Essays on Self-Employment in Africa

Jonathan Lain

Lincoln College, University of Oxford

A thesis submitted for the degree of D.Phil. in Economics

Trinity Term 2015

Abstract

Informal sectors in developing countries provide a substantial pool of jobs for some of the world's poorest people. Self-employment comprises a large portion of the job opportunities available to individuals working in these sectors. This thesis is concerned with the factors that drive people to become self-employed and determine their welfare as an entrepreneur, with a special emphasis on differences between women and men.

In Chapter 1, we explain the Ghanaian context to which this thesis relates and outline the contribution of each main chapter and the common themes.

In Chapters 2 and 3, we examine the trade-off between domestic work, such as caring for children and household chores, and market work. In Chapter 2, we consider the extent to which individuals are able to substitute between these two tasks to adjust to short-run variation in domestic productivity brought about by outages in electricity. We find that self-employed workers adjust non-monotonically to changes in domestic productivity, initially increasing their levels of domestic work to preserve consumption levels, but then substituting towards market work when power outages become more severe. We show that this relationship is heterogeneous by sex, and build a model of time allocation to demonstrate the theoretical mechanisms behind these results.

In Chapter 3 we examine whether the factors that drive occupational selection differ by sex. It is often argued that women choose jobs in self-employment because this allows them to balance income-generation with childcare and other domestic work. We test the plausibility of this claim and its implications for labour market outcomes. First, we use a simple model of occupational choice to clarify our ideas about which notions of 'job flexibility' are important for the Ghanaian context. Second, we examine whether differential selection forces between women and men may explain the raw sex earnings gaps that appear to persist in various sectors, using a multinomial logit model to adjust for non-random occupational selection. We find that controlling for selection substantially widens the earnings gap amongst the self-employed, but shrinks it for the wage-employed. Third, we interrogate our selection equations and show that domestic obligations increase women's likelihood of entering low-input self-employment jobs more than men. We assess the importance of endogeneity using a maximum simulated likelihood estimator to couch the idea that selection on observables can be used as a guide for selection on unobservables, focussing on the discrete choice made over occupation.

In Chapter 4, we turn to theory to try and resolve some of the empirical puzzles that remain from Chapter 3. In particular, we attempt to reconcile the fact that female participation in self-employment is so high even when the average differences in potential earnings are large. To do this, we construct a search model, which allows for individual heterogeneity and participation in both self- and wage-employment, as well as discrimination against female workers in the wage sector. We numerically solve and simulate this model, using calibrations from the existing literature, to explain a set of stylised facts generated from a longitudinal dataset of workers in urban Ghana. We show that wage sector discrimination leads to average earnings gaps in *all* sectors of the economy, even if the underlying ability distribution is the same for both sexes. We also conduct a series of experiments to examine how women and men may be affected differently by government policy.

Finally, in Chapter 5 we connect our main findings to policy and make some suggestions for future work.

[Approximate Word Count: 78,000]

Acknowledgements

This thesis has only been possible thanks to the generous support of many different people. I am most indebted to Prof. Margaret Stevens for her patient and insightful supervision throughout the course of both my graduate and undergraduate studies. Margaret has always made time to help me, offering clear and practical suggestions even during the more difficult times of doing a D.Phil. I sincerely hope that I apply her approach to economics in this thesis and in my life after Oxford.

I must also thank members of the Centre for the Study of African Economies (CSAE), who have provided feedback on my work and always been willing to discuss tricky econometric issues. I am particularly grateful to Dr. Francis Teal, who continued to support my research even after his retirement. Dr. Simon Quinn also deserves special mention, not only for his one-on-one guidance, but also for co-ordinating the Firms and Development Research Group, where I presented my work several times. The participants in this group have been a continual source of useful research ideas, Stata know-how, and other friendly advice. Finally, CSAE provided me with the chance to travel to Ghana to work on collecting some of the data used in this thesis. This experience was invaluable, not only in shaping my ideas about African labour markets, but also in helping me understand the opportunities and challenges associated with doing fieldwork.

In addition, I am thankful for the financial support of the Economic and Social Research Council, without which my graduate studies would have been impossible.

Finally, my friends in Lincoln College JCR and MCR and my family have continually spurred me on throughout my work on this thesis. I extend my sincere gratitude for their constant encouragement.

List of Tables

List of Figures

Contents

Chapter 1

Introduction

1.1 Motivation

Labour is the main productive asset available to the world's poorest people. Thus, understanding how individuals choose their occupation, allocate their work effort, and are subsequently rewarded, is vital for guiding strategies to reduce poverty. Informal sectors, which we define below, employ a substantial pool of workers in developing countries, with self-employment comprising a large portion of these jobs. It is therefore no surprise that labour economists have paid significant attention to the factors that drive people to become self-employed and determine their welfare as an entrepreneur. These are the broad topics which this thesis will examine, focussing in particular on how self-employment differs between women and men.

1.2 Key Concepts

Although they are related, the concepts of 'self-employment' and 'informality' are not synonymous. Since they are invoked throughout this thesis, we begin by considering how best to understand these terms.

Before doing so, however, it is useful to clarify what we mean by 'employment', given the rarity of concrete labour contracts in developing countries (Teal, 2012). Sen's (1975) framework is widely used to discuss this question, which suggests that there are three potential aspects of employment:

1. The *income* aspect. Individuals are employed if and only if some part of their income is conditional on their working.

2. The *production* aspect. Individuals are employed to produce an output, which can be sold for income or given directly to the household. This output would not be produced if the individual did not work.

3. The *recognition* aspect. Individuals are employed if and only if they perceive themselves as such. Thus, workers' employment status depends on whether they believe their job is appropriate for them, given their preferences and human capital.

For simplicity, we focus on the 'income' aspect of employment throughout this thesis. Thus, individuals are employed if and only if they work for some money income, which would not be available without their job.

There are two key properties that we use to characterise self-employment, which match the definition used during the data collection for Chapters 2 and 4. In particular, workers are classified as self-employed if: (1) they have main control over their business activity, and (2) they use their own tools or other physical capital, if any are used at all. For almost all workers, this provides a parsimonious and coherent method for distinguishing the self-employed from the wage-employed.

The 'informal sector' is broadly understood to encompass activities where regulation and support from the government are lacking. The self-employed in developing countries tend to operate in the informal sector, insofar as their businesses are not registered with government agencies. However, the concept of informality has been expanded to include some individuals who work in wage jobs. For example, employees in unregistered enterprises, casual day labourers, undeclared workers, or temporary/part-time workers may all be regarded as working in the informal sector (Chen, 2006; Fields, 2013). Since there are many possible dimensions to informality, delineating the informal sector immediately becomes a difficult task, and there is little consensus about precisely which workers should be included and excluded. By contrast, focussing on the self-employed offers a simple approach for analysing a key portion of this complex sector.

1.3 Segmentation and Sorting

In spite of the importance of self-employment, and informal sectors more generally, for developing countries, the debate about how these sectors should be understood remains unresolved.

African labour markets have traditionally been modelled as 'segmented', whereby barriers to entry and rationing in the formal sector ensure that earnings premiums persist relative to the informal sector. Thus, differences in earnings are determined mainly by structural features of the labour market rather than by worker heterogeneity. One of the earliest attempts to capture this idea was Harris and Todaro's (1970) model of rural-urban migration. Individuals face a choice between the certainty of a low wage in rural areas and the lottery of migrating to the urban area, where they either gain a high wage in a formal sector job or fall into unemployment. Formal sector wages are fixed above the competitive level by a number of potential factors, including minimum wage legislation, trade union activity, and firms' desire to pay efficiency wages in the absence of perfect information on worker effort (Shapiro and Stiglitz, 1984). Thus, the economy *must* have some unemployment, in order to equilibrate *expected* returns in rural and urban areas.

There are several possibilities for modifying the Harris and Todaro (1970) framework to explicitly include an informal sector. One option is to simply assume that the informal sector and unemployment are one and the same (Kingdon et al., 2006). This approach is somewhat restrictive, however, as it does not explicitly capture the earnings of informal workers and there may be meaningful differences between the unemployed and those employed in the informal sector. As such, a number of attempts have been made to extend the Harris and Todaro framework, to add a distinct informal sector. For example, in Fields' (1975) model, migrants from rural areas can either divert all their time to searching for an urban formal sector job or work in the urban informal sector, which affords them some income *and* the possibility, albeit reduced, of finding a formal sector job. This results in a substantial urban informal sector, in which earnings are lower than the certain wage in rural areas, and are therefore significantly less than urban formal sector wages. Thus, self-employment, as the main activity in the informal sector, is an involuntary response to structural problems with the labour market.

It may also be possible to couch labour market segmentation in terms of the search literature. In these models, there is friction in the process by which workers and firms meet and match, creating barriers to entry to formal sector work. We explore this modelling approach extensively in Chapter 4.

Whilst models of segmentation emphasise the importance of labour market frictions in determining inter-sector differences in earnings, the 'sorting' framework suggests that individuals move into their optimal jobs on the basis of their productivity, preferences, and other characteristics. The crux of the sorting framework was captured by Roy's (1951) model, in which individuals choose between hunting and fishing on the basis of their comparative advantage in each activity. This has been the dominant view of occupational selection — be it into self-employment, the informal sector, or otherwise — in developed country settings. If formal sector workers earn more, it is because they are more productive, or they are being compensated for some non-wage aspect of their job.

A number of sorting models, which focus in particular on self-employment in developed countries, have explicitly captured the decision to leave wage work and become an entrepreneur. For example Lucas (1978) suggests that there may be extra non-pecuniary sources of utility in self-employment, and that individuals are endowed with a certain level of 'managerial talent', about which they have perfect information. The distribution of managerial ability and the stock of capital and labour in the economy then determine how many individuals become entrepreneurs, how large their businesses are, and thus how many people remain wage-employed as they work for these entrepreneurs. Jovanovic (1982) adds a dynamic element to Lucas' model by dropping the assumption that individuals have perfect information about their entrepreneurial ability. Instead, individuals enter self-employment on the basis of *expected* profits. Enterprise output adjusts as ability is revealed noisily through observations of profit, which are subject to stochastic shocks. Thus, Jovanovic's model predicts that the self-employment sector will contain both enterprises that are destined to survive and those that are doomed to fail. We investigate similar heterogeneity within the self-employment sector, as well as the factors that guide selection into entrepreneurship, in Chapter 3.

A substantial empirical literature has emerged, considering whether sorting or segmentation are more important in shaping informal sectors and self-employment in developing countries. In some contexts, especially in Latin America, it appears that the sorting framework is more appropriate (Maloney, 2004). For example, Magnac (1991) uses a maximum likelihood approach to correct earning functions for selection across multiple sectors in urban Colombia, rejecting the hypothesis that the formal sector is characterised by costs of entry. After controlling for a series of human capital and household demographic variables, he shows that the married women in his sample select, on average, into the sector in which their potential earnings are highest. Cunningham and Maloney (2001) apply cluster and factor analysis to investigate heterogeneity amongst the self-employed in Mexico, and find that almost all individuals enter self-employment as it is optimal for them to do so, given their characteristics.

Vitally, the factors that drive sorting into different jobs may be determined by restrictions outside the labour market. Thus, even if individuals choose where to work, they do so subject to constraints over how much human capital they can attain, how much domestic work they must do for the household, or how much physical capital they can access (Maloney, 2004; Heintz and Pickbourn, 2012). For example, workers with low levels of schooling optimally select informal sector jobs because that is where they can earn the highest wages, given their poor access to education. It would be sub-optimal for them to try and obtain a formal sector job, because they would spend most of their time unemployed, searching unsuccessfully.

Even with this broad notion of sorting, however, the evidence for Africa, which we discuss in Section 1.4.2, suggests that it is a mistake to ignore the logic of segmentation entirely (Günther and Launov, 2012). As such, the forces of segmentation and sorting may coexist to create a heterogeneous self-employment sector with 'upper' and 'lower' tiers (Fields, 1990). In this thesis, we consider the importance of both mechanisms, as we attempt to document and explain differences between self-employed women and men.

1.4 Context

Although the balance of theoretical and empirical work varies across the central chapters of this thesis, we are guided throughout by evidence from Ghanaian labour markets. This presents three key advantages. Firstly, as we explain in each chapter, the data available for Ghana are varied and of comparatively high quality. Secondly, there is a thorough empirical literature on Ghana, on which this thesis can build. Finally, the policy background in Ghana, which we explain below, is similar to a number of other countries in sub-Saharan Africa, especially Côte d'Ivoire, Kenya, Tanzania, and Uganda (Kingdon et al., 2006; Sandefur et al., 2006; Sandefur, 2010). As such, our main findings may be applicable in a variety of settings.

1.4.1 Macro Evidence

Beginning in the mid-1980s, and with support from the International Monetary Fund (IMF) and the World Bank, Ghana undertook a series of reforms to fiscal, monetary, and trade policy, which substantially changed the performance and structure of the economy. In particular, the public sector underwent significant contraction and the taxation net was widened, interest rates and banking reserve ratios were increased, and trade was liberalised and the exchange rate was allowed to float (Aryeetey and McKay, 2004; Aryeetey and Baah-Boateng, 2007; Kolavalli et al., 2012; Obeng-Odoom, 2012). Following the reforms, Ghana has enjoyed positive and steady growth and, with incomes among unskilled workers in small-scale enterprises rising, poverty was halved in the period 1991–2005 (Nsowah-Nuamah et al., 2010).

Concurrently, the Ghanaian economy has undergone sizeable changes to the structure of industry and employment, which motivate the analysis of self-employment in this thesis. The urban informal sector has expanded for two key reasons. Firstly, the number of jobs in the public sector has been reduced so as to improve fiscal stability (Aryeetey and McKay, 2004; Kolavalli et al., 2012). Secondly, although job creation has kept pace with the growth of the labour force, these have generally been in small informal enterprises, which have limited possibilities for growth into medium- or large-scale firms (Nsowah-Nuamah et al., 2010; Sandefur, 2010). In part, this is because growth has been driven by low value-added sectors, such as in cocoa and gold, and more recently, oil and natural gas, rather than in manufacturing. The proceeds of this growth have then been distributed through small-scale buying and selling (Aryeetey and Baah-Boateng, 2007; Obeng-Odoom, 2012). In this way, Ghana has not followed the structural transformation path first described by Lewis (1954) in which the informal sector shrinks as the economy grows.

1.4.2 Labour Markets

There is a wealth of micro-evidence on the Ghanaian labour market, which demonstrates that both sorting and segmentation may determine individuals' occupation and their outcomes in certain jobs.

Substantial earnings differences persist between formal and informal sectors in Ghana, not all of which can be attributed to individuals with heterogeneous earnings potential sorting between jobs. Heintz and Slonimczyk (2007) show that there are a number of dimensions along which the Ghanaian labour market may be segmented. Not only are the formal-informal and wage- versus self-employment distinctions important for earnings, but there are also large differences between the public and private sector, and those operating in rural and urban areas.[1] Kolavalli et al. (2012) argue that Ghana's legislative environment has contributed to this cross-cutting segmentation. For example, redundancy costs are especially high, with Ghanaian firms paying an average of 178 weeks salary when making a worker redundant, and the premium for unionised workers is amongst the highest in sub-Saharan Africa (Kingdon et al., 2006; Lejárraga, 2010).

[1] Heintz and Slonimczyk (2007) use a broad definition of informality, which includes any enterprises not registered with the government, and any wage workers lacking an employment contract.

One dimension of formality, which strongly influences the earnings of both the self- and wage-employed in Ghana, is firm size. Analysis of a number of different panel datasets has shown that workers and employers in large firms enjoy a sizeable premium on their earnings, even after controlling for time-invariant unobservables (Söderbom et al., 2004; Falco et al., 2011). That is, the size effect does not just result from more able individuals working in larger firms. To help interpret this finding, Teal (1996) further shows that firms' profitability drives employees' earnings, suggesting that workers are able to capture some of the rents gained by larger firms.

The firm size effect appears to be compounded by binding constraints on small enterprises' ability to grow. In particular, the mix of small and large firms in the economy results mainly from selection rather than expansion, insofar as large firms are 'born big' and micro-enterprises stay small (Sandefur, 2010; Davies and Kerr, 2015). A number of possible explanations emerge for the persistent small scale of new enterprises. Vitally, the returns to hired labour for micro-enterprises appear to be quite low and, although the returns to capital are initially high they quickly tail off given the extreme concavity of their production functions (Falco, 2011). There is also experimental evidence suggesting that micro-entrepreneurs may be misinformed about the productivity of potential extra workers (Caria and Falco, 2013).

Despite the evidence suggesting that the Ghanaian labour market is segmented, there is significant heterogeneity *within* sectors as well as between them. Indeed, this results in there being a great deal of overlap between the earnings distributions for the self- and wage-employed (Sandefur et al., 2006). However, despite this intra-sector variation in earnings, the returns to measures of human capital — such as education, experience, and tenure — differ across sectors. For example, the returns to education appear to be largest in large-firm wage-employment, whilst in self-employment the returns are small and convex. Moreover, individuals' human capital is more important for determining which sector to work in, than it is for explaining earnings variation within each sector (Kingdon and Söderbom, 2007; Rankin et al., 2010). As such, the factors that determine success for the self-employed are clearly different to those for the wage-employed.

As well as having a large and varied informal sector, the Ghanaian economy is also characterised by relatively high participation from women, making sex one of the most important dimensions of heterogeneity in the labour market. In 2013, 67.3 percent of females aged 15 or above participated in the labour force, whilst the average female participation rate for lower-middle income countries was just 47.6 percent. However, Ghana is less of an outlier in the context of sub-Saharan Africa, where the average female participation rate is up to 63.8 percent (World Bank, 2015). Thus, our understanding of women's outcomes in the Ghanaian labour market may be reflective of similar countries in the region, but should not be directly applied to contexts in Asia and Latin America.

There are both macro- and micro-level explanations for the relatively high female labour force participation in African labour markets. On the one hand, some have argued that women in developing countries work out of necessity, as households simply cannot afford to have certain members abstain from income-generating activities in their efforts to avoid poverty (Verick, 2014). However, there may also be household-level factors which drive women in Ghana, and other parts of sub-Saharan Africa, to work. Rather than pooling incomes and optimising task allocation jointly, it appears that household members, including spouses, tend to separate their activities according to social norms. Since women are responsible for feeding and clothing children, they must generate income to provide any goods the household itself cannot produce (Warner et al., 1997; Goldstein, 2004).

In any case, women tend to be concentrated in jobs where the returns are low. In particular, female participation is disproportionately weighted towards non-farm self-employment and casual wage work, where poverty rates tend to be higher than in other occupations. Even within sectors, women often work in industries where earnings are lower and rates of poverty are higher, such as retail and petty trading (Heintz, 2007). Also, female and male entrepreneurs appear to use different technologies, with women far less likely to deploy either capital- or labour-intensive methods to generate profits (Falco, 2011). We discuss the differences in the selection of occupation and technology between women and men in more detail in Chapter 3.

A growing experimental literature has documented striking heterogeneity in the returns to physical assets gained by female- and male-owned micro-enterprises in developing countries (de Mel et al., 2009, 2014). These findings have been replicated in Accra and neighbouring Tema, one of Ghana's largest urban centres, by Fafchamps et al. (2014). In their study, both female and male entrepreneurs are randomly treated with 'capital drops', which come in two forms — (1) cash grants or (2) in kind resources that would benefit their businesses. Whilst all male entrepreneurs tend to obtain some kind of return on these extra assets coming in either form, the results are far more mixed for women. In particular, it is only *wealthy* female entrepreneurs treated with the in kind capital who increase their profits. The cash treatment has no effect on any female-owned enterprise and women running *subsistence* businesses are unable to derive a return on assets delivered either in cash or in kind. The authors couch their results in terms of a 'flypaper effect', whereby capital given in kind sticks to the business rather than being used for other purposes, but the overall disparity between women and men remains unexplained. This further reinforces the importance of understanding how the motivations behind self-employment work may differ by sex.

Notwithstanding the important dimensions of heterogeneity discussed above, evidence suggests that the majority of labour market participants in Ghana are vulnerable and that this has severe implications for their welfare. Workers struggle to obtain and maintain good jobs, especially in the early part of their life-cycle, and simply biding time before taking a job does not appear to improve outcomes after starting work (Falco and Teal, 2012). Similarly, Falco (2010) assesses income uncertainty in the informal sector relative to the formal sector, linking individuals' occupational choice to their risk preferences. He shows that more risk averse individuals are less willing to tolerate the income variability in the informal sector, and thus queue until they can gain relative security in a formal sector wage job. Clearly, if workers suffer negative shocks to their income, this in itself will reduce their welfare, but the losses may be even more severe. Using subjective well-being data documenting individuals' life satisfaction, Caria and Falco (2011) also show that this vulnerability directly affects people's welfare, *over and above* the basic income effect. It is thus important to acknowledge the variability of incomes in informal sector jobs, when evaluating the factors which drive occupational choice.

The Ghanaian labour market, with its large and heterogeneous informal sector, is emblematic of the situation in many sub-Saharan African economies, but there are some important exceptions, which help to illustrate potential sources of labour market segmentation that do not affect Ghana. In South Africa, for example, self-employment is somewhat rare whilst the proportion of the labour force that is *unemployed* is large, at around 40 percent for the population at large and up to 50 percent for the young (Rankin and Roberts, 2011). These trends suggest that there are barriers to entry into self-employment in South Africa, which are not present in Ghana. Kingdon et al. (2006) argue that entrepreneurial activities are limited by the legislative environment in South Africa, where the penalties for failing to comply with regulations are high. Others have emphasised that individuals in South Africa lack the business know-how, access to credit, and tools for managing extra risks, especially crime, to establish profitable micro-enterprises (Chandra et al., 2002; Cichello et al., 2011). Understanding the conditions under which self-employment will and and will not prevail is vital for assessing the Ghanaian context, where such a large portion of the labour force run their own small businesses.

1.5 Outline and Contribution

Although there are key linkages between them, the three central chapters of this thesis are mainly presented as distinct research papers.

The first two main chapters examine the trade-off between domestic work, such as caring for children and household chores, and market work. In Chapter 2, we consider the extent to which individuals are able to substitute between these two tasks to adjust to short-run variation in domestic productivity brought about by outages in electricity. In doing so, we investigate whether domestic obligations may limit individuals' ability to participate in income-generating activities. We find that self-employed workers adjust non-monotonically to changes in domestic productivity, initially increasing their levels of domestic work to preserve consumption levels, but then substituting towards market work when power outages become more severe. Interestingly, self-employed women are less able to adjust their time-use than self-employed men. By contrast, there are no robust relationships between power outages and time allocation for the wage-employed. We also build and simulate a model of time allocation to help understand the theoretical mechanisms behind these results.

The primary contribution of Chapter 2 is empirical. To our knowledge this is the first attempt to use high-frequency data to document short-run volatility in the provision of public services and then link this variation to labour market participation. We build on an existing literature, which examines whether improvements to domestic productivity may allow women to reduce the time they dedicate to domestic work. However, we clearly focus on a different periodicity and our findings must be interpreted accordingly. We believe our results help us to understand how workers are able to manage their time between income-generation and other activities week-by-week. In particular, we demonstrate the extent to which individuals are really *obliged* to undertake certain domestic tasks each week and consider what this implies for welfare and how people choose their jobs.

In Chapter 3 we examine whether the factors that drive occupational selection differ by sex. It is often argued that women become self-employed because this allows them to balance income-generation with childcare and other domestic work. We test the plausibility of this claim and its implications for labour market outcomes. This links directly to our analysis of the substitutability between domestic tasks and income-generation from Chapter 2. First, we use a simple model of occupational choice to clarify our ideas about which notions of 'job flexibility' are important for the Ghanaian context. Second, we examine whether differential selection forces between women and men may explain the raw sex earnings gaps that appear to persist in various sectors, using a multinomial logit model to adjust for non-random occupational selection. We find that controlling for selection substantially widens the earnings gap amongst the self-employed, but shrinks it for the wage-employed. We then consider what this implies for the unobservables that influence income-generation in each sector. Third, we interrogate our selection equations and show that domestic obligations, measured by the structure of the household, affect women's likelihood of entering low-input self-employment jobs more than men. We demonstrate that this finding is robust to concerns about endogeneity.

Chapter 3 is also mainly empirical in its contribution. On the one hand, we build on a large literature documenting differences in occupational selection between women and men, as well as a growing body of work that decomposes earnings gaps in developing country contexts. However, the literature is somewhat sparse on the specific role that job flexibility plays in determining outcomes for female entrepreneurs. We try to address this shortcoming. We also develop a new technique to assess the endogeneity of household structure to occupational choice, as we wish to know whether the former has a causal effect on the latter. Specifically, we use a maximum simulated likelihood estimator to capture the idea that selection on observables can be used as a guide for selection on unobservables, focussing on the discrete choice made over occupation. Although this method has been widely used on continuous and binary outcome variables, we believe this is the first attempt to apply it to a multinomial problem.

In Chapter 4, we turn to theory to try and resolve some of the empirical puzzles that remain from Chapter 3. In particular, we attempt to reconcile the fact that female participation in self-employment is so high even when the average sex earnings gaps in that sector are large. To do this, we construct a search model, which allows for individual heterogeneity and participation in both self- and wage-employment, as well as discrimination against female workers in the wage sector. We numerically solve and simulate this model, using calibrations from the existing literature. This enables us to explain a set of stylised facts that we generate from a longitudinal dataset of workers in urban Ghana. We show that wage sector discrimination leads to average earnings gaps in *all* sectors of the economy, even if the underlying ability distribution is the same for both sexes. Our model captures the logic of the Roy (1951) model, insofar as individuals' comparative advantage partially determines sorting between occupations, but also allows for labour market segmentation to differentially affect women and men. In addition, we conduct a series of experiments to examine how the effects of government policy may differ according to sex.

Chapter 4 is largely a theoretical exercise, as we investigate the impact of extra frictions on women and men the labour market. Although there have been numerous papers using search modelling to analyse multi-sector economies and discrimination separately, this is the first attempt to bring these two strands of the literature together. We link our model to data on broad sectors in urban Ghana, but we believe the logic of our results can be applied to a number of settings. Firstly, the stylised facts to which we match our model hold in other countries in sub-Saharan Africa. However, we also believe the logic of our framework can be applied to different *levels* of occupational sorting. Even *within* the informal sector, there is a great deal of heterogeneity over who gets what type of job — such as manufacturing, services, or retail — and our model can be modified to capture some key features of this sorting process.

Finally, in Chapter 5 we connect our main findings to policy and make some suggestions for future work.

1.6 Common Themes

Although each chapter tackles slightly different research questions, there are a three common themes which run throughout this thesis.

Firstly, we are interested in how people allocate their time. We recognise that individuals do not simply face a choice over working in the market or taking leisure, but they may also undertake other domestic tasks, which do not yield monetary rewards. These trade-offs are especially important in developing countries, because the returns to income-generating activities are generally low. Also, social norms strongly influence how tasks are allocated across members of the household, and this has direct consequences for the labour market outcomes of women and men. The increased availability of time-use data has intensified economists' focus on questions of task allocation in recent years, and this thesis attempts to complement this literature (Aguiar et al., 2012).

Secondly, we explore the factors that shape occupational choice. This is closely linked to questions about time allocation, insofar as individuals' management of non-market activities influences the type of market work that they are able to do. Since different jobs clearly have different monetary and non-monetary returns, understanding why women and men are employed in particular sectors is vital for addressing inequality. Recent advances in the quality of micro-data from developing countries have shown the importance of certain individual characteristics for determining individuals' welfare, and we regard occupational selection as one vital channel. We recognise, however, that choosing a job is about more than just moving between all-encompassing categories like self-employment and wage-employment. Crucially, there is extra sorting within these broad sectors, across junctures based around industry and technology. We want to know which of these junctures matter.

Thirdly, we acknowledge the existing evidence on Ghana and sub-Saharan Africa at large, by allowing for the possibility that labour markets are non-competitive. The literature on labour market segmentation highlights the possibility that some workers may be disadvantaged by barriers to entry into certain sectors of the labour market. We try to ascertain who these disadvantaged workers are likely to be. Thus, by examining how individual characteristics determine the extent to which workers are affected by labour market frictions, we take the logic of both segmentation and sorting seriously.

Chapter 2

Are Market and Domestic Work Substitutable?

2.1 Introduction

The trade-off between domestic work, such as caring for children and undertaking household chores, and market work is important. Firstly, in a basic sense, domestic obligations may limit individuals' ability to generate market income for their household. Secondly, women may suffer a disproportionate limit to their market activities if, either due to social norms or intra-household comparative advantage, their obligations to perform domestic tasks are greater. Finally, if individuals are unable to substitute between domestic work and market work, then welfare will be far more susceptible to negative shocks to the productivity of these activities.

Whilst the trade-off between domestic work and market work has been the subject of many household models (Becker, 1965; Gronau, 1977), theory alone yields ambiguous predictions about the impact of changes in exogenous factors, such as wage rates, domestic productivity, and unearned income, on the time allocated to particular tasks.

Subsequently, an empirical literature has emerged to examine these questions. This chapter draws on numerous papers from both developed and developing countries contexts, which examine whether improvements in domestic task productivity, either from the proliferation of time-saving appliances or improvements in infrastructure, allow individuals to reduce the time allocated to domestic work in favour of market work.

In this chapter, we exploit week-on-week variation in the quality of electricity provision in Accra, Kumasi, and Cape Coast, three of Ghana's largest cities, to examine substitutability between market work, domestic work, and leisure. In particular, we utilise data collected through weekly interviews during the Ghana High-Frequency Labour Survey 2013, in which individuals were able to report the time they allocated to various tasks, as well as the duration and timing of power outages. As such, the periodicity of our data is different to the existing evidence, allowing us to address similar questions from a new angle.

We find that individuals respond heterogeneously to power outages depending on their occupational status. For self-employed individuals, there is a significant non-monotonic relationship between our best proxy for domestic productivity — total duration of power outage at home in a given week — and hours of domestic and market work. Initially, extra hours of power outage result in work being re-allocated from the market to domestic tasks. As outages become more severe, however, individuals begin to substitute away from domestic work towards market work. This relationship appears to be heterogeneous by sex, with women less able to adjust than men. In contrast, the wage-employed display virtually no response to power outages. Our results therefore allow us to ascertain which individuals are truly *obliged* to undertake domestic tasks.

We also build a simple model of weekly time allocation to explain these findings. By calibrating and simulating this model, we are able to rationalise the data, and understand the combinations of 'demand effects' and 'supply effects' required to produce the relationships we observe.

This chapter proceeds as follows. In Section 2.2 we review the related literature. In Section 2.3 we outline the conceptual framework used to examine substitutability between market and domestic work. In Section 2.4 we explain the context in which our data was collected, and report some key descriptive statistics. In Section 2.5 we outline the empirical strategy. We report our main results, and conduct a series of robustness checks in Sections 2.6 and 2.7 respectively. In Section 2.8 we return to our theoretical model and calibrate it so as to rationalise the data. We conclude in Section 2.9.

2.2 Related Literature

2.2.1 Theoretical Literature

The empirical literature on time allocation is theoretically motivated, in the main, by 'unitary' models of the household, in which it is assumed that household preferences can be represented by a single well-defined utility function. The unitary assumption has been justified in a number of ways. Becker's 'rotten kid theorem', for example, suggests that households contain some altruistic benefactor (a 'parent'), who may transfer goods to some selfish 'kids' (Becker, 1974, 1981). This set-up causes the selfish 'kids' to act in an ostensibly unselfish way, insofar as the altruistic benefactor can adjust transfers so as to incentivise all household members to maximise household income. As such, the preferences of the household and the altruistic benefactor converge. The implications of Becker's 'rotten kid theorem' therefore coincide with Samuelson's consensus model, in which household members agree to pool their resources and maximise some joint social welfare function of all individuals' utilities subject to a joint budget constraint (Samuelson, 1956).

Unitary models should be distinguished from 'collective' models of the household, in which individuals have their own distinct utility functions, which cannot be captured by an aggregate set of preferences for the household. Some collective models impose a specific structure on the interaction that occurs between household members. McElroy and Horney (1981), for example, outline a model where individual outcomes are determined by Nash bargaining, in which household members' outside options determine their bargaining power. Other collective models are less restrictive about the structure of intra-household interactions. Chiappori's (1997) model makes one key assumption — intra-household outcomes will be Pareto efficient. Under this assumption, any interaction between household members can be captured by simply distributing household resources to each member according to particular welfare weights, and then allowing each individual to solve their own utility maximisation problem.[1]

Although a number of empirical studies have sought to reject the unitary model (Thomas, 1990; Alderman et al., 1995), the trade-off between market work, domestic work, and leisure is far more tractable if the household can be characterised by one set of preferences and production technology. Gronau's (1977) adaptation of the neoclassical household model was amongst the first attempts to delineate domestic work — such as childcare, cleaning, repair work, cooking, and washing — from leisure. In his model, households are able to buy market goods using earnings from market work, whilst they can produce domestic goods by allocating time to domestic work.

[1] Chiappori's model effectively applies the Second Fundamental Welfare Theorem at the intra-household level.

Problematically, however, the model restricts household preferences over domestic and market goods, such that they are perfect substitutes.

The vital insights from Gronau are generalised in de Janvry, Fafchamps, and Sadoulet's (1991) household model, in which households derive utility from multiple goods. In this model, goods may be either tradeable or untradeable, which captures an important feature of domestic goods that underpins much of the empirical literature. Namely, goods such as childcare and other domestic chores do not have close substitutes from outside the household and, as such, cannot simply be bought in.

2.2.2 Domestic Productivity and Time Allocation

Domestic productivity is likely to be endogenous to decisions about time allocation, insofar as both are jointly driven by unobserved household characteristics. The existing empirical literature has therefore exploited a number of sources of exogenous variation to try and identify the causal impact of changes in the determinants of domestic productivity on the time spent doing domestic work and market work.

In a developed country context, economists have consistently argued that the proliferation of household appliances — the so-called 'Engines of Liberation' — enabled women to participate more in the labour market in the post-War era (Greenwood et al., 2005; Cavalcanti and Tavares, 2008; Cardia, 2008; Coen-Pirani et al., 2010). For example, Greenwood et al. (2005) build a dynamic model, in which individuals make lifetime decisions in markets for labour, capital, and domestic appliances. Utilising data on changes in the price of a number of household appliances and simulating their model, they argue that more than half the observed rise in female labour force participation in the United States can be explained by the introduction of new or improved household technologies.

The evidence linking increases in domestic productivity and labour market participation in developing countries, which focusses mostly on improvements in infrastructure and public services is, in contrast, far more mixed.

In their assessment of the impact of water infrastructure on women's labour supply, Koolwal and van de Walle (2013), aggregate household data to the community-level and use extensive geographic controls to purge their estimates of the endogeneity described above. Utilising rural household surveys from nine countries in sub-Saharan Africa, South Asia, and the Middle East and North Africa (MENA), they find that the time gains enjoyed when infrastructure is improved rarely result in extra female participation in income-generating activities. Using similar techniques on cross-sectional data on access to water in Pakistan, Ilahi and Grimard (2000) also find that, for more than half of their sample, household members substitute any time saved in domestic work for leisure rather extra market work.

Given the possibility of time-invariant confounds, which may not be tackled through the inclusion of controls, some have used panel techniques in conjunction with natural or artificial experiments to assess the impact of improved public services on task allocation. Using data from rural Kyrgyzstan, Meeks (2012) exploits variation in the timing of the construction of shared public water taps combined with a difference-in-differences approach to assess how improved water access influences household time allocation. For both men and women, although the majority of the time saved in childcare and water collection is re-allocated to leisure, an extra hour per day is also devoted to the primary income-generating activity in the Kyrgyz sample, working on the household farm.

Devoto et al. (2012) adopt a similar approach in urban Morocco, but rather than relying upon a natural experiment for exogenous variation in infrastructural quality, they utilise a randomised controlled trial. In the treatment group, they provide households with extra information and help with the administrative procedures required to apply for private water connections on interest free credit, in a context where water is ordinarily taken from public taps. The effects of this treatment were substantial, with 69 percent of respondents in the treatment group purchasing the private water connection, compared with just 10 percent of the control group. However, they also find that time saved in domestic work from owning a private tap was used almost entirely for 'leisure and socializing' rather than extra income-generating activity (Devoto et al., 2012).

Since the results from developing country contexts are so mixed, we are presented with an open empirical question which may be tested using information from urban Ghana. The periodicity of our analysis is somewhat different to the existing evidence, so our results must be interpreted accordingly, but the questions that we are able to answer with our data are closely linked to this literature.

2.2.3 The Economics of Power Outages

The literature on power outages has focussed largely on the costs endured by medium- and large-sized firms and their resultant investment responses, rather than assessing their impact on households. There are, however, some lessons to be taken from this literature insofar as households adjust their behaviour to power cuts in similar ways to firms, especially in the urban Ghanaian context, where self-employment is so prevalent (Steinbuks and Foster, 2010; Falco et al., 2011).

Some authors have estimated the losses firms suffer during power cuts, both in terms of lost output and damage to equipment, through a 'revealed preference approach', by assessing the adjustments made to cope with such costs (Bental and Ravid, 1982; Beenstock, 1990). Using data on generator uptake in Nigeria, Adenikinju (2003) estimates the costs of production during power outages, which necessitate private provision of electricity, to be roughly three times as large as when electricity is publicly provided. Pasha et al. (1989) take an especially holistic approach to estimating the costs of electricity outages on data from Pakistan by allowing for both short- and long-run adjustments by firms. For example, in the short-run, firms may require their staff to work overtime to make up for lost output, whilst in the long-run, they may permanently renegotiate the timing of shifts to minimise the risk that workers find themselves at work with the power out (Shangvi, 1982; Pasha et al., 1989).

Others have modelled the implications of electricity and water outages more explicitly, to consider the response of investment in both 'productive capital' (such as tools and machines) and private 'complementary capital' (such as electricity, water, telephone lines). Reinikka and Svensson (2002) build a model in which the public provision of complementary capital is ex ante uncertain, but firms learn about potential infrastructure reform before deciding how much productive capital and private complementary capital (such as generators) to invest in. Although the return to productive capital depends on the provision of complementary capital, investment in productive capital is crowded out if firms choose to provide complementary capital privately.[2] The predictions of the model are then tested using firm survey data from 171 Ugandan firms.

2.2.4 Intertemporal Wage-Elasticities and High-Frequency Data

Whilst the literature on short-term labour supply decisions is not directly analogous to this chapter per se, there are a number of key themes that complement our approach. Firstly, a number of papers demonstrate the value of using short-run exogenous variation to help identify

[2] The model implies; (1) that firms with low expectations of publicly provided complementary capital will invest in private replacements (such as generators), (2) that for firms providing their own complementary capital, the quality of public complementary capital will have no impact on investment in productive capital, (3) that firms relying on public provision of complementary capital will make less productive investment when the public provision is low quality, and (4) that when the quality of publicly provided complementary capital is high, firms that install private substitutes may invest less than firms relying on public provision.

key parameters. For example, in his analysis of vendors at sports stadiums in the US, Oettinger (1999) uses the determinants of baseball game attendance to instrument for wages, and hence estimate vendors' labour supply elasticities. Similarly, Connolly (2008), examines how individuals intertemporally substitute between work and leisure, when the return to the latter changes with shocks to the weather, using extensive time-use and meteorological data, again in the US.

Secondly, we draw upon the growing behavioural economics literature, which uncovers negative labour supply elasticities in certain types of labour markets, and attempts to reconcile these findings theoretically.[3] Using data on taxi-drivers in New York, Camerer et al. (1997) present a set of empirical results, which suggest that individuals may make decisions about their labour supply 'one day at a time'.[4] They attribute these findings to the behavioural theory of 'narrow bracketing', where individuals isolate particular decisions, rather than taking the entire stream of potential decisions together (Read and Loewenstein, 1996). Similar results are demonstrated in a developing country context by Dupas and Robinson (2013) in their study of Kenyan bicycle taxi-drivers. We believe that the concept of narrow bracketing supports our modelling approach, which initially assumes that decisions are made 'one week at a time'.

2.3 Conceptual Framework

Following Koolwal and van de Valle (2013), we underpin our empirical analysis with a simple model of time allocation.

Individuals choose their levels of domestic work (t_d), market work (t_m), and leisure (t_l) to maximise utility week-by-week. Utility is derived from the consumption of market goods (x_m) and domestic goods (x_d), as well as leisure. The consumption of market goods is constrained by the time devoted to market work and its return (w), in addition to unearned endowments of market goods (τ_m). Analogously, domestic goods consumption is constrained by the time in domestic work and domestic productivity (ρ), and an endowment of domestic goods (τ_d).[5] In the

[3] These studies generally focus on the elasticity of labour supply with respect to wages.

[4] In particular, Camerer et al. (1997) suggest that labour supply decisions may be characterised by daily targets. There is a substantial literature trying to understand how these targets can be theoretically rationalised, and whether they relate to income, hours worked, or both (Farber, 2005; Kőszegi and Rabin, 2006; Crawford and Meng, 2011).

[5] We specify the model at the individual rather than the household level as this better suits our data. Our model does not explicitly capture intra-household interactions, but by including endowments of market and domestic goods in each individuals budget constraint, we are effectively allowing that individuals maximise their utility taking the contributions and/or demands of other household members as given. Thus, we allow τ_m and τ_d to be positive or negative. This is in-keeping with the ethnographic literature on Ghana, which suggests that responsibility for domestic tasks and income-generation is separated across household members (Goldstein, 2004; Warner et al., 1997).

empirical analysis, we will interpret our main explanatory variable, namely the total duration of power outages within a week, as a proxy for domestic productivity. We assume that an increase in the duration of power outages corresponds to a fall in ρ.

The maximisation problem for the individual can thus be written as:

$$\max_{x_m, x_d, t_l} \; U = u(x_m, x_d, t_l) \tag{2.1}$$

$$\text{subject to } t_m + t_d + t_l = 1 \tag{2.2}$$

$$x_m = wt_m + \tau_m \tag{2.3}$$

$$x_d = \rho t_d + \tau_d \tag{2.4}$$

$$t_k \geq 0 \; \forall k = m, d, l. \tag{2.5}$$

This problem can be solved for a system of Marshallian demands, which depend on the exogenous parameters of the model.

$$\begin{pmatrix} x_m^* \\ x_d^* \\ t_l^* \end{pmatrix} = \begin{pmatrix} x_m(w, \rho, \tau_m, \tau_d) \\ x_d(w, \rho, \tau_m, \tau_d) \\ t_l(w, \rho, \tau_m, \tau_d) \end{pmatrix} \tag{2.6}$$

This solution can also expressed in terms of time-use, by rearranging constraints (2.3) and (2.4) for t_m and t_d and substituting in the Marshallian demands for goods, x_m^* and x_d^*.

$$\begin{pmatrix} t_m^* \\ t_d^* \\ t_l^* \end{pmatrix} = \begin{pmatrix} t_m(w, \rho, \tau_m, \tau_d) \\ t_d(w, \rho, \tau_m, \tau_d) \\ t_l(w, \rho, \tau_m, \tau_d) \end{pmatrix} = \begin{pmatrix} \frac{x_m^* - \tau_m}{w} \\ \frac{x_d^* - \tau_d}{\rho} \\ t_l^* \end{pmatrix} = \begin{pmatrix} \frac{x_m(w,\rho,\tau_m,\tau_d) - \tau_m}{w} \\ \frac{x_d(w,\rho,\tau_m,\tau_d) - \tau_d}{\rho} \\ t_l(w, \rho, \tau_m, \tau_d) \end{pmatrix} \tag{2.7}$$

It is the response of these time allocations to changes in ρ that we will investigate in our empirical work.

Since the Marshallian demands for time are simple rearrangements of the Marshallian demands for goods, we can describe some of the mechanisms through which ρ affects time allocation, *without* adding functional form or finding explicit solutions for Equation (2.7).

First, we note that the constraints in Equations (2.2), (2.3), and (2.4) can be combined without loss of generality for:

$$x_m + \left(\frac{w}{\rho}\right) x_d + w t_l = \tau_m + w + \left(\frac{w}{\rho}\right) \tau_d \tag{2.8}$$

Thus, we effectively have a single budget constraint for the three goods. On the left-hand-side we can see that the price of market goods is normalised to one, whilst the relative prices of domestic goods and leisure are $\frac{w}{\rho}$ and w respectively. These relative prices are also reflected in the endowments of market goods, time, and domestic goods on the right-hand-side.

Also, focussing on domestic work, we can differentiate t_d^* in Equation (2.7) with respect to ρ to obtain:

$$\frac{\partial t_d^*}{\partial \rho} = \underbrace{\frac{1}{\rho}\frac{\partial x_d^*}{\partial \rho}}_{\text{`Demand Effects'}} - \underbrace{\frac{(x_d^* - \tau_d)}{\rho^2}}_{\text{`Supply Effects'}} \tag{2.9}$$

Equation (2.9) helps elucidate the sources of ambiguity in the relationship between ρ and t_d^*. We begin by distinguishing between what we label 'demand effects' and 'supply effects'. The first term on the right-hand-side of Equation (2.9) relates to how demand for the domestic good adjusts when its (relative) price changes, so we label these 'demand effects'. By contrast, the second term on the right-hand-side of Equation (2.9) relates to how individuals' ability to provide a *given* level of domestic goods changes when domestic productivity fluctuates, so we label these 'supply effects'. We argue that, whilst the supply effects must be non-positive by construction, the overall sign of the demand effects, and hence of $\frac{\partial t_d^*}{\partial \rho}$, is ambiguous.

Inspection of Equation (2.8) reveals three potential demand effects that ρ may exert on x_d^*.

Firstly, there is a regular substitution effect. As ρ increases, the relative price of x_d falls, which, holding all else constant, incentivises greater consumption of domestic goods.

Secondly, there is a regular income effect, which enters through the $\left(\frac{w}{\rho}\right) x_d$ term in Equation (2.8). If domestic goods are normal, this will reinforce the regular substitution effect. That is, an increase in ρ gives individuals extra purchasing power, increasing demand for x_d. There is, however, no a priori reason to assume away the possibility that domestic goods are inferior. If this is the case, the regular income effect would place downward pressure on demand for x_d, when ρ increases.

Thirdly, a change in ρ produces an extra income effect, or 'endowment effect', through the $\left(\frac{w}{\rho}\right)\tau_d$ term in Equation (2.8). The sign of this endowment effect will depend on whether τ_d is positive or negative. Assuming that $\tau_d > 0$, an increase in ρ reduces the relative size of the endowment of domestic goods. Interpreting this as an income shock alone, this should place downward pressure on the demand for all three goods, x_m, x_d, and t_l, assuming they are normal. As such, the endowment effect must have an opposite sign to the regular income effect and, assuming x_d is normal, the substitution effect. Thus, although the substitution effect and the regular income effect are likely to result in $\frac{\partial x_d^*}{\partial \rho} > 0$, the endowment effect may make the sign of $\frac{\partial x_d^*}{\partial \rho}$ ambiguous.

Even if $\frac{\partial x_d^*}{\partial \rho} > 0$ and demand effects are positive, the overall sign of $\frac{\partial t_d^*}{\partial \rho}$ is ambiguous, given the presence of what we label supply effects. These capture how changes in ρ affect individuals' ability to meet a given net demand for domestic goods, $(x_d^* - \tau_d)$. Increases in domestic productivity enable individuals to meet their desired level of x_d with less time spent in domestic work. The downward pressure this places on the optimal level of t_d is captured by the $-\frac{(x_d^* - \tau_d)}{\rho^2}$ term in Equation (2.9).[6]

As such, the combination of demand effects and supply effects results in an ambiguous relationship between ρ and t_d^* (as well as t_m^* and t_l^*). This ambiguity can only be resolved by estimating these relationships empirically. We then rationalise the results by adding functional form to our model and simulating outcomes in terms of both consumption and time-use.

[6] We note that $x_d^* \geq \tau_d$, regardless of the sign of τ_d. For $\tau_d < 0$, the τ_d term effectively captures some lower bound on consumption of domestic goods, much like the subsistence parameters in a Stone-Geary utility function (Geary, 1950; Stone, 1954). For $\tau_d > 0$, the τ_d term is effectively an endowment of domestic goods. However, the individual cannot trade this endowment of domestic goods for extra x_m or t_l so, assuming $\frac{\partial U}{\partial x_d} > 0$, they should set x_d at least as high as τ_d.

2.4 Data and Context

2.4.1 Data

The data are taken from the Ghana High-Frequency Labour Survey (GHFLS) 2013, which collected information on nearly 1600 individuals living in Accra, Kumasi, and Cape Coast between July and November 2013. Households were randomly selected for the high-frequency survey from the existing Ghana Household Urban Panel Survey (GHUPS), collected by the Centre for the Study of African Economies since 2004.

Respondents in the GHFLS 2013 were randomly assigned to three treatment and two control groups. These groups only differ insofar as data were collected from the respondents using different methods:

1. *Control Without Phone* — respondents were interviewed face-to-face at baseline and then three months later at endline. They were paid a small monetary incentive to participate each time they were interviewed.
2. *Control With Phone* — respondents were interviewed face-to-face at baseline and given a mobile phone. They were then interviewed face-to-face three months later, and paid a small monetary incentive to participate.
3. *Called Once a Week* — respondents were interviewed face-to-face at baseline and given a mobile phone. Using this phone, or another if the respondent preferred, interviews were conducted once a week, asking about the previous seven days. If the respondent successfully completed their scheduled interview the phone was topped up with credit and a small monetary incentive was set aside to be distributed at endline. The endline interview was conducted face-to-face.
4. *Called Three Times a Week* — the main details were the same as the 'Called Once a Week' treatment, but three interviews were scheduled every week, each time asking questions about the previous two days.
5. *Face-to-Face (Once a Week)* — identical to the 'Called Once a Week' treatment, but the respondent was not given a phone, and weekly interviews happened during face-to-face meetings between enumerator and respondent.

Each high-frequency interview was comprised of a series of questions about individuals' recent occupation, earnings, and time-use. Individuals were also asked to report on the extent and timing of interruptions to their water and electricity provision.

The randomisation occurred at the household level, so individuals within a given household all had data collected in the same way. However, not all individuals in a treated households took part in the survey. The mean number of individuals from a particular household included in the sample is 2.60.[7]

The number of individuals in each treatment group at baseline is shown in Table 2.1.

Table 2.1: Breakdown of Treatment Groups by City

	City			
Treatment	Accra	Cape Coast	Kumasi	Total
Called Once a Week	168	20	127	315
Called Three Times a Week	182	21	118	321
Control With Phone	172	19	123	314
Control Without Phone	168	21	122	311
Face-to-Face (Once a Week)	158	28	132	318
Total	848	109	622	1579

For reasons explained in Section 2.5, we focus mainly on individuals who are interviewed once a week, either face-to-face or over the phone. After cleaning this sample, this leaves 629 respondents for whom we have up to 12 observations.[8] The data cleaning process is explained in Appendix 2.B.

2.4.2 Electricity Provision in Ghana

The vast increase in demand for electricity in Ghana has not been matched by adequate reform of the power sector to keep pace and provide reliable service. *Alternative Energy Africa* estimate the rate of increase in demand for electricity in Ghana to be approximately 10–15 percent per year between 1990 and 2010 (Kemausuor et al., 2011).

The supply of energy is managed by the public sector through the following key institutions:

[7]We show household-level descriptive statistics in Appendix 2.A to illustrate how households were assigned across treatment groups.

[8]Weekly interviews were carried out for three months, with minimal 'full attrition' — although not all interviews were completed, very few individuals from the three treatment groups actually quit the survey.

- *Ministry of Energy* — responsible for formulating and implementing policies related to fuel and electricity.
- *Public Utilities Regulatory Commission* — organisation for setting tariffs and dealing with customer services.
- *Volta River Authority (VRA)* — generates and supplies electricity to large industrial/mining units as well as two electricity distribution companies:
    1. *The Electricity Company of Ghana (ECG)* distribute electricity to southern Ghana.
    2. *Northern Electricity Department (NED)* distribute electricity to northern Ghana.
- Various other private generation projects, such as the *Bui Development Authority* and *Takoradi International Company.*

As such, a number of large organisations are linked in a complex way with political economy constraints on reform, which may explain why supply of electricity in Ghana is lacking (Williams and Ghanadan, 2006).

In the months leading up to the start of the GHFLS 2013, power outages were partially governed by so-called 'load-shedding' guides. Under this system, ECG would produce a schedule at the start of the week, published in local newspapers and broadcast on the radio, detailing in advance when the electricity would be cut to certain areas of Ghana, to try and deal with the energy shortfall systematically. However, during the period of the GHFLS 2013, load-shedding was not being implemented. Power outages resulted from capacity shortages, as well as malfunction of various distribution equipment such as transformers and power lines (Nkwetta et al., 2010).[9]

Since the period of the survey, power cuts in Ghana have intensified both in terms of length and frequency, and load-shedding has been restored. Indeed, the summer of 2013 was, if anything, an uncharacteristically stable period of electricity provision. We discuss the implications of this in Chapter 5.

2.4.3 Sample Characteristics

The composition of the sample in terms of sex, marital status, age, and education is summarised in Table 2.2.

[9]These were the explanations provided during informal interviews with the Public Relations team at ECG as well as others reporting on the state of the Ghanaian energy sector using media/social media.

Table 2.2: Sample Characteristics

	N	Mean	S.Dev.	Min.	Median	Max.
Male? (1=Y, 0=N)	629	0.42	0.49	0	0	1
Married? (1=Y, 0=N)	629	0.47	0.50	0	0	1
Years of Education	629	10.14	3.44	0	10	18
Age	626	35.14	10.51	16	33	65
No. Children	629	1.89	2.05	0	2	14
Observations	629					

There are slightly more women and unmarried individuals in the sample, reflecting the underlying characteristics of the GHUPS from which the high-frequency respondents were drawn.[10] There is also significant variation in the age and education levels of the sample. We account for these dimensions of individual heterogeneity in our econometric approach.

2.4.4 Occupation

Individuals' adjustment to short-run productivity shocks may depend on their occupational status. Respondents reported whether they were typically unemployed or not working, self-employed, or wage-employed during the baseline interview. The breakdown of typical occupation by sex is shown in Table 2.3.

Table 2.3: Occupational Status by Sex

	Sex					
Occupational Status	Female		Male		Total	
	N	%	N	%	N	%
Unemployed/Not Working	125	34.34	74	27.92	199	31.64
Self-Employed	164	45.05	94	35.47	258	41.02
Wage-Employed	75	20.60	97	36.60	172	27.34
Total	364	100.00	265	100.00	629	100.00

41 percent of the respondents in our sample are typically engaged in self-employment, and a little over 60 percent of the self-employed are women. Only 27 percent of the respondents typically do wage work, and approximately 55 percent of the wage-employed are men. Therefore, we anticipate that female and male workers may adjust to power outages differently.

We also report intra-sector heterogeneity in the types of jobs people do in Table 2.4. Self-

[10] We discuss the sampling and composition of the GHUPS in more detail in Chapter 4, where the long panel is used more extensively.

employment is dominated by trading, especially amongst the women working in that sector. Such activities typically involve buying and selling small consumer goods such as food, clothes, and toiletries. For the wage-employed, however, service industries are the most prevalent type of job.

Table 2.4: Detailed Occupational Breakdown by Sex

	Sex					
	Female		Male		Total	
	No.	%	No.	%	No.	%
Self-Employed						
Manufacturing	10	3.88	11	4.26	21	8.14
Services	18	6.98	45	17.44	63	24.42
Trading	136	52.71	38	14.73	174	67.44
Total	164	63.57	94	36.43	258	100.00
Wage-Employed						
Manufacturing	7	4.07	10	5.81	17	9.88
Mining/Fish/Agric	0	0.00	3	1.74	3	1.74
Public Sector	11	6.40	10	5.81	21	12.21
Retail/Trade	10	5.81	13	7.56	23	13.37
Services	47	27.33	61	35.47	108	62.79
Total	75	43.60	97	56.40	172	100.00

2.4.5 Time-Use

The key outcome variables are the time that individuals allocate to domestic work and market work. The questionnaire was worded so as to encourage respondents to focus on primary activities by asking how much time was 'mainly' allocated to particular activities.[11,12] Domestic work includes household chores and childcare, whilst market work is defined as work done for pay, profit, or gain, in either self- or wage- employment. During the interviews, respondents were given the same examples of household chores — including cleaning, repair work, mending, cooking, and washing — so as to fix ideas equally across the sample.

Decisions about domestic and market work are taken on both an extensive and intensive margin. Each week, individuals can choose *whether or not* to undertake domestic and/or market work. If they choose to participate in either activity, then can then decide *how much* work to do.

In Table 2.5 we show how individuals of different sexes and occupations choose their domestic

[11] This idea was reinforced throughout the enumerators' training.

[12] Although a full account of time-use, including secondary and tertiary activities may have been preferable, this set-up provides a plausible signal of how individuals allocated their time during a given week, even in the presence of multi-tasking. We allow for multi-tasking explicitly in Chapter 3.

and market work along the extensive margin. Pooling the observations across all rounds of the high-frequency survey, we can see that self-employed and wage-employed individuals undertook both domestic and market work approximately 65 percent of the time. For the self-employed, we can see that participation in domestic work is higher amongst women than men, but this disparity is not as strong amongst the wage-employed. Participation in market work is diminished for the full sample, because this includes respondents who were typically unemployed or not working.

Table 2.5: Extensive Work Margin by Sex and Occupation

	Sex					
	Female		Male		Total	
	No.	%	No.	%	No.	%
Self-Employed						
Both	1225	45.76	463	17.30	1688	63.06
Domestic Only	344	12.85	90	3.36	434	16.21
Market Only	93	3.47	356	13.30	449	16.77
No Work	45	1.68	61	2.28	106	3.96
Total	1707	63.77	970	36.23	2677	100.00
Wage-Employed						
Both	576	32.62	560	31.71	1136	64.33
Domestic Only	143	8.10	89	5.04	232	13.14
Market Only	40	2.27	280	15.86	320	18.12
No Work	11	0.62	67	3.79	78	4.42
Total	770	43.60	996	56.40	1766	100.00
Full Sample						
Both	1943	29.90	1101	16.94	3044	46.84
Domestic Only	1550	23.85	625	9.62	2175	33.47
Market Only	142	2.18	678	10.43	820	12.62
No Work	129	1.98	331	5.09	460	7.08
Total	3764	57.92	2735	42.08	6499	100.00

Turning to the intensive margin, the distribution of hours dedicated to domestic work in a given week is shown in Figure 2.1, pooling across occupation and sex. This includes variation both across individuals, and within individuals over time. However, the within-individual variation in hours of domestic work per week is also large — the mean of the within-individual standard deviation is 6.24 hours. In Panel A we show the raw data, including all outliers. We then drop outliers in Panel B.[13] Focussing on the intensive margin, we show the distribution without zeros in Panel C. Finally, we show the distribution of the log of domestic work hours in Panel D. We use the logarithmically transformed version of work hours in our estimating equations, as we explain in Section 2.5.

We repeat these charts for market work in Figure 2.2. Again, although these histograms show all datapoints, across both individuals and time, there is also substantial within-individual variation. The mean of the within-individual standard deviation is 9.07 hours.

[13]Outliers are dropped at the 95th percentile.

Figure 2.1: Distribution of Hours of Domestic Work per Week

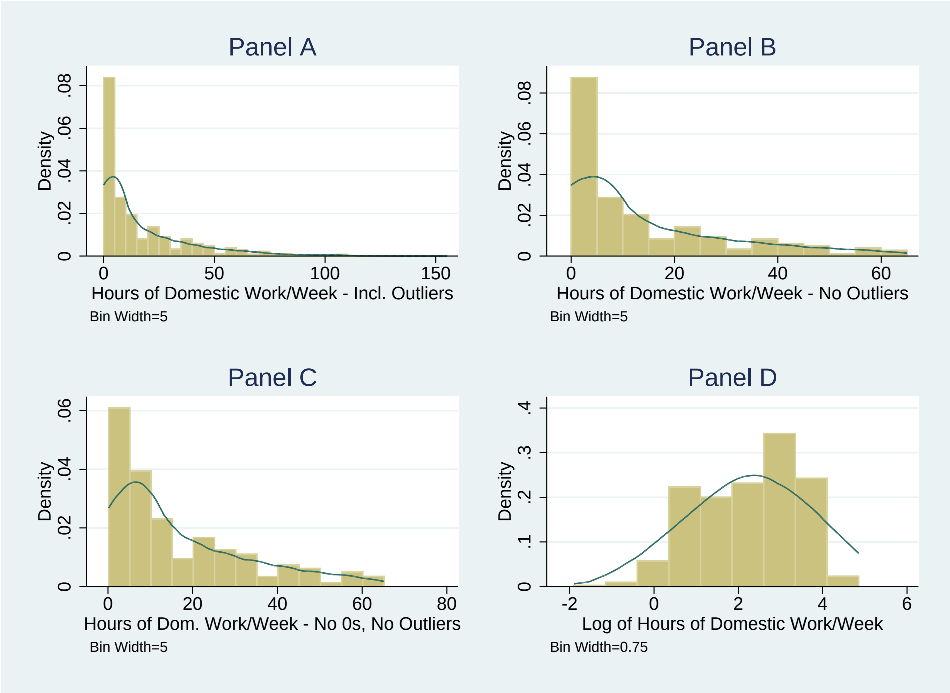
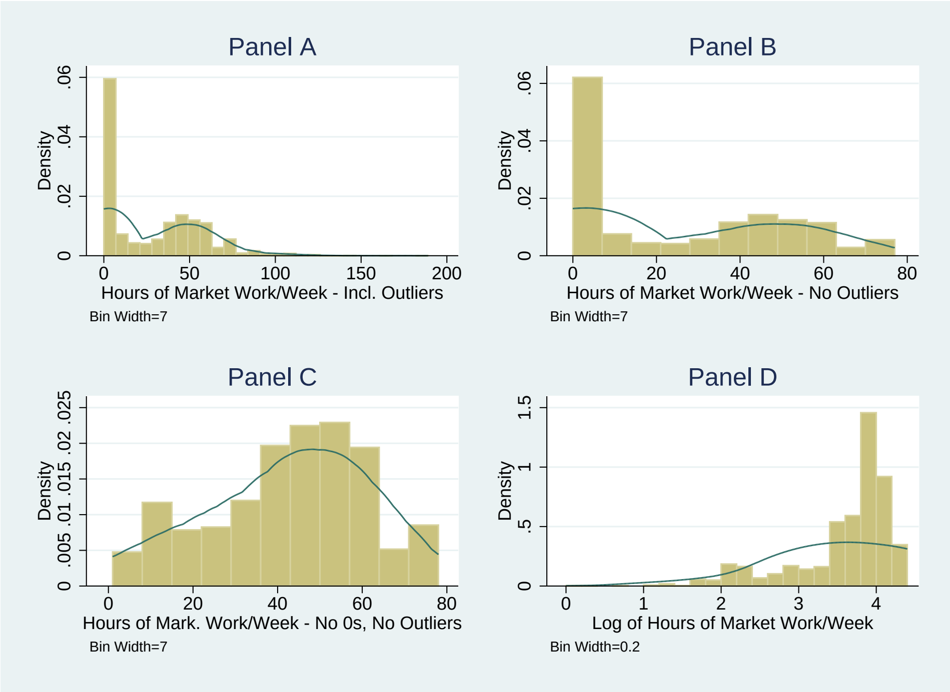


Figure 2.2: Distribution of Hours of Market Work per Week



Summary statistics for time allocation, broken down by sex, are presented in Appendix 2.C.

2.4.6 Use of Electricity

In order to link the data to our conceptual framework, we use power outages as a proxy for productivity in domestic work, ρ. The ways in which individuals use electricity will determine the suitability of this proxy.

At the baseline interview, respondents were asked whether or not their household was connected to the electricity grid, and whether or not they owned a generator. Table 2.6 demonstrates that the majority of households were connected to the electricity grid, but that they did not have back-up generation. As such, households generally lack insurance against any effects power outages may have on productivity at home.

Table 2.6: Electricity at Home

	Household connected to electricity grid?					
Household has a generator?	No		Yes		Total	
	N	%	N	%	N	%
No	7	1.11	604	96.03	611	97.14
Yes	0	0.00	18	2.86	18	2.86
Total	7	1.11	622	98.89	629	100.00

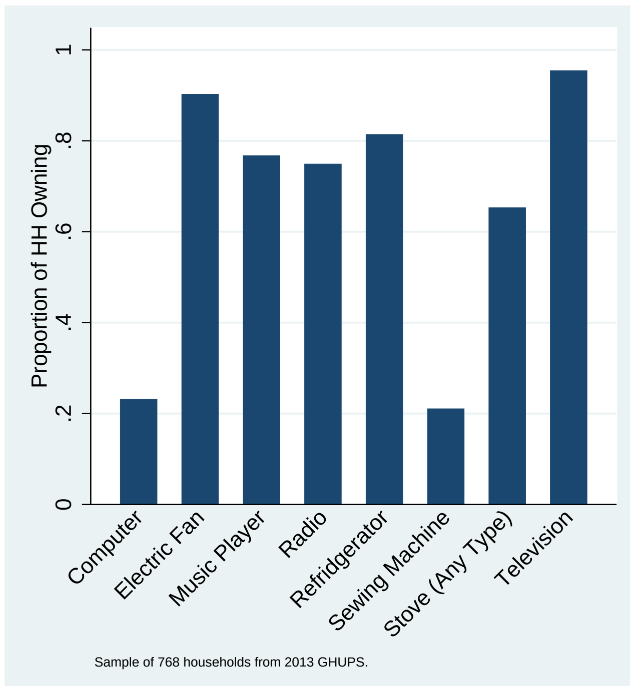
During the endline interview, respondents were asked whether or not they used electricity (and water) to help with household chores and childcare. Table 2.7 summarises individuals' responses to this question.

Table 2.7: Electricity and Water Usage in Domestic Work

Use electricity/water to help with household chores or childcare?	N	%
No	45	7.15
Yes, both	544	86.49
Yes, just electricity	17	2.70
Yes, just water	23	3.66
Total	629	100.00

Approximately 90 percent of the sample use electricity to help with household chores and childcare so the assumption that power outages may proxy for domestic productivity is plausible. There are two key channels through which electricity provision may influence the productivity of domestic work. Firstly, individuals may use appliances, such as stoves, refrigerators and sewing machines, which save time. Secondly, electricity may make the *environment* in which domestic tasks are undertaken easier to work in, by providing light, cooling, and distraction of children (such as with television or radio). The pattern of household asset ownership in Ghana, which we show using data from the 2013 wave of the GHUPS in Figure 2.3, supports the idea that both these channels may be plausible.

Figure 2.3: Ownership of Household Assets



In contrast to domestic work, market work is far less exposed to the impact of power outages for the majority of the sample. As Table 2.8 demonstrates, approximately 25 percent of the individuals typically working in the market, use electricity for their work without back-up generation.

Table 2.8: Electricity at Work

	Typically uses electricity for work?					
Main workplace has generator?	No		Yes		Total	
	N	%	N	%	N	%
No	225	52.33	111	25.81	336	78.14
Yes	13	3.02	81	18.84	94	21.86
Total	238	55.35	192	44.65	430	100.00

For individuals that work in the market with electricity but without back-up generation, power outages may proxy for market work productivity or wages, w, rather than domestic work productivity, ρ. This may present a problem in linking our conceptual framework to the data. However, if individuals work in a different location from their home, then power outages at home proxy far better for ρ than w. Furthermore, it is unlikely that the wage-employed will experience changes in their wages, insofar as wage contracts do not undergo the same short-run variation experienced by the self-employed. We tabulate work location and occupational status for the sub-sample of individuals using electricity at work without back-up generation in Table 2.9.

Table 2.9: Work Location and Occupational Status

	Work location close to home?					
Occupational Status	No		Yes		Total	
	N	%	N	%	N	%
Self-Employed	14	12.61	51	45.95	65	58.56
Wage-Employed	25	22.52	21	18.92	46	41.44
Total	39	35.14	72	64.86	111	100.00

2.4.7 Power Outages

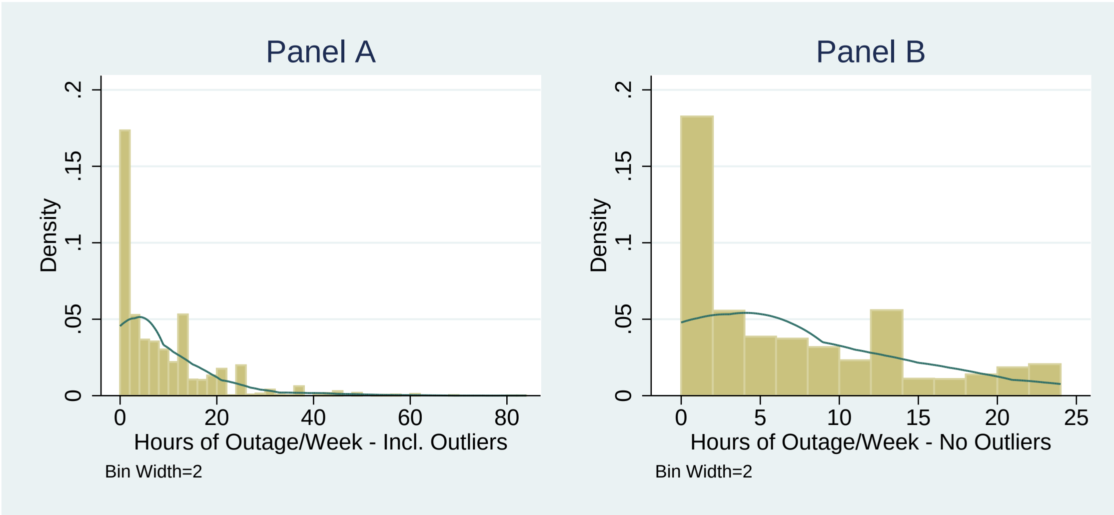
Our main proxy for domestic productivity is the number of hours in a particular week when the electricity was cut. Although questions relating to power outages were asked at the individual level during the weekly interviews, we harmonise these reports to the household level as we explain in Appendix 2.B.

In Figure 2.4, we plot histograms for the hours of outage per week variable, both with outliers (Panel A) and without (Panel B). Whilst we exclude the outliers in our final specifications, we maintain the zeros and do not logarithmically transform this variable.

Although Figure 2.4 demonstrates that the dispersion in the severity of power outages for particular household-weeks is high, our identification strategy relies on this variation occurring across both time *and* space. Using Global Positioning System (GPS) data from previous waves of the GHUPS, we can show the geographical spread of power outages for Accra and Kumasi.[14] During the weekly interviews, individuals were asked which days of the previous seven were affected by power outages. We aggregate this data to the Enumeration Area (EA) level, by calculating the proportion of households within a particular EA that experienced outages on a

[14]Reliable GPS data is not available for Cape Coast.

Figure 2.4: Distribution of Hours of Power Outage per Week



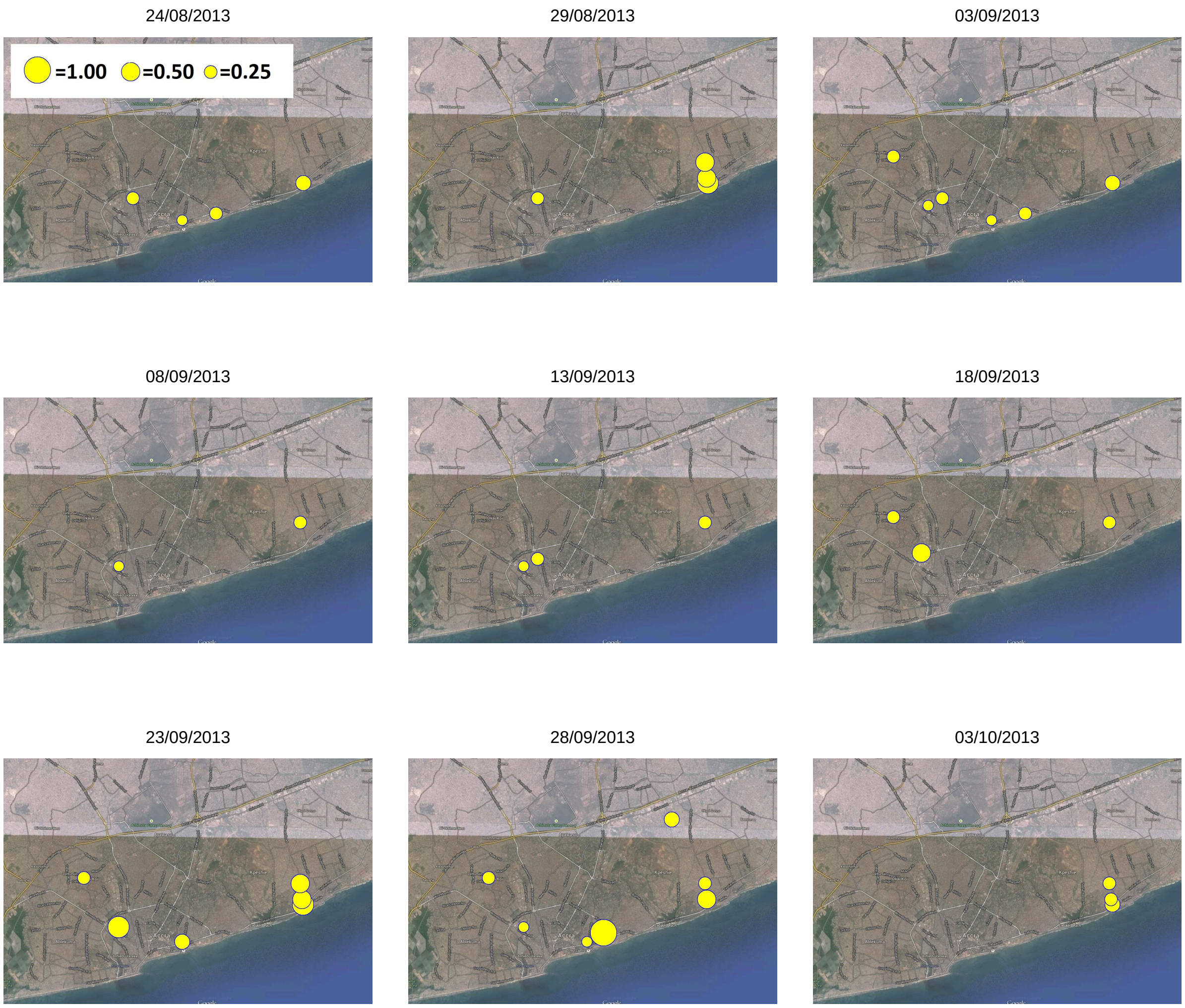
particular day.[15] The size of the circles in Figure 2.5, which focusses on Accra, are proportional to the percentage of surveyed households within each EA that experienced power outages on particular days in the survey.[16]

It appears that the effects of power outages are dispersed across both time and space. We show the analogous maps for Kumasi in Figure 2.6.

[15]EAs are typically identified by a local landmark that enumerators can use to try and locate households. They thus provide an approximate location for households, but do they do not have strict borders in the same way as 'municipalities' or 'districts'.

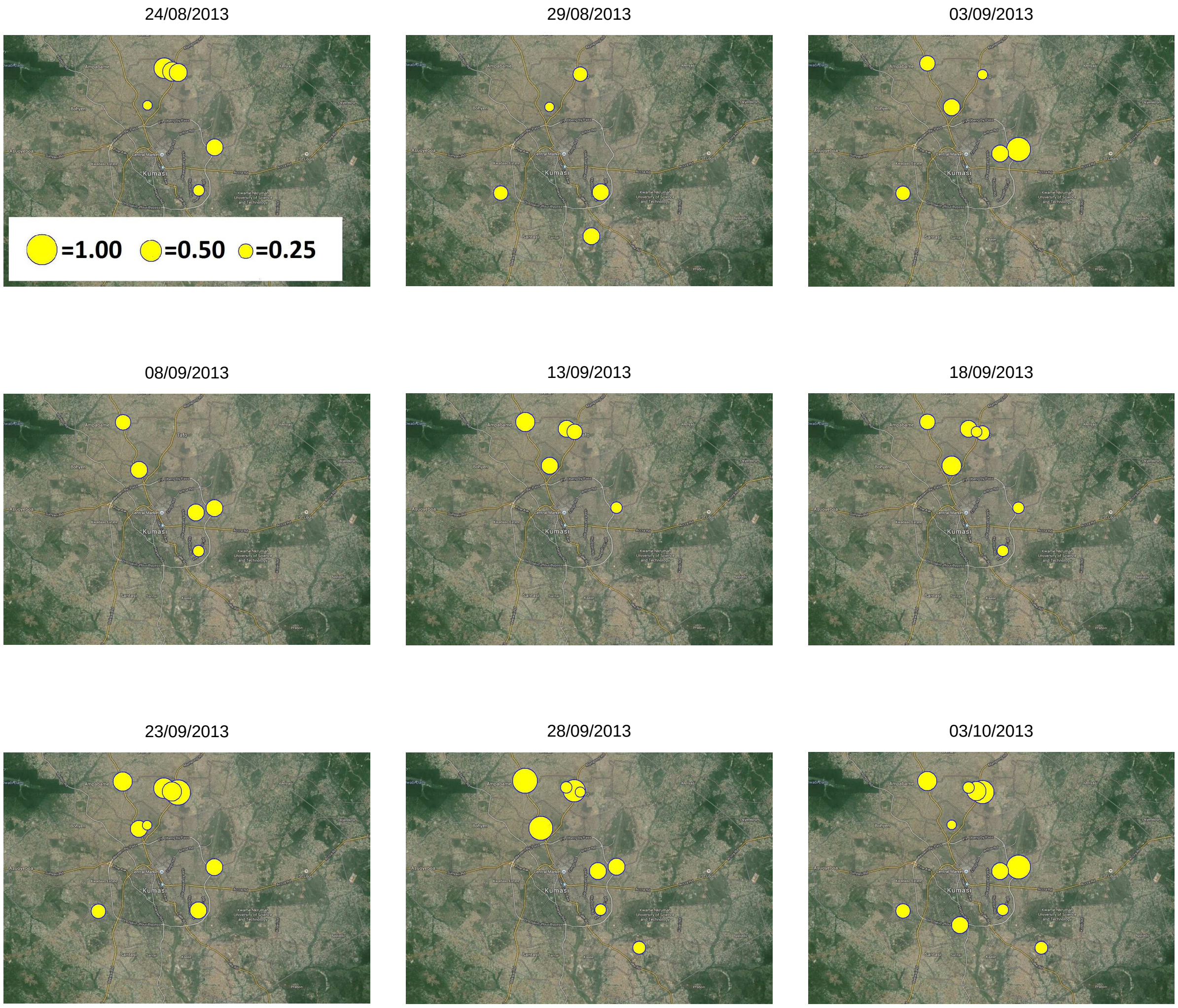
[16]Here we show nine days, each five days apart. A moving map, representing the entire length of the survey, is available.

Figure 2.5: Distribution of Power Outages across Accra



Circles show the proportion of surveyed households within an Enumeration Area affected by power cuts on that day. Key shown in 24/08/2013 map. Base maps taken from *Google Maps*.

Figure 2.6: Distribution of Power Outages across Kumasi



Circles show the proportion of surveyed households within an Enumeration Area affected by power cuts on that day. Key shown in 24/08/2013 map. Base maps taken from *Google Maps*.

2.5 Empirical Strategy

2.5.1 Regression Equations

To assess the impact that changes in domestic productivity have on time allocation, we use weekly variation in the extent of power outages. We regress the log of hours spent doing domestic work ($\ln(Domestic)_{iht}$) and the log of hours spent doing market work ($\ln(Market)_{iht}$) on the number of hours of power outage the household reported for that week. As we explain below, we restrict our regressions to datapoints where positive hours of both domestic work and market work were undertaken. In particular, for individual i in household h at week t we estimate Equations (2.10) and (2.11).[17]

$$\ln(Domestic)_{iht} = \beta_0^D + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{\beta_1^D} + \boldsymbol{(Ind)_{ih}} \cdot \boldsymbol{\beta_2^D} + \gamma_t + \mu_h + \epsilon_{iht}^D \tag{2.10}$$

$$\ln(Market)_{iht} = \beta_0^M + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{\beta_1^M} + \boldsymbol{(Ind)_{ih}} \cdot \boldsymbol{\beta_2^M} + \gamma_t + \mu_h + \epsilon_{iht}^M \tag{2.11}$$

$\boldsymbol{(Out)_{ht}}$ is a vector of hours of outage variables. We include polynomial terms to estimate the relationship with some flexibility.

$\boldsymbol{(Ind)_{ih}}$ is a vector of time-invariant individual controls, including sex, marital status, age, and education.

γ_t and μ_h are week- and household-specific effects respectively.

We include week fixed-effects in our preferred specifications to control for common shocks, which simultaneously affect all individuals in the survey. Time-invariant differences between households are absorbed by adding household fixed-effects. Our results will therefore be robust to factors like household composition and wealth insofar as they are stable throughout the 3 months of the high-frequency survey. In addition, the inclusion of household fixed-effects also prevents any time-invariant differences between EAs and treatment groups from confounding our analysis.

[17] We use the subscripts for activity $k = D, M$ to distinguish parameters associated with domestic and market work respectively.

We elect not to estimate these equations using the Seemingly Unrelated Regression (SUR) method. Since the set of regressors in Equations (2.10) and (2.11) is the same, the point estimates from SUR would be identical to those we report in Section 2.6. However, by estimating Equations (2.10) and (2.11) separately we can more easily adjust our standard errors for clustering and the presence of fixed-effects.

The error term, ϵ_{iht}^k, is decomposed in Equation (2.12).

$$\epsilon_{iht}^k = \phi_{ht}^k + \pi_{ih}^k + \upsilon_{iht}^k \tag{2.12}$$

Firstly, there are household-level time-varying factors ϕ_{ht}^k, which may confound the analysis. Secondly, although we include some important individual controls, there is still time-invariant individual variation, π_{ih}^k, in the estimating equation. Finally, there may be time-variant individual factors, υ_{iht}^k, although since we have harmonised our power outage variables to the level of the household, the extent to which these factors bias the estimation of our parameters of interest in $\boldsymbol{\beta_1^k}$ should be limited.

2.5.2 Intensive and Extensive Margins

We focus our main analysis on the intensive margin. This means we restrict our equations to datapoints (individual-weeks) where *both* domestic work *and* market work were undertaken. We demonstrate in Appendix 2.F that, by adding individuals controls to our main equation, the inherent sample selection in this approach occurs on the basis of observables rather than unobservables, eliminating potential biases. The inclusion of these controls also improves the efficiency of our estimates.

We estimate our equations with a logarithmically transformed outcome variable, because this substantially improves the fit of our models and the efficiency of our estimates. In particular, the R^2 for our main domestic work equation improves from 0.2742 to 0.3874 by logarithmically transforming the outcome variable, whilst for the market work equation the R^2 jumps from 0.0478 to 0.0512. This is intuitive, given the positive skew we observed for the outcome variables in Section 2.4, especially domestic work.

Although we estimate our main relationship in logs, we can exponentially transform the results for a *given* worker to calculate the absolute time allocated to domestic work and market work, according to our regression results. In-keeping with our conceptual framework, we can then predict the total time allocated to leisure under the assumption that t_m, t_d, and t_l must add up to the total time available each week.

2.5.3 Adjustment

Thus far, we have implicitly assumed that individuals are able to observe domestic productivity, ρ, (proxied by power outages) and adjust their time-use accordingly within that week. We now consider the plausibility of this assumption.

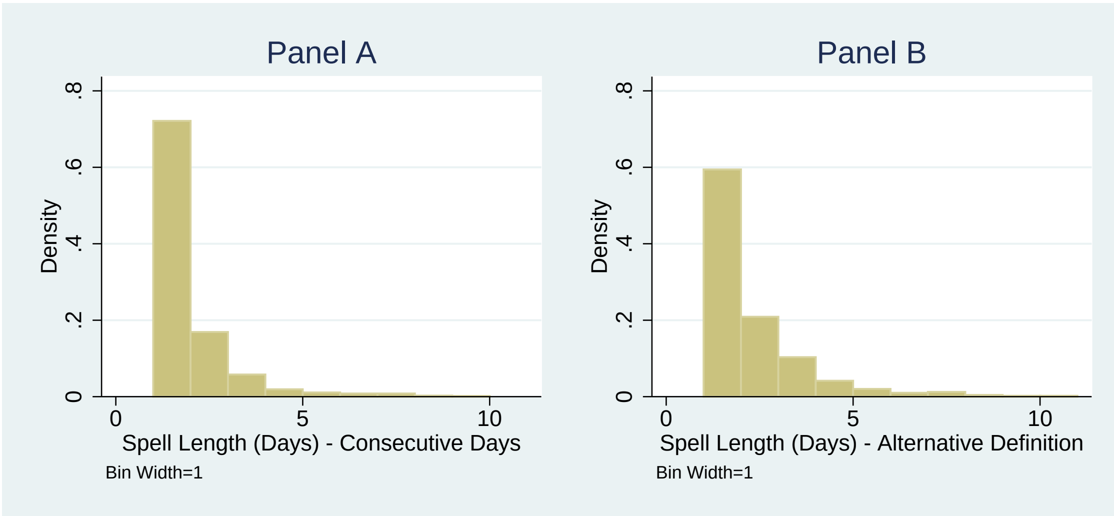
Firstly, it is helpful to draw a conceptual distinction between instantaneous 'knee-jerk' reactions, which individuals make in direct response to power outages, and adjustments to their *planned* time-use, which may operate over a longer time-horizon. Our analysis focusses on the second of these two concepts. 'Knee-jerk' responses, such as suddenly being forced to abandon cooking on an electric stove, tell us little about the individuals' willingness to substitute between different activities. However, if individuals believe a particular period is characterised by better or worse domestic productivity, ρ, and subsequently re-optimise and change their plans accordingly, then their resulting time allocation is highly informative about their preferences and the nature of their maximisation problem.

The time-horizon of our data is well-suited to focussing on this notion of adjustment. Over the period of a week, individuals have sufficient time to make observations of domestic productivity, ρ, and change their planned time allocation *for that week* in response.

By transforming our data, we can investigate the time-series properties of power outages, which further support the interpretation described above. Since respondents report the particular days each week that were affected by power outages, we are able to reshape our dataset into daily data with a small number of basic variables.

Firstly, the spells of time affected by power outages appear to be relatively short. In Figure 2.7, we plot the length of outage spells in our sample, defined in two ways. In Panel A, a 'spell' corresponds to consecutive days affected by power outage. In Panel B, we relax the definition somewhat, allowing 'spells' to contain gaps of up to 48 hours without outage before ending. Using both definitions, the vast majority of spells are less than five days long.

Figure 2.7: Power Outage Spell Length



Secondly, the dummy variable for whether or not a particular household-day was affected by power outages shows far greater persistence over the first few days than over a week or more. We begin by testing whether this variable is auto-correlated AR(1) using the Wooldridge test for serial correlation in panel data (Wooldridge, 2002; Drukker, 2003).[18] The null hypothesis of no auto-correlation is strongly rejected with a p-value of 0.0090, so there is some persistence across days in the incidence of power outages.

We also regress the dummy variable for whether or not a particular household-day was affected by power outages on its lags, and report the results in Table 2.10. We estimate these models using pooled OLS, in which we expect the coefficients on the lags of the dependent variable to be biased up, and fixed-effects regressions, in which we expect the coefficients to be biased downwards. We also use the Arellano-Bond System-GMM estimator for these equations, but given the Sargan statistics reported at the bottom of Table 2.10, it is clear these results should be interpreted with some caution.[19] In spite of the limitations of these regressions, it appears that outages are more persistent *within* weeks than *across* weeks.

Importantly, the strong persistence of outages within the period of a few days suggests that individuals have time to react and adjust their planned time-use within the same week. However, since spells are short and the lags become less significant after approximately the seventh day, focussing on the adjustment to outages within the same week is a suitable approach.[20]

[18] We implement this test with Stata's `xtserial` command.

[19] Rejecting the null hypothesis for the Sargan test of over-identifying restrictions, as we do in all three models shown, questions the validity of using extra lags of the dependent variable as instruments.

[20] We consider the implications of relaxing this assumption in Section 2.7, by including lags and leads of our

Table 2.10: Lag Structure of Power Outages at the Household-Day Level

	(1) POLS	(2) FE	(3) GMM	(4) POLS	(5) FE	(6) GMM	(7) POLS	(8) FE	(9) GMM
L.Outage at HH	0.1974***	0.1169***	0.1208***	0.1578***	0.1033***	0.1445***	0.1569***	0.1053***	0.1426***
	(0.0174)	(0.0138)	(0.0086)	(0.0150)	(0.0138)	(0.0086)	(0.0138)	(0.0136)	(0.0086)
L2.Outage at HH	0.1527***	0.0772***	0.0790***	0.1198***	0.0699***	0.1073***	0.1130***	0.0689***	0.1004***
	(0.0144)	(0.0127)	(0.0086)	(0.0101)	(0.0126)	(0.0086)	(0.0102)	(0.0127)	(0.0086)
L3.Outage at HH	0.1125***	0.0349**	0.0340***	0.0670***	0.0222	0.0553***	0.0644***	0.0250*	0.0526***
	(0.0153)	(0.0144)	(0.0087)	(0.0139)	(0.0141)	(0.0086)	(0.0142)	(0.0143)	(0.0087)
L4.Outage at HH				0.0460***	0.0032	0.0347***	0.0457***	0.0044	0.0333***
				(0.0112)	(0.0087)	(0.0086)	(0.0108)	(0.0091)	(0.0087)
L5.Outage at HH				0.0584***	0.0157	0.0466***	0.0578***	0.0156	0.0447***
				(0.0120)	(0.0125)	(0.0086)	(0.0118)	(0.0131)	(0.0087)
L6.Outage at HH				0.0787***	0.0345***	0.0660***	0.0765***	0.0370***	0.0638***
				(0.0138)	(0.0122)	(0.0087)	(0.0135)	(0.0120)	(0.0086)
L7.Outage at HH				0.1100***	0.0628***	0.0959***	0.1000***	0.0639***	0.0880***
				(0.0110)	(0.0112)	(0.0087)	(0.0112)	(0.0111)	(0.0086)
L8.Outage at HH							0.0247**	-0.0091	0.0132
							(0.0107)	(0.0108)	(0.0086)
L9.Outage at HH							0.0180**	-0.0191*	0.0048
							(0.0081)	(0.0099)	(0.0087)
L10.Outage at HH							0.0132*	-0.0294***	-0.0028
							(0.0073)	(0.0091)	(0.0088)
N	16994	16994	16994	15426	15426	15426	14419	14419	14419
R^2	0.110	0.175		0.140	0.183		0.144	0.187	
Arellano-Bond AR(1) p-value			0.0000			0.0000			0.0000
Arelano-Bond AR(2) p-value			0.5921			0.5990			0.8251
Sargan p-value			0.0000			0.0000			0.0000

Standard errors in parentheses

Standard errors clustered two-way by household and day in the POLS and FE models

Up to three lags are used as instruments in the System-GMM estimates

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6 Results

2.6.1 Main Results

We begin by estimating Equation (2.10) for domestic work in Table 2.11 and Equation (2.11) for market work in Table 2.12, using both a linear and quadratic specification. We include household and week fixed-effects, as discussed in Section 2.5, as well as a vector of individual characteristics. Initially, we use two way clustered standard errors, by individual and week, and then consider the implications of clustering standard errors differently in our robustness checks (Cameron et al., 2011). Vitally, we split our sample on the basis of occupational status for each model, showing first the results for the full sample, and then for the self-employed and wage-employed separately.

Table 2.11: Main Results — Domestic Work

	(1) Full Sample	(2) SE	(3) WE	(4) Full Sample	(5) SE	(6) WE
Hours of Outage/Week	-0.0050	-0.0065	-0.0004	0.0093	0.0201**	0.0091
	(0.0035)	(0.0040)	(0.0043)	(0.0063)	(0.0085)	(0.0112)
Hours of Outage/Week Squared (x 100)				-0.0745**	-0.1396***	-0.0481
				(0.0354)	(0.0482)	(0.0542)
Male? (1=Y, 0=N)	-0.8799***	-1.0744***	-0.5427***	-0.8820***	-1.0780***	-0.5473***
	(0.1161)	(0.2777)	(0.1047)	(0.1158)	(0.2786)	(0.1052)
Married? (1=Y, 0=N)	0.6623***	0.6771***	0.9031**	0.6607***	0.6830***	0.9014**
	(0.1317)	(0.1643)	(0.3656)	(0.1314)	(0.1636)	(0.3674)
Male x Married	-0.6042***	-0.7443**	-0.6010*	-0.6017***	-0.7361**	-0.5978*
	(0.1524)	(0.3109)	(0.3249)	(0.1515)	(0.3107)	(0.3265)
Years of Education	0.0225	0.0594	0.0548	0.0237	0.0623	0.0561
	(0.0397)	(0.0659)	(0.0562)	(0.0396)	(0.0652)	(0.0558)
Years of Education Squared	-0.0005	-0.0030	-0.0032	-0.0005	-0.0031	-0.0033
	(0.0022)	(0.0043)	(0.0027)	(0.0022)	(0.0043)	(0.0027)
Age (Years)	0.1134***	0.1134***	0.1900***	0.1141***	0.1138***	0.1897***
	(0.0301)	(0.0421)	(0.0319)	(0.0302)	(0.0420)	(0.0319)
Age (Years) Squared	-0.0014***	-0.0016***	-0.0026***	-0.0014***	-0.0016***	-0.0026***
	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0005)	(0.0004)
Observations	2556	1363	996	2556	1363	996
R^2	0.317	0.383	0.159	0.318	0.387	0.159

Standard errors in parentheses
Dependent variable: Hours of Domestic Work/Week.
Standard errors clustered two way by individual and week. Columns (1) to (6) include week and household fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In Table 2.11, there is a positive quadratic relationship between hours of outage and the time allocated to domestic work in a given week for the self-employed only. At the 25$^{\text{th}}$ percentile

outage variables to examine any dynamic adjustment to changes in domestic productivity.

of the hours of outage distribution, an extra hour of outage *increases* domestic work for that week by 2.01 percent. At the 75$^{\text{th}}$ percentile, however, an extra hour of outage *reduces* domestic work by 1.34 percent. As anticipated, the relationship appears to be far less significant, both economically and statistically, for the wage-employed.

Looking at the controls, even amongst this sub-sample of observations where both domestic and market work are being undertaken, women do more than double the hours of domestic work compared to men. Intuitively, marriage also increases the amount of domestic work undertaken by around two-thirds. Whilst there are no clear relationships between domestic work and years of education, age appears to have a positive but diminishing effect on domestic work hours, especially for the wage-employed.

Table 2.12: Main Results — Market Work

	(1) Full Sample	(2) SE	(3) WE	(4) Full Sample	(5) SE	(6) WE
Hours of Outage/Week	0.0030*	0.0048**	0.0012	-0.0074**	-0.0125**	0.0046
	(0.0016)	(0.0019)	(0.0028)	(0.0032)	(0.0053)	(0.0075)
Hours of Outage/Week Squared (x 100)				0.0544***	0.0913***	-0.0172
				(0.0165)	(0.0275)	(0.0291)
Male? (1=Y, 0=N)	-0.0603	-0.0805	-0.1701***	-0.0588	-0.0782	-0.1718***
	(0.0655)	(0.1165)	(0.0496)	(0.0655)	(0.1151)	(0.0492)
Married? (1=Y, 0=N)	-0.1743*	-0.3334***	0.1767	-0.1731*	-0.3372***	0.1761
	(0.0933)	(0.1138)	(0.1307)	(0.0928)	(0.1126)	(0.1303)
Male x Married	0.0702	0.0813	0.1352	0.0685	0.0759	0.1364
	(0.0939)	(0.1404)	(0.2100)	(0.0934)	(0.1393)	(0.2096)
Years of Education	-0.0029	-0.0528	0.0889**	-0.0038	-0.0547	0.0894**
	(0.0219)	(0.0539)	(0.0380)	(0.0218)	(0.0539)	(0.0379)
Years of Education Squared	0.0016	0.0042	-0.0033**	0.0017	0.0043	-0.0034**
	(0.0012)	(0.0039)	(0.0016)	(0.0012)	(0.0039)	(0.0016)
Age (Years)	0.0664***	0.0682***	-0.0674**	0.0660***	0.0679***	-0.0675**
	(0.0240)	(0.0249)	(0.0268)	(0.0239)	(0.0245)	(0.0268)
Age (Years) Squared	-0.0008***	-0.0007**	0.0009**	-0.0008***	-0.0007**	0.0009**
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	2556	1363	996	2556	1363	996
R^2	0.037	0.045	0.044	0.038	0.051	0.044

Standard errors in parentheses
Dependent variable: Hours of Market Work/Week.
Standard errors clustered two way by individual and week. Columns (1) to (6) include week and household fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In Table 2.12, we observe both a positive linear and a negative quadratic relationship between hours of outage and the time spent in market work each week. We elect to focus on the quadratic specifications as this improves the goodness of fit of our models.[21] Looking again to the self-

[21] We can see in Table 2.12 that the R^2 improves for the self-employed by specifying the model quadratically rather than linearly. Moreover, we conduct a Ramsey RESET test on both the linear and the quadratic model (leaving out the fixed-effects). The null hypothesis of no omitted variables is strongly rejected for the linear model,

employed sub-sample, at the 25^{th} percentile of the hours of outage distribution, an extra hour of outage *reduces* domestic work for that week by 1.25 percent. At the 75^{th} percentile, however, an extra hour of outage *increases* domestic work by 0.94 percent. Again, however, there is no clear relationship for the wage-employed.

The coefficients on the individual controls demonstrate that wage-employed women work 17 percent more than wage-employed men, but that there is little difference in the market hours worked by the self-employed by sex. Amongst the self-employed, however, married individuals work around a third less than the unmarried. Our human capital measures show the clearest relationships in the wage-employed sub-sample. Years of education has a positive quadratic effect on market work hours, whilst age has a negative quadratic impact.

Linking these results back to our model from Section 2.3, we can calculate the implied relationship between power outages and leisure by simply subtracting the total time allocated to work ($t_d + t_m$) from the total time available in a week. Since the equations in Tables 2.11 and 2.12 are specified with logarithmically transformed hours worked, it is clear these relationships will be different for individuals who do more or less work, on average. We plot our findings for the average self-employed worker in Figure 2.8.[22]

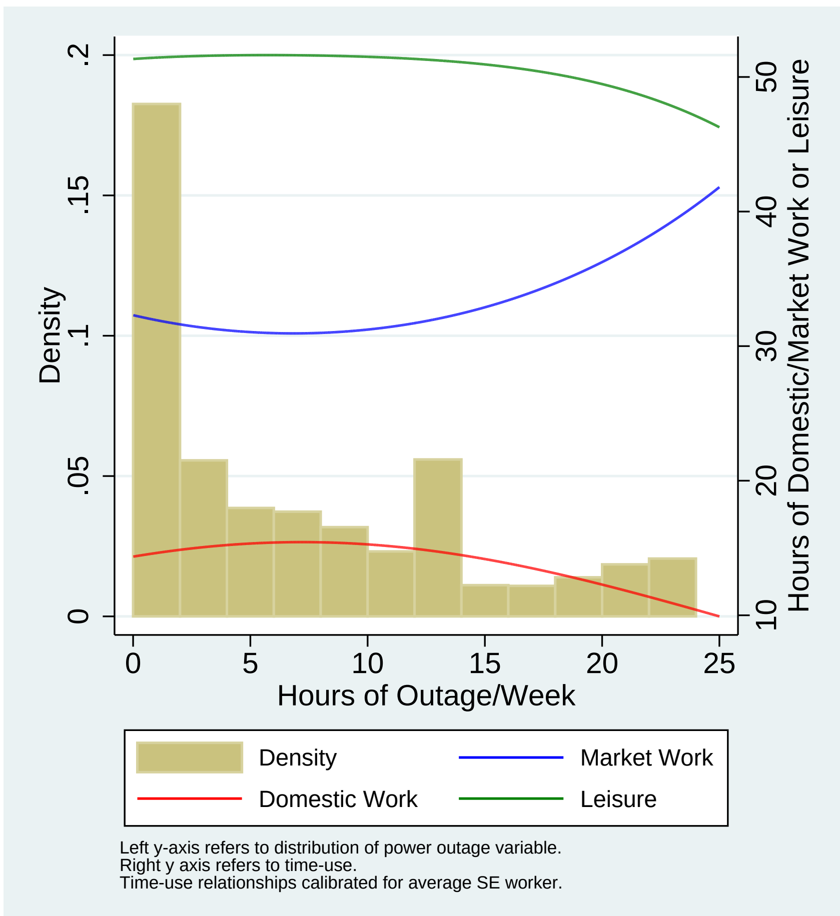
The relationship for leisure is echoed if we regress the log of *total* work, $\ln(t_d + t_m)$, on power outages. In spite of the functional form issues with this type of regression, the results show a small but statistically significant negative quadratic relationship between power outages and total work. This implies outages have a positive quadratic relationship with leisure, as we see in Figure 2.8. We display these results in Appendix 2.D. In our conceptual framework, time is simply shared between t_m, t_d, and t_l, so focussing on the domestic and market work equations provides all the information we need — this is the approach we adopt from now on.

There is, however, one key result from the total work regressions, which we believe is worth highlighting. Namely, taking domestic and market tasks together, women do significantly more hours of work each week than men. Self-employed women work 30 percent more than their male counterparts, whilst for the wage-employed sub-sample the difference drops to 16 percent. Although this is not the focus on this chapter per se, we believe this result should be borne in mind when evaluating the implications of both task allocation and occupational choice for male and female welfare.

with a p-value of 0.0016. For the quadratic model, however, the p-value is 0.5108, suggesting this is a tenable functional form.

[22] We assume that individuals have 100 hours of time to allocate each week.

Figure 2.8: Impact of Power Outages on Time Allocation



We re-estimate the domestic work and market work equations with and without individual-level controls and changing the level of the fixed-effects. These results are reported in Appendix 2.E. Overall, our estimates are robust to changing the set of regressors in this way. The coefficients generally become larger when household and week fixed-effects are removed, and we also see that removing the individual controls worsens the fit of our models, especially for domestic work.

We also try estimating our equations with extra fixed-effects. Our results appear to be robust to adding city-specific time effects. However, although the signs of our estimates remain the same, we do not have sufficient statistical power in the presence of individual fixed-effects.

2.6.2 Heterogeneity

We now consider potential sources of heterogeneity in the relationship between power outages and time allocation. Since our most coherent results are for the self-employed, we focus on this sub-sample. We re-estimate Equations (2.10) and (2.11), adding interactions between the power outage variables and sex. The results are shown for domestic work in Table 2.13 and for market work in Table 2.14.

Table 2.13: Heterogeneity by Sex — Domestic Work

	(1)	(2)	(3)	(4)	(5)
Hours of Outage/Week	-0.0065	-0.0008	0.0201**	0.0242***	0.0310***
	(0.0040)	(0.0059)	(0.0085)	(0.0087)	(0.0114)
Hours of Outage/Week Squared (x 100)			-0.1396***	-0.1332***	-0.1714**
			(0.0482)	(0.0477)	(0.0709)
Male? (1=Y, 0=N)	-1.0744***	-0.9671***	-1.0780***	-0.9774***	-0.9473***
	(0.2777)	(0.2591)	(0.2786)	(0.2623)	(0.2546)
Male x Married	-0.7443**	-0.7450**	-0.7361**	-0.7371**	-0.7186**
	(0.3109)	(0.2994)	(0.3107)	(0.2998)	(0.3004)
Outage Var. x Male Dummy		-0.0179**		-0.0167*	-0.0387
		(0.0087)		(0.0092)	(0.0243)
Outage Var. Sq. x Male Dummy					0.1183
					(0.1406)
Observations	1363	1363	1363	1363	1363
R^2	0.383	0.386	0.387	0.390	0.391

Standard errors in parentheses
Dependent variable: Log of Hours of Domestic Work/Week.
Standard errors clustered two way by individual and week.
All columns include week and individual fixed effects, and individual controls.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Whilst women adjust from market work towards domestic work more when hours of outage per week are low, they are less able to substitute away from domestic work when power outages

Table 2.14: Heterogeneity by Sex — Market Work

	(1)	(2)	(3)	(4)	(5)
Hours of Outage/Week	0.0048**	-0.0001	-0.0125**	-0.0161***	-0.0210***
	(0.0019)	(0.0026)	(0.0053)	(0.0054)	(0.0058)
Hours of Outage/Week Squared (x 100)			0.0913***	0.0857***	0.1130***
			(0.0275)	(0.0280)	(0.0331)
Male? (1=Y, 0=N)	-0.0805	-0.1728	-0.0782	-0.1661	-0.1877
	(0.1165)	(0.1325)	(0.1151)	(0.1310)	(0.1319)
Male x Married	0.0813	0.0818	0.0759	0.0768	0.0636
	(0.1404)	(0.1438)	(0.1393)	(0.1426)	(0.1422)
Outage Var. x Male Dummy		0.0154***		0.0146**	0.0303***
		(0.0058)		(0.0058)	(0.0108)
Outage Var. Sq. x Male Dummy					-0.0846
					(0.0553)
Observations	1363	1363	1363	1363	1363
R^2	0.045	0.053	0.051	0.058	0.060

Standard errors in parentheses
Dependent variable: Log of Hours of Market Work/Week.
Standard errors clustered two way by individual and week.
All columns include week and individual fixed effects, and individual controls.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

become more severe. This supports the notion that the obligation for women to undertake domestic work is stronger than for men. We take the predictions from Column (4) in Tables 2.13 and 2.14 and illustrate these relationships in Figure 2.9, showing the adjustment paths for women with the dashed line and for men with the dotted line. We suppress the disparity in the intercepts between women and men for domestic and market work so as to make the different adjustment paths clearer.[23]

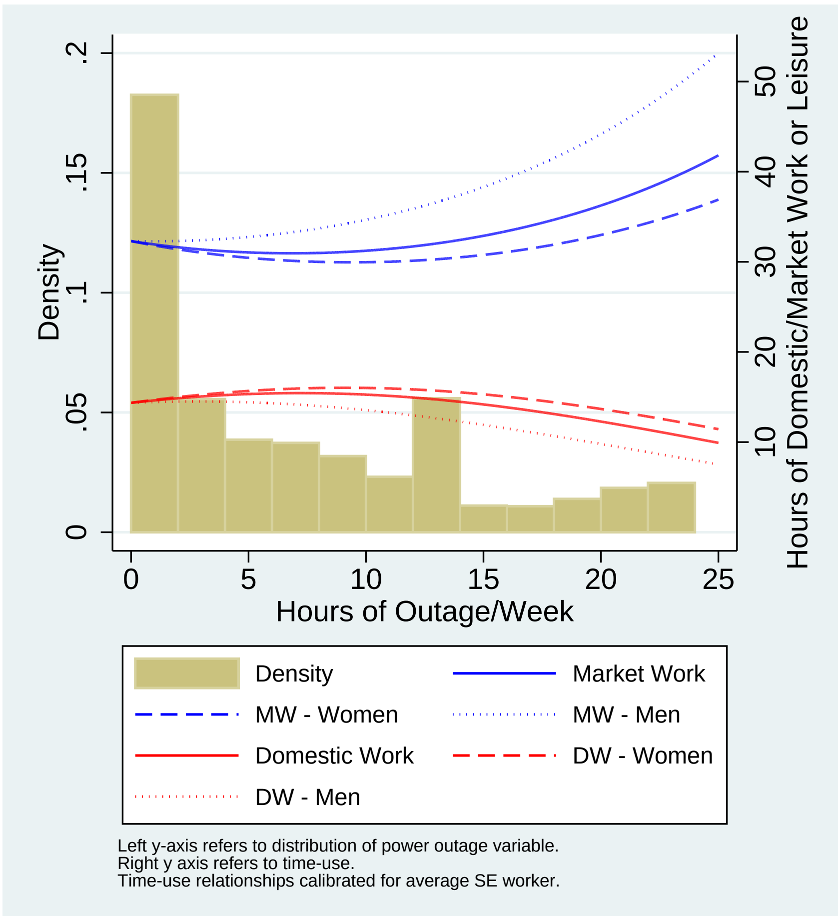
We also assess heterogeneity by marriage, age, and education, using the same approach, but find no robust patterns.

2.7 Robustness

In our robustness checks, we test two key assumptions of our analysis. Firstly, we consider whether our interpretation of hours of outage per week as a proxy for domestic productivity, ρ, is plausible. We then examine our assumption that the relationship between time-use and domestic productivity can be considered one week at a time, by allowing for dynamic effects in the impact of power outages. For completeness, we also explore different approaches to clustering

[23]In reality, however, self-employed women do significantly more domestic work, even when Hours of Outage/Week = 0, as Table 2.11 demonstrates.

Figure 2.9: Impact of Power Outages on Time Allocation with Heterogeneity by Sex



our standard errors. Since our most coherent results from Section 2.6 are for the self-employed, we focus our robustness checks on this sub-sample.

2.7.1 Interpreting Power Outages

We noted in Section 2.4 that, for some individuals, power outages at home might proxy for *market work* productivity, w, rather than just productivity in domestic work, ρ. This issue is most likely to generate bias for self-employed individuals working and living in the same location, and using electricity in their place of work. To examine these factors, we augment Equations (2.10) and (2.11) as follows:

$$\begin{aligned}\ln(Domestic)_{iht} = \; & \beta_0^D + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{\beta_1^D} + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{(Work)_{ih}} \cdot \boldsymbol{\beta_2^D} + \boldsymbol{(Work)_{ih}} \cdot \boldsymbol{\beta_3^D} + \\ & \boldsymbol{(Ind)_{ih}} \cdot \boldsymbol{\beta_4^D} + \gamma_t + \mu_h + \epsilon_{iht}^D\end{aligned} \tag{2.13}$$

$$\begin{aligned}\ln(Market)_{iht} = \; & \beta_0^M + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{\beta_1^M} + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{(Work)_{ih}} \cdot \boldsymbol{\beta_2^M} + \boldsymbol{(Work)_{ih}} \cdot \boldsymbol{\beta_3^M} + \\ & \boldsymbol{(Ind)_{ih}} \cdot \boldsymbol{\beta_4^M} + \gamma_t + \mu_h + \epsilon_{iht}^M\end{aligned} \tag{2.14}$$

$\boldsymbol{(Work)_{ih}}$ is a vector of variables pertaining to individuals' typical work practices, including dummy variables for whether or not an individual typically uses electricity for their work, and whether or not they work close to their home. Although the coefficients on these variables cannot be given a causal interpretation per se, their inclusion may remove individual-level time-invariant factors from the error term (π_{ih}^k).

We show the results for domestic work in Table 2.15 and for market work in Table 2.16. In Column (1) we report our original main results. We then add the work practice controls without interactions with the outage variable in Column (2). Finally, in Columns (3)–(6) we add the work practice variables interacted with outage variables, focussing first on electricity use, then work location, then both together.

For both domestic and market work, the addition of these extra controls leaves the signs of the coefficients on our outage variables unchanged, with the magnitudes also remaining fairly stable. The only equations in which we lose statistical significance on hours of outage/week at

Table 2.15: Distinguishing ρ from w with Controls — Domestic Work

	(1)	(2)	(3)	(4)	(5)	(6)
Hours of Outage/Week	0.0201**	0.0193**	0.0226***	0.0197**	0.0225***	0.0210**
	(0.0085)	(0.0086)	(0.0075)	(0.0089)	(0.0077)	(0.0087)
Hours of Outage/Week Squared (x 100)	-0.1396***	-0.1382***	-0.1313***	-0.1413***	-0.1332**	-0.1315**
	(0.0482)	(0.0498)	(0.0490)	(0.0511)	(0.0523)	(0.0528)
Use Elec. at Work? (1=Y, 0=N)		0.0708	0.0788		0.0415	0.2105
		(0.1908)	(0.1418)		(0.1862)	(0.2128)
Work/Live Same EA? (1=Y, 0=N)		0.1336		0.0327	0.0705	0.1135
		(0.2088)		(0.1525)	(0.2005)	(0.2270)
Work Elec. x Work Location		-0.2483				-0.2550
		(0.3114)				(0.3155)
Outage Var. x Work Elec.			-0.0143*		-0.0139	-0.0141
			(0.0086)		(0.0085)	(0.0176)
Outage Var. x Work Location				0.0009	0.0004	0.0014
				(0.0083)	(0.0088)	(0.0099)
Outage Var. x Work Elc. x Work Loc.						0.0002
						(0.0203)
Observations	1363	1363	1363	1363	1363	1363
R^2	0.387	0.389	0.389	0.387	0.389	0.390

Standard errors in parentheses

Dependent variable: Log of Hours of Domestic Work/Week.

Standard errors clustered two way by individual and week. All columns include week and household fixed effects, and individual controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.16: Distinguishing ρ from w with Controls — Market Work

	(1)	(2)	(3)	(4)	(5)	(6)
Hours of Outage/Week	-0.0125**	-0.0108**	-0.0128**	-0.0084	-0.0088	-0.0053
	(0.0053)	(0.0053)	(0.0053)	(0.0052)	(0.0054)	(0.0059)
Hours of Outage/Week Squared (x 100)	0.0913***	0.0857***	0.0925***	0.1041***	0.1045***	0.1025***
	(0.0275)	(0.0287)	(0.0286)	(0.0302)	(0.0310)	(0.0321)
Use Elec. at Work? (1=Y, 0=N)		-0.2176	0.1193		0.0820	-0.2145
		(0.1791)	(0.1528)		(0.1674)	(0.1938)
Work/Live Same EA? (1=Y, 0=N)		-0.0484		0.1742	0.1335	0.0592
		(0.1447)		(0.1059)	(0.1242)	(0.1417)
Work Elec. x Work Location		0.4660*				0.4474*
		(0.2522)				(0.2601)
Outage Var. x Work Elec.			-0.0011		-0.0007	-0.0039
			(0.0051)		(0.0047)	(0.0072)
Outage Var. x Work Location				-0.0108*	-0.0102*	-0.0135**
				(0.0058)	(0.0060)	(0.0068)
Outage Var. x Work Elc. x Work Loc.						0.0051
						(0.0100)
Observations	1363	1363	1363	1363	1363	1363
R^2	0.051	0.065	0.054	0.056	0.057	0.069

Standard errors in parentheses

Dependent variable: Log of Hours of Market Work/Week.

Standard errors clustered two way by individual and week. All columns include week and household fixed effects, and individual controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the 5 percent level are on the linear terms in Columns (4) to (6) of Table 2.16.

Outages seem to have a different effect on market work, depending on the respondents' typical work location. However, since the three way interaction in Column (6) is statistically insignificant in both Tables 2.15 and 2.16, it does not appear that those typically working close to their home *and* using electricity react to outages differently. This supports our initial interpretation of power outages as a better proxy for ρ than w.

An alternative strategy for testing our interpretation of the power outage variable is to control for week-on-week variation in w directly. Adding individual weekly earnings into the regressions would lead to endogeneity, so instead we adapt the approach taken by Camerer et al. (1997). They instrument for daily earnings of New York City taxi-drivers using the summary statistics for the earnings all *other* drivers working on the same shift that day. These summary statistics provide a proxy for market conditions orthogonal to idiosyncratic unobservable factors determining each individuals' own earnings.

Rather than using the same instrumental variable technique as such, we instead add the analogous summary statistics directly into our regressions to absorb variation in w. In particular we use summary statistics for the earnings of all other individuals working in the same EA in a particular week, including the mean, the 25^{th} percentile, the median, and the 75^{th} percentile.[24] We include statistics from both the present period, and one period before to allow for the possibility that individuals adjust their time-use in response to changing market conditions with a lag. Vitally, we assume that these summary statistics capture variation in earnings orthogonally to variation in our hours of outage variable. Including these summary statistics in the vector $\boldsymbol{(Earnings)_{iht}}$ we estimate:

$$\ln(Domestic)_{iht} = \beta_0^D + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{\beta_1^D} + \boldsymbol{(Earnings)_{iht}} \cdot \boldsymbol{\beta_2^D} + \boldsymbol{(Ind)_{ih}} \cdot \boldsymbol{\beta_3^D} + \gamma_t + \mu_h + \epsilon_{iht}^D \tag{2.15}$$

$$\ln(Market)_{iht} = \beta_0^M + \boldsymbol{(Out)_{ht}} \cdot \boldsymbol{\beta_1^M} + \boldsymbol{(Earnings)_{iht}} \cdot \boldsymbol{\beta_2^M} + \boldsymbol{(Ind)_{ih}} \cdot \boldsymbol{\beta_3^M} + \gamma_t + \mu_h + \epsilon_{iht}^M \tag{2.16}$$

We estimate Equation (2.15) for domestic work in Table 2.17 and Equation (2.16) for market work in 2.18. In Column (1) we present our original results. Columns (2)–(4) add summary

[24] In Appendix 2.G, we show descriptive statistics for the earnings variable, as well as a set of pseudo-first-stage regressions of individual earnings on the summary statistics for the rest of the EA.

statistics from the current week, whilst Columns (5)–(7) add the same statistics lagged one period. Again, the signs on the key coefficients for the outage variables remain the same for both domestic work and market work. Also, the coefficients remain statistically significant, except for the linear outage term in the market work equation, when the current period quartiles are included. These results, therefore, further support our intuition that power outages proxy for ρ rather than w.

Table 2.17: Controlling Directly for w — Domestic Work

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hours of Outage/Week	0.0201**	0.0140*	0.0183**	0.0182**	0.0172*	0.0227**	0.0228**
	(0.0085)	(0.0078)	(0.0087)	(0.0088)	(0.0100)	(0.0107)	(0.0106)
Hours of Outage/Week Squared (x 100)	-0.1396***	-0.1051**	-0.1256**	-0.1251**	-0.1283**	-0.1610**	-0.1612***
	(0.0482)	(0.0427)	(0.0548)	(0.0552)	(0.0537)	(0.0627)	(0.0625)
Log of Mean Earnings (others in EA)		-0.0328		0.0161			
		(0.0453)		(0.1214)			
Log of 25th Percentile Earnings (others in EA)			0.0232	0.0209			
			(0.0622)	(0.0730)			
Log of Median Percentile Earnings (others in EA)			0.1109*	0.1089**			
			(0.0612)	(0.0538)			
Log of 75th Percentile Earnings (others in EA)			-0.1682**	-0.1778*			
			(0.0685)	(0.1026)			
Lagged Log of Mean Earnings (others in EA)					-0.0432		-0.0153
					(0.0538)		(0.0587)
Lagged Log of 25th Percentile Earnings (others in EA)						0.0065	0.0088
						(0.0529)	(0.0518)
Lagged Log of Median Percentile Earnings (others in EA)						0.0772	0.0787
						(0.0896)	(0.0921)
Lagged Log of 75th Percentile Earnings (others in EA)						-0.1307*	-0.1213*
						(0.0775)	(0.0636)
Observations	1363	1250	1093	1093	1155	1004	1004
R^2	0.387	0.359	0.368	0.368	0.379	0.397	0.397

Standard errors in parentheses

Dependent variable: Log of Hours of Domestic Work/Week.

Standard errors clustered two way by individual and week. All columns include week and household fixed effects, and individual controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.18: Controlling Directly for w — Market Work

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hours of Outage/Week	-0.0125**	-0.0101*	-0.0099*	-0.0096	-0.0181***	-0.0199***	-0.0202***
	(0.0053)	(0.0059)	(0.0059)	(0.0060)	(0.0055)	(0.0060)	(0.0058)
Hours of Outage/Week Squared (x 100)	0.0913***	0.0788***	0.0786**	0.0774**	0.1072***	0.1142***	0.1174***
	(0.0275)	(0.0282)	(0.0317)	(0.0320)	(0.0295)	(0.0322)	(0.0307)
Log of Mean Earnings (others in EA)		0.0214		-0.0381			
		(0.0284)		(0.0531)			
Log of 25th Percentile Earnings (others in EA)			-0.0097	-0.0043			
			(0.0226)	(0.0235)			
Log of Median Percentile Earnings (others in EA)			0.0126	0.0175			
			(0.0484)	(0.0439)			
Log of 75th Percentile Earnings (others in EA)			0.0351	0.0577			
			(0.0480)	(0.0576)			
Lagged Log of Mean Earnings (others in EA)					0.0350		0.1955**
					(0.0330)		(0.0765)
Lagged Log of 25th Percentile Earnings (others in EA)						-0.0408	-0.0707***
						(0.0252)	(0.0271)
Lagged Log of Median Percentile Earnings (others in EA)						0.0451	0.0257
						(0.0320)	(0.0332)
Lagged Log of 75th Percentile Earnings (others in EA)						0.0304	-0.0901
						(0.0543)	(0.0573)
Observations	1363	1250	1093	1093	1155	1004	1004
R^2	0.051	0.054	0.052	0.052	0.053	0.055	0.065

Standard errors in parentheses

Dependent variable: Log of Hours of Market Work/Week.

Standard errors clustered two way by individual and week. All columns include week and household fixed effects, and individual controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As a final check on our interpretation of power outages, we consider the implications of restricting the sample on the basis of individuals' domestic work practices. In particular, we drop any individuals claiming not to typically use electricity for household chores or childcare, and then re-estimate Equations (2.10) and (2.11). We demonstrate these results for both market and domestic work in Table 2.19. In Columns (1) and (2) we show the domestic work equation first with the complete sample of self-employed individuals, then removing those who do not typically use electricity for household chores or childcare. The analogous results for the market work equation are shown in Columns (3) and (4).

Table 2.19: Restricting the Sample for Domestic Work Practices

	Domestic Work		Market Work	
	(1) Original	(2) Restricted	(3) Original	(4) Restricted
Hours of Outage/Week	0.0201**	0.0181*	-0.0125**	-0.0141**
	(0.0085)	(0.0098)	(0.0053)	(0.0058)
Hours of Outage/Week Squared (x 100)	-0.1396***	-0.1300**	0.0913***	0.0994***
	(0.0482)	(0.0536)	(0.0275)	(0.0279)
N	1363	1248	1363	1248
R^2	0.387	0.377	0.051	0.042
Test Equal Coefs.		0.8746		0.5025

Standard errors in parentheses
Dependent variable: Log of Hours of Domestic or Market Work/Week.
Standard errors clustered two way by individual and week.
All columns include week and household fixed effects, and individual controls.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficients on the outage variables in both the domestic and market work equations remain the same sign and around the same magnitude when restricting the sample in this way. In the footer of Table 2.19 we report p-values for cross-equation Wald Tests for the coefficients on 'Hours of Outage/Week' and 'Hours of Outage/Week Squared', and we are unable to reject the null hypotheses that the coefficients are unchanged, even at the 10 percent level.

2.7.2 Dynamic Effects

To test for the presence of dynamic effects in individuals' adjustment to power outages, we include up to two lags and leads of the outage variables in our regressions. In doing this, we are examining our initial assumption that decisions about time-use are made one week at a time. If lagged power outages affect current time-use, this suggests that individuals' responses may be delayed. Conversely, the leads of power outages may inform us about potential anticipation effects.

We therefore modify Equations (2.10) and (2.11) for:

$$
\begin{aligned}
\ln(Domestic)_{iht} =& \beta_0^D + \boldsymbol{(Out)_{ht} \cdot \beta_1^D} + \boldsymbol{(Out)_{ht-1} \cdot \beta_2^D} + \boldsymbol{(Out)_{ht-2} \cdot \beta_3^D} + \\
& \boldsymbol{(Out)_{ht+1} \cdot \beta_4^D} + \boldsymbol{(Out)_{ht+2} \cdot \beta_5^D} + \boldsymbol{(Ind)_{ih} \cdot \beta_6^D} + \gamma_t + \mu_h + \epsilon_{iht}^D
\end{aligned} \tag{2.17}
$$

$$
\begin{aligned}
\ln(Market)_{iht} =& \beta_0^M + \boldsymbol{(Out)_{ht} \cdot \beta_1^M} + \boldsymbol{(Out)_{ht-1} \cdot \beta_2^M} + \boldsymbol{(Out)_{ht-2} \cdot \beta_3^M} + \\
& \boldsymbol{(Out)_{ht+1} \cdot \beta_4^M} + \boldsymbol{(Out)_{ht+2} \cdot \beta_5^M} + \boldsymbol{(Ind)_{ih} \cdot \beta_6^M} + \gamma_t + \mu_h + \epsilon_{iht}^M
\end{aligned} \tag{2.18}
$$

We show the results for domestic work in Table 2.20 and for market work in Table 2.21. In Column (1) we show our original results. In Columns (2) and (3) we add one lag and one lead of the outage variables, first linearly, then quadratically. Since adding lags and leads necessarily reduces the number of datapoints, we re-estimate Equations (2.10) and (2.11), excluding lags and leads, but on the same restricted sample in Column (4). We repeat this process, adding up to two lags and leads in Columns (5)–(7).

Table 2.20: Dynamic Effects — Domestic Work

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hours of Outage/Week	0.0201**	0.0220**	0.0182*	0.0209**	0.0325***	0.0371***	0.0316***
	(0.0085)	(0.0109)	(0.0105)	(0.0103)	(0.0113)	(0.0126)	(0.0110)
Hours of Outage/Week Squared (x 100)	-0.1396***	-0.1260**	-0.1045**	-0.1348***	-0.2015***	-0.2272***	-0.2068***
	(0.0482)	(0.0506)	(0.0479)	(0.0499)	(0.0614)	(0.0672)	(0.0643)
Lag. Outage		-0.0094***	0.0113*		-0.0090***	0.0114	
		(0.0021)	(0.0059)		(0.0021)	(0.0109)	
Lag. Outage Sq. (x100)			-0.1137***			-0.1137**	
			(0.0208)			(0.0501)	
Lag. 2 Outage					-0.0038	0.0269*	
					(0.0036)	(0.0141)	
Lag.2 Outage Sq. (x100)						-0.1751***	
						(0.0662)	
Lead. Outage		-0.0066	0.0126		-0.0037	0.0154*	
		(0.0040)	(0.0099)		(0.0037)	(0.0088)	
Lead. Outage Sq. (x 100)			-0.1019*			-0.1066**	
			(0.0600)			(0.0496)	
Lead. 2 Outage					-0.0043	0.0034	
					(0.0048)	(0.0144)	
Lead. 2 Outage Sq. (x100)						-0.0302	
						(0.0577)	
Observations	1363	1003	1003	1003	741	741	741
R^2	0.387	0.411	0.415	0.407	0.436	0.444	0.431

Standard errors in parentheses
Dependent variable: Log of Hours of Market Work/Week.
Standard errors clustered two way by individual and week.
All columns include week and household fixed effects, and individual controls.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.21: Dynamic Effects — Market Work

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hours of Outage/Week	-0.0125**	-0.0222***	-0.0189**	-0.0219***	-0.0269***	-0.0296***	-0.0274***
	(0.0053)	(0.0085)	(0.0077)	(0.0083)	(0.0084)	(0.0096)	(0.0082)
Hours of Outage/Week Squared (x 100)	0.0913***	0.1250***	0.1065***	0.1311***	0.1651***	0.1783***	0.1677***
	(0.0275)	(0.0400)	(0.0359)	(0.0413)	(0.0447)	(0.0500)	(0.0489)
Lag. Outage		-0.0001	-0.0184***		-0.0039	-0.0239***	
		(0.0024)	(0.0053)		(0.0024)	(0.0083)	
Lag. Outage Sq. (x100)			0.1003***			0.1147***	
			(0.0153)			(0.0390)	
Lag. 2 Outage					0.0069***	-0.0063	
					(0.0026)	(0.0055)	
Lag.2 Outage Sq. (x100)						0.0722***	
						(0.0226)	
Lead. Outage		0.0069**	-0.0093*		0.0037	-0.0076	
		(0.0030)	(0.0055)		(0.0031)	(0.0062)	
Lead. Outage Sq. (x 100)			0.0860**			0.0658*	
			(0.0377)			(0.0392)	
Lead. 2 Outage					0.0078***	-0.0075	
					(0.0029)	(0.0095)	
Lead. 2 Outage Sq. (x100)						0.0780*	
						(0.0434)	
Observations	1363	1003	1003	1003	741	741	741
R^2	0.051	0.059	0.070	0.054	0.084	0.099	0.069

Standard errors in parentheses

Dependent variable: Log of Hours of Market Work/Week.

Standard errors clustered two way by individual and week.

All columns include week and household fixed effects, and individual controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The strongest effects on time-use continue to arise from the *current* period's outage variables. Indeed, the sign of the linear and quadratic outage coefficients remain the same when adding lags and leads and, if anything, they become more statistically significant. We cannot, however, rule out the possibility of delayed responses and anticipation effects. For both domestic and market work, these effects generally have the same sign as the current period effects, but are not as large and not always statistically significant.

As such, by omitting dynamic effects, our conceptual framework is certainly a simplification of individuals' true patterns of response to power outages. However, since the most robust effects occur in the current period, we believe it still serves as a useful model of behaviour. Individuals react mostly to their observations of the current week's electricity provision. Anticipation of future week's power outages and delayed responses appear to be less important.

2.7.3 Clustering

As a final robustness check, we consider the implications of clustering our standard errors differently. Thus far, our standard errors have generally been clustered two way, by individual and by week. In Table 2.22 we re-estimate Equation (2.10) for domestic work (Columns (1)–(3)) and Equation (2.11) for market work (Columns (4)–(6)). Initially we cluster our standard errors as in the main specifications, by individual and week. We then cluster two way by household and week, and also one way by household.

Table 2.22: Clustering

	Domestic Work			Market Work		
	(1)	(2)	(3)	(4)	(5)	(6)
	ID-Week	HH-Week	HH	ID-Week	HH-Week	HH
Hours of Outage/Week	0.0201**	0.0201**	0.0201*	-0.0125**	-0.0125**	-0.0125*
	(0.0085)	(0.0092)	(0.0107)	(0.0053)	(0.0063)	(0.0067)
	[2.3578]	[2.1751]	[1.8781]	[-2.3762]	[-2.0055]	[-1.8650]
Hours of Outage/Week Squared (x 100)	-0.1396***	-0.1396**	-0.1396**	0.0913***	0.0913***	0.0913***
	(0.0482)	(0.0565)	(0.0615)	(0.0275)	(0.0322)	(0.0345)
	[-2.8955]	[-2.4712]	[-2.2717]	[3.3199]	[2.8343]	[2.6439]
Observations	1363	1363	1363	1363	1363	1363
R^2	0.387	0.387	0.387	0.051	0.051	0.051

Standard errors in regular parentheses, t-statistics in square parentheses
Dependent variable: Log of Hours of Domestic or Market Work/Week
All columns include week and household fixed effects, and individual controls
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As expected, the statistical significance of our key coefficients declines when clustering at the household level (both two way and one way). However, even if the t-statistics increase with more conservative clustering approaches, they do not jump hugely. This suggests our initial specification is tenable.

2.8 Simulation

2.8.1 Functional Form and Calibration

In this section, we add functional form and simulate our simple model to rationalise the results for the self-employed. In particular, we utilise a Nested Constant Elasticity of Substitution (CES) utility function. The results suggest that substitution between market work and domestic work, and hence market goods and domestic goods, occurs in a separate nest from leisure, so this is the structure we adopt. To increase flexibility, we also allow for the possibility of subsistence constraints for domestic goods, market goods, and leisure, by adding Stone-Geary parameters to each term in the utility function (Geary, 1950; Stone, 1954). Maintaining the same notation as Section 2.3, we can thus write the utility function:

$$U = u(x_m, x_d, t_l) = (\beta_G(\alpha_m(x_m - \gamma_m)^{-\epsilon_G} + \alpha_d(x_d - \gamma_d)^{-\epsilon_G})^{\frac{\epsilon_L}{\epsilon_G}} + \beta_L(t_L - \gamma_l)^{-\epsilon_L})^{-\frac{1}{\epsilon_L}} \quad (2.19)$$

The relative weights that individuals place on goods (G), either domestic or market, and leisure (L) are governed by β_G and β_L respectively. The willingness to substitute between these two nests is determined by the parameter ϵ_L, which constitutes a transformation of the elasticity of substitution between goods and leisure, σ_L, of the following form:

$$\epsilon_L = \frac{1 - \sigma_L}{\sigma_L} \quad (2.20)$$

Within the goods nest, the α_m and α_d are the relative weights on market and domestic goods respectively. The elasticity of substitution between these two goods, σ_G, enters through the ϵ_G term as follows:

$$\epsilon_G = \frac{1 - \sigma_G}{\sigma_G} \quad (2.21)$$

The lower bounds on market goods, domestic goods, and leisure are given by γ_m, γ_d, and γ_l. However, we drop γ_m and γ_d given the presence of the unearned domestic and market goods terms (τ_d and τ_m) in the maximisation problem set out in Equations (2.1)–(2.5). We can therefore simplify Equation (2.19) to:

$$U = u(x_m, x_d, t_l) = (\beta_G(\alpha_m {x_m}^{-\epsilon_G} + \alpha_d {x_d}^{-\epsilon_G})^{\frac{\epsilon_L}{\epsilon_G}} + \beta_L(t_L - \gamma_l)^{-\epsilon_L})^{-\frac{1}{\epsilon_L}} \tag{2.22}$$

The individual optimisation problem thus requires maximising Equation (2.22) subject to Equations (2.2), (2.3), (2.4), and (2.5). By simulating this problem with different parameter values, we can better understand the mechanisms that drive our results for the self-employed.

We begin by calibrating the model to approximate the relationships observed in the data. We normalise market work productivity, w, to 1 and simulate for different values of domestic work productivity, ρ.

The initial values of the other parameters are shown in Table 2.23. These are chosen both to rationalise the data, and provide a suitable baseline for the simulations that follow. We focus utility heavily on the goods nest rather than the leisure nest by selecting a high value of β_G compared to β_L, as our results suggest the main margin of adjustment is between domestic and market work. However, we calibrate with a high value on γ_l to ensure individuals still devote some time to leisure. However, our choices of the weights within the goods nest, α_m and α_d, are neutral.

We begin with the assumption that $\sigma_G = \sigma_L = 1$, which corresponds to Cobb-Douglas preferences. However, as we explain below, we must increase the level of friction in substituting between domestic goods and market goods (σ_G) to best illustrate the non-monotonic relationships between outages and time allocation observed in the data. We also show below that these empirical patterns rely on $\tau_d > 0$.

We simulate the model for different values of domestic productivity, ρ, at intervals of 0.5 between 0 and 30. The units for these simulations are simplified by the fact that $w = 1$. Under this assumption, ρ is the rate at which time can be converted into domestic goods, relative to the rate for *both* market goods *and* leisure. Similarly, $\frac{1}{\rho}$ is the price of domestic goods relative to market goods and leisure. Higher levels of ρ imply it is easier to produce domestic goods, so that each unit of x_d, and indeed the endowment τ_d, is valued less. Thus, we can think of ρ in terms of 'quantity of domestic goods produced per hour' or the rate at which domestic and market goods

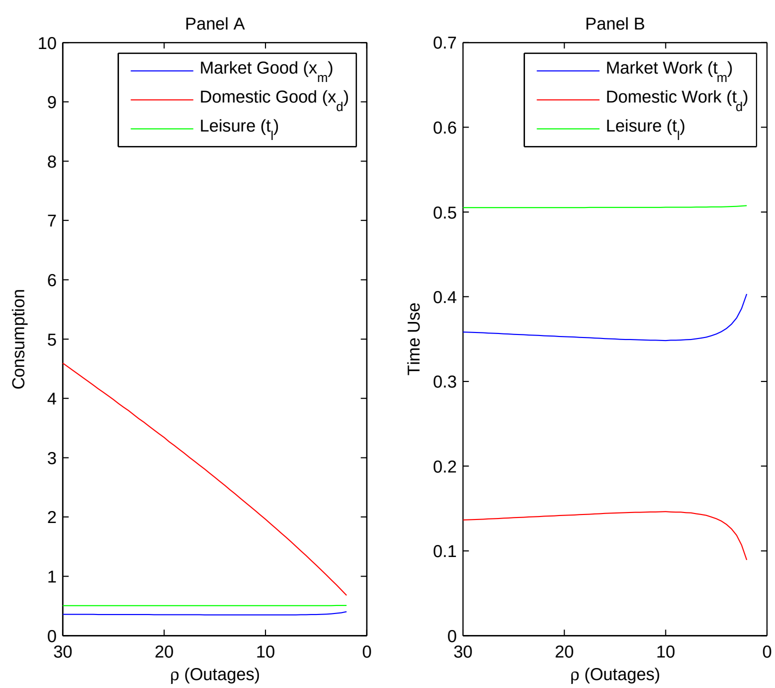
Table 2.23: Calibration

Parameter	Calibration Value
τ_m	0.0
τ_d	0.5
β_G	0.99
β_L	0.01
α_m	0.5
α_d	0.5
σ_G	0.75
σ_L	1.0
γ_l	0.5

could be traded *if* it were actually possible to exchange them.

In Figure 2.10, we plot the outcomes for consumption (Panel A) and time-use (Panel B). We reverse the x-axes, so that the graphs are directly analogous to the empirical results displayed in Figure 2.8.

Figure 2.10: Base Simulation of the Model



The non-monotonicity in the ρ-t_d and ρ-t_m relationships result from a combination of what we labelled 'supply effects' and 'demand effects' in Section 2.3. In Panel A, we see a monotonically increasing relationship between ρ and x_d, consistent with the substitution effect receiving support from the regular income effect. As such, it is unlikely that demand effects alone are resulting in the non-monotonic relationships we see in Panel B. Instead, it appears that for higher values of ρ, or when outages are less severe, the supply effect dominates. The decrease in domestic productivity that occurs when outages hit mean that more domestic work is required to meet demand for domestic goods (in spite of the fall in x_d). Thus, t_d initially rises with hours of outage per week. As outages become more severe, however, supply effects start to wane relative to demand effects. Eventually demand for domestic goods approaches the endowment of domestic goods, and $|(x_d - \tau_d)|$ approaches zero (see Equation (2.9)). This coincides with a substitution away from t_d towards t_m, as well as increased consumption of market goods, x_m.

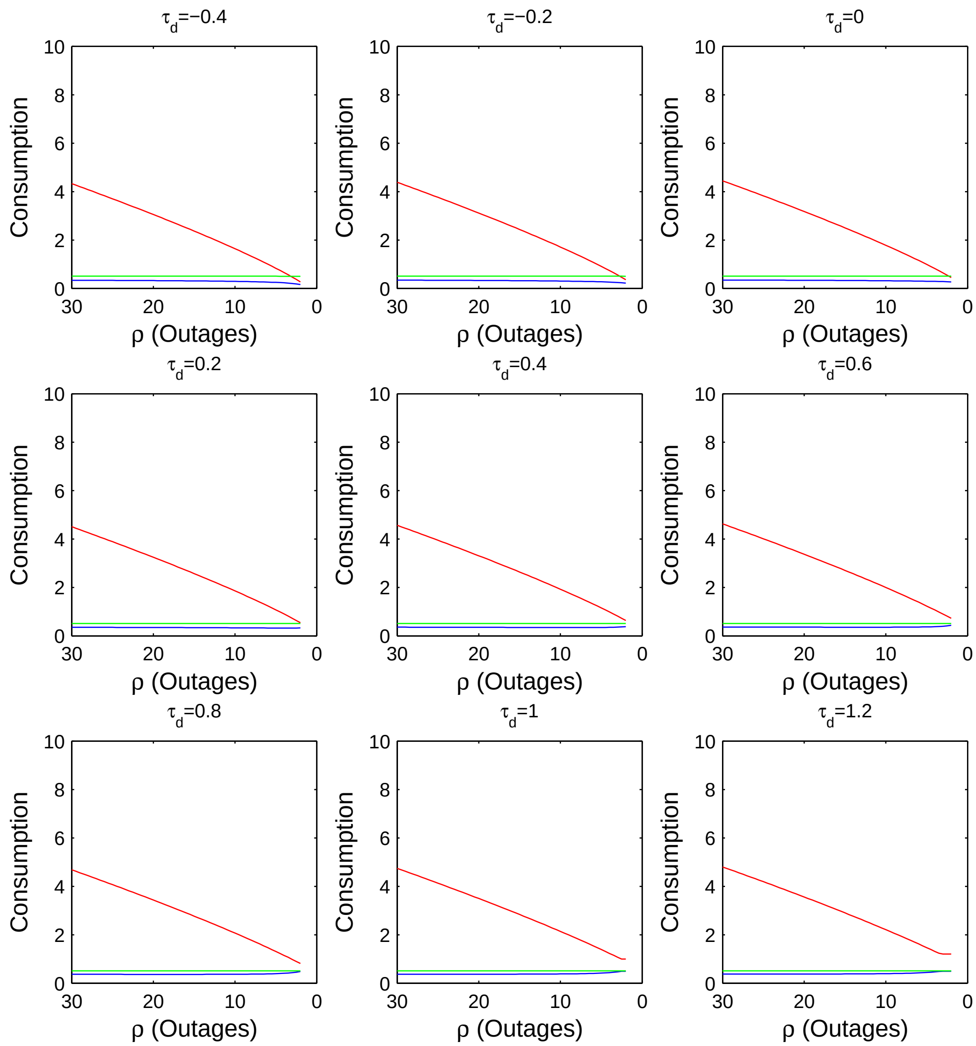
2.8.2 Unearned Domestic Goods: τ_d

The endowment of domestic goods that individuals receive from (or must provide for) the household, τ_d, plays an important role in generating the relationships between outages and time-use we observe in the data. To illustrate this, we simulate the relationship between ρ and domestic work, market work, and leisure for different values of τ_d. We show the results in terms of consumption in Figure 2.11 and time-use in Figure 2.12.

Consumption patterns remain fairly stable when changing the value of τ_d. As expected, an increase in τ_d shifts the x_d curve upwards, ceteris parabis.

By contrast, patterns of time-use are highly sensitive to the sign and size of τ_d, in accordance with our interpretation in terms of demand and supply effects. When $\tau_d < 0$, demand for domestic goods cannot approach the endowment/lower bound on domestic good provision. As such, the absolute value of $(x_d - \tau_d)$ remains large, and thus supply effects continue to place upward pressure on t_d when outages hit for all values of ρ. When $\tau_d > 0$, however, supply effects eventually become small because, when outages are high and ρ is low, individuals can meet their demand for domestic goods using their endowment. This means the absolute value of $(x_d - \tau_d)$ is small. As such the positive sign on $\frac{\partial x_d}{\partial \rho}$ results in a positive sign on $\frac{\partial t_d}{\partial \rho}$, the demand effects dominate, and individuals devote less time to domestic work as outages become more severe. It follows that a greater endowment of domestic goods causes this turning point to arrive for lower values of the outage variable, or higher values of ρ.

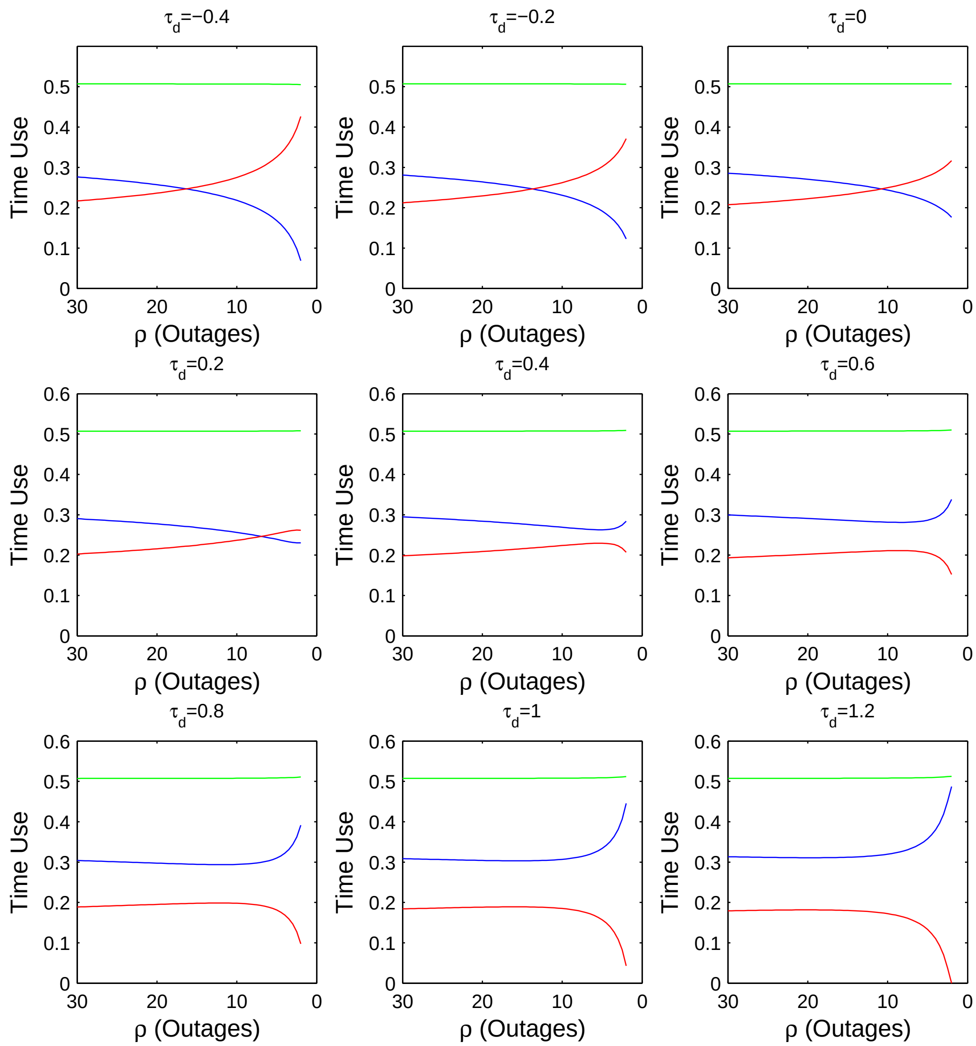
Figure 2.11: Adjusting τ_d — Consumption Patterns



2.8.3 Elasticity of Substitution between Domestic and Market Goods: σ_G

The importance of the balance between demand effects and supply effects is further emphasised by changing the elasticity of substitution between domestic and market goods, σ_G. We once again simulate our model, holding everything besides σ_G constant. We show the consumption

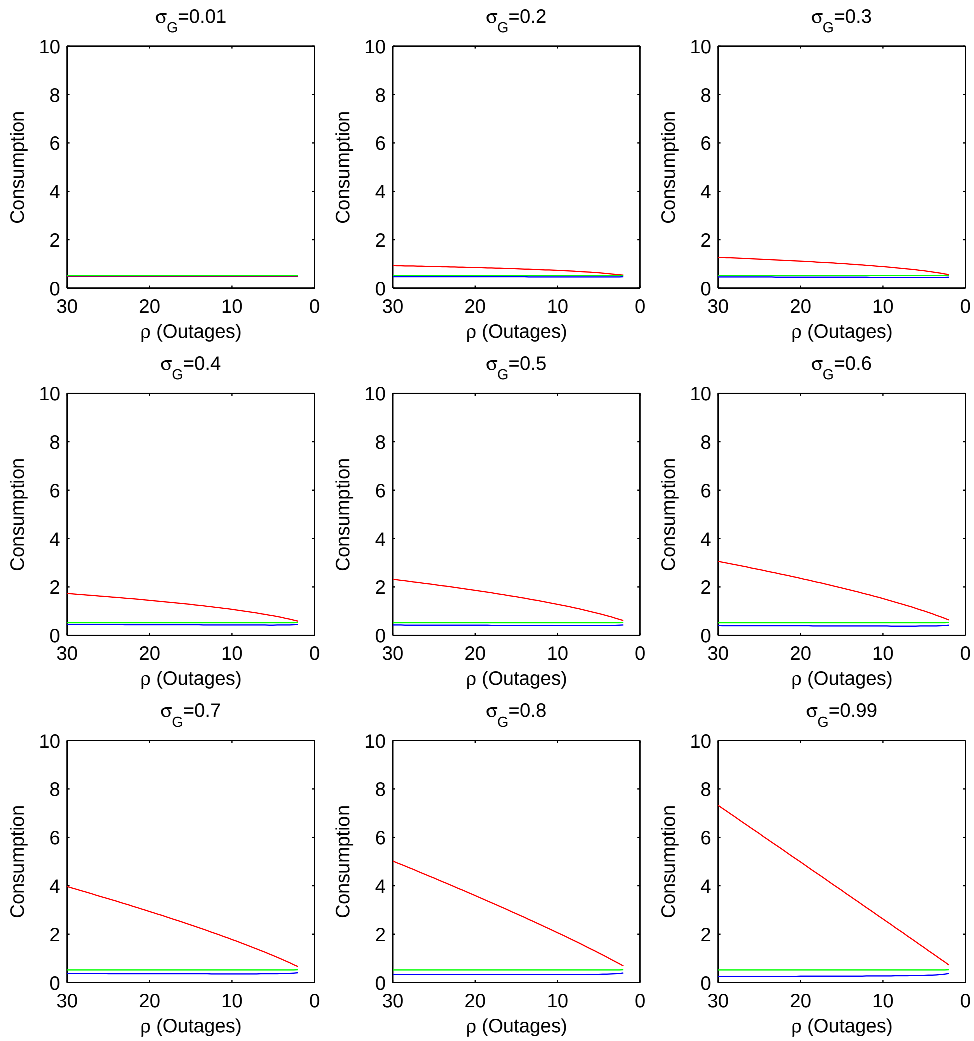
Figure 2.12: Adjusting τ_d — Time-Use Patterns



paths in Figure 2.13 and time allocation in Figure 2.14.

Firstly, a higher level of σ_G implies that individuals consume more domestic goods unless ρ is very low. As such, if preferences over the mix of x_d and x_m are very inelastic, workers gain little from expanding their consumption of domestic goods when they are easier to produce, that is, when ρ is high and power cuts are less severe. This has direct implications for welfare. Lower values of σ_G result in individuals being worse off.

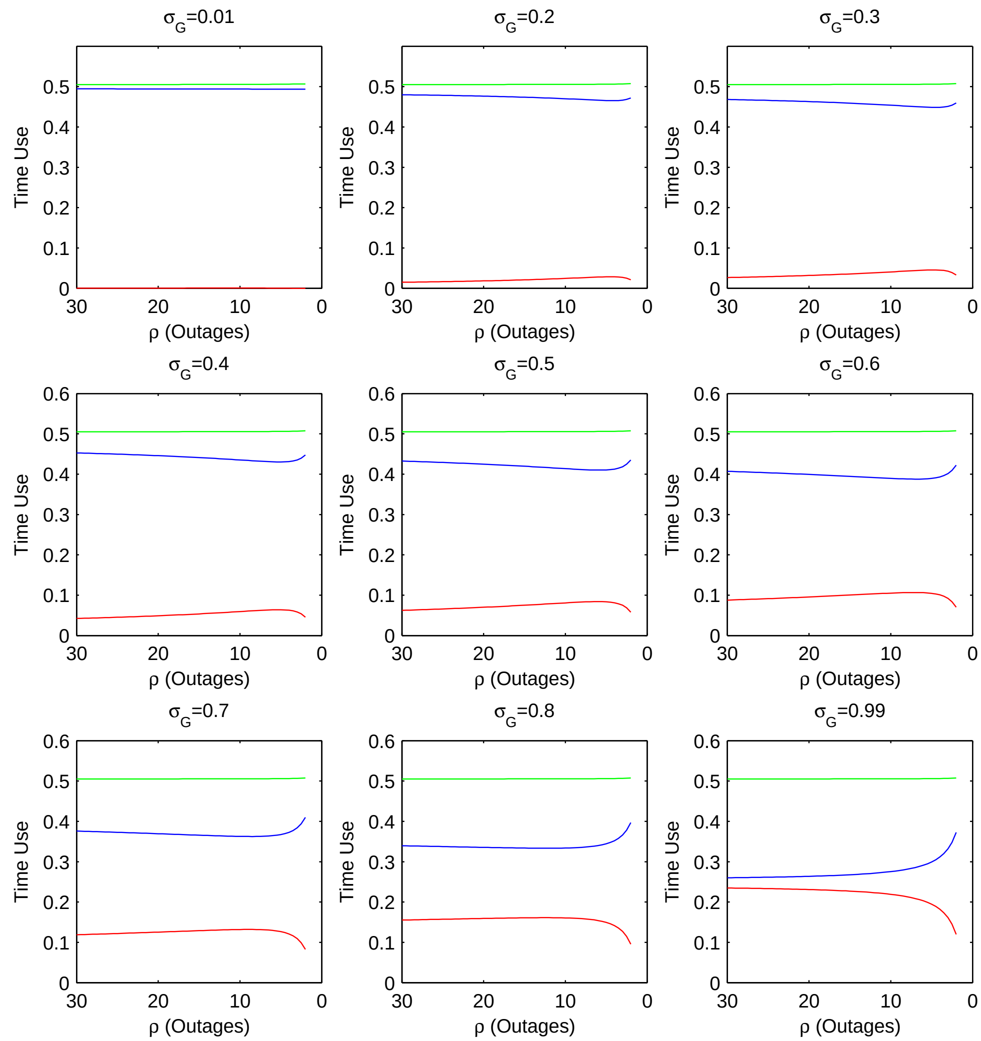
Figure 2.13: Adjusting σ_G — Consumption Patterns



The consumption patterns also demonstrate that individuals substitute away from domestic goods faster as ρ decreases, when σ_G is higher. The rate at which individuals substitute away from x_d determines whether or not outages and time-use are related non-monotonically.

If individuals are totally unwilling to substitute between domestic and market goods (as σ_G tends to zero) outages have no effect on consumption and, in turn, time-use. This is shown by the case where $\sigma_G = 0.01$. Given the other parameter values, in this scenario we observe individuals

Figure 2.14: Adjusting σ_G — Time-Use Patterns



simply consuming their endowment of domestic goods, and splitting their time between market work and leisure. As such, there are no demand effects, since $\frac{\partial x_d}{\partial \rho} = 0$, and no supply effects, since $(x_d - \tau_d) = 0$.

However, if market and domestic goods are too substitutable, demand effects may dominate supply effects for all values of ρ. In particular, as we see when $\sigma_G = 0.99$, individuals wish to substitute away from x_d so much when ρ decreases, that t_d also falls accordingly. Although

$(x_d - \tau_d) > 0$, supply effects are insufficient to counteract the large changes in demand. Thus, we require some stickiness in individuals' willingness to substitute between domestic and market goods to produce the non-monotonic empirical relationships between outages and t_d and t_m that we observe in Section 2.6.

2.8.4 External Validity

One way of checking the external validity of our results is to consider how our theoretical predictions might change if individuals' returns to market work vary. In a sense, this may be thought of as a proxy for the level of economic development, insofar as wages are higher in more developed countries. However, simulating our model with different market earnings rates, w, makes little difference to the relationships between time-use and ρ in our baseline model. As such our findings are robust to changing the price of leisure, as well as the relative price of domestic and market goods ($\frac{w}{\rho}$). This in itself suggests our model may be suitable for contexts besides urban Ghana, where workers have better or worse options to work in the market.

However, the empirical patterns, on which our model calibration is based, are likely to be somewhat context-specific. Firstly, our main findings are based on a sample of self-employed individuals, most of whom are retailers. Whilst workers of this type are prevalent in urban Ghana, the composition of the workforce may be different in other labour markets, especially those outside sub-Saharan Africa. Secondly, as our descriptive statistics in Section 2.4 demonstrate, power outages are a continuous threat for the respondents in our sample. Thus, it is likely that individuals have adopted adaptive strategies to minimise the welfare losses incurred when there are shocks to ρ. Thus, it is only in developing countries, where electrification is similarly extensive but unreliable, that these strategies may be required and subsequently deployed.

2.8.5 Theoretical Robustness

Although we believe our basic model accurately captures how individuals respond to power cuts in urban Ghana, there may be alternative stories which are worth considering. In particular, we may be concerned that power outages proxy for returns to leisure rather than returns to market work, ρ.

Overall, however, the data appear to fit our modelling approach better. Firstly, anecdotal evidence suggests that individuals are flexible about how they take their leisure, such that power outages have little effect on how much time is spent away from work (Darko, 2015). More importantly, the main margin of adjustment, especially below the 75^{th} percentile of the outages variable, appears to be between market work and domestic work. Constructing a model in which these patterns are driven by changing returns to leisure requires some rather heroic assumptions about the relationships between x_m, x_d, and t_l in the utility function. As such, we believe our theoretical framework, and our approach for linking it to the data, are tenable.

2.9 Conclusion

Using high-frequency labour force data, we find a significant non-monotonic relationship between the severity of power outages in a particular week, and the time that individuals devote to domestic and market work, for the sub-sample of individuals typically working in self-employment. At the 25^{th} percentile in the distribution of power outages, an extra hour of outage for that week increases time spent in domestic work by 2.01 percent, and reduces the time spent in market work by 1.25 percent for the average self-employed worker. Conversely, at the 75^{th} percentile, an extra hour of outage reduces domestic work by 1.34 percent and increases market work by 0.94 percent. We find no robust relationships between time-use and power outages for the wage-employed.

We interpret hours of power outage per week as a proxy for domestic productivity. This interpretation is supported by our robustness checks in Section 2.7. To begin, we recognise that for some self-employed individuals, especially those that use electricity for their employment and live and work in close proximity, power outages may proxy for market productivity as well as domestic productivity. However, our key coefficients are robust to controlling for these factors. Moreover, accounting directly for variation in market productivity does not substantially change our results.

By adding functional form and simulating the simple model outlined in Section 2.3, we examine the combinations of 'demand effects', which result from changing consumption patterns, and 'supply effects', which relate to changes in individuals' ability to produce a given level of domestic goods, that operate to produce the relationships we observe empirically.

Simulating the model for different values of τ_d, we show that positive endowments of domestic income are required to induce the non-monotonicity of the relationship between outages and domestic/market work. This is because larger endowments of domestic goods make individuals less sensitive to supply effects, even when outages are less severe. Intuitively, when individuals receive more unearned domestic goods, productivity in producing those goods matters less for meeting demand. This result is interesting, insofar as it implies there are positive transfers of domestic goods between individuals within the urban households in our sample.

We also demonstrate that the level of substitutability between market and domestic goods is vital for producing the non-monotonic relationships between ρ and time-use, by simulating our model for different values of σ_G. As domestic goods and market goods approach perfect complementarity, neither consumption nor time-use react to any changes in domestic productivity. However, if the goods are too substitutable, then demand effects dominate supply effects for all values of ρ, yielding a positive relationship between ρ and t_d, which is at odds with the data.

Our results reveal some important heterogeneity between the adjustment paths of women and men in the self-employed sub-sample. In particular, mild outages cause women to devote extra time to domestic work far more than men, whilst serious outages cause women to substitute back towards market work far less. In addition, self-employed women do at least double the amount of domestic work each week compared to self-employed men. These findings reinforce the idea that domestic tasks may be better understood as *obligations* for women, compared with men. This links with our model of occupational choice in Chapter 3.

Overall, our results suggest that individuals' ability to substitute between domestic and market work is complex. When negative shocks to domestic productivity are mild, domestic work rises at the expense of market work to try and partially maintain existing consumption levels. However, more severe negative shocks result in a substitution away from domestic work towards market work. As such, it appears that if the variation in domestic productivity is sufficiently large, domestic work and market work can be substituted. Vitally, these relationships differ by sex, helping us understand the extent to which different household members are truly obliged to undertake domestic work.

2.A Household-Level Treatment Assignment

Table 2.24: Breakdown of Treatment Groups by City — Household Level

	City			
Treatment	Accra	Cape Coast	Kumasi	Total
Called Once a Week	60	7	50	117
Called Three Times a Week	70	7	44	121
Control With Phone	62	8	46	116
Control Without Phone	62	8	50	120
Face-to-Face (Once a Week)	70	7	47	124
Total	324	37	237	598

Table 2.25: Number of Individuals Interviewed from Weekly Sample Households

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
No. Sampled Individuals in HH	242	2.60	1.55	1	1	2	4	7
Observations	242							

2.B Data Cleaning

To begin, data from the high-frequency survey (GHFLS) were merged with data from the existing labour force panel (GHUPS), to bring together all available information on the respondents. This allowed individual- and household-level characteristics, which were missing in the GHFLS to be recovered and used in the analysis. At this point the merged data were cleaned such that implausible values (e.g. more than 24 hours in a day), were either 'winsorised' or set to missing.

The most important data cleaning occurred on the outage variables, which were reported by each household member separately, but subsequently harmonised to the household level. This helps alleviate any reporting biases that may occur if individuals within the household work in different occupations and therefore spend more or less of their time at home.

Harmonising the outage variables to the household level was complex, because household members were sometimes interviewed on different days of the week. Thus, when asking about the 'last seven days', which was the format of many questions in the questionnaire, respondents within a household in a given week may have been talking about different periods. To get around this difficulty, we expanded the dataset from the weekly level to the daily level, using the *specific date* of the interview to ascertain the days of the week to which each respondent was referring. The outage variables could then be harmonised across individuals to the household-day level. We used the maximum reported outage for each household-day, because under-reporting of outages at home was the most serious potential problem, given individuals' work practices.

The outage variables could then be calculated for each individuals' reference seven days, regardless of when in the week they were interviewed, using all available data from the household. By harmonising the data in this way, we are effectively able to repeat the interview as though the respondent had the best information available from all household members in the survey.

We show in Table 2.26 that our main results are not sensitive to harmonising the data in this way. 'Method 1' corresponds to the approach described above, and thus Columns (1) and (4) recapitulate our main results for domestic and market work respectively. For 'Method 2' we simply drop any observations where individuals within the same household were interviewed on different days of a particular week and re-estimate our main equations. The magnitude of our main results actually increases in this sub-sample. Finally, for 'Method 3', we simply aggregate across weeks, without accounting for whether interviews of respondents within the same household occurred on different days of the week. This is clearly a much coarser approach and, as a result, it is unsurprising that the standard errors on the coefficients increase. Nonetheless, the signs and magnitudes of the coefficients are somewhat stable.

Table 2.26: Changing the Method of Harmonising Outage Reports to the Household Level

	Domestic Work			Market Work		
	(1)	(2)	(3)	(4)	(5)	(6)
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
Hours of Outage/Week	0.0201**	0.0243**	0.0158*	-0.0125**	-0.0145**	-0.0183*
	(0.0085)	(0.0122)	(0.0095)	(0.0053)	(0.0070)	(0.0098)
Hours of Outage/Week Squared (x 100)	-0.1396***	-0.1624***	-0.1082	0.0913***	0.0987***	0.1383*
	(0.0482)	(0.0598)	(0.0804)	(0.0275)	(0.0334)	(0.0728)
Observations	1363	1090	1400	1363	1090	1400
R^2	0.387	0.422	0.385	0.051	0.048	0.035

Standard errors in parentheses

Dependent variable: Log of Hours of Domestic or Market Work/Week

Standard errors clustered two way by individual and week

All columns include week and household fixed effects, and individual controls

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.C Alternative Summary Statistics for Time-Use and Outage Variables

Table 2.27: Time-Use Summary Statistics by Sex

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Full Sample								
Hours of Domestic Work/Week	6221	13.24	15.78	0.00	2.00	6.50	21.00	63.00
Hours of Market Work/Week	6231	24.46	25.13	0.00	0.00	16.50	48.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	4941	16.66	16.01	0.15	4.00	10.00	26.00	63.00
Hours of Market Work/Week (Excluding Zeros)	3596	42.39	18.28	1.00	30.00	45.00	55.00	72.00
Female								
Hours of Domestic Work/Week	3488	20.45	17.13	0.00	6.00	16.00	32.00	63.00
Hours of Market Work/Week	3612	22.13	24.51	0.00	0.00	9.00	45.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	3217	22.18	16.73	0.15	8.00	20.00	35.00	63.00
Hours of Market Work/Week (Excluding Zeros)	1933	41.35	18.09	1.00	30.00	45.00	54.00	72.00
Male								
Hours of Domestic Work/Week	2733	4.02	6.39	0.00	0.00	2.00	6.00	63.00
Hours of Market Work/Week	2619	27.68	25.62	0.00	0.00	30.00	52.50	72.00
Hours of Domestic Work/Week (Excluding Zeros)	1724	6.38	7.06	0.15	2.00	4.00	8.00	63.00
Hours of Market Work/Week (Excluding Zeros)	1663	43.60	18.43	1.00	32.00	48.00	60.00	72.00
Observations	6480							

Table 2.28: Time-Use Summary Statistics by Sex — Self-Employed Only

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Full Sample								
Hours of Domestic Work/Week	2539	14.36	16.11	0.00	1.50	8.00	22.00	63.00
Hours of Market Work/Week	2516	32.30	23.90	0.00	8.00	36.00	54.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	1984	18.38	16.07	0.20	5.00	14.00	28.00	63.00
Hours of Market Work/Week (Excluding Zeros)	1976	41.13	19.09	1.33	28.00	45.00	55.80	72.00
Female								
Hours of Domestic Work/Week	1569	20.87	16.94	0.00	7.00	18.00	34.00	63.00
Hours of Market Work/Week	1605	30.57	24.05	0.00	4.67	32.00	54.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	1431	22.88	16.39	0.50	8.00	21.00	35.00	63.00
Hours of Market Work/Week (Excluding Zeros)	1216	40.35	19.21	1.33	25.00	44.00	55.40	72.00
Male								
Hours of Domestic Work/Week	970	3.84	6.04	0.00	0.00	1.75	6.00	57.00
Hours of Market Work/Week	911	35.35	23.33	0.00	10.00	40.00	54.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	553	6.74	6.67	0.20	2.00	4.00	8.50	57.00
Hours of Market Work/Week (Excluding Zeros)	760	42.37	18.84	2.00	30.00	48.00	56.00	72.00
Observations	2659							

Table 2.29: Time-Use Summary Statistics by Sex — Wage-Employed Only

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Full Sample								
Hours of Domestic Work/Week	1744	9.92	13.08	0.00	1.00	5.00	14.00	63.00
Hours of Market Work/Week	1669	36.32	23.00	0.00	15.00	40.00	54.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	1346	12.86	13.56	0.15	3.00	7.00	16.10	63.00
Hours of Market Work/Week (Excluding Zeros)	1359	44.61	16.72	1.00	36.00	45.00	55.00	72.00
Female								
Hours of Domestic Work/Week	750	16.71	15.49	0.00	5.00	12.00	24.50	63.00
Hours of Market Work/Week	723	34.56	22.67	0.00	12.00	40.00	54.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	699	17.93	15.35	0.40	6.00	14.00	28.00	63.00
Hours of Market Work/Week (Excluding Zeros)	569	43.92	15.55	3.00	36.00	45.00	54.00	72.00
Male								
Hours of Domestic Work/Week	994	4.80	7.63	0.00	0.00	2.00	6.00	63.00
Hours of Market Work/Week	946	37.67	23.17	0.00	16.00	44.00	54.00	72.00
Hours of Domestic Work/Week (Excluding Zeros)	647	7.38	8.39	0.15	2.00	4.00	8.00	63.00
Hours of Market Work/Week (Excluding Zeros)	790	45.11	17.51	1.00	36.00	47.75	60.00	72.00
Observations	1765							

2.D Total Work Regressions

Table 2.30: Main Results — Total Work

	(1) Full Sample	(2) SE	(3) WE	(4) Full Sample	(5) SE	(6) WE
Hours of Outage/Week	0.0003	0.0022	0.0018	-0.0034*	-0.0094**	0.0105*
	(0.0012)	(0.0015)	(0.0018)	(0.0019)	(0.0044)	(0.0057)
Hours of Outage/Week Squared (x 100)				0.0191	0.0610***	-0.0445*
				(0.0128)	(0.0218)	(0.0263)
Male? (1=Y, 0=N)	-0.2919***	-0.2993**	-0.1576***	-0.2913***	-0.2978**	-0.1619***
	(0.0560)	(0.1207)	(0.0383)	(0.0559)	(0.1194)	(0.0377)
Married? (1=Y, 0=N)	0.0418	-0.0675	0.2582*	0.0422	-0.0701	0.2566*
	(0.0715)	(0.0824)	(0.1323)	(0.0714)	(0.0818)	(0.1324)
Male x Married	-0.1051	-0.1747	-0.0191	-0.1057	-0.1782	-0.0161
	(0.0787)	(0.1268)	(0.2021)	(0.0785)	(0.1257)	(0.2019)
Years of Education	0.0013	-0.0070	0.0753**	0.0009	-0.0083	0.0765**
	(0.0175)	(0.0372)	(0.0336)	(0.0175)	(0.0373)	(0.0336)
Years of Education Squared	0.0009	-0.0001	-0.0034**	0.0010	-0.0000	-0.0035**
	(0.0010)	(0.0028)	(0.0015)	(0.0010)	(0.0028)	(0.0015)
Age (Years)	0.0558***	0.0468**	-0.0067	0.0557***	0.0466**	-0.0070
	(0.0165)	(0.0201)	(0.0230)	(0.0164)	(0.0200)	(0.0229)
Age (Years) Squared	-0.0007***	-0.0005**	0.0001	-0.0007***	-0.0005**	0.0001
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0003)
Observations	2556	1363	996	2556	1363	996
R^2	0.129	0.159	0.064	0.129	0.163	0.067

Standard errors in parentheses
Dependent variable: Log of Hours of Total Work/Week
Standard errors clustered two way by individual and week
Columns (1) to (6) include week and household fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.E Changing Fixed-Effects and Controls

Table 2.31: Changing Fixed-Effects and Controls — Domestic Work

	(1)	(2)	(3)	(4)	(5)	(6)
Hours of Outage/Week	0.0201**	0.0416***	0.0414***	0.0543***	0.0177**	0.0149*
	(0.0085)	(0.0129)	(0.0125)	(0.0197)	(0.0082)	(0.0090)
Hours of Outage/Week Squared (x 100)	-0.1396***	-0.2326***	-0.2270***	-0.2955***	-0.1377***	-0.0809*
	(0.0482)	(0.0663)	(0.0659)	(0.0974)	(0.0448)	(0.0433)
N	1363	1363	1363	1363	1363	1363
R^2	0.387	0.350	0.343	0.015	0.401	0.003
Household Fixed Effects	✓	✗	✗	✗	✓	✓
Week Fixed Effects	✓	✓	✗	✗	✓	✓
Individual Controls	✓	✓	✓	✗	✓	✗
City-Week Fixed Effects	✗	✗	✗	✗	✓	✗
Individual Fixed Effects	✗	✗	✗	✗	✗	✓

Standard errors in parentheses
Dependent variable: Log of Hours of Domestic Work/Week
Standard errors clustered two way by individual and week
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.32: Changing Fixed-Effects and Controls — Market Work

	(1)	(2)	(3)	(4)	(5)	(6)
Hours of Outage/Week	-0.0125**	-0.0407***	-0.0404***	-0.0393***	-0.0121**	-0.0030
	(0.0053)	(0.0101)	(0.0101)	(0.0099)	(0.0056)	(0.0043)
Hours of Outage/Week Squared (x 100)	0.0913***	0.1483***	0.1525***	0.1384**	0.0887***	0.0286
	(0.0275)	(0.0555)	(0.0564)	(0.0552)	(0.0290)	(0.0218)
N	1363	1363	1363	1363	1363	1363
R^2	0.051	0.040	0.038	0.026	0.050	0.002
Household Fixed Effects	✓	✗	✗	✗	✓	✓
Week Fixed Effects	✓	✓	✗	✗	✓	✓
Individual Controls	✓	✓	✓	✗	✓	✗
City-Week Fixed Effects	✗	✗	✗	✗	✓	✗
Individual Fixed Effects	✗	✗	✗	✗	✗	✓

Standard errors in parentheses
Dependent variable: Log of Hours of Market Work/Week
Standard errors clustered two way by individual and week
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.F Extensive Margin

Our main analysis focusses on the intensive margin, that is, on individuals' decisions about *how much* domestic or market work to do, given that they have elected to work. In this appendix we argue that our results regarding the intensive margin are not biased by unobservables associated with the extensive margin. We focus on the sub-sample of self-employed individuals.

First, we estimate a set of probit regressions to consider if the binary decision over whether or not to work in domestic or market work in a given week is influenced by power outages. The results are reported in Tables 2.33 and 2.34. All these equations are estimated without household and week fixed-effects so as to avoid the Incidental Parameters Problem, but the models are estimated with and without individual controls — sex, marital status, education, and age. We test whether the controls are jointly significant, when included, and report the p-value for this test at the bottom of each table.[25]

Table 2.33: Marginal Effects for Extensive Margin Probit Regression — Domestic Work

	Linear		Quadratic	
	(1)	(2)	(3)	(4)
Positive Domestic Work? (1=Y, 0=N)				
Hours of Outage/Week	-0.0042	-0.0001	0.0326	0.0304
	(0.0090)	(0.0087)	(0.0206)	(0.0216)
Hours of Outage/Week Squared (x 100)			-0.1876*	-0.1546
			(0.1079)	(0.1011)
N	2206	2206	2206	2206
Biprobit: p-value on ρ	0.871	0.344	0.829	0.376
Individual Controls	✗	✓	✗	✓
p-value on Controls		0.0000		0.0000

Standard errors in parentheses
Standard errors clustered at the individual level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

If *and only if* the individual controls are included, power outages do not appear to affect the extensive margin for either domestic or market work. However, the controls themselves do appear to determine whether or not individuals do any work. Thus, their inclusion in the main *intensive* margin equations may help to alleviate concerns about sample selection.

Pushing this logic further, we also re-estimate our main regression equations from Section 2.6.1, using the Heckman method to correct for the fact that we use a selected sample. The main issue with this approach is finding exclusion restrictions, which could plausibly influence whether

[25] We also use a biprobit equation to test whether the extensive margin for domestic work and market work can be estimated separately, and report the p-values. We are unable to reject the null hypothesis that the equations are independent across the specifications.

Table 2.34: Marginal Effects for Extensive Margin Probit Regression — Market Work

	Linear		Quadratic	
	(1)	(2)	(3)	(4)
Positive Market Work? (1=Y, 0=N)				
Hours of Outage/Week	0.0006	0.0017	0.0122	0.0252
	(0.0084)	(0.0090)	(0.0208)	(0.0215)
Hours of Outage/Week Squared (x 100)			-0.0601	-0.1209
			(0.1132)	(0.1186)
N	2206	2206	2206	2206
Biprobit: p-value on ρ	0.871	0.344	0.829	0.376
Individual Controls	✗	✓	✗	✓
p-value on Controls		0.1261		0.0620

Standard errors in parentheses
Standard errors clustered at the individual level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

or not individuals do domestic or market work, without influencing the intensive margin. We use variables relating to the number of children of the respondent and the quality of water provision for the household that week.[26] Given the questionable excludability of these variables, we are somewhat cautious about interpreting the selection-corrected results causally, but we believe they provide a useful benchmark for assessing the impact of sample selection.

The results for the domestic work equation are shown in Columns (1)–(3) of Table 2.35, whilst those for market work are shown in Columns (4)–(6). Since there is some missing data on our exclusion restrictions, we first re-estimate our main regression equations, including all fixed-effects and controls, in Columns (1) and (4), but on the slightly reduced sample. We then omit household fixed-effects in Columns (2) and (5), and then estimate the selection-corrected models in Columns (3) and (6). We also test whether the exclusion restrictions are jointly significant and report the p-value (0.0018) at the bottom of the table.

The Inverse Mills Ratio is not statistically significant, even at the 10 percent level, in either the domestic or market work equations. This suggests that sample selection may not create too much bias in the intensive margin estimates. In addition, we are also unable to reject the null hypotheses that the power outage variables have the same coefficients in the OLS and Heckman models.

Moreover, Columns (2), (3), (5), and (6) omit household fixed-effects. These may further control for unobservable factors that could bias the intensive margin equations due to sample selection. That is, with both the individual controls *and* the full set of fixed-effects it is even more plausible that selection occurs on the basis of observables rather than unobservables.

[26] We informally test the excludability of these variables by including them in the main outcome equations and find that they have no effect.

Table 2.35: Re-Estimating the Main Results with Selection Correction

	Domestic Work			Market Work		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	Heckman	OLS	OLS	Heckman
Hours of Outage/Week	0.0187*	0.0460***	0.0362**	-0.0103*	-0.0414***	-0.0473***
	(0.0108)	(0.0145)	(0.0170)	(0.0062)	(0.0112)	(0.0125)
Hours of Outage/Week Squared (x 100)	-0.1263**	-0.2583***	-0.2069**	0.0873***	0.1469**	0.1780***
	(0.0589)	(0.0752)	(0.0891)	(0.0331)	(0.0609)	(0.0685)
Inverse Mills Ratio			-0.8736			-0.5285
			(0.8315)			(0.5394)
N	1281	1281	1281	1281	1281	1281
R^2	0.388	0.361	0.363	0.050	0.046	0.049
Test Equal Coefs.		0.5639			0.6140	
Test Exclusion Restrictions			0.0018			0.0018
Household Fixed Effects	✓	✗	✗	✓	✗	✗
Week Fixed Effects	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓

Standard errors in parentheses
Standard errors clustered at the individual level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Whilst there are definitely limitations to the methods we use to assess the importance of the extensive margin, this evidence suggests that focussing on the intensive margin and accounting for selection by adding observable control variables is a tenable approach.

2.G Earnings: Descriptive Statistics and Pseudo-First-Stage Regressions

Figure 2.15: Distribution of Weekly Earnings

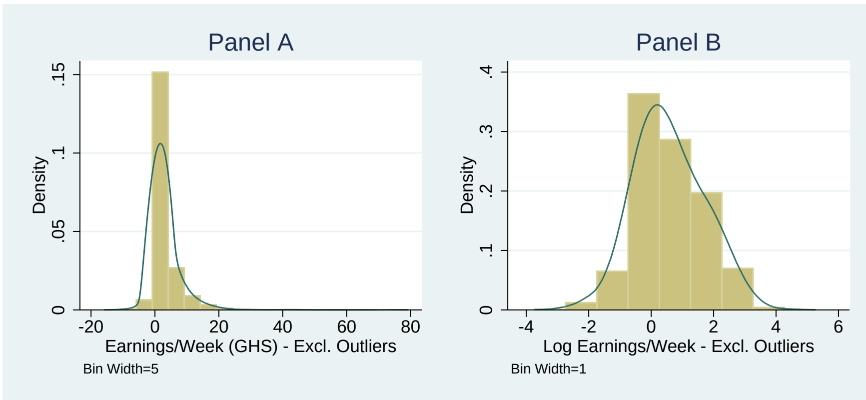


Table 2.36: Earnings Psuedo-First-Stage

	(1)	(2)	(3)	(4)	(5)	(6)
Log of Mean Earnings (others in EA)	-0.1650***		-0.1509*			
	(0.0408)		(0.0831)			
Log of 25th Percentile Earnings (others in EA)		-0.1438**	-0.1242**			
		(0.0598)	(0.0563)			
Log of Median Percentile Earnings (others in EA)		0.0479	0.0705			
		(0.0666)	(0.0712)			
Log of 75th Percentile Earnings (others in EA)		-0.1638***	-0.0744			
		(0.0529)	(0.0745)			
Lagged Log of Mean Earnings (others in EA)				-0.2059***		-0.1016
				(0.0510)		(0.0685)
Lagged Log of 25th Percentile Earnings (others in EA)					-0.1543***	-0.1413***
					(0.0341)	(0.0322)
Lagged Log of Median Percentile Earnings (others in EA)					-0.0157	-0.0009
					(0.0526)	(0.0544)
Lagged Log of 75th Percentile Earnings (others in EA)					-0.1477***	-0.0884**
					(0.0414)	(0.0421)
N	2931	2703	2703	2660	2433	2433
R^2	0.011	0.022	0.024	0.018	0.031	0.032
F-statistic	14.0042	5.0141	4.2989	13.7681	7.7139	5.9584

Standard errors in parentheses
Dependent variable: Log of Earnings/Week
Standard errors clustered two way by individual and week
All columns include week and household fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Female-Male Earnings Gaps, Job Flexibility, and Occupational Selection

3.1 Introduction

The non-pecuniary benefits of certain jobs — such as job security, working conditions, and 'warm glow' — may influence occupational selection, alongside the desire to maximise earnings. Women and men may value these extra benefits differently, and this may help to explain differences in labour market outcomes. For example, it is often argued that women select out of wage work into self-employment because the extra flexibility permits balancing a career with raising children. Using nationally representative data from Ghana we examine the importance of this logic by addressing two related questions. First, we assess what factors determine female-male earnings gaps, focussing in particular on the role of occupational selection. Second, we investigate whether the non-pecuniary benefit of 'job flexibility' draws women into informal low-input jobs more than men.

To analyse the sources of sex earnings differentials, we build on the widely applied methods of Blinder (1973) and Oaxaca (1973). We supplement this analysis using a multinomial logit model to control for selection across multiple sectors, to determine how unobservables may obscure or augment the *observed* earnings differential, as compared to the difference in *potential* earnings. We find that controlling for selection widens the earnings gap to nearly 60 percent for own account self-employed individuals, but leaves the earnings gap amongst the wage-employed small and statistically insignificant.

We focus our analysis of occupational selection in particular on the hypothesis that women have a preference for the 'flexibility' of self-employment, which allows them to meet domestic obligations that are not similarly imposed upon men, such as undertaking household chores and childcare. As discussed in Chapter 1, Ghana provides an especially useful context in which to investigate this question because female labour force participation is so high. Vitally, we examine heterogeneity *within* self-employment, focussing in particular on the technology choice taken over whether to use others' labour. We disaggregate between 'own account' workers — self-employed individuals who work alone — and 'employers' — those who employ others in their business — showing that flexibility may differ between these types of jobs. We also refer to these jobs as 'low-input' and 'high-input' self-employment respectively throughout this chapter.

We use a simple model to outline three potential ways of formalising the concept of 'job flexibility', and predict how it influences occupational selection in the presence of domestic obligations. We couch this framework in terms of time allocation, linking directly with the approach taken in Chapter 2. Firstly, we consider how some jobs may allow for 'multi-tasking', in which market work and domestic work may be undertaken concurrently. Secondly, we discuss the implications of certain jobs, especially in the formal sector, having minimum working hours. Thirdly, we build a model in which individuals face costs for adjusting their work hours, which depend on their chosen occupation. Such adjustment costs may be important when domestic obligations increase suddenly, for example, when a family member becomes ill. We consider the plausibility of these three models in the context of Ghana, and ultimately conclude that the multi-tasking and minimum hours stories are the most important.

To assess the implications of these models empirically, we interrogate the selection equations used to adjust our earnings decompositions, thus treating occupational choice as a discrete multinomial choice problem. We proxy for domestic obligations using the demographic characteristics of the household, and investigate whether this heterogeneity across households affects female and male occupational selection differently. We find that a 1 standard deviation increase in the dependency ratio increases the probability of participation in own account self-employment by 3 percent for women, but only 0.8 percent for men. Conversely, there are no robust positive effects on participation in high-input self-employment or wage-employment from an increase in the dependency ratio for either sex.

Household demographics may be endogenous to occupational selection, insofar as there are unobservable factors that drive decisions about both family structure and employment. We address this issue by developing a new estimator, which incorporates the logic of Altonji et al. (2005) into a discrete choice problem using maximum simulated likelihood. The main idea is to use selection (of the endogenous variable) on *observables* as a guide to selection on *unobservables*. Although this method has been widely used on continuous and binary outcome variables, we believe this is the first attempt to apply it to a multinomial problem.

This chapter proceeds as follows. In Section 3.2 we review the related literature. In Section 3.3, we outline three simple approaches relating occupational choice to job flexibility and individuals' domestic obligations. In Section 3.4 we describe our data. In Section 3.5 we outline our econometric approach, explaining how we decompose sex earnings gaps and investigate the determinants of occupational selection. In Section 3.6 we report our main results, assessing the impact of the potential endogeneity of household demographics, using our new estimator based on the logic of Altonji et al. (2005). In Section 3.7 we conduct some other useful robustness checks, and consider possible heterogeneity in our results. In Section 3.8 we conclude.

3.2 Related Literature

3.2.1 Decomposing Female-Male Earnings Gaps

There is a large literature, building on the techniques developed by Blinder (1973) and Oaxaca (1973), decomposing sex earnings gaps into 'explained' and 'unexplained' components, especially for developed countries (Psacharopoulos and Tzannatos, 1992; Horton, 1996; Appleton et al., 1999; Fortin et al., 2011). The explained component, also known as the 'composition effect', relates to the differences in earnings that can be accounted for by differences in individual characteristics such as education, skills, and access to capital.[1] The unexplained component, also known as the 'wage structure effect', measures the difference in returns to these individual characteristics. This may result from discrimination in the labour market, or unobserved aspects of heterogeneity.

[1] Levels of human capital between women and men may differ due to heterogeneity in innate skills or preferences. However, 'pre-market discrimination', which restricts female access to education and training, may also account for some of this differential.

To reduce the ambiguity in the content of the wage structure effect, economists have attempted to control for individuals' selection into different sectors of the labour market, beginning with Reimers' (1983) study into labour market discrimination against Hispanic and black men in the United States. Following Heckman (1979), Reimers notes that the sample of individuals for which wage data is available is a selected sample, insofar as participation in wage-employment is non-random. As such, correlation between the unobservables affecting participation and the unobservables affecting wages biases estimates of the wage structure effect, and drives a wedge between the *observed* wage gap and the gap in wage *offers*.

A number of studies have followed Reimers (1983), modelling the selection process as a binary problem, with individuals being categorised as either in or out of employment. For example, Deininger at al. (2013), in their comparison of gender and caste discrimination in rural India, use a simple probit equation including exclusion restrictions based on household structure and land ownership to control for selection into the wage sector and adjust their earnings decompositions. They find that, after controlling for selection, the unexplained component of the caste wage gap is statistically insignificant. However, wage discrimination by sex remains high, especially in the agricultural sector.

Given the importance of informal sectors in developing countries, some authors have treated labour market status as a multinomial problem, in order to better correct earnings decompositions for selection (Glick and Sahn, 1997; Appleton et al., 1999; Tansel, 2001; Ben Yahmed, 2013). Appleton et al. (1999), for example, focus on individuals' selection between three sectors — private wage-employment, public sector wage-employment, and not wage-employed. Correcting their decompositions of earnings differentials in Côte d'Ivoire, Ethiopia, and Uganda using Lee's (1983) method, they find that controlling for selection significantly reduces the sex earnings gap in both public and private wage-employment.

We believe, however, that analyses of sex earnings gaps within *informal* sectors of the economy are relatively rare, especially for Africa. This motivates our focus on self-employed workers in Ghana.

3.2.2 Sex-Specific Occupational Choice and Non-Wage Job Characteristics

There is significant evidence from developed countries arguing that the selection forces driving women and men into self-employment are different, suggesting in particular that female participation in entrepreneurship is driven by a demand for non-wage job characteristics. For example, Lombard (2001) uses a structural approach to disaggregate the drivers of occupational choice into differences in earnings and differences in other job attributes between wage- and self-employment. Using data on married women from the United States, she shows that women's chances of participating in self-employment rise not only with potential earnings in that sector, but also with demand for a non-standard work week and demand for flexibility to vary one's work schedule. The latter two attributes are also found to be strongly linked to the demographic composition of the household, especially the number of children present. Estimating separate selection-corrected wage equations for men and women on United States Census data, Clain (2000) shows that the difference in potential earnings between wage- and self-employed women is far greater than difference between wage- and self-employed men, and that selection on the basis of non-wage job attributes can help explain these patterns.

A recent Ghanaian literature has emerged, which also explores the links between fertility, education, and female labour force participation. Taken together, these studies show that education has positive effects on women's involvement in the labour market, for both wage-employment and self-employment (Sackey, 2005). However, the effects of fertility appear to be far more mixed. Whilst having extra children increases the chances of women participating in income-generating activities in general, it emerges that similar effects prevail for men (Ackah et al., 2009; Baah-Boateng et al., 2013). Moreover, fertility only appears to increase the likelihood of doing non-farm self-employment when it is undertaken as a secondary activity alongside other work in agriculture or wage-employment (Heintz and Pickbourn, 2012; Ackah, 2013). In part, we believe these equivocal results stem from the fact that most of these studies try and reduce occupational selection down to a series of binomial choices — such as between participating in self-employment or not — rather than allowing for the multinomial choice between three or more job types. We address this shortcoming in the current chapter.

Problematically, many studies do not explicitly allow for the fact that occupational choice may be endogenous to the factors that drive demand for non-wage job attributes, such as household structure, number of children, and marital status. For example, Becker (1985) argues that occupational selection affects family stability and hence the likelihood of having children, insofar as higher earnings make marriage, as a source of income, relatively less desirable. Also, Rosenzweig and Wolpin (1980) argue that children (or indeed other family members) may be added to increase the earnings of the household at large, especially if individuals work in jobs with low returns. Additionally, occupation may be correlated with a wide range of variables related to family planning, such as knowledge and understanding of contraception, which may not be easily observed and measured.

Some economists have used natural experiments to try and address these endogeneity concerns. For example, Angrist and Evans (1998) use an instrumental variable approach to assess the impact of family size on individuals' labour supply, exploiting exogenous variation in child sex and parental preferences for a mixed sibling-sex composition. In similar vein, the birth of twins may also be treated as a shock to fertility, which generates an exogenous change in household demographics (Rosenzweig and Wolpin, 1980; Bronars and Grogger, 1994). However, although these techniques have been applied to the labour supply decision, their use in the literature on occupational selection, with multiple employment sectors, is limited.

Others have used panel techniques to try and examine how potentially endogenous variables related to household demographics, may influence transitions into self-employment (Evans and Leighton, 1989; Fajnzylber et al., 2006). Focussing on white married women in the United States, Wellington (2006) examines the impact of family size on individuals' chances of being self-employed. Estimating both a cross-section model, and a 'longitudinal' model, in which time-invariant factors are absorbed, she finds that the impact of extra children is somewhat mixed and dependent on women's level of education.

We hope to complement the existing literature by adopting a new approach to assessing endogeneity and by allowing for *multinomial* choice across different occupations *and* technologies.

3.3 Formalising Job Flexibility and Occupational Choice

The notion of 'flexibility' is often invoked in the empirical literature to explain why certain individuals, especially women, may select into self-employment or other informal labour market activities. This concept, however, is rarely formalised explicitly. Starting with a basic framework, we suggest three possible ways of thinking about job flexibility, and consider what this implies for occupational choice. Whilst we do not formally test between these different models, it is useful to fix ideas to help interpret the data we present for the Ghanaian context in our descriptive statistics and main results.

3.3.1 Basic Framework

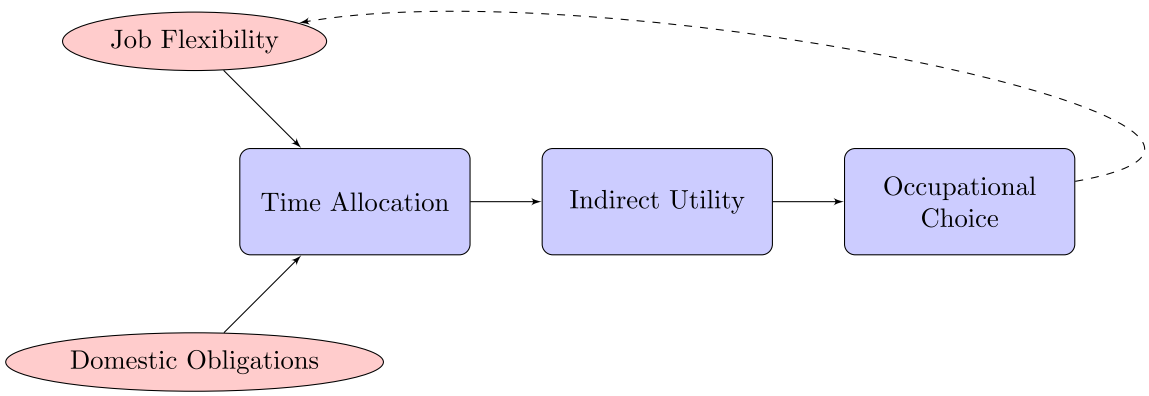
Our aim in this section is to construct a plausible framework for occupational choice. To do this, we split individuals' decisions into two steps.

First, we write down a simplified version of the time allocation model from Chapter 2. We add individual heterogeneity — in terms of domestic obligations — and job flexibility into the model by altering to the set up of the constraints. This allows us to calculate indirect utility for each type of individual working in each different type of job.

In the second step, individuals choose the occupation that gives them the most welfare, *given* the results of their time allocation problem.

The overall structure of this framework is shown in Figure 3.1.

Figure 3.1: Time Allocation and Occupational Choice



In the time allocation problem, individuals derive utility from market and domestic goods, x_m and x_d respectively. These are produced separately through market work (t_m), with return w, and domestic work (t_d), with return ρ. Individuals are only endowed with their time, which adds up to 1. For simplicity, we assume Cobb-Douglas preferences over domestic and market goods.

Taking these components of the model together, the maximisation problem for the individual can be written:

$$\max_{x_m, x_d} \; U = u(x_m, x_d) = x_m^{\beta} x_d^{(1-\beta)} \tag{3.1}$$

$$\text{subject to } t_m + t_d = 1 \tag{3.2}$$

$$x_m = w t_m \tag{3.3}$$

$$x_d = \rho t_d \tag{3.4}$$

$$t_k \geq 0 \; \forall k = m, d. \tag{3.5}$$

We capture individual heterogeneity in terms of domestic obligations, by adding a lower bound on the domestic goods that must be provided. As we now know from Chapter 2, this is something of a simplification given that individuals, especially self-employed men, are somewhat willing to substitute between market and domestic work week-by-week. However, the time-horizon that is relevant to occupational choice is likely to be longer than a week. Thus, we believe this captures the imperative to provide domestic goods in a plausible way.

$$x_d \geq \gamma \tag{3.6}$$

As such, individuals with greater domestic obligations are potentially more constrained in their allocation of time and their choice of goods, which may leave them worse off. We now modify the set of constraints to consider the types of occupations that individuals with different domestic obligations would select.

3.3.2 Multi-Tasking

Certain types of self-employment activities may be undertaken concurrently with domestic work. For example, self-employed retailers may still be able to run their stall, whilst watching over their children. We anticipate, however, that multi-tasking in this way comes at the expense of productivity in market work activities.

To formalise this, we add a parameter to the basic model, $0 < \pi \leq 1$, which captures these *two* effects of job flexibility. Firstly, we assume that π governs the extent to which domestic work draws down the endowment of time. As such, π measures the proportion of time spent on domestic work, which can simultaneously be devoted to market work. Secondly, we wish to incorporate the idea that more flexible jobs have lower returns. To do this, we write w as a positive linear function of π, such that $w = \hat{w}\pi$, where the parameter $\hat{w}$ measures how strongly job flexibility and market returns are associated.[2] Since π is actually *lower* for jobs with higher flexibility, we think of π as 'job rigidity'.

The initial time allocation problem for the individual may now be written:

$$\max_{x_m, x_d} \; U = u(x_m, x_d) = x_m^{\beta} x_d^{(1-\beta)} \tag{3.7}$$

$$\text{subject to } t_m + \pi t_d = 1 \tag{3.8}$$

$$x_m = \hat{w}\pi t_m \tag{3.9}$$

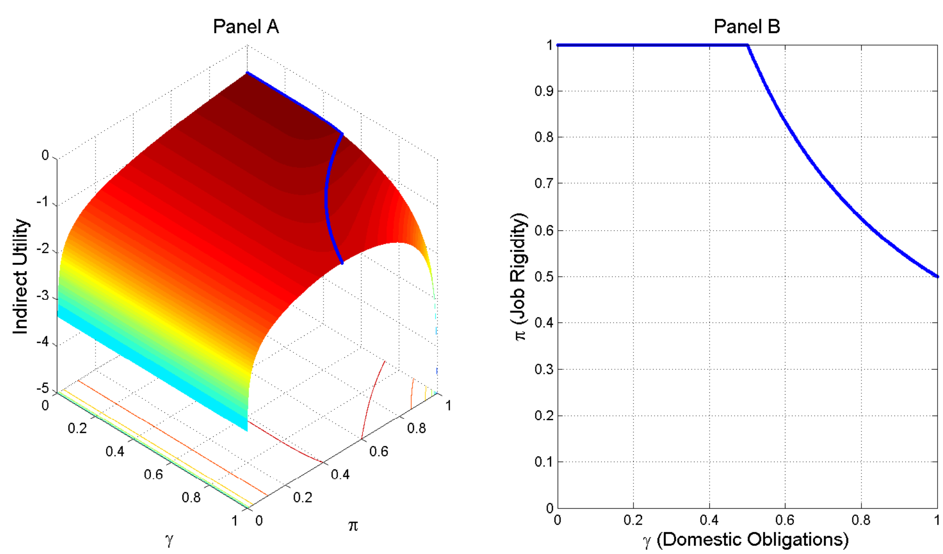
$$x_d = \rho t_d \tag{3.10}$$

$$t_k \geq 0 \; \forall k = m, d. \tag{3.11}$$

$$x_d \geq \gamma \tag{3.12}$$

We derive analytical expressions for indirect utility in the presence of multi-tasking in Appendix 3.A.1.

Figure 3.2: Occupational Selection in a Multi-Tasking Model



[2]The findings of the model are not sensitive to the assumption of a linear functional form.

To illustrate the main intuition of the model, we plot indirect utility for individuals with different levels of domestic obligations, γ, and job rigidity/flexibility, π. This is shown in Panel A of Figure 3.2.[3] Indirect utility is, on average, decreasing in γ, as anticipated. This is because individuals with greater domestic obligations are more constrained in their time-use, and hence their consumption of goods. For individuals with low domestic obligations, indirect utility is always increasing in π. Intuitively, individuals with fewer domestic obligations are willing to endure extra job rigidity in pursuit of higher returns to market work. However, for individuals with high domestic obligations, indirect utility has an 'inverted-U' shape in π. At first, jobs with higher π raise welfare, because of the greater returns to market work. However, a higher level of π makes it more likely that the domestic obligations constraint will bind, which eventually reduces welfare. The higher the level of γ, the lower the level of job rigidity π needed for the domestic obligations constraint to start binding and bringing down indirect utility.

The shape of the indirect utility function, which results from our initial time allocation problem, also tells us the occupations that individuals with different levels of domestic obligations would choose. In particular, an individual with a given level of γ, will select the job type π that yields the highest indirect utility. This is shown by the blue ridge in Panel A of Figure 3.2, and then recast as the optimal job schedule in two dimensions in Panel B. This clarifies the relationship between between γ and π — individuals with sufficiently high levels of domestic obligations choose less rigid/more flexible jobs.

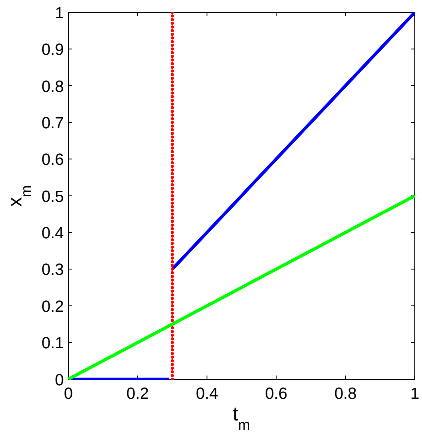
3.3.3 Minimum Hours

Alternatively, flexibility may relate to the *minimum* number of hours of market work that must be supplied in order to work in a particular occupation. Formal wage sector jobs tend to require a certain number of hours to be committed each week, whereas the choice over how much to work in an informal self-employment job may be far less constrained.

To illustrate this idea, we show two possible earnings schedules in Figure 3.3. The blue earnings function, which represents a typical formal wage-employment job, is characterised by greater market work productivity (higher gradient), but also a lower bound on the hours that must be worked in order to participate. In contrast, the green earnings function, which may approximate informal self-employment work, has lower market work productivity, but any number of work hours is possible.

[3] We set $\rho = \hat{w} = 1$ and $\beta = 0.7$. As we show in Appendix 3.A.1, β must be sufficiently high for the domestic obligations constraint to bind, and therefore affect indirect utility in the model.

Figure 3.3: Possible Wage Schedules with Minimum Work Hours



As before, we build this into our basic framework by changing the set of constraints in the model. Job flexibility once again has *two* effects on the time allocation problem, which we capture with a parameter τ. Firstly, τ places a lower bound on the time that must be spent doing market work in order to work in that job. However, we also incorporate the idea that jobs with minimum hours may have higher returns by relating w to τ. Again we adopt a simple linear functional form such that $w = k\tau$, where the parameter k captures the association between minimum hours and the earnings rate.[4] As such, τ captures *both* of the relevant aspects of job flexibility. Once again, by writing down the time allocation problem and finding out the indirect utility for different values of γ and τ, we can recover the types of jobs that individuals with different domestic obligations would prefer. We also understand τ in terms of 'job rigidity', since a *greater* value of τ implies a *less* flexible job.

The maximisation problem may now be written:

[4]Once again, the main intuition of the model is robust to different functional form assumptions about this relationship.

$$\max_{x_m, x_d} \; U = u(x_m, x_d) = x_m^{\beta} x_d^{(1-\beta)} \tag{3.13}$$

$$\text{subject to } t_m + t_d = 1 \tag{3.14}$$

$$x_m = k\tau t_m \text{ if } t_m \geq \tau \tag{3.15}$$

$$x_d = \rho t_d \tag{3.16}$$

$$t_k \geq 0 \; \forall k = m, d. \tag{3.17}$$

$$x_d \geq \gamma \tag{3.18}$$

$$\tag{3.19}$$

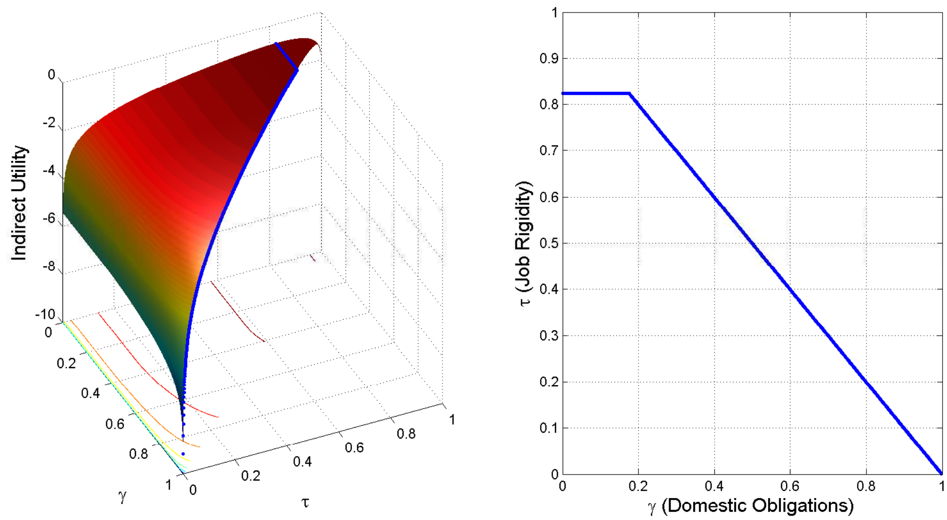
We derive the analytical expressions for indirect utility that emerge from this problem in Appendix 3.A.2.

Plotting the indirect utility function for possible values of γ and τ allows us to link domestic obligations to occupational selection.[5] This is shown in Panel A of Figure 3.4. The surface only goes up to the line $\tau = 1 - \frac{\gamma}{\rho}$, because it is impossible for individuals with high levels of γ to choose jobs with a high level of τ — the domestic obligations constraint and the lower bound on minimum hours could not be met simultaneously. Aside from this restriction at $\tau = 1 - \frac{\gamma}{\rho}$, indirect utility has an 'inverted-U' shape in job rigidity τ, for *all* levels of domestic obligations γ. At first, individuals are better off with higher τ as this increases their returns to market work, and hence their potential consumption of market goods. However, if τ becomes too high, the returns to market work are not sufficient to compensate individuals for the restricted choice they have over their allocation of time, and hence mix of consumption of market and domestic goods. Thus, if τ becomes too high, extra job rigidity begins to reduce welfare.

Individuals select the occupation that maximises their indirect utility, after having solved the time allocation problem. The optimum level of τ for each individual is shown with the blue ridge in Panel A of Figure 3.4. This is plotted in two dimensions in Panel B. Individuals with low domestic obligations are unconstrained in their occupational choice, so they pick the level of τ at the top of the 'inverted-U'. However, for individuals with higher levels of γ, they simply select the occupation with the highest level of τ, in which they can still meet their domestic obligations. This runs along the line $\tau = 1 - \frac{\gamma}{\rho}$.

[5] As before, we set $\beta = 0.7$ and $\rho = k = 1$.

Figure 3.4: Occupational Selection in a Minimum Work Hours Model



3.3.4 Adjustment Costs

Finally, we consider the possibility that it may be easier in some jobs than others to adjust to *shocks* to domestic obligations, such as caring for sick family members. We are able to capture this intuition by adding a stochastic element to a simplified version of our basic model.

Firstly, we assume that individuals derive utility only from market goods, $x_m = wt_m$. However, the actual number of hours they will be able to work in the market is constrained according to some stochastic parameter which, maintaining notation, we label γ.[6] The parameters of the *ex ante* distribution of γ are labelled θ. Vitally, the individual chooses the amount of market work they wish to undertake, $\hat{t_m}$, *before* the shock to domestic obligations is realised. In the context of wage-employment, we can think of $\hat{t_m}$ as the 'contracted' number of hours. The analogous interpretation for the self-employed may be the number of hours to which the individual has, in some sense, committed prior to working. As such, the initial problem is now one of choosing the optimal level of $\hat{t_m}$, rather than allocating time between market and domestic work per se.

Once again, job flexibility, which we capture with the parameter c, has *two* related components. Firstly, departure from the pre-committed level of market work incurs some convex adjustment penalty to wages, the size of which is determined by c. However, c also determines

[6] We have implicitly made the normalisation $\rho = 1$.

the wage rate that would be paid, *absent* any difference between $\hat{t_m}$ and t_m, which we label w_0. We make the simplifying assumption that w_0 and c are linearly associated through some parameter h, such that $w_0 = hc$. Thus c relates to both the returns to market work and the costs of adjustment when plans must change.

Making the assumption of log utility, we can write the modified individual maximisation problem.[7]

$$\max_{\hat{t_m}} \ \mathrm{E}[U] = \mathrm{E}[\ln(x_m)] = \mathrm{E}[\ln(wt_m)] \tag{3.20}$$

$$\text{subject to } t_m \leq 1 - \gamma \tag{3.21}$$

$$\gamma \sim f(\gamma;\theta) \tag{3.22}$$

$$w = w_0\left(1 - \frac{c}{2}(\hat{t_m} - t_m)^2\right) = hc\left(1 - \frac{c}{2}(\hat{t_m} - t_m)^2\right) \tag{3.23}$$

We can write the problem more succinctly by substituting the constraints into the maximand. This reinforces the idea that the initial problem is now one of choosing a 'contracted' number of hours, $\hat{t_m}$, rather than time allocation per se. After solving this problem, individuals with different domestic obligations, parameterised by θ, can then choose the level of c which gives them the highest level of expected (indirect) utility.

$$\max_{\hat{t_m}} \ \mathrm{E}[U] = \int_{-\infty}^{\infty} \left[\ln\left(hc\big(1 - \frac{c}{2}(\hat{t_m} - t_m)^2\big)t_m \right) \right] f(\gamma;\theta)d\gamma \tag{3.24}$$

In order to further the analytical treatment of this problem, we suggest a functional form for the distribution of γ. We assume that there are just two states of the world. In the 'bad' state, which occurs with probability q, the individual faces a lower bound on domestic work of $\tilde{\gamma}$.[8] In the 'good' state of the world, however, the individual devotes no time to domestic work, such that $\gamma = 0$. We assume that the individual has perfect information about their own values of $\tilde{\gamma}$

[7] We want to set up the model so that individuals are risk averse.

[8] We assume $0 \leq q \leq 1$ and $0 \leq \tilde{\gamma} < 1$.

and q. Summarising this formulation, the distribution of γ may be written:

$$\gamma(q,\tilde{\gamma}) = \begin{cases} \tilde{\gamma} & \text{w.p. } q \\ 0 & \text{w.p. } (1-q) \end{cases} \tag{3.25}$$

As such, individuals' susceptibility to shocks to domestic obligations is captured by q and $\tilde{\gamma}$.

To solve the model, we note that the problem effectively spans two periods. Whilst the level of $\hat{t_m}$ is chosen *before* the value of γ has been realised, the individual chooses t_m *after*. Therefore, in Period 1, the individual chooses their contracted hours, $\hat{t_m}$, on the basis of predictions about domestic obligations (q and $\tilde{\gamma}$), and the characteristics of their job, c. Then, in Period 2, the individual chooses the level of t_m to maximise consumption $x_m = wt_m$ *conditional* on $\hat{t_m}$, as well as the revealed value of γ.

We solve the model using backwards induction. First, we find a 'reaction function' for the optimal level of t_m given the parameters $\Gamma = \{\hat{t_m}, q, \tilde{\gamma}, h, c\}$. We then substitute $t_m(\Gamma)$ back into the original problem, of the form in Equation (3.24) to solve for the optimum $\hat{t_m}^*$ and hence the expected utility, for different levels of the job characteristic and domestic obligation parameters.

A full description of the solution is provided in Appendix 3.A.3.

We plot expected utility for individuals with different levels of domestic obligations and different jobs in Panel A of Figure 3.5. In particular, we focus on heterogeneity in $\tilde{\gamma}$.[9] We hold q constant, setting $q = 0.5$ and $h = 1$, and simulating the model for all possible values of $\tilde{\gamma}$.[10]

Looking to the shape of the surface in Panel A, we can see that expected utility is everywhere decreasing in $\tilde{\gamma}$, as expected. For low values of $\tilde{\gamma}$, expected utility is increasing in c, as this corresponds to higher returns to returns to market work, and hence higher consumption. When domestic obligations are larger, however, expected utility is at first increasing but then *decreasing* in c. Individuals face more variation in their outcomes when $\tilde{\gamma}$ is higher.[11] The extra adjustment

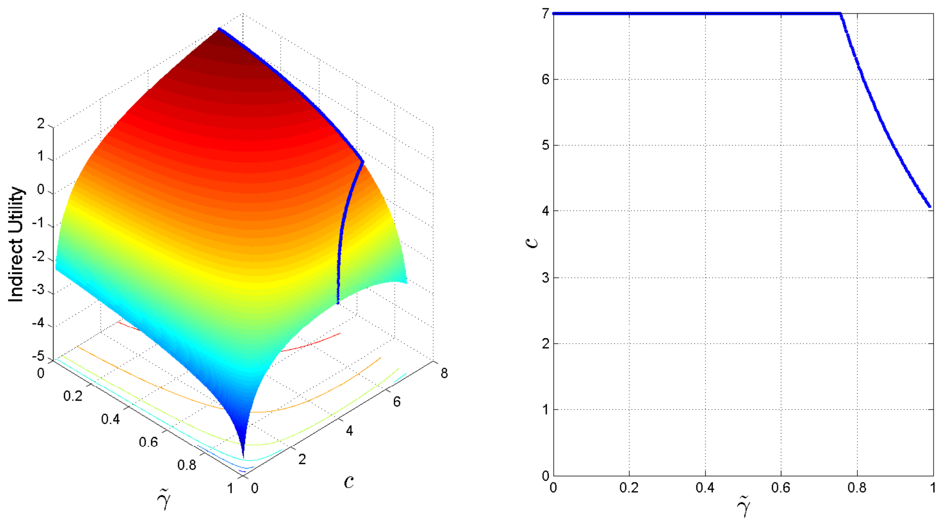
[9]The implications of changing q and holding $\tilde{\gamma}$ constant are somewhat more complex. Overall, it emerges that the optimal level of job rigidity, c, is a non-monotonic function of q. These predictions are discussed in Appendix 3.B.

[10]In order to maintain positive consumption levels, consistent with log utility, we ensure $c < \frac{2}{(t_m - \hat{t_m})^2)}$. We show the results for $0 < c \leq 7$, since as c becomes too large, the model is insoluble for all values of $\tilde{\gamma}$.

[11]In particular, both the mean and the variance of γ are increasing in $\tilde{\gamma}$. $E(\gamma) = q\tilde{\gamma}$, therefore $\frac{\partial E(\gamma)}{\partial \tilde{\gamma}} = q \geq 0$. $\text{Var}(\gamma) = q\tilde{\gamma}^2 - q^2\tilde{\gamma}^2$, therefore $\frac{\partial \text{Var}(\gamma)}{\partial \tilde{\gamma}} = 2\tilde{\gamma}q(1-q) \geq 0$.

penalties that come with an increase in c therefore make the risk averse individuals in the model worse off, in expectation.

Figure 3.5: Occupational Selection in an Adjustment Costs Model — Changing $\tilde{\gamma}$



The expected utility function translates into a model of occupational choice, insofar as individuals with different levels of domestic obligations choose the level of c that maximises their (expected) welfare. This is shown by the blue ridge in Panel A, and then plotted in two dimensions in Panel B of Figure 3.5. Individuals facing a higher potential shock to domestic obligations, $\tilde{\gamma}$, have a lower optimal level of job rigidity, c. When $\tilde{\gamma}$ is low, the expected involuntary adjustment of t_m away from $\hat{t_m}$ is also low. Thus, individuals will expose themselves to higher adjustment penalties, in pursuit of higher initial wages, hc. The converse is true when $\tilde{\gamma}$ is high, and individuals' expected involuntary adjustments of t_m away from $\hat{t_m}$ are also high. It is then optimal to choose occupations with lower c, even if this means foregoing higher initial wages hc.

3.3.5 Theoretical Predictions

In each of our three possible interpretations, there is some range of the parameter values for which individuals with greater domestic obligations optimally choose occupations with more flexibility, even at the expense of reduced market work productivity. Primarily, we want to test this relationship by examining whether individuals from households with more dependent members choose low-input self-employment activities instead of choosing high-input self-employment activities or wage-employment. However, writing down the three stories formally enables us to ascertain which interpretations might be more important for the Ghanaian context. We do this using our descriptive statistics in Section 3.4 and disaggregating *which* dependents matter for occupational choice in Section 3.7.

3.4 Data and Descriptive Statistics

3.4.1 Sample Composition

Our data are taken from the fifth wave of the Ghana Living Standards Survey (GLSS5+), a nationally representative survey sampling approximately 8,000 households across the ten regions of Ghana.[12] Information on every individual within each surveyed household was collected, either through direct interview of the given respondent, or from the household head. After merging different sections of the survey, we have data on 37,098 individuals, 20,651 of whom are working age (15–65 years).

We show the regional distribution of our sample in Table 3.1, splitting up location by rural and urban areas. We control for regional and rural-urban variation, where possible, to reduce the impact of unobserved heterogeneity on our estimates.

Besides earnings, the key outcome variable on which our analysis focusses is occupational choice. We show the breakdown of occupation, by sex, in Table 3.2. We restrict our analysis to primary occupation, defined as the job that the individual had undertaken most recently, and spent most time doing. Importantly, we separate out the self-employed sample into enterprises that operate with and without others' labour. We label those that do not use labour besides their own 'Own Account' or 'low-input' self-employed, whereas those which use others' labour are labelled 'Employer' or 'high-input' self-employed.[13] This is a broad dimension on which to split the sample, but we believe it captures a key technology choice taken by entrepreneurs, particularly because the employment of labour is correlated with the use of other factors, such as capital (Woodruff, 2006).[14] By dividing the sample in this way, we are able to investigate heterogeneity *within* the self-employment sector, especially as job flexibility may differ between low- and high-input activities.

[12]The collection of GLSS5+ was undertaken by the Ghana Statistical Service (GSS) in conjunction with the World Bank. The survey is unique in focussing especially on non-farm household enterprises.

[13]The employers in our sample use labour from inside or outside the household for their businesses.

[14]The sample correlation coefficient between log labour and log capital amongst non-farm enterprises is 0.31.

Table 3.1: Sample Location

Region	Location					
	Rural		Urban		Total	
	N	%	N	%	N	%
Western	1138	5.51	655	3.17	1793	8.68
Central	860	4.16	517	2.50	1377	6.67
Greater Accra	278	1.35	2502	12.12	2780	13.46
Volta	1151	5.57	442	2.14	1593	7.71
Eastern	1272	6.16	671	3.25	1943	9.41
Ashanti	1760	8.52	1695	8.21	3455	16.73
Brong Ahafo	1112	5.38	678	3.28	1790	8.67
Northern	1818	8.80	549	2.66	2367	11.46
Upper East	1560	7.55	230	1.11	1790	8.67
Upper West	1616	7.83	147	0.71	1763	8.54
Total	12565	60.84	8086	39.16	20651	100.00

Sample of individuals of working age (15–65)

Table 3.2: Occupational Choice by Sex

Occupation	Sex					
	Female		Male		Total	
	N	%	N	%	N	%
Wage Employed	658	3.19	1768	8.56	2426	11.75
SE - Agriculture	1181	5.72	2736	13.25	3917	18.97
Non-Farm SE (Own Account)	2640	12.78	813	3.94	3453	16.72
Non-Farm SE (Employer)	807	3.91	526	2.55	1333	6.45
Unemployed	467	2.26	347	1.68	814	3.94
Out of LF/Other	5173	25.05	3535	17.12	8708	42.17
Total	10926	52.91	9725	47.09	20651	100.00

Sample of individuals of working age (15–65)

Overall, 64 percent of working age men participate in the labour market, compared with 51 percent of women. This difference is somewhat larger than expected, mainly because of the composition of the 'Out of LF/Other' category. This includes not only those who are not working and not searching for work, but also apprentices, domestic workers (that is, 'househelp'), and unpaid family workers. Whether or not this latter group count as in or out of the labour force differs from survey to survey, but following our definition of 'employment' set out in Chapter 1 we classify those workers who do not receive money income as not employed (Sen, 1975).

The most striking differences in terms of occupational selection arise in low-input non-farm self-employment, where 76 percent of the participants are female. The sample of employers, however, is far more balanced where only 61 percent of the sample are women. This fact, in itself, suggests that the selection forces behind female and male participation in low- and high-input self-employment may differ. Also, women comprise just 27 percent of the wage-employed, further emphasising a disparity in the factors driving occupational choice.

In Tables 3.3 and 3.4 we report detailed summary statistics for women and men separately.

Table 3.3: Summary Statistics (Females)

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Age (Years)	10926	33.22	13.56	15.00	22.00	31.00	43.00	65.00
Education (Years)	10926	4.86	4.90	0.00	0.00	5.00	9.00	18.00
Tenure (Months)	10926	64.22	107.97	0.00	0.00	0.00	84.00	540.00
Married? (1=Y, 0=N)	10926	0.49	0.50	0.00	0.00	0.00	1.00	1.00
Household Size	10926	5.93	3.43	1.00	4.00	5.00	7.00	29.00
Dependency Ratio	10926	0.85	0.75	0.00	0.33	0.67	1.00	7.00
Active Ratio in HH (excl. Individual)	10926	0.31	0.37	0.00	0.13	0.25	0.40	6.00
HH Land (Acres)	10926	8.60	33.60	0.00	0.00	0.00	6.00	320.00
HH Owns Land? (1=Y, 0=No)	10926	0.47	0.50	0.00	0.00	0.00	1.00	1.00
Farmer Father? (1=Y, 0=N)	10926	0.58	0.49	0.00	0.00	1.00	1.00	1.00
Professional Father? (1=Y, 0=N)	10926	0.05	0.21	0.00	0.00	0.00	0.00	1.00
Service Sector Father? (1=Y, 0=N)	10926	0.10	0.31	0.00	0.00	0.00	0.00	1.00
Farmer Mother? (1=Y, 0=N)	10926	0.50	0.50	0.00	0.00	0.00	1.00	1.00
Professional Mother? (1=Y, 0=N)	10926	0.01	0.10	0.00	0.00	0.00	0.00	1.00
Service Sector Mother? (1=Y, 0=N)	10926	0.02	0.14	0.00	0.00	0.00	0.00	1.00
HH Unearned Income? (1=Y, 0=N)	10926	0.01	0.09	0.00	0.00	0.00	0.00	1.00
Observations	10926							

Sample of individuals of working age (15–65)

Focussing on human capital, whilst age is, as expected, approximately balanced between the samples of women and men, there is some male advantage in terms of education and tenure at the median.

The key household demographic variables, which we use to proxy domestic obligations, are the individual's marital status and household size, as well as the 'Dependency Ratio'.[15] The

[15]We also show the 'Active Ratio' in our descriptive statistics, which calculates the ratio of individuals in the

Table 3.4: Summary Statistics (Males)

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Age (Years)	9725	32.58	13.85	15.00	20.00	30.00	43.00	65.00
Education (Years)	9725	6.69	5.08	0.00	0.00	8.00	10.00	18.00
Tenure (Months)	9725	100.21	127.97	0.00	0.00	40.00	180.00	540.00
Married? (1=Y, 0=N)	9725	0.44	0.50	0.00	0.00	0.00	1.00	1.00
Household Size	9725	5.70	3.46	1.00	3.00	5.00	7.00	29.00
Dependency Ratio	9725	0.69	0.66	0.00	0.17	0.50	1.00	7.00
Active Ratio in HH (excl. Individual)	9725	0.28	0.35	0.00	0.00	0.20	0.33	5.00
HH Land (Acres)	9725	8.20	31.43	0.00	0.00	0.00	6.00	320.00
HH Owns Land? (1=Y, 0=No)	9725	0.48	0.50	0.00	0.00	0.00	1.00	1.00
Farmer Father? (1=Y, 0=N)	9725	0.50	0.50	0.00	0.00	1.00	1.00	1.00
Professional Father? (1=Y, 0=N)	9725	0.05	0.22	0.00	0.00	0.00	0.00	1.00
Service Sector Father? (1=Y, 0=N)	9725	0.10	0.30	0.00	0.00	0.00	0.00	1.00
Farmer Mother? (1=Y, 0=N)	9725	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Professional Mother? (1=Y, 0=N)	9725	0.01	0.10	0.00	0.00	0.00	0.00	1.00
Service Sector Mother? (1=Y, 0=N)	9725	0.02	0.14	0.00	0.00	0.00	0.00	1.00
HH Unearned Income? (1=Y, 0=N)	9725	0.02	0.12	0.00	0.00	0.00	0.00	1.00
Observations	9725							

Sample of individuals of working age (15–65)

dependency ratio is calculated in the usual way, with ‘dependents’ defined as anyone aged under 15 or over 65:

$$\text{Dependency Ratio} = \frac{\text{No. Dependents}}{\text{Household Size} - \text{No. Dependents}} \tag{3.26}$$

Our equations of occupational choice also include information on household land ownership, parental occupation, and unearned household income, coming from social security, retirement benefits, or dowry or inheritance payments.[16]

In Appendix 3.C we report further summary statistics broken down by sex *and* occupation.

3.4.2 Raw Earnings Gaps

Our measures of earnings are calculated from the amount of money an individual receives from a job, including bonuses, commissions, allowances, or tips. We use this definition of earnings, as opposed to measures of enterprise value-added to reduce bias in cross-sectoral comparisons of earnings. We convert our measures of earnings from Second Ghana Cedis (GHC) to 2005 United

household that do and do not participate in the labour force.

[16]For the binary variables reported in these tables, we can read the mean as the proportion of individuals for which the variable is 1.

States Dollars (USD).

Total monthly earnings are shown by sex and occupation in Table 3.5.

Table 3.5: Monthly Earnings Gaps

	Male				Female			
	N	Mean	S.Dev.	Median	N	Mean	S.Dev.	Median
Full Sample	5092	80.26	99.26	47.74	5110	55.57	82.52	28.65
Wage Employed	1728	113.52	110.04	87.65	645	91.73	98.70	55.09
Non-Farm SE (Own Account)	788	84.59	108.73	47.74	2615	54.11	79.35	28.65
Non-Farm SE (Employer)	510	106.55	115.98	66.11	791	72.80	103.04	38.20
Observations	5092				5110			

Outliers trimmed at the 1^{st} and 99^{th} percentiles
Sample of individuals of working age (15–65)

As expected, for the three sectors shown, median monthly earnings are highest in wage-employment, and lowest in low-input self-employment. Also, for the full sample, and all three occupations shown, median monthly earnings are higher for men than women.

We adjust our earnings measures for the hours actually worked per month, and report the summary statistics in Table 3.6. Some economists have argued it is inappropriate to account for working hours in this way because it may not be possible for casual workers to scale up the time they work in a given month (Günther and Launov, 2012). However, given our focus on job flexibility, where work hours may be an important component, we believe this adjustment is important.

Table 3.6: Hourly Earnings Gaps

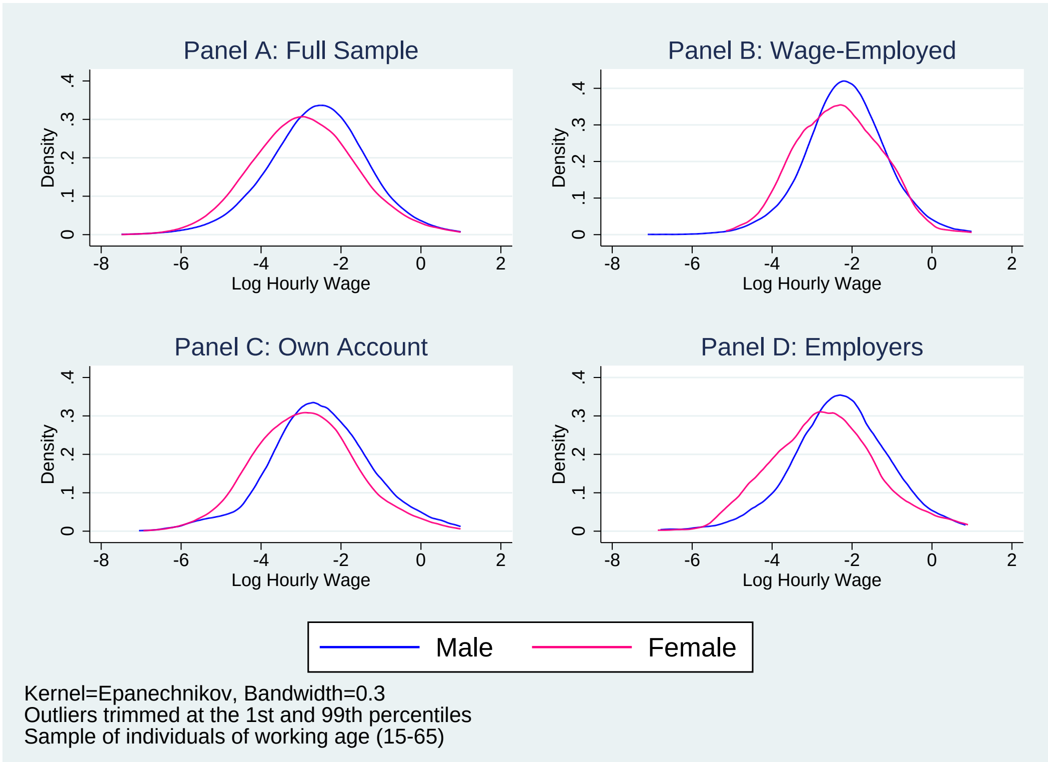
	Male				Female			
	N	Mean	S.Dev.	Median	N	Mean	S.Dev.	Median
Full Sample	5110	0.54	0.89	0.27	5090	0.44	0.86	0.18
Wage Employed	1726	0.68	0.95	0.41	641	0.60	0.96	0.34
Non-Farm SE (Own Account)	795	0.61	1.04	0.26	2598	0.44	0.85	0.18
Non-Farm SE (Employer)	522	0.67	0.98	0.34	797	0.57	1.09	0.23
Observations	5110				5090			

Outliers trimmed at the 1^{st} and 99^{th} percentiles
Sample of individuals of working age (15–65)

Virtually the same patterns emerge in the summary statistics for hourly earnings, as compared to monthly earnings. At the median, men earn more than women across all sectors. The wage-employed enjoy the highest earnings, whilst the low-input self-employed earn the least.

To visualise the sex earnings gap across the entire distribution of earnings, we plot kernel densities for women and men separately in Figures 3.6. We first adjust our figures for hourly earnings for cross-regional price differences using the local Consumer Price Index (CPI), and then take logs.

Figure 3.6: Log of Hourly Earnings by Sex



3.4.3 Job Characteristics

We present descriptive statistics on the characteristics of each occupation for two reasons. Firstly, this enables us to test our prior that low-input self-employment jobs are the most flexible type of work. Secondly, we can consider which dimensions of flexibility are most important in the Ghanaian data, providing some guidance over which of the interpretations set out in Section 3.3 are most suitable.

3.4.3.1 Job Location

In Figure 3.7 we show the proportion of individuals in wage-employment, low-input self-employment, and high-input self-employment, working in particular locations. The blue bars represent individuals working at home or on their own land, the red bars represent individuals working without a fixed location, such as street vendors or those engaged in transport services, whilst the green bars represent individuals working in a fixed location away from home, such as an office, workshop, or stationary stall.

As expected, the wage-employed appear to work mainly in fixed locations away from home. There also appear to be small differences between low- and high-input self-employment activities. In particular, the proportion of individuals working in fixed locations locations away from home appears to be somewhat larger for the entrepreneurs that employ labour besides their own. Insofar as job location is a sufficiently relevant proxy, these trends confirm our priors that own account self-employment work may be more flexible.

Figure 3.7: Primary Work Location by Occupation

3.4.3.2 Time-Use

In Figure 3.8, we show the time that individuals devote to activities outside of market work each week, in different occupations. We separate the time devoted to caring for other household members (green bars) from the time spent on other household chores, such as washing, cleaning, and running errands (red bars). Total time spent on *all* domestic work is shown by the blue bars. We show the median time spent on domestic work in Panel A, and then separate out the means for all domestic work, non-care work, and care work in Panels B, C, and D respectively, allowing us to add error bars to our estimates.

Figure 3.8: Time Spent on Domestic Work by Occupation

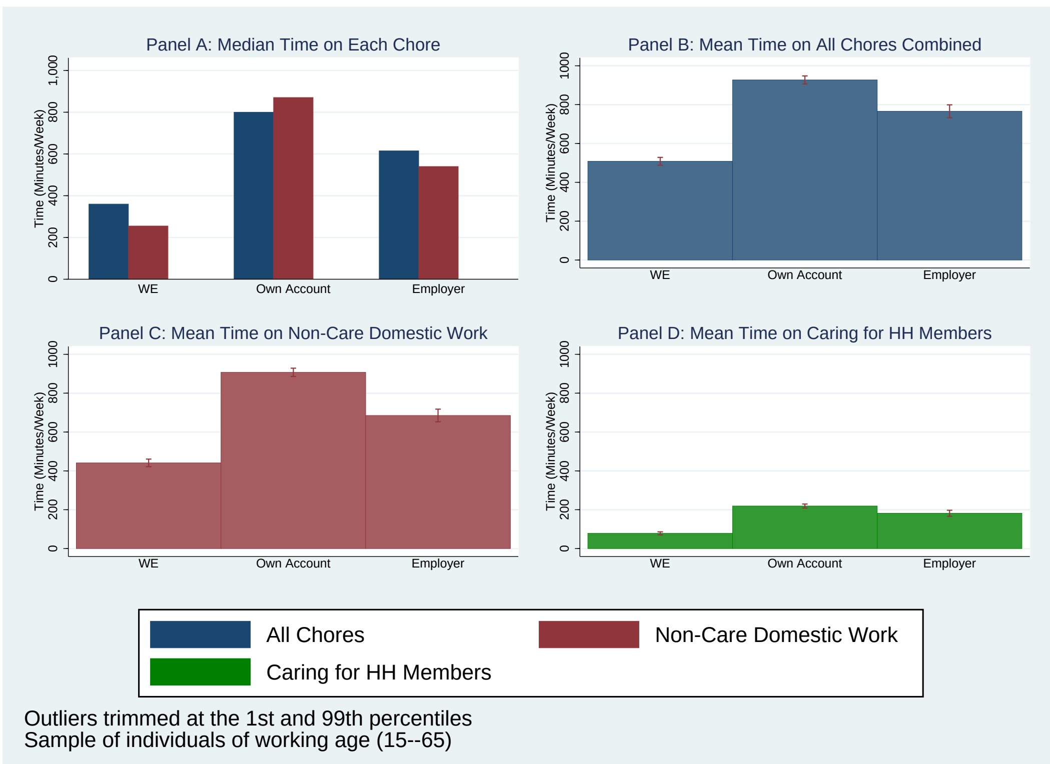
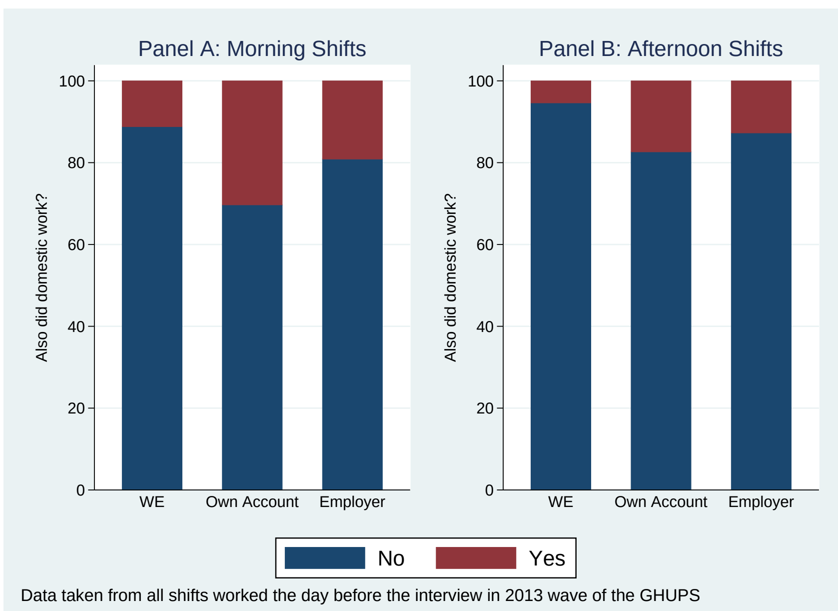


Figure 3.8 shows that the wage-employed do the least domestic work, whilst the low-input self-employed individuals do the most. This in consistent with the notion that job flexibility may be highest for entrepreneurs that do not employ labour besides their own. We also see that caring for other household members makes a relatively small contribution to the total time spent on domestic work in our sample. Since care may be the most unpredictable type of domestic work, this suggests that the model of adjustment costs set out in Section 3.3.4 has less importance for the Ghanaian context.

Whilst the prevalence of multi-tasking is not measured in the GLSS5+ data, the 2013 wave of the GHUPS contains a small time-use module, where individuals were able to *list* all the activities they undertook during the morning, afternoon, and evening. We show the proportion of morning and afternoon market work shifts in which the respondent also reported doing domestic work in Panels A and B of Figure 3.9 respectively.

Figure 3.9: Prevalence of Multi-Tasking by Occupation

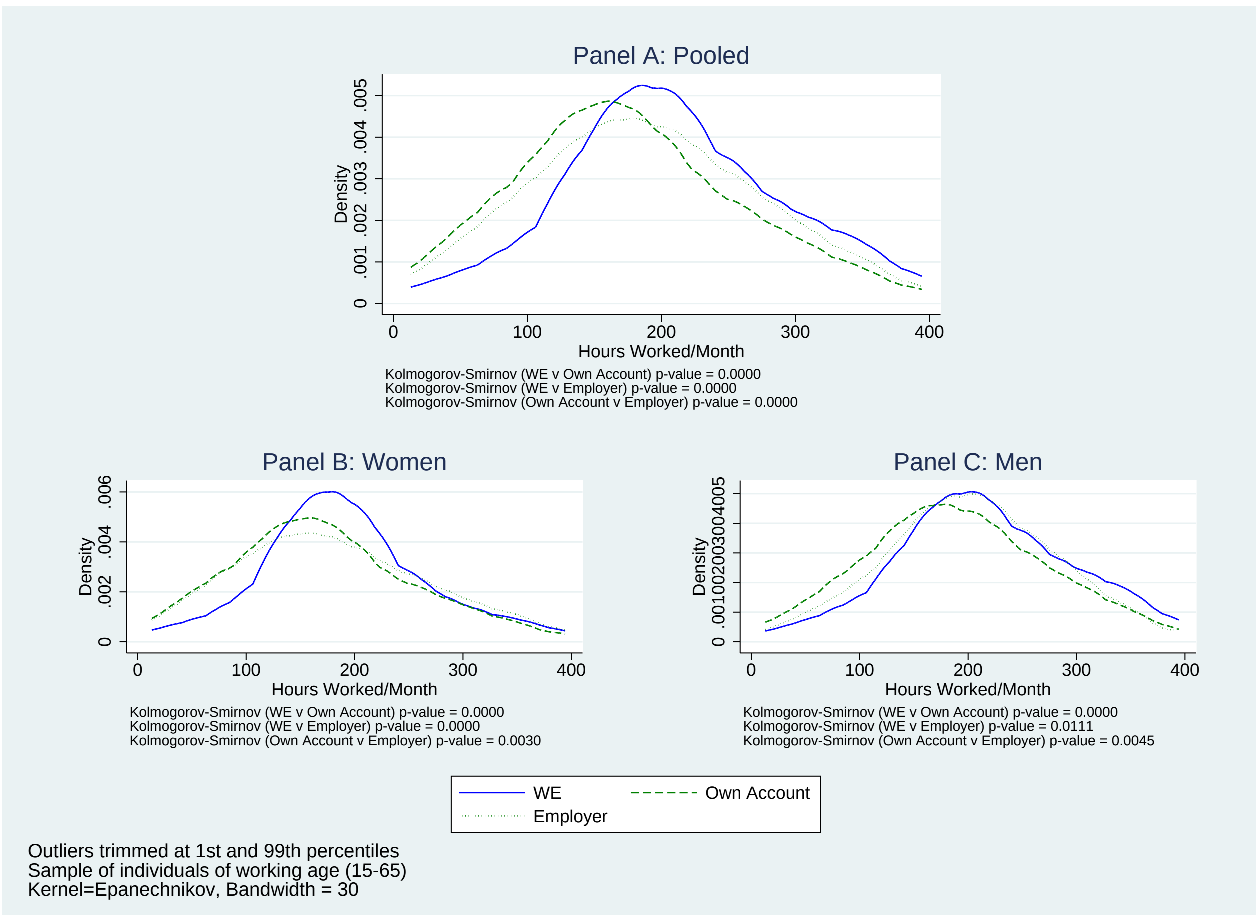


In both the morning and afternoon shifts, domestic work is undertaken concurrently with market work more in self-employment, especially amongst the own account workers. Again this supports our priors that low-input self-employment offers more job flexibility. However, this also suggests that the multi-tasking model outlined in Section 3.3.2 may be a tenable way of interpreting our data on occupational choice.

3.4.3.3 Work Hours

Finally, we plot the hours worked per month for each occupation to investigate whether any of the occupations display evidence of sharp lower bounds on the amount of time that must be dedicated to market work. In Panel A of Figure 3.10 we show the distributions for the full sample, and then we split by sex in Panels B and C.

Figure 3.10: Distributions of Hours Worked per Month by Occupation and Sex



The distributions differ significantly across all three sectors, for the full sample, and for women and men separately. This is tested formally using the Kolmogorov-Smirnov method, with the p-values reported under each graph. The most pronounced differences appear to be at the bottom of the distribution, where there are far fewer wage jobs than self-employment jobs, especially for the sub-sample of women. This is consistent with the notion that formal wage sector jobs are characterised by minimum hours, linking with the model outlined in Section 3.3.3. However, there is significant *intra*-sector heterogeneity, and there do appear to be some wage and high-input self-employment jobs, where individuals can participate but still work very few hours per month.

3.4.3.4 Job Flexibility in the Data

The data broadly confirm our priors that job flexibility, which we measure in various ways, is greatest in own account work and lowest in wage-employment. There are also important differences between low- and high-input self-employment, further reinforcing the need to consider heterogeneity amongst the self-employed. It appears that the multi-tasking and minimum hours stories are borne out most in the descriptive statistics. Perhaps surprisingly, the time-use data suggest that caring for other household members is a relatively small proportion of domestic work, and that household chores comprise the largest share. Since we expect household chores to vary less over time than caring for household members, this suggests that the adjustment costs model may be less important.

We further examine which of the models outlined in Section 3.3 reflects the patterns in our data, when we present our results in Section 3.6 and 3.7.

3.5 Econometric Approach

Our econometric approach comes in two parts. First, we attempt to decompose the earnings gaps in each sector, controlling for a number of observable variables and selection. Second, we interrogate our selection equations to investigate whether domestic obligations may determine occupational choice as our theoretical models from Section 3.3 predict.

3.5.1 Male-Female Earnings Gap Decompositions

We use a modified version of the method developed by Blinder (1973) and Oaxaca (1973) in order to decompose the raw male-female earnings gap in each sector.

One upfront difficulty with this approach is that the portion of the earnings gap that can be explained by differences in individual characteristics is highly sensitive to the ‘reference wage structure' (Oaxaca and Ransom, 1994; Jann, 2008). Intuitively, this problem relates to whether the earnings gap is evaluated in terms of male advantage or female disadvantage. To overcome this issue, we estimate the reference wage structure with a *pooled model* in which male advantage and female disadvantage are restricted to be equal (Fortin, 2008; Fortin et al., 2011).

To implement this approach, we estimate three earnings equations for each sector, j. We first estimate earnings equations for women (f) and men (m), as outlined in Equations (3.27) and (3.28) respectively. We also estimate a pooled earnings equation (p), with the restriction that male advantage and female disadvantage are equal in Equation (3.29). This restriction is operationalised by adding a both a male dummy (M_i) *and* a female dummy (F_i) to the pooled earnings equation, and then imposing the constraint that the coefficients on these dummies are equal and opposite. Individual i's earnings are determined not only by their sex, but also by a vector of individual characteristics $\boldsymbol{x}_i$ and an unobservable term, denoted u with the relevant subscripts.

$$\ln w_{ifj} = \boldsymbol{x_i}'\boldsymbol{\beta_{fj}} + u_{ifj} \tag{3.27}$$

$$\ln w_{imj} = \boldsymbol{x_i}'\boldsymbol{\beta_{mj}} + u_{imj} \tag{3.28}$$

$$\begin{aligned} \ln w_{ipj} \ =& \gamma_{pfj} F_i + \gamma_{pmj} M_i + \boldsymbol{x_i}'\boldsymbol{\beta_{pj}} + u_{ipj} \\ & \text{subject to } \gamma_{pfj} + \gamma_{pmj} = 0 \end{aligned} \tag{3.29}$$

The earnings gap can then be decomposed using Equation (3.30).

$$\overline{\ln w_{mj}} - \overline{\ln w_{fj}} = (\overline{\boldsymbol{x_m}}' - \overline{\boldsymbol{x_f}}')\widehat{\boldsymbol{\beta_{pj}}} + \overline{\boldsymbol{x_m}}'(\widehat{\boldsymbol{\beta_{mj}}} - \widehat{\boldsymbol{\beta_{pj}}}) + \overline{\boldsymbol{x_f}}'(\widehat{\boldsymbol{\beta_{pj}}} - \widehat{\boldsymbol{\beta_{fj}}}) \tag{3.30}$$

The first term on the right-hand-side relates to differences in the individual characteristics between women and men, given that they are rewarded according to the estimated parameters from the pooled earnings equation. This is the explained portion of the earnings gap, also known as the 'composition effect'.

The second and third term capture differences in rewards to particular characteristics between women and men, using the pooled earnings equation as the reference point. This unexplained component, also known as the 'wage structure effect', may result from discrimination or unobserved heterogeneity.

The individual characteristics in $\boldsymbol{x_i}$ on which we focus are mainly related to human capital. In particular, we include education, age, and, tenure to measure individuals' earnings potential, in the spirit of Mincer (1974). However, we also control for a set of demographic variables, namely marital status, household size, and the household dependency ratio, as well as heterogeneity in location and ethnicity.[17] Although this chapter is not focussed on estimating the causal impact of dimensions of human capital, we try to control for potential confounds where possible, because failing to do so may bias the decomposition in Equation (3.30). For example, our demographic and ethnicity variables should capture important variation in family background, which may influence the estimated returns to education (Oster, 2013). We initially controlled for ability directly by using responses to the literacy and numeracy skills questions in the GLSS5+ data. However, since these had virtually no impact on our results, they are omitted from the set of controls.

3.5.2 Controlling for Selection

If selection into different sectors is non-random, the parameters of the earnings equations for women, men, and the pooled sample may be biased. In particular, if the unobservables that determine occupational selection are correlated with the unobservables that determine earnings, this will induce a correlation between the unobservables and observables in the earnings equations. In turn, the estimation of the wage structure effect will be biased.

We therefore correct our estimates by including terms akin to the Inverse Mills Ratio to control for selection, as though it were a problem of omitted variable bias (Heckman, 1979). To allow for selection both into non-farm self-employment, and technology selection *within* that sector, we use six different occupational categories:

1. Wage-Employed

[17] Specifically, we include dummy variables for the ten regions available in the sample interacted with the urban dummy. We also include dummy variables for the three main ethnic groups in our data; Ashanti, Ewe, and Fante. We use the routine outlined by Jann (2008) to ensure our results are not sensitive to the choice of the omitted base category.

2. Self-Employed — Agriculture
3. Non-Farm Self-Employed (Own Account)
4. Non-Farm Self-Employed (Employer)
5. Unemployed
6. Out of the Labour Force

We focus on earnings amongst the wage-employed and the non-farm self-employed, splitting the latter by whether or not they employ labour besides their own. We distinguish 'Unemployed' from 'Out of the Labour Force' on the basis of the individual's recent search activity, and we treat those self-employed in agriculture as a different category.

We estimate separate selection models for the female, male, and pooled, samples to provide separate correction terms for Equations (3.27), (3.28), and (3.29), utilising two different selection corrections based on the multinomial logit model. First, we employ the widely used Lee (1983) method, which generates just one correction term. Lee's approach is somewhat restrictive, however, requiring that the covariance between the outcome equation error terms and the selection equation error terms is the same for all sectors. To overcome this, we also use the more flexible Dubin and McFadden (1984) method, which generates a correction term for each occupational category.[18]

The relative merits of these methods are evaluated extensively using Monte Carlo simulations by Bourguignon et al. (2007). They show that both the Lee and the Dubin-McFadden approaches are robust to the failure of the Independent of Irrelevant Alternatives (IIA) assumption, with the latter performing slightly better. However, the Lee (1983) is more appropriate when the overall sample size is small relative to the number of categories. In these contexts, the selection terms generated by the Dubin-McFadden technique may be collinear, reducing efficiency in the estimates of the outcome equation parameters. For this reason, we focus on the Lee method in our results, and then use Dubin-McFadden to check the signs and magnitudes.

To decompose the earnings gap whilst controlling for selection, we add a vector of correction

[18] In particular, we adopt the modified version of the Dubin-McFadden approach described by Bourguignon et al. (2007).

terms, $\boldsymbol{\lambda_i}$, to our three earnings equations.[19]

$$\ln w_{ifj} = \boldsymbol{x_i}'\boldsymbol{\beta_{fj}} + \boldsymbol{\lambda_i}'\boldsymbol{c_{fj}} + u_{ifj} \tag{3.31}$$

$$\ln w_{imj} = \boldsymbol{x_i}'\boldsymbol{\beta_{mj}} + \boldsymbol{\lambda_i}'\boldsymbol{c_{mj}} + u_{imj} \tag{3.32}$$

$$\begin{aligned} \ln w_{ipj} = \ & \gamma_{pfj} F_i + \gamma_{pmj} M_i + \boldsymbol{x_i}'\boldsymbol{\beta_{pj}} + \boldsymbol{\lambda_i}'\boldsymbol{c_{pj}} + u_{ipj} \\ & \text{subject to } \gamma_{pfj} + \gamma_{pmj} = 0 \end{aligned} \tag{3.33}$$

For the decomposition, we follow Reimers (1983) in moving the selection terms to the left-hand side. In Reimers' terminology, adjusting the earnings gap for selection yields the difference in wage *offers*, as opposed to the difference in *observed* wages. Given the prevalence of self-employment in our data, thinking in terms of 'wage offers' is somewhat inappropriate, so we label the selection-corrected earnings gap as the difference in 'potential earnings'. We show the selection-corrected decomposition in Equation (3.34).

$$\overline{\ln w_{mj}} - \overline{\ln w_{fj}} - [\overline{\boldsymbol{\lambda_m}}'\widehat{\boldsymbol{c_{mj}}} - \overline{\boldsymbol{\lambda_f}}'\widehat{\boldsymbol{c_{fj}}}] = (\overline{\boldsymbol{x_m}}' - \overline{\boldsymbol{x_f}}')\widehat{\boldsymbol{\beta_{pj}}} + \overline{\boldsymbol{x_m}}'(\widehat{\boldsymbol{\beta_{mj}}} - \widehat{\boldsymbol{\beta_{pj}}}) + \overline{\boldsymbol{x_f}}'(\widehat{\boldsymbol{\beta_{pj}}} - \widehat{\boldsymbol{\beta_{fj}}}) \tag{3.34}$$

The performance of the multinomial logit based selection methods is sensitive to finding suitable exclusion restrictions. We require variables that plausibly influence occupational status without directly influencing earnings. Firstly, we use ownership of land and receipts of unearned income from social security, retirement benefits, or dowry or inheritance payments. There is much existing evidence suggesting that these variables may affect access to capital, and that this is an important determinant of the type of self-employment opportunities available to individuals (Blanchflower and Oswald, 1998; Blanchflower, 2004; Deininger et al., 2013). Secondly, we can use the detailed information available in the GLSS5+ on the jobs undertaken by other family members (Appleton et al., 1999). This may have two key effects on individuals' occupational selection.

[19] We do this for our pooled equation too, even though the selection terms are not directly used in the decomposition.

First, there may be network effects, which give individuals better access to certain occupations if, for example, the same job is done by their father. Second, the occupational statuses of other family or household members may affect individuals' roles within the household, and hence their demand for domestic or market work.

3.5.3 Endogeneity of Domestic Obligations

We investigate whether domestic obligations may explain occupational choice for females differently from males, given different preferences for job flexibility, by focussing on the household demographic variables in our selection equations. In particular, we examine whether a higher dependency ratio pushes women into low-input self-employment, more than men. We begin by using the same multinomial logit models used to correct our earnings decompositions for selection.

However, as we argued in Section 3.2.2, it may be that family demographic variables are endogenous to an individual's choice of job. Unobserved characteristics may jointly influence decisions about employment and family planning. Moreover, individuals' occupations may affect family stability and thus fertility, whilst having extra children could also be a rational means of boosting household income for those working in jobs with low returns (Becker, 1985; Rosenzweig and Wolpin, 1980).

We want to assess whether our results are sensitive to this potential endogeneity of family demographics to occupational selection. To do this, we use selection on observables as a guide for selection on unobservables, exploiting the logic of the Altonji et al. (2005). *We are now concerned about the selection of the potentially endogenous demographic variable, rather than occupation.* We modify this method for the context of a discrete multinomial choice problem, using the multinomial probit model.

To explain our technique, we begin by setting up the problem as an Additive Random Utility Model.[20] Utility for individual (i) in occupation ($j = 1, 2, ..., J$) is a function of individual characteristics ($\boldsymbol{x}_i$) as well as some potentially endogenous family demographic variable, D_i. Specifically, this will be measured by the dependency ratio. We observe alternative k being chosen by individual i if $u_{ik} > u_{il}$ for $\forall l \neq k$. Although $\boldsymbol{x}_i$ and D_i vary only at the individual level, and not across occupations, the coefficients and error terms are occupation-specific. Utility

[20] This motivation is implicit in the multinomial logit model, but the assumptions about the distribution of the error terms are different.

in the J sectors can thus be written:

$$\begin{aligned} u_{i1} &= \boldsymbol{x_i'\alpha_1} + \gamma_1 D_i + \xi_{i1} \\ u_{i2} &= \boldsymbol{x_i'\alpha_2} + \gamma_2 D_i + \xi_{i2} \\ &\vdots \\ u_{iJ} &= \boldsymbol{x_i'\alpha_J} + \gamma_J D_i + \xi_{iJ} \end{aligned} \tag{3.35}$$

We assume that the error terms are distributed multivariate Standard Normal, as in Equation (3.36).

$$\begin{pmatrix} \xi_{i1} \\ \xi_{i2} \\ \vdots \\ \xi_{iJ} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right] \tag{3.36}$$

Compared with the multinomial logit, specifying the model with a multivariate Normal error variance-covariance matrix adds computational complexity for very little gain, especially as the off-diagonal terms are all set to zero (Cameron and Trivedi, 2005; Long and Freese, 2006).[21] However, we eventually use this functional form assumption to allow for the presence of an endogenous regressor.

In order to estimate multinomial problems of this type, we must first respecify the model in terms of the difference between utility in each occupation compared to some base category. Following the convention in Train (2009), we use the first category as the base category.[22] We operationalise this by setting $u_{i1} = 0$. The *relative* utilities $\forall j = 2, 3, ..., J$ can then be written:

$$\begin{aligned} y^*_{ij'} &= u_{ij} - u_{i1} \\ &= \boldsymbol{x_i'}(\boldsymbol{\alpha_j} - \boldsymbol{\alpha_1}) + (\gamma_j - \gamma_1) D_i + (\xi_{ij} - \xi_{i1}) \\ &= \boldsymbol{x_i'\beta_{j'}} + \psi_{j'} D_i + \epsilon_{ij'} \end{aligned} \tag{3.37}$$

[21] Indeed, this is analogous to the IIA assumption imposed on the multinomial logit model.

[22] Although the choice of the base category will influence the parameters in $\boldsymbol{\alpha_j}$ and γ_j, this does not affect the marginal effects.

The new error terms, $\epsilon_{ij'}$, are thus distributed as in Equation (3.38).

$$\begin{pmatrix} \epsilon_{i2'} \\ \epsilon_{i3'} \\ \vdots \\ \epsilon_{iJ'} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{pmatrix} \right] \tag{3.38}$$

We cater for endogeneity by assuming a specific structure for the unexplained component of our endogenous variable, D_i. First, we imagine a 'first-stage' selection equation, which relates our endogenous variable D_i to the other observable variables in $\boldsymbol{x}_i$. We do not include any excluded instrumental variables in this equation. Instead, we tackle endogeneity by considering different assumptions about the error term υ_i.

$$D_i = \boldsymbol{x}_i^{'}\boldsymbol{\pi} + \upsilon_i \tag{3.39}$$

Under the assumption that D_i is exogenous, the error terms in Equations (3.37) and (3.39), $\epsilon_{ij'}$ and υ_i, are uncorrelated. Endogeneity arises when this correlation is non-zero. We follow Rosenbaum and Rubin (1983) and Greene (2003) and formalise this potential endogeneity using a multivariate Normal distribution for the error terms.

$$\begin{pmatrix} \epsilon_{i2'} \\ \epsilon_{i3'} \\ \vdots \\ \epsilon_{iJ'} \\ \upsilon_i \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 & \cdots & 1 & \rho\sigma \\ 1 & 2 & \cdots & 1 & \rho\sigma \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 2 & \rho\sigma \\ \rho\sigma & \rho\sigma & \cdots & \rho\sigma & \sigma^2 \end{pmatrix} \right] \tag{3.40}$$

The correlation between υ_i and *all* the $\epsilon_{ij'}$ terms is governed by the parameter ρ.[23] Under the assumption that D_i is exogenous, $\rho = 0$. However, by estimating the model with different values of ρ, we can test the sensitivity of our results to relaxing this exogeneity assumption. Therefore, we do not estimate ρ, as one might with a regular Heckman selection model with

[23] The assumption that this correlation is the same for all of the latent variable error terms is strong. However, since ρ is a parameter to be altered by the econometrician, we believe this approach is still informative about the impact of endogeneity. Allowing for different values of ρ for each sector would dramatically increase the number of possible assumptions about endogeneity, and is beyond the scope of this chapter.

excluded instruments, but rather change it manually. As such, the only free parameter in the error variance-covariance matrix is the standard deviation of υ_i, labelled σ.[24]

In order to write the likelihood function, we first need to derive the distribution of $(\epsilon_{i2'}, \epsilon_{i3'}, ...\epsilon_{iJ'})$ conditional on υ_i. Using the properties of a Multivariate Normal distribution we can write:

$$\left(\begin{array}{c|c} \epsilon_{i2'} & \\ \epsilon_{i3'} & \upsilon_i \\ \vdots & \\ \epsilon_{iJ'} & \end{array} \right) \sim N \left[\begin{pmatrix} \rho\sigma^{-1}\upsilon_i \\ \rho\sigma^{-1}\upsilon_i \\ \vdots \\ \rho\sigma^{-1}\upsilon_i \end{pmatrix}, \begin{pmatrix} (2-\rho^2) & (1-\rho^2) & \cdots & (1-\rho^2) \\ (1-\rho^2) & (2-\rho^2) & \cdots & (1-\rho^2) \\ \vdots & \vdots & \ddots & \vdots \\ (1-\rho^2) & (1-\rho^2) & \cdots & (2-\rho^2) \end{pmatrix} \right] \tag{3.41}$$

Following Greene (2003), we rewrite the latent variable equations such that the error terms are distributed with a 0 vector for the mean.

$$\begin{aligned} y^*_{ij'} &= \boldsymbol{x_i}'\boldsymbol{\beta_{j'}} + \psi_{j'} D_i + \epsilon_{ij'} \\ &= \boldsymbol{x_i}'\boldsymbol{\beta_{j'}} + \psi_{j'} D_i + \rho\sigma^{-1}\upsilon_i + \zeta_{ij'} \end{aligned} \tag{3.42}$$

The error variance-covariance matrix may now be written:

$$\left(\begin{array}{c|c} \zeta_{i2'} & \\ \zeta_{i3'} & \upsilon_i \\ \vdots & \\ \zeta_{iJ'} & \end{array} \right) \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} (2-\rho^2) & (1-\rho^2) & \cdots & (1-\rho^2) \\ (1-\rho^2) & (2-\rho^2) & \cdots & (1-\rho^2) \\ \vdots & \vdots & \ddots & \vdots \\ (1-\rho^2) & (1-\rho^2) & \cdots & (2-\rho^2) \end{pmatrix} \right] \tag{3.43}$$

With this error structure in place, we can now begin to write the likelihood function. Labelling the observed occupation for individual i as y_i, and the parameter vector $\boldsymbol{\theta}$, we can write the likelihood function for the individual (Train, 2009).[25]

$$L_i(\boldsymbol{\theta}; y_i \mid \boldsymbol{x_i}, D_i) = \prod_{j=1}^{J} \left\{ \mathbb{1}(y_i = j) \times \Pr(y_i = j \mid \boldsymbol{x_i}, D_i) \times \frac{1}{\sigma}\phi\left(\frac{\upsilon_i}{\sigma}\right) \right\} \tag{3.44}$$

[24]To simplify computation in our empirical results section, we normalise our variables such that $\sigma = 1$.
[25]We denote the Standard Normal distribution using ϕ and Φ for the PDF and CDF respectively.

Replacing υ_i using the observables in Equation (3.39), and taking logs, we can write the log likelihood for the individual.

$$l_i(\boldsymbol{\theta}; y_i \mid \boldsymbol{x_i}, D_i) = \sum_{j=1}^{J} \left\{ \mathbb{1}(y_i = j) \times \ln \left[\Pr(y_i = j \mid \boldsymbol{x_i}, D_i) \right] + \ln \left[\frac{1}{\sigma} \phi \left(\frac{D_i - \boldsymbol{x_i}' \boldsymbol{\pi}}{\sigma} \right) \right] \right\} \quad (3.45)$$

For the sample as a whole, we can write:

$$l(\boldsymbol{\theta}; \boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{D}) = \sum_{i=1}^{N} \sum_{j=1}^{J} \left\{ \mathbb{1}(y_i = j) \times \ln \left[\Pr(y_i = j \mid \boldsymbol{x_i}, D_i) \right] + \ln \left[\frac{1}{\sigma} \phi \left(\frac{D_i - \boldsymbol{x_i}' \boldsymbol{\pi}}{\sigma} \right) \right] \right\} \quad (3.46)$$

We can now estimate the parameters in $\boldsymbol{\theta}$ using maximum simulated likelihood, given our distributional assumptions about the error terms in Equation (3.43).[26] We calculate the simulated probability for each individual using a 'logit-smoothed' simulator (Train, 2009).[27] Importantly, we can do this for different values of ρ, thus testing different assumptions about the endogeneity of D_i.

Greater values of ρ imply more selection of D_i on the basis of unobservables, and hence more endogeneity. However, it is difficult to assess the magnitude of ρ. The question remains over *how large* ρ can become before we are content our results are robust. Put differently, how much selection of D_i on the basis of unobservables are we willing to allow for? One possibility is to assume that the selection (of D_i) on unobservables is *equal* to the selection on observables (Altonji et al., 2005, 2008; Oster, 2013). In reality, we hope that our control variables are sufficiently relevant to explain household demographics *more* than the unobservables, so the 'equal selection' condition can be understood as something of an upper bound on how bad endogeneity could become. Thus, if our main results survive under this condition, this supports the notion that they are robust to concerns about endogeneity.

We label the value of ρ, which implies 'equal selection', $\tilde{\rho}_{j'}$. Since there are different sets of unobservables for each occupational category, this value must be indexed by j'. This reflects the

[26] We prefer this method to using our preliminary assumptions in Equation (3.36) to derive a closed-form likelihood function. We show, however, that this may be possible in Appendix 3.D.

[27] This approach helps overcome some of the short-comings of a simple 'accept-reject' simulator, whilst maintaining parsimony for coding the estimator. We simulate with 1000 repetitions.

fact that the factors driving individuals to select wage-employment, relative to the base category, are different from the factors for self-employment. Thus, when we conduct our sensitivity analysis, we assess our results as though the *common* ρ had been increased to the $\tilde{\rho}_{j'}$ *for that sector* as well as for the sector where $\tilde{\rho}_{j'}$ is highest. Although, evaluating our results in terms of a *common* ρ for all occupational categories is somewhat restrictive, it still allows us to use selection on observables as a guide to selection on unobservables in a parsimonious way.

Following Altonji et al. (2005), we can write the equal selection condition as in Equation (3.47).[28]

$$\tilde{\rho}_{j'} = \left| \frac{\mathrm{Cov}(\boldsymbol{x_i'\beta_{j'}}, \boldsymbol{x_i'\pi})}{\mathrm{Var}(\boldsymbol{x_i'\beta_{j'}})} \right| \tag{3.47}$$

Thus, by checking whether our effects survive when ρ is pushed up to $\tilde{\rho}_j$ and beyond, we directly capture the idea that selection on observables serves as a guide for selection on unobservables.

3.6 Results

In this section we present our main results. We begin in Section 3.6.1 by reporting the main wage equations that we use to decompose the sex earnings gaps in different sectors, both with and without correcting for occupational selection. We then show the analogous decompositions in Section 3.6.2. In Section 3.6.3 we report the multinomial selection models used to adjust our wage equations, focussing in particular on the impact that household demographics have on occupational choice. Finally, we assess the sensitivity of these results in Section 3.6.4, using the simulated maximum likelihood technique described above.

3.6.1 Wage Equations

We begin by presenting the estimates for the earnings equations used to conduct our Blinder-Oaxaca decompositions in each sector. We first estimate these using OLS, and then include

[28] We must take the absolute value of this correlation, because of the way we have set up utility *relative* to a base category in our model.

the Lee and Dubin-McFadden correction terms described in Section 3.5.2.[29] We bootstrap our standard errors to cater for the presence of the generated regressor(s), resampling from household-level clusters. We test the joint significance of the selection terms and present the p-value for this F-test for each sector. We also present a formal test of the difference between the coefficients on the 'DMF 3' and 'DMF 4' correction terms for the Dubin-McFadden model, as explained below.

The earnings equations for women are shown in Table 3.7. Overall, education is positively related to female earnings in all three sectors, with the most economically and statistically significant results emerging amongst the wage-employed. The impact of age is somewhat mixed, however, depending on occupation. Amongst the self-employed, age has a negative effect on the earnings of own account entrepreneurs, but a positive effect for those employing labour besides their own. This supports the interpretation that low-input self-employment partially acts as a 'sink' sector, where older individuals end up working after losing previous jobs (Sandefur et al., 2006).

Controlling for occupational selection seems to alter the coefficients on the main human capital variables. The age-earnings elasticity in low-input self-employment rises using both the Lee and modified Dubin-McFadden correction methods. Moreover, the impact of education on earnings for own account self-employed women becomes statistically insignificant, when using the modified Dubin-McFadden technique. This may be because women with high unobserved self-employment ability, but also *low* education, select into the low-input self-employment sector.[30] This sets up a positive correlation between education and unobserved ability in that sub-sample, biasing upwards the coefficient on education when the estimates are not corrected for selection. In contrast, for wage-employed women, the tenure-wage elasticity remains positive at approximately 0.13, regardless of whether or not we control for selection.

The analogous male earnings equations are shown in Table 3.8. Once again, education has a positive impact on earnings across the three sectors, with the strongest effects appearing amongst the wage-employed. Age only seems to affect the wage-employed, with a positive and significant age-earnings elasticity in all three models, and there are no significant tenure effects in any sector. Although the F-tests show that the selection terms are significant, especially in the Dubin-McFadden specification, the coefficients on the human capital variables appear to be somewhat more robust to controlling for selection than in the sample of women.

The question remains over how we should interpret the signs and magnitudes of the selection

[29] We implement the multinomial logit selection corrections using `selmlog`, a user-written Stata command provided by Fournier and Gurgand (2005).

[30] These patterns of selection are confirmed by the marginal effects for the multinomial logit reported in Section 3.6.3.

correction terms in Tables 3.7 and 3.8. We start with the Lee method, since interpreting the single adjustment term is relatively straightforward. First, we denote the correlation between the unobservables that determine selection into each sector j and the unobservables that determine earnings in the sector of interest 'r_j'. For the Lee method, we only have one term for the sector in which the earnings equation is being estimated, so the Lee correction terms are directly analogous to the Inverse Mills Ratio created for the two-step Heckman method (Heckman, 1979). It turns out, however, that the Lee correction term enters the wage equations with the *opposite* sign to r_j. We can see that the Lee term for own account self-employment in both Tables 3.7 and 3.8 has a positive sign. This indicates that, for both women and men, the unobservables related to selection into own account self-employment are *negatively* correlated with the unobservables determining earnings in that sector. The estimates of both female and male earnings in low-input self-employment are biased downwards, because individuals with low earnings potential select disproportionately into that sector. However, the size of the coefficient on this Lee correction term is larger for men (0.3118) than for women (0.2502), so this downward bias is more serious in the male sample. Therefore, we expect the gap between male and female earnings to get larger when we control for this occupational selection.

We can use exactly the same logic to evaluate the Lee terms in high-input self-employment and wage-employment. For the employers, the coefficients on the Lee terms are positive in both the female and male earnings equations but, once again, the magnitude is greater for the men (0.7017) than women (0.1237). This indicates that the negative correlation between the unobservables associated with selecting into high-input self-employment and the unobservables associated with earnings in that sector is stronger for men than women. Thus, earnings are biased downward more for the males in that sector than the females, such that controlling for selection should widen the earnings gap between men and women. The converse, however, is true for wage-employment. Here, the Lee term enters with a positive coefficient for women (0.0795) and a negative coefficient for men (-0.0412). Thus the relevant r_j is negative for women and positive for men. This implies that occupational choice biases female wages downwards and male wages upwards, such that controlling for selection should *reduce* the earnings premium for men.

Table 3.7: Female Wage Equations

	OLS			Lee			Dubin-McFadden		
	Own Account	Employer	WE	Own Account	Employer	WE	Own Account	Employer	WE
Education (Years)	0.0208***	0.0389***	0.0874***	0.0200***	0.0381***	0.0811***	0.0069	0.0425**	0.0754***
	(0.0068)	(0.0128)	(0.0074)	(0.0058)	(0.0122)	(0.0152)	(0.0089)	(0.0189)	(0.0183)
Log of Age	-0.1742*	0.4127***	0.1345	-0.2693**	0.3342	0.1193	-0.4473***	0.4871	0.2243
	(0.0935)	(0.1498)	(0.1624)	(0.1180)	(0.2506)	(0.1569)	(0.1592)	(0.3315)	(0.2969)
Log of Tenure	0.0261	-0.0339	0.1364***	0.0286	-0.0330	0.1360***	0.0383**	-0.0354	0.1206***
	(0.0216)	(0.0387)	(0.0361)	(0.0207)	(0.0395)	(0.0316)	(0.0183)	(0.0446)	(0.0348)
Urban	0.3001**	-0.2068	-0.3031	0.2189	-0.2211	-0.3123	0.1875	-0.2462	-0.0572
	(0.1440)	(0.3070)	(0.2440)	(0.1585)	(0.3143)	(0.2556)	(0.1629)	(0.3708)	(0.3014)
Lee				0.2502*	0.1237	0.0795			
				(0.1385)	(0.2226)	(0.1451)			
DMF 1							-0.3754	-0.1303	0.0654
							(0.4132)	(0.8260)	(0.0840)
DMF 2							0.5612*	0.1593	-0.5151
							(0.2910)	(0.6283)	(0.5669)
DMF 3							0.0272	-0.7787	1.8818***
							(0.1046)	(0.7192)	(0.5272)
DMF 4							-1.8327***	-0.0293	-0.6268
							(0.4982)	(0.1341)	(0.8237)
DMF 5							0.7929	0.6378	-0.4492
							(0.5447)	(1.0208)	(0.8314)
DMF 6							-0.1503	-0.6420	0.7220*
							(0.3583)	(0.6476)	(0.4156)
Ethnicity Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	2586	791	641	2586	791	641	2586	791	641
R^2	0.148	0.227	0.345	0.149	0.227	0.346	0.158	0.231	0.365
Selection F-Test				0.0709	0.5783	0.5838	0.0001	0.6535	0.0044
Test DMF 3 v DMF 4							0.0005	0.3351	0.0403

Standard errors in parentheses

Dependent variable: Log of Hourly Wages

Bootstrapped standard errors to account for generated regressors, replicated with household clusters

F-Test for joint significance of selection terms

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Male Wage Equations

	OLS			Lee			Dubin-McFadden		
	Own Account	Employer	WE	Own Account	Employer	WE	Own Account	Employer	WE
Education (Years)	0.0346***	0.0391***	0.0851***	0.0367***	0.0343**	0.0873***	0.0225	0.0528**	0.0760***
	(0.0119)	(0.0127)	(0.0064)	(0.0117)	(0.0144)	(0.0071)	(0.0188)	(0.0237)	(0.0093)
Log of Age	-0.0176	0.4462*	0.2899***	-0.0973	0.3746	0.3042***	0.0427	0.2879	0.3693***
	(0.1990)	(0.2511)	(0.1072)	(0.1887)	(0.2489)	(0.1011)	(0.2903)	(0.3983)	(0.1260)
Log of Tenure	0.0017	-0.0468	0.0238	-0.0000	-0.0670	0.0246	-0.0012	-0.0623	0.0206
	(0.0463)	(0.0616)	(0.0195)	(0.0478)	(0.0652)	(0.0216)	(0.0464)	(0.0605)	(0.0199)
Urban	0.0329	0.3079	0.0712	-0.0745	0.0398	0.0740	-0.2568	0.3282	0.2285*
	(0.3044)	(0.3515)	(0.1130)	(0.2800)	(0.3605)	(0.1213)	(0.2862)	(0.4356)	(0.1323)
Lee				0.3118	0.7017***	-0.0412			
				(0.2079)	(0.2334)	(0.0856)			
DMF 1							0.0552	0.2485	-0.0388
							(0.8079)	(0.8702)	(0.0733)
DMF 2							0.9670	-0.8209	0.0005
							(0.5991)	(0.7050)	(0.2397)
DMF 3							-0.0312	0.5461	1.5237***
							(0.0998)	(0.8009)	(0.3732)
DMF 4							-0.6822	-0.4069***	-1.1120***
							(0.7809)	(0.1293)	(0.3801)
DMF 5							0.6404	-0.4956	0.8548***
							(0.8887)	(1.2247)	(0.2727)
DMF 6							0.2053	-0.2193	-0.1607
							(0.7089)	(0.9003)	(0.2964)
Ethnicity Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	788	520	1726	788	520	1726	788	520	1726
R^2	0.148	0.129	0.213	0.151	0.143	0.213	0.159	0.154	0.229
Selection F-Test				0.1336	0.0026	0.6303	0.2608	0.0272	0.0000
Test DMF 3 v DMF 4							0.4184	0.2502	0.0000

Standard errors in parentheses

Dependent variable: Log of Hourly Wages

Bootstrapped standard errors to account for generated regressors, replicated with household clusters

F-Test for joint significance of selection terms

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The Dubin-McFadden correction terms are somewhat more difficult to interpret, because they allow for different correlations between the unobservables in the earnings equation and the unobservables associated with selection into *each* sector. For example, the 'DMF 1' term in the own account earnings equation tells us the correlation between the unobservable determinants of low-input self-employment earnings and the unobservables associated with selection into the *wage*-employment sector. As in the Lee model, almost all these terms enter with the opposite sign to r_j.[31] Interpreting the Dubin-McFadden results is further complicated by the fact that each selection term is weighted according to the probability of being in a particular sector.[32] Thus, it is not sufficient to simply add up the terms to ascertain whether workers are positively or negatively selecting into a particular employment sector. However, as we see in the next section, our decomposition estimates move in similar directions, whether we use the Lee or the Dubin-McFadden method. This suggests that, taking all the Dubin-McFadden terms together, the overall correlation between the unobservables associated with earnings in a given sector and the unobservables associated with selection into that sector is not sensitive to the correction technique used.

Nonetheless, there are some important differences between the Lee and the Dubin-McFadden approaches, which are shown by Tables 3.7 and 3.8, that we wish to emphasise. Firstly, on the methodology, the Dubin-McFadden results show that the correlation structure between the unobservable determinants of earnings and the unobservables associated with selection are not the same for each sector. Thus, this restriction in the Lee method is somewhat at odds with the data, so the Lee and the Dubin-McFadden techniques must be interpreted together. Secondly, we can see that there are significant differences between the Dubin-McFadden correction terms for own account self-employment and employer self-employment. This means that the unobservables in the earnings function have a different relationship with the unobservables that result in selection into low-input self-employment and high-input self-employment. Since these selection forces are so different, we believe our approach of splitting the self-employed into those that do and do not employ labour besides their own is important.

How should these wage equations be interpreted from the perspective of our theoretical models of job flexibility in Section 3.3? One possibility is that a preference for job flexibility helps to explain why the correlation between earnings equation unobservables and selection unobservables is less negative for women than men for the own account self-employment sector. For both sexes, unobserved earnings ability is lower in the sub-sample of low-input self-employed

[31] The exception is the correction term for the sector in which we are estimating the earnings equation, which enters with the *same* sign as r_j. For example, in the own account earnings equations, 'DMF 3' enters with the same sign as the correlation between the unobservables associated with selection into low-input self-employment and the unobservables associated with earnings in that sector.

[32] These weights are explained by Fournier and Gurgand (2005).

people, but this is somewhat offset by the fact that comparatively more productive women also have a preference for flexible jobs. Indeed, *relative* to males, females select *positively* into own account self-employment, which is consistent with the story that this type of work presents non-pecuniary benefits to women. We are able to investigate this idea more directly by looking at the multinomial logit models used to adjust our wage equations in Section 3.6.3.

For completeness, we show the pooled sample estimates of Equation (3.33) in Appendix 3.E, since these estimates are used for the reference wage structure in our Blinder-Oaxaca decompositions.

3.6.2 Earnings Gap Decompositions

Our main earnings decompositions are presented in Table 3.9. Again, we show the unadjusted estimates first, then control for selection using the Lee and the modified Dubin-McFadden methods. For each column, we present z-scores to assess the difference in the earnings gaps between sectors. These are based on the adjusted earnings gap for the selection-corrected models.[33]

For the unadjusted model, the raw earnings gap appears to be significantly smaller for the wage-employed compared to both the self-employed sectors. Amongst the wage-employed, women earn approximately 19 percent less than men, whilst the difference is 36 percent for the low-input self-employed and 45 percent for the high-input self-employed. In all three sectors, a substantial proportion of the differential is not explained by observables factors. However, the unexplained component of the earnings gap is most important for the own account self-employed.

The Lee method provides the most coherent picture of how adjusting for selection alters the earnings decompositions. As anticipated, the earnings gap amongst the wage-employed virtually disappears, falling to just 3 percent and no longer being statistically significant. This reflects the fact that, before controlling for selection, female earnings are biased down relative to male earnings, because women with lower income-generating ability select disproportionately into the wage sector. In fact, the composition effect remains at around 8 percent in favour of men, but there is a 5 percent wage structure effect in favour of women, which pushes the overall wage gap closer to zero. This suggests that, if anything, women are rewarded slightly more for a *given* set of individual characteristics than men in the wage sector.

[33] 'Difference' = $\overline{\ln w_{mj}} - \overline{\ln w_{fj}}$. 'Adjusted' = $\overline{\ln w_{mj}} - \overline{\ln w_{fj}} - [\overline{\boldsymbol{\lambda_m}}' \widehat{\boldsymbol{c_{mj}}} - \overline{\boldsymbol{\lambda_f}}' \widehat{\boldsymbol{c_{fj}}}]$. 'Explained' = $(\overline{\boldsymbol{x_m}}' - \overline{\boldsymbol{x_f}}')\widehat{\boldsymbol{\beta_{pj}}}$. 'Unexplained' = $\overline{\boldsymbol{x_m}}'(\widehat{\boldsymbol{\beta_{mj}}} - \widehat{\boldsymbol{\beta_{pj}}}) + \overline{\boldsymbol{x_f}}'(\widehat{\boldsymbol{\beta_{pj}}} - \widehat{\boldsymbol{\beta_{fj}}})$.

The opposite story holds for the self-employed. Focussing on the own account workers, after controlling for selection, the gap in potential earnings rises to approximately 60 percent. This is entirely intuitive given the selection terms in our earnings equations. Earnings estimates are biased downwards for both men and women, because the individuals selecting into that sector have lower income-generating ability. However, this effect is stronger for men, so failing to control for selection obscures a larger gap in potential earnings. The composition effect remains low at around 8 percent, so most of the selection-corrected earnings gap is not explained by observable factors. Thus, women with a given set of individual characteristics are rewarded less than men in the low-input self-employment sector.

These differences are even stronger for the high-input self-employed, for whom the earnings gap widens even further after controlling for selection. Using the Lee method, self-employed women that employ labour besides their own earn approximately 150 percent less than their male counterparts. Again, only a small portion of this gap can be explained by observable factors. The extent of this differential may reflect the fact that these entrepreneurs interact with both output *and* input markets far more than the low-input self-employed, giving greater opportunity for factors, like discrimination, to influence earnings.

We showed in Section 3.6.1 that it is useful to consider both the Lee and the Dubin-McFadden method, but in terms of our earnings decompositions, the two methods tell a similar story. Since the Lee method produces just one selection term, it is more robust to small samples and is somewhat easier to interpret, but it imposes much more structure on the correlation between the unobservables associated with occupational selection and the unobservables that determine earnings. As expected, the decompositions based on the Dubin-McFadden method are somewhat noisier, so we focus mainly on the direction in which the estimates move when we control for selection. As with the Lee method, the earnings gap amongst the wage-employed appears to close after controlling for selection, whilst for the self-employed it appears to widen. Again, for the wage-employed the wage structure effect favours women, cutting in the opposite direction to the composition effect. However, for both the self-employed sectors, most of the adjusted earnings gap is unexplained by observable factors.

Table 3.9: Blinder-Oaxaca Decompositions by Sector

	OLS			Lee			Dubin-McFadden		
	Own Account	Employer	WE	Own Account	Employer	WE	Own Account	Employer	WE
Differential									
Female Prediction	-2.900***	-2.745***	-2.342***	-2.901***	-2.746***	-2.341***	-2.899***	-2.744***	-2.336***
	(0.0258)	(0.0556)	(0.0420)	(0.0258)	(0.0564)	(0.0420)	(0.0259)	(0.0563)	(0.0419)
Male Prediction	-2.540***	-2.298***	-2.155***	-2.542***	-2.302***	-2.155***	-2.539***	-2.304***	-2.154***
	(0.0475)	(0.0556)	(0.0242)	(0.0471)	(0.0548)	(0.0242)	(0.0474)	(0.0547)	(0.0243)
Difference	-0.359***	-0.447***	-0.187***	-0.359***	-0.445***	-0.186***	-0.360***	-0.440***	-0.182***
	(0.0541)	(0.0787)	(0.0485)	(0.0537)	(0.0786)	(0.0485)	(0.0540)	(0.0785)	(0.0485)
Adjusted				-0.581*	-1.490***	-0.0304	-1.217	-1.807	0.307
				(0.342)	(0.570)	(0.161)	(0.802)	(1.207)	(0.291)
Decomposition									
Explained	-0.0808***	-0.250***	-0.0774***	-0.0830***	-0.252***	-0.0826***	-0.0256	-0.254***	-0.0555*
	(0.0267)	(0.0428)	(0.0276)	(0.0264)	(0.0416)	(0.0298)	(0.0357)	(0.0650)	(0.0317)
Unexplained	-0.278***	-0.197**	-0.110***	-0.498	-1.238**	0.0522	-1.192	-1.554	0.362
	(0.0532)	(0.0786)	(0.0426)	(0.342)	(0.569)	(0.160)	(0.802)	(1.210)	(0.290)
N	3374	1311	2367	3374	1311	2367	3374	1311	2367
Z-Score v Own Account		-0.9261	2.3682		-1.9719	1.9388		-0.9774	1.9759
Z-Score v Employer	0.9261		2.8178	1.9719		2.4619	0.9774		1.7022
Z-Score v WE	-2.3682	-2.8178		-1.9388	-2.4619		-1.9759	-1.7022	

Standard errors in parentheses

Z-Score calculated for adjusted differences in selection-corrected models

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6.3 Multinomial Logit Selection Equations

We report marginal effects for the family demographic variables included in the multinomial logit selection equations in Table 3.10.[34]

Table 3.10: Main Marginal Effects on Occupational Selection

	Female			Male		
	Own Account	Employer	WE	Own Account	Employer	WE
Married? (1=Y, 0=N)	0.0620***	0.0223***	-0.0085	0.0019	0.0177***	0.0142
	(0.0102)	(0.0068)	(0.0056)	(0.0071)	(0.0065)	(0.0089)
Dependency Ratio	0.0390***	0.0046	0.0040	0.0128***	0.0037	-0.0218***
	(0.0054)	(0.0034)	(0.0035)	(0.0049)	(0.0039)	(0.0068)
Household Size	-0.0160***	0.0021*	-0.0056***	-0.0082***	0.0021	-0.0093***
	(0.0027)	(0.0013)	(0.0016)	(0.0020)	(0.0013)	(0.0023)
Education (Years)	0.0010	0.0012*	0.0099***	-0.0014**	-0.0007	0.0130***
	(0.0010)	(0.0006)	(0.0006)	(0.0006)	(0.0005)	(0.0009)
Log of Age	0.1448***	0.0944***	0.0299***	0.0411***	0.0187**	0.0860***
	(0.0110)	(0.0073)	(0.0060)	(0.0089)	(0.0077)	(0.0114)
N	10926	10926	10926	9725	9725	9725
Log-Likelihood	-12175.7068	-12175.7068	-12175.7068	-9061.3860	-9061.3860	-9061.3860
Pseudo-R^2	0.2229	0.2229	0.2229	0.3865	0.3865	0.3865

Standard errors in parentheses
Base category is 'Out of the Labour Force'
Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported
All decomposition and selection variables included
Standard errors clustered at the household level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Higher domestic obligations push women into own account self-employment more than other occupations. Using our descriptive statistics in conjunction with Table 3.10, a 1 standard deviation increase in the dependency ratio, which we believe is our best proxy for domestic obligations, implies women are approximately 3 percent more likely to enter own account self-employment. In contrast, the dependency ratio has no statistically or economically significant effects on entry into wage-employment or high-input self-employment. The results for the other household demographic variables tell a similar story. In particular, marriage appears to make entry into low-input self-employment much more likely, so it is women with their own families that choose this type of work.

The picture is somewhat more mixed for men. The impact of the dependency ratio on male

[34] To do this, we use Hausman and McFadden's (1984) method to test whether our results are sensitive to the IIA assumption. We re-estimate the multinomial logit, omitting each category in turn, and examine whether the coefficients on all the variables change significantly. For both the female and male sub-samples, it is only when we omit the 'Out of the Labour Force' category that our results change substantially, causing us to reject the null of the Hausman-McFadden test. This suggests imposing the IIA, as is implied by the multinomial logit, may not be too restrictive.

selection into own account self-employment is smaller than for women. Using the descriptive statistics once again, a 1 standard deviation in the dependency ratio increases the probability of entry into low-input self-employment by just 0.8 percent. Moreover, the effects of the other household demographic variables on selection into own account self-employment appear to be weaker, with marriage having no statistically significant effects.

These results therefore suggest that increased domestic obligations drive women into flexible jobs, such as low-input self-employment, more than men.

We also report the marginal effects for education and age to help link our selection results to the wage equations in Section 3.6.1. The findings echo the existing literature documenting the effect of human capital on occupational choice in Ghana. Firstly, extra education increases the chances of working in wage-employment, for both women and men. However, the marginal effects of education on entry into own account work are either statistically insignificant, or slightly negative. Thus, access to schooling may serve as a barrier to entry into wage work, but not low-input self-employment. There are also positive marginal effects for age for all three of the occupational categories shown, reflecting the fact that non-participation is more likely among the young.

3.6.4 Sensitivity Analysis

We now test the sensitivity of these occupational choice results to different assumptions about endogeneity, using the framework outlined in Section 3.5.3. Since this technique is computationally intensive we make three modifications to our original selection equations. First, we reduce the number of categories from six to four, by collapsing 'Self-Employed — Agriculture', 'Unemployed', and 'Out of the Labour Force' into one category. Second, we reduce the number of variables included as controls. We retain age, education, land holdings, and some household characteristics, but are unable to include the ethnicity and regional indicator variables. Third, we normalise our variables, such that for each variable x_i, $\mathrm{E}(x_i) = 0$ and $\mathrm{Var}(x_i) = 1$.

The sensitivity of the marginal effect on the dependency ratio to different assumptions about endogeneity is shown in Table 3.11. In the leftmost column, we show the results from a multinomial logit model, having made the changes to the data described above. We then report the analogous results derived from an independent multinomial probit model, calculated with a closed-form likelihood function.[35] In the remaining columns, we report our maximum simulated

[35] We use Stata's `mprobit` to derive these results. A closed-form for the log-likelihood is found using the result

likelihood results, under different assumptions about ρ. We also show $\tilde{\rho}_{j'}$ for each sector in the final column, which indicates the level of ρ that would imply that the dependency ratio is selected *equally* by observables and the unobservables for that sector.

Table 3.11: Sensitivity Analysis

	MNL	MNP	Simulated Maximum Likelihood MNP						$\tilde{\rho}_{j'}$
			$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.2$	$\rho = 0.25$	
Female									
Own Account	0.0426	0.0428	0.0416	0.0347	0.0265	0.0194	0.0119	0.0065	0.1783
	(11.91)	(11.79)	(11.66)	(9.69)	(7.36)	(5.36)	(3.28)	(1.80)	
Employer	0.0051	0.0046	0.0047	0.0027	-0.0005	-0.0026	-0.0050	-0.0085	0.0605
	(2.25)	(2.00)	(2.39)	(1.16)	(-0.20)	(-1.08)	(-2.09)	(-3.48)	
WE	-0.0006	-0.0002	-0.0021	-0.0048	-0.0061	-0.0082	-0.0115	-0.0125	0.2110
	(-0.29)	(-0.11)	(-0.95)	(-2.29)	(-2.84)	(-3.71)	(-5.13)	(-5.47)	
Male									
Own Account	0.0133	0.0119	0.0114	0.0074	0.0056	0.0001	-0.0032	-0.0083	0.1904
	(4.07)	(3.65)	(3.56)	(2.33)	(1.75)	(0.02)	(-1.01)	(-2.57)	
Employer	0.0067	0.0066	0.0077	0.0051	0.0044	-0.0010	-0.0030	-0.0044	0.1096
	(2.65)	(2.55)	(2.93)	(1.93)	(1.66)	(-0.37)	(-1.14)	(-1.66)	
WE	-0.0192	-0.0182	-0.0183	-0.0250	-0.0299	-0.0375	-0.0458	-0.0523	0.1427
	(-4.08)	(-4.05)	(-4.00)	(-5.45)	(-6.51)	(-8.22)	(-9.86)	(-11.26)	

t-statistics in parentheses
Base category is all other working age individuals
Standard errors for simulated maximum likelihood calculated using Stata's `mprobit` command

Firstly, transforming the data leaves the main selection equation results largely unchanged, with the largest positive marginal effects pushing women into low-input self-employment. This, in itself, demonstrates the robustness of our results. Making different assumptions about endogeneity reduces the key marginal effects, as anticipated. However, the marginal effect on female selection of own account work remains positive and significant at the 10 percent level when the *common* $\rho = 0.25$. For the unobservables associated with own account self-employment, $\tilde{\rho}_{j'} = 0.1783$, so the positive effects would survive even if we assumed the dependency ratio was selected *equally* by observables and these unobservables. Indeed, even if the common ρ were set at the highest level of $\tilde{\rho}_{j'}$ (0.2110 for the wage-employed) the key marginal effect would still be positive and significant.

The extent to which we believe that equal selection on observables and unobservables is sufficient to demonstrate that our results are robust to endogeneity depends on the effectiveness of our control set at explaining the dependency ratio. If the control variables are strongly related to the dependency ratio, then the restriction that unobservables have an equal effect may be

due to Dunnett (1989). As in the multinomial logit, and indeed our simulations, this estimator assumes there is no correlation between the error terms for utility in each occupation, as in Equation (3.36).

quite conservative. However, if the control variables do little to explain the dependency ratio, we may expect unobservables to do more selection than observables (Oster, 2013). Although we are somewhat constrained by the computational intensity of this estimator, we have maintained control variables on age, education, land holdings, and other household characteristics (such as household size and spouse education). We believe these controls drive the dependency ratio sufficiently for the 'equal selection' condition, encapsulated by $\tilde{\rho}_{j'}$, to provide a useful benchmark against which to judge our results' robustness to endogeneity.

In the next section, we consider sub-samples of the population in which we believe endogeneity is likely to have less of an effect. However, even in the full sample, it appears our results are somewhat robust to concerns about the endogeneity of domestic obligations.

3.7 Robustness and Heterogeneity

3.7.1 Exclusion Restrictions

We begin by assessing whether our earnings decompositions are robust to using different exclusion restrictions in the Lee and modified Dubin-McFadden selection equations. Overall, our results remain largely stable to alternative combinations of the excluded variables.

One specific critique of the exclusion restrictions we deploy may be that land proxies for the *level* of capital available for running a business, rather than simply the ability to finance start-up costs. Insofar as the level as capital affects enterprise profits, this means land should not be excluded from our earnings equations. We show that this does not substantially change our results by re-estimating our earnings decompositions, but moving land from the excluded set into the earnings equations itself. The new decompositions are reported in Table 3.20 of Appendix 3.F.

In fact, this appears to strengthen our main results. Looking at the Lee specifications, the earnings differential in own account self-employment now jumps up to around 70 percent when controlling for selection, whilst the in wage-employment, the gap is, if anything, *positive* in favour of women. These findings are also reflected in the Dubin-McFadden adjusted equations.

3.7.2 Controlling for Industry

By splitting the sample of self-employed workers into those that do and do not employ labour besides their own, we try to capture an important component of *intra*-sector heterogeneity. However, it may be that there are important differences in the types of jobs that men and women do, even amongst the own account workers and the employers. As such, the unexplained portion of the earnings gap may simply reflect the fact that women end up working in less lucrative industries than men, and that we have failed to control for this variation.

To examine this issue, we use the International Standard Industrial Classification (ISIC) codes reported for each respondent's main job to split the data into four categories — 'Manufacturing', 'Primary/Agriculture', 'Retail', and 'Services/Other' — and then control for this variation.[36] As we show in Table 3.17 in Appendix 3.C, amongst the self-employed, women tend more towards retail activities, whilst men are better represented in services. We take the residuals from our original wage equations (estimated by OLS), and plot these separately for each industry in Figure 3.11. It emerges that, for both women and men in low- and high-input self-employment, there are vital differences between the distributions of these residuals by industry. Overall, services and manufacturing enterprises appear to outperform retailers, conditional on observable human capital. Thus, given women's high representation in retail, we would expect that controlling for industry would cause our earnings equations to explain far more of the differences between women and men, and therefore reduce the proportion of the earnings gap which is unexplained.

To test this, we add dummy variables for each main industrial category to our earnings equations and hence the set of observable variables entering our decompositions. These new decompositions are reported in Table 3.21 of Appendix 3.F. Contrary to our priors, however, controlling for industry has little impact on the proportion of the sex earnings gap that we can explain, for both low- and high-input self-employment.

We believe that simply controlling for industry does little to explain the aggregate sex earnings gaps for the low- and high-input self-employed, because earnings differentials persist, to a greater or lesser degree, across the different industries. In general, the earnings gap appears to be wider in retail and primary/agriculture activities compared with services and manufacturing. To illustrate this, we use the same Blinder-Oaxaca decomposition technique, focussing on the sub-samples for each industry amongst the own account self-employed. These results are reported in

[36]Even though we focus on the non-farm self-employed, there are still some primary industries and other activities related to agriculture, such as forestry, fishing, and animal husbandry.

Figure 3.11: Residuals from OLS Wage Equations Separated by Industry

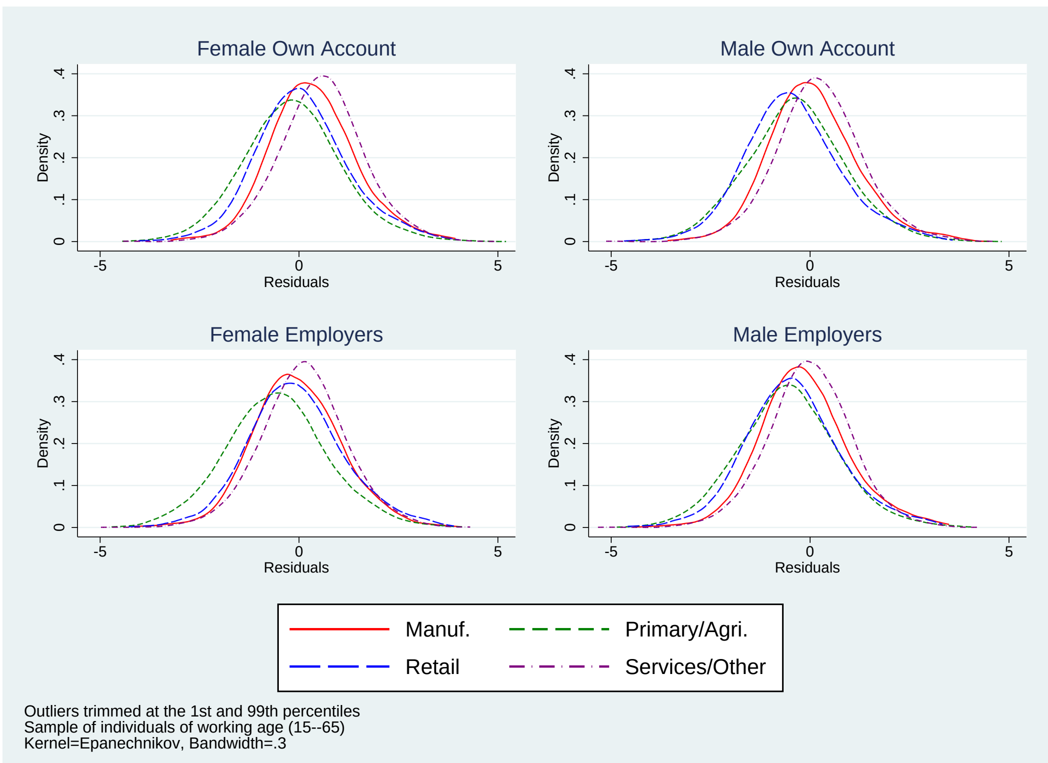


Table 3.12 [37] We cannot tenably control for selection across different industries, so the earnings equations behind these decompositions are estimated using OLS.

Table 3.12: Blinder-Oaxaca Decompositions by Industry for Own Account Workers

	All	Manuf.	Primary/Agri.	Retail	Services/Other
Differential					
Female Prediction	-2.898***	-2.786***	-3.415***	-2.741***	-2.648***
	(0.0258)	(0.0559)	(0.0595)	(0.0360)	(0.0830)
Male Prediction	-2.536***	-2.533***	-2.959***	-2.174***	-2.251***
	(0.0475)	(0.0991)	(0.0949)	(0.102)	(0.0867)
Difference	-0.363***	-0.253**	-0.456***	-0.567***	-0.397***
	(0.0540)	(0.114)	(0.112)	(0.108)	(0.120)
Decomposition					
Explained	-0.0607*	-0.143***	-0.218***	-0.0949***	-0.189***
	(0.0333)	(0.0448)	(0.0367)	(0.0317)	(0.0708)
Unexplained	-0.302***	-0.110	-0.238**	-0.472***	-0.208
	(0.0557)	(0.110)	(0.109)	(0.102)	(0.132)
N	3374	700	846	1385	443
Z-Score v Manuf.			-1.2733	-2.0009	-0.8731
Z-Score v Primary/Agri.		1.2733		-0.7097	0.3588
Z-Score v Retail		2.0009	0.7097		1.0490
Z-Score v Services/Other		0.8731	-0.3588	-1.0490	

Standard errors in parentheses
All specifications estimated by OLS, without selection correction
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results in Table 3.12 also demonstrate that the proportion of the sex earnings gap that can be explained by observable characteristics differs substantially between industries. Once again, the most striking differences are between manufacturing and retail. In the former, nearly 60 percent of the earnings gap can be accounted for by observable factors, whereas in the latter, just 15 percent of the differential can be explained.

As such, although controlling for industry does little to help us explain the aggregate earnings gap in low- and high-input self-employment, this dimension of intra-sector heterogeneity appears to be important. This is an issue to which we return in Chapter 4.

3.7.3 Respecifying Domestic Obligations

Thus far, we have mainly proxied for domestic obligations using the household dependency ratio, whilst also considering the size of the household and each individual's marital status. We now

[37] Even though the size of the raw earnings gap differs between each industry, we can only reject the null that the earnings differentials are the same when comparing retail and manufacturing. This is shown by the z-scores reported in Table 3.12. We may lack statistical power because the sample sizes become small at this level of granularity regarding job-type.

try and disentangle which household dependents drive our results by respecifying the selection equations, including the number of infants (aged < 2), young children (aged 2–4), other children (aged 5–14), and elders (aged > 65), as shown in Table 3.13.

Table 3.13: Marginal Effects for Occupational Selection:
Respecifying Domestic Obligations

	Female			Male		
	Own Account	Employer	WE	Own Account	Employer	WE
Married? (1=Y, 0=N)	0.0611***	0.0238***	-0.0089	0.0016	0.0160**	0.0091
	(0.0102)	(0.0068)	(0.0056)	(0.0072)	(0.0065)	(0.0089)
No. Infants (< 2 years) in HH	0.0350***	-0.0017	0.0073	0.0143**	0.0072	0.0198**
	(0.0087)	(0.0053)	(0.0070)	(0.0065)	(0.0052)	(0.0086)
No. Young Children (2–4 years) in HH	0.0426***	-0.0086**	0.0036	0.0116**	0.0131***	-0.0064
	(0.0074)	(0.0044)	(0.0043)	(0.0055)	(0.0039)	(0.0070)
No. Older Children (5–14 years) in HH	0.0282***	0.0059**	0.0038	0.0096**	0.0037	-0.0058
	(0.0048)	(0.0026)	(0.0029)	(0.0038)	(0.0029)	(0.0045)
No. Elders (> 65 years) in HH	0.0014	-0.0126*	-0.0012	0.0115	0.0028	-0.0336**
	(0.0119)	(0.0070)	(0.0080)	(0.0108)	(0.0086)	(0.0144)
Household Size	-0.0300***	0.0007	-0.0072***	-0.0119***	-0.0004	-0.0093***
	(0.0036)	(0.0018)	(0.0021)	(0.0026)	(0.0020)	(0.0030)
N	10926	10926	10926	9725	9725	9725
Log-Likelihood	-12127.0104	-12127.0104	-12127.0104	-8999.3215	-8999.3215	-8999.3215
Pseudo-R^2	0.2260	0.2260	0.2260	0.3907	0.3907	0.3907

Standard errors in parentheses
Base category is 'Out of the Labour Force'
Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported
All decomposition and non-household demographic selection variables included
Standard errors clustered at the household level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These results echo our initial findings in terms of the dependency ratio from Table 3.10. Moreover, we see that the number of children has the largest marginal effects on female entry into low-input self-employment, whilst the number of elders in the household matters less. A 1 standard deviation increase in the number of young children in the household increases the likelihood of female participation in own account self-employment by around 3 percent, whilst the effects on entry into wage-employment and high-input self-employment are minimal. In contrast, the marginal effect on the number of elders for female entry into low-input self-employment is small and statistically insignificant. Also, as before, the effects are far weaker for men.

Respecifying the selection equations using alternative proxies of domestic obligations not only reinforces our main results, but also allows us to unpack how job flexibility may influence occupational selection. It appears to be extra young children, which push women into self-employment rather than the presence of elderly relatives. Insofar as young children require some level of *constant* care, this suggests that the multi-tasking and minimum hours models outlined in Section 3.3 are more suitable for interpreting the data. In contrast, the number of elders, whose care requirements are more likely to vary over time, has little effect. Coping with *volatile* domestic obligations as in the adjustment costs model, thus appears to be less important.

3.7.4 Heterogeneity by Marital Status

The potential endogeneity of domestic obligations may be less important in certain sub-samples of the population. For example, we expect that choices about family structure are likely to be decided more by married individuals rather than unmarried individuals. As such, endogeneity issues should be less severe for the unmarried sub-sample.

To test this hypothesis, we begin by re-estimating our main multinomial logit results, splitting the sample by marital status. These results are reported in Tables 3.14 and 3.15. The dependency ratio appears to have stronger effects for the unmarried sub-sample, for both women and men.

Table 3.14: Main Marginal Effects on Female Job Selection by Marital Status

	Unmarried			Married		
	Own Account	Employer	WE	Own Account	Employer	WE
Household Size	-0.0186***	-0.0001	-0.0055***	-0.0049**	0.0044***	-0.0011
	(0.0023)	(0.0015)	(0.0020)	(0.0025)	(0.0013)	(0.0012)
Dependency Ratio	0.0520***	0.0068*	0.0022	0.0363***	0.0069	0.0044
	(0.0058)	(0.0038)	(0.0058)	(0.0087)	(0.0057)	(0.0043)
N	5558	5558	5558	5368	5368	5368
Log-Likelihood	-5619.7711	-5619.7711	-5619.7711	-6783.2227	-6783.2227	-6783.2227
Pseudo-R^2	0.2544	0.2544	0.2544	0.1369	0.1369	0.1369

Standard errors in parentheses
Base category is 'Out of the Labour Force'
Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported
Standard errors clustered at the household level when computationally feasible
Ethnicity dummies and exclusion restrictions omitted to aid computation, but all other decomposition variables included
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To test our prior that the endogeneity of domestic obligations is less problematic for the unmarried sub-sample, we repeat our sensitivity analysis using maximum simulated likelihood for different assumptions about the error variance-covariance matrix. As before, we collapse the number of categories from six to four, reduce the number of controls, and normalise our variables to aid computation. The results are shown in Table 3.16.

The results for the unmarried sub-sample are more robust to different assumptions about endogeneity, than the full sample. Now the marginal effect of the dependency ratio female on entry into own account self-employment remains positive and significant at the 1 percent level

Table 3.15: Main Marginal Effects on Male Job Selection by Marital Status

	Unmarried			Married		
	Own Account	Employer	WE	Own Account	Employer	WE
Household Size	-0.0079***	-0.0030***	-0.0109***	-0.0036*	0.0047***	0.0049**
	(0.0016)	(0.0012)	(0.0023)	(0.0021)	(0.0016)	(0.0025)
Dependency Ratio	0.0169***	0.0074*	-0.0301***	0.0112	0.0075	-0.0222**
	(0.0052)	(0.0038)	(0.0098)	(0.0080)	(0.0068)	(0.0095)
N	5467	5467	5467	4258	4258	4258
Log-Likelihood	-4653.6852	-4653.6852	-4653.6852	-4650.1200	-4650.1200	-4650.1200
Pseudo-R^2	0.3276	0.3276	0.3276	0.2236	0.2236	0.2236

Standard errors in parentheses
Base category is 'Out of the Labour Force'
Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported
Standard errors clustered at the household level when computationally feasible
Ethnicity dummies and exclusion restrictions omitted to aid computation, but all other decomposition variables included
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.16: Sensitivity Analysis for the Unmarried Sub-Sample

	MNL	MNP	Simulated Maximum Likelihood MNP						$\tilde{\rho}_{j'}$
			$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.2$	$\rho = 0.25$	
Female									
Own Account	0.0398	0.0391	0.0418	0.0352	0.0300	0.0236	0.0171	0.0130	0.1419
	(10.38)	(9.91)	(10.31)	(8.60)	(7.32)	(5.75)	(4.16)	(3.14)	
Employer	0.0038	0.0032	0.0062	0.0044	0.0033	0.0006	-0.0010	-0.0028	0.0859
	(1.60)	(1.35)	(2.00)	(1.80)	(1.40)	(0.26)	(-0.40)	(-1.17)	
WE	-0.0022	-0.0022	-0.0056	-0.0066	-0.0107	-0.0122	-0.0142	-0.0194	0.1898
	(-0.63)	(-0.67)	(-1.82)	(-2.12)	(-3.43)	(-3.79)	(-4.41)	(-5.86)	
Male									
Own Account	0.0114	0.0105	0.0136	0.0114	0.0076	0.0066	0.0002	-0.0016	0.2840
	(3.09)	(2.76)	(3.48)	(2.93)	(1.96)	(1.73)	(0.06)	(-0.40)	
Employer	0.0022	0.0011	0.0020	0.0002	-0.0006	-0.0017	-0.0036	-0.0052	0.2424
	(0.78)	(0.39)	(0.79)	(0.09)	(-0.24)	(-0.68)	(-1.29)	(-1.91)	
WE	-0.0311	-0.0319	-0.0326	-0.0400	-0.0454	-0.0530	-0.0575	-0.0679	0.2150
	(-4.47)	(-4.99)	(-5.10)	(-6.23)	(-7.02)	(-8.12)	(-8.73)	(-10.12)	

t-statistics in parentheses
Base category is all other working age individuals
Standard errors for simulated maximum likelihood calculated using Stata's `mprobit` command

and is more than double the size in the full sample when the common ρ is set at 0.25. Also, $\tilde{\rho}_{j'}$ is slightly lower for all three occupations for the unmarried sub-sample, focussing on the women. This suggests that the lower bound estimates of the marginal effects, under the assumption that selection of the dependency ratio by observables and unobservables is equal, are somewhat higher than in the full sample.

3.8 Conclusion

In this chapter, we investigate whether non-monetary characteristics of certain jobs cause female and male workers to select occupations differently. In particular, we test the hypothesis that women are drawn into low-input self-employment activities more than men, as this allows them to balance their career with domestic obligations, such as caring for children. We also examine what these patterns of occupational choice imply for labour market outcomes, by decomposing earnings gaps in different economic sectors whilst controlling for selection.

We use a simple model, couched in terms of time allocation, to fix ideas about how the concept of 'job flexibility' should be understood, and how this may influence occupational choice. We explore three possibilities. Firstly, we consider the idea that certain occupations may allow individuals to multi-task such that market and domestic work may, to some extent, be undertaken simultaneously. Secondly, we examine the implications of minimum work hours for more formal job activities. Finally, we build a simple model in which domestic obligations are imposed stochastically, and individuals face occupation-specific costs of adjusting their 'contracted' allocation of time. The data suggest that the first two stories, namely multi-tasking and minimum work hours, may be more important for Ghana.

Turning to the empirical results, we find that controlling for selection leaves the earnings gap amongst the wage-employed small and statistically insignificant. This suggests that in wage-employment, unobservable aspects of income-generating ability are higher amongst men than among women. In contrast, adjusting for selection drastically widens the earnings gap amongst the self-employed. Women have higher unobservable income-generating ability than men in self-employment, such that the observed earnings gap actually obfuscates the true difference in *potential* earnings.

Naturally, these findings beg the question over what explains the large gaps in potential earnings amongst the self-employed. One innovation in our approach is to allow for heterogeneity in the self-employment sector, depending on whether entrepreneurs use labour besides their own. However, it appears that there may be even finer level sorting across jobs within each occupational category, since even within low- and high-input self-employment, the *industries* in which businesses engage differs between women and men. In particular, female participation in self-employment is dominated by retail activities, whilst the picture for manufacturing and services is far more balanced by sex. Nonetheless, simply controlling for this variation in industry does little to explain why the earnings of women and men differ so much for low- and high-input self-employment as a whole. It appears that, in part, this may be because sex earnings gaps persist, to a varying degree, even within each industry.

We interrogate the selection equations used to correct our earnings decompositions to test whether domestic obligations, proxied using household demographic variables, affect occupational choice. We find that females with greater domestic obligations are more likely to select into low-input self-employment, which we show are characterised by greater flexibility. The analogous effects are much smaller for men. It appears that extra young children drive most of this effect, and that the number of elders in the household has little effect on women's occupational choice.

Household structure may be endogenous to occupational choice, as we anticipate that there may be unobservable factors — such as family stability, understanding of family planning, and social norms — which drive both variables. We develop a new estimator based on the idea of using selection on observables as a guide to selection on unobservables to tackle this concern. Our main finding, that women with greater domestic obligations are pushed into own account work, is robust to very conservative assumptions about the nature of endogeneity. We also find that this potential endogeneity is likely to be less problematic for the sub-sample of unmarried individuals. We argue that this is because these individuals have less control over the structure of their household.

The fact that women participate most in sectors where they are rewarded substantially less than men presents something of an empirical puzzle. This chapter partially helps to reconcile these patterns. We find that women may be drawn towards certain types of self-employment for the non-pecuniary benefit of job flexibility. This may outweigh the low returns available to them in these activities. Also, although we control for selection across one key juncture in the self-employment sector — namely whether or not to employ others' labour — our results suggest there may be extra sorting across industry and job-type.

However, even accounting for these issues, large differences in potential earnings coexist with high female participation in self-employment. This chapter emphasises *sorting*, according to productivity and preferences, as the main determinant of occupational choice. However, as we argued in Chapter 1, if labour markets in sub-Saharan Africa are *segmented*, this approach may be incomplete. Thus, in Chapter 4 we turn to theory to assess the how frictions in the process of matching individuals to jobs may differentially influence the labour market outcomes of women and men.

3.A Analytical Solutions

3.A.1 Multi-Tasking

To derive analytical expressions for indirect utility in the multi-tasking model, we begin by writing the Lagrangian for the time allocation problem. We log linearise the utility function for simplicity.

$$\mathcal{L} = \beta \ln(x_m) + (1-\beta)\ln(x_d) - \lambda_1 \left[x_m + \left(\frac{\hat{w}\pi^2}{\rho} \right) x_d - \hat{w}\pi \right] + \lambda_2 \left[x_d - \gamma \right] \tag{3.48}$$

The resulting Kuhn-Tucker conditions may then be written:

$$\frac{\partial \mathcal{L}}{x_m} = \frac{\beta}{x_m} - \lambda_1 = 0 \tag{3.49}$$

$$\frac{\partial \mathcal{L}}{x_d} = \frac{1-\beta}{x_d} - \lambda_1 \left(\frac{\hat{w}\pi^2}{\rho} \right) + \lambda_2 = 0 \tag{3.50}$$

$$\frac{\partial \mathcal{L}}{\lambda_1} = -\left[x_m + \left(\frac{\hat{w}\pi^2}{\rho} \right) x_d - \hat{w}\pi \right] = 0 \tag{3.51}$$

$$\frac{\partial \mathcal{L}}{\lambda_2} = x_d - \gamma = 0 \tag{3.52}$$

Given the monotonicity of the utility function, the full income constraint necessarily binds, such that $\lambda_1 > 0$. However, the resulting choices for x_m and x_d will depend on whether the domestic obligations constraint binds ($\lambda_2 > 0$) or is slack ($\lambda_2 = 0$). There are thus two scenarios to consider.

Scenario 1: $\lambda_2 = 0$

If $\lambda_2 = 0$ and the domestic obligations constraint is slack, the individual simply chooses x_m and x_d subject to the income constraint — this is a typical maximisation problem. The

Marshallian demands that result are:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} \beta \hat{w} \pi \\ (1-\beta)\frac{\rho}{\pi} \end{pmatrix} \tag{3.53}$$

Scenario 2: $\lambda_2 > 0$

If $\lambda_2 > 0$ and the domestic obligations constraint binds, the consumption of domestic goods is fixed at γ. This, in turn, determines the amount of time left over for doing market work. Thus, the resulting consumptions may be written:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} \hat{w}\pi\left(1 - \frac{\pi\gamma}{\rho}\right) \\ \gamma \end{pmatrix} \tag{3.54}$$

The domestic obligations constraint will only bind if $(1-\beta)\frac{\rho}{\pi} \leq \gamma$. Thus, we can write the indirect utility function as:

$$V(\beta, \pi, \gamma, \omega(\pi), \rho) = \begin{cases} \beta \ln[\beta \hat{w} \pi] + (1-\beta)\ln[(1-\beta)\frac{\rho}{\pi}] & \text{if } (1-\beta)\frac{\rho}{\pi} > \gamma \\ \beta \ln\left[\hat{w}\pi\left(1 - \frac{\pi\gamma}{\rho}\right)\right] + (1-\beta)\ln[\gamma] & \text{if } (1-\beta)\frac{\rho}{\pi} \leq \gamma \end{cases} \tag{3.55}$$

We use this formulation of the indirect utility function to evaluate the welfare of individuals with different levels of domestic obligations in different jobs, allowing us to predict their occupational choice.

3.A.2 Minimum Hours

To derive indirect utility in the minimum hours model, we once again begin by writing out the Lagrangian.

$$\mathcal{L} = \beta \ln(x_m) + (1-\beta)\ln(x_d) - \lambda_1 \left[x_m + \left(\frac{k\tau}{\rho}\right) x_d - k\tau \right] + \lambda_2 \left[x_d - \gamma \right] + \lambda_3 \left[\frac{x_m}{k\tau} - \tau \right] \quad (3.56)$$

The Kuhn-Tucker conditions may then be written:

$$\frac{\partial \mathcal{L}}{x_m} = \frac{\beta}{x_m} - \lambda_1 + \lambda_3 = 0 \quad (3.57)$$

$$\frac{\partial \mathcal{L}}{x_d} = \frac{1-\beta}{x_d} - \lambda_1 \left(\frac{k\tau}{\rho}\right) + \lambda_2 = 0 \quad (3.58)$$

$$\frac{\partial \mathcal{L}}{\lambda_1} = -\left[x_m + \left(\frac{k\tau}{\rho}\right) x_d - k\tau \right] = 0 \quad (3.59)$$

$$\frac{\partial \mathcal{L}}{\lambda_2} = x_d - \gamma = 0 \quad (3.60)$$

$$\frac{\partial \mathcal{L}}{\lambda_3} = \frac{x_m}{k\tau} - \tau = 0 \quad (3.61)$$

Since the utility function is monotonic, we know that the income constraint binds and $\lambda_1 > 0$. We also know that it is not possible for both the domestic obligations constraint and the lower bound on market work to simultaneously bind, so it cannot be that $\lambda_2 > 0$ and $\lambda_3 > 0$. Thus we are left with three potential scenarios to consider.

Scenario 1: $\lambda_2 = \lambda_3 = 0$

If both the domestic obligations constraint and the lower bound on market work are slack, then the individual simply maximises utility subject to the income constraint. This produces regular Marshallian demands of the form:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} \beta k\tau \\ (1-\beta)\rho \end{pmatrix} \tag{3.62}$$

This situation will prevail if, both $\beta > \tau$ and $(1-\beta)\rho > \gamma$, as this implies both extra constraints are too small to affect time allocation and consumption patterns.

Scenario 2: $\lambda_2 > 0; \lambda_3 = 0$

In this scenario, both the domestic obligations constraint and the income constraint bind. This occurs, if $\beta > \tau$ and $(1-\beta)\rho \leq \gamma$. The resulting outcomes for market and domestic goods are:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} k\tau\left(1 - \frac{\gamma}{\rho}\right) \\ \gamma \end{pmatrix} \tag{3.63}$$

Scenario 3: $\lambda_2 = 0; \lambda_3 > 0$

Finally, it may be that the income constraint and the lower bound on market work hours bind. This will happen if $\beta \leq \tau$ and $(1-\beta)\rho > \gamma$. The resulting consumption outcomes are:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} k\tau^2 \\ \rho(1-\tau) \end{pmatrix} \tag{3.64}$$

Thus, we can now write down the indirect utility function, taking these three potential scenarios into account, as in Equation (3.65). This is the function we plot, to recover the occupational choice of individuals with different domestic obligations.

$$V(\beta,\tau,\gamma,k,\rho) = \begin{cases} \beta \ln\big[\beta k\tau\big] + (1-\beta)\ln\big[(1-\beta)\rho)\big] & \text{if } \tau \text{ and } (1-\beta)\rho > \gamma \text{ and } \beta > \tau \\ \beta \ln\left[k\tau\left(1-\frac{\gamma}{\rho}\right)\right] + (1-\beta)\ln[\gamma] & \text{if } \tau \text{ and } (1-\beta)\rho \leq \gamma \text{ and } \beta > \tau \\ \beta \ln[k\tau^2] + (1-\beta)\ln\big[\rho(1-\tau)\big] & \text{if } \tau \text{ and } (1-\beta)\rho > \gamma \text{ and } \beta \leq \tau \end{cases} \tag{3.65}$$

3.A.3 Adjustment Costs

In this appendix we show how to find the reaction function for actual time spent on market work, $t_m(\Gamma)$.

To do this, we write the maximisation problem that faces the individual in Period 2. We recall that, by Period 2, the level of γ has been realised as either $\tilde{\gamma}$ or 0, and $\hat{t_m}$ has been chosen.

$$\max_{t_m} \; x_m = t_m hc\left(1 - \frac{c}{2}(t_m - \hat{t_m})^2\right) \tag{3.66}$$

$$\text{subject to } t_m \leq 1 - \gamma \tag{3.67}$$

The Lagrangian for the problem is:

$$\mathcal{L} = t_m hc\left(1 - \frac{c}{2}(t_m - \hat{t_m})^2\right) - \lambda[t_m - 1 + \gamma] \tag{3.68}$$

We can then write the resulting Kuhn-Tucker conditions.

$$\frac{\partial \mathcal{L}}{t_m} = hc - \frac{hc^2}{2}\left(3{t_m}^2 - 4t_m\hat{t_m} + \hat{t_m}^2\right) - \lambda = 0 \tag{3.69}$$

$$\frac{\partial \mathcal{L}}{\lambda} = -[t_m - 1 + \gamma] = 0 \tag{3.70}$$

As such, there are two possible scenarios depending on whether or not the constraint binds. If the constraint does in fact bind, and $\lambda > 0$, the optimal choice of t_m is simply $1 - \gamma$. However, if the constraint does not bind, we use the quadratic formula to find the optimal choice of t_m, which we label ${t_m}^*$, as a function of $\hat{t_m}$ and c.

$$t_m(\Gamma) = \begin{cases} 1 - \gamma & \text{if } {t_m}^* \geq 1 - \gamma \\ {t_m}^* = \frac{2c\hat{t_m} + \sqrt{c^2 \hat{t_m}^2 + 6c}}{3c} & \text{otherwise} \end{cases} \tag{3.71}$$

We can now substitute the reaction function $t_m(\Gamma)$ back into the Period 1 objective function. This allows us to write an unconstrained maximisation, which the individual solves to choose their preferred contracted hours, $\hat{t_m}$, given the parameters of the model.

$$\begin{aligned} \max_{\hat{t_m}} \ \mathrm{E}[U] = q \ln & \left[hc\left(1 - \frac{c}{2}\big(\hat{t_m} - t_m(\Gamma|\gamma = \tilde{\gamma})\big)^2\right)(t_m(\Gamma|\gamma = \tilde{\gamma})\right] \\ + (1-q) \ln & \left[hc\left(1 - \frac{c}{2}\big(\hat{t_m} - t_m(\Gamma|\gamma = 0)\big)^2\right)(t_m(\Gamma|\gamma = 0)\right] \end{aligned} \tag{3.72}$$

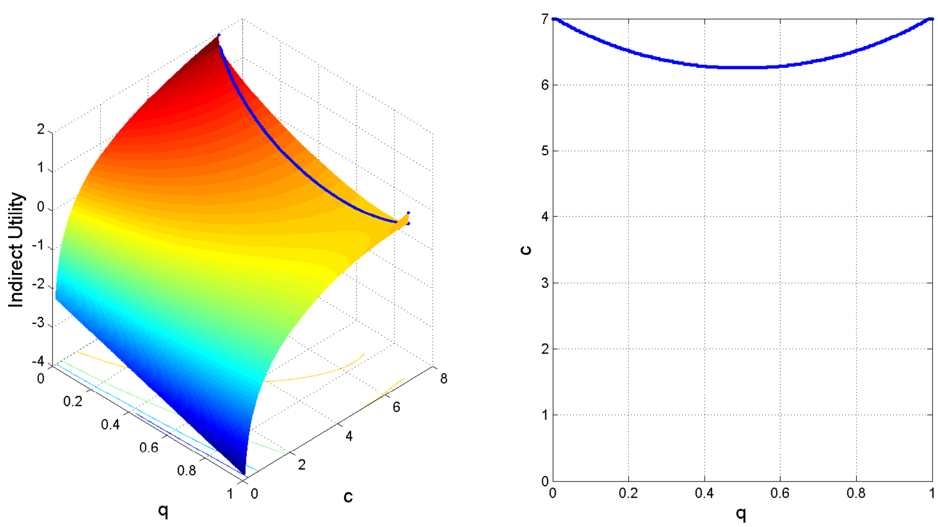
Once we solve for the optimal $\hat{t_m}$, we can recover the expected utility from different jobs (c) for individuals with different domestic obligations, ($\tilde{\gamma}$ and q). Insofar as individuals select the jobs which give highest welfare ex ante, this enables us examine occupational choice. Since solving the model analytically is not straightforward, we use the MATLAB program `fmincon` to maximise Equation (3.72) with respect to $\hat{t_m}$, and calculate expected utility.

3.B Changing q in the Adjustment Costs Model

In Section 3.3.4 we showed that, holding q constant, individuals with higher levels of $\tilde{\gamma}$ prefer jobs with lower adjustment costs in spite of the lower initial returns to market work. We now take a different approach to assessing the impact of domestic obligations on occupational choice. In particular, we hold $\tilde{\gamma}$ constant and vary the level of q. That is, we change the likelihood that a shock to domestic obligations, of a given size, affects individuals' time allocation.

We plot expected utility, holding $\tilde{\gamma} = 0.8$ and varying q in Panel A of Figure 3.12. For all values of $0 < q < 1$, expected utility is at first increasing but then decreasing in c. As in Section 3.3.4 individuals are at first attracted by the higher initial wages of jobs with higher c, but eventually the threat of larger adjustment penalties limits the extra expected utility from more rigid jobs. It turns out this turning point arrives at lower values of c as $q \to 0.5$. This implies the optimal level of job flexibility has a non-linear relationship with domestic obligations. This is shown by the blue ridge in Panel A of 3.12, and plotted again in two dimensions in Panel B.

Figure 3.12: Occupational Selection in an Adjustment Costs Model — Changing q



The non-linear relationship we see in Figure 3.12 stems from the fact that a higher q at first increases, but then decreases, the variance of γ. In particular, $\frac{\partial \mathrm{Var}(\gamma)}{\partial q} = \tilde{\gamma}^2(1-2q)$. Thus, for $q < 0.5$, a higher level of q means a higher variance for γ. Given the assumption of risk averse preferences, these individuals try to reduce the variation in their potential outcomes by choosing an occupation in which job rigidity, c, is lower, even if this means foregoing base wages, hc. For $q > 0.5$, however, a higher value for q means a higher variance for γ. Thus, these individuals are willing to expose themselves to extra adjustment penalties, in pursuit of higher base wages, hc, in spite of the increased likelihood that shocks to domestic obligations will actually occur.[38]

We do not focus on individual heterogeneity in q in our main analysis, largely because these results are specific to our assumptions about the form of the shocks to domestic obligations. In particular, the mean of γ is monontonically increasing in q, but the variance has a non-monotonic relationship.

3.C Extra Summary Statistics

Table 3.17: Detailed Occupational Breakdown for the Self-Employed by Sex

	Sex					
	Female		Male		Total	
	No.	%	No.	%	No.	%
Own Account						
Manufacturing	551	15.96	161	4.66	712	20.62
Primary/Agriculture	584	16.91	278	8.05	862	24.96
Retail	1244	36.03	164	4.75	1408	40.78
Services/Other	261	7.56	210	6.08	471	13.64
Total	2640	76.46	813	23.54	3453	100.00
Employer						
Manufacturing	257	19.28	120	9.00	377	28.28
Primary/Agriculture	213	15.98	170	12.75	383	28.73
Retail	208	15.60	70	5.25	278	20.86
Services/Other	129	9.68	166	12.45	295	22.13
Total	807	60.54	526	39.46	1333	100.00

[38] $\frac{\partial \mathrm{E}(\gamma)}{\partial q} = \tilde{\gamma} \geq 0$

Table 3.18: Summary Statistics by Sex and Occupation

	Wage Employed				Non-Farm SE (Own Account)				Non-Farm SE (Employer)			
	Male		Female		Male		Female		Male		Female	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Age (Years)	37.85	36.00	34.36	32.00	37.94	37.00	37.21	36.00	39.15	38.00	40.12	40.00
Education (Years)	10.47	10.00	10.62	11.00	7.29	9.00	5.03	5.00	7.44	9.00	4.58	3.00
Tenure (Months)	112.81	72.00	89.98	43.00	141.52	108.00	105.01	63.50	153.99	120.00	146.53	120.00
Married? (1=Y, 0=N)	0.57	1.00	0.41	0.00	0.61	1.00	0.63	1.00	0.73	1.00	0.67	1.00
Household Size	4.22	4.00	4.52	4.00	4.53	4.00	5.23	5.00	5.19	5.00	6.15	6.00
Dependency Ratio	0.52	0.33	0.64	0.50	0.70	0.50	0.96	1.00	0.78	0.67	0.92	0.75
Active Ratio in HH (excl. Individual)	0.24	0.17	0.33	0.20	0.24	0.17	0.26	0.20	0.23	0.14	0.25	0.20
HH Land (Acres)	5.24	0.00	2.71	0.00	6.10	0.00	8.06	0.00	6.09	0.00	7.05	0.00
HH Owns Land? (1=Y, 0=No)	0.27	0.00	0.19	0.00	0.39	0.00	0.40	0.00	0.42	0.00	0.43	0.00
Farmer Father? (1=Y, 0=N)	0.46	0.00	0.34	0.00	0.59	1.00	0.62	1.00	0.60	1.00	0.65	1.00
Professional Father? (1=Y, 0=N)	0.10	0.00	0.13	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00
Service Sector Father? (1=Y, 0=N)	0.19	0.00	0.23	0.00	0.13	0.00	0.14	0.00	0.16	0.00	0.14	0.00
Farmer Mother? (1=Y, 0=N)	0.38	0.00	0.28	0.00	0.50	0.00	0.53	1.00	0.51	1.00	0.56	1.00
Professional Mother? (1=Y, 0=N)	0.02	0.00	0.04	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Service Sector Mother? (1=Y, 0=N)	0.05	0.00	0.06	0.00	0.04	0.00	0.03	0.00	0.04	0.00	0.03	0.00
HH Unearned Income? (1=Y, 0=N)	0.03	0.00	0.02	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.01	0.00
Observations	1768		658		813		2640		526		807	

Sample of individuals of working age (15–65)

3.D Closed-Form Log-Likelihood Function

In order to estimate the log-likelihood function without simulation, we need to be able to calculate the probability for each outcome, $\Pr(y_i = j \mid \boldsymbol{x_i}, D_i)$. For a general error variance-covariance matrix, this would require evaluating a complicated multidimensional integral. However, since the distribution of $(\zeta_{i2'}, \zeta_{i3'}, ...\zeta_{iJ'} \mid \upsilon_i)$ is *exchangeable* (or characterised by compound symmetry), we can use the result due to Dunnett (1989) to reduce this multidimensional integral to one dimension. We can thus write the probability that individual i chooses k as:[39]

$$Pr(y_i = k \mid \boldsymbol{x_i}, D_i) = \frac{1}{\sqrt{\pi}} \int_0^{\infty} \Bigg\{ \prod_{j=1}^{J-1} \Phi\Bigg(\frac{-\big[z\sqrt{2}\sqrt{1-\rho^2}\big] - \lambda_{ij}}{\rho} \Bigg) + \prod_{j=1}^{J-1} \Phi\Bigg(\frac{\big[z\sqrt{2}\sqrt{1-\rho^2}\big] - \lambda_{ij}}{\rho} \Bigg) \Bigg\} e^{-z^2} dz \tag{3.73}$$

Empirically, we substitute for λ_{ij} using:

$$\lambda_{ij} = \boldsymbol{x_i^\intercal \beta_j} + \psi_j D_i + \rho\sigma^{-1}(D_i - \boldsymbol{x_i^{'}\pi}) \tag{3.74}$$

Substituting Equation (3.73) into Equation (3.46) results in a closed-form for the log likelihood.

[39]Following the help file for Stata's `mprobit` command, the Dunnett (1989) result can be approximated using Gaussian quadrature.

3.E Pooled Earnings Equations

Table 3.19: Pooled Wage Equations

	OLS			Lee			Dubin-McFadden		
	Own Account	Employer	WE	Own Account	Employer	WE	Own Account	Employer	WE
Male Dummy	0.1425***	0.0873**	0.0477**	0.1461***	0.0885**	0.0479**	0.1343***	0.0750*	0.0475**
	(0.0260)	(0.0369)	(0.0194)	(0.0251)	(0.0371)	(0.0192)	(0.0245)	(0.0403)	(0.0187)
Female Dummy	-0.1425***	-0.0873**	-0.0477**	-0.1461***	-0.0885**	-0.0479**	-0.1343***	-0.0750*	-0.0475**
	(0.0260)	(0.0369)	(0.0194)	(0.0251)	(0.0371)	(0.0192)	(0.0245)	(0.0403)	(0.0187)
Education (Years)	0.0236***	0.0380***	0.0852***	0.0263***	0.0371***	0.0944***	0.0005	0.0317	0.0917***
	(0.0054)	(0.0082)	(0.0038)	(0.0052)	(0.0088)	(0.0082)	(0.0120)	(0.0195)	(0.0108)
Log of Age	-0.1372**	0.4274***	0.2728***	-0.1938**	0.2518	0.3207***	-0.4342***	0.0082	0.1401
	(0.0697)	(0.1538)	(0.0747)	(0.0909)	(0.1875)	(0.0844)	(0.1288)	(0.2204)	(0.1222)
Log of Tenure	0.0216	-0.0315	0.0651***	0.0232	-0.0311	0.0654***	0.0277	-0.0333	0.0607***
	(0.0190)	(0.0353)	(0.0160)	(0.0222)	(0.0355)	(0.0175)	(0.0200)	(0.0297)	(0.0160)
Urban	0.2398*	0.0073	0.0106	0.1587	-0.0707	0.0146	0.1465	0.1363	0.0464
	(0.1300)	(0.2100)	(0.0964)	(0.1431)	(0.2539)	(0.1020)	(0.1604)	(0.2482)	(0.1250)
Lee				0.2033*	0.3798*	-0.1268			
				(0.1227)	(0.2201)	(0.0918)			
DMF 1							-0.5449	-0.2184	0.1694***
							(0.3711)	(0.7040)	(0.0593)
DMF 2							0.3422	-0.5630	0.2873
							(0.2256)	(0.5034)	(0.1930)
DMF 3							0.1647*	0.3564	0.7639**
							(0.0884)	(0.8468)	(0.3292)
DMF 4							-2.0783***	-0.2714**	-0.9259**
							(0.5934)	(0.1357)	(0.4366)
DMF 5							0.6672	0.3407	0.6681*
							(0.4280)	(0.8148)	(0.3649)
DMF 6							0.1179	-0.0649	0.4962**
							(0.2784)	(0.6100)	(0.2235)
Ethnicity Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	3374	1311	2367	3374	1311	2367	3374	1311	2367
R^2									
Selection F-Test				0.0975	0.0844	0.1671	0.0000	0.1512	0.0003
Test DMF 3 v DMF 4							0.0007	0.5038	0.0147

Standard errors in parentheses

Dependent variable: Log of Hourly Wages

Bootstrapped standard errors to account for generated regressors, replicated with household clusters

F-Test for joint significance of selection terms

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.F Robustness of Earnings Decompositions

Table 3.20: Blinder-Oaxaca Decompositions by Sector — Land Not Excluded

	OLS			Lee			Dubin-McFadden		
	Own Account	Employer	WE	Own Account	Employer	WE	Own Account	Employer	WE
Differential									
Female Prediction	-2.901***	-2.748***	-2.339***	-2.902***	-2.748***	-2.337***	-2.898***	-2.747***	-2.333***
	(0.0259)	(0.0560)	(0.0418)	(0.0259)	(0.0564)	(0.0418)	(0.0259)	(0.0563)	(0.0419)
Male Prediction	-2.542***	-2.298***	-2.154***	-2.543***	-2.304***	-2.154***	-2.540***	-2.306***	-2.154***
	(0.0478)	(0.0560)	(0.0242)	(0.0471)	(0.0548)	(0.0242)	(0.0474)	(0.0546)	(0.0243)
Difference	-0.359***	-0.450***	-0.185***	-0.359***	-0.445***	-0.183***	-0.358***	-0.441***	-0.179***
	(0.0543)	(0.0792)	(0.0483)	(0.0537)	(0.0786)	(0.0483)	(0.0540)	(0.0784)	(0.0485)
Adjusted				-0.699**	-1.888***	0.0500	-1.334	-2.119*	0.314
				(0.346)	(0.607)	(0.170)	(0.843)	(1.242)	(0.293)
Decomposition									
Explained	-0.0843***	-0.245***	-0.0765***	-0.0866***	-0.246***	-0.0782***	-0.0101	-0.256***	-0.0729**
	(0.0271)	(0.0432)	(0.0281)	(0.0267)	(0.0420)	(0.0290)	(0.0375)	(0.0684)	(0.0321)
Unexplained	-0.275***	-0.205***	-0.108**	-0.612*	-1.642***	0.128	-1.323	-1.862	0.387
	(0.0534)	(0.0791)	(0.0423)	(0.346)	(0.606)	(0.169)	(0.842)	(1.243)	(0.293)
N	3374	1311	2367	3374	1311	2367	3374	1311	2367
Z-Score v Own Account		-0.9392	2.4009		-2.5090	2.2849		-1.1011	2.2334
Z-Score v Employer	0.9392		2.8536	2.5090		3.0723	1.1011		1.9062
Z-Score v WE	-2.4009	-2.8536		-2.2849	-3.0723		-2.2334	-1.9062	

Standard errors in parentheses

Z-Score calculated for adjusted differences in selection-corrected models

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.21: Blinder-Oaxaca Decompositions by Sector — Controlling for Industry

	OLS			Lee			Dubin-McFadden		
	Own Account	Employer	WE	Own Account	Employer	WE	Own Account	Employer	WE
Differential									
Female Prediction	-2.898***	-2.750***	-2.341***	-2.899***	-2.751***	-2.340***	-2.898***	-2.750***	-2.336***
	(0.0258)	(0.0557)	(0.0421)	(0.0258)	(0.0563)	(0.0420)	(0.0259)	(0.0563)	(0.0420)
Male Prediction	-2.536***	-2.294***	-2.155***	-2.538***	-2.297***	-2.155***	-2.537***	-2.300***	-2.155***
	(0.0474)	(0.0556)	(0.0243)	(0.0468)	(0.0544)	(0.0243)	(0.0472)	(0.0544)	(0.0244)
Difference	-0.362***	-0.457***	-0.186***	-0.361***	-0.454***	-0.185***	-0.360***	-0.450***	-0.181***
	(0.0539)	(0.0787)	(0.0486)	(0.0534)	(0.0783)	(0.0486)	(0.0538)	(0.0783)	(0.0485)
Adjusted				-0.674**	-1.370**	-0.0485	-1.211	-2.032*	0.296
				(0.338)	(0.570)	(0.164)	(0.793)	(1.176)	(0.292)
Decomposition									
Explained	-0.0372	-0.221***	-0.0821***	-0.0392	-0.225***	-0.0880***	-0.00489	-0.259***	-0.0616*
	(0.0308)	(0.0473)	(0.0281)	(0.0305)	(0.0473)	(0.0306)	(0.0382)	(0.0692)	(0.0324)
Unexplained	-0.325***	-0.235***	-0.104**	-0.635*	-1.145**	0.0395	-1.207	-1.773	0.358
	(0.0544)	(0.0790)	(0.0425)	(0.339)	(0.569)	(0.162)	(0.792)	(1.179)	(0.292)
N	3374	1311	2367	3374	1311	2367	3374	1311	2367
Z-Score v Own Account		-0.9918	2.4262		-1.7631	1.8092		-1.1095	2.1695
Z-Score v Employer	0.9918		2.9275	1.7631		2.2281	1.1095		1.9209
Z-Score v WE	-2.4262	-2.9275		-1.8092	-2.2281		-2.1695	-1.9209	

Standard errors in parentheses

Z-Score calculated for adjusted differences in selection-corrected models

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.G Summary Statistics for Alternative Household Demographics

Table 3.22: Alternative Summary Statistics (Females)

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
No. Infants (< 2 years) in HH	10926	0.28	0.52	0.00	0.00	0.00	1.00	4.00
No. Young Children (2–4 years) in HH	10926	0.44	0.66	0.00	0.00	0.00	1.00	5.00
No. Older Children (5–14 years) in HH	10926	1.62	1.63	0.00	0.00	1.00	2.00	12.00
No. Elders (> 65 years) in HH	10926	0.16	0.41	0.00	0.00	0.00	0.00	5.00
Observations	10926							

Sample of individuals of working age (15–65)

Table 3.23: Alternative Summary Statistics (Males)

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
No. Infants (< 2 years) in HH	9725	0.24	0.48	0.00	0.00	0.00	0.00	4.00
No. Young Children (2–4 years) in HH	9725	0.39	0.63	0.00	0.00	0.00	1.00	5.00
No. Older Children (5–14 years) in HH	9725	1.49	1.58	0.00	0.00	1.00	2.00	12.00
No. Elders (> 65 years) in HH	9725	0.12	0.37	0.00	0.00	0.00	0.00	5.00
Observations	9725							

Sample of individuals of working age (15–65)

Chapter 4

Discrimination in a Search-Match Model with Self-Employment

4.1 Introduction

Female participation in self-employment is substantial in urban Africa. For the urban Ghanaian sample used in this chapter, self-employed women outnumber self-employed men by a ratio of nearly 3:1. However, the difference between male and female earnings is far larger in self-employment than in the wage sector. For the self-employed, male earnings are approximately double female earnings at the median. For the wage-employed, however, the difference is around 25 percent. As we showed using a nationally-representative Ghanaian sample in Chapter 3, controlling for individual and job characteristics, as well as selection across employment sectors, can only partially account for these empirical patterns.

In this chapter, we turn to theory to try to reconcile this puzzle. In particular, we develop a search-match model, which allows for discrimination in the wage sector. In our model, unemployed workers can choose to search for wage-employment, self-employment, or both, depending on their returns in each sector. Workers are heterogeneous along two dimensions. Their 'Type', which translates to sex in an empirical setting, determines whether they will face discrimination in trying to obtain a wage sector job. However, workers also differ by their productivity in self-employment, which ultimately generates earnings dispersion for both sexes in both types of jobs.

The search literature provides an especially useful framework within which to examine discrimination. Firstly, matching frictions afford firms monopsony power, which they may exploit to worsen the outcomes of women relative to men. Secondly, matching models allow us to consider the outcomes of the unemployed, which would not be possible if workers could freely move between wage- and self-employment. Finally, we are able to couch discrimination in terms of the matching function itself, by adding extra friction to the process by which women can obtain wage sector jobs compared to men. Since our model directly accounts for the these sources of non-competition in the labour market, as well as heterogeneity in individuals' productivity, we capture the intuition of both segmentation *and* sorting.

By numerically solving and simulating our model, we are able to explain the differences in average earnings between men and women in *both* wage- *and* self-employment, *without* assuming any differences in the innate ability distribution of each sex. Our model echoes many of the ideas from Roy's classic model of occupational selection (Roy, 1951). In short, the frictions inherent in the search-match framework prevent workers from selecting occupations entirely on the basis of comparative advantage, which reduces average productivity in all sectors. Since these frictions are worse for women, due to discrimination, they sort even less efficiently than men, driving a wedge between male and female earnings amongst the self- and wage-employed.

Our model also generates a series of predictions about the numbers of women and men participating in particular occupations. As expected, discrimination drives women away from wage-employment, which in turn increases the number of unemployed and self-employed women. However, we also find that there are spillover effects on men. In particular, male wage-employment rises and self-employment falls as discrimination becomes more severe.

To our knowledge, this is the first attempt to analyse discrimination in a search-match model that allows for both wage- and self-employment. We build on two key strands of the theoretical literature. Firstly, we complement a growing body of work, which uses the search framework to assess the implications of large informal sectors on labour market outcomes in developing countries (Albrecht et al., 2009; Saatchi and Temple, 2009; Kerr, 2012; Narita, 2012). Secondly, we draw on a number of studies which capture either employer or employee discrimination in the presence of search frictions (Black, 1995; Bowlus, 1997; Sasaki, 1999; Rosén, 2003). Unlike in competitive models, where discrimination is eroded by competition and segregation, the monopsony power afforded to firms in search models allows firms to 'indulge a taste for discrimination' such that discriminatory outcomes may persist (Becker, 1957; Black, 1995; Altonji and Blank, 1999).

We also provide descriptive evidence on the urban Ghanaian labour market using a large panel survey collected by the Centre for the Study of African Economies between 2004 and 2013. These data help to guide our modelling approach and verify the logic of our simulations. As such, we contribute to the empirical literature on informality in urban Africa (Hart, 1973; Kingdon et al., 2006; Falco et al., 2011; Fields, 2011; Falco and Quinn, 2012).

This chapter proceeds as follows. In Section 4.2 we review the related literature. In Section 4.3 we describe our data and present evidence on the urban Ghanaian labour market. We outline our model and then describe how we solve for the equilibrium in Sections 4.4 and 4.5 respectively. In Section 4.6 we present our simulation results. We discuss our findings and conclude in Section 4.7.

4.2 Related Literature

4.2.1 Models of Discrimination

Competitive models broadly predict that discriminatory outcomes cannot survive in equilibrium. This idea can be traced back to Becker's (1957) classic model of 'taste-based' discrimination. Becker assumes that there are two types of potential employees, Type-A and Type-B, and some subset of firms that are prejudiced, such that they incur a utility cost from employing Type-B individuals. If the proportion of prejudiced firms is relatively low, the market will simply segregate such that Type-B employees only work for unprejudiced firms. However, even if the share of prejudiced firms is initially relatively large, an equilibrium with wage discrimination cannot be sustained. Any Type-B employees working at prejudiced firms for lower wages, can be attracted by expanding unprejudiced firms, who offer higher wages to exploit the arbitrage opportunity. Thus, the size or prevalence of prejudiced firms must fall until this type of arbitrage is no longer possible, and segregation occurs.

This logic has been widely applied not only to employer discrimination, but also employee discrimination as well as consumer discrimination (Altonji and Blank, 1999; Autor, 2003; Stevens, 2012). For example, in a competitive model of employee discrimination, we might assume that Type-A workers must be compensated with higher wages in order to work alongside Type-B workers. However, even if firms with a mixed A-B workforce exist initially, they will be unable to compete with firms employing only one type of employee in the long-run, since the one-Type firms will not have to offer extra compensation to Type-A workers. Thus, a segregated equilibrium eventually results, and in spite of the prejudiced preferences of Type-A workers, discrimination does not affect Type-Bs' outcomes.

Discriminatory outcomes can, however, persist if at least one of the competitive market assumptions is dropped. This is an idea on which models of statistical discrimination rely. In particular, without making any assumptions about prejudiced preferences per se, these models suggest that imperfect information may lead to outcomes in which Type-B workers suffer.

Aigner and Cain (1977) consider a variety of models of statistical discrimination, in which there are differences between the signals of ability provided to employers by Type-B and Type-A workers. In their framework, when potential employees apply to jobs, employers must make wage offers based only on their observation of employee Type as well as some noisy signal of ability. Employers weight the mean ability of each Type, which they know, against the individual worker's signal. Assuming Type-Bs have lower ability on average, this means that high-ability Type-B workers may obtain a worse wage offer than equally productive Type-A workers, even with a high signal. The extent of this problem will, of course, depend on the signal-to-noise ratio, which, in turn, is determined by the variance of the signal for each Type. However, each worker Type, on average, will not face discrimination, insofar as employers know their mean abilities.[1]

More recent models of statistical discrimination have endogenised workers' decisions over whether to invest in human capital, such that employers' responses to imperfect information may be enough on their own to detour Type-B workers from gaining education and skills. In these models, there are no innate differences between the abilities or the distributions of signals between Type-A and Type-B workers. For example, in Coate and Loury's (1993) model, potential workers must choose whether or not to become 'qualified' before applying to jobs. This results in some noisy signal, which employers then observe.[2] Employers wish to assign only qualified workers to some high-skill (high-wage) task, so they set a particular 'standard' for obtaining such jobs, based on their priors about workers of a particular Type as well as the value of the signal.[3] It turns out that employers' priors about potential workers' investment in the qualification ends up being true. If employers have more pessimistic priors, then workers' noisy signals must meet a higher standard in order to gain assignment to the high-skill task. This implies investing in the

[1] In Aigner and Cain's (1977) framework, Type-B workers may, in fact, suffer lower wage offers on average, if the variance of Type-B's signal is higher and employers are risk-averse, even if the means of the Type-A signal and Type-B signal are the same.

[2] Employers do not observe the initial qualification.

[3] The optimal standard is set using Bayes' Law.

qualification has a lower return to workers since, even after becoming qualified, they have less chance of being assigned to the high-skill task. In turn, fewer workers invest in the qualification, and employers' priors are confirmed. Insofar as employers' prior beliefs differ by observable characteristics such as race and sex, this mechanism may result in a self-fulfilling discriminatory equilibrium.[4]

By dropping the assumption of perfect information, models of statistical discrimination provide a coherent framework for explaining the persistence of discrimination in equilibrium. However, similar results may obtain if other competitive market assumptions are relaxed. In particular, by couching our model in terms of search frictions, we focus on the possibility that employers have monopsony power.

4.2.2 Analysing Discrimination using the Search Framework

Search models explicitly capture the frictions involved in the employer-employee meeting process, giving employers monopsony power and thus potentially allowing discriminatory equilibria to persist. For clarity, we distinguish between two main types of search models that pervade both the theoretical and empirical literature (Mortensen and Pissarides, 1999). In 'matching models', the rate at which workers and firms meet is modelled through some function of 'labour market tightness', or the ratio of unemployment to unfilled job vacancies. Wages may be determined by a bargain between worker and firm over the surplus accruing from a successful match. By contrast, in 'wage-posting' models, firms choose wages prior to matches, engaging in a strategic game with all other labour market participants (Burdett and Mortensen, 1998; Rogerson et al., 2005).

Black (1995) builds a simple search model in the wage-posting framework, in which a taste for discrimination limits, but does not eliminate, the existence of prejudiced firms. Two types of workers randomly meet with either unprejudiced firms, who are willing to employ both Type-A and Type-B workers, and prejudiced firms, who are only willing to employ Type-As. Wage offers, conditional on worker type, are posted by both firms prior to meeting, but workers also derive some stochastic match-specific utility from a particular match, which cannot be observed by

[4] A number of extensions to the Coate and Loury (1993) model have been proposed. Moro and Norman (2004) examine statistical discrimination in a general equilibrium context, suggesting that particular groups may end up specialising in certain tasks. This couches the idea that certain groups may try to exploit others in terms of informational externalities. Fryer (2007) develops a dynamic model of statistical discrimination, which explores the outcomes of discriminated individuals that are able to overcome the first hurdle of actually getting a job. He shows that employers may eventually favour such individuals, since their success in obtaining the job is made less likely by their observable type, implying their initially unobservable ability is high.

the firm.[5] Interestingly, even the 'unprejudiced' firms offer Type-B workers lower wages because, given the presence of prejudiced firms, Type-Bs have lower outside options, and hence reservation wages, than Type-As. Black also explicitly models the entry decision of potential entrepreneurs, who are heterogeneous in both prejudice and ability. Prejudiced entrepreneurs find entry less profitable than unprejudiced entrepreneurs of the same ability. However, prejudiced firms still enjoy sufficient monopsony power, such that they can use their profits to pay for discriminatory behaviour.[6]

Employer-employee frictions are not the only source of monopsony power that allow discrimination to persist. Borjas and Bronars (1989) build a model in which some subset of consumers derive dis-utility from buying goods from self-employed sellers of a certain Type. However, matches are random and consumers initially have imperfect information about the 'type-price' quote of the sellers. They must undertake costly search in order to extract this information. Sellers can allocate their time either to producing goods according to some heterogeneous productivity parameter, or contacting potential consumers to make sales.[7] Firstly, it emerges that both the mean and the variance of Type-Bs' self-employment income distribution are lower than Type-As'. Moreover, the returns to productive ability are lower for Type-Bs than Type-As. If they choose to integrate prejudiced consumers, they must lower their offer prices. However, if they choose to reject prejudiced consumers, more matches are wasted.

As Sasaki's (1999) model shows, discrimination may also persist within the context of a search model even if it is potential co-workers that have a taste for discrimination. In his model, which focusses on sex, male workers suffer a utility cost if they work for a firm with a mixed female-male workforce. This both gives firms an incentive to segregate and affords them the opportunity to offer females lower wages, since the outside options of female workers are reduced by the presence of male-only firms. Interestingly, the model predicts that the gap between the outcomes of men and women shrinks if female participation in the labour market increases, since the profitability of male-only firms is reduced.

A number of other studies have used the search framework to investigate the extent to which discrimination can explain female-male differences in earnings and labour market dynamics empirically (Bowlus, 1997; Eckstein and Wolpin, 1999; Borowczyk-Martins et al., 2012). In a recent example, Flabbi (2010) estimates a matching model in the spirit of Mortensen and

[5]This hidden match-specific utility guarantees that not all matches are successful.

[6]The reduced profitability of prejudiced firms implies that the proportion of prejudiced firms is lower than proportion of prejudiced entrepreneurs.

[7]Sellers must choose how many hours they wish to allocate to production, and the set of consumers — prejudiced only, unprejudiced only, or both — that they wish to serve. There is no correlation between productive ability and type.

Pissarides (1994), which attempts to isolate the effects of female-male productivity differences and discrimination using data from the United States. The maximum likelihood results suggest that as much as two thirds of the female-male earnings gap is explained by employers' taste for discrimination.

Overall, however, aside from the Borjas and Bronars (1989) model, few attempts have been made to incorporate the extra option of self-employment into search models that allow for discrimination. It is this gap in the literature, which we hope to address.

4.2.3 Search Models and Informality

Given the importance of self-employment and informal sectors more generally, search models that focus only on formal wage-employment are insufficient for understanding labour markets in developing countries. There is a growing literature attempting to address this shortcoming.

Some economists have used the wage-posting framework to model informality. Narita (2012), for example builds such a model, which allows for heterogeneous firms operating in a formal and an informal wage sector, *as well as* a self-employment sector and unemployment. Whilst firms operating in the formal sector must pay a series of taxes on profits, payroll, and severance, in the informal sector they face the possibility of being fined for their activity. Noting that it may take time for individuals to learn about good business ideas to pursue in self-employment, Narita's approach allows the values of being in each sector and the probabilities of transitioning between jobs to vary with total work experience. The model is solved using backwards induction from retirement to derive a series of moment conditions which link the observed earnings distributions with the predicted earnings distributions. The estimated model successfully reproduces the composition of the Brazilian labour force over workers' life cycles, finding that transitioning into self-employment is significantly more likely for more experienced workers.

More commonly, however, economists have used Mortensen and Pissarides' (1994) random match and bargaining model as a starting point for search models that incorporate self-employment and informality. For example, Saatchi and Temple (2009) build a model where an urban formal sector, in which firms and workers randomly match and then bargain over wages, coexists with an urban informal sector. They also allow for an agricultural sector, from which potential urban workers can migrate. Calibrating their model with Mexican data, they are able to show that formal sector frictions create a reasonably sized informal sector, even if agriculture remains a viable option. Their model is used to simulate the impact of productivity shocks to different sectors, changes in tax policy, and adjustments to the efficiency of matching and the wage-bargaining parameters.

By capturing heterogeneity in individuals' ability, some models allow us to analyse variation in worker's welfare *within* particular job sectors. For example, in Albrecht et al.'s (2009) model of large Latin American economies, individuals work in either the formal sector, in which their productivity is heterogeneous, or the informal sector, in which productivity is homogeneous.[8] In the formal sector, productivity is also subject to shocks and, providing the match survives, wages are constantly renegotiated. In the informal sector, by contrast, there is no earnings dispersion, as individuals simply receive their marginal product. Kerr (2012) modifies Albrecht et al. (2009) framework to make it more suitable for urban African labour markets. In his model, workers are heterogeneous in their *informal* sector productivity instead. This generates earnings dispersion in both sectors, however, because workers' outside options at the formal sector wage bargain differs depending on their welfare in informal work.

We build on this literature, by constructing a matching model which allows for a self-employment sector as well as heterogeneity in (1) worker ability *and* (2) the frictions they face in trying to obtain wage jobs. This enables us to examine the effects of discrimination, when self-employment provides workers with a plausible outside option.

4.3 Evidence from Ghana

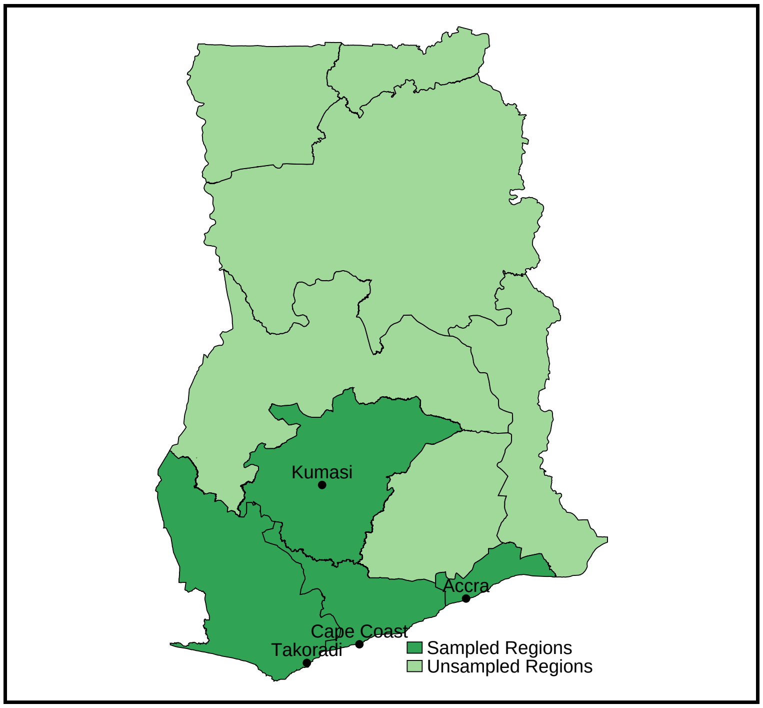
In this section, we present a series of summary statistics and generate a number of stylised facts using panel data from urban Ghana, to motivate our search-match model.

4.3.1 Sample

Our data are taken from the Ghana Household Urban Panel Survey (GHUPS), a longitudinal dataset collected in four of Ghana's largest cities — Accra, Kumasi, Cape Coast, and Takoradi — by the Centre for the Study of African Economies between 2004 and 2013. We show the location of these cities in Figure 4.1.

[8] They also have the option to remain unemployed, and wait for either formal or informal sector matches to come along. There are no transitions between formal and informal sector or vice-versa.

Figure 4.1: Sample Location



We drop the 2004 wave for our analysis, because the survey instrument used was significantly different from the subsequent waves. The number of observations available, across time and space, is shown in Table 4.1.[9]

Table 4.1: No. Observations over Time and Space

Wave	City							
	Accra		CC/Takoradi		Kumasi		Total	
	No.	%	No.	%	No.	%	No.	%
2005	571	4.23	215	1.59	459	3.40	1245	9.23
2006	716	5.31	266	1.97	575	4.26	1557	11.54
2008	584	4.33	224	1.66	413	3.06	1221	9.05
2009	651	4.83	227	1.68	436	3.23	1314	9.74
2010	928	6.88	421	3.12	767	5.68	2116	15.68
2012	1320	9.78	627	4.65	1127	8.35	3074	22.78
2013	1312	9.72	613	4.54	1040	7.71	2965	21.98
Total	6082	45.08	2593	19.22	4817	35.70	13492	100.00

Pooling these observations across all available waves, we present some basic summary statistics for the full sample in Table 4.2. Our sample is not perfectly balanced by sex, as we discuss in Appendix 4.A, but we have sufficient numbers of men and women to allow for cross-sex comparisons. We restrict our sample to individuals of working age, between 15 and 65 years old. The

[9] No survey was conducted in 2007 or 2011.

median worker is 29 years old and has 9 years of education.

Table 4.2: Summary Statistics (Pooled Sample)

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Full Sample								
Sex (0=Female, 1=Male)	13490	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Education (years)	13487	9.26	3.59	0.00	9.00	9.00	12.00	20.00
Age (years)	13491	31.82	12.30	15.00	22.00	29.00	40.00	65.00
Hours worked per week	7028	47.46	21.48	0.00	38.00	48.00	60.00	120.00
Tenure (months)	7009	90.17	100.58	0.00	17.00	52.00	129.00	608.00
Female								
Education (years)	7780	8.71	3.80	0.00	8.00	9.00	11.00	20.00
Age (years)	7782	32.24	12.33	15.00	22.00	30.00	40.00	65.00
Hours worked per week	3937	46.85	21.79	0.00	35.00	48.00	60.00	120.00
Tenure (months)	3919	86.19	100.41	0.00	16.00	48.00	122.00	608.00
Male								
Education (years)	5706	10.00	3.12	0.00	9.00	9.00	12.00	18.00
Age (years)	5708	31.24	12.24	15.00	21.00	28.00	39.00	65.00
Hours worked per week	3090	48.24	21.05	0.00	40.00	48.00	60.00	120.00
Tenure (months)	3089	95.23	100.60	0.00	19.00	57.00	141.00	598.00
Observations	13491							

Sample of individuals of working age (15–65)

Given the focus of this chapter on differences between men and women, we break down these summary statistics by sex in lower two panels of Tables 4.2. We note that there appear to be small differences in typical Mincerian measures of human capital. Women appear to have fewer years of education than men on average. Also, of the sub-sample working in wage- or self-employment, women appear to have less tenure in their current job than men.

We test the differences in means for these variables formally using the pooled sample in Table 4.3.[10] The differences between men and women in the key human capital variables at the mean are statistically significant.

Table 4.3: Differences in Observable Human Capital

	Women	Men	p-value	N
Education (years)	8.71	10.00	0.0000	13486
Age (years)	32.24	31.24	0.0000	13490
Hours worked per week	46.85	48.24	0.0067	7027
Tenure (months)	86.19	95.23	0.0002	7008
Observations	13490			

Sample of individuals of working age (15–65)

[10]These tests were conducted using the 'two-group mean-comparison test' in Stata's `ttest` command.

4.3.2 Occupations

Occupational selection for the pooled sample is shown in Table 4.4. Any individuals engaged in full-time schooling or unpaid casual work, *without* searching for paid work, are classified as out of the labour force. The unemployed, by contrast, are actively searching for wage work or potential business activities. We also split the sub-sample of individuals in private wage-employment by firm size.[11] In this chapter, however, we do not divide the self-employed into those that do and do not employ labour besides their own.[12]

Table 4.4: Occupational Breakdown by Sex (Pooled)

Occupation	Sex					
	Female		Male		Total	
	No.	%	No.	%	No.	%
Apprentice	315	2.34	305	2.26	620	4.60
Out of LF	2325	17.23	1543	11.44	3868	28.67
Private WE (Large Firm)	313	2.32	734	5.44	1047	7.76
Private WE (Small Firm)	835	6.19	1200	8.90	2035	15.09
Public WE	217	1.61	241	1.79	458	3.40
Self-Employed	2937	21.77	1132	8.39	4069	30.16
Unemployed	840	6.23	553	4.10	1393	10.33
Total	7782	57.69	5708	42.31	13490	100.00

Women outnumber men substantially in self-employment. Indeed, there are nearly three times as many women as men in the self-employed sub-sample. The opposite, however, is true for wage-employment, where there are nearly double the number of men compared to women. The difference is particularly profound amongst workers employed in privately run large firms.

In Table 4.5, we use the pooled sample to consider a more detailed breakdown of the industries *within* self- and wage-employment in which men and women engage. Amongst the self-employed, we see the starkest differences between the sexes in the 'Retail / Trade' category, in which female participation is over five times greater than male participation. The most notable differences amongst the wage-employed are in 'Manufacturing' and primary industries ('Mining / Fishing / Agriculture'), where men outnumber women significantly.

The heterogeneity displayed in Table 4.5 is not the focus of this Chapter per se. However, we argue that if women face different frictions in obtaining particular jobs *within* self-employment, say, then the logic we present in Sections 4.4, 4.5, and 4.6 may help explain these intra-sector patterns.

[11] Large firms have > 20 employees.

[12] The sample size for each wave is much smaller than the cross-sectional GLSS5+ data used in Chapter 3. Thus, we simply have too few employers to make this kind of distinction based on technology meaningful.

Table 4.5: Detailed Occupational Breakdown by Sex (Pooled)

	Sex					
	Female		Male		Total	
	No.	%	No.	%	No.	%
Self-Employed						
Manufacturing	425	12.25	151	4.35	576	16.60
Retail / Trade	1756	50.62	337	9.71	2093	60.33
Services	370	10.67	430	12.40	800	23.06
Total	2551	73.54	918	26.46	3469	100.00
Wage-Employed						
Banking / Finance	5	0.15	6	0.18	11	0.34
Construction	3	0.09	28	0.86	31	0.95
Education	15	0.46	8	0.24	23	0.70
Entertainment	0	0.00	16	0.49	16	0.49
Manufacturing	165	5.05	348	10.65	513	15.70
Mining / Fishing / Agriculture	10	0.31	83	2.54	93	2.85
Public Sector	107	3.28	111	3.40	218	6.67
Religious	0	0.00	8	0.24	8	0.24
Retail / Trade	255	7.81	246	7.53	501	15.34
Services	708	21.67	1145	35.05	1853	56.72
Total	1268	38.81	1999	61.19	3267	100.00

4.3.3 Earnings

We present the distribution of (log) earnings in each sector using the pooled sample in Figure 4.2. The data for the wage-employed relate to before tax payments to the employee, excluding any bonuses or allowances. The data for the self-employed capture profits accruing to the entrepreneur, net of all costs (including labour costs). We construct earnings figures for a standard working week.[13]

We separate out the earnings distributions for the wage-employed by the public-private distinction and by firm-size, but also aggregate the wage-employment data in the 'WE (All)' category. Earnings dispersion appears to be greater in self-employment than in wage-employment, although there is substantial variation within each sector. Moreover, the earnings distributions for each sector overlap. This suggests that both *intra*-sector and *inter*-sector heterogeneity are important.[14]

In Table 4.6 we present median earnings for men and women between 2010 and 2013. The data are deflated to 2010 Ghana Cedis (GHS).

[13]We adjust the raw weekly data for the number of hours *actually* worked that week. We then multiply by 48 hours, the median number of hours worked by individuals in self- or wage-employment, to give a standard working week.

[14]We are somewhat cautious about making comparisons of the *level* of earnings between the self- and wage-employed on the basis of this data. This is because the questionnaire has separate sections for self- and wage-employment, such that different questions were asked to elicit earnings data for workers in each sector.

Figure 4.2: Sectoral Earnings

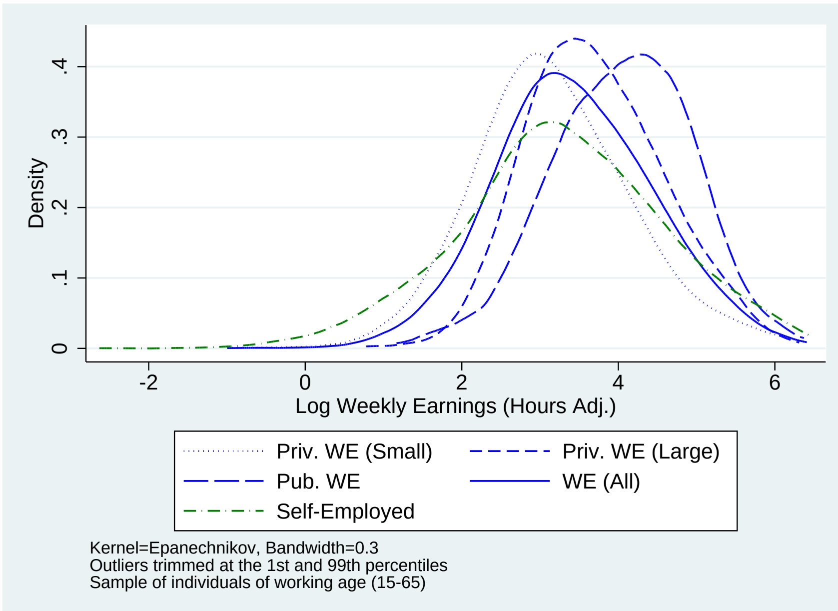


Table 4.6: Median Weekly Earnings over Time (2010–2013)

	2010		2012		2013	
	Male	Female	Male	Female	Male	Female
Private WE (Small Firm)	25.73	15.44	38.77	20.51	50.33	31.06
	(73.28)	(29.35)	(134.68)	(609.11)	(261.92)	(78.06)
Private WE (Large Firm)	42.70	32.24	49.89	56.33	67.94	65.32
	(63.97)	(81.10)	(157.64)	(101.50)	(127.44)	(86.75)
Public WE	67.27	80.31	90.88	86.30	104.78	105.53
	(165.39)	(94.51)	(84.25)	(174.91)	(149.43)	(265.10)
Wage-Employed (All)	36.18	23.16	47.28	35.92	57.48	44.29
	(88.12)	(67.94)	(135.93)	(468.71)	(219.14)	(133.13)
Self-Employed	54.45	23.92	77.22	31.36	97.61	50.24
	(362.08)	(134.23)	(247.72)	(181.40)	(586.09)	(1123.16)
Observations	472	692	721	914	715	873

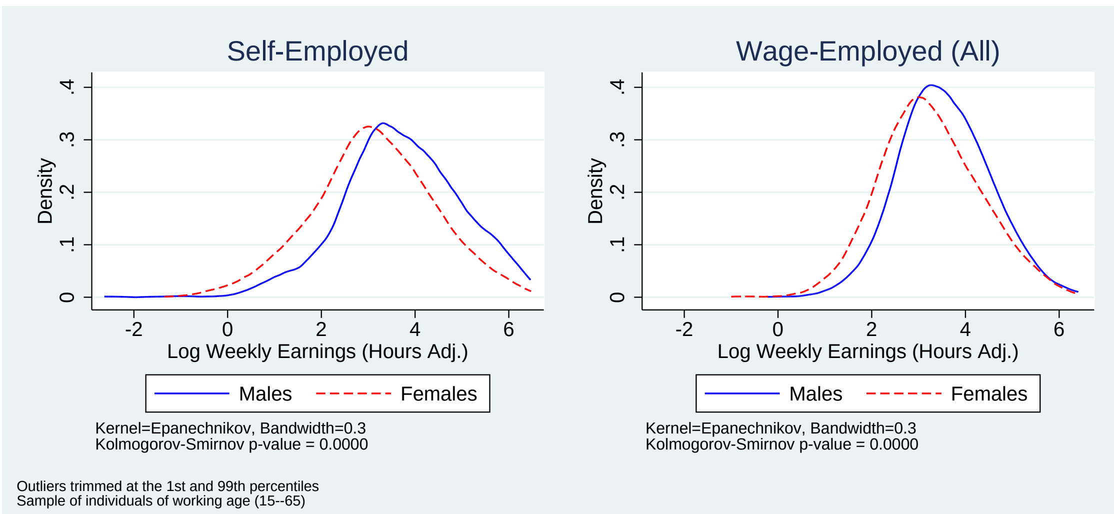
Sample of individuals of working age (15–65)

Standard deviations in parentheses

We can see that differences between men and women are largest in self-employment, where male earnings are approximately double those of women at the median. Further, the overall earnings gap in wage-employment largely seems to be driven by individuals employed in small private firms. Insofar as firm size proxies for formality, the data suggest that sex earnings differentials tend to be larger in less formal sectors.

To emphasise this point, we plot the distributions of earnings for men and women in self- and wage-employment (all) in Figure 4.3.

Figure 4.3: Earnings Distributions by Sector and Sex



We test the equality of the earnings distributions for men and women using the non-parametric Kolmogorov-Smirnov test. For both the self- and wage-employed, we reject the null hypothesis that the underlying distributions are the same. As we show in Appendix 4.B, however, if we break down the wage sector by public and private as well as firm-size, we are unable to reject this null hypothesis for the sub-samples of workers employed privately in large firms or in the public sector. This echoes the notion that average earnings differentials between men and women are larger in less formal sectors of the economy.

4.3.4 Transitions

Finally, we examine the extent to which workers move between different sectors. In Tables 4.7 and 4.8 we show the 2012–2013 transition matrices for men and women respectively, collapsing all types of wage jobs into one category. We show more detailed transition matrices, where the wage sector is disaggregated by public versus private and firm size, in Appendix 4.C.

Table 4.7: Transition Matrix for Men (2012–2013)

2012 Occupation	2013 Occupation Out of LF %	 SE %	 UE %	 WE %	 Total %
Out of LF	20.10	1.11	5.93	6.33	33.47
SE	0.70	11.46	0.80	4.12	17.09
UE	2.51	0.90	2.01	2.71	8.14
WE	4.12	4.32	3.32	29.55	41.31
Total	27.44	17.79	12.06	42.71	100.00

Table 4.8: Transition Matrix for Women (2012–2013)

2012 Occupation	2013 Occupation Out of LF %	 SE %	 UE %	 WE %	 Total %
Out of LF	20.52	3.90	5.64	4.55	34.61
SE	3.90	25.58	2.24	2.82	34.54
UE	3.47	2.24	3.03	1.59	10.33
WE	3.25	2.89	2.24	12.14	20.52
Total	31.14	34.61	13.15	21.10	100.00

Transitions between self- and wage-employment occur for both men and women. Approximately 10 percent of males and 14 percent of females engaged in wage-employment in 2012 are engaged in self-employment by 2013. Conversely, nearly a quarter of males and around 8 percent of females engaged in self-employment in 2012 have wage jobs by 2013. As such, individuals do not fully specialise in either self- or wage-employment for their entire life-cycle and some are willing to do both.

4.3.5 Stylised Facts

We summarise our empirical analyses from Section 4.3 with five key stylised facts:

1. Most of the wage-employed are men. Most of the self-employed are women.
2. Sex earnings gaps persist in both self- and wage-employment.
3. Average earnings differentials between men and women are smaller in more formal sectors.
4. Earnings dispersion is higher in self-employment than wage-employment for both sexes, but there is substantial variation in both sectors.

5. Inter-sector transitions occur. Some individuals do not fully specialise in either self- or wage-employment for their entire life-cycle.

We use this evidence to motivate and guide the theoretical model that follows.

4.4 Allowing for Discrimination in a Search-Match Model with Self-Employment

In this section, we outline the key features of our theoretical model. Our approach allows for discrimination against a certain sub-section of the population in trying to obtain wage jobs. We also capture the idea that workers have outside options in self-employment, enabling them to avoid simply queueing in unemployment if they do not have a wage job.

4.4.1 Model Basics

The population of workers, of size N, is divided into two, such that some proportion π^A are Type-As and some proportion π^B are Type-Bs, the latter of whom are affected by discrimination. In particular we assume that $\pi^A = \pi$ and $\pi^B = (1-\pi)$. Individuals are also heterogeneous in their self-employment ability, $y \sim \text{Uniform}(0,1)$. We assume this ability distribution is the same for Type-As and Type-Bs.

To model discrimination, we assume that Type-Bs face the possibility of some extra shock which immediately destroys their match to a firm.[15] This shock hits a potential match with probability λ. By modelling this logic stochastically, we can maintain the assumption that firms are ex ante homogeneous, such that there is no inherent heterogeneity in firms' preferences. Instead, λ effectively proxies for the degree of ex post discrimination in the market.[16]

Intuitively, we can understand discrimination in our model in the following way. We suppose that, at each firm, some proportion λ of the 'interviewers' of job candidates are prejudiced and therefore unwilling to employ Type-Bs. On being matched with a firm, potential Type-B workers

[15] In particular, the shock can be understood as some prohibitively high extra cost, which forces firms to immediately terminate the job. This is analogous to the idea in Black (1995) that some prejudiced firms are simply unwilling to employ Type-B workers.

[16] This concept is analogous to the proportion of prejudiced firms in the market, from other models.

therefore have probability λ that their match is immediately destroyed.[17] As such, our model treats discrimination as an *exogenous* feature of the matching process. Unlike Borjas and Bronars (1989), Black (1995), Sasaki (1999) and other search-match models that allow for discrimination, we do not attempt to endogenise the process through which prejudiced firms survive per se.

To build the model, we first consider individuals' behaviour, separating out the value functions for Type-As and Type-Bs. We then consider the problem faced by firms as they try to match with workers and make profit. Finally, we outline the way in which wages are determined.

4.4.2 Type-A Value Functions

We begin by specifying the problem facing Type-A workers. Both Types of individuals can work in either wage- or self-employment, queuing for both types of jobs in unemployment. We also assume that all workers are risk neutral.

In self-employment, individuals simply earn their marginal product, determined directly by their ability y. They also face some exogenous possibility that their self-employment activity will end, which occurs with probability q_s. Given the interest rate, r, we can write the flow value of self-employment for Type-As, $rV_s^A(y)$.

$$rV_s^A(y) = y + q_s(V_u^A(y) - V_s^A(y)) \tag{4.1}$$

The asset equation for unemployment is determined by some flow of unearned income, b, as well as the options of entering self- and wage-employment. The probability of a self-employment opportunity arriving is exogenously determined by the parameter α. However, the probability of a wage-employment opportunity is derived endogenously from the matching function, $m(.)$, which itself depends on labour market tightness, θ.[18] The max operators capture the fact that a worker may forego an opportunity in either self- or wage-employment if, given their ability, it is in their best interest to do so. We write the flow value of unemployment for Type-As, $rV_u^A(y)$,

[17] Informal interviews with Ghanaian workers suggests these types of mechanisms may be realistic.

[18] In Section 4.5 we impose properties on the matching function to guarantee the existence and uniqueness of the equilibrium.

as:

$$rV_u^A(y) = b + \alpha \max[(V_s^A(y) - V_u^A(y)), 0] + m(\theta) \max[(V_e^A(y) - V_u^A(y)), 0] \tag{4.2}$$

Tightness in the labour market is dependent on the number of unemployed Type-As and Type-Bs, which we label U^A and U^B respectively, as well as the number of vacancies, v.[19]

$$\theta = \frac{v}{U} = \frac{v}{U^A + U^B} \tag{4.3}$$

Finally, the flow value of wage-employment for Type-As, $rV_e^A(y)$, is determined by the bargained wage, $w^A(y)$, as well as the possibility that a match ends. Separations are exogenous, occurring with some probability q_e each period.

$$rV_e^A(y) = w^A(y) + q_e(V_u^A(y) - V_e^A(y)) \tag{4.4}$$

4.4.3 Type-B Value Functions

The value functions for self-employment, unemployment, and wage-employment are almost identical for Type-Bs. The only difference arises in the flow value for unemployment, where the chances of a successful match are adjusted by $(1-\lambda)$. We assume that the exogenous parameters r, α, q_e and q_s are the same for Type-As and Type-Bs.

$$rV_s^B(y) = y + q_s(V_u^B(y) - V_s^B(y)) \tag{4.5}$$

$$rV_u^B(y) = b + \alpha \max[(V_s^B(y) - V_u^B(y)), 0] + m(\theta)(1-\lambda) \max[(V_e^B(y) - V_u^B(y)), 0] \tag{4.6}$$

[19] We label the total number of unemployed individuals U.

$$rV_e^B(y) = w^B(y) + q_e(V_u^B(y) - V_e^B(y)) \tag{4.7}$$

4.4.4 Firm Value Functions

Firms can employ either one or zero workers. We label the former case a 'filled job' and the latter case a 'vacancy'. Additionally, we assume that firms are risk neutral.

The value of a filled job will differ, depending on whether firms successfully match with a Type-A or a Type-B worker. As we noted above, we do *not* assume there are any differences in the underlying ability distributions for self-employment productivity, y, between Type-As and Type-Bs, whilst wage-employment productivity, z_0, is assumed to be homogeneous.[20] As such, the value of a match for Type-As and Type-Bs only differs because the wage function is dependent on Type. This arises because Type-Bs' outside options are reduced by the extra possibility that matches may simply terminate immediately after formation, weakening their wage-bargaining position.

We can thus write the flow value of matching with a Type-A worker with ability y, which we label $r\Pi_e^A(y)$, as:

$$r\Pi_e^A(y) = z_0 - w^A(y) + q_e(\Pi_v - \Pi_e^A(y)) \tag{4.8}$$

The analogous flow value of matching with a Type-B worker, $r\Pi_e^B(y)$, may be written:

$$r\Pi_e^B(y) = z_0 - w^B(y) + q_e(\Pi_v - \Pi_e^B(y)) \tag{4.9}$$

The asset equation for posting a vacancy is complicated by the possibility of matching with either a Type-A or a Type-B worker. The probability of doing this is determined by the proportion of Type-As and Type-Bs amongst the unemployed pool, which is captured by the $\frac{U^A}{U^A+U^B}$ and $\frac{U^B}{U^A+U^B}$ terms. Given the per period cost of advertising a job, c, we can write the flow value

[20] We assume that $z_0 \in [0, 1]$ to make productivity in wage- and self-employment comparable. We also add the restriction $z_0 > b$. As we show in Appendix 4.D, this enables us to determine the type of equilibrium that prevails in our model.

of posting a vacancy, $r\Pi_v$, as:

$$
\begin{aligned}
r\Pi_v = & - c + \frac{m(\theta)}{\theta}\frac{U^A}{U^A+U^B}\mathrm{E}\Big\{\max\left[(\Pi_e^A(y)-\Pi_v),0\right]\Big\} \\
& + \frac{m(\theta)}{\theta}\frac{U^B}{U^A+U^B}(1-\lambda)\mathrm{E}\Big\{\max\left[(\Pi_e^B(y)-\Pi_v),0\right]\Big\}
\end{aligned} \tag{4.10}
$$

4.4.5 Wage Determination

To complete the set-up of the model, we must specify the mechanism through which wages are determined. Since the form of this problem is the same for Type-As and Type-Bs, these equations are written for both Types, $k = A, B$.

We assume that wages are determined through a Nash bargain, which divides up the surplus derived from a successful match. Bargaining power is parameterised by γ, such that wages are negotiated to solve:

$$
\max_{w^k(y)} \left[V_e^k(y) - V_u^k(y)\right]^{\gamma}\left[\Pi_e^k(y) - \Pi_v\right]^{(1-\gamma)} \; \forall k = A, B \tag{4.11}
$$

In Section 4.5 we note that the equilibrium will be characterised by free-entry of firms, such that the value of posting a vacancy, Π_v, is driven down to zero. Imposing this restriction, Equation (4.11) can be solved for the wage-equation for each Type of worker, with ability y.

$$
w^k(y) = \gamma z_0 + (1-\gamma) r V_u^k(y) \; \forall k = A, B \tag{4.12}
$$

The differences in wages between Type-As and Type-Bs enter through the unemployment term, $V_u^k(y)$. Conditional on y, this is lower for Type-Bs because their chances of successful matches to take them out of unemployment into wage-employment are reduced by discrimination.

4.5 Solving the Model

The model specified in Section 4.4 is currently written as many equations in terms of many endogenous variables. To solve for the equilibrium, we wish to summarise the model in just *one* equation in terms of *one* endogenous variable, namely θ.

To proceed, we note that the equilibrium can be defined by the following conditions:

1. Firms enter freely, such that the value of maintaining a vacancy is zero.
2. Matches are consummated if and only if it is in the interests of worker and firm to do so.
3. Steady state flows into and out of unemployment, self-employment, and wage-employment are equal, for each Type.
4. Individuals only take jobs which are worth their while, given their Type and their ability y.

Firstly, we use Condition 4 to derive some cut-off levels of productivity y, above and below which workers will forego work in either wage- or self-employment. Secondly, we use Condition 3 to derive the steady state level of unemployment in the model. Finally, we bring these components together by using Condition 1 and rewriting the asset equation for posting a vacancy, initially specified in Equation (4.10).

4.5.1 Productivity Cut-Offs

4.5.1.1 Definition and Interpretation

The job options that individuals are willing to take depend on their self-employment productivity, y, as well as their Type, $k = A, B$. Workers will take wage-employment only if its flow value is greater than unemployment, such that $V_e^k(y) > V_u^k(y)$. Similarly, workers will participate in self-employment only if $V_s^k(y) > V_u^k(y)$. This implies there may be some cut-off values of y, for each Type, above/below which certain jobs will not be taken.[21]

[21] These potential cut-offs will arise if there exist some levels of y for which $V_s^k(y) = V_u^k(y)$ or $V_e^k(y) = V_u^k(y)$.

To derive expressions for the cut-offs and consider how they may be interpreted, we begin by defining two functions which capture the difference in value between working in self- and wage-employment and being unemployed. For Type $k = A, B$, we can write:

$$X_s^k(y) \equiv V_s^k(y) - V_u^k(y) \tag{4.13}$$

$$X_e^k(y) \equiv V_e^k(y) - V_u^k(y) \tag{4.14}$$

Since $X_s^k(y)$ and $X_e^k(y)$ only differ between Type-As and Type-Bs due to the possibility of discriminatory shocks to the matching function, parameterised by λ, we define a general matching function for Type-k in Equation (4.15).

$$m^k(\theta) = \begin{cases} m(\theta) & \text{if } k = A \\ m(\theta)(1-\lambda) & \text{if } k = B \end{cases} \tag{4.15}$$

We can then use the original value functions for unemployment, self-employment, and wage-employment to express $X_s^k(y)$ and $X_e^k(y)$ (for a Type-k) as follows:

$$rX_s^k(y) = y - q_s X_s^k(y) - b - \alpha \max[X_s^k(y), 0] - m^k(\theta) \max[X_e^k(y), 0] \tag{4.16}$$

$$rX_e^k(y) = w^k(y) - q_e X_e^k(y) - b - \alpha \max[X_s^k(y), 0] - m^k(\theta) \max[X_e^k(y), 0] \tag{4.17}$$

Working with Equations (4.16) and (4.17) is complicated by the presence of the max operators. We first consider the case where both $X_s^k(y) > 0$ and $X_e^k(y) > 0$ (the 'Specific Case'), then show that the problem can be handled in exactly the same way when $X_s^k(y) \leq 0$ and/or $X_e^k(y) \leq 0$ (the 'General Case').

Specific Case

We note that all the information relating to individuals' occupational choice can be summarised in three simple equations. The first two compare the values of being in self- and wage-employment with being unemployed. The third re-expresses the wage schedule from Section 4.4.5.

$$(r + q_s + \alpha)X_s^k(y) = y - b - m^k(\theta)X_e^k(y) \tag{4.18}$$

$$(r + q_s + m^k(\theta))X_e^k(y) = w^k(y) - b - \alpha X_s^k(y) \tag{4.19}$$

$$w^k(y) = \gamma z_0 + (1-\gamma)\big[b + \alpha X_s^k(y) + m^k(\theta)X_e^k(y)\big] \tag{4.20}$$

In order to interpret the cut-off values, where $X_s^k(y) = 0$ or $X_e^k(y) = 0$, we need to understand how $X_s^k(y)$ and $X_e^k(y)$ change in y. In particular, we need to evaluate the signs of $\frac{\partial X_s^k(y)}{\partial y}$ and $\frac{\partial X_e^k(y)}{\partial y}$. Differentiating Equations (4.18), (4.19), and (4.20), we can write:

$$(r + q_s + \alpha)\frac{\partial X_s^k(y)}{\partial y} = 1 - m^k(\theta)\frac{\partial X_e^k(y)}{\partial y} \tag{4.21}$$

$$(r + q_s + m^k(\theta))\frac{\partial X_e^k(y)}{\partial y} = \frac{\partial w^k(y)}{\partial y} - \alpha\frac{\partial X_s^k(y)}{\partial y} \tag{4.22}$$

$$\frac{\partial w^k(y)}{\partial y} = (1-\gamma)\big[\alpha\frac{\partial X_s^k(y)}{\partial y} + m^k(\theta)\frac{\partial X_e^k(y)}{\partial y}\big] \tag{4.23}$$

Rearranging, we can show that $\frac{\partial X_s^k(y)}{\partial y}$ and $\frac{\partial X_e^k(y)}{\partial y}$ have opposite signs.

$$(r + q_e + \gamma m^k(\theta))\frac{\partial X_e^k(y)}{\partial y} = -\alpha\frac{\partial X_s^k(y)}{\partial y} \tag{4.24}$$

Using Equations (4.21) and (4.24), we can also write an explicit expression for $\frac{\partial X_s^k(y)}{\partial y}$, showing that $\frac{\partial X_s^k(y)}{\partial y} > 0$.

$$\frac{\partial X_s^k(y)}{\partial y} = \left[r + q_s + \alpha\left(\frac{r + q_e}{r + q_e + m^k(\theta)\gamma}\right)\right]^{-1} \tag{4.25}$$

General Case

When $X_s^k(y) \leq 0$ and/or $X_e^k(y) \leq 0$, exactly the same intuition applies, but some of the terms drop out of Equations (4.21)–(4.23). We summarise this, using indicator variables which take either 0 or 1, in the equations below.

$$(r + q_s + \mathbb{1}_{[X_s^k(y)>0]}\alpha)\frac{\partial X_s^k(y)}{\partial y} = 1 - \mathbb{1}_{[X_e^k(y)>0]}m^k(\theta)\frac{\partial X_e^k(y)}{\partial y} \tag{4.26}$$

$$(r + q_s + \mathbb{1}_{[X_e^k(y)>0]}m^k(\theta))\frac{\partial X_e^k(y)}{\partial y} = \frac{\partial w^k(y)}{\partial y} - \mathbb{1}_{[X_s^k(y)>0]}\alpha\frac{\partial X_s^k(y)}{\partial y} \tag{4.27}$$

$$\frac{\partial w^k(y)}{\partial y} = (1-\gamma)\left[\mathbb{1}_{[X_s^k(y)>0]}\alpha\frac{\partial X_s^k(y)}{\partial y} + \mathbb{1}_{[X_e^k(y)>0]}m^k(\theta)\frac{\partial X_e^k(y)}{\partial y}\right] \tag{4.28}$$

As such, we can adjust (4.24) accordingly for:

$$(r + q_e + \gamma\mathbb{1}_{[X_e^k(y)>0]}m^k(\theta))\frac{\partial X_e^k(y)}{\partial y} = -\mathbb{1}_{[X_s^k(y)>0]}\alpha\frac{\partial X_s^k(y)}{\partial y} \tag{4.29}$$

As such, $\frac{\partial X_s^k(y)}{\partial y}$ is always positive, whilst the sign of $\frac{\partial X_e^k(y)}{\partial y}$ depends on whether $X_s^k(y) > 0$ or $X_s^k(y) \leq 0$.

$$\frac{\partial X_e^k(y)}{\partial y}\begin{cases} < 0 & \text{if } X_s^k(y) > 0 \\ = 0 & \text{if } X_s^k(y) \leq 0 \end{cases} \tag{4.30}$$

These derivatives enable us to define and interpret the cut-off values of y, above/below which certain types of employment will not be taken.

First, we consider the potential cut-off where $X_s^k(y) = 0$, which we label y^{*k}. Since $X_s^k(y)$ slopes upwards, unemployment is more valuable than self-employment below y^{*k}. This means workers with $y < y^{*k}$ would never accept self-employment opportunities. Intuitively, if y is low, the returns to self-employment are insufficient to tempt workers away from queueing for wage jobs, where they could achieve far higher earnings. Conversely, if $y > y^{*k}$, unemployed workers would accept self-employment opportunities. For higher productivity workers, the returns to self-employment are enough to draw them out of unemployment. Thus, at y^{*k}, unemployed individuals are just indifferent between accepting and rejecting a self-employment opportunity.

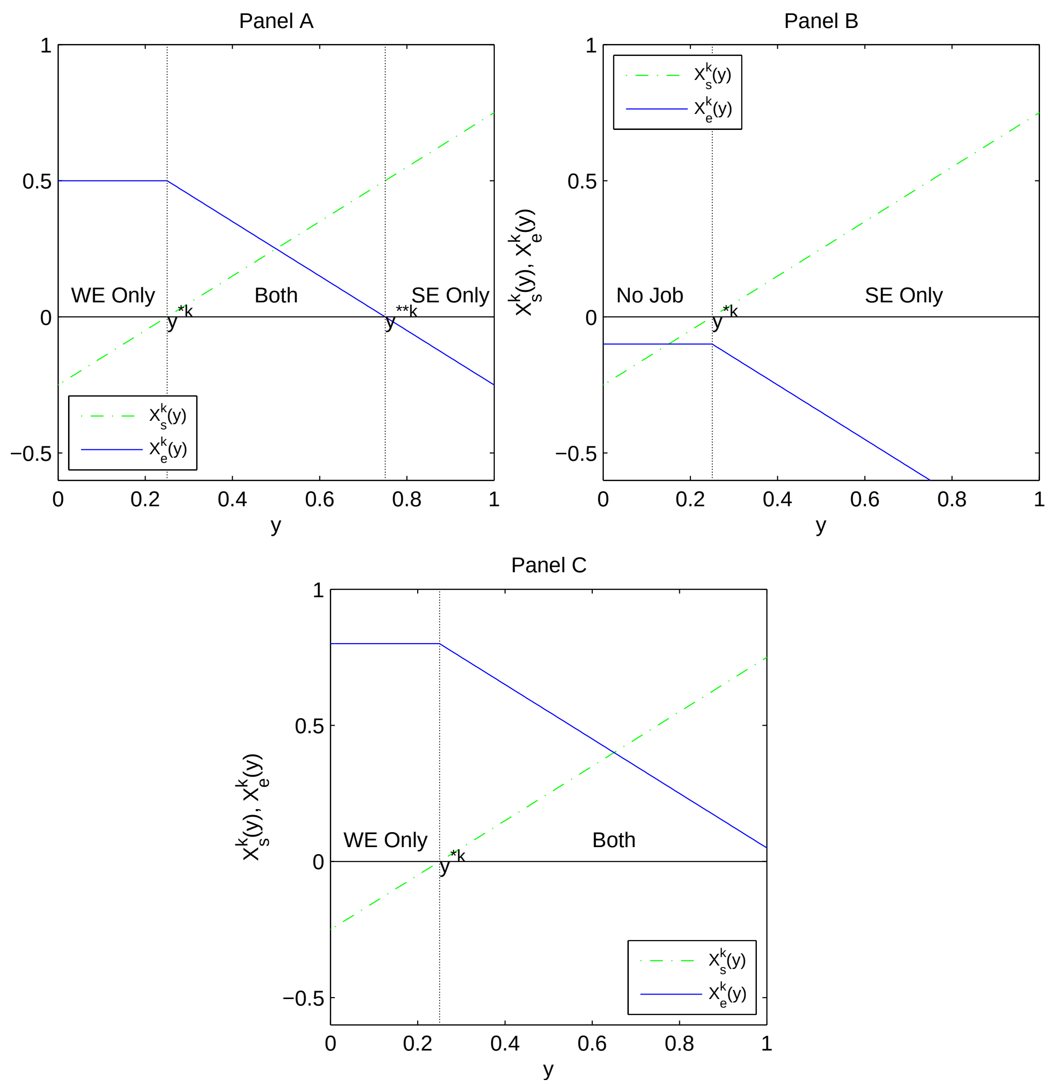
We also note that the cut-off y^{*k} has implications for the shape of the $X_e^k(y)$ schedule. In particular, we know from Equation (4.29) that $X_e^k(y)$ is flat for $X_s^k(y) \leq 0$. Intuitively, if y is so low that workers would never choose self-employment, the difference between the welfare of a wage-employed worker and an unemployed worker is invariant to y.

There is also a potential cut-off where $X_e^k(y) = 0$, which we label y^{**k}. Since $X_e^k(y)$ is weakly downward sloping, unemployment is more valuable than wage-employment *above* y^{**k}, such that workers with $y > y^{**k}$ would never accept wage-employment offers. Even if wage-employment jobs come along, these workers are better off biding their time in unemployment, and waiting for a self-employment job, in which they have high productivity and their earnings would be greater. If $y < y^{**k}$, however, unemployed workers would accept wage-employment offers. It is not in their interest to just wait for a self-employment opportunity to come along. Again, at y^{**k}, unemployed individuals are just indifferent between accepting and rejecting wage-employment offers.

We are now able to outline the types of equilibria that may emerge. Given the derivatives in Equations (4.26)–(4.29), and assuming that the lower cut-off, y^{*k}, exists, we know there are just *three* potential equilibria to consider. These are illustrated in Figure 4.4. The equilibrium that prevails will depend largely on the relative values of wage sector productivity z_0, and the flow benefit received in unemployment, b. We discuss this, and the existence of y^{*k} and y^{**k} in Appendix 4.D.

In Panel A, we show the case where both cut-offs, y^{*k} and y^{**k}, exist. Workers with very low levels of y will only ever accept opportunities in wage-employment. Since their returns to self-employment are so low, they would rather queue for wage jobs than become self-employed.

Figure 4.4: Possible Equilibria and Productivity Cut-Offs



Functions shown are stylised, and not to scale. We represent the upward and downward sloping sections of $X_s^k(y)$ and $X_e^k(y)$ as linear for simplicity.

By contrast, workers with very high productivity y will only ever accept self-employment opportunities. Their relative returns in the wage sector are too low to tempt them away from waiting for a self-employment job to come along. For individuals with $y^{*k} \leq y < y^{**k}$, however, both self- and wage-employment jobs would be accepted. These individuals would not fully specialise in either wage- or self-employment, and would transition (via unemployment) between different sectors throughout their lifetime.

In Panel B, by contrast, we see that the difference in the value of wage work and unemployment, $X_e^k(y)$, is shifted down, such that the upper cut-off y^{**k} does not exist. This situation will only arise, however, if $b \geq z_0$, which we show in Appendix 4.D. In fact we ruled out this possibility with our restriction that $b < z_0$ in Section 4.4.3. If b is high relative to z_0, the flow value of unemployment is increased so much that no worker would ever be willing to take wage jobs. As we can see in Panel B, this implies that, for workers with very low y it is never rational to exit unemployment. The returns to self-employment for these workers are insufficient to overcome the flow of unemployment benefits, b. Thus, depending on their ability, workers either take self-employment work only, or no work at all.

Finally, In Panel C we show the situation where $X_e^k(y)$ is shifted too high for y^{**k} to exist. We show in Appendix 4.D that this occurs if z_0 becomes too large relative to b. In this type of equilibrium, all individuals would be willing to take wage-employment opportunities. The lower cut-off, y^{*k}, simply divides the workers into those that would also be willing to work in self-employment ($y \geq y^{*k}$) and those that would not ($y < y^{*k}$).

In the simulations that follow in Section 4.6 we restrict the parameters z_0 and b to ensure equilibria of the type shown in Panel A of Figure 4.4 prevail.

4.5.1.2 Solving for y^{*k} and y^{**k}

We first find the lower cut-off for each Type, by noting that, at y^{*k}, the values of unemployment and self-employment are equal ($V_u^k(y^{*k}) = V_s^k(y^{*k})$). This enables us to rewrite the asset equation for the unemployed.

$$rV_u^k(y^{*k}) = b + m^k(\theta)(V_e^k(y^{*k}) - V_u^k(y^{*k})) \tag{4.31}$$

By rearranging the asset equation for a wage-employment job and substituting into Equation (4.31), we can write:

$$rV_u^k(y^{*k}) = \frac{b(r+q_e) + m^k(\theta)w^k(y^{*k})}{r+q_e+m^k(\theta)} \tag{4.32}$$

We note that, at the cut-off y^{*k}, $rV_u^k(y^{*k}) = rV_s^k(y^{*k}) = y^{*k}$. Intuitively, individuals who are just indifferent between self-employment and unemployment are not concerned about the possibility that a self-employment job might end, so the flow value to them of being self-employed would simply be their marginal product.

$$y^{*k} = \frac{b(r+q_e) + m^k(\theta)w^k(y^{*k})}{r+q_e+m^k(\theta)} \tag{4.33}$$

Finally, we can use the wage schedule in Equation (4.12) to write down an expression for y^{*k} in terms of the exogenous parameters and θ.

$$y^{*k} = \frac{b(r+q_e) + \gamma m^k(\theta)z_0}{r+q_e+\gamma m^k(\theta)} \tag{4.34}$$

To find the upper cut-off for each Type, we repeat the logic above by noting that, at y^{**k}, the values of unemployment and wage-employment are equal ($V_u^k(y^{**k}) = V_e^k(y^{**k})$). Rewriting the asset equation for the unemployed:

$$rV_u^k(y^{**k}) = b + \alpha(V_s^k(y^{**k}) - V_u^k(y^{**k})) \tag{4.35}$$

We can then rearrange the asset equation for self-employment and substitute into Equation (4.35).

$$rV_u^k(y^{**k}) = \frac{b(r+q_s) + \alpha y^{**k}}{r+q_e+\alpha} \tag{4.36}$$

Finally, we note that at the cut-off, $rV_u^k(y^{**k}) = rV_e^k(y^{**k}) = w^k(y^{**k}) = z_0$. The flow value of wage-employment for individuals who are just indifferent between wage-employment and unemployment will simply equal their wage, since the possibility of the match ending has no impact on their welfare. Also, such individuals will be able to bargain for the highest possible wage, z_0, since, given their high ability in self-employment, they have the maximum outside options possible for any individual involved in a wage bargain. We can therefore write an equation for the upper cut-off.

$$y^{**k} = \frac{z_0(\alpha + q_s + r) - b(r + q_s)}{\alpha} \tag{4.37}$$

Since the $m^k(\theta)$ and $w^k(y)$ terms have been eliminated, we can immediately see that $y^{**A} = y^{**B}$. This is because workers who are most able in self-employment are not affected by discrimination in the wage sector, regardless of their Type. They simply forego wage-employment opportunities.

4.5.2 Steady State Employment Flows

The feasible flows between self-employment, wage-employment, and unemployment, will depend on workers' ability relative to the cut-offs, y^{*k} and y^{**k}. The form of these employment flows will be the same for Type-As and Type-Bs, so we take a general approach for Type-k, once again using $m^k(\theta)$ to capture the different matching probabilities.

Firstly, we note that the proportions of time a Type-k individual of ability y spends in unemployment, self-employment, and wage-employment must add up to 1.

$$u^k(y) + n_s^k(y) + n_e^k(y) = 1 \tag{4.38}$$

Individuals with $y < y^{*k}$ will never take self-employment jobs, so the only relevant flow is between wage-employment and unemployment. This implies that:

$$q_e(1 - u^k(y)) = q_e n_e^k(y) = m^k(\theta)u^k(y) \tag{4.39}$$

By rearranging Equation (4.39) we can derive the steady state employment levels for individuals with $y < y^{*k}$.

$$u^k(y) = \frac{q_e}{m^k(\theta) + q_e} \tag{4.40}$$

$$n_s^k(y) = 0 \tag{4.41}$$

$$n_e^k(y) = \frac{m^k(\theta)}{m^k(\theta) + q_e} \tag{4.42}$$

Individuals with middling levels of self-employment ability, $y^{*k} \leq y < y^{**k}$, are willing to take both wage- and self-employment jobs. For these individuals:

$$m^k(\theta)u^k(y) = q_e n_e^k(y) \tag{4.43}$$

$$\alpha u^k(y) = q_s n_s^k(y) \tag{4.44}$$

Rearranging these equations allows us to write:

$$u^k(y) = \frac{q_e}{m^k(\theta) + q_e + \alpha\frac{q_e}{q_s}} \tag{4.45}$$

$$n_s^k(y) = \frac{\alpha\frac{q_e}{q_s}}{m^k(\theta) + q_e + \alpha\frac{q_e}{q_s}} \tag{4.46}$$

$$n_e^k(y) = \frac{m^k(\theta)}{m^k(\theta) + q_e + \alpha\frac{q_e}{q_s}} \tag{4.47}$$

Finally, for $y \geq y^{**k}$, individuals will never take wage-employment. This restricts employment flows to:

$$\alpha u^k(y) = q_s n_s^k(y) = q_s(1 - u^k(y)) \tag{4.48}$$

Thus, we can write the equilibrium employment levels for $y \geq y^{**k}$ individuals.

$$u^k(y) = \frac{q_s}{\alpha + q_s} \tag{4.49}$$

$$n_s^k(y) = \frac{\alpha}{\alpha + q_s} \tag{4.50}$$

$$n_e^k(y) = 0 \tag{4.51}$$

This reinforces the notion that, for individuals with very high self-employment productivity, discrimination in the wage sector does not affect their occupational choice. The $m^k(\theta)$ term is absent from Equations (4.49), (4.50), and (4.51).

We calculate the total number of Type-k individuals that are unemployed in two stages. First, we calculate the *average* unemployment rate, $\mathrm{E}[u^k(y)]$, for Type-k, given Equations (4.40), (4.45), and (4.49). This can be written:

$$\mathrm{E}[u^k(y)] = \int_0^{y^{*k}} u^k(y)f(y)dy + \int_{y^{*k}}^{y^{**k}} u^k(y)f(y)dy + \int_{y^{**k}}^{1} u^k(y)f(y)dy \tag{4.52}$$

As can be seen, this is the just the regular expression for the mean, modified because $u^k(y)$ takes different values depending on y^{*k} and y^{**k}. As such, $\mathrm{E}[u^k(y)]$ is a *rate* rather than a number, and must lie between 0 and 1.

The total number of unemployed Type-ks can then be calculated by simply multiplying by the number of Type-ks in the population.

$$U^k = N\pi^k \mathrm{E}[u^k(y)] \tag{4.53}$$

Total unemployment can then be written:

$$U = U^A + U^B \tag{4.54}$$

4.5.3 The Free-Entry Condition

To find the equilibrium, we write a single equation in which the only unknown is labour market tightness, θ. We do this by rewriting the asset equation for posting a vacancy in Equation (4.10), assuming that the free-entry of firms drives Π_v down to 0, and incorporating all the information recovered in Sections 4.5.1 and 4.5.2.

Setting $\Pi_v = 0$ in Equation (4.10), and noting that the max operators are redundant because firms will not accept matches that do not provide them with a positive profit, we can write:

$$c = \frac{m(\theta)}{\theta}\frac{U^A}{U^A+U^B}\mathrm{E}\Big[\Pi_e^A(y)\Big] + \frac{m(\theta)}{\theta}\frac{U^B}{U^A+U^B}(1-\lambda)\mathrm{E}\Big[\Pi_e^B(y)\Big] \tag{4.55}$$

The expectations operators in Equation (4.55) are taken over all the possible values of y, separating out Type-As and Type-Bs, with which a firm could match. Noting that firms can only match with unemployed workers, we need to apply Bayes' Law to the productivity distribution for the population.[22] The underlying productivity distribution is the same for Type-As and Type-Bs, so we label this $f(y)$. We label the productivity distribution for unemployed Type-ks, which *will* differ by Type, $f_u^k(y)$.

$$f_u^k(y) = \frac{u^k(y)f(y)}{\left({}^{U^k}\!/_{N\pi^k}\right)} \tag{4.56}$$

[22]In fact, firms can only match with a *subset* of unemployed workers with $y < y^{**k}$. However, this restriction to the sample is captured later, by bounding the integrals in Equation (4.57).

We can therefore write the free-entry condition without the expectations operators.[23]

$$
\begin{aligned}
c = & \frac{m(\theta)}{\theta} \frac{U^A}{U^A + U^B} \int_0^{y^{**A}} \Pi_e^A(y) \frac{u^A(y) f(y)}{\left({}^{U^A}/_{N\pi} \right)} dy \\
& + \frac{m(\theta)}{\theta} \frac{U^B}{U^A + U^B} (1-\lambda) \int_0^{y^{**B}} \Pi_e^B(y) \frac{u^B(y) f(y)}{\left({}^{U^B}/_{N(1-\pi)} \right)} dy
\end{aligned}
\tag{4.57}
$$

Finally, we use the asset equations for a filled job with a Type-A and a Type-B worker, as well as the relevant wage schedules, to write a final version of the free-entry condition. Since the integrands for $y < y^{*k}$ and $y^{*k} \leq y < y^{**k}$ differ, we split each integral from Equation (4.57) at y^{*A} and y^{*B}.

We are now able to write down *all* the equations required to describe the equilibrium. These are presented in Equations (4.58)–(4.64). We begin with the free-entry condition in Equation (4.58). Any unknowns in this equation are then defined in terms of the exogenous parameters in Equations (4.59)–(4.64).

$$
\begin{aligned}
c = & \frac{m(\theta)}{\theta} \frac{(1-\gamma)}{(r+q_e)} \left[\frac{U^A}{U^A + U^B} \left[\int_0^{y^{*A}} (z_0 - rV_u^A(y)) \frac{u^A(y) f(y)}{\left({}^{U^A}/_{N\pi} \right)} dy + \right.\right. \\
& \left. \int_{y^{*A}}^{y^{**A}} (z_0 - rV_u^A(y)) \frac{u^A(y) f(y)}{\left({}^{U^A}/_{N\pi} \right)} dy \right] \\
& + \frac{U^B}{U^A + U^B} (1-\lambda) \left[\int_0^{y^{*B}} (z_0 - rV_u^B(y)) \frac{u^B(y) f(y)}{\left({}^{U^B}/_{N(1-\pi)} \right)} dy + \right. \\
& \left.\left. \int_{y^{*B}}^{y^{**B}} (z_0 - rV_u^B(y)) \frac{u^B(y) f(y)}{\left({}^{U^B}/_{N(1-\pi)} \right)} dy \right] \right]
\end{aligned}
\tag{4.58}
$$

The only unknowns besides θ in Equation (4.58), aside from the unemployment rates and levels and the productivity cut-offs covered in the previous sections, are the unemployment flow value terms, $rV_u^A(y)$ and $rV_u^B(y)$. These will differ for individuals either side of y^{*k}. For workers with $y^{*k} \leq y < y^{**k}$, for whom both self- and wage-employment are an option, we can use the

[23] We are aware that the population size parameter, N, could be cancelled out by writing Equation (4.56) in terms of unemployment *rates* rather than *numbers*. However, we continue to work in absolute values to maintain clarity.

asset equations for all three occupational statuses to derive:

$$rV_u^k(y) = \frac{b(r+q_s)(r+q_e) + \alpha y(r+q_e) + \gamma m^k(\theta)z_0(r+q_s)}{(r+q_s)(r+q_e) + \alpha(r+q_e) + \gamma m^k(\theta)(r+q_s)} \tag{4.59}$$

Similarly, for workers with $y < y^{*k}$, who take only wage-employment opportunities, we can write:

$$rV_u^k(y) = \frac{b(r+q_e) + \gamma m^k(\theta)z_0}{(r+q_e) + \gamma m^k(\theta)} \tag{4.60}$$

We also restate some of the previous equations defining unemployment rates and levels and productivity cut-offs, which are present in Equation (4.58).

Firstly, the total number of unemployed Type-k individuals was given by Equations (4.52) and (4.53). We combine these for:

$$U^k = N\pi^k \left[\int_0^{y^{*k}} u^k(y)f(y)dy + \int_{y^{*k}}^{y^{**k}} u^k(y)f(y)dy + \int_{y^{**k}}^{1} u^k(y)f(y)dy \right] \tag{4.61}$$

The $u^k(y)$ terms, which give the proportion of time spent unemployed (or the unemployment rate) for a Type-k worker with productivity y, enter Equation (4.58) both directly, and through Equation (4.61). Recapitulating Section 4.5.2, we can write:

$$u^k(y) = \begin{cases} \frac{q_e}{m^k(\theta)+q_e} & \text{if } y < y^{*k} \\ \frac{q_e}{m^k(\theta)+q_e+\alpha\frac{q_e}{q_s}} & \text{if } y^{*k} \le y < y^{**k} \\ \frac{q_s}{\alpha+q_s} & \text{if } y \ge y^{**k} \end{cases} \tag{4.62}$$

Finally, we also require the productivity cut-offs, for each Type, y^{*k} and y^{**k}. We recall from Section 4.5.1 that these may be written:

$$y^{*k} = \frac{b(r+q_e)+\gamma m^k(\theta)z_0}{r+q_e+\gamma m^k(\theta)} \tag{4.63}$$

$$y^{**k} = \frac{z_0(\alpha+q_s+r)-b(r+q_s)}{\alpha} \tag{4.64}$$

As such, Equation (4.58) is an expression with only one unknown, namely θ. By imposing a functional form on the matching function and adding exogenous parameter values, we can solve this equation using numerical methods.

4.5.4 Existence and Uniqueness

4.5.4.1 Existence

Until now, we have been somewhat agnostic about the form of the matching function $m(\theta)$. For simplicity, we impose the following standard conditions on $m(\theta)$ to guarantee the existence of the equilibrium:

1. $m(\theta)$ is increasing in θ.
2. $\frac{m(\theta)}{\theta}$ is decreasing in θ.
3. $\lim_{\theta\to 0} m(\theta) = 0$ and $\lim_{\theta\to\infty} m(\theta) = \infty$.
4. $\lim_{\theta\to 0} \frac{m(\theta)}{\theta} = \infty$ and $\lim_{\theta\to\infty} \frac{m(\theta)}{\theta} = 0$.

To demonstrate that an equilibrium exists, we must show that the right-hand-side of the free-entry condition in Equation (4.58) tends to ∞ as $\theta \to 0$ and that it tends to 0 as $\theta \to \infty$. Given the assumptions made about the matching function, this is guaranteed providing the term in the large square brackets does not tend to ∞ as $\theta \to \infty$ and does not tend 0 as $\theta \to 0$.

Firstly, we note that the unemployment terms are all bounded at 0 and N, abstracting from the population parameter, π.[24] The ability distribution, $f(y)$ is independent of θ, as are z_0, r, and λ.

The only remaining terms to consider are the flow values of unemployment for each type, $rV_u^A(y)$ and $rV_u^B(y)$. Firstly, we argue that $rV_u^k(y)$ will not explode as $\theta \to \infty$. With a very high level of θ, jobs immediately become available to those who want them. Wages, however, will still be restricted by the output that matches provide to firms, z_0. This places an upper bound on the value of transitioning from unemployment to wage-employment and therefore limits $rV_u^k(y)$.

Secondly, as $\theta \to 0$, $rV_u^k(y)$ should not to tend to 0. Even as the labour market becomes very tight and individuals cannot find wage-employment jobs, they still receive unemployment flow income b, the value of which is independent of θ. They also retain the option value of transitioning into self-employment. The probability of doing this is determined entirely by α, and the flow value of self-employment is governed by y and q_e.

Since the terms in the square brackets neither explode as $\theta \to \infty$ nor tend to 0 as $\theta \to 0$, the right-hand-side of the free-entry condition in Equation (4.58) tends to ∞ as $\theta \to 0$ and tends to 0 as $\theta \to \infty$. This guarantees the existence of at least one equilibrium.

4.5.4.2 Uniqueness

The equilibrium will be unique if the right-hand-side of the free-entry condition is monotonically decreasing for all values of θ. We now consider the conditions, which will guarantee this. First,

[24] Unemployment can approach, but never reach 0, because each period some wage- and self-employment workers will lose their jobs providing $q_e > 0$ and $q_s > 0$.

we rearrange Equation (4.58) for:

$$
\begin{aligned}
c = \frac{m(\theta)}{\theta}\frac{N}{U}\frac{(1-\gamma)}{(r+q_e)}\Bigg[\pi\Bigg[\int_0^{y^{*A}} (z_0 - rV_u^A(y))u^A(y)f(y)dy + \\
\int_{y^{*A}}^{y^{**A}} (z_0 - rV_u^A(y))u^A(y)f(y)dy\Bigg] \\
+ (1-\pi)(1-\lambda)\Bigg[\int_0^{y^{*B}} (z_0 - rV_u^B(y))u^B(y)f(y)dy + \\
\int_{y^{*B}}^{y^{**B}} (z_0 - rV_u^B(y))u^B(y)f(y)dy\Bigg]\Bigg]
\end{aligned}
\tag{4.65}
$$

We consider Equation (4.65) term by term to establish that its right-hand-side is monotonically decreasing in θ, generalising for Type-k.[25]

First, we can see directly from Equation (4.62) that $u^k(y)$ is decreasing in θ both above and below the cut-off y^{*k}.

Second, since $rV_u^k(y)$ enter Equation (4.65) negatively, we show that $\frac{\partial[rV_u^k(y)]}{\partial\theta} > 0$, for all values of y.

For $0 \leq y < y^{*k}$, we can write:

$$
\frac{\partial[rV_u^k(y)]}{\partial\theta} = (z_0 - b)\frac{\gamma(q_e + r)}{(q_e + r + m^k(\theta))^2}\frac{\partial m^k(\theta)}{\partial\theta} \tag{4.66}
$$

Under the assumption $z_0 > b$, which we impose in Appendix 4.D, and given the properties of the matching function, we can see that $\frac{\partial[rV_u^k(y)]}{\partial\theta} > 0$.

For $y^{*k} \leq y < y^{**k}$

$$
\frac{\partial[rV_u^k(y)]}{\partial\theta} = \frac{\gamma(q_e + r)(q_s + r)\big[(\alpha + q_s + r)z_0 - \alpha y - b(q_s + r)\big]}{(q_s + r)(q_e + r) + \alpha(q_e + r) + \gamma m^k(\theta)(q_s + r)} \tag{4.67}
$$

[25]It is sufficient to show that each term is monotonically decreasing in θ, but not necessary.

By inspecting the definition of the cut-off in Equation (4.64) we can immediately see that for individuals with productivity below y^{**k}, $b(q_s + r) + \alpha y < z_0(\alpha + q_s + r)$. This implies that $\frac{\partial[rV_u^k(y)]}{\partial[m(\theta)]} > 0$, and hence $\frac{\partial[rV_u^k(y)]}{\partial\theta} > 0$.

The final term to consider in Equation (4.65) is $\frac{m(\theta)}{\theta}\frac{1}{U}$.[26] It turns out that $\frac{m(\theta)}{\theta}\frac{1}{U}$ will be decreasing in θ if $m(\theta)$ is 'sufficiently concave'. We formalise this concept in Appendix 4.E.

As such, we can guarantee that the equilibrium exists and is unique.

4.6 Simulations

4.6.1 Baseline Model

To help understand the key properties of the equilibrium, we begin by simulating a baseline version of our model, adapting parameter and functional form assumptions from previous work.[27]

The assumptions made for the baseline model are outlined in Table 4.9.

Table 4.9: Exogenous Parameters and Functional Forms

Parameter	Calibration Value
r	0.05
b	0.15
$m(\theta)$	$2.5\ \theta^{\left(1/2\right)}$
α	2.5
q_s	0.5
q_e	0.5
c	0.3
γ	0.5
N	1
π	0.5
λ	0.5

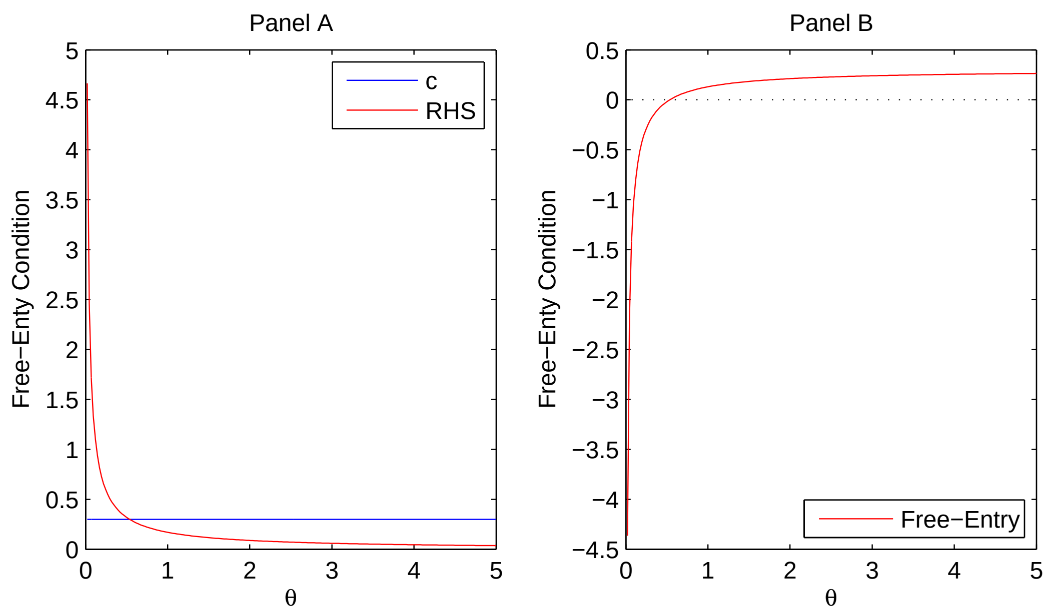
Vitally, we set $\lambda > 0$ so the baseline model captures the presence of discrimination.

[26] We take $\frac{m(\theta)}{\theta}$ and $\frac{1}{U}$ together because, on its own, $\frac{1}{U}$ is clearly increasing in θ. As the labour market becomes slacker, θ rises, and this pushes down U.

[27] In particular we use Albrecht et al. (2009) and Kerr (2012) as a starting point, and then impose the restrictions we derive in Appendices 4.D and 4.E.

We begin by plotting the free-entry condition to illustrate how we solve for the equilibrium. In Panel A of Figure 4.5, the red line shows the right-hand-side of the free-entry condition and the blue line shows the flow cost of posting a vacancy: the equilibrium level of θ is at the intersection. Solving for the equilibrium level of θ can therefore easily be recast as a root finding problem, by subtracting the right-hand-side of the free-entry condition from c, as we show in Panel B of Figure 4.5.

Figure 4.5: Plotting the Free-Entry Condition



We solve for the equilibrium level of θ with Quasi-Newton methods, in particular using Broyden's approach (Broyden, 1965; Miranda and Fackler, 2002).[28] Having solved for θ we can then derive the endogenously determined features of the equilibrium in which we are interested, especially those relating to ability, earnings, and the size of each sector.

The different distributions of ability and earnings of Type-As and Type-Bs are of particular interest for analysing the effects of discrimination in the baseline model. We define the $h_j^k(y)$ as the density of ability in sector $j = u, e, s$, for a Type-k individual.

We demonstrate how to calculate the ability distribution for the unemployed as an example.[29] Bayes' Law allows us to write $h_u^k(y)$ in terms of individuals' unemployment rates, $u^k(y)$, which depend on productivity y, as well as the underlying distribution of productivity in the population,

[28] We use the MATLAB function `broyden`, which is available as part of the CompEcon Toolbox downloaded from `http://www4.ncsu.edu/ pfackler/compecon/toolbox.html`.

[29] This is analogous to the calculations made in Section 4.5.3.

$f(y)$, and the average unemployment rate for Type-ks $\left({}^{U^k}/_{N\pi^k}\right)$.

$$h_u^k(y) = \frac{u^k(y)f(y)}{\left({}^{U^k}/_{N\pi^k}\right)} \tag{4.68}$$

Since $u^k(y)$ takes different values either side of the cut-offs y^{*k} and y^{**k}, we can thus summarise the Probability Density Function (PDF) for $h_u^k(y)$ in Equation (4.69).

$$h_u^k(y) = \begin{cases} \frac{q_e}{m^k(\theta)+q_e}\frac{N\pi^k}{U^k} & \text{if } y < y^{*k} \\ \\ \frac{q_e}{m^k(\theta)+q_e+\alpha\frac{q_e}{q_s}}\frac{N\pi^k}{U^k} & \text{if } y^{*k} \leq y < y^{**k} \\ \\ \frac{q_s}{\alpha+q_s}\frac{N\pi^k}{U^k} & \text{if } y \geq y^{**k} \end{cases} \tag{4.69}$$

We plot the analogous Cumulative Distribution Function (CDF) for ability amongst the unemployed, $H_u^k(y)$, in Figure 4.6. The distribution for the Type-As is shown by the hard blue line, whilst the distribution for the Type-Bs is shown by the dashed red line.

The CDFs in Figure 4.6 have three components because unemployed individuals may plausibly have productivity levels below y^{*k} and above y^{**k} (as well as between the two cut-offs). However, when we calculate the relevant ability distributions for the self-employed, $H_s^k(y)$, and wage-employed, $H_e^k(y)$, we find the CDFs have only two components. This is because there are no wage-employed individuals with $y \geq y^{**k}$, and no self-employed individuals with $y < y^{*k}$. The ability CDFs for self- and wage-employment are plotted in Figures 4.7 and 4.8 respectively.

In self-employment, individuals simply earn their marginal product, such that Figure 4.7 can be interpreted in terms of both ability and earnings. This is not, however, the case amongst the wage-employed. The wage distribution, which we label $m(w)$ only varies between $w^k(y^{*k})$ and $w^k(y^{**k})$. All individuals with $y < y^{*k}$ simply receive $w^k(0) = w^k(y^{*k})$. This is because wage-employment productivity, z_0, is homogenous, and the outside option of unemployment, valued at $rV_u^k(y)$, is constant for these individuals.[30] However, for individuals with $y^{*k} \leq y < y^{**k}$, this is not the case since, for them, $rV_u^k(y)$ varies in y (see Equation (4.59)). We equate the probability mass associated with the 'block' of $y^{*k} \leq y < y^{**k}$ in the ability distribution with

[30] We recall that $w^k(y) = \gamma z_0 + (1-\gamma)rV_u^k(y)$.

Figure 4.6: Unemployment Ability Distribution

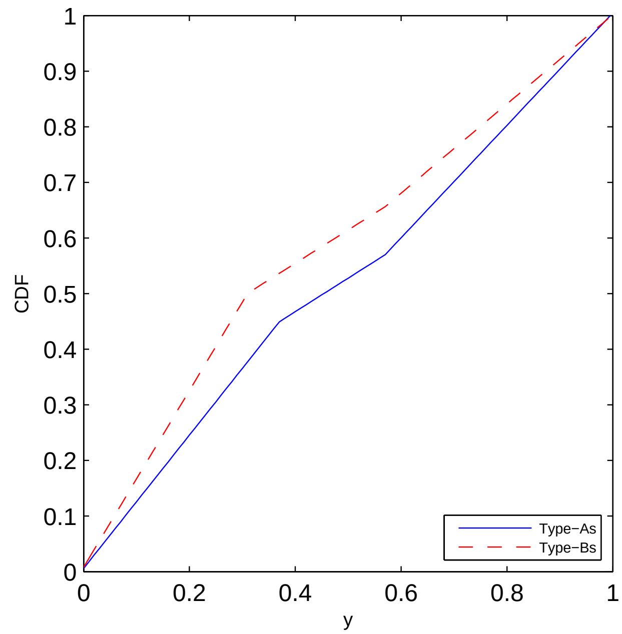


Figure 4.7: Self-Employment Ability Distribution

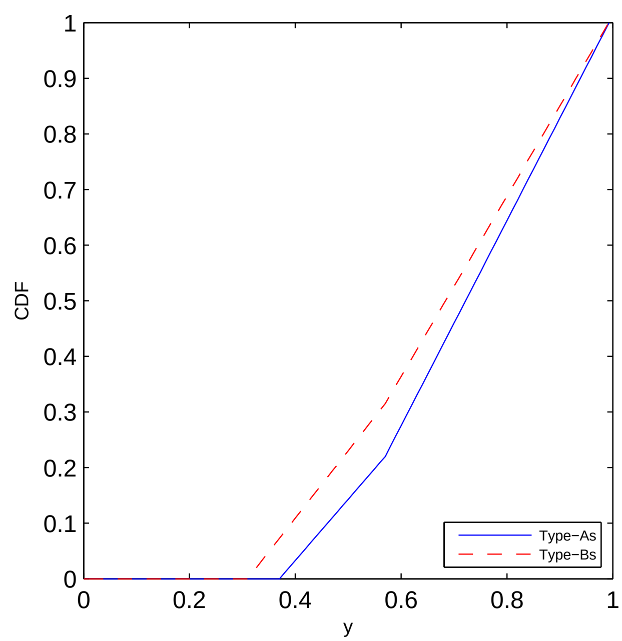
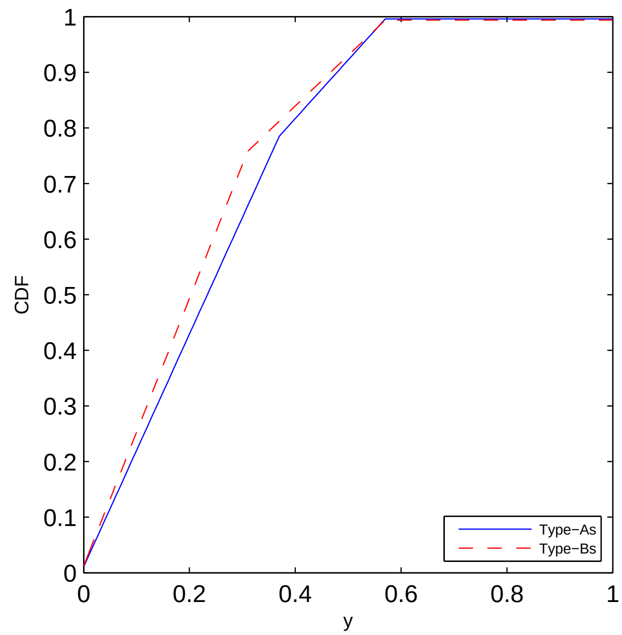


Figure 4.8: Wage-Employment Ability Distribution



that in the earnings distribution. As such, we can plot the CDFs for the wages of both Type-As and Type-Bs as shown in in Figure 4.9.

For each sector, the ability and earnings CDFs for the Type-As appear to First-Order Stochastically Dominate (FOSD) the CDFs for the Type-Bs. The implication, however, that mean ability is lower for Type-Bs than Type-As in *all* three sectors is prima facie somewhat puzzling, since both Types have the same underlying ability distribution, $f(y)$. To reconcile this finding, we must consider how discrimination ($\lambda > 0$) affects the labour market sorting of Type-As and Type-Bs differently.

To do this, we first present the other key properties of the baseline model's equilibrium in Table 4.10. We can now examine unemployment, self-employment, and wage-employment in turn to understand how the average ability of Type-Bs can be lower in all three sectors.

There are more unemployed Type-Bs than Type-As, resulting directly from the extra frictions they face in obtaining wage jobs. This drives down average ability for the unemployed Type-Bs because *lower* productivity workers are less able to exit unemployment than when there was no discrimination. The behaviour of the higher productivity unemployed Type-Bs, however, is

Figure 4.9: Wage-Employment Earnings Distribution

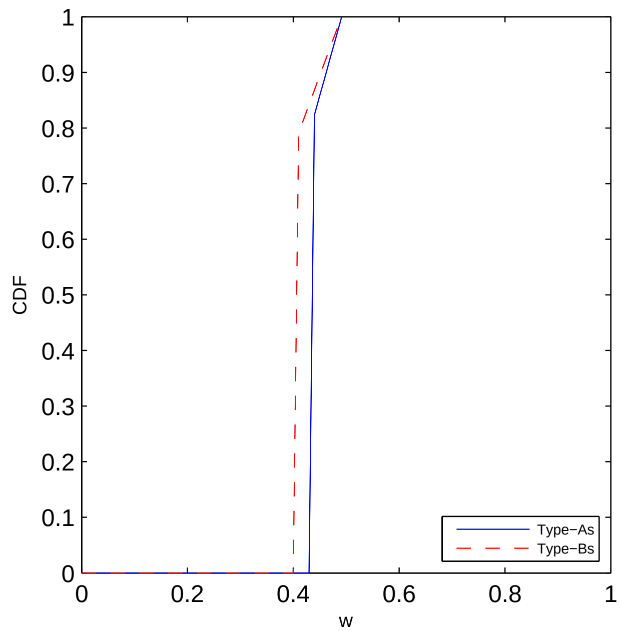


Table 4.10: Baseline Model Equilibrium

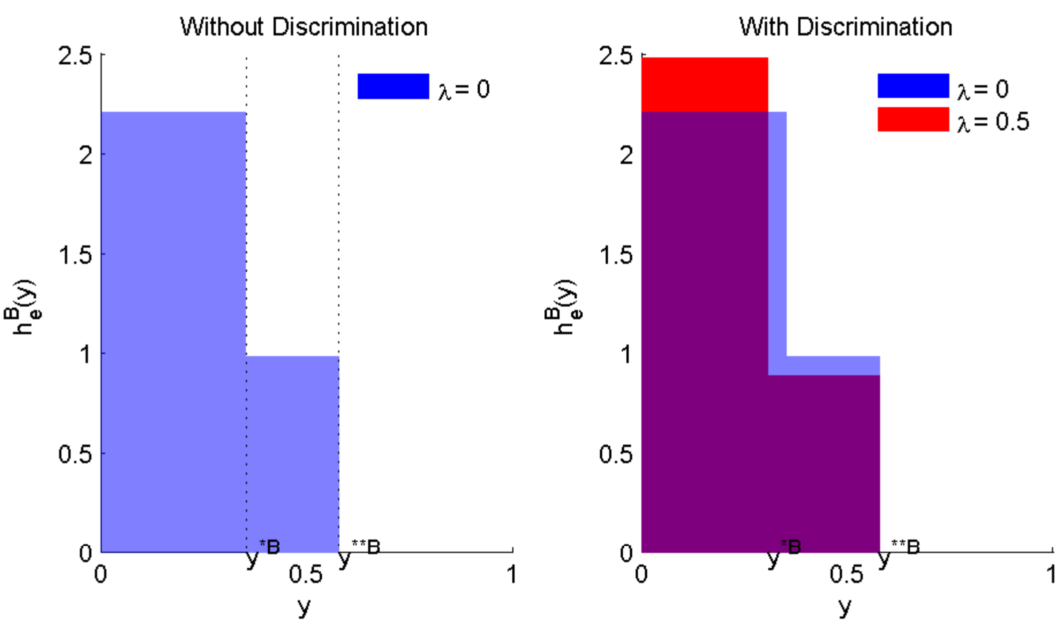
	Type-As	**Type-Bs**
u^k	0.0824	0.1038
n_s^k	0.2260	0.2569
n_e^k	0.1916	0.1393
$E[h_u^k(y)]$	0.4798	0.4175
$E[h_s^k(y)]$	0.7200	0.6814
$E[h_e^k(y)]$	0.2493	0.2268
$E[m^k(w)]$	0.4450	0.4197
y^{*k}	0.3770	0.3180
y^{**k}	0.5770	0.5770

unchanged.

There are also more self-employed Type-Bs than Type-As. The extra frictions in trying to obtain wage jobs mean that more lower ability Type-Bs are willing to take both wage- and self-employment. This is reflected in the fact that $y^{*A} > y^{*B}$ in Table 4.10. Thus, the extra Type-B workers in self-employment are of lower ability, on average, than those who would be self-employed were there no discrimination.

Finally, Type-Bs' average ability in wage-employment is lower than Type-As' because it is the *higher* ability wage-employed Type-Bs who are induced to take both self- and wage-employment by the extra frictions associated with discrimination. Thus, even though there are fewer wage-employed Type-Bs, these workers have lower productivity, on average. This intuition is reflected in Figure 4.10, in which we draw the PDF of ability in the wage sector first without (blue), then with (red) discrimination. Importantly, a high level of λ reduces y^{*B}. This means individuals with productivity just below y^{*B} (when there is no discrimination) become much less likely to be wage-employed when $\lambda > 0$. Put differently, it is the most able wage-employed workers that are pushed out of the wage sector by discrimination.

Figure 4.10: Wage Sector Ability PDF With and Without Discrimination



Taking these effects together, we can see how it is possible for average ability to be lower for Type-As than Type-Bs in all three sectors. To reinforce this idea, we decompose average ability for Type-Bs in the three sectors in Equation (4.70). The average ability in each sector

$\mathrm{E}\big[h_j^B(y)\big]\forall j = u, s, e$ is weighted by the proportion of Type-B individuals in that sector. The underbraces in Equation (4.70) show how each term responds when discrimination is added into the model.

$$\mathrm{E}^B[y] = \underbrace{\mathrm{E}\big[h_u^B(y)\big]}_{-}\underbrace{\left(\frac{U^B}{N(1-\pi)}\right)}_{+} + \underbrace{\mathrm{E}\big[h_e^B(y)\big]}_{-}\underbrace{\left(\frac{n_e^B}{N(1-\pi)}\right)}_{-} + \underbrace{\mathrm{E}\big[h_s^B(y)\big]}_{-}\underbrace{\left(\frac{n_s^B}{N(1-\pi)}\right)}_{+} \tag{4.70}$$

The first and last term have an overall negative sign, whilst the second term has an overall positive sign. Thus, the sign of the overall expression is unchanged by discrimination, reflecting the fact that the ability distribution of Type-Bs remains the same as Type-As, even though average Type-B ability is reduced in all three sectors.

As such, it is plausible that discrimination lowers average ability for Type-Bs in all three sectors. This reduces average earnings in both self- and wage-employment. Labour market frictions prevent individuals from choosing jobs on the basis of comparative advantage alone. These imperfections in the sorting process lower the average ability for both Types in each sector. However, since Type-Bs face extra frictions due to discrimination, their occupational selection departs even further from their comparative advantage, resulting in the earnings differential between Type-As and Type-Bs we observe in all three sectors.[31]

This analysis also provides some insight into the distributional effects of discrimination. In Table 4.10, we can see that higher levels of λ only affect the lower cut-off for Type-Bs, y^{*B}. Thus the sorting behaviour of high-ability Type-Bs is unaffected by the presence of discrimination, since these individuals would never choose wage-employment anyway, leaving their welfare robust to $\lambda > 0$. The opposite, however, is true for low-ability Type-Bs, who continue to try obtaining wage jobs, in spite of the extra frictions and the increased chance of being unemployed. Overall, this reflects the simple logic that those individuals with the best outside options fare better when discrimination occurs.

Additionally, if we interpret Type-As as men and Type-Bs as women, we can see that the baseline model matches the stylised facts from the data in Section 4.3. Men are disproportionately represented in wage-employment, whilst women are over-represented in self-employment.

[31] We emphasise that there are gaps between Type-Bs and Type-As in terms of both average earnings and average ability. Thus even if individuals were paid their marginal product, as in a competitive labour market, earnings differentials would persist.

Average earnings differentials persist in both sectors, but are larger for the self-employed than the wage-employed.[32] Also, earnings are more dispersed in self-employment than wage-employment. Finally, since for both Types y^{**k} exceeds y^{*k} significantly, there are some individuals who do not specialise in either wage- or self-employment. This means our model can explain the occurrence of inter-sector transitions.

4.6.2 Comparative Statics by Simulation

4.6.2.1 Changing λ

We now simulate the model using the same assumptions made in Table 4.9, but varying the value of the discrimination parameter λ between 0 (where there is no discrimination) and 1 (where discrimination rules out wage-employment for Type-Bs).

Firstly, we consider how the size of each sector changes as discrimination becomes more severe. This is shown aggregating Type-As and Type-Bs in Figure 4.11 and is then broken down by Type in Figure 4.12. The hard blue line represents wage-employment, the dashed red line represents unemployment, and the dash-dot green line represents self-employment. As expected, the increased frictions push workers away from wage-employment towards self-employment, on average. The number of unemployed individuals also rises, but more modestly.

This aggregate trend is entirely driven by the Type-Bs. Their participation in wage-employment is driven to 0 as $\lambda \to 1$, such that they are split between self-employment and unemployment. By contrast, the Type-As are drawn out of unemployment and self-employment towards wage-employment when λ increases. This is because they effectively face a slacker labour market when they are better insulated from competition from the Type-Bs, increasing their chances of successfully matching with a firm and gaining a wage-employment job. Thus, our model predicts that discrimination exerts *positive* spillovers on Type-As.

We consider the impact that changing λ has on the ability distribution in Figure 4.13. The hard blue, dashed red, and dash-dot green lines represent mean ability in wage-employment, unemployment, and self-employment respectively. Mean wages are also shown by the dotted blue line. Average ability and earnings in all three sectors are increasing in λ for Type-As and decreasing in λ for Type-Bs, echoing the key insights from the baseline model above. This

[32] This would be true even if the wage-employed were paid their marginal product.

Figure 4.11: Changing λ and Sectoral Size

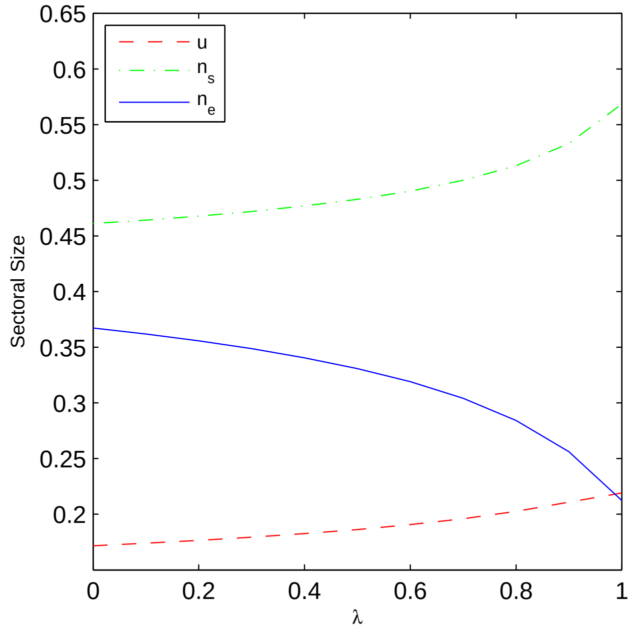
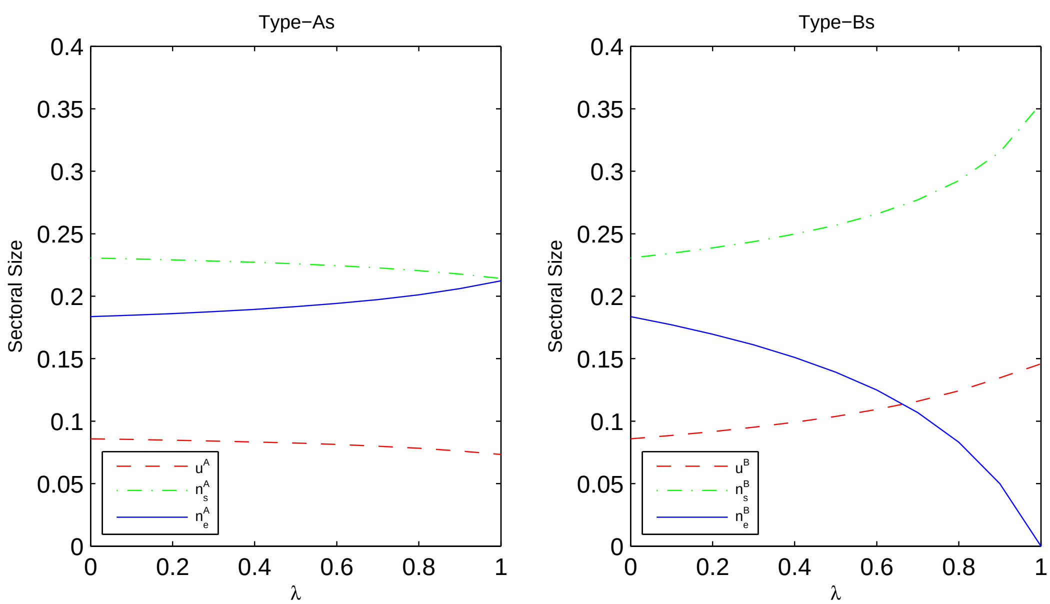
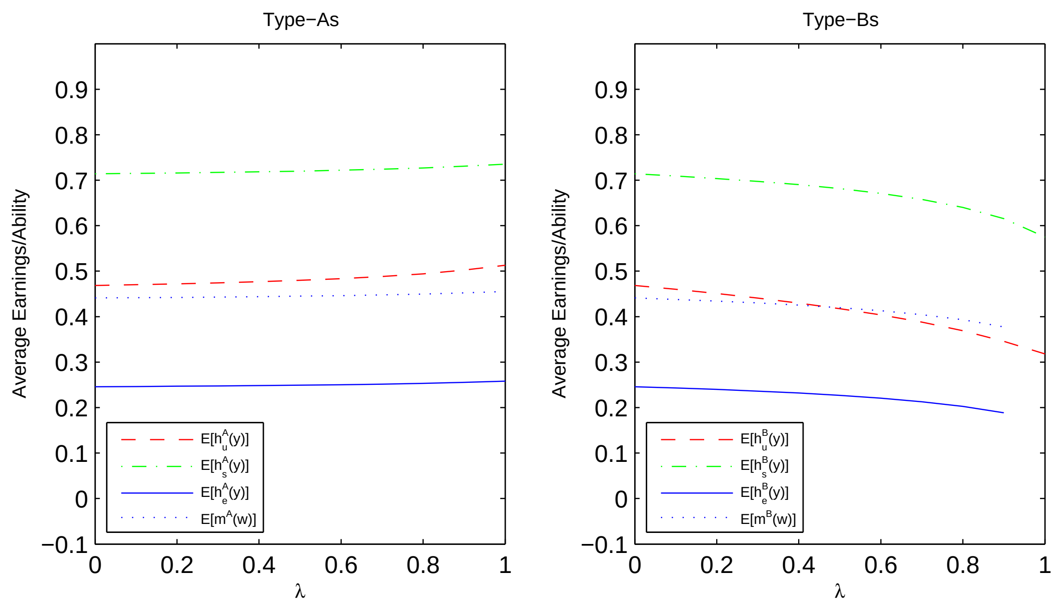


Figure 4.12: Changing λ and Sectoral Size by Type



further emphasises the positive spillover effects, which discrimination against Type-Bs exerts on Type-As.

Figure 4.13: Changing λ and Ability/Earnings by Type



We also plot the response of the cut-offs y^{*k} and y^{**k} to changing λ, to confirm the intuition we presented for the baseline model. These results are shown in Figure 4.14. Importantly, increasing λ substantially reduces y^{*B}. This means individuals with productivity just below y^{*B} (when there is no discrimination) become much less likely to be wage-employed when $\lambda > 0$, as they are suddenly willing to take both wage- and self-employment. This is the key effect which reduces the average ability of wage-employed Type-Bs in the presence of discrimination.

4.6.2.2 Growth

Increases in wage sector productivity, z_0, can be interpreted as a proxy for economic growth, insofar as incomes and expenditures in the economy are determined by firms' output. To consider the impact that growth might have, we simulate our model for different levels of z_0, maintaining the restrictions described in Appendix 4.D and the assumption that discrimination persists ($\lambda = 0.5$).

In Figure 4.15, we plot the size of each sector by Type for different values of z_0.

Figure 4.14: Changing λ and the Cut-Offs by Type

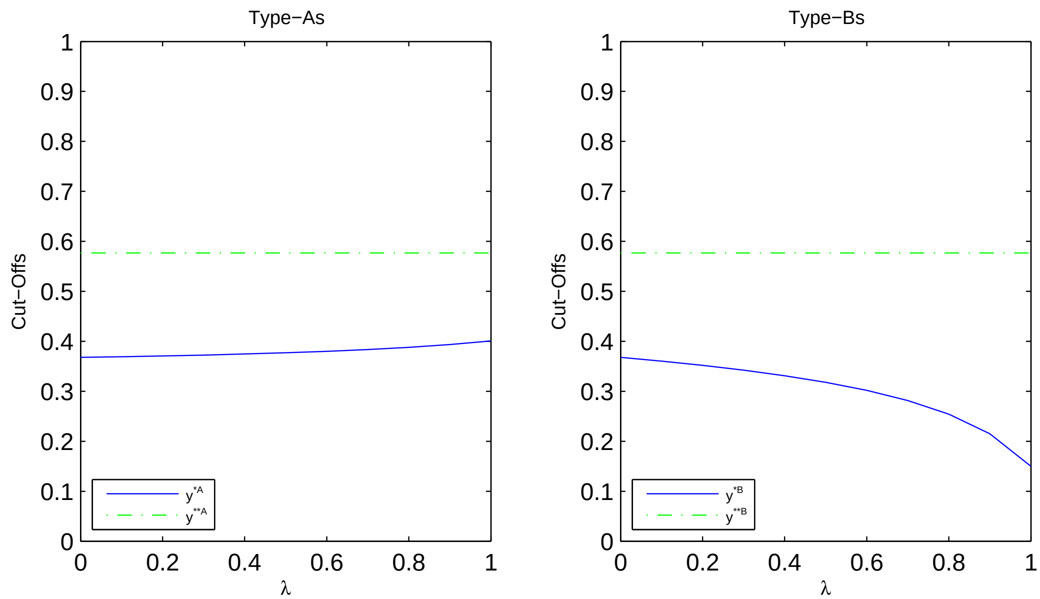
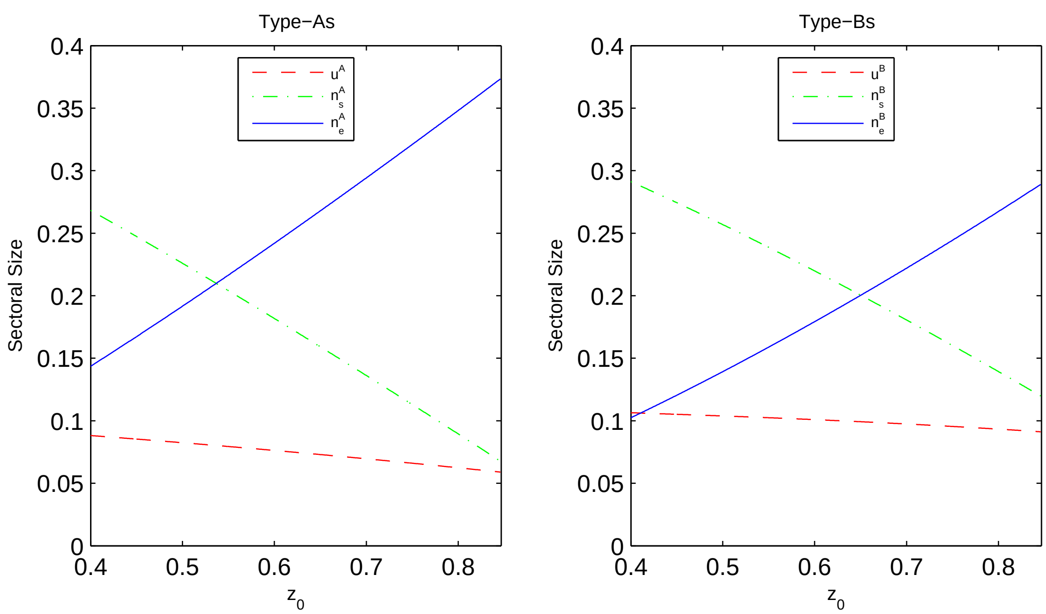
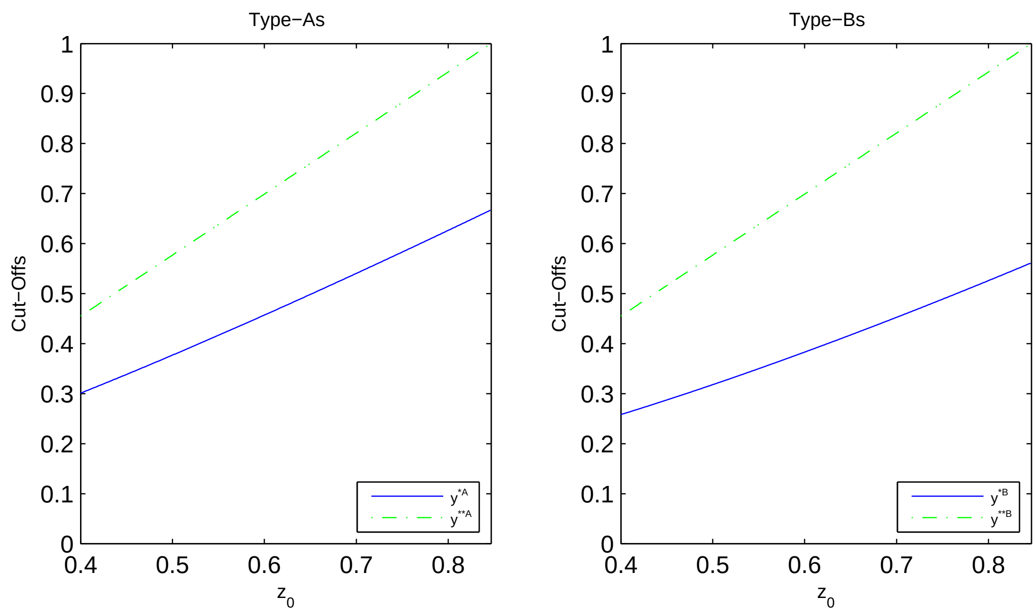


Figure 4.15: Changing z_0 and Sectoral Size by Type



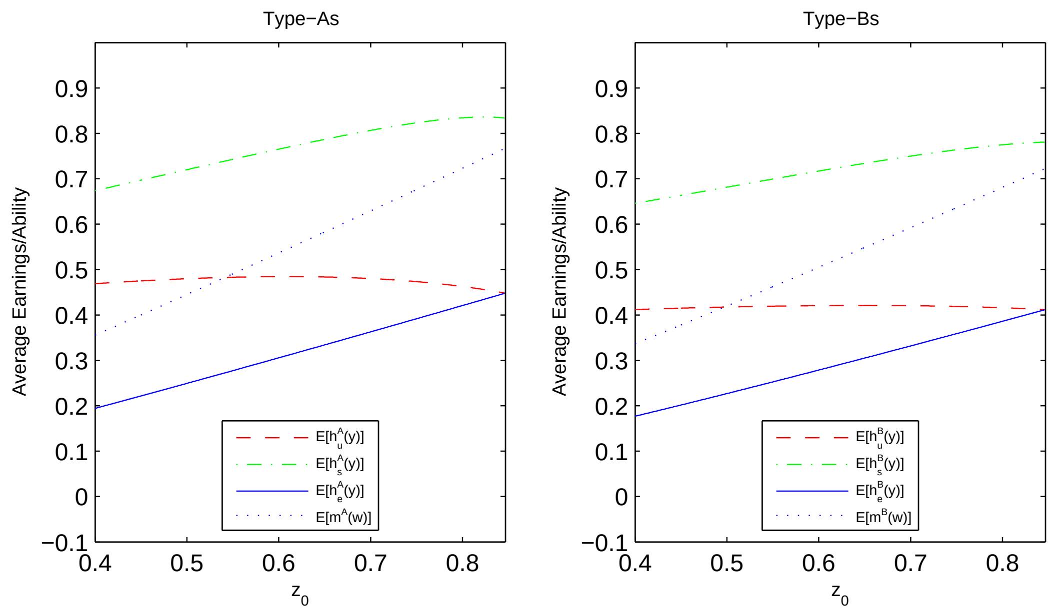
As expected, the size of the wage sector for both Types grows, because productivity in and returns to wage work are increased. Interestingly, these individuals are drawn out of both self-employment and, to a lesser extent, unemployment. The decline in self-employment is intuitive, since the returns in that sector become relatively lower for both Types. The fall in unemployment can be understood with reference to the cut-offs, which we plot in Figure 4.17. Growth increases both cut-offs, but with the baseline calibration, y^{**k} rises faster. As such, the proportion of individuals who are willing to take both self- and wage-employment rises. As the number of workers who are not fully specialised falls, so does the level of unemployment.

Figure 4.16: Changing z_0 and the Cut-Offs by Type



We also plot the average ability in each sector by Type in Figure 4.17. Growth increases the earnings in self- and wage-employment for both Type-As and Type-Bs. The rise in average ability and earnings in the wage sector is intuitive. The increase in z_0 raises match surplus and hence wages directly, such that higher ability workers are more likely to undertake wage work and less likely to fully specialise in self-employment. The simultaneous increase in average ability in self-employment is explained by the fact that it is the *lowest* ability self-employed workers that are drawn towards wage work first. As such, self-employment becomes a smaller and more specialised sector as the economy grows.

Figure 4.17: Changing z_0 and Ability/Earnings by Type



4.6.2.3 Social Security

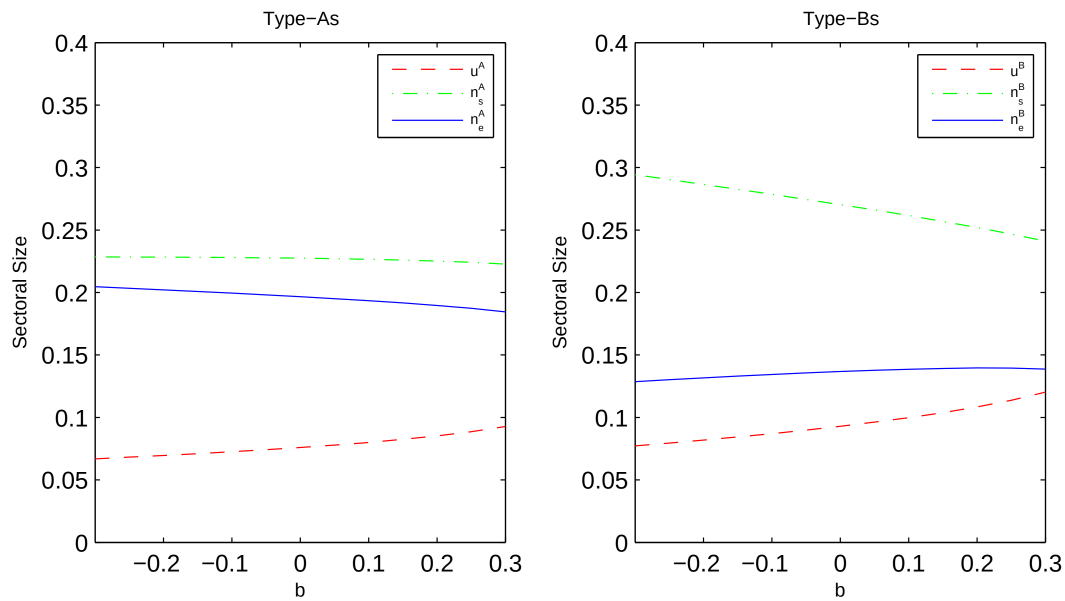
Finally, we consider the implications of changing the flow value of being unemployed, b, which can be understood as a measure of the government's provision of social security. We plot the size of each sector by Type in 4.18.[33]

Improved social security provision increases unemployment and reduces the size of the self-employment sector for both Type-As and Type-Bs. These effects are far more pronounced on the Type-Bs, because their initial levels of self-employment and unemployment are higher. Interestingly, increasing b has opposing effects on the size of the wage sector for Type-As and Type-Bs. For both Types, social security increases the extent to which individuals specialise in either self- or wage-employment, since individuals can afford to wait for jobs that better suit their comparative advantage.[34] When the market is characterised by discrimination, the frictions for entry into self-employment *relative* to those for wage-employment are greater for Type-As than Type-Bs. Thus, as waiting becomes less costly, Type-As may be drawn out of wage work, into unemployment with the hope of finding a self-employment job. Conversely, the extra willingness of Type-Bs to wait allows them to enter the wage sector, in a way that was prohibitively costly

[33] We provide supplementary plots showing the response of average ability and the productivity cut-offs in Appendix 4.G.

[34] See Figure 4.26 in Appendix 4.G.

Figure 4.18: Changing b and Sectoral Size by Type



when b was low.

Overall, therefore, social security reduces the costs associated with unemployment, and helps individuals overcome the initial frictions in the market to select occupations which better suit their comparative advantage. The distinct responses of Type-As and Type-Bs arise because the initial frictions they face are different.

4.7 Discussion

4.7.1 Heterogeneous Wage-Employment Productivity

The logic of Roy's (1951) classic model of occupational selection is present throughout our analysis. However, there is one simplifying assumption in our approach, which makes a potentially important departure from Roy's framework. In particular, we assume that productivity in wage-employment is homogeneous and fixed at z_0. By contrast, Roy's model allows for some joint distribution of wage- and self-employment ability, allowing for a flexible correlation structure between them.[35] In this section, we consider the implications of allowing for heterogeneous productivity in both sectors.

To do this, we adopt a very simple relationship between wage-employment productivity $z(y)$ and self-employment productivity, y. We assume that this structure is the same for Type-As and Type-Bs.

$$z(y) = \beta_0 + \beta_1 y \tag{4.71}$$

The modification feeds directly into the wage schedule for both Types, which may now be written:

$$w^k(y) = \gamma z(y) + (1-\gamma) r V_u^k(y) = \gamma(\beta_0 + \beta_1 y) + (1-\gamma) r V_u^k(y) \tag{4.72}$$

First, we note that our initial formulation of the model is a special case of this more general model, where $\beta_0 = z_0$ and $\beta_1 = 0$. In fact, we can show that our original logic remains intact, providing β_1 does not become too high. In particular, high ability individuals with $y \geq y^{**k}$ will only accept self-employment jobs, low ability individuals with $y < y^{*k}$ will only accept wage work, and those in the middle will do both.

We also note that, even if β_1 is sufficiently low and the original model remains intact, *both* the productivity cut-offs, y^{*k} and y^{**k} are increasing in *both* of the productivity parameters, β_0 and β_1.[36] This is somewhat intuitive, since as individuals become more productive in wage-employment, either on average (β_0) or relative to their self-employment ability (β_1), their chances of selecting into wage work, or selecting away from self-employment, should rise. The extent to which β_1 departing from 0 affects the amount of specialisation in the economy, depends on whether y^{*k} or y^{**k} increases faster in β_1, which in turn depends on the exogenous parameters

[35] Roy's original formulation was couched in terms of hunters and fishermen.

[36] Specifically, we can write:

$$y^{*k} = \frac{b(r+q_e) + \gamma m^k(\theta)\beta_0}{r + q_e + \gamma m^k(\theta)(1-\beta_1)} \tag{4.73}$$

$$y^{**k} = \frac{\beta_0(\alpha + q_s + r) - b(q_s + r)}{\alpha - \beta_1(\alpha + q_s + r)} \tag{4.74}$$

and the matching function.[37] If the upper cut-off is more responsive to β_1, then increasing β_1 widens the gap between the two productivity cut-offs, and increases the proportion of the population who are willing to take both self- and wage-employment jobs. The converse is true, however, if the lower cut-off is more responsive to β_1.

We recognise, however, that if β_1 becomes large then the interpretation of the productivity cut-offs is reversed. In contrast to the original model, it may be that high-ability individuals select into wage-employment, and low-ability individuals select into self-employment. Using the same notation as in Section 4.5.1 we label the difference between the value of self-employment and unemployment $X_s^k(y)$ and the difference between the value of wage-employment and unemployment $X_e^k(y)$. We also maintain the cut-off definitions: $X_s^k(y^{*k}) = 0$ and $X_e^k(y^{**k}) = 0$.

We display the four main types of equilibria that may arise with a high values of β_1 in Figure 4.19. A more formal treatment of this issue is reserved for Appendix 4.H.

In Panel A, we show the most plausible equilibrium, where the cut-offs from the original model have simply been reversed. Individuals with productivity $y \geq y^{*k}$ will take only wage-employment, those with $y < y^{**k}$ will take only self-employment, whilst those in the middle will accept both types of jobs. Moving up the y distribution has a greater impact on wages than self-employment earnings, so the most able individuals are drawn towards the wage sector.

As in the original model, these equilibria break down if $X_e^k(y)$ is too low (Panel B) or too high (Panel C). This is determined by the parameters of the productivity distribution, β_0 and β_1 relative to the level of unemployment benefit, b. There is also a fourth possibility to consider, since $X_e^k(y)$ is no longer flat for $X_s^k(y) \leq 0$, which is shown in Panel D. In this scenario, workers with $y^{*k} \leq y < y^{**k}$ end up rejecting both types of jobs, even though very high- and very low-ability individuals accept wage- and self-employment jobs respectively.

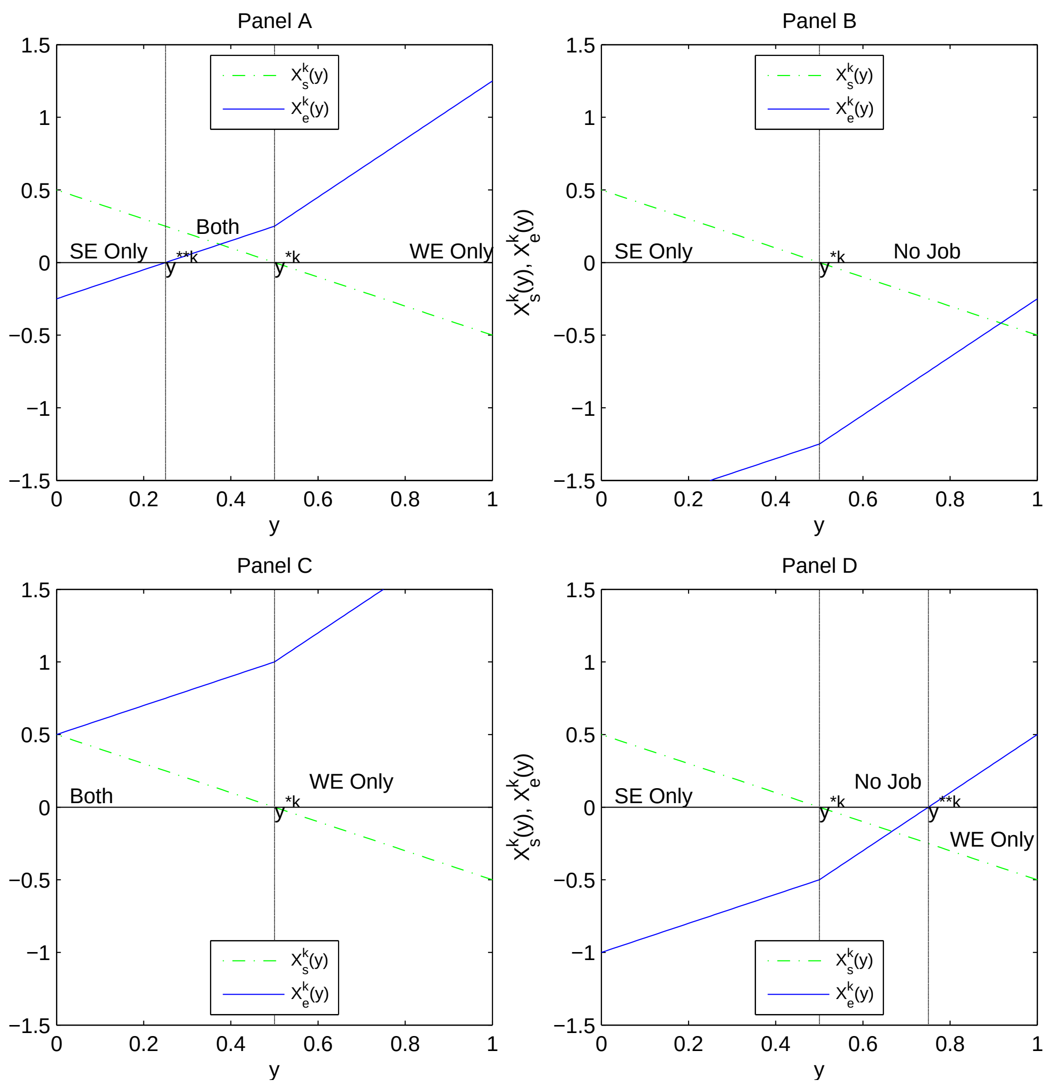
Our model can thus be straightforwardly extended to allow for the flexible relationship between ability in different occupations, inherent in the Roy (1951) model. However, we believe that the simpler approach we outlined originally is sufficient to investigate the impact of discrimination on earnings and labour market sorting, which is the main aim of this chapter. In

[37] Differentiating equations (4.73) and (4.74) with respect to β_1 (and treating θ as a constant), we have:

$$\frac{\partial y^{*k}}{\partial \beta_1} = \frac{\gamma m^k(\theta)\big[b(r+q_e) + \gamma m^k(\theta)\beta_0\big]}{\big[r + q_e + \gamma m^k(\theta)(1-\beta_1)\big]^2} \tag{4.75}$$

$$\frac{\partial y^{**k}}{\partial \beta_1} = \frac{(\alpha + q_s + r)\big[\beta_0(\alpha + q_s + r) - b(q_s + r)\big]}{\big[\alpha - \beta_1(\alpha + q_s + r)\big]^2} \tag{4.76}$$

Figure 4.19: Possible Equilibria and Productivity Cut-Offs with High β_1



Functions shown are stylised, and not to scale. We represent $X_e^k(y)$ and $X_s^k(y)$ as linear for simplicity.

particular, high values of β_1 suggest that variation is greater in wage-employment productivity than in self-employment productivity. However, insofar as ability proxies for earnings, this pattern of dispersion is at odds with the data. Thus, not only is our initial framework more parsimonious, but it may also better fit the Ghanaian economy.

4.7.2 Intra-Sector Analysis

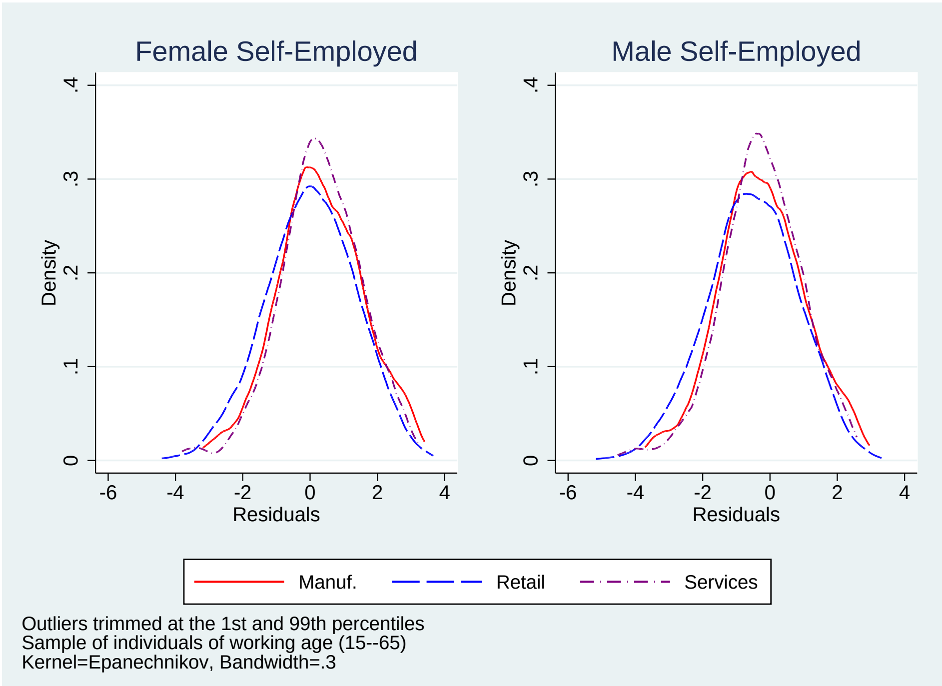
In Chapter 3 we demonstrated that intra-sector heterogeneity — such as across industry and technology — is vital for understanding labour market outcomes. The descriptive statistics in Section 4.3 confirm that, for our urban sample, there are differences in how women and men are allocated to different types of jobs *within* self- and wage-employment. For example, focussing on the self-employed, women are concentrated mainly in retail/trade. In contrast, men are better represented in services, both for the self- and wage-employed sub-samples.

As we observed in Chapter 3, workers in certain self-employment industries earn more than others, for reasons other than observable human capital. Similar patterns emerge for the urban sample used in this chapter. Focussing on the self-employed, if we regress the log of earnings on education, age, and tenure, using our pooled sample, and then plot the residuals, we can see that, although the differences are small, services and manufacturing once again outperform retail.[38] These plots are shown in Figure 4.20. We also note that in retail, earnings are somewhat more dispersed and the sex earnings gap is slightly larger, when compared with services. The earnings distributions for the self-employed retailers and service sector workers are shown, by sex, in Appendix 4.I. Thus, the empirical patterns for services and retail amongst the self-employed, are analogous to the empirical patterns for wage- and self-employment for the economy at large.

To the extent that individuals choose between industries on the basis of their comparative advantage and the relative frictions they face in obtaining certain jobs, our model could also be applied at the intra-sector level. If we abstract from the process of matching workers and firms momentarily, our model simply assumes that wage sector jobs are harder to obtain than self-employment jobs. Individuals are less likely to transit from unemployment to wage-employment (than they are to self-employment), and women have even less chance than men. Although self-employed workers do not have to deal with employers per se, it is certainly possible that some jobs *within self-employment* are characterised by more sex-specific frictions than others, perhaps due to social norms or discrimination in input or output markets. We thus believe

[38] As in Chapter 3, we run these regressions using OLS. We also add dummy variables for the survey wave and city, and control for marital status and number of children. These regressions are reported in Table 4.14 in Appendix 4.I.

Figure 4.20: Residuals from OLS Earnings Equations for the Self-Employed



our model could be straightforwardly modified to explain why self-employed women are over-represented in retail activities, even though the earnings gaps here are large. Namely, the frictions associated with starting service sector businesses are larger than for retail, and even worse for female entrepreneurs.

4.7.3 Conclusion

In this chapter, we construct a search-match model of the labour market, in which individuals can participate in both self- and wage-employment, where the latter is characterised by discrimination against some sub-sample of the workforce. We allow for worker heterogeneity both in terms of productivity and in terms of the extent to which they face discrimination. In contrast, firms are homogeneous ex ante, but each firm contains an exogenously determined proportion of prejudiced 'interviewers', which captures the level of discrimination in the economy. This set-up allows us to analyse the distributional impacts of the extra frictions women may face in trying to obtain wage jobs. Our model performs well when compared with a large longitudinal dataset collected in urban Ghana between 2004 and 2013. In particular, we can explain the different occupational selection patterns of men and women, as well as the relative sizes of the earnings gaps in each sector.

Vitally, our model can account for the fact that average male earnings exceed average female earnings in self-employment, wage-employment, and unemployment, *without* assuming any differences between the ability distributions for each sex. Drawing heavily on the logic of the Roy (1951) model, we show that extra labour market frictions push individuals further away from the occupations that best match their comparative advantage. Since discrimination worsens these frictions for women, their occupational selection is distorted more than the choices of men, driving down their average ability in each sector.

We also predict that discrimination may result in positive spillover effects for men. Higher sex-specific frictions insulate men against competition from women in the labour market, which increases the likelihood that they obtain and maintain wage jobs. This directly reduces the number of men in unemployment and self-employment. More importantly, however, it allows men to select occupations more in accordance with their comparative advantage, which raises average male earnings in both self- and wage-employment.

We also use our model to investigate the implications of policy on a labour market characterised by informality and discrimination. Growth, which we capture using wage sector productivity z_0, draws individuals out of self-employment towards wage work, and raises the average earnings of women and men in both sectors. Under our calibration, specialisation in the economy also falls, since the number of individuals that are only willing to work in self-employment falls faster than the number of wage-only workers rises.

Our model produces somewhat surprising predictions about the impact of social security on labour market sorting. As anticipated, higher levels of b increase the number of unemployed and reduce the number of self-employed workers for both Types. However, the size of the wage sector responds differently, depending on whether workers face discrimination. Indeed, our model predicts that extra social security actually raises wage-employment among women, because waiting becomes less costly, which, in turn, helps to overcome the added initial frictions they face in getting wage work.

We also show that our simple model can be modified in one key way, to allow for heterogeneity in wage work productivity, as well as ability in self-employment. By adding a more flexible structure to the relationship between productivity in each sector, our model converges even closer on the ideas of the Roy model. We show that under certain formulations of this relationship, the intuition of our original model remains largely the same. However, the way we interpret the occupational selection problem is altered, if there is sufficient variation in wage-employment ability compared with self-employment ability. In spite of this potential modification, we argue that our original simple model is suitable for capturing discrimination in a developing country labour market, both in terms of trends in the data and from the perspective of parsimony.

Although our model focusses on the distinction between wage- and self-employment, we believe the same logic can also be applied at the intra-sector level. It may be that some self-employment jobs are harder to obtain than others, and that the extent of these extra frictions are sex-specific. With this set-up, our model could be modified to help explain some of the more nuanced heterogeneity *within* wage- and self-employment observed in the data.

Overall, this chapter develops a model which can explain some of the key differences between men and women in the urban Ghanaian labour market. The extra frictions that female workers face in obtaining wage jobs can generate earnings differentials in both wage- and self-employment, and account for the prevalence of entrepreneurship amongst women. Although we adopt a stylised approach, our model provides a series of useful insights into the gender dimensions of labour market segmentation.

4.A Sampling and Attrition

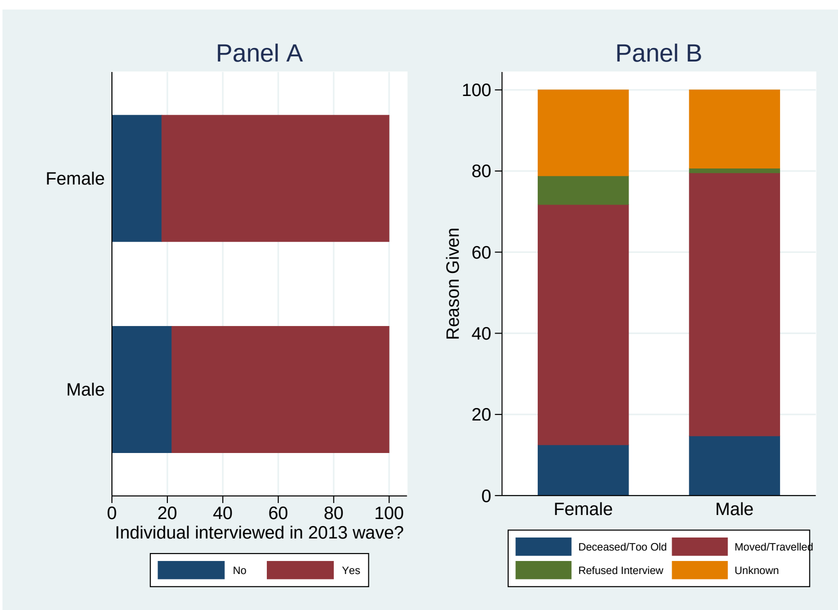
The data for this chapter are taken from the Ghana Household Urban Panel Survey, using all waves available between 2005 and 2013. We omit the 2004 data because the survey instrument used was so different in that year.

The initial samples were based on a stratified random sample of urban households in Accra (including Tema), Kumasi, Takoradi, and Cape Coast, using the 2000 Ghanaian Census. Although sampling occurred at the household level, interviews were undertaken at the individual level, so any household members of working age (15–60) were tracked (Teal, 2008).

We noted in Section 4.3 that there were more women then men in the sample, pooling across all waves 2005-2013. We now consider why this pattern emerges.

One possibility is that men attrite from the survey more than women. In Panel A of Figure 4.21, we show the number of individuals of each sex who were successfully tracked and interviewed in the 2013 wave. Approximately 28 percent of men and 22 percent of women for whom interviews were scheduled in 2013 could not be tracked, suggesting differential attrition may be important.

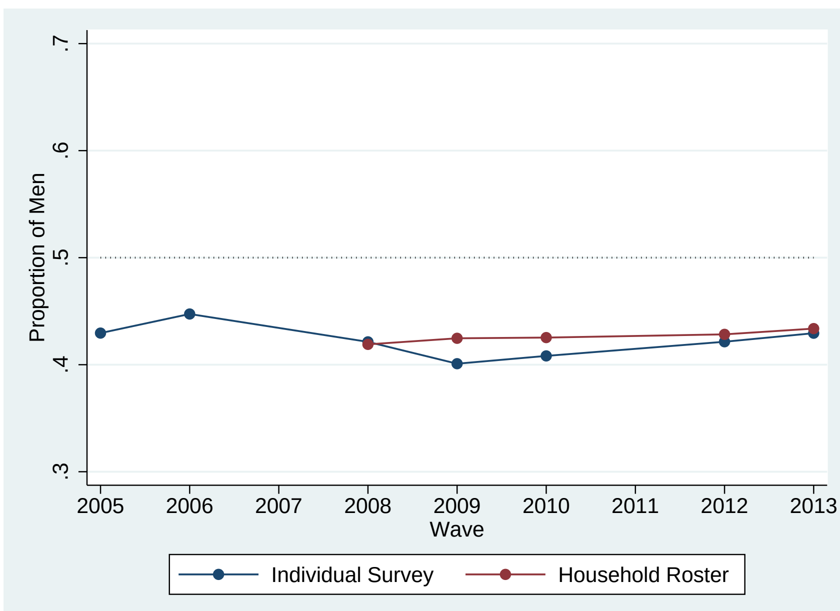
Figure 4.21: Attrition in 2013 by Sex



There are also slight differences in the reasons given by other household members for *why* potential respondents of each sex could not be tracked and interviewed. For example, it appears that men are somewhat more likely to have deceased or moved away from the survey area than women. This is shown in Panel B of Figure 4.21.

However, if we look at the proportion of men successfully interviewed in the survey over time, we do not see a clear downward trend. The proportions of men and women of working age present on the household roster are shown by the red line in Figure 4.22. The household roster, which is conducted with the household head, lists *all* individuals in each household, even if they cannot be interviewed, so this should be representative of the potential population from which respondents may be drawn. The blue line shows the proportion of individuals who actually responded to the individual labour force survey that year. Although fewer men answered the individual interview than were present in the household roster, these differences have, if anything decreased since 2009, as tracking procedures improved.

Figure 4.22: Proportion of Men Interviewed over Time



The fact that the proportion of working age men in the household roster is also below 50 percent suggests there may be underlying differences between the number of working age men and women in the population. This is partially confirmed by looking at data from the 2010 Ghanaian Census, as shown in Table 4.11.

We can see that around 52 percent of the 15–64 year olds in the 2010 Ghanaian Census sample were female. This is still some way below the 57 percent we see for working age individuals on the household roster, but there do appear to be differences between the number of women and men in Ghana as a whole.

Table 4.11: Sex and Age Breakdown in 2010 Ghanaian Census

Age	Sex					
	Female		Male		Total	
	No.	%	No.	%	No.	%
0–14	4651454	18.86	4798944	19.46	9450398	38.32
15–64	7312945	29.66	6727948	27.28	14040893	56.94
65+	669579	2.72	497953	2.02	1167532	4.73
Total	12633978	51.24	12024845	48.76	24658823	100.00

4.B Earnings Distributions in Wage-Employment

Figure 4.23: Small-Firm Wage-Employment Earnings by Sex

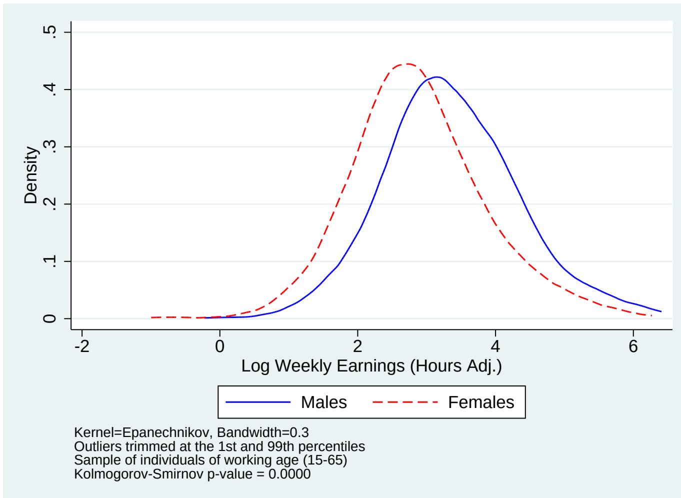


Figure 4.24: Large-Firm Wage-Employment Earnings by Sex

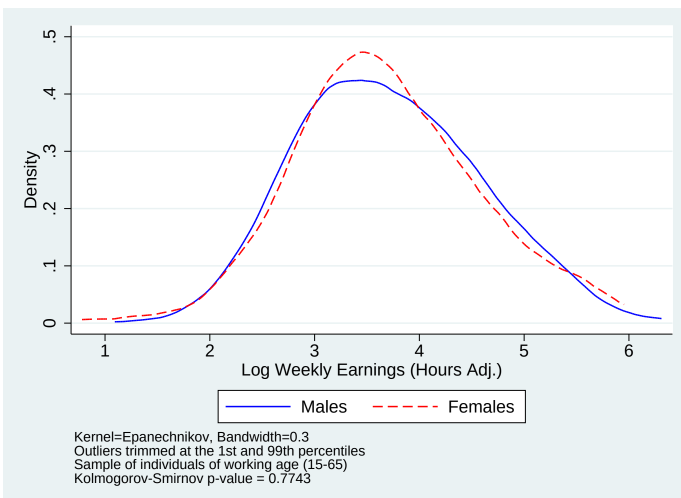
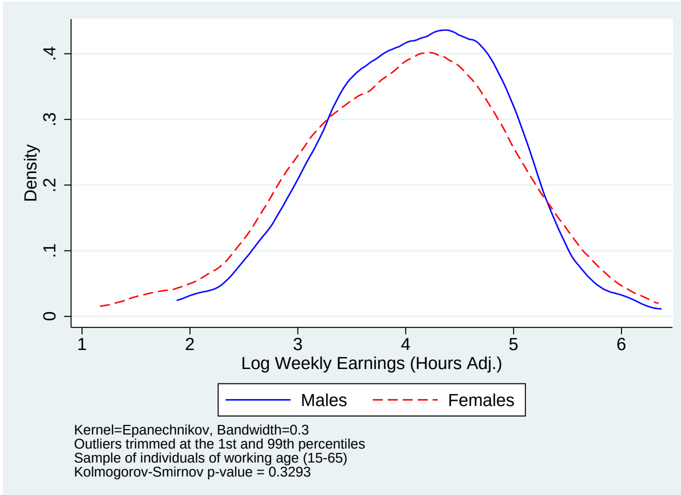


Figure 4.25: Public Wage-Employment Earnings by Sex



4.C Detailed Transition Matrices

Table 4.12: Detailed Transition Matrix for Men (2012–2013)

	2013 Occupation							
2012 Occupation	Apprentice %	Out of LF %	Priv. WE (Large) %	Priv. WE (Small) %	Pub. WE %	SE %	UE %	Total %
Apprentice	1.91	0.30	0.40	1.31	0.00	0.00	0.60	4.52
Out of LF	0.60	17.29	1.21	2.91	0.50	1.11	5.33	28.94
Priv. WE (Large)	0.00	0.70	4.82	3.32	1.41	0.60	0.60	11.46
Priv. WE (Small)	0.40	2.61	2.81	12.16	0.60	3.72	2.21	24.52
Pub. WE	0.00	0.40	1.31	0.60	2.51	0.00	0.50	5.33
SE	0.00	0.70	0.70	3.42	0.00	11.46	0.80	17.09
UE	0.30	2.21	1.01	1.71	0.00	0.90	2.01	8.14
Total	3.22	24.22	12.26	25.43	5.03	17.79	12.06	100.00

Table 4.13: Detailed Transition Matrix for Women (2012–2013)

	2013 Occupation							
2012 Occupation	Apprentice %	Out of LF %	Priv. WE (Large) %	Priv. WE (Small) %	Pub. WE %	SE %	UE %	Total %
Apprentice	1.45	0.36	0.07	0.87	0.00	0.43	0.51	3.68
Out of LF	0.58	18.14	0.72	2.31	0.58	3.47	5.13	30.92
Priv. WE (Large)	0.07	0.29	1.37	1.23	0.36	0.29	0.36	3.97
Priv. WE (Small)	0.29	1.95	0.72	5.06	0.14	2.46	1.52	12.14
Pub. WE	0.07	0.58	1.08	0.36	1.81	0.14	0.36	4.41
SE	0.07	3.83	0.22	2.46	0.14	25.58	2.24	34.54
UE	0.14	3.32	0.58	1.01	0.00	2.24	3.03	10.33
Total	2.67	28.47	4.77	13.29	3.03	34.61	13.15	100.00

4.D Existence of the Productivity Cut-Offs

The productivity cut-offs will only exist if they lie between 0 and 1, given our assumption for both Type-As and Type-Bs that productivity $y \sim \text{Uniform}(0, 1)$. In this appendix, we consider the conditions that will guarantee the existence of both cut-offs.

4.D.1 Existence of y^{*k}

We begin by recalling that the lower cut-off for each Type, y^{*k}, may be written:

$$y^{*k} = \frac{b(r + q_e) + \gamma m^k(\theta) z_0}{r + q_e + \gamma m^k(\theta)} \tag{4.77}$$

Since all the terms on the right-hand-side of Equation (4.77) are non-negative by construction, it is clear that the condition that $y^{*k} \geq 0$ is met trivially.

We can also see that $y^{*k} \leq 1$ providing b does not become too large. In particular, we require:

$$b \leq \frac{r + q_e + \gamma m^k(\theta)(1 - z_0)}{r + q_e} \tag{4.78}$$

By comparing the numerator and the denominator in Equation (4.77), we can see $y^{*k} \leq 1$ is guaranteed by simply imposing that $b \leq 1$. Moreover, since $0 \leq z_0 \leq 1$, this will hold if $b \leq z_0$.

4.D.2 Existence of y^{**k}

To guarantee the existence of the upper cut-off, we derive two conditions which rule out the equilibria shown in Panel B and Panel C of Figure 4.4 (see Section 4.5.1.1).

To exclude the possibility that Panel B prevails, we require that $X_e^k(y^{*k}) > 0$. Since

$X_s^k(y^{*k}) = 0$ by definition, we can write:

$$X_e^k(y^{*k}) = \frac{\gamma(z_0 - b)}{r + q_e + \gamma \mathbb{1}_{[X_e^k(y)>0]} m^k(\theta)} \tag{4.79}$$

Thus, $X_e^k(y^{*k}) > 0$ if $z_0 > b$.

To ensure that $X_e^k(y)$ is not too high, ruling out the situation in Panel C of Figure 4.4, we require that $y^{**k} \leq 1$. In particular, this requires

$$z_0 < \frac{\alpha + b(r + q_s)}{\alpha + q_s + r} \tag{4.80}$$

This condition restricts the values of z_0 (relative to b), since otherwise the surplus and hence earnings from matches in wage-employment are too large for any individuals to forego wage-employed jobs. Also, greater values of α make this condition more likely to be met.

4.D.3 Simulation Conditions

Summarising the lessons from Appendix 4.D, we impose two restrictions:

1. $z_0 > b$
2. $z_0 < \frac{\alpha+b(r+q_s)}{\alpha+q_s+r}$

These are sufficient to guarantee the existence of both cut-offs, y^{*k} and y^{**k}.

4.E Functional Form of $m(\theta)$ and Uniqueness of the Equilibrium

To establish the uniqueness of the equilibrium, it is sufficient to show that that $\frac{m(\theta)}{\theta}\frac{1}{U}$ is decreasing in θ. In this appendix, we consider the properties of the matching function that will guarantee this.

We begin with the simplifying assumption that there are no Type-Bs in the model, such that $U = U^A$. We also note that:

$$\frac{\partial}{\partial\theta}\left[\frac{\theta}{m(\theta)}U\right] > 0 \Rightarrow \frac{\partial}{\partial\theta}\left[\frac{m(\theta)}{\theta}\frac{1}{U}\right] < 0 \tag{4.81}$$

We define $I(\theta) \equiv \frac{\theta}{m(\theta)}U$. Using the expression for U derived in Section 4.5.2 and the properties of the Uniform distribution, we can write an expression for $I(\theta)$.[39]

$$\begin{aligned}
I(\theta) =& \frac{\theta}{m(\theta)}\left[\frac{1}{m(\theta)+\bar{q_e}}\left[\frac{f(\theta)}{m(\theta)+q_e}+C\right]+D\right] \\
&\text{where} \\
&\bar{q_e} = q_e + \frac{\alpha q_e}{q_s} \\
&f(\theta) = \frac{\alpha q_e^2}{q_s}y^*(\theta) \\
&C = q_e y^{**} \\
&D = \frac{q_s}{\alpha+q_s}(1-y^{**})
\end{aligned} \tag{4.82}$$

It is convenient to rewrite Equation (4.82) as:

$$I(\theta) = \underbrace{\frac{\theta f(\theta)}{m(\theta)(m(\theta)+\bar{q_e})(m(\theta)+q_e)}}_{①} + \underbrace{\frac{\theta C}{m(\theta)(m(\theta)+\bar{q_e}}}_{②}) + \underbrace{\frac{\theta D}{m(\theta)}}_{③} \tag{4.83}$$

[39] We discuss how to use the Uniform distribution to write U in Appendix 4.F.

Differentiating each term separately, we have:

$$\frac{\partial \textcircled{1}}{\partial\theta} = \frac{m(\theta)(m(\theta)+\bar{q_e})(m(\theta)+q_e)(\theta f^{'}(\theta)+f(\theta)) - \theta f(\theta)m^{'}(\theta)\big[3m(\theta)^2+2m(\theta)(q_e+\bar{q_e})+q_e\bar{q_e}\big]}{\big[m(\theta)(m(\theta)+\bar{q_e})(m(\theta)+q_e)\big]^2} \tag{4.84}$$

$$\frac{\partial \textcircled{2}}{\partial\theta} = \frac{C\Big[m(\theta)(m(\theta)+\bar{q_e}) - \theta m^{'}(\theta)\big[2m(\theta)+\bar{q_e}\big]\Big]}{\big[m(\theta)(m(\theta)+\bar{q_e})\big]^2} \tag{4.85}$$

$$\frac{\partial \textcircled{3}}{\partial\theta} = \frac{D\Big[m(\theta)-\theta m^{'}(\theta)\Big]}{\big[m(\theta)\big]^2} \tag{4.86}$$

To illustrate the idea that Equations (4.84), (4.85), and (4.86) will be positive if the matching function is 'sufficiently concave', we now add functional form to $m(\theta)$. In particular, we assume that:

$$m(\theta) = s\theta^{(1/p)} \tag{4.87}$$

Under this functional form assumption, we can rewrite Equations (4.84), (4.85), and (4.86) as:

$$\frac{\partial \textcircled{1}}{\partial\theta} = \frac{\Bigg[m(\theta)^3+m(\theta)^2(q_e+\bar{q_e})+m(\theta)\bar{q_e}q_e - \Big[\big(3/p\big)m(\theta)^3+\big(2/p\big)m(\theta)^2(q_e+\bar{q_e})+\big(1/p\big)m(\theta)\bar{q_e}q_e\Big]\Bigg]f(\theta)+K}{\big[m(\theta)(m(\theta)+\bar{q_e})(m(\theta)+q_e)\big]^2}$$

where

$$K = \theta f^{'}(\theta)m(\theta)(m(\theta)+\bar{q_e})(m(\theta)+q_e) \tag{4.88}$$

$$\frac{\partial \textcircled{2}}{\partial \theta} = \frac{\Big[m(\theta)\big(1-\left(^2/_p\right)\big) + \bar{q_e}\big(1-\left(^1/_p\right)\big)\Big]m(\theta)C}{\big[m(\theta)(m(\theta)+\bar{q_e})\big]^2} \tag{4.89}$$

$$\frac{\partial \textcircled{3}}{\partial \theta} = \frac{\big(1-\left(^1/_p\right)\big)m(\theta)D}{\big[m(\theta)\big]^2} \tag{4.90}$$

Since $f(\theta)$ and $f^{'}(\theta)$ are positive, under the condition that $z_0 > b$, we can see that it is sufficient to set $p \geq 3$ to guarantee that $\frac{\partial \textcircled{1}}{\partial \theta}$, $\frac{\partial \textcircled{2}}{\partial \theta}$, and $\frac{\partial \textcircled{3}}{\partial \theta}$ are all positive, and hence that $I^{'}(\theta) > 0$. In turn this implies that $\frac{\partial}{\partial \theta}\left[\frac{m(\theta)}{\theta}\frac{1}{U}\right] < 0$.

Setting $p \geq 3$ is a sufficient, but not a necessary condition for a unique equilibrium.

For our simulations in Section 4.6, we use $p = 2$ as this fits the existing empirical literature. We check that the right-hand-side of the free-entry condition is monotonically decreasing for all values of θ, given the parameters of the model, in each simulation.

4.F Uniform Distribution

Since we have assumed a standard uniform distribution for ability, $y \sim \text{Uniform}(0,1)$, we can rewrite the total unemployment terms for each type, U^k.

$$\begin{aligned}
U^k &= N\pi^k\left[\int_0^{y^{*k}} u^k(y)f(y)dy + \int_{y^{*k}}^{y^{**k}} u^k(y)f(y)dy + \int_{y^{**k}}^{1} u^k(y)f(y)dy\right] \\
&= N\pi^k\left[\int_0^{y^{*k}} \frac{q_e}{\kappa^k(\theta)+q_e}f(y)dy + \int_{y^{*k}}^{y^{**k}} \frac{q_e}{\kappa^k(\theta)+q_e+\alpha\frac{q_e}{q_s}}f(y)dy + \int_{y^{**k}}^{1} \frac{q_s}{\alpha+q_s}f(y)dy\right] \\
&= N\pi^k\left[\frac{q_e}{\kappa^k(\theta)+q_e}\int_0^{y^{*k}} f(y)dy + \frac{q_e}{\kappa^k(\theta)+q_e+\alpha\frac{q_e}{q_s}}\int_{y^{*k}}^{y^{**k}} f(y)dy + \frac{q_s}{\alpha+q_s}\int_{y^{**k}}^{1} f(y)dy\right] \\
&= N\pi^k\left[\frac{q_e}{\kappa^k(\theta)+q_e}y^{*k} + \frac{q_e}{\kappa^k(\theta)+q_e+\alpha\frac{q_e}{q_s}}(y^{**k}-y^{*k}) + \frac{q_s}{\alpha+q_s}(1-y^{**k})\right]
\end{aligned} \tag{4.91}$$

4.G Social Security Comparative Statics

Figure 4.26: Changing b and the Cut-Offs by Type

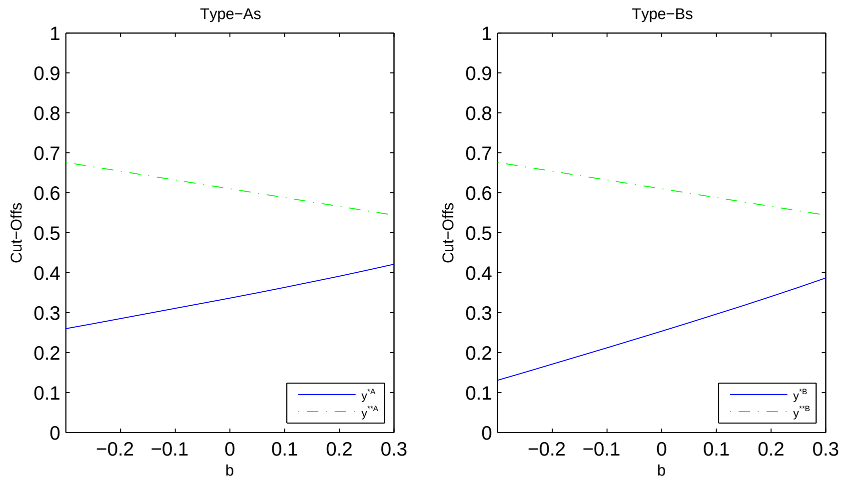
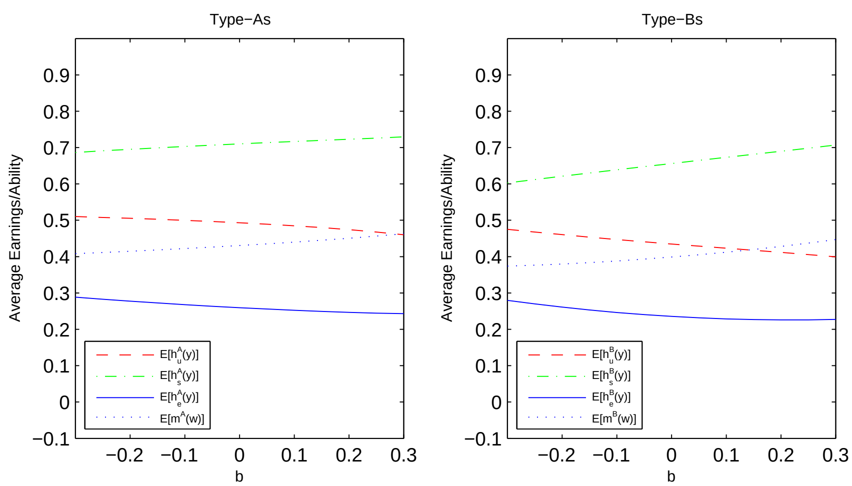


Figure 4.27: Changing b and Ability/Earnings by Type



4.H Cut-Offs and Heterogeneous Wage-Employment Productivity

As in the original model, occupational choice for each Type can be captured by three key equations. Equation (4.92) gives the wage schedule for an individual with ability y. We also require the values of self- and wage-employment relative to unemployment, which are given by $X_s^k(y)$ in Equation (4.93) and $X_e^k(y)$ in Equation (4.94) respectively. We use indicator functions to show when particular terms do and do not enter these equations.

$$w^k(y) = \gamma z(y)+(1-\gamma)rV_u^k(y) = \gamma(\beta_0+\beta_1 y)+(1-\gamma)\left[b+\mathbb{1}_{[X_s^k(y)>0]}\alpha X_s^k(y)+\mathbb{1}_{[X_s^k(y)>0]}m^k(\theta)X_e^k(y)\right] \tag{4.92}$$

$$(r+q_s+\mathbb{1}_{[X_s^k(y)>0]}\alpha)X_s^k(y) = y-b-\mathbb{1}_{[X_e^k(y)>0]}m^k(\theta)X_e^k(y) \tag{4.93}$$

$$(r+q_s+\mathbb{1}_{[X_e^k(y)>0]}m^k(\theta))X_e^k(y) = w^k(y)-b-\mathbb{1}_{[X_s^k(y)>0]}\alpha X_s^k(y) \tag{4.94}$$

Differentiating Equations (4.92), (4.93), and (4.94) gives:

$$\frac{\partial w^k(y)}{\partial y} = \gamma\beta_1+(1-\gamma)\left[\mathbb{1}_{[X_s^k(y)>0]}\alpha\frac{\partial X_s^k(y)}{\partial y}+\mathbb{1}_{[X_e^k(y)>0]}m^k(\theta)\frac{\partial X_e^k(y)}{\partial y}\right] \tag{4.95}$$

$$(r+q_s+\mathbb{1}_{[X_s^k(y)>0]}\alpha)\frac{\partial X_s^k(y)}{\partial y} = 1-\mathbb{1}_{[X_e^k(y)>0]}m^k(\theta)\frac{\partial X_e^k(y)}{\partial y} \tag{4.96}$$

$$(r+q_s+\mathbb{1}_{[X_e^k(y)>0]}m^k(\theta))\frac{\partial X_e^k(y)}{\partial y} = \frac{\partial w^k(y)}{\partial y}-\mathbb{1}_{[X_s^k(y)>0]}\alpha\frac{\partial X_s^k(y)}{\partial y} \tag{4.97}$$

We can then write an expression relating the slopes of $X_e^k(y)$ and $X_s^k(y)$.

$$\frac{r + q_e + \gamma \mathbb{1}_{[X_e^k(y)>0]} m^k(\theta)}{\gamma} \frac{\partial X_e^k(y)}{\partial y} = \beta_1 - \mathbb{1}_{[X_s^k(y)>0]} \alpha \frac{\partial X_s^k(y)}{\partial y} \tag{4.98}$$

Finally, we can solve for $\frac{\partial X_s^k(y)}{\partial y}$ to examine how β_1 impacts the paths of $X_e^k(y)$ and $X_s^k(y)$.

$$\frac{\partial X_s^k(y)}{\partial y} = \left(\frac{r + q_e + \gamma \mathbb{1}_{[X_e^k(y)>0]} m^k(\theta) - \gamma \beta_1}{r + q_e + \gamma \mathbb{1}_{[X_e^k(y)>0]} m^k(\theta)} \right) \left[r + q_s + \mathbb{1}_{[X_s^k(y)>0]} \alpha \left(\frac{r + q_e}{r + q_e + \gamma \mathbb{1}_{[X_e^k(y)>0]} m^k(\theta)} \right) \right]^{-1} \tag{4.99}$$

Thus, we can see that if $\beta_1 > \frac{r + q_e + \gamma \mathbb{1}_{[X_e^k(y)>0]} m^k(\theta)}{\gamma}$ then $\frac{\partial X_s^k(y)}{\partial y} < 0$ and $\frac{\partial X_s^k(y)}{\partial y} > 0$. This reverses the interpretation of the productivity cut-offs from the original model.

4.I Intra-Sector Empirical Analysis

Table 4.14: OLS Earnings Equations for the Self-Employed (Pooled Sample)

	(1) Women	(2) Men
Education (Years)	0.0203**	0.0365**
	(0.0086)	(0.0174)
Log of Age	-0.0137	0.0798
	(0.1443)	(0.2536)
Log of Tenure	0.0940***	0.0734
	(0.0250)	(0.0528)
Married? (1=Y, 0=N)	0.1998***	0.1190
	(0.0635)	(0.1325)
No. Children (Any Age)	-0.0087	0.0068
	(0.0189)	(0.0312)
Constant	2.4883***	2.5062***
	(0.4862)	(0.8224)
N	1841	655
R^2	0.061	0.051
Wave Fixed Effects	✓	✓
City Fixed Effects	✓	✓

Standard errors in parentheses
Dependent variable: Log of Earnings per Week
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4.28: Sex Earnings Gaps for the Self-Employed in Retail and Services

Chapter 5

Conclusion

5.1 Summary

In this thesis, we examine the factors which determine participation and performance in self-employment, focussing in particular on differences between women and men. We use multiple data sources from Ghana but, as we argue in Chapter 1, we believe our results are relevant in other contexts, especially in sub-Saharan Africa.

In Chapter 2 we show that the self-employed are able to adjust to short-run variation in domestic productivity, increasing their domestic work during periods of mild power outages to try and maintain consumption levels, but then substituting back towards market work when outages become more severe. We also find that women not only undertake far more domestic work than men, but that they are also less able to substitute towards market work when outages become severe. As such, Chapter 2 demonstrates that domestic tasks are truly an *obligation* for women more than they are for men. We rationalise the non-linear relationships observed in our data by simulating a simple model of time allocation, emphasising the importance of intra-household transfers of domestic goods and the elasticity of substitution between domestic and market goods.

In Chapter 3 we test the hypothesis that women choose to work in self-employment because the flexibility of these jobs allows them to balance income-generation with domestic tasks. We also consider what these patterns of occupational selection imply for average earnings gaps in different sectors of the economy. We find that greater domestic obligations — proxied by the dependency ratio of the household — increase the likelihood that women become self-employed, especially in low-input activities where they do not employ any labour besides their own. The analogous effects for men are far weaker. These differential selection forces imply that the potential earnings gap is *overestimated* in wage-employment but *underestimated* in self-employment, if we fail to control for occupational choice. Although some of the remaining gaps in potential earnings can be explained by human capital a large portion of the difference between women and men remains unexplained, even accounting for finer level sorting across industries.

In Chapter 4 we use theory to try and reconcile the observation that female participation in self-employment is so high, even though that is where potential earnings gaps are largest. In particular, we build a search-match model, which allows for both self- and wage-employment, where the latter is characterised by discrimination. Although firms are homogeneous ex ante, the matching process for women is characterised by extra frictions, which may be understood in terms of there being prejudiced 'interviewers' within each firm. Our model shows that these extra frictions can account for earnings gaps in all sectors of the economy, *without* assuming any differences in the underlying ability distributions of women and men. We argue that these mechanisms may be important not just for sorting between wage- and self-employment, but also across industries within broader sectors of the economy.

5.2 Policy

This thesis presents five related policy messages, which we wish to emphasise.

Firstly, differences in how women and men allocate their work hours are important, so focussing solely on earnings gaps misses a key component of sex inequality. The fact that women do more domestic work is well documented, but in our urban sample it also emerges that women do significantly more work *in total*, especially if they are self-employed. This is especially concerning for welfare, since women appear to be far more constrained in how they can adjust their time-use in response to exogenous variation in productivity.

Secondly, female workers would benefit from policies to decouple the flexibility of jobs from their monetary returns. In Chapter 3, our modelling approach assumes that extra job flexibility comes at cost of lower market returns, such that individuals with greater domestic obligations optimally choose occupations with lower wages. This mechanism is then borne out in the empirical work. If, however, formal sector wage jobs enable workers to balance childcare and other domestic work with market work more easily, the data suggest that females would participate more in these kinds of activities and less in low-input self-employment. From the analysis in this thesis and the previous work on Ghana outlined in Chapter 1, we know that small-scale informal sectors tend to have lower, more volatile returns. Thus, altering the parameters of the occupational selection model as described would have positive effects on female welfare.

Thirdly, earnings gaps between women and men appear to be larger in less formal sectors. In Chapter 3 we show that even after controlling for occupational selection, human capital, and industry, female own account self-employed workers earn around 50 percent less than their male equivalents. In wage-employment, by contrast, the small 3 percent earnings gap can be accounted for by pre-market disparity in human capital. As we show in Chapter 4, it is not only workers' occupation which determines their susceptibility to earnings gaps, but also their position in the ability distribution. More productive individuals, who have better outside options, are less sensitive to extra barriers to obtaining certain types of jobs. Thus, there is important heterogeneity across the economy in terms of where earnings gaps bite.

Fourthly, although sex inequality appears to be more severe in informal sector jobs, policies to address this may be better targeted at the formal sector. This emerges directly from Chapter 4, where discrimination occurs in wage-employment but a larger earnings gap emerges in self-employment. In a sense, this simply extends the idea that selection across occupations is as important, if not more important, than heterogeneity within occupations. Although extra frictions in obtaining wage jobs affect the composition of, and earnings in, the wage sector, the main impacts are on those individuals who miss out and end up becoming self-employed.

Finally, heterogeneity within self-employment matters. However, observable human capital, especially education, tends to explain very little of the earnings variation for the self-employed. Instead we emphasise three factors in this thesis, which appear to differ substantially by sex. First, entrepreneurs employ different *technologies*. In Chapter 3, we show that there are key differences between enterprises that do and do not use labour besides their own, not only in terms of earnings and non-wage characteristics in those jobs, but also the forces which guide selection into these activities. Second, although *industry* does little to help explain the overall earnings gap for the self-employed, there are some key asymmetries between the types of jobs undertaken by women, who tend towards retail, and men, who tend towards services. Third, on a more conceptual level, the self-employment sector results from a mix of *sorting* and *segmentation*. In our model in Chapter 4, some self-employed individuals truly have comparative advantage in their work and would choose to become entrepreneurs even if labour markets were fully competitive, whereas others would work in the wage sector were it not for search frictions. Since the predictions of the model match the data, we believe this is an accurate portrayal of the Ghanaian economy. As such, emphasising only the sorting framework and suggesting that intra-sector heterogeneity stems largely from differences in human capital, which appears to be the dominant view in the Latin American literature, is not appropriate for sub-Saharan Africa.

5.3 Future Directions

There are several ways in which new data and alternative modelling approaches could extend the research presented in the three main chapters of this thesis.

Firstly, the short-run variation in domestic productivity we exploited in Chapter 2, based on power outages, has intensified in Ghana since our data were collected in the summer of 2013. Indeed, the period during which our survey was implemented was an uncharacteristically stable period of electricity provision in the country. Now, not only are the lights out more often and for longer, but also the location and timing of the cuts are, to some extent, planned and publicised by the Ghanaian energy authorities. High-frequency data collected in these conditions could help to answer two key questions. Firstly, we could ascertain whether the results from Chapter 2 are externally valid. In particular, it would be useful to know if the coping strategies, which determine the adjustments in time allocation that we observe, are robust to more volatile provision of electricity. Secondly, since the outages are now being governed partially through 'load-shedding', whereby local radio and newspapers publish where and when power cuts are planned in advance, it would be possible to investigate how individuals gain information and update their subjective beliefs. This would enable us to unpack the anticipation effects and lagged responses we observe in our data.

Chapter 2 could also be extended with more detailed measurement of time-use. Our data do not allow us to examine how the particular components of domestic work, market work, and leisure adjust to power outages. This would substantially aid our understanding of the intricacies of how individuals allocate their time in response to exogenous shocks. Moreover, without full time diaries allowing for secondary and tertiary activities to be recorded, we cannot easily allow for multi-tasking. This is not solely a question of extra data, however, as there is an open *theoretical* question about how best to understand the time allocation problem, when certain activities can be undertaken simultaneously. We begin to tackle this issue with the simple model outlined in Chapter 3, but believe future research may provide a more generalisable framework.

In Chapter 3 we present a new technique for assessing endogeneity in multinomial choice problems. Although we believe our approach is a useful starting point, especially for the question we investigate, the logic of our method could be transported to other settings and made more flexible. In particular, it would, in principal, be possible to relax the IIA assumption and allow for different correlation structures between the unobservables associated with each occupational category and those associated with domestic obligations. However, even though the multinomial probit and other maximum simulated likelihood techniques can be estimated much faster as computers become more advanced, we still face something of a trade-off between the complexity and feasibility of potential estimators.

The analysis in Chapter 3 also provokes a conceptual question about how far discrete multinomial models can be pushed as a method for examining occupational selection. In our Ghanaian sample, it emerges that not only is employment sector important for understanding individuals' outcomes, but also industry, technology, and potentially other junctures. As the number of cross-cutting dimensions increases, and thus the number of 'bins' becomes large, occupation begins to look more like a continuous rather than a discrete variable. We could nest the set of choices that result in a particular kind of job, but this adds substantial structure to the decisions facing workers and indeed may not be enough if we require an especially fine level of granularity. For example, even amongst the own account retailers in Accra, it would be interesting to know why some choose to sell pure water and others choose to sell paintings, when demand for the former appears to be substantially higher and more stable. Whether these sort of problems are truly the domain of discrete multinomial methods is unresolved.

The treatment of discrimination in Chapter 4 yields several potential avenues for future work. Firstly, since it is exogenous in our model, finding some way of endogenising discrimination could shed light on the dynamic question of whether unequal outcomes between women and men are likely to persist or abate. For example, firms may be heterogeneous according to some innate level of prejudice, which in turn determines their chances of survival in the market. Alternatively, in the spirit of statistical discrimination models, firms may have incomplete information about potential workers, driving a wedge between the true ability distributions for women and men. Indeed, if the effects of discrimination spill over to benefit men, there may even be political economy reasons to suppose sex inequality would continue. Secondly, it may be useful to examine potential sources of discrimination within *self*-employment. Although entrepreneurs do not have to deal with employers, they must participate in both input and output markets, where frictions differentially affect women and men. These mechanisms may further drive the asymmetries we see in terms of industry and technology choice between female and male self-employed workers.

5.4 Final Remarks

Unless there are substantial changes to the current trends in employment and industry, self-employment will remain a vital source of jobs for people in Ghana, and sub-Saharan Africa at large, for many years to come. This thesis has attempted to explain why outcomes for women and men working in these sectors are so different. Since a large part of this story is about occupational selection, we believe labour market interventions need to do more than simply target the informal sector by, for example, trying to extend and improve micro-finance or providing capital directly to entrepreneurs. Instead, we argue that sex-specific barriers to entry into formal wage sector jobs — stemming from discrimination, domestic obligations, and access to human capital — must be eliminated if women and men are to fare equally in self-employment. Although this thesis emphasises the importance of these extra frictions facing female workers, the question over what specific interventions could remove them is unanswered. This presents a crucial topic for future research. However, the design and evaluation of policy programmes can only succeed with a more holistic understanding of labour market dynamics. We hope this thesis helps elucidate some of the opportunities and challenges associated with creating fair and efficient labour market outcomes in developing countries.

Bibliography

C. Ackah. Non-farm employment and incomes in rural Ghana. *Journal of International Development*, 2013.

C. Ackah, C. Ahiadeke, and A. Fenny. Determinants of female labour force participation in Ghana. Working paper, Global Development Network, 2009.

A. Adenikinju. Electric infrastructure failures in Nigeria: a survey-based analysis of the costs and adjustment responses. *Energy Policy*, 2003.

M. Aguiar, E. Hurst, and L. Karabarbounis. Recent developments in the economics of time use. *Annual Review of Economics*, 2012.

D. J. Aigner and G. G. Cain. Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 1977.

J. Albrecht, L. Navarro, and S. Vroman. The effects of labour market policies in an economy with an informal sector. *The Economic Journal*, 2009.

H. Alderman, P. A. Chiappori, L. Haddad, J. Hoddinott, and R. Kanbur. Unitary versus collective models of the household: Is it time to shift the burden of proof? *The World Bank Research Observer*, 1995.

J. G. Altonji and R. M. Blank. Race and gender in the labor market. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3. Elsevier, 1999.

J. G. Altonji, T. E. Elder, and C. R. Taber. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *The Journal of Political Economy*, 2005.

J. G. Altonji, T. E. Elder, and C. R. Taber. Using selection on observed variables to assess bias from unobservables when evaluating Swan-Ganz catheterization. *American Economic Review*, 2008.

J. D. Angrist and W. N. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 1998.

S. Appleton, J. Hoddinott, and P. Krishnan. The gender wage gap in three African countries. *Economic Development and Cultural Change*, 1999.

E. Aryeetey and W. Baah-Boateng. Growth, investment and employment in Ghana. Working paper, International Labour Office, 2007.

E. Aryeetey and A. McKay. Operationalising pro-poor growth: A country case study on Ghana. Operationalising pro-poor growth case study, World Bank, 2004.

D. H. Autor. Lecture note: The economics of discrimination. MIT Lecture, 2003.

W. Baah-Boateng, E. Nketiah-Amponsah, and R. Frempong. The effect of fertility and education on female labour force participation in Ghana. *Ghanaian Journal of Economics*, 2013.

G. S. Becker. *The Economics of Discrimination*. University of Chicago Press, 1957.

G. S. Becker. A theory of the allocation of time. *The Economic Journal*, 1965.

G. S. Becker. A theory of social interactions. *The Journal of Political Economy*, 1974.

G. S. Becker. *A Treatise on the Family*. Havard University Press, 1981.

G. S. Becker. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 1985.

M. Beenstock. Generators and the cost of electricity outages. *Energy Economics*, 1990.

S. Ben Yahmed. *Inégalité entre Hommes et Femmes sur le Marché du Travail, les Rôles du Commerce Interational et du Secteur Informel*. PhD thesis, Aix-Marseille Unversité, 2013.

B. Bental and A. Ravid. A simple method for evaluating the marginal cost of unsupplied electricity. *The Bell Journal of Economics*, 1982.

D. A. Black. Discrimination in an equilibrium search model. *Journal of Labor Economics*, 1995.

D. G. Blanchflower. Self-employment in OECD countries. *Labour Economics*, 2004.

D. G. Blanchflower and A. J. Oswald. What makes and entrepreneur? *Journal of Labor Economics*, 1998.

A. S. Blinder. Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 1973.

G. J. Borjas and S. G. Bronars. Consumer discrimination and self-employment. *The Journal of Political Economy*, 1989.

D. Borowczyk-Martins, J. Bradley, and L. Tarasonis. Assortative matching, search and discrimination. Working paper, University of Bristol, 2012.

F. Bourguignon, M. Fournier, and M. Gurgand. Selection bias corrections based on the multinomial logit model: Monte carlo comparisons. *Journal of Economic Surveys*, 2007.

A. J. Bowlus. A search interpretation of male-female wage differentials. *Journal of Labor Economics*, 1997.

S. G. Bronars and J. Grogger. The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review*, 1994.

C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 1965.

K. Burdett and D. T. Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, 1998.

C. Camerer, L. Babcock, G. Loewenstein, and R. Thaler. Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 1997.

A. Cameron, J. Gelbach, and D. Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 2011.

A. C. Cameron and P. K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005.

E. Cardia. Household technology: Was it the engine of liberation? 2008 Meeting Papers 826, Society for Economic Dynamics, 2008.

S. Caria and P. Falco. Not so happy with vulnerability: Panel evidence from urban Ghana on the effect of vulnerability on life satisfaction. Working paper, Centre for the Study of African Economies, 2011.

S. Caria and P. Falco. Do employers trust workers too little? An experimental study of trust in the labour market. Working paper, Centre for the Study of African Economies, 2013.

T. Cavalcanti and J. Tavares. Assessing the 'Engines of Liberation': Home appliances and female labor force participation. *The Review of Economics and Statistics*, 2008.

V. Chandra, J. Nganou, and C. Noel. Constraints to growth in Johannesburg's black informal sector: Evidence from the 1999 informal sector survey. Working paper, World Bank, 2002.

M. A. Chen. Rethinking the informal economy: Linkages with the formal economy and the formal regulatory environment. In B. Guha-Khasnobis, R. Kanbur, and E. Ostrom, editors, *Linking the Formal and Informal Economy: Concpets and Policies*. Oxford University Press, 2006.

P. A. Chiappori. Collective models of household behavior: The sharing rule approach. In L. Haddad, J. Hoddinott, and H. Alderman, editors, *Intrahousehold Resource Allocation in Developing Countries*, pages 39–52. The John Hopkins University Press, 1997.

P. L. Cichello, C. Almeleh, L. Mncube, and M. Oosthuizen. Perceived barriers to entry into self-employment in Khayelitsha, South Africa: Crime, risk, and start-up capital dominate profit concerns. Working paper, Centre for Social Science Research: Aids and Society Research Unit, 2011.

S. H. Clain. Gender differences in full-time self-employment. *Journal of Economics and Business*, 2000.

S. Coate and G. C. Loury. Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 1993.

D. Coen-Pirani, A. León, and S. Lugauer. The effect of household appliances on female labor force participation: Evidence from microdata. *Labour Economics*, 2010.

M. Connolly. Here comes the rain again: Weather and the intertemporal substitution of leisure. *Journal of Labor Economics*, 2008.

V. Crawford and J. Meng. New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review*, 2011.

W. V. Cunningham and W. F. Maloney. Heterogeneity among mexico's microenterprises: An application of factor and cluster analysis. *Economic Development and Cultural Change*, 2001.

S. Darko. Eight surprising consequences of Ghana's power outages. *BBC Africa News Website*, 2015.

E. Davies and A. Kerr. Firm survival and change in Ghana, 2003–2013. Working paper, Centre for the Study of African Economies, 2015.

A. de Janvry, M. Fafchamps, and E. Sadoulet. Peasant household behaviour with missing markets: Some paradoxes explained. *The Economic Journal*, 1991.

S. de Mel, D. McKenzie, and C. Woodruff. Are women more credit constrained? Experimental evidence on gender and microenterprise returns. *American Economic Journal: Applied Economics*, 2009.

S. de Mel, D. McKenzie, and C. Woodruff. Business training and female enterprise start-up, growth, and dynamics: Experiental evidence from Sri Lanka. *Journal of Development Economics*, 2014.

K. Deininger, J. Songqing, and H. Nagarajan. Wage discrimination in India's informal labor markets: Exploring the impact of caste and gender. *Review of Development Economics*, 2013.

F. Devoto, E. Duflo, P. Dupas, W. Parienté, and V. Pons. Happiness on tap: Piped water adoption in urban morocco. *American Economic Journal: Economic Policy*, 2012.

D. M. Drukker. Testing for serial correlation in linear panel-data models. *Stata Journal*, 2003.

J. A. Dubin and D. L. McFadden. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 1984.

C. W. Dunnett. Algorithm AS 251: Multivariate normal probability integrals with production correlation structure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1989.

P. Dupas and J. Robinson. Daily needs, income targets and labor supply: Evidence from Kenya. Working paper, National Bureau of Economic Research, 2013.

Z. Eckstein and K. I. Wolpin. Estimating the effect of racial discrimination on first job wage offers. *The Review of Economics and Statistics*, 1999.

D. S. Evans and L. S. Leighton. Some empirical aspects of entrepreneurship. *The American Economic Review*, 1989.

M. Fafchamps, D. McKenzie, S. Quinn, and C. Woodruff. Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana. *Journal of Development Economics*, 2014.

P. Fajnzylber, W. F. Maloney, and G. Rojas. Microenterprise dynamics in developing countries: How similar are they to those in the industrialized world? Evidence from Mexico. *World Bank Economic Review*, 2006.

P. Falco. Risk-aversion and occupational choices: Evidence from matched field experiments and survey data in urban Ghana. Working paper, Centre for the Study of African Economies, 2010.

P. Falco. Determinants of income in informal self-employment: New evidence from a long African panel. Working paper, Centre for the Study of African Economies, 2011.

P. Falco and S. Quinn. Identifying income vulnerability in urban Ghana: A dynamic binary model with maximal heterogeneity. Working paper, Centre for the Study of African Economies, 2012.

P. Falco and F. Teal. Unemployment, not working and jobs for the young in urban Africa: Evidence from Ghana. Working paper, Centre for the Study of African Economies, 2012.

P. Falco, A. Kerr, N. Rankin, J. Sandefur, and F. Teal. The returns to formality and informality in urban Africa. *Labour Economics*, 2011.

H. Farber. Is tomorrow another day? The labor supply of New York City cabdrivers. *Journal of Political Economy*, 2005.

G. S. Fields. Rural-urban migration, urban unemployment and underemployment and job search activitiy in LDCs. *Journal of Development Economics*, 1975.

G. S. Fields. Labour market modelling and the urban informal sector: Theory and evidence. In D. Turnham, B. Salomé, and A. Schwarz, editors, *The Informal Sector Revsitied*, pages 49–69. Organistation for Economic Co-operation and Development, 1990.

G. S. Fields. Labor market analysis for developing countries. *Labour Economics*, 2011.

G. S. Fields. Self-employment in the developing world: A report to the high-level panel of eminent persons. Background research paper, Cornell University and IZA, 2013.

L. Flabbi. Gender discrimination estimation in a search model with matching and bargaining. *International Economic Review*, 2010.

N. M. Fortin. The gender wage gap among young adults in the United States: The importance of money versus people. *Journal of Human Resources*, 2008.

N. M. Fortin, T. Lemieux, and S. Firpo. Decomposition methods in economics. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 4a, pages 1–102. Elsevier, 2011.

M. Fournier and M. Gurgand. Details of methods and formulas for Stata ado file selmlog.ado. Stata command description, Paris School of Economics, 2005.

R. G. Fryer Jr. Belief flipping in a dynamic model of statistical discrimination. *Journal of Public Economics*, 2007.

R. Geary. A note on 'a constant-utility index of the cost of living'. *Review of Economic Studies*, 1950.

P. Glick and D. E. Sahn. Gender and education impacts on employment and earnings in West Africa: Evidence from Guinea. *Economic Development and Cultural Change*, 1997.

M. Goldstein. Intrahousehold efficiency and individual insurance in Ghana. Development economics discussion paper, London School of Economics, 2004.

W. H. Greene. *Econometric Analysis: Fifth Edition.* Pearson Education, 2003.

J. Greenwood, A. Seshadri, and M. Yorukoglu. Engines of liberation. *The Review of Economic Studies*, 2005.

R. Gronau. Leisure, home production, and work – the theory of the allocation of time revisited. *The Journal of Political Economy*, 1977.

I. Günther and A. Launov. Informal employment in developing countries: Opportunity or last resort? *Journal of Development Economics*, 2012.

J. Harris and M. Todaro. Migration, unemployment, and development: A two sector analysis. *American Economic Review*, 1970.

K. Hart. Informal income opportunities and urban employment in Ghana. *The Journal of Modern African Studies*, 1973.

J. A. Hausman and D. L. McFadden. Specification tests for the multinomial logit model. *Econometrica*, 1984.

J. J. Heckman. Sample selection bias as a specification error. *Econometrica*, 1979.

J. Heintz. Employment, poverty, and gender in Ghana. Working paper, University of Massachusetts Amherst, 2007.

J. Heintz and L. Pickbourn. The determinants of selection into non-agricultural self-employment in Ghana. *The Journal of Applied Economic Research*, 2012.

J. Heintz and F. Slonimczyk. Beyond dualism: Multisegmented labor markets in Ghana. Working paper, University of Massachusetts Amherst, 2007.

S. Horton. *Women and Industrialization in Asia*. Routledge, 1996.

N. Ilahi and F. Grimard. Public infrastructure and private costs: Water supply and time allocation of women in rural Pakistan. *Economic Development and Cultural Change*, 2000.

B. Jann. The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 2008.

B. Jovanovic. Selection and the evolution of industry. *Econometrica*, 1982.

F. Kemausuor, G. Obeng, A. Brew-Hammond, and A. Duker. A review of trends, policies and plans for increasing energy access in Ghana. *Renewable and Sustainable Energy Reviews*, 2011.

A. Kerr. A model of comparative advantage with matching in the urban Tanzanian labour market. Working paper, Centre for the Study of African Economies, 2012.

B. Kőszegi and M. Rabin. A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 2006.

G. Kingdon and M. Söderbom. Education, skills, and labor market outcomes: Evidence from Ghana. Working paper, Centre for the Study of African Economies, 2007.

G. Kingdon, J. Sandefur, and F. Teal. Labour market flexibility, wages and incomes in Sub-Saharan Africa in the 1990s. *African Development Review*, 2006.

S. Kolavalli, E. Robinson, X. Diao, V. Alpuerto, R. Folledo, M. Slavova, G. Ngeleza, and F. Asante. Economic transformation in Ghana: Where will the path lead? Discussion paper, International Food Policy Research Institute, 2012.

G. Koolwal and D. van de Walle. Access to water, women's work, and child outcomes. *Economic Development and Cultural Change*, 2013.

L. F. Lee. Generalized econometric models with selectivity. *Econometrica*, 1983.

I. Lejárraga. Roaring tiger or purring pussycat: A growth diagnostics study of Ghana. Annual meeting paper, American Economic Association, 2010.

W. A. Lewis. Economic development with unlimited supplies of labour. *The Manchester School*, 1954.

K. V. Lombard. Female self-employment and demand for flexible, non-standard work schedules. *Economic Inquiry*, 2001.

J. S. Long and J. Freese. *Regression Models for Categorical Dependent Variables using Stata*. Stata Press, 2006.

R. E. Lucas. On the size distribution of business firms. *The Bell Journal of Economics*, 1978.

T. Magnac. Segmented or competitive labor markets. *Econometrica*, 1991.

W. F. Maloney. Informality revisited. *World Development*, 2004.

M. B. McElroy and M. J. Horney. Nash-bargained household decisions: Toward a generalization of the theory of demand. *International Economic Review*, 1981.

R. Meeks. Water works: The economic impact of water infrastructure. Discussion Paper 2012-35, Harvard Environmental Economics Program, 2012.

J. A. Mincer. Schooling, experience and earnings. Working paper, National Bureau of Economic Research, 1974.

M. J. Miranda and P. L. Fackler. *Applied Computational Economics and Finance*. MIT Press, 2002.

A. Moro and P. Norman. A general equilibrium model of statistical discrimination. *Journal of Economic Theory*, 2004.

D. T. Mortensen and C. A. Pissarides. Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, 1994.

D. T. Mortensen and C. A. Pissarides. New developments in models of search in the labor market. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3. Elsevier, 1999.

R. Narita. Self-employment in developing countries: a search-equilibrium approach. Working paper, World Bank, 2012.

D. Nkwetta, M. Smyth, V. van Thong, J. Driesen, and R. Belmans. Electricity supply, irregularities, and the propsect for solar energy and energy sustainability in sub-Saharan Africa. *Journal of Renewable and Sustainable Energy*, 2010.

N. Nsowah-Nuamah, F. Teal, and M. Awoonor-Williams. Jobs, skills and incomes in Ghana: How was poverty halved? Working paper, Centre for the Study of African Economies, 2010.

R. L. Oaxaca. Wage differentials in urban labor markets. *International Economic Review*, 1973.

R. L. Oaxaca and M. R. Ransom. On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 1994.

F. Obeng-Odoom. Neoliberalism and the urban economy in Ghana: Urban employment, inequality, and poverty. *Growth and Change*, 2012.

G. Oettinger. An empirical analysis of the daily labor supply of stadium vendors. *Journal of Political Economy*, 1999.

E. Oster. Unobservable selection and coefficient stability: Theory and evidence. Working paper, National Bureau of Economic Research, 2013.

H. Pasha, A. Ghaus, and S. Malik. The economic cost of power outages in the industrial sector of Pakistan. *Energy Economics*, 1989.

G. Psacharopoulos and Z. Tzannatos. *Women's Employment and Pay in Latin America: Overview and Methodology*. World Bank, 1992.

N. Rankin and G. Roberts. Youth unemployment, firm size and reservation wages in South Africa. *South African Journal of Economics*, 2011.

N. Rankin, J. Sandefur, and F. Teal. Learning and earning in Africa: Where are the returns to education high? Working paper, Centre for the Study of African Economies, 2010.

D. Read and G. Loewenstein. Temporal bracketing of choice: Discrepancies between simultaneous and sequential choice. Working paper, Carnegie Mellon University, Department of Social and Decision Sciences, 1996.

C. W. Reimers. Labor market discrimination against hispanic and black men. *The Review of Economics and Statistics*, 1983.

R. Reinikka and J. Svensson. Coping with poor public capital. *Journal of Development Economics*, 2002.

R. Rogerson, R. Shimer, and R. Wright. Search-theoretic models of the labor market: A survey. *Journal of Economic Literature*, 2005.

A. Rosén. Search, bargaining, and employer discrimination. *Journal of Labor Economics*, 2003.

P. R. Rosenbaum and D. B. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1983.

M. R. Rosenzweig and K. I. Wolpin. Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy*, 1980.

A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 1951.

M. Saatchi and J. Temple. Labor markets and productivity in developing countries. *Review of Economic Dynamics*, 2009.

H. A. Sackey. Female labour force participation in Ghana: The effects of education. Research paper, African Economic Research Consortium, 2005.

P. Samuelson. Social indifference curves. *Quarterly Journal of Economics*, 1956.

J. Sandefur. On the evolution of the firm size distribution in an African economy. Working paper, Centre for the Study of African Economies, 2010.

J. Sandefur, P. Serneels, and F. Teal. Poverty and earnings mobility in three African countries. In P. Paci and P. Serneels, editors, *Employment and Shared Growth: Rethinking the Role of Labor Mobility for Development*. World Bank, 2006.

M. Sasaki. An equilibrium search model with coworker discrimination. *Journal of Labor Economics*, 1999.

A. Sen. Employment, institutions and technology: Some policy issues. *International Labour Review*, 1975.

A. Shangvi. Economic costs of electricity supply interruptions: US and foreign experience. *Energy Economics*, 1982.

C. Shapiro and J. E. Stiglitz. Equilibrium unemployment as a worker discipline. *The American Economic Review*, 1984.

M. Söderbom, F. Teal, and A. Wambugu. Unobserved heterogeneity and the relation between earnings and firm size: Evidence from two developing countries. *Economics Letters*, 2004.

J. Steinbuks and V. Foster. When do firms generate? Evidence on in-house electricity supply in Africa. *Energy Economics*, 2010.

M. Stevens. Labour economics I: Competitive and imperfectively competitive labour markets. Oxford University Lecture, 2012.

J. Stone. Linear expenditure systems and demand analysis: An application to the pattern of British demand. *Economic Journal*, 1954.

A. Tansel. Wage earners, self employed and gender in the informal sector in turkey. Working paper, Economic Research Forum, 2001.

F. Teal. The size and sources of economic rents in a developing country manufacturing labour market. *The Economic Journal*, 1996.

F. Teal. The Ghana and Tanzania Urban Household Panel Surveys: 2004–2006. Online note, Centre for the Study of African Economies, 2008.

F. Teal. Labour and employment in Africa. In E. Aryeetey, S. Devarajan, R. Kanbur, and L. Kasekende, editors, *The Oxford Companion to the Economics of Africa*. Oxford University Press, 2012.

D. Thomas. Intra-household resource allocation: An inferential approach. *Journal of Human Resources*, 1990.

K. Train. *Discrete Choice Methods with Simulation: Second Edition*. Cambridge University Press, 2009.

S. Verick. Female labor force participation in developing countries. IZA world of labor article, IZA, 2014.

M. W. Warner, R. M. Al-Hassan, and J. G. Kydd. Beyond gender roles? Conceptualizing the social and economic lives of rural peoples in sub-Saharan Africa. *Development and Change*, 1997.

A. J. Wellington. Self-employment: the new solution for balancing family and career? *Labour Economics*, 2006.

J. Williams and R. Ghanadan. Electricity reform in developing and transition countries: A reappraisal. *Energy*, 2006.

C. Woodruff. Self-employment: Engine of growth or self-help safety net? In P. Paci and P. Serneels, editors, *Employment and Shared Growth: Rethinking the Role of Labor Mobility for Development*. World Bank, 2006.

J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

World Bank. World Development Indicators. http://data.worldbank.org/, 2015.