Salaried employment and earnings in Indonesia: New evidence on the selection bias *

Margherita Comola †
Paris School of Economics

Luiz de Mello‡
OECD Economics Department

September 2011

Abstract

This paper uses household survey data to estimate the determinants of earnings in Indonesia, a country where non-salaried work is widespread and earnings data are available for salaried employees only. We deal with the selection bias by estimating a full-information maximum likelihood system of equations, where selection into the labour market is modelled in a multinomial setting. We find that some estimated parameters of the earnings equation differ from a binomial selection procedure *à la* Heckman (1979), in particular for those variables with the strongest impact on the selection into the different labour-market statuses. However, the estimated returns to education are unaffected, even when we deal with the endogeneity of educational attainments following Duflo (2001). Overall, our findings show that the choice of the selection rule affects the estimates of the earnings determinants in the Indonesian labour market.

Keywords: Indonesia; employment; earnings; multinomial selection
JEL codes: J21; J23; J31

*The authors are indebted to Esther Duflo for sharing her data.

†Corresponding author: Margherita Comola, Paris School of Economics, 48 bl. Jourdan 75014 Paris, France. Tel.: (+33-1) 43136303, Fax: (+33-1) 43136310, email: `comola@pse.ens.fr`

‡OECD Economics Department. Email: `luiz.demello@oecd.org`

1 Introduction

This paper uses data from the 2004 Indonesian labour market survey (*Sakernas*) to estimate the determinants of employment and earnings in Indonesia. Our aim is to explore how individual characteristics, such as age, place of residence and educational attainment, affect a worker's labour market status and earnings in a standard Mincerian setting. We need to deal with the fact that earnings data are available only for salaried workers in *Sakernas*, but not for the self employed and household workers, who account for the bulk of employment in Indonesia. It is well known that a truncated earnings distribution poses a problem for the estimation of employment and earnings equations. Conventional techniques, such as the Heckman (1979) binomial selection procedure, can be used to deal with this issue, but we argue that a binomial selection rule would be too simple to cover all relevant labour market outcomes in a segmented labour market, such as Indonesia's. Following Hill (1983), we recognize that individuals face three options in the Indonesian labour market: they may be inactive (*i.e.*, out of the labour force), they may work as wage-earners or they may work in non-salaried jobs. In this paper, we deal with the selection bias by estimating a full-information maximum likelihood system of equations, where wages are observed for wage-earners (salaried employees) and selection into different labour-market statuses is modelled in a multinomial choice setting.

Among the earnings determinants, education deserves particular attention. Educational attainment should be treated as endogenous, because it is likely to be correlated with unobserved variables, such as family background or innate ability. We instrument educational attainment in both the earnings and selection equations, which is to our knowledge an innovation in the empirical literature. We use the same identification strategy as Duflo (2001), which consists of instrumenting educational attainment using the information available on a large-scale government-sponsored school construction programme (*Sekolar Dasah IMPRES*), which was implemented between 1973-74 and 1978-79 and resulted in the construction of over 61 000 schools nation-wide. If the school construction programme is assumed to have no effect on wages other than to increase educational attainment, exposure to the programme (gauged by the intensity of school construction activity in an individual's district of birth and his/her age when the programme was launched) can be used to estimate the impact of educational attainment on wages. Duflo shows that individuals born in districts that benefited from the programme are more likely to stay longer at school and to earn more once joining the labour force.

We check whether or not estimates of the determinants of earnings in Indonesia change if we apply our maximum likelihood procedure, which models selection as multinomial choice, relative to an OLS regression, which ignores the selection bias, and a simpler (binomial) selection procedure *à la* Heckman (1979). We show that some estimates of the earnings determinants do differ across model specifications. The parameter estimates for the wage equation that differ the most are precisely those most relevant for individual selection into different labour-market statuses. For instance, living in rural areas is positively significant or insignificant in the OLS and binomial selection equations but has a strong negative effect on wages when the selection rule is multinomial. The finding that rural workers have a lower probability of working as salaried employees is consistent with our priors. Also, an interaction term between gender (female) and marital status (married) is insignificant in the binomial selection framework, while it is positively signed and significant under multinomial selection. This reflects the fact that married women have a much higher probability of being inactive than to work in non-salaried jobs, a characteristic of the Indonesian labour market that the binomial selection rule fails to capture.

As for educational attainment, we find that the multinomial selection procedure does not affect the estimated returns to education. This is true even when we deal with a possible endogeneity bias by instrumenting schooling in both the selection and wage equations. This result is in line with Duflo (2001) using a cross-section of male workers from the 1995 intercensal survey of Indonesia.[1] Our results are very similar whether or not we apply a multinomial selection procedure or treat educational attainment as endogenous: we estimate the returns to education at between 9 and 10.8%, while Duflo reports coefficients in the range of 6.8-10.6%.

Although our results suggest the absence of a selection bias as far as educational attainment is concerned, we find significant bias in the estimated coefficients of other important variables. We therefore conclude that a binomial setting is too crude an approximation of the selection process in the Indonesian labour market and that the choice of the selection rule affects the estimates of the salaried earnings determinants.

[1] Duflo recognized the existence of a selection problem, since wage-earners make up only 45% of her sample, and investigated the size of the bias using two alternative procedures, which nevertheless do not model selection explicitly: she first conditions the probability of selection in the second stage given the instruments (Heckman and Hotz, 1989) and then imputes the income of self-employed individuals to include them in the sample. In both cases, she does not find any significant impact of selection on the estimated coefficients on education.

The paper is organized as follows. Section 2 is devoted to the literature review, Section 3 describes the data, the methodology and the variables definition, and Section 4 reports the empirical findings. Conclusions are presented in Section 5.

2 Literature review

This paper follows the empirical literature on the estimation of Mincerian wage equation (Mincer, 1974) to gauge the effect on earnings of individual characteristics, such as age, educational attainment and marital status, among others (Willis, 1987; Card, 1999; Heckman *et al.*, 2003). Several methodological extensions have been proposed to deal with the limitations of the conventional Mincerian model: for instance, Ichino and Winter Ebmer (1999) show how the choice of instruments affects the estimated returns to education, and Björklund and Kjellström (2002) discuss how well the schooling coefficients of standard Mincer equations approximate the rate of return to education. Empirical evidence is now available for a host of developing and emerging market countries, including Panama (Heckman and Hotz, 1986), Mexico (Brown, Pagan and Rodriguez Oreggia, 1999), Colombia (Gaston and Tenjo, 1992) and Brazil (Dickerson, Green and Arbache, 2001).

A seminal extension to the empirical literature is the Heckman selection model, which deals with truncations in the earnings distribution (Heckman, 1979). This is the case, for example, of the data used in this paper, where information on earnings is available only for salaried workers.[2] The literature has proposed alternative methods for dealing with multinomial selectivity, as in the case where labour market status cannot be described by just two alternatives. Different methods were developed by Lee (1983) and Dubin and McFadden (1984), and a non parametric alternative was proposed by Dahl (2002). These multinomial selection models have been applied in different settings, including the study of self selection into technical training (Trost and Lee, 1984), firm-size wage differentials (Brunello and Colusso, 1998) and the estimation of demand for electricity (Dubin and McFadden, 1984). We follow Dubin and McFadden's step modelling selection into the labour market in a multinomial manner, which we estimate jointly with the wage equation in a full-information maximum likelihood setting. Our selection is similar in spirit to the one proposed by Pradhan and van Soest (1995). However, they estimate two wage equations (for formal and informal workers)

[2] In this case OLS estimates are inconsistent and propensity score matching (Rosenbaum and Rubin, 1984) is not a valid alternative.

with the aim of comparing two selectivity models: ordered probit and multinomial logit *à la* Lee (1983), while we estimate only one earnings equation for wage-earners to compare our full-information maximum likelihood procedure with OLS and the two-step procedure proposed by Heckman (1979). Our strategy also allows us to address the endogeneity of education, both in the multinomial selection and the earning equations. Few methodological papers have dealt specifically with the issue of regressors' endogeneity in sample selection models, namely Das, Newey and Vella (2003) in a non-parametric context, and Kim (2006) for a common endogenous dummy.

While the empirical literature on employment is relatively rich for Indonesia (Lim, 1997; Islam and Nazara, 2000; Suryadarma, Suryahadi and Sumarto, 2007; Islam and Chowdhury, 2007), evidence is considerably more limited on the determinants of earnings. Among the few contributions available to date, Pirmana (2006) uses four waves of *Sakernas* to estimate earning differentials among groups of workers. He concludes that socio demographic factors, human capital and place of residence are powerful determinants of individual earnings, and that only 42% of the earnings differential between males and females is caused by differences in individual characteristics. Suryahadi, Sumarto and Maxwell (2001) use a panel of *Sakenas* data from 1988 to 2000 to gauge the impact of changes in the minimum wage on earnings and employment, and find that the elasticity of average wages with respect to the minimum wage is positive but statistically insignificant. Skoufias and Suryahadi (1999) use a synthetic cohort approach and show that the decline in real wages induced by the financial crisis of 1997-98 was evenly distributed across cohorts, while the impact of the crisis on wage inequality within cohorts was mixed. Deolalikar (1993) uses National Socio Economic Survey (*Susenas*) data to estimate a wage equation and the returns to schooling for different groups. His approach is comparable to ours in that he acknowledges the problem of selectivity. But he deals with it on the basis of a dichotomic selection rule (*i.e.* individuals may work as wage earners or not), while we argue that a multinomial selection is more appropriate.

3 Data, methodology and variable definition

3.1 The data

We use the 2004 wave of data from the Indonesian National Labour Force Survey (*Sakernas*), which started to be collected in 1976 and is currently carried out on an annual basis. In the 2004 wave 75 371 households (237 290 individuals) were surveyed.

Data on earnings and employment are reported in *Sakernas* as follows. Each family member belonging to the working age population (those aged 15 years and above) is classified as employed or unemployed depending on his/her activities during the previous week. Employed individuals are classified as wage-earners (salaried workers), self employed (with or without assistance) or unpaid/family/casual workers. While *Sakernas* data are overall considered to be of good quality, earnings data are collected for employees only, thus excluding the large number of workers.

Table 1 reports labour force participation, employment, unemployment and the incidence of non-salaried work for 2004. Labour force participation is about two-thirds for individuals aged at least 15 years. It is higher in rural areas and for males, and tends to rise with educational attainment. Employment patterns are comparable to those of labour supply: it is higher for males, residents in rural areas and among prime age individuals. Unemployment is particularly high for youths and, somewhat surprisingly, for workers with upper-secondary and tertiary education, who would otherwise be best equipped to work as salaried employees. When faced with a job loss, these individuals may prefer to wait for another salaried employment opportunity, instead of changing their labour-market status, so long as they can support themselves and their families in the meantime (queuing unemployment). Non-salaried work, including self employment (own account workers, with or without assistance) and unpaid/casual/household work, accounted for about 70% of employment in 2004 and is more widespread among women than men, workers living in rural than urban areas, and among older individuals.[3] The incidence of non-salaried work declines with educational attainment.

[3] The (already high) estimates of non-salaried work may in fact be biased downwards, as individuals working independently as non-wage earners may define themselves as employees.

Table 1: Labour force participation, unemployment, employment and the incidence of non-salaried work, 2004
(In per cent, individuals aged 15 years and above)

	Participation	Employment	Unemployment	Non-salaried work
Total	**65**	**60.7**	**6.7**	**69.6**
By gender				
Males	83.5	78.6	5.8	67.7
Females	46.7	42.9	8.2	72.9
By age				
15-24	50	39	22.1	58.8
25-54	74.2	71.8	3.2	68.5
55-64	63.5	63.1	0.6	88.4
65+	39.7	39.6	0.2	95.5
By residence				
Rural	69.8	67.1	3.9	86.3
Urban	60.1	54.2	9.9	48.7
By education				
No schooling	63.5	62.8	1.2	92.2
Primary	66.6	64.9	2.6	84.4
Lower secondary	55.9	51.7	7.5	72.2
Upper secondary	68.9	58.7	14.8	41.0
Tertiary	85.3	77.3	9.4	15.9

Note: non-salaried work is expressed in % of employment, and the unemployment rate is defined as the number of unemployed individuals divided by number of unemployed and working individuals.
Source: Sakernas

3.2 The methodology

Because earnings data are available only for wage-earners, estimation of a Mincerian equation by OLS would produce biased estimators if, as expected, selection into different job market statuses were correlated with potential determinants of earnings. In an influential paper, Heckman (1979) proposes a two step procedure based on the non selection hazard ratio (*i.e.*, the ratio of the probability density function over the cumulative density function of a distribution) to obtain consistent estimators in the presence of dichotomous sample selection. Analogous results can be obtained by jointly estimating the selection and the earnings equations by full-information maximum likelihood. For a recent review of the available parametric and semi-parametric estimation methods to address sample selection issues see Lee (2003).[4]

This paper aims to compare estimates of wage determinants using a multinomial selection rule against those obtained by OLS and under binomial selection. To ensure comparability, a full-information maximum likelihood (FIML) technique is used to estimate three models: a single continuous-variable earnings equation; a two-equation system for the binomial selection model, including a wage equation with a continuous censored dependent variable and a selection equation with a binomial dependent variable; and a multiple-equation system for the multinomial model, including a wage equation with a continuous censored dependent variable and separate equations for each alternative labour-market status.

The multinomial selection model, where individuals can chose among M alternatives, can be defined as:

$$\begin{aligned} y_1 &= x\beta_1 + \epsilon_1 \\ y_s^* &= z_s\gamma_s + v_s \end{aligned} \tag{1}$$

where $s = 1, ..., M$ and the wage outcome y_1^* is observed if and only if $y_1^* = max_{j \neq 1} y_j^*$, so that category 1 (salaried work) is chosen. As shown by Mac Fadden (1973), under the Independence of Irrelevant Alternatives (IIA) hypothesis, Equation (1) reduces to a multinomial logit model.[5] We estimate the two equations for y_1 and y_s^* jointly to take account of the correlation between the error terms, which is equivalent to estimating a recursive system of generalized linear models with a Gaussian error distribution: in the ML-based seemingly unrelated regression model (SUR), all equations are independent, but the underlying errors are jointly normally distributed. In the multinomial selection equation, each choice other than the base alternative $s = 2, ..., M$ is represented separately by an equation. Since multinomial choice depends on the same set of regressors for all alternatives, we have to impose the IIA condition through constraints on the covariance among the errors of the $M - 1$ equations

[4]Sample-selection models can be also used to estimate counterfactual outcomes and opportunity costs for forgone choices (in our case, the earnings of non-salaried workers). A welfare analysis of this kind is out of the scope of our paper, but we remand to Lee (2003) for a discussion of how matching techniques can be used to evaluate counterfactual outcomes in self-selected samples.

[5]See Bourguignon, Fournier and Gurgand (2007) for a survey of the available methods to obtain consistent estimates of β_s and γ_s with a two-step procedure.

representing the selection alternatives.[6]

In some specifications we also address the problem of endogeneity of educational attainment by instrumenting the individual's years of schooling, which is a right hand-side variable in both the wage and the selection equations, by the number of new schools built in his/her district of birth between 1973-74 and 1978-79. To do so, we add a reduced-form equation to the FIML system(s) with years of schooling as the dependent variable and all exogenous variables and the instrument as regressors. Since the instrumentation strategy imposes recursiveness, in this case only the second-step coefficients are structural (limited-information maximum likelihood).

3.3 Definition of the variables

Under binomial selection, individuals are either employed as salaried workers (and hence we observe their wages), or they are not. In the multinomial selection framework, we assume that workers can select themselves into three labour-market statuses: inactivity, employment as a wage-earner and non-salaried work.[7] The set of exogenous explanatory variables is the same for both selection rules (binomial and multinomial) and includes: *age*, *age squared*, a place of residence dummy (*rural*), a gender dummy (*female*) and a marital status dummy (*married*). We also include an interaction term (*female*married*) and the *dependency ratio* (computed as the number of household members who are younger than 15 or older than 65 divided by the number of household members aged 15-65) on its own and interacted with gender (*female*dependency ratio*). Educational attainment is measured as years of schooling.[8] Finally, we control for the average years of schooling of the other adult household members, which proxies for an individual's socio-economic background. Provincial dummies (the omitted province is Aceh) are included in all regressions.

[6] For a detailed description of the required IIA parameterization see Roodman (2009).

[7] For both selection rules, individuals who are currently unemployed and actively looking for a job (around 7% of the overall population) are excluded from the sample.

[8] Because *Sakernas* reports the highest educational qualification achieved, we transformed the reported achievements into the minimum number of years required to obtain the corresponding qualification in Indonesia. For instance, primary educational attainment is coded as 6 years of schooling, while Diploma III (which corresponds to a Bachelors' degree) corresponds to 15 years of schooling. We assigned a score of 3 to those individuals who declared to have started but not finished primary education.

Following the practice of using household composition variables as identifying instruments, we exclude the dependency ratio and its interaction with the gender dummy from the earnings equation. All the other regressors in the selection equation also appear in the earnings equation. As an additional robustness check, we control for the worker's sector of activity (agriculture, manufacturing or services, with trade as the omitted category) in Table 3, columns 7 to 9.[9]

To deal with the endogeneity of educational attainment, we instrument *years of education* by exposure to *Sekolah Dasar INPRES*, measured as the intensity of school construction in an individual's district of birth and his/her age when the programme was launched (Table 3, columns 4 to 9). Following Duflo (2001), we define district-level *programme intensity* as the number of schools built in a district between 1973-74 and 1978-79 divided by the number of children aged 5-14 years living in that district in 1971 (in thousands). Since most Indonesian children attend primary school between the age of 6 and 12, we assume that children benefit from the construction of schools only if they are aged 11 or less at the time the school is built. Therefore, our instrument *programme exposure* is equal to *programme intensity* in the individual's district of birth if he/she was aged 11 or less in 1974, and zero otherwise.[10] Duflo shows that this instrument has good explanatory power and that individuals born in districts that benefited more from the program were more likely to stay longer at school and to earn more once joining the labour force.

The descriptive statistics are reported in Table 2.

[9]The original classification follows the ISIC rev. 3 codes. We aggregate all sectors under four macro labels: agriculture (agriculture, fishing), manufacturing (mining, manufacturing, electricity, construction), trade (trade, hotels, transports) and services (finance, real estate, government, education, health, other services).

[10]Although it is not obvious to assume that the district of residence is also the district where pupils attend primary school, Duflo reports that 91.5% children surveyed in the Indonesian Family Life Survey were still living in the district of birth at age 12.

Table 2: Descriptive Statistics

variable	obs.	mean	min.	max.	s.d.
Log. hourly wage	38551	8.24	4.83	13	0.74
rural	189605	0.51	0	1	0.5
age	189605	35	15	65	13.1
age squared	189605	1400	225	4225	999.54
female	189605	0.5	0	1	0.5
married	189605	0.69	0	1	0.46
dependency ratio	189605	0.33	0	5	0.3
years of education	189605	7.96	0	16	3.64
household education	189605	7.84	0	16	3.44
sector: agriculture	123043	0.4	0	1	0.49
sector: manifacture	123043	0.17	0	1	0.38
sector: services	123043	0.15	0	1	0.35
programme intensity	181483	2.01	0.59	8.6	1.12
programme exposure	183493	1.37	0	8.6	1.33

Source: Sakernas

4 Results

4.1 The determinants of earnings

The results of the estimation of a Mincerian wage equation for 2004 are reported in Table 3. The sample includes all individuals aged 15-65 years who worked at least one hour as salaried workers in the previous week. The dependent variable is the logarithm of hourly wages.[11] Nine different specifications are reported: educational attainment is treated as exogenous in the first set of results (Table 3, columns 1 to 3) and is instrumented by *programme exposure* in the second set of results (columns 4 to 6). In the third set of results (columns 7 to 9) we instrument educational attainment and we control for workers' sector of activity (columns 7 to 9). For each set of results, three specifications are presented: OLS, which ignores the selection bias (column 1, 4 and 7); binomial selection, where inactivity and non-salaried work fall in the same category (column 2, 5 and 8); and the multinomial selection process described above with three different outcomes: salaried work, non-salaried work and inactivity (column 3, 6, 9).

[11]Respondents are asked the number of hours worked during the previous week and their average monthly wage as employees. For those employees who are temporarily out of work at the time the survey is conducted, the number of hours worked in the previous week is computed as the mean of the sample distribution.

The results of the instrumenting equation for columns 4 to 9 (*i.e.*, the regression of *years of schooling* on *programme exposure* and the other controls) are not reported to economize on space, but they are available upon request. We find that the instrument has good explanatory power, with point estimates very similar to those found by Duflo (2001) using a different dataset. For instance, for each new school built per 1000 children we found an average increase of 0.11 years of education, while Duflo finds a coefficient of 0.15.

The results of Table 3 show that most parameter estimates (including returns to education) are consistent regardless of the selection correction, but some other coefficients differ a great deal across specifications. For instance, the *rural* dummy is positive signed or insignificant in the OLS and binomial selection specifications, while it is negative and highly significant under multinomial selection, which takes into account the fact that salaried work is very infrequent in rural areas. Likewise, the interaction *female**married* is insignificant under binomial selection, but positive and significant under multinomial selection. The magnitude of the estimated coefficient on the interaction term suggests that being married, which yields a wage premium, offsets in part the negative effect of being female, which is probably related to the fact that very few married women work as salaried workers.[12] It is worth noticing that the regressors whose estimated effects on wage vary the most across specifications are the ones that have the strongest impact on multinomial selection into the labour market: *rural*, *married* and *female**married* (see below). These variables help to discriminate among those two statuses (non-salaried work and inactivity) that are treated as a single outcome under binomial selection and whose coefficients are therefore biased in the associated wage equation. These findings suggest that a binomial rule is too crude for describing selection in the Indonesian labour market, and that the choice of the selection rule affects the estimates of the earnings determinants.

All other coefficients are comparable in sign and magnitude across specifications, regardless of the selection rule chosen. For instance, wages rise with educational attainment and age (albeit for age in a non linear manner), and women are paid less than men. Socio-economic background, proxied by the average years of schooling of all other adult household members, is positively signed and significant, as expected. Moreover, all else equal, workers in trade are paid less than in the other sectors, while the highest wages are in manufacturing.

[12]In our sample, 52% of unmarried women and 54% of married women are inactive. On the other hand, 40% of unmarried men but only 4% of married man are out of the labour force.

Table 3. Wage equation, 2004 (Dep. Var.: Logarithm of hourly wage)

	(1) Minceriand	(2) Heckman selection	(3) Multinomial selection	(4) Mincerian	(5) Heckman selection	(6) Multinomial selection	(7) Mincerian	(8) Heckman selection	(9) Multinomial selection
rural	0.0333***	-0.0102	-0.0284***	0.0271**	-0.0199	-0.0390***	0.0240**	-0.0242*	-0.0422***
	(0.008)	(0.010)	(0.009)	(0.012)	(0.014)	(0.013)	(0.012)	(0.014)	(0.013)
age	0.0437***	0.0567***	0.0379***	0.0460***	0.0579***	0.0392***	0.0465***	0.0583***	0.0383***
	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
age squared	-0.0003***	-0.0005***	-0.0003***	-0.0004***	-0.0005***	-0.0003***	-0.0004***	-0.0005***	-0.0003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	-0.1996***	-0.2014***	-0.1717***	-0.2031***	-0.2060***	-0.1769***	-0.1990***	-0.2017***	-0.1726***
	(0.009)	(0.009)	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)
married	0.0788***	0.0837***	0.0527***	0.0724***	0.0833***	0.0524***	0.0627***	0.0765***	0.0439***
	(0.009)	(0.009)	(0.011)	(0.013)	(0.009)	(0.011)	(0.013)	(0.009)	(0.011)
female*married	0.0675***	-0.0251	0.0849***	0.0701***	-0.0308	0.0782***	0.0657***	-0.0414**	0.0762***
	(0.013)	(0.020)	(0.026)	(0.013)	(0.020)	(0.027)	(0.013)	(0.021)	(0.029)
years of education	0.1032***	0.1161***	0.1121***	0.0965***	0.1084***	0.1036***	0.0949***	0.1079***	0.1021***
	(0.001)	(0.002)	(0.002)	(0.010)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
household education	0.0064***	0.0066***	0.0103***	0.0096**	0.0103***	0.0144***	0.0115**	0.0120***	0.0162***
	(0.001)	(0.001)	(0.001)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
sector: agriculture							0.0598***	0.0545***	0.0510***
							(0.014)	(0.014)	(0.014)
sector: manifacturing							0.1613***	0.1611***	0.1619***
							(0.007)	(0.007)	(0.007)
sector: services							0.0753***	0.0735***	0.0800***
							(0.008)	(0.008)	(0.008)
Provincial dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Education instrumented	NO	NO	NO	YES	YES	YES	YES	YES	YES
Constant	6.0426***	5.5217***	5.9520***	6.0573***	5.5570***	5.9899***	5.9715***	5.4580***	5.9317***
	(0.043)	(0.099)	(0.102)	(0.048)	(0.105)	(0.108)	(0.047)	(0.109)	(0.116)
Observations	38551	38551	38551	38551	38551	38551	38551	38551	38551

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors in parentheses.

Note: Education attainment is instrumented using the product of programme intensity in the individuals' district of birth and a dummy equal to one if the individual was aged 11 or less in 1974, and 0 otherwise.

Our estimate for the returns to education ranges from 9 to 10.8%, which is comparable to the interval of 6.8-10.6% reported by Duflo (2001) on the basis of a cross-section of male workers from the 1995 Inter-Census Survey of Indonesia. The estimated coefficients do not change significantly whether educational attainment is instrumented or not, which underscores Duflo's finding that OLS coefficients do not appear to be biased upwards, as argued by Behrman (1990) in the context of developing countries. Also, our results do not change depending on the selection rule chosen. This is in line with the conclusions of Duflo (2001), who investigated the selection bias using two alternative procedures which do not model selection explicitly.

4.2 Selection into the labour market

The results of the selection equation(s) are reported in Table 4. The estimations carried out under binomial selection are reported in column (1), where the estimates refer to the probability of non-salaried work or inactivity (salaried work is the omitted category). Columns (2) and (3) report the multinomial selection results: column (2) refers to the probability of non-salaried work, and column (3) refers to the probability of inactivity (salaried work is the omitted category). In columns (4) to (6) educational attainment is instrumented as described above. Again, column (4) reports the binomial selection coefficients, while columns (5) and (6) refer to the multinomial selection equations.

The estimation results shed some light on the differences between non-salaried work and inactivity. The rural dummy is always positive in columns (1) to (3), but the magnitude of the effect is much bigger for non-salaried workers. This suggests that individuals living in rural areas, who are on average less educated but have a higher participation rate (from Table 1), are more likely to work in non-salaried jobs than being inactive and to work as salaried employees, an effect that is not captured by the binomial selection rule, which averages out non-salaried and inactive workers. However, when educational attainment is instrumented, the rural dummy for inactive workers under multinomial selection is not significant. The effect of age on labour-market status is, as expected, non linear. Older workers are more experienced and therefore more likely to work as salaried employees, although the effect is counterbalanced by a quadratic term, which is positively signed. Under the multinomial rule, the female dummy is negatively signed for non-salaried workers but positively signed for inactive individuals, and women tend to choose inactivity much more frequently than men (also from Table 1).

Marital status also matters. The married dummy is negatively signed under binominal selection, although married individuals are more likely to have non-salaried jobs and less likely to be inactive than single individuals under the multinomial rule. The combined sign and magnitude of the interaction terms suggests that married women have a slightly higher probability of having a non-salaried job than working as salaried workers and a much higher probability of being inactive.

Under multinomial selection, a higher dependency ratio seems to discourage workers from remaining inactive and to push them into non-salaried jobs, while the distinction is not captured under binomial selection. As for the interaction *female*dependency ratio*, females living in a household with a high dependency ratio are less likely to work as a salaried employee and more likely to be inactive than those living in a low dependency household. The effect is overall positive, but greater in magnitude for non-participants under multinomial selection. The finding is robust to instrumentation of years of schooling.

Educational attainment seems to be a powerful predictor of labour-market outcomes: an additional year of education decreases the probability of non-salaried work and inactivity with respect to salaried work across all specification, and the negative effect is more pronounced when educational attainment is instrumented (columns 4 to 6). Finally, the average years of schooling of the individual's household raises his/her probability to be inactive relative to having a salaried or non-salaried job. This seems to suggest that members of highly educated households tend not to accept low quality non-salaried jobs. The effect is stronger once the endogeneity of educational attainment is taken into account.

5 Conclusions

This paper used household survey (*Sakernas*) data to estimate the determinants of earnings in Indonesia. The Indonesian labour market is segmented, with a majority of workers engaged in non-salaried occupations, and earnings data are available only for salaried workers. This poses problems for the estimation of wage equations, because selection into different labour-market statuses is likely to be non random. In addition, correcting for this selection bias using a binomial rule would not be appropriate, because it would not encompass the different

Table 4: Selection Equation

	Heckman selection	Multinomial	selection	Heckman selection	Multinomial	selection
	(1) Non-salaried inactive workers	**(2)** Non-salaried workers	**(3)** Inactive workers	**(4)** Non-salaried inactive workers	**(5)** Non-salaried workers	**(6)** Inactive workers
rural	0.0862***	0.2175***	0.0291***	0.0812***	0.2118***	0.0150
	(0.002)	(0.004)	(0.005)	(0.003)	(0.005)	(0.009)
age	-0.0285***	-0.0099***	-0.1208***	-0.0280***	-0.0091***	-0.1191***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
age squared	0.0004***	0.0002***	0.0016***	0.0004***	0.0002***	0.0015***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	-0.0018	-0.1343***	0.0481***	-0.0044	-0.1377***	0.0405***
	(0.003)	(0.006)	(0.008)	(0.003)	(0.006)	(0.009)
married	-0.0137***	0.0410***	-0.3670***	-0.0140***	0.0408***	-0.3674***
	(0.003)	(0.006)	(0.006)	(0.003)	(0.006)	(0.006)
female*married	0.1696***	0.1797***	0.6838***	0.1674***	0.1767***	0.6800***
	(0.003)	(0.006)	(0.005)	(0.003)	(0.006)	(0.005)
dependency ratio	0.0282***	0.0230***	-0.0788***	0.0282***	0.0230***	-0.0788***
	(0.004)	(0.007)	(0.014)	(0.004)	(0.007)	(0.014)
female*dependency ratio	0.0111*	0.0441***	0.2217***	0.0122**	0.0454***	0.2246***
	(0.006)	(0.011)	(0.018)	(0.006)	(0.012)	(0.018)
years of education	-0.0277***	-0.0484***	-0.0542***	-0.0317***	-0.0534***	-0.0658***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.003)	(0.006)
household education	-0.0002	-0.0121***	0.0150***	0.0017*	-0.0010***	0.0206***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Provincial dummies	YES	YES	YES	YES	YES	YES
Education instrumented	NO	NO	NO	YES	YES	YES
Observations	189605	189605	189605	189605	189605	189605

*** p<0.01, ** p<0.05, * p<0.1

Standard errors in parentheses

labour market outcomes, which include not only salaried and non-salaried work, but also the possibility that individuals may drop out of the labour force altogether. An additional empirical problem we faced is related to the likely endogeneity of educational attainment in both the earnings and selection equations. Endogeneity arises from the fact that an individual's attainment correlates with unobservable variables, such as family background, which in turn affect earnings and selection into different labour market statuses. Endogeneity is often dealt with in earnings equations but not both in earnings and selection equations, as we do in this paper.

We propose an innovative procedure to deal with the selection bias by estimating a full-information maximum likelihood system of equations, where wages are observed for salaried employees only and selection into different labour-market statuses is modelled in a multinomial setting. The use of a FIML technique is to ensure comparability in the estimation of the three different models: simple OLS, two-equation binary selection and three-equation multinomial selection. We also deal with the endogeneity of educational attainment by instrumenting years of schooling in both the selection and earnings equations. We do so by adding a reduced-form equation to the FIML system with years of schooling as the dependent variable. Our identification strategy, which is the same as in Duflo (2001), is based on exposure to a large-scale school construction programme implemented in the 1970s.

Comparison of the results obtained under multinomial selection with those estimated by standard OLS, which ignores the selection bias, and a binomial procedure *à la* Heckman (1979) shows that several parameter estimates differ when multinomial selection is allowed in the estimation of the wage equation. In addition, the regressors whose estimated effects vary the most are those with the strongest impact on individual selection into different labour-market statuses. For instance, in absence of a multinomial selection rule the negative effect of living in rural areas on wages is underestimated. This finding has important implications for the design of public policies in that the disadvantage of living in rural areas, where informal work is widespread, may actually be stronger than previously believed when more complex selection mechanisms are taken into account.Another important finding is that the estimated returns to education do not seem to be effected by the selection bias. This is in line with the conclusions of Duflo (2001), who investigated the issue without explicitly modelling selection. Again, this finding is interesting for empirical analysis, because it allows future research to focus on novel identification strategies to deal with the endogeneity of education attainment in earnings equations independently of the complexity of underlying selection process, if any.

Overall, our findings show that the binomial selection rule does not properly account for the complexity of the Indonesian labour market, and that the choice of the selection rule affects the estimates of the earnings determinants. The main implication of our findings for future research on labour market issues in economies with large informal sectors, such as developing and emerging-market economies, is that the complexities arising from labour informality should not be underestimated.

References

Behrman, J.R. (1990), "The Action of Human Resources and Poverty on One Another: What We Have Yet to Learn." World Bank Living Standards Measurement Studies Working Paper No. 74.

Björklund, A. and C. Kjellström (2002), "Estimating the Return to Investments in Education: How Useful is the Standard Mincer Equation?", *Economics of Education Review*, Vol. 21, pp. 195-210.

Brown, C., J. Pagan and E. Rodriguez Oreggia (1999), "Occupational Attainment and Gender Earnings Differentials in Mexico", *Industrial and Labor Relations Review*, Vol. 53, pp. 123-35.

Brunello, G. and A. Colusso (1998), "The employer size-wage effect: evidence from Italy", *Labour Economics*, Vol. 5, pp. 217-230.

Bourguignon, F., M. Fournier and M. Gurgand (2007), "Selection Bias Correction Based on the Multinomial Logit Model: Monte Carlo Comparisons", *Journal of Economic Surveys*, Vol. 21, pp. 174-205.

Card, D. (1999), "The Causal Effect of Education on Earnings", in O. Ashenfelter and D. Card (eds.), *Handbook of Labour Economics*, Elsevier, Ch. 30, pp. 1801-63.

Das, M., W. Newey and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models", *Review of Economic Studies*, Vol. 70, pp. 33-58.

Dahl, G.B. (2002), "Mobility and the Returns to Education: Testing a Roy Model with Multiple Markets", *Econometrica*, Vol. 70, pp. 2367-2420.

Deolalikar, A. (1993), "Gender Differences in the Returns to Schooling and in School Enrolment Rates in Indonesia", *Journal of Human Resources*, Vol. 28, pp. 899-932.

Dickerson, A., F. Green and J.S. Arbache (2001), "Trade Liberalization and the Returns to Education: A Pseudo panel Approach", Working Paper, No. 0114, University of Kent, Canterbury.

Dubin, J.A. and D.L. McFadden (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", *Econometrica*, Vol. 52, pp. 345-62.

Duflo, E. (2001), "Schooling and Labour Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review*, Vol. 91, pp. 795-813.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, pp. 153-61.

Heckman, J. and V.J. Hotz (1986), "An Investigation of the Labour Market Earnings of Panamanian Males Evaluating the Sources of Inequality", *Journal of Human Resources*, Vol. 21, pp. 507-42.

Heckman, J. and V.J. Hotz (1989), "Choosing Among Alternative Non Experimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training", *Journal of the American Statistical Association*, Vol. 84, pp. 862-74.

Heckman, J., J. Lance, J. Lochner and P. Todd (2003), "Fifty Years of Mincer Earnings Regressions", NBER Working Paper, No. 9732, NBER, Cambridge, MA.

Hill, H. (1983), "Female Labour Force Participation in Developing and Developed Countries - Consideration of the Informal Sector", *Review of Economics and Statistics*, Vol. 65, pp. 459-68.

Ichino, A. and R. Winter Ebmer (1999), "Lower and Upper Bounds of Returns to Schooling: An Exercise in IV Estimation with Different Instruments", *European Economic Review*, Vol. 43, pp. 889-901.

Islam, A. and S. Nazara (2000), "Estimating Employment Elasticity for the Indonesian Economy", International Labour Organization (ILO), Jakarta.

Islam, A. and A. Chowdhury (2007), "Indonesia's Employment Challenges - Growth, Structural Change and Labour Market Rigidity", Report for the International Labour Organisation (ILO), Geneva.

Kim, K. (2006), "Sample Selection Models with a Common Dummy Endogenous Regressor in Simultaneous Equations: A Simple Two-Step Estimation", *Economic Letters*, Vol. 91, pp. 80-86.

Jaffe, A.J. and K. Azumi (1960), "The Birth Rate and Cottage Industry in Underdeveloped Countries", *Economic Development and Cultural Change*, Vol. 9, pp. 52-63.

Lee, L. (1983), "Generalized Econometric Models with Selectivity", *Econometrica*, Vol. 51, pp. 507-12.

Lee, L. (2003), "Self-Selection" in Baltagi (ed.), *A Companion to Theoretical Econometrics*, ch. 18.

Lim, D. (1997), "Forecasting Employment Growth in Indonesia", *Bulletin of Indonesian Economic Studies*, Vol. 33, pp. 111-19.

McFadden D.L. (1973), "Conditional Logit Analysis of Qualitative Choice Behaviour", in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press.

Mincer, J. (1974), "*Schooling, Experience, and Earnings*", NBER, Columbia University Press, New York, N.Y.

Pirmana, V. (2006), "Earnings Differential between Male Female in Indonesia: Evidence from Sakernas Data", Padjadjaran University Working Paper, Jakarta.

Pradhan, M. and A. van Soest (1995), "Formal and informal sector employment in urban areas of Bolivia", *Labour Economics*, vol. 2, pp. 275-297.

Roodman, D. (2009), "Estimating fully observed recursive mixed-process models with cmp", Working Paper 168. Center for Global Development. Washington, DC.

Rosenbaum, P. and D. Rubin (1984), "Reducing Bias in Observational Studies Using

Subclassification on the Propensity Score," *Journal of the American Statistical Association* vol. 79, pp. 516-524.

Suryadarma, D., A. Suryahadi and S. Sumarto (2007), "Reducing Unemployment in Indonesia: Results from a Growth Employment Elasticity Model", SMERU Working Paper, Jakarta.

Suryahadi, A., S. Sumarto and J. Maxwell (2001), "Wage and Employment Effects of Minimum Wage Policy in the Indonesian Urban Labour Market", SMERU Working Paper, SMERU, Jakarta.

Suryahadi, A., W. Widyanti, D. Perwira and S. Sumarto (2003), "Minimum Wage Policy and Its Impact on Employment in the Urban Formal Sector", *Bulletin of Indonesian Economic Studies*, Vol. 39, pp. 29-50.

Skoufias, E. and A. Suryahadi (1999), "Growth and Crisis Impacts on Formal Sector Wages in Indonesia", SMERU Working Paper, SMERU, Jakarta.

Trost, R. and L. Lee (1984), "Technical Training and Earnings: A Polychotomous Choice Model with Selectivity", *Review of Economics and Statistics*, Vol. 66, pp. 151-156.

Willis, R. (1987), "Wage Determinants: A survey and Reinterpretation of Human Capital Earnings Functions", in O. Ashenfelter, R. Layard and D. Card (eds.), *Handbook of Labour Economics*, Elsevier, Ch. 10, pp. 525-602.