Journal of International Development
J. Int. Dev. **30**, 760–789 (2018)
Published online 5 September 2017 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jid.3310

RATIONING OF FORMAL SECTOR JOBS AND INFORMALITY: THE COLOMBIAN CASE

GUSTAVO A. GARCÍA[1]* and ERIKA RAQUEL BADILLO[2]
[1]*Research in Spatial Economics (RiSE-group), Universidad EAFIT, Medellin, Colombia*
[2]*GINVECO Research Group, Faculty of Economics, Universidad Autónoma Latinoamericana, Medellin, Colombia*

Abstract: We analyse the factors that affect the decision to find employment in the formal or informal sectors taking into account the fact that there is a rationing of formal sector jobs and hence a job queue to enter that sector. Using data from Colombia, we confirm that formal job rationing does exist and that it affects around 62 per cent of the workers that find themselves in the informal labour market. Further factors determining actively seeking formal employment include personal traits, the characteristics of human capital and of the household and the wage premiums associated with working in the formal sector. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: labour informality; rationing of formal sector jobs; endogenous switching regression model; labour segmentation; formal sector queuing

1 INTRODUCTION

The neoclassical view of the labour market is based on the implicit assumption that the demand for labour is broad and diverse and that, as such, it is the worker's characteristics that are the main determinant of employment decisions. Structural economic factors and labour demand factors are both deemed irrelevant; the only relevant factors are those associated with the conditions of supply that determine whether individuals opt for participation or inactivity, employment or unemployment and whether they work in the formal or informal sectors.

However, because this view does not consider the demand factors and hence the restrictions that exist in the creation of jobs (above all in the modern sector of the

*Correspondence to: Gustavo A. García, Research in Spatial Economics (RiSE-group), Department of Economics, Universidad EAFIT, Medellin, Colombia.
E-mail: ggarci24@eafit.edu.co

economy), it has significant limitations when it comes to explaining the heterogeneity present in labour markets, especially the markets of developing countries.

In this paper, we provide an alternative view of the labour market in which employment decisions are considered as being determined by factors of both supply and demand. Thus, although individuals have the opportunity to choose and to decide whether to work or not and to select the type of work they wish to perform, there are restrictions on the range of options available to them. These restrictions correspond to the limitations imposed by the demand side of the labour market, such as the few good quality jobs that are being created.

As Table 1 shows, the labour markets of a group of developing regions and countries present a marked heterogeneity characterized by the scarcity of employment opportunities and the poor quality of the jobs that are available. The informal market provides between 50 and 70 per cent of jobs and is responsible for between 25 and 40 per cent of total output, while the unemployed represent about 7 per cent of the labour force, exceeding 10 per cent in extreme cases such as those of North Africa and Colombia.

Given this great weight of informal economic activities, as well as the fact that the quality of employment constitutes an essential element of life quality and of socioeconomic welfare, in addition to being a mechanism for countering social inequalities, this paper seeks to analyse the basic characteristics and determinants of finding employment in either the formal or informal labour markets, by applying an alternative methodology that assumes the rationing of formal jobs. To date, the literature has adopted models that fail to take into account formal job rationing, so that being employed in the formal or informal sectors is deemed an exclusively individual decision dependent solely on that individual's personal characteristics and preferences. The alternative methodology proposed here is based on the so-called 'job queue' approach and takes into account the criteria applied by employers in the job placement process when recruiting applicants for a limited number of jobs. The job placement process, then, can be seen to comprise two mechanisms: one is the workers' choice to actually join the job queue for a formal job and the other is the mechanism applied by employers in choosing workers from that job queue to fill a vacancy.

This analysis of the determinants of job placement (informality vs formality) is based on the estimation of an endogenous switching regression model. This model, in addition to taking into consideration the dual mechanism outlined in the preceding texts in the job finding process, includes as additional determinants the unbiased predictions of the wage premiums resulting from working in the formal sector. In our analysis, we consider workers to be informally employed if they are not covered by the health insurance and

Table 1. Informality and unemployment in developing regions (per cent)

Region	Informality	Unemployment	Informal GDP
Asia (south–southeast)	69.9	4.6	32.3
North Africa	47.3	11.3	28.0
Latin America	54.2	6.2	41.1
Argentina	46.9	7.5	25.3
Brazil	38.4	5.9	39.0
Mexico	54.2	4.9	30.0
Ecuador	52.2	4.2	32.4
Colombia	56.8	10.5	37.3

Source: informality: Jütting and De Laiglesia (2009) and ILO (2012). Unemployment: World Bank data, referring to 2013. Informal GDP: Schneider, Buehn and Montenegro (2010).

DOI: 10.1002/jid

pension system, which is a legalistic definition regularly used to analyse the informal sector in developing countries (see for instance, Maloney, 1999; Saavedra & Chong, 1999; Pratap & Quintin, 2006; Herrera-Idárraga, López-Bazo & Motellón, 2015; García, 2017). The estimations are conducted for the job market in Colombia, which presents an interesting case for analysing heterogeneity in the informal sector in developing countries. As we have seen in Table 1, Colombia provides rich evidence from a large and heterogeneous informal sector and suffers marked quantitative and qualitative imbalances, that is, significant levels of informality (56.8 per cent) combined with a high rate of unemployment (10.5 per cent). These characteristics of the labour market in Colombia show that there is a limited creation of formal jobs in the economy, which can affect the decisions that individuals make about the type of employment (formal–informal) they choose.

The rest of the paper is organized as follows. Section 2 discusses the econometric model used. Section 3 describes the estimation procedure and discusses some statistical tests to validate the results. In section 4, the data are presented, and the results of the econometric model are discussed in section 5. Finally, section 6 presents our conclusions.

2 ECONOMETRIC MODEL

From a neoclassical perspective, labour market decisions are determined by the worker's characteristics, that is, the conditions of supply. Thus, agents voluntarily choose from the options available to them: participation or inactivity, employment or unemployment, formal or informal work, etc, following a process of welfare maximization subject to a number of restrictions imposed by the individual, such as the availability of time, initial wealth, human capital endowment, among others. From this perspective, the structure of the economy and the characteristics of labour demand are not taken into account among the factors that determine employment or unemployment.

An alternative approach to the labour market might allow the workers to choose, to decide for themselves, in line with neoclassical thinking, but here the range of options is determined by the constraints imposed by the demand side. This alternative perspective is not that labour demand is broad and diverse, as is assumed in neoclassical theory, but on the contrary, that there are restrictions on jobs, primarily formal jobs, as the modern sector only demands the jobs that it needs (Archibald, 1977; Dickens & Lang, 1985; García, 2017; Lang & Dickens, 1987; Uribe & Ortiz, 2006; Uribe, Ortiz & García, 2007).

Because there are restrictions on the number of jobs, the decision of wanting to work does not necessarily mean the individual will be contracted; that is, the probability of wanting a job in a given sector is not equal to the probability of being employed in that sector. Moreover, taking into account the possibility of excess demand for a type of job ('a job queue'), finding a worker employed in a particular job depends on the job rationing rules used by employers as well as the preferences of workers for that post.

The possibility that there is a formal sector queuing in developing countries is consistent with the arguments of the seminal works of Roy (1951), Harris and Todaro (1970), Mincer (1976) and Dickens and Lang (1985). These authors suggested the existence of segmented labour markets, where there are two distinct sectors of the labour market with different wage-setting mechanisms. The first sector is characterized by relatively high wages, attractive employment conditions and opportunities for advancement into higher paying jobs (primary or modern jobs), whereas in the second sector, the wages are low, there

DOI: 10.1002/jid

are bad working conditions and little opportunities for advancement and the employments are unstable (secondary or precarious jobs). In the Harris–Todaro (1970) model, the primary sector represents the formal sector and the secondary sector is the informal sector, and the latter sector arises as a consequence of fixing wages above the market-clearing rate, which can lead to unemployment in equilibrium, where unemployment goes undercover in the informal sector. The existence of higher wages in the formal sector may attract workers from the informal sector who may not find a position in the formal sector due to formal jobs rationing in the economy. This leads to a formal sector queuing and informality acts as an escape from unemployment.

In more recent research, Maloney (1998) in Mexico; Pisani and Pagán (2003) in Nicaragua; Soares (2004) in Brazil; Co, Gang and Yun (2005) in Hungary; and Puentes and Contreras (2009), Contreras, De Mello and Puentes (2008) and Contreras, Gillmore and Puentes (2017) in Chile have shown evidence for the presence of a queue for entry into the formal sector. In these papers, several explanations are mentioned for the presence of a formal sector queuing. For instance, Maloney (1998) states that there is an important heterogeneity in the informal sector, as part of this sector follows the dualistic conception of the labour market, remaining unprotected, with bad labour conditions and queues for entry into the formal sector, while another fraction of the individuals in the informal sector are informal by choice, and this status does not imply inferior work to them. Pisani and Pagán (2003) suggest that given that in developing countries there is no unemployment insurance supporting the job search, potential formal sector job seekers must wait in the informal sector (or queue for formal sector employment) until a formal sector position becomes available. Soares (2004) and Asea (1996) argue that a formal sector queuing is evidence of possible rigidities and market imperfections in the labour market due to excessive regulation.

In order to model the rationing of formal jobs, this paper follows the model of job queues proposed by Abowd and Farber (1982) and extended by Mengistae (1999). In the first of these papers, a model is developed to analyse the determinants of the union status of workers that enable the possibility of queuing for union jobs, while the second, which employs Abowd and Farber's model, analyses the existence of a job queue in the public sector and the wage differential between this and the private sector. Taking this same approach to the analysis of labour market constraints and the existence of job queues for formal jobs, we find the works of Maloney (1998), Soares (2004), Puentes and Contreras (2009), Contreras *et al*. (2008) and Contreras *et al*. (2017).

From this perspective, it is assumed that workers freely choose the sector of the labour market in which they want to work; however, this choice is constrained by the employment options available in the market. The structural constraints in the economy limit the number of formal jobs, and as these are the ones that most workers prefer, given that they offer the best work conditions (stability, social security, opportunities for promotion and rewards for human capital accumulation), a queue of applicants forms for these jobs (Dickens & Lang, 1985; Lang & Dickens, 1987). The existence of a job queue for the formal sector relaxes the assumption that the probability of wanting to be employed in the formal sector is equal to the probability of obtaining a formal job. Seen from this perspective, a worker's employment in the formal sector depends on two decision processes: the worker's decision to join the queue for a formal job and the employer's decision to hire a worker who is in the queue.

Let us analyse the first of these decisions: joining the job queue for formal employment. A worker's decision to join the queue to obtain a formal job depends on the utility that

DOI: 10.1002/jid

such a post represents to him or her. A major determinant in this decision and in utility maximization is the wage premium to be gained from being employed in one sector or the other. Let W_{1i} be the hourly wage earned by individual i in the formal sector and W_{2i} the hourly wage earned in the informal sector. We also define U_{1i} as the maximum utility individual i can be obtain in the formal sector and U_{2i} as the maximum utility the individual can be obtain in the informal sector. Thus, it is assumed that individual i will prefer a formal job and will queue for such a job if $V_{1i}^{*} = U_{1i} - U_{2i} > 0$. Assuming that V_{1i}^{*} depends linearly on the wage premium provided by the formal sector ($\ln W_{1i}$- $\ln W_{2i}$) and on a set of individual traits and characteristics of his or her working environment and family (X_{1i}), the following job queue equation can be written:

$$V_{1i}^{*} = \alpha_1(\ln W_{1i} - \ln W_{2i}) + X_{1i}^{'}\beta_1 + u_{1i} \tag{1}$$

where α_1 is a constant to be estimated, β_1 is a vector of coefficients to be estimated, and u_1 is a random error term distributed with mean zero and variance $\sigma_{u_1}^2$.

Because V_{1i}^{*} is not directly observable and only its sign can be observed, a binary variable I_{1i} can be defined that captures an individual's willingness to obtain a formal job, as follows:

$$I_{1i} = 1 \text{ if } V_{1i}^{*} = \alpha_1(\ln W_{1i} - \ln W_{2i}) + X_{1i}^{'}\beta_1 + u_{1i} > 0$$

$$I_{1i} = 0 \text{ if } V_{1i}^{*} = \alpha_1(\ln W_{1i} - \ln W_{2i}) + X_{1i}^{'}\beta_1 + u_{1i} < 0 \tag{2}$$

Hence, I_1 is a binary variable equal to 1 if the worker joins the queue for a formal job and equal to 0 if the worker prefers not to join the queue. This equation is normally noted as being in the queue (IQ).

The second decision, that is, that taken by employers when selecting workers from the job queue to fill the formal vacancy, can also be modelled on the basis of a latent variable V_{2i}^{*}. It is assumed that the objective of the employers in the formal sector is to maximize worker productivity per unit spent on this factor. Thus, the employer will hire a given individual ($V_{2i}^{*} > 0$), taking into consideration characteristics of productivity (X_{2i}) and the absolute costs the employer expects to incur for hiring said individual ($E(\ln W_{1i})$). This can be expressed as

$$V_{2i}^{*} = \alpha_2 E\left(\ln W_{1i} | V_{1i}^{*} > 0\right) + X_{2i}^{'}\beta_2 + u_{2i} \tag{3}$$

where i represents the group of workers queuing for a job, α_2 is a constant to be estimated, $E\left(\ln W_{1i} | V_{1i}^{*} > 0\right)$ represents the expected wage a worker could earn in the formal sector after having joined the queue, β_2 is a vector of coefficients to be estimated, and u_2 is a random error term distributed with mean zero and variance $\sigma_{u_2}^2$. This equation is often known as being chosen from the queue (CFQ).

As in the job queue equation, a binary variable I_{2i} can be defined that captures the sign of V_{2i}^{*} and which also captures the selection process undertaken by the employer when choosing a worker from the queue, which is

$$I_{2i} = 1 \text{ if } V_{2i}^{*} = \alpha_2 E\left(\ln W_{1i} | V_{1i}^{*} > 0\right) + X_{2i}^{'}\beta_2 + u_{2i} > 0$$

$$I_{2i} = 0 \text{ if } V_{2i}^{*} = \alpha_2 E\left(\ln W_{1i} | V_{1i}^{*} > 0\right) + X_{2i}^{'}\beta_2 + u_{2i} < 0 \tag{4}$$

Then, I_2 is a dummy variable equal to 1 if the worker is chosen from the queue to fill the post available in the formal sector or equal to 0 if not selected.

DOI: 10.1002/jid

Together with Equations (1)–(3), the system is completed with the wage equations for the formal and informal sectors:

$$\ln W_{1i} = Z_{1i}^{'}\gamma_1 + \upsilon_{1i} \text{ if } V_{1i}^{*} > 0 \text{ and } V_{2i}^{*} > 0 \tag{5}$$

$$\ln W_{2i} = Z_{2i}^{'}\gamma_2 + \upsilon_{2i} \text{otherwise}, \tag{6}$$

where Z_{ji} is a vector of observable individuals and labour characteristics; γ_1 and γ_2 are vectors of coefficients to estimated; and υ_{1i} and υ_{2i} are *iid* normal with mean zero and variances $\sigma_{\upsilon_1}^2$ and $\sigma_{\upsilon_2}^2$, respectively. Assuming the joint distribution of the error terms $(u_{1i}, u_{2i}, \upsilon_{1i}, \upsilon_{2i})$, Mengistae (1999) shows that the covariance matrix has the form:

$$\begin{bmatrix} \sigma_{u_1}^2 & \sigma_{1u_2} & \sigma_{1\upsilon_1} & \sigma_{1\upsilon_2} \\ & \sigma_{u_2}^2 & \sigma_{2\upsilon_1} & \sigma_{2\upsilon_2} \\ & & \sigma_{\upsilon_1}^2 & \sigma_{\upsilon_1\upsilon_2} \\ & & & \sigma_{\upsilon_2}^2 \end{bmatrix}$$

where $\text{Cov}(u_{1i}, \upsilon_{1i}) = \sigma_{1\upsilon_1}$, $\text{Cov}(u_{1i}, \upsilon_{2i}) = \sigma_{1\upsilon_2}$, $\text{Cov}(u_{2i}, \upsilon_{1i}) = \sigma_{2\upsilon_1}$, $\text{Cov}(u_{2i}, \upsilon_{2i}) = \sigma_{2\upsilon_2}$, $\text{Cov}(\upsilon_{1i}, \upsilon_{2i}) = \sigma_{1\upsilon_1}\sigma_{\upsilon_1\upsilon_2}$ and $\text{Cov}(u_{1i}, u_{2i}) = \sigma_{1u_2}$.[1]

The OLS estimation of the wage equations in the preceding texts, regardless of the sector in which the workers choose to seek work and the employers' selection decisions, can lead to bias. $\ln W_1$ and $\ln W_2$ are censored variables because the former is only observed for formal workers and the latter is solely for informal workers, so that the disturbance terms υ_{1i} and υ_{2i} in Equations (5) and (6), respectively, have a truncated distribution and the OLS estimates would not be consistent for the existence of a selection bias problem (Heckman, 1979; Huguet, 1996).

In order to fully identify the system of equations, we first need to deduce the wage that a worker in the queue might expect to earn in the formal sector [i.e. the expression $E\left(\ln W_{1i} | V_{1i}^{*} > 0\right)$ in the Equation (3)]. Using Equations (1)–(5), we obtain:

$$E\left(\ln W_{1i} | V_{1i}^{*} > 0\right) = Z_{1i}^{'}\gamma_1 + \sigma_{1\upsilon_1}\lambda_{1i} \tag{7}$$

where λ_{1i} is analogous to the inverse Mills ratio and its structure depends on whether ρ is equal to or different from zero. Its expression is developed in the following section.

The full model, therefore, comprises Equations (1), (3), (5) and (6). The system is endogenous given that in the job-finding equations [Equations (1) and (3)], wages are included [Equations (5) and (6)] as additional determinants, while the wage equations are in turn determined by the job-finding decisions of the workers and the employers' selection decisions. Such models receive the name of endogenous switching regression models (Maddala y Nelson, 1975). The following section provides a discussion of the estimation strategy for models of this type, which depends on the degree of observability of individuals who are in the job queue and the assumption regarding the decision-making process: whether it is sequential or simultaneous.

[1]In section 3.1, σ_{1u_2} is referred to ρ. Mengistae (1999) assumes that this parameter is equal to zero, and therefore, his estimation follows a sequential structure.

DOI: 10.1002/jid

3 ESTIMATION PROCEDURE AND THE TEST ON THE EXISTENCE OF THE JOB QUEUE

In order to estimate the endogenous switching model, a three-step procedure is followed. In the first step, the reduced-form bivariate probit model (which does not include the wage differential between sectors) is estimated and the residuals are taken to approximate the expectations of v_1 and v_2 conditional upon u_1 and u_2. These approximations, which are nothing other than the inverse Mills ratios, can enter the wage equation for each of the sectors and correct the selection bias in the second step. Finally, in the third step, the wage differential, estimated in the second step, is included as an additional regressor in the 'in-queue' (IQ) equation. Likewise, the wage that a worker could earn in the formal sector, as estimated from the wage equations, is included in the CFQ equation.

To identify the system, a number of exclusion restrictions need to be considered for the model. On the one hand, to be able to identify the parameters of the IQ and CFQ equations, variables that affect the former decision but not the latter, and vice versa, need to be incorporated. On the other hand, to identify the parameters of the wage equations, at least one variable that determines wages but does not affect the IQ and CFQ decisions needs to be included.

Another important aspect to consider when estimating the bivariate probit model is the partial observability of the data.[2] Household surveys usually only provide information about the workers' current situation, that is, whether they are employed in the informal or formal sector, but they provide no details about the decision process that leads an individual to opt for formality or informality. This means that it is not possible to observe the different types of informal workers that might emerge from the decision process, that is, workers that choose the informal sector and who do not queue for formal employment or workers who are in the queue but are not selected and have to opt for the informal sector. In terms of the equations described in the previous section, no information is available regarding the IQ (I_1) and CFQ (I_2) decisions. Only the product of these two decisions can be observed ($I = I_1 I_2$). If this product is equal to 1, the worker is in the formal sector, and if it is equal to 0, the worker is in the informal sector.

To overcome this lack of information, Poirier (1980) and Abowd and Farber (1982) propose using the same dependent variable for the IQ and CFQ equations and differentiating between the characteristics that affect each of the equations. Thus, the exclusion restrictions acquire considerable importance for identifying the effects.

One aspect that differentiates Poirier's method from that of Abowd–Farber is the assumption concerning the decision-making process. The first author assumes that both decisions are made simultaneously, while the latter assumes that the decision process is sequential. According to Maddala (1983), the choice of one of these models or the other depends on the purpose for which it is required. If the analysis seeks to examine the factors that influence the employers' decision to hire a specific type of worker, the simultaneous model would be the most appropriate, whereas if it seeks to analyse the whole process of worker placement between the formal and informal sectors, taking into account the factors that affect both the preferences of workers and employers, the sequential model would be the most appropriate. The second approach, therefore, seems to be the most

[2]A more detailed explanation on the partial observability of the data can be found in Abowd and Farber (1982), Meng and Schmidt (1985) and Tunali (1986).

DOI: 10.1002/jid

suitable for the case we deal with here; however, as far as possible, both methods will be followed so as to observe the extent to which their respective results are consistent.

3.1 Estimation

As mentioned in the preceding texts, the estimation of the model depends on the assumption that it is made regarding the decision-making process: whether it is sequential or simultaneous. In the Abowd–Farber model, it is assumed that decisions are made sequentially, which implies that these decisions are independent. This in turn suggests that the joint distribution function of the disturbance terms of the IQ and CFQ equations (u_1 and u_2, respectively) follows a bivariate normal distribution with zero means and unit variances and covariances equal to zero. Given this structure, the likelihood function with partial observability to be maximized is (Meng & Schmidt, 1985; Tunali, 1986)

$$L_1 = \prod_{I=1} \left[\Phi\left(X'_{1i}\beta_1\right)\Phi\left(X'_{2i}\beta_2\right)\right] \prod_{I=0} \left[1 - \Phi\left(X'_{1i}\beta_1\right)\Phi\left(X'_{2i}\beta_2\right)\right] \tag{8}$$

where Φ is the normal cumulative distribution and X_1 and X_2 are the explanatory variables of the IQ and CFQ equations, respectively. As in the first step, the reduced-form bivariate probit model is estimated; we also include as explanatory variables those variables assumed only to affect wages.

By contrast, Poirier's model assumes that the process of choosing between the formal and informal sectors is simultaneous, which implies that the correlation between the disturbance terms of the two equations is different from zero. Under this assumption, therefore, the likelihood function to be maximized has the following form (Meng & Schmidt, 1985; Tunali, 1986):

$$L_2 = \prod_{I=1} \left[\Phi_2\left(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho\right)\right] \prod_{I=0} \left[1 - \Phi_2\left(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho\right)\right] \tag{9}$$

where Φ_2 is the bivariate normal cumulative distribution, ρ is the covariance between u_1 and u_2, and we assume that $\sigma^2_{u_1} = \sigma^2_{u_2} = 1$.

Note that in the absence of a correlation between the disturbance terms of the two equations, ρ is equal to zero in Equation (9) and the simultaneous model becomes a sequential model.

Continuing therefore to the second step in the estimation (as outlined at the beginning of this section), the wage equations must be estimated next, while correcting for bias due to the censoring of the data generated by workers choosing to work in either the formal or informal sectors. To correct for this selection bias, the procedure proposed by Maddala (1983, pp. 278–283), which is an extension of the Heckman–Lee two-stage estimation method (Heckman, 1976; Heckman, 1979; Lee, 1978), is adopted. Assuming the disturbances in the formal workers' wage equation to be normally distributed, the first moment of the truncated distribution is given by the following expression (Mohanty, 2001; Tunali, 1986):

$$E(v_{1i}|I_{1i} = 1, I_{2i} = 1) = \sigma_{1v_1}\lambda_{1i} + \sigma_{2v_1}\lambda_{2i} \tag{10}$$

where λ_{1i} and λ_{2i} are analogous to the inverse Mills ratios in the univariate context. These ratio values depend on the existence or otherwise of a correlation between the IQ and CFQ

DOI: 10.1002/jid

decisions, that is, whether or not $\mathrm{Cov}(u_{1i}, u_{2i}) = \rho = 0$. Assuming that the preceding expression is different from zero, then (Tunali, 1986):

$$\lambda_{1i} = \frac{\phi\left(X'_{1i}\beta_1\right)\Phi\left(X'_{2i}\beta_2 - \rho X'_{1i}\beta_1\right)/\sqrt{1-\rho^2}}{\Phi_2\left(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho\right)} \tag{11}$$

$$\lambda_{2i} = \frac{\phi\left(X'_{2i}\beta_2\right)\Phi\left(X'_{1i}\beta_1 - \rho X'_{2i}\beta_2\right)/\sqrt{1-\rho^2}}{\Phi_2\left(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho\right)} \tag{12}$$

Now, assuming independence between these decisions, then ρ = 0, so that the expressions in the preceding texts can be reduced to

$$\lambda_{1i} = \frac{\phi\left(X'_{1i}\beta_1\right)}{\Phi\left(X'_{1i}\beta_1\right)} \tag{13}$$

$$\lambda_{2i} = \frac{\phi\left(X'_{2i}\beta_2\right)}{\Phi\left(X'_{2i}\beta_2\right)} \tag{14}$$

Thus, the consistent estimation of the wage equation for workers in the formal sector is obtained by incorporating λ_{1i} and λ_{2i} as additional regressors, that is, by estimating the following equation:

$$\ln W_{1i} = Z'_{1i}\gamma_1 + \sigma_{1v_1}\lambda_{1i} + \sigma_{2v_1}\lambda_{2i} + \varepsilon_{1i} \tag{15}$$

where $\varepsilon_{1i} = v_{1i} - \sigma_{1v_1}\lambda_{1i} - \sigma_{2v_1}\lambda_{2i}$ and $E(\varepsilon_{1i}|I_i = 1) = 0$. Following the Heckman–Lee method, an approximation of the expressions of λ_{1i} and λ_{2i} are the estimates from the bivariate probit, which gives consistent estimates of γ_1. That is, applying OLS to the following expression:

$$\ln W_{1i} = Z'_{1i}\gamma_1 + \sigma_{1v_1}\hat{\lambda}_{1i} + \sigma_{2v_1}\hat{\lambda}_{2i} + \eta_{1i} \tag{16}$$

where $\eta_{1i} = \varepsilon_{1i} + \sigma_{1v_1}\left(\lambda_{1i} - \hat{\lambda}_{1i}\right) + \sigma_{2v_1}\left(\lambda_{2i} - \hat{\lambda}_{2i}\right)$.

By following the Heckman–Lee two-stage estimation method for the informal workers' wage equation ($I = 0$), problems arise in terms of identifying the different types of informal workers; as a result, the sample selection rules are more complicated than they are for the case of formal workers. Using Poirier's model (simultaneous case), two types of informal worker can, in principle, be differentiated: those that are in the job queue but who are not chosen by the employers ($E(v_{2i}|I_{1i} = 1$ and $I_{2i} = 0)$) and those that have opted not to join the queue ($E(v_{2i}|I_{1i} = 0)$). In addition, there is another type of informal worker, in this case not very clearly defined, who while not having joined the queue may yet be selected for a formal job ($E(v_{2i}|I_{1i} = 0$ and $I_{2i} = 1)$).[3] Given this problem of identifying types of informal workers, it is not possible to either calculate the inverse Mills ratio or therefore correct the selection of the wage equation of the informal workers with Poirier's model.

[3]It is possible to determine the sample selection rules and therefore to calculate the inverse Mills ratios, in the case of formal workers, that is, when $I = I_1 I_2 = 1$. This product is equal to 1 only when the worker is in the job queue and is chosen by the employer to fill a formal position ($I_1 = 1$ and $I_2 = 1$). However, in the case of informal workers, the sample selection rules cannot be identified. We observe an informal worker when $I = I_1 I_2 = 0$. Now, if we assume that the decision-making process is simultaneous, there are three possible combinations for which this last product is equal to zero. These are (i) when a worker opts not to join the queue, and he or she is not chosen by the employers ($I_1 = 0$ and $I_2 = 0$); (ii) when a worker has opted to join the queue, but he or she

DOI: 10.1002/jid

In the case of the sequential model (Abowd–Farber's model), Mengistae (1999, pp. 11–12) proposes a solution for correcting the selection bias in the wage equation of the informal workers. The author shows that the expected wage of informal workers ($I=0$) can be estimated as the weighted average of the wage of workers that are informal because they have not joined the job queue ($I_1=0$) and the expected wage of workers who are informal as they have not been chosen from the job queue despite having joined it ($I_1=1$ and $I_2=0$). That is:

$$(1-\pi)E\left(W_{2i}^{1}|I_{1i}=0\right)+\pi E\left(W_{2i}^{2}|I_{1i}=1 \text{ e } I_{2i}=0\right) \tag{17}$$

with π is the proportion of informal workers that are in the job queue and that are not chosen and $(1-\pi)$ is the proportion of informal workers that are not in the queue. The expression in the preceding texts can be written as

$$\ln W_{2i}=Z_{2i}^{'}\gamma_2+\sigma_{1v_2}\lambda_{3i}+\delta_1\lambda_{1i}^{*}+\delta_2\lambda_{4i}+\varepsilon_{2i} \tag{18}$$

where $\lambda_{3i}=-\phi\left(X_{1i}^{'}\beta_1\right)/\left[1-\Phi\left(X_{1i}^{'}\beta_1\right)\right]$, $\lambda_{4i}=-\phi\left(X_{2i}^{'}\beta_2\right)/\left[1-\Phi\left(X_{2i}^{'}\beta_2\right)\right]$, $\lambda_{1i}^{*}=\lambda_{1i}-\lambda_{3i}$, $\delta_1=\pi\sigma_{1v_2}$, $\delta_2=\pi\sigma_{2v_2}$, $\varepsilon_{2i}=(1-\pi)\varepsilon_{2i}^{1}+\pi\varepsilon_{2i}^{2}$, ε_{2i}^{1} and ε_{2i}^{2} are the disturbances in the wage equation of $I_{1i}=0$, $I_{1i}=1$ and $I_{2i}=0$, respectively.

Consistent estimates of γ_2, σ_{1v_2}, δ_1 and δ_2 can be obtained by estimating by OLS the following equation:

$$\ln W_{2i}=Z_{2i}^{'}\gamma_2+\sigma_{1v_2}\hat{\lambda}_{3i}+\delta_1\hat{\lambda}_{1i}^{*}+\delta_2\hat{\lambda}_{4i}+\eta_{2i} \tag{19}$$

where $\eta_{2i}=\varepsilon_{2i}+\sigma_{1v_2}\left(\lambda_{3i}-\hat{\lambda}_{3i}\right)+\delta_1\left(\lambda_{1i}^{*}-\hat{\lambda}_{1i}^{*}\right)+\delta_2\left(\lambda_{4i}-\hat{\lambda}_{4i}\right)$ and the inverse Mills ratios can be approximated by estimating the bivariate probit Abowd–Farber model.

With the Heckman–Lee two-stage procedure, consistent estimates of the parameters are obtained; however, if there is selection bias, the standard errors of these estimates will not be consistent. We used bootstrap techniques to calculate these standard errors. Notice that Tunali (1986) and Mengistae (1999) both propose methods for correcting the standard errors, but their procedure is cumbersome.

In the third step of the endogenous switching model, the IQ and CFQ equations are estimated with a structural bivariate probit model. As mentioned earlier, in the first equation, the estimated difference in wages is included as an additional regressor, and in the second equation, the estimated wages in the formal sector are added. The structural bivariate probit can only be estimated for the Abowd–Farber's model because for Poirier's model, estimates corrected by selectivity of the informal sector wages cannot be obtained.

A final aspect to take into account in estimating the endogenous switching model is that of exclusion restrictions. There are two sets of restrictions: the first ones are those that need to be imposed to distinguish between the IQ and CFQ equations, and the second ones are those that make the identification of the wage equations possible.

For the first set of restrictions, it is assumed that the factors that influence the decision to join the queue for a formal job are closely related to personal characteristics, above all to those related to family and household. At the same time, the employers' hiring decisions

is not chosen by the employers ($I_1=1$ and $I_2=0$); and (iii) when a worker has not opted to join the queue, but he or she is chosen by the employers ($I_1=0$ and $I_2=1$). The two first sample selection rules can be identified; that is, it is possible to identify these individuals in the data. Nevertheless, it is not possible to identify the last sample selection rule because, if the selection process is simultaneous, these individuals cannot be detected in the data, with which the inverse Mills ratios in the case of Poirier's model cannot be calculated and the correction of the sample selection of the wage equation is not possible.

DOI: 10.1002/jid

are assumed to be more closely affected by the endowment of human capital, including such factors as education, previous employment status, the degree of commitment to work and the specific job sector. Thus, the variables excluded from the second equation but included in the first are number of years in the current job (i.e. tenure) and a number of household variables, including number of children, occupational status of other members of the household and educational level of the household.

Let us examine the assumptions underpinning the inclusion or exclusion of some of the aforementioned variables. The variable of tenure is excluded from the CFQ equation. This exclusion assumes that employers consider the experience a worker has acquired throughout the worker's life and not just what was gained in one particular job (Abowd & Farber, 1982; Farber, 1983; Mengistae, 1999; Soares, 2004). On the other hand, the variables capturing a worker's position within the household and his or her marital status are included so as to measure two distinct effects. In the CFQ equation, these variables represent the worker's degree of commitment to work, to the extent that the responsibilities assumed by heads of households can be related to a greater commitment to their work; and in the IQ equation, they represent the traditional characteristics of the individual, which, depending on his or her position and responsibilities within the family structure, will reflect their commitment to the taking on of a particular job.

At the same time, the inclusion of household variables in the first equation seeks to account for the costs incurred by an individual that opts to work in either the formal or informal sectors. The number of children is a proxy for the level of household responsibility, especially as far as women are concerned. Thus, an informal job might be considered more beneficial, as it is more likely to offer a flexible timetable so that the workers can distribute their time more easily between childcare and work.

As for the second set of exclusion restrictions used for identifying wage equations, it is assumed that household characteristics and previous employment status have no effect on the wage levels of formal and informal workers. Likewise, it is assumed for identification that the specific job sector only affects wage levels, and so these variables are excluded from the structural bivariate probit model.

3.2 The Existence and Length of the Formal Job Queue

In order to test for the existence of a formal job queue, we calculate the tests proposed by Abowd and Farber (1982) and Mengistae (1999). These tests, aside from validating the robustness of the bivariate model with respect to the univariate model, provide evidence of the barriers to obtaining a formal job. In other words, the existence of a job queue means that not all the workers who want to work in the formal sector can do so, due to the restrictions imposed by their own characteristics and the limitation on the number of formal jobs that are in fact available.

Abowd and Farber (1982) propose evaluating the existence of a job queue based on the estimates of a first-stage bivariate probit model. In the absence of a job queue, the parameters of the CFQ equation are restricted to zero with the exception of the constant term. Based on these restricted and unrestricted estimates of the bivariate probit model, the null hypothesis of the nonexistence of the queue based on a likelihood ratio (LR) statistic can be tested. A further test that could be applied (the opposite to that of the no-queue model) is that of the existence of a universal queue. A universal queue is equivalent to the restriction that all the parameters of the IQ equation are zero with the exception of

DOI: 10.1002/jid

the constant term. As in the first test, we can test the null hypothesis of the existence of a universal queue with an LR statistic.

Mengistae (1999) proposes tests based on estimates of the conditional wage equations. The tests seek to examine hypotheses about the parameter π of Equation (18), which represents the proportion of informal workers that have been rationed out of the formal sector. Testing the nonexistence of a job queue involves testing the null hypothesis of $\pi = 0$, while testing $\pi = 1$ implies corroborating the existence of a universal queue. The first test is performed with a joint significance test of δ_1 and δ_2 (Ho: $\delta_1 = \delta_2 = 0$), while the second the restriction of $\sigma_{1v_2} = \delta_1$ is tested by using an F-statistic. The limitation of this test is that it cannot be applied to Poirier's model because, as mentioned in the previous section, the wage equation of the informal workers cannot be estimated.

The length of the job queue can be calculated from the probabilities resulting from the first-stage bivariate probit estimation. Farber (1983) and Venti (1987) suggest that the length of the job queue (q) is calculated as the inverse of the average probability of being chosen from the queue given that the worker is in the queue [P(CFQ)]. Likewise, the probabilities of joining the queue [P(IQ)] and of working in the formal sector [P(formal)] can also be deduced. Depending on the structure of the model (sequential or simultaneous), the probabilities in the preceding texts have different structures, namely

	Sequential	Simultaneous
$\mathrm{P(IQ)} = \mathrm{Prob}\left(V_{1i}^* > 0\right)$:	$\Phi\left(X_{1i}'\hat{\beta}_1\right)$	$\Phi\left(X_{1i}'\hat{\beta}_1\right)$
$\mathrm{P(CFQ)} = \mathrm{Prob}\left(V_{2i}^* > 0 \mid V_{1i}^* > 0\right)$:	$\Phi\left(X_{2i}'\hat{\beta}_2\right)$	$\Phi_2\left(X_{1i}'\hat{\beta}_1, X_{2i}'\hat{\beta}_2, \rho\right)/\Phi\left(X_{1i}'\hat{\beta}_1\right)$
$\mathrm{P(formal)} = \mathrm{Prob}\left(V_{1i}^* > 0 \text{ y } V_{2i}^* > 0\right)$:	$\Phi\left(X_{1i}'\hat{\beta}_1\right)\Phi\left(X_{2i}'\hat{\beta}_2\right)$	$\Phi_2\left(X_{1i}'\hat{\beta}_1, X_{2i}'\hat{\beta}_2, \rho\right)$
$\mathrm{q} = 1/\frac{1}{n}\sum_{i=1}^{n}\mathrm{P(CFQ)}$:	$1/\frac{1}{n}\sum_{i=1}^{n}\Phi\left(X_{2i}'\hat{\beta}_2\right)$	$1/\frac{1}{n}\sum_{i=1}^{n}\Phi_2\left(X_{1i}'\hat{\beta}_1, X_{2i}'\hat{\beta}_2, \rho\right)/\Phi\left(X_{1i}'\hat{\beta}_1\right)$

4 DATA AND DESCRIPTIVE EVIDENCE

In this study, we use data from Colombia. It is a medium-income country located in northwestern South America, which has had positive and stable economic growth accompanied by high levels of poverty and inequality, low levels of human capital and poor labour conditions. The annual GDP growth rate was 3.1 per cent in 2015, but the proportion of people living below the poverty line (US$1.90 purchasing power parity per day) was 5.7 per cent and the Gini index was 53 per cent. In terms of education, approximately 57 per cent of the population ages 25 and older in Colombia has reached at least a secondary level of education, which is low when compared with other countries in Latin America. Regarding the labour market indicators, Colombia presents a marked heterogeneity characterized by high levels of labour informality and unemployment. According to the International Labour Organization reports and the World Bank database, Colombia reports an informality rate of around 57 per cent and an unemployment rate of 10.5 per cent, which are the highest in Latin America. For an international comparison of these main economic and social indicators, see Table 2.

DOI: 10.1002/jid

Table 2. Economic and social indicators in Colombia and other countries

	Colombia	Ecuador	Brazil	Argentina	Mexico	USA	UK
GDP per capita (2011 PPP, US$)[a]	12 988	10 776	14 455	19 126	16 490	52 704	38 519
GDP growth (annual, %)[b]	3.1	4.2	−3.8	3.0	2.5	2.6	2.2
Gini index[c]	53.5	45.4	51.5	42.7	48.2	41.1	32.6
Population living below US$1.90 PPP per day (%)[d]	5.7	3.8	3.7	1.7	3.0	—	—
Population with at least some secondary education (% ages 25 and older)[e]	56.3	39.8	53.6	56.9	58.0	95.0	99.9
Informality rate (%)[f]	56.8	52.2	38.4	46.9	54.2	—	—
Unemployment rate (%)[g]	10.5	4.2	5.9	7.5	4.9	7.4	7.5

PPP, purchasing power parity.

[a]Data refer to 2015 (World Bank data).
[b]Refer to 2015 (World Bank data).
[c]Data refer to GINI index calculated by World Bank to the most recent year available during the period 2011–2014 (World Bank data).
[d]Data refer to the most recent year available during the period 2011–2014 (World Bank data).
[e]Data refer to the most recent year available during the period 2005–2013 (UNDP, 2015).
[f]Data refers to 2011 (ILO, 2012).
[g]Refer to 2013 (World Bank data).

The data used in this paper originate from the Great Integrated Household Survey, carried out by the National Administrative Statistics Department and includes all monthly data from the year 2009. This cross-sectional survey has information at a micro-level on labour force, unemployment and informality in 13 major Colombian cities with their metropolitan areas.[4] The sample considered in this paper is composed of individuals between 20 and 60 years old that were not carrying out formal studies, and we further excluded agricultural workers. Our final sample comprises 58 654 individuals.[5]

In the literature, there are at least two approaches frequently used to define informality: the productivity approach and the legalistic approach. The former refers to the characteristics of the production unit and includes such factors as productivity, technology, type of employment, firm size, among others (Hussmanns, 2004), while the latter corresponds to the lack of coverage of workers by mandated labour protection or firms avoiding taxation or other legal regulation (Saavedra & Chong, 1999). In this study, we opt to follow the legalistic definition for two reasons. First, the approach followed here is based on a framework for examining decisions regarding the type of employment that are mainly determined by the characteristics of the individual and the labour supply. By contrast, the definition of informality from a productivity perspective is determined by the characteristics of the production unit, not concerning the individual. Second, the

[4]Namely, Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Cúcuta, Ibagué, Manizales, Medellín, Montería, Pasto, Pereira and Villavicencio. These metropolitan areas represent 45 per cent of the total population and about 60 per cent of the urban population according to the 2005 Population Census.
[5]Notice that we excluded government employees, employers and self-employed people. Given this exclusion, the informality rate may differ from that reported by Jütting and De Laiglesia (2009) or the International Labour Organization (ILO, 2012).

DOI: 10.1002/jid

legalistic definition is much broader because it includes the possible presence of informal employment within large firms, which is not possible in the productivity definition. This can imply that using the productivity definition can yield lower informality rates than those obtained by using the legalistic definition.[6]

In this study, we classify informal workers as those workers who do not have access to the social security system to receive healthcare and a pension. There are several definitions of informality based on the legal definitions, for instance, lack of a signed contract or not having a set of rights and benefits guaranteed by the legal framework (transportation subsidy, right to severance pay, paid vacation, among others).[7] However, we opt here for the social protection definition because it is the most common definition of informality used in the literature given the availability of similar data sources across countries (Bargain & Kwenda, 2014; García, 2017; Khamis, 2012; Maloney, 2004; Marcouiller, Ruiz de Castilla & Woodruff, 1997; Saavedra & Chong, 1999), which allow us to contrast our results with other international studies. Additionally, Bernal (2009) and Galvis (2012) (for the Colombian case) and Khamis (2012) (for the case of Mexico) show that in terms of correlating lack of social security coverage and other legal definitions of informality, there is a significant correlation between these dimensions of informality at a point in time and over time, which suggests that there is a substantial overlap between the social security definition and other legal measures of informality.

Following this social security definition of informality, we have 35 817 (61.1 per cent) formal workers and 22 837 (38.9 per cent) informal workers. We now turn to examine the characteristics of the formal and informal workers, taking into consideration their personal characteristics as well as the characteristics of the households they inhabit and the jobs they perform. Tables 3 and 4 show that formal workers are on average better educated than their informal counterparts, there being a difference of 2.5 years of schooling. Most formal workers have completed secondary education (51 per cent) and tertiary education (37 per cent). Similarly, most informal workers have completed secondary school (56 per cent), but a good proportion only has completed their primary education (28 per cent). As far as gender is concerned, the majority of informal workers are women (52 per cent), while the majority of formal workers are men (56 per cent). It can additionally be noted that most informal workers are not heads of households (61 per cent), which might indicate that the search for a secondary source of household income is conducted in the informal sector.

In terms of age (Table 4), it can be seen that there are no marked differences between formal and informal workers. By contrast, in terms of the number of years of experience acquired in the current job, differences do exist: While formal workers have around 5 years of tenure, informal workers just have 3. As for hourly wages, formal workers earn 30 per cent more than their informal counterparts, which is indicative of the precarious and low levels of productivity that exist in the informal sector.

In the case of the household variables (Tables 4 and 5), it can be seen that informal workers are more likely to live in households with lower levels of education. Likewise,

[6]In fact, the informality rate using the productivity definition proposed by the ILO, which expands informal sector employment to include all own-account workers (excluding administrative workers, professionals and technicians), unpaid family workers and employers and employees working in establishments with fewer than five people, was around six percentage points lower than the rate calculated on the legal definition (65.93 vs 76.95 per cent).

[7]A more detailed discussion can be found in Khamis (2009).

DOI: 10.1002/jid

Table 3. Personal and employment characteristics

	Formal	Informal	Total sample
Education			
Less than primary	0.32%	1.66%	0.77%
Primary	11.43%	27.66%	16.80%
Secondary	51.08%	55.63%	52.59%
Tertiary	37.16%	15.05%	29.84%
Total	100%	100%	100%
Gender			
Female	44.30%	52.25%	42.93%
Male	55.70%	47.75%	53.07%
Total	100%	100%	100%
Head of household			
Nonhead	56.20%	61.39%	57.92%
Head	43.80%	38.61%	42.08%
Total	100%	100%	100%
Marital status			
Unmarried	45.62%	49.39%	46.87%
Married	54.38%	50.61%	53.13%
Total	100%	100%	100%
Occupational status before the current job			
Employed	85.65%	81.17%	84.16%
Unemployed	14.35%	18.83%	15.84%
Total	100%	100%	100%
Sector			
Industry	26.42%	18.98%	23.95%
Construction	4.09%	10.00%	6.04%
Commerce and hotel	22.31%	32.79%	25.78%
Transport and telecommunications	9.65%	5.47%	8.27%
Financial and education	24.53%	8.40%	19.20%
Service	13.00%	24.36%	16.76%
Total	100%	100%	100%

Note: We used personal sampling weight available in the database.

Table 4. Wages and education

	Formal		Informal		Total sample	
	Mean	SD	Mean	SD	Mean	SD
Age	34.57	9.77	34.29	10.36	34.48	9.97
Education (years)	11.07	3.58	8.61	3.78	10.26	3.83
Tenure at job (years)	4.83	5.71	2.90	4.04	4.19	5.30
Education of household (other members)	5.19	2.50	4.54	2.41	5.00	2.49
Real hourly wage	3347.9	1710.0	2353.8	1910.4	2018.9	1876.9

Note: We used personal sampling weight available in the database. The wages are in Colombian pesos. We use the consumer price indices of the biggest cities that were obtained from the National Administrative Statistics Department. In Ibagué, the consumer prices index is not calculated by the National Administrative Statistics Department, so we used the consumer prices index of Pereira given the similarities in population and social and cultural characteristics, as well as proximity between these cities. In December 2009, the exchange rate was 2935 Colombian pesos per euro.

a high proportion of informal workers live in households with school-age children (aged between 6 and 17). An equally important factor characterizing informality is the presence

DOI: 10.1002/jid

Table 5. Household characteristics

	Formal	Informal	Total sample
Presence of children between 0 and 2 years old	10.40%	12.03%	10.94%
Presence of children between 3 and 5 years old	11.97%	13.82%	12.58%
Presence of children between 6 and 10 years old	20.62%	23.35%	21.52%
Presence of children between 11 and 17 years old	26.92%	31.49%	28.43%
Presence of other relatives working as formal	46.86%	30.76%	41.53%
Presence of other relatives working as informal	41.54%	58.40%	47.12%

Note: We used personal sampling weight available in the database.

of other informal workers among members of the same household: 58 per cent of informal workers live in households in which there are more informal workers.

The distribution by economic sector (Table 3) shows that formal workers are concentrated primarily in the industrial (26 per cent) and the financial and education sectors (24 per cent), while there is a prevalence of informal workers in the retail trade and the hotel sector (33 per cent) and in the personal services sector (24 per cent). Finally, according to previous employment status records, around 19 per cent of informal workers were unemployed before entering their current job.

The characterization in the preceding texts shows that informal workers are mainly unmarried women with an average age of 34, nonheads of households and with an educational level that does not extend beyond secondary schooling. The households where informal workers reside are characterized by a high presence of school-age children, poor levels of education (with an average of 4.5 years of schooling) and the presence of other household members working in the informal labour market. As for employment, it has been found that the hourly wage of informal workers is 30 per cent less than the equivalent salary paid to a formal worker and that more than half work in the commerce, hotel and personal service sectors. Informality, therefore, is a sector with precarious working conditions, characterized by an abundance of unskilled labour, concentrated in sectors with a low level of technical development.

5 RESULTS

5.1 Univariate Probit Model: Absence of a Job Queue

The probit estimation in the absence of a job queue offers a good description of the variables to be included among the determinants of the workers' choice of labour sector. The results of this estimation are shown in Table 6. It can be seen that as the workers' level of education improves, the probability of their working in the formal sector rises and that those who have successfully completed higher education have the greatest probability of finding work in this sector. This probability also increases with the number of years of work experience and tenure but at a decreasing rate in the case of the first of these two variables. The coefficient of the gender variable was positive and statistically significant, which is indicative of a certain degree of discrimination in the process of deciding between the formal and informal sectors. In the case of the heads of households and civil status variables, the estimates show that being a head of the household or being married means a worker is more likely to find work in the formal sector. These positive effects may reflect

DOI: 10.1002/jid

Table 6. Estimates of univariate probit model ($Y = 1$ formal, 0 informal)

	Coefficients	z	Marginal effects	z
Personal characteristics				
Primary	0.331[a]	5.24	0.118[a]	5.24
Secondary	0.944[a]	15.09	0.346[a]	15.09
Tertiary	1.525[a]	23.89	0.456[a]	23.89
Age	0.071[a]	15.80	0.027[a]	15.80
Age2	−0.001[a]	−16.58	−0.001[a]	−16.58
Tenure	0.065[a]	38.90	0.024[a]	38.90
Male	0.154[a]	11.57	0.058[a]	11.57
Head of household	0.201[a]	13.98	0.075[a]	13.98
Married	0.123[a]	9.33	0.046[a]	9.33
Unemployed before the current job	−0.279[a]	−15.87	−0.107[a]	−15.87
Household characteristics				
# kids 0–2	−0.068[a]	−3.74	−0.026[a]	−3.74
# kids 3–5	−0.078[a]	−4.68	−0.029[a]	−4.68
# kids 6–10	−0.053[a]	−4.52	−0.020[a]	−4.52
# kids 11–17	−0.095[a]	−10.63	−0.036[a]	−10.63
# of relatives working as formal	0.219[a]	21.46	0.082[a]	21.46
Education of household	0.016[a]	4.38	0.006[a]	5.38
Sector				
Industry	0.353[a]	22.47	0.126[a]	22.47
Transport and telecommunications	0.498[a]	21.29	0.168[a]	21.29
Financial and education	0.571[a]	29.95	0.194[a]	29.95
Services	−0.088[a]	−5.24	−0.033[a]	−5.24
Constant	−2.116[a]	−20.52	—	—
N		58 654		
Log L		−31 341.85		
Pseudo R^2		0.201		
AIC		62 749.7		
BIC		63 046.0		

Note:
[a]Denotes significance at 1 per cent.
[b]Denotes significance at 5 per cent.
[c]Denotes significance at 10 per cent. Robust standard errors. Less than primary education and commerce, hotel and construction sectors as references. We include city dummies.

the fact that the greater responsibilities associated with a household are in turn associated with a greater need for a good job or are a good indication of a worker's commitment to their workplace and hence their suitability to fill a formal position.

The household variables show that the presence of children, especially those in preschool ages (3–5 years) and those that are older (11–17 years), has a negative effect on the probability of finding formal employment. By contrast, the variables associated with a higher educational level and the number of formal workers residing in the household have a positive effect on finding formal employment.

As for the workers' previous employment status, being unemployed prior to finding their current job has a negative effect on the attainment of formal employment, compared with having been previously employed (base category). Finally, the economic sector variable shows that finding work in sectors other than commerce, hotel and construction increases the probability of finding formal employment, with the exception of the personal service sector, which has a negative effect on the probability of being formally employed.

DOI: 10.1002/jid

5.2 Bivariate Probit Model, Reduced Form: The Existence of a Queue for Formal Jobs

The univariate specification implies that the process of being either formally or informally employed in fact omits the selection mechanism used by employers and, as such, the possible formal job rationing that exists. As mentioned in the preceding texts, a more appropriate specification for describing the process of finding work in the presence of job restrictions is the estimation of a bivariate model. The estimates of both the sequential and simultaneous bivariate probit models with partial observability are shown in Table 7.

Interesting patterns are observed for the IQ and CFQ equations. A higher level of education increases the likelihood of joining the queue and of being chosen for a formal job, with the most marked effect being found for those that have successfully completed higher education. The other human capital variables of age and tenure have a positive and decreasing effect on the likelihood of joining the queue.

Regarding the gender variable, being male increases both the likelihood of joining the job queue and of being chosen from it. In the case of the head of household variable, positive effects were recorded in both the worker's decision to seek formal employment and in the employers' selection process. As mentioned in the preceding texts, the head of household variable appears to be associated to a greater degree of work commitment, and so employers consider this factor positively when recruiting. Finally, it is noted that a married individual is more likely than a single person to seek formal employment, and the former is also more likely to be selected for formal employment.

As in the univariate specification, the variable of being unemployed prior to current employment is not positively related to joining the job queue. This outcome might imply that a phase of prior unemployment results in greater pressure to find employment of any kind, so that individuals would rather obtain an informal job quickly than have to go through the process of finding formal employment. Similarly, we find that workers in the commerce, hotel and construction sectors are more likely to be employed informally either because they have not joined the job queue or because they have not been selected from it.

In the case of the variables that are only included in the IQ equation, the results do not differ greatly from those obtained in the univariate specification. The presence of children at both ends of the age spectrum (3–5 and 11–17) in the household has a negative effect, while a higher level of education in the household and the presence of other household members employed in the formal sector have a positive impact on finding work in the formal sector.

The simultaneous estimation shows that the correlation between the two equations is negative (-0.807) and statistically significant. This result indicates that finding employment in either the formal or informal sectors is more of a simultaneous process than a sequential one and that there are unobserved factors that affect the process. Additionally, we test that all exclusion variables are zero at the same time. The results are shown at the bottom of Table 7 and indicate that we reject at the 1 per cent level that all variables have jointly no effect on the IQ equation, which provides evidence that the exclusion restrictions are jointly relevant and determine that an individual queue for a formal job.

5.3 The Formal and Informal Wage Equations

The formal and informal wage equation estimates are shown in Table 8. It should be noted that all the estimated coefficients of the selection correction are statistically significant,

DOI: 10.1002/jid

Table 7. Estimates of bivariate probit model reduced form (Y_j = 1 formal, 0 informal; j = IQ, CFQ)

	Abowd–Farber model (sequential)				Poirier model (simultaneous)			
	IQ		CFQ		IQ		CFQ	
	Coeff	*z*	Coeff	*z*	Coeff	*z*	Coeff	*z*
Personal characteristics								
Primary	0.023	0.13	0.409[a]	3.50	−0.204	−1.12	0.403[a]	3.43
Secondary	0.449[a]	2.60	1.018[a]	8.79	−0.077	−0.42	1.016[a]	8.69
Tertiary	0.902[a]	5.08	1.574[a]	13.20	0.127	0.68	1.646[a]	13.75
Age	0.070[a]	5.53	0.030[a]	2.80	0.080[a]	5.63	0.001	0.932
Age2	−0.001[a]	−7.77	−0.001[b]	−1.66	−0.001[a]	−7.66	0.001	0.369
Tenure	0.437[a]	17.03	—	—	0.345[a]	15.00	—	—
Male	0.244[a]	7.14	0.073[a]	2.76	0.155[a]	4.37	0.095[a]	3.37
Head of household	0.315[a]	9.18	0.024[c]	2.53	0.258[a]	7.51	0.017[b]	1.67
Married	0.091[a]	2.74	0.105[a]	4.38	0.037	1.12	0.110[a]	4.44
Unemployed before the current job	−0.254[a]	−5.33	−0.236[a]	−8.63	−0.131[a]	−2.86	−0.237[a]	−8.18
Household characteristics								
# kids 0–2	−0.110[a]	−3.38	—	—	−0.085[a]	−3.18	—	—
# kids 3–5	−0.110[a]	−3.68	—	—	−0.090[a]	−3.64	—	—
# kids 6–10	−0.062[a]	−2.91	—	—	−0.054[a]	−3.04	—	—
# kids 11–17	−0.136[a]	−8.52	—	—	−0.109[a]	−8.26	—	—
# of relatives working as formal	0.346[a]	15.44	—	—	0.294[a]	16.02	—	—
Education of household	0.031[a]	5.40	—	—	0.025[a]	5.25	—	—
Sector								
Industry	0.236[a]	6.36	0.330[a]	11.19	0.134[a]	3.43	0.300[a]	9.64
Transport and telecommunications	0.381[a]	7.00	0.456[a]	10.08	0.208[a]	3.88	0.468[a]	9.88
Financial and education	0.648[a]	11.35	0.402[a]	10.80	0.411[a]	7.95	0.443[a]	10.82
Services	0.002	0.06	−0.154[a]	−4.96	0.109[c]	2.48	−0.200[a]	−6.11
Constant	−1.788[a]	−6.16	−0.672[a]	−2.83	−1.105[a]	−3.45	−0.097	−0.35
Rho	—	—	—	—	−0.807[a]	−27.18	—	—
N	58 654				58 654			
Log L	−30 809.82				−30 762.91			
AIC	61 737.64				61 737.64			
BIC	61 645.82				61 645.82			
Wald test over exclusion vars = 0	653.93				568.74			
p-value	0.000				0.000			

Note:
[a]Denotes significance at 1 per cent.
[b]Denotes significance at 5 per cent.
[c],Denotes significance at 10 per cent. Robust standard errors. Less than primary education and commerce, hotel and construction sectors as references. We include city dummies. IQ, 'in-queue' equation; CFQ, 'chosen from the queue' equation.

indicating the importance of considering the selection bias in the wage equations. The main implication of the selection bias is that the rates of return to human capital, in general, and to education, in particular, are overestimated with OLS, hence the importance of correcting it when such bias exists.

DOI: 10.1002/jid

Table 8. Wage equations for formal and informal workers with correction for selectivity based on the bivariate probit models (Y = Log real hourly wage)

	OLS without correction				Correction base on the Abowd–Farber's model				Correction base on the Poirier's model	
	Formal		Informal		Formal		Informal		Formal	
	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Primary	0.027	1.24	0.096[a]	5.20	0.051[b]	2.05	0.202[a]	9.11	0.049[b]	2.13
Secondary	0.173[a]	7.96	0.200[a]	10.79	0.237[a]	8.24	0.442[a]	12.25	0.219[a]	8.18
Tertiary	0.542[a]	24.63	0.517[a]	24.90	0.633[a]	19.08	0.926[a]	17.12	0.606[a]	19.30
Age	0.019[a]	13.01	0.016[a]	7.84	0.019[a]	12.46	0.022[a]	9.12	0.018[a]	12.12
Age2	−0.0002[a]	−11.04	−0.0001[a]	−7.01	−0.0002[a]	−10.94	−0.0002[a]	−6.45	−0.0002[a]	−9.74
Tenure	0.011[a]	27.16	0.011[a]	15.32	0.010[a]	19.81	−0.023[a]	−4.79	0.010[a]	18.54
Male	0.037[a]	9.20	0.083[a]	12.47	0.050[a]	11.80	0.155[a]	20.15	0.040[a]	9.25
Head of household	0.020[a]	4.72	0.026[a]	4.31	0.026[a]	6.03	0.065[a]	4.58	0.020[a]	4.77
Married	0.016[a]	4.03	0.041[a]	7.29	0.029[a]	6.81	0.108[a]	15.90	0.020[a]	4.74
Industry	0.069[a]	14.69	−0.006	−0.89	0.071[a]	10.80	0.089[a]	6.49	0.079[a]	12.88
Transport and telecommunication	−0.013[c]	−1.83	−0.069[a]	−4.94	0.0001	0.02	0.032	1.50	0.001	0.12
Financial and education	0.128[a]	23.42	0.144[a]	11.72	0.145[a]	18.87	0.221[a]	11.34	0.139[a]	18.20
Services	0.080[a]	13.02	0.027[a]	3.44	0.034[a]	5.00	−0.241[a]	−25.00	0.070[a]	10.20
Constant	7.233[a]	216.70	7.147[a]	173.45	7.090[a]	134.99	7.194[a]	150.00	7.173[a]	157.66
σ_{1v_1}	—	—	—	—	0.017[c]	1.92	—	—	−0.037[a]	−2.90
σ_{2v_1}	—	—	—	—	0.042[b]	2.33	—	—	0.097[a]	3.07
σ_{1v_2}	—	—	—	—	—	—	−0.183[a]	−10.79	—	—
δ_1	—	—	—	—	—	—	−0.114[a]	−9.62	—	—
δ_2	—	—	—	—	—	—	0.372[a]	7.69	—	—
N	35 817		22 837		35 817		22 837		35 817	
R^2 adjust	0.273		0.180		0.39		0.310		0.273	

Note

[a]Denotes significance at 1 per cent.

[b]Denotes significance at 5 per cent.

[c],Denotes significance at 10 per cent. Robust standard errors. Less than primary education and commerce, hotel and construction sectors as references. All models include city dummies.

J. Int. Dev. 30, 760–789 (2018)
DOI: 10.1002/jid

The corrected estimates obtained with the sequential model show that the human capital variables are important positive factors for determining wages both in the formal and informal sectors. Schooling, especially tertiary education, has the greatest effect. In the informal sector, this positive effect might indicate that the jobs of some of these workers are not secondary jobs; therefore, returns to education exist. Turning to personal characteristics, we find that there are gender wage differences in both sectors, in which heads of households and married people enjoy a wage premium with respect to nonheads of households and to those who are single and that in this last case, the premium is higher in the informal sector.

With the simultaneous model (Poirier's model), it was only possible to correct the formal wage equation (section 3), and it should be noted that there are no marked differences between this and the sequential model correction. The selection coefficients are equally significant, there being positive returns on education, while age and tenure both present positive but concave returns. Positive wage premiums are also found for being male, the head of a household and married.

In addition to being used to corroborate the existence of the queue for formal jobs (next section), the estimates in the preceding texts can be used to calculate the expected wages paid in the formal and informal sectors, which are incorporated as additional regressors in the structural bivariate probit model (section 5.5). In the next section, we analyse the existence of the job queue.

5.4 The Existence and Length of the Formal Job Queue: Results

In order to test the validity of the results obtained with the bivariate probit models and hence the existence of a formal job queue, we calculated the Abowd–Farber and Mengistae tests as described in section 3.2. Based on an LR test, Abowd and Farber (1982) examine the null hypothesis of the no-queue scenario and the existence of a universal queue by using a bivariate probit model. The first hypothesis implies that the coefficients of the CFQ equation are zero with the exception of the constant, while the second restricts the coefficients of the IQ equation to zero. Employing a different test, Mengistae (1999) examines the existence or otherwise of formal job rationing by conducting tests on the π parameter that appears in the conditional wage equations. The null hypothesis of the existence of a job queue is $\pi = 0$ and the null hypothesis of a universal queue is $\pi = 1$. The first hypothesis is checked by using a test of joint significance of $\delta_1 = \delta_2 = 0$, and the second is verified with an F-test of $\sigma_{1v_1} = \delta_1$. These tests are presented in Table 9.

Table 9. Test for universal queue and no queue hypothesis

	Abowd–Farber test		Mengistae test	
Ho	Abowd–Farber model	Poirier model	Abowd–Farber model	Poirier model
No queue	LR $\chi^2(25) = 2170.49$ Prob $> \chi^2 = 0.000$	LR $\chi^2(23) = 2552.84$ Prob $> \chi^2 = 0.000$	$F(2,22\,576) = 68.98$ Prob $> F = 0.000$	—
Universal queue	LR $\chi^2(32) = 4054.03$ Prob $> \chi^2 = 0.000$	LR $\chi^2(32) = 4127.50$ Prob $> \chi^2 = 0.000$	$F(1,22\,576) = 32.15$ Prob $> F = 0.000$	—

Note: As mentioned, Mengistae's test for the Poirier model is not available because the informal wage equation cannot be estimated (see section 3.1).

DOI: 10.1002/jid

The test results show that at a significance level of 5 per cent, both the null hypothesis of the nonexistence of a job queue and the existence of a universal queue are rejected. This provides evidence in support of a bivariate as opposed to a univariate specification to describe the process of obtaining a formal or informal job. Furthermore, the existence of a job queue indicates a rationing of formal jobs. Hence, not everyone who wants to be employed in the formal sector has this possibility because of the restriction on the creation of formal jobs. In line with Mengistae (1999), from Equation (18), it can be deduced that the estimated proportion of informal workers who have been rationed out of the formal sector is 62 per cent ($\pi = \hat{\delta}_1/\hat{\sigma}_{1v_2}$). At the same time, the rejection of a universal queue indicates that not all workers seek employment in the formal sector. Thus, two types of informal workers emerge: those who benefit, or earn more, from being employed in the informal sector and those that given their low level of qualifications opt for this sector as the only way of subsisting and escaping unemployment.

Next, the length of the job queue is calculated together with the probabilities that can be deduced from the models. Table 10 shows the results for the mean values of the

Table 10. Probabilities and queue length from the Abowd–Farber (A-F) and Poirier bivariate models

	Total sample		Male		Female	
	A-F	Poirier	A-F	Poirier	A-F	Poirier
P(IQ)	0.785	0.840	0.815	0.856	0.753	0.823
P(CFQ)	0.750	0.760	0.770	0.783	0.728	0.734
P(formal)	0.610	0.647	0.644	0.678	0.572	0.614
q	1.332	1.316	1.298	1.277	1.373	1.362

	Less than primary		Primary		Secondary		Tertiary	
	A-F	Poirier	A-F	Poirier	A-F	Poirier	A-F	Poirier
P(IQ)	0.506	0.740	0.621	0.762	0.780	0.832	0.907	0.907
P(CFQ)	0.383	0.388	0.557	0.560	0.751	0.759	0.880	0.897
P(formal)	0.207	0.290	0.360	0.431	0.594	0.636	0.802	0.815
q	2.604	2.578	1.794	1.786	1.331	1.318	1.136	1.114

	Head of household		Nonhead of household		Nonmarried		Married	
	A-F	Poirier	A-F	Poirier	A-F	Poirier	A-F	Poirier
P(IQ)	0.809	0.859	0.764	0.826	0.764	0.826	0.805	0.853
P(CFQ)	0.764	0.769	0.731	0.743	0.731	0.743	0.767	0.775
P(formal)	0.637	0.669	0.581	0.623	0.581	0.623	0.636	0.669

	Unemployed before the current job		Employed before the current job	
	A-F	Poirier	A-F	Poirier
P(IQ)	0.825	0.879	0.777	0.832
P(CFQ)	0.668	0.682	0.767	0.775
P(formal)	0.570	0.608	0.618	0.655
q	1.496	1.465	1.304	1.290

IQ, 'in the queue'; CFQ, 'chosen from the queue'; q, queue length.

DOI: 10.1002/jid

probabilities in the preceding texts. It can be seen that the conditional probability of being a formal worker lies between 61 and 65 per cent and that only individuals who have tertiary education have a probability greater than 80 per cent of being employed in the formal sector. With regard to the length of the queue, this is calculated as being around 1.33; that is, for each 100 workers in the formal sector, there are a further 133 who wish to obtain a formal job. This value is high in comparison with the lengths of the formal job queues in Brazil (1.19) and Chile (1.06) (Soares, 2004; Puentes & Contreras, 2009).

An important aspect of the analysis of probabilities is to see how these vary for different characteristics. Thus, it can be seen that women are less likely to seek formal employment (75 per cent) than men are (81 per cent). In the case of education, individuals with low levels of education are more likely to work in the informal sector. By contrast, the more highly educated are more likely to seek formal employment and to receive more offers of this type of work. As for heads of households, it can be noted that they are more likely to seek formal employment than nonheads of households, yet the latter have similar chances of receiving formal job offers as those enjoyed by the former. Regarding the married/single variable, married individuals are more likely to find formal employment than single individuals, and at the same time, the former are more likely to seek formal employment and to receive a greater number of offers of this type of work.

Regarding the length of the queue, workers with no schooling form the longest job queue: for every 100 formal employees, there are 260 employees without schooling that want a formal job. Likewise, women, nonheads of households and those that are not married find themselves in a longer job queue for formal employment: for every female worker, nonhead of household or single individual in the formal sector, there are 1.37 workers that want a formal job. A period of prior unemployment also lengthens the job queue, in this case to 1.50 workers.

5.5 Determinants of the Decision of Being a Formal or Informal Worker: Structural Bivariate Probit Model

Equation (1) suggests that an individual's preference for formal employment depends on the wage premium offered by this sector ($\ln W_f - \ln W_{inf}$). Under this framework, a percentage change in $\ln W_f$ has the same effect as a percentage change of equal magnitude in$\ln W_{inf}$. The different valuation of wages between sectors may result from the fact that there are nonwage benefits that are directly associated with wages. Thus, an increase in formal sector wages implies access to a higher pension, as well as subsidies for transport, food, recreation, housing and education, among other benefits. Therefore, it seems reasonable to think that in the decision to join the queue for formal employment, the wages paid in the formal and informal sectors have asymmetric effects. This means that instead of including the wage differential between sectors, wages should be included separately in the IQ equation.[8]

A further important question to bear in mind pertains to the types of wage equation estimates to be included in the structural bivariate probit. In practice, there are two types: unconditional and conditional. The former are estimates of the wage equation that do not take into account the type of job performed by the workers and are therefore free of the effects of wage selection bias (the inverse Mills ratios). The estimates of the conditional wage equation, by contrast, are based on the relative weight of the workers' wages once

[8]This alternative was proposed by Venti (1987).

DOI: 10.1002/jid

the specific job sector has been determined. This wage estimate differs from the unconditional wage equation estimate insofar as the selection effects are due to differences in the rates of return of the unobserved characteristics of the workers and not to the levels of these characteristics (Gyourko & Tracy, 1988; Mengistae, 1999). Here, the unconditional wage equation estimates are used because we are interested in determining whether there is a wage premium at which workers choose a formal job.

The maximum likelihood estimates of the structural and sequential probit models with partial observability are shown in Table 11. Note that for the reduced estimate (Table 7), the education variables have not been considered. In fact, they have been eliminated, as they are also used as predictors of wage levels and so the possibility of the double-counting of effects is avoided. However, in Table A1 of the appendix, these estimates are conducted with the education variables and, as can be seen, no marked differences are found.

It can be noted, therefore, that formal sector wages have a marked positive effect on the probability of joining the queue for formal employment, while this same variable reduces the probability of being selected for a formal vacancy. This latter effect is attributable to the need employers have to minimize their production costs. In turn, the predicted wages

Table 11. Estimated of Abowd–Farber structural bivariate probit model (Y_j = 1 formal, 0 informal; j = IQ, CFQ)

	IQ		CFQ	
	Coeff	z	Coeff	z
Expected wages of the formal/informal sector				
$\ln\hat{W}_f$	1.191[a]	4.47	−0.546[b]	1.82
$\ln\hat{W}_{inf}$	−0.647[a]	−17.24	—	—
Personal characteristics				
Age	0.067[a]	4.66	0.096[a]	3.79
Age2	−0.0004[c]	−2.12	−0.001[a]	−3.71
Tenure	0.198[a]	4.38	—	—
Male	0.041[c]	2.13	0.041[c]	2.20
Head of household	0.220[a]	5.61	0.357	1.12
Married	−0.045	−1.31	0.003	0.09
Unemployed before the current job	−0.284[a]	−4.51	−0.217[a]	−3.77
Household characteristics				
# kids 0–2	−0.105[a]	−3.41	—	—
# kids 3–5	−0.124[a]	−4.44	—	—
# kids 6–10	−0.068[a]	−3.57	—	—
# kids 11–17	−0.140[a]	−8.58	—	—
# of relatives working as formal	0.358[a]	12.46	—	—
Education of household	0.041[a]	6.73	—	—
Constant	−23.718[a]	−16.44	−4.864[c]	−2.39
N		58 654		
Log L		−31 272.44		
AIC		62 640.87		
BIC		63 071.88		

Note:
[a]Denotes significance at 1 per cent.
[b]Denotes significance at 5 per cent.
[c]Denotes significance at 10 per cent. Robust standard errors. Less than primary education and commerce, hotel and construction sectors as references. We include city dummies. IQ, 'in the queue' equation; CFQ, 'chosen from the queue' equation.

DOI: 10.1002/jid

in the informal sector have a negative effect on the search for a formal job, and this effect differs in magnitude from that generated by formal sector wages. The foregoing discussion validates the hypothesis of asymmetric effects in the decision to join the queue due to changes in the wages of the formal and informal sectors.

Turning to the effect of the other variables, it can be seen that there are no marked differences in the estimates of the reduced model. Being male increases the likelihood of joining the queue for a formal job and of being selected from it for a formal vacancy. Being the head of the household exerts a positive effect on the probability of seeking formal employment. Finally, a prior period of unemployment has a negative effect on the whole process of finding formal employment.

In short, the estimates confirm the existence of formal job rationing. Such rationing is a reflection of the limited capacity of the formal, modern sector of the economy to generate sufficient jobs, thus resulting in a job queue and the mechanisms that employers use to fill these few vacancies. Given these two selection mechanisms that characterize the labour market, finding a better job depends to a considerable degree on the wage premium offered by the formal sector and the characteristics of the workers (especially their human capital), both of which serve to enhance the job search and to increase the number of job offers that a worker might receive.

The wage a worker can expect to receive in formal employment, along with the extent of the individual's work commitment and responsibility and the quality of his or her work history, are the main factors that determine whether a worker finds employment in the formal sector or not. Likewise, the household variables play a key role in determining whether work is sought in the formal or informal sectors. The financial pressure that having young children puts on a household has a negative effect on finding formal work. The need for resources increases the opportunity cost of finding formal employment, and so individuals prefer to generate income more rapidly and with fewer legal obstacles. By contrast, a better educational environment and a greater number of formal workers in the household have a positive effect on helping individuals find formal employment.

5.6 Distinguishing Results by Gender

In our analysis, we do not distinguish by gender, which could translate into selection bias. It is known that women and men present significant differences in the labour market. In particular, female labour participation could be low compared with male participation, which could lead to sample selection bias in our models.[9] Therefore, in order to check if the results derived so far for the whole sample of workers remain for both genders, we estimate all the equations for men and women workers separately. Due to space constraints, we only report the estimates of the structural and sequential probit model, but the other estimates are available upon request. The results are shown in Table 12.

The results show that the main findings observed for the whole sample are also detected when the model is estimated separately for female and male workers. We can note that formal sector wages have a positive effect on the probability of joining the queue for formal employment for both men and women, but this effect is more marked for men. In the case of the influence of this same variable on the probability of being selected for a

[9]We would like to thank an anonymous referee for highlighting this potential sample selection bias. We assumed that any concern about sample selection would be mitigated when the equations are distinguished by gender.

DOI: 10.1002/jid

Table 12. Estimated of Abowd–Farber structural bivariate probit model by gender ($Y_j = 1$ formal, 0 informal; j = IQ, CFQ)

	Male				Female			
	IQ		CFQ		IQ		CFQ	
	Coeff	z	Coeff	z	Coeff	z	Coeff	z
Expected wages of the formal/informal sector								
$\ln\hat{W}_f$	2.417[a]	3.20	−0.419[a]	−6.73	0.467[a]	3.02	−0.021[b]	−1.95
$\ln\hat{W}_{inf}$	−0.247[c]	−2.24	—	—	−0.407[a]	−17.94	—	—
Personal characteristics								
Age	0.003	0.09	0.024	1.51	0.071[a]	4.64	0.176[a]	6.49
Age2	−0.0003	−0.89	−0.0002	−1.30	−0.0004[c]	−2.55	−0.002[a]	−6.47
Tenure	0.611[a]	4.90	—	—	0.172[a]	13.97	—	—
Head of household	0.377[a]	6.63	0.005	0.15	0.029	0.73	0.031	0.54
Married	0.056	0.80	0.087[c]	2.43	−0.148[a]	−3.89	−0.088[b]	−1.93
Unemployed before the current job	−0.255[a]	−3.51	−0.286[a]	−7.97	−0.372[a]	−8.23	−0.284[a]	−4.51
Household characteristics								
# kids 0–2	−0.1089[b]	−1.77	—	—	−0.104[c]	−256	—	—
# kids 3–5	−0.011	−0.23	—	—	−0.207[a]	−5.88	—	—
# kids 6–10	−0.010	−0.28	—	—	−0.105[a]	−4.38	—	—
# kids 11–17	−0.093[a]	−3.50	—	—	−0.168[a]	−9.24	—	—
# of relatives working as formal	0.600[a]	9.61	—	—	0.245[a]	10.52	—	—
Education of household	0.046[a]	3.87	—	—	0.029[a]	4.46	—	—
Constant	−16.821[a]	−3.97	−10.44[a]	−7.41	−21.122[a]	−16.15	−1.819	−1.23
N	30 535				—		28 119	
Log L	−16 359.06				—		−14 566.96	
AIC	32 819.15				—		29 225.92	
BIC	33 193.15				—		29 605.15	

Note:
[a]Denotes significance at 1 per cent.
[b]Denotes significance at 5 per cent.
[c]Denotes significance at 10 per cent, respectively. Robust standard errors. Less than primary education and commerce, hotel and construction sectors as references. We include city dummies. IQ, 'in the queue' equation; CFQ, 'chosen from the queue' equation.

formal vacancy, this effect is negative and slightly higher for women. This last effect may be associated to discrimination factors where employers could prefer to hire a man instead of a woman to minimize their production costs due to, for instance, maternity factors and higher sick leaves, among others. With regard to the predicted wages in the informal sector, we can observe that when the wages increase in the informal sector, it is less likely for people to join the job queue for a formal job. Interestingly, this negative effect is higher for women than for men.

As far as the particular effect of the other variables, beyond specific differences between the two genders, results in Table 12 are similar to those achieved when the whole sample is used. The human capital variable associated to age has a positive and decreasing effect on the likelihood of joining the queue, but only for women, while tenure presents a significant

DOI: 10.1002/jid

and positive effect in both genders, being higher for men. In the case of the head of household and marriage variable, a significant and positive effect was found in the probability of joining the queue for men but not for women, which could suggest that men who are heads of households present a greater need for a good job than women in a similar position within the household. In terms of the marital status variable, the estimates show that for men, being married positively affects only the probability of being chosen from the job queue. Interestingly, for women, the condition of being married affects negatively both the decision to seek formal employment and the employers' selection process. This latter result is consistent with the fact that married women present worse indicators in the labour market than single women: less labour participation, higher unemployment rates, lower wages and higher informality rates (Arango, Castellani & Lora, 2016).

In terms of household variables, the differential effect by gender of the presence of children in the household is worth highlighting. The estimates show that for women, children at home at any age spectrum negatively affect the decision to seek a job in the formal sector, while for men, only children between the ages of 11 and 17 affect such a decision. These results could indicate the bias in Colombian households regarding the greater responsibilities in child rearing that women have in comparison with men's responsibilities, which affect the labour market outputs of women.

6 CONCLUSIONS

Adopting an alternative methodology to those more typically used in the analysis of the factors influencing the decision to enter the formal or informal job sectors in Colombia, this paper has presented evidence of formal job rationing in the Colombian labour market. The relative paucity of formal employment owing to the lack of development of the modern, formal sector has led to the generation of a dual selection mechanism for finding work in the formal sector. First, individuals must be prepared to join a queue for formal jobs. Joining this queue depends mainly on the wage premium offered by the formal sector and the level of responsibility held by the individual in the household.

In the second stage, with the individuals now in the queue for formal employment, they must await a job offer. Such job offers depend primarily on the absolute recruitment costs of the individuals, the degree of responsibility and work commitment they show and their work history. Thus, workers that manifest a high degree of responsibility and can show a good work history (or who were in employment in the last period) enjoy a greater likelihood of being selected for a formal job.

In quantitative terms, this study has been able to determine that around 62 per cent of workers employed in the informal sector owe their position to the rationing of formal jobs. This result indicates that there are two types of informal workers: those who do not have no alternative other than to be employed in the informal sector and those who are able to derive certain benefits from being employed in this sector. This latter group sees informality as a sector in which they can accumulate human capital and where they can reap returns on their education; in other words, they, unlike the former group, do not see informality as a secondary job sector. However, the characteristics of this second type of informal worker need to be examined in greater depth because the outcomes reported here may well be biased as the preferences of workers for switching from one sector to another have not been taken into account. This limitation is inherent to the survey used.[10]

DOI: 10.1002/jid

Finally, the confirmation of the existence of a dual selection mechanism operating in the labour market for finding employment in the formal or informal sectors is evidence of a process of labour segmentation that characterizes Colombia today.

ACKNOWLEDGEMENT

The authors wish to thank José Luis Roig, Josep Lluís Raymond, and an anonymous reviewer for their suggestions and comments.

REFERENCES

Abowd J, Farber H. 1982. Job queue and the union status of workers. *Industrial and Labor Relations Review* **35**(3): 354–367.

Arango LE, Castellani F, Lora E. 2016. *Desempleo Femenino en Colombia*. Banco de la República-BID: Bogotá-Colombia.

Archibald RB. 1977. Labor queues and involuntary unemployment. *Economic Inquiry* **15**(1): 33–50.

Asea P. 1996. The informal sector: baby or bath water? A comment. *Carnegie-Rochester Conference Series on Public Policy* **45**: 163–171.

Bargain O, Kwenda P. 2014. The informal sector wage gap: new evidence using quantile estimations on panel data. *Economic Development and Cultural Change* **63**(1): 117–153.

Bernal R. 2009. The informal labor market in Colombia: identification and characterization. *Desarrollo y Sociedad* **63**: 145–208.

Co C, Gang I, Yun M. 2005. Self-employment and wage earning in Hungary. *Review of Development Economics* **9**(2): 150–165.

Contreras D, De Mello L, Puentes E. 2008. "Tackling business and labour informality in Chile", OECD Economics Department Working Papers, No 607, OECD Publishing.

Contreras D, Gillmore R, Puentes E. 2017. Self-employment and queues for wage work: evidence from Chile. *Journal of International Development* **29**(4): 473–499.

Dickens W, Lang K. 1985. A test of dual labor market theory. *The American Economic Review* **75**(4): 792–805.

Farber H. 1983. The determination of the union status of workers. *Econometrica* **51**(5): 1417–1437.

Galvis L. 2012. Informalidad Laboral en las Áreas Urbanas de Colombia. *Coyuntura Económica* **42**: 15–51.

García G. 2017. Labor informality: choice or sign of segmentation? A quantile regression approach at the regional level for Colombia. *Review of Development Economics, Forthcoming* https://doi.org/10.1111/rode.12317

Gyourko J, Tracy J. 1988. An analysis of public and private sector wages allowing for endogenous choices of both government and union status. *Journal of Labor Economics* **6**: 229–253.

Harris JR, Todaro MP. 1970. Migration, unemployment and development: a two-sector analysis. *American Economic Review* **60**(1): 126–142.

Heckman J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simpler estimator for such models. *Annals of Economic and Social Measurement* **5**(4): 475–492.

[10]For the Brazilian case, Soares (2004) can distinguish between informal workers who want to change their current job for a formal job. This information makes it possible to take into account a greater diversity of informal workers.

DOI: 10.1002/jid

Heckman J. 1979. Sample selection bias as a specification error. *Econometrica* **47**(1): 153–161.
Herrera-Idárraga P, López-Bazo E, Motellón E. 2015. Double penalty in returns to education: informality and educational mismatch in the Colombian labour market. *Journal of Development Studies* **51**(12): 1683–1701.
Huguet A. 1996. Dualidad en el Mercado de Trabajo Español. *Revista de Economía Aplicada* **4**(11): 81–104.
Hussmanns R. 2004. Defining and measuring informal employment. In *International Labour Office Working Paper*. ILO: Geneva.
ILO 2012. 2012 Labour overview. Latin America and the Caribbean, Lima.
Jütting J, De Laiglesia JR. 2009. Is informal normal? Towards more and better jobs in developing countries. OECD Development Centre, 163 pages.
Khamis M. 2012. A note on informality in the labor market. *Journal of International Development* **24**(7): 894–908.
Lang K, Dickens W. 1987. "Neoclassical and sociological perspectives on segmented labor markets", NBER Working Paper Series, No 2127, NBER.
Lee LF. 1978. Unionism and wage rates: a simultaneous equation model with qualitative and limited dependent variables. *International Economic Review* **19**(2): 415–433.
Maddala GS. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.
Maddala GS, Nelson F. 1975. "Switching regression models with exogenous and endogenous switching", Proceedings of the American Statistical Association (Business and Economics Section), 423–426.
Maloney, W. 1998. "The structure of labor markets in developing countries. Time series evidence on competing views", Policy Research Working Paper, No 1940, The World Bank.
Maloney W. 1999. Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *The World Bank Economic Review* **13**(2): 275–302.
Maloney W. 2004. Informality revisited. *World Development* **32**(7): 1159–1178.
Marcouiller D, Ruiz de Castilla V, Woodruff C. 1997. Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru. *Economic Development and Cultural Change* **45**(2): 367–392.
Meng C, Schmidt P. 1985. On the cost of partial observability in the bivariate probit model. *International Economic Review* **26**(1): 71–85.
Mengistae T. 1999. "Wage rates and job queue: does the public sector overpay in Ethiopia", Policy Research Working Paper, No 2105. The World Bank.
Mincer J. 1976. Unemployment effects of minimum wages. *Journal of Political Economy* **84**(4): S87–S104.
Mohanty M. 2001. Testing for the specification of the wage equation: double selection approach or single selection approach. *Applied Economics Letter* **8**(8): 525–529.
Pisani M, Pagán J. 2003. Sectoral queuing in a transitional economy: the case of Nicaragua in the 1990s. *Labour* **17**(4): 571–597.
Poirier DJ. 1980. Partial observability in bivariate probit models. *Journal of Econometrics* **12**(2): 209–217.
Pratap S, Quintin E. 2006. Are labor market segmented in developing countries? A semiparametric approach. *European Economic Review* **50**(7): 1817–1841.
Puentes E, Contreras D. 2009. "Informal jobs and contribution to social security: evidence from a double selection model", Serie Documentos de Trabajo, No 307, Universidad de Chile, Santiago, December.
Roy AD. 1951. Some thoughts on the distribution of earnings. *Oxford Economics Papers* **3**: 135–146.
Saavedra J, Chong A. 1999. Structural reform, institutions and earnings: evidence from the formal and informal sectors in urban Peru. *Journal of Development Studies* **35**(4): 95–116.
Schneider F, Buehn A, Montenegro CE. 2010. New estimates for the shadow economies all over the world. *International Economic Journal* **24**(4): 443–461.

DOI: 10.1002/jid

Soares F. 2004. "Do informal workers queue for formal jobs in Brazil?", Working Paper, No 1021, IPEA, Brasilia.

Tunali I. 1986. A general structure for models of double-selection and an application to joint migration/earning process with remigration. In *Research in Labor Economics*, Ehrenberg R (ed), **8**, Part B. JAI Press: Greenwich. MA; 235–282.

Uribe J, Ortiz C. 2006. *Informalidad Laboral En Colombia 1988–2000: Evolución, Teorías y Modelos*. Programa Editorial Universidad del Valle: Cali-Colombia.

Uribe J, Ortiz C, García G. 2007. La Segmentación del Mercado Laboral Colombiano en la Década de los Noventa. *Revista de Economía Institucional* **9**(16): 189–221.

Venti SF. 1987. Wages in the federal and private sectors. In *Public Sector Payrolls*, Wise DA (ed). University of Chicago Press. Chicago; 147–182.

A : APPENDIX

Table A1. Estimated of Abowd–Farber structural bivariate probit model (Y_j = 1 formal, 0 informal; j = IQ, CFQ)

	IQ		CFQ	
	Coeff	z	Coeff	z
Expected wages of the formal/informal sector				
$\ln\hat{W}_f$	2.349[a]	3.63	−0.371[a]	−7.18
$\ln\hat{W}_{inf}$	−0.765[a]	−13.29	—	—
Personal characteristics				
Primary	0.263[a]	2.45	0.318[a]	3.01
Secondary	0.505[b]	2.45	0.523[a]	4.33
Tertiary	0.934[a]	3.70	0.148	0.81
Age	−0.027	−0.87	0.0006	0.04
Age2	−0.0001	−0.49	−0.0001	−0.05
Tenure	0.561[a]	−5.58	—	—
Male	0.120[a]	6.27	0.019	0.70
Head of household	0.253[a]	5.39	−0.021	−0.85
Married	0.154[a]	3.51	0.048[b]	2.05
Unemployed before the current job	−0.121[c]	−1.89	−0.329[a]	−12.25
Household characteristics				
# kids 0–2	−0.129[a]	−3.39	—	—
# kids 3–5	−0.136[a]	−4.03	—	—
# kids 6–10	−0.063[b]	−2.45	—	—
# kids 11–17	−0.170[a]	−9.20	—	—
# of relatives working as formal	0.388[a]	13.52	—	—
Education of household	0.046[a]	6.04	—	—
Constant	−30.953[a]	−6.54	−16.677[a]	−7.96
N	58 654			
Log L	−31 066.53			
AIC	62 241.07			
BIC	62 725.96			

Note
[a]Denotes significance at 1 per cent.
[b]Denotes significance at 5 per cent.
[c],Denotes significance at 10 per cent, respectively. Robust standard errors. Less than primary education and commerce, hotel and construction sectors as references. We include city dummies. IQ, 'in the queue' equation; CFQ, 'chosen from the queue' equation.

DOI: 10.1002/jid