Competitive and Segmented Informal Labor Markets

Isabel Günther
University of Göttingen

Andrey Launov
University of Würzburg and IZA, Bonn

This version: January 31, 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. Although several authors have stressed this hypothesis, it has so far not received satisfactory empirical treatment. Hence, we formulate a model that allows to test econometrically the hypothesis that the informal labor market comprises two distinct segments. Our model integrates Heckman regression with sample selection into a finite mixture setting, which takes into account both selectivity bias and the fact that sector affiliation of an individual in the informal sector is generally unobservable. 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. Empirical tests for segmentation have often relied on the comparison of earnings and earning equations for the formal and informal sector estimated with ordinary least squares techniques.

However, these tests of labor market segmentation have been frequently criticized by several authors (e.g. Dickens and Lang, 1985; Heckman and Hotz, 1986). Whereas the empirically shown differences in earnings and earning equations for the formal and informal sector have not been questioned, it has been claimed that the mere existence of lower wages as well as lower returns to education and experience in the informal sector does not imply labor market segmentation.

From an econometric point of view, estimations applying OLS might strongly be biased by (sector) selectivity bias as the active population is not randomly drawn from the population as a whole (Heckman, 1979) and as individuals found in a certain sectors are not randomly drawn from the active population as a whole (Gindling, 1991; Heckman and Hotz, 1986; Magnac, 1991).

From an economic point of view, 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 their 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. Several empirical studies have tried to test these opposing views (see e.g. Gindling, 1991; Magnac, 1991). However, these analysis can often not provide robust results, i.e. they cannot rule out any of the two hypothesis.

Most recent theory has combined these polar views and emphasized the heterogeneity or internal duality of the informal sector, with an “upper-tier” and “lower-tier” informal sector (Fields, 2004) or a “voluntary entry” and “involuntary entry” informal sector (Maloney, 2004). The “upper-tier” might then represent the “competitive” informal sector into which individuals enter voluntarily because, given their specific characteristics, they expect to earn more than they would in the formal sector. And the “lower-tier” informal sector would then represent the “segmented” informal sector into which workers are pushed because of entry-barriers into the formal sector (and possibly also the “upper-tier” informal sector.

This theory is quite appealing as it first of all could explain the inconclusive outcomes of studies which have tried to evaluate whether formal-informal labor markets in developing countries are segmented or competitive. Second, from own experience in developing countries, we intuitively know of such a duality within the informal labor market; i.e. we would expect that the “informal” taxi driver earns more than he would in the formal sector, whereas we would expect that it is not the particular choice of the street vendor to earn his income from the sale of tissues.

Despite the above described theory of the structure of the informal sector, the hypothesis of a duality within the informal sector has so far not received satisfactory empirical treatment. The difficulty in testing such an hypothesis is that the affiliation of any single individual in the informal sector to either part of the informal sector is unobserved and an arbitrary division of the informal labor market based on observed individual characteristics or earnings would lead to biased estimates (Dickens and Lang, 1985; Heckman and Hotz 1986). On top of that, to get reliable estimates of expected
earnings in both the formal sector as well as in any unobserved section of the informal labor market the selection bias that arises due to the non-random active population (see Heckman, 1979) and due to the non-random population in each sector (e.g. Gindling, 1991; Heckman and Hotz, 1986; Magnac, 1991) should be taken into account.

Cunningham and Maloney (2001) is to our knowledge the only empirical study that has so far tested for informal sector heterogeneity. Applying cluster and factor analysis to firm life, education, experience, capital intensity and earnings of Mexican micro-firms, their study represents the informal sector as a mixture of “upper-tier” and “lower-tier” entreprises. In the formulation of Cunningham and Maloney (2001) one can therefore not draw any conclusion on individual employment nor make any statement about voluntary or involuntary stay in either part of the informal sector, as the formal sector is excluded of the study and only entreprises are analyzed. As a result the detected heterogeneity of the informal market is rather a statistical outcome.

Hence, we formulate a statistical model to study the structure and determinants of informal sector employment. We analyze whether the hypothesis of a dichotomous informal sector can be supported by the data, and, in case yes, analyze the determinants of earnings in the two segments of the informal labor market. Furthermore, we try to address the question if part of informal sector employment is indeed the result of comparative advantage considerations whereas the other part is the result of entry-barriers into the formal (and eventually also the competitive informal) labor market.

The paper is structured as follows. In section 2 we derive the econometric model to test the hypothesis of an heterogenous informal labor market with two distinct segments, where part of informal employment is competitive to formal sector employment and part of informal employment is the result of market segmentation. Section 3 presents the data, the estimation results and relates our econometric model to other empirical literature. Section 4 summarizes and concludes.

2 Econometric Model

In the model below we first integrate Heckman (1979) regression with sample selection into a finite mixture framework. By doing so we cannot only test for informal sector heterogeneity, where sector affiliation of an individual, i.e. whether an individual belongs to a “lower-tier” or “upper-tier” informal sector, is generally unobservable. But
we also appropriately control for selection bias that could bias estimated coefficients in our wage regressions because of the non-randomness of the active population Heckman (1979). In a second step we formulate a test that allows to analyze whether the two detected sectors of the informal labor market are the result of segmentation or comparative advantage considerations of individuals, i.e. whether they constitute “involuntary entry” or “voluntary entry” informal sector employment (Maloney, 2004).

2.1 Specification

Finite Mixture We presume that the labor market consists of different segments with distinct independent distribution of earnings in each of them. In other words, we assume that the earnings sample consists of $J$ disjoint segments $Y_j : Y = \bigcup_{j=1}^{J} Y_j$, and that the probability that a certain individual observation $y_i$ belongs to $Y_j$ is equal to $\pi_j$. Furthermore we assume that for any segment $Y_j$ the data generation process is distinct and independent of the others. Since it is in general not known which of $J$ processes generates the observed realization $y_i$, the density of individual earnings is given by

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

Thus our basic specification is a conventional mixture model. Next assume that the earning equation for any segment $Y_j$ is

$$y_i = x_i \beta_j + u_i, \quad u_i \sim N(0, \sigma_j^2 | x_i, y_i \in Y_j), \tag{2}$$

where $x_i$ represents a set of individual 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 $\mu_r (y_i) = \sum_{j=1}^{J} \mu_r (y_i | x_i, \theta_j) \pi_j$. Given this result we get the earning regression for the whole active population.

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

Sample Selection The regression in (3) may however not be the ultimate specification of our model. The reason being sample selection bias (Heckman, 1979), which

---

1In our particular case these are one formal and two informal segments.
arises in the observed sample because the wage of any individual is only observed if it exceeds his reservation wage. Or in other words, wages can only be observed if an individual decides to participate in the labor market, i.e. if the offered market wage exceeds his reservation wage, and hence the observed sample on earnings is non-random.

Let the reservation wage of every individual depend on a set of personal characteristics \( z_i \). Then, by writing down the selection equation

\[
y_{is} = z_i \gamma + u_{is}, \quad u_{is} \sim N(0, 1),
\]

in which \( z_i \gamma \) reflects the individual decision to work, we can 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 \( Y_j \)-specific equation (2) and the selection equation (4) follow a bivariate Normal distribution

\[
\begin{bmatrix}
    u_i \\
    u_{is}
\end{bmatrix}
\sim N
\left(
    \begin{bmatrix}
        0 \\
        0
    \end{bmatrix},
    \begin{bmatrix}
        \sigma_j^2 & \rho_j \sigma_j \\
        \rho_j \sigma_j & 1
    \end{bmatrix}
\right)
\bigg| y_i \in Y_j
\]

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

\[
E(y_i | y_{is} > 0) = E(y_i) + \sum_{j=1}^{J} E(u_i | u_{is} > -z_i \gamma, x_i, \theta_j) \pi_j,
\]

where \( E(u_{ij} | u_{is} > -z_i \gamma) \neq 0 \) 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

\[
h(v_i | \theta_j) = \sum_{j=1}^{J} \left[ \frac{\sigma_j^{-1}}{\Phi(z_i \gamma)} \varphi \left( \frac{y_i - x_i \beta_j}{\sigma_j} \right) \Phi \left( \frac{z_i \gamma + \rho_j \sigma_j^{-1} [y_i - x_i \beta_j]}{(1 - \rho_j^2)^{1/2}} \right) \right] \pi_j,
\]

where \( \varphi \) and \( \Phi \) are the standard normal density and distribution functions.\(^2\)

\(^2\)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.
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, because 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 that generate the 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 \( \{Z, \gamma\} \) 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, \( \gamma \) 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 becomes

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

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 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** If the distribution of individuals across sectors is induced by comparative advantage considerations, or in other words if individuals are employed in the
sector where, given their individual characteristics, they maximize their earnings, distribution of individuals across different sectors would be in line with competitive theory, even if estimated returns to individual characteristics are lower in certain sectors. To test whether the observed distribution across sectors reflect individual sector choice or not (and is hence the result of entry barriers into certain sectors) we undertake the following analysis.

Assume that log-earnings are completely specified by \( x_i \beta_j \) (i.e. there exists no unobserved component which we cannot account for). Then competitive theory would imply that the probability of an individual choosing sector \( j \) is equal to

\[
P_i (y_i \in Y_j | x_i) = \prod_{l=1, l \neq j}^J P \left( \ln (y_{il}^j | x_i) > \ln (y_{lj}^l | x_i) \right) \]

\[
= \prod_{l=1, l \neq j}^J P \left( (\beta_j - \beta_l) x_i + (\varepsilon_{il} - \varepsilon_{ij}) > 0 \right). \tag{9}
\]

In the context of only two sectors Dickens and Lang (1985) notice that if there are no entry barriers to the formal sector, the difference in the returns to individual characteristics in the two wage equations must be equal to the corresponding coefficients in the equation that determines the individual-specific probability of sector membership. Heckman and Hotz (1986) suggest the “proportionality test”, where, for a standard regression with sample selection, in absence of barriers coefficients in the participation equation have to be proportional to coefficients in the wage equation.

With regard to our model, despite it is easy to let sector affiliation probabilities \( \pi_j \) in (8) be dependent on individual characteristics, with \( J > 2 \) the parametrization of \( \pi_j \) will be non-linear and neither the equality result of Dickens and Lang (1985) nor the proportionality result of Heckman and Hotz (1986) will carry over. Instead we assume that knowing the returns in all sectors, an individual will choose the sector where the expected earnings given his specific characteristics are maximized. Thus the probability of an agent choosing sector \( j \) can be written down as

\[
P (y_i \in Y_j | x_i) = P \left( E \left[ \ln (y_{ij}^j | x_i) \right] = \max_{l, l \neq j} \left\{ E \left[ \ln (y_{il}^l | x_i) \right] \right\} \right). \tag{10}
\]

Equation (10) provides us with the expected distribution of individuals over all segments of the labor market if free sector choice existed, i.e. where the distribution of individuals across sectors were exclusively determined by returns to individual characteristics. At the same time, the estimated distribution of agents across sectors in (8) is
given by $\{\pi_j\}_{j=1}^J$. This fact provides grounds for the test of free entry into the desired sector. If $\{\pi_j\}_{j=1}^J$ in (8) and the estimated 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. The analysis of sector choice and further issues connected with the implied above 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 $\gamma$ in (4) running Probit.

2. On the second step use $z_i\hat{\gamma}$ as consistent estimates of $z_i\gamma$ to estimate the mixture model.

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

$$\ln \mathcal{L} = \sum_{i=1}^N \ln \left( \sum_{j=1}^J h_i(\theta_j, \rho|x_i, z_i\hat{\gamma}) \pi_j \right),$$

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 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 $\{Y_F\}$. Then (11) modifies to

$$\ln \mathcal{L} = \sum_{i \in \{Y_F\}} \ln h_i(\theta_F, \rho|x_i, z_i\hat{\gamma}) - N_F \ln \pi_F$$

$$\quad + \sum_{i \notin \{Y_F\}} \left[ \ln \left( \sum_{j=1}^J h_i(\theta_{I,j}, \rho|x_i, z_i\hat{\gamma}) \pi_{I,j} \right) \right],$$

where $h_i(\theta_j, \rho)$ is given in (8).
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 \( \xi = \{\theta_j\}_{j=1}^J, \rho, \{\pi_j\}_{j=1}^{J-1} \) is given by

\[
V(\xi) = D^{-1}(\xi) + D^{-1}(\xi)M(\xi, \gamma)D^{-1}(\xi),
\]

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

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

\[
\ln L = \sum_{i \in \{Y\}} \ln [\ell_i(\xi, \gamma| y_i, x_i, w_i, z_i)\Phi(z_i^\gamma)] + \sum_{i \in \{Y^c\}} \ln (1 - \Phi(z_i^\gamma)),
\]

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

\section{Empirical Application}

\subsection{Data and Estimation Method}

The data we use is drawn from the Ivorian household survey, the \textit{Enquete de Niveau de Vie}, of 1998 which was undertaken by the \textit{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

\(^3\)For exact form of \( M(\xi, \gamma) \) see Murphy and Topel (1985).
Table 1: Descriptive Statistics of the Labor Market

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Inactive</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Informal</td>
</tr>
<tr>
<td>Sample</td>
<td>100%</td>
<td>52.6%</td>
<td>31.3%</td>
</tr>
<tr>
<td>Monthly Wage</td>
<td>98,815.0</td>
<td>–</td>
<td>64,837.8</td>
</tr>
<tr>
<td>Males</td>
<td>49.7%</td>
<td>40.6%</td>
<td>49.0%</td>
</tr>
<tr>
<td>Age</td>
<td>30.0</td>
<td>25.2</td>
<td>34.7</td>
</tr>
<tr>
<td>Education (years)</td>
<td>5.3</td>
<td>5.8</td>
<td>2.9</td>
</tr>
<tr>
<td>Literacy rate</td>
<td>64.1%</td>
<td>69.8%</td>
<td>44.4%</td>
</tr>
<tr>
<td>Training after schooling</td>
<td>17.6%</td>
<td>11.1%</td>
<td>14.7%</td>
</tr>
<tr>
<td>Religion:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– muslim</td>
<td>43.4%</td>
<td>38.3%</td>
<td>56.8%</td>
</tr>
<tr>
<td>– christian</td>
<td>42.2%</td>
<td>46.2%</td>
<td>30.6%</td>
</tr>
<tr>
<td>– other</td>
<td>14.4%</td>
<td>15.5%</td>
<td>12.6%</td>
</tr>
<tr>
<td>Living in Abijan</td>
<td>49.6%</td>
<td>50.4%</td>
<td>42.2%</td>
</tr>
</tbody>
</table>

*Note: Monthly Wage in CFA Francs.*

in 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
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

---

4For the whole sample, the average length of education among males is more than 60% higher than among females.
any of them may not necessarily be observable. Second, and even more 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.5

The results on model selection are presented in Table 2. First of all, the values of the Andrews $\chi^2$ 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.

Table 2: Model Selection

<table>
<thead>
<tr>
<th></th>
<th>Homogeneous Informal Market</th>
<th>Two-Segment Informal Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>10689.85</td>
<td>10580.23</td>
</tr>
<tr>
<td>CAIC</td>
<td>10879.05</td>
<td>10864.03</td>
</tr>
<tr>
<td>SBC</td>
<td>10855.05</td>
<td>10828.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Cr. Value</th>
<th>Test Statistic</th>
<th>Cr. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrews’ $\chi^2$-Test</td>
<td>155.26</td>
<td>51.00</td>
<td>143.35</td>
</tr>
</tbody>
</table>

5Anders (1988) shows that if $P(\Gamma)$ is the empirical measure and $F(\Gamma, \theta)$ is the conditional empirical measure defined on a partition $\mathbf{Y} \times \mathbf{X} = \bigcup \gamma_i$, and $v(\Gamma, \theta) \equiv \sqrt{n} (P(\Gamma) - F(\Gamma, \theta))$, then: $v(\Gamma, \theta)' \Sigma^2 v(\Gamma, \theta) \sim \chi^2_{n k | \Sigma^1}$, where $\Sigma$ is the covariance matrix of $v(\Gamma, \theta)$. Three different estimators of $\Sigma$ 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 $\mathbf{Y}$.
Thus the labor market under study consists of at least three distinct parts: the formal sector and two latent segments of the informal sector.

Even though cell frequencies generated by the “better” 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.

In a next step, we 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. Eventually, it is important to notice the significance of correlation coefficient $\rho$, which underlines the necessity of accounting for sample selection bias when estimating slope coefficients in segment-specific wage equations.6

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

---

6Furthermore, 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.
to individual characteristics. Hence, among the different theories on labor market composition (as described in the introduction), the labor market structure proposed by Fields (2005) and Maloney (2004) seems to be supported most by our empirical estimates.

However, 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 fragmented 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.8. 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 (9), where the sector-specific expected wage for every individual is given by

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

If no barriers of entry to either sector exist, the distribution in (9) must be the same as the mixing distribution \( \{ \pi_j \}^f_{j=1} \). To the contrary, if there are certain institutional rigidities or statistical discrimination on the employers’ side the individuals will be heaped in undesired sectors and hence there will be a mismatch of the estimated distribution in (8) and the distribution of individuals across sectors that would occur.
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 (9) probabilities for being affiliated with every sector. From this figure one can already see that the 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 $\pi_j$ are known, the easiest way of setting up the test would be to take the expected frequencies implied by (9) as given and formulate a Wald test of their joint equality to $\pi_j$. 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 $\max\{}$-operator in (9), which makes Taylor approximation inapplicable. This is also the reason why we cannot perform the LR test: by virtue of $\max\{}$-operator the likelihood function under null is not everywhere differentiable and hence the distribution of the likelihood ratio is unknown.
Table 3: Distribution of Agents across Sectors

<table>
<thead>
<tr>
<th></th>
<th>Formal</th>
<th>Informal-1</th>
<th>Informal-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
</tr>
<tr>
<td>$\hat{\pi}_j$</td>
<td>0.3392</td>
<td>0.3767</td>
<td>0.2840</td>
</tr>
<tr>
<td>$\tilde{\pi}_j$</td>
<td>0.6136</td>
<td>0.2929</td>
<td>0.0935</td>
</tr>
<tr>
<td>$\hat{\pi}_j/\tilde{\pi}_j$</td>
<td>0.5528</td>
<td>1.2863</td>
<td>3.0385</td>
</tr>
</tbody>
</table>

To suggest one more alternative, we bootstrap the test. In Table 3 we report the bootstrap confidence intervals for the estimated and implied mass points ($\hat{\pi}$ and $\tilde{\pi}$, respectively) and for their ratio. The hypothesis of the equality of the two distributions is rejected when the ratio of the mass points significantly departs from unity. 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. Taken together with the fact that 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, implies 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 (or
duality) 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 (with more than two segments) in the market,
(ii) false distributional assumptions,
(iii) the fact that agents are utility maximizers rather than earnings maximizers.

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 (including Gindling, 1991; Heckman and Sedlacek, 1985, and Magnac, 1990), as these latter models are confined to only two sectors with observed sector membership and supposed homogeneity of the informal sector.

On the other hand, studies that have assumed a latent structure of the labor market (Dickens and Lang, 1984; Cunningham and Maloney, 2001) have ignored the issue of sample selection bias induced by the labor market participation decision of individuals. Hence the, in this paper detected, significance of sample selection bias provides evidence of possible misspecification of these models.

Concerning (ii), 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 this view the developed framework in the present paper without doubts fills some of
the gaps in the empirical literature on informal sector heterogeneity as well as on labor market segmentation.

The relative disadvantages of our model is that we only analyze “ex-post” the self-selection of individuals into a specific sector, since in the generalized Roy models sector choice parameters are introduced explicitly as in Gindling (1991) or Magnac (1990). Also we 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). These extensions are 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. Thus, the model is an integration of Heckman regression with sample selection into a finite mixture setting. We apply this model to learn about the composition of the urban labor market in Côte d’Ivoire.

First of all, our estimation 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 characteristic actually chose to be in 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 their earning possibilities.

With regard to the “lower”-tier informal sector, policy interventions have to reduce 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.

Our prospects for future research include generalization of the model to utility-maximizing agents and the treatment of sectoral selection bias. Moreover, it would be very interesting to compare market structures across countries, since quite a number of considerations about differences in the composition of the informal sector between Africa and Latin America exists (see Fields, 2005).

References


Appendix

Component Density of the Error Term

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

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

Since joint distribution of $(u_i, u_{is})$ is bivariate normal, conditional density $f(u_{is} > -z_i\gamma|u_i, \theta_j)$ 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

$$f(u_i|u_{is} > -z_i\gamma, \theta_j) = P\left(\frac{u_{is} - z_i\gamma - \rho_j\sigma_j^{-1} u_i}{\sqrt{1 - \rho_j^2}} > \frac{-z_i\gamma - \rho_j\sigma_j^{-1} u_i}{\sqrt{1 - \rho_j^2}}\right) \frac{f(u_i|\theta_j)}{P(u_{is} > -z_i\gamma)}$$

$$= \Phi\left(\frac{z_i\gamma + \rho_j\sigma_j^{-1} (y_i - x_i\beta_j)}{\sqrt{1 - \rho_j^2}}\right) \frac{1}{\sigma_j} \frac{1}{\Phi(\gamma)}$$

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)

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

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

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

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

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

$$= \frac{e^{\frac{t^2\sigma_j^2 - t\mu_j}}{\Phi(a)} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \Phi\left(\frac{a + \rho_j z}{\sqrt{1 - \rho_j^2}}\right) dz.}$$
For convenience of the argument to follow use integration by parts to rewrite \( \phi_j \) as

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

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

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

(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 \( t_* \in S_j \).

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

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

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

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

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

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

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

23
= \varphi (a + t\sigma_j \rho_j) \int_{-\infty}^{+\infty} \varphi \left( \frac{z + \left[a \rho_j + t\sigma_j (1 - \rho_j^2)\right]}{\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

$$
\lim_{t \to +\infty} \frac{\hat{\phi}_l(t)}{\phi_j(t)} = \lim_{t \to +\infty} e^{\frac{1}{2} t^2 (\sigma_l^2 - \sigma_j^2) - t(\mu_l - \mu_j)} \lim_{t \to +\infty} \frac{\varphi (a - t\sigma_j \rho_l)}{\phi_j(t)} \left[ \frac{\rho_l \sigma_l}{\rho_j \sigma_j} \right] 
$$

$$
= \lim_{t \to +\infty} e^{\frac{1}{2} t^2 (\sigma_l^2 - \sigma_j^2) - t(\mu_l - \mu_j)}, \quad \lim_{t \to +\infty} e^{-\frac{1}{2} t^2 (\sigma_l^2 \rho_l^2 - \sigma_j^2 \rho_j^2) + ta(\sigma_l - \sigma_j \rho_l)} \left[ \frac{\rho_l \sigma_l}{\rho_j \sigma_j} \right].
$$

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 $t_s \to +\infty$.

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 $h_j(y; \mu_j, \sigma_j, \rho) < h_j(y; \mu_l, \sigma_l, \rho)$ if $\sigma_l < \sigma_j$ and $\mu_l > \mu_j$ when $\sigma_l = \sigma_j$. Then for $t_s = +\infty$, $t_s \in S_j$ we get

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

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 $\rho$ is identifiable.

**Remark** From the Proof above immediately follows that allowing for a sector-specific selection rule (i.e. letting $a$ be $a_j = \zeta_j$) leads to an unidentifiable model, since the limit of ratio writes down as

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

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 $\sigma_l = \sigma_j$. 

24
Estimation Results

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

<table>
<thead>
<tr>
<th>Formal</th>
<th>Coef.</th>
<th>(Std.Error)</th>
<th>Informal</th>
<th>Coef.</th>
<th>(Std.Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept *</td>
<td>7.0595</td>
<td>0.3797</td>
<td>Intercept *</td>
<td>7.5028</td>
<td>0.2378</td>
</tr>
<tr>
<td>Sex *</td>
<td>0.3443</td>
<td>0.0732</td>
<td>Sex *</td>
<td>0.5734</td>
<td>0.0538</td>
</tr>
<tr>
<td>Age *</td>
<td>0.1300</td>
<td>0.0196</td>
<td>Age *</td>
<td>0.1062</td>
<td>0.0127</td>
</tr>
<tr>
<td>Age²/100 *</td>
<td>-0.1184</td>
<td>0.0258</td>
<td>Age²/100 *</td>
<td>-0.1215</td>
<td>0.0165</td>
</tr>
<tr>
<td>Education *</td>
<td>0.1058</td>
<td>0.0091</td>
<td>Education *</td>
<td>0.0421</td>
<td>0.0105</td>
</tr>
<tr>
<td>Literacy</td>
<td>-0.1420</td>
<td>0.1140</td>
<td>Literacy</td>
<td>-0.0466</td>
<td>0.0844</td>
</tr>
<tr>
<td>Training *</td>
<td>0.1598</td>
<td>0.0626</td>
<td>Training *</td>
<td>0.2006</td>
<td>0.0802</td>
</tr>
<tr>
<td>Muslim</td>
<td>0.1542</td>
<td>0.0896</td>
<td>Muslim *</td>
<td>0.2580</td>
<td>0.0781</td>
</tr>
<tr>
<td>Christian</td>
<td>-0.0185</td>
<td>0.0849</td>
<td>Christian</td>
<td>0.1225</td>
<td>0.0831</td>
</tr>
<tr>
<td>Abijan</td>
<td>0.0809</td>
<td>0.0576</td>
<td>Abijan *</td>
<td>0.2273</td>
<td>0.0506</td>
</tr>
<tr>
<td>σF *</td>
<td>0.8288</td>
<td>0.0192</td>
<td>σI *</td>
<td>1.0261</td>
<td>0.0174</td>
</tr>
<tr>
<td>ρ *</td>
<td>0.0953</td>
<td>0.0467</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \pi_F^* : 0.3392 \quad 0.0092 \quad \pi_I^* : 0.6608 \quad 0.0092 \]

Expected log-Wage: 11.3524
Expected Wage: 105084.42
Expected log-Wage: 10.3183
Expected Wage: 33816.37

Selection Equation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>(Std.Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.0422</td>
<td>0.0400</td>
</tr>
<tr>
<td>Sex *</td>
<td>0.5682</td>
<td>0.0374</td>
</tr>
<tr>
<td>Infants *</td>
<td>0.2705</td>
<td>0.0196</td>
</tr>
<tr>
<td>Children *</td>
<td>0.2677</td>
<td>0.0162</td>
</tr>
<tr>
<td>Old</td>
<td>-0.0518</td>
<td>0.0439</td>
</tr>
<tr>
<td>HH Size *</td>
<td>-0.2693</td>
<td>0.0092</td>
</tr>
<tr>
<td>Active Members *</td>
<td>0.4709</td>
<td>0.0157</td>
</tr>
</tbody>
</table>

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”

<table>
<thead>
<tr>
<th>Formal</th>
<th>Coeff. (Std.Error)</th>
<th>Informal 1</th>
<th>Coeff. (Std.Error)</th>
<th>Informal 2</th>
<th>Coeff. (Std.Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept *</td>
<td>7.0516 (0.3799)</td>
<td>Intercept *</td>
<td>7.5818 (0.3225)</td>
<td>Intercept *</td>
<td>7.4643 (0.5803)</td>
</tr>
<tr>
<td>Sex *</td>
<td>0.3476 (0.0734)</td>
<td>Sex *</td>
<td>0.6659 (0.0700)</td>
<td>Sex *</td>
<td>0.4417 (0.1257)</td>
</tr>
<tr>
<td>Age *</td>
<td>0.1301 (0.0196)</td>
<td>Age *</td>
<td>0.1199 (0.0169)</td>
<td>Age *</td>
<td>0.0816 (0.0307)</td>
</tr>
<tr>
<td>Age^2/100 *</td>
<td>-0.1187 (0.0258)</td>
<td>Age^2/100 *</td>
<td>-0.1285 (0.0221)</td>
<td>Age^2/100 *</td>
<td>-0.1012 (0.0397)</td>
</tr>
<tr>
<td>Education *</td>
<td>0.1058 (0.0091)</td>
<td>Education *</td>
<td>0.0577 (0.0160)</td>
<td>Education</td>
<td>0.0210 (0.0261)</td>
</tr>
<tr>
<td>Literacy</td>
<td>-0.1420 (0.1140)</td>
<td>Literacy</td>
<td>-0.1405 (0.1103)</td>
<td>Literacy</td>
<td>0.0706 (0.1958)</td>
</tr>
<tr>
<td>Training *</td>
<td>0.1600 (0.0626)</td>
<td>Training</td>
<td>-0.1190 (0.1063)</td>
<td>Training *</td>
<td>0.6664 (0.2031)</td>
</tr>
<tr>
<td>Muslim</td>
<td>0.1550 (0.0896)</td>
<td>Muslim</td>
<td>-0.0923 (0.0979)</td>
<td>Muslim *</td>
<td>0.7532 (0.2103)</td>
</tr>
<tr>
<td>Christian</td>
<td>-0.0185 (0.0850)</td>
<td>Christian</td>
<td>-0.0505 (0.1025)</td>
<td>Christian</td>
<td>0.4026 (0.2150)</td>
</tr>
<tr>
<td>Abijan</td>
<td>0.0807 (0.0576)</td>
<td>Abijan *</td>
<td>0.1871 (0.0683)</td>
<td>Abijan *</td>
<td>0.2530 (0.1225)</td>
</tr>
<tr>
<td>Education *</td>
<td>0.8294 (0.0192)</td>
<td>Education</td>
<td>0.6556 (0.0388)</td>
<td>Education</td>
<td>1.2960 (0.0574)</td>
</tr>
<tr>
<td>( \rho ) *</td>
<td>0.1058 (0.0497)</td>
<td>( \rho ) *</td>
<td>0.1058 (0.0497)</td>
<td>( \rho ) *</td>
<td>0.1058 (0.0497)</td>
</tr>
</tbody>
</table>

\( \pi_{F}^{*} : 0.3392 (0.0092) \)  \( \pi_{I.1}^{*} : 0.3767 (0.0403) \)  \( \pi_{I.2}^{*} : 0.2840 (0.0401) \)

Expected Wage: 105095.04  Expected Wage: 40992.12  Expected Wage: 28054.92

Selection Equation

<table>
<thead>
<tr>
<th></th>
<th>Coeff. (Std.Error)</th>
<th>Number of Obs. (cens):</th>
<th>2939</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.0422 (0.0400)</td>
<td>Number of Obs. (mix):</td>
<td>2653</td>
</tr>
<tr>
<td>Sex *</td>
<td>0.5682 (0.0374)</td>
<td>Log-Likelihood:</td>
<td>-5272.11</td>
</tr>
<tr>
<td>Infants *</td>
<td>0.2705 (0.0196)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children *</td>
<td>0.2677 (0.0142)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>-0.0518 (0.0439)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH Size *</td>
<td>-0.2693 (0.0092)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active Members *</td>
<td>0.4709 (0.0157)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>