

Lain, Jonathan

**Article**

## Discrimination in a search and matching model with self-employment

IZA Journal of Development and Migration

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Lain, Jonathan (2019) : Discrimination in a search and matching model with self-employment, IZA Journal of Development and Migration, ISSN 2520-1786, Springer, Heidelberg, Vol. 9, Iss. 6, pp. 1-35,  
<https://doi.org/10.1186/s40176-019-0140-3>

This Version is available at:

<https://hdl.handle.net/10419/222179>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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



<https://creativecommons.org/licenses/by/4.0/>

ORIGINAL ARTICLE

Open Access



# Discrimination in a search and matching model with self-employment

Jonathan Lain 

Correspondence:

[jonathan.LAIN@oecd.org](mailto:jonathan.LAIN@oecd.org)

This paper was written prior to my joining the OECD and the views expressed should not be taken as representing the views of the OECD or its member countries. I am also responsible for any errors.

Organisation for Economic Co-operation and Development, 2 Rue André Pascal, 75775 Paris CEDEX 16, France

## Abstract

In many urban African labour markets, women outnumber men in self-employment, even though gender earnings gaps are larger for the self-employed than the wage-employed. In this paper, we suggest an explanation for this pattern using a search and matching model that allows for individual heterogeneity and participation in both self- and wage-employment as well as discrimination against women in the wage sector. We show that discrimination in wage-employment can generate gender earnings gaps across all sectors of the economy, even if the underlying ability distributions for women and men are identical. This result arises because discrimination creates extra frictions for women, making it harder for them to select occupations according to their comparative advantage. Moreover, the earnings gaps that arise turn out to be larger for the self-employed, even though it is not the self-employment sector that is characterised by discrimination. Policy makers may therefore need to target the wage sector in order to address gender inequality for the self-employed.

**JEL classifications:** J46, J60, J71

**Keywords:** Search models, Discrimination, Comparative advantage, Self-employment, Informality, Urban Africa

## 1 Introduction

Female participation in self-employment is prevalent in urban Africa. In urban Ghana, self-employed women outnumber self-employed men by a ratio of nearly 3:1. However, gender earning gaps are far larger in self-employment than in the wage sector. For the self-employed, men's earnings are approximately double those of women at the median, and observable worker characteristics do little to explain this gap. For the wage-employed, however, the earnings differential at the median is around 25%.

There are several potential reasons why these patterns might arise. Self-employment may be characterised by more gender-based discrimination than the wage sector, driving a larger wedge between the earnings of women and men. For example, women may face additional credit constraints when trying to start or expand their businesses or they may struggle to sell their goods to prejudiced consumers. At the same time, there may be non-pecuniary benefits—such as flexible working hours or independence—which attract women to self-employment in spite of these drawbacks. Additionally, there may be productivity differences between women and men that generate gender earnings gaps across all sectors of the economy, but which cannot be easily observed, especially for the self-employed. The variables typically included in Mincer regressions are likely to carry less

explanatory power in self-employment, where entrepreneurial talent, motivation, and other characteristics that are difficult to measure may matter more.

In this paper, we offer an alternative explanation, which assumes no discrimination in the self-employment sector and no productivity differences between women and men. Instead, gender earnings gaps in all sectors of the economy are driven solely by discrimination in the wage sector. Our model adopts a search and matching framework in the spirit of Mortensen and Pissarides (1994)—also known as random matching and bargaining—where unemployed individuals can choose to search for jobs in wage-employment, self-employment, or both, depending on their returns in each sector. Workers are heterogeneous along two dimensions. Firstly, their gender determines whether they will face discrimination in trying to obtain wage sector jobs. However, workers also differ in terms of their *self*-employment productivity. By allowing for search frictions as well as individual heterogeneity, our model captures both segmentation and sorting in the labour market.

The search and matching framework provides an apt arena within which to examine discrimination. Firstly, matching frictions afford firms monopsony power, which they may exploit to satisfy any preferences they have for discrimination. In competitive labour market models, discriminatory outcomes—such as gender earnings gaps—typically cannot persist, because prejudiced firms will be disadvantaged compared to unprejudiced firms when trying to recruit workers (Becker 1957). Secondly, matching models allow us to consider the outcomes of the unemployed, which would not be possible if workers could move freely between wage- and self-employment. Finally, we are able to couch discrimination in terms of the matching function itself, by adding extra friction into the process through which women obtain wage sector jobs compared to men.

In order to guide our modelling approach and verify our simulation results, we present evidence from a panel dataset collected in urban Ghana between 2004 and 2013. This builds on a large empirical literature on informality in urban Africa (Hart 1973; Kingdon et al. 2006; Falco et al. 2011; Fields 2011). Ghana provides a useful context within which to investigate gender differences in occupational selection since the labour force participation of women is relatively high.<sup>1</sup>

Our model is able to explain differences in average earnings between men and women in both wage- and self-employment—in line with the Ghanaian data—without assuming any differences between the innate ability distributions of each gender. The frictions inherent in the job matching process prevent workers from selecting occupations entirely on the basis of comparative advantage, which reduces average productivity in all sectors. With discrimination, these frictions are more severe for women than for men, pushing women even further away from the jobs in which they have comparative advantage. The results therefore echo Roy's (1951) classic model of occupational selection.

Additionally, our model mirrors the observation that gender earnings gaps are larger for the self-employed than the wage-employed. This is, in part, because the earnings of the self-employed are determined directly by their productivity, whereas wage sector earnings are governed by bargaining between workers and firms.

Our model's predictions about the sectoral composition of the economy also match what is observed in the Ghanaian data. As anticipated, wage sector discrimination drives women away from wage-employment, which in turn increases the number of unemployed and self-employed women. However, it also emerges that there are spillover effects

on men, with their participation in wage-employment rising at the expense of self-employment when discrimination is introduced into the model. Discrimination increases men's chances of successfully matching with a firm, because they are insulated from competition for jobs from women, and because firms create more vacancies to ensure they can still operate.

To our knowledge, this is the first attempt to analyse discrimination in a search and matching model that allows for both wage- and self-employment. We build on two key strands of the theoretical literature. Firstly, we complement a growing body of work, which uses search models to assess the impacts that informal sectors or self-employment sectors have on labour market outcomes in developing countries. Our approach draws most directly on the model of Albrecht et al. (2009) and later adapted by Kerr (2012), to allow for self-employment as well as wage-employment in a random matching and bargaining model. Secondly, the paper also draws upon a wide range of existing studies, which use search models to analyse the effects of discrimination on those participating in the labour market. However, unlike previous models, we do not make any additional assumptions about employers', employees', or consumers' preferences, nor do we explicitly examine the conditions under which firms with a 'taste for discrimination' survive. Instead, we treat discrimination as an economy-wide stochastic process that affects the matching function. This is sufficient for testing the effects of wage-sector discrimination across different employment sectors and may be a tenable approach if—as appears to be the case in Ghana—discrimination is prevalent across many firms operating in many industries.

This paper proceeds as follows. In Section 2, we orientate this paper within the related literature. In Section 3, we describe the key properties of the urban Ghanaian labour market. We outline our model and then explain how we solve for the equilibrium in Sections 4 and 5 respectively. In Section 6, we present our main simulation results. In Section 7, we consider the comparative statics, linking the model to policy. In Section 8, we conclude.

## 2 Related literature and contribution

Search models have become a key tool in the analysis of labour markets. Building the notion that it takes time and potentially other resources for workers to find jobs and for employers to fill vacancies into a coherent framework has enabled economists to address many important research questions, providing key insights into what causes certain individuals to choose to remain unemployed, how unemployment and unfilled vacancies can coexist, and why wages may vary among homogeneous workers (Rogerson et al. 2005). This paper sits within two specific strands of the search literature: (1) the use of search models to better understand informality and self-employment and (2) the use of search models to examine labour market discrimination.

### 2.1 Search models, informality, and self-employment

Typically, search models focus on workers finding formal wage-employment jobs, either from unemployment or from an existing wage job through on-the-job search. However, only allowing for formal wage-employment is insufficient for modelling labour markets in developing countries, where informality and self-employment are prevalent. A number of recent papers have attempted to address this shortcoming.

Some economists have used wage-posting models in the spirit of Burdett and Mortensen (1998)—where employers post the wages associated with particular vacancies—to model informality. Narita (2012), for example, constructs a model in which workers can match with either formal- or informal-sector firms, as well as working in self-employment (or remaining unemployed).<sup>2</sup> At the same time, firms are able to choose between the obligation to pay taxes in the formal sector and the risk of being fined in the informal sector. Narita's model is also dynamic, in the sense that workers have the potential to learn about good business opportunities throughout their life cycle. When estimated using Brazilian data, the model not only successfully reproduces the composition of the Brazilian labour force, but also demonstrates that the chances of transitioning into self-employment are significantly higher for more experienced workers. While work experience is not an aspect of heterogeneity considered in this paper and we do not use the wage-posting framework, Narita's model is a helpful counterpoint to the class of models on which we primarily draw.

More commonly, Mortensen and Pissarides' (1994) random matching and bargaining framework has been used as the basis for search models that allow for self-employment or an informal sector. In one example, Saatchi and Temple (2009) build a model with an urban formal sector—in which firms and workers randomly match and then bargain over wages—that coexists with an urban informal sector—in which workers are self-employed and earn some fixed level of output. They also allow for an agricultural sector, from which potential workers can migrate. When calibrated to Mexican data, Saatchi and Temple's model produces a reasonably sized informal sector in the presence of formal sector frictions. Given its relative simplicity, the model can also be used to examine analytically the effects of shocks to sector-specific productivity, worker bargaining power, and tax policies, without requiring numerical methods. However, Saatchi and Temple's model does not allow for worker heterogeneity: this is a vital element of our modelling approach, since (1) we want to allow for the possibility that only some workers face discrimination and (2) we are interested in the resulting distribution of worker productivity and earnings in different sectors of the economy.

Our paper therefore draws most directly on the model by Albrecht et al. (2009) and its adaptation by Kerr (2012), which allows for both informality and worker heterogeneity within a random matching and bargaining framework. In Albrecht et al.'s (2009) paper, unemployed individuals may transition into the formal sector—where productivity is heterogeneous and subject to shocks that lead matches to end or wages to be renegotiated—or the informal sector—where productivity and earnings are homogeneous. Noting that earnings dispersion is higher in the informal sector in African labour markets, Kerr (2012) modifies Albrecht et al.'s (2009) model by assuming that workers are heterogeneous in their *informal* sector productivity instead, with earnings dispersion arising in both sectors because workers' outside options (their informal sector welfare) during the formal sector wage bargain differ. This paper develops these models further by incorporating a further element of worker heterogeneity: their exposure to discrimination.

## 2.2 Search models and discrimination

Competitive labour market models, in which there are no frictions associated with the coming together of employers and employees, often predict that discriminatory outcomes

cannot persist in equilibrium. As Becker (1957) describes, in a simple model in which there are two types of workers ('Type-As' and 'Type-Bs') and some subset of prejudiced firms that incur a utility cost from employing Type-Bs, there are two possibilities, neither of which lead to Type-Bs faring worse than Type-As. If the number of prejudiced firms is initially low, the market instantaneously and fully segregates, such that Type-Bs only work for unprejudiced firms. Even if the share of prejudiced firms is initially large, any Type-Bs working at prejudiced firms for lower wages can be attracted by expanding unprejudiced firms, who offer higher wages to exploit the arbitrage opportunity: prejudiced firms go out of business until these arbitrage opportunities no longer exist, and the market fully segregates.

As such, to allow for the possibility that discriminatory outcomes persist in equilibrium, one of the assumptions underpinning the competitive labour market model is typically dropped. For example, in models of 'statistical discrimination,' the assumption that employers have perfect information about the true underlying ability of potential employees is relaxed. Employers instead base their hiring decisions on observable characteristics, which they believe proxy for ability, such as gender or race.<sup>3</sup> Search and matching models, by contrast, depart from the competitive labour market model by adding frictions into the process through which employers and employees meet: this affords employers monopsony power and allows discriminatory outcomes to persist in equilibrium.

A number of papers—typically using the wage-posting framework—explore the specific conditions under which discriminatory outcomes can persist in search models in equilibrium. Black (1995), for example, constructs a model in which firms are heterogeneous in terms of both their entrepreneurial ability and their 'taste for discrimination,' where unprejudiced firms are willing to employ both Type-A and Type-B workers, but prejudiced firms will only employ Type-Bs.<sup>4</sup> In Black's model, it is only those prejudiced firms with higher entrepreneurial ability that are sufficiently competitive to survive while indulging their taste for discrimination. Interestingly, even though only unprejudiced firms will actually employ Type-Bs, such firms still exploit the fact that Type-B workers have worse outside options when posting wages, driving Type-dependent earnings gaps in the model. In a related model, Sasaki (1999) shows that prejudice among potential co-workers—meaning that Type-A workers prefer to work with other Type-As rather than Type-Bs—can also drive discriminatory outcomes in equilibrium, as firms have to compensate Type-As for working alongside Type-Bs. Sasaki also shows that an increased taste for discrimination among Type-As (the extent of their unwillingness to work with Type-Bs) increases their wages. As such, Type-As have an incentive to ensure discrimination persists.

As in this paper, some economists have even tried to consider what discrimination might mean in search models that explicitly capture self-employment. Borjas and Bronars (1989), for example, build a model in which certain consumers derive dis-utility from buying goods from Type-B self-employed sellers (but not Type-As). However, since matches are random, consumers initially have imperfect information about the Type and price 'quoted' by self-employed sellers, and contacting other sellers is costly. This not only drives down the average earnings of self-employed Type-Bs, but also lowers the incentives for high-ability Type-Bs to become self-employed in the first place. However, while Borjas and Bronars' (1989) model is crucial for understanding discrimination within self-employment,

the approach taken in this paper is different, insofar as we focus on the effects that *wage*-sector discrimination has the self-employed.

This paper is closest to a series of studies that use a random matching and bargaining framework to try and empirically estimate the extent to which discrimination may drive race or gender earnings gaps observed in the labour market. In one example, Flabbi (2010) constructs a model in which certain firms incur some heterogeneous flow of dis-utility from hiring Type-Bs, while for workers, productivity, job arrival rates, and termination rates are all Type-specific.<sup>5</sup> Estimating the model to understand gender labour market inequality in the USA, Flabbi shows that as many as half of firms may be prejudiced and that discrimination accounts for around two thirds of the gender earnings gap.

To our knowledge, however, no existing search models that incorporate discrimination also allow for both self- and wage-employment. This constitutes one key contribution that this paper aims to make. Additionally, rather than assuming firms are heterogeneous like much of the existing literature, we characterise discrimination using a simple stochastic process that directly affects the matching function. This appears to be a sufficient and tenable approach for understanding how discrimination in the wage sector affects the self-employed.

### 3 Data: evidence from Ghana

Between 2004 and 2013, the Centre for the Study of African Economies (CSAE) collected longitudinal data from a sample of approximately 3000 individuals in four of Ghana's largest cities: Accra, Kumasi, Cape Coast, and Takoradi. This dataset—known as the Ghana Household Urban Panel Survey (GHUPS)—allows us to generate a set of stylised facts describing the labour market outcomes of women and men in urban Ghana.<sup>6</sup>

*Firstly*, as Table 1 shows, women substantially outnumber men in self-employment, while the converse is true in wage-employment. In the GHUPS, workers are classified as self-employed if (1) they have principal control over their income-generating activity and (2) they use their own tools or other physical capital, if any are used at all. There are approximately three times as many self-employed women as there are self-employed men, whereas in the wage sector there are nearly the double the number of men compared to women.

*Secondly*, while gender earnings gaps persist in both self- and wage-employment, the differential is larger for the self-employed. The measurement of earnings in self- and

**Table 1** Occupational breakdown by gender

Occupation	Gender					
	Female		Male		Total	
	No.	%	No.	%	No.	%
Apprentice	315	2.34	305	2.26	620	4.60
Out of the labour force	2325	17.23	1543	11.44	3868	28.67
Private sector wage-employed (large firm)	313	2.32	734	5.44	1047	7.76
Private sector wage-employed (small firm)	835	6.19	1200	8.90	2035	15.09
Public sector wage-employed	217	1.61	241	1.79	458	3.40
Self-employed	2937	21.77	1132	8.39	4069	30.16
Unemployed	840	6.23	553	4.10	1393	10.33
Total	7782	57.69	5708	42.31	13490	100.00

Pooled Sample of working age individuals (15–65) for the years 2005–2013

Large firms have >20 employees

**Table 2** Median weekly earnings for the self- and wage-employed 2010–2013 (2010 Ghana Cedis)

	2010		2012		2013	
	Women	Men	Women	Men	Women	Men
Private sector wage-employed (small firm)	15.44 (29.35)	25.73 (73.28)	20.51 (45.45)	38.73 (108.63)	31.06 (78.06)	49.90 (112.52)
Private sector wage-employed (large firm)	32.24 (81.10)	42.70 (63.97)	56.33 (101.50)	48.76 (83.11)	65.32 (86.75)	67.23 (109.16)
Public sector wage-employed	80.31 (94.51)	65.03 (53.06)	83.44 (131.34)	90.88 (84.25)	92.91 (130.77)	102.78 (110.90)
All wage-employed	23.16 (67.94)	36.14 (69.05)	35.66 (91.17)	46.99 (99.34)	43.48 (93.25)	56.62 (111.99)
Self-employed	23.95 (92.06)	54.43 (159.56)	31.20 (116.02)	76.46 (140.14)	49.40 (131.09)	92.98 (168.64)
Observations	687	463	896	710	852	696

Sample of individuals of working age (15–65)

Standard deviations in parentheses

Earnings deflated to 2010 Ghana Cedis and adjusted to a standard working week using data on hours worked

Outliers trimmed at the 1st and 99th percentiles

Large firms have >20 employees

wage-employment in the GHUPS is discussed in Appendix 2. As Table 2 shows, in 2013, men earned approximately 30% more than women at the median in the wage sector. In contrast, the median earnings of self-employed men were nearly double those of self-employed women.

Moreover, the gender earnings gap for the self-employed is less well explained by observable differences in human capital and individual characteristics than the gender earnings gap for the wage-employed. By regressing self- and wage-employment earnings on gender with and without a set of controls including education and age, we can decompose the raw gender earnings gap into ‘explained’ factors—that is, differences between women and men in terms of observable human capital—and ‘unexplained’ factors—or differences between the returns to a given level of human capital, in the spirit of Blinder and Oaxaca (Blinder 1973; Oaxaca 1973). The coefficient on gender in the regressions without controls can be interpreted directly as the raw gender earnings gap, the analogous coefficient in the regressions with controls comprises the ‘unexplained’ portion, while the difference between the two is the ‘explained’ portion (Elder et al. 2010). The results, shown in Appendix 3, demonstrate that around 20% of the gender earnings gap in self-employment can be attributed to differences in observable human capital, compared with 27 percent for the wage-employed.

It is also important to consider how the raw gender earnings gaps that arise in both self- and wage-employment may be influenced by unobservable drivers of occupational selection, alongside the observable measures of human capital that are included in the earnings equations in Appendix 3. This helps to characterise what the unexplained components of the gender earnings gap might be. To assess the impact of such unobservable selection on the results, we can estimate a multinomial logit model for individuals’ choice between the main five occupational statuses in the GHUPS—(1) self-employment, (2) wage-employment, (3) out of the labour force, (4) unemployment (not working but searching), and (5) being an apprentice—and then calculate selection correction terms based on the method described by Dubin and McFadden (1984).<sup>7</sup> In effect, Dubin and McFadden’s approach applies the insights of the Heckman model for sample selection to



a setting with multiple categories (Heckman 1979). As such, the model relies on finding ‘exclusion restrictions’ or variables that can be added to the multinomial logit but not the final earnings equation, which can plausibly influence occupational selection without directly influencing earnings. In this analysis, the exclusion restrictions comprise variables that capture sources of unearned income and parental characteristics.<sup>8</sup>

Using the Heckman-type correction methods described above suggests that both observed and unobserved occupational selection drive the gender earnings gaps seen in the data. In wage-employment, the point estimates for the raw, unexplained, and explained gender earnings gaps all drop to less than 5% when the Dubin and McFadden selection correction terms are included. We can also reject the null hypothesis that the coefficient on the gender dummy is equal in the models with and without the selection correction terms. In self-employment, the picture is somewhat more complex. Even though the point estimates for the raw, unexplained, and explained gender earnings gaps fall substantially when correcting for unobservable selection, the standard errors for the selection-corrected model with controls in the earnings equation are relatively high. As such, we are unable to reject the null hypothesis that the unexplained gender earnings gap is equivalent in the model, which has controls, but which does not have the selection correction terms, and the model which includes both controls and the selection correction terms.<sup>9</sup>

Since adjusting for unobservable selection appears to reduce the gender earnings gap in both self- and wage-employment, it may be that women are ‘negatively’ selecting into both of these sectors: this helps to characterise the unobservable drivers of selection. The data are consistent with a story where women, to a greater extent than men, are unable to select into the occupation in which they have comparative advantage and therefore in which their relative productivity and earnings would be highest. This paper suggests discrimination as a potential source of additional friction that prevents women from selecting



**Table 3** Transition matrix for women (2012–2013)

2012 Occupation	2013 Occupation				
	Out of the labour force %	Self-employed %	Unemployed %	Wage-employed %	Total %
Out of the labour force	20.52	3.90	5.64	4.55	34.61
Self-employed	3.90	25.58	2.24	2.82	34.54
Unemployed	3.47	2.24	3.03	1.59	10.33
Wage-employed	3.25	2.89	2.24	12.14	20.52
Total	31.14	34.61	13.15	21.10	100.00

Sample of working age women (15–65) present in the 2012 and 2013 waves of the GHUPS

occupations according to their comparative advantage. This is not to rule out other drivers of selection. For example, there exists a wide literature documenting the impact of family structure on women's labour market engagement—including a paper by Heath (2017) that uses the same GHUPS data as this paper—which could explain why women select into occupations that do not necessarily maximise their earnings.<sup>10</sup> The aim instead is to present wage-sector discrimination as a mechanism that appears plausible given how women and men are distributed across different occupations and given what they earn in those occupations.

*Thirdly*, variation in earnings is greater for the self-employed than for the wage-employed. Much of this extra dispersion appears to be among relatively low-return activities, as shown by the long left tail on the self-employment earnings distribution in Fig. 1. However, there is substantial variation in the earnings of women and men in both occupations.

*Fourthly*, as Tables 3 and 4 show, transitions between self- and wage-employment occur for both women and men. Approximately 14% of women and 10% of men engaged in wage-employment in 2012 were engaged in self-employment by 2013. Conversely, around 8% of women and nearly a quarter of men engaged in self-employment in 2012 had wage jobs by 2013. As such, individuals do not specialise fully in either self- or wage-employment: some will end up working in both sectors during their life cycle.

The model that follows aims to explain these disparities between the labour market outcomes of women and men, without assuming underlying gender differences in terms of productivity and without introducing additional constraints on women—such as lack of access to credit—directly into the self-employment sector.

**Table 4** Transition matrix for men (2012–2013)

2012 Occupation	2013 Occupation				
	Out of the labour force %	Self-employed %	Unemployed %	Wage-employed %	Total %
Out of the labour force	20.10	1.11	5.93	6.33	33.47
Self-employed	0.70	11.46	0.80	4.12	17.09
Unemployed	2.51	0.90	2.01	2.71	8.14
Wage-employed	4.12	4.32	3.32	29.55	41.31
Total	27.44	17.79	12.06	42.71	100.00

Sample of working age men (15–65) present in the 2012 and 2013 waves of the GHUPS

## 4 Model

### 4.1 Background

The population of workers, of size  $N$ , is divided into two, such that some proportion  $\pi^A$  are Type-As (men) and some proportion  $\pi^B$  are Type-Bs (women). The latter are affected by discrimination. We assume that  $\pi^A = \pi$  and  $\pi^B = (1 - \pi)$ . Individuals are also heterogeneous in their self-employment ability,  $y \sim \text{Uniform}(0, 1)$ . This ability distribution is the same for Type-As and Type-Bs.

To model discrimination, we assume that Type-Bs face the possibility of some extra shock that immediately destroys a match with a particular firm. This shock hits potential matches with probability  $\lambda$ . By modelling discrimination as an exogenous and stochastic feature of the matching process, we can maintain the assumption that firms are ex ante homogeneous, such that there is no inherent heterogeneity in firms' preferences.

Our approach is therefore a departure from the models of Borjas and Bronars (1989), Black (1995), Sasaki (1999), and other search and matching models that allow for discrimination, as we do not attempt to endogenise the process through which prejudiced firms survive. The plausibility of this setup hinges on discrimination being widespread and similar in intensity across employers. Intuitively,  $\lambda$  can be understood as the proportion of 'interviewers' at each firm that are prejudiced and therefore unwilling to employ Type-B individuals. Direct evidence on this issue is somewhat scarce. Bowlus and Eckstein (2002) show that more than half of firms (56 percent) in the USA have a 'dis-utility factor' associated with employing black workers, but it is not clear these findings translate to gender discrimination in urban Ghana. Nevertheless, descriptive evidence indicates that gender gaps in terms of access to formal jobs and earnings persist across almost all industries in the Ghanaian economy (Baah-Boateng 2012). This suggests that  $\lambda > 0$  for at least some firms in each industry. Thus, although treating firms as homogenous is a strong assumption, modelling discrimination as an exogenous process in this way provides a tenable basis for our model.

### 4.2 Worker value functions

Both types of individuals can work in either wage- or self-employment, queuing for jobs in each sector in unemployment. Following Mortensen and Pissarides (1994), workers are assumed to be risk neutral, have infinite lifespans, and discount the future at rate  $r$ .

In self-employment, individuals earn their marginal product determined directly by their ability  $y$ .<sup>11</sup> They also face some exogenous possibility that their self-employment activity will end, which occurs with probability  $q_s$ . The flow value of self-employment for workers of Type- $k = A, B$ , denoted  $rV_s^k(y)$ , can be written:

$$rV_s^k(y) = y + q_s \left( V_u^k(y) - V_s^k(y) \right) \quad (1)$$

The asset equation for unemployment is determined by some flow of unearned income,  $b$ , as well as the options of entering self- and wage-employment. The probability of a self-employment opportunity arriving is exogenously determined by the parameter  $\alpha$ . However, the probability of obtaining a wage-employment opportunity is derived endogenously from the matching function,  $m^k(\cdot)$ , which itself depends on labour market tightness,  $\theta$ . We write the flow value of unemployment for each type,  $rV_u^k(y)$ , as:

$$rV_u^k(y) = b + \alpha \max \left[ \left( V_s^k(y) - V_u^k(y) \right), 0 \right] + m^k(\theta) \max \left[ \left( V_e^k(y) - V_u^k(y) \right), 0 \right] \quad (2)$$

Tightness in the labour market is dependent on the total number of unemployed Type-As and Type-Bs, labelled  $U^A$  and  $U^B$  respectively, as well as the number of vacancies,  $v$ .

$$\theta = \frac{v}{U^A + U^B} \quad (3)$$

Although the underlying matching function  $m(\theta)$  is the same for both types, the probability that Type-Bs actually secure a wage job is adjusted for the fact that some matches are immediately destroyed due to discrimination by a factor of  $(1 - \lambda)$ .

$$m^k(\theta) = \begin{cases} m(\theta) & \text{if } k = A \\ m(\theta)(1 - \lambda) & \text{if } k = B \end{cases} \quad (4)$$

Finally, the flow value of wage-employment for both types is determined by the bargained wage,  $w^k(y)$ , as well as the possibility that a match ends. Separations are exogenous, occurring with some probability  $q_e$  each period.

$$rV_e^k(y) = w^k(y) + q_e (V_u^k(y) - V_e^k(y)) \quad (5)$$

Despite what the transition matrices for urban Ghana show, we do not allow for direct transitions between self- and wage-employment, as doing so adds substantial complication to the model. Moreover such transitions are still captured by workers ability to move between self- and wage-employed jobs through unemployment.<sup>12</sup>

### 4.3 Firm value functions

Firms, assumed to be risk neutral, can employ either one or zero workers. The value of a filled job will differ, depending on whether firms successfully match with a Type-A or a Type-B worker. Productivity in a filled job,  $z_0$ , is assumed to be homogeneous.<sup>13</sup> As such, the value of a match for Type-As and Type-Bs only differs because the wage function is dependent on type. This arises because Type-Bs' outside options are reduced by the extra possibility that matches simply terminate immediately after formation, weakening their wage bargaining position. The flow value of matching with a Type- $k = A, B$  worker with ability  $y$ , which we label  $r\Pi_e^k(y)$ , can therefore be written:

$$r\Pi_e^k(y) = z_0 - w^k(y) + q_e (\Pi_v - \Pi_e^k(y)) \quad (6)$$

The probability of matching with a Type-A or a Type-B worker is determined by the proportion of Type-As and Type-Bs among the unemployed. This is captured by the  $\frac{U^A}{U^A + U^B}$  and  $\frac{U^B}{U^A + U^B}$  terms below. Given the per period cost of advertising a job,  $c$ , the flow value of posting a vacancy,  $r\Pi_v$ , is:

$$r\Pi_v = -c + \frac{m^A(\theta)}{\theta} \frac{U^A}{U^A + U^B} E \left\{ \max \left[ \left( \Pi_e^A(y) - \Pi_v \right), 0 \right] \right\} \\ + \frac{m^B(\theta)}{\theta} \frac{U^B}{U^A + U^B} E \left\{ \max \left[ \left( \Pi_e^B(y) - \Pi_v \right), 0 \right] \right\} \quad (7)$$

### 4.4 Wage determination

Wages are determined through a Nash bargain, which divides up the surplus derived from a successful match. Bargaining power is parameterised by  $\gamma$ , such that wages are negotiated to solve:

$$\max_{w^k(y)} \left[ V_e^k(y) - V_u^k(y) \right]^\gamma \left[ \Pi_e^k(y) - \Pi_v \right]^{(1-\gamma)} \quad \forall k = A, B \quad (8)$$

Assuming that the equilibrium will be characterised by free-entry of firms, such that the value of posting a vacancy,  $\Pi_v$ , is driven to zero, Eq. (8) can be solved for:

$$w^k(y) = \gamma z_0 + (1 - \gamma)rV_u^k(y) \forall k = A, B \quad (9)$$

The differences in wages between Type-As and Type-Bs enter through the unemployment term,  $V_u^k(y)$ . Conditional on  $y$ , this is lower for Type-Bs because their probability of a successful match that takes them out of unemployment into wage-employment is reduced by discrimination.

## 5 Solution

The equilibrium can be characterised by reducing the many equations in terms of many endogenous variables specified above, into just *one* equation in terms of *one* endogenous variable, namely  $\theta$ . Four conditions define the equilibrium: (1) firms enter freely, such that the value of maintaining a vacancy is zero, (2) matches in the wage sector are consummated if and only if it is in the interests of worker and firm to do so, (3) steady state flows into and out of unemployment, self-employment, and wage-employment are equal, for each type, and (4) individuals only take jobs—in either self- or wage-employment—that are worth their while, given their type and their ability  $y$ . To find the equilibrium, we first use Condition (4) to derive some cut-off levels of productivity  $y$ , above and below which workers will forego work in either wage- or self-employment. Secondly, we use Condition (3) to derive the steady-state level of unemployment in the model. Finally, we bring these components together by using Condition (1) and rewriting the asset equation for posting a vacancy.

### 5.1 Productivity cut-offs

As in Roy's (1951) classic model of occupational choice, the jobs that individuals are willing to take depend on their self-employment productivity,  $y$ , as well as their type,  $k = A, B$ . Defining  $X_e^k(y) \equiv V_e^k(y) - V_u^k(y)$ , workers will take wage-employment only if its flow value is greater than unemployment, such that  $X_e^k(y) > 0$ . Similarly, defining  $X_s^k(y) \equiv V_s^k(y) - V_u^k(y)$ , workers will participate in self-employment only if  $X_s^k(y) > 0$ . This implies there may be some type-specific cut-off values of  $y$ , where  $X_e^k(y) = 0$  and  $X_s^k(y) = 0$ , above and below which certain jobs will not be taken.

In order to interpret these cut-offs, we need to know how workers value wage- and self-employment differently—relative to unemployment—according to their productivity. That is, we need to understand how  $X_e^k(y)$  and  $X_s^k(y)$  change in  $y$ . Rearranging the worker value functions along with the wage schedule and differentiating, we can express  $\frac{\partial X_e^k(y)}{\partial y}$  in terms of  $\frac{\partial X_s^k(y)}{\partial y}$ , showing that these two differentials have (weakly) opposite signs.

$$\left( r + q_e + \gamma \mathbb{1}_{[X_e^k(y) > 0]} m^k(\theta) \right) \frac{\partial X_e^k(y)}{\partial y} = -\mathbb{1}_{[X_s^k(y) > 0]} \alpha \gamma \frac{\partial X_s^k(y)}{\partial y} \quad (10)$$

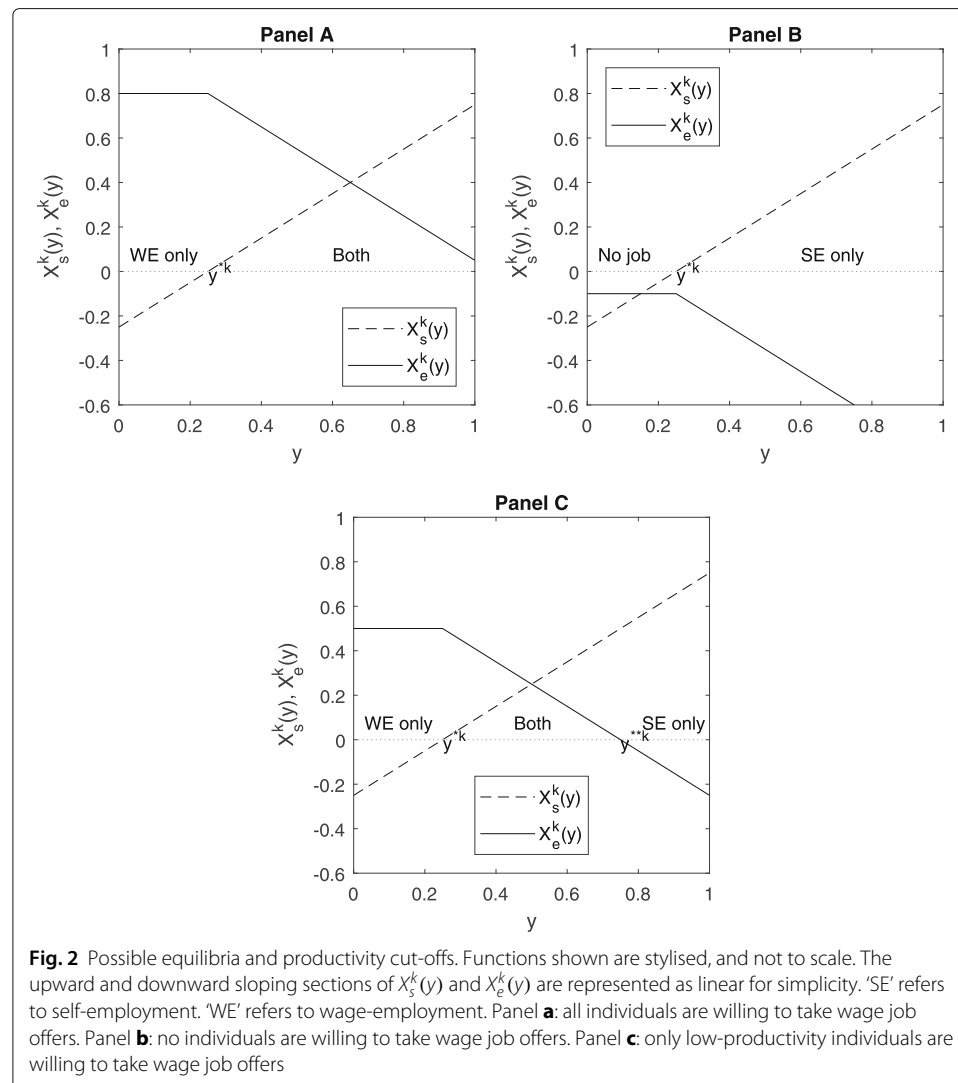
It can further be shown that  $\frac{\partial X_s^k(y)}{\partial y}$  is always positive, while the sign of  $\frac{\partial X_e^k(y)}{\partial y}$  depends on whether  $X_s^k(y) > 0$  or  $X_s^k(y) \leq 0$ .

$$\frac{\partial X_e^k(y)}{\partial y} \begin{cases} < 0 & \text{if } X_s^k(y) > 0 \\ = 0 & \text{if } X_s^k(y) \leq 0 \end{cases} \quad (11)$$

At the first potential cut-off where  $X_s^k(y) = 0$ , unemployed individuals are just indifferent between accepting and rejecting a self-employment job: we label the level of productivity at which this occurs  $y^{*k}$ . Since  $X_s^k(y)$  slopes upwards, unemployment is more valuable than self-employment when productivity is below  $y^{*k}$ . This means workers with  $y < y^{*k}$  would never accept self-employment opportunities. Intuitively, if  $y$  is low, the returns to self-employment are insufficient to tempt workers away from queueing for wage jobs, where they could achieve higher earnings in the future.<sup>14</sup>

There is also a potential cut-off—which we label  $y^{**k}$ —where  $X_e^k(y) = 0$  and unemployed workers are just indifferent between accepting and rejecting wage job offers. Since  $X_e^k(y)$  is weakly downward sloping, unemployment is more valuable than wage-employment when productivity is above  $y^{**k}$ . Workers with  $y > y^{**k}$  will never accept wage-employment offers, as they are better off bidding their time in unemployment and waiting for a self-employment job, in which they have high productivity and therefore greater earnings.

Providing the lower productivity cut-off  $y^{*k}$  exists, which is guaranteed under relatively mild assumptions about  $b$ , three types of equilibria may prevail (see Fig. 2). This depends



**Fig. 2** Possible equilibria and productivity cut-offs. Functions shown are stylised, and not to scale. The upward and downward sloping sections of  $X_s^k(y)$  and  $X_e^k(y)$  are represented as linear for simplicity. 'SE' refers to self-employment. 'WE' refers to wage-employment. Panel **a**: all individuals are willing to take wage job offers. Panel **b**: no individuals are willing to take wage job offers. Panel **c**: only low-productivity individuals are willing to take wage job offers

on the upper productivity cut-off  $y^{**k}$ . The existence of both productivity cut-offs is discussed in Appendix 4.

When  $z_0$  is large relative to  $b$ , all individuals are willing to take wage job offers, so the upper cut-off at  $y^{**k}$  does not exist (Panel A). In this case, the lower cut-off at  $y^{*k}$  simply divides individuals into those that would also be willing to work in self-employment ( $y \geq y^{*k}$ ) and those that would not ( $y < y^{*k}$ ).

If, however,  $b$  is large relative to  $z_0$ , the flow value of unemployment is so high that no individuals are willing to take wage job offers, again meaning the upper cut-off at  $y^{**k}$  does not exist (Panel B). In this type of equilibrium, individuals with low productivity levels ( $y < y^{*k}$ ) never exit unemployment, because the returns to self-employment are insufficient to tempt them away from simply receiving  $b$  each period. However, individuals with higher productivity ( $y \geq y^{*k}$ ) enter self-employment if such jobs arrive.

In the final case, where both cut-offs  $y^{*k}$  and  $y^{**k}$  exist, individuals can be divided into three groups (Panel C). Individuals with very low productivity only ever accept opportunities in wage-employment, while individuals with very high productivity accept only self-employment opportunities. However, individuals with  $y^{*k} \leq y < y^{**k}$ , accept both self- and wage-employment jobs. These individuals do not fully specialise in either wage- or self-employment and may transition (via unemployment) between different sectors throughout their lifetime. In the analysis that follows in Section 6, we restrict the values of  $z_0$  and  $b$  to ensure the existence of both productivity cut-offs, such that an equilibrium of this type prevails.

Using the worker value functions and the wage schedule to solve for  $y^{*k}$  in  $X_s^k(y^{*k}) = 0$ , the lower cut-off can be expressed in terms of the model's exogenous parameters and the matching function:

$$y^{*k} = \frac{b(r + q_e) + \gamma m^k(\theta)z_0}{(r + q_e + \gamma m^k(\theta))} \quad (12)$$

The worker value functions and the wage schedule can also be used to solve for  $y^{**k}$  in  $X_e^k(y^{**k}) = 0$ . Since the  $m^k(\theta)$  and  $w^k(y)$  terms are eliminated,  $y^{**k}$  can be expressed as a function of the exogenous parameters of the model, meaning that  $y^{**A} = y^{**B}$ , and the upper cut-off is unaffected by type. This is because workers with high self-employment productivity are not affected by discrimination in the wage sector, even if they are Type-Bs, as they forego wage-employment opportunities anyway.

$$y^{**k} = \frac{z_0(\alpha + q_s + r) - b(r + q_s)}{\alpha} \quad (13)$$

## 5.2 Steady-state employment flows

The flows between unemployment, self-employment, and wage-employment, for Type- $k = A, B$  workers depend on their productivity relative to the cut-offs,  $y^{*k}$  and  $y^{**k}$ . Individuals with low productivity ( $y < y^{*k}$ ) will never take self-employment jobs, so the only relevant flow for them is between wage-employment and unemployment. By contrast, individuals with high productivity ( $y > y^{**k}$ ) will never take wage-employment, so they only move between self-employment and unemployment. It is only individuals with middling ability ( $y^{*k} \leq y < y^{**k}$ ) that are willing to become both wage- and self-employed, and therefore flow from unemployment to self- and wage-employment (and back). In a steady state, these employment flows must balance so that the proportion

of time that a Type- $k$ , ability  $y$  individual is unemployed ( $u^k(y)$ ), self-employed ( $n_s^k(y)$ ), or wage-employed ( $n_e^k(y)$ ) remains unchanged and adds up to 1. The proportion of time spent by Type- $k$  individuals with ability  $y$  in unemployment, self-employment, and wage-employment can be written:

$$u^k(y) = \begin{cases} \frac{q_e}{m^k(\theta) + q_e} & \text{if } y < y^{*k} \\ \frac{q_e}{m^k(\theta) + q_e + \alpha \frac{q_e}{q_s}} & \text{if } y^{*k} \leq y < y^{**k} \\ \frac{q_s}{\alpha + q_s} & \text{if } y \geq y^{**k} \end{cases} \quad (14)$$

$$n_s^k(y) = \begin{cases} 0 & \text{if } y < y^{*k} \\ \frac{\alpha \frac{q_e}{q_s}}{m^k(\theta) + q_e + \alpha \frac{q_e}{q_s}} & \text{if } y^{*k} \leq y < y^{**k} \\ \frac{\alpha}{\alpha + q_s} & \text{if } y \geq y^{**k} \end{cases} \quad (15)$$

$$n_e^k(y) = \begin{cases} \frac{m^k(\theta)}{m^k(\theta) + q_e} & \text{if } y < y^{*k} \\ \frac{m^k(\theta)}{m^k(\theta) + q_e + \alpha \frac{q_e}{q_s}} & \text{if } y^{*k} \leq y < y^{**k} \\ 0 & \text{if } y \geq y^{**k} \end{cases} \quad (16)$$

The total number of unemployed Type- $k$ s, which is needed to pin down  $\theta$ , can be written as the product of the *average* proportion of time that a Type- $k$  individual spends in unemployment ( $E[u^k(y)]$ ) and the total number of Type- $k$ s in the population<sup>15</sup>. The total number of unemployed individuals ( $U$ ) is then simply the sum of  $U^A$  and  $U^B$ . The Type-specific numbers of wage-employed and self-employed people ( $N_e^k$ ) and ( $N_s^k$ ) and the total numbers of wage-employed and self-employed people ( $N_e$ ) and ( $N_s$ ) can be calculated in the same way.

$$U^k = N\pi^k E[u^k(y)] = N\pi^k \left[ \int_0^{y^{*k}} u^k(y)f(y)dy + \int_{y^{*k}}^{y^{**k}} u^k(y)f(y)dy + \int_{y^{**k}}^1 u^k(y)f(y)dy \right] \quad (17)$$

### 5.3 The free-entry condition

To find the model's equilibrium, we write a single equation in which the only unknown is labour market tightness,  $\theta$ . We do this by rewriting the asset equation for posting a vacancy in Eq. (7), assuming that the free-entry of firms drives  $\Pi_v$  down to 0, and incorporating all the information on occupational choice and steady-state employment flows from Sections 5.1 and 5.2. Setting  $\Pi_v = 0$  and noting that the max operators are redundant because firms will not accept matches that do not provide them with a positive profit, we can write:

$$c = \frac{m^A(\theta)}{\theta} \frac{U^A}{U^A + U^B} E[\Pi_e^A(y)] + \frac{m^B(\theta)}{\theta} \frac{U^B}{U^A + U^B} E[\Pi_e^B(y)] \quad (18)$$

The expectations operators in Eq. (18) are taken over all the possible values of  $y$ , separating out Type-As and Type-Bs, with which a firm could match. Since firms can only match with unemployed workers, we apply Bayes' Law to adjust the productivity distribution for the entire population<sup>16</sup>. The underlying productivity distribution for the full population,  $f(y)$ , is the same for Type-As and Type-Bs, but the productivity distribution for



unemployed individuals,  $h_u^k(y)$ , differs according to type, because Type-As and Type-Bs do not have equal chances of being in unemployment.

$$h_u^k(y) = \frac{u^k(y)f(y)}{(U^k/N\pi^k)} \quad (19)$$

Using the asset equations for filled jobs and the relevant wage schedules, we can therefore write the free-entry condition without the expectations operators as:

$$\begin{aligned} c = & \frac{m(\theta)}{\theta} \frac{(1-\gamma)}{(r+q_e)} \left[ \frac{U^A}{U^A+U^B} \left[ \int_0^{y^{*A}} (z_0 - rV_u^A(y)) \frac{u^A(y)f(y)}{(U^A/N\pi)} dy \right. \right. \\ & \left. \left. + \int_{y^{*A}}^{y^{**A}} (z_0 - rV_u^A(y)) \frac{u^A(y)f(y)}{(U^A/N\pi)} dy \right] \right. \\ & \left. + \frac{U^B}{U^A+U^B} (1-\lambda) \left[ \int_0^{y^{*B}} (z_0 - rV_u^B(y)) \frac{u^B(y)f(y)}{(U^B/N(1-\pi))} dy \right. \right. \\ & \left. \left. + \int_{y^{*B}}^{y^{**B}} (z_0 - rV_u^B(y)) \frac{u^B(y)f(y)}{(U^B/N(1-\pi))} dy \right] \right] \end{aligned} \quad (20)$$

The only unknowns besides  $\theta$  in Eq. (20)—aside from the unemployment rates and levels and the productivity cut-offs covered in the previous sections—are the unemployment flow value terms,  $rV_u^A(y)$  and  $rV_u^B(y)$ . These will differ for individuals either side of  $y^{*k}$ . For workers with  $y^{*k} \leq y < y^{**k}$ , for whom both self- and wage-employment are an option, the worker value functions can be rearranged for:

$$rV_u^k(y) = \frac{b(r+q_s)(r+q_e) + \alpha y(r+q_e) + \gamma m^k(\theta)z_0(r+q_s)}{(r+q_s)(r+q_e) + \alpha(r+q_e) + \gamma m^k(\theta)(r+q_s)} \quad (21)$$

Similarly, for workers with  $y < y^{*k}$ , who take only wage-employment opportunities, we can write:

$$rV_u^k(y) = \frac{b(r+q_e) + \gamma m^k(\theta)z_0}{(r+q_e) + \gamma m^k(\theta)} \quad (22)$$

As such, Eq. (20) is an expression with only one unknown:  $\theta$ . By imposing a functional form on the matching function and finding values for the exogenous parameters, we can solve this equation using numerical methods.

As shown in Appendix 5, the following standard conditions for the matching function ( $m(\theta)$ ) guarantee the existence of the equilibrium:

1.  $m(\theta)$  is increasing in  $\theta$ .
2.  $\frac{m(\theta)}{\theta}$  is decreasing in  $\theta$ .
3.  $\lim_{\theta \rightarrow 0} m(\theta) = 0$  and  $\lim_{\theta \rightarrow \infty} m(\theta) = \infty$ .
4.  $\lim_{\theta \rightarrow 0} \frac{m(\theta)}{\theta} = \infty$  and  $\lim_{\theta \rightarrow \infty} \frac{m(\theta)}{\theta} = 0$ .

As in other models that incorporate worker heterogeneity into the Mortensen and Pissarides (1994) framework, the uniqueness of the equilibrium cannot be guaranteed analytically (Chéron et al. 2011; Albrecht et al. 2017). We therefore follow Albrecht et al. (2009) and numerically verify that the right-hand-side of the free-entry condition in Eq. (20) is monotonically decreasing for all values of  $\theta$  in the simulations that follow.

## 6 Simulation results

To examine the effects of discrimination, we simulate the model for different values of  $\lambda$ , using the parameter values shown in Table 5. Taking 1 year to be the implicit unit of time, the chances of exiting self- and wage-employment— $q_s$  and  $q_e$ —can be calculated directly from the transition matrices in Tables 3 and 4 above.<sup>17</sup> The precise functional form for  $m(\theta)$  and the value of  $\alpha$  cannot be directly estimated from the GHUPS data, not only because we lack information on vacancies in the wage sector, but also because certain workers ( $y < y^{*k}$ ) will never take self-employment jobs and others ( $y \geq y^{**k}$ ) will never take wage-employment jobs. We therefore adopt the same Cobb-Douglas matching function used by Albrecht et al. (2009) for Latin America. We set  $\alpha$  to 1 to ensure that the higher productivity cut-off,  $y^{**k}$ , exists (see Appendix 4). Moreover, assuming that individuals have at least one opportunity to enter self-employment each year seems reasonable, as the barriers to entry to many entrepreneurial activities in urban Ghana appear to be low.<sup>18</sup>

Returns in each occupation are also, where possible, guided by the data. Given the substantial overlap in the distributions of earnings in self- and wage-employment, we set  $z_0$  to 0.5, so the homogeneous productivity parameter in wage-employment is the mean and median of the distribution of self-employment productivity,  $y$ . We set  $b$  to 0 because no unemployed individuals received periodic social security unemployment benefits in Ghana in 2013 (International Labour Organization 2018). This also ensures that both productivity cut-offs exist. The real interest rate and the cost of creating vacancies are set at 0.09 and 0.3 following similar work by Yassin and Langot (2018) on informality in Egypt. The bargaining parameter  $\gamma$  is set at 0.5: given that this equals the matching elasticity, the Hosios condition is satisfied.

Solving for the equilibrium level of  $\theta$  can be recast as a root finding problem, by subtracting the right-hand-side of the free-entry condition in Eq. (20) from  $c$ . We solve for the equilibrium level of  $\theta$  with Quasi-Newton methods, in particular using Broyden's approach (Broyden 1965; Miranda and Fackler 2002).

Turning to the equilibrium ability and earnings distributions, we define  $h_j^k(y)$  as the density of ability in sector  $j = u, e, s$ , for a Type- $k$  individual, and demonstrate how to calculate the ability distribution for the unemployed as an example. As shown in Eq. (19),

**Table 5** Exogenous parameters and functional forms in the baseline model

Parameter	Calibration value	
	Without discrimination	With discrimination
$r$	0.09	0.09
$b$	0	0
$m(\theta)$	$4\theta^{0.5}$	$4\theta^{0.5}$
$\alpha$	1	1
$q_s$	0.29	0.29
$q_e$	0.35	0.35
$c$	0.3	0.3
$\gamma$	0.5	0.5
$N$	1	1
$\pi$	0.5	0.5
$\lambda$	0	0.5

Bayes' Law allows us to write  $h_u^k(y)$  in terms of the proportion of time that Type- $k$ , ability  $y$  individuals spend in unemployment,  $u^k(y)$ , as well as the underlying distribution of productivity in the population,  $f(y)$ , and the average unemployment rate for Type- $k$ s ( $U^k/N\pi^k$ ). Since  $u^k(y)$  takes different values either side of the cut-offs  $y^{*k}$  and  $y^{**k}$ , the probability density function (PDF) for  $h_u^k(y)$  has three discontinuous portions as shown in Eq. (23). The PDFs of ability for the self-employed,  $h_s^k(y)$ , and the wage-employed,  $h_e^k(y)$ , can be recovered in exactly the same way.

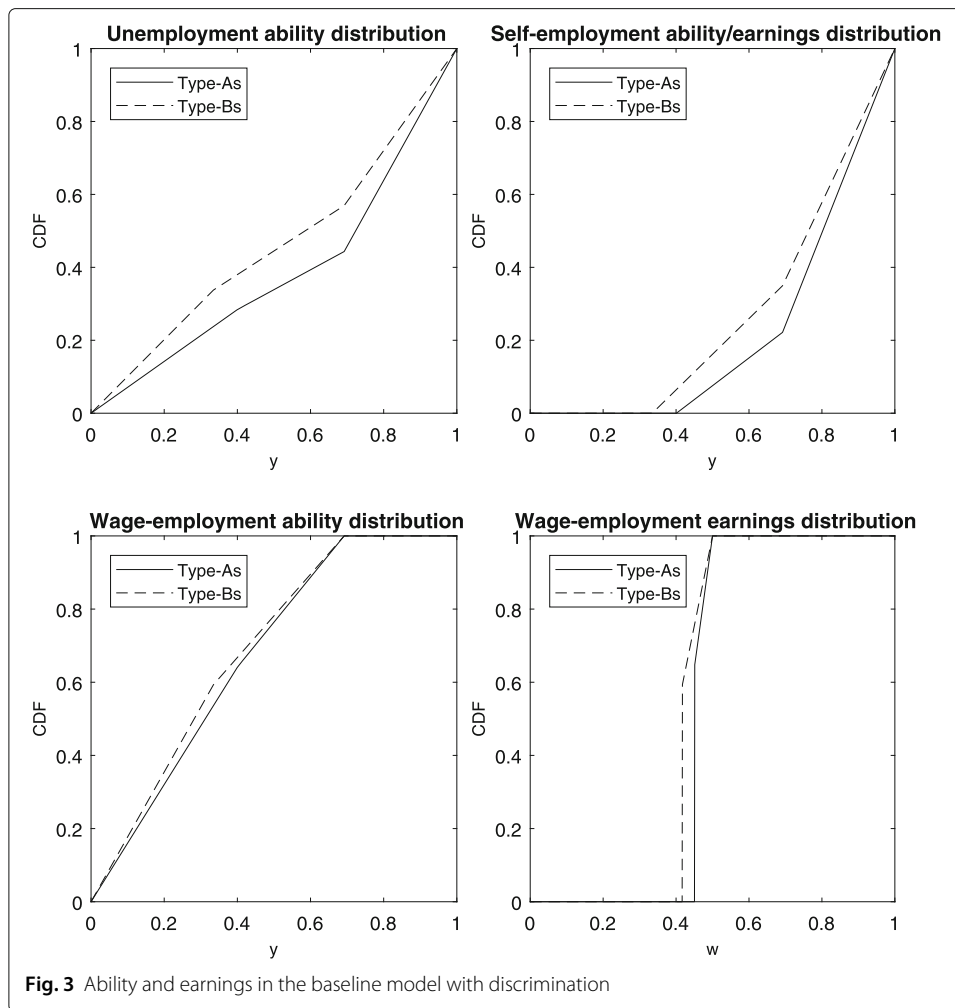
$$h_u^k(y) = \begin{cases} \frac{q_e}{m^k(\theta)+q_e} \frac{N\pi^k}{U^k} & \text{if } y < y^{*k} \\ \frac{q_e}{m^k(\theta)+q_e+\alpha} \frac{q_e}{q_s} \frac{N\pi^k}{U^k} & \text{if } y^{*k} \leq y < y^{**k} \\ \frac{q_s}{\alpha+q_s} \frac{N\pi^k}{U^k} & \text{if } y \geq y^{**k} \end{cases} \quad (23)$$

In self-employment, the PDF for ability can equally be interpreted in terms of earnings, but this is not the case for the wage-employed. The wage distribution, which we label  $g^k(w)$ , only varies between  $w^k(y^{*k})$  and  $w^k(y^{**k})$ . All individuals with  $y < y^{*k}$  simply receive  $w^k(0) = w^k(y^{*k})$ . This is because wage-employment productivity,  $z_0$ , is homogenous and the outside option of unemployment—which determines workers' bargaining position—is constant in  $y$  for these individuals, since they will never become self-employed. This reduces dispersion in earnings for the wage-employed. However, there is variation in wages for individuals with  $y^{*k} \leq y < y^{**k}$  since, they are willing to take self-employment jobs, so the outside option of unemployment varies in  $y$  (see Eq. (21)). The highest wage paid is  $w^k(y^{**k})$  because, any individuals with  $y > y^{**k}$  cannot be incentivised by firms to take wage jobs given their productivity and hence their earnings in self-employment.

Across unemployment, self-employment, and wage-employment, the ability and earnings distributions for Type-As First-Order Stochastically Dominate the distributions for Type-Bs in the baseline model with discrimination, as shown in Fig. 3. The implication, however, that mean ability is lower for Type-Bs than Type-As in *all* three sectors is *prima facie* somewhat puzzling since both Types have the same underlying ability distribution. This begs the question of where the 'missing' ability for Type-Bs actually goes.

To reconcile this puzzle, we need to consider how discrimination affects the labour market sorting of Type-As and Type-Bs differently. Firstly, as shown in Table 6, when  $\lambda > 0$  there are more unemployed Type-Bs than Type-As in the equilibrium of the baseline model with discrimination, resulting directly from the extra frictions that Type-Bs face in obtaining wage jobs. This drives down average ability among the unemployed Type-Bs because it is *lower* productivity individuals that struggle to exit unemployment more with wage sector discrimination. The outcomes of the higher productivity unemployed Type-Bs, however, are unchanged, because their opportunities to work in self-employment remain. This emphasises the fact that the effects of discrimination are not equal across the whole distribution of Type-Bs.

There are also more self-employed Type-Bs than Type-As, because the extra wage sector frictions associated with discrimination result in less productive Type-Bs accepting both wage- and self-employment. This is reflected in the fact that  $y^{*B}$  is lower when



**Table 6** Outcomes in the baseline model equilibrium

	$\lambda = 0$		$\lambda = 0.5$	
	Type-As	Type-Bs	Type-As	Type-Bs
$U^k$	0.0681	0.0681	0.0629	0.0812
$N_S^k$	0.1608	0.1608	0.1529	0.1830
$N_e^k$	0.2711	0.2711	0.2842	0.2358
$E[h_U^k(y)]$	0.5901	0.5901	0.6147	0.5398
$\text{Var}[h_U^k(y)]$	0.0920	0.0920	0.0887	0.0964
$E[h_S^k(y)]$	0.7657	0.7657	0.7796	0.7297
$\text{Var}[h_S^k(y)]$	0.0263	0.0263	0.0232	0.0340
$E[h_e^k(y)]$	0.3197	0.3197	0.3242	0.3080
$\text{Var}[h_e^k(y)]$	0.0386	0.0386	0.0386	0.0388
$E[g^k(w)]$	0.4521	0.4521	0.4593	0.4342
$\text{Var}[g^k(w)]$	0.0003	0.0003	0.0002	0.0006
$y^{*k}$	0.3824	0.3824	0.4010	0.3347
$y^{**k}$	0.5770	0.5770	0.6920	0.6920
$v$	0.0687	0.0687	0.1128	0.1128
$\theta$	0.5046	0.5046	0.7827	0.7827

discrimination is introduced into the model. Thus, the extra Type-B workers in self-employment are of lower ability, on average, than those who would be self-employed were there no discrimination.

Finally, Type-Bs' average ability in wage-employment is lower than Type-As' because it is the *higher* ability wage-employed Type-Bs who are induced to accept both self- and wage-employment by discrimination. Thus, even though there are fewer wage-employed Type-Bs, these workers have lower productivity, on average. This intuition is reflected in Fig. 4, which shows the PDF of ability for Type-Bs in the wage sector both with and without discrimination. Vitrally, setting  $\lambda > 0$  reduces the lower cut-off, as shown by the shift from  $y^{*B}$  to  $y^{*B'}$ . As such, it is individuals with productivity just below  $y^{*B}$  that suddenly become less likely to enter the wage sector when  $\lambda > 0$ : it is the most able wage-employed workers that are pushed out of the wage jobs first by discrimination.

Decomposing Type-Bs' average ability,  $E^B[y]$ , we can see how these three effects combine to create ability and earnings gaps in all three sectors, despite the distribution of ability being identical for Type-As and Type-Bs. This is shown in Eq. (24) by writing  $E^B[y]$  in terms of average ability in each sector,  $E[h_j^B(y)] \forall j = u, s, e$ , and the proportion of Type-B individuals that are unemployed, self-employed, and wage-employed. The underbraces show the impact of adding discrimination into the model. The negative effect that introducing discrimination has on the first and second terms on the right-hand side is entirely offset by the positive effect on the third term.

$$E^B[y] = \underbrace{E[h_u^B(y)]}_{-} \underbrace{\left(\frac{U^B}{N(1-\pi)}\right)}_{+} + \underbrace{E[h_s^B(y)]}_{-} \underbrace{\left(\frac{N_s^B}{N(1-\pi)}\right)}_{+} + \underbrace{E[h_e^B(y)]}_{-} \underbrace{\left(\frac{N_e^B}{N(1-\pi)}\right)}_{-} \quad (24)$$

The intuition behind this puzzling result can also be couched in terms of the Roy (1951) model of occupational choice. Labour market frictions prevent individuals from selecting jobs on the basis of comparative advantage alone. These imperfections in the sorting

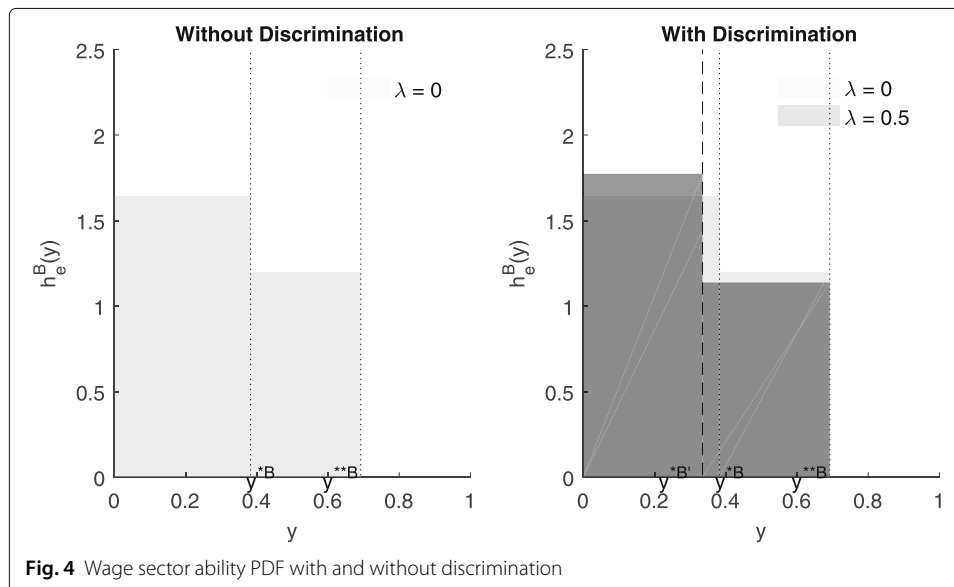


Fig. 4 Wage sector ability PDF with and without discrimination

process lower the average ability for both types in each sector. However, since Type-*Bs* face extra frictions due to discrimination, their occupational selection departs even further from their comparative advantage. This drives a wedge between the values of the lower cut-off for Type-*As* and Type-*Bs* and explains why ability and earnings differentials can persist across all three sectors.

The model also suggests that discrimination may exert spillover effects on Type-*As*. In particular, discrimination increases Type-*As*' chances of gaining wage jobs, reducing the number who are unemployed. This is because they face a slacker labour market. Firstly, they are better insulated against competition from the Type-*Bs*, which increases their chances of successfully matching with a firm. At the same time, firms actually create more vacancies as they need to compensate for the fact the discrimination instantly destroys some proportion of potential matches.

Interpreting Type-*As* as men and Type-*Bs* as women, the baseline model with discrimination matches the stylised facts emanating from the GHUPS in Section 3. Men dominate wage-employment, while women are over-represented in self-employment. Average gender earnings gaps persist in both sectors, but are larger for the self-employed than the wage-employed. Also, earnings are more dispersed for the self-employed than the wage-employed, as bargaining between worker and firm makes the link between ability and earnings less direct in the wage sector. Finally, since  $y^{**k}$  exceeds  $y^{*k}$  for both types, there are some individuals who do not fully specialise in either wage- or self-employment.

## 7 Comparative statics

This section considers the comparative statics for three key parameters to demonstrate the links between wage sector discrimination and policy. Productivity in wage-employment,  $z_0$ , serves as a proxy for growth. The flow of unearned income in unemployment,  $b$ , captures government-provided unemployment benefit. The rate at which unemployed individuals receive opportunities to become self-employed,  $\alpha$ , illustrates the role the government may play in alleviating barriers to entry into self-employment, for example, by providing better access to inputs or credit. As in previous models incorporating worker heterogeneity and self-employment or informality into Mortensen and Pissarides' (1994) random matching and bargaining model, it is not possible to review the comparative statics purely analytically. As such, we take the simulations from Section 6, in the case with discrimination with  $\lambda = 0.5$ , and investigate the comparative statics numerically. Nevertheless, it is possible to differentiate some of the key components of the equilibrium—such as the productivity cut-offs, the size of each sector, and the ability distributions—while holding  $\theta$  constant. This turns out to be a useful approximation for the overall effects and illustrates some of the main mechanisms that cause sectors to expand and contract and average earnings to rise and fall differently for Type-*As* and Type-*Bs* as the key parameters are altered.

### 7.1 Varying $z_0$

Higher levels of productivity in wage-employment—which can be seen as a proxy for economic growth—draw both Type-*As* and Type-*Bs* into wage-employment and out of self-employment. This is clear in the numerical simulations shown in Figure 5 in

Appendix 6, yet it also emerges that  $\left. \frac{\partial N_e^k}{\partial z_0} \right|_{\theta=\theta^*} < 0$  and  $\left. \frac{\partial N_s^k}{\partial z_0} \right|_{\theta=\theta^*} > 0$ . This results from the fact that both of the productivity cut-offs are increasing in  $z_0$ .

While the numerical simulations calibrated using the GHUPS data show that increasing  $z_0$  decreases unemployment for both types, we can see analytically that the effects on unemployment are ambiguous and dependent on the other parameters in the model. Differentiating  $U^k$  with respect to  $z_0$ , while holding  $\theta$  constant, we can write:

$$\left. \frac{\partial U^k}{\partial z_0} \right|_{\theta=\theta^*} = \frac{N\pi^k}{m^k(\theta) + q_e + \alpha \frac{q_e}{q_s}} \left[ \frac{\alpha \frac{q_e}{q_s}}{m^k(\theta) + q_e} \left. \frac{\partial y^{*k}}{\partial z_0} \right|_{\theta=\theta^*} - \frac{q_s m^k(\theta)}{\alpha + q_s} \left. \frac{\partial y^{**k}}{\partial z_0} \right|_{\theta=\theta^*} \right] \quad (25)$$

Increasing  $z_0$  thus produces two countervailing effects. On the one hand, the resulting rise in  $y^{*k}$  reduces the proportion of individuals who are willing to work in both self- and wage-employment: this places upward pressure on unemployment because more low-productivity individuals rule out becoming self-employed. On the other hand, the resulting rise in  $y^{**k}$  increases the proportion of people who are willing to work in both self- and wage-employment: this suppresses unemployment because more high-productivity individuals consider becoming wage-employed as well as self-employed.

The balance of these two countervailing effects depends on  $m^k(\theta)$  and  $\alpha$ , such that discrimination makes it more likely that a rise in  $z_0$  will increase unemployment for Type-Bs. A higher level of  $\alpha$  boosts the upward pressure that the resulting rise in  $y^{*k}$  has on unemployment while also reducing the downward pressure from the resulting rise in  $y^{**k}$ . Intuitively, if self-employment jobs are easier to find, the increase in  $z_0$  will induce fewer high-productivity to consider wage-employment as well as self-employment. The converse is true for  $m^k(\theta)$ . In particular, if it is easier to match with a wage job, high-productivity workers are more likely to consider wage-employment alongside self-employment and some relatively high-productivity workers start excluding the option of self-employment altogether. This is, however, less likely for Type-Bs if discrimination is more widespread as, with a higher  $\lambda$ ,  $m^B(\theta)$  will be lower. Thus, discrimination increases the likelihood that unemployment among Type-Bs will rise when the economy grows.

Average ability and earnings in both wage- and self-employment also rise if  $z_0$  is increased, but discrimination slows this process for Type-Bs. Again, these effects emerge because both of the productivity cut-offs are increasing in  $z_0$ . Higher-productivity workers are less inclined to exclude the possibility of wage work and some relatively high-productivity individuals begin to exclude self-employment, so average ability and earnings rise among the wage-employed. At the same time, this leaves a smaller pool of individuals with an even higher level of  $\gamma$  in self-employment. However, the effects of the shifting cut-offs are suppressed when  $m^B(\theta)$  is lowered by discrimination. To illustrate this, we can use the properties of the uniform distribution to write expressions for average ability and earnings in each sector and then differentiate with respect to  $z_0$  while holding  $\theta$  constant. For example, for average ability in wage-employment, we can write:

$$\begin{aligned} \frac{\partial E[h_e^k(y)]}{\partial z_0} \Big|_{\theta=\theta^*} &= \frac{N\pi^k}{2N_e^{k2} \left( m^k(\theta) + q_e + \alpha \frac{q_e}{q_s} \right)} \left[ \frac{2m^k(\theta)\alpha \frac{q_e}{q_s} N_e^k y^{**k}}{m^k(\theta) + q_e} \frac{\partial y^{**k}}{\partial z_0} \Big|_{\theta=\theta^*} \right. \\ &\quad - \left. \left[ \frac{m^k(\theta)\alpha \frac{q_e}{q_s} y^{**k2}}{m^k(\theta) + q_e} + m^k(\theta)y^{**k2} \right] \frac{\partial N_e^k}{\partial z_0} \Big|_{\theta=\theta^*} \right. \\ &\quad \left. + 2m^k(\theta)N_e^k y^{**k} \frac{\partial y^{**k}}{\partial z_0} \Big|_{\theta=\theta^*} \right] \end{aligned} \quad (26)$$

Since  $m^k(\theta)$  enters the terms before  $\frac{\partial y^{**k}}{\partial z_0} \Big|_{\theta=\theta^*}$  and  $\frac{\partial y^{**k}}{\partial z_0} \Big|_{\theta=\theta^*}$  positively, the extent of discrimination,  $\lambda$  enters negatively. As such, discrimination may potentially slow the distribution of the proceeds of growth.

## 7.2 Varying $b$

While raising unemployment benefit,  $b$ , increases unemployment and self-employment at the expense of wage-employment in the numerical simulations shown in Figure 6 in Appendix 6, looking analytically at the effects of raising  $b$  once again reveals that these results are sensitive to the other parameters in the model. This arises because  $b$  moves the productivity cut-offs in different directions:  $\frac{\partial y^{**k}}{\partial b} \Big|_{\theta=\theta^*} > 0$  while  $\frac{\partial y^{**k}}{\partial b} \Big|_{\theta=\theta^*} < 0$ . Intuitively, higher unemployment benefits afford workers the opportunity to wait for longer before accepting a job, reducing the need to take any job that comes along. In turn, workers can specialise more in either wage- or self-employment in accordance with their productivity. In the numerical simulations, increasing  $b$  causes  $y^{**k}$  to fall faster than  $y^{*k}$  rises for both types. As such, the increase in the proportion of high-productivity workers who are only willing to do self-employment rises faster than the fall in the proportion of low-productivity workers who are willing to do wage-employment only. However, this need not be the case, as differentiating  $N_e^k$  and  $N_s^k$  with respect to  $b$  demonstrates.<sup>19</sup>

$$\frac{\partial N_e^k}{\partial b} \Big|_{\theta=\theta^*} = \frac{N\pi^k}{m^k(\theta) + q_e + \alpha \frac{q_e}{q_s}} \left[ \frac{\alpha \frac{q_e}{q_s}}{m^k(\theta) + q_e} \frac{\partial y^{**k}}{\partial b} \Big|_{\theta=\theta^*} + m^k(\theta) \frac{\partial y^{**k}}{\partial b} \Big|_{\theta=\theta^*} \right] \quad (27)$$

$$\frac{\partial N_s^k}{\partial b} \Big|_{\theta=\theta^*} = \frac{N\pi^k}{m^k(\theta) + q_e + \alpha \frac{q_e}{q_s}} \left[ -\alpha \frac{q_e}{q_s} \frac{\partial y^{**k}}{\partial b} \Big|_{\theta=\theta^*} - \frac{\alpha m^k(\theta)}{\alpha + q_s} \frac{\partial y^{**k}}{\partial b} \Big|_{\theta=\theta^*} \right] \quad (28)$$

The likelihood that increasing  $b$  leads to a rise in wage-employment (and a decline in self-employment) is therefore higher when  $m^k(\theta)$  is lower. For Type-Bs, this happens when discrimination is more widespread ( $\lambda$  is higher). As such, unemployment benefits



may offer a way of shifting employment from self-employment to the wage sector, especially for Type-Bs.

The effects of changing  $b$  on average ability and earnings in each sector are similarly ambiguous, although—as the simulations show—an interesting wedge emerges between earnings and ability in wage-employment. This arises because unemployment becomes a more tenable outside option, affording wage workers more bargaining power.

### 7.3 Varying $\alpha$

While Section 6 considers how altering the matching function for wage-employment—by introducing discrimination—alters Type-As' and Type-Bs' labour market outcomes, there is a supplementary question over the impact of changing the arrival rate for opportunities in self-employment. As the simulations in Figure 7 in Appendix 6 show, the number of individuals in self-employment rises at the expense of wage-employment if  $\alpha$  is increased. These changes are almost entirely driven by the top of the productivity distribution. Since  $\left. \frac{\partial y^{*k}}{\partial \alpha} \right|_{\theta=\theta^*} = 0$ , there are only very small changes in  $y^{*k}$ , brought about by the adjustment in labour market tightness  $\theta$ . The upper cut-off, by contrast, declines relatively sharply with higher levels of  $\alpha$ . When self-employment opportunities come along more frequently, higher-productivity workers of both types are more willing to forego wage-employment and pursue a job in which they have comparative advantage.

Correspondingly, there are moderate declines in average ability and earnings in all sectors of the economy, as  $\alpha$  is increased. As the the upper cut-off falls, more individuals with lower productivity are drawn into self-employment. At the same time, it is those individuals with the highest ability among those who were previously willing to take both self- and wage-employment, that start to ignore the option of becoming wage-employed, with higher levels of  $\alpha$ . Indeed, the effects of changing  $\alpha$  are analogous to the effects of changing  $\lambda$  (as in Section 6), but the mechanism operates around the higher productivity cut-off, rather than the lower one.

## 8 Conclusion

In this paper, we develop a search and matching model of the labour market, which gives workers the option of self-employment and allows for gender-based discrimination in the wage sector. Discrimination is modelled without assuming firms are heterogeneous: instead we assume that there is some stochastic shock to the matching function for all female job seekers, which affects all firms in the economy. This can be thought of as each firm containing some proportion of prejudiced interviewers, whom job seekers may encounter. This approach may be tenable in countries like Ghana, where workers' outcomes depend on their gender across many industries and many sectors. The model also includes individual heterogeneity in terms of self-employment productivity: this affects wage workers' outside options generating earnings dispersion in the wage sector too.

We show that earnings and ability gaps can persist in all sectors of the economy, even if the underlying ability distributions for women and men are identical. Drawing on the logic of the Roy (1951) model, this result arises because labour

market frictions push individuals away from the jobs that best match their comparative advantage. In the presence of discrimination, women's occupational choices are even more distorted than those of men, driving down their average ability in each sector.

Our model therefore suggests an alternative explanation for why earnings gaps are so large for the self-employed, without assuming that there are underlying differences between the productivity distributions of women and men and without building discrimination into the self-employment sector itself. This has crucial policy implications for countries like Ghana, where self-employment is prevalent among women. Previous programmes have solely targeted self-employed women, either by trying to build their entrepreneurial ability through training and mentoring or by attempting to relax constraints to business growth by improving access to credit and other inputs. However, this paper suggests that improving the outcomes of women in self-employment requires a more holistic approach, including policies that target discrimination or other frictions in wage-employment. Gender inequality in the labour market cannot be eliminated one sector at a time.

## Endnotes

<sup>1</sup> The female and male labour force participation rates in 2013 were 74 and 79 percent respectively (International Labour Organization 2018).

<sup>2</sup> A later version of this model, which does not allow for self-employment but which focusses more directly on firm heterogeneity, is discussed in Meghir et al. (2015).

<sup>3</sup> For examples see, Aigner and Cain (1977), Coate and Loury (1993), and Moro and Norman (2004).

<sup>4</sup> For an extension with a third type of firm that employs both Type-As and Type-Bs and pays them equal wages, see Lanning (2014).

<sup>5</sup> Flabbi's (2010) builds directly on work by Bowlus (1997), Eckstein and Wolpin (1999), and Bowlus and Eckstein (2002).

<sup>6</sup> Details of the sample are provided in Appendix 1.

<sup>7</sup> The relative merits of different approaches for correcting for selection using the multinomial logit in different settings are discussed at length by Bourguignon et al. (2007). The method due to Lee (1983) offers a potential alternative to the Dubin and McFadden approach. Unlike Dubin and McFadden's approach, Lee's method makes the restrictive assumption that the covariance between the earnings equation error terms and the selection equation error terms is the same for all occupations, so it is not preferred for the analysis in the paper. However, in certain circumstances — especially where the sample size is small relative to the number of categories — Lee's approach may be more appropriate, since the multiple correction terms generated by the Dubin and McFadden approach may be collinear, reducing the efficiency with which the earnings equation parameters can be estimated. As such, we focus on the results from Dubin and McFadden approach, but check the results are robust to using the Lee method and highlight any differences that arise.

<sup>8</sup> Specifically, the exclusion restrictions for unearned income comprise two dummy variables for whether the individual lived in a household with agricultural land or non-agricultural land and a dummy variable for whether the individual received any

sources of income unrelated to their present job (such as remittances, gifts, or payments from a non-governmental organisation). The exclusion restrictions for parental characteristics comprise a set of dummy variables for whether the mother and father had attained primary, lower secondary, upper secondary, or higher education and whether the mother/father had spent most of their life working in self- or wage-employment.

<sup>9</sup> If the Lee method is used, correcting for unobservable selection appears to have even less effect on the point estimates for the raw, unexplained, and explained gender earnings gaps in self-employment.

<sup>10</sup> For a range of other studies looking at drivers of occupational selection in Ghana, see Sackey (2005), Ackah et al. (2009), Heintz and Pickbourn (2012), Ackah (2013), Baah-Boateng et al. (2013), and Lain (2016).

<sup>11</sup> It is often observed that the earnings of self-employed or informal sector workers vary over time (as Falco (2014) shows, also using the GHUPS data). In previous iterations of the model, we allowed for stochastic shocks to self-employment earnings, but since this had no effect on the model's main predictions, we maintain the assumption that self-employment earnings are time-invariant for parsimony's sake.

<sup>12</sup> Cano-Urbina (2015) constructs a model for the Mexican labour market that allows for transitions from the informal to the formal sector but not from the formal to the informal sector.

<sup>13</sup> While treating wage-employment productivity as constant (at  $z_0$ ) is a strong assumption, the data indicate that earnings dispersion for the self-employed is generally higher than for the wage-employed. Moreover, previous iterations of the model allowed for heterogeneous productivity in both self- and wage-employment, but this only served to increase computational complexity, without substantively changing the results. As such, the simpler approach taken here is preferred.

<sup>14</sup> We also note that the cut-off  $y^{*k}$  determines the shape of the  $X_e^k(y)$  schedule, since  $\frac{\partial X_e^k(y)}{\partial y} = 0$  when  $X_s^k(y) \leq 0$ . If productivity is so low that workers would never choose self-employment, the difference between the welfare of a wage-employed worker and an unemployed worker is invariant to  $y$ .

<sup>15</sup>  $E[u^k(y)]$  can also be thought of as the unemployment rate.

<sup>16</sup> In fact, firms can only match with a *subset* of unemployed workers with  $y < y^{**k}$ . However, this restriction to the sample is captured later, by bounding the integrals in Eq. (20).

<sup>17</sup> We assume these transitions follow a simple Markov chain process (Bosch and Maloney 2007). Kavuma et al. (2015) find similar results using a comparable dataset from Uganda, although job-to-job transitions are somewhat less frequent in Kerr's (2012) analysis of Tanzania. We classify any direct switches between self- and wage-employment, and vice-versa, as job exits, since individuals can only reach a new job via a spell in unemployment in our model.

<sup>18</sup> In the GHUPS data, 60% of self-employed workers are in retail or trade, typically requiring no factors of production other than time and a stock of goods to sell. Setting higher values for  $\alpha$  does not substantially affect the main findings of the model.

<sup>19</sup> It emerges that  $\frac{\partial U^k}{\partial b} \Big|_{\theta=\theta^*} > 0$  unambiguously.

<sup>20</sup> Unemployment can approach, but never reach 0, because each period some wage- and self-employment workers will lose their jobs providing  $q_e > 0$  and  $q_s > 0$ .

## Appendix

### Appendix 1: Sample characteristics

**Table 7** Number of observations over time and space

Wave	City							
	Accra		CC/Takoradi		Kumasi		Total	
	No.	%	No.	%	No.	%	No.	%
2005	566	4.27	213	1.61	456	3.44	1235	9.32
2006	699	5.28	257	1.94	562	4.24	1518	11.46
2008	573	4.32	221	1.67	400	3.02	1194	9.01
2009	641	4.84	223	1.68	418	3.15	1282	9.68
2010	910	6.87	412	3.11	748	5.65	2070	15.62
2012	1297	9.79	616	4.65	1110	8.38	3023	22.82
2013	1295	9.77	606	4.57	1026	7.74	2927	22.09
Total	5981	45.14	2548	19.23	4720	35.63	13249	100.00

Sample of working age Individuals (15–65) for the years 2005–2013  
 'CC' refers to Cape Coast

**Table 8** Summary statistics

	<i>N</i>	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
<i>Full sample</i>								
Sex (0 = Female, 1 = Male)	13490	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Education (years)	13487	9.26	3.59	0.00	9.00	9.00	12.00	20.00
Age (years)	13491	31.82	12.30	15.00	22.00	29.00	40.00	65.00
Hours worked per week	7028	47.46	21.48	0.00	38.00	48.00	60.00	120.00
Tenure (months)	7009	90.17	100.58	0.00	17.00	52.00	129.00	608.00
<i>Female</i>								
Education (years)	7780	8.71	3.80	0.00	8.00	9.00	11.00	20.00
Age (years)	7782	32.24	12.33	15.00	22.00	30.00	40.00	65.00
Hours worked per week	3937	46.85	21.79	0.00	35.00	48.00	60.00	120.00
Tenure (months)	3919	86.19	100.41	0.00	16.00	48.00	122.00	608.00
<i>Male</i>								
Education (years)	5706	10.00	3.12	0.00	9.00	9.00	12.00	18.00
Age (years)	5708	31.24	12.24	15.00	21.00	28.00	39.00	65.00
Hours worked per week	3090	48.24	21.05	0.00	40.00	48.00	60.00	120.00
Tenure (months)	3089	95.23	100.60	0.00	19.00	57.00	141.00	598.00
Observations	13,491							

Pooled sample of individuals of working age (15–65) for the years 2005–2013

## Appendix 2: Measurement of earnings

Despite its importance for determining consumption and poverty, measuring earnings is no easy task (Teal 2011). There are two particular challenges associated with measuring earnings using the GHUPS, to which this paper must respond.

Firstly, while the GHUPS adopted a different strategy for measuring earnings in wage- and self-employment, this should not drastically affect the analysis undertaken in this paper because the key earnings comparisons are between women and men, rather than between the wage- and self-employed. Earnings in wage-employment were measured by asking workers to report all pre-tax monetary payments from employers to employees, excluding bonuses or allowances. Earnings in self-employment were measured by asking self-employed individuals to report their profits, calculated as revenues minus all costs (including the costs of hired labour). Unfortunately, there is no general question that asks about take-home earnings in the same way for both wage- and self-employed individuals as there has been in other surveys conducted in Ghana (see for, example, the Ghana Living Standards Survey). That said, even in surveys conducted in similar contexts where the self-employed have been asked a general take-home earnings question alongside questions about revenues, costs, and profits, the correlation between the responses to the two types of questions has tended to be quite low (Vijverberg 1991; 1992). Nevertheless, there is a sizeable overlap between the earnings distributions for the self- and the wage-employed, especially for those wage-employed individuals working in small firms (Sandefur et al. 2006). This suggests that the figures obtained from the separate self- and wage-employment sections of the GHUPS could be plausible, even if there were no inter-occupation frictions. Thus, for this paper—in which the analysis does not depend directly on comparisons between earnings in wage- and self-employment—having two different approaches for measuring earnings in the two occupations should not have a large effect on the results.

Secondly, focussing just on the self-employed, several steps were taken to measure self-employment earnings (through profits) as accurately as possible. Since self-employed individuals in developing countries rarely keep formal accounts of their revenues and costs, and differentiating between business and personal expenditures may be difficult, it is sometimes suggested that any estimates of profits will provide inaccurate signals of earnings (Rankin et al. 2010). When the GHUPS data were collected, however, the enumerators carefully explained the concepts of ‘revenues’, ‘business costs’, and ‘profits’ to the respondents to ensure they were understood. The surveys also included a long list of questions about costs—including on raw materials, others’ labour, rental of buildings and vehicles, and utilities—to ensure all aspects of self-employed individuals’ businesses were covered. At the same time, the surveys were conducted with hand-held computers, which flagged any potential inconsistencies in the reported answers (such as costs exceeding revenues) to the enumerators in real time, so they could verify with respondents. Responses from each day of the survey were also reviewed by field supervisors and the survey team. Thus, even before analysis began, many measures had been undertaken to ensure the earnings data for the self-employed were of high quality.

We also ensure that the analysis—in particular for the gender earnings gaps—is robust to the inclusion/exclusion of outliers. In the main results, outliers are dropped at the 5th and 95th percentiles, but varying these percentiles and even leaving all outliers in does not qualitatively affect the main results.

**Appendix 3: Decomposition of gender earnings gap for self- and wage-employment**

	Self-employed: OLS		Self-employed: DMF		Wage-employed (All): OLS		Wage-employed (All): DMF	
	No controls	With controls	No controls	With controls	No controls	With controls	No controls	With controls
Male? (1 = Y, 0 = N)	0.6363*** (0.0593)	0.5341*** (0.0713)	0.3344** (0.1324)	0.4435 (0.2830)	0.3748*** (0.0467)	0.2823*** (0.0427)	-0.0192 (0.0795)	0.0308 (0.1545)
Education (years)		0.0436*** (0.0100)		0.0449* (0.0242)		0.0742*** (0.0074)		0.0560*** (0.0147)
Age (years)		0.0277 (0.0192)		-0.0292 (0.0883)		0.0653*** (0.0133)		0.0289 (0.0593)
Age (years) squared		-0.0003 (0.0002)		0.0004 (0.0011)		-0.0007*** (0.0002)		-0.0002 (0.0007)
Married? (1 = Y, 0 = N)		0.1525* (0.0784)		0.1589 (0.1287)		0.1396** (0.0601)		0.2768*** (0.0699)
Household size		-0.0054 (0.0107)		-0.0056 (0.0098)		-0.0058 (0.0040)		-0.0072* (0.0039)
Constant	3.5053*** (0.0454)	2.2458*** (0.3431)	3.1174*** (0.2098)	3.3382* (1.9232)	3.4766*** (0.0305)	1.1850*** (0.2472)	4.8154*** (0.2178)	2.8219* (1.4996)
City dummies	x	✓	x	✓	x	✓	x	✓
Wave dummies	x	✓	x	✓	x	✓	x	✓
DMF correction	x	x	✓	✓	x	x	✓	✓
Explained gender earnings gap		0.1023		-0.1091		0.0925		-0.0500
p-value on explained gap		0.0001		0.6953		0.0000		0.7204
p-value on male coefficient versus uncorrected model			0.0243	0.7716			0.0000	0.0872
N	1761	1760	1760	1760	1717	1717	1717	1717
R <sup>2</sup>	0.0614	0.1128	0.0777	0.1146	0.0539	0.2339	0.1632	0.2377

Standard errors in parentheses

Dependent variable: Log of Weekly Earnings

\*DMF refers to Dubin and McFadden model

Sample of individuals of working age (15–65) from the 2010, 2012, and 2013 waves of the GHUPS

Outliers trimmed at the 5th and 95th percentiles for each sector

Standard errors cluster bootstrapped by enumeration area

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

#### Appendix 4: Existence of the productivity cut-offs

The productivity cut-offs  $y^{*k}$  and  $y^{**k}$  will only exist if they lie between 0 and 1, given our assumption for both Type-As and Type-Bs that productivity  $y \sim \text{Uniform}(0, 1)$ . In this appendix, we consider the conditions that will guarantee the existence of both cut-offs.

##### Existence of $y^{*k}$

We begin by recalling that the lower cut-off for each Type,  $y^{*k}$ , may be written:

$$y^{*k} = \frac{b(r + q_e) + \gamma m^k(\theta) z_0}{r + q_e + \gamma m^k(\theta)} \quad (29)$$

Since all the terms on the right-hand side of Eq. (29) are non-negative by construction, it is clear that the condition that  $y^{*k} \geq 0$  is met trivially.

We can also see that  $y^{*k} \leq 1$  providing  $b$  does not become too large. In particular, we require:

$$b \leq \frac{r + q_e + \gamma m^k(\theta)(1 - z_0)}{r + q_e} \quad (30)$$

By comparing the numerator and the denominator in Eq. (29), we can see  $y^{*k} \leq 1$  is guaranteed by simply imposing that  $b \leq 1$ . Moreover, since  $0 \leq z_0 \leq 1$ , this will hold if  $b \leq z_0$ .

##### Existence of $y^{**k}$

To guarantee the existence of the upper cut-off, we derive two conditions which rule out the equilibria shown in panels A and B of Fig. 2 (see Section 5.1).

To ensure that  $X_e^k(y)$  is not too high, ruling out the situation in panel A of Fig. 2, we require that  $y^{**k} \leq 1$ . In particular, this requires:

$$z_0 \leq \frac{\alpha + b(r + q_s)}{\alpha + q_s + r} \Leftrightarrow \alpha \geq \frac{(z_0 - b)(q_s + r)}{1 - z_0} \quad (31)$$

This condition restricts the values of  $z_0$  (relative to  $b$ ), since otherwise the surplus and hence earnings from matches in wage-employment are too large for any individuals to forego wage-employed jobs. The condition also places a lower bound on  $\alpha$ .

To exclude the possibility that panel B prevails, we require that  $X_e^k(y^{*k}) > 0$ . Since  $X_s^k(y^{*k}) = 0$  by definition, we can write:

$$X_e^k(y^{*k}) = \frac{\gamma(z_0 - b)}{r + q_e + \gamma \mathbb{1}_{[X_e^k(y) > 0]} m^k(\theta)} \quad (32)$$

Thus,  $X_e^k(y^{*k}) > 0$  if  $z_0 > b$ .

##### Simulation conditions

Summarising the lessons from Appendix 4, we impose two restrictions:

1.  $z_0 > b$
2.  $z_0 < \frac{\alpha + b(r + q_s)}{\alpha + q_s + r} \Leftrightarrow \alpha \geq \frac{(z_0 - b)(q_s + r)}{1 - z_0}$

These are sufficient to guarantee the existence of both cut-offs,  $y^{*k}$  and  $y^{**k}$ .

#### Appendix 5: Existence of the equilibrium

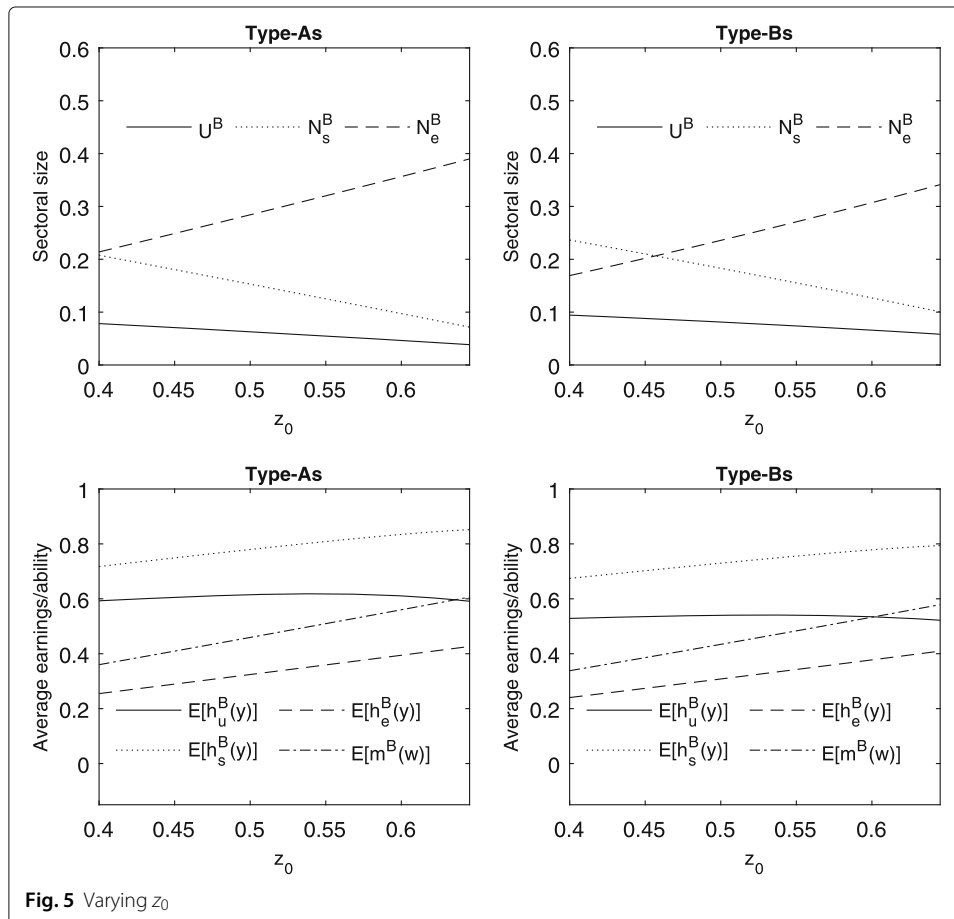
To demonstrate that an equilibrium exists, we must show that the right-hand side of the free-entry condition in Eq. (20) tends to  $\infty$  as  $\theta \rightarrow 0$  and that it tends to 0 as  $\theta \rightarrow \infty$ . Given the assumptions made about the matching function, this is guaranteed providing the term in the large square brackets does not tend to  $\infty$  as  $\theta \rightarrow \infty$  and does not tend 0 as  $\theta \rightarrow 0$ .

Firstly, we note that the unemployment terms are all bounded at 0 and  $\pi^k N \forall k = A, B$ .<sup>20</sup>The ability distribution,  $f(y)$  is independent of  $\theta$ , as are  $z_0$ ,  $r$ , and  $\lambda$ .

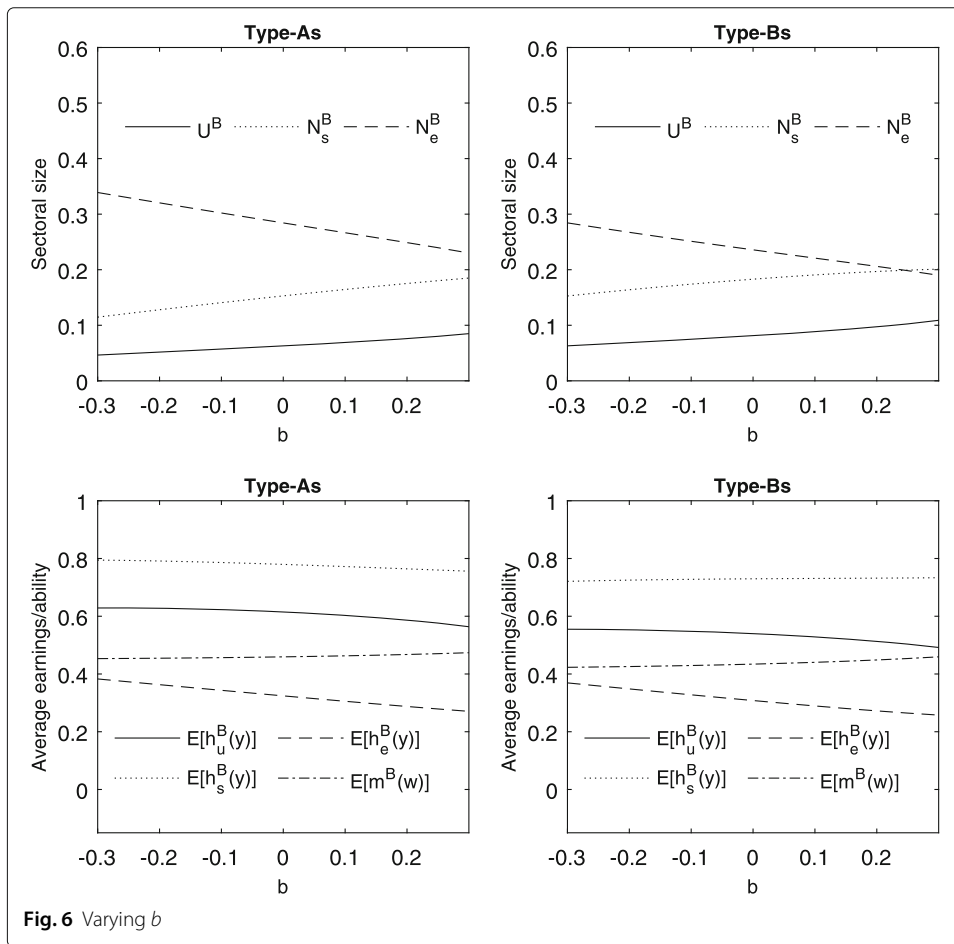
The only remaining terms to consider are the flow values of unemployment for each type,  $rV_u^A(y)$  and  $rV_u^B(y)$ . Firstly,  $rV_u^k(y)$  will not explode as  $\theta \rightarrow \infty$ . With a very high level of  $\theta$ , jobs immediately become available to those who want them. Wages, however, will still be restricted by the output that matches provide to firms,  $z_0$ . This places an upper bound on the value of transitioning from unemployment to wage-employment and therefore limits  $rV_u^k(y)$ . Secondly, as  $\theta \rightarrow 0$ ,  $rV_u^k(y)$  should not tend to 0. Even as the labour market becomes very tight and individuals cannot find wage jobs, they still receive unemployment flow income  $b$ , the value of which is independent of  $\theta$ . They also retain the option value of transitioning into self-employment. The probability of doing this is determined entirely by  $\alpha$ , and the flow value of self-employment is governed by  $y$  and  $q_e$ .

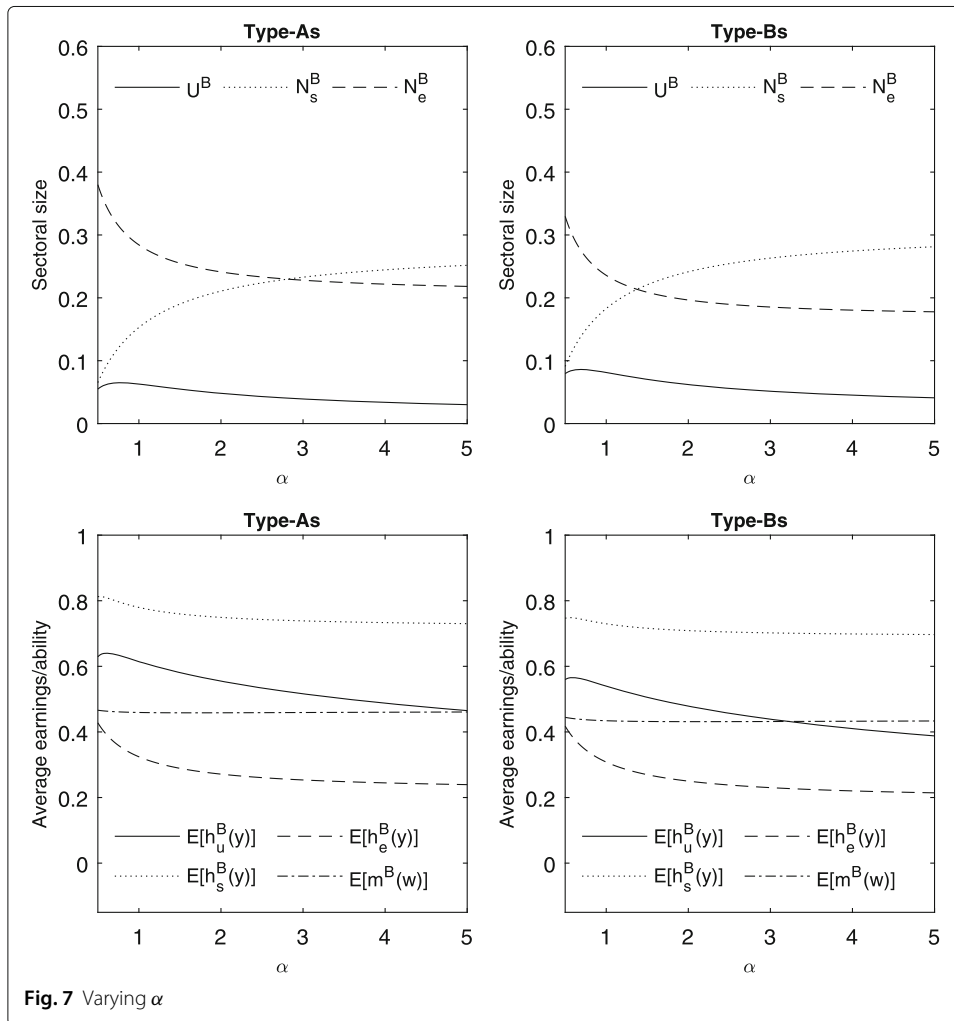
Since the terms in the square brackets neither explode as  $\theta \rightarrow \infty$  nor tend to 0 as  $\theta \rightarrow 0$ , the right-hand-side of the free-entry condition in Eq. (20) tends to  $\infty$  as  $\theta \rightarrow 0$  and tends to 0 as  $\theta \rightarrow \infty$ . This guarantees the existence of at least one equilibrium.

**Appendix 6: Comparative statics: numerical simulations**









### Acknowledgements

I would like to thank my supervisor, Margaret Stevens, for her generous support and guidance on this paper. I am also grateful to Ian Crawford, Mohit Dalwadi, Andrew Kerr, Simon Quinn, Neil Rankin, and the members of the Firms and Development Research Group at the University of Oxford for their important inputs. Additionally, I would like to thank the anonymous referee and the responsible editor for their vital feedback and advice. This research was funded by the Economic and Social Research Council, and the data were collected by the Centre for the Study of African Economies. This paper was written prior to my joining the OECD and the views expressed should not be taken as representing the views of the OECD or its member countries. I am also responsible for any errors.  
Responsible editor: Hartmut F. Lehmann

### Funding

This research was funded by the Economic and Social Research Council, through a doctoral studentship.

### Availability of data and materials

The data used for this paper were collected by the Centre for the Study of African Economies and may be requested from them: <http://www.csae.ox.ac.uk/data>.

### Authors' contributions

JL is the sole author of this paper. JL developed the theoretical model, analysed the data, and wrote the manuscript. The author read and approved the final manuscript.

### Competing interests

The author declares no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 June 2018 Accepted: 1 January 2019

Published online: 13 March 2019

### References

- Ackah C (2013) Non-farm employment and incomes in rural Ghana. *J Int Dev* 25(3):325–339
- Ackah C, Ahiadeke C, Fenny AP (2009) Determinants of female labour force participation in Ghana. Working Paper 14, Global Development Network
- Aigner DJ, Cain GG (1977) Statistical theories of discrimination in labor markets. *Ind Labor Relat Rev* 30(2):175–187
- Albrecht J, Navarro L, Vroman S (2009) The effects of labour market policies in an economy with an informal sector. *Econ J* 119:1105–1129
- Albrecht J, Robayo-Abril M, Vroman S (2017) Public-sector employment in an equilibrium search and matching model. Discussion Paper 10467, IZA Institute of Labor Economics
- Baah-Boateng W (2012) Labour Market Discrimination in Ghana: A Gender Dimension. LAP LAMBERT Academic Publishing, Saarbrücken
- Baah-Boateng W, Nketiah-Amponsah E, Frempong R (2013) The effect of fertility and education on female labour force participation in Ghana. *Ghanaian J Econ* 1(1):119–137
- Becker GS (1957) *Econ Discrimination*. University of Chicago Press, Chicago
- Black DA (1995) Discrimination in an equilibrium search model. *J Labor Econ* 13(2):309–334
- Blinder AS (1973) Wage discrimination: Reduced form and structural estimates. *J Hum Resour* 8(4):436–455
- Borjas GJ, Bronars SG (1989) Consumer discrimination and self-employment. *J Polit Econ* 97(3):581–605
- Bosch M, Maloney W (2007) Comparative analysis of labor market dynamics using Markov processes: An application to informality. In: Policy Research Working Paper 4429, World Bank
- Bourguignon F, Fournier M, Gurgand M (2007) Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons. *J Econ Surv* 21(1):174–205
- Bowlus AJ (1997) A search interpretation of male-female wage differentials. *J Labor Econ* 15(4):625–657
- Bowlus AJ, Eckstein Z (2002) Discrimination and search differences in an equilibrium search model. *Int Econ Rev* 43(4):1309–1345
- Broyden CG (1965) A class of methods for solving nonlinear simultaneous equations. *Math Comput* 19(92):577–593
- Burdett K, Mortensen DT (1998) Wage differentials, employer size, and unemployment. *Int Econ Rev* 39(2):257–273
- Cano-Urbina J (2015) An equilibrium search model with coworker discrimination. *J Dev Econ* 112:33–55
- Chéron A, Hairault J-O, Langot F (2011) Age-dependent employment protection. *Econ J* 121:1477–1504
- Coate S, Loury GC (1993) Will affirmative-action policies eliminate negative stereotypes? *Am Econ Rev* 83(5):1220–1240
- Dubin JA, McFadden DL (1984) An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* 52(2):345–362
- Eckstein Z, Wolpin KI (1999) Estimating the effect of racial discrimination on first job wage offers. *Rev Econ Stat* 81(3):384–392
- Elder TE, Goddeeris JH, Haider SJ (2010) Unexplained gaps and Oaxaca-Blinder decompositions. *Labour Econ* 17(1):284–290
- Falco P (2014) Does risk matter for occupational choices? experimental evidence from an African labour market. *Labour Econ* 28(C):96–109
- Falco P, Kerr A, Rankin N, Sandefur J, Teal F (2011) The returns to formality and informality in urban Africa. *Labour Econ* 18(S1):23–31
- Fields GS (2011) Labor market analysis for developing countries. *Labour Econ* 18(S1):16–22

- Flabbi L (2010) Gender discrimination estimation in a search model with matching and bargaining. *Int Econ Rev* 51(3):745–783
- Hart K (1973) Informal income opportunities and urban employment in Ghana. *J Mod Afr Stud* 11(1):61–89
- Heath R (2017) Fertility at work: Children and women's labor market outcomes in urban Ghana. *J Dev Econ* 126(C):190–214
- Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica* 47(1):153–161
- Heintz J, Pickbourn L (2012) The determinants of selection into non-agricultural self-employment in Ghana. *Margin J Appl Econ Res* 6(2):181–209
- International Labour Organization (2018) ILOSTAT. Retrieved from: <http://www.ilo.org/ilostat>
- Kavuma S, Morrissey O, Upward R (2015) Worker flows and the impact of labour transitions on earnings in Uganda. In: Research Paper 15/01, CREDIT
- Kerr A (2012) A model of comparative advantage with matching in the urban Tanzanian labour market. In: Working Paper 2012–21, Centre for the Study of African Economies
- Kingdon G, Sandefur J, Teal F (2006) Labour market flexibility, wages and incomes in Sub-Saharan Africa in the 1990s. *Afr Dev Rev* 18(3):392–427
- Lain J (2016) Job flexibility and occupational selection: An application of maximum simulated likelihood using data from Ghana Working Paper 2016–34, Centre for the Study of African Economies
- Lanning J (2014) A search model with endogenous job destruction and discrimination: Why equal wage policies may not eliminate wage disparity. *Labour Econ* 26(1):55–71
- Lee L-F (1983) Generalized econometric models with selectivity. *Econometrica* 51(2):507–512
- Meghir C, Narita R, Robin J-M (2015) Wages and informality in developing countries. *Am Econ Rev* 105(4):1509–1546
- Miranda MJ, Fackler PL (2002) *Applied Computational Economics and Finance*. MIT Press, Cambridge
- Moro A, Norman P (2004) A general equilibrium model of statistical discrimination. *J Econ Theory* 114:1–30
- Mortensen DT, Pissarides CA (1994) Job creation and job destruction in the theory of unemployment. *Rev Econ Stud* 61(3):397–415
- Narita R (2012) Self-employment in developing countries: a search-equilibrium approach. Working paper, World Bank, Washington DC
- Oaxaca R (1973) Male-female wage differentials in urban labor markets. *Int Econ Rev* 14(3):693–709
- Rankin N, Sandefur J, Teal F (2010) Learning and earning in Africa: Where are the returns to education high? Working Paper 2010–02, Centre for the Study of African Economies
- Rogerson R, Shimer R, Wright R (2005) Search-theoretic models of the labor market: A survey. *J Econ Lit* 43:959–988
- Roy AD (1951) Some thoughts on the distribution of earnings. *Oxf Econ Pap* 3(2):135–146
- Saatchi M, Temple J (2009) Labor markets and productivity in developing countries. *Rev Econ Dyn* 12(1):183–204
- Sackey HA (2005) Female labour force participation in Ghana: The effects of education. Research Paper 150, African Economic Research Consortium
- Sandefur J, Serneels P, Teal F (2006) Self-employment: Engine of growth or self-help safety net? In: Paci P, Serneels P (eds). *Employment and Shared Growth: Rethinking the Role of Labor Mobility for Development*. World Bank, Washington DC
- Sasaki M (1999) An equilibrium search model with coworker discrimination. *J Labor Econ* 17(2):377–407
- Teal F (2011) The price of labour and understanding the causes of poverty. *Labour Econ* 18(S1):7–15
- Vijverberg WPM (1991) Profits from self-employment: The case of Côte d'Ivoire. *World Dev* 19(6):683–696
- Vijverberg WPM (1992) Measuring income from family enterprises with household surveys. *Small Bus Econ* 4(4):287–305
- Yassin S, Langot F (2018) Informality, public employment and employment protection in developing countries. *J Comp Econ* 46(1):326–348

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---