A Model of Comparative Advantage with Matching in the Urban Tanzanian Labour Market

Andrew Kerr*

September 12, 2011

1 Introduction

African labour market analysis has often emphasised average wage differentials across sectors. A high-paying "protected" or formal sector that is usually assumed to comprise unionised jobs in large firms and public sector employment is assumed to exist together with a low paying "sink", informal or "murky" sector comprised of self-employment and wage employment in unregistered small firms. A labour market with these characteristics has been described as "segmented", and much effort has been exerted in testing whether labour markets in a variety of countries are indeed segmented. In this chapter I suggest an alternative approach to modeling the Tanzanian labour market and focus on the distinction between wage and self-employment and the role of comparative advantage and individual heterogeneity in determining where to work and earnings in each of these sectors.

In this chapter I argue that describing the Tanzanian labour market as segmented does not adequately capture some important empirical features of this market. In Tanzania there are very large overlaps in the earnings distributions for jobs traditionally considered to be in the formal and informal sectors. There is also a large variance in the distribution of earnings *within* what has traditionally been considered the informal sector. These facts require an explanation not to be found in segmented labour market theories.

Economic theory suggests that a focus on the differences between self and wage employment may be important for understanding the workings of the labour market in low income countries like Tanzania with large amounts of self-employment. In the recent search theory literature both ex-ante wage posting and ex-post wage bargaining models suggest workers may earn less than their marginal product due to the monopsony power of firms. Self-employment, however, which is very common in urban Tanzania, does not involve interaction with a firm and thus the self-employed can capture all the returns to their labour

*Centre for the Study of African Economies, Department of Economics and St Antony's College, University of Oxford (DPhil Candidate) and DataFirst, School of Economics, University of Cape Town: andrew.kerr@uct.ac.za

and the return to any capital they have accumulated. This is an important distinction which has not been captured in models of African labour markets to date.

I use these theoretical and empirical insights to develop a matching model of the urban labour market in Tanzania. In this model the productivity of some individuals in wage employment is low relative to their self-employment productivity. Variation in self-employment productivity generates an opportunity cost of wage employment that is increasing in self-employment productivity, thus incorporating insights about comparative advantage in different sectors that date back to Roy (1951).

Section 2 reviews the literature on African labour markets, other developing countries and the recent literature on the role of sorting, comparative advantage and heterogeneity, which has been explored in the developed country context as well as in Latin America. Section 3 describes the data I use in this paper. Section 4 explores the data and how they speak to different models of the Tanzanian labour market. In section 5 I develop a model which captures some of the features presented in the empirical exploration. Section 6 explores how the numerical solution to the model matches the empirical evidence and section 7 concludes.

2 Modeling African Labour Markets

Much of the recent work on African labour markets attempts to provide evidence on whether or not labour markets are segmented, inspired by Fields's (1975) extension of the Harris and Todaro (1970) model of migration. In this paper I use the insight that individual heterogeneity is an important determinant of labour market outcomes to model the effects of comparative advantage and sorting in the Tanzanian labour market, in the tradition of Roy (1951). In this section I review the debate about wage differentials that the Harris and Todaro (1970) model has generated as well as describing alternative models of comparative advantage and sorting.

2.1 Segmented Labour Markets

The Harris and Todaro (1970) model and the extension by Fields (1975) have provided the basis for much of the empirical research on earnings determination in labour markets in Africa, Latin America and in developing countries in other regions over the last 40 years.

Heintz and Posel (2008) and Badaoui et al. (2008) are two recent examples of papers testing for segmentation in South Africa that are representative of part of the debate that has occurred in the literature. Heintz and Posel (2008) use cross-sectional evidence from South Africa and find large wage differentials between the formal and informal sectors, using a definition of informality suggested by the ILO that includes workers in informal enterprises as well as workers in informal jobs working for formal firms. This leads them to conclude that the labour market is segmented between the formal and informal sectors[1]. Badaoui et al. (2008), however, come to the opposite conclusion using similar data from South Africa and the same definition of informality, though their analysis is limited to men only, in order that their results are not influenced by female labour force participation decisions. The authors also exclude the self-employed and those in the public sector to focus on the effects of informality on employees of profit maximising firms. They find that, although there is a difference in average earnings in the formal sector and informal sector when using the cross-section and controlling for observable human capital, there is no difference once they take into account tax payments in the formal sector, as well as unobserved heterogeneity, using the panel element of the data.

2.2 Moving away from the representative worker

In 2000 James Heckman was jointly awarded the Nobel Prize in economics with Daniel McFadden, the first time the award had been given to individuals working in microeconometrics. In the conclusion to his Nobel Prize lecture Heckman explained the rapid evolution of microeconometrics in noting that "[i]n the past half century, economics has been enriched by vast new resources of microeconomic data. The data have opened the eyes of economists to the diversity and heterogeneity of economic life" (Heckman 2001, pg. 734). The exploration of the implications of this heterogeneity for economic theory, econometrics, and analysis of public policy continues and much of Heckman's Nobel lecture was devoted to summarising the frontier of this research.

The importance of heterogeneity in economic life has stimulated research in many areas of economics. In econometrics the explosion of the treatment effects literature and the new tools developed to assess these effects has been a response to the heterogeneity of responses to treatment that plagues estimation of average treatment effects. The heterogeneity in individuals' choices of whether to work and how much to work motivated Heckman's early econometric work on female labour supply (Heckman 1974). The low explanatory power of human capital regressions is suggestive of the importance of unobserved individual heterogeneity. This observation has also led researchers to develop models of the labour market which can generate wage differentials even with identical workers (Mortensen 2003), as I explore in more detail in the next section.

The focus on heterogeneity has also influenced the debate on segmentation in developing countries in several ways. As I discussed above, the importance of unobserved individual heterogeneity has been offered as an alternative explanation of sectoral earnings differentials to segmentation, partly as a result of the emergence of panel data in developing countries.

Bill Maloney has taken a slightly different tack, suggesting that the utility of different jobs is affected by the unobserved characteristics of different types of employment. Maloney (1999) argues that researchers should take account of,

[1] They also suggest there are different types of informal employment, some of which have higher barriers to entry, which means that *within* the informal sector there may be different segments paying different wages on average.

or at least be aware of these unobserved characteristics, but that in practice they have mostly failed to do this in studies arguing for or against segmentation on the basis of earnings differentials. Maloney's conclusion is that "traditional earnings differentials cannot prove or disprove segmentation in the developing-country context" (Maloney 1999, pg. 275). Maloney has also suggested an alternative method of testing for segmentation. He argues that in a segmented labour market, movement between the formal and informal sectors would be limited. If formal sector jobs could be held for life (barring any bad conduct), as has been argued to be the case in some Latin American countries, then there should be movement from the informal sector to the formal sector but very little in the opposite direction.

Instead of thinking about a segmented labour market, Maloney (2004) suggests that self-employment in the informal sector represents a viable alternative to salaried work in the formal sector. This is as a result of the low firm productivity in many developing countries and the benefit of being one's own boss. Maloney (2004) thus argues that workers' sector choices may represent the best outcomes possible given their low education levels. This suggests a model of comparative advantage, similar in spirit to Roy (1951), in which workers choose between salaried employment in a low productivity firm and self-employment. I use this as the basis for my model of the Tanzanian labour market.

As a result of the importance of the returns to education within labour economics, and as a key parameter for policy makers, there is a body of literature attempting to estimate the returns to education in African economies. Very little of this work, however, acknowledges any of the difficulties that individual heterogeneity creates in estimating an accurate return to education. In a review of previous studies of the estimates of the return to education in South Africa, for example, Keswell and Poswell (2004) do not include any studies that use instrumental variables as a result of education being endogenous and only one of the studies they review attempted to deal with ability bias (Moll 1998) and then only by including a measure of cognitive ability in the earnings function.

There has been some research emphasising the role of individual heterogeneity in determining individual earnings in African economies. Sandefur et al. (2007) show that the earnings distribution *within* different sectors has a large variance and also that there is a substantial overlap of the sectoral earnings distributions for Ghana, Tanzania and Ethiopia. Falco et al. (2010) have subsequently shown that this overlap still holds even when controlling for observable human capital. This analysis uses the Urban Panel Surveys for Ghana and Tanzania, though this analysis is still in the context of exploring sectoral differentials. There is also some research using manufacturing firm and worker data that attempts to explore the role of individual, as well as firm, heterogeneity in wage determination. Fafchamps et al. (2009) use data on matched manufacturing firms and workers from 11 African countries to explore whether sorting across firms and occupations can explain the return to education found in earnings regressions. Firm and occupational controls lower the estimated returns to education leading the authors to conclude that sorting occurs; through individuals with higher education levels both matching with more productive firms and being in more productive occupations.

Despite the existence of some empirical research on the role of heterogeneity I have not come across a single theoretical model of African labour markets that takes into account heterogeneity and comparative advantage. This paper attempts to address this lacuna and the following section explores the class of models I use to do this.

2.3 Equilibrium Search Models

In addition to the acknowledgement of the "diversity of economic life", another outcome of the collection of large amounts of micro data was the realisation that the frictionless competitive model of the labour market could not explain many of the features these data suggested were important (Mortensen and Pissarides 1999). This led to the emergence of equilibrium search and matching models that incorporated labour market frictions and the necessity of search by both firms and workers. Important contributions to this literature have come from Peter Diamond, Dale Mortensen and Christopher Pissarides, who were jointly awarded the 2010 Nobel Prize in economics for their work.

Mortensen and Pissarides (1999) have noted there are two main areas of theoretical research within the equilibrium search framework that have traditionally been employed to explain different sets of stylised facts. Matching models with search and recruiting frictions have been used to explore the flows of workers between employment, unemployment and non-participation as well as job creation and destruction flows (Mortensen and Pissarides 1994). The second main area is the "wage posting" literature, which has been used to explain the distribution of wages in an economy as the outcome of a strategic game where firms post wages in a labour market with search frictions (Burdett and Mortensen 1998)[2].

In both matching and wage posting models, the labour market is assumed to be characterised by search frictions for both workers and firms. Thus information about profitable opportunities is not instantly available to all agents in the economy, unlike in the frictionless competitive model. It takes time, therefore, for workers to find firms with vacant jobs and for firms to fill these jobs. In matching models the rate at which firms and workers meet is modeled using a matching function, an aggregate representation of how many workers and firms match at any point in time (Cahuc and Zylberberg 2004). In matching models, search frictions imply that firms have monopsony power over the workers they meet, and can exploit this to pay workers less than their marginal product. Firms and workers are assumed to bargain over the wage based on the value of the match created by the firm and worker, where the value may vary across matches if firms or workers are heterogeneous.

Recent research has used the equilibrium search framework in a number of ways to explain features of developing country labour markets. Satchi and Temple (2009) develop a macro matching model of the Mexican labour market

[2]Equilibrium search models with wage posting can generate different wages paid to identical individuals (Mortensen 2003). Interestingly, Fields (2004) sees this reflecting the influence of the earlier segmentation literature.

where individuals choose employment in either the rural, urban formal or urban informal sectors. The model allows for the presence of frictions in the formal sector and assumes their absence in the informal sector, which helps to generate a realistically sized informal sector. They also assume, however, that the level of informal sector productivity is similar to the formal sector wage, which they argue is backed up microeconometric studies using Mexican survey data. This assumption has the effect of making the frictionless informal sector more attractive and thus also larger. Because theirs is a macro model, Satchi and Temple (2009) focus on average wage differentials between sectors, as well as the relative size of the sectors. Individuals, however, are still homogeneous and as a result wages are also identical within sectors, implying that the model does not allow for any of the heterogeneity Heckman (2001) has noted is so important. Thus whilst it provides an explanation for the size of the informal sector, it cannot help to explain the distribution of earnings.

Albrecht et al. (2009) build a two-sided search model of the Mexican labour market based on the Mortensen and Pissarides (1994) matching model but allow for two sectors: the formal and informal sectors. This model does help to explain the observed earnings distribution. Workers differ in their formal sector productivity, which generates a distribution of formal sector earnings, however, informal sector productivity and earnings are assumed to be the same for all individuals. Informal sector productivity is calibrated to be 70 percent of the average level of formal sector productivity, with 65 percent of the population assumed to be more productive in the formal sector. The model thus sets up variation in the opportunity cost of informal employment. Some individuals have a low opportunity cost of informal sector employment as a result of their low productivity in the formal sector. These individuals take up informal employment whilst more productive individuals only accept formal employment opportunities. As a result earnings are higher in the formal sector, which contradicts the assertions of Satchi and Temple (2009), mentioned above, that earnings are roughly similar in the two sectors.

The assumption of wage bargaining in the formal sector means individuals are not paid their marginal product. The model incorporates ideas from Roy (1951) about comparative advantage in the labour market, but with ability varying in only one of the sectors. I use this as a basis for the model I build of the Tanzanian labour market but address some of its weaknesses, including the lack of a distribution of wages in one of the sectors.

2.4 Differing Conceptions of the Source and Characteristics of the Informal Sector

The divergent models of segmentation and comparative advantage in developing countries have very different conceptions of the source and characteristics of the informal sector. As regards characteristics, in Fields's (1975) model the informal sector offers both low wages and hence low utility to individuals who are unable to obtain a formal sector job. Wages are set by bargaining with unions or by non-profit maximising public sector firms in the formal sector. No individuals would prefer the low paying informal sector in these models. By contrast models of comparative advantage assume that there are different abilities required in each sector and that some individuals are better at work in one sector than others. This means that most people prefer to work in either one sector or the other.

In Satchi and Temple (2009) the informal sector and the formal sector also differ in the levels of frictions. A justification for the assumed difference in sectoral frictions is not provided by the authors. Zenou (2008) argues that the informal sector is mostly comprised of self-employment where there can be no matching with a firm and informal work for friends or relatives, in which case co-ordination failures are much less likely to occur. He concludes that frictions are likely to be much lower in the informal sector and that this can be approximated by assuming there are no frictions in the informal sector.

In the Albrecht et al. (2009) model, the formal sector differs from the informal sector because there is wage bargaining in the formal sector, whereas informal earnings are simply the common informal sector productivity. In my model I take these distinctions to their logical conclusion by calling the two sectors wage and self-employment, with workers to choose between them based on their comparative advantage. Clearly, earnings determination differs between these two kinds of employment and the levels of frictions may also differ. In the following section I show that a distinction between wage and self-employment helps to make sense of the empirical evidence from Tanzania.

3 Data Sources

I use two sources of data from Tanzania in this paper. The first is the nationally representative 2001 and 2006 Integrated Labour Force Surveys (ILFS) collected by the Tanzanian National Bureau of Statistics (NBS). I also make use of the 2004 and 2005 rounds of the Tanzanian Urban Panel Survey (TUPS) to explore individual movement between different types of employment. The ILFSs collected detailed individual member and household information from households in both urban and rural Tanzania. Population weights were calculated by NBS for each of the surveys but since the survey weights for 2001 have not been released publicly by NBS I have not weighted any of these data.

The TUPS data were collected by members of the Centre for the Study of African Economies (CSAE) at the University of Oxford. They were collected in the main urban centre of Tanzania and provide a small but representative sample of workers from urban Tanzania. The 2004 round was an urban sub-sample of those individuals surveyed by NBS for the 2000-2001 Household Budget Survey (HBS) and the 2005 round attempted to re-interview those interviewed a year before.

The Tanzanian UPS (along with a similar survey undertaken by CSAE in Ghana) provide one of the few individual panel surveys in Africa. Information was collected on both wage and self-employment, and thus the TUPS provides evidence on individual movement between wage and self-employment as well as between large and small firms and the public sector. The survey did not track those who had moved or could not be found, resulting in a 25 percent rate of attrition between 2004 and 2005, and implying the sample may not be representative of the 2005 urban population. I do not explore the effects of attrition in this paper.

3.1 Measurement error in ILFS earnings data

Self-employment income has been considered difficult to measure and is thus often not collected in household surveys (Deaton 1997). Some recent work using the TUPS has, however, used self-employment income to explore the effects of experience, education and sector of employment on earnings (Rankin et al. 2010). In this paper I also use of self-employment earnings data in Tanzania from the ILFSs and the TUPS.

In the Tanzanian ILFSs each survey first asked about gross income and the expenses incurred to earn it in the period in question (either a week or a month) for those engaged in self-employment. The surveys then asked about net income for the business.

Subtracting expenses from gross income produces a net income value in 2001 identical to the net income question asked directly, suggesting that the Tanzanian National Bureau of Statistics (NBS) forced the equivalence of these two measures in its post-survey cleaning. In 2006, however, the direct question on net income yields radically different answers from those calculated by taking gross income and subtracting expenses. The direct income question yields self-employment incomes that are on average six times higher than the measure using gross income and subtracting expenses.

The data from this income question also have some other problematic features, such as the amounts reported for weekly earnings being higher on average than those for monthly earnings, even before adjustment to a common period. In addition, when comparing the 2001 earnings data with inflation adjusted earnings in 2006 the direct income question in 2006 yields average increases in self-employment income over 5 years that are not matched by increases for employees in either the private or public sector. Further investigation revealed that NBS had again forced the equivalence of net income and gross income minus expenses by altering the net income measure but without taking into account whether individuals reported the net income or expenses figures as daily, weekly or monthly. This meant that if individuals reported weekly net income but monthly gross income and costs, which the survey instrument allowed, NBS forced the weekly net income to equal the monthly gross income. This means that when I convert net weekly income to a monthly figure I obtain an income figure much higher than actual income. As a result I use gross income minus expenses as the measure of self-employment income in both surveys.

Inflation makes comparison between the 2001 and 2006 surveys more difficult. A consumer price index (CPI) is collected monthly by the Tanzanian National Bureau of Statistics but there have been concerns raised about the accuracy of this index. I discuss deflating incomes in section 8.A in the Appendix to this paper.

4 Evidence for distinguishing between Models of the Tanzanian Labour Market

In this section I present empirical evidence from the Tanzanian labour market in an attempt to discriminate between competing ways of modelling this market. It is possible to think about the labour market as segmented, using Fields's (1975) update of the Harris and Todaro (1970) model, or to emphasise a comparative advantage model as first suggested by Roy (1951) and recently explored by Albrecht et al. (2009) in a developing country context.

4.1 Employment

I begin by exploring the distribution of employment in urban Tanzania in Table 1, using the ILFS data for 2001 and 2006. Table 1 shows that self-employment is the most common type of employment in urban Tanzania. Over 80 percent of the self-employed are own account workers with the remainder mostly employing a small number of employees: in 2001, conditional on having paid employees, 85 percent of the self-employed had 5 or fewer employees. Employment in the public sector is common and constituted 60 percent of urban wage employment in 2001, which is consistent with a picture of a segmented labour market in which there is a substantial amount of employment in firms that pay wages that are not profit maximising (though there is evidence that public sector employment declined relative to the size of most other kinds of employment between 2001 and 2006). Table 1 also shows that agriculture is still fairly important even within urban Tanzania.

The data show a substantial increase in the proportion of large firm employees between 2001 and 2006. Part of this may be explained by the change in the way firm size was captured in the two surveys. The 2006 ILFS survey only asked individuals whether the firm in which they worked had less than 10 employees or 10 or more. Thus the measure of firm size used is a crude one and I must assume any firm with 10 or more employees is a "large" firm. The 2001 ILFS asked if an individual worked in a firm with 1-5 employees, 5-10 and more than 10. I treat large firms in 2001 as those with more than 10 employees. Thus the definitions in 2001 and 2006 are not exactly comparable. This by itself does not seem to be a sufficient explanation for the jump in the number of individuals reported to be working for large firms between 2001 and 2006, however. It might be partly explained by different weightings used in each of the surveys but since I have been unable to access the 2001 population weights I cannot confirm this. My analysis suggests there are some issues of comparability between the surveys but in the absence of detailed reports on these surveys and how they were conducted, as well as surveys weights, it is difficult to explore this in any further detail.

Table 1 also shows the extent to which those who have obtained high levels of education obtain public sector employment. Over half of those who are working and have completed O levels (11 years of education under the new education system, 12 under the old) are working in the public sector. Those unable to obtain O levels seem to be disadvantaged as regards obtaining public sector employment, which I show in the next section is well paid relative to other forms of employment in Tanzania.

4.2 Earnings

The main focus in testing for segmentation has been on whether earnings differentials exists between sectors. Table 2 shows that median earnings in the ILFS are low for those in self-employment with no employees and in small firms, and are higher in large firms and the public sector, which seems consistent with a segmented labour market hypothesis. Wage employees were asked about their gross income in both ILFSs. It is not clear what impact the non-payment of taxes would have on sectoral income differentials. It is likely, though, that public sector workers paid tax but that small firm employees and the self-employed did not. Median earnings are also much higher for entrepreneurs with employees, who own what seem to be mostly informal businesses, with one third of these employers having written accounts and only 15 percent having accessed credit in the last year. Table 2 also shows low or negative earnings growth across the sectors between 2001 and 2006. This is surprising because this occurred at a time when Tanzania recorded real GDP growth of around 6 percent per year. It is consistent, however, with the Household Budget Survey analytical report (National Bureau of Statistics, Tanzania 2007), which suggests there was no decline in poverty between 2000 and 2007 as recorded in the HBSs in these years.

Differences in median earnings tell only part of story however. Figure 1 shows the real earnings densities in wage employment for the same categories as in Table 2 from the pooled 2001 and 2006 ILFS data. There is clearly much more variance in earnings in self-employment compared to wage employment[3]. Despite having similar medians, the density of earnings for the self-employed without employees has more mass in the upper tail of the density, compared to the density for employees in small firms, which has much more mass in the centre and roughly the same in the lower tail.

There are also large overlaps between what are generally considered formal and informal sector jobs. Around a third of employees in firms that have more than 10 employees earn less than the median in self-employment for those without employees. The differences in median earnings suggests that a focus on average earnings differentials is important but the large variation within sectors suggests a model of earnings determination should also account for this heterogeneity in earnings.

[3]This may partly be explained by measurement error, since the self-employed are asked about and must therefore calculate their profits, whereas wage employees simply report their earnings, as explained above.

One obvious explanation for earnings variation is that there are differences in observable human capital. I explore this possibility in Figure 2 which plots the density of the unexplained part of earnings after a regression of log earnings that includes basic human capital controls (age, education, tenure), as well as gender, which is reported in Table 9 in the appendix to this paper. Figure 2 is similar to Figure 1, although the variance of earnings in each sector is lower. This similarity is perhaps suprising, given that education seemed important in obtaining access to public sector employment, but makes more sense when considering that even in the public sector, human capital can only explain about 25 percent of the variance in earnings (see Table 9 in the Appendix to this paper). This analysis suggests that the substantial overlap between wage and self-employment is not due to differences in observable human capital.

Figure 3 makes the distinction between wage and self-employment in graphing the earnings densities but includes public employment separately. There is clearly a large premium for those in the public sector, and I have shown above that this seems difficult to obtain without at least having obtained O levels. Within the private sector, however, there is not much difference between median earnings in self and wage employment, though there is substantially higher variance in self-employment, as I noted above.

A final piece of evidence is shown in Table 3, which gives earnings regressions with detailed sectoral controls, and controls for firm and own enterprise size using the TUPS. The TUPS is much smaller than the ILFS but allows a more substantial set of controls than the ILFS. As shown in Table 9 in the Appendix to this paper using the ILFS data, and as discussed above, education, experience and gender are all important determinants of earnings. Table 3 also shows the large premia for public sector earnings and that those in larger firms earn more. Probably most interesting is the fact that individuals in the smallest firms actually earn less on average than those in self-employment. I return to this point when discussing the results of my model.

4.3 Employment Transitions

In the segmentation literature, following the tradition of the Harris and Todaro (1970) model, individuals are assumed not to move from the formal sector to the informal sector, usually as a result of labour legislation or union strength that prevent workers being fired. Maloney (1999) shows that in Mexico, transitions between the formal sector and the informal sector occur in both directions, thus questioning whether the Mexican labour market is segmented.

Table 4 uses longitudinal data from the 2004 and 2005 rounds of the Tanzanian Urban Panel Survey to explore how individuals move between different types of employment. Two things stand out from this table. The first is that there does seem to be little movement from the formal to informal sectors, with public sector workers and workers from larger firms very unlikely to move into self-employment, as the segmented labour market hypothesis would suggest. The second is that there is also very little movement between self-employment and wage employment, though there is movement within these categories between self-employment with, and without, employees, as well as some between small and large firm wage employment. This suggests that comparative advantage in wage and self-employment may play an important role in decisions about which sector to work in. It is possible that individuals are choosing to work either for themselves or for a firm, and that this may depend on their comparative advantage in each of these options.

4.4 Summary of empirical results

I have noted that some of the of the evidence presented above suggests that the urban Tanzanian labour market fits broadly into the Fields (1975) view. Wages are high in the formal sector, particularly the public sector, and are low in an informal sector comprised of wage employment in smaller firms and the self-employed. There is also almost no movement from employment in the public sector into self-employment or small firms. I have also shown, however, that there is a large diversity in earnings within self-employment and private wage employment and that those owning informal firms who employ others have relatively high earnings, comparable with those earning by employees in larger firms and the public sector. In addition, wage employees in small firms earn less on average than the self-employed without employees. The Tanzanian UPS transitions data also suggest that there is very little movement in either direction between wage employment (in small or larger firms or the public sector) and self-employment.

I have also noted that there are obvious differences between wage and self-employment in the way earnings are determined in each. The empirical evidence above, as well as the conceptual differences in earnings determination, suggest that a more fruitful way of modelling the labour market may be to assume that individuals choose between wage and self-employment based on their comparative advantage, which I undertake in the following section.

5 A Model of the Tanzanian Labour Market

5.1 Model Outline

Much of the labour economics literature on developing countries divides the labour market into a formal and informal sector where earnings are higher in the formal sector. Some of the evidence I showed in the previous section, however, does not support this view. As a result of this evidence, as well as intrinsic differences in earnings determination for wage and self-employment, I build a two-sector model of comparative advantage in the Tanzanian labour market that also captures the difference in the way earnings are determined in wage and self-employment. The empirical analysis above suggested the public sector provides a large part of the incentive to acquire advanced secondary or tertiary education by rewarding those who acquire jobs in it with high earnings. I do not include it in my model, however, and instead focus on the choice between private wage employment and self-employment. The model I build generates outcomes that are broadly consistent with the empirical evidence from Tanzania.

I use a matching framework in a Roy-type model (Roy 1951) of comparative advantage, in which individuals choose either wage or self-employment opportunities based on their expected earnings in each sector. Frictions limit the rate at which workers and firms meet each other. Frictions are also modeled in finding self-employment, so that workers do not instantaneously find self-employment opportunities. Thus I do not follow Zenou (2008) or Satchi and Temple (2009) in assuming that one of the sectors has no frictions, but this does not affect my main conclusions. Individuals earn their marginal product in self-employment, though wage employees do not as they are assumed to bargain with the firms they meet.

Homogeneity in ability would mean earnings would only differ across sectors but not within sectors. As a result of the variation in earnings within sectors that I highlighted above, I assume that individuals vary in self-employment productivity but that there is constant productivity across all firms and workers in wage employment. Despite there being variation only in self-employment productivity, my model also generates earnings variation in the bargained wage, as a result of the variation in workers' outside options due to their differences in self-employment productivity.

This incorporates ideas from Roy (1951), who assumes individuals choose to work in the sector that maximises their income and that this income ability differs in each sector, as well as including ideas from Maloney (1999) and Albrecht et al. (2009) who argue that wage employment has a low opportunity cost for individuals with low self-employment productivity and that self-employment becomes a relevant alternative when firm productivity is low.

As a result of the differences in self-employment productivity, individuals with low ability take only wage employment, some individuals take both, and some take only self-employment. All experience unemployment when their jobs are destroyed, and they then go back to searching for either wage opportunities or self-employment opportunities or both[4].

A more complex model might perhaps include variation in firm productivity (Postel-Vinay and Robin 2002), ability differences in both sectors (Roy 1951) or productivity shocks to matches (Mortensen and Pissarides 1994). I have built what I think is the simplest model that can explain some of the empirical evidence from the Tanzanian labour market, however, particularly earnings

[4]Some readers may find it odd that I include a job destruction rate and frictions in self-employment, since this is often assumed to be a residual sector that those who lose their wage employment move to until they find another wage job. I have not motivated my choice with empirical evidence but it is possible to think of situations in which this may indeed be a plausible assumption. For example, someone selling items to drivers stuck in traffic jams may find his customers disappear if the road is improved. An individual selling things from an illegally sited kiosk may find it destroyed by government officials. In both cases the individual concerned would have to look for a new job, in the sense of finding a site to sell goods at. This would take time and could be modeled by assuming there are frictions in self-employment. Frictions may well be lower in self-employment, as argued in Zenou (2008) and I allow for this possibility, exploring the effects of varying the level of frictions in section 6.

dispersion within wage and self-employment as well as a lack of movement between wage and self-employment that suggests a role for comparative advantage. The basis for the model is the Albrecht et al. (2009) matching model of a Latin American labour market, but I modify the model in ways that generate earnings distributions that better fit the evidence presented above.

5.2 Individual Behaviour

I begin by describing individual behaviour in the model. Individuals can choose to work in either wage or self-employment. I allow only self-employment productivity to vary across individuals. I assume that the population is constant and that self-employment productivity is distributed over a population of individuals whose size is normalised to 1. In much of the labour economics literature both employment for workers and the filled jobs of firms are thought of as assets that bring an instantaneous return, in the form of wage or self-employment income for workers and profits for firms. The capital value of these assets may change in the future as a result of shocks to the match which result in it being destroyed. In my model the probability that there will be a change in the capital value of the asset is the probability that the match will end. For workers, the new value of the employment asset is assumed to be the value of unemployment that results from the job ending. For the firm, the new value is the (negative) value of holding a vacancy open.

An individual with productivity y in self-employment earns y but faces a probability q_s that this self-employment opportunity will be destroyed, that she will enter unemployment and that there will be a change in the value of the self-employment asset. This means that the asset equation for the self-employed is[5]

$$rV_s(y) = y + q_s(V_u(y) - V_s(y)) \quad (1)$$

where V_u is the value of unemployment, V_s is the value of self-employment and r is the interest rate[6].

Unemployed workers search for employment opportunities and whilst doing this receive unemployment income $z(y)$, which is a function of their self-employment ability, as in Wong's (2003) model of the US labour market[7].

$$rV_u(y) = z(y) + \alpha \max[(V_s(y) - V_u(y)), 0] + m(\theta)\max[(V_e(y) - V_u(y)), 0]. \quad (2)$$

α is the probability that an unemployed worker will find a self-employment opportunity. $m(\theta)$ is the rate at which wage employment opportunities arrive

[5]I present an intuitive explanation of the asset equations that will be used throughout the model in section 8.B in the Appendix to this paper.

[6]For reference I list and provide a brief explanation of all the model parameters and variables in Table 5.

[7]Assortative mating by income, education or assets is a possible explanation for unemployment earnings depending on productivity, with a spouse providing income in a period of unemployment, but neither the household nor education are taken into account explicitly in this model. Fafchamps and Quisumbing (2005) provide evidence that sorting on human capital occurs in the Ethiopian marriage market.

and is a function of θ, a measure of labour market tightness. $\theta = v/u$, where v is the number of vacancies and u is the number of unemployed persons. When the labour market is tight v is low relative to u, θ is thus low and workers will have a harder time finding firms to match with, implying that the matching rate will be lower. The matching rate in self-employment is assumed to be exogenous. When an individual finds a self-employment opportunity or a wage employment opportunity, if she matches with a firm, this results in a change in the capital value of the unemployment asset ($V_u(y)$) that an unemployed person is thought of as holding.

I noted above that variation in self-employment productivity creates an opportunity cost of taking wage employment that is increasing in self-employment productivity, as in the model of Albrecht et al. (2009). Thus workers with low self-employment productivity choose employment in the wage sector and reject potential self-employment options, because they would earn much less in self-employment and are better off searching for wage opportunities. Similarly, workers with high self-employment productivity reject wage employment offers. Leaving unemployment, and entering wage employment, is worse than staying in unemployment and waiting for future self-employment opportunities for individuals with high self-employment productivity. This is captured by the max operator in the last part of equation (2). Similarly, taking self-employment opportunities for unemployed individuals with low self-employment productivity, is worse than waiting for future wage employment opportunities, and again the max operator in the second to last part of equation (2) captures the fact that individuals do not take opportunities that leave them worse off than their current situation.

The asset equation for wage employees is

$$rV_e(y) = w(y) + q_e(V_u(y) - V_e(y)), \tag{3}$$

where the wage w depends on y and V_e is the value of wage employment. As in self-employment, wage employment is assumed to end with an exogenous probability (q_e in wage employment) in which case the asset value $V_e(y)$ changes and the worker returns to unemployment.

Individuals only differ in their ability in self-employment. Despite this, the model generates variation in earnings in both sectors, shown in the dependence of w on productivity y in equation (3). This dependence of the wage on ability is a result of variation in the outside options of workers who are assumed to bargain with the firms with which they match. I explain the bargaining process in Section 5.4 below.

5.3 Firms

Firms are assumed to have homogeneous productivity y_w and to employ only one worker. They match with workers at the rate $m(\theta)/\theta$. When a match occurs the firm produces a continuous stream of output y_w and pays a wage $w(y)$ to the worker it matches with. The wage paid depends on worker ability because ability affects the worker's outside option. Filled jobs can also be thought of as assets that firms hold, meaning that the asset equation for a filled job is

$$r\prod_e(y) = y_w - w(y) + q_e(\prod_v - \prod_e(y)). \tag{4}$$

Matches are assumed to end with exogenous probability q_e, which brings about a change in the capital value of the filled job asset for the firm $\prod_e(y)$. Firms whose matches end are assumed to return to search for workers who are also searching for employment. While this search occurs, firms are assumed to incur a cost of holding a job open, meaning that the asset equation for a vacant job is

$$r\prod_v = -c + (m(\theta)/\theta)E[\max[(\prod_e(y) - \prod_v, 0]]. \tag{5}$$

In equation (5), c is the cost of holding a vacant job open per unit of time and $m(\theta)/\theta$ is the rate at which firms match with workers. Following much of the literature, I also assume that firms are risk neutral.

5.4 Wage Determination

The match between a firm and a worker generates a gain for both parties compared to what they would have earned if the match did not occur. These gains are called rents and the sum of the rents of the firm and worker are termed the surplus of the match in the matching literature. When workers and firms match they are assumed to bargain over how this surplus is distributed between them. Setting up a simple non-cooperative bargaining game with some straightforward assumptions can generate a "surplus sharing rule" (Cahuc and Zylberberg 2004).

In the model I have set out, a worker's outside option is the value of unemployment and a firm's is the value of keeping a vacancy open. Thus the rents are $(V_e(y) - V_u(y))$ for a worker and $\prod_e - \prod_v$ for a firm. This means that the surplus is $S = \prod_e - \prod_v - (V_e(y) - V_u(y))$.

I assume that with free entry firms compete away expected profits, so that $\prod_v = 0$ and hence that the firm's outside option has an expected value of zero. Thus $S = \prod_e - (V_e(y) - V_u(y))$. Following the matching literature I assume that the bargaining outcome is the maximum of the Nash criterion (Nash (1953), Binmore et al. (1986)). Thus the negotiated wage, w(y), maximises

$$[V_e(y) - V_u(y)]^\gamma[\prod_e(y)]^{(1-\gamma)}. \tag{6}$$

The first order conditions from this problem give $\prod_e(y) = (1-\gamma)S$ and $V_e(y) - V_u(y) = \gamma S$, the surplus sharing rule. Thus firms and workers share the surplus according to their bargaining power, which is determined outside the model. Substituting the value of $\prod_e$ from equation (4) it is possible to show that

$$w(y) = \gamma y_w + (1-\gamma)rV_u(y)). \tag{7}$$

Equation (7) shows a dependence of the negotiated wage w on y because workers' outside options, $V_u(y)$), vary with self-employment ability y, despite identical productivity for all employees. There are two sources of this variation. The first is that unemployment income depends on self-employment ability y. The second is that even if unemployment income was constant across workers, the value of the unemployment asset $V_u(y)$) differs for different kinds of workers, as I explore in section 5.5.

5.5 Productivity Cut Offs

The model setup and the variation in self-employment ability implies an opportunity cost of taking wage employment that is increasing in self-employment ability. Workers with high self-employment productivity prefer unemployment to wage employment and will reject wage employment offers because taking these would mean foregoing high self-employment income. Workers with low self-employment ability prefer unemployment to self-employment and will reject self-employment opportunities because these mean foregoing higher income in wage employment. Some individuals will take both kinds of employment because the incomes they earn in each are similar. The equilibrium of this model will thus be one in which there are two cutoff values in ability, y^* and y^{**} with $y^* < y^{**}$. These correspond respectively to the self-employment productivity at which unemployment is no longer preferred to self-employment, and at which wage employment is no longer preferred to unemployment.

Individual with ability $y^* < y < y^{**}$ take both self-employment and wage employment, whilst workers with ability $y < y^*$ do not take self-employment. This means that $y^* < y < y^{**}$ workers have an extra term in their unemployment asset equation (31), where their self-employment ability y enters directly. This affects the outside options of these workers and hence their wages have an extra source of wage variation compared to the $y < y^*$ workers , whose self-employment ability does not enter directly into their outside option in the wage bargain, since they are assumed never to take self-employment in this equilibrium.

I have thus injected a further degree of plausibility into the model by generating earnings dispersion in both sectors, despite only having one type of ability, in self-employment. In the next sections I follow Albrecht et al. (2009) in first solving the model, and then showing that the equilibrium is unique.

5.6 Solving the Model

Solving the model requires obtaining the cutoff values y^* and y^{**} and solving for the equilibrium value for θ, labour market tightness. I begin by finding an expression for y^* in terms of the exogenous model parameters and θ. Below y^*, $V_u(y) > V_s(y)$ (unemployment is preferred to self-employment). This explains why low ability individuals take only wage employment opportunities. Individuals with productivity above y^*, but below y^{**}, take both wage and self-employment. At y^*, $V_u(y^*) = V_s(y^*)$. This implies the max operator in the unemployment asset equation is now irrelevant for individuals with productivity y^*, because taking a wage job does not imply a decrease in the capital value of the unemployment asset. Thus equation (2) becomes

$$rV_u(y^*) = z(y^*) + m(\theta)(V_e(y^*) - V_u(y^*)). \tag{8}$$

It is then possible obtain $V_e(y^*)$ in terms of $V_u(y^*)$ from equation (3), the wage employment asset equation:

$$V_e(y^*) = \frac{w(y^*) + q_e V_u(y^*)}{r + q_e}, \tag{9}$$

and then solve for $V_u(y^*)$:

$$rV_u(y^*) = \frac{z(y^*)(r + q_e) + m(\theta)w(y^*)}{r + q_e + m(\theta)}. \tag{10}$$

Since $V_u(y^*) = V_s(y^*)$ at $y = y^*$, equation (1), the self-employment asset equation, then implies $rV_s(y^*) = y^*$. Thus using equation (10) I can express y^* in terms of θ and the exogenous parameters of the model:

$$y^* = \frac{z(y^*)(r + q_e) + m(\theta)w(y^*)}{r + q_e + m(\theta)}. \tag{11}$$

In a similar fashion I also find the value of y^{**}, the level of productivity above which individuals take only self-employment. Above y^{**}, $V_u(y) > V_e(y)$, meaning unemployment is preferred to wage employment. At y^{**}, $V_u(y^{**}) = V_e(y^{**})$ and hence equation (2), the unemployment asset equation, can be written as

$$rV_u(y^{**}) = z(y^{**}) + \alpha(V_s(y^{**}) - V_u(y^{**})). \tag{12}$$

Equation (1), the self-employment asset equation, then implies that

$$V_s(y^{**}) = \frac{y^{**} + q_s V_u(y^{**})}{r + q_s}. \tag{13}$$

I can then use equations (12) and (13) to substitute out $V_s(y^{**})$ and solve for $V_u(y^{**})$:

$$rV_u(y^{**}) = \frac{z(y^{**})(r + q_s) + \alpha y^{**}}{r + q_s + \alpha}. \tag{14}$$

y^{**} is the value of y such that $V_u(y^{**}) = V_e(y^{**})$ and this combined with equation (3), the wage employment asset equation, implies that $rV_e(y^{**}) = w(y^{**})$. This and equation (14) imply that

$$w(y^{**}) = \frac{z(y^{**})(r + q_s) + \alpha y^{**}}{r + q_s + \alpha}. \tag{15}$$

Now the bargained wage equation (7), with $V_u(y^{**}) = V_e(y^{**})$ and $rV_e(y^{**}) = w(y^{**})$, implies that $w(y^{**}) = y_w$. This is intuitive because the highest wage firms can pay must be paid to workers with the highest ability in self-employment who still take wage employment, and this is the wage which leaves the firm indifferent between agreeing to the match and returning to search for another worker, where the expected return is zero in equilibrium as a result of free entry.

I am then able to form an equation for y^{**} in terms of the exogenous parameters of the model:

$$\alpha y^{**} + z(y^{**})(r + q_s) = y_w(\alpha + q_s + r). \tag{16}$$

With a functional form for $z(y)$ I can solve directly for y^{**}. I also have one equation with two unknowns y^* and θ, and require a second to be able to solve for these variables. To do this it is necessary to first solve for the steady state employment flows, which I undertake in the next section.

5.7 Equilibrium Employment Flows

In the steady state equilibrium of this model the unemployment rate cannot be changing, implying that the flows into and out of unemployment must be equal. These flows differ by the workers' ability type, and by the sector of employment they choose. There are thus different steady state conditions for individuals with $y < y^*$, who take only wage employment, for those with $y^* < y < y^{**}$, who take both wage and self-employment opportunities, and for those with $y > y^{**}$, who take only self-employment. These steady state conditions give the unemployment rates for the different groups of workers and the overall unemployment rate for the population.

Following Albrecht et al. (2009), I let $u(y)$ be the fraction of time an individual of productivity y spends unemployed, $n_e(y)$ be the fraction of time spent in wage employment and $n_s(y)$ be the fraction of time spent in self-employment, meaning that $u(y) + n_e(y) + n_s(y) = 1$[8].

Individuals with productivity $y < y^*$ only move between wage employment and unemployment, implying that $n_s(y) = 0$. This means there is only one steady state condition for this group, which is that the flow of workers out of wage employment must equal the flow into wage employment. q_e is the rate at which wage jobs are destroyed, and $m(\theta)$ is the rate at which unemployed workers match with firms. The steady state condition is thus $q_e(1 - u(y)) = m(\theta)u(y)$, with

$$u(y) = \frac{q_e}{m(\theta) + q_e} \tag{17}$$

and

$$n_e(y) = 1 - u(y) = \frac{m(\theta)}{m(\theta) + q_e}. \tag{18}$$

Individuals with $y^* < y < y^{**}$ take both wage and self-employment opportunities, meaning there are two steady state conditions for these individuals. Firstly, the flows into wage employment equal the flows into unemployment

[8] I do not allow for non-participation in this model.

from wage employment, and secondly, the flows out of self-employment equal the flows into unemployment from self-employment. Hence $m(\theta)u(y) = q_e n_e(y)$ and $\alpha u(y) = q_s n_s(y)$. Combining these two implies

$$u(y) = \frac{q_e}{m(\theta) + q_e + \frac{\alpha q_e}{qs}}, \tag{19}$$

$$n_e(y) = \frac{m(\theta)}{m(\theta) + q_e + \frac{\alpha q_e}{qs}} \tag{20}$$

and

$$n_s(y) = \frac{q_e \alpha}{q_s(m(\theta) + q_e + \frac{\alpha q_e}{qs})}. \tag{21}$$

Individuals with $y > y^{**}$ take only self-employment so there is again only one steady state condition, that the flow from unemployment to self-employment equals the reverse flow from self-employment into unemployment. This implies $\alpha u(y) = q_s n_s(y)$, so that $n_s(y) = 1 - u(y)$, $n_e(y) = 0$,

$$u(y) = \frac{q_s}{\alpha + q_s} \tag{22}$$

and

$$n_s = \frac{\alpha}{\alpha + q_s}. \tag{23}$$

Total unemployment can then be calculated as

$$u = \int_0^{y^*} u(y)f(y)dy + \int_{y^*}^{y^{**}} u(y)f(y)dy + \int_{y^{**}}^{1} u(y)f(y)dy, \tag{24}$$

where $u(y)$ varies for each of the three categories of worker, and $f(y)$ is the ability distribution in the population.

5.8 Finding the Equilibrium

The model can be solved by finding a second equation containing the two unknowns y^* and θ, which can be obtained from the free entry condition $\prod_v(y) = 0$. This, together with equation (5), the vacant job asset equation, implies

$$c = \frac{m(\theta)}{\theta} E[\max[(\prod_e(y), 0]]. \tag{25}$$

Now in the equilibrium $\prod_e(y) \geqslant 0$ because firms only match with workers that they do not make a loss from hiring and hence

$$c = \frac{m(\theta)}{\theta} E[\prod_e(y)]. \tag{26}$$

The value of a filled job to a firm is uncertain because the wage paid depends on the ability of the worker employed. In equilibrium only workers with productivity $y < y^{**}$ take wage employment offers. To calculate the expected value of a filled job, $E[\prod_e(y)]$, I also need to know the distribution of workers with which firms can potentially match. I know the distribution of worker ability in the population but this is different to the distribution of unemployed workers who would take wage employment offers from firms. Following Albrecht et al. (2009), I denote the density of ability for those in unemployment as $f_u(y)$ and then use Bayes' Law to write

$$f_u(y) = \frac{u(y)f(y)}{u}. \tag{27}$$

Hence I can rewrite the free entry condition (25) as

$$c = \frac{m(\theta)}{\theta} \int_0^{y^{**}} \prod_e(y) \frac{u(y)f(y)}{u} dy. \tag{28}$$

The equation for a filled job, equation (4), with $\prod_v = 0$, implies that $\prod_e(y) = \frac{y_w - w(y)}{r+q_e}$. I also know the value of $w(y)$ from equation (7), the outcome of the Nash Bargain with free entry. Hence I can write the free entry condition as

$$c = \frac{m(\theta)(1-\gamma)}{\theta(r+q_e)} \int_0^{y^{**}} (y_w - V_u(y)) \frac{u(y)f(y)}{u} dy. \tag{29}$$

I now have the second of the two equations in y^* and θ, which can be solved to find the equilibrium values of these variables, once I have values for $V_u(y)$ for those workers with productivity below y^{**}, who take wage employment and who can be hired by firms.

The values of $V_u(y)$ differ for individuals with productivity above and below y^*. Workers with $y^* < y < y^{**}$ take both wage and self-employment, meaning that the self-employment and wage employment asset equations (equations (1) and (3)) still apply and the unemployment asset equation (equation (2)) becomes

$$rV_u(y) = z(y) + \alpha(V_s(y) - V_u(y)) + m(\theta)(V_e(y) - V_u(y)). \tag{30}$$

Solving these three equations for rV_u implies that

$$rV_u(y) = \frac{z(y)(r+q_s)(r+q_e) + \alpha y(r+q_e) + \gamma m(\theta) y_w (r+q_s)}{(r+q_s)(r+q_e) + \alpha(r+q_e) + \gamma m(\theta)(r+q_s)}. \tag{31}$$

For workers with $y < y^*$ only the wage job asset equation (3) still applies and the unemployment asset equation (2) becomes

$$rV_u(y) = z(y) + m(\theta)(V_u(y) - V_e(y)), \tag{32}$$

since unemployment is preferred to self-employment for $y < y^*$ individuals. Solving equations (3) and (32) implies that

$$rV_u(y) = \frac{z(y)(r+q_e) + \gamma m(\theta) y_w}{r + q_e + \gamma m(\theta)}. \tag{33}$$

I now have the values for $rV_u(y)$ in terms of the exogenous parameters. I can then also substitute the values of y^* from equation (11) and y^{**} from equation (16) into the free entry condition (29). This is a non-linear equation in theta, meaning I am required to solve it numerically, which I explain in more detail below, after proving that a unique equilibrium to the model exists.

5.9 The Existence of a Unique Equilibrium

A steady state equilibrium is one in which the expected value of maintaining a vacancy is zero, matches are consummated only if it is in the interests of the firm and individual to do so, the steady state employment flows conditions hold, wage employment is not worthwhile for individuals with $y > y^{**}$ and self-employment is not worthwhile for individuals with $y < y^*$ (Albrecht et al. 2009). A value of θ together with cutoffs y^* and y^{**} and an unemployment rate $u(y)$ at which the above conditions are satisfied is an equilibrium of the model.

Since I have shown above that $u(y)$ and y^* are uniquely determined by θ, and that y^{**} is determined by the exogenous parameters of the model, a unique equilibrium exists if there is a unique value of θ that solves the free entry condition (28). If I assume decreasing returns to scale in the matching function m, a standard assumption in the literature (Cahuc and Zylberberg 2004), then $\frac{m(\theta)}{\theta}$ is strictly decreasing in θ. $u(y)$ has different values for $y < y^*$ and $y^* < y < y^{**}$ but, as can be seen from equations (17) and (22), both are strictly decreasing in θ. Since u is the sum of these two terms for $u(y)$ that are both strictly decreasing in θ, it is also strictly decreasing in θ (u also includes a term for $u(y)$ above y^{**} which does not depend on θ).

To show that the free entry condition is decreasing in θ I now only need show to show that $rV_u(y)$ is increasing in θ. Firstly, for $y < y^*$, equation (33) can be written as

$$rV_u(y) = \frac{z(y)A + By_w}{A + B}, \tag{34}$$

where only B depends on θ. I can differentiate $rV_u(y)$ and find that

$$\frac{\partial rV_u(y)}{\partial \theta} = \frac{(y_w - z(y))AB'}{(A + B)^2}. \tag{35}$$

Since $y < y^*$ this means $y < y_w$ and since $z(y) < y$ this implies $y_w - z(y)$ is always positive. The other terms are clearly also all positive so $rV_u(y)$ is increasing in θ for $y < y^*$.

For $y^* < y < y^{**}$ the value of $rV_u(y)$ is found in equation (31), which can be written as

$$rV_u(y) = \frac{z(y)D + Ey + Fy_w}{D + E + F}, \tag{36}$$

where only F depends on θ. To determine when $rV_u(y)$ is increasing in θ I calculate

$$\frac{\partial rV_u(y)}{\partial \theta} = \frac{F'[D(y_w - z(y)) + E(y_w - y))]}{(D + E + F)^2}. \tag{37}$$

Since F' and $(D + E + F)^2$ are always positive this implies that $rV_u(y)$ is increasing in θ if $D(y_w - z(y)) + E(y_w - y)) > 0$. Since $y^* < y < y^{**}$ this means $y_w - z(y) \geq 0$ and $y_w - y \geq 0$. And since D and E are both positive this implies $rV_u(y)$ is increasing in θ for $y^* < y < y^{**}$.

Having shown that the right hand side of the free entry condition is strictly decreasing in θ, to prove that the equilibrium is unique I must show that as $\theta \to \infty$ the right hand side of the free entry condition tends to zero and that it tends to ∞ as $\theta \to 0$. $\frac{m(\theta)(1-\gamma)}{\theta(r+q_e)}$ clearly tends to ∞ as $\theta \to 0$ and tends to zero as $\theta \to \infty$. Thus I only need to show that the definite integral is bounded to show that the equilibrium is unique. u and $u(y)$ are unemployment rates that are bounded between 0 and 1 for any value of θ. $f(y)$ is the distribution of abilities in the population and so does not depend on θ, and neither does y_w. $V_u(y)$ is the value of unemployment. As $\theta \to 0$ the labour market becomes very tight and matching with a firm increasingly likely. Thus, in the limit, individuals would not find wage employment. But as the model has been set up they can still find self-employment, where the arrival rate does not vary with θ. Thus the value of unemployment is still bounded in this case.

A similar argument holds when $\theta \to \infty$ and wage jobs are immediately available for those who become unemployed, except that some individuals still take self-employment, even though wage jobs are available instantly, because their employers would not pay a wage greater than the common level of productivity. Thus the definite integral is bounded and there is a unique value of θ that solves the free entry condition.

I have outlined a unique equilibrium above in which there are two cutoffs y^* and y^{**}. Though the equilibrium will always be unique there may actually be two, one or zero cutoffs, depending primarily on the exogenous value of y_w but also on the other exogenous parameters. At values of y_w close to 1 there are no individuals who only take self-employment and some who take both wage and self-employment. If y_w takes a value above 1, however, taking self-employment becomes relatively less lucrative and at a high enough value of y_w eventually all individuals choose only wage employment. Due to the high levels of self-employment in Tanzania I focus on the two cutoff type of equilibrium in this paper, though I do discuss the implications of varying y_w and some of the other exogenous parameters below.

5.10 Equilibrium Earnings distributions in Wage and Self-Employment

In the model I have outlined wages depend on worker's outside options, which are themselves determined by individuals' self-employment productivities, and the self-employed earn their productivity. This gives rise to a distribution of wages and also of self-employment earnings. I now calculate these distributions in equilibrium, which then enables me to calculate the average wage paid in each sector.

$f(y)$ is the density of abilities in the population. If $h(y)$ is the density of ability in wage employment and $P(E)$ is the probability of being in wage employment, Bayes Law then implies that

$$h(y) = \frac{P(E|y)f(y)}{P(E)} = \frac{n_e(y)f(y)}{\int_0^1 n_e(y)f(y)dy}. \tag{38}$$

For simplicity I assume a uniform distribution of ability between 0 and 1. It is clear from equations (18) and (20) above that the probability of wage employment, n_e, differs for different values of y. This means

$$\begin{aligned} h(y) &= \frac{\frac{m(\theta)}{m(\theta)+q_e}}{\frac{y^* m(\theta)}{m(\theta)+q_e} + (y^{**} - y^*)\frac{m(\theta)}{m(\theta)+q_e+\frac{\alpha q_e}{q_s}}} && 0 \leq y < y^* \\ &= \frac{\frac{m(\theta)}{m(\theta)+q_e+\frac{\alpha q_e}{q_s}}}{\frac{y^* m(\theta)}{m(\theta)+q_e} + (y^{**} - y^*)\frac{m(\theta)}{m(\theta)+q_e+\frac{\alpha q_e}{q_s}}} && y^* < y < y^{**} \\ &= 0 && y^{**} < y \leq 1 \end{aligned}$$

Since $h(y)$ takes a constant value for $0 \leq y < y^*$, another constant value on $y^* < y < y^{**}$ and is zero beyond y^{**} calculating the distribution of wages $m(w)$ is relatively simple. The distribution will also have two constant parts on the domain of w, but with a much narrower base than the ability distribution. If I let the height of the first block of the $h(y)$ distribution be a then the probability mass within this block is ay^*. The area of the first block of the distribution of w will be the same and thus the height of this block will be $\frac{ay^*}{w(y^*)-w(0)}$. Using similar reasoning, the height of the second block will be $\frac{b(y^{**}-y^*)}{w(y^{**})-w(y^*)}$. Solving the model numerically I will be able to calculate the equilibrium wages at $y = 0$, $y = y^*$ and $y = y^{**}$, obtain the density and then be able to solve for the average wage paid in the equilibrium. The average wage is calculated as $\int_0^{y^{**}} wm(w)dw$.

Calculating the density of earnings in self-employment is also simple. Again, the density will differ for those with ability y above and below y^{**} since the probability of self-employment varies with ability. Denoting the probability of self-employment as $P(S)$ I can write the distribution of ability in self-employment, again using Bayes' rule, as

$$g(y) = \frac{P(S|y)f(y)}{P(S)} = \frac{n_s(y)f(y)}{\int_0^1 n_s(y)f(y)dy}. \tag{39}$$

From equations (21) and (23) I have the probability of being in self-employment for individuals with $y > y^*$ and I know that $P(S) = 0$ if $y < y^*$. Thus it is possible to write down the density of ability in self-employment as

$$
\begin{aligned}
g(y) &= 0 & 0 < y < y^* \\
&= \frac{\frac{q_e\alpha}{q_s(m(\theta)+q_e+\frac{\alpha q_e}{q_s})}}{((y^{**}-y^*)\frac{q_e\alpha}{q_s(m(\theta)+q_e+\frac{\alpha q_e}{q_s})})+(1-y^{**})\frac{\alpha}{\alpha+q_s}} & y^* < y < y^{**} \\
&= \frac{\frac{\alpha}{\alpha+q_s}}{((y^{**}-y^*)\frac{q_e\alpha}{q_s(m(\theta)+q_e+\frac{\alpha q_e}{q_s})})+(1-y^{**})\frac{\alpha}{\alpha+q_s}} & y^{**} < y \leq 1
\end{aligned}
$$

Since earnings in self-employment are simply the value of individual productivity in self-employment, the ability distribution is also the earnings distribution in self-employment.

6 Numerical Solutions to the Model

Solving for two unknowns θ and y^* in two non-linear equations requires numerical methods. I use Broyden's method of solving nonlinear equations numerically, a quasi-Newton method first outlined by Broyden (1965). This was implemented using the Compecon Matlab package, a companion package to the book *Applied Computational Economics and Finance* (Miranda and Fackler 2002).

The two non-linear equations to be solved are the free entry condition (29) and the solution for y^* (11). A solution requires assuming values for the exogenous parameters and functional forms of the model, or estimating the model and parameters using data. As far as I am aware, there have been no matching models estimated on African, or possibly even any developing country data, and as a result there are no empirical estimates of most of the parameters for which I require values. I do not attempt to prove the parameters are identified or estimate my model on the Tanzanian Urban Panel Survey data here either; rather, I simply follow other papers in using what seem to be reasonable parameter values and then explore what the model predicts about the earnings distributions across wage and self-employment, transitions between the two sectors, and average wages in each sector.

6.1 Exogenous Parameter Values

The values of the exogenous parameters which I need in order to solve the model are those of the interest rate, the job destruction rate in self-employment and wage employment, the arrival rate of self-employment opportunities, firm productivity, the cost to a firm of holding a job open and the bargaining power parameter. I also require a functional form for $z(y)$, which shows how unemployment income depends on ability as well as a functional form for the matching function $m(\theta)$. I assume that the unit of time is one year.

Providing empirical estimates of the rest of the exogenous parameters and functional forms is challenging and I do not attempt to explore the conditions under which the parameters would be identified. I thus follow some of the assumptions of Albrecht et al. (2009), the only other two sector developing country microeconomic matching model I am aware of, in assuming values for these parameters and the functional forms, but modify these slightly for the Tanzanian context in some cases. I assume the matching function to be of the form $m = 4\theta^{.5}$, the cost of holding a job open $c = .3$, the arrival rate of self-employment offers $\alpha = 3$, the real interest rate $r = 0.05$ and a uniform distribution of self-employment productivity between 0 and 1. I assume that $y_w = .5$ which, given $y \sim \text{uniform}(0,1)$, is the mean and median of this productivity. I discuss the implications of other potential underlying distributions below. I finally assume that $z = .3y$, so that individuals have unemployment income that is 30 percent of their self-employment productivity.

6.2 Equilibrium Solution

Having assumed the value of parameter values, as well as of functional forms, I now explore the solution to the baseline model, before undertaking some comparative statics.

The equilibrium outcome with the chosen parameter values and functional forms is shown in the third column of Table 6. In this baseline equilibrium $y^* = .3936$, $y^{**} = .5608$ and $\theta = .51$. This means that 39 percent of the population choose wage employment only, around 17 percent choose either wage or self-employment and 46 percent choose only self-employment. Total unemployment is 13.5 percent, with unemployment being 15 percent for individuals with $y < y^*$, 7.8 percent for individuals with $y^* < y < y^{**}$ and 14 percent for high productivity individuals with $y > y^{**}$. Individuals with productivity $y^* < y < y^{**}$ take both wage and self-employment opportunities so their unemployment rate is much lower than for those with $y < y^*$ or $y > y^{**}$ who take only wage and self-employment offers respectively. The vacancy rate $v = u\theta$ so v =6.75 percent. The average wage paid is .44 and the average earnings in self-employment is .73, meaning that the model generates earnings in self-employment that are much higher than those in wage employment. I discuss this further in section 6.2.4 below.

6.2.1 Earnings variation in wage employment

Those who take only wage employment in the equilibrium ($0 < y < y^*$) have an extremely low variance of earnings since $w(0) = .43$ whilst $w(y^*) = w(.39) = .447$. This narrow distribution is a result of identical firm productivity and the small variation in the outside options of those who choose only wage employment in the equilibrium. For these individuals, their self-employment productivity does not enter directly into their outside option in the wage bargain with firms, because they never actually exercise their self-employment option in the equilibrium of the model. Thus only the difference in workers' unemployment income, which is assumed to be a function of self-employment productivity, generates wage differentials amongst these workers.

There is more wage variation over a narrower range of abilities for those who also take self-employment opportunities ($y^* < y < y^{**}$) since $w(y^*) = w(.39) = .447$ and $w(y^{**}) = w(.56) = .5$. This is because these individuals do exercise their self-employment options, which enter the wage bargain and generate greater variation in individuals' outside options.

Having shown how it is possible to generate the distribution of earnings conditional on wage employment in Section 5.10 above, I now undertake this and describe the equilibrium wage density. It consists of two rectangular blocks, one between $w(0) = .43$ and $w(y^*) = .447$ containing mass .82 (in the row of Table 6 labeled "wmass 1") and the other between $w(y^*) = .447$ and $w(y^{**}) = .5$ containing mass .18 (in the row of Table 6 labeled "wmass 2"). The corresponding cdf is shown in Figure 4 (labeled baseline), which emphasises the lack of variation in wages. Allowing unemployment income to vary by self-employment ability thus buys some degree of realism at the lower end of the earnings distribution, in that there is some variation in the bargained wage for individuals taking only wage employment, which would not be the case if unemployment income was fixed (there would be a mass point). Figure 4 suggests, however, that the amount of variation is limited.

6.2.2 Earnings variation in self-employment

Those taking only self-employment simply earn their self-employment productivity, which means that there is a much larger amount of variation in self-employment earnings compared to wages earned working for firms with identical productivity. The minimum amount earned in self-employment is y^*, whilst the maximum is the highest productivity in the distribution of y, which are .39 and 1 in the baseline model. Thus the maximum and minimum earnings for the population are respectively higher and lower than the maximum and minimum earnings in wage employment.

The probability of self-employment differs above and below y^{**}, because unemployed individuals with productivity $y < y^{**}$ also take wage employment offers if these arrive before self-employment opportunities. This means the density of earnings conditional on being self-employed also consists of two rectangular blocks, the first between y^* and y^{**} and the second between y^{**} and 1. The first contains mass .17 (the row labeled "smass 1" in Table 6) and the second .83 in the baseline equilibrium equilibrium (in the row of Table 6 labeled "smass 2").

6.2.3 Transitions

The comparative advantage of individuals in one or other sector in the model means that most individuals prefer to take opportunities in only one of the sectors. Only those individuals with self-employment productivity near the constant level of firm productivity take both wage and self-employment opportunities. This is consistent with the evidence from the TUPS data, shown in Table 4, which suggests that there is little movement between wage and self-employment.

6.2.4 Analysis of the Model Solution

The ILFS data suggest that earnings in self-employment exhibit a higher variance than those in wage employment but that the medians of the two distributions are roughly equal. The TUPS OLS regressions suggested earnings in wage employment in small firms was lower than in self-employment after controlling for basic observable human capital. This goes against the view that self-employment is a worse outcome than wage employment but accords with the view of Maloney (2004) that in Latin America, informal salaried workers are generally considered to be the worst off of the employed, below even the informally self-employed.

In the model I have built, average self-employment earnings are much higher than those in wage employment as a result of high productivity individuals choosing self-employment. This is partly as a result of the assumed underlying productivity distribution, which was chosen for its simplicity rather than for its realistic properties. The differential would be much lower with an ability distribution that had more mass near the centre. This would result in a higher proportion of individuals taking both wage and self opportunities and give less weight to those individuals with higher self-employment ability in calculating the average. My model will always have average self-employment earnings higher than the average wage paid, however, without more fundamental changes, some of which I discuss below.

The model I have built above can be thought of as an assignment model in the tradition of Roy (1951), where workers have a choice of sector in which to seek employment. In a review of assignment models, Sattinger (1993) notes that different models generate earnings distributions that can amplify or compress the underlying ability distribution assumed to be present in the population. In my model the identical productivity across all firms compresses the underlying ability distribution very substantially, so that there is much less inequality in wages than in ability. There is no such compression in self-employment, where earnings are equivalent to productivity. The compressed wage distribution is unrealistic but qualitatively, the distribution results are in line with the data I have presented from Tanzania, in that self-employment earnings exhibit a higher variance than wage employment earnings and that the lowest and highest earnings observed are in self-employment. This difference in variance is much higher in my model than in the data, however. One possible way of generating more variation in the wages of those only taking wage opportunities is to increase frictions in the model, which I show below. By changing the functional form of the matching function in wage employment and the exogenous arrival rate of self-employment opportunities, it is possible to increase frictions, make differences in unemployment income more important and thus increase the variation in wages paid.

Although I have shown some weaknesses in my model, it does generate realistic predictions for the transitions between wage and self-employment and the relative size of the variances of earnings in the two sectors without imposing the cost of extra complexity. It also compares well with the Albrecht et al. (2009)

model which has constant earnings in the informal sector.

6.3 Comparative Statics

Having explored the equilibrium outcomes in the baseline model, I now explore how the equilibrium changes as I vary some of the exogenous parameters. To start with, I explore the effect of raising firm productivity, shown in Table 6, where y_w varies across the forth row. This could be thought of as exogenous development progress, where firms become more productive. This could possibly as a result of increased product market competition as trade barriers are lowered and imports increase, or as local production by more productive foreign-owned firms increases. As firm productivity rises, the percentage of workers willing to take only wage opportunities increases, as does the self-employment ability of the worker with productivity y^*, who is indifferent between unemployment and self-employment. The percentage of individuals willing to take both wage and self-employment opportunities first rises and then falls as productivity increases whilst the percentage of the population in self-employment declines as productivity increases. The variance of wages in wage employment is constant and then decreases at high levels of productivity. Both wage and self-employment average earnings rise as productivity increases: in wage employment this is because firm productivity rises, and in self-employment because rising productivity pushes lower productivity individuals who were self-employed into wage employment and thus raises the average productivity of those still in self-employment. Unemployment decreases as firm productivity rises, mainly as a result of decreased unemployment amongst the low ability types who take only wage employment.

It is also possible to explore how changing the matching function and the exogenous arrival rate of self-employment opportunities affects the wage and self-employment earnings distributions. I noted above that increased frictions would raise the importance of unemployment income and hence possibly generate more variation in the wage employment distribution. Table 7 shows the results of varying both the matching function for wage employment and the exogenous arrival rate of self-employment opportunities α. The matching function is of the form $a\theta^{.5}$ and I allow a to take on the values 1,2 and 3. The variation in wages for those who always take only wage employment is proportionately much higher when the matching rate in wage employment is lower. Figure 4 shows the cdf for $a = 1$ and $\alpha = .5$. The difference between $w(0)$ and $w(y^*)$ doubles when a is only 1 (compared to $a = 4$ in the baseline equilibrium). In absolute terms it is still low, however, only around a maximum of 5 percent of the variation in abilities of the individuals who take only wage employment. Unfortunately, increased frictions come at a high cost as the unemployment rate that results from large frictions is very high, much higher than what is seen in the data. I discuss other ways of increasing variation in wages in the following section.

6.4 Potential extensions

The model I have built is relatively simple since it assumes identical firms and only allows individual heterogeneity in one sector. Despite this, it generates some realistic outcomes in predicting the pattern of transitions, and also the relative size of the variation in wage and self-employment earnings. There are several possible extensions which would enhance the predictions of the model but which I have not pursued in this paper. Varying firm productivity would be one possibility for improving the model since this would increase the variation in wages across individuals in wage employment. Another way to increase wage variation without increasing the amount of heterogeneity, would be to follow the original Mortensen and Pissarides (1994) model and include random productivity shocks that lower the productivity of the match and hence the wage paid. This is also the way match break up is modeled in the formal sector in Albrecht et al. (2009). This assumption would mean that the matches that exist in the equilibrium have existed for different lengths of time, have experienced different numbers of productivity shocks and will thus pay many different wages, even amongst those having the same initial productivity. This requires introducing an endogenous reservation wage that depends on the tightness of the labour market so that when a productivity shock lowers the wage below some level, the match dissolves and the firm and the worker look for other opportunities. This would be more realistic since it endogenises the separation rate rather than having an exogenous rate like in my model.

Productivity shocks or heterogeneous firm productivity would generate more realistic variation in wages but this would not get around high ability individuals still taking only self-employment. One potential way of addressing this issue is to allow for individual heterogeneity in both sectors, as in the Roy (1951) model. There would then be a question of how different assumptions about the correlation of individual ability in the two sectors affects the model's predictions. Positively correlated ability would likely result in many more transitions between wage and self-employment than are shown in the data.

My model has different assumptions about arrivals of opportunities in each sector. The rate of matching for wage employment is assumed to vary with labour market tightness but the arrival rate of self-employment opportunities is exogenously determined. Although I have not presented any empirical evidence of this, it would seem preferable to have the self-employment opportunity arrival rate also depend on labour market tightness. It seems intuitive that the self-employment arrival rate is higher the tighter the labour market is, although it is not clear what effect this would have on the uniqueness of the equilibrium in the current model.

The question of the origins of the firms that workers match with has not been addressed in this paper. This is an important issue in a labour market where self-employment is so prevalent. Future work should incorporate the possibility of the self-employed owning firms and employing other workers, as in Banerjee and Newman's (1993) model of occupational choice and economic development.

7 Conclusion

In this paper I have presented some stylised facts from the Tanzanian labour market and built a model to explain these. I have moved beyond a focus on average wage differentials in the choice between wage and self-employment in Tanzania and presented a model which posits a central role for individual heterogeneity in both the employment choice and earnings determination. I have shown that there is a large amount of heterogeneity in earnings both within and between what have traditionally been called the formal and informal sectors. I have also shown that there are also very few transitions between small firm wage employment and self-employment, despite these having traditionally been lumped together as the informal sector. I have also presented some evidence that small firm employees earn less than the self-employed in Tanzania, challenging the notion that self-employment is a residual sector for those unable to find wage employment.

The model of the urban Tanzanian labour market that I have built can explain some of these outcomes, using a matching framework, and allowing for variation in self-employment productivity that implies an opportunity cost of taking wage employment which is increasing in self-employment productivity. Workers' outside options include this opportunity cost and so bargained wages also vary as a result, despite individuals having constant productivity in wage employment.

The model generates distributions of earnings in wage and self-employment that exhibit characteristics that are broadly similar to the empirical distribution, including higher variance in self-employment earnings than wage employment earnings. The transitions predicted by the model are also similar to those observed empirically, in that there is little movement between wage and self-employment, as a result of most individuals in the population working only in the sector where they have a comparative advantage. As a result of selection by high ability individuals into self-employment, the model I have built generates average earnings in self-employment that are much higher than those in wage employment. This is not too dissimilar from the evidence I presented that earnings in small firms are roughly 15 percent lower than in self-employment.

I have suggested several ways of increasing the complexity of the model and generating patterns more similar to those observed empirically. In particular, allowing heterogeneous firms, a more complicated ability distribution in self-employment and possibly adding ability in wage employment would generate outcomes closer to those observed in the data. I believe the strength of the model is its relative simplicity in capturing the distinction between wage and self-employment in the matching framework and its prediction of some of the stylised facts from the urban Tanzanian labour market.

Table 1: Urban Employment by Sector and education

Employment Categories	**2001** **N**	**2006** **N**	**Pooled** < **O levels**	**O levels or Higher**
Self Emp no Employees	2168	3294	4985	477
Self Emp with Employees	458	522	729	251
Small Firm Wage Employment	470	660	1012	118
Large Firm Wage Employment	238	718	681	275
Public Employment	1080	1055	868	1267
Agriculture	1853	2785	4530	108
Total	6815	9736	13732	2819

Source: ILFS 2001 and 2006. O levels represents 11 years of education under the new education system and 12 under the old system

Table 2: Median Monthly Earnings

Employment Categories	**2001**	**2006**
Self Emp no Employees	30000	29464
	(98007)	(172078)
Self Emp with Employees	78571	56250
	(492279)	(259446)
Small Firm Wage Employment	30000	28125
	(42826)	(69443)
Large Firm Wage Employment	45000	45313
	(49195)	(136469)
Public Employment	78000	75000
	(192792)	(258443)
Frequency	4962	6951

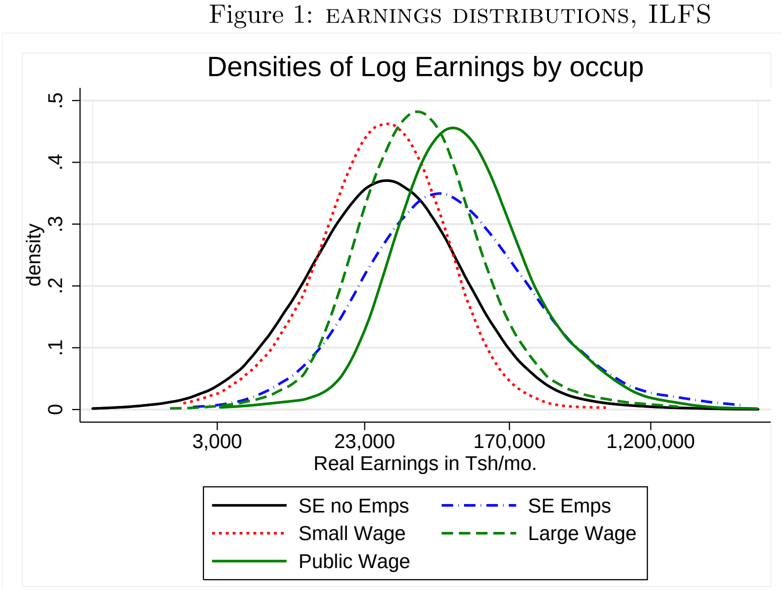
Standard Deviations in Parenthesis. Earnings expressed in monthly constant 2001 Tanzanian Shillings. The average exchange rate was Tsh 916 per US dollar in 2001 and 1261 in 2006 (International Monetary Fund 2010)

Table 3: OLS earnings regressions from the TUPS

	2004	2005	Pooled
	(1)	(2)	(3)
Total Exp	0.024** (0.012)	0.014 (0.01)	0.027*** (0.008)
$exp^2/100$	-.030 (0.021)	-.015 (0.019)	-.036*** (0.013)
Ln(Hours)	0.025 (0.117)	0.298** (0.136)	0.028 (0.073)
Ln(firsiz)	0.18*** (0.033)	0.168*** (0.048)	0.149*** (0.022)
Ln(emps)	0.615*** (0.123)	0.492*** (0.135)	0.555*** (0.076)
Tenure	-.006 (0.013)	0.017 (0.012)	0.006 (0.007)
$Tenu^2/100$	0.011 (0.033)	-.040 (0.037)	-.012 (0.02)
Male	0.217*** (0.065)	0.299*** (0.065)	0.313*** (0.044)
Educ	0.0004 (0.024)	-.008 (0.026)	0.01 (0.019)
$Educ^2/100$	0.634*** (0.183)	0.496*** (0.172)	0.436*** (0.131)
Pub Ent	-.085 (0.215)	0.193 (0.236)	0.112 (0.135)
Civ Ser	0.985*** (0.17)	0.908*** (0.125)	0.826*** (0.073)
Priv Wag	-.241* (0.124)	-.177 (0.116)	-.121 (0.074)
Const.	9.626*** (0.483)	8.735*** (0.569)	9.634*** (0.316)
Obs.	568	579	1643
R^2	0.354	0.35	0.327

Source: Tanzanian UPS 2004-2005. A year dummy for 2004 was included but not reported.

Figure 1: earnings distributions, ILFS



Source: pooled earnings data from Tanzanian ILFS 2001 and 2006. The kernel density was estimated using Stata.

Table 4: Occupation Transitions

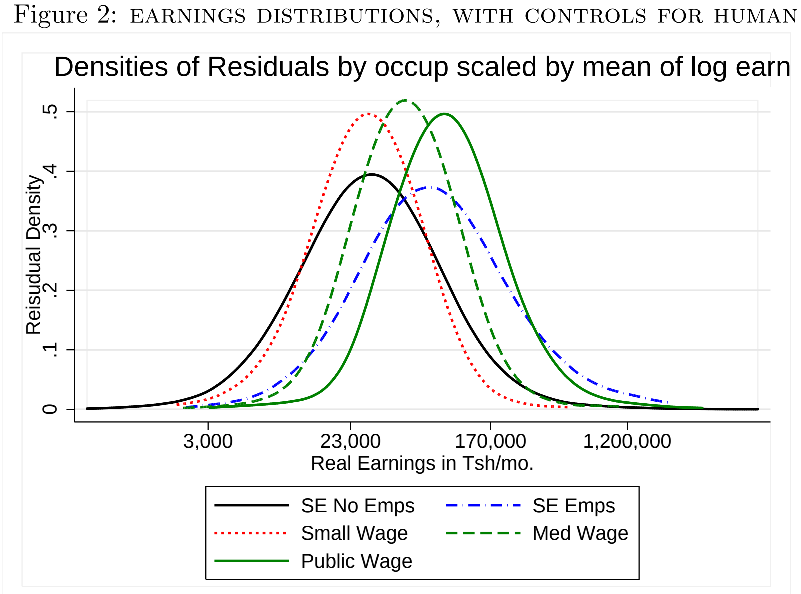
	Occupation						
Occupation	S/E no emps	S/E with emps	Small firm	Large firm	Public	none	Total
	%	%	%	%	%	%	%
S/E no emps	87	4	1	0	1	7	100
S/E with emps	43	48	0	2	0	7	100
Small firm	2	2	49	22	10	15	100
Large firm	2	0	6	34	52	6	100
Public	4	0	4	11	70	11	100
none	55	16	2	2	4	20	100
Total	55	9	6	8	13	9	100

Source: 2004 and 2005 rounds of the Tanzanian Urban Panel Survey. Total number of observations observed in both 2004 and 2005 in this table is 457

Table 5: Model Parameters

Parameter		Description
Exogenous	r	Interest rate
	y	Self-employment productivity
	$z(y)$	Unemployment income
	α	Matching rate in self employment
	m	Matching function in wage employment
	y_w	Firm productivity
	q_e	Wage job destruction rate
	q_s	Self-employment job destruction rate
	c	Cost of holding a vacant job open
	γ	Bargaining power parameter
Endogenous	V_u	Value of unemployment asset
	V_s	Value of self-employment asset
	V_e	the value of wage employment asset
	θ	Tightness of the labour market
	w	Wage in wage employment
	u	Unemployment rate
	v	Vacancy rate
	$\prod_e$	Value of a filled job asset
	$\prod_v$	Value of a vacant job asset

Figure 2: EARNINGS DISTRIBUTIONS, WITH CONTROLS FOR HUMAN CAPITAL



Source: pooled earnings data from Tanzanian ILFS 2001 and 2006. The kernel density was estimated using Stata.

Table 6: VARYING FIRM PRODUCTIVITY

θ	0.18534	0.32621	0.50762	0.73923	1.039	1.4413	1.9975	2.3829
y^*	0.20731	0.29917	0.39364	0.49024	0.5888	0.68945	0.79211	0.88912
y^{**}	0.33649	0.44866	0.56082	0.67299	0.78515	0.89731	1	1
y_w	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$w(0)$	0.24153	0.335	0.43038	0.5273	0.62564	0.72545	0.8267	0.92439
$w(y^*)$	0.25365	0.34958	0.44682	0.54512	0.6444	0.74473	0.84606	0.94456
$w(y^{**})$	0.3	0.4	0.5	0.6	0.7	0.8	0.89765	0.971
avg wage	0.25372	0.34859	0.44497	0.54256	0.64125	0.74111	0.8417	0.93632
wmass 1	0.79041	0.8061	0.81696	0.82535	0.83233	0.83854	0.85003	0.92078
wmass2	0.20959	0.1939	0.18304	0.17465	0.16767	0.16146	0.14997	0.079219
avg self earnings	0.62249	0.67494	0.72783	0.78045	0.83155	0.87716	0.89606	0.94456
smass 1	0.11543	0.14093	0.17343	0.21989	0.29683	0.46044	1	1
smass2	0.88457	0.85907	0.82657	0.78011	0.70317	0.53956	0	0
u	0.1538	0.1454	0.13466	0.12211	0.10797	0.092203	0.07572	0.072335
u1	0.22502	0.17956	0.14926	0.12693	0.10924	0.0943	0.081256	0.07491
u2	0.095748	0.086436	0.078741	0.072055	0.065987	0.060225	0.054625	0.051682
u3	0.14286	0.14286	0.14286	0.14286	0.14286	0.14286	NaN	NaN

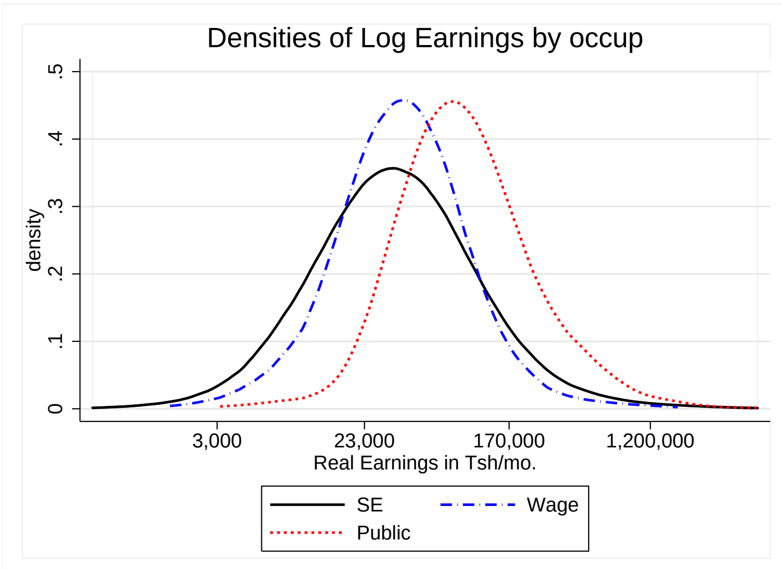
Note: each column represents a model solution, which changes as firm productivity varies, shown in the forth row. wmass1 is the percentage of wage employees who only take wage employment in the equilibrium and wmass2 is the percentage of wage employees who also take self-employment in the equilibrium. smass1 is the percentage of the self-employed who also take wage employment in the equilibrium and smass2 is the percentage of the self-employed who only take self-employment in the equilibrium. u is total unemployment, $u1$ is the unemployment rate for $y < y^*$ types, $u2$ is the unemployment rate for $y^* < y < y^{**}$ types and $u3$ is the unemployment rate for $y > y^{**}$ types.

Table 7: Varying the Level of Frictions in Wage and Self Employment

θ	0.2253	0.36047	0.39689	0.28277	0.43905	0.476	0.35455	0.53396	0.57886
y^*	0.19068	0.30465	0.35526	0.20425	0.31625	0.36443	0.21804	0.32747	0.37387
y^{**}	0.78947	0.78947	0.78947	0.57223	0.57223	0.57223	0.54126	0.54126	0.54126
α	0.5	0.5	0.5	2.5	2.5	2.5	4.5	4.5	4.5
m	1	2	3	1	2	3	1	2	3
y_w	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$w(0)$	0.32536	0.38048	0.40803	0.33147	0.38661	0.41324	0.3378	0.39264	0.4187
$w(y^*)$	0.34534	0.40232	0.42763	0.35212	0.40812	0.43221	0.35902	0.41373	0.43694
$w(y^{**})$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
avg wage	0.39428	0.42436	0.44095	0.37085	0.4118	0.43245	0.36665	0.41207	0.43376
wmass 1	0.32514	0.44845	0.49732	0.65516	0.74539	0.77577	0.77506	0.83459	0.85392
wmass2	0.67486	0.55155	0.50268	0.34484	0.25461	0.22423	0.22494	0.16541	0.14608
avg self earnings	0.62824	0.71696	0.76048	0.61813	0.68584	0.71516	0.61957	0.6815	0.70791
smass 1	0.65856	0.51134	0.41646	0.42221	0.29332	0.22328	0.38636	0.26505	0.20034
smass2	0.34144	0.48866	0.58354	0.57779	0.70668	0.77672	0.61364	0.73495	0.79966
u	0.40611	0.30497	0.25471	0.22237	0.18752	0.1627	0.17428	0.14589	0.12455
u1	0.513	0.29398	0.20921	0.48461	0.27394	0.19457	0.45644	0.25491	0.1797
u2	0.33906	0.22719	0.17301	0.14157	0.1156	0.098623	0.089358	0.077382	0.068658
u3	0.5	0.5	0.5	0.16667	0.16667	0.16667	0.1	0.1	0.1

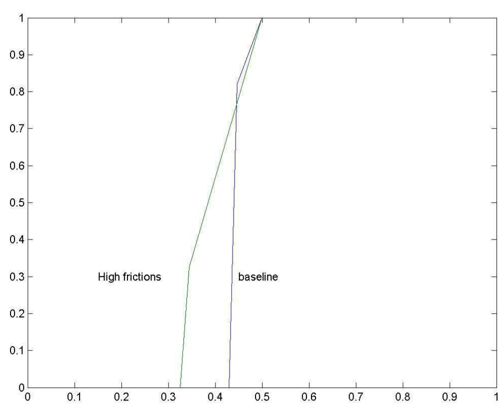
Note: each column represents a model solution, which changes as frictions vary, shown in the forth and fifth rows. wmass1 is the percentage of wage employees who only take wage employment in the equilibrium and wmass2 is the percentage of wage employees who also take self-employment in the equilibrium. smass1 is the percentage of the self-employed who also take wage employment in the equilibrium and smass2 is the percentage of the self-employed who only take self-employment in the equilibrium. u is total unemployment, $u1$ is the unemployment rate for $y < y^*$ types, $u2$ is the unemployment rate for $y^* < y < y^{**}$ types and $u3$ is the unemployment rate for $y > y^{**}$ types.

Figure 3: EARNINGS DISTRIBUTIONS, ILFS



Source: pooled earnings data from Tanzanian ILFS 2001 and 2006. The kernel density was estimated using Stata.

Figure 4: WAGE CDFS FOR BASELINE AND HIGH FRICTIONS MODEL



8 Appendix

8.A Measuring Inflation

Price increases over the period makes comparing incomes at different times more difficult. A consumer price index (CPI) is collected monthly by the Tanzanian National Bureau of Statistics and is shown in Table 8. It indicates there was inflation of around 31 percent over the 5 year period between the two surveys. However some concerns have been raised about its accuracy (National Bureau of Statistics, Tanzania 2007)[9]. It is a Laspeyres index based on the 2000 Household Budget Survey, which means it does not take into account changes in consumption patterns that occurred post-2000. If relative prices or patterns of expenditure have changed over the period then this will not be an accurate measure of how the prices households face have changed (Deaton and Tarozzi 2000).

It is possible to calculate alternative price indices based on the Household Budget Surveys of 2000 and 2007, nationally representative samples of households in these years, which collected prices faced by households and household expenditure. Using these data a Fisher ideal index (which measures price changes over an average consumption basket across the two survey years (Deaton and Tarozzi 2000)) is constructed for both food and non food inflation in the 2007 HBS report (National Bureau of Statistics, Tanzania 2007). This suggests that food inflation was actually nearly 100 percent over this 7 year period, whilst the CPI puts food inflation at 52 percent over the same period. A similar trend is found in non-food inflation: the HBSs indicate inflation was over 100% but the CPI implies it was 29 percent (National Bureau of Statistics, Tanzania 2007). A Fisher ideal index is then calculated using weights for food and non-food inflation of 0.72 and 0.28 respectively, based on the share of food and non-food expenditure for the poorest 25 percent. The resultant index of price increases between 2000 and 2007 is 1.93, compared to 1.47 using the CPI.

Unfortunately the HBS years do not correspond to the years the ILFS was conducted, meaning some interpolation will be required to obtain the estimated price inflation over the period. To do this I assume a constant rate of inflation over the period and I then wish to find a π that satisfies $(1+\pi)^7 = 1.93$. This generates a value of π of 9.8 percent compared to an average yearly inflation figure of roughly 5.5 percent from the CPI. This interpolation means I estimate that the increase in prices between 2001 and 2006 was 60 percent, nearly double the 31 percent calculated from the CPI. Despite the assumption of constant inflation and the weighting of the index using the poorest 25 percent of the population, given the problems with the CPI, I believe this is more reliable than the CPI data from the National Bureau of Statistics and this is what I use to deflate earnings in the surveys I use.

[9] NBS collects its own CPI data but outsourced the 2007 HBS analytical report to a group of UK researchers, which explains how two NBS sources give different numbers for inflation.

8.B The Evolution of the value of the asset

Following Cahuc and Zylberberg (2004) I provide an intuitive explanation for the value of an asset which gives a flow income $\omega(x)$ over a small interval of time dt and which can evolve to a new state with net present value $\bar{\Pi}(t+dt)$ over the small time interval dt with a probability of $\lambda(t)dt$. It is possible to write the value of the asset as

$$\Pi(t) = \frac{1}{1+rdt}(\omega(t) + \lambda(t)dt\bar{\Pi}(t+dt) + (1-\lambda(t)dt)\Pi(t+dt)) \tag{40}$$

Rearranging it is possible to obtain

$$r\Pi(t) = \omega(t) + \lambda(t)(\bar{\Pi}(t+dt) - \Pi(t+dt)) + \frac{\Pi(t+dt)-\Pi(t)}{dt} \tag{41}$$

Letting $t \to 0$ we then have

$$r\Pi(t) = \omega(t) + \lambda(t)(\bar{\Pi}(t) - \Pi(t)) + \dot{\Pi}(t) \tag{42}$$

The time derivative of the asset $\dot{\Pi}(t)$ is assumed to be zero in the asset equation for a job in my model and equation (42) is the basis of the asset equations I use in this paper.

Table 8: CPI as measured by the National Bureau of Statistics

2000	2001	2002	2003	2004	2005	2006	2007
77.9523	81.9648	86.3236	90.9018	95.2067	100	107.251	114.786

Source: International Monetary Fund (2010)

8.C Human Capital Regressions

Table 9: Basic Human Capital Regressions from the 2006 ILFS

	occup1	occup2	occup3	occup4	occup5
	(1)	(2)	(3)	(4)	(5)
male	0.614	0.478	0.435	0.308	0.09
	(0.029)***	(0.08)***	(0.053)***	(0.054)***	(0.038)**
age	0.025	0.021	0.049	0.065	0.069
	(0.012)**	(0.029)	(0.019)**	(0.019)***	(0.017)***
agesq	-.029	-.009	-.042	-.062	-.052
	(0.014)**	(0.034)	(0.024)*	(0.023)***	(0.02)***
years of education	0.029	0.027	0.092	-.002	0.045
	(0.014)**	(0.04)	(0.027)***	(0.025)	(0.025)*
educsq	0.251	0.4	0.006	0.698	0.311
	(0.112)**	(0.24)*	(0.194)	(0.144)***	(0.122)**
D2001	0.022	0.228	0.16	-.089	0.119
	(0.029)	(0.075)***	(0.048)***	(0.05)*	(0.036)***
Obs.	4429	831	875	886	1646
R^2	0.131	0.16	0.193	0.31	0.259
F statistic	110.921	26.227	34.523	65.908	95.572

Notes: dependent variable is log(income) in all columns. *, **, *** denote significance at 10%, 5% and 1% levels.

References

Albrecht, James, Lucas Navarro, and Susan Vroman, "The Effects of Labour Market Policies in an Economy with an Informal Sector," *Economic Journal*, July 2009, *119* (539), 1105–1129.

Badaoui, Eliane, Eric Strobl, and Frank Walsh, "Is There An Informal Sector Wage Penalty? Evidence from South Africa," *Economic Development and Cultural Change*, 2008, *56*, 683–710.

Banerjee, Abhijit V. and Andrew F. Newman, "Occupational Choice and the Process of Development," *Journal of Political Economy*, April 1993, *101* (2), 274–98.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky, "The Nash Bargaining Solution in Economic Modelling," *Rand Journal of Economics*, 1986, *17* (1), 176–188.

Broyden, C G, "A Class of Methods for Solving Nonlinear Simultaneous Equations," *Mathematics of Computation*, 1965, *19* (92), 577–593.

Burdett, Kenneth and Dale T. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, May 1998, *39* (2), 257–73.

Cahuc, Pierre and Andre Zylberberg, *Labor Economics*, The MIT Press, 2004.

Deaton, Angus, *The Analysis of Household Surveys – A Microeconomic Approach to Development Policy*, Baltimore: Johns Hopkins University Press, 1997.

_ and Alessandro Tarozzi, "Prices and poverty in India," Working Paper 213, Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies. July 2000.

Fafchamps, Marcel and Agnes R. Quisumbing, "Marriage, Bequest, and Assortative Matching in Rural Ethiopia," *Economic Development and Cultural Change*, January 2005, *53* (2), 347–80.

_ , Mans Sderbom, and Najy Benhassine, "Wage Gaps and Job Sorting in African Manufacturing," *Journal of African Economies*, November 2009, *18* (5), 824–868.

Falco, Paolo, Andrew Kerr, Neil Rankin, Justin Sandefur, and Francis Teal, "The returns to formality and informality in urban Africa.," 2010. CSAE Working Paper Series, WPS/2010-03.

Fields, Gary S., "Rural-Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDCs," *Journal of Development Economics*, 1975, *2* (2), 165–187.

_ , "Dualism In The Labor Market: A Perspective On The Lewis Model After Half A Century," *Manchester School*, December 2004, *72* (6), 724–735.

Harris, John and Michael Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 1970, *60* (1), 126–142.

Heckman, James J., "Effects of Child-Care Programs on Women's Work Effort," *Journal of Political Economy*, 1974, *82* (2), S136–S163.

_ , "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, August 2001, *109* (4), 673–748.

Heintz, James and Dori Posel, "Revisiting informal employment and segmentation in the South African labour market," *South African Journal of Economics*, 2008, *76* (1), 26–44.

International Monetary Fund, "Internation Financial Statistics April 2010, Series ..DE.ZF," 2010.

Keswell, Malcolm and Laura Poswell, "Returns To Education In South Africa: A Retrospective Sensitivity Analysis Of The Available Evidence," *South African Journal of Economics*, 2004, *72* (4), 834–860.

Maloney, William, "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico," *The World Bank Economic Review*, 1999, *13* (2), 275–302.

_ , "Informality Revisited," *World Development*, July 2004, *32* (7), 1159–1178.

Miranda, Mario and Paul Fackler, *Applied Computational Economics and Finance*, MIT Press, 2002.

Moll, Peter, "Primary Schooling, Cognitive Skills and Wages in South Africa," *Economica*, May 1998, *65* (258), 263–84.

Mortensen, Dale T., *Wage dispersion : why are similar workers paid differently*, Cambridge, MA. London: MIT Press, 2003.

_ **and Christopher A. Pissarides**, "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, July 1994, *61* (3), 397–415.

_ **and** _ , "New developments in models of search in the labor market," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3 of *Handbook of Labor Economics*, Elsevier, April 1999, chapter 39, pp. 2567–2627.

Nash, John, "Two-person cooperative games," *Econometrica*, 1953, *21* (1), 128–140.

National Bureau of Statistics, Tanzania, "Household Budget Survey 2007 Analytical Report," 2007.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, November 2002, *70* (6), 2295–2350.

Rankin, Neil, Justin Sandefur, and Francis Teal, "Learning and earning in Africa: Where are the returns to education high?," 2010. CSAE Working Paper WPS/2010-02.

Roy, A, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, *3* (2), 135–146.

Sandefur, Justin, Pieter Serneels, and Francis Teal, "African Poverty through the Lens of Labor Economics: Earnings and Mobility in Three Countries," in P Paci and P Serneels, eds., *Employment and Shared Growth*, World Bank, 2007.

Satchi, Mathan and Jonathan Temple, "Labor Markets and Productivity in Developing Countries," *Review of Economic Dynamics*, January 2009, *12* (1), 183–204.

Sattinger, Michael, "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, June 1993, *31* (2), 831–80.

Wong, Linda Y., "Can the Mortensen-Pissarides Model with Productivity Changes Explain U.S. Wage Inequality?," *Journal of Labor Economics*, January 2003, *21* (1), 70–105.

Zenou, Yves, "Job search and mobility in developing countries. Theory and policy implications," *Journal of Development Economics*, June 2008, *86* (2), 336–355.