

# Competitive and Segmented Informal Labor Markets\*

Isabel Günther

*University of Göttingen*

Andrey Launov

*University of Würzburg and IZA, Bonn*

January 11, 2007

## **Abstract**

It has been recently argued that the informal sector of the labor market in a developing economy shows a dual structure with one part of it being competitive to the formal sector and another part being the result of market segmentation. To test this hypothesis we formulate an econometric model which allows for a heterogeneous informal sector with unobserved individual affiliation and which takes into account selection bias induced by the employment decision of individuals. Our test results for the urban labor market in Côte d'Ivoire indeed show existence of both competitive and segmented employment in the informal sector.

JEL CODES: J42, O17

KEYWORDS: developing economy, informal labor market, segmentation, comparative advantage, selection bias, latent structure, finite mixture models.

---

\*We would like to thank Michael Grimm, Stephan Klasen, Klaus Wälde and participants of the IZA/World Bank Conference on Employment and Development, ESPE 2006 and of the Verein für Socialpolitik Annual Meeting for very useful comments. Any errors remain our responsibility.

# 1 Introduction

One often observed characteristic of urban labor markets in developing countries is the coexistence of a small well-organized formal sector with relatively high wages and attractive employment conditions with a large informal sector with low as well as 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 wages and working conditions in the formal and informal sector are the result of labor market segmentation or whether competitive labor market theories still hold.

Traditional dual labor market theories 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) for either institutional or efficiency-wage reasons (Stiglitz, 1976). Hence it is argued that workers in the informal sector earn less than identical workers in the formal sector and if no entry barriers existed, workers from the informal sector would enter the formal sector.

Whereas the empirically shown differences between earnings in the formal and informal sectors have not been questioned, it has been claimed that the mere existence of lower wages and lower returns to education and experience in the informal sector does not yet imply market segmentation (see e.g. Dickens and Lang, 1985, Heckman and Hotz, 1986, and Rosenzweig, 1988). More precisely, a labor market with two distinct wage equations does not constitute a segmented labor market as long as individuals are free to move between the two sectors (see e.g. Dickens and Lang, 1985). An explanation for the existence of a formal and informal segment in the labor market would rather be that a large number of those working in the informal sector choose doing so voluntarily, either because the informal sector has desirable non-wage features (Maloney, 2004) and individuals maximize their utility rather than their earnings, or because workers, given their characteristics, have a comparative advantage in the informal sector and would not do any better in the formal sector (e.g. Gindling, 1991).

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 tried to test which of these opposing views holds empirically. Among the most no-

table empirical contributions one can list Magnac (1990) and Gindling (1991). Magnac (1990) addresses the hypothesis of competitiveness in the framework of an extended Roy model whereas the paper of Gindling (1991) considers the same question of competitiveness in a framework of generalized regression with sample selection introduced by Lee (1983). Both find weak evidence of a competitive rather than a segmented labor market structure.

Most recent theory on urban labor markets in developing countries has combined the polar views of competitive and segmented labor markets and emphasized a more complex structure of the informal sector, with an “upper-tier” and “lower-tier” (Fields, 2005) or a “voluntary entry” and “involuntary entry” (Maloney, 2004) informal sector. The “upper-tier” represents the competitive part into which individuals enter voluntarily because, given their specific characteristics, they expect to earn more than they would earn in the formal sector. The “lower-tier”, to the contrary, is the part that consists of individuals which were rationed out of the formal (and, possibly, “upper-tier” informal) labor market.

This latest hypothesis about the structure of the urban informal labor market in developing economies has however not received satisfactory empirical treatment. The difficulty of testing such a hypothesis is that the affiliation of any given individual to either tier of the informal sector is unobservable, i.e. data on the causes of informal employment is in most cases missing; and an arbitrary division of the informal labor market based on observed individual characteristics or earnings would lead to biased estimates. A first step undertaken by Cunningham and Maloney (2001) was to represent the informal sector as a mixture of “upper-tier” and “lower-tier” enterprises. But as Cunningham and Maloney (2001) consider only informal entrepreneurs an option of choosing formal sector employment does not even exist in their model. Moreover, unlike Magnac (1990) and Gindling (1991), Cunningham and Maloney (2001) do not consider selection bias induced by the employment decision of individuals.<sup>1</sup>

In this paper we suggest a simple econometric framework which is able to model the possible unobserved heterogenous structure of the informal labor market as Cunningham and Maloney (2001) and at the same time considers sample selection bias as Magnac (1990) and Gindling (1991).

Following the argument of Maloney (2004) and Fields (2005) we first let the in-

---

<sup>1</sup>Magnac (1990) and Gindling (1991), on the other hand, ignore the hypothetical latent structure of the informal sector.

formal sector consist of a finite number of segments with unobservable affiliation of individuals and distinct earnings equations in each segment. Hence, the whole labor market is represented as a mixture model with both observable (for the formal sector) and unobservable (for the informal sector) membership. As the individual employment decision is also influenced by the outside option of being non-employed, we let the component densities depend on the market participation decision as in Heckman (1979). This leads to a finite mixture with sample selection, which allows us to estimate the distribution of individuals across the different segments of the labor market as well as unbiased earnings equation in each of them.

We then test, whether the possible detected heterogenous structure of the informal labor market is the result of labor market segmentation or not. Here, the finite mixture setting of the suggested model offers an intuitively appealing test. The rationale of this test is that under the assumption that individuals are earnings maximizers and can freely move between different parts of the labor market, the distribution of individuals across sectors induced by the earnings maximizing decision would be the same as the estimated distribution of individuals across sectors. Rejection of the equality of these two distributions will imply existence of entry barriers between different segments of the labor market, i.e. market segmentation.

The paper is structured as follows. In Section 2 we outline the econometric model and construct the test for market segmentation. Section 3 presents the data and the discussion of the estimation results. Here we also relate our model to existing empirical literature on segmented and competitive labor markets. Section 4 summarizes and concludes.

## 2 Econometric Model

### 2.1 Specification

**Finite Mixture** Assume that the labor market  $\mathcal{Y}$  consists of  $J$  disjoint segments  $\mathcal{Y}_j$  such that  $\mathcal{Y} = \bigcup_{j=1}^J \mathcal{Y}_j$ . Let earnings in each segment  $\mathcal{Y}_j$  of the labor market be the outcome of a random variable  $Y_j$  with a probability distribution of  $F(y|\theta_j)$ , where for all  $j \in [1, J]$ ,  $F(y|\theta_j)$  are distinct and independent of each other. Next, assume that the affiliation of a given individual earning  $y$  with any of the segments  $\mathcal{Y}_j$  is unobservable. However, it is known that the probability of any individual earning  $y$  belonging to

$\mathcal{Y}_j$  is given by  $P(y \in \mathcal{Y}_j) = \pi_j$ . With these assumptions we can write the density of individual earnings  $y_i$  as

$$f(y_i) = \sum_{j=1}^J f(y_i|\theta_j) \pi_j. \quad (1)$$

In other words, we suggest that the labor market consists of an arbitrary number of segments with a distinct earnings distribution in each of them. Our basic specification is thus a conventional mixture model.

Next, assume that in any segment  $\mathcal{Y}_j$  of the labor market  $\mathcal{Y}$ , the sector specific earnings are given by

$$y_i = \mathbf{x}_i\beta_j + u_i, \quad u_i \sim N(0, \sigma_j^2 | \mathbf{x}_i, y_i \in \mathcal{Y}_j), \quad (2)$$

where  $\mathbf{x}_i$  represents a set of personal characteristics that determine individual earnings  $y_i$ . Using (1) and (2) it is easy to show that the expected earnings of any individual drawn from the whole population  $\mathcal{Y}$  is given by  $E(y_i) = \sum_{j=1}^J [\mathbf{x}_i\beta_j] \pi_j$ . We can hence write down the population regression

$$y_i = E(y_i) + v_i, \quad y_i \in \mathcal{Y}, \quad (3)$$

in which the density of the error term  $v_i$  is a mixture of standard normal densities

$$h(v_i) = \sum_{j=1}^J \sigma_j^{-1} \varphi(\sigma_j^{-1} [y_i - \mathbf{x}_i\beta_j]) \pi_j. \quad (4)$$

**Sample Selection and Identification** One of the reasons why the regression in (3) might be misspecified is that earnings  $y_i$  are only observed if they exceed individuals' reservation wage (see Heckman, 1979). Being influenced by a subjective employment decision, the observed earnings sample may not necessarily be representative for the whole population. This gives rise to "sample selection" bias.

Let the employment decision of individuals depend on a set of personal characteristics  $\mathbf{z}_i$ . Writing down the selection equation

$$y_{is} = \mathbf{z}_i\gamma + u_{is}, \quad u_{is} \sim N(0, 1), \quad (5)$$

in which  $\mathbf{z}_i\gamma$  reflects the individual decision to work, we state that wages  $y_i$  in equation (2) are observed only if the realization of the selection variable  $y_{is}$  is positive.

Assuming that the errors of the  $\mathcal{Y}_j$ -specific earnings equation (2) and the selection equation (5) follow a bivariate normal distribution with  $Cov(u_i, u_{is}) = \rho_j \sigma_j$ , we can represent the sample selection bias as an omitted variable in (3):

$$E(y_i | y_{is} > 0) = E(y_i) + \sum_{j=1}^J E(u_i | u_{is} > -\mathbf{z}_i \gamma, \mathbf{x}_i, \theta_j) \pi_j,$$

where  $E(u_i | u_{is} > -\mathbf{z}_i \gamma) \neq 0$  unless  $\rho_j = 0$ . Since the second term in  $E(y_i | y_{is} > 0)$  is in general not equal to zero, the expected value of the error term  $v_i$  in (3) is not equal to zero. Hence the density of the error term in (4) is misspecified. The selected-sample counterpart of the regression in (3) reads

$$y_i = E(y_i | y_{is} > 0) + v_i, \quad \{y_i \in \mathcal{Y} : y_{is} > 0\}. \quad (6)$$

One can show that the error term  $v_i$  in this regression on a selected sample has a mixture density of

$$\begin{aligned} h(v_i | y_{is} > 0) &= \sum_{j=1}^J h(u_i | \theta_j, y_{is} > 0) \pi_j \\ &= \sum_{j=1}^J \left[ \frac{\sigma_j^{-1}}{\Phi(\mathbf{z}_i \gamma)} \varphi \left( \frac{y_i - \mathbf{x}_i \beta_j}{\sigma_j} \right) \Phi \left( \frac{\mathbf{z}_i \gamma + \rho_j \sigma_j^{-1} [y_i - \mathbf{x}_i \beta_j]}{(1 - \rho_j^2)^{1/2}} \right) \right] \pi_j, \quad (7) \end{aligned}$$

where  $\varphi$  and  $\Phi$  are the standard normal density and distribution functions.<sup>2</sup>

The above mixture model is a generalization of the Heckman regression with sample selection that allows for  $J$  different conditional distributions of the dependent variable instead of only one. From the very outset we assume that the work decision rule is the same across all sectors (i.e.  $\gamma_j = \gamma, \forall j$ ). This assumption however is not restrictive. It just implies that if all individuals were identical, they would have the same reservation wage.

Our next result demonstrates under which conditions the model in (7) rules out the existence of two distinct mixtures that have the same probability law for the observed dependent variable. The proof relies on the sufficient condition for identifiability of finite mixtures provided by Teicher (1963).

**Proposition 1** *For any given selection rule  $\{\mathbf{Z}, \gamma\}$ , the finite mixture (7) is identifiable if  $\rho_j = \rho, \forall j = 1, \dots, J$ .*

---

<sup>2</sup>Derivation of the component density of this mixture is presented in Appendix 1.

**Proof.** (See Appendix 1) ■

From the above proposition we see that the general class of finite mixtures with sample selection is not identifiable. So the attention should be narrowed to a sub-class where the correlation between the selection and earnings equations is the same for every segment of the labor market. In Appendix 1 it is also shown that the assumption of the common selection rule  $\gamma_j = \gamma, \forall j$  directly follows from the proof of Proposition 1. Finally, identifiability result of Proposition 1 is conditional on the agents' employment decision. However,  $\gamma$  is always identified from the data set that contains both employed and non-employed agents.

Given the identifiability restriction of Proposition 1 the ultimate specification of the error distribution becomes

$$h(v_i|\theta, \rho) = \sum_{j=1}^J \left[ \frac{\sigma_j^{-1}}{\Phi(\mathbf{z}_i\gamma)} \varphi\left(\frac{y_i - \mathbf{x}_i\beta_j}{\sigma_j}\right) \Phi\left(\frac{\mathbf{z}_i\gamma + \rho\sigma_j^{-1}[y_i - \mathbf{x}_i\beta_j]}{(1 - \rho^2)^{1/2}}\right) \right] \pi_j, \quad (8)$$

where  $\theta = \{\beta_j, \sigma_j\}_{j=1}^J$ . This completes the formulation of the model that allows for multiple segments with unobserved individual affiliation and accounts for employment decision of individuals.

As we will show next, the above formulated model suggests a simple testing procedure that allows telling whether the structure of the labor market under study is the result of market segmentation or whether the competitive sector choice mechanism still works. Thereby we become able to answer whether latent heterogeneity of the informal market, as suggested by Fields (2005), can explain observed earnings better than traditional models with a homogenous informal sector.

**Sector Choice** Assume that agents are earnings maximizers and log-earnings are completely specified by  $\mathbf{x}_i\beta_j$  (i.e. there exists no unobserved component for which we cannot account for). Let  $y_i^j$  denote the earnings of individual  $i$  in sector  $j$ . Competitive theory would imply that the individual-specific probability of choosing sector  $j$  is equal to

$$\begin{aligned} P(y_i \in \mathcal{Y}_j|\mathbf{x}_i) &= \prod_{l=1, l \neq j}^J P(\ln(y_i^j|\mathbf{x}_i) > \ln(y_i^l|\mathbf{x}_i)) \\ &= \prod_{l=1, l \neq j}^J P((\beta_j - \beta_l)\mathbf{x}_i + (\varepsilon_{il} - \varepsilon_{ij}) > 0). \end{aligned} \quad (9)$$

In the context of only two sectors Dickens and Lang (1985) notice that if there are no entry barriers to the formal sector, the difference in returns to individual characteristics in the two wage equations must be equal to the corresponding coefficients in the equation that determines the individual probability of sector membership. In our model it is straightforward to let sector affiliation probabilities  $\pi_j$  in (8) depend on individual characteristics. Though with  $J > 2$  the parametrization of  $\pi_j$  will be non-linear and therefore the result of Dickens and Lang (1985) will not carry over. So, instead of considering the individual-specific sector choice probabilities, we concentrate on the distribution of individuals across all possible sectors.

Assume that, knowing the returns in all sectors, an individual will choose the sector where his expected earnings given his personal characteristics are maximized. Then the probability distribution of individuals across sectors can be written as

$$P(y \in \mathcal{Y}_j) = P\left(E_{\mathbf{x}}[\ln(y_i^j | \mathbf{x}_i)] = \max_{l, l \neq j} \{E_{\mathbf{x}}[\ln(y_i^l | \mathbf{x}_i)]\}\right) \quad (10)$$

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

$$E[\ln(y_i^j | y_{is} > 0, \mathbf{x}_i)] = \mathbf{x}_i \hat{\beta}_j + \hat{\rho} \hat{\sigma}_j \frac{\varphi(-\mathbf{z}_i \hat{\gamma})}{1 - \Phi(-\mathbf{z}_i \hat{\gamma})}. \quad (11)$$

Equation (10) assumes free sector mobility and therefore provides us with the expected distribution of individuals across sectors if the market were competitive. It is also important to note that equation (10) does not exclude that returns to certain individual characteristics are lower in the chosen sector than in the alternative ones. Consequently, a sole comparison of estimated coefficients in sector-specific earnings equations cannot be informative about the labor market structure.

So, on the one hand, the distribution of workers across sectors under the assumption of no entry barriers is given by (10). On the other hand, the distribution of agents across sectors is also given by  $\{\pi_j\}_{j=1}^J$  in (8). This fact creates a basis for the test of free entry into the desired sector. If mixing probabilities  $\{\pi_j\}_{j=1}^J$  and the probabilities in (10), induced by the privately optimal sector choice, are not significantly different from each other, we obtain evidence of no entry barriers between the various segments of the labor market. Rejection of the equality of these two distributions will point at the existence of entry-barriers between sectors.



## 2.2 Implementation

For the above formulated model the following two-step estimation procedure may be suggested:

1. Estimate  $\gamma$  in the selection equation (5) by running Probit.
2. Use  $\mathbf{z}_i \hat{\gamma}$  as consistent estimates of  $\mathbf{z}_i \gamma$  to estimate the mixture model in (8).

This estimation approach fits into to the two-step framework of Murphy and Topel (1985) who demonstrate that under standard regularity conditions for the likelihood functions on both steps such a two-step procedure provides consistent estimates of the full set of parameters of interest.

In the second part of the suggested procedure parameters are estimated via maximum likelihood. For a general case of unobserved sector affiliation the appropriate log-likelihood function is

$$\ln \mathcal{L} = \sum_{i=1}^N \ln \left( \sum_{j=1}^J h_i(\theta_j, \rho | \mathbf{x}_i, \mathbf{z}_i \hat{\gamma}) \pi_j \right), \quad (12)$$

where  $h_i(\theta_j, \rho)$  is given in (8).

Typically, and this is also true for the present application, it is possible to observe from empirical data on labor markets in developing countries whether an individual belongs to the formal or informal sector. Hence, only the affiliation with any possible segment of the informal market remains unobservable. Denote the set of earnings outcomes in the formal sector by  $\mathcal{Y}_F$ . Then (12) becomes

$$\begin{aligned} \ln \mathcal{L} &= \sum_{i \in \mathcal{Y}_F} \ln h_i(\theta_F, \rho | \mathbf{x}_i, \mathbf{z}_i \hat{\gamma}) - N_F \ln \pi_F \\ &\quad + \sum_{i \notin \mathcal{Y}_F} \left[ \ln \left( \sum_{j=1}^{J-1} h_i(\theta_{I,j}, \rho | \mathbf{x}_i, \mathbf{z}_i \hat{\gamma}) \pi_{I,j} \right) \right], \end{aligned} \quad (13)$$

where  $N_F$  is the size of the formal sector. It is also straightforward to show that ML estimate of the fraction of formal workers in the economy is equal to their observed sample proportion.

Asymptotic covariance matrix of the estimated on the second step vector of parameters  $\xi = \left\{ \{\theta_j\}_{j=1}^J, \rho, \{\pi_j\}_{j=1}^{J-1} \right\}$  is given by

$$V(\xi) = D^{-1}(\xi) + D^{-1}(\xi) M(\xi, \gamma) D^{-1}(\xi), \quad (14)$$

where  $D(\xi)$  is the expected negative Hessian and  $M(\xi, \gamma)$  is the matrix constructed using scores from the first and second steps.<sup>3</sup>

Finally, we notice that the suggested two-step procedure is used merely for the reduction of computational complexity. Alternatively, one can take a full information approach. The likelihood function will then be

$$\ln \mathcal{L} = \sum_{i \in \mathcal{Y}} \ln [\ell_i(\xi, \gamma | y_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i) \Phi(\mathbf{z}_i \gamma)] + \sum_{i \in \mathcal{Y}^c} \ln (1 - \Phi(\mathbf{z}_i \gamma)), \quad (15)$$

where  $\ell_i$  stands for the individual contribution to the likelihood function in (12), or (13), if applicable, and  $\mathcal{Y}^c$  denotes the complementary set of non-employed individuals. In this case, the parameter space of the former model augments by  $\gamma$  which is to be estimated together with  $\xi$ .

When estimating the model we opt for the two-step approach. This ensures a well-behaved numerical problem that converges from a wide range of starting values. The model is estimated using BFGS method with analytical derivatives.

## 3 Empirical Application

### 3.1 Data Description

The data we use is drawn from the 1998 Ivoirian household survey, the *Enquete de Niveau de Vie*, which was undertaken by the *Institut National de la Statistique de la Cote d'Ivoire* (INSD) and the World Bank. We focus our analysis on the urban population and limit our sample to individuals who are between the age of 15 and 65 years. This leaves us with a sample of 5592 observations. Among these, we consider as inactive the individuals who voluntarily stay out of the labor market as well as those who are involuntarily unemployed, as this is only a negligible proportion of the inactive population.

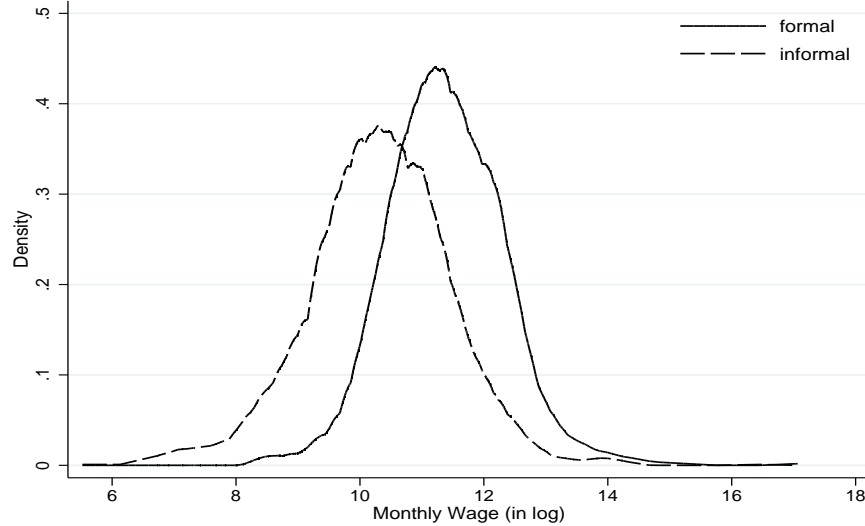
The active population is classified into the informal and formal sector according to reported primary employment.<sup>4</sup> The formal sector includes individuals working in the

---

<sup>3</sup>For the exact form of  $M(\xi, \gamma)$ , see Murphy and Topel (1985).

<sup>4</sup>Consideration of secondary informal employment of employees in the formal sector, which is an often observed characteristic of urban labor markets in developing countries, would imply that the earnings distributions in  $\mathcal{Y}_F$  and  $\mathcal{Y}_I$  are no longer disjoint. An extension of the model that incorporates this fact is left for future research.

Figure 1: Densities of Monthly Earnings



public sector as well as wage workers and self-employed in the formal private sector. As formal private we consider being employed in an enterprise which either pursues formal bookkeeping or offers written contracts and/or pay slips. The informal sector comprises the active population which is neither employed in the public nor in the private formal sector.

In Table 1 and Figure 1 we present the sample means and kernel densities of monthly formal and informal earnings. We use monthly wages instead of hourly wages because given the irregular and often constraint working hours in the informal sector we think that monthly wages reflect earning opportunities in the informal sector better than hourly wages. As expected, there is a large earnings differential between informal and formal workers. However, Figure 1 also demonstrates that despite the considerable difference in mean earnings, the densities of informal and formal earnings overlap to a large extent, indicating that not all informal employment is inferior to formal employment.

Table 1 also displays summary statistics on the variables used in the earnings 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 respectively. As expected, the educational level is the highest in the formal sector (8.1 years), with

Table 1: Descriptive Statistics of the Labor Market

	Total	Inactive	Active	
			Informal	Formal
Sample	100%	52.6%	31.3%	16.1%
Monthly earnings	98,815.0	–	64,837.8	164,995.1
Males	49.7%	40.6%	49.0%	80.6%
Age	30.0	25.2	34.7	36.6
Education (in years)	5.3	5.8	2.9	8.1
Literacy rate	64.1%	69.8%	44.4%	84.0%
Training after schooling	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%
– Other	14.4%	15.5%	12.6%	14.0%
Living in Abijan	49.6%	50.4%	42.2%	61.7%

*Note: Monthly earnings level in CFA Francs.*

somewhat lower and much lower educational attainment among inactive (5.8 years) and informal (2.9 years) workers. With regard to age, we find the youngest individuals among the inactive (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 a privilege of males, who constitute 80.6% of formal employees, which is most likely explained by the gender-specific education gap.<sup>5</sup> In contrast, only 49.0% of informal workers and 40.6% of inactive individuals are males. Finally an interesting observation can be made about the distribution of religious groups in the active population: despite the fraction of Muslims and Christians in the entire sample is almost the same, formal sector is dominated by Christians whereas the informal sector is dominated by Muslims. This can be explained by the specific composition of the government, i.e. the public sector, which constitutes a large part of the formal sector and which is dominated by Christians.

<sup>5</sup>For the whole sample, the average years of education for males is more than 60% higher than for females.

Hence, there are considerable differences in characteristics between both the inactive and active population as well as between workers employed in the informal and formal sectors. Systematic differences between active and inactive individuals highlights the danger of sample selection bias, that may arise if we ignore the employment decision of individuals in our model. The nature of systematic differences in characteristics of formal and informal workers is a bit less clear. It might be the result of self-selection of employees into the sectors where they maximize their earnings as well as the result of employers' discrimination based on workers' characteristics.

To specify the selection equation we use other variables such as the number of infants in the household, the number of children under 14 in the household, the number of old household members, household size and the number of active members in the household. The reason for this choice is twofold. First, the above listed variables filter out non-individual reasons for making job decision, such as family and environment matters, so that the magnitude of the earnings could be later explained by only individual qualities. Second, these variables provide sufficient exclusion restrictions to ensure identifiability of the slope coefficients in wage equations, while necessity of having exclusion restrictions in Heckman regression is strongly advocated by Olsen (1980) and Little (1985).

### **3.2 Composition of the Labor Market**

We start with an analysis of the sector composition of the labor market. The model in (8) allows for an arbitrary number of labor market segments where individual affiliation to any of them may not necessarily be observable. Moreover, the model takes into account selectivity induced by individuals' employment decision, which ensures consistent estimation of marginal returns to individual characteristics.

Initially we estimate two specifications: a model with a homogeneous informal sector and a model with an informal sector that consists of two segments with unobservable individual affiliation. Estimation results for both models are provided in Tables A1 and A2 of Appendix 2.

To decide on the ultimate number of segments in the labor market we use information criteria: Akaike (AIC), Schwarz (SBC), consistent Akaike (CAIC) and Hannan-Quinn (H-Q). In addition to these we use the Andrews (1988) goodness of fit test based

Table 2: Model Selection

	Homogeneous Informal Market	Two-Segment Informal Market	Three-Segment Informal Market
AIC	10713.85	<b>10616.23</b>	10616.46
SBC	10855.05	<b>10828.03</b>	10898.87
CAIC	10879.05	<b>10864.03</b>	10946.87
H-Q	10764.96	<b>10692.89</b>	10718.68
$\ln \mathcal{L}$	-5332.92	-5272.11	-5260.23

	Test Stat.	Cr.Value	Test Stat.	Cr.Value
Andrews Test	155.26	51.00	143.35	51.00

on the difference between observed and predicted cell frequencies.<sup>6</sup>

The results on model selection are presented in Table 2. The values of the Andrews  $\chi^2$ -test statistics imply a clear rejection of a homogenous informal sector. In addition to that, all information criteria uniformly show that the specification with a dual informal sector is superior to the model with a homogeneous informal sector. Thus the labor market under study consists of at least three distinct parts: the formal sector and two latent segments of the informal sector.

The results of the  $\chi^2$ -tests also show that fitted values from the model with two latent informal segments are still significantly different from the observed ones. Yet, extending the model to a three-segment informal sector does not bring any improvement in terms of information criteria. From the last column of Table 2 we see that all information criteria show that such a specification would overparameterize the model. In addition, the extended three-segment model would place a very low probability on the third segment of the informal market. The estimated size of this additional segment would make only 4.1% of the informal sector and 2.7% of the whole labor

<sup>6</sup>Partitioning of  $\mathbf{X}$  is made with respect to sex and formal sector membership. The covariance matrix of the difference between actual and predicted cell frequencies accounts for the fact that in our application  $\hat{\xi}$  is asymptotically not fully efficient (see Andrews 1988, p.1431-1432, for details).

market respectively. We hence conclude that the specification with a two-segment informal labor market is the best fitting and at the same time the most parsimonious model.<sup>7</sup>

Next very important finding is the significance of the correlation coefficient  $\rho$  (see Table A.2 in Appendix 2), which underlines the necessity of accounting for sample selection bias when estimating slope coefficients in segment-specific wage equations.

With regard to the properties of each segment of the labor market Table A.2 in Appendix 2 suggests that the two unobserved informal segments make 57.6% and 42.4% of the informal sector respectively, which shows that each of them constitutes a significant part of the informal sector. Expected wages in both informal segments are clearly below the expected wage in the formal sector; but there is an additional significant differential between expected earnings in the “upper”-tier (Informal-1) and “lower”-tier (Informal-2) informal sectors.

Wage equations across the three segments are quite diverse. As expected, returns to education and experience (measured in years of age) are high in the formal sector. Also, in the better-paid segment of the informal sector education and experience have a high and significant impact on earnings. But, whereas returns to experience in this segment are the same as in the formal sector, returns to education are already almost twice as low. In the lower-paid part of the informal sector returns to experience are only two thirds of the returns to experience in the formal and higher-paid informal sectors, and education appears to have no returns at all. Hence, workers in the lower-paid informal sector are stuck with very low wages almost independent of their abilities. Furthermore, whereas gender has a significant impact on earnings in all segments of the labor market, the male-female wage gap in the two informal segments is wider than in the formal sector.

Thus we do not only find that the labor market under study consists of three different segments. We also that each of these segments shows a quite distinct pattern of returns to individual characteristics. On a first glance, only the most general labor market structure, like the one proposed by Fields (2005) and Maloney (2004) and discussed in the Introduction, may seem to be supported by our estimates. Nevertheless,

---

<sup>7</sup>One should keep in mind that using information criteria is not only the simplest but also the only feasible way to decide on the appropriate specification. Since the component densities in (8) do not belong to the exponential family, neither the residual-based methods for selecting the optimal mixture (see Lindsay and Roeder, 1992) nor NPMLE of Heckman and Singer (1984) can be applied here.

even a significant diversity in the characteristics of each labor market segment should not necessarily mean that the labor market does not fit into either the segmented or the competitive labor market model. Rephrasing Basu (1997), it is beyond doubts that the labor market may be split into several segments. But as long as these segments possess the properties attributable to a competitive market, the whole labor market can still be treated as competitive. Alternatively, if entry barriers between some detected fragments could be found, the market would be segmented. Thus to learn about the nature of competition on the market certain sector choice mechanisms have to be explicitly addressed. This is done in the next Section.

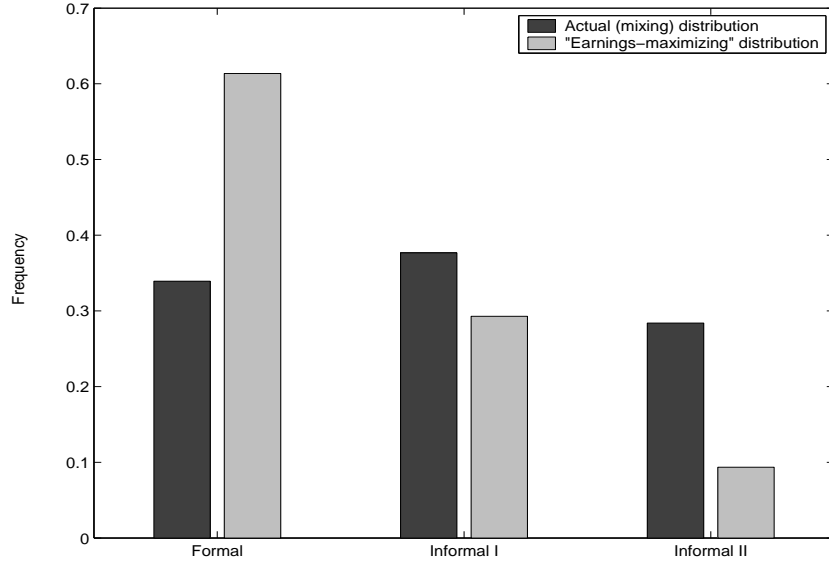
### 3.3 Segmented or Competitive Market?

We seek to answer whether employment in the two informal segments is the result of own comparative advantage considerations or a result of entry-barriers into the formal market, i.e. market segmentation. The theoretical argument for the analysis is presented in Section 2.1. Assuming that individuals are earnings maximizers, they will choose the sector where, given their characteristics, their expected earnings will be maximized. This induces the distribution of individuals across sectors which is formulated in (10). If no entry-barriers to either sector existed, the distribution in (10) would be the same as the actual distribution of individuals across sectors  $\{\hat{\pi}_j\}_{j=1}^J$  estimated from (8). If however certain entry-barriers were in place, individuals would be over-represented in undesired sectors and under-represented in sectors where they would actually maximize their expected earnings. As a result there should be a statistically significant difference between the estimated actual distribution of individuals across sectors  $\{\hat{\pi}_j\}_{j=1}^J$  and the distribution (10) induced by the earnings-maximizing sector choice of individuals.

Let  $\{\tilde{\pi}_j\}_{j=1}^J$  denote distribution (10) implied by the earnings-maximizing choice of individuals (for the rest of the paper we will call  $\{\tilde{\pi}_j\}_{j=1}^J$  the “earnings-maximizing” distribution). Figure 2 presents  $\{\hat{\pi}_j\}_{j=1}^J$  and  $\{\tilde{\pi}_j\}_{j=1}^J$  for each sector of the labor market. From this figure we see that the fraction of those who, conditional on their personal characteristics, would be better off in the formal sector almost doubles the actual share of formal sector employees. The contrary can be observed for the lower-paid informal segment, where the actual number of workers is almost three times as high as the number of workers that would choose to be employed in this segment for comparative



Figure 2: Distribution of Individuals across Sectors



advantage considerations.

In Table 3 we report bootstrap confidence intervals for the estimated sector affiliation probabilities  $\{\hat{\pi}_j\}_{j=1}^J$  and the predicted earnings-maximizing distribution  $\{\tilde{\pi}_j\}_{j=1}^J$ . In addition to that, for each sector we report the ratios of corresponding point pass values. The hypothesis of equality of the estimated mixing and predicted earnings-

Table 3: Distribution of Individuals across Sectors

	Formal		Informal-1		Informal-2	
	Value	[95% Conf.Interval]	Value	[95% Conf.Interval]	Value	[95% Conf.Interval]
$\hat{\pi}_j$	0.3392	[0.3224, 0.3554]	0.3767	[0.2325, 0.4867]	0.2840	[0.1717, 0.4279]
$\tilde{\pi}_j$	0.6136	[0.3727, 0.7740]	0.2929	[0.1425, 0.5237]	0.0935	[0.0337, 0.1813]
$\hat{\pi}_j/\tilde{\pi}_j$	0.5528	[0.4348, 0.9284]	1.2863	[0.5251, 3.1431]	3.0385	[1.2043, 8.5987]

maximizing distributions is rejected when at least one of the  $\hat{\pi}_j/\tilde{\pi}_j$  ratios significantly departs from unity. The results in Table 3 indicate that only for the “upper”-tier informal sector the estimated sector affiliation probability is not significantly different from the predicted earnings-maximizing sector size, i.e.  $\hat{\pi}_j/\tilde{\pi}_j = 1$ . For both formal and “lower”-tier informal sectors the hypothesis of equality of sector sizes is rejected, as, at 5% level,  $\hat{\pi}_j/\tilde{\pi}_j < 1$  for the formal sector and  $\hat{\pi}_j/\tilde{\pi}_j > 1$  for the “lower”-tier informal sector. This implies:

- (i) the share of workers that would chose to enter the formal sector is significantly higher than the share of workers actually employed in the formal sector,
- (ii) the share of individuals affiliated to the “upper”-tier of the informal sector is equal to the share of those who would chose to work in this sector,
- (iii) the actual share of workers in the “lower”-tier informal sector is significantly higher than the share of workers that would voluntarily choose to stay in this sector.

If we assume that workers can freely move between the two segments of the informal labor market, the three statements above imply that an entry-barrier between the formal and the “lower”-tier informal sector and no entry-barrier between the formal and the “upper”-tier informal sector exist. This supports the hypothesis of Fields (2005) and Maloney (2004). If we do however assume an entry-barrier between the two segments of the informal labor market, the three statements above could also imply an entry-barrier between the “upper” -tier informal and the formal sector. From an empirical perspective we cannot distinguish between these two cases. But given the nature of the informal market in developing economies, and in particular the fact that capital costs of setting up an informal business can be considered as negligible, the assumption of no entry-barriers between the two segments of the informal market seems to be quite realistic. Anyway, in both cases the hypothesis of unlimited inter-sectoral mobility with no entry-barriers between any sectors, and consequently fully competitive market, is rejected. This is the main result of this paper.

Last, if we assume that people are utility- rather than earnings-maximizers, it is also possible to argue that our empirical results are a consequence of non-wage preferences for the “lower” -tier informal sector and not an evidence of entry-barriers into the formal sector. Given the significantly lower earnings in the “lower” -tier informal sector, this would mean that being employed in this sector brings along considerable

non-wage advantages that the formal sector does not offer. Reflecting on possible non-wage attributes of the formal and informal sectors, it does however not seem that the informal sector has more positive non-wage features than the formal one. Whereas the informal sector offers just more flexibility, the formal sector provides access to employment certainty, social security, medical insurance, pension funds etc. Hence, the alternative of treating individuals as earnings maximizers should not bias results.

### 3.4 Individual Sector Affiliation

The mixture approach to modelling the informal sector provides us only with the probability distribution of agents across sectors. This means that we are not able to learn about the exact affiliation of a given individual to any segment of the market. At the same time, for policy targeting it would be important to know the specific characteristics of individuals found in the “lower”-tier and “upper”-tier informal sector. In this section we therefore make an attempt to describe a typical worker of each labor market segment. To learn about average individual characteristics in each segment we consider the following heuristic approach. Let the mixing distribution (i.e. the sector affiliation probabilities) in (8) depend on certain individual characteristics  $\mathbf{w}_i$ , so that  $\pi_j$  becomes

$$\pi_{ij} = \frac{\exp\{\alpha_j \mathbf{w}_i\}}{\sum_{j=1}^J \exp\{\alpha_j \mathbf{w}_i\}}, \quad (16)$$

where  $\alpha_j$  for the formal market is normalized to zero. The parametrization in (16) provides individual-specific probabilities of being a member of any sector of the labor market. Using these individual-specific probabilities as weights we can compute the weighted average for any individual characteristic  $\mathbf{w}_i$  in any sector  $j$ . The only information we need for that is the vector  $\hat{\alpha}$ , which is obtained by estimating the model in (8) with (16) instead of  $\pi_j$ .

In Table 4 we report the weighted characteristics for all sectors (notice that for the formal sector these estimated characteristics are the same as the observed characteristics presented in Table 1). As can be observed, members of the “lower” -tier informal sector are, on average, the oldest and the least educated of the working population. We might hence conclude that low levels of human capital as well as high age can also act as some form of entry-barrier to the formal sector.

Table 4: Expected Individual Qualities in each Sector <sup>a)</sup>

	Formal	Informal-1	Informal-2
Sex [“male”=1]	0.807	0.508	0.464
Age [years]	36.558	31.047	40.004
Years of education	8.096	3.311	2.309
Literacy rate	0.840	0.514	0.343

<sup>a)</sup> Prediction based on individual-specific affiliation probabilities

### 3.5 Contribution to Existing Empirical Literature

In this concluding section we briefly discuss the contribution of our analysis to the existing empirical literature in both developing and industrialized economies.

A benchmark paper in the empirical literature on testing for the dual structure of the labor market is the study of Dickens and Lang (1985). With U.S. data, Dickens and Lang (1985) were the first to account for unobservability of sector affiliation by using a switching regime regression. However, the follow up paper of Heckman and Hotz (1986) has provided a fundamental critique addressed not only to Dickens and Lang (1985), but also to the general framework of conducting such tests. Namely, Heckman and Hotz (1986) argue that such approaches suffer from the following potential sources of misspecification:

- (i) there may be multiple (not just two) sectors in the labor market,
- (ii) individuals are utility maximizers rather than earnings maximizers,
- (iii) inability to separate costs of moving between sectors from entry barriers,
- (iv) false distributional assumptions.

Subsequent papers have addressed only some of these shortcomings. In particular, Heckman and Sedlacek (1985) explicitly introduce non-wage valuation of the sectors and thereby tackle (ii); Magnac (1990) considers the cost of entry and resolves (iii).

In this paper we consistently discuss (i), developing a model that allows for sector multiplicity. Explicit introduction of heterogeneity in a form of distinct segments with unobserved affiliation provides a relative advantage in comparison to all models that originate from the Roy framework, as these models (including both Heckman and Sedlacek, 1985, and Magnac, 1990) are confined only to sectors with observed sector membership, out of which homogeneity of the informal market follows.<sup>8</sup>

Concerning (iv), with exception of Heckman and Sedlacek (1985), all existing models are not robust to distributional assumptions. One possible advantage of our framework in this respect is that by increasing the number of unobserved classes one can reduce the severity of misspecification, which is a positive feature of all mixture models.

In addition to that, we find significant sample selection bias induced by the employment decision of individuals. This means that the studies that consider a latent structure of the labor market (e.g. Dickens and Lang, 1985, and Cunningham and Maloney, 2001) but ignore sample selection may potentially suffer from this type of misspecification.

Taken together, the above overview shows that the framework created in the present paper certainly fills gaps in the empirical literature on informal sector heterogeneity and labor market segmentation in a developing economy.

## 4 Summary and Conclusions

In this paper we formulate an econometric model that allows for sector multiplicity when sector affiliation of any particular individual is not necessarily observable. Additionally, the model accounts for sample selection due to individuals' employment decision. The finite mixture setting of our model suggests a straightforward test for entry barriers to the formal sector of the labor market.

We apply the model to study the composition of the urban labor market in Côte d'Ivoire. Our results support the hypothesis that informal labor markets in developing countries are composed of two segments with a distinct wage equation in each of them. We further state that both informal sectors are considerable in size and make up 60% and 40% of informal employment respectively. In addition, we show that one segment

---

<sup>8</sup>Though, the framework of Magnac (1990) has definitely a great advantage in modelling entry costs and richer specification of nonparticipation.

of the informal sector (the “upper”-tier informal sector) is superior to the other (the “lower”-tier informal sector) in terms of significantly higher 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 formal sector, or rather a result of comparative advantage considerations of workers. We find that the “lower”-tier informal sector is the result of market segmentation whereas comparative advantage considerations seem to be the cause for the existence of the “upper”-tier informal sector. Hence, the informal sector comprises both, individuals who are voluntarily informal and individuals for whom the informal sector is a strategy of last resort to escape involuntary unemployment.

For the theoretical modelling of labor markets in developing economies this means that there may exist cases in which neither solely competitive theories, nor exclusively segmented labor market theories will provide a satisfactory explanation of labor market interactions. For the empirical literature our results clearly show that testing for labor market competitiveness in developing economies can be misspecified by either ignoring the employment decision of individuals (leading to selection bias) or the heterogeneity of the informal sector.

From a policy perspective, it is important to take into account the latent structure of the informal labor market as recommendations for the two informal segments should be quite different. Individuals who voluntarily participate in the informal sector just realize an opportunity to earn more than they would in the formal sector. Still, they have much lower earnings than employees in the formal sector, so policies should address their individual endowments to improve their earnings possibilities.<sup>9</sup>

With regard to the “lower” -tier informal sector, policy interventions have to counter entry-barriers into the formal sector. Moreover, agents found in this part of the informal market show especially low earnings. Hence, if the policy objective is to address the most disadvantaged, the “lower”-tier informal sector should receive highest priority. Our results indicate that especially the oldest and least educated as well as rather women than men are found in this most “disadvantaged” sector of the labor market.

---

<sup>9</sup>This part of the informal sector probably also constitutes a considerable part of tax evasion in developing countries and hence measures to improve tax collection in this part of the informal sector might be enhanced.

## References

- ANDREWS, D., "Chi-Square Diagnostic Tests for Econometric Models: Theory", *Econometrica*, 1988, p.1419-1453.
- BASU, A., "Analytical Development Economics", (Cambridge, MA: MIT Press, 1997).
- CUNNINGHAM, W. AND W.F. MALONEY, "Heterogeneity in the Mexican Micro-enterprise Sector: An Application of Factor and Cluster Analysis ", *Economic Development and Cultural Change*, 2001, p.131-156.
- DICKENS, W.T. AND K. LANG, "A Test of Dual Labor Market Theory", *American Economic Review*, 1985, p.792-805.
- FIELDS, G., S., "Labour Market Modelling and the Urban Informal Sector: Theory and Evidence", in Thurnham, D., Salome, B., Schwarz, A., eds, *The Informal Sector Revisited*, (OECD, 1990).
- FIELDS, G., S., "A Guide to Multisector Labor Market Models", Paper prepared for the World Bank Labor Market Conference, 2005.
- GINDLING, T., "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*, 1991, p.585-603.
- HARRIS, J.R. AND M.P. TODARO "Migration, unemployment and development: A two-sector analysis", *American Economic Review*, 1970, p.126-142.
- HECKMAN, J., "Sample Selection Bias as a Specification Error", *Econometrica*, 1979, p.153-161.
- HECKMAN, J. AND V. HOTZ "An Investigation of the Labor Market Earnings of Panamian Males", *Journal of Human Resources*, 1986, p.507-542.
- HECKMAN, J. AND G. SEDLACEK "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market", *Journal of Political Economy*, 1985, p.1077-1125.

- HECKMAN, J. AND B. SINGER, "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 1984, p.271-320.
- LEE, L.-F., "Generalized Econometric models with Selectivity", *Econometrica*, 1983, p.507-512.
- LINDSAY, B., AND K. ROEDER, "Residual Diagnostics for Mixture Models", *Journal of the American Statistical Association* , 1992, p.785-794.
- LITTLE, R., "A Note about Models for Selectivity Bias", *Econometrica*, 1985, p.1469-1474.
- MAGNAC, T., "Segmented or Competitive Labor Markets ", *Econometrica*, 1991, p.165-187.
- MALONEY, W.,F., "Informality Revisited ", *World Development*, 2004, p.1159-1178.
- MURPHY, K., AND R. TOPEL, "Estimation and Inference in Two Step Econometric Models", *Journal of Business and Economics Statistics*, 1985, p.370-379.
- OLSEN, R., "A Least Squares Correction for Selectivity Bias", *Econometrica*, 1980, p.1815-1820.
- ROSENZWEIG, M., "Labor Markets in Low Income Countries", in: Chenery H. and T.N. Srinivasan, eds., *Handbook of Development Economics, Volume 1*, (Amsterdam: North Holland, 1988).
- STIGLITZ, J., E., "The Efficiency Wage Hypothesis, surplus Labor, and the Distribution of Labour in LDCs ", *Oxford Economic Papers*, 1976, p.185-207.
- TEICHER, H., "Identifiability of Finite Mixtures", *Annals of Mathematical Statistics*, 1963, p.1265-1269.



# Appendix 1

## Component Density of the Error Term

Consider a component density  $f(u_i|u_{is} > -\mathbf{z}_i\gamma, \theta_j)$ . Using Bayes rule (for simplicity of notation we suppress conditioning on  $y_i \in Y_j$ ) we get

$$f(u_i|u_{is} > -\mathbf{z}_i\gamma, \theta_j) = \frac{P(u_{is} > -\mathbf{z}_i\gamma|u_i, \theta_j)f(u_i|\theta_j)}{P(u_{is} > -\mathbf{z}_i\gamma)}$$

Since joint distribution of  $(u_i, u_{is})$  is bivariate normal, conditional density  $f(u_{is} > -\mathbf{z}_i\gamma|u_i, \theta_j)$  follows  $N(\frac{\rho_j}{\sigma_j}u_i, 1 - \rho_j^2)$  and marginal density  $f(u_i|\theta_j) \sim N(0, \sigma_j^2)$ . Thus

$$\begin{aligned} f(u_i|u_{is} > -\mathbf{z}_i\gamma, \theta_j) &= P\left(\frac{u_{is} - \mathbf{z}_i\gamma - \rho_j\sigma_j^{-1}u_i}{\sqrt{1 - \rho_j^2}} > \frac{-\mathbf{z}_i\gamma - \rho_j\sigma_j^{-1}u_i}{\sqrt{1 - \rho_j^2}}\right) \frac{f(u_i|\theta_j)}{P(u_{is} > -\mathbf{z}_i\gamma)} \\ &= \Phi\left(\frac{\mathbf{z}_i\gamma + \rho_j\sigma_j^{-1}[y_i - \mathbf{x}_i\beta_j]}{\sqrt{1 - \rho_j^2}}\right) \frac{1}{\sigma_j} \varphi\left(\frac{y_i - \mathbf{x}_i\beta_j}{\sigma_j}\right) \frac{1}{\Phi(\mathbf{z}_i\gamma)} \end{aligned}$$

where  $\theta_j = \{\beta_j, \sigma_j, \rho_j\}$  and  $\varphi$  and  $\Phi$  are the probability density and distribution functions of the Standard Normal distribution.

**Proof of Proposition 1.** Consider the component density of (7)

$$h_j(y|\mu_j, \sigma_j, \rho_j) = \frac{\varphi(\sigma_j^{-1}[y - \mu_j])}{\sigma_j\Phi(a)} \Phi\left(\frac{a + \rho_j\sigma_j^{-1}[y - \mu_j]}{\sqrt{1 - \rho_j^2}}\right),$$

where  $\mu_j = \mathbf{x}\beta_j$  and  $a = \mathbf{z}\gamma$ . Bilateral Laplace transform of this density is given by

$$\begin{aligned} \phi_j[h(y)](t) &= \int_{-\infty}^{+\infty} e^{-ty} \frac{\varphi(\sigma_j^{-1}[y - \mu_j])}{\sigma_j\Phi(a)} \Phi\left(\frac{a + \rho_j\sigma_j^{-1}[y - \mu_j]}{\sqrt{1 - \rho_j^2}}\right) dy \\ &= \frac{1}{\Phi(a)} \int_{-\infty}^{+\infty} e^{-t(\sigma_j z + \mu_j)} \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \Phi\left(\frac{a + \rho_j z}{\sqrt{1 - \rho_j^2}}\right) dz \\ &= \frac{e^{-t\mu_j}}{\Phi(a)} \int_{-\infty}^{+\infty} \frac{e^{-t\sigma_j z - \frac{1}{2}z^2}}{\sqrt{2\pi}} \Phi\left(\frac{a + \rho_j z}{\sqrt{1 - \rho_j^2}}\right) dz \\ &= \frac{e^{\frac{1}{2}t^2\sigma_j^2 - t\mu_j}}{\Phi(a)} \int_{-\infty}^{+\infty} \frac{e^{-\frac{1}{2}(z + t\sigma_j)^2}}{\sqrt{2\pi}} \Phi\left(\frac{a + \rho_j z}{\sqrt{1 - \rho_j^2}}\right) dz. \end{aligned}$$

Using integration by parts yields

$$\begin{aligned}
\phi_j[h(y)](t) &= \frac{e^{\frac{1}{2}t^2\sigma_j^2-t\mu_j}}{\Phi(a)} \int_{-\infty}^{+\infty} \varphi(z+t\sigma_j) \Phi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) dz \\
&= \frac{e^{\frac{1}{2}t^2\sigma_j^2-t\mu_j}}{\Phi(a)} \left[ \Phi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \Phi(z+t\sigma_j) \Big|_{-\infty}^{+\infty} - \frac{\rho_j}{\sqrt{1-\rho_j^2}} \int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \Phi(z+t\sigma_j) dz \right] \\
&\stackrel{\rho_j \neq 0}{=} \frac{e^{\frac{1}{2}t^2\sigma_j^2-t\mu_j}}{\Phi(a)} \left[ 1 - \frac{\rho_j}{\sqrt{1-\rho_j^2}} \int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \Phi(z+t\sigma_j) dz \right].
\end{aligned}$$

Also notice that for  $\rho_j = 0$  the transform reduces to that of the Normal distribution.

Let  $S_j$  denote the domain of definition of  $\phi_j(t)$ . First, for any  $l, j$  we get  $S_j \subseteq S_l$ . This fulfills the first requirement of Theorem 2 in Teicher (1963).

Next, we seek for a limiting behavior of  $\phi_l(t)/\phi_j(t)$  once  $t \rightarrow t_*$  for some  $t_* \in \bar{S}_j$ .

$$\lim_{t \rightarrow +\infty} \frac{\phi_l(t)}{\phi_j(t)} = \lim_{t \rightarrow +\infty} \frac{e^{\frac{1}{2}t^2\sigma_l^2-t\mu_l}}{e^{\frac{1}{2}t^2\sigma_j^2-t\mu_j}} \lim_{t \rightarrow +\infty} \frac{1 - \frac{\rho_l}{\sqrt{1-\rho_l^2}} \int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_l z}{\sqrt{1-\rho_l^2}}\right) \Phi(z+t\sigma_l) dz}{1 - \frac{\rho_j}{\sqrt{1-\rho_j^2}} \int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \Phi(z+t\sigma_j) dz},$$

where, applying l'Hospital's rule to the second limit, we get

$$\lim_{t \rightarrow +\infty} \frac{\phi_l(t)}{\phi_j(t)} = \lim_{t \rightarrow +\infty} e^{\frac{1}{2}t^2(\sigma_l^2-\sigma_j^2)-t(\mu_l-\mu_j)} \lim_{t \rightarrow +\infty} \frac{\int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_l z}{\sqrt{1-\rho_l^2}}\right) \varphi(z+t\sigma_l) dz}{\int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \varphi(z+t\sigma_j) dz} \left[ \frac{\rho_l \sigma_l \sqrt{1-\rho_j^2}}{\rho_j \sigma_j \sqrt{1-\rho_l^2}} \right].$$

For the integral in the ratio above, omitting intermediate steps, it can be shown that

$$\begin{aligned}
\int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \varphi(z+t\sigma_j) dz &= \int_{-\infty}^{+\infty} \frac{\exp\left\{-\frac{1}{2} \frac{(a+\rho_j z)^2}{1-\rho_j^2}\right\}}{\sqrt{2\pi}} \frac{\exp\left\{-\frac{1}{2}(z+t\sigma_j)^2\right\}}{\sqrt{2\pi}} dz \\
&= \int_{-\infty}^{+\infty} \frac{\exp\left\{-\frac{1}{2} \left[ \frac{1}{1-\rho_j^2} (a+\rho_j z)^2 + (z+t\sigma_j)^2 \right]\right\}}{2\pi} dz \\
&= \int_{-\infty}^{+\infty} \frac{\exp\left\{-\frac{1}{2} \frac{(z+[a\rho_j+t\sigma_j(1-\rho_j^2)])^2}{1-\rho_j^2}\right\}}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(a-t\sigma_j\rho_j)^2\right\}}{\sqrt{2\pi}} dz
\end{aligned}$$

$$= \varphi(a - t\sigma_j\rho_j) \int_{-\infty}^{+\infty} \varphi\left(\frac{z + [a\rho_j + t\sigma_j(1 - \rho_j^2)]}{\sqrt{1 - \rho_j^2}}\right) dz = \varphi(a - t\sigma_j\rho_j) \sqrt{1 - \rho_j^2},$$

where the last equality obtains recognizing that the integral one step before is a Gaussian kernel.

Thus the limit of the ratio of the two transforms becomes

$$\begin{aligned} \lim_{t \rightarrow +\infty} \frac{\phi_l(t)}{\phi_j(t)} &= \lim_{t \rightarrow +\infty} e^{\frac{1}{2}t^2(\sigma_l^2 - \sigma_j^2) - t(\mu_l - \mu_j)} \lim_{t \rightarrow +\infty} \frac{\varphi(a - t\sigma_l\rho_l)}{\varphi(a - t\sigma_j\rho_j)} \left[ \frac{\rho_l\sigma_l}{\rho_j\sigma_j} \right] \\ &= \lim_{t \rightarrow +\infty} e^{\frac{1}{2}t^2(\sigma_l^2 - \sigma_j^2) - t(\mu_l - \mu_j)} \lim_{t \rightarrow +\infty} e^{-\frac{1}{2}t^2(\sigma_l^2\rho_l^2 - \sigma_j^2\rho_j^2) + ta(\sigma_l\rho_l - \sigma_j\rho_j)} \left[ \frac{\rho_l\sigma_l}{\rho_j\sigma_j} \right] \\ &= \lim_{t \rightarrow +\infty} e^{\frac{1}{2}t^2(\sigma_l^2[1 - \rho_l^2] - \sigma_j^2[1 - \rho_j^2]) - t([\mu_l - \mu_j] - a[\sigma_l\rho_l - \sigma_j\rho_j])} \left[ \frac{\rho_l\sigma_l}{\rho_j\sigma_j} \right] \end{aligned}$$

Repeating the ordering argument of Teicher (1963) we see that the general class of mixtures (7) is not identifiable because there is no lexicographic order  $h_j(y) \prec_{\sigma, \rho} h_l(y)$  that can insure that the leading term in the exponent will always converge to zero as  $t_* \rightarrow +\infty$ .

However, restricting the attention to a sub-class, in which  $\rho_l = \rho_j \forall l, j \in [1, J]$  we obtain the claimed result. For any  $l, j \in [1, J]$  let  $\rho_l = \rho_j$  and order the subfamily lexicographically so that  $h_j(y; \mu_j, \sigma_j, \rho) \prec h_l(y; \mu_l, \sigma_l, \rho)$  if  $\sigma_l < \sigma_j$  and  $\mu_l > \mu_j$  when  $\sigma_l = \sigma_j$ . Then for  $t_* = +\infty$ ,  $t_* \in \bar{S}_j$  we get

$$\lim_{t \rightarrow t_*} \phi_l(t)/\phi_j(t) = 0,$$

which fulfills the second and the last requirement of Theorem 2 in Teicher (1963).

Since the sufficient condition of Teicher (1963) applies, the sub-class of finite mixtures (7) with common  $\rho$  is identifiable. ■

**Remark** From the proof above it immediately follows that allowing for a sector-specific selection rule (i.e. letting  $a$  be  $a_j = \mathbf{z}\gamma_j$ ) leads to an unidentifiable model, since the limit of ratio writes

$$\lim_{t \rightarrow +\infty} \frac{\phi_l(t)}{\phi_j(t)} = \lim_{t \rightarrow +\infty} e^{\frac{1}{2}t^2(\sigma_l^2[1 - \rho_l^2] - \sigma_j^2[1 - \rho_j^2]) - t([\mu_l - \mu_j] - [a_l\sigma_l\rho_l - a_j\sigma_j\rho_j])} \left[ \frac{\rho_l\sigma_l\Phi(a_j)}{\rho_j\sigma_j\Phi(a_j)} e^{-\frac{1}{2}(a_l^2 - a_j^2)} \right]$$

and even within the considered sub-class of  $\rho_l = \rho_j = \rho$  there is no ordering over  $\{\mu\}$  which will insure that this limit is zero once  $\sigma_l = \sigma_j$ .

# Appendix 2

## Estimation Results

Table A.1: “The Model with the Homogeneous Informal Sector” §

<b>Formal</b>			<b>Informal</b>			
	Coeff.	(Std.Error)		Coeff.	(Std.Error)	
Intercept *	7.0595	0.3797	Intercept *	7.5028	0.2378	
Sex *	0.3443	0.0732	Sex *	0.5734	0.0538	
Age *	0.1300	0.0196	Age *	0.1062	0.0127	
Age <sup>2</sup> /100 *	-0.1184	0.0258	Age <sup>2</sup> /100 *	-0.1215	0.0165	
Education *	0.1058	0.0091	Education *	0.0421	0.0105	
Literacy	-0.1420	0.1140	Literacy	-0.0466	0.0844	
Training *	0.1598	0.0626	Training *	0.2006	0.0802	
Muslim	0.1542	0.0896	Muslim *	0.2580	0.0781	
Christian	-0.0185	0.0849	Christian	0.1225	0.0831	
Abijan	0.0809	0.0576	Abijan *	0.2273	0.0506	
$\sigma_F$ *	0.8288	0.0192	$\sigma_I$ *	1.0261	0.0174	
$\rho$ *	0.0953	0.0467				
	$\pi_F^*$ :	0.3392	0.0092			
				$\pi_I^*$ :	0.6608	0.0092
Expected log-Wage:		11.3524	Expected log-Wage:		10.3183	
Expected Wage:		105084.42	Expected Wage:		33816.37	
<b>Selection Equation</b>			Number of Obs. (missing): 2939			
			Number of Obs. (mixture): 2653			
Intercept	-0.0422	0.0400	Log-Likelihood: -5332.92			
Sex *	0.5682	0.0374				
Infants *	0.2705	0.0196				
Children *	0.2677	0.0162				
Old	-0.0518	0.0439				
HH Size *	-0.2693	0.0092				
Active Members *	0.4709	0.0157				

§ Here and henceforward asterisk indicates significance at 5% level.

Table A.2: “The Model with the Two-Component Informal Sector”

<b>Formal</b>	Coeff.	(Std.Error)	<b>Informal 1</b>	Coeff.	(Std.Error)	<b>Informal 2</b>	Coeff.	(Std.Error)			
Intercept *	7.0516	0.3799	Intercept *	7.5818	0.3225	Intercept *	7.4643	0.5803			
Sex *	0.3476	0.0734	Sex *	0.6659	0.0700	Sex *	0.4417	0.1257			
Age *	0.1301	0.0196	Age *	0.1199	0.0169	Age *	0.0816	0.0307			
Age <sup>2</sup> /100 *	-0.1187	0.0258	Age <sup>2</sup> /100 *	-0.1285	0.0221	Age <sup>2</sup> /100 *	-0.1012	0.0397			
Education *	0.1058	0.0091	Education *	0.0577	0.0160	Education	0.0210	0.0261			
Literacy	-0.1420	0.1140	Literacy	-0.1405	0.1103	Literacy	0.0706	0.1958			
Training *	0.1600	0.0626	Training	-0.1190	0.1063	Training *	0.6664	0.2031			
Muslim	0.1550	0.0896	Muslim	-0.0923	0.0979	Muslim *	0.7532	0.2103			
Christian	-0.0185	0.0850	Christian	-0.0505	0.1025	Christian	0.4026	0.2150			
Abijan	0.0807	0.0576	Abijan *	0.1871	0.0683	Abijan *	0.2530	0.1225			
$\sigma_F$ *	0.8294	0.0192	$\sigma_{I,1}$ *	0.6556	0.0388	$\sigma_{I,2}$ *	1.2960	0.0574			
$\rho$ *	0.1058	0.0497									
	$\pi_F^*$ :	0.3392	0.0092		$\pi_{I,1}^*$ :	0.3767	0.0403		$\pi_{I,2}^*$ :	0.2840	0.0401
Expected log-Wage:		11.3524	Expected log-Wage:		10.4956	Expected log-Wage:		10.0964			
Expected Wage:		105095.04	Expected Wage:		40992.12	Expected Wage:		28054.92			
<b>Selection Equation</b>											
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:		-5272.11			
Old	-0.0518	0.0439									
HH Size *	-0.2693	0.0092									
Active Members *	0.4709	0.0157									