Journal of Development Economics 97 (2012) 88–98

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Informal employment in developing countries Opportunity or last resort?☆

Isabel Günther [a,*], Andrey Launov [b,c]

[a] ETH Zürich, Switzerland
[b] University of Mainz, Germany
[c] Université catholique de Louvain (Louvain-la-Neuve), Belgium

ARTICLE INFO

Article history:
Received 5 October 2009
Received in revised form 20 December 2010
Accepted 5 January 2011

JEL classification:
O17
J42

Keywords:
Informal labor market
Segmentation
Comparative advantage
Selection bias
Finite mixture model

ABSTRACT

There is an ongoing debate among researchers and policy makers, whether informal sector employment is a result of competitive market forces or labor market segmentation. More recently it has been argued that none of the two theories sufficiently explains informal employment, but that the informal sector shows a heterogenous structure. For some workers the informal sector is an attractive employment opportunity, whereas for others – rationed out of the formal sector – the informal sector is a strategy of last resort. To test the empirical relevance of this hypothesis we formulate an econometric model which allows for several unobserved segments within the informal sector and apply it to the urban labor market in Côte d'Ivoire.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In the 1970s, when it became a stylized fact that the informal sector is often the most important employment opportunity in developing countries, the evolution of the informal sector in the course of economic development became an extensively researched topic (Livingstone, 1971; Hart, 1973; Fields, 1975). The topic was somewhat neglected in the 1980s. But at the end of the 1990s, with international development policy focusing on poverty reduction, the informal sector – often considered as the economy of the poor – has reemerged and become omnipresent on the policy and research agenda. [1]

One often observed feature of urban labor markets in developing countries is the coexistence of a small, well-organized formal sector characterized by relatively high earnings and attractive employment conditions with a large informal sector characterized by low and volatile earnings. An important question for both the understanding of the labor market as well as for policy recommendations, is whether the observed differences in earnings in the formal and informal sectors are the result of market segmentation or whether competitive labor market theories hold despite the observed wage differences. In other words, the question is whether the informal sector is voluntarily chosen by individuals as an employment opportunity or if individuals are pushed into informal employment because of entry barriers into the formal sector. Another related question is whether individuals are poor because they are employed in the informal sector (an implication of a segmented labor market); or alternatively, whether they are employed in the informal sector because they are poor(ly endowed) with characteristics that generate high returns in the formal sector (an implication of a competitive labor market).

Traditional dual labor market theories, starting with Lewis (1954), assert that the informal sector is the disadvantaged sector into which workers enter to escape unemployment once they are rationed out of the formal sector where wages are set above market-clearing prices (Harris and Todaro, 1970; Fields, 1990; Stiglitz, 1976). It is argued that workers in the informal sector earn less than identical workers in the

☆ We would like to thank Gary Fields, Michael Grimm, Stephan Klasen, William Maloney, Walter Zucchini, two anonymous referees and the Editor as well as seminar and conference participants in Barcelona, Berlin, Bertinoro, Göttingen, Verona and Warwick for a number of very useful comments and discussions. Any errors remain our responsibility.

* Corresponding author.
E-mail address: isabel.guenther@nadel.ethz.ch (I. Günther).

[1] The most recent contributions to theoretical research on the causes and consequences of informality are Maloney (2004), Fields (2005), Loayza and Rigolini (2006), Amaral and Quintin (2006), Bennett and Estrin (2007) and Bennett (2008). See also World Bank (2007) for a wider review and call for future research.

0304-3878/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jdeveco.2011.01.001

formal sector. If no entry barriers existed, workers from the informal sector would enter the formal sector.

While the sizable differences in earnings between the formal and informal sectors are not a matter of debate, the mere existence of lower wages in combination with lower returns to education and experience does not necessarily imply market segmentation (see e.g. Dickens and Lang, 1985; Heckman and Hotz, 1986; Rosenzweig, 1988; Magnac, 1991; Pratap and Quintin, 2006). A labor market with two distinct sectors and wage equations does not constitute a segmented labor market as long as individuals are free to move between these two sectors (see e.g. Dickens and Lang, 1985; Basu, 1997). This suggests a voluntary employment of individuals in the informal sector (Rosenzweig, 1988; Maloney, 2004). Voluntary employment in the informal sector can, for example, be due to desirable non-wage features of the informal sector where individuals maximize their utility rather than their earnings (Maloney, 2004). Moreover, workers may have an individual comparative advantage in the informal sector and would not do any better in the formal sector (Gindling, 1991; Maloney, 1999, 2004).

Hence, two opposing theories exist: The segmentation hypothesis sees informal employment as a strategy of last resort to escape involuntary unemployment, whereas the comparative advantage hypothesis sees informal employment as a voluntary choice of workers based on income or utility maximization. Several empirical studies have addressed the question of competitive versus segmented labor markets in developed and developing economies (see e.g. Dickens and Lang, 1985; Heckman and Hotz, 1986; Rosenzweig, 1988; Magnac, 1991; Gindling, 1991; Pratap and Quintin, 2006). For the case of developing countries, Magnac (1991) tests for competitiveness in the framework of an extended Roy model, Gindling (1991) considers the same question in a generalized regression with sample selection, and Pratap and Quintin (2006) address labor market segmentation taking a semi-parametric approach. All three papers find evidence of a competitive labor market, with informal sector employment being a choice of comparative advantage considerations, rather than a strategy of last resort.

This empirical literature is, however, dominated by the *a priori* assumption that the informal sector is homogenous. This assumption has recently been challenged by several authors (for example Fields, 2005; Paulson and Townsend, 2005; Guha-Khasnobis et al., 2006). The alternative hypothesis on urban labor markets in developing countries is a combination of the polar views of segmented and competitive labor markets and emphasizes a more complex structure of the informal sector. Fields (2005) argues that the informal sector consists of two distinct parts: the 'upper' tier and the 'lower' tier. The upper tier represents a competitive part into which individuals enter voluntarily because, given their specific characteristics, they expect to earn more in the informal than in the formal sector. The lower tier consists of individuals rationed out of the formal labor market. Similarly, but from a dynamic perspective, Paulson and Townsend (2005) show that individuals who start informal entrepreneurial activities during economic crises (to escape unemployment) are not only very distinct in characteristics, but also earn a much lower profit than informal entrepreneurs who started their business during normal times (often with much higher start-up investments).

However, few studies exist that have analyzed this theory of a heterogenous informal labor market from an empirical perspective. The main problem is that it is not easy to detect a dual structure within the informal sector as we usually only have data on informal affiliation. Information on the motivation of individuals to enter the informal sector is not available. As a consequence, we cannot directly observe, whether all or only a part of informal employees see the informal sector as an opportunity or as a strategy of last resort. To the best of our knowledge, Cunningham and Maloney (2001), who represent informal enterprises as a mixture of upper-tier and lower-tier enterprises, have written the only paper that allows for unobserved heterogeneity in informal activity. The limitation of this study is that it only looks at informal entrepreneurs, which is a quite narrow selection of employment opportunities. The options of formal as well as informal, salaried employment or staying out of the labor market are not considered in their analysis.

The objective of our paper is to help close this gap in the empirical literature. We develop an econometric model which is both able to detect unobserved heterogeneity in the composition of the informal sector (as Cunningham and Maloney, 2001), and to take into account all employment options in the labor market (as Magnac, 1991). The proposed framework can be classified as a finite mixture regression with sample selection. Several aspects of our framework are new to the empirical literature on informality: first, it provides us with consistent estimates of returns to individual characteristics within any of the segments of the informal sector, accounting for selection bias into the labor market (Heckman, 1979; Dickens and Lang, 1985; Magnac, 1991). Second, it provides us with an intuitive approach to identify the size of voluntary and/or involuntary employment in the informal sector (Maloney, 2004; Fields, 2005).

We apply our model to household survey data from Côte d'Ivoire at the end of the 1990s. We deliberately chose a general household (and not a labor market) survey from a Sub-Saharan African country for the following reasons: first, our proposed method is applicable to any cross-sectional household survey that collects data on general individual characteristics (including employment and wages), so it can be used for a wide range of countries and time periods. Second, focusing on employment data availability and quality has, at least in the past, largely led to a neglect of Sub-Saharan African labor markets. Most studies on the informal sector, as well as most work cited in this paper, have analyzed Latin American countries, where labor market dynamics might be quite different from Sub-Saharan African countries. Third, even panel data on labor markets does usually not provide enough information to test the hypothesis of involuntary and/or voluntary informal employment.[2]

We find that the hypothesis of a dual structure of the informal sector, with both voluntary and involuntary employment, best describes the empirical data on the Ivorian urban labor market. Policies therefore have to take into account the fact that the informal sector consists of both individuals who would like to switch to a formal job and individuals who currently have no incentive of doing so. From a poverty perspective, this means that informality is both a cause and a result of poverty.

The paper is structured as follows: In Section 2 we develop the econometric model. Section 3 presents the data and discusses the estimation results. Section 4 concludes.

2. Econometric model

Individuals can be employed in the formal or informal sector of the labor market or they can stay out of the labor market. Given the objective of the paper, we assume that the formal sector is homogeneous, whereas the informal sector can be heterogenous.[3] This means that the informal sector can – if supported by the data – consist of a number of (latent) segments, with no ad-hoc assumption whether these segments comprise a voluntary or involuntary employees. Each segment of the labor market is distinguished by its own unique wage function.

Household surveys usually only provide information on whether an individual belongs to the formal or informal sector, but it is impossible to observe affiliation to a specific segment of the informal sector. Once we assume that the informal sector has a heterogenous structure we could –

[2] For instance, panel data on work flows across sectors, available for some Latin American countries, would not help to identify involuntary and voluntary informal employment: If no work flow existed between the informal and formal sector this could be the result of entry barriers into the formal sector or because employees would not be better off in the formal sector and hence choose to stay within the informal sector.

[3] The model can easily be extended to a model where also the formal sector consists of several segments with unobservable segment affiliation. For example one could distinguish between private and public formal employment and earnings potentials.

at least hypothetically – also try to identify segments within the informal sector using observable variables (such as industries, wages, etc.). The problem is that we would have to make various ad-hoc assumptions without any evidence about whether these assumptions are close to the true structure of the labor market. We would first have to decide which variable(s) (e.g. industries or wages) actually identify whether a worker belongs to a specific informal segment. Second, we would have to know the exact number of segments within the informal sector. Last, we would have to have information on the fraction of informal workers that are paid according each wage equation. This information is unavailable, and trying to make a sophisticated guess can lead to substantial biases (see Dickens and Lang, 1985).

In the next section, we therefore develop an econometric model that allows us to analyze a heterogenous informal labor market with *unobservable* sector affiliation of individuals, using minimal *a-priori* assumptions: we let the data determine the number and size of informal segments that best describe the data.

2.1. Latent segments and earnings distribution

The entire labor market Y consists of J segments Y_j, such that $Y=\cup_{j=1}^{J} Y_j$. We assume that within any given segment Y_j log-earnings are described by a wage equation

$$lny_{ij} = \mathbf{x}_i'\beta_j + u_{ij}, \qquad i \in \mathcal{Y}_j, \tag{1}$$

where y_{ij} are the earnings of an individual i in segment j. The error term follows a normal distribution with zero mean and variance σ_j^2, $u_{ij} \sim N(0,\sigma_j)$, and errors are uncorrelated across segments. Thus, our assumption is that earning distributions within all segments are distinct and independent of each other. Furthermore, the wage function, i.e. the returns (β_j) to individual characteristics ($\mathbf{x}_i$) varies from segment to segment.

It is well-known (Heckman, 1979) that the distribution of observed earnings is influenced by the decision of individuals to enter (or not enter) the labor market. In other words, the observed sample of workers is a non-random sample of all individuals because of self-selection into the labor market. We assume that individuals' employment decision is a function of a set of personal characteristics $\mathbf{z}_i$:

$$y_{is} = \mathbf{z}_i'\gamma + u_{is}, \qquad u_{is} \sim N(0,1), \tag{2}$$

such that the earning y_{ij} is observed only if the outcome of the selection Eq. (2) is positive. Assume that the errors of the segment-specific wage Eq. (1) and the selection Eq. (2) follow a bivariate normal distribution with correlation coefficient ρ_j. It is easy to show that the distribution of observed wages in the j-th segment of the labor market is given by

$$f\left(y_{ij}|y_{is}>0\right) = \frac{\varphi\left(\left(\ln y_{ij} - \mathbf{x}_i'\beta_j\right)/\sigma_j\right)}{\sigma_j \Phi\left(\mathbf{z}_i'\gamma\right)} \Phi\left(\frac{\mathbf{z}_i'\gamma + \left(\rho_j/\sigma_j\right)\left[\ln y_{ij} - \mathbf{x}_i'\beta_j\right]}{\sqrt{1-\rho_j^2}}\right), \tag{3}$$

where φ and Φ denote the density and the cumulative density functions of the standard normal distribution, respectively (see Appendix, Lemma 1).[4]

Last, whereas a typical household survey frequently provides information on whether an individual works in the formal or the informal sector, few provide information on either the existence or number of segments within the informal sector let alone the assignment of individuals within the informal sector to each segment.

Treating segment affiliation as unobserved, we estimate the probability $P(i \in Y_j) = \pi_j$ of any individual i belonging to any segment Y_j, and the distribution of observed wages in the entire labor market is:

$$f(y_i) = \sum_{j=1}^{J} \pi_j f\left(y_i|y_{is}>0,\theta_j\right). \tag{4}$$

$f(y|y_s>0,\theta_j)$ is given in Eq. (3), and $\theta_j \equiv \{\beta_j,\sigma_j,\rho_j\}$.

The model developed above can be described as a finite mixture with sample selection. It is easy to see that this model has two commonly known special cases. Assuming that $J=1$ (i.e. there is no unobserved sector heterogeneity) it reduces to the original Heckman (1979) selection model. Assuming that $\rho_j=0$ (i.e. there is no self-selection into the labor market) it reduces to a finite mixture. The model's advantage over the original Heckman model is that (log-) normality is assumed only at the segment-specific level, whereas on the aggregate level the distribution $f(y_i)$ has a flexible form.[5] Its advantage over a simple finite mixture model is that sample selection is taken explicitly into account, which allows a consistent estimation of segment-specific returns to individual characteristics.

Any data set will only support the mixture model described in Eq. (4) if there is sufficient heterogeneity in the dependent variable y_i. Otherwise the mixture cannot be estimated successfully and would collapse into a single component density. Thus, a successful estimation of a mixture indicates the presence of latent subgroups that come from different earning distributions.[6]

Moreover, even if the observed earning outcomes are indeed generated by numerous underlying distributions, we still need to make sure that the mixture model developed in Eq. (4) is identifiable. The identification approach taken in this paper is fully parametric. To understand parametric identification, consider the definition of a non-identifiable mixture: for a given family of distributions, the mixture model is not identified if another mixture model exists that implies the same probability of observing an outcome y_i for any i. Hence, we have to find a parametric form for the component density to ensure that two distinct mixture models can never provide the same probability of observing y_i. All possible densities that allow for identification of a finite mixture are distinguished by some particular property of their shape, one of which is described by Teicher (1963).[7] The density of our model is given in Eq. (3). We therefore only need to analyze whether this density belongs to the admissible set of distributions that lead to an identifiable mixture. The following proposition establishes the necessary result.

Proposition 1. For any given selection rule $\{\mathbf{z},\gamma\}$, the model (Eq. (4)) is identifiable if $\rho_j=\rho$, $\forall\, j=1,...,J$.

Proof. The proof verifies the sufficient condition for identifiability of finite mixtures given in Teicher (1963). See Appendix. □

Proposition 1 means that, as a consequence of self-selection into the labor market we have to focus on a sub-class of mixtures, where the correlation between the errors of the wage and the selection equations is set to be the same for every segment. We argue that, from

[4] Apart from being demonstrated in the Appendix, Eq. (3)is known as a backbone for the maximum likelihood estimation of the original Heckman (1979) regression with sample selection.

[5] The advantage of choosing log-normal segment-specific earning distributions is that we obtain an analytical identification result for the fully-parametric estimation. While the assumption of log-normality is criticized at the aggregate labor market, there is no argument about the distribution of segment-specific earnings, leaving some freedom to choose a convenient parametric form. The results that stretch beyond log-normality at the segment level are desirable for future research.

[6] For example, bimodality, too heavy tail(s), unusual skewness with multiple points of inflection that could not be smoothed out by kernel density estimates, too steep increase of the probability mass close to the mean relative to too flat tops etc. in the aggregate distribution of y_i speak for outcomes that come from two or more underlying latent distributions.

[7] Teicher (1963) formulates the sufficient condition for identification in terms of Laplace transforms of the mixture distribution. By uniqueness of a Laplace transform for any given probability distribution, identifiability of the transformed model immediately implies identifiability of the original model.

an economic point of view, this necessary condition is not too restrictive. The economic meaning of a positive (negative) correlation of error terms in a regression with sample selection is that a worker with a higher (lower) valuation of employment has a higher probability to be found at the upper (lower) end of the earnings distribution, if the selection mechanism reflects such a valuation of employment. The identification restriction $\rho_j=\rho$ implied by Proposition 1 therefore means that a person with a higher valuation of employment is more likely to be found at the higher end of the earning distribution, irrespective of the segment. These probabilities need, however, not be equal across segments even if $\rho_j=\rho$, because the probability of occupying a higher position within the segment-specific earning distributions also depends on the variance parameters of log-earnings σ_j, which are segment-specific.

2.2. Employment opportunity or last resort?

The finite mixture setting also allows us to analyze whether the distribution of individuals across the sectors of the labor market is a result of market segmentation or a result of comparative advantage considerations. We assume that workers are earning maximizers who know the wage function, and hence also their expected earnings given their own characteristics, for each segment of the labor market. In this case competitive theory would imply that any individual will be found in the sector where his expected earnings, given his personal characteristics, are the highest. The *hypothetical* distribution of individuals across sectors in a competitive market will therefore be given by

$$P\left(i \in Y_j\right) = P\left(E\left[\ln y_{ij} \,|\, y_{is} > 0;\, \mathbf{x}_i\right] = \max_{l, l \in [1,J]} \{E[\ln y_{il} \,|\, y_{is} > 0;\, \mathbf{x}_i]\}\right), \quad (5)$$

$j=1,\ldots,J$.[8] This distribution is conditional on both individuals' characteristics and the returns to these characteristics in each segment of the labor market, and assumes that there are no entry barriers to any sector. Individuals are always found in the segment that pays the highest expected wage for them.

On the other hand, the *actual* distribution of individuals across sectors is given by the distribution $P(i \in Y_j)=\pi_j, j=1,\ldots,J$, in Eq. (4). If we cannot reject that the actual distribution (Eq. (4)) and the hypothetical distribution of workers across sectors (Eq. (5)) are equal, there is an indication of perfect sectoral mobility. In other words, individuals would be found in the sector where, given their specific characteristics, they have the highest earning opportunity. Such a market can be considered competitive. Rejection of equality of these two distributions of workers across sectors would indicate the existence of some sort of entry barrier that prevents certain individuals from being in the sector that pays the highest expected wage to them.[9] To avoid unemployment, individuals choose a strategy of last resort, entering a sector with lower earning opportunities, i.e. the market is segmented.[10]

2.3. Implementation

For the estimation of our model we suggest the following two-step procedure:

Step 1 Estimate γ from the selection Eq. (2) by running a Probit on the data for employed and non-employed individuals.

Step 2 Use $\mathbf{z}_i'\hat{\gamma}$ as consistent estimates of $\mathbf{z}_i'\gamma$ to estimate the model in Eq. (4) on the data for all employed individuals.

As mentioned before, we know whether an individual belongs to the formal sector or the (heterogenous) informal sector. We do not know, however, whether an individual works in one or the other segment within the heterogenous informal sector. We denote the set of earning outcomes in the formal sector by Y_F, and the number of observations in the formal sector by N_F. The total log-likelihood can then be written as

$$\begin{aligned}\ln\mathcal{L} = &\sum_{i \in \mathcal{Y}_F} \ln f\left(\theta_F, \rho \,|\, y_{iF}, y_{is} > 0, \mathbf{x}_i, \mathbf{z}_i'\hat{\gamma}\right) - N_F \ln \pi_F \\ &+ \sum_{i \notin \mathcal{Y}_F}\left[\ln\left(\sum_{j=1}^{J-1} f\left(\theta_{I_j}, \rho \,|\, y_{il}, y_{is} > 0, \mathbf{x}_i, \mathbf{z}_i'\hat{\gamma}\right)\pi_{I_j}\right)\right],\end{aligned} \quad (6)$$

where π_F is the probability of belonging to the formal sector, π_{I_j} is the probability of belonging to the j-th segment of the informal sector and $f(\cdot)$ is the component density function given in Eq. (3) with the relevant j-specific parameter vector θ_{I_j}. The asymptotic covariance matrix of the parameter estimates on the second step is given by

$$V(\xi) = D^{-1}(\xi) + D^{-1}(\xi)M(\xi,\gamma)D^{-1}(\xi), \quad (7)$$

where $\xi=\{\{\theta_j\}_{j=1}^{J}, \rho, \{\pi_{I_j}\}_{j=1}^{J-1}\}$ is the parameter vector, $D(\xi)$ is the expected negative Hessian from the second step and $M(\xi,\gamma)$ is the matrix constructed using scores from the first and second steps (for the exact form of $M(\xi,\gamma)$, see Murphy and Topel, 1985). Finally, it is straightforward to show that the maximum likelihood estimate of π_F is equal to the observed share of formal workers in the sample.[11]

3. Empirical application

3.1. Data description

The model developed in the previous sections is applied to the case of Côte d'Ivoire at the end of the 1990s. After independence, Côte d'Ivoire retained – like most other West-African countries – the stringent labor market policies and institutions from the French colonial administration. Formal labor market restrictions were mainly characterized by a reliance on minimum wages, labor unions linked to the government, centralized hiring mechanisms and considerable constraints on firing employees (Rama, 1998).[12] With the large devaluation of the CFA Franc in 1994, labor market policies in Côte d'Ivoire have started to experience significant changes, especially with regard to the government monopoly over hiring decisions.

According to a World Bank survey in the 1990s, governmental labor market regulations were not seen as a major problem for growth and/or employment by most private firms. Unfortunately, no such survey on labor market regulations and their impact on employment is available for employees, in particular with regard to entering the public sector, which still constituted about 40% of the formal sector in the 1990s. From an institutional analysis, the magnitude of the impact of labor market restrictions, i.e. entry barriers into the formal sector, on informal sector employment is hence unclear.

The data we use is the 1998 Ivorian household survey, the *Enquê te de Niveau de Vie*, undertaken by the *Institut National de la Statistique de la Côte d'Ivoire* and the World Bank.[13] We focus our analysis on the urban population (to exclude agricultural employment) and limit our sample to individuals between the age of 15 and 65 years. This leaves

[8] Details on the computation of this distribution are presented in the Appendix.

[9] Note again that with a mixture model we cannot directly observe which individuals are found in which sector. We can only estimate the share of individuals that are found in each sector and compare this estimation with the shares we would obtain if every individual were found in the sector where he would earn – given his characteristics – the highest expected wage.

[10] It is important to note that the assumption that sector choice is only based on expected earnings might be too simple to reflect the actual complexity of individuals' entry decisions into various segments of the labor market. Modeling more complex sector choice mechanisms is an avenue for future research.

[11] The outlined two step procedure is a limited information maximum likelihood approach. Full information maximum likelihood estimation is equally possible, although computationally more demanding.

[12] Payroll taxes and social benefits were up to 2000 and in comparison to other developing countries rather low in Côte d'Ivoire.

[13] We used a rather dated survey of Côte d'Ivoire to exclude the specific effects of the Ivorian crisis since 2001.

us with a sample of 5592 observations. Among these, we consider individuals who voluntarily stay out of the labor market, as well as those who are involuntarily unemployed, as inactive. Only 11.3% of the inactive population are actively looking for employment and we therefore decided that we can include them into the group of the voluntary inactive without loss of information.

The active population is classified into the informal and formal sector according to reported primary employment. Information on secondary employment is also available in the data and dual employment in the informal and formal sector is indeed a phenomenon often observed in developing countries. One could integrate this additional data in our model without a large change in the structure of the model and analyzing dual job holdings would be most interesting for future research.

All individuals who work in the public sector or who have a written contract with a company that pursues formal bookkeeping are considered as formal employees. We are aware of the fact that the public formal sector and the private formal sector might also be quite distinct with regard to earning opportunities and/or earning equations. The focus of this paper is, however, informal sector heterogeneity and for simplicity we combined the two sectors. We also decided to not exclude the public sector from our analysis. We think that the public sector is important when analyzing whether the informal sector is an opportunity (with regard to any formal employment) or a strategy of last resort; especially in Sub-Saharan African countries, where a large part of the formal sector is public sector employment. Moreover, in many Sub-Saharan African countries, including Côte d'Ivoire (Rama, 1998), governments still have a large influence on formal private employment and, hence, the distinction between formal private and formal public is not always clear-cut.

We define the active population which is either self-employed or which is employed without a written contract or with a company without formal bookkeeping as the informal sector. Hence, we follow a legal definition of informal employment as proposed by the International Labor Organization (ILO) in 1993 (see e.g. Hussmanns, 2004). We are aware that various other definitions of informal employment have emerged within the last years (Guha-Khasnobis et al., 2006), but for our analysis a legal definition seems to be most appropriate.

Fig. 1 illustrates kernel density estimates of monthly formal and informal earnings. Table 1 reports the corresponding sample means. We use monthly wages instead of hourly wages. Given the often constrained working hours in the informal sector we think that monthly wages reflect earning opportunities in the informal sector better than hourly wages.[14] Fig. 1 also demonstrates that despite the considerable difference in mean earnings, the densities of informal and formal log-earnings overlap to a large extent. This already shows that not all informal employment is inferior to formal employment with regard to wages.

Table 1 displays summary statistics of further variables used in the earning and selection equations. The information is provided for the population as a whole as well as separately for inactive workers and workers in the informal and formal sectors. Educational levels are the highest in the formal sector (8.1 years), with somewhat lower and much lower levels among inactive (5.8 years) and informal (2.9 years) workers, respectively.[15] With regard to age, we find that the inactive population is the youngest (mean age of 25.2 years) followed by informal (34.7 years) and formal (36.6 years) employees. In addition, membership in the formal sector is dominated by males, who constitute 80.6% of all formal employees. In contrast, only 49.0% of informal workers and 40.6% of inactive individuals are males.

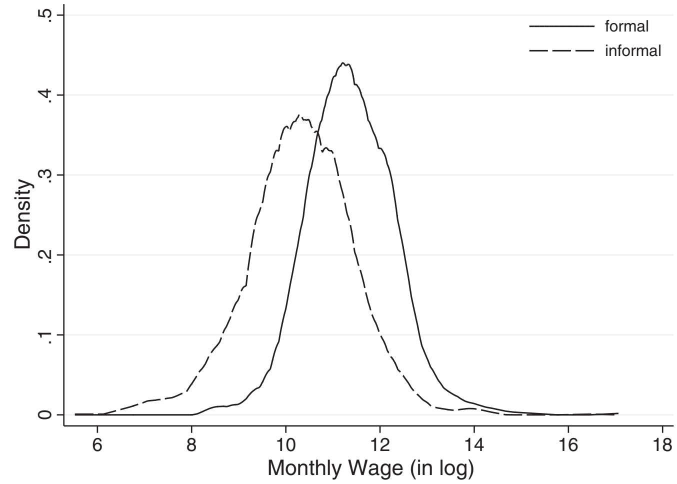

Fig. 1. Densities of monthly log-earnings. *Source: Enquête de Niveau de Vie*, 1998. Computations by the authors.

Finally, an interesting observation can be made about the distribution of religious groups. The formal sector is dominated by Christians whereas the informal sector is dominated by Muslims. This may be the result of the specific composition of the government (the public sector) which is dominated by Christians and which constitutes a large part of the formal sector. An alternative 'geographic' explanation is that formal employment is predominantly concentrated in the southern cities (especially in Abidjan) where most Christians live, whereas Muslims are predominantly living in the north.

To specify the selection equation of entering the labor market, we use additional variables: The number of infants in the household, the number of children under the age of 14 in the household, the number of old household members (members in their sixties), household size and the number of active members in the household. The reason for choosing this set of variables is twofold. First, we argue that selection variables should collect household-specific reasons which influence the decision to participate in the labor market, by determining the opportunity cost of staying out of the labor market, but at the same time have no direct impact on the earning potentials of individuals. Second, the chosen variables for a selection equation should provide sufficient exclusion restrictions (Olsen, 1980; Little, 1985).

From the descriptive statistics in Table 1 it is obvious that considerable differences in characteristics between both the inactive and the active population as well as between workers employed in the informal and formal sectors exist. Systematic differences between active and inactive individuals highlight the possibility of self-selection into the labor market. Systematic differences between individuals employed in the formal and informal sector might indicate comparative advantages in either sector.[16]

3.2. Segmented informal labor market

We start with an analysis of the number of segments within the informal labor market. The econometric model described in Section 2 allows for any number of unobserved labor market segments. A test for existence of an additional segment is an extremely complex statistical problem because the null hypothesis $H_0: \pi_{j+1} = 0$ is on the boundary of the parameter space for π_{j+1}, and because the parameter vector θ_{j+1} is

[14] If individuals in the informal sector only worked few hours (or days) per months we calculated monthly income based on the actual hours worked and did not scale up income by "theoretical" working hours per months. Many activities in the informal sector are part-time jobs by nature and employees in the informal sector could not easily increase their income by simply providing more labor hours per month.

[15] To reflect potential sector-specific non-linearities in returns to education, we specify education by education levels in our estimations. We define 1 to 6 years of education as low, 7–10 years of education as medium, and more than 10 years of education as high (see Table 1). The residual group none has zero years of education (and is almost identical to one minus the literacy rate in the population).

[16] An alternative explanation for systematic differences in characteristics would be employers' discrimination based on workers' characteristics.

Table 1
Summary statistics of the urban labor market.

	Total[a]	Inactive	Active	
			Informal	Formal
Sample (%)	100	52.6	31.3	16.1
Monthly earnings	98,815	–	64,837	164,995
Variables in wage equation				
Sex (male = 1, %)	49.7	40.6	49.0	80.6
Age (years)	30.0	25.2	34.7	36.6
Education (years)	5.3	5.8	2.9	8.1
Education (%)				
None	38.2	31.5	59.7	18.2
Low	21.3	21.6	22.5	17.8
Medium	23.8	29.6	11.6	28.4
High	16.7	17.3	6.2	35.6
Literacy (yes = 1, %)	64.1	69.8	44.4	84.0
Training after school (yes = 1, %)	17.6	11.1	14.7	44.3
Religion (%)				
Muslim	43.4	38.3	56.8	33.8
Christian	42.2	46.2	30.6	52.2
Indigenous	14.4	15.5	12.6	14.0
Living in Abidjan (yes = 1, %)	49.6	50.4	42.2	61.7
Exclusion variables in selection equation				
Infants in HH	0.96	0.97	0.98	0.86
Children in HH	2.09	2.31	1.85	1.85
Old members of HH	0.20	0.24	0.18	0.10
HH size	8.24	9.23	7.19	7.01
Active HH members	2.26	1.98	2.77	2.17

Source: Enquête de Niveau de Vie, 1998. Computations by the authors.
Notes: Monthly earnings in CFA Francs. [a] "Total" refers to individuals between 15 and 65

Table 2
Model selection.

	Information criteria		
Informal sector	SBC	cAIC	H–Q
Homogeneous	13861.5	13949.7	13792.7
2-segment	13840.0	13941.2	13722.4
3-segment	13916.8	14031.0	13750.4

not defined under H_0. This issue was first pointed out by Heckman and Hotz (1986).[17] For practical considerations we decide on the number of segments on the basis of information criteria. Leroux (1992) shows that information criteria consistently estimate the number of segments within a finite mixture.[18] We estimate the model with a homogeneous, two-segment and three-segment informal sector. Results for model selection are reported in Table 2.

Table 2 shows that the specification with a two-segment informal sector describes the data best, and that the three-segment specification already overparameterizes the model. Moreover, the three-segment specification shows a very low probability of belonging to the third segment of the informal sector: The third segment of informal employment would only make up 3.1% of the entire labor market. We therefore conclude that the specification with two wage equations, and hence with two distinct earnings segments within the informal sector, provides us with the best description of the urban labor market. The estimated parameters of the wage and selection equations for a labor market with two unobservable segments within the informal sector are shown in Table 3.

Our first important finding is the statistical significance of the correlation coefficient ρ. This underlines the necessity of accounting for sample selection into the labor market when estimating coefficients of segment-specific wage equations. Furthermore, since the estimation of the distribution of individuals across sectors based on individuals' optimal sector choice (Eq. (5)) is also dependent on slope coefficients, accounting for sample selection is crucial for a correct estimation of Eq. (5) and conclusions based on its results (see Section 3.3).

Our second important finding is that the expected wages in both informal segments are significantly below the expected wage in the formal sector. Moreover, there is also a significant difference between expected earnings in the higher-paid (Informal-1) and lower-paid (Informal-2) informal segments. Informal-1 pays, on average, 50% higher wages than Informal-2. Finally, both informal segments are considerable in size, making up 36.6% (Informal-1) and 29.5% (Informal-2) of the whole labor market, or 55.4% and 44.6% of the informal labor market, respectively.

Last, we find different wage equations in each of the three segments. Returns to education and experience (measured in years of age) are high in the formal sector. In the higher-paid segment of the informal sector (Informal-1) education and experience also have a significant impact on individuals' earnings. But returns to education are only half of the returns in the formal sector and low education levels show even no returns. In the lower-paid informal sector (Informal-2) returns to experience are only two thirds of the returns to experience in the other two segments of the labor market, and education appears to have no influence at all. Hence, workers in the lower-paid informal sector are left with very low wages almost independent of their skills.

Furthermore, whereas gender has a significant impact on earnings in all segments of the labor market, the male–female wage gap within each informal segment is wider than within the formal sector. One explanation is that legal enforcement in the formal sector prevents high sex-specific wage discrimination. Alternatively, it is possible that only the most skilled females enter the formal labor market, which contributes to a lower difference in earnings. Location and religion do not have an impact on earnings in the formal sector but have a significant impact on earnings in both informal segments of the labor market.

Thus, not only do we find that the urban labor market in Côte d'Ivoire consists of one formal and two latent informal earnings segments; we also find that each of these segments shows a distinct pattern of returns to individual characteristics.

Significant variations in returns to individual characteristics do, however, not necessarily imply that a labor market is segmented (in the sense of informal employees being rationed out of the formal labor market), as long as no entry barriers can be found between any segments of the labor market. To understand the reasons behind the observed wage differentials across segments, we have to analyze whether the observed distribution of individuals across the formal and informal segments is the result of free sectoral mobility (competitive market) or the result of entry barriers into one or several of the sectors (segmented market), which we do in the next section.

3.3. Employment opportunity and last resort

If no entry barriers between sectors existed an individual that maximized his expected earnings would enter the sector where his expected earnings (given personal characteristics) are the highest, leading to the distribution of individuals across sectors as formulated in Eq. (5). Without entry barriers this hypothetical distribution – which we will call earnings-maximizing distribution in the following – should be the same as the actual distribution of individuals across sectors in Eq. (4). If, however, entry barriers are in place, individuals will be under-represented in the segments where they would earn the highest expected wage and we will observe a statistically significant difference between the actual distribution of workers across sectors and the hypothetical earnings-maximizing distribution.

Let $\left\{\hat{\pi}_j\right\}_{j=1}^{J}$ denote the estimate of the actual distribution and let $\left\{\tilde{\pi}_j\right\}_{j=1}^{J}$ denote the estimate of the earnings-maximizing distribution.

[17] Andrews (2001) is the first to provide a theory of testing under such circumstances, but implementation of the test is nontrivial.

[18] We use Schwarz (SBC), consistent Akaike (cAIC) and Hannan–Quinn (H–Q) information criteria.

Table 3
Model with a two-segment informal sector.

Formal	Coeff.	(Std. error)	Informal 1	Coeff.	(Std. error)	Informal 2	Coeff.	(Std. error)
Intercept*	7.1584	0.3858	Intercept*	7.5951	0.3298	Intercept*	7.4615	0.5624
Sex*	0.3570	0.0738	Sex*	0.6511	0.0709	Sex*	0.4683	0.1212
Age*	0.1281	0.0199	Age*	0.1206	0.0173	Age*	0.0833	0.0297
Age^2/100*	−0.1175	0.0261	Age^2/100*	−0.1286	0.0225	Age^2/100*	−0.1038	0.0383
Education/Low*	0.3014	0.0986	Education/Low	0.1014	0.0918	Education/Low	0.1757	0.1645
Education/Medium*	0.6851	0.0930	Education/Medium*	0.2853	0.1209	Education/Medium	0.1265	0.2136
Education/High*	1.1867	0.0921	Education/High*	0.7811	0.1842	Education/High	0.2747	0.2734
Training*	0.1902	0.0633	Training	−0.0862	0.1084	Training*	0.6439	0.2003
Muslim	0.1378	0.0902	Muslim	−0.1232	0.0986	Muslim*	0.7320	0.2036
Christian	−0.0196	0.0853	Christian	−0.0731	0.1030	Christian	0.3996	0.2069
Abidjan	0.0712	0.0580	Abidjan*	0.1806	0.0692	Abidjan*	0.2566	0.1179
σ_F^*	0.8343	0.0194	$\sigma_{I_1}^*$	0.6450	0.0411	$\sigma_{I_2}^*$	1.2837	0.0567
ρ^*	0.1168	0.0493						
π_F*:	0.3392	0.0092	π_{I_1}*:	0.3661	0.0417	π_{I_2}*:	0.2947	0.0438
Expected log-wage:	11.3524		Expected log-Wage:	10.5051		Expected log-wage:	10.1013	
Expected wage:	105,085.84		Expected wage:	41,755.55		Expected wage:	28,026.71	
Selection equation:								
Intercept	−0.0422	0.0400		Number of obs. (cens):	2939			
Sex*	0.5682	0.0374		Number of obs. (mix):	2653			
Infants*	0.2705	0.0196						
Children*	0.2677	0.0162		Log-likelihood:	−5276.07			
Old	−0.0518	0.0439						
HH size*	−0.2693	0.0092						
Active members*	0.4709	0.0157						

Notes: Asterisk indicates significance at 5% level. Monthly wages in CFA Francs.

Both of these estimated distributions are plotted in Fig. 2. We observe that the fraction of individuals who, conditional on their personal characteristics, would be better off in the formal sector is almost double the actual share of formal sector employees. The contrary can be observed for the lower-paid informal segment (Informal-II), where the actual number of workers is three times higher than the number of workers who would choose to be employed in this segment for comparative advantage reasons. Table 4 reports the estimated probabilities for the actual, $\left\{\hat{\pi}_j\right\}_{j=1}^J$, and earnings-maximizing, $\left\{\tilde{\pi}_j\right\}_{j=1}^J$, distributions as well as the ratios of these estimated probabilities, $\hat{\pi}_j/\tilde{\pi}_j$, for each segment. We also report the respective bootstrap confidence intervals.

The results in Table 4 indicate that at a 5% significance level the actual probability of workers to be found in the formal or the lower-paid informal sector (Informal-2) is significantly different from the hypothetical earnings-maximizing probability. It is only equal for the higher-paid informal sector (Informal-1). We find $\hat{\pi}_j/\tilde{\pi}_j<1$ for the formal sector, $\hat{\pi}_j/\tilde{\pi}_j=1$ for the higher-paid informal sector and $\hat{\pi}_j/\tilde{\pi}_j>1$ for the lower-paid informal sector, which implies that:

(i) the share of workers who are actually employed in the formal sector is significantly lower than the share of workers that would choose to work there if no entry barriers existed,
(ii) the actual share of individuals found in Informal-1 is equal to the share of workers who would choose to work in this sector,
(iii) the actual share of workers in Informal-2 is significantly higher than the share of workers that would voluntarily stay in this sector.

We can therefore reject the hypothesis of unlimited sector mobility and a fully competitive labor market. The labor marked studied here includes involuntary employment at least in the lower-paid informal segment, where a significant fraction of workers would do better if working in another segment of the labor market. Furthermore, our results do neither support the hypothesis of full labor market segmentation, that considers all informal employment as a strategy of last resort. Fig. 2 and Table 4 clearly show that both informal segments (mainly Informal-1 but also Informal-2) contain a significant fraction of individuals who would not be better off in any other sector — at least with regard to expected earnings. We therefore conclude that the informal market consists of both voluntary and involuntary employment.

With the results of Table 4 we can also quantify the relative size of informal employment 'by choice' and 'by force'. The share of individuals who voluntary work in the informal sector and who would not be better off in the formal sector, is the sum of the shares of

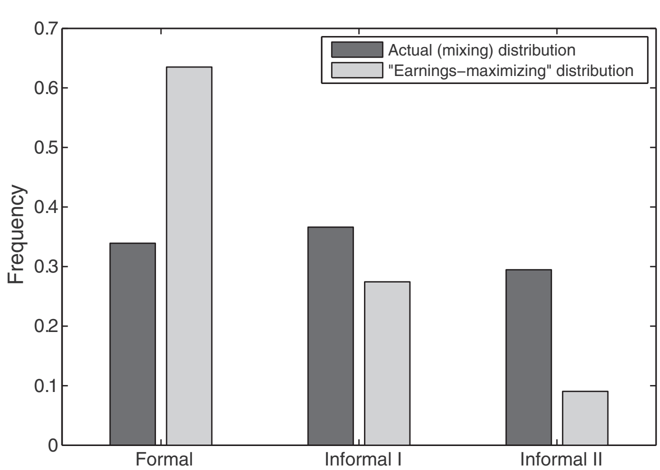

Fig. 2. Distribution of individuals across sectors.

Table 4
Distribution of individuals across sectors.

	Formal		Informal-1		Informal-2	
	Value	[95% conf. int.]	Value	[95% conf. int.]	Value	[95% conf. int.]
$\hat{\pi}_j$	0.3392	[0.3196, 0.3567]	0.3661	[0.2314, 0.4930]	0.2947	[0.1818, 0.4305]
$\tilde{\pi}_j$	0.6351	[0.3666, 0.7753]	0.2744	[0.1287, 0.5278]	0.0905	[0.0429, 0.1869]
$\hat{\pi}_j/\tilde{\pi}_j$	0.6038	[0.4270, 0.9369]	1.3999	[0.4855, 3.4732]	3.3766	[1.2218, 8.0180]

individuals found in each segment of the informal sector according to the earnings-maximizing distribution. The difference between the actual and hypothetical share of all informal employment is the share of informal employment that is a strategy of last resort to escape unemployment. According to our estimates, 44.8% of informal employment is not voluntarily, and it is mainly, but not exclusively, found in the lower-paid informal segment. Individuals who have a comparative advantage in the informal sector make up the remaining 55.2% of the informal labor market, and they are predominantly found in the higher-paid informal segment.[19]

One might question the robustness of the presented results with regard to the assumption made that individuals are earnings-maximizers rather than utility-maximizers. It is indeed possible to argue that our empirical results are a consequence of non-wage preferences for the informal sector and not the evidence of entry barriers into the formal sector. Given the significantly lower earnings in the informal sector, this would mean that the informal sector brings along considerable non-wage advantages that the formal sector cannot offer. However, we think that the informal sector should not have more (and we would argue rather less) positive non-wage features than the formal sector.[20] Hence, assuming that individuals are earnings-maximizers should not considerably bias our results. De Paula and Sheinkman (2008) further stress that informal activity is mainly driven by tax avoidance, which would provide an additional argument why individuals prefer being employed in the informal sector despite considerably lower wages (given the same individual characteristics). The income measure in this paper is after tax income so that a tax advantage of the informal sector should not have an influence on our results.

Our result that voluntary and involuntary informal employment co-exist is particularly important for the design of policies which aim to decrease informal activities to improve individuals' low and volatile earnings, and to increase fiscal revenues. For individuals for whom informal employment – at least until now – is the best opportunity, policies should aim at improving individuals' ‘poor’ endowments to enhance their earning possibilities in the formal sector. Measures to increase the attractiveness of formalization of enterprises might also be considered for this part of the informal sector. For individuals or enterprises who consider the informal sector as a strategy of last resort, policy interventions should counter entry barriers into the formal sector, for example, by easing formal registration or by lowering minimum wages in the formal sector.

4. Conclusion

In this paper we formulate an econometric model of the labor market that allows for unobserved earnings heterogeneity and accounts for sample selection caused by individuals' self-selection into the labor market. The model furthermore provides an intuitive approach for analyzing whether employment in the informal sector of the labor market is voluntary or a strategy of last resort.

We apply the model to study the structure of the urban labor market in pre-conflict Côte d'Ivoire. Our results show that the informal sector is composed of two segments with a distinct wage equation in each segment. We further find that both segments are considerable in size, each making about half of informal employment. In addition, we show that one segment of the informal sector is superior to the other in terms of significantly higher average earnings as well as higher returns to education and experience.

We also test whether the detected structure of the informal sector is a result of market segmentation that deters individuals from entering the desired sector, or a result of comparative advantage considerations of workers. Our results reject the hypothesis of unlimited sector mobility of workers, detecting a considerable number of informal workers that would be better off in another segment of the labor market (most, but not all, of them found in the lower-paid informal segment). However, a larger share of informal employees (mainly, but not solely, found within the higher-paid informal segment) seem to have a comparative advantage in the informal sector. Hence, the informal sector includes both individuals for whom informality is a strategy of last resort to escape unemployment and individuals who have a comparative advantage in the informal sector. The estimated size of involuntary informal employment is about 45% of the entire informal sector.

Our findings are particularly important for policies which aim to combat informal employment for employee protection or tax collection purposes, as they have to address both individuals who would like to switch to a formal job and individuals who currently have no incentive of doing so. For the theoretical modeling of labor markets in developing economies we conclude that neither solely competitive nor exclusively segmented labor market theories provide an appropriate explanation of labor market dynamics. Moreover, we suggest that we might even have to rethink the general assumption of no entry barriers into the informal sector, and acknowledge the possibility of entry barriers into its higher-paid segment. An interesting follow-up of this analysis could be to apply our developed model to various other developing economies for a cross-country comparison of the relative sizes of voluntary and involuntary informal employment. A regional comparison between Sub-Saharan African countries and Latin American countries seems to be especially interesting.

From an econometric perspective, our results show that tests for labor market segmentation in developing economies can be misspecified by ignoring the employment decision of individuals (e.g. Dickens and Lang, 1985; Cunningham and Maloney, 2001) or a possible latent heterogeneous structure of the labor market (e.g. Heckman and Sedlacek, 1985; Magnac, 1991; Pratap and Quintin, 2006).

Our paper also opens avenues for future methodological research. Our current framework allows to study latent segments within the informal labor market and to cautiously conclude about the voluntary and/or involuntary nature of informal employment. The proposed model is only a baseline framework and further extensions are possible. The most interesting one would be to explicitly model selection into each of the observed and/or unobserved segments, allowing error terms of segment-specific wage equations to be correlated across the labor market in addition to selection into employment in general. A known special case of this environment is a Roy model with multiple markets which we would obtain under perfect observability of sector affiliation. We acknowledge that such an extension could have an influence on the wage equations estimated within each single sector of the labor market, and hence also alter our conclusions about the voluntary/involuntary composition of the informal sector.[21]

[19] Note, that the magnitude of voluntary and involuntary informal employment is not necessarily equal to the size of high (Informal-1) and low (Informal-2) remuneration within the informal sector. Moreover, more than two unobserved segments with distinct wage equations (e.g. Informal-3, Informal-4 etc.) are possible, but still only two types of employment – voluntary and involuntary – can exist.

[20] Whereas the informal sector offers more flexibility for employees and less regulations for the self-employed, the formal sector provides employment rights and access to social security, medical insurance and pension funds for employees and legal protection, easier access to (and possibly lower cost of) capital and potential increases in business size and hence economies of scale for employers.

[21] The essential problem of this extended model is identification. Known identification results (Dahl, 2002) rely on observability of sector affiliation, which is incompatible with an informal labor market comprising involuntary and voluntary employment that cannot be distinguished by observable variables. Therefore, without at least some proxy measures of informal sector affiliation in the data, new identification results need to be worked out, which is a research project in itself.

Appendix. Auxiliary results and proofs

Lemma 1. Let regression equations

$$lny = \mathbf{x}'\beta + u, \quad u{\sim}N(0,\sigma), \qquad and \qquad y_s = \mathbf{z}'\gamma + u_s, \quad u_s{\sim}N(0,1),$$

describe the outcomes (y,y_s) of the wage and selection variables respectively and let (u,u_s) follow a bivariate normal distribution with correlation coefficient ρ. Then

$$f(y|y_s>0) = \frac{\varphi((\ln y - \mathbf{x}'\beta)/\sigma)}{\sigma\Phi(\mathbf{z}'\gamma)}\Phi\left(\frac{\mathbf{z}'\gamma + (\rho/\sigma)[\ln y - \mathbf{x}'\beta]}{\sqrt{1-\rho^2}}\right)$$

is the probability density of the observed wage outcomes given the selection rule $y_s>0$.

Proof. Using Bayes formula, we get

$$f(y|y_s>0) = \frac{P(y_s>0|y)f(y)}{P(y_s>0)}.$$

Once (u,u_s) are jointly normally distributed with the parameters as above, the density of u_s conditional on u is a normal density with mean $(\rho/\sigma)u$ and variance $1-\rho^2$. Thus

$$\begin{aligned} P(y_s>0|y) &= P(\mathbf{z}'\gamma + u_s>0|y) = P(u_s > -\mathbf{z}'\gamma|y) \\ &= P\left(\frac{u_s-(\rho/\sigma)u}{\sqrt{1-\rho^2}} > \frac{-\mathbf{z}'\gamma-(\rho/\sigma)u}{\sqrt{1-\rho^2}}|y\right) = 1-\Phi\left(\frac{-\mathbf{z}'\gamma-(\rho/\sigma)u}{\sqrt{1-\rho^2}}|y\right) \\ &= \Phi\left(\frac{\mathbf{z}'\gamma+(\rho/\sigma)u}{\sqrt{1-\rho^2}}|y\right) = \Phi\left(\frac{\mathbf{z}'\gamma+(\rho/\sigma)[lny-\mathbf{x}'\beta]}{\sqrt{1-\rho^2}}\right). \end{aligned}$$

Since the marginal density of u_s is normal with mean 0 and variance 1, we finally have

$$\begin{aligned} f(y|y_s>0) &= \Phi\left(\frac{\mathbf{z}'\gamma+(\rho/\sigma)[lny-\mathbf{x}'\beta]}{\sqrt{1-\rho^2}}\right)\frac{f(y)}{P(y_s>0)} \\ &= \Phi\left(\frac{\mathbf{z}'\gamma+(\rho/\sigma)[lny-\mathbf{x}'\beta]}{\sqrt{1-\rho^2}}\right)\frac{\frac{1}{\sigma}\varphi((lny-\mathbf{x}'\beta)/\sigma)}{P(u_s>-\mathbf{z}'\gamma)} \\ &= \Phi\left(\frac{\mathbf{z}'\gamma+(\rho/\sigma)[lny-\mathbf{x}'\beta]}{\sqrt{1-\rho^2}}\right)\frac{\frac{1}{\sigma}\varphi((lny-\mathbf{x}'\beta)/\sigma)}{1-\Phi(-\mathbf{z}'\gamma)} \\ &= \frac{\varphi((lny-\mathbf{x}'\beta)/\sigma)}{\sigma\Phi(\mathbf{z}'\gamma)}\Phi\left(\frac{\mathbf{z}'\gamma+(\rho/\sigma)[lny-\mathbf{x}'\beta]}{\sqrt{1-\rho^2}}\right), \end{aligned}$$

which completes the proof. □

Lemma 2. Let

$$f(x) = \frac{\varphi((x-\mu)/\sigma)}{\sigma\Phi(a)}\Phi\left(\frac{a+(\rho/\sigma)[x-\mu]}{\sqrt{1-\rho^2}}\right)$$

be a density of a random variable X defined on $\mathbb{R}$. Then bilateral Laplace transform $\phi\{f(x)\}(t)$ of this density is given by

$$\phi\{f(x)\}(t) = e^{\frac{1}{2}t^2\sigma^2 - t\mu}\frac{\Phi(a-t\sigma\rho)}{\Phi(a)}.$$

Proof. By definition of the bilateral Laplace transform,

$$\phi(t) = \int_{-\infty}^{+\infty} e^{-xt}f(x)dx = \int_{-\infty}^{+\infty}\frac{e^{-xt}\varphi((x-\mu)/\sigma)}{\sigma\Phi(a)}\Phi\left(\frac{a+(\rho/\sigma)[x-\mu]}{\sqrt{1-\rho^2}}\right)dx. \tag{A1}$$

Considering $e^{xt}\varphi((x-\mu)/\sigma)$ we notice that

$$\begin{aligned} e^{-xt}\varphi((x-\mu)/\sigma) &= e^{-xt}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} = \frac{1}{\sqrt{2\pi}}exp\left\{-xt-\frac{x^2-2x\mu+\mu^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}}exp\left\{-\frac{x^2-2x\left(\mu-t\sigma^2\right)+\mu^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}}exp\left\{-\frac{\left(x-\left(\mu-t\sigma^2\right)\right)^2}{2\sigma^2}+\frac{\left(\mu-t\sigma^2\right)^2-\mu^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}}exp\left\{-\frac{\left(x-\left(\mu-t\sigma^2\right)\right)^2}{2\sigma^2}+\left(t^2\sigma^2/2-\mu t\right)\right\} \\ &= e^{\frac{1}{2}t^2\sigma^2-t\mu}\varphi\left(\left(x-\left(\mu-t\sigma^2\right)\right)/\sigma\right). \end{aligned}$$

Substituting this back into (A1) and rearranging we get:

$$\begin{aligned} \phi(t) &= \int_{-\infty}^{+\infty} e^{\frac{1}{2}t^2\sigma^2-t\mu}\frac{\varphi\left(\left(x-\left(\mu-t\sigma^2\right)\right)/\sigma\right)}{\sigma\Phi(a)}\Phi\left(\frac{a+(\rho/\sigma)[x-\mu]}{\sqrt{1-\rho^2}}\right)dx \\ &= \frac{e^{\frac{1}{2}t^2\sigma^2-t\mu}}{\Phi(a)}\int_{-\infty}^{+\infty}\frac{1}{\sigma}\varphi\left(\frac{x-\left(\mu-t\sigma^2\right)}{\sigma}\right)\Phi\left(\frac{a+(\rho/\sigma)[x-\mu]}{\sqrt{1-\rho^2}}\right)dx \end{aligned}$$

and finally, for $\rho\neq 0$,

$$\phi(t) = \frac{e^{\frac{1}{2}t^2\sigma^2-t\mu}}{\Phi(a)}\int_{-\infty}^{+\infty}\Phi\left(\frac{x+\left(\frac{a\sigma}{\rho}-\mu\right)}{\frac{\sigma}{\rho}\sqrt{1-\rho^2}}\right)\frac{1}{\sigma}\varphi\left(\frac{x-\left(\mu-t\sigma^2\right)}{\sigma}\right)dx. \tag{A2}$$

For the integral in (A2) consider the change of variable $s=b-x$, $(ds=-dx)$, where b is an arbitrary constant. This implies

$$\begin{aligned} \phi(t) &= \frac{e^{\frac{1}{2}t^2\sigma^2-t\mu}}{\Phi(a)}\int_{-\infty}^{+\infty}\Phi\left(\frac{b-s+\left(\frac{a\sigma}{\rho}-\mu\right)}{\frac{\sigma}{\rho}\sqrt{1-\rho^2}}\right)\frac{1}{\sigma}\varphi\left(\frac{b-s-\left(\mu-t\sigma^2\right)}{\sigma}\right)ds \\ &= \frac{e^{\frac{1}{2}t^2\sigma^2-t\mu}}{\Phi(a)}\int_{-\infty}^{+\infty}\Phi\left(\frac{(b-s)-\left(\mu-\frac{a\sigma}{\rho}\right)}{\frac{\sigma}{\rho}\sqrt{1-\rho^2}}\right)\frac{1}{\sigma}\varphi\left(-\frac{s-b+\left(\mu-t\sigma^2\right)}{\sigma}\right)ds \\ &= \frac{e^{\frac{1}{2}t^2\sigma^2-t\mu}}{\Phi(a)}\int_{-\infty}^{+\infty}\Phi\left(\frac{(b-s)-\left(\mu-\frac{a\sigma}{\rho}\right)}{\frac{\sigma}{\rho}\sqrt{1-\rho^2}}\right)\frac{1}{\sigma}\varphi\left(\frac{s-\left(b-\left(\mu-t\sigma^2\right)\right)}{\sigma}\right)ds, \end{aligned} \tag{A3}$$

where the last step obtains by virtue of the symmetry of the normal probability density function. Let $g(x;\mu_x,\sigma_x)$ and $G(x;\mu_x,\sigma_x)$ denote the density and cumulative density functions of the normal distribution with mean μ_x and variance σ_x^2, respectively. Rewriting (A3) in terms of $g(x)$ and $G(x)$ we get

$$\phi(t) = \frac{e^{\frac{1}{2}t^2\sigma^2-t\mu}}{\Phi(a)}\int_{-\infty}^{+\infty}G_1\left(b-s;\mu-\frac{a\sigma}{\rho},\frac{\sigma}{\rho}\sqrt{1-\rho^2}\right)g_2\left(s;b-\left(\mu-t\sigma^2\right),\sigma\right)ds. \tag{A4}$$

The integral in (A4) is a convolution of two normally distributed variables. We know that for any two independent random variables $U{\sim}H_1$ and $V{\sim}H_2$ their sum $W=U+V$ follows the distribution $H(w)$ which is a convolution

$$H(w) = \int_{-\infty}^{+\infty}H_1(w-s)dH_2(s),$$

(see, e.g., Cramér, 1999, p.189–190). Furthermore, we know that the sum $W=U+V$ of two independent normally distributed variables $U\sim G_1(u;\mu_u,\sigma_u)$ and $V\sim G_2(u;\mu_v,\sigma_v)$ follows normal distribution $W\sim G\left(w;\mu_u+\mu_v,\sqrt{\sigma_u^2+\sigma_v^2}\right)$. Thus

$$\int_{-\infty}^{+\infty} G_1\left(b-s;\mu-\frac{a\sigma}{\rho},\frac{\sigma}{\rho}\sqrt{1-\rho^2}\right)g_2\left(s;b-\left(\mu-t\sigma^2\right),\sigma\right)ds$$

$$= G\left(b-s+s;\mu-\frac{a\sigma}{\rho}+b-\left(\mu-t\sigma^2\right),\sqrt{\frac{\sigma^2}{\rho^2}\left(1-\rho^2\right)+\sigma^2}\right) = G\left(b;b-\frac{a\sigma}{\rho}+t\sigma^2,\frac{\sigma}{\rho}\right)$$

$$= \Phi\left(\frac{b-\left(b-\frac{a\sigma}{\rho}+t\sigma^2\right)}{\sigma/\rho}\right) = \Phi\left(\frac{\frac{a\sigma}{\rho}-t\sigma^2}{\sigma/\rho}\right) = \Phi\left(\frac{\frac{\sigma}{\rho}(a-t\sigma\rho)}{\sigma/\rho}\right) = \Phi(a-t\sigma\rho).$$

Substituting this result into (A4) we finally have[22]: $\phi(t)=e^{\frac{1}{2}t^2\sigma^2-t\mu}\Phi(a-t\sigma\rho)/\Phi(a)$, which completes the proof. □

Proof of Proposition 1. We verify the necessary conditions for identifiability, provided by Teicher (1963), Theorem 2.

Let $a\equiv\mathbf{z}'\gamma$, $\mu j\equiv\mathbf{x}'\beta_j$ and let $\mathcal{H}=\{H\}$ be a family of the distributions with the density functions

$$h\left(y;\mu_j,\sigma_j,\rho_j\right)=\frac{\varphi\left(\left(lny-\mu_j\right)/\sigma_j\right)}{\sigma_j\Phi(a)}\Phi\left(\frac{a+\left(\rho_j/\sigma_j\right)\left[lny-\mu_j\right]}{\sqrt{1-\rho_j^2}}\right). \quad (A5)$$

Using Lemma 2, bilateral Laplace transform of (A5) is given by

$$\phi_j(t)=e^{\frac{1}{2}t^2\sigma_j^2-t\mu_j}\frac{\Phi\left(a-t\sigma_j\rho_j\right)}{\Phi(a)}.$$

Let S_{ϕ_j} denote the domain of definition of the transform $\phi_j(t)$. We need to show that there exists an ordering $\prec$ of the family H such that $H(y;\mu_1,\sigma_1,\rho_1)\prec H(y;\mu_2,\sigma_2,\rho_2)$ implies: (i) $S_{\phi_1}\subseteq S_{\phi_2}$, and (ii) there exists some $t_1\in\bar{S}_{\phi_1}$ such that $\lim_{t\to t_1}\phi_2(t)/\phi_1(t)=0$. Consider

$$\frac{\phi_2(t)}{\phi_1(t)}=e^{\frac{1}{2}t^2\left(\sigma_2^2-\sigma_1^2\right)-t(\mu_2-\mu_1)}\frac{\Phi(a-t\sigma_2\rho_2)}{\Phi(a-t\sigma_1\rho_1)}.$$

First we show that a general class of the finite mixtures with component densities (A5) is not identifiable.

- Order the family $\mathcal{H}$ of lexicographically, so that $H(y;\mu_1,\sigma_1,\rho_1)\prec_{\sigma,\mu,\rho}H(y;\mu_2,\sigma_2,\rho_2)$ means that: (a) $\sigma_1>\sigma_2$, (b) if $\sigma_1=\sigma_2$, then $\mu_1>\mu_2$ (or $\mu_1<\mu_2$), and (c) if $\sigma_1=\sigma_2$ and $\mu_1=\mu_2$, then $\rho_1>\rho_2$. Then (i) $S_{\phi_1}\subseteq S_{\phi_2}$ is fulfilled. However, (ii) for $\sigma_1=\sigma_2$ and $\mu_1=\mu_2$ we see that $\lim_{t\to+\infty}\Phi(a-t\sigma\rho_2)/\Phi(a-t\sigma\rho_1)$ is always undetermined whenever $\rho_1>\rho_2>0$ and $\lim_{t\to-\infty}\Phi(a-t\sigma\rho_2)/\Phi(a-t\sigma\rho_1)$ is always undetermined whenever $0>\rho_1>\rho_2$.

Applying the l'Hospital's rule to solve the indeterminacy we see that

$$\lim_{t\to\infty}\frac{\phi_2(t)}{\phi_1(t)}=\lim_{t\to\infty}e^{\frac{1}{2}t^2\left(\sigma_2^2-\sigma_1^2\right)-t(\mu_2-\mu_1)}\lim_{t\to\infty}\frac{\Phi(a-t\sigma_2\rho_2)}{\Phi(a-t\sigma_1\rho_1)}$$

$$=\lim_{t\to\infty}e^{\frac{1}{2}t^2\left(\sigma_2^2-\sigma_1^2\right)-t(\mu_2-\mu_1)}\lim_{t\to\infty}\left[\frac{\sigma_2\rho_2}{\sigma_1\rho_1}\right]\frac{\varphi(a-t\sigma_2\rho_2)}{\varphi(a-t\sigma_1\rho_1)}$$

$$=\lim_{t\to\infty}e^{\frac{1}{2}t^2\left(\sigma_2^2-\sigma_1^2\right)-t(\mu_2-\mu_1)}$$

$$\times\lim_{t\to\infty}\left[\frac{\sigma_2\rho_2}{\sigma_1\rho_1}\right]e^{-\frac{1}{2}t^2\left(\sigma_2^2\rho_2^2-\sigma_1^2\rho_1^2\right)+ta(\sigma_2\rho_2-\sigma_1\rho_1)}$$

$$=\left[\frac{\sigma_2\rho_2}{\sigma_1\rho_1}\right]\lim_{t\to\infty}e^{\frac{1}{2}t^2\left(\sigma_2^2\left[1-\rho_2^2\right]-\sigma_1^2\left[1-\rho_1^2\right]\right)-t([\mu_2-\mu_1]-a[\sigma_2\rho_2-\sigma_1\rho_1])}$$

and $\rho_1>\rho_2$ does not insure that the leading term in the exponent will always be negative for $\sigma_1<\sigma_2$, implying $\lim_{t\to\infty}\phi_2(t)/\phi_1(t)\neq 0$. The same result obtains if the ordering with respect to ρ is $\rho_1<\rho_2$.

Thus, there is no ordering that leads to $\lim_{t\to t_1}\phi_2(t)/\phi_1(t)=0$ at least for some $t_1\in\bar{S}_{\phi_1}$. Since the second condition fails, a class of finite mixtures of the family $\mathcal{H}$ is not identifiable.

Next we show that a general class of the finite mixtures of a subfamily of $\mathcal{H}$ is identifiable.

- Consider a subfamily $\mathcal{H}'\subset\mathcal{H}$ in which $\rho_j=\rho\;\forall j$. Order H' lexicographically, so that $H(y;\mu_1,\sigma_1)\prec_{\sigma,\mu}H(y;\mu_2,\sigma_2)$ means that: (a) $\sigma_1>\sigma_2$, (b) if $\sigma_1=\sigma_2$, then $\mu_1>\mu_2$. Then (i) $S_{\phi_1}\subseteq S_{\phi_2}$ is fulfilled, and (ii) for $t\to-\infty$ we see that $\lim_{t\to-\infty}\phi_2(t)/\phi_1(t)=0$, which satisfies the second condition of Theorem 2 in Teicher (1963).

Thus a general class of finite mixtures of a subfamily H' with the common correlation coefficient ρ is identifiable. □

Estimation of the hypothetical distribution of workers across sectors

The distribution in question is given by

$$P\left(i\in\mathcal{Y}_j\right)=P\left(E\left[\ln y_{ij}|y_{is}>0;\mathbf{x}_i\right]=\max_{l,l\in[1,J]}\left\{E\left[\ln y_{ij}|y_{is}>0;\mathbf{x}_i\right]\right\}\right),$$

$j=1,...,J$. Define the indicator function $I(y_{ij})$ such that

$$I\left(y_{ij}\right)=\begin{cases}1, & \text{if } E\left[\ln y_{ij}|y_{is}>0;\mathbf{x}_i\right]=\max_{l,l\in[1,J]}\left\{E\left[\ln y_{ij}|y_{is}>0;\mathbf{x}_i\right]\right\}\\ 0, & \text{otherwise}\end{cases}$$

Then the above distribution is estimated by

$$P\left(i\in\mathcal{Y}_j\right)=n^{-1}\sum_{i=1}^{n}\hat{I}\left(y_{ij}\right),$$

where the estimated sector-specific expected log-wage for every individual is given by

$$\hat{E}\left[lny_{ij}|y_{is}>0,\mathbf{x}_i\right]=\mathbf{x}_i'\hat{\beta}_j+\hat{\rho}\hat{\sigma}_j\frac{\varphi\left(-\mathbf{z}_i'\hat{\gamma}\right)}{1-\Phi\left(-\mathbf{z}_i'\hat{\gamma}\right)}.$$

[22] The same approach is used e.g. for the derivation of the Gini coefficient of the lognormal distribution (see Aitchison and Brown, 1963, [Aitchison, J., and J., Brown, "The Lognormal Distribution", Cambridge: Cambridge University press, 1963]).

References

Amaral, P., Quintin, E., 2006. A competitive model of the informal sector. Journal of Monetary Economics 53 (7), 1541–1553.

Andrews, D., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69 (3), 683–734.

Basu, A., 1997. Analytical Development Economics. MIT Press, Cambridge, MA.

Bennett, J., 2008. Formality, informality, and social welfare. IZA Discussion Paper No.3550.

Bennett, J., Estrin, S., 2007. Informality as a stepping stone: entrepreneurial entry in a developing economy. IZA Discussion Paper No.2950.

Cramér, H., 1999. Mathematical Methods of Statistics. Princeton University Press, Princeton, 19th Printing.

Cunningham, W., Maloney, W.F., 2001. Heterogeneity in the Mexican micro-enterprise sector: an application of factor and cluster analysis. Economic Development and Cultural Change 50 (1), 131–156.

Dahl, G., 2002. Mobility and the return to education: testing a Roy model with multiple markets. Econometrica 70 (6), 2367–2420.

De Paula, A., Sheinkman, J.A., 2008. The informal sector. *Penn Institute for Economic Research Working Paper*, 08-018.

Dickens, W.T., Lang, K., 1985. A test of dual labor market theory. The American Economic Review 67 (2), 792–805.

Fields, G.S., 1975. Rural–urban migration, urban unemployment and underemployment, and job-search activity in LDCs. Journal of Development Economics 2 (2), 165–187.

Fields, G.S., 1990. Labour market modelling and the urban informal sector: theory and evidence. In: Thurnham, D., Salome, B., Schwarz, A. (Eds.), The Informal Sector Revisited. Paris, OECD.

Fields, G.S., 2005. A guide to multisector labor market models. Paper prepared for the World Bank Labor Market Conferencehttp://www.ilr.cornell.edu/directory/gsf2/downloads/Multisector_mar05.pdf.

Gindling, T., 1991. Labor market segmentation and the determination of wages in the public, private-formal and informal sectors in San-Jose, Costa-Rica. Economic Development and Cultural Change 39 (3), 585–603.

Guha-Khasnobis, B., Kanbur, R., Ostrom, E., 2006. Linking the Formal and Informal Economy: Concepts and Policies. Oxford University Press, New York.

Harris, J.R., Todaro, M.P., 1970. Migration, unemployment and development: a two-sector analysis. The American Economic Review 60 (1), 126–142.

Hart, K., 1973. Informal income opportunities and urban employment in Ghana. Journal of Modern African Studies 11 (1), 61–89.

Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.

Heckman, J., Hotz, V., 1986. An investigation of the labor market earnings of Panamian males. The Journal of Human Resources 21 (4), 507–542.

Heckman, J., Sedlacek, G., 1985. Heterogeneity, aggregation and market wage functions: an empirical model of self-selection in the labor market. Journal of Political Economy 93 (6), 1077–1125.

Hussmanns, R., 2004. Defining and Measuring Informal Employment. Bureau of Statistics, ILO.

Leroux, B., 1992. Consistent estimation of a mixing distribution. Annals of Statistics 20 (3), 1350–1360.

Lewis, W.A., 1954. Economic development with unlimited supplies of labour. Manchester School 22 (2), 139–191.

Little, R., 1985. A note about models for selectivity bias. Econometrica 53 (6), 1469–1474.

Livingstone, I., 1971. Economic Policy for Development: Selected Readings. Penguin Books, Harmonds-worth.

Loayza, N., Rigolini, J., 2006. Informality trends and cycles. World Bank Policy Research Working Paper No.4078.

Magnac, T., 1991. Segmented or competitive labor markets. Econometrica 59 (1), 165–187.

Maloney, W.F., 1999. Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. World Bank Economic Review 13 (2), 275–302.

Maloney, W.F., 2004. Informality revisited. World Development 32 (7), 1159–1178.

Murphy, K., Topel, R., 1985. Estimation and inference in two step econometric models. Journal of Business and Economics Statistics 3 (4), 370–379.

Olsen, R., 1980. A least squares correction for selectivity bias. Econometrica 48 (7), 1815–1820.

Paulson, A.L., Townsend, R.M., 2005. Financial constraints and entrepreneurship: evidence from the Thai financial crisis. Economic Perspectives 3Q, 34–48.

Pratap, S., Quintin, E., 2006. Are labor markets segmented in developing countries? A semiparametric approach. European Economic Review 50 (7), 1817–1841.

Rama, M., 1998. Wage misalignment in CFA Countries. *World Bank Policy Research Working Paper*, WPS 1873.

Rosenzweig, M., 1988. Labor Markets in Low Income Countries: In: Chenery, H., Srinivasan, T.N. (Eds.), Handbook of Development Economics, vol. 1. North Holland Publishing, Amsterdam.

Stiglitz, J.E., 1976. The efficiency wage hypothesis, surplus labor, and the distribution of labour in LDCs. Oxford Economic Papers 28 (2), 185–207.

Teicher, H., 1963. Identifiability of finite mixtures. Annals of Mathematical Statistics 34 (4), 1265–1269.

World Bank, 2007. Informality: Exit and Exclusion. World Bank, Washington D.C.