FEMALE ACTIVITY CHOICE IN A DUAL CONTEXT: AN INTEGRATED MODEL FOR FORMAL AND INFORMAL SECTORS IN CAMEROON*

Christophe Muller

WP-AD 2003-39

Correspondence to: Christophe Muller, Universidad de Alicante, Departamento de Fundamentos del Análisis Económico, Apartado de Correos, 99. 03080 Alicante (Spain). E-mail: cmuller@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Diciembre 2003
Depósito Legal: V-5481-2003

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

* This research has been funded by the programme PARADI at the Université Laval. I am grateful to G. Lanot for his collaboration on a previous project of which this paper is a continuation. I would like to thank participants in seminars at Journées PARADI and Université Laval in Canada; CREST and DIAL in Paris; University of Oxford, University of Keele and University of Dundee in United Kingdom; and particularly I. Walker, D. Margolis, B. Fortin, T. Lemieux, S. Bond and two referees for useful comments. The responsibility for errors is mine. I am grateful to DIAL-DSCN, which made the data available for me.

FEMALE ACTIVITY CHOICE IN A DUAL CONTEXT: AN INTEGRATED MODEL FOR FORMAL AND INFORMAL SECTORS IN CAMEROON

Christophe Muller

ABSTRACT

The estimation of models of activity choice in LDCs (Less Developed Countries) is complicated by specific features of labour markets. In particular, entry into activity sectors is often restricted and worked hours are rationed. Moreover, in the informal sector workers may obtain their work rewards directly through technology with decreasing returns to labour.

In this situation, complex structural modelling is often reputed infeasible. In this paper, we show that this statement is exaggerated and that it is possible to simultaneously estimate preferences, informal technology and sector entry costs.

We estimate a structural non-separable model of activity choice and labour supply of female workers for formal, informal and inactivity sectors in the capital city of Cameroon. Our results provide evidence for the concavity in hours of the earnings function in the informal sector. We also estimate significant effects for several explanatory variables of the preferences and the earnings function. This allows us to contrast the labour supply elasticities for different sectors in a structural framework.

However, estimation using a small sample is difficult and requires strong restrictions. In particular, there exists a trade-off between incorporating many explanatory variables in the model, and accounting for the non-linearity and the consistency of the structural explanation. These difficulties raise questions about the interpretation of simple models that are not simultaneously based on preferences, technology and entry costs.

JEL codes: J22, J23, O17.

Keywords: Structural Models, Self-Employment, Activity Choice, Formal and Informal Sectors.

1. Introduction

Modelling activity choice in LDCs (Less Developed Countries) is made difficult by specific features of labour markets. Labour markets in developing countries are more often characterised by dualism[1] than by perfect competition. The existence of activities with diminishing return to labour in the traditional sector and entry costs in the modern sector is a specific feature of this dualism, especially when the informal sector is involved. Moreover, the large gap between real wages in both sectors induces workers in the traditional sector to migrate to the modern sector. In view of this, when modelling the activity choice of workers, one should compare the rewards obtained in different sectors. These rewards result from different technologies or markets, and may involve non-linearities in worked hours. Several attempts at estimating such models have been carried out in the context of labour markets[2].

The aim of this article is to extend the structural modelling of activity choice and labour supply based on comparison of wage rates, to more general cases involving entry costs, rationing of hours and non-linear earning functions[3]. In such a situation, the Roy model (Roy, 1951) and the separation of decisions are no longer valid and an integrated structural model of preferences, technologies and entry costs must be estimated. Our goal is to use new modelling methods to deal with limitations of past approaches. We describe our integrated non-separable model in Section 2. In Section 3, we discuss the data used for the estimation. We derive in Section 4 the likelihood function of the structural model, discuss the explanatory variables and present the estimation results. Our results provide evidence for the concavity in hours of the earnings function in the informal sector. We can also estimate significant effects for several explanatory variables of the preferences and the earnings function. This allows us to contrast the

---

[1] Lewis (1954), Ranis and Fei (1961), Sen (1966), Myint (1985), Behrman (1999).

[2] Benjamin (1992), Jacoby (1993), Lemieux, Fortin and Fréchette (1994), Newman and Gertler (1994).

[3] These three elements may be considered as stemming from market imperfections. However, calling these features "imperfections" is merely a way of collecting together several characteristics differing from the standard labour supply model. In particular, it is debatable whether the lack of flexibility of the hours in the informal sector is evidence of imperfections. Rigid hours schedules in modern industries may result from the constraints of a complex technology, in contrast with informal jobs that are much less dependent on fixed capital. Thus, entry costs, diminishing returns to labour in the informal earnings function and rationing of hours may be associated with market imperfections by some economists and not by others.

labour supply elasticities for different sectors in a structural framework. Finally, we conclude in Section 5.

2. The Model

2.1. The approach

In LDCs, urban workers can choose between varied work and earnings perspectives. The recent empirical literature often distinguishes the formal sector (or primary sector), the informal sector (or secondary sector), and the non-participation[4]. Although our data do not allow the estimation of a full model with four sectors, a fourth important sector would be unemployment, which we regroup here with non-participation. Another reason for this regrouping is that it is unclear how to define unemployment in a poor LDC without a developed social insurance system.

Almost by definition in the segmented market literature, the primary sector is characterised by high wages, high returns to education and to on-the-job training, and a concentrated weekly schedule of hours. In contrast, in the secondary sector, wages are generally low. Moreover, in this sector earnings may be a non-linear function of the worked hours when they directly originate from produced output. The rates of return to education or experience are expected to be low or non-existent in the secondary sector. Moreover, the weekly hours are much more variable than in the primary sector.

We account for these characteristics in our structural model. First, we describe the rationing of the job supply in the primary sector by entry costs and by rationing of hours. To simplify the model we assume that entry costs only exist in the primary sector[5]. This does not imply that every worker in the secondary sector is necessarily rationed since she may have chosen the secondary sector even without the entry costs

---

[4] See Dickens and Lang (1985, 1992), Hill (1988), Chiswick (1988), Magnac (1991), Coate and Tennyson (1992), Vijverberg (1993), Lemieux, Fortin, Fréchette (1994), Tiefenthaler (1994), Hill (1994), Pradhan (1994). The case of a large number of sectors is dealt with in Dahl (2002).

[5] One could also imagine that there are different entry costs in the primary sector for workers coming from unemployment and the secondary sector. We do not treat the rationing process in such detail and we consider that the job rationing corresponding to unemployed workers is exogenous to focus on the links between primary and secondary labour markets.

because her productivity is higher in this sector. Second, we also avoid representing activity choices resulting from comparisons of wage (or shadow wage) linear functions[6]. Finally, the literature generally deals with endogeneity problems with recursivity assumptions or with two-stage estimation methods. In contrast, we specify a non-recursive and non-linear model of activity choice.

Some non-linearity in our model arises from diminishing returns to labour. In most applications, the informal sector wage is measured with the average hourly income. This is incorrect if the earnings function is a non-linear function of the worked hours and if the informal sector is quasi-autarkic in labour, with the tasks for the worker's firm carried out by the worker herself. In this situation, in the informal sector the marginal productivity of labour can be very different from the market wage rates of workers of the same qualification, since these workers are not perfect substitutes of family members in the informal firm. Consequently, for the informal sector, we specify a labour earnings function that is concave in the worked hours because many workers directly get the returns from their production[7]. This is in contrast with most of the literature on the informal sector where an earnings function linear in hours is estimated[8]. To simplify we neglect the intra-family interactions[9]. When the earnings function in the labour-autarkic informal sector is non-linear in hours, or when there are entry costs and hours rationing in the formal sector, wages are no longer the sole information needed to explain sector choices. Moreover, sector choice and labour supply are no longer separable. They are obtained by the comparison of indirect utility levels, specific to each sector.

One simplifying strategy, different from the separability assumption, is to carry out the estimation in two steps to avoid a simultaneous non-linear estimation[10]. Indeed, the FIML of a non-separable non-linear model with several sectors is often deemed 'unfeasible' and to our best knowledge, there has never been a full information estimation of it. In this respect, it is difficult to understand, especially when explicit

[6] Roy (1951), Todaro (1971), Chiswick (1988), Magnac (1991), Jacoby (1993), Pradhan (1994), Dahl (2002).

[7] like in Rosenzweig (1988), Lemieux, Fortin and Fréchette (1994).

[8] Vijverberg (1993), Hill (1994).

[9] like in Newman and Gertler (1994).

[10] e.g., in Kooreman and Kapteyn (1987), Jacoby (1993), Newman and Gertler (1994).

distribution assumptions are made, how the properties of the usual reduced-form estimators can be considered satisfactory when the FIML efficient approach has failed. What we endeavour in this paper is the FIML estimation of a complex model of activity choice. Because such estimation is difficult and to facilitate the maximisation of the likelihood of the model, we focus on the choices of female active members in the household[11], as it is recognised that female labour supply is more flexible than male labour supply[12].

Following our foregoing discussion, we propose a model of activity choice and labour supply with entry cost and hours rationing in the primary sector, and a concave earnings function in the secondary sector. As usual in the applied literature, the entry cost is a stylised way to treat the rationing process in the formal sector. Is it the right way to do it? We choose this approach first because it corresponds to clear intuitions (a fixed bribe to access some advantages, a fixed amount of clothing expenditure to be able to apply for a formal job); second because it is related to standard labour supply models based on fixed costs, and third because it leads to tractable estimation.

The error terms of the model are entered in the preferences, the fixed cost and the concave earnings function of the informal sector. These errors can be correlated because of unobserved heterogeneity and measurement problems jointly affecting these three elements of the model. As a counterpart to the increased sophistication of the model, we keep the number of considered sectors at a minimum and we assume that the wage rate in the primary sector is exogenous because of institutional constraints and to simplify the estimation[13]. We now turn to the optimisation program of the female worker.

2.2. Definition of the optimisation regimes

Three undamental structures determine the optimisation regimes: the preferences of the worker, her earnings function in the secondary sector and the entry cost in the

[11] Female members working simultaneously in two different jobs are rare in our sample (8 percent of workers) and we ignore this element of the activity choice.

[12] Killingsworth (1983).

[13] Coate and Tennyson (1992) propose a theoretical model with labour supply and self-employment where the choice of skill investment is endogenous. However, this model is not accompanied by any empirical application.

primary sector. The preferences of the individual over consumption (c) and worked hours (h) are represented either with a direct utility function (U(c, h)), or with the corresponding indirect utility function (Ψ(w, y)), where w is the wage rate and y is the exogenous income. We assume as usual that the indirect utility function is strictly monotonically increasing in y and strictly monotonically increasing and strictly quasi-convex in w. The Marshallian labour supply function obtained from Ψ(w, y) through Roy's identity is denoted h(w, y).

The labour earnings function in the secondary sector, R(h), is such that R'(h) > 0 and R"(h) < 0, allowing for decreasing return to labour. Consistently with the quasi-autarkic imperfect labour market associated with informal activities, only the labour of the individual enters the subjacent production function.

To simplify, the entry cost in the primary sector is assumed to be fixed. More complex specifications did not lead to a converging estimation. We consider that only the entry cost (f) in the primary sector matters, and we neglect a possible entry cost in the secondary sector. Moreover, consistently with the descriptive statistics shown below, any job in the primary sector is supposed to correspond to fixed hours, $\bar{h}$, and to a given hourly wage rate w.

In these conditions, the sector choice is derived from the comparison of the sector-dependent indirect utilities, as in Hausman (1985). Thus, the activity and labour decisions of workers have a similar structure to those of non-linear discrete-continuous choice models (Hanemann, 1984). The observed worker position is summarised in (s^{obs}, h^{obs}, R^{obs}) where s^{obs} is the observed sector of activity (1 = primary sector, 2 = secondary sector, 3 = non-participation and unemployment), h^{obs} is the observed labour supply, and R^{obs} is the observed non-wage activity income. (s^{obs}, h^{obs}, R^{obs}) is related to the latent model in the following paragraph where the unobserved variables are replaced by a zero.

The worker works in the primary sector if her utility level in the primary sector is above her utility level in the other two sectors, i.e. $U(c,^{-}, h,^{-}) \geq \Psi(w^v(h^*), \mu^v(h^*) + y_0)$ and $U(c,^{-}, h,^{-}) \geq U(y_0, 0)$, where $w^v(h) = R'(h)$ and $\mu^v(h) = R(h) - h\, w^v(h)$, are respectively the virtual wage and the net virtual activity income associated with the shadow linear budget constraint in the secondary sector for a level h of worked hours. y_0 is the level of unearned income, w is the observed wage in the primary sector and f is the fixed entry cost in the primary sector. $\bar{h}$ is the level of rationed hours and

$\bar{c} = w\,\bar{h} + y_0 - f$ is the consumption expenditure in the primary sector. The optimal labour supply in the secondary sector obtained from Roy's identity is $h^* = h(w^v(h^*), \mu^v(h^*) + y_0)$ and $R^* = R(h^*)$. In that case, $(s^{obs}, h^{obs}, R^{obs}) = (1, \bar{h}, 0)$. The worker works in the secondary sector if $\Psi\,(w^v(h^*), \mu^v(h^*) + y_0) \geq U(\bar{c}, \bar{h})$ and $\Psi\,(w^v(h^*), \mu^v(h^*) + y_0) \geq U(y_0, 0)$. In that case, $(s^{obs}, h^{obs}, R^{obs}) = (2, h^*, R^*)$. The worker is non-participant if $U(y_0, 0) > \Psi\,(w^v(h^*), \mu^v(h^*) + y_0)$ and $U(y_0, 0) > U(\bar{c}, \bar{h})$. In that case, $(s^{obs}, h^{obs}, R^{obs}) = (3, 0, 0)$.

To simplify the estimation process, we consider that the case of unemployed workers is exterior to this structural model. Indeed, comparing indirect utilities implies that without job rationing, these workers would be included among workers in the primary or secondary sectors. Then, the corresponding observations should satisfy $U(\bar{c}, \bar{h}) \geq U(y_0, 0)$ or $\Psi(w^v(h^*), \mu^v(h^*) + y_0) \geq U(y_0, 0)$. However, explaining their unemployment status would require entry costs specific to these workers, in the primary and in the secondary sectors. Indeed, one may expect that it is more difficult to find a job when coming from unemployment status, which may be a signal of low productivity, and that therefore unemployed workers would suffer high entry costs into both activity sectors. Even when it is not the case, including unemployed workers in the model would hamper an already difficult estimation. Several estimation attempts have shown us that our data do not allow us to estimate such a full model[14]. We choose rather to consider the job rationing process leading to unemployment as exogenous, which is equivalent to excluding unemployed workers from the sample used in the estimation. In the next section, we describe the utility function.

2.3. The utility function

Several specifications of the utility function have been used in the labour supply literature. The linear form in income and wage (Hausman, 1981, Stern, 1986) presents deficiencies: low income individuals may exhibit a negative marginal rate of substitution between consumption and hours of work, and for a sufficiently large wage rate, consumption becomes an inferior good. As a simple alternative to Hausman's form, we make use of the following class of utility functions, suggested by Blundell (1990) and used in Creedy and Duncan (1998).

---

[14] Lanot and Muller (1996) is an example of an extended model that did not converge.

$$(1) \qquad \Psi(\mathrm{w}, \mathrm{y}_0) = \left[\mathrm{y}_0 + \frac{\alpha}{\gamma+1}\mathrm{w} + \frac{\beta}{\gamma+1}\ \mathrm{w} \left(\ln(\mathrm{w}) - \frac{1}{\gamma+1} \right) \right] \mathrm{w}^{\gamma},$$

where imposing $\beta > 0$, $0 > \gamma > -1$, is sufficient to ensure the convexity of preferences, a strictly positive marginal rate of substitution, and the normality of consumption good. Using Roy's identity we obtain the following labour supply function.

$$(2) \qquad \mathrm{h} = \alpha + \beta \ln(\mathrm{w}) + \gamma \frac{\mathrm{y}_0}{\mathrm{w}}.$$

Heterogeneity of preferences is introduced through the intercept term α. When the labour supply is strictly positive, the marginal rate of substitution $\omega(c, h)$ and the virtual income $\mu(c, h)$, associated with the separating hyperplane of the optimisation programme, are solutions to the system of equations:

$$(3) \qquad \mathrm{h} = \alpha + \beta \ln(\omega(\mathrm{c}, \mathrm{h})) + \gamma \frac{\mu(\mathrm{c}, \mathrm{h})}{\omega(\mathrm{c}, \mathrm{h})} \text{ and}$$

$$(4) \qquad c = \omega(c, h)\, h + \mu(c, h).$$

The direct utility function can be obtained by

$$(5) \qquad U(c, h) = \Psi(\omega(c, h), \mu(c, h)),$$

which yields the marginal utilities:

$$(6) \qquad \frac{\partial \mathrm{U}(\mathrm{c}, \mathrm{h})}{\partial \mathrm{c}} = \omega(\mathrm{c}, \mathrm{h})^{\gamma} \text{ and } \frac{\partial \mathrm{U}(\mathrm{c}, \mathrm{h})}{\partial \mathrm{h}} = -\, \omega(\mathrm{c}, \mathrm{h})^{\gamma+1}.$$

The second important functional form to specify is that of the earning function in the secondary sector. We now attend to this task.

2.4. The earnings function in the informal sector

The functional form must first allow for decreasing returns to labour over a sufficiently wide range of hours to obtain a unique equilibrium in the secondary sector. Second, it must permit non-participation in the secondary sector, the marginal return to labour at zero hours must be bounded[15]. We use the earnings function introduced by Rosen (1976):

(7) $R(h) = h \exp(\xi + \delta h)$,

where ξ and δ are parameters. Constant returns are obtained when $\delta = 0$, and provided the individual time endowment is bound by a level T, positive marginal returns to labour and decreasing returns to labour are obtained if $-1/T < \delta < 0$. This yields a linear log-average-wage equation:

(8) $$\ln\left(\frac{R(h)}{h}\right) = \xi + \delta h .$$

The implicit wage rate in the secondary sector is $R'(h) = (1 + \delta h)R(h)/h$. We add exogenous explanatory variables and an error term to ξ, to incorporate deterministic and stochastic variations into (7). Finally, the virtual net activity income is obtained by substracting the shadow value of labour input.

(9) $$\mu^v(h) = R(h) - h\,R'(h) = -\delta h\,R(h) .$$

We are now ready to present the data used for the estimation.

3. The Data

Cameroon in 1993 is a lower-middle income economy (770 US$ of GNP per capita) in Africa. The data used in this paper are taken from a survey about the employment conditions of the population of Yaoundé, capital of Cameroon (DIAL-DSCN, 1993). The survey was conducted by DIAL and the DSCN at the Finance Ministry of the Republic of Cameroon. 11,172 people in 1961 households were selected using a sampling scheme based on spatial tele-detection. The collection took place in January-February 1993. The survey provides information about the labour market status and the activity of every household member aged over 10. Among 7,865 individuals of

[15] While the Cobb-Douglas specification meets the first requirement, it does not satisfy the second one since marginal return to labour tends to infinity when hours tend to zero. Non-parametric estimation of non-linear budget sets has been investigated by Blomquist and Newey (2002). In this paper, we follow a parametric approach.

more than 9 years old, 2,947 are occupied active members and 832 are unemployed. For the estimation, we focus on the married women or single female heads of household aged between 18 and 55. This leaves us with 1,225 individuals.

These individuals in the primary sector (27 percent of the population) are wage earners, employed in the public sector, either in administration (64 percent) or in a state firm (12 percent), or employed in a private firm (24 percent). The secondary sector workers (18 percent of the population) are those who claim to be working either as small firm employers (less than 10 percent of the secondary sector workers) or as artisans. Unemployed workers (41 percent of the population) are: individuals reporting to have been searching for a job at least once during the four weeks before the interview, individuals claiming unavailability of jobs as a reason for not searching and individuals reporting to suffer from under-employment because of bad economic conditions. 14 percent of female workers decided not to participate in the labour market.

We summarise this situation by assuming that salaried workers belong to the primary sector, whereas the secondary sector is composed of self-employed workers. Because of the importance of administration in Yaoundé, most (76 percent of) salaried workers are civil servants. The majority of these individuals work around 40 hours a week. Primary sector jobs in the private sector often share similar characteristics. In the secondary sector, the worker determines her labour supply that is equal to the work input of her firm.

Table 1 gives details on the socio-economic characteristics of the sample by activity sector. The values obtained compare well with what is known of the labour market in Yaoundé (DIAL-DSCN, 1993a, 1993b). The mean level of household assets is slightly higher in the secondary sector than in the primary sector, consistently with the presence of physical capital to support informal activities. Non-working women correspond to a much lower level of household assets.

The variance of the worked hours in the two sectors corresponds to what is recognised in the literature as the relatively stronger rigidity of hours in the primary sector. In the primary sector, 81 percent of the population works from 30 to 48 hours a week, with a strong mode at about 40 hours where we observe 65 percent of the sample. This justifies considering the hours worked per week rationed in the primary sector. In the secondary sector, we find 38 percent of the population working from 30 to 48 hours a week and the mode is at 60 hours where we observe 10 percent of the observations.

Table 1: Descriptive statistics by sector of activity, for the female workers

	Primary sector	Secondary sector	Non participant	Unemployed	Inactivity sector	Whole population
Age	36.37 (6.8)	35.16 (8.07)	32.01 (9.2)	28.77 (6.7)	29.58 (7.54)	32.43 (8.1)
Years of experience	11.45 (6.65)	8.40 (7.98)	1.73 (4.56)	1.33 (4.01)	1.43 (4.16)	5.41 (7.29)
Years of education	11.92 (4.0)	6.48 (3.78)	7.83 (4.6)	8.78 (3.3)	8.54 (3.70)	9.07 (4.2)
Husband's age	27.7 (20.4)	28.36 (21.04)	37.02 (13.4)	31.64 (14.4)	32.99 (14.38)	30.71 (17.66)
Husband's years of education	9.60 (7.9)	5.85 (5.59)	9.52 (5.70)	9.12 (5.4)	9.22 (5.45)	8.70 (6.4)
Primary education	0.11 (0.32)	0.62 (0.49)	0.46 (0.50)	0.34 (0.48)	0.37 (0.48)	0.35 (0.48)
Higher education	0.28 (0.45)	0.027 (0.16)	0.10 (0.30)	0.072 (0.26)	0.079 (0.27)	0.12 (0.33)
No female members 13-18 years	0.67 (0.86)	0.50 (0.73)	0.35 (0.66)	0.37 (0.70)	0.37 (0.69)	0.47 (0.75)
No female members 0-12 years	1.15 (1.28)	1.20 (1.25)	1.03 (1.24)	1.18 (1.23)	1.14 (1.24)	1.15 (1.25)
No children under 4 years old	0.47 (0.73)	0.53 (0.71)	0.74 (0.84)	0.71 (0.75)	0.72 (0.77)	0.62 (0.75)
No children 4-10 years old	1.16 (1.19)	1.24 (1.23)	1.02 (1.18)	1.24 (1.32)	1.18 (1.29)	1.19 (1.25)
Number of dependent members	1.85 (2.37)	2.76 (2.16)	1.04 (1.68)	1.36 (1.80)	1.28 (1.77)	1.47 (2.11)
Dummy for married woman	0.67 (0.47)	0.68 (0.47)	0.93 (0.25)	0.86 (0.34)	0.88 (0.32)	0.79 (0.41)
Weeks per year	45.23 (8.30)	41.37 (14.02)	4.57 (13.96)	2.71 (9.69)	3.17 (10.94)	21.52 (22.98)
Hours per week	38.94 (8.47)	42.25 (21.78)	0 (0)	9.73 (10.48)	9.73 (10.48)	39.69 (15.88)
Assets (*)	878.50 (2720)	989.58 (3943)	301.40 (2079)	171.96 (1006)	204.42 (1356.3)	530.22 (2443)
Unearned income (predicted) (*)	61.02 (64.9)	24.34 (27.01)	40.74 (43.37)	32.49 (26.21)	34.56 (31.57)	39.81 (44.64)
Secondary sector wage (pred.) (*)	0.49 (0.47)	0.19 (0.16)	0.23 (0.23)	0.19 (0.17)	0.20 (0.19)	0.28 (0.31)
Primary sector wage (pred.) (*)	0.79 (0.54)	0.35 (0.25)	0.34 (0.26)	0.34 (0.20)	0.34 (0.22)	0.46 (0.40)
Observed wage (*)	67.02 (112.07)	-	-	-	-	45.77 (94.72)
Number of observations	330	225	168	502	670	1225

* : in thousands of FCFA.

Equations in Tables 2 and 3 are used for generating the predicted variables so as to replace missing values.

Dependents are defined as individuals surveyed living with the household who are neither the head of household, nor his/her spouse nor the children of the household head.

Are those activity sectors chosen by the workers? DIAL-DSCN (1993a, 1993b) suggests that the workers often choose their activity sector for themselves. Indeed, 25 percent of unemployed people who lost their job chose to quit voluntarily. Furthermore, less than 20 percent of the informal independent workers search for work in the informal sector because they had not found a salaried job in the modern sector.

We now present the estimation results of wage equations for the two sectors, as a preliminary statistical analysis of the data and to provide predictions of these wages for workers not observed in the considered sector. For the primary sector, we estimate a log-wage equation on the continuous observations using an instrumental variable method to account for the possible endogeneity of experience. Inverse Mills' ratios included to investigate the selectivity bias for the participation in the considered sector are not significant and have been excluded. The log-wage equation is estimated for the secondary sector with the average hourly income instead of the wage. Table 2 shows the OLS estimation results of the log-wage equations.[16] These equations are used as instrumental equations for missing values and workers observed in another sector. As expected, returns to education and experience are higher in the primary sector than in the secondary sector[17].

Statistics of the predictions from the OLS estimates of the wage equations are shown in Table 1. The mean predicted unearned income and the mean predicted weekly wage rate of the female worker are higher in the primary sector than in the secondary one. Whatever the activity sector of the worker, on average her predicted primary sector wage rate is higher than her predicted secondary sector wage rate, which is consistent

---

[16] The number of observations differ here from that of Table 1 because we used only the most precise indicators for wages and unearned incomes in the estimation, systematically eliminating the "zero" observations for incomes and inaccurate data (by classes) for wages. Indeed, we believe that, for this type of data, measurement error is a major problem and if not treated would lead to meaningless estimates.That is another reason for not treating explicitly the endogenous selectivity of the sample in these equations since some attrition bias already results from the calculus of indicators. There are two selection processes with those who do not work in the corresponding sector and those who work in this sector but did not answer their wage. Unfortunately, in both cases accurate instruments are not available to seriously identify these selection processes.

[17] The 2SLS estimates are rejected because of the results of Hausman tests which are shown in Table 7. In all cases, the exogeneity of education and experience cannot be rejected at usual level. This is also the case when the exogeneity of education and experience have been tested jointly with selectivity. The endogeneity of household composition is also rejected in these data. Table 6b shows that we are not in the case of weak instruments since the F-statistics of the first-stage regressions are always substantial. Selection correction and non-instrumentation of experience did not significantly change the results.

with dualism. However, primary workers gain more by being in this sector than other workers.

Table 2: Log-wage equations

	Primary sector OLS	Secondary sector OLS
Constant	-1.914 (5.3)	-3.355 (5.9)
Age	-0.0020 (0.3)	0.0228 (1.7)
Years of education	0.0707 (3.60)	0.0483 (1.3)
Experience	0.0398 (3.78)	0.0119 (1.0)
Primary education	-0.206 (1.2)	0.0123 (0.05)
Higher education	0.653 (4.4)	1.362 (2.6)
R^2	0.536	0.205
Number of observations	188	108

Student's t in parentheses. Wages are measured in thousand of FCFA per week. Only observations with exact wages and exact earnings are used in these regressions. Unaccounted missing information are dropped since they yield meaningless estimates. No selection correction for these missing values has been implemented because instrumental variables are not available to identify it with accuracy this selectivity. Data collected by using classes are converted by using mean values in these classes.

Another important determinant of participation and activity choice should be unearned income since women who already dispose of a large unearned income have little incentive to reduce their leisure time. Observed unearned income is the sum of the woman's unearned income and the total income of the husband and of the dependents. Some women are observed with no unearned income. We assume that this is due to non-observation rather than to non-existence of the unearned income. We provide predictions for this unobserved information ("predicted unearned income") by estimating an unearned income equation. The results of this estimation are presented in Table 3. The existence of a significant selection effect using a conventional Heckman

Table 3: Non-labour income equations

	Heckman two-steps estimates		Ordinary Least Squares	
Variables	*Estimate*	*(t-value)*	*Estimate*	*(t-value)*
Constant	2.582	(2.0)	1.067	(3.3)
Number of children	0.071	(1.6)	0.099	(3.2)
Number of children (0 to 3 years)	-0.194	(2.5)	-0.146	(2.7)
Number of children (4 to 10 years)	-0.0928	(1.7)	-0.086	(1.9)
Number of dependents	0.008	(0.4)	0.027	(1.7)
Husband's age	0.0256	(2.3)	0.016	(2.4)
Husband's years of education	0.0857	(3.7)	0.096	(5.6)
Husband's primary education	-0.115	(0.7)	-0.145	(1.1)
Husband's higher education	0.146	(0.77)	0.233	(1.7)
Husband's father employed	-0.406	(1.6)	-0.259	(1.4)
Husband's father self-employed	0.128	(1.2)	0.157	(1.9)
Husband's father employed in public sector	0.252	(1.1)	0.052	(0.4)
Husband's father employed in a small private firm	-0.731	(2.0)	-0.673	(2.4)
Husband's father employed in a large private firm	0.347	(2.0)	0.322	(2.3)
Age	-0.0007	(0.08)	0.003	(0.5)
Years of education	0.035	(1.5)	0.021	(1.2)
Primary education	0.032	(0.2)	-0.046	(-0.4)
Higher education	0.208	(1.1)	0.227	(1.5)
Father employed	0.0209	(0.11)	0.132	(1.0)
Father self-employed	-0.0424	(0.4)	-0.047	(0.6)
Father employed in public sector	-0.137	(0.9)	-0.165	(1.3)
Father employed in a small private firm	-0.181	(0.5)	0.035	(0.1)
Father employed in a large private firm	0.0470	(0.3)	0.0041	(0.03)
Dummy for married woman	-1.534	(1.4)	-0.281	(0.8)
Inverse Mill's ratio	-1.492	(1.2)	-	-
R^2	0.421		0.420	
Number of observations	956		956	

Because of missing values for non-labour income, the sample size (956) is smaller than that of the whole sample (1225).

2-steps method is rejected. The endogeneity of household composition and head's characteristics cannot be rejected by Hausman's tests (see Table 7). However, the 2SLS estimates of the coefficients are almost all insignificant and family background variables may not be valid instruments since they may be correlated with unobserved ability. Then, the predictions for unearned income are obtained using the OLS estimates on complete observations only. Now that we have presented the method for generating the predictions for wages and incomes, we proceed to the estimation of the structural model.

4. Estimation of the Structural Model

4.1. Some econometric difficulties

Several econometric difficulties have to be dealt with for the functional and stochastic specification of the model. First, the selectivity bias has been well documented for the standard labour supply model (Heckman and MaCurdy, 1986). Chiswick (1988) proposes to incorporate, in an analysis of earnings, those labour force participants who are self-employed, and thus he avoids the selection bias in the earning functions estimates that arises when the sample is restricted to wage earners. We go further with all sector choices simultaneously estimated. Nonetheless, the possible endogeneity of household composition may stochastically link this composition with activity sector and labour supply. For example, new members from outside may join the richer households, whose female workers are predominantly active in the primary sector. Unfortunately, our data do not enable us to correct for these parasite links, and we ignore them.

Second, identification conditions have to be imposed in switching response models. In our case, because of the highly non-linear character of our model, it is not possible to globally identify many parameters with accuracy. In view of this, we limit the number of exogenous variables in this model to ease parameter identification. Additionally, the correct derivation of the likelihood from the errors distribution involves the imposition of logical coherency to the model (Gouriéroux, Laffont, Monfort, 1980). Logical coherency is satisfied if the parameter values correspond to a unique optimum for each value of the errors and vice versa. This condition is obtained by assuming the strict concavity of the utility function and of the earnings function[18], and by the fact that the stochastic state defined by the equality of utility levels in two different sectors has a null probability.

Third, one limitation of the model is that the identification is obtained from a specific choice of functional forms and distribution assumptions. In particular, the unobserved heterogeneity is incorporated in the three error terms. Clearly, the choice of functional forms and distributions matters for the model characteristics. However, in the absence of a priori information, we try to select the most appropriate specification in terms of simplicity and reasonable properties. This approach is a first step towards more realistic, but less tractable specifications.

Fourth, we adopt the following simplifying strategy to deal with wages, rationed hours and activity incomes when they are unobserved. The definition of the rationing level in the primary sector for workers who are observed in this sector is the observed level of hours. The rationed hours in the primary sector for workers not observed in this sector are predicted at the mean of hours observed in the primary sector. The wage rate in the primary sector of workers unobserved in this sector are predicted using the estimated log-wage equation in Table 2. The non-labour income for workers for which it is unobserved is predicted using the estimated income equation in Table 3. Naturally, this approach ignores the potential selectivity in these equations[19]. However, since including these equations in the global model would be intractable, we can thus obtain reasonable predictions for these variables. The secondary sector earnings for those who do not report it are directly predicted using eq. (7) of the model, which is estimated simultaneously with the preferences and the entry cost. With these econometric difficulties in mind, we proceed with the calculation of the likelihood of the model.

4.2. The likelihood function and the exogenous variables

The full model is given by the inequalities between the utility levels in the different sectors, as defined in Section 2.2 with the utility function (1) and the earnings function (7) in the secondary sector. The maximum likelihood method used in the

---

[18] As noticed by Heim and Meyer (2003) for Hausman's labour supply model under nonlinear budget constraint.

[19] Two-stage Heckman estimates including Mill ratios have induced us to reject the selectivity, although in a non-structural way.

estimation is discussed in Appendix 1. The three stochastic terms are related to the structural parameters of the model: the heterogeneity of preferences, α, the scale of the earnings function, ξ, and the fixed cost, f. Parameters α, ξ and f are supposed to be linear in the error terms and in some individual characteristics, and to be distributed according to a trivariate normal distribution.

However, more restrictions must be imposed on this model to make its estimation tractable. Indeed, it is not possible to identify with accuracy all the parameters related to the explanatory variables when they are simultaneously included in the preferences, the technology and the entry cost. Therefore, we introduce each exogenous explanatory variable only in one of these structures. Earnings (or wage rates) are explained by the usual variables in the human capital theory (age, education level, experience), and also, in the case of the informal sector, by the capital stock of the individual firm, approximated by household assets. We are also interested in children whose presence generates incentives to the mother to take care of them, and in other female members who might substitute themselves for the female workers in domestic activities. Then, we assume that the preferences of the female worker depend on the number of children and on the number of female members (both categories split up in two age classes), on the woman's age and on a dummy variable for married women. Indeed, age is generally considered a major determinant of preferences and the presence of a husband is expected to influence female worker decisions in Cameroon. As for the entry cost, our attempts at statistically identifying explanatory variables have failed. Therefore, we assume that it is a constant parameter whose measurement unit is Franc CFA. Its variance may also be considered a measure of the heterogeneity of the budget constraint specific to the primary sector. A dual interpretation of the fixed cost suggests that it may partially account for the unobserved heterogeneity in hours rationing in the primary sector. We now examine the estimation results.

4.3. *Estimation results*

Table 4 shows one typical estimate of the model.[20] The parameters β and $(1/\gamma)$, which are related to the economic variables in the implicit labour supply equation, i.e. the predicted implicit wage and the predicted unearned income, are well determined[21]. The number of young children under 4 years of age has a negative effect on the propensity to work, which may reflect their need for caring by the female worker. In contrast, the number of children between 4 and 10 years of age, the number of female members and the fact that the individual is married, correspond to incentives to work when they are high. The latter effect may be related to the adverse economic conjuncture with a steady rise of unemployment in the studied period, since married women may be losing advantages linked to the husband's job, while non-married workers do not suffer from this shock. The presence of other female members may ease the substitutability of the female worker's labour in domestic activities. Age has no significant effect on work decisions, perhaps because of compensation effects in the life cycle. We cannot capture the latter effects in our model because age and age squared are too correlated for being accurately identified.

The fixed costs are identified by the structure that we impose on preferences and technology. Because of the high non-linearity of the model, we have not been able to obtain estimates for the variance-covariance matrix of the parameters when explanatory variables have been included in the fixed costs, even when the convergence is reached.[22]

---

[20] Because of all the non-linearities involved in the model, the computation is time consuming. The precision required for the integration and for the root calculation makes it variable. One week of calculation time with GAUSS on a SUN-SPARC station is a typical duration for one estimation. The region of convergence had also to be reached by grid method and likelihood concentration methods so as to compensate the inaccuracy of the gradient method near the optimum. Note also that the algorithm for a similar model with hours freely chosen in the primary sector does not converge. An alternative method to improve the estimation of the integrals in the likelihood could have been to use simulated likelihood estimators like in Pradhan (1999).

[21] The standard errors shown in the table do not account for the fact that some data on wages and non-labour income are predicted.

[22] Note that to the best of our knowledge, nobody has yet succeeded in estimating fixed cost functions in such non-linear models. This may explain why Newman and Gertler (1994) eliminate the budget constraint in their model and rely on comparisons of implicit log-wages based on linear specifications. One reason for this difficulty is that only very rich data sets could identify with accuracy both the parameters in the linear budget constraint and in a non-linear production function included in this budget constraint. Indeed, the corresponding Fisher information matrix has columns that are likely to be very correlated, leading to its quasi-singularity and the impossibility of estimating the variance-covariance matrix of parameters.

Table 4: Maximum likelihood estimates

α	**Constant**	52.71 (6.5)
	Number of children (0-3)	-5.52 (2.9)
	Number of children (4-10)	4.25 (3.1)
	Married woman	7.71 (2.2)
	Number of female members (over 13 of age)	4.65 (2.4)
	Number of female members (0-13)	4.15 (2.9)
	Age	-1.00 (0.05)
	β	6.25 (5.4)
	1/γ	-12.24 (11.8)
ξ	**Constant**	-3.05 (17.6)
	Assets	0.030 (2.1)
	Education	0.110 (11.4)
	Experience	0.0044 (5.9)
	δ	-0.0022 (2.9)
	f	2.69 (0.4)
Variance-Covariance Matrix	σ_α	32.8 (17.2)
	σ_ξ	0.89 (10.5)
	σ_f	109.9 (12.4)
	$\rho_{\alpha,\xi}$	-0.44 (5.4)
	$\rho_{\alpha,f}$	0.23 (2.9)
	$\rho_{\xi,f}$	-0.66 (2.6)

Mean log-likelihood = -3.64. Number of observations = 1206. Absolute t-values in parentheses.

A few outliers (19) have been discarded from the sample so as to reach the convergence point. To ease the optimisation algorithm behaviour, some variables have been rescaled: hours are per week, wages in thousands of FCFA, assets in millions of FCFA, education and age in years. This must be taken into account when interpreting the above estimates.

Moreover, the constant fixed cost coefficient is not significant at 5% level. However, its estimate is 4.4 percent of the non-labour income for primary sector workers, 11 percent for workers in secondary sectors, and respectively 6.6 percent and 8.3 percent for non-participants and unemployed workers. This is consistent with the contribution of the entry cost to the selection of workers since it corresponds to a relatively lower financial burden for actually observed primary workers. Workers whose fixed cost is a relatively larger financial constraint, compared to their financial capacity without wage income, and who may not access to the primary sector, seem to be rejected more often in the secondary sector than in inactivity. The large variance σ^2_f indicates that the amount of fixed costs any individual face can be very variable.

In view of these difficulties in estimating the entry cost, one may wonder if the introduction of an entry cost is justified. We tried to estimate the model without the entry cost, but in that case it is not even possible to reach the convergence of the likelihood optimisation, except towards nonsensical estimates, because the simplified model does not fit the data well. A fortiori the estimation of a model without hours rationing does not converge. Without an entry cost in the primary sector, it is impossible to explain, with the model and the available exogenous variables, why workers whose productivity in the secondary sector and reservation wages are low are not observed as working in the primary sector. Indeed, without barriers to sector entry the only reason for a worker observed as non-participating or in the informal sector not to enter the primary sector would be either that the wage rate is too low compared to its reservation wage (or to the shadow wage that she could get from informal activities), or that the level of rationed hours is too far from her desired level. This is unlikely in our context, since primary jobs are much better remunerated than secondary jobs and observed workers are generally poor enough to be willing to work if the primary sector is available. Clearly, the entry cost is needed to explain job rationing in the primary sector, although our data do not enable us to describe well the heterogeneity of situations associated with this rationing. Here, the non-significance of the fixed cost parameter does not mean that it should be zero, but rather that more information is needed to identify it with accuracy.

However, one may want to simplify the model more. What would happen without the entry cost, the hours rationing and the non-linearity of the earnings function? In this situation, we would be back to a model based on the comparison of the reservation wage with the wage rates in the two sectors. In that case, most observed workers should want to work in the primary sector, when the estimated wage equations are used for predicting their wage in the primary sector. Then, such a simple model would fail to explain the existence of a large secondary sector and a large unemployment level. Also, a model with a linear earnings function in the secondary sector would not fit the data well. For example, such a model would exaggerate the rewards of working in the secondary sector for workers wishing to work a large number of hours. Besides, the existence of diminishing returns to labour in this sector is of major interest in the study of the informal sector. Introducing only the non-linearity in the earnings function and the hours rationing in the primary sector would not substantially modify the relative attractiveness of the primary sector. Therefore, the entry cost, or an equivalent device, seems to be essential to the model.

The third estimated structure is the earnings function in the secondary sector. Household assets, which are a proxy for the capital stock of the firm, are statistically significant, suggesting the inadequacy of the usual wage rate approach. The experience and the education of the female worker positively influence the productivity, which was not the case for the direct estimates of the wage rate equation of the secondary sector[23]. Parameter δ is significantly negative, which implies that the earnings function is strictly concave. However, the estimated curvature of the earning function is very moderate. This may be caused by two features of the studied context. First, the informal sector is an aggregate of heterogeneous activities, with some workers receiving very variable remunerations and other ones paid at almost a fixed wage rate. The contribution of these last workers in the likelihood pushes the estimates towards a zero curvature. Second, as in organisational dualism theories, the earnings of the female worker may be shared with other members of the family, and the average shadow wage or a fraction of the shadow wage may sometimes be closer to the characteristics of the decision process, than the actual shadow earnings function that is therefore imprecisely determined by our individualistic model. Nonetheless, the log-linear shadow wage equations, which are generally used in the literature, are likely to be misspecified.

We can compare the returns to education and experience obtained from the earnings function estimation by using the average shadow wage, R(h)/h, with the returns obtained in the wage equations. Indeed, wages in wage equations are also calculated as means over hours. The comparison with the marginal shadow wages (R'(h)) can also be made by multiplying the returns obtained for the average shadow

[23] However, as we indicated above, this may also come from the impossibility of identifying the effects of these variables in preferences.

wage by $(1+\delta h)$, with h chosen, for example, at the mean hours for the population of interest. In practice, this makes only a small difference because the value of parameter δ is close to zero.

The estimated returns to education in the secondary sector are much larger than had been obtained in the wage equation (11 % instead of 4.8 %), and even more than the return to education in the wage equations from the primary sector, suggesting that the global model overestimates these returns, or that the wage equation underestimates them. The source of this possible overestimation is unclear, but it may be that part of the influence of the education variable should be attributed to preferences shift rather than to productivity differences. For example, more educated workers in the informal sector could prefer to work fewer hours and then have a larger productivity in presence of diminishing returns to labour.

Conversely, the return to experience calculated from the estimated earnings function appears much lower than the estimates for the wage equation in the secondary sector (0.4 % instead of 1.2 %). The difficulty of introducing explanatory variables of the entry barriers in the formal sector may be at the origin of this. Indeed, workers with a high level of experience should be less subject to such barriers. Then, workers with high experience observed in the secondary sector often have bad unobserved characteristics that both explain why they cannot access the primary sector and why their productivity is low in the secondary sector.

Finally, the correlation coefficients between preferences, fixed cost and technology are all significant. Owing to the identification restrictions that were necessary for the convergence of the estimation, these correlation coefficients partly reveal the influence of exogenous variables that should be introduced in several of the three main structures, rather than in only one of them as is the case now. A second interpretation of these coefficients is that they reflect unobserved heterogeneity common to preferences, informal technology and entry cost. The estimated magnitude of the correlation coefficient of the entry cost and of the informal technology, $\rho_{\xi,f}$, is substantial, while the sign of the coefficient is negative. This suggests that unobserved characteristics may simultaneously explain worker productivity in the informal sector and low levels of entry costs in the formal sector. Then, when possible, one should incorporate in the model more information on qualification, human capital and characteristics that are related to productivity gains in both sectors. What is described as an entry barrier may sometimes come from unobserved differences in productivity, rather than for example arbitrary discrimination. In that sense, education and experience appear to be insufficient characterisation of the productivity characteristics of workers.

The coefficients of correlations of ξ and f with the preference parameter α are also interesting, even if their magnitudes are weaker. Preferences and productivity in the informal sector are negatively correlated, while preferences are positively correlated with the entry cost. This is consistent with the presence of unobserved characteristics of human capital simultaneously explaining differences in tastes, and productivity in the two activity sectors. Thus, the results support joint modelling as a response to simultaneously unobserved heterogeneity of preferences, technology and entry costs.

Since the data do not allow us to estimate the parameters of the fixed cost with accuracy, one could wonder if the specification of a fixed cost is justified for treating the rationing in the formal sector. Other modelling approaches could be tried, such as exogenous rationing, introduction of a probability of being rationed, and a shift in the preferences when entering the primary sector. All these possible solutions share unattractive features in that they introduce exogenous determination of the sector choice and/or correspond to unintuitive representations of barriers to entry in the primary sector. We mentioned earlier the difficulties involved in the elimination of the fixed cost altogether.

On the whole, we believe that the three new introduced elements (fixed costs, hours rationing, non-linearity of the earnings function) are important. However, linear reduced-form models remain useful approaches because they allow a large number of explanatory variables even if the interpretation is difficult since these variables may originate from different underlying structures. The impossibility of including all explanatory variables is the reason why we do not attempt to compare our results directly with a linear reduced-form approach. We prefer to focus on the structural properties of our model rather than on the robustness of the effects. It seems that except when very rich data are available, a choice between the structural relevance of the model and a complete set of exogenous explanatory variables is inevitable. In that sense, there exists a trade-off in terms of theoretical relevance against goodness-of-fit, results of statistical tests and capacity to incorporate many correlates. In the state of the present modelling knowledge, a complex non-linear model cannot compete yet in terms of pure statistical adequation with linear regressions incorporating many regressors. However, as opposed to simple equations, it provides specific qualitative features of interest (concavity, diminishing returns, barrier to entry) and a framework of interpretation. Moreover, endogeneity problems in reduced-forms and the insufficiencies arising from the methods of instrumental variables (Staiger and Stock, 1997) raise doubts about the results obtained in reduced-form regressions. Investigations such as the one presented in this paper are likely to guide future specifications. We now examine the predicted elasticities generated by the estimated model.

4.4. Predicted elasticities

Table 5 provides several measures of elasticities based on the separation structure of the optimisation programme, for the mean worker in each of the three sectors of interest. They are estimates of elasticities of labour supply and elasticities of earnings[24]. Moreover, we also calculated predicted actual and virtual wages and incomes for these workers. All these estimates characterise the potential labour supply responses and the corresponding revenues for the mean worker in each sector, under the assumption that there is no change of regime, and in reference to simplified constraints (the implicit separating hyperplane associated with the optimisation programme). It must be remembered that the above elasticities characterise latent supply functions rather than observable elasticities that should account for the non-separability of the system and the changes in regime. They provide, however, useful analytical information about the estimated system. We calculate these elasticities by using as the level of worked hours h the mean observed labour supply, and by using the other mean characteristics, in each sub-population. For the non-participant and the unemployed, we account for the fact that their reservation rate should be higher than the proposed shadow wage rate (in the absence of fixed costs and hours rationing) by using the highest of the predicted wage rates for the primary and the secondary sectors[25]. Let us now comment on these calculated elasticities, which are all statistically significant at 5% level, as the standard errors show, except in the case of the secondary sector for the elasticities of hours with respect to preferences characteristics and of earnings with respect to technology characteristics.

---

[24] The obtained elasticities of labour demand are too negative (about –10) to be credible and are not shown. Clearly, more information, perhaps about technology characteristics, is needed to estimate labour demand elasticities with accuracy in this data.

[25] Direct uses of behavioural micro-simulations such as the ones surveyed in Creedy and Duncan (1998), would be possible although cumbersome and computation time consuming. We do not follow this approach because the estimation results are not accurate enough to justify to overcome these obstacles, keeping in mind that the prediction errors in such simulations would be very large and would probably include a notable component coming from misspecification problems.

Table 5: Elasticities, wages and incomes for different sectors

Statistics	Primary	Secondary	Inactivity	Whole Sample
Elasticities of:				
Labour supply w.r.t. (with respect to) wage rate	0.371 (0.0117)	0.497 (0.0124)	+∞	0.46 (0.0180)
Labour supply w.r.t. char. of preferences	1.903 (0.803)	1.907 (5.492)	+∞	1.98 (1.024)
Labour supply w.r.t. nonlabour income	-0.186 (0.0037)	-0.311 (0.0040)	-∞	-0.27 (0.00705)
Earnings w.r.t. worked hours	0.925 (0.00602)	0.926 (0.00984)	0.999 (0.00226)	0.93 (0.00755)
Earnings w.r.t. char. technology	-1.209 (0.155)	-1.938 (4.67)	-2.042 (1.210)	-1.79 (0.149)
Wages and Incomes:				
Wage rate in the primary sector	0.79	0.35	0.34	0.46
Wage rate in the secondary sector	0.49	0.19	0.20	0.28
Shadow wage rate in the secondary sector	0.257 (0.0965)	0.124	0.130	0.143 (0.0662)
Non-labour income when inactive	61.02	24.34	34.56	39.81
Non-labour income minus entry cost	58.33 (1.577)	21.65 (2.067)	31.87 (2.06)	37.12 (2.067)
Potential activity income in the secondary sector	9.39 (2.66)	4.49	0	5.00 (1.693)

∞ means that the elasticities tend to the infinity when the worked hours tend to zero. Standard errors based on the estimation are in parentheses. In some cases, the wages and earnings are merely sample means that do not depend on the result of the maximum likelihood estimation, which explain why standard errors are not provided for these cases.

The wage elasticities of the labour supply in the primary sector (0.371) or in the secondary sector (0.497) compare well with traditional figures (Killingsworth, 1983, Blundell, 1992). They correspond to virtual situations where only the implicit budget constraint (without fixed costs or hours rationing) is considered. Since $\gamma < 0$, this elasticity is higher for low exogenous non-labour incomes and lower for high labour supply levels, other variables fixed. It is also higher for workers observed in the secondary sector because their wage rate is much lower, which is far from being fully compensated by a lower exogenous income and larger worked hours. So, informal and less wealthy workers would react more to wage changes if there were no constraint by entry and technology barriers.

The elasticity of the labour supply with respect to the exogenous unearned income is lower in absolute value in the primary sector (-0.186) than in the secondary sector (-0.311). This is due, as previously, to a higher wage rate in the primary sector. For the non-participants and the unemployed, the wage elasticity of labour supply tends to plus infinity and the exogenous income elasticity of labour supply tends to minus infinity, when the worked hours tend to zero. This is because the number of worked hours appear at the denominator in the formulae of these elasticities since entry costs are neglected when focusing on one regime. Such a feature illustrates the importance of entry costs. Alternatively, different functional forms (more difficult to estimate) could have been chosen to avoid this situation.

Because of the moderate concavity of the earnings function, the elasticity of the potential earnings with respect to worked hours in the informal sector is close to one in all sectors for the mean household. The potential earnings in the informal sector result from the latent labour demand of the individual firm in this sector. In each sector, the shadow wage rate of the informal sector income is lower than the mean predicted wage rate. It is much higher for workers observed in the formal sector than for workers observed in the informal sector, because of the higher education, experience and assets of the former. Then, the incentives to work in the secondary sector for primary workers would be higher than for secondary workers, if the primary sector did not exist or were to disappear. In a time of employment crisis with liquidation of part of the public sector, such as the studied period in Cameroon, this is consistent with massive migration of dismissed workers from the primary towards the secondary sector. That is indeed what has been recorded in employment surveys (DIAL-DSCN, 1993a, b)

Finally, for each sector, the activity income provided by the potential informal activities of the female worker is on average relatively small as compared to her exogenous income, even when substracting the entry cost in the formal sector. Indeed, in addition to the non-activity incomes of the households, the labour incomes of the other members of the family, and particularly of the husband, constitute most of the family income. This implies that a more complete analysis of female activity decisions would benefit from considering the activity decisions of other members.

5. Conclusion

Modelling labour supply and activity choices in the dualistic context of developing countries raises complex econometric issues, especially when one allows for the presence of a non-linear earning function in the informal sector, allied to quasi-autarky of labour in this sector, and entry costs and hours rationing in the formal sector. This explains why no model dealing jointly with these characteristics has been estimated until now. This paper is a first attempt at such a task.

Using a cross-section sample of female workers from Yaoundé (Cameroon), we estimate a joint model of activity choice and labour supply, where the error terms in preferences, entry cost and informal technology are correlated. This approach is proven tractable and our results provide evidence for the concavity of the earnings function in the informal sector with respect to the hours. We estimate the effects of several explanatory variables for the preferences and for the earnings function in this simultaneous structural estimation, and we derive elasticities for different sectors.

However, the approach shows its limits in that strong identifying restrictions must be imposed on the list of exogenous variables in each structure of the model. In this sense, there exists an inevitable trade-off between, on the one hand the consistency and completeness of the economic model, and on the other hand, the incorporation of relevant information about the worker characteristics through exogenous variables.

The difficulty of obtaining accurate estimates for structural models with joint estimation of preferences, entry costs and technology raises doubts about the interpretation of the estimates of reduced-form models[26]. Indeed, if one believes that the structural model is correct, or that its possible misspecification is of little importance, as compared to the roughness of the reduced-form approach, then its estimation should provide more efficient, and perhaps more consistent, estimates than reduced-forms. Then, the discrepancy between high correlation results obtained in reduced-form models and low significance obtained with the joint structural modelling can be attributed to several causes. First, the structural model is too imperfect and these imperfections do not appear in the reduced-form specification. The problem here is that the meaning of

---

[26] Recently, Belzil and Hansen (2002) showed estimation results of a structural model of education and work choices for US workers that contradict the usual estimated returns for schooling in reduced-form wage equations.

the reduced-form specification is often implicitly or explicitly interpreted in reference to a metaphoric structural economic model. This raises the question: How can the structural model be wrong and its reduced-form consequence right? Another possibility is that the general structural model is correct, but its implementation is faulty. For example, badly specified error terms, inadequate functional forms or incorrect estimation methods may have been used. This type of error is possible in the present state of the difficult estimation of structural models. But again, why would these problems affect only the structural model and not the reduced-form model? Third, the bulk of the problem may come from the way reduced-form estimates are considered. Take for example the basic labour supply equation incorporating wage rate, exogenous income, household composition, age, experience and education. It may be that the interpretation of the estimated coefficients in this equation should be more complex than is usually supposed. The effects of age and education are generally considered as characterising the female worker preferences, or labour market characteristics. However, in LDCs with imperfect labour market, such variables may be correlated with productivity in informal technology, levels of productive assets and entry costs in various activity sectors. Therefore, all these elements should be taken into account when interpreting the results of reduced-form models. Unfortunately, mixing explanations does not generally clarify the analysis of estimates. We leave this discussion here.

What we should aim for is to reach the same type of empirical results, at least qualitatively, for complex activity choice models and simple reduced-form models, so as to cross-validate the estimation results of both approaches. For this, more progress in the estimation of structural models is needed.

References

BEHRMAN, J. R. (1999), "Labor Markets in Developing Countries," chapter 13 in O. Ashenfelter and D. Card (Eds.), "*Handbook of Labour Economics, Volume 2*," Elsevier Science B.V.

BELZIL, C. and J. HANSEN, "Unobserved Ability and the Return to Schooling," *Econometrica*, Vol. 70, No. 5, 2075-2091, September 2002.

BENJAMIN, D. (1992), "Household Composition, Labour Markets, and Labour Demand: Testing for Separation in Agricultural Households Model," *Econometrica*, 60, 287-322.

BLOMQUIST, S. and W. NEWEY (2002), "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica*, Vol. 70, No. 6, 2455-2480.

BLUNDELL, R. (1990), "Evaluating Structural Microeconometric Models of Labour Supply," *Working Paper University College London*.

BLUNDELL, R. (1992), "Labour Supply and Taxation: a Survey," *Fiscal Studies*, 13, 15-40.

CHISWICK, C.U. (1988), "Analysis of Earnings from Household Enterprises: Methodology and Application to Thailand," *Review of Economics and Statistics*, February.

COATE, S. and S. TENNYSON (1992), "Labor Market Discrimination, Imperfect Information and Self-Employment," *Oxford Economic Papers*, 44, 272-288.

CREEDY, J. and A. DUNCAN (1998), "Welfare, Non-Linear Budget Constraints and Behavioural Microsimulations," mimeo University of Melbourne.

DAHL, G. B. (2002), "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, Vol. 70, No. 6, 2367-2420.

DIAL-DSCN (1993a), "Conditions d'Activité de la Population de Yaoundé: Premiers Résultats," *Technical Report*.

DIAL-DSCN (1993b), "Le Secteur Informel à Yaoundé, Enquête 1-2-3. Phase 2," *Technical Report*, November.

DICKENS, W.T. and K. LANG (1985), "A Test of Dual Labour Market Theory," *American Economic Review*, 75, 792-805.

DICKENS, W.T. and K. LANG (1988), "The Reemergence of Segmented Labour Market Theory," *AEA Papers and Proceedings*, May.

DICKENS, W.T. and K. LANG (1992), "Labour Market Segmentation Theory: Reconsidering the Evidence," *National Bureau of Economic Research Working Paper*, 4087.

GOURIEROUX, C., J.-J. LAFFONT and A. MONFORT (1980), "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes," *Econometrica*, 48, 675-695.

HANEMANN, W.M. (1984), "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, 541-61.

HAUSMAN, J.A. (1980), "The Effect of Wages, Taxes and Fixed Costs on Women's Labour Force Participation," *Journal of Public Economics*, 14, 161-194.

HAUSMAN, J.A. (1985), "The Econometrics of Nonlinear Budget Sets," *Econometrica*, 53, 1255-1882.

HECKMAN, J. and B. HONORÉ (1990), "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121-1149.

HECKMAN, J. and T.E. MACURDY (1986), "Labour Econometrics," chapter 32 in Z. Griliches and M.D. Intriligator, "Handbook of Econometrics," Elsevier Science Publishers.

HECKMAN, J. and G. SEDLACEK (1985), "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labour Market," *Journal of Political Economy*, 93, 1077-1125.

HEIM, B.T. and B.D. MEYER, "Structural Labor Supply Models when Budget Constraints are Nonlinear," mimeo Northwestern University, June 2003.

HILL, M.A. (1988), "Female Labour Force participation in Developing and Developed Countries - Consideration of the Informal Sector," *Review of Economics and Statistics*, February.

HILL, M.A. (1994), "Female Labour Supply in Japan. Implications of the Informal Sector for Labour Force Participation and Hours of Work," *Journal of Human Resources*, XXIV, 1.

JACOBY, H.J. (1993), "Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra," *The Review of Economics Studies*, 60, 903-921.

KILLINGSWORTH, M.R. (1983), "Labour Supply," Cambridge University Press.

KILLINGSWORTH, M.R. and J. HECKMAN, (1986), "Female Labour Supply: A Survey," chapter 2 in O. Ashenfelter and R. Layard, "Handbook of Labour Economics," Elsevier.

KOOREMAN, P. and KAPTEYN, A., (1987), "A Disaggregated Analysis of the Allocation of Time within the Household," *Journal of Political Economy*, Vol. 95, No. 2.

LANOT, G. and C. MULLER, "Dualistic Sector Choice and Female Labour Supply," Mimeo University of Oxford, 1996.

LEMIEUX, T., B. FORTIN and P. FRÉCHETTE (1994), "The Effect of Taxes on Labour Supply in the Underground Economy," *American Economic Review*, 84, 231-254.

MAGNAC, T. (1991), "Segmented or Competitive Labour Markets," *Econometrica*, 59, 165-187.

MYINT, H., (1985), "Organizational Dualism and Economic Development," *Asian Development Review*, vol. 3, no. 1, pp. 25-42.

NEWMAN, J.L. and P.J. GERTLER (1994), "Family Productivity, Labor Supply, and Welfare in a Low Income Country," *Journal of Human Resources*, XXIX, N. 4.

PRADHAN, M. (1994), "Labour supply in Urban areas of Bolivia and the role of the informal sector," Ph.D.

ROSEN, H.S. (1976), "Taxes in a Labour Supply Model with Joint Wages-Hours Determination," *Econometrica*, 44, 485-507.

ROSENZWEIG, M.R. (1988), "Labour Markets in Low-Income Countries," chapter 15, in H. Chenery and T.N. Srinivasan, "*Handbook of Development Economics*," Elsevier Science Publishers.

ROY, A. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135-146.

SEN, A.K. (1966), "Peasants and Dualism with or without Surplus Labour," *Journal of Political Economy*, October.

STAIGER, D. and J.H. STOCK (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

STERN, N. (1986), "On the Specification of Labour Supply Functions," in Blundell, R. & I. Walker eds, "*Unemployment, Search and Labour Supply*," Cambridge University Press, Cambridge.

THOMAS, J.J. (1992), "*Informal Economic Activity*," *L.S.E. Handbooks in Economics*, Harvester Wheatsheaf.

TIEFENTHALER, J. (1994), "A Multisector Model of Female Labour Free Participation: Empirical Evidence from Cebu Island, Philippines," *Economic Development and Cultural Change*, July.

TODARO, M.P. (1971), "Income Expectations, Rural-Urban Migrations and Employment in Africa," *International Labour Review*, vol. 104, no. 5, November.

VIJVERBERG, W. P. M. (1993), "Educational Investments and Returns for Women and Men in Cote d'Ivoire," *The Journal of Human Resources*, XXVIII, 4.

WILLIS, R. (1986), "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," chap. 10 in O. Ashenfelter and R. Layard, "*Handbook of Labour Economics*," Elsevier Science Publishers.

Appendix 1: The Likelihood

We describe the contributions to the likelihood of each regime. The stochastic terms are normal and introduced as follows.

$$(10)\qquad \begin{pmatrix}\alpha\\ \xi\\ \mathrm{f}\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}\bar{\alpha}\\ \bar{\xi}\\ \bar{\mathrm{f}}\end{pmatrix}, \Sigma \equiv \begin{pmatrix}\sigma_\alpha^2 & \rho_{\alpha\xi}\,\sigma_\alpha\,\sigma_\xi & \rho_{\alpha\mathrm{f}}\,\rho_\alpha\,\rho_\mathrm{f}\\ \rho_{\alpha\xi}\,\rho_\alpha\,\rho_\xi & \sigma_\xi^2 & \rho_{\xi\mathrm{f}}\,\sigma_\xi\,\sigma_\mathrm{f}\\ \rho_{\alpha\mathrm{f}}\,\sigma_\alpha\,\sigma_\mathrm{f} & \rho_{\xi\mathrm{f}}\,\sigma_\xi\,\sigma_\mathrm{f} & \sigma_\mathrm{f}^2\end{pmatrix}\right)$$

where $\bar{\alpha}$, $\bar{\mathrm{f}}$, $\bar{\xi}$ are respectively the means of the parameters α, f and ξ for the preferences, the secondary sector earnings function and the fixed cost. § is the variance-covariance matrix of the three random parameters. We start with the contribution to the likelihood of female workers observed in the primary sector.

Primary Sector Workers:

This regime is such that the following inequalities are fulfilled:

$$(11)\qquad \mathrm{U}(\bar{\mathrm{c}}(\mathrm{f}), \bar{\mathrm{h}}; \alpha) \geq \Psi(\mathrm{w}^{\mathrm{v}}(\mathrm{h}^*(\alpha,\xi);\xi), \mu^{\mathrm{v}}(\mathrm{h}^*(\alpha,\xi);\xi) + \mathrm{y}_0; \alpha),$$

with $\mathrm{h}^*(\alpha, \xi)$ such that:

$$\mathrm{h}^*(\alpha,\xi) = \mathrm{h}(\mathrm{w}^{\mathrm{v}}(\mathrm{h}^*(\alpha,\xi);\xi), \mu^{\mathrm{v}}(\mathrm{h}^*(\alpha,\xi);\xi) + \mathrm{y}_0; \alpha) \text{ and}$$

$$\mathrm{U}(\bar{\mathrm{c}}(\mathrm{f}), \bar{\mathrm{h}}; \alpha) \geq \mathrm{U}(\mathrm{y}_0, 0; \alpha),$$

where we have extended our notation to take into account the dependence on the error terms in α, ξ and f and where w^{v} and μ^{v} are short notations for ω(.,.) and μ(.,.) defined as the solution of (3) and (4). Then, $\mathrm{h}^*(\mathrm{w}^{\mathrm{v}}(\mathrm{h}^*(\alpha,\xi);\xi), \mu^{\mathrm{v}}(\mathrm{h}^*(\alpha,\xi);\xi)\ ;\alpha)$ is obtained from (2). The consumption level in the primary sector, $\bar{\mathrm{c}}(\mathrm{f})$, corresponds to the fixed cost f and the worked hours rationed at $\bar{\mathrm{h}}$. Then, $\bar{\mathrm{c}}(\mathrm{f}) = \mathrm{w}\ \bar{\mathrm{h}} + \mathrm{y}_0 - \mathrm{f}$. Given a value of α, the second inequality gives us an upper bound for the fixed cost, $\mathrm{f}_1^{\mathrm{M}}(\alpha)$. This quantity does not depend on the scale of production described by ξ. Provided the Slutsky condition is satisfied, this bound is unique. Therefore, for any value of the fixed cost less than $\mathrm{f}_1^{\mathrm{M}}(\alpha)$ and any value of α, the upper bound for the production scale can be defined, $\xi_1^{\mathrm{M}}(\alpha, \mathrm{f})$.

In the optimisation programme, the upper and lower bounds in the integrals (in this regime and the other ones) are numerically computed using a combination of bissection and Newton-Raphson method.

The probability to observe a primary sector worker is:

$$(12) \qquad L_1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{f_1^M(\alpha)} \phi_2\left[\frac{e_\alpha}{\sigma_\alpha}, \frac{e_f}{\sigma_f}, \rho_{\alpha,f}\right] \Phi\left[\frac{e_{\xi 1}^M - [e_\alpha, e_f]\Sigma_{11}^{-1}\Sigma_{12}}{\sqrt{\sigma_\xi^2 - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}}\right] \frac{1}{\sigma_\alpha \sigma_f} df\, d\alpha,$$

where $e_\alpha = \alpha - \overline{\alpha}$, $\quad e_f = f - \overline{f}$, $\quad e_{\xi 1}^M = \xi_1^M(\alpha, f) - \overline{\xi}$.

The parameters with an upper bar denote the linear combination of the exogenous variables and of the coefficients to estimate.

Φ (respectively ϕ) is the standard normal cumulative distribution function (respectively probability density function) and $\phi_2[., ., \rho]$ is the bivariate standard normal probability density function. The matrices Σ_{11}, Σ_{12}, Σ_{21} are relevant sub-matrices of Σ corresponding to well-known results on the conditional distribution of multivariate normal random variables. We now turn to the regime corresponding to the secondary sector workers.

Secondary Sector Workers

For the secondary sector workers, we assume that the earnings function depends non-linearly on the labour supplied, which is endogenous to our model. The description of the probability for this regime follows a similar approach to that of the previous regime. The following equations characterise a self-employed worker in the secondary sector:

$$(13) \qquad \Psi(w^v(h^*(\alpha,\xi);\xi), \mu^v(h^*(\alpha,\xi);\xi + y_0;\alpha) \geq U(\overline{c}(f), \overline{h};\alpha),$$

$$h^*(\alpha,\xi) = h(w^v(h^*(\alpha,\xi);\xi), \mu^v(h^*(\alpha,\xi);\xi);\alpha),$$

$R^*(\alpha,\xi) = R(h^*(\alpha,\xi);\xi)$ defines the activity income in the secondary sector.

Given the values of α and ξ that solve the latter two equations, the inequality enables us to define a unique lower bound for the value of the fixed cost, $f_2^m(\alpha, \xi)$. The contribution to the likelihood is

$$(14) \qquad L_2 = \Phi\left[-\frac{e_{f,2}^m - [e_\alpha, e_\xi]\Sigma_{11}^{-1}\Sigma_{12}}{\sqrt{\sigma_f^2 - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}}\right] \phi_2\left[\frac{e_\alpha}{\sigma_\alpha}, \frac{e_\xi}{\sigma_\xi}, \rho_\alpha, \xi\right] J^{-1},$$

where $e_\alpha = \alpha - \bar{\alpha}$, $e_\xi = \xi - \bar{\xi}$, $e^m_{f,2} = f^m_2(\alpha,\xi) - \bar{f}$,

and where J is the determinant of the Jacobian matrix of the transformation of the error terms in α and ξ into the observed values h^* and R^*. For other regimes, the Jacobian is simple enough to be directly incorporated in the formulae of the contributions to the likelihood. While the sub-matrices Σ_{ij} (i = 1; j = 1, 2) are different from their counterparts of the previous section, they are defined with the same convention where i corresponds to the line number of the block and j to its column number.

$$(15)\qquad J = \frac{\Delta^2 h^* (1+\delta h^*)^2}{\Delta(1+\delta h^*)((1+\delta h^*)(1+\gamma) - \beta\delta(2+\delta h^*)) - \Delta\gamma + \delta\gamma y_0 (2+\delta h^*)},$$

with $\Delta = e^{\xi + \delta h^*}$.

The likelihood of the self-employed workers can be slightly changed to account for the fact that the observed revenue lies inside a given range. Since the scale of the revenue function lies in a range $[\xi^m, \xi^M]$, the likelihood L_{2b} is such that

$$(16)\qquad L_{2b} = \int_{\xi^m}^{\xi^M} \Phi\left[-\frac{e^m_{f,2} - [e_\alpha, e_\xi]\Sigma_{11}^{-1}\Sigma_{12}}{\sqrt{\sigma_f^2 - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}} \right] \phi_2\left[\frac{e_\alpha}{\sigma_\alpha}, \frac{e_\xi}{\sigma_\xi}, \rho_{\alpha\xi} \right] J_{2b}^{-1}\, d\xi,$$

where J_{2b} results from the change in variables from α to h^*.

<u>Non-Participants</u>

The regime is described by the following inequalities

$$(17)\qquad U(y_0, 0; \alpha) > U(\bar{c}(f), \bar{h}; \alpha) \text{ and } \omega(y_0, 0; \alpha) > w^v(0; \xi).$$

Given α, it is possible to define a lower bound for the fixed costs, $f_3^m(\alpha)$, and an upper bound for the scale of the revenue function, $\xi_3^M(\alpha)$. Therefore, the likelihood becomes:

$$(18)\qquad L_3 = \int_{-\infty}^{+\infty} \Phi_2\left[\frac{e^M_{\xi,3}/\sigma_\xi - \rho_{\alpha\xi}\, e_\alpha/\sigma_\alpha}{\sqrt{1-\rho^2_{\alpha,\xi}}}, -\frac{e^m_{fc,3}/\sigma_f - \rho_{\alpha,f}\, e_\alpha/\sigma_\alpha}{\sqrt{1-\rho_{\alpha,f}}}, \right.$$

$$\left. -\frac{\rho_{f,\xi} - \rho_{\alpha,f}\,\rho_{\alpha,\xi}}{\sqrt{1-\rho^2_{\alpha,f}}\sqrt{1-\rho^2_{\alpha,\xi}}} \right] \phi\left[\frac{e_\alpha}{\sigma_\alpha}\right] \frac{1}{\sigma_\alpha} d\alpha \quad \text{where} \quad e^m_{f,3} = f^m_3(\alpha) - \bar{f}, e^M_{\xi,3} = \xi^M_3(\alpha) - \bar{\xi} \quad \text{and}$$

with obvious notations.

Appendix 2

Table 6a: 2SLS estimates of the wage equations

	Primary sector 2SLS	Secondary sector 2SLS
Constant	-1.468 (0.05)	-5.731 (2.3)
Age	-0.057 (1.15)	0.0229 (0.8)
Years of education	0.140 (0.6)	0.268 (1.1)
Experience	0.113 (1.7)	0.049 (0.7)
Primary education	-0.295 (0.2)	1.127 (0.8)
Higher education	0.101 (0.06)	-0.161 (0.08)
Number of observations	188	108

t-values in parentheses. The instruments are the number of children, the number of children aged from 0 to 3, from 4 to 10, from 11 to 15, the education level of parents, the occupation of the father, the age and education of the husband. One could argue that children cannot be considered as exogenous if they are simultaneously determined with education and work decisions. Finally, family background is likely to be correlated with unobserved ability.

Table 6b: Instrumental regressions

Primary Sector	Education level	Primary Education	Superior Education	Experience
Intercept	10.258 (0.000)	0.169 (0.004)	0.0784 (0.227)	10.893 (0.00)
Husband's age	-0.0554 (0.008)	0.00232 (0.235)	-0.00652 (0.007)	-0.00311 (0.93)
Husband's education	0.295 (0.000)	-0.139 (0.006)	0.0313 (0.000)	-0.0352 (0.72)
Children 0-3 years	-0.399 (0.328)	0.0194 (0.613)	-0.0758 (0.106)	-1.848 (0.015)
Children 4-10 years	-0.618 (0.105)	0.0153 (0.669)	-0.0731 (0.095)	-0.116 (0.869)
Children 11-15 years	-0.734 (0.034)	0.0284 (0.379)	-0.0495 (0.210)	1.750 (0.006)
Total number of children	0.267 (0.283)	-0.0171 (0.464)	0.0314 (0.272)	0.368 (0.42)
Education of the father	0.161 (0.004)	-0.00583 (0.265)	0.0197 (0.002)	0.0291 (0.77)
Profession of the father	0.0129 (0.872)	0.0100 (0.184)	0.00683 (0.459)	-0.0699 (0.64)
R^2	0.294	0.101	0.250	0.165
F	9.70 (0.0000)	2.62 (0.0096)	7.77 (0.0000)	4.44 (0.0001)
Number of observations	195	195	195	195

Secondary Sector	Education level	Primary Education	Superior Education	Experience
Intercept	6.472 (0.000)	0.647 (0.000)	-0.0170 (0.649)	6.188 (0.001)
Husband's age	-0.706 (0.000)	0.00531 (0.031)	-0.00127 (0.162)	0.119 (0.006)
Husband's education	0.410 (0.000)	-0.0389 (0.000)	0.0134 (0.000)	-0.343 (0.055)
Children 0-3 years	0.153 (0.797)	-0.140 (0.073)	-0.0571 (0.051)	-0.375 (0.785)
Children 4-10 years	0.257 (0.596)	-0.0348 (0.582)	0.0222 (0.347)	1.335 (0.253)
Children 11-15 years	-0.213 (0.452)	0.00991 (0.789)	0.00832 (0.546)	1.574 (0.017)
Total number of children	-0.386 (0.249)	0.0770 (0.080)	-0.00127 (0.938)	-0.463 (0.552)
Education of the father	0.132 (0.150)	-0.00161 (0.893)	0.0198 (0.000)	0.124 (0.391)
Profession of the father	-0.0334 (0.709)	0.00550 (0.638)	-0.00630 (0.149)	-0.178 (0.001)
R^2	0.325	0.229	0.338	0.187
F (Prob > F)	6.98 (0.000)	3.80 (0.0006)	6.52 (0.000)	2.86 (0.0067)
Number of observations	111	111	111	111

P-values in parentheses. The values of the R^2 and F statistics shows that we are not in the case of weak instruments in the sense of Staiger and Stock (1997).

Table 7: Hausman Tests

Test	P-value for the wage or non-labour income equation in the informal sector	P-value for the wage or non-labour income equation in the formal sector
A	0.982	0.986
B	0.484	0.904
C	0.214	0.617
D	0.152	0.449
E	0.264	0.828
F	0.0006	0.0017

A,B, C, D, E are exogeneity tests for the wage equation. F is an exogeneity test of the non-labour equation.

A: Exogeneity tests of equations where the selectivity has been controled by including an inverse Mill's ratio. Then, selectivity and exogeneity can be simultaneously tested. Age is considered as exogenous. The possibly endogenous regressors are: education and experience and the Mills's ratio. The instruments are: household composition, education and activity sectors of the parents.

B: Idem than A with experience assumed exogenous.

C: Exogeneity test of education and experience without Mill's ratio.

D: Idem than C with experience assumed exogenous.

E: : Idem than A with household composition possibly endogenous.

F: Exogeneity test of the education and household composition variables in the non-labour equation The instruments are the characteristics of the two parents: education, experience, ethnic group, region of origin, ability to speak French, sector and activity, household size of the family in which they were raised.