reasonably draw a link between profitability and firm growth over time? Answering this question would partly rely on being able to test workers intertemporal preferences over saving and consumption and their degree of self-control, in conjunction with an improved understanding of the degree of asset integration between households and businesses. Fafchamps, McKenzie, Quinn, and Woodruff (2011), in a recent experiment offering random capital drops in the form of cash and in-kind grants to a sample of Ghanaian entrepreneurs similar to ours, find that cash grants were commonly used for household consumption and transfers rather than investment.

Table 3: Distribution of *Output/Capital* and *Returns to Capital*

	1st	**5th**	**25th**	**50th**	**75th**	**95th**	**99th**
$\frac{ValAdd}{(K+R)}$	.01	.04	.15	.31	.74	4.31	20.8
$\frac{\partial ValAdd}{\partial(K+R)}$	.001	.006	.02	.05	.12	.72	3.49

NOTE: Returns to Capital computed using 2-Step Difference-GMM estimate of $\alpha = .168$

Figure 5: Returns to Capital as real income gains

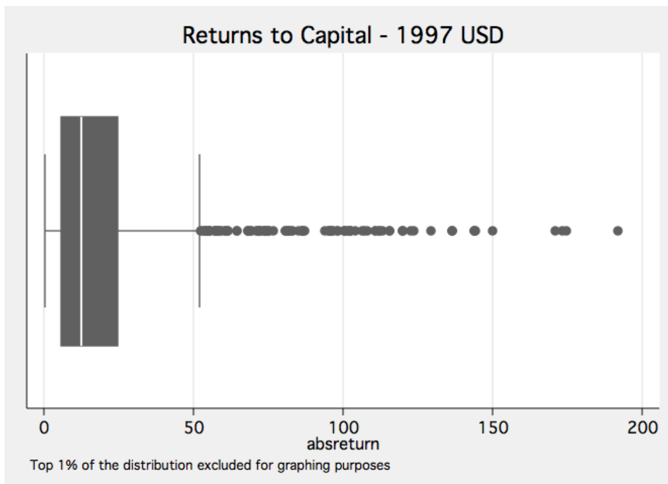
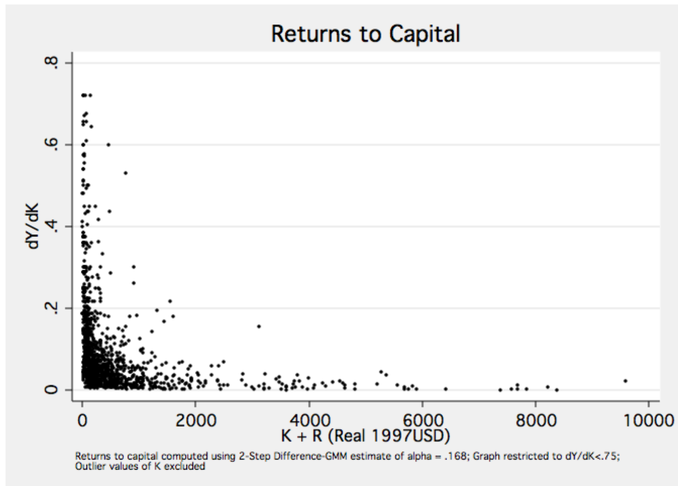


Figure 6: Marginal Returns to Capital



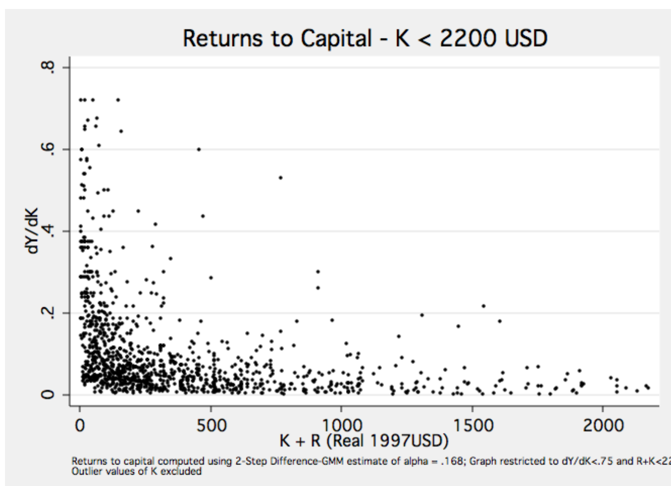
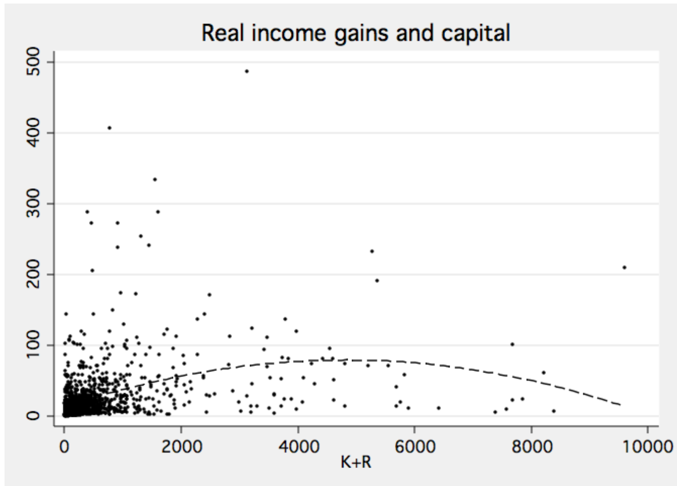
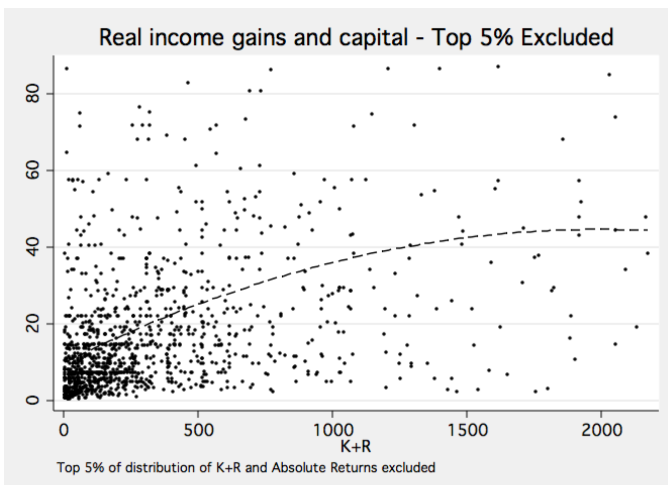


Figure 7: Real income gains and Capital





5 Endogenous choice of production technologies

As I discussed in section 3, my sample of self-employed workers is segmented into a fraction who have positive value of machinery, tools and equipment ($K > 0$) and another group that only uses raw materials without any sizable form of capital that our survey can detect ($K = 0$ and $R > 0$); the latter group is formed predominantly by traders and constitutes about 25% of the sample. In this section I will explicitly allow for the possibility that selection into these two different production technologies is endogenous and I will test the robustness of my results upon controlling for such selection. My strategy will be to re-estimate the production function limiting the sample to entrepreneurs with $K > 0$ and using a first-stage selection model à la Heckman (1979) to control for selection into such *capital-intensive technology*. In Appendix C I will further extend the analysis to a multinomial setting where workers can sort into one of 4 types of technologies, defined by the four possible combinations of zero/positive levels of equipment (K) and zero/positive levels of hired labour (L) in addition to the entrepreneur himself (see table 7 in the appendix). The econometric framework in this second part will be based on the multinomial selection-correction model designed by Dubin and McFadden (1984) and further developed by Bourguignon, Fournier, and Gurgand (2004). Crucially, both the binary and the multinomial model hinge upon the existence of exclusion restrictions that yield valid instruments for selection in the first stage (discussed below). Despite their limitations, however, these strategies will begin to inform our understanding of the heterogeneity that may have been concealed in my previous analysis by lumping together technology types that are potentially very different (e.g. traders and manufacturers), a choice that was dictated by the size of our sample, coupled with the demands of complex GMM estimators. The availability of additional waves of GHUPS in the future will enable me to explore different dimensions of heterogeneity more thoroughly.

First, I begin by augmenting my model of production by adding the following selection equation:

$$D_{K>0,i,t} = 1(Z_{it}\delta + v_{it} \geq 0) \tag{6}$$

where $D_{K>0,i,t} = 1$ if we observe $K > 0$ and zero otherwise, Z_{it} is a vector of variables which comprises X_{it} plus additional instruments for selection and v_{it} is an error term assumed to be independent of Z_{it}. A standard assumption, which I will make, is that Z_{it} is exogenous in (4), such that

$$E(u_{it}|X_{it}, Z_{it}) = 0 \tag{7}$$

If this assumption holds, it follows that

$$\begin{aligned}
E[y_{it}|D_{Z>0,it} = 1] &= E[y_{it}|v_{it} > -Z_{st}\delta] \\
&= \alpha k_{it} + \beta l_{it} + \gamma H_i + \theta X_{it} + E(u_{it}|v_{it} > -z_{it}\delta) && (8) \\
&= \alpha k_{it} + \beta l_{it} + \gamma H_i + \theta X_{it} + \rho E(v_{it}|v_{it} > -z_{it}\delta) && (9) \\
&= \alpha k_{it} + \beta l_{it} + \gamma H_i + \theta X_{it} + \rho\lambda(z_{it}\delta) && (10)
\end{aligned}$$

where I assume joint normality of u_{it} and v_{it} to move from (8) to (9) and λ is the inverse Mills ratio when $D_{K>0,i} = 1$. From the normality assumptions it results that $D_{K>0,i}$ given Z follows a probit model such that:

$$Pr(D_{K>0,it} = 1) = \Phi(Z_{it}\delta) \quad (11)$$

which can be used to derive the Mills ratio to be included in our principal equation as a control for selection.

Estimating this model on the entire sample allows me to obtain $\hat{\delta}$ and compute individual values of the inverse Mills ratio $\hat{\lambda}_{it} = \lambda(Z_{it}\hat{\delta})$, which I can include in the earnings model on the selected sample to correct for the bias. If the model assumptions hold, my estimates of $\alpha, \beta, \gamma, \theta$ should now be robust to endogenous selection into the capital intensive technology. As it is well known, this procedure provides a simple tool to test for the presence of selection bias. Namely, if the coefficient on $\hat{\lambda}$ in the selection-corrected model (ρ) is not significantly different from 0, I will conclude that selection is not a major cause of concern.[12]

The main challenge for this analysis is that our dataset only provides a limited number of potential instruments for sample selection. First, I will use a workers' *marital status* and second, a measure of *unexpected expenses or losses of income/assets* over the year prior to the interview. The rationale for the former instrument is that marriage may contribute to relaxing credit constraints by giving workers access to the assets of their spouse's family and to a new support network, without necessarily affecting his/her productivity. Such intuition appears consistent with first-stage results (see Table 4), where marriage

[12] This conclusion will hinge upon the validity of my instrument set.

is positively correlated with the probability of working in a capital-intensive business. One of the main problems with this instrument is the potential endogeneity of marriage with respect to prior wealth. Thus, I next introduce the latter two instruments, which I believe constitute a more robust engine of exogenous variation in the selection equation. The first-stage results confirm that workers who faced an unexpected loss of assets/income (due to damages to their property, theft, perished inventories, etc.), are less likely to be producing with a capital intensive technology $K > 0$ in the current period. The result is in line with the hypothesis that negative shocks deplete workers' capital. Being the result of unexpected events, such shocks can be held to be exogenous in the earnings equation (though potential biases in self-reporting should be acknowledged). The main limitation with the use of these variables is the fact that they were not recorded in the 2007 wave of the survey, and therefore we are forced to drop a year of data when using them (in addition to the fact that missing data for some individuals causes a further drop in sample size). Interestingly, the first stage model also reveals that men are more likely to have capital intensive technologies than women, and that age correlates positively (albeit weakly from a statistical point of view) with capital utilization, in line with the hypothesis that workers are able to accumulate wealth over their lifetime.

The results of the second stage, in table 5, show that the consequences of controlling for selection are minimal. Comparing OLS (col 1) to the selection-corrected specifications (col 2 - 4) reveals evidence of only a negligible (positive) bias in the OLS returns to capital and labour, detected upon employing all the three instruments described above, but one that cannot be clearly attributed to the selection correction due to the drop in sample size resulting from the exclusion of the 2007 wave. In fact, the insignificant coefficient on λ tells us that selection is not playing a strong role in the equation. And even when the coefficient is significant at the 15% level (HECK 4), the results do not change considerably. While I cannot confidently rule out the hypothesis that the first-stage selection model is unable to control for selection effectively, I take these results as prima-facie evidence that selection into production technologies that employ a positive level of capital, as defined by tools, machinery and equipment, does not bias our estimates of the returns to productive assets.[13] This result is further corroborated by the multinomial analysis in appendix C, where I allow for selection correction á la Dubin and McFadden (1984) into four types of technology, and I am still unable to detect any significant bias in the results. Though this evidence clearly requires further scrutiny as more data becomes available, my overall prima-facie conclusion is that selection into production technologies that are more capital (and labour) intensive does not seem to occur for endogenous reasons that are directly related to productivity and that may bias my estimates of the production model.

Finally, it is important to acknowledge that while I explicitly attempted to control for endogenous choice of technology *within* self-employment, my analysis does not tackle the issue of endogenous *selection into* self-employment. This should not necessarily affect the consistency of the results (i.e. their internal

[13] A further source of improvement on this approach will be to estimate the model via Full-Information Maximum Likelihood that re-estimates the first and second stage equation jointly and therefore makes more efficient use of the available information. The advantage, though, comes at the cost of stricter assumptions on the joint distribution of the error terms.

validity), but it may pose some limitations to their external validity, as these results may in fact only apply to the set of people who have selected into self-employment under prevailing market conditions.

Table 4: Selection into capital intensive technologies

	HECK1	HECK2	HECK3	HECK4
	(1)	(2)	(3)	(4)
Educ	-.022 (.043)	.007 (.048)	.041 (.053)	.039 (.053)
Educ2	.0004 (.002)	-.0003 (.003)	-.003 (.003)	-.003 (.003)
Age	.044 (.022)**	.038 (.025)	.041 (.028)	.040 (.028)
Age2	-.0007 (.0003)***	-.0005 (.0003)*	-.0005 (.0003)	-.0005 (.0003)
Male	.215 (.083)***	.230 (.093)**	.217 (.099)**	.218 (.099)**
2007	.083 (.098)			
2008	.117 (.097)	.117 (.096)	.083 (.104)	.090 (.105)
2009	.560 (.100)***	.560 (.100)***	.562 (.107)***	.562 (.107)***
Married	.209 (.075)***	.181 (.085)**	.168 (.091)*	.169 (.091)*
Finan.Loss			-.209 (.119)*	-.180 (.127)
Unexp.Exp.				-.061 (.092)
Const.	-.319 (.486)	-.325 (.550)	-.390 (.601)	-.348 (.605)
Obs.	1304	1057	941	941

Dependent Variable: $I(K > 0) = 1$; ***Confidence:* *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%**.;

Table 5: Determinants of value-added with *endogenous technology selection*

	OLS	HECK1	HECK2	HECK3	HECK4
	(1)	(2)	(3)	(4)	(5)
K+R	.257	.258	.254	.239	.239
	(.019)***	(.019)***	(.021)***	(.022)***	(.022)***
L	.189	.190	.207	.176	.175
	(.044)***	(.043)***	(.049)***	(.052)***	(.052)***
Educ	-.013	-.022	-.043	-.040	-.039
	(.037)	(.040)	(.046)	(.055)	(.056)
Educ2	-.002	-.001	-.0004	-.0007	-.0008
	(.002)	(.002)	(.002)	(.003)	(.003)
Age	.048	.067	.066	.078	.080
	(.020)**	(.029)**	(.033)**	(.033)**	(.033)**
Age2	-.0005	-.0008	-.0008	-.001	-.001
	(.0002)**	(.0004)**	(.0004)**	(.0004)**	(.0004)**
Male	.498	.557	.548	.561	.570
	(.067)***	(.095)***	(.122)***	(.118)***	(.119)***
2007	.221	.247			
	(.087)**	(.095)***			
2008	.178	.212	.226	.203	.208
	(.085)**	(.096)**	(.105)**	(.108)*	(.110)*
2009	-.132	.022	.085	.115	.137
	(.079)*	(.182)	(.235)	(.213)	(.211)
Const.	-.146	-.767	-.820	-.897	-.981
	(.469)	(.814)	(1.026)	(1.003)	(.995)
$\hat{\lambda}$		.601	.828	.988	1.075
		(.631)	(.848)	(.736)	(.722)
Obs. (uncens.)	996	996	803	712	712
Obs. (cens.)		308	254	229	229
R^2	.302				

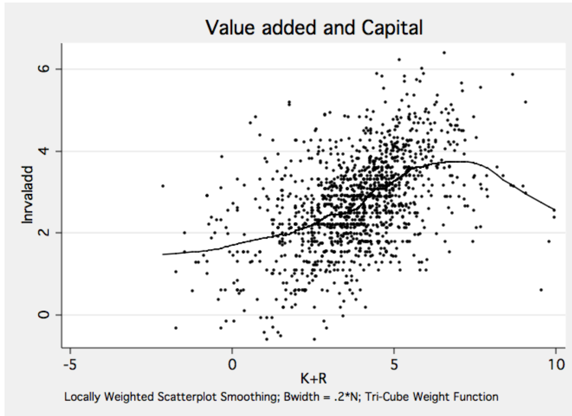
***Confidence:* *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%**.; Obs. (uncens) : $K > 0$; Obs. (cens): $K = 0$

6 Non-convexities in production

A large literature in development economics has focused on the interaction between production non-convexities and capital market imperfections as a potential source of poverty traps. In the presence of imperfect credit markets, the existence of fixed costs and minimum scale requirements that are higher than the wealth available to workers may prevent entry of able but credit constraints entrepreneurs, or it may force them into inefficient production technologies with low returns on their investment. Banerjee and Newman (1993) develop a compelling theoretical argument to describe how capital constraints that induce people into non-capital intensive occupations may lead the economy to a low-growth equilibrium. Minimum investment levels for entrepreneurial activities are also a feature of the models by Aghion and Bolton (1997). McKenzie and Woodruff (2006) discuss the link between poverty traps and non-convexities in production, while finding no evidence of the latter in their Mexican dataset. In particular, they explain how the co-existence of production non-convexities and poorly functioning capital markets may lead to poverty traps, as workers are unable to *borrow* nor *bootstrap* (via savings) their way out of poverty. Conversely, in the absence of non-convexities, even when capital markets function poorly, poverty traps may cease to exist.

In the spirit of the work by McKenzie and Woodruff (2006) and on the basis of the results in the previous sections, one can empirically test the two necessary premises for the existence of poverty traps, namely that (a) minimum scale requirements are high relative to wealth and (b) that returns to capital are low at low levels of investment (and increase thereafter). The first premise is strongly challenged by the results in section 4, where I have shown that a large proportion of entrepreneurs in our sample operates with very low levels of capital and sometimes with no fixed capital at all. In the light of this evidence, it would appear unrealistic to believe in a model where minimum scale requirements keep credit-constrained entrepreneurs out of 'any kind' of entrepreneurship. More plausibly, I am inclined to believe that entrepreneurs who are credit constrained may be 'pushed' into specific forms of entrepreneurship, and if those are the least profitable ones, the result may be a poverty trap. To assess this hypothesis, it becomes important to evaluate the second premise, which has also been challenged by the results in section 4 (where marginal rates of return on capital were estimated very high at low capital levels and rapidly decreasing thereafter), and which I will now subject to further testing. In fact, my empirical model, derived from a log-linearisation of a Cobb-Douglas production function, has so far imposed a linear relationship between log-capital and log-earnings. I will now relax this assumption and allow for greater flexibility in the shape of the function with the aim to detect regions of increasing returns to capital. My point of departure is figure 8, where I plot a locally weighted scatterplot smoothing of earnings against capital (K + R). The graph is suggestive of the hypothesis that returns to capital might be lower both at the low and at the top end of the capital distribution (a pattern that survives the exclusion of some apparent outliers). This is an interesting fact, which *may* point to the existence of the kind of non-convexities described above. In order to explore the shape of the production set more formally, I first re-estimate the production model on each tertile of the capital distribution separately, with a view to

Figure 8: Marginal Returns to Capital



assess differences in the magnitude of the estimated effects. The results are reported in table 6, and they show some interesting patterns. In businesses with medium levels of capital, the production technology is closer to one with constant returns to scale, labour plays a stronger role in production and the share of value added that is attributed to capital is highest. The coefficients on capital, however, are not statistically significant from each other (more clearly so in the specification with fixed effects - *WG*). Interestingly, it also appears that the most capital intensive businesses are characterised by strongest degree of decreasing returns to scale. A potential explanation for this finding is that higher income and higher capital levels may be associated with higher measurement error in the data, and therefore, precision in identifying the effects of the factors of production drops. This hypothesis is supported by the drop in the R^2 in column 5 of table 6.

Finally, by introducing even greater flexibility in the estimator, I will try to refine my search for regions of increasing marginal returns to capital. My

Table 6: Income by (K+R)-tertile

	OLSQ1	WGQ1	OLSQ2	WGQ2	OLSQ3	WGQ3
	(1)	(2)	(3)	(4)	(5)	(6)
K+R	.21 (.05)***	.11 (.08)	.53 (.11)***	.50 (.18)***	.09 (.05)*	.04 (.09)
L	.26 (.06)***	.0002 (.11)	.22 (.06)***	.14 (.10)	.13 (.07)*	-.16 (.13)
Educ	.003 (.01)		-.01 (.01)		.01 (.01)	
Age	.09 (.02)***	1.05 (.81)	.04 (.03)	.36 (.65)	.09 (.04)**	.57 (1.18)
Age2	-.001 (.0003)***	-.0004 (.002)	-.0004 (.0003)	-.004 (.002)**	-.001 (.0004)**	-.004 (.003)
Male	.78 (.11)***		.44 (.10)***		.35 (.10)***	
2007	.24 (.14)*	-.61 (.73)	-.14 (.12)	-.16 (.60)	.43 (.14)***	.47 (1.12)
2008	.19 (.13)	-1.82 (1.64)	-.09 (.11)	-.19 (1.33)	.29 (.14)**	.03 (2.50)
2009	-.08 (.13)	-3.18 (2.57)	-.15 (.10)	-.24 (2.10)	-.02 (.13)	-.19 (3.91)
Const.	-1.34 (.49)***	-34.38 (27.87)	-1.04 (.70)	-6.60 (22.97)	.24 (.80)	-12.66 (44.96)
Obs.	419	419	445	445	440	440
R^2	.23	.08	.15	.17	.1	.14

Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%.; The constant term in the WG estiamtor is set-up to be the average of the fixed effects;

preferred method is the estimation of fractional polynomial regressions of the income-generating process. First, I specify the following general family of non-linear polynomials:

$$y_{it} = \sum_{m=1}^{M} \alpha_m k_{it}^{p_m} + \beta l_{it} + \gamma H_i + \theta X_{it} + (\gamma_0 + \delta_t + \eta_i + \omega_{it}) \qquad (12)$$

where each power p_m is chosen from a pre-defined set.[14]. Then, all combinations of powers are fitted to the data and the best performing model is selected based on goodness of fit. Figure 9 (part a) plots the best-fitting (logarithmic) production function obtained with this method, coupled with a plot of its first

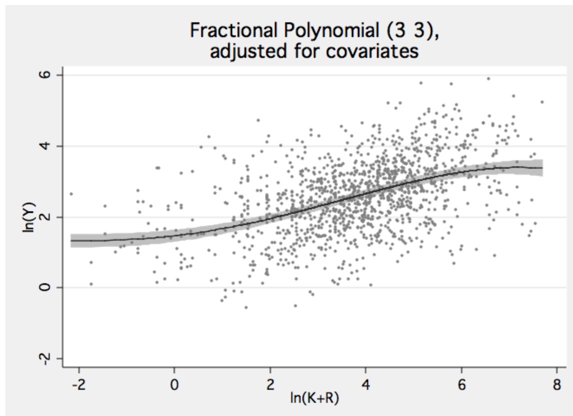
[14] The algorithm I choose searches over the following powers of p_m: -2, -1, -.5, 0, .5, 1, 2, 3

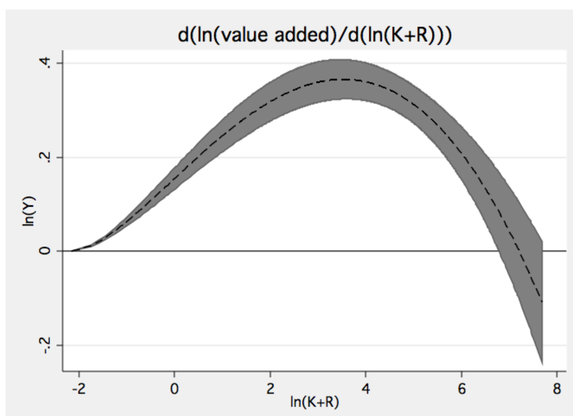
derivative, evaluated at different levels of capital (part b).[15]

Despite its mainly descriptive value, this exercise shows that even after we allow for greater flexibility in the income-generating process, we are far from detecting, *ceteris paribus*, any non-convexity in the production set. The share of value added attributed to capital is highest in the middle of the capital distribution, but the implied marginal rate of return is still decreasing all along the observed spectrum of the capital stock (albeit possibly decreasing at different speeds). Moreover, the change in the magnitude of the (non-linear) coefficient on log-capital (bottom panel, figure 9) is not significant over the observed capital range. These findings, coupled with evidence of small start-up costs reported by the entrepreneurs in the sample, run counter to the hypothesised existence of poverty traps and they are hence in line with the evidence on Mexico presented by McKenzie and Woodruff (2006).

[15] Such first derivative effectively corresponds to α in the linear model estimated in previous sections and it should take values greater than 1 in regions of increasing returns to capital.

Figure 9: Fractional Polynomial Estimation





7 Conclusions

This article has investigated the returns to workers' productive assets (in the form of labour, physical capital and human capital) in African micro-enterprises. From a theoretical standpoint, I have argued a case for abridging the gap between the analysis of individual earnings and the study of firms' value-added, using a model of the income-generating process that is grounded in the study of enterprises' production functions. From an empirical perspective, I have attempted identification of the returns to labour, physical and human capital by means of a new 'long' african panel dataset, collected by CSAE from 2004 to 2009. The panel dimension of the data has allowed me to employ panel estimators that are suitable to address concerns of endogeneity in input selection due to both time-varying and time-invariant unobservables.

The results I obtain show that physical capital and labour market experience play the strongest role in the income generating process of the self-employed, while the share of value-added attributed to labour is small and imprecisely estimated. The latter result can be partly attributed to the difficulty of measuring labour inputs in household based enterprises where hiring is generally informal and wages paid irregularly. It may also reflect production processes that are genuinely inefficient, with more labour being employed than it would be optimal. The overall result is a production technology that appears to have decreasing returns to scale. Most strikingly, the productivity-enhancing effect of formal education in self-employment appears to be negligible. Learning on the job seems to be a more important dimension of human capital than formal schooling. This result may be viewed as evidence of the limited effectiveness of universal education policies in economies where the majority of available earning opportunities is in informal self-employment; and it suggests that while education may be granting workers access to desirable wage-opportunities (e.g. the public sector), its capacity to increase their productivity in informal self-employment remains limited. Estimated returns to capital investment, on the other hand, are found to be highly heterogeneous and often very large for a substantial proportion of the sample (median estimated returns are .05 per month). Both these findings are in line with the existing literature. When I control for endogenous selection into capital intensive technologies using a first stage selection model, my core results do not change significantly, but the robustness of these finding is weakened by the limitations of the available instruments.

Finally, when I explore the shape of the production function over the range of capital observed, I find a highly concave technology. Marginal returns to investment are high at very low capital levels (it is not uncommon in our sample to find businesses that operate with capital value equal to 10USD), but they decrease very rapidly. The implication of this result are two-fold. On the one hand, coupled with evidence of low entry costs, these findings point against the existence of non-convexities in the production technology driven by minimum-scale requirements or regions of convex technology. On the other hand, the real income gains that result from high marginal returns are modest as they are produced from very small capital stocks. Whether high returns to investment will be conducive to firm growth as firms re-invest their profits and attempt to *bootstrap* themselves out of poverty remains therefore open to debate and it will partly depend on the workers' inter-temporal preferences.

In conclusion, a robust assessment of returns to micro-entrepreneurship indirectly allows us to shed light on the effectiveness of policies aimed at relaxing workers' credit constraints in developing countries. In particular, the proliferation of micro-credit as a poverty alleviation tool is grounded in the belief that profitable investment opportunities are available to the poor, but cannot be taken advantage of, due to the existence of binding credit constraints. The spread of microfinance in Ghana over the last few decades was largely based on this argument. My results show that this view is apparently justified by the existence of high marginal returns to capital at very low capital stocks (similar in magnitude to the capital-stocks at which micro-finance operates). However, I am rather sceptic on the effectiveness of micro-investments as an engine of growth, since the lack of efficient saving mechanisms coupled with preferences that are skewed towards immediate consumption in a context of widespread poverty appears to be an important missing link between high returns to capital and reinvestment that leads to growth and sustainable poverty alleviation. It is true, though, that whether these micro-enterprises will be able to grow larger over time is ultimately an empirical research question that we are aiming to investigate as more data becomes available, with a particular focus on inter-temporal saving and consumption decisions and on the interconnections between households and business choices.

References

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2006): "Structural Identification of Production Functions," *Unpublished manuscript*.

AGHION, P., AND P. BOLTON (1997): "A Theory of Trickle-Down Growth and Development," *Review of Economic Studies*, 64(2), 151–72.

ANDERSON, T. W., AND C. HSIAO (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18(1), 47–82.

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58(2), 277–97.

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, vol. 1 of *Handbook of Economic Growth*, chap. 7, pp. 473–552. Elsevier.

BANERJEE, A. V., AND A. F. NEWMAN (1993): "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101(2), 274–98.

BASU, S., AND J. FERNALD (1995): "Are Apparent Productive Spillovers a Figment of Speci

cation Error," *Journal of Monetary Economics*, 36(1), 165–188.

——— (1997): "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 105(2), 249–83.

BLUNDELL, R., AND S. BOND (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87(1), 115–143.

BOND, S., AND M. SODERBOM (2005): "Adjustment costs and the identification of Cobb Douglas production functions," IFS Working Papers W05/04, Institute for Fiscal Studies.

BOURGUIGNON, F., M. FOURNIER, AND M. GURGAND (2004): "Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons," Discussion paper.

CARD, D. (1999): "Chapter 30 The causal effect of education on earnings," vol. 3, Part A of *Handbook of Labor Economics*, pp. 1801 – 1863. Elsevier.

DE MEL, S., D. MCKENZIE, AND C. WOODRUFF (2008): "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 123(4), 1329–1372.

——— (2010): "Wage Subsidies for Microenterprises," *American Economic Review*, 100(2), 614–18.

DE MEL, S., D. J. MCKENZIE, AND C. WOODRUFF (2009): "Measuring microenterprise profits: Must we ask how the sausage is made?," *Journal of Development Economics*, 88(1), 19–31.

DUBIN, J. A., AND D. L. MCFADDEN (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52(2), 345–62.

EBERHARDT, M., AND C. HELMERS (2010): "Addressing Transmission Bias in Micro Production Function Models: A Survey for Practitioners," *Working Paper*.

FAFCHAMPS, M., D. MCKENZIE, S. QUINN, AND C. WOODRUFF (2010): "Using PDA consistency checks to increase the precision of profits and sales measurement in panels," Discussion paper.

——— (2011): "When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana," .

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56(6), 1371–95.

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(3), 392–427.

LEVINSOHN, J., AND A. PETRIN (2000): "Estimating Production Functions Using Inputs to Control for Unobservables," NBER Working Papers 7819, National Bureau of Economic Research, Inc.

MCKENZIE, D., AND C. WOODRUFF (2008): "Experimental Evidence on Returns to Capital and Access to Finance in Mexico," *World Bank Economic Review*, 22(3), 457–482.

McKenzie, D. J., and C. Woodruff (2006): "Do Entry Costs Provide an Empirical Basis for Poverty Traps? Evidence from Mexican Microenterprises," *Economic Development and Cultural Change*, 55, 3–42.

Monk, C., J. Sandefur, and F. Teal (2008): "Does Doing an Apprenticeship Pay Off? Evidence from Ghana," .

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–1426.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–97.

Roodman, D. (2009a): "How to do xtabond2: An introduction to difference and system GMM in Stata," *Stata Journal*, 9(1), 86–136.

——— (2009b): "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.

Udry, C., and S. Anagol (2006): "The Return to Capital in Ghana," *American Economic Review*, 96(2), 388–393.

Windmeijer, F. (2005): "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of Econometrics*, 126(1), 25–51.

APPENDIX

A Panel Estimators

In this section I will outline the structure and briefly discuss the properties of the panel estimators used to identify the effect of different productive assets in the income generating process.

Recall the empirical analog of the income model:

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma H_i + \theta X_{it} + (\gamma_0 + \delta_t + \eta_i + \omega_{it}) \tag{13}$$

The standard Within Group estimator (WG) is based on the following *transformed* equation:

$$\tilde{y}_{it} = \alpha \tilde{k}_{it} + \beta \tilde{l}_{it} + \theta \tilde{X}_{it} + (\tilde{\delta}_t + \tilde{\omega}_{it}) \tag{14}$$

where:

$$\begin{aligned} \tilde{y}_{it} &= \frac{1}{T}\sum_{t=1}^{T} y_{it}; & \tilde{k}_{it} &= \frac{1}{T}\sum_{t=1}^{T} k_{it} \\ \tilde{l}_{it} &= \frac{1}{T}\sum_{t=1}^{T} l_{it}; & \tilde{\delta}_{t} &= \frac{1}{T}\sum_{t=1}^{T} \delta_{t} \\ \tilde{\omega}_{it} &= \frac{1}{T}\sum_{t=1}^{T} \omega_{it} \end{aligned} \tag{15}$$

The first-differenced model, on the other hand, is based on the following transformation:

$$\Delta y_{it} = \alpha \Delta k_{it} + \beta \Delta l_{it} + \theta \Delta X_{it} + (\Delta \delta_t + \Delta \omega_{it}) \tag{16}$$

where $\Delta(.)$ is standard notation for the first-difference of each variable. Due to the structure of the transformed error term, the WG estimator generally suffers from the so-called *Nickell bias* (see Nickell (1981)). First-Differencing, on the other hand, is likely to be characterised by lower precision due to reduced sample size, especially in unbalanced panels where attrition is not an absorbing state (i.e. where respondents who are not interviewed in one wave may 're-appear' in subsequent waves, which can result in a 'patchy' dataset). Since for every missing observation *two* first differences are lost, sample size may indeed shrink dramatically in those circumstances.

The Anderson-Hsiao (1982) Instrumental Variable approach introduces the lags of the regressors as suitable instruments to overcome issues of endogeneity. In matrix notation, the estimator can be expressed as follows:

$$\Theta_{AH} = (\Delta X' Z (Z'Z)^{-1} Z' \Delta X)(\Delta X' Z (Z'Z)^{-1} Z' \Delta Y) \tag{17}$$

where:

$$\Theta = \begin{pmatrix} \alpha \\ \beta \\ \theta \\ \delta_t \end{pmatrix}$$

is the vector of coefficients we are aiming to estimate. ΔY is the stacked $N(T-s)\times 1$ vector of observations on Δy_{it}, ΔX is the stacked $N(T-s)\times m$ vector of observations on ΔX_{it} and Z is the stacked $N(T-s)\times 1$ vector of observations on $x_{i,t-s)}$. m is the number of individual characteristics included in X, while s may be equal to 1 or 2, depending on whether we assume pre-determinedness or endogeneity of capital and labour respectively.

The Anderson - Hsiao estimator is a specific case of a more general class of GMM estimators, which can be expressed as follows:

$$\Theta_{AH} = (\Delta X' Z W Z' \Delta X)(\Delta X' Z W Z' \Delta Y) \tag{18}$$

where W is a *weighting matrix*. Optimal GMM sets:

$$W = (Z'\hat{\Omega}Z) \tag{19}$$

Weighting the data by the inverse of Ω (the variance-covariance matrix of Δu_{it}), enables us to make a more efficient use of the available information by attributing less weight to noisier signals. Implicitly, therefore, the Anderson-Hsiao estimator (equivalent to 2-stage least squares) assumes that $\Omega = \sigma^2 I$ and is only robust if the error terms are homoskedastic.

To improve efficiency, Holtz-Eakin, Newey, and Rosen (1988) extend the Anderson - Hsiao approach to use deeper lags of the endogenous regressors as additional instruments. In order to do that, they have to overcome the problem that deeper lags cause dramatic reductions in sample-size, as additional time-periods must be dropped. For example, as explained in Roodman (2009a), standard Anderson-Hsiao (2SLS) estimators, would enter the instrument k_{it-1} in a single column of Z, as a stack of:

$$Z_i = \begin{pmatrix} . \\ k_1 \\ \vdots \\ k_{T-2} \end{pmatrix}$$

The "." represents a missing value, which forces the estimator to drop the first row in the dataset. The way around this problem is to build a set of instruments from the twice-lag of k, one for each time period, and substitute zeros for missing observations, resulting in 'GMM-style instruments.

$$Z_i = \begin{pmatrix} 0 & 0 & \dots & 0 \\ k_{i1} & 0 & \cdots & 0 \\ \vdots & k_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & k_{i,T-2} \end{pmatrix}$$

In unbalanced panels, as in our case, one also substitutes zeroes for missing values. Once we have overcome the trade-off between lag-depth and sample-depth, we can include all *valid* lags of the endogenous variables as additional instruments. For endogenous variables, these are all lags up to $t-2$, while for pre-determined variables, we can add the extra $t-1$ lag. Since it makes use of additional information from deeper lags, the Holtz-Eakin, Newey and Rosen (1988) estimator is more efficient than AH. However, it still implicitly assumes homoskedastic error terms, which is often implausible, especially after first-differencing (when the first-differenced error terms are far from *i.i.d.* (Roodman (2009a)), and it often remains poorly behaved. As they allow for more complex patterns in the covariance of the error terms that form the weighting matrix W, GMM estimators á la Arellano and Bond (1991) overcome this problem. In the body of the article, I have reported the results form a two-step estimation. In the first step, I chose an arbitrary specified weighting matrix (which is in fact based on the assumption that the disturbances are *i.i.d.*) and using the first-stage residuals I obtain an estimate of $\hat{\Omega}$. Then, using this estimate in the optimal second-stage weighting matrix I arrive at the consistent and robust estimate of Θ_{GMM}, discussed in the text. The standard errors I presented were further corrected using Windmeijer (2005) small-sample correction method.

B Data Construction

The Ghana Household Urban Panel Survey (GHUPS) is the result of long-lasting efforts by CSAE researchers. Based on Census data, a representative sample of the Ghanaian urban population was identified in 2004 and, since then, the same respondents (with the obvious caveat of attrition) have been interviewed (together with their families) at yearly intervals. Self-employed workers, who are the focus of this paper, are all those individuals who report working in trading, manufacturing and service businesses where they own their means of production, do not perceive a wage from an external employer and ultimately bear the risks of the production process. In this section we describe how the variables used in this study were generated from the raw GHUPS dataset.

Age

The GHUPS dataset records respondents' year and month of birth. Using this information and assuming, as a convention, that every respondent is born exactly in the middle of the month, we compute a continuous age variable measured in years with 3 decimal places. This variable is assumed to vary by exactly one between every two waves of the dataset.[16]

Education

GHUPS respondents are asked to report the total number of years they spent in school, as well as the highest school-grade completed. Upon examining these two measures I concluded that the second one is less prone to issues of measurement error. For the purpose of my statistical analysis, I translated it into a continuous value of years in formal education, assigning to every grade the corresponding number of years in the Ghanaian school system.

Capital

The discussion on proxies for capital was already provided in the text. To that, we should add that among the capital proxies available in the data, value of tools and equipment appears to be the most precisely estimated and the most relevant for the identification of the production technology. Using the replacement value of the tools (defined as the amount of money for which which the tools in possession of the entrepreneur could be sold in the current period) seems to be the most accurate method for measuring the capital stock in an economy where the exchange of second-hand tools is frequent and capital changes hands very fluidly (e.g. used cars bought to be transformed into taxis are the norm, rather than the exception).

Working capital is measured via a more complex set of questions and live cal-

[16] Effectively assuming that all respondents are interviewed on the same day. This is not an unreasonable assumption, given that the duration of each wave of the survey was usually contained in a period of a 4-5 weeks.

culations performed by means of handheld computers. Respondents are asked to list all the raw materials purchased for production or unprocessed re-sale during the course of a week. The computer, operated by a trained enumerator, sums up these values, taking into account the appropriate unit of measurement as well as the unit prices, and computes a total, which the respondent is asked to confirm before moving on to the next section of the questionnaire.

Labour Hours

Labour is measured as the total number of hours dedicated to the business during the course of a normal week. Respondents are directly asked to report this figure. As discussed in the text, entrepreneurs who report having employees were not also asked to report the number of hours these employees worked. As a convention, I chose to multiply the number of employees by a flat rate of 40 hours per week to obtain total hours of hired labour, which I added to the hours worked by the entrepreneur to obtain L.

C Multinomial selection of technology

In this section I refine my analysis of the potential endogeneity in the choice of the production technology by constructing a multinomial first-stage selection model, whereby workers sort into one of *four* types of production mechanisms, as summarised in Table 7.

Table 7: Multinomial Production Technologies

	$L = 1$	$L > 1$
$K = 0$	TECH 1	TECH 2
$K > 0$	TECH 3	TECH 4

where L only in the table above is exceptionally used to define the number of employees in the business, and not the number of hours they work (as it has been the case throughout the regression analysis).

In addition to whether or not the firm uses positive values of K, this model allows for the selection into using hired labour, $L > 1$ (in addition to the entrepreneur's own time). As most of our sample is constituted of firms with $L = 1$ it is especially interesting to analyse the 'rare' decision of becoming an 'employer' (against remaining a one-worker firm);[17] and if labour-intensive technologies are chosen for endogenous reasons related to productivity, explicitly modeling the process of selection will add robustness to the analysis.

[17] This is especially true in the light of recent experimental evidence from Sri-Lanka using randomized wage-subsidies to suggest that important constraints to optimal hiring decisions may exist in developing countries (de Mel, McKenzie, and Woodruff (2010))

A detailed treatment of the estimation method is beyond the scope of this chapter and the reader is referred to the original work by Dubin and McFadden (1984) and Bourguignon, Fournier, and Gurgand (2004). For the purpose of this analysis, it will suffice to say that I will employ a first stage selection model based on a multinomial logit of the probability of being in one of the four technologies above. In the second stage I will re-estimate the income model, controlling for selection by means of the selection terms generated from the first stage estimates (see Bourguignon, Fournier, and Gurgand (2004) for a feasible methodology to implement this estimator). As in the previous attempt to control for selection, I will use marital status, unexpected expenses and losses of income/assets over the year prior to the survey as instruments for selection.

The results are reported in table 8 and 9.[18] As in the binary case, the first stage analysis in table 8 shows a strong effect of marriage on the allocation into capital intensive technologies (TECH3 and TECH4). Moreover, marriage strongly correlates with using a technology that is both labour and capital intensive (TECH 4); and, as in the previous section, gender displays a strong correlation with the type of technology chosen.

The results of the second stage estimation are reported in table 9, where I choose to focus on Technology 3 and 4, which are the ones that employ positive levels of K and therefore lend themselves to direct comparison with the results of the Hackman model in the previous section. As a benchmark, I report the OLS results re-estimated on the selected samples. As in the previous

[18] For easiness of exposition, I choose to report only the results from the regressions that include marital status as the sole instrument. My findings do not change significantly upon adding the remaining two variables, while undergoing further loss of precision due to reduction in sample sizes.

section, controlling for selection causes only minor departures from the OLS estimates of returns to capital and labour; but confining the sample to the most capital and labour intensive technology (TECH 4) causes a sharp increase in the returns to labour. This result says that, as one might expect, the largest, most capital endowed and perhaps more formal firms are the ones where labour productivity is highest. However, we are unable to draw strong conclusions on whether selection, as captured by our model, matters statistically, since the coefficient on the selection correction terms in the second stage (m1 - m3) are not significant. Therefore, as I fully acknowledge the limitations of the available instrument set, I am inclined to conclude that this evidence should still be considered prima-facie and will require further investigation as new data becomes available.

Table 8: Multinomial Technology Selection

	TECH2	TECH3	TECH4
Educ	.012 (.039)	-.023 (.020)	.023 (.024)
Age	.073 (.084)	.022 (.044)	.003 (.053)
Age2	-.0005 (.001)	-.0005 (.0005)	-1.00e-05 (.0006)
Male	1.219 (.332)***	.766 (.196)***	1.119 (.221)***
2007	1.267 (.387)***	.414 (.220)*	1.526 (.266)***
2008	.766 (.385)**	.265 (.197)	1.164 (.250)***
2009	.137 (.511)	.934 (.213)***	1.854 (.259)***
Married	-.205 (.304)	.296 (.156)*	.652 (.192)***
Const.	-4.428 (1.736)**	-4.428 (1.736)**	-4.428 (1.736)**
Obs.	1304	1304	1304

Confidence: ***** ↔ 99%, ** ↔ 95%, * ↔ 90%**.; Multinomial logit of selection with 4 categories (TECH1, TECH2, TECH3, TECH4), base category: TECH 1

Table 9: Determinants of value-added with multinomial technology selection

	OLS - T3	DMF - T3	OLS - T4	DMF - T4
K + R	.272 (.022)***	.270 (.022)***	.229 (.038)***	.240 (.039)***
L	.079 (.062)	.069 (.062)	.468 (.135)***	.479 (.135)***
Educ	-.007 (.009)	-.045 (.017)***	.003 (.016)	-.062 (.085)
Age	.071 (.021)***	.097 (.025)***	.032 (.039)	.124 (.099)
Age2	-.001 (.001)***	-.001 (.0004)***	-.0002 (.0005)	-.002 (.002)
Male	.594 (.079)***	.783 (.294)***	.214 (.129)*	.007 (.022)***
2007	.271 (.104)*	-.051 (.371)	.035 (.182)	-.666 (1.250)
2008	.191 (.098)*	-.068 (.254)	.060 (.184)	-.479 (1.012)
2009	-.173 (.091)	-.179 (.254)	-.140 (.171)	-.027 (.810)
m1		-1.359 (1.395)		-.746 (4.002)
m2		.852 (1.599)		-4.119 (3.431)
m3		-1.254 (.59)**		4.907 (4.907)
Const.	-.215 (.459)	-1.876 (1.034)**	-1.064 (.985)	1.230 (5.112)
Obs.	706	706	283	283

***Confidence:* *** ↔ 99%, ** ↔ 95%, * ↔ 90%**.; Base Category: TECH 1; OLS-T3 and OLS-T4 report Ordinary Least Squares estimates confined to the samples of workers using technology 3 and 4 respectively; DMF-T3 and DMF-T4 report selection-corrected estimates using the Dubin-McFadden (1984) methodology to model selection into technology 3 and 4 respectively;

C.1 Relaxing pre-determinedness of labour

Table 10: Relaxing pre-determinedness of labour

	OLS	WG	AH	HNR	DIFF-2S
	(1)	(2)	(3)	(4)	(5)
K+R	.272	.196	.194	.172	.152
	(.017)***	(.026)***	(.122)	(.053)***	(.061)**
L	.197	.108	-1.080	-.030	.143
	(.038)***	(.051)**	(.935)	(.239)	(.217)
Educ	.002				
	(.007)				
Age	.074				
	(.016)***				
Age2	-.0008	-.002	.0007	-.0009	-.001
	(.0002)***	(.001)**	(.003)	(.002)	(.002)
Male	.504				
	(.060)***				
2007	.222	.464	.518	.453	.437
	(.076)***	(.109)***	(.266)*	(.151)***	(.143)***
2008	.130	.627	.399	.472	.460
	(.073)*	(.188)***	(.527)	(.301)	(.251)*
2009	-.090	.717	.083	.400	.452
	(.070)	(.283)**	(.862)	(.474)	(.403)
Const.	-.799	4.980			
	(.330)**	(1.545)***			
Obs.	1304	1304	334	459	459
R^2	.313	.165	.	.	
AR(2) test (p-val)					.028
Hansen test (p-val)					.756
Num. of Ins.					19

***Confidence:* *** ↔ 99%, ** ↔ 95%, * ↔ 90%**.; DIFF-2S uses 2-step difference GMM with optimal weighting allowing for arbitrary patterns of heteroskedasticity and Widnmeijer (2005) small sample correction for se;

C.2 Robustness to outliers

Table 11: Value Added - Hours - Outliers excluded

	OLS	WG	AH	HNR	DIFF-2S
	(1)	(2)	(3)	(4)	(5)
K+R	.311 (.018)***	.249 (.029)***	.153 (.061)**	.164 (.052)***	.179 (.061)***
L	.184 (.037)***	.115 (.050)**	.089 (.117)	.090 (.094)	.017 (.094)
Educ	.0009 (.007)				
Age	.073 (.015)***				
Age2	-.0008 (.0002)***	-.002 (.001)**	-.002 (.002)	-.002 (.002)	-.001 (.001)
Male	.477 (.060)***				
2007	.199 (.075)***	.402 (.110)***	.503 (.152)***	.496 (.150)***	.457 (.145)***
2008	.117 (.072)	.554 (.188)***	.584 (.301)*	.579 (.300)*	.530 (.247)**
2009	-.119 (.069)*	.621 (.282)**	.607 (.468)	.599 (.466)	.539 (.392)
Const.	-.827 (.326)**	4.499 (1.539)***			
Obs.	1281	1281	449	449	449
R^2	.335	.192	.	.	
AR(2) test (p-val)					.032
Hansen test (p-val)					.74
Num. of Ins.					22

***Confidence:* *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%**.; DIFF-2S uses 2-step difference GMM with optimal weighting allowing for arbitrary patterns of heteroskedasticity and Widnmeijer (2005) small sample correction for se;

Chapter 3:

Income Vulnerability and Happiness: panel evidence on the determinants of life satisfaction in Ghana

1 Introduction

The study of happiness and of the mechanisms that generate it has been increasingly attracting the attention of economists in recent years. Partly due to wider availability of survey data that provide "direct" measures of life-satisfaction, many researchers have explored the link between pecuniary and non-pecuniary aspects of life and people's happiness.

Particular efforts have been dedicated to investigate the link between *income and life-satisfaction*. An established result in the existing literature is that such link exists at a cross-sectional level, where richer people are generally observed to report higher levels of happiness, but is not found in aggregate time-series, where generalised income gains fail to deliver generalised improvements in life-satisfaction (see the next section for a comprehensive survey of the literature). This body of work is particularly valuable, as it directly tests the important economic tenet according to which income is closely linked to well-being. Since welfare goes well beyond material wealth, measuring life satisfaction and documenting such link is a step towards strengthening the foundations of economic analysis.

However, if the connection between happiness and income has been heavily researched, the one between *happiness and income-vulnerability* (defined in this paper as the probability of falling below a low-income threshold, alternatively referred to as 'poverty line') is still unexplored. This is partly due to the challenges of estimating the probability distribution of income convincingly. At the same time, it appears to be a very important area of research, especially in developing countries, where the lack of formal insurance often leaves workers exposed to severe income shocks that may have dramatic impact on their welfare. In particular, we may expect that being 'vulnerable' has a direct impact on utility and ultimately on life-satisfaction, *over and above any income effect*. The main objective of this study is to investigate this unexplored link. The focus of the analysis will be the urban labour market of Ghana, but given the novelty of the question and the testing strategy, the results may extend beyond the context of developing countries.

My empirical approach is motivated by a large literature (discussed in the next section) that identifies uninsured downside risk as a fundamental element of workers' livelihoods and highlights that aversion to losses may be an appropriate characterisation of workers' preferences. In doing so, it justifies the definition of vulnerability used in this paper, one that concentrates on *downside income risk*. The measures of vulnerability I use build on the work by Chaudhuri (2003) and Chaudhuri, Jalan, and Suryahadi (2002), who propose two indices of vulnerability that are amenable to empirical estimation and are based on panel and cross-sectional variation respectively. Using data from the Ghana Household Urban Panel Survey (GHUPS), a long panel dataset gathered by the Centre for the Study of African Economies in urban Ghana, I will obtain estimates of these two indices for a sample of working age Ghanaian earners and I will study the relationship between vulnerability and life-satisfaction. I will focus more extensively on the first of these two measures (panel measure), since it enables me to exploit the longitudinal dimension of the data to model variation in earnings at the level of the individual and estimate individual-specific income uncertainty. The second measure, based on cross-sectional variation will mainly serve as a benchmark, allowing me to assess the feasibility of this type of analysis in the absence of panel data (which, is still rare in developing countries).

After constructing the measure of individual vulnerability, my testing strategy will proceed in four steps. First, I will estimate the main relationships of interest, between income, income-vulnerability and life-satisfaction. Given the panel dimension of the data, I will be in a position to control for individual fixed effects in the happiness model, ruling out biases that may be due to unobserved determinants of happiness such as innate personality traits. The model will also control for working-conditions and relative-income, allowing me to exclude potentially confounding effects. Second, I will explore the hypothesis of adaptive income aspirations, which has caught the interest of a growing literature as it may help explain why societies that grow richer do not grow happier. Third, I will present prima facie evidence documenting the interaction between vulnerability to low income and attitudes to risk in determining happiness, using the same experimental measure of risk-aversion employed in chapter 1. Fourth, I will test whether vulnerability to downward income losses (as defined by the vulnerability index) has a different impact on happiness compared to general (two-sided) uncertainty (as defined by simpler measures of income variance).

My main result is a strong and significant effect of vulnerability to low income on workers' happiness *over and above* a sheer income effect. Rather interestingly, the result only becomes significant once individual fixed effects, that account for unobserved personality traits, are controlled for. Moreover, when I bootstrap the entire estimation sequence to explicitly account for the imprecision in the estimation of my measure of vulnerability in the first stage, the results do not change. Second, when I extend the model to include both current and past income, I find that the latter has a significant negative impact on current happiness, indicating that income aspirations may indeed adapt quite rapidly, effectively 'raising the bar' for the level of income that is necessary to attain a given level of happiness. Upon testing the role of two-sided uncertainty as opposed to downward income losses, I find that the effect of downward vulnerability on happiness is more evident, in line with the strand of the literature that argues in favor of loss-aversion as an appropriate characterization of workers' preferences. Finally, when I contrast the results obtained with the panel and the cross-sectional measure of vulnerability, I find that the most important patters are confirmed.

Aside sheer scientific interest in documenting a potentially important and so far unexplored mechanism, the results of this analysis bear important policy implications that may generalise well-beyond the African context. In fact, most of the happiness literature focuses on developed countries and the novelty of this analysis will certainly contribute to the general debate. in particu-

lar, uncovering whether income vulnerability has a direct impact on people's life-satisfaction will help motivate policy interventions whose aim is to reduce people's exposure to risk. Conversely, if workers directly suffer from being exposed to downside risks, non-Rawlsian models of growth where "someone may be left behind" could fail to enhance general welfare if the risk of "falling behind" is high enough and sufficiently widespread among the population.
The chapter is structured as follows. Section 2 outlines the existing empirical literature on life-satisfaction, with the aim to motivate the main hypotheses we set out to test. Section 3 presents the dataset and reports descriptive statistics for the variables of interest, including our experimental measure of risk-aversion. Section 4 outlines the empirical strategy in two steps. First, it explains the methodology to estimate income vulnerability; second, it outlines the happiness model and how it changes according to the hypothesis we are aiming to test. Section 5 presents and discusses the results. Section 6 concludes.

2 Related literature

There now exists a substantial body of literatures that investigates the covariates of life satisfaction, in both developed and developing countries. Both pecuniary and non-pecuniary factors have been explored. Among the latter, health, marriage, unemployment and institutional arrangements have received particular attention (Layard, 2005; Frey and Stutzer, 2002). Among the former, the literature has focused with particular emphasis on the role of income. The question of whether income matters for happiness, in fact, has been inspiring social and economic research for a long time (Powdthavee, 2010; Blanchflower and Oswald, 2002).

Two separate stylized facts, with opposing prima facie interpretation, have shaped the modern debate. First, cross-sectional studies almost always report a positive correlation between income and life satisfaction (Frey and Stutzer, 2002). De Neve and Cooper (1998), for example, summarizing 85 independent studies on the subject quote a mean positive correlation coefficient of about 0.17. Second, aggregate life satisfaction time series for rich, growing economies tend to be fairly flat (Easterlin, 1995; Blanchflower and Oswald, 2002; Layard, 2005). If, at one point in time, richer individuals tend to be more satisfied with their lives, why do generalized income gains fail to deliver improvements in generalized life satisfaction? One possible answer to this question is to debate the causal interpretation of the coefficient on income on estimated life satisfaction equations. Ferrer-i Carbonell and Frijters (2004), using data from the German Socio-Economic Panel, emphasize the importance of time-invariant un-observables as drivers of the income-life satisfaction correlation. Controlling for individual fixed effects reduces their estimated coefficient on income by two thirds. Similarly, Powdthavee (2010) finds his FE estimates to be much below those from pooled OLS for respondents of the British Household Panel Survey. "Personality bias" is often quoted to explain these results: individuals who earn higher income tend to show a number of "positive" psychological traits, which are also conducive to life satisfaction. Failure to account for un-observed fixed personality may significantly bias the coefficient on income in an OLS regression.

Increases in income may also be associated with time-varying omitted variables, for example more stressful working conditions or less leisure (Powdthavee, 2010). Cassar (2010) presents evidence for both procedural preferences and lack of facilities determining Chilean's workers job-satisfaction, findings that had been previously established for workers in developed economies (Blanchflower, 2000). Hence it is not surprising that when income is instrumented, IV and FE-IV estimates are larger than OLS (Powdthavee, 2010; Knight, 2008), although, clearly, the validity of the instruments for income is an issue. In general, the question of time-varying omitted variables is an important one the researcher has to confront.

A second possibility to explain the two stylized facts mentioned above is to invoke a model whereby life satisfaction is determined by the the gap between realized and endogenously determined aspired income. Income aspirations can be formed by comparison with a relevant reference group, or with own past or current levels of income (Easterlin, 2001; Frey and Stutzer, 2002). At a given point in time, a rise in income will reduce the "aspiration gap" between actual and desired income and it will hence have a positive effect on life satisfaction. But this effect will not be long lasting- and hence it will fail to show in time series of aggregated life satisfaction- if the desired level of income will soon increase due to adaptation to current income levels or if the mean income of the reference group is also increasing. Evidence for adaptation to current income has been presented by Di Tella, Haisken-De New, and MacCulloch (2007) for Germany and by Knight (2008) for China. In particular, Di Tella, Haisken-De New, and MacCulloch (2007) argue that *"the size of adaptation is sufficiently large that no significant income effects on happiness remain after the fourth year" (pp.2-3)*. Similarly, evidence on the importance of relative income to life satisfaction has been presented both for developed (Blanchflower and Oswald, 2002; Luttmer, 2005) and developing (Kingdon and Knight, 2004) countries.

Despite the fast pace at which the literature is growing, the effect on life satisfaction of the uncertainty and downside risk associated with future outcomes, and in particular income, has not yet been explored. This is perhaps not surprising, given the difficulties of estimating the probability distribution of stochastic outcomes. Nevertheless, it stands out as an important area for research, especially in a developing country context, given what we know about the high variability of income of poor people; the widespread preference for less risky and loss-safe outcomes and the importance that risk and vulnerability have in determining a number of livelihood, technology and asset holding choices. Let us look at these three issues in turn.

First, the poor in developing countries face substantial risk, especially income risk (Fafchamps, 2009). The increasing availability of panel data for developing countries has showed researchers widespread idiosyncratic income variability and substantial movement across time of the same households in and out of income poverty (Dercon, 2004). Evidence for substantial downside risk has been collected in a number of papers showing persistent income effects following shocks (Dercon, Hoddinott, and Woldehanna, 2005).

Second, there exists extensive evidence on the importance of risk- and (perhaps more interestingly for our purposes) loss-aversion, that is aversion to downside risk, among the poor in developing countries. Binswanger (1980) pioneering field experiments with Indian farmers first documented risk aversion among the poor in a developing country. Liu (2008), Falco (2010) (chapter 1 of this thesis) and Caria (2009) present recent evidence of (surprisingly similar levels) of risk aversion among rural Chinese and both urban and rural Ghanaians. Bellemare, Barrett, and Just (2009) estimate price risk aversion for Ethiopian farmers and argue that farmers would be willing to pay a substantial amount of their income in order to stabilize prices at their mean level. Liu (2008) also experimentally documents loss aversion. Tversky (1991) and Fafchamps (2009) argue that individual preferences are more driven by loss than risk aversion.

Finally, risk and downside risk have been shown to influence livelihood strategies, technology adoption and asset holding strategies. Dercon (1996) shows how Tanzanian farmers with less protection against shocks prefer to plant drought resistant crops. Dercon and Christiaensen (2007) find evidence, again in Tanzania, of consumption risk discouraging technology adoption. Rosenzweig and Binswanger (2003) provide evidence for Indian farmers that the composition of productive and non productive asset holdings varies with idiosyncratic weather risk. Fafchamps (2009, 2003) provides a summary of the literature and some further insights. Holzmann and Jorgensen (2000) and Chaudhuri, Jalan, and Suryahadi (2002) make good normative and instrumental cases for considering vulnerability reduction, and in particular, the promotion of comprehensive safety nets, as a priority for development policy.

The overall conclusion that emerges from this literature is that workers in developing countries are exposed to widespread and often severe downside risk, which they generally face in the absence of formal insurance mechanisms. The second important conclusion is that loss-aversion is an important aspect of their decision making (as highlighted by Fafchamps (2009)). These two observations taken together motivate the approach followed in this paper, where a measure of *downward risk* in workers' earnings will be directly constructed and its impact on workers well-being will be explicitly analysed. In the last section I will also assess whether two-sided risk, rather than downward losses, have a different impact on workers' livelihood.

3 Data

The Ghana Urban Household Panel Survey ('GUHPS') has been conducted by the Centre for the Study of African Economies in the cities of Accra, Kumasi, Takoradi and Cape-Coast since 2004. It has run annually since then and it now spans six years.[1] Panel datasets of this length are unusual in developing countries, and are particularly uncommon for Africa.

A module on subjective well-being was added to the GHUPS questionnaire in 2005 and it was administered in every subsequent wave with the exception of 2007. The questions that compose the module were designed to be in line with the existing literature on subjective well-being. For the purpose of this analysis, we will focus on the answers to the following two questions: (a) *"All*

[1] There was one exception: the survey was not conducted in 2007, but information for that year was collected in 2008 as a 'recall' questionnaire. However, due to the low reliability of retrospective questions on subjective well-being, the happiness module was not part of this recall questionnaire.

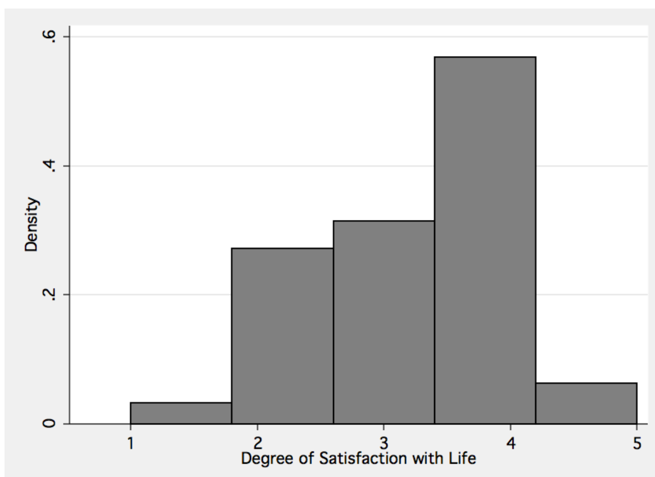
things considered, how satisfied are you with your life as a whole these days?"

(b) *"All things considered, how satisfied are you with your current work?"*

In both cases, the options given to respondents were: *"1.Very Dissatisfied, 2. Dissatisfied, 3. Neither Satisfied Nor Dissatisfied, 4. Satisfied, 5. Very Satisfied"*. Figure 1 depicts the distribution of answers. Responses appear to be skewed towards positive values. For our quantitative analysis, we attribute numerical values on a scale from 1 to 5 to these answers, where 1 corresponds to "Very Dissatisfied" and 5 to "Very Satisfied".

A selection of key summary statistics for the pooled sample over all survey waves is presented in Table 9 of the appendix.[2]

Figure 1: Distribution of Life-Satisfaction



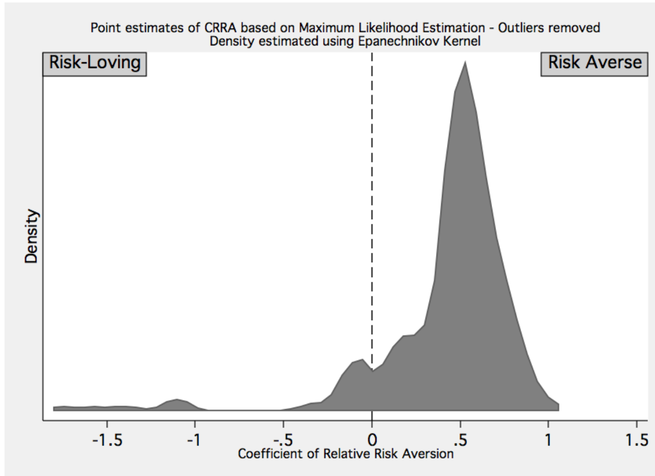
Moreover, this analysis will make use of an experimental data component that distinguishes the GUHPS panel from most available household data-sets from

[2] Our analysis will focus on paid workers, for whom income is observed and for whom we are able to construct a measure of income vulnerability. Workers who are unemployed, unpaid or out of the labour force are therefore excluded from the current analysis and will deserve full attention in a subsequent study that will attempt to incorporate vulnerability to unemployment.

both developed and developing countries. In 2007, a random sub-sample of GUHPS respondents was invited to participate in a behavioural experiment designed to elicit their attitudes to risk. The experiment, already extensively described in Chapter 1, consisted of 21 choices between pairs of monetary lotteries. Each 'game' was framed as a choice between two opaque urns containing marbles of different colours (and, correspondingly, different monetary values).[3] After being shown the composition of each urn, respondents were asked to choose the one from which they would prefer to draw a marble. Prior to making their choices, they were informed that at the end of the game one of their 21 preferred lotteries would be randomly selected and played out. The winnings of that game would then be paid to the respondent. Monetary incentives of this kind are used to induce truthful revelation of preferences. Following the same procedure as in chapter 1, we use this data to estimate a choice model based on an underlying utility function (assumed CRRA) using maximum likelihood techniques. This allows us to obtain an estimate of the individual coefficients of risk-aversion that best fit the individual series of 21 choices. Details on the estimation procedure are further outlined in Harrison (2008) and the first chapter of this thesis. Figure 2 shows the distribution of estimated coefficients of relative risk-aversion. Values below 0 indicate risk-loving behaviour, while values between 0 and 1 indicate moderate risk-aversion. These estimates are comparable to the results obtained by similar studies both in Ghana (Caria, 2009) and elsewhere (e.g. Liu (2008) on China).

[3] A detailed description of the experimental setup is contained in (Barr, 2007). "Attitudes to Risk in Ghana: Field Manual." Unpublished.

Figure 2: Distribution of Risk-Aversion



4 Empirical methodology

4.1 Constructing a vulnerability indicator

This section outlines the methodology to construct the vulnerability indicators used in the remainder of the analysis. For a detailed discussion of the relative merits of different vulnerability indices, the reader is referred to the survey paper by Ligon and Schechter (2004). The analysis in this chapter will mainly draw on the two measures proposed by Chaudhuri (2003) and Chaudhuri, Jalan, and Suryahadi (2002). The former relies on time-series variation in individual earnings and suits particularly well the characteristics of our dataset, where subjective well-being is recorded for the same individuals over a number of consecutive years, in addition to income and other worker characteristics from which we can model vulnerability. The latter method will produce a benchmark index that attempts to model cross-sectional variation and infer from it the degree of individual vulnerability. Ligon and Schechter (2004) compare the performance of these two (and several other) vulnerability indices via Monte Carlo simulations and their conclusion is in favour of the panel approach as the best performing indicator of actual vulnerability.[4]

4.1.1 PANEL Approach - *Chaudhuri (2003)*

The income vulnerability of a worker at time t is defined as the probability that the worker's income will fall below a certain threshold (z) next period. Let ν_{it} be the inverse of vulnerability, that is, i's likelihood at t of earning an income above z at $t+1$:

$$\nu_{i,t} = Pr(y_{i,t+1} > z) \tag{1}$$

Following standard Mincerian earnings analysis, assume that income is generated by the following process:

$$ln(y_{i,t}) = \delta X_{i,t} + \eta_i + \tau_t + e_{i,t} \tag{2}$$

where X_{it} is a bundle of observable characteristics, η_i is an individual unobservable fixed effect, τ_t captures time-effects that are common across workers (e.g. aggregate income growth factors and common shocks) and e_{it} is a stochastic component.

[4] This is a sensible conclusion, considering the likely presence of unobserved individual fixed effects that cannot be controlled for in a cross sectional model of earnings and could therefore mislead the analysis of vulnerability. As part of my future research I intend to explore how the results of this analysis will change upon using new vulnerability measures, including Ligon and Shechter's own index of vulnerability (see Ligon and Schechter (2002))

Second, following Chaudhuri (2003), I will explicitly model the heteroskedasticity in the data and assume the variance of e_{it} to be a function of worker and household characteristics. As in his study, I model the log of the variance as follows:

$$ln(\sigma^2_{lny_{i,t}}) = \theta K_{i,t} + \xi_i \tag{3}$$

where $K_{i,t}$ may or may not contain additional workers' characteristics, outside the set $X_{i,t}$, depending on our priors on the role that certain workers' traits will play in determining earnings volatility over and above earnings levels, and ξ_i is an individual fixed-effect in the model of income variance.

The variance of the stochastic component can be modeled empirically using the log of first-stage residuals from the earnings model:

$$ln(\hat{e}^2_{i,t}) = \theta K_{i,t} + \xi_i + \omega_{i,t} \tag{4}$$

given that:

$$\frac{1}{T}\sum_{t=1}^{T} \hat{e}^2_{i,t} \rightarrow_p \sigma^2_{lny_{i,t}} \tag{5}$$

Assuming income to be (log)normally distributed and Φ to be the cumulative distribution function of the log normal distribution, I can now calculate the individual vulnerability index using the following expression. [5]

$$\nu_{it} = Pr(ln(y_{it}) > ln(z)|X_{it,}K_{it}, \hat{\delta}, \hat{\theta}) = 1 - \Phi\left[\frac{ln(z) - \hat{\delta}X_{it}}{\hat{\theta}K_{it}}\right] \quad (6)$$

The above measure of vulnerability is meant to capture income-fluctuations without explicitly trying to differentiate transitory from permanent shocks. In the presence of an effective saving technology, the two are likely to have different impact on welfare, since transitory fluctuations can be smoothed out through precautionary savings. This is unlikely to be the case in an economy where saving and formal insurance devices are generally lacking and accumulated wealth is limited. In such an economy transitory shocks can be expected to have a significant impact on consumption and the use of the vulnerability index above appears to be justified. Moreover, attempting to explicitly separate the permanent from the transitory component of income variation (e.g. following the approach by Meghir and Pistaferri (2004)) would pose major challenges given the (short) length of the GHUPS panel, though it clearly remains an open alley for future research.[6]

[5] The reader should note that, differently from the definition in 1, our estimates of vulnerability are obtained as the probability of falling below the poverty line given worker characteristics at t, rather than $t + 1$. This choice was made following the intuition that workers are most likely to assess their future prospects on the basis of their current characteristics, some of which might themselves be stochastic and subject to unpredictability.

[6] Similarly, it will be interesting to extend the current analysis to incorporate data on savings and wealth that will become available in the next-coming waves of GHUPS, which may enable us to control for precautionary savings.

4.1.2 CS Approach - *Chaudhuri et al. 2002*

This section outline an alternative cross-sectional approach to estimating vulnerability, based on Chaudhuri, Jalan, and Suryahadi (2002), which will serve as a useful benchmark in the remainder of the paper. The method is very similar to the panel approach outlined earlier, except for the fact that it relies on cross-sectional variation in earnings to obtain measures of income vulnerability. It may therefore suffer from inconsistency due to confounding individual fixed effects in the earnings model. Comparing the results with the ones we obtained from the panel-approach, our aim is precisely to assess the strength of such concerns. The main advantage of this methodology is that it allows us to gain precision, using a three-step feasible generalised least squares estimator (FGLS).

The definition of income vulnerability, as well as the income and variance models remain the same as in the previous section. Differently from above, I now take each of the available cross sections in isolation. For each one of them, consistent estimates of δ and θ can be obtained with a three-step feasible generalized least squares estimator (FGLS) (Chaudhuri, Jalan, and Suryahadi (2002)). First, I estimate equation (2) using Ordinary Least Squares (on each cross-section separately). Next, I obtain the squared residuals from the first stage (which now contain both the idiosyncratic component and the fixed effect, i.e. $u_{it}^2 = (\hat{\eta}_i + \hat{e}_{it})^2$) and use them as the dependent variable in a second stage model of the variance:

$$\hat{u}_{it}^2 = \theta K_{it} + \omega_{it} \tag{7}$$

At this point I can obtain predicted values for the variance of income - $\hat{\theta}K_{it}$ and use them to adjust equation (7) as follows:

$$\frac{\hat{u}_{it}^2}{\hat{\theta}K_{it}} = \frac{\theta K_{it}}{\hat{\theta}K_{it}} + \frac{\omega_{it}}{\hat{\theta}K_{it}} \tag{8}$$

From this adjusted equation, Chaudhuri, Jalan, and Suryahadi (2002) show that one can estimate the asymptotically efficient FGLS estimator - $\hat{\theta}_{FGLS}$. The next step is to obtain a consistent prediction for the variance of income, $\hat{\sigma}^2 = (\hat{\theta}_{FGLS}K_{it})$ and use it to adjust equation (2):

$$\frac{ln(y_{it})}{\hat{\sigma}^2} = \frac{\delta X_{it}}{\hat{\sigma}^2} + \frac{e_{it}}{\hat{\sigma}^2} \tag{9}$$

which delivers $\hat{\delta}_{FGLS}$.

As above, having obtained asymptotically efficient estimates of δ and θ, I can now obtain consistent predictions for the first two moments of the income distribution:

$$E(ln(y_{it})|X_it) = \hat{\delta}_{FGLS}X_{it} \tag{10}$$

$$V(ln(y_{it})|K_it) = \hat{\theta}_{FGLS}K_{it} \tag{11}$$

and given the same assumption of (log)normally distributed earnings, we can obtain vulnerability as in (6).

4.2 A model of happiness

Having constructed a measure of vulnerability, I can now explore its relation with subjective-well being. The following equation describes my workhorse model of happiness:

$$h_{i,t} = \beta y_{i,t} + \gamma \nu_{i,t} + \delta Z_{i,t} + \kappa_i + \epsilon_{i,t} \tag{12}$$

where h_{it} is worker i's level of life satisfaction in period t, y_{it} is income at time t and ν_{it} is the index of (the inverse of) vulnerability in the same period; Z_{it} is a vector of worker characteristics that are expected to be correlated with life-satisfaction. κ_i is an unobserved happiness fixed-effect that accounts for unobserved traits that make an individual naturally more (or less) prone to be satisfied with his/her life (e.g. optimism). My main hypothesis is that β and γ are positive (recall that $\nu_{i,t}$ is the inverse of vulnerability and, hence a 'good' in this specification): *increasing income and decreasing vulnerability enhances life satisfaction*. In order to test it, I will attempt to overcome a number of econometric issues that complicate identification.

First, a number of both time-varying and time-invariant determinants of happiness may be correlated with income and vulnerability. If omitted from the analysis, those variables may bias the results. Among the time-invariant factors, one can think of *personality traits* or endowments of *social and human capital*, which may have a direct impact on life-satisfaction. More extroverted and optimistic individuals, for instance, may be both 'naturally' satisfied with their life and more likely to find good, secure employment, or, equally plausibly, more willing to face the risks and uncertainty of entrepreneurship. The same may hold for educated or well-connected people. Among the time varying unobservables, *working conditions* are a first, obvious source of bias.

Powdthavee (2010) argues that income gains are often correlated with deterioration in the conditions of work and the latter may have an important influence on life satisfaction. Vulnerability might also be correlated with working conditions, though I have no strong a-priori evidence of the sign of such correlation. *Relative income* is a third potentially confounding factor. Extensive empirical evidence has been generated showing that relative income is correlated with the life satisfaction of individuals in both developed and developing countries (Blanchflower and Oswald, 2002; Luttmer, 2005; Kingdon and Knight, 2004) and it is natural to assume that relative income will be correlated with absolute income and vulnerability. In this analysis I will control for these three potential sources of bias, by including in the model controls for working-conditions (proxied by a measure of satisfaction with work) and for a worker's position in the income distribution. Most importantly, the main contribution of this paper consists in the use of the panel dimension of the data to control for the unobservable and time-invariant personality traits that may constitute an important source of bias in happiness regressions. By means of *within group* and *differenced* estimators, I am able to exclude the possibility that personality traits that are time-invariant, such as innate optimism, are the drivers of the relationship between happiness and income, and between happiness and vulnerability. And, perhaps more interestingly for the advancement of the literature, I can study directly how such personality traits may bias the results if they are not controlled for.

Second, *aspirations* may play an important role in the generation of happiness. In particular, income aspirations may adapt to current conditions rapidly enough to reduce mid-run life satisfaction effects of income gains (Easterlin, 2001; Frey and Stutzer, 2002). If individuals have adaptive income aspirations that are relevant to life-satisfaction (i.e. distance from an 'desired' level of income drives their happiness), we would expect past income to negatively influence current happiness once current income is controlled for. A higher level of past income would in fact raise income aspirations and hence, ceteris paribus, *reduce* life-satisfaction for a given level of current income. As a test for the presence of aspiration effects, I will include lagged income in the following model specification:

$$h_{it} = \beta y_{i,t} + \gamma \nu_{i,t} + \beta'' y_{i,t-1} + \delta X_{i,t} + \kappa_i + \epsilon_{i,t} \tag{13}$$

Under the adaptive income aspirations hypothesis, I would expect β'' to be negative.

The third problem is methodological: life-satisfaction is generally recorded in datasets like GHUPS as a categorical variable. Modeling it as a discrete (ordered) outcome would, therefore, appear to be the most appropriate approach. However, such approach would not easily lend itself to controlling for those time invariant unobservables that I have argued are of great relevance in the determination of life satisfaction. To address this issue Ferrer-i Carbonell and Frijters (2004) develop a conditional estimator for the fixed effects logit model. Their findings show that *"it makes virtually no difference whether one assumes ordinality or cardinality of happiness answers, whilst allowing for fixed effects does change results substantially"* (Ferrer-i Carbonell and Frijters, 2004). It therefore seems justifiable to assume cardinality of the life satisfaction indicator and use the corresponding estimators.

Fourth, issues of reverse causality may arise in the analysis. High levels of life satisfaction may help individuals earn higher incomes or reduce their income vulnerability. Such effects may again bias the estimated coefficients β and γ. In order to fully address this problem, I would be required to specify an FE-IV regression approach. However, doubts are often raised about the validity of the instruments proposed by the authors who have attempted the IV or FE-IV approach for income such as Knight (2008); Powdthavee (2010).[7] Hence, I do not attempt to instrument vulnerability, while fully acknowledging the possibility that these concerns might be important.

As the last step in the analysis I will explore the interaction between vulnerability and attitudes to risk in generating happiness. Ceteris paribus, I would expect more risk averse people to experience lower life satisfaction when faced with a higher likelihood of poverty in the future. In order to explore this question, I augment the model as follows:

$$h_{i,t} = \beta y_{i,t} + \gamma \nu_{i,t} + \rho(\nu_{i,t} * r_i) + \delta X_{i,t} + \kappa_i + \epsilon_{i,t} \tag{14}$$

where r_i will be captured by the experimental measure of risk-aversion described in section 3. In the empirical analysis I will estimate (14) using FE and first differences; and if vulnerability was indeed affecting happiness through

[7] Furthermore, the vulnerability variable has been constructed as a deterministic function of the predicted values of an earnings model, which would complicate an IV strategy.

risk-aversion I would will expect ρ to be positive.[8]

It should finally be remarked that the vulnerability index is a non-linear function of the first two moments of the earnings distribution, which are both modeled as functions of household and individual characteristics in the first stage of the estimation. It follows that the happiness model (where we include both income and vulnerability on the right-hand side) contains *two* functions of those characteristics among the regressors. Separate identification of these two functions implicitly relies on assumptions regarding the relationship between income and well-being. Existing studies have often imposed *linearity* on the relationship and, for comparability, we choose the same approach. Fafchamps and Shilpi (2008, 2009) report non-parametric results that show a linear relationship between consumption expenditures and subjective satisfaction with consumption levels, lending empirical support to this modeling choice.

5 Results

5.1 Vulnerability

5.1.1 PANEL Approach - *Chaudhuri (2003)*

Table 1 shows the results from estimating the earnings and variance models used to predict vulnerability later in the analysis. For easiness of exposition

[8] The reader should be alert to the fact that throughout the analysis I have treated risk-aversion as a time-invariant individual characteristic that is innate or acquired early in life. Evidence for this hypothesis exists and was discussed in chapter 1, but I am aware of the potentially important issues of reverse causality (from vulnerability to risk-attitudes) that may confound the results. Data-limitations hinder my current ability to deal with this issue, which will receive further attention in future research, as new experimental data becomes available.

I have chosen to report only my preferred specification for the income model (col 1), while I document the results from several specifications of the variance model that may be informative to the reader (col 2 -5). The first feature of the results is that while the income model (col 1) shows a relatively high predictive power, trying to predict the variance of earnings proves to be a much more challenging exercise. This is to be expected, given that part of what appears to be true variation in earnings may in fact be due to random measurement error. Upon experimenting with different specifications, I conclude that the best model is one that controls for individual fixed effects (col 5) and for a set of key time-varying covariates, the choice of which is grounded in a long-established literature on mincerian earnings regressions (see Rankin, Sandefur, and Teal (2010) for an application on Ghana using the GHUPS dataset). The results in col 1 confirm a number of standard patterns observed in related studies of earnings in Sub-Saharan Africa. First, I find a statistically significant effect of firm-size on earnings (captured by positive coefficients on the log of firm-size for wage-employees and on the log of the number of hired employees for the self-employed). Second, I detect a sizable civil service premium and a positive premium for longer tenure in the job; and while the linear effect of age cannot be estimated when time-trends are also controlled for, I am able instead to capture the concavity of the age-earnings profile (albeit the coefficient appears to be insignificant). The reader should notice that since estimation in col 1 is carried out with controls for individual fixed effects, it is not possible to separately identify the coefficients on time-invariant characteristics such as education and gender.

It is also worthwhile highlighting that, upon carefully scanning the information available in the GHUPS dataset, I identified two sets of worker characteristics that may theoretically drive the variance of earnings. The first one is *ethnicity*, potentially reflecting the idea that social networks provide an important buffer against negative shocks and can 'insulate' one's earnings through several channels (e.g. informal lending to cover variable business costs). The strength of one's network largely depends on his family ties, which in the Ghanaian context are highly intertwined with tribal and ethnic background, and different ethnic groups may be able to count on support networks of different strength. Second, respondents' *marital status* may also be thought to have an impact on earnings variability, since being married presumably changes the degree of income uncertainty a person faces. For instance, forming a family is likely to changes the risk-management strategies of income earners as they become responsible for a larger group of people. Interestingly, I find that ethnic ties are indeed a significant predictor of the variance of earnings, while marriage does not appear to play a significant role in the process (col 2, table 1). As both of those traits are time-invariant in our datasets (only a handful of people changed their marital status throughout the period), they will be effectively controlled for by simply including individual fixed effects (as in col 5); yet, this discussion may help shed some light on the nature of time-invariant unobservables.

Given the estimates from the earnings and earnings variance models, before I can obtain an estimate of vulnerability I will need to define the low-earning threshold (alternatively referred to, with what I acknowledge might be an abuse of terminology, as 'poverty line'), z. Figure 3 shows the percentage of people who, in every year, are below different income thresholds, while figure 4 shows the resulting cumulative distribution of the vulnerability index. As one would expect in the presence of real earnings growth, the percentage of people below this threshold fell from above 40 % in 2004 to about 35 % in 2009. My choice for the remainder of the paper is be to set z = 10 real (1997) GHC, which approximately translates into 40 (1997) USD (approx. 1.5 USD per day). While ultimately being an arbitrary decision, this choice is grounded in the proximity to the widely used measure of 1 Dollar per day. When we experimented with alternative lines in the vicinity of this value, the main patterns in our results did not change substantially. [9]

[9] These figures are unadjusted for PPP; and the reader should be alerted to the fact that in 2007 the Ghana Cedi was converted into the New Ghana Cedi at a rate of 10,000 Ghana Cedi to 1 New Ghana Cedi. All the analysis in this paper is conducted in New Ghana Cedi (loosely referred to as Ghana Cedi in the remainder of the paper), into which also the 1997 (pre-reform) figures have been converted for uniformity.

Table 1: Model of Earnings and Earnings Variance - PANEL METHOD

	lnY	lnVAR-K	lnVAR-X	lnVAR-XSQ	lnVAR-XFE
	(1)	(2)	(3)	(4)	(5)
Age		.035	.019		-.0005
		(.028)	(.026)		(.001)
Age2	-.0003	-.0004	-.0002	-5.49e-08	-.0005
	(.0004)	(.0003)	(.0003)	(4.76e-08)	(.001)
Educ		-.028	-.055		
		(.034)	(.032)*		
Educ2		.003	.004	7.17e-06	
		(.002)	(.002)*	(8.15e-06)	
Male		.134	.133	.127	
		(.095)	(.092)	(.092)	
Priv Wage	-.147	-.991	-.944	-.967	-.234
	(.074)**	(.149)***	(.145)***	(.121)***	(.177)
Civil or Pubent	.196	-1.342	-1.319	-1.313	-.120
	(.112)*	(.184)***	(.180)***	(.177)***	(.269)
Ln(employees)	.187	.008	.040	.044	-.030
	(.051)***	(.115)	(.113)	(.057)	(.121)
Ln(firmsize)	.055	.005	-.013	-.0003	.062
	(.021)**	(.042)	(.042)	(.008)	(.051)
Yrs since curr. job start	.005	.003	.004	.0002	.005
	(.003)*	(.006)	(.006)	(.0002)	(.007)
Married		-.052			
		(.095)			
Eth.: Ga-Dangme		-.037			
		(.119)			
Eth.: Ewe		.392			
		(.171)**			
Eth.: Mole Dag. - Hausa		.578			
		(.155)***			
Other ethnicity		-.181			
		(.162)			
IndFE	Yes				Yes
Time Dum.	Yes	Yes	Yes	Yes	Yes
Const.	2.588	-3.041	-2.642	-2.514	-3.293
	(1.346)*	(.519)***	(.491)***	(.190)***	(1.294)**
Obs.	3659	3014	3110	3110	3110
R^2	.685	.073	.065	.064	.627

lnY: Log of Real Earnings; lnVAR-K: LogVariance modeled as a function of K; lnVAR-X: LogVariance modeled as a function of X; lnVAR-XSQ: LogVariance modeled as a function of X^2 and lnVAR-XFE: LogVariance modeled as a function of X plus individual fixed effects (used to compute vulnerability subsequently in the paper); X is the set of key regressors in the income model, K is an augmented set of regressors to include potential determinants of the variance; omitted occupational category = self-employed; omitted ethnicity = Akan; *Confidence:* *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%; Robust standard errors in parentheses.

Figure 3: Percentage of employed with $y < z$

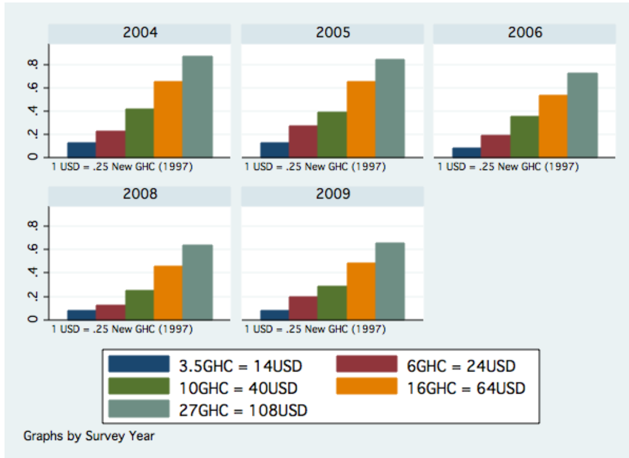
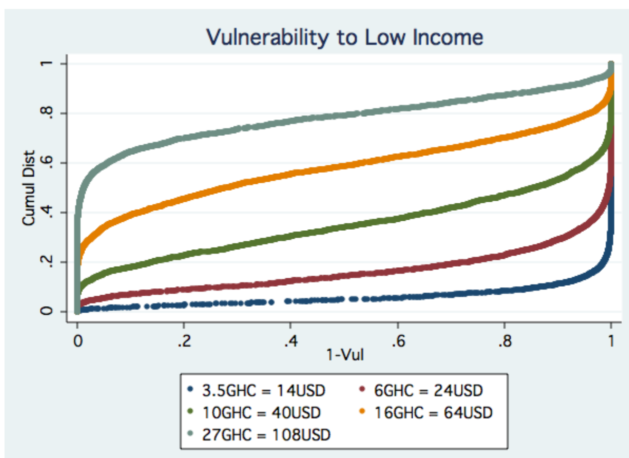


Figure 4: Cumul. Dist. of Vulnerability for different poverty lines (z_t)



5.1.2 CS Approach - *Chaudhuri et al. 2002*

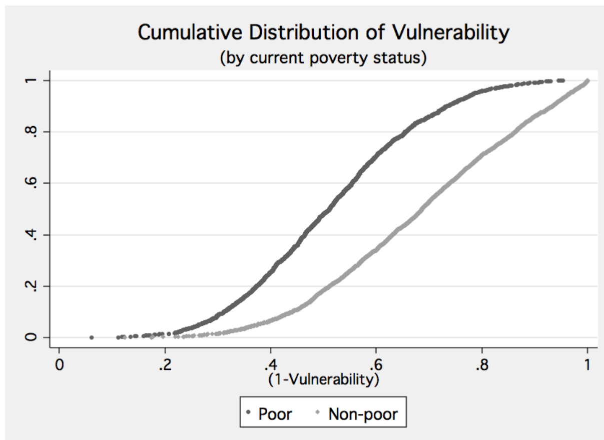
Detailed results from the four estimation steps in the CS approach for the estimation of vulnerability are reported in the appendix (Table 10 to 13) and in the interest of conciseness I will not fully discuss them. It will suffice to note that my choice of regressors for the earnings and the variance model follows the same principles as in the panel approach and, as in the previous section, I find that while earnings regressions (Step 1 and 3) show a relatively high predictive power, trying to predict the variance of earnings (Step 2 and 4) is considerably more difficult.

More interestingly, when I plot the cumulative distribution of the estimated levels of vulnerability for z = 10 (1997) Cedis per month (Figure 5) and I compare it to the same distribution in the previous section (the central line in Figure 4), I find that vulnerability to being poor next period is now much more widespread (both among the current poor and the current non-poor [10]) than it was according to the estimates in the previous section (where a sizable proportion of respondents was clustered at $\nu_{it} = 0$ and $\nu_{i,t} = 1$, as shown by the lower and upper tails of the lines in Figure 4). This is in itself an interesting result, drawing a clear distinction between the two methodologies, though one that might be expected, considering that vulnerability estimated through a cross-section compounds true idiosyncratic variation and unobserved heterogeneity across individuals that cannot be controlled by means of fixed-effects estima-

[10] Vulnerability is obviously lower among the currently high-earners, but still considerable, indicating that the risk of falling into poverty between periods is widespread across the labour market (50% of the currently poor face a likelihood of 50% or higher to be poor next period; the same level of risk is faced by approximately 20% of the currently non-poor).

tors. As mentioned above, one of the goals of this study is to assess whether not controlling for such potentially confounding effects will make a significant difference in the results of our happiness model.

Figure 5: Cumul. Dist. of Vulnerability for z_t = 10 (1997) GhCedis - *CS approach*



5.2 Happiness

This section will present the results from estimating the happiness model. Figure 6 plots the histogram of happiness responses that was presented in section 3, after now splitting the sample by low/high income relative to the poverty line (referred to as 'income poverty' in the figures). The histogram shows prima-facie evidence for the link between income and happiness that I am attempting to formally test, with people who are above the low-income threshold more likely to report to be "satisfied" with their life. More interestingly perhaps, figure 7 plots changes in happiness against changes in (the inverse of) vulnerability (obtained through the Panel Approach), showing a positive, albeit weak, relationship, that I take as a point of departure.

Figure 6: Happiness and Income

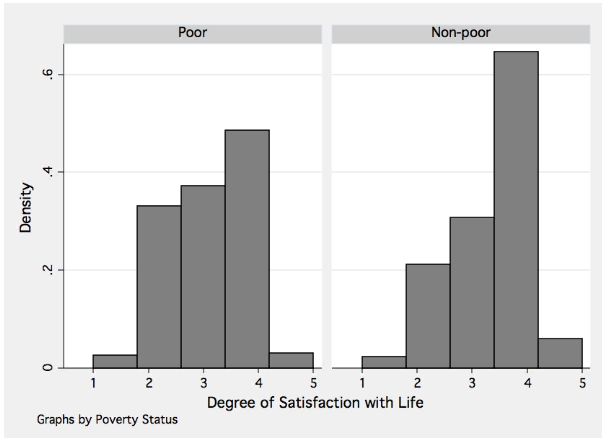
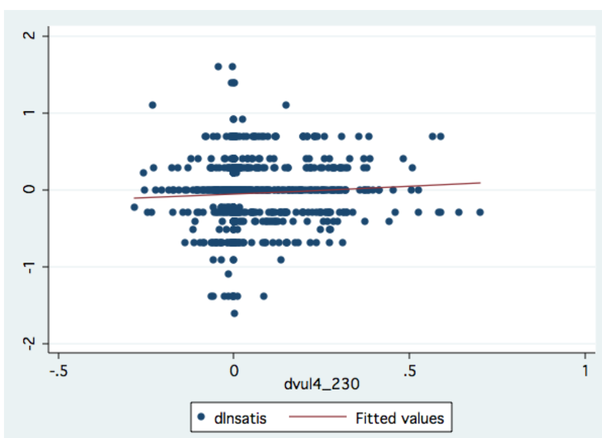


Figure 7: Changes in Happiness and Vulnerability



5.2.1 PANEL Approach - *Chaudhuri (2003)*

Table 2 reports the results from estimating the workhorse model of happiness (equation 12) using first OLS (col 1-2) and then controlling for fixed effects by means of a 'Within Group' (col 3-4) and a 'First Difference' transformation (col 5-6).

My first result is a positive and significant effect of absolute income on life-satisfaction, in line with the existing literature (De Neve and Cooper, 1998). This relationship is evident in OLS and, rather strikingly, it does not change significantly once I control for fixed effects. It appears, therefore, that time-invariant unobservables correlated with earnings are not biasing upwards the effect of income on happiness as one may expect. This constitutes evidence against the hypothesis that individuals who are 'naturally' more positive and optimistic (and hence tend to be 'naturally happier') tend to achieve higher earnings. Interestingly, the size of the estimated coefficient on the log of income under WG (1) - 0.017 - is remarkably close to that estimated by Powdthavee (2010) using data from the British Households Panel Survey and a fixed-effect estimator- 0.019.

On the other hand, individual fixed effects appear to play a crucial role in the relationship between vulnerability and life-satisfaction. The most important finding in this paper is a strong negative effect of vulnerability *over and above* the income effect just described (notice that in the regression tables this is reported as a positive relationship between the *opposite of vulnerability* and happiness). Such effect is absent in OLS, where we record an insignificant coefficient, but it becomes strong and significant once fixed effects are controlled for. As outlined in the previous section, my measure of vulnerability is both individual-specific and time-varying. It is therefore possible to control for individual fixed effects in Table 2 (moving from col 1 to col 3) and still be able to estimate the effect of vulnerability on earnings. The fact that the main result only appears upon controlling for fixed effects is an indication that unobserved time-invariant worker characteristics driving life-satisfaction are also playing a role in determining vulnerability, albeit one that appears difficult to characterise and may seem counter-intuitive at first. The negative bias in the coefficient of the (inverse of) vulnerability in table 2 should be read as evidence that the time-invariant traits that induce people to be 'naturally' happier (high κ_i) correlate positively with the amount of volatility attached to one's earnings.[11] This could be the result, for instance, of innately optimistic people seeking employment opportunities that are riskier, a hypothesis that we do not deem unreasonable. The estimation includes controls for work satisfaction (proxying changes in working conditions), income quartile, age and its square and marital status. Work satisfaction is closely correlated with life satisfaction and shows by far the biggest positive coefficient in the life satisfaction regression. On the other hand, the income quartile the respondent belongs to does not show a significant effect.[12] Furthermore, absolute income remains a significant driver of happiness and the strong role of vulnerability changes negligibly. The result suggests that it is not a respondent's rank in the income distribution, but rather his/her level of earnings what really matters for life-satisfaction. This

[11] They correlate negatively with the 'inverse' of vulnerability. Hence, they correlate positively with vulnerability

[12] In addition to what is reported in the table, I experimented with finer quantile disaggregation (quintiles and deciles) and the main results outlined above did not change.

contradicts some of the established evidence on the role of relative income for life satisfaction. However, it should be remarked that the relevant reference group may be a subset of the whole sample, cutting across income quantiles. Urban Ghanaians may, for instance, compare their income to that of people in the same neighborhood, social class or ethnicity. If so, the position in the overall distribution may not matter significantly. Having recently been extended to include data on self-reported tendencies to compare oneself to others, the GHUPS should soon enable us to test these hypotheses more directly.

The vulnerability index has been constructed using estimates from a first-stage model of earnings. Hence, it carries a certain degree of statistical imprecision that could pose a challenge to the significance of my estimates in the second stage model of happiness. In order to check the robustness of my results to such concern, I have bootstrapped the entire estimation sequence (including the first stage to construct the vulnerability index), sampling with replacement to obtain 200 replications of the original sample. The results are summarised in Figure 8 and 9, where I have plotted the distribution of the bootstrapped coefficients on Income and on (the inverse of) Vulnerability, and they are consistent with the discussion so far. The effect of income on happiness remains statistically different from 0 in every specification. The effect of vulnerability, on the other hand, is not significant in OLS and it becomes significant once the fixed effects in the happiness model are controlled for.

Table 2: Life-Satisfaction and vulnerability - *Panel Approach*

	OLS1	OLS2	WG1	WG2	FD1	FD2
	(1)	(2)	(3)	(4)	(5)	(6)
(1-Vul)	-.006	.005	.087	.117	.154	.135
	(.017)	(.018)	(.036)**	(.044)***	(.082)*	(.085)
LnRealEarn	.013	.033	.017	.051	.024	.042
	(.006)**	(.012)***	(.008)**	(.017)***	(.011)**	(.023)*
LnWorkSatis	.618	.615	.588	.587	.566	.566
	(.014)***	(.014)***	(.025)***	(.025)***	(.026)***	(.026)***
Married		.027		.023		.020
		(.010)**		(.020)		(.031)
Age		-.006		-.019		-.049
		(.003)*		(.014)		(.093)
Age2		.00007		.0002		-.0002
		(.00004)*		(.0002)		(.0005)
EarnQuart=2		-.0009		-.026		-.016
		(.018)		(.025)		(.033)
EarnQuart=3		-.048		-.069		-.034
		(.023)**		(.034)**		(.045)
EarnQuart=4		-.060		-.099		-.065
		(.032)*		(.048)**		(.063)
Const.	.426	.489	.388	.784	-.022	.044
	(.019)***	(.059)***	(.036)***	(.273)***	(.013)*	(.089)
Obs.	2507	2507	2507	2507	832	832
R^2	.45	.454	.422	.425	.38	.382

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%; Robust standard errors in parentheses;

Figure 8: Bootstrapped distribution of the coeff on *LnRealEarn*

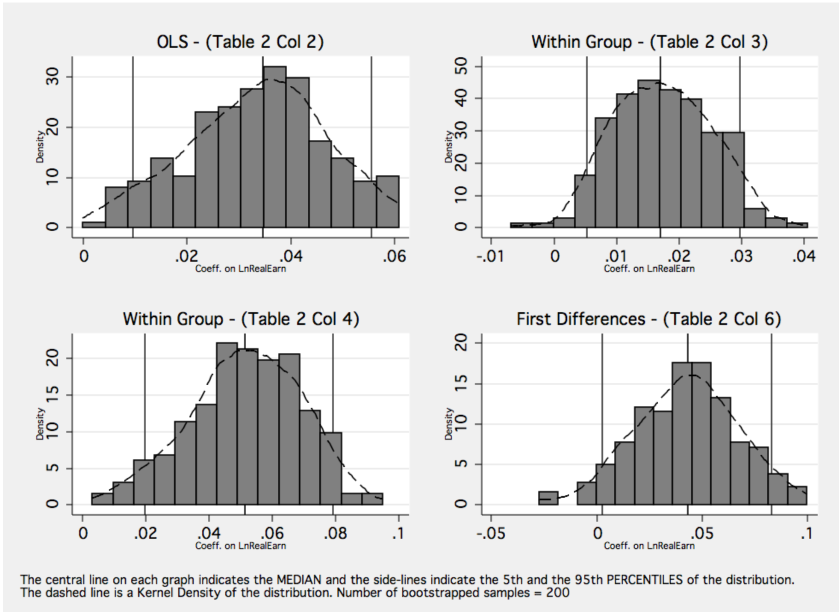
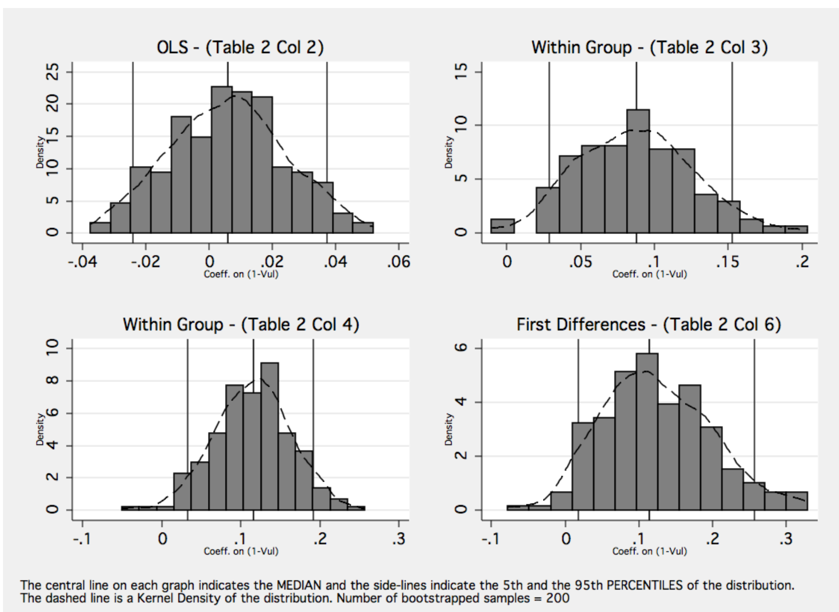


Figure 9: Bootstrapped distribution of the coeff. on *(1-Vul)*



As an extension to the main set of results, I next re-estimate the model with a lagged income term. The goal is to test whether the mechanisms of *income aspirations adaption* described in the previous section can be detected. Table 3 reports the estimates from the empirical model with a minimal set of key control variables.[13] The main result is that the coefficient on lagged income is negative and statistically significant after controlling for individual fixed effects. The result is not (statistically) very strong, which is perhaps to be expected as a one-year period may be too short to allow for sufficient adaptation. Upon experimenting with longer lags, however, the resulting drop in sample size was generally too large to draw any robust conclusions. I also attempted to test the hypothesis that people's aspirations may adapt to current vulnerability levels, but adding a lagged term for estimated vulnerability complicated the interpretation of the results and proved to be overly demanding for the data at my disposal. In conclusion, the empirical evidence lends mild support to the adaptation hypothesis, with an estimated negative effect of 1-period lagged income on life satisfaction capturing a rather fast process of aspirations revision. This prima facie evidence will need to be explored further, looking with enhanced attention at the time-series properties of the happiness series.

What drives the effect of vulnerability on happiness? In this last step of the analysis I begin to explore the plausible hypothesis that respondents' *risk aversion* plays a role in generating happiness when income is uncertain. If that was true, we would expect that for positive values of risk-aversion, higher income-

[13] This choice of regressors is dictated by the fact that upon including lagged income the available sample size drops dramatically (due to attrition across any two consecutive years), and since the aim of this analysis is mainly to provide prima-facie evidence of the relationship between lagged income and happiness, we reserve the task to investigate this hypothesis more thoroughly when new data becomes available.

uncertainty should be conducive to lower life-satisfaction. Prima-facie evidence of the potential existence of this mechanism is the estimated distribution of risk-aversion coefficients reported in figure 2, where the large majority of our sample appeared to be in the range of positive risk-aversion values (and would therefore be expected to suffer from increases in vulnerability). A formal test of this hypothesis can be performed by adding to the happiness model an interaction term between the experimental measure of risk-aversion described above and the index of vulnerability (eq. 14). The rationale is to test whether individuals with higher risk-aversion undergo heavier happiness losses for a given increase in vulnerability, ceteris paribus. The results, reported in table 4, are severely impaired in their precision by the small size of the sample for whom the measure of risk-aversion is available. The interaction effect is negative in OLS (col 1), where unaccounted fixed effects are presumably playing a similar confounding role as in our previous estimation and it becomes positive when such fixed effects are controlled for (col 2). As the coefficient is not significant, this is only a tentative indication that people with stronger aversion to risk, as measured in our behavioural experiments, enjoy higher utility gains from a given reduction in vulnerability. It is plausible, of course, to believe that risk-aversion may itself have an impact on vulnerability if workers sort into different occupations on the basis of their risk-preferences (as it was shown in the first chapter of this thesis), and in doing so they can influence *directly* the amount of vulnerability they face. One should therefore consider the possibility that causality running from risk-aversion to vulnerability (and then to happiness) may hinder our ability to separately identify the role of attitudes to risk. This is a legitimate concern, which may partly explain the weakness of the result. Similarly, endogenous self-selection into less risky jobs needs to be taken into consideration when assessing the external validity of the estimates above. Chapter 1 has shown that workers who are more risk-averse *self-select* into occupations where income uncertainty is lower, since they would suffer the most from higher uncertainty. Hence, if those workers were to find themselves in riskier jobs, they would presumably suffer from a greater utility loss than I have estimated.

In conclusion, the evidence is still insufficient to make robust conclusions on the direct role of risk-aversion and will require further investigation with additional data and new waves of experimental evidence that will be gathered in the years to come. It should also be remarked that a similar analysis should be carried out using specific measures of loss-aversion, which may be arguably more relevant if happiness is driven by fear of *downside* risk, as I will discuss in the next session. Using the experimental data to construct a measure of loss-aversion that would be amenable to an analysis like the one in this paper constitutes an open alley for future research.

Table 3: Life-satisfaction and vulnerability with lagged income - *Panel Approach*

	OLS1	OLS2	WG1	WG2
	(1)	(2)	(3)	(4)
(1-Vul)	.011	.031	.224	.225
	(.023)	(.038)	(.062)***	(.128)*
LnRealEarn	.049	.050	.049	.073
	(.009)***	(.012)***	(.011)***	(.018)***
L.lnrearn		-.015		-.040
		(.013)		(.021)*
Age	-.012	-.010	-.035	-.072
	(.004)***	(.006)*	(.018)*	(.039)*
Age2	.0002	.0001	.0002	.0004
	(.00005)***	(.00007)*	(.0002)	(.0004)
Const.	1.234	1.198	1.807	2.878
	(.072)***	(.104)***	(.364)***	(.806)***
Obs.	2507	1235	2507	1235
R^2	.027	.028	.025	.064

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%; Robust standard errors in parentheses;

Table 4: Life-satisfaction and vulnerability with risk aversion - *Panel Approach*

	OLS1	OLS2	WG1	WG2	WG3
	(1)	(2)	(3)	(4)	(5)
(1-Vul)	.008	.041	.048	-.037	
	(.063)	(.068)	(.133)	(.163)	
LnRealEarn	.056	.056	.060	.059	.059
	(.023)**	(.023)**	(.029)**	(.029)**	(.029)**
RA*(1-Vul)		-.091		.247	.200
		(.071)		(.333)	(.271)
Age	-.008	-.008	-.063	-.065	-.067
	(.012)	(.012)	(.053)	(.052)	(.049)
Age2	.00007	.00008	.0004	.0004	.0004
	(.0002)	(.0002)	(.0007)	(.0007)	(.0007)
Const.	1.155	1.156	2.621	2.648	2.697
	(.207)***	(.207)***	(.921)***	(.907)***	(.846)***
Obs.	381	381	381	381	381
R^2	.026	.031	.046	.048	.048

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%; Robust standard errors in parentheses;

5.2.2 CS Approach - *Chaudhuri et al. 2002*

In this section I will employ the measure of vulnerability obtained with the CS methodology to re-estimate the happiness model (equation 12). My main interest is to explore the role of (time-invariant) unobserved heterogeneity *from the vulnerability specification* in potentially confounding the results of the happiness regressions. The reader should pay attention not to confuse such fixed effects from the vulnerability model (which are the ones the CS approach cannot control for, η_i and ξ_i) with the fixed effects in the happiness model mentioned hereafter (κ_i).

Table 5 reports the estimates from my happiness model with vulnerability obtained through the CS methodology. The results are qualitatively very similar to the ones in the previous section. A positive and significant effect of absolute income on life-satisfaction is estimated, as well as a strong negative effect of vulnerability on life satisfaction, *over and above* the income effect.[14] As in the previous section, I next estimate the model with lagged income and vulnerability and again I find a negative and significant effect of lagged income, which supports the income adaptation hypothesis, while the strong estimated effect of vulnerability does not appear to be affected (Table 6). When I finally explored the role of risk-aversion as a potential channel through which vulnerability has an impact on life-satisfaction (as before, by means of an interaction term between risk-aversion and vulnerability in the happiness model), I again

[14] It should further be remarked how the estimated effect of vulnerability grows larger as we move from the WG to the FD estimator. This is at least partly due to the well known downward bias that affects WG estimators (the well-known 'Nickell Bias', see Nickell (1981)). However, the fact that both these estimation strategies deliver a considerably large and significant vulnerability effect is a per se important indication that vulnerability plays an important role in driving happiness.

found very tenuous prima facie evidence in support of the main hypothesis (the results are not reported for the sake of conciseness and as they do not add significantly to the discussion). As before, the precision of the results is limited by the size of the sample.

Most importantly, from comparing the main set of results in Tables 2 and 5 I conclude that the two methodologies for the construction of vulnerability (Panel and CS) deliver very similar results in the context of this happiness analysis, pointing towards the conclusion that unobserved (time-invariant) heterogeneity *in the earnings model* (again, not to be confounded with time-invariant heterogeneity *in the happiness model*, which is on the other hand very important for our results and controlled for under both approaches) does not interfere significantly with the measurement of vulnerability, at least insofar as it does not appear to make a significant difference for our second stage model of happiness. This may partly be expected, since, by definition, the individual fixed effects in the earnings model (η_i) are constant over time and hence do not form part of the idiosyncratic variation that constitutes ex-ante vulnerability. Hence, we may not expect them to have an impact on our analysis of how ex-ante vulnerability impacts happiness. This finding should contribute to the methodological debate led by Ligon and Schechter (2002, 2004) on the relative merits of different approaches to computing vulnerability.

Table 5: Life - Satisfaction and vulnerability - *CS Approach*

	OLS1	OLS2	WG1	WG2	FD1	FD2
	(1)	(2)	(3)	(4)	(5)	(6)
(1-Vul)	-.024	-.006	.141	.237	.362	.362
	(.027)	(.032)	(.055)**	(.080)***	(.111)***	(.121)***
LnRealEarn	.019	.037	.016	.048	.021	.040
	(.005)***	(.011)***	(.008)*	(.017)***	(.011)*	(.023)*
LnWorkSatis	.601	.599	.590	.589	.567	.566
	(.013)***	(.013)***	(.026)***	(.026)***	(.026)***	(.026)***
Married		.031		.025		.017
		(.010)***		(.020)		(.031)
Age		-.007		-.035		-.043
		(.003)**		(.016)**		(.092)
Age2		.00008		.0003		.0002
		(.00004)**		(.0002)*		(.0006)
EarnQuart=2		.002		-.021		-.015
		(.017)		(.026)		(.033)
EarnQuart=3		-.040		-.066		-.034
		(.021)*		(.035)*		(.045)
EarnQuart=4		-.052		-.094		-.064
		(.030)*		(.049)*		(.063)
Const.	.440	.516	.354	1.019	-.031	-.006
	(.021)***	(.054)***	(.041)***	(.313)***	(.013)**	(.091)
Obs.	2814	2814	2814	2814	830	830
R^2	.441	.445	.424	.429	.385	.386

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%; Robust standard errors in parentheses;

Table 6: Life-satisfaction and vulnerability with lagged income - *CS Approach*

	OLS1	OLS2	WG1	WG2
	(1)	(2)	(3)	(4)
(1-Vul)	-.064	-.103	.260	.046
	(.041)	(.063)*	(.104)**	(.191)
LnRealEarn	.064	.062	.049	.073
	(.007)***	(.011)***	(.011)***	(.018)***
L.lnrearn		-.005		-.037
		(.011)		(.021)*
Age	-.008	-.005	-.036	-.061
	(.004)**	(.006)	(.020)*	(.041)
Age2	.0001	.00007	.0003	.0004
	(.00005)**	(.00007)	(.0002)	(.0005)
Const.	1.180	1.127	1.746	2.684
	(.063)***	(.104)***	(.383)***	(.797)***
Obs.	2833	1262	2833	1262
R^2	.035	.03	.021	.062

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%; Robust standard errors in parentheses;

5.2.3 Aversion to *downside risk* or Aversion to *two-sided uncertainty*?

The measures of vulnerability I have employed in this paper are restricted to the notion of *exposure to downside risk*, since they are based on the probability of falling below a certain income threshold, but leave no role for the converse probability of positive outcomes, above average. Those measures were the most appropriate for the analysis I had in mind, to the extent that they enabled me to isolate downside risk (which is most commonly associated to the notion of vulnerability) from overall uncertainty. In this section I will make a further step towards disentangling the two by replacing the vulnerability measure with *symmetric measures* of variation in the happiness analysis. The results are shown in table 7 and 8. First, I use the raw squared residual $\hat{e}_{it}^2$ from a first stage earnings regression with fixed effects as a proxy of income volatility and find no significant relationship with happiness (table 7), despite the sign of the estimated effect is always negative, as we would expect if workers are risk-averse (and our experimental measure shows that is indeed the case). The lack of statistical significance might be due to the fact that ex-post realizations of the shock are a noisy realization of the expected degree of vulnerability workers perceive (and affected by). A way to circumvent the problem is to model the variance of these residuals, as I already did in section 4.1, and use the predicted value as a measure of expected variance. The results are reported in table 8, where I use the predicted standard deviation of e_{it} (obtained, using the results in table 1). As in the previous analysis, I document a negative effect, which is also statistically significant when I use the First-Differenced estimator. However, given the considerable reduction in sample size that accompanies first-differencing, we are unable to draw strong conclusions. Overall, this evidence seems to point to the overall conclusion that vulnerability to downside income risk as captured in the previous analysis plays a clearer role in income determination than symmetric volatility. Moreover, it is worth remarking that fixed effects in the happiness-generating process appear, once again, to be crucial for the identification of the effect of income vulnerability on happiness; and I believe the results in this section bring this conclusion to the fore even more compellingly, given that they now rest on a simpler (non-composite) measure of uncertainty.

Table 7: Vulnerable to Downside Risk or Averse to Uncertainty? (Residual)

	OLS1	OLS2	WG1	WG2	FD1	FD2
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{e}^2_{lny}$	-.003	-.003	-.009	-.011	-.035	-.034
	(.007)	(.007)	(.010)	(.010)	(.016)**	(.017)**
LnRealEarn	.016	.036	.021	.051	.024	.041
	(.004)***	(.010)***	(.008)***	(.017)***	(.011)**	(.023)*
LnWorkSatis	.611	.609	.588	.588	.566	.566
	(.013)***	(.013)***	(.025)***	(.025)***	(.026)***	(.026)***
Married		.027		.018		.015
		(.010)***		(.020)		(.031)
Age		-.006		-.007		-.080
		(.003)**		(.013)		(.092)
Age2		.00008		.0001		-.0002
		(.00003)**		(.0002)		(.0005)
EarnQuart=2		-.003		-.027		-.026
		(.016)		(.026)		(.033)
EarnQuart=3		-.040		-.068		-.040
		(.021)*		(.035)*		(.045)
EarnQuart=4		-.053		-.096		-.061
		(.029)*		(.049)**		(.063)
Const.	.424	.499	.437	.523	-.011	.087
	(.017)***	(.050)***	(.031)***	(.245)**	(.012)	(.086)
Obs.	2978	2978	2978	2978	832	832
R^2	.452	.456	.42	.423	.381	.383

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%; Robust standard errors in parentheses;

Table 8: Vulnerable to Downside Risk or Averse to Uncertainty? (Predicted Variance)

	OLS1	OLS2	WG1	WG2	FD1	FD2
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\sigma}_{lny}$	-.012	-.014	-.009	-.127	-.268	-.351
	(.015)	(.016)	(.058)	(.084)	(.165)	(.176)**
LnRealEarn	.013	.032	.021	.049	.024	.047
	(.005)**	(.012)***	(.008)**	(.018)***	(.012)**	(.024)*
LnWorkSatis	.600	.596	.580	.580	.569	.566
	(.015)***	(.015)***	(.028)***	(.028)***	(.029)***	(.029)***
Married		.033		.021		.055
		(.011)***		(.023)		(.035)
Age		-.008		-.003		-.115
		(.003)**		(.014)		(.113)
Age2		.0001		.0001		-.0002
		(.00004)**		(.0002)		(.0006)
EarnQuart=2		.006		-.024		-.035
		(.019)		(.026)		(.036)
EarnQuart=3		-.041		-.065		-.045
		(.025)		(.036)*		(.048)
EarnQuart=4		-.051		-.096		-.089
		(.035)		(.051)*		(.068)
Const.	.452	.555	.454	.411	-.005	.131
	(.023)***	(.063)***	(.044)***	(.280)	(.015)	(.109)
Obs.	2144	2144	2144	2144	641	641
R^2	.437	.442	.417	.421	.39	.396

Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%; Robust standard errors in parentheses;

6 Conclusions

This chapter has explored the direct relationship between income and well-being in Africa with a particular focus on the unexplored link between *income vulnerability and happiness*. Using unique longitudinal data on happiness and labour market outcomes from a representative household survey in Ghana, I estimate measures of vulnerability at the level of the individual worker and explore their relationship with life-satisfaction. Upon controlling for unobservable individual fixed effects, work-satisfaction, relative income and for a set of worker characteristics, I find a sizable impact of income on happiness, as generally found in cross-sectional analyses, but rarely explored in a panel data context. Moreover, and uniquely to this study, I find a significant effect of *vulnerability to low income* on life-satisfaction. Respondents who face higher expected vulnerability to falling below a low-income threshold in the next period are less satisfied with their life, even upon controlling for income and a number of other key regressors.

As an extension to the main result, I set out to explore the potential mechanisms driving my findings. First, I test the role of 'adaptive income aspirations' and find some evidence that aspiration adaptation to current income may be partly driving the estimated income effect, which could be interpreted as only transitory, while reversion to the mean takes place as aspirations grow. Second, I explore the 'risk-aversion channel', testing whether individuals with higher risk-aversion suffer from larger welfare losses, ceteris paribus, for a give increase in vulnerability. Using a direct measure of risk-aversion obtained from field-experiments, I test this hypothesis directly, but I am unable to draw strong conclusions, at least partly due to the limited size of the sample for which the experimental measure of risk-aversion is available. There is only a tentative indication that when I control for individual fixed effects, more risk-averse people appear to suffer from greater losses of utility for the same increase in vulnerability, but the effect is not statistically significant. Moreover, issues of confounding causality running from risk-aversion to vulnerability via job-sorting and risk-coping strategies further complicate the interpretation of this evidence and will require further attention as new data becomes available. Third, I attempt to disentangle the effect of vulnerability to downside risk on happiness from the effect of *symmetric uncertainty* and find mixed evidence, overall leading to the conclusion that two-sided uncertainty has a less clear-cut impact on subjective well-being than vulnerability to downside risk. My overarching conclusion, borne out of the availability of panel data for this analysis, is that failing to control for individual fixed effects in the study of subjective well-being may lead to strong bias and misleading conclusions.

Aside their scientific value for documenting a generally assumed but not previously tested link between income vulnerability (or, income uncertainty) and individual well-being, the results in this paper bear important policy-implications. In particular, they lend clear support to policy interventions that aim to reduce earnings uncertainty and vulnerability to poverty, as we would expect such policies to have a direct 'impact' on agents' life-satisfaction via a reduced perception of the downside risk they face. Moreover, my results suggest that non-Rawlsian models of growth whereby "someone may be left behind" may fail to enhance general welfare if the risk of falling behind is high enough and sufficiently widespread among the population (which the results in this paper indicate as a clear possibility).

Several leads for future research emerge from this work. Here are just three examples. First, further refinements in the analysis should aim to disentangle the 'risk-aversion' from the 'loss-aversion' channel through which vulnerability may be affecting happiness. This is particularly important as both established evidence and casual observation suggest that downside and upside risks may be weighed differently in workers' perceptions. The next waves of GHUPS, designed to contain additional experiments to elicit risk-aversion (and loss aversion) will be the ground upon which I will strengthen the current analysis. Second, the time-series properties of the mechanisms under scrutiny will need to be formally analysed and the modeling framework adjusted accordingly. Third, collecting data on income-expectations should both allow us to assess how aware agents are of their level of vulnerability and to test the robustness of the vulnerability model.

References

BARR, A. (2007): "Attitudes to Risk in Ghana: Field Manual," Unpublished.

BELLEMARE, M. F., C. B. BARRETT, AND D. R. JUST (2009): "The welfare impacts of price stabilization: evidence from rural Ethiopia," unpublished working paper.

BINSWANGER, H. (1980): "Attitudes Towards Risk: Experimental Measurement Evidence in Rural India," *American Journal of Agricultural Economics*, 62(3), 395–407.

BLANCHFLOWER, D. (2000): "Self-employment in OECD countries," *Labour Economics*, 7, 471–505.

BLANCHFLOWER, D., AND A. OSWALD (2002): "Well-being over time in Britain and the USA," *Journal of Public Economics*, 88 (7-8), 1359–1386.

CARIA, A. S. (2009): "Risk Attitudes and The Formation of Endogenous Learning Networks among Cocoa Farmers in Ghana," unpublished Mphil thesis.

CASSAR, C. (2010): "Quality of employment and job satisfaction: evidence from Chile," Unpublished.

CHAUDHURI, S. (2003): "Assessing vulnerability to poverty: concepts, empirical methods and illustrative examples," Columbia University, unpublished manuscript.

CHAUDHURI, S., J. JALAN, AND A. SURYAHADI (2002): "Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from Indonesia," Columbia University, Department of Economics Discussion Paper 0102-52.

DE NEVE, K., AND H. COOPER (1998): "The happy personality: a meta analysis of 127 personality traits of subjective well being," *Psychological Bulletin*, 125, 197–229.

DERCON, S. (1996): "Risk, crop choice and savings: evidence from Tanzania," *Economic Development and Cultural Change*, 44(3), 485–514.

——— (2004): "Growth and shocks: evidence from rural Ethiopia," *Journal of Development Economics*, 74(2), 309–329.

DERCON, S., AND L. CHRISTIAENSEN (2007): "Consumption risk, technology adoption and poverty traps: evidence from Ethiopia," CSAE Working Paper.

DERCON, S., J. HODDINOTT, AND T. WOLDEHANNA (2005): "Shocks and consumption in 15 Ethiopian villages, 1999-2004," *Special Issue on Risk, Poverty and Vulnerability in Africa, Journal of African Economies*, 14(4), 559–585.

DI TELLA, R., J. HAISKEN-DE NEW, AND R. MACCULLOCH (2007): "Happiness adaptation to income and to status in an individual panel," NBER working paper 13159.

EASTERLIN, R. (1995): "Will raising the incomes of all increase the happiness of all?," *Journal of Economic Behaviour and Organisation*, 27, 35–47.

——— (2001): "Income and happiness: towards a unified theory," *Economic Journal*, 111(473), 465–484.

FAFCHAMPS, M. (2003): *Rural Poverty, Risk, and Development*. Elgar Publishing.

——— (2009): "Vulnerability, Risk Management, and Agricultural Development," paper presented at the AERC Conference on Agriculture and Development held in Mombasa, Kenya, on May 28-30, 2009.

FAFCHAMPS, M., AND F. SHILPI (2008): "Subjective welfare, isolation, and relative consumption," *Journal of Development Economics*, 86(1), 43 – 60.

——— (2009): "Isolation and Subjective Welfare: Evidence from South Asia," *Economic Development and Cultural Change*, 57(4), 641–683.

FALCO, P. (2010): "Risk Aversion and Occupational Choices: evidence from matched field experiments and survey data in urban Ghana," paper presented at the WB-IZA Conference on 'Employment and Development' (Cape Town, May 2010).

FERRER-I CARBONELL, A., AND P. FRIJTERS (2004): "How important is methodology for the estimates of the determinants of happiness?," *Economic Journal*, 114(497), 641–659.

FREY, B., AND A. STUTZER (2002): "What can economists learn from happiness research?," *Journal of Economic Literature*, 40(2), (402:435).

HARRISON, G. W. (2008): "Maximum Likelihood Estimation of Utility Functions using Stata," *University of Central Florida, Working Paper 06-12*.

HOLZMANN, R., AND S. JORGENSEN (2000): "Social risk management: a new conceptual framework for social protection and beyond," Social Protection Discussion Paper series.

KINGDON, G., AND J. KNIGHT (2004): "Community, Comparisons and Subjective Well-being in a Divided Society," CSAE Working Paper.

KNIGHT, J.AND GUNATILAKA, R. (2008): "Aspirations, Adaptation and Subjective Well-Being of Rural-Urban Migrants in China," Department of Economics Discussion Paper.

LAYARD, R. (2005): *Happiness: lessons from a new science*. Penguin.

LIGON, E., AND L. SCHECHTER (2004): "Evaluating different approaches to estimating vulnerability," (30159).

LIU, E. (2008): "Time to Change What to Sow: Risk Preferences and Technology Adoption Decisions of Cotton Farmers in China," Unpublished working paper.

LUTTMER, E. (2005): "Neighbours as negatives: relative earnings and well-being," *Quarterly Journal of Economics*, 120, 963–1002.

MEGHIR, C., AND L. PISTAFERRI (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–1426.

POWDTHAVEE, N. (2010): "How much does money really matter? Estimating the causal effects of income on happiness," *Empirical Economics*, 39(1), 77–92.

RANKIN, N., J. SANDEFUR, AND J. TEAL (2010): "Learning and Earning in Africa: Where are the Returns to Education High?," *Working Paper*.

ROSENZWEIG, M., AND H. BINSWANGER (2003): "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," *Economic Journal*, 103(416), 56–78.

TVERSKY, A. (1991): "Loss aversion in riskless choice: a reference-dependent model," *Quarterly Journal of Economics*, 106(4), 1039–1061.

APPENDIX

A Summary Statistics

Table 9: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Age	36.513	10.552	14.102	69.659	2569
Educ	8.220	3.82	0	19	2569
Male	0.431	0.495	0	1	2569
Priv Wage	0.328	0.47	0	1	2569
Civil or Pubent	0.069	0.253	0	1	2569
Ln(employees)	0.132	0.385	0	5.017	2569
Ln(firmsize)	0.915	1.534	0	5.011	2569
Years since started current job	9.368	9.341	0	65.583	2569
Married	0.529	0.499	0	1	2569
Ga-Dangme	0.161	0.367	0	1	2507
Ewe	0.08	0.272	0	1	2507
Mole-Dagbani and Hausa	0.103	0.304	0	1	2507
Other ethnicity	0.063	0.244	0	1	2507

NOTE: The sample restricted to workers who enter the Happiness Regressions.

B CS Approach - First Stage and Estimated Vul.

Table 10: Earnings Regression by year (STEP1 - CS Approach)

	Y2004	Y2005	Y2006	Y2008	Y2009
	(1)	(2)	(3)	(4)	(5)
Age	.064	.079	.100	.107	.097
	(.022)***	(.016)***	(.015)***	(.024)***	(.023)***
Age2	-.0008	-.0008	-.001	-.001	-.001
	(.0003)***	(.0002)***	(.0002)***	(.0003)***	(.0003)***
Educ	-.058	-.026	-.052	-.021	-.031
	(.025)**	(.020)	(.020)**	(.031)	(.030)
Educ2	.005	.005	.006	.005	.005
	(.002)***	(.001)***	(.002)***	(.002)**	(.002)**
Male	.223	.281	.264	.373	.442
	(.069)***	(.061)***	(.059)***	(.091)***	(.089)***
Priv Wage	-.157	-.258	-.186	-.164	-.161
	(.109)	(.098)***	(.092)**	(.144)	(.154)
Civil or Pubent	.304	.242	.446	.310	.338
	(.138)**	(.134)*	(.117)***	(.174)*	(.156)**
Ln(employees)	.441	.180	.271	.262	.280
	(.113)***	(.072)**	(.096)***	(.104)**	(.088)***
Ln(firmsize)	.154	.134	.122	.146	.111
	(.030)***	(.028)***	(.030)***	(.041)***	(.045)**
Years since started current job	.011	.008	.023	.015	.018
	(.005)**	(.004)**	(.004)***	(.006)**	(.006)***
Const.	.898	.331	.291	.147	.389
	(.388)**	(.299)	(.280)	(.462)	(.444)
Obs.	619	826	1007	593	614
R^2	.212	.261	.243	.208	.188

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%

Table 11: Residual Regression by year (STEP2 - CS Approach)

	Y2004	Y2005	Y2006	Y2008	Y2009
	(1)	(2)	(3)	(4)	(5)
Age	-.049	-.030	-.056	.008	.020
	(.028)*	(.019)	(.023)**	(.037)	(.038)
Age2	.0007	.0003	.0007	.0001	-.0003
	(.0004)*	(.0002)	(.0003)**	(.0005)	(.0005)
Educ	.071	-.012	-.030	-.126	-.060
	(.030)**	(.023)	(.030)	(.046)***	(.050)
Educ2	-.004	.001	.002	.010	.004
	(.002)*	(.002)	(.002)	(.003)***	(.004)
Male	.087	.024	-.049	-.030	.151
	(.083)	(.068)	(.083)	(.134)	(.146)
Priv Wage	-.324	-.110	-.474	-.820	-.788
	(.131)**	(.108)	(.129)***	(.213)***	(.249)***
Civil or Pubent	-.347	-.387	-.317	-.611	-.546
	(.165)**	(.146)***	(.167)*	(.255)**	(.255)**
Ln(employees)	-.070	-.138	-.098	-.043	.218
	(.135)	(.080)*	(.133)	(.152)	(.141)
Ln(firmsize)	-.028	.002	.002	-.019	.029
	(.036)	(.030)	(.043)	(.061)	(.073)
Years since started current job	.003	.004	-.006	-.017	-.005
	(.006)	(.004)	(.006)	(.009)*	(.009)
Married	-.180	.033	.126	-.127	-.126
	(.088)**	(.069)	(.084)	(.136)	(.145)
Ga-Dangme	.232	-.084	.060	-.243	-.143
	(.110)**	(.089)	(.106)	(.168)	(.165)
Ewe	.408	.245	.170	.097	.098
	(.151)***	(.133)*	(.155)	(.209)	(.245)
Mole-Dagbani and Hausa	-.006	-.048	.358	-.168	-.550
	(.150)	(.111)	(.130)***	(.201)	(.268)**
Other ethnicity	.255	-.048	.160	-.088	-.029
	(.138)*	(.104)	(.132)	(.286)	(.286)
Const.	1.279	1.300	1.993	1.282	1.079
	(.478)***	(.346)***	(.420)***	(.716)*	(.738)
Obs.	617	813	852	591	595
R^2	.069	.028	.07	.106	.059

Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%

Table 12: Weighted Residual Regressions by year (STEP3 - CS Approach)

	Y2004	Y2005	Y2006	Y2008	Y2009
	(1)	(2)	(3)	(4)	(5)
Age	-.056	-.024	-.047	.005	-.010
	(.026)**	(.019)	(.021)**	(.026)	(.032)
Age2	.0007	.0002	.0006	.0001	.0001
	(.0003)**	(.0002)	(.0003)**	(.0003)	(.0004)
Educ	.067	-.016	-.035	-.137	-.024
	(.022)***	(.022)	(.029)	(.040)***	(.047)
Educ2	-.003	.001	.002	.010	.003
	(.002)*	(.002)	(.002)	(.003)***	(.003)
Male	.080	.026	-.022	.022	.098
	(.067)	(.064)	(.070)	(.099)	(.127)
Priv Wage	-.199	-.132	-.448	-.819	-.687
	(.098)**	(.098)	(.088)***	(.146)***	(.183)***
Civil or Pubent	-.273	-.342	-.239	-.643	-.549
	(.127)**	(.111)***	(.102)**	(.206)***	(.209)***
Ln(employees)	-.023	-.124	-.117	.027	.123
	(.117)	(.068)*	(.128)	(.153)	(.168)
Ln(firmsize)	-.047	.004	-.0002	.021	.046
	(.024)*	(.025)	(.029)	(.034)	(.052)
Years since started current job	-.0001	.001	-.004	-.012	.004
	(.005)	(.004)	(.005)	(.007)*	(.008)
Married	-.127	.018	.061	-.054	.003
	(.076)*	(.064)	(.071)	(.103)	(.129)
Ga-Dangme	.195	-.089	.067	-.209	-.064
	(.105)*	(.078)	(.088)	(.100)**	(.139)
Ewe	.287	.201	.127	.103	.160
	(.162)*	(.148)	(.148)	(.172)	(.253)
Mole-Dagbani and Hausa	.073	-.013	.324	-.056	-.213
	(.108)	(.107)	(.137)**	(.109)	(.179)
Other ethnicity	.186	-.060	.188	-.172	-.068
	(.133)	(.099)	(.125)	(.204)	(.265)
-con	1.406	1.216	1.861	1.386	1.308
	(.441)***	(.345)***	(.381)***	(.487)***	(.624)**
Obs.	614	813	851	588	591
R^2	.114	.03	.101	.173	.051

Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%

Table 13: Weighted Earnings Regression by year (STEP4 - CS Approach)

	Y2004	Y2005	Y2006	Y2008	Y2009
	(1)	(2)	(3)	(4)	(5)
Age	.068	.079	.099	.099	.095
	(.022)***	(.017)***	(.017)***	(.022)***	(.022)***
Age2	-.0009	-.0008	-.001	-.001	-.001
	(.0003)***	(.0002)***	(.0002)***	(.0003)***	(.0003)***
Educ	-.050	-.024	-.055	-.037	-.034
	(.024)**	(.020)	(.024)**	(.032)	(.030)
Educ2	.005	.005	.007	.006	.005
	(.002)***	(.001)***	(.002)***	(.002)***	(.002)**
Male	.233	.290	.233	.351	.427
	(.067)***	(.061)***	(.063)***	(.084)***	(.089)***
Priv Wage	-.138	-.214	-.155	-.178	-.139
	(.104)	(.095)**	(.092)*	(.124)	(.142)
Civil or Pubent	.311	.265	.435	.310	.317
	(.130)**	(.121)**	(.121)***	(.160)*	(.149)**
Ln(employees)	.437	.147	.277	.254	.297
	(.110)***	(.068)**	(.105)***	(.108)**	(.096)***
Ln(firmsize)	.149	.116	.112	.143	.105
	(.027)***	(.026)***	(.030)***	(.035)***	(.042)**
Years since started current job	.011	.007	.022	.014	.018
	(.005)**	(.004)**	(.004)***	(.006)**	(.006)***
Const.	.805	.324	.271	.341	.413
	(.392)**	(.304)	(.318)	(.427)	(.440)
Obs.	614	813	851	588	591
R^2	.254	.271	.242	.229	.193

Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%

Summary and overview

This thesis has investigated the occupational choices faced by workers in Sub-Saharan Africa, and both the pecuniary and non-pecuniary outcomes of those choices. Combining data from a uniquely long and comprehensive household panel dataset with behavioural evidence from field-experiments, I have been able to focus 'explicitly' on the role of workers' psychological traits in decision making, such as their risk-aversion, and on direct measures of well-being, tackling questions that had previously remained unanswered, most prominently in the literature on development economics. In summary, the thesis makes three main contributions to the advancement of our knowledge on the functioning of labour markets in developing countries. Given the novelty of the analytical tools employed, however, my findings will certainly be of interest beyond the developing world. Moreover, the conclusions I have reached carry several implications for policy and future research.

First, by documenting the link between earnings volatility, risk-aversion and occupational choices through novel experimental techniques, the thesis shows that concerns of earnings stability play a crucial role in workers' choice over different occupational paths. In so doing, it implicitly suggests that job-creation policies should not be uncoupled from social protection policies and, more generally, that creating jobs that are characterised by a high degree of uncertainty may not stimulate employment as much as one could wish in dual labour markets where workers have the option to queue for the more secure occupations. This type of perverse effects were already hypothesised in the original dual-economy framework proposed by Harris and Todaro (1970), which this thesis has extended to explicitly test the role of income uncertainty. Furthermore, the thesis shows that social protection policies (e.g. unemployment benefits) may assist risk-averse workers in their career choices and, as such, they should not be viewed as 'mere' transfer payments, but as powerful development tools.

Second, this thesis shows that while returns to capital in small enterprises are high, and not the artifact of unobserved time-varying heterogeneity, returns to formal education are low in informal self-employment. From a policy angle, the first of these findings is in agreement with a large empirical literature documenting very high marginal returns to micro-investment in various parts of the developing world. However, it leaves an important question unanswered, which I am planning to address in my future research: why do micro-enterprises generally fail to grow despite such high returns to investment? Part of the answer will most likely lie in a careful study of inter-temporal preferences over saving and consumption, coupled with an improved understanding of the degree of asset integration between businesses and households, and of their interrelated decisions. The second result should be subject to further scrutiny, especially since it has been documented elsewhere, including in Ghana, that while education may fail to have an impact on earnings conditional on a given occupational category, it plays a significant role in determining which jobs people do/do not undertake. However, taken at face value this result calls into question the efficacy of education policy that is not carefully geared towards the need of the labour market. In particular, a labour market that is largely populated by informal businesses may require specific skills that are not necessarily offered by the standard curricula. This idea is further corroborated by the results of other studies on the Ghanaian labour market, which have found professional apprenticeships to be conducive to higher earnings despite years in formal school had no detectable effect.

Third, by documenting the direct link between income volatility and subjective welfare, this thesis lends strong support to the creation of earning safety nets, as an instrument to raise social welfare by insulating workers from the distress that facing volatile earnings may cause. Most importantly, it suggests that societies where growth is accompanied by higher and sufficiently widespread uncertainty may not be societies where overall welfare increases. This is a very important result that extends well-beyond developing countries, to the more general context of labour market policies that affect job security. For instance, to draw an interesting parallel, this kind of evidence could inform the fierce debate currently raging in Europe on the need for more job flexibility in some of the European Union countries to advance competitiveness. If my results were corroborated, the crucial question would be whether the additional growth that may result from the added flexibility would more than compensate the loss of welfare resulting from decreased job security, and how quickly workers would adapt to the higher level of uncertainty that such flexibility entails. It will be interesting in the future to test whether these considerations can be extended to the risk of losing one's job, as the current analysis only focuses on earnings fluctuations within employment.

Finally, this thesis leaves unanswered an overarching question, to which I am planning to dedicate my attention over the coming years: how do attitudes to risk and uncertainty originate? And how do they correlate with poverty? As argued by Banerjee in a paper titled 'The Two Poverties' (2000), economic theory has ambiguous predictions on the direction of such correlation. The poor may equally be viewed as 'desperate' and hence more risk-prone or 'vulnerable', and hence more risk-averse, though the two conclusions clearly carry very different implications for anti-poverty measures. Studying the evolution of risk-attitudes as a function of job-experience, poverty and other workers' characteristics is a promising alley to tackle this research question. More rounds of data collection and new field-experiments within the context of a uniquely long panel dataset like the GHUPS should provide the ideal framework to continue this fascinating research program.