Competitive and Segmented Informal Labor Markets

Isabel Günther
University of Göttingen

Andrey Launov
University of Würzburg and IZA, Bonn

February 26, 2006

Abstract

It has recently been argued that the informal sector in developing countries shows a dual structure, with part of the informal sector being competitive to the formal sector and part of the informal sector being the result of market segmentation. We formulate an econometric model to test this hypothesis. The model allows for sector multiplicity with unobserved sector affiliation in the informal sector and takes into account sample selection bias induced by the employment decision of individuals. An estimation of the model for the urban labor market in Côte d'Ivoire shows that the informal labor market is indeed composed of two segments with both competitive as well as segmented employment.

JEL Codes: J42, O17

Keywords: informal labor market, segmentation, comparative advantage, selection bias, latent structure, finite mixture.

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. The important question for both the understanding of the labor market and policy recommendations is whether this phenomenon is due to labor market segmentation or if competitive labor market theories still hold despite the observed differences in wages and working conditions in the formal and informal sector.

Traditional dualistic 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 for either institutional (Fields, 1990) or efficiency-wage reasons (Stiglitz, 1976). Hence it is argued that workers in the informal sector earn less than observationally identical workers in the formal sector and if no entry barriers existed, workers from the informal sector would enter the formal one.

Whereas the empirically shown differences between earnings in the formal and informal sectors have not been questioned, it has been claimed that mere existence of lower wages and lower returns to education and experience in the informal sector does not imply market segmentation.[1] In particular, a labor market with two distinct wage equations does not constitute a segmented labor market as long as freedom of choice between the two sectors is given (e.g. Dickens and Lang, 1985). An alternative explanation for the existence of two segments in the labor market would then rather assert that a large number of those working in the informal sector choose to do so voluntarily, either because the informal sector has desirable non-wage features (Maloney, 2004) and individuals maximize their utility rather than there earnings, or because workers 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 the informal employment as a voluntary choice of workers' based on income or utility maximization.

Most recent theory has combined these polar views and emphasized a more complex

[1] See Rosenzweig (1988) for literature review.

structure of the informal sector. Fields (2005) suggests that the informal sector is most likely to consist of two latent groups: the "upper-tier" and "lower-tier" informal market. 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 do in the formal sector. The "lower-tier", to the contrary, is the part that consists of the workers rationed out of the formal sector. Considerations of the same type can be found in Maloney (2004), who calls the two groups "voluntary entry" and "involuntary entry" informal sectors.

Despite the variety of the above described views on the structure of the labor market in a developing economy, neither of them has so far received a satisfactory empirical treatment. Among the most notable empirical contributions one can list Magnac (1990), who addresses the hypothesis of competitiveness in the framework of an extended Roy model. Despite finding evidence of a competitive rather than a dual labor market structure, the model of Magnac (1990) considers only a homogeneous informal sector and therefore cannot provide us with information about the validity of the most recent theoretical view of Maloney (2004) and Fields (2005). The paper of Gindling (1991) addresses the same question of competitiveness in a framework of generalized regression with sample selection introduced by Lee (1983). Though, like in Magnac (1990), homogeneity of the informal sector is again a drawback. Cunningham and Maloney (2001) offer possibly a first attempt to model the latent structure of the informal market explicitly, representing the informal sector as a mixture of "upper-tier" and "lower-tier" enterprises. However, Cunningham and Maloney (2001) consider only informal entrepreneurs so an option of choosing formal sector employment does not even exist in their model. Finally, unlike Magnac (1990) and Ginndling (1991), Cunningham and Maloney (2001) do not consider selection bias induced by the employment decision.

In this paper we suggest a relatively simple econometric framework, that is able to model for the sample selection bias, as Magnac (1990) and Gindling (1991), and at the same time consider the latent structure of the informal labor market, as in Cunningham and Maloney (2001). Following Maloney (2004) and Fields (2005) we let the informal sector consist of a finite number of groups with unobservable affiliation and distinct earnings equations in each group. As a result, the whole labor market is represented as a mixture model with both observable (for the formal sector) and unobservable (for the informal sector) group membership. As long as irrespective of group affiliation the individual employment decision is always influenced by the outside option of being non-employed, as in Heckman (1979), we make the component densities being dependent on this decision. This leads us to a finite mixture with sample selection, which is a generalization of the original model of Heckman (1979).

The finite mixture setting of the suggested model also offers an intuitively appealing test for the existence of entry barriers between different sectors (see Section 3 for other literature on testing duality). The rationale of this test is that under the assumption that agents are earnings maximizers and can freely enter different sectors, the distribution of agents across sectors induced by the earnings maximizing decision would be the same as the estimated mixing distribution. Rejection of the equality of these two distributions will imply existence of entry barriers, i.e. market segmentation.

The paper is structured as follows. In section 2 we outline the econometric model and discuss its features. Section 3 presents the data, the estimation results and relates our model to the existing empirical literature. Section 4 summarizes and concludes.

2 Econometric Model

2.1 Specification

Finite Mixture Assume that the labor market $\mathcal{Y}$ consists of J disjoint sets $\mathcal{Y}_j$ such that $\mathcal{Y} = \bigcup_{j=1}^{J} \mathcal{Y}_j$. Let earnings in each $\mathcal{Y}_j$ be outcomes of a random variable Y_j with probability distribution $F(y_i|\theta_j)$ such that for all j $F(y_i|\theta_j)$ are distinct and independent of each other. Next assume that for a given earnings outcome y_i affiliation with any of $\mathcal{Y}_j$ is unobservable, but it is known that $P(y_i \in \mathcal{Y}_j) = \pi_j$. Then the population density of individual earnings y_i will be

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

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

Assume that in any segment $\mathcal{Y}_j$ the wage equation is given by

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

where $\mathbf{x}_i$ represents a set of personal characteristics that determine individual earnings y_i. Johnson et al. (1992) show that r-th raw moment of any finite mixture can be computed by $\mu_r(y_i) = \sum_{j=1}^{J} \mu_r(y_i|\mathbf{x}_i, \theta_j)\pi_j$. Given this result we get the population regression

$$E(y_i) = \sum_{j=1}^{J} E(y_i|\mathbf{x}_i, \theta_j)\pi_j = \sum_{j=1}^{J} [\mathbf{x}_i\beta_j + E(u_i|\mathbf{x}_i)]\pi_j = \sum_{j=1}^{J} [\mathbf{x}_i\beta_j]\pi_j. \tag{3}$$

Sample Selection The reason why the regression in (3) may not be the ultimate specification is given in Heckman (1979). Namely, wages are observed only if they exceed the individual reservation wage. Consequently, being influenced by the subjective employment decision, the observed earnings sample need not necessarily be representative of the whole population.

Let the reservation wage of every individual depend on a set of personal characteristics $\mathbf{z}_i$. Writing down the selection equation

$$y_{is} = \mathbf{z}_i\gamma + u_{is}, \quad u_{is} \sim N(0,1), \tag{4}$$

in which $\mathbf{z}_i\gamma$ reflects the individual decision to work, we state that wages y_i in equation (2) are only observed if the realization of the selection variable y_{is} is, without loss of generality, positive.

Assume that the errors of the $\mathcal{Y}_j$-specific equation (2) and the selection equation (4) follow a bivariate Normal distribution

$$\begin{bmatrix} u_i \\ u_{is} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_j^2 & \rho_j\sigma_j \\ \rho_j\sigma_j & 1 \end{bmatrix} \middle| y_i \in \mathcal{Y}_j\right). \tag{5}$$

Repeating the argument of Heckman (1979), the sample counterpart of the population regression in (3) becomes

$$\begin{aligned} E(y_i|y_{is} &> 0) = \\ &= \sum_{j=1}^{J} E(y_i|y_{is} > 0, \mathbf{x}_i, \theta_j)\pi_j = \sum_{j=1}^{J} \left[\mathbf{x}_i\beta_j + E(u_i|u_{is} > -\mathbf{z}_i\gamma, \mathbf{x}_i, \theta_j)\right]\pi_j \\ &= E(y_i) + \sum_{j=1}^{J} E(u_i|u_{is} > -\mathbf{z}_i\gamma, \mathbf{x}_i, \theta_j)\pi_j, \end{aligned} \tag{6}$$

where $E(u_{ij}|u_{is} > -\mathbf{z}_i\gamma) \neq 0$ unless $\rho_j = 0$. Both (5) and (6) imply that as a consequence of selection the error term v_i in the regression on the observed sample

$$y_i = E(y_i) + v_i$$

will follow a mixture distribution

$$h(v_i|\theta_j) = \sum_{j=1}^{J} \left[\frac{\sigma_j^{-1}}{\Phi(\mathbf{z}_i\gamma)} \varphi\left(\frac{y_i - \mathbf{x}_i\beta_j}{\sigma_j}\right) \Phi\left(\frac{\mathbf{z}_i\gamma + \rho_j\sigma_j^{-1}[y_i - \mathbf{x}_i\beta_j]}{\left(1-\rho_j^2\right)^{1/2}}\right)\right] \pi_j, \tag{7}$$

where φ and Φ are the standard normal density and distribution functions.[2]

The above mixture model is a generalization of Heckman regression with sample selection that allows for J different generation processes of the dependent variable instead of only one, as in the classical model. 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 is, however, by no means restrictive. It just implies that if all individuals were identical, they would have had the same reservation wage.

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

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

Proof. (See Appendix) ■

From the above proposition we see that the general class of finite mixtures with sample selection is not identifiable. So the attention should be restricted to a sub-class in which correlation between selection and wage equations is the same in every segment. Additionally, as shown in the Appendix, the assumption of the common selection rule $\gamma_j = \gamma$, $\forall\, j$ follows from the proof. Finally, identifiability result of Proposition 1 is conditional on the agents' employment decision. However, γ 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 becomes

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

where $\theta_j = \{\beta_j, \sigma_j\}$. This specification is rich enough to provide us with exact results about market structure in presence of unobserved sector affiliation. Thereby it enables

[2] Derivation of the component density in (7) replicates the derivation of the likelihood function for the standard Heckman selection model. For completeness, we also present it in the Appendix.

us to answer if the model with heterogenous informal market, as suggested by Fields (2005), can explain more than the traditional dual models.

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 which we cannot account for). Competitive theory would imply that the individual-specific probability of choosing sector j is equal to

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

In the context of only two sectors Dickens and Lang (1985) notice that if there are no entry barriers to the formal sector, the vector of 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, despite it is easy to let sector affiliation probabilities π_j in (8) be dependent on individual characteristics, with $J > 2$ the parametrization of π_j will be non-linear and the equality result of Dickens and Lang (1985) will not carry over. Therefore, instead of considering the individual-specific sector choice probabilities, we concentrate on the distribution of agents over all possible sectors.

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

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

Equation (10) assumes free sector mobility and therefore provides us with the expected distribution of individuals on a competitive market.[3] 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 $\{\pi_j\}_{j=1}^J$ and the estimated

[3] Also note that this fact does not exclude that returns to certain individual characteristics in the chosen sector are lower then in the alternative ones. Consequently, plain comparison of estimated coefficients in sector-specific earnings equations cannot be informative about the market structure.

probabilities in (10) are not significantly different from each other, one obtains the equivalence between privately optimal and actual distributions of individuals over sectors, hence, the indication of no entry barriers between the segments of the market. Rejection of the equality of these two distributions will point at existence of certain barriers of an unknown form.

The analysis of sector choice and further issues connected with the above implied test are discussed in detail in Section 3.3.

2.2 Implementation

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

1. On the first step estimate γ in (4) running Probit.
2. On the second step use $\mathbf{z}_i\hat{\gamma}$ as consistent estimates of $\mathbf{z}_i\gamma$ to estimate the mixture model in (8).

This approach to estimation of the model 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 two-step procedure provides consistent estimates of the full set of the parameters of interest.

On the second step of the suggested procedure parameters of the mixture model are estimated by maximum likelihood. For a general case of unobserved sector affiliation the appropriate log-likelihood function is

$$\ln \mathcal{L} = \sum_{i=1}^{N} \ln \left(\sum_{j=1}^{J} h_i \left(\theta_j, \rho | \mathbf{x}_i, \mathbf{z}_i\hat{\gamma}\right) \pi_j \right), \tag{11}$$

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

Typically, and this is also true for the present application, it is possible to observe from the data whether an agent belongs to the formal sector. So 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 (11) modifies to

$$\begin{aligned} \ln \mathcal{L} &= \sum_{i \in \mathcal{Y}_F} \ln h_i \left(\theta_F, \rho | \mathbf{x}_i, \mathbf{z}_i\hat{\gamma}\right) - N_F \ln \pi_F \\ &+ \sum_{i \notin \mathcal{Y}_F} \left[\ln \left(\sum_{j=1}^{J-1} h_i \left(\theta_{I.j}, \rho | \mathbf{x}_i, \mathbf{z}_i\hat{\gamma}\right) \pi_{I.j} \right) \right], \end{aligned} \tag{12}$$

where N_F is the size of the formal sector. It is also straightforward to show that MLE 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

$$V(\xi) = D^{-1}(\xi) + D^{-1}(\xi)M(\xi,\gamma)D^{-1}(\xi), \tag{13}$$

where $D(\xi)$ is the expected negative Hessian and $M(\xi,\gamma)$ is the matrix constructed using scores from the first and second steps.[4]

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

$$\ln \mathcal{L} = \sum_{i\in\{Y\}} \ln \left[\ell_i(\xi,\gamma|y_i,\mathbf{x}_i,\mathbf{w}_i,\mathbf{z}_i)\Phi(\mathbf{z}_i\gamma)\right] + \sum_{i\in\mathcal{Y}^c} \ln\left(1-\Phi(\mathbf{z}_i\gamma)\right), \tag{14}$$

where ℓ_i stands for the individual contribution to the likelihood function in (11) [(12), if applies] and $\mathcal{Y}^c$ denotes the complementary set of non-employed individuals. In this case the parameter space of the former model augments by γ which has to be estimated together with ξ.

3 Empirical Application

3.1 Data and Estimation Method

The data we use is drawn from the Ivorian household survey, the *Enquete de Niveau de Vie*, of 1998 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 between 15 and 65 years old. This leaves us with a sample size of 5592 individuals. Among these, we consider as inactive individuals who voluntarily stay out of the labor market as well as those who are involuntarily unemployed (which is however a negligible proportion of the inactive population).

The active population is classified into the informal and formal sector. The formal sector includes individuals working in the public sector as well as wage workers and self-employed in the formal private sector. As formal private we consider being employed in

[4]For exact form of $M(\xi,\gamma)$ see Murphy and Topel (1985).

Table 1: Descriptive Statistics of the Labor Market

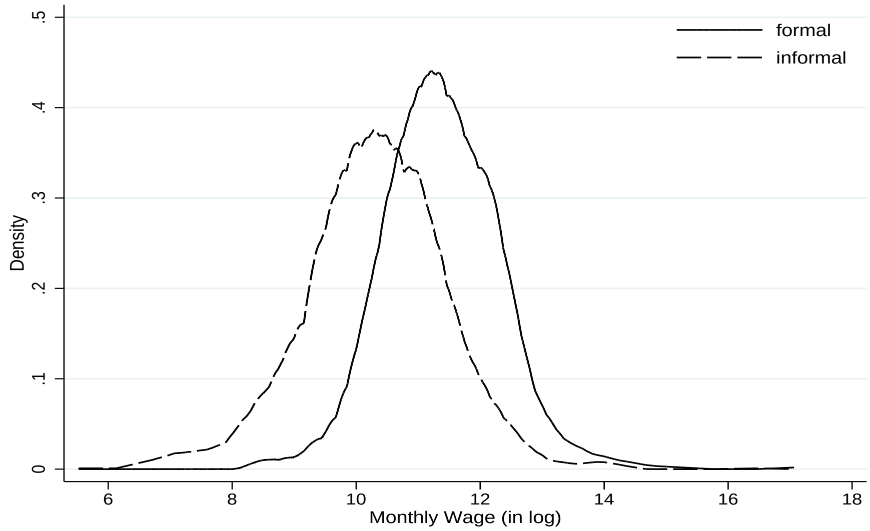
	Total	Inactive	Active	
			Informal	Formal
Sample	100%	52.6%	31.3%	16.1%
Monthly Wage	98,815.0	–	64,837.8	164,995.1
Males	49.7%	40.6%	49.0%	80.6%
Age	30.0	25.2	34.7	36.6
Education (years)	5.3	5.8	2.9	8.1
Literacy rate	64.1%	69.8%	44.4%	84.0%
Training after schooling	17.6%	11.1%	14.7%	44.3%
Religion:				
– 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 Wage in CFA Francs.

an enterprise which either pursues formal bookkeeping or offers written contracts or pay slips. The informal sector comprises the active population which is neither employed in the public nor in the private formal sector. The survey contains data on monthly wages as well as detailed information on socio-economic and demographic characteristics of individuals. In Table 1 we present summary statistics of the variables used for the earning equations for the population as a whole, as well as for its inactive and the "informal" and "formal" parts. As expected, there is a large earnings differential between informal and formal workers. However, Figure 1 also demonstrates that despite the big difference in mean earnings the densities of informal and formal monthly labor earnings overlap to a large extent, indicating that not all informal work is inferior to formal employment.

Also, as expected, education level and literacy rates are the highest in the formal sector. In addition membership in the formal sector is a privilege of males, who constitute 80% of formal employees, which is most likely explained by the gender-specific

Figure 1: Densities of Monthly Wages



education gap.[5] Finally an interesting observation can be made about the distribution of religion 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 informal sector is dominated by Muslims.

To specify the selection equation of the model (see p.6) we use further 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. When estimating the model we opt for the two-step approach described on p.9. This ensures a well-behaved numerical problem that converges from a wide range of starting values. The model is estimated using BFGS algorithm with analytical derivatives.

3.2 Composition of the Labor Market

We first analyze the sector composition of the labor market. The developed model in (8) first of all allows for an arbitrary number of segments where individual affiliation to any

[5]For the whole sample, the average length of education among males is more than 60% higher than among females.

Table 2: Model Selection

	Homogeneous Informal Market		Two-Segment Informal Market	
AIC	10689.85		10580.23	
CAIC	10879.05		10864.03	
SBC	10855.05		10828.03	
	Test Statistic	Cr.Value	Test Statistic	Cr.Value
Andrews' χ^2-Test	155.26	51.00	143.35	51.00

of them may not necessarily be observable. Second, and not less important, the model takes into account selectivity induced by employment decision, which ensures consistent estimation of conditional means of the segment-specific earnings distributions.

We estimate two specifications: the model with homogeneous informal sector and the model with an informal sector that consists of two latent groups. Estimation results for both models are provided in Tables A1-A2 of the Appendix. To decide on the ultimate number of segments on the market we use information criteria (Akaike, consistent Akaike and Schwarz) and Andrews (1988) goodness of fit test based on the difference between observed and predicted cell frequencies.[6]

The results on model selection are presented in Table 2. First of all, the values of the Andrews χ^2 test statistics indicate clear rejection of the homogeneity of the informal sector. In addition to that, all information criteria uniformly show that the specification with dichotomous informal sector is superior to the homogeneous model. Thus the labor market under study consists of at least three distinct parts: the formal

[6] Andrews (1988) shows that if $P(\Gamma)$ is the empirical measure and $F(\Gamma, \theta)$ is the conditional empirical measure defined on a partition $\mathbf{Y} \times \mathbf{X} = \cup_i \gamma_i$ and $v(\Gamma, \theta) \equiv \sqrt{n}\,(P(\Gamma) - F(\Gamma, \theta))$, then: $v(\Gamma, \theta)'\Sigma^{+}v(\Gamma, \theta) \sim \chi^2_{rk|\Sigma|}$, where Σ is the covariance matrix of $v(\Gamma, \theta)$. Three different estimators of Σ are offered. Here we use a $\hat{\Sigma}_{2n}$-estimator for the case when $\hat{\theta}$ is asymptotically not fully efficient, which is true for our two-step procedure (see Andrews 1988, p.1431-1432). Finally, for $\mathbf{Y} \times \mathbf{X}$ we partition $\mathbf{X}$ with respect to sex and formal sector membership and for each group form cells for $\mathbf{Y}$.

sector and two latent segments of the informal sector.

Even though cell frequencies generated by the better-fitting model are still significantly different from the observed ones, consideration of the specification with the three-part informal sector does not bring any improvement in terms of information criteria. As the attempt to further refine heterogeneity of the informal sector leads to the unnecessary overparametrization of the model, we conclude that the specification with the dichotomous informal market is the best fitting and at the same time the most parsimonious one.

Let us analyze the properties of each segment of the labor market in more detail. From the results reported in Table A.2 one can infer that the two latent informal segments make 57.5% and 48.5% 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 in addition to that there is a significant earnings differential between the mean earnings in the two informal sectors.

Wage equations across the three segments are also quite diverse. As expected, returns to education and experience are high in the formal sector. In the better-paid informal sector experience as well as education have also a high and significant impact on wages. But whereas returns to experience are the same as in the formal sector, returns to education are almost twice as low as in the formal sector. In contrast, in the lower-paid informal sector returns to experience are only two thirds of the returns to experience in the formal and higher-paid informal sector and there are no returns to education at all. Workers in this sector are hence stuck with very low wages almost independent of their abilities.[7] Eventually, it is important to notice the significance of correlation coefficient ρ, which underlines the necessity of accounting for sample selection bias when estimating slope coefficients in segment-specific wage equations.

Thus, we do not only find that the labor market under study consists of three different segments, but that these segments also have quite distinct patterns of returns to individual characteristics. On the first glance, among the different theories on labor

[7]Furthermore, gender has a significant impact on earnings in all parts of the market, but the male-female wage gap is wider in the two informal sectors than in the formal sector. In addition, living in the capital city Abijan has a positive impact on wages in both informal segments and no influence on formal earnings; being a Muslim has only a significant positive impact on wages in the low-paid informal segment.

market composition (as described in the introduction), the labor market structure proposed by Fields (2005) and Maloney (2004) seems to be the closest our empirical estimates. Though, even such obvious diversity in the characteristics of the segments does not automatically mean that the labor market may not fit into either the dualistic or the competitive labor model. Rephrasing Basu (1997, p.151-152), it is beyond doubts that the market may be split into several segments. But if all these segments possess the properties attributable to a competitive market, the whole labor market can be as well treated as competitive. Alternatively, if the detected fragments can be categorized as two groups between which entry barriers exist, the market will be dual. Therefore, to attribute the correct properties to the above described parts of the market, one has to consider whether the observed distribution of individuals across segments is the result of sector choice (competitive market) or entry-barriers into sectors (segmented market).

3.3 Entry Barriers or Comparative Advantage?

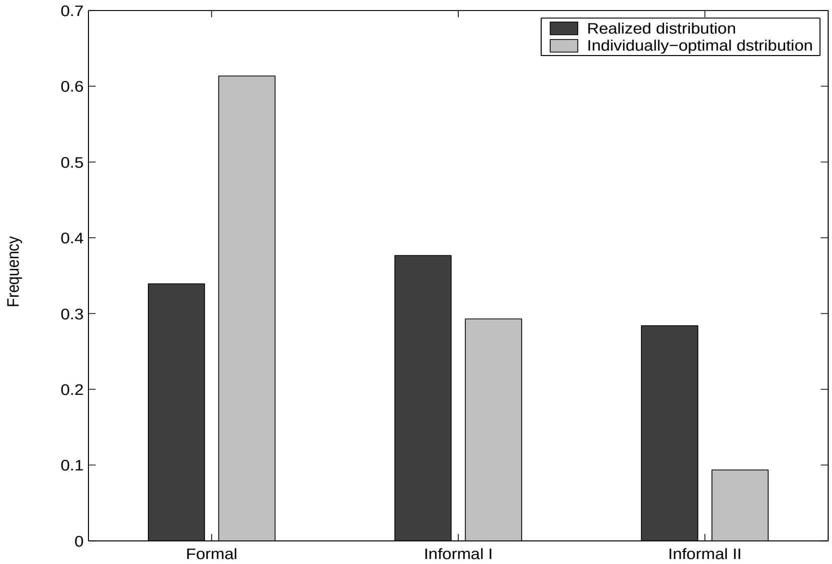
We seek to answer whether employment in the two informal segments is the result of own comparative advantage considerations or a result of entry-barriers into the formal market. The basic argument for the analysis to follow is presented in Section 2.1, p.7. Assuming that agents are earnings maximizers and there is no unobserved components for which we cannot account in our model, the agents will choose the sector in which the expected earnings given their personal characteristics are maximized. This sector choice mechanism induces a probability distribution of agents across sectors formulated in (10), where the sector-specific expected wage for every individual is given by

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

If no barriers of entry to either sector exist, the distribution in (10) must be the same as the mixing distribution $\{\pi_j\}_{j=1}^J$. To the contrary, if there are certain institutional rigidities or statistical discrimination on the employers' side the individuals will be heaped in undesired sectors. As a result there will be a mismatch between the estimated $\hat{\pi}_j$-s and the distribution of individuals that would obtain if individuals were found in the sector where (given their characteristics) they would maximize their earnings.

In Figure 2 we present the estimated by $\{\hat{\pi}_j\}_{j=1}^J$ and implied by (10) probabilities for being affiliated with every sector. Form this figure one can already see that the

Figure 2: Distribution of Agents across Sectors



fraction of those who, conditional on their personal characteristics, expect to be better off in formal sector almost doubles the actual share of the formal sector in the market. On the other hand, the opposite situation can be seen for the "lower-paid"-informal segment (Informal II).

Since the variances of the estimated point mass values π_j are known, the easiest way of setting up the test would be to take the expected frequencies implied by (10) as given and formulate a Wald test of their joint equality to π_j. Even though such test will overreject, the respective test statistic of *895.17* clearly indicates that even with the knowledge of the variances of the implied point mass values we would get a rejection. Estimation of the covariance matrix is complicated by $\mathsf{max}\{\}$-operator in (10), which makes Taylor approximation inapplicable. This is also the reason why we cannot perform the LR test: by virtue of $\mathsf{max}\{\}$-operator the likelihood function under null is not everywhere differentiable and hence the distribution of the likelihood ratio is unknown.

To suggest an additional alternative, we bootstrap the test. In Table 3 we report the bootstrap confidence intervals for the estimated and implied probability mass values ($\hat{\pi}$ and $\tilde{\pi}$, respectively) and for their ratio. The hypothesis of the equality of the two distributions is rejected when the ratio of these values significantly departs from unity.

Table 3: Distribution of Agents 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]

We find that for the formal sector and the "lower"-informal sector significant departure from unity is indeed the case. So the hypothesis of unlimited intersectoral mobility and, consequently, competitiveness, is once again rejected.

To sum up: The amount of workers that would chose to enter the formal sector is significantly higher than the amount of workers actually employed in the formal sector. At the same time the amount of workers in the "lower"-tier of the informal sector is almost three times as high as the amount of workers that would voluntarily choose staying in this segment. Finally, the number of individuals affiliated with the "upper" -tier of the informal sector is the same as the number of those who would chose to be in this sector. Since the workers are free to move between any segments of the informal market, the three statements above imply that there exists an entry barrier between formal and "lower"-tier informal sectors.

This result establishes empirical relevance of the dichotomous structure of the informal market, as suggested by Fields (2005) and Maloney (2004). For the theoretical modelling of the labor market in a developing economy this means that there may exist cases in which neither solely competitive theories, nor exclusively dual frameworks will provide satisfactory approximation of market interactions. For the empirical literature our results are even more important, as we find that testing for competitiveness in the context of the developing economy can be misspecified by either ignoring the employment decision (i.e. selection bias) or, which is more alerting, ignoring the heterogeneity of the informal sector. Empirical contribution of our model, as well as its shortcomings are briefly discussed in the next section.

3.4 Empirical Models for Dual and Competitive Markets

An acknowledged benchmark in the empirical literature on testing duality versus competitiveness is a paper of Dickens and Lang (1985), who were the first to account for unobservability of sector affiliation by implementing 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) state such potential sources of misspecification as:

(i) sector multiplicity in the market,

(ii) the fact that agents are utility maximizers rather than earnings maximizers,

(iii) inability to separate mobility costs across sectors from entry barriers,

(iv) false distributional assumptions.

All the papers that followed, have dealt only with selected number of points. Heckman and Sedlacek (1985) explicitly introduce non-wage valuation of the sector and thereby tackle (ii); Magnac (1990) considers cost of entry and resolves (iii).

In this paper we consistently discuss (i), developing a model that allows both for sample selection and 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 to only two sectors with observed sector membership, out of which homogeneity of the informal sector follows.[8]

In addition to that, we find significance of the that sample selection bias induced by employment decision. This means that the studies that consider a latent structure of the labor market (e.g. Dickens and Lang, 1984, and Cunningham and Maloney, 2001) but ignore sample selection may potentially suffer from this type of misspecification.

Concerning (ii), with exception of Heckman and Sedlacek (1985), all existing models are not robust to distributional assumptions. One possible advantage of our framework

[8] Although the framework of Magnac (1990) has definitely a great advantage in modelling entry costs and richer specification of nonparticipation.

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.

From this perspective the framework developed in the present paper certainly fills some of the gaps in the empirical literature on informal sector heterogeneity and labor market segmentation.

To relative disadvantages of our model one can add the "ex-post" nature of the sector choice, once the employment decision is made. Our model does not make any statement about the exact mechanism of the self-selection, whereas even in the simplest Roy model without employment decision this mechanism is modelled explicitly (see Borjas, 1987). We also need to admit that, unlike in Heckman and Sedlacek (1985), our model in its present formulation does not consider agents as utility-maximizers to comply with (iii). Extension to utility-maximizing agents invokes a more complicated identifiability problem and is reserved for future work.

4 Summary and Conclusions

In this paper we formulate an econometric model that accounts for sample selection and sector multiplicity when sector affiliation of any particular observation is not necessarily observable. We apply this model to learn about the composition of the urban labor market in Côte d'Ivoire.

First, our results support the hypothesis that the informal labor market has a dichotomous structure with distinct wage equations and therefore should not be regarded as one homogenous sector. Moreover, we show that one part of the informal sector is superior over the other in terms of significantly higher earnings as well as higher returns to education and experience.

Next we test whether the detected latent 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, where individuals given their specific characteristics voluntarily enter the informal sector. The outcome we get points at the existence of entry barrier to the formal sector for the "lower"-tier informal sector, 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.

From a policy point of view, it is important to take into account the latent structure of the informal labor market, because recommendations for the two distinct informal sectors are clearly different. Individuals who voluntary participate in the informal sector just realize an opportunity to earn more than they would in the formal sector. But as they still have much lower earnings than employees in the formal sector, policies have to address their individual endowments to improve earning possibilities.

With regard to the "lower"-tier informal sector, policy interventions have to counter entry barriers to the formal sector. Moreover, agents found in the "involuntary" part of informal market show especially low earnings which are also much lower than earnings in the "voluntary" informal part. So if the policy objective is to address the most disadvantaged, the "lower"-tier informal sector should receive the highest priority.

References

[1] Andrews, D., "Chi-Square Diagnostic Tests for Econometric Models: Theory", *Econometrica*, 1988, p.1419-1453.

[2] Basu, A., "Analytical Development Economics", (Cambridge, MA: MIT Press, 1997).

[3] Borjas, G., "Self-Selection and the Earnings of Immigrants", *American Economic Review*, 1987, p.531-553.

[4] Cunningham, W. and W.F. Maloney, "Heterogeneity in the Mexican Microentreprise Sector: An Application of Factor and Cluster Analysis ", *Economic Development and Cultural Change*, 2001, p.131-156.

[5] Dickens, W.T. and K. Lang, "A Test of Dual Labor Market Theory", *American Economic Review*, 1985, p.792-805.

[6] Fields, G., S., "Labour Market Modeling and the Urban Informal Sector: Theory and Evidence", in Thurnham, D., Salome, B., Schwarz, A., eds, *The Informal Sector Revisited*, (OECD, 1990).

[7] Fields, G., S., "A Guide to Multisector Labor Market Models", Paper prepared for the World Bank Labor Market Conference, 2005.

[8] 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.

[9] Heckman, J., “Sample Selection Bias as a Specification Error”, *Econometrica*, 1979, p.153-161.

[10] Heckman, J. and V. Hotz “An Investigation of the Labor Market Earnings of Panamian Males”, *Journal of Human Resources*, 1986, p.507-542.

[11] 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.

[12] Johnsosn, N., Kotz, S., and A. Kemp, “Univariate Distributions”. (Wiley, 2nd edition, 1992).

[13] Lee, L.-F., “Generalized Econometric models with Selectivity”, *Econometrica*, 1983, p.507-512.

[14] Magnac, T., “Segmented or Competitive Labor Markets ”, *Econometrica*, 1991, p.165-187.

[15] Maloney, W.,F., “Informality Revisited ”, *World Development*, 2004, p.1159-1178.

[16] Murphy, K., and R. Topel, “Estimation and Inference in Two Step Econometric Models”, *Journal of Business and Economics Statistics*, 1985, p.370-379.

[17] 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).

[18] Stiglitz, J., E., “The Efficiency Wage Hypothesis, surplus Labor, and the Distribution of Labour in LDCs ”, *Oxford Economic Papers*, 1976, p.185-207.

[19] Teicher, H., “Identifiability of Finite Mixtures”, *Annals of Mathematical Statistics*, 1963, p.1265-1269.

Appendix

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 φ and Φ are the probability density and distribution functions of the Standard Normal distribution.

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

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

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

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

For convenience of the argument to follow use integration by parts to rewrite ϕ_j as

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

$$= \frac{e^{\frac{1}{2}t^2\sigma_j^2 - t\mu_j}}{\Phi(a)} \left[\Phi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \Phi\left(z+t\sigma_j\right)\Big|_{-\infty}^{+\infty} - \frac{\rho_j}{\sqrt{1-\rho_j^2}} \int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \Phi\left(z+t\sigma_j\right) dz \right]$$

$$\overset{\rho_j \neq 0}{=} \frac{e^{\frac{1}{2}t^2\sigma_j^2 - t\mu_j}}{\Phi(a)} \left[1 - \frac{\rho_j}{\sqrt{1-\rho_j^2}} \int_{-\infty}^{+\infty} \varphi\left(\frac{a+\rho_j z}{\sqrt{1-\rho_j^2}}\right) \Phi\left(z+t\sigma_j\right) dz \right]$$

(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, $S_j \subseteq S_l$, which fulfills the first requirement of Theorem 2 of Teicher (1963).

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

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

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

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

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

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

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

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

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

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

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

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 $t_* \to +\infty$.

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 $h_j(y; \mu_j, \sigma_j, \rho) \prec h_j(y; \mu_l, \sigma_l, \rho)$ 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

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

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

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

Remark From the Proof above 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 down as

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

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 $\sigma_l = \sigma_j$.

Estimation Results

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

Formal	Coeff.	(Std.Error)	**Informal**	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
$Age^2/100$ *	−0.1184	0.0258	$Age^2/100$ *	−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
σ_F *	0.8288	0.0192	σ_I *	1.0261	0.0174
ρ^*	0.0953	0.0467			
π_F^* :	0.3392	0.0092	π_I^* :	0.6608	0.0092
Expected log-Wage:	11.3524		Expected log-Wage:	10.3183	
Expected Wage:	105084.42		Expected Wage:	33816.37	

Selection Equation		
Intercept	−0.0422	0.0400
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

Number of Obs. (missing):	2939
Number of Obs. (mixture):	2653
Log-Likelihood:	−5332.92

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

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

Formal	Coeff.	(Std.Error)	Informal 1	Coeff.	(Std.Error)	Informal 2	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
$Age^2/100$ *	−0.1187	0.0258	$Age^2/100$ *	−0.1285	0.0221	$Age^2/100$ *	−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
σ_F *	0.8294	0.0192	$\sigma_{I.1}$ *	0.6556	0.0388	$\sigma_{I.2}$ *	1.2960	0.0574
ρ *	0.1058	0.0497						
π^*_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	

Selection Equation	Coeff.	(Std.Error)		
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		