Società italiana di
economia pubblica

XIX CONFERENZA

# DROPPING THE BOOKS AND WORKING OFF THE BOOKS

RITA CAPPARIELLO AND ROBERTA ZIZZA

# DROPPING THE BOOKS AND
# WORKING OFF THE BOOKS

Rita Cappariello* and Roberta Zizza*

## Abstract

This paper empirically tests the relationship between underground labour and schooling achievement for Italy, a country characterised by both a huge size of the shadow sector and a bad ranking in terms of education when compared to other high-income economies. We exploit in particular the information on individuals' social security positions available from the Bank of Italy's Survey on Household Income and Wealth to identify the underground workers. We show, after controlling for a wide range of socio-demographic and economic variables and addressing potential endogeneity and selection issues, that having a low level of education sizeably and significantly increases the probability of working underground. Switching from completing the compulsory school to graduating at college more than halves this probability for both men and women. The gain is slightly higher when one completes the compulsory track with respect to those with no formal education at all.

JEL classification: I21, J24, O17, R23.

Keywords: irregular employment, underground economy, occupational choice, education, school drop-out.

## 1. Motivation of the paper and literature review

It is widely recognized that the size of the Italian underground labour, namely labour which is deliberately concealed from the public authorities to avoid payment of taxes or complying with regulations (OECD (2002), is huge, especially when compared to other industrialised economies. In 2004, according to official estimates, it accounted for 11.5 per cent of the total input of labour, measured in terms of full time equivalent units; the vast majority of EU15 members report figures close to 5 per cent (European Commission, 2004). Moreover, Italy is characterised by a pronounced duality between the Northern and the Southern parts of the country that strikingly emerges from regional estimates of the underground labour: in the same year irregular labour ranged from 8 per cent of total employment in Lombardia (in the North) to 33 per cent in Calabria (in the South). This geographical divide stands at odds with a substantial homogeneity, at least de jure, in terms of tax and audit/punishment policies, burden of regulation and bureaucracy, the standard arguments the literature on noncompliance puts forward to explain *cross-country* differences in the extent of the underground sector (Schneider and Enste, 2000).

Also in terms of education Italy ranks badly with respect to other high-income economies. In 2005 only 37.5 per cent of the population aged between 25 and 64 years completed high school, about 8 percentage points less than the OECD average. The gap was even wider for the share of those having a college degree (12 per cent, about a half the OECD average). Only 76 teenagers out of 100 get a diploma, one of the lowest values among advanced economies (OECD, 2006). Data from the 2001 Population Census confirm the duality of the country in education achievement of the labour force: the Southern regions are lagging behind according to many indicators of schooling (e.g. enrolment rates in the compulsory and secondary school, percentage of people with no formal education; Istat, 2005).

Several studies investigate the relationship between underground labour and educational attainment in different countries from an empirical standpoint. Most of them are based on household surveys, a source traditionally considered more appropriate than firm level data for measuring the extent of informal labour. In fact, a surplus of labour input derived from surveys of households over that from

enterprise sources is considered an indication of non observed labour (OECD, 2002).

For the Spanish labour market Ahn and de la Rica (1997) find that a higher level of education increases the probability of working in the regular sector. Moreover, the wage returns to education and tenure are much higher for those holding a formal job. Similar results are obtained by Gong and van Soest (2001) for the Mexican economy. In particular, they show that wage differentials between the formal and the informal sector increase with the education level.

Barth and Ognedal (2005), by proxying the supply of and the demand for unreported work respectively with the willingness to receive unreported income and the actual provision of unreported work, show for the Norwegian economy a negative effect on both demand and supply of irregular labour of different human capital variables. A study for Germany of the Rockwool Foundation (2005) finds that education exerts a significant negative effect on the probability of participating in black activities though only for men. The fact that workers engaged in shadow employment have, on average, lower educational attainments than regular workers and / or hold jobs requiring unskilled workers has been documented for Brazil (Boeri and Garibaldi, 2005) and for Bulgaria and Colombia (Peracchi *et al.*, 2007).

Boeri and Garibaldi (2002), exploiting an ad hoc survey carried out in Sicily, an Italian region where the underground economy is traditionally flourishing, show that irregular jobs involve mainly workers at the lower end of the skill distribution, as the proportion of workers with a primary or lower level of education is much larger in the shadow sector. They argue from these results, as well as from those emerging from a survey conducted by ISAE (ISAE, 2002), that shadow jobs are to a large extent low productivity jobs. These results have been later extended nationwide by the same authors (Boeri and Garibaldi, 2005).

In this paper we empirically test for Italy the relationship between engagement in shadow labour and schooling attainments. This can shed light on the role played by education in shaping the heterogeneity across regions in the diffusion of the irregular labour. Moreover, it can provide evidence in favour of the argument that a huge number of irregular firms hiring non-skilled workers, the low incentives for firms and households to invest in human capital and the lack of innovation are interconnected phenomena, able to undermine the development of an economic system (Scarlato, 2001; Carillo and Papagni, 2002; Carillo and Pugno, 2002).

We exploit two different sources of information. First, we rely upon official measures of hidden labour available at regional level; second, we use individual data from the Bank of Italy's Survey on Household Income and Wealth to identify the underground workers. To our knowledge only two papers provided

descriptive evidence based on the same survey (Brandolini and D'Alessio, 2002; Boeri and Garibaldi, 2005).

The rest of the paper is organised as follows. In Section 2 we discuss on a theoretical ground the relationship between informality and level of education. Section 3 presents the preliminary evidence based on macro data at regional level. In Section 4 we model the probability of working underground by relying on individual data. Section 5 concludes.

## 2. The theoretical framework

There are two competing views concerning the existence of informal labour. On the demand side, working in the underground sector can be regarded as the only option for individuals who cannot find a job in the formal sector. In presence of wage and non wage rigidities in the formal labour market, such as minimum wage and firing costs, the underground sector represents a 'free-entry sector of last resort' (Fields, 2005). On the supply side, a job in the informal sector could be voluntarily chosen by the worker as a better alternative to a formal job, trading social protections off for other dimensions of job quality (Maloney, 2004). The implied relationship between education and underground labour are clearly different in the two cases.

The first line of reasoning provides a simple explanation to the observation that less educated people tend to work in the underground economy. Under the assumption that education is a signal of an individual's ability, the presence of an informal sector allows individuals whose productivity is below the minimum wage to have a job. In areas characterised by a low level of education of the labour force firms may decide to increase revenues by evading taxes and eluding laws rather than by hiring more productive workers and adopting new technologies. One possible suggestion to the policy maker would be to introduce policies targeted at fostering education and improving entrepreneurship, as they would have the beneficial side effect of reducing the size of the underground economy. If instead working in the underground economy is a matter of supply, the implications of the education level on individual's probability of working underground are less clear. Although working in the underground sector prevents individuals from gaining access to some Social Security provisions, in particular sick leave and pension coverage after retirement, some of them may still prefer working in the informal sector if they can receive immediate monetary compensation for this renounce. This is likely to hold especially for the self-employed, irrespective of the level of education.

The framework is further complicated by the fact that the direction of causality may be the opposite. It might be the opportunity to work in the irregular sector

4

that influences the choice to acquire higher education. Kolm and Larsen (2003) assume that the effort put in increasing the level of schooling is worthwhile only for high ability workers, as the educational costs they pay are a decreasing function of their ability; these workers apply only for formal jobs. The other workers, referred to as "manual", choose not to attain higher education and face job opportunities in both the formal and the informal segments of the labour market.

Against this background, the following analysis attempt at assessing the existence, the magnitude and the direction of the link between education and informal labour.

## 3. Preliminary evidence based on regional data

Our first empirical test consists of regressing a measure of shadow employment (*IRR*) on an indicator of education (*EDUC*) and on some controls (the *X*s). Hidden labour regional estimates used in this section are the official ones, supplied by Istat within the framework of national accounts in order to ensure the exhaustiveness of GDP estimates. They are obtained by comparing several sources of information (such as surveys on firms and households, census, administrative data, etc.) and, after allowing for pertinent conceptual differences, giving economic meaning to discrepancies detected (Istat, 2006).

As we have data for twenty Italian regions in the time span ranging from 1995 to 2004, we can use panel regressions and include fixed effects for regions and years ($x_i$ and $z_t$, respectively), the former able to take into account structural characteristics in the economies at the local level that are stable over time and not captured by our control variables. Thus the basic equation to be estimated is:

$$[1] \qquad IRR_{it} = \alpha + \Sigma_j \beta_j X_{jit} + \gamma EDUC_{it} + x_i + z_t + e_{it.}$$

In principle several measures of schooling could be relevant. The average number of years spent at school; an enrolment rate, for example the ratio between boys and girls enrolled in the secondary school divided by the reference population (individuals between 14 and 18 years old); an indicator of dropout. In our view the use of flow data (such as years of schooling and dropouts) is to be preferred to the enrolment rates, as the latter tend to be quite inertial.

Due to data availability we have chosen the rate of dropout at the second year of the secondary school (*DROPOUT*). The rate of dropout at the first year (corresponding to an age of 14 according to the Italian education system) is also available, but since in Italy – in compliance with the ILO standards - labour is forbidden to those aged less than 15, we preferred to deal with a more consistent

measure, accounting for the education choices of those already eligible for employment.[1]

The set of control variables (see table 1 for details and descriptive statistics) includes: the rate of unemployment (UNEM), as a proxy for labour demand restrictions in the formal sector, and a measure of the efficiency of the judiciary system (ENF). An efficient functioning of the legal system can have a deterrence effect on perpetrating irregularities, as it introduces less uncertainty on the actual punishment in case a break of the law is detected. Moreover, per capita income (*GDPPC*), in order to take into account of different productivity at the regional level, and sectoral composition in terms of value added (*SHAG, SHCO, SHSE* respectively) have been included with a one year lag to get rid of circularity issues as the measure of irregular labour enters the estimate of GDP.[2]

Due to the complexity of the phenomenon it might have been reasonable to include other factors to explain the development of the underground labour. However, given the relatively short period of time considered, it is likely that most of the variables that could be relevant to our scopes are basically invariant over time (e.g. firm size, social capital), and hence captured quite well by the fixed effects. The availability of credit to firms, that could also potentially be blamed for resorting to irregular labour, has been shown not to be relevant for Italy (Gobbi and Zizza, 2007).

Dropping out of secondary school has a positive and significant impact on the extent of the irregular employment (table 2). Among the control variables, only *ENF*, *SHAG* and *SHCO* display coefficients which are statistically different from zero at standard confidence level and whose sign is conform to theory: in particular, the size of the underground labour is the bigger, the less efficient is the judiciary system. Results in terms of both sign and magnitude of the coefficient of *DROPOUT* are robust to the exclusion of control variables and of regional fixed effects.

As already hinted, the human capital variable in equation *[1]* is likely to be endogenous as it might be the opportunity to work in the irregular sector that influences the choice to acquire higher education. In regions characterised by a developed underground segment, incentives to drop out of school can be strong. So the causation can flow in the opposite direction. To address the issue of endogeneity one would be willing to adopt an instrumental variable strategy.

---

[1] Gobbi and Zizza (2007) use the same variable as an instrument for the underground labour, in turn supposed to be a determinant of the development of the credit market across the Italian regions.

[2] Besides, using the contemporaneous values of these variables does not change our results.

Finding a legitimate instrument is in general not easy. The task becomes even harder as we are dealing with geographical areas where institutions and rules in force (for example, the length of compulsory education) are in principle the same. One potentially appealing candidate instrument for *DROPOUT* seemed a variable related to culture which is not dependent on the level of income[3], such as the percentage of people having read at least a book in a reference year. This instrument however emerged as not relevant for *DROPOUT*. This "macro" exercise hence fails to address the issue of endogeneity (if any) of education choices, and in this regard cannot be considered as entirely satisfactory. An approach based on individual data seems more suitable to our scopes, and is then attempted in the next section.

## 4. Exploiting individual data: evidence from the Bank of Italy's Survey on Household Income and Wealth

### 4.1    Identification of the underground workers

The Bank of Italy carries out every two years the Survey on Household Income and Wealth (SHIW thereafter); the sample is composed of about 8000 households, representative of the whole Italian population (Banca d'Italia, 2006).

Using data on self-reported individuals' social security positions available from the SHIW it is possible to identify the underground workers. In the questionnaire it is asked whether the person itself, or one of his employers, has ever paid social security contributions. Quoting the SHIW questionnaire:

*"Considering the lifetime work experience of ..... (name) , did he/she ever pay, or his/her employer pay, pension contributions, even for a short period (and even if long ago)?"*

As the questionnaire is very comprehensive and replying is not compulsory, it is likely that respondents do not feel threatened or suspicious and would hence reply truthfully. However, misreporting cannot be ruled out. Apart from fear of being detected, there is in fact presumably a stigma associated with holding an irregular job. Many underground workers could be "hidden" among those who declared themselves as unemployed, job seekers or inactive at the moment of the interview. Boeri and Garibaldi (2002), using the Italian Labour Force Survey, estimate that about 45 per cent of those classified as unemployed and 10 per cent of those with an inactive status are actually involved in irregular activities.

---

[3] Income is arguably low for many of those engaged in the shadow sector. Books have usually affordable prices and in many cases can be borrowed for free from public libraries or friends. The same would not hold if we looked at data on cinema, theatre or museum attendance.

Due to the formulation of the question, if an individual replies negatively stating, at the same time, that he has been working, this means he has been working off the books *throughout his entire career*. Those who are currently irregular but have cumulated pension contributions in the past, or those whose main occupation is regular but hold an irregular second job are not labelled as underground workers. Against this background, this measure is conceivably on the conservative side, and we refer to it as a *narrow* estimate of irregular labour.

A further question deals with the number of years (or months) the individuals (or their employers) have been paying social security contributions up to the time of interview (YCONTR). By comparing this information with that on the number of years worked (i.e. with experience; EXPER) it would be possible to retrieve an indicator of social contribution evasion, that occurs when YCONTR < EXPER.[4] Unfortunately the SHIW does not ask exactly the number of years worked, but records both the age at the time of the interview (AGE) and the age when the interviewed started working (AGESTART). The difference between the two measures the potential experience (POTEXPER = AGE-AGESTART) and corresponds exactly to EXPER only if the worker has had a very regular work history. There can be several reasons for this not to be true, such as unemployment spells or on-leave periods; moreover, not necessarily leaving a job implies an interruption of the payment of social contributions, as it could be either the worker who voluntarily continues to pay them or there could be some forms of unemployment treatment providing also for the payment of contributions (Brandolini and D'Alessio, 2002).

Given this limitation to the data, some "tolerance" criteria to identify those who spent a fraction of their working life underground are needed. Given that we are interested to permanent effects of the low level of education we are willing to select only those individuals who spent a *significant* fraction of their working life underground, that is those who paid too little in comparison with their working experience. In particular, we label as underground workers those satisfying the following relative criterion[5]:

*[2]*        YCONTR / POTEXPER < p25

i.e. the pension coverage must be below a threshold given by the first quartile of the distribution of the coverage across individuals of the same gender reporting

---

[4] It might be also the case that YCONTR > EXPER, for example if the worker has made extra contributions referred to the years spent at the university or to the period in which he was doing the military service.

[5] To allow for some imprecision in the replies and since we are willing to identify only those whose working history has been heavily characterised by the irregularity status we also require the difference between POTEXPER and YCONTR to exceed ten years (absolute criterion).

the same number of jobs held. It is in fact conceivable that the lower the number of jobs held, the better the variable POTEXPER is proxying the number of years *actually* worked.[6] We take care of differences by gender as women are likely to present more discontinuous career profiles due to maternity and childcare.

Adding these partially underground workers to those included in the narrow measure we get a *broad* measure of underground labour. It is worthy noticing that even this broad measure is not comparable with that provided by Istat and used in Section 3, as it includes workers who were "chronically" underground in the past but are regular today and excludes those irregular today but possibly regular for most of their life. Despite this, we believe that our measure is suitable for the scope of investigating the long term consequences of schooling on labour status.

We pool the observations in the four surveys conducted every two years from 1998 to 2004, removing those who appeared more than once.[7] The sample is also restricted to those aged between 14 and 65. The incidence of irregular labour in selected groups, according to both narrow and broad measures, is reported in table 3a.

### 4.2    *Modelling the probability of working underground*

We model the probability of working underground through pooled probit regressions, where our dependent variable takes a value of one if the individual is underground, and zero if not. Being underground is set to depend on a wide range of socio-demographic and economic variables at the individual level provided by the SHIW. The characteristics of our sample by occupational status are summarized in table 3b. First of all, as this is the object of our analysis, the educational attainment. The SHIW asks for the level of qualification achieved, allowing for the following breakdown: no education (zero years), primary school (5 years), junior high school (8 years), vocational high school (11 years), general high school (13 years), short college degree (16 years), college degree (18 years), graduate studies (more than 20 years). As in Attanasio, Guiso and Jappelli (2002), we convert this information in terms of number of years of schooling, in particular

---

[6] For those having had only one employer, EXPER and POTEXPER should in principle be coincident. Even among them, however, we detect conspicuous differences between years worked and pension contributions.

[7] In this way we are in practice reducing ourselves to a cross-section framework. This is suitable to the kind of phenomenon we are investigating, as we are looking at long lasting consequences of education in terms of workers' permanence in the underground sector for a relevant fraction of their working history.

by considering the minimum number of years that it takes to achieve a certain qualification.[8]

Given that we are looking at workers who have been underground for all or most of their working life, the other explanatory variables have been selected among those reflecting invariant or at least highly permanent conditions (gender; province of residence; marital status; having children; kind of occupation; sector of economic activity) or among those accounting for the past working history (experience and its square; number of jobs). Most of these independent variables are pretty standard in empirical models attempting at describing working decisions. In our specific context, it is reasonable to assume that individuals who value more social security provisions are more likely to hold a job in the formal sector (Ahn and de la Rica, 1997). For example, being married or having children are likely to be associated with a stronger favour for a regular position.

The results of this baseline specification are the following (table 4, column A). The marginal effect of years spent at school on the probability of working underground is, as expected, negative and is strongly significant. Switching from the compulsory schooling level[9] to a college degree decreases this probability for men and women respectively by 11 and 14 percentage points, which compares to a sample average probability respectively equal to 16 and 19 per cent. Moving from no formal education to the compulsory schooling age lowers the risk of holding an irregular job by respectively 15 and 18 percentage points (picture 1). Men are on average less likely to work irregularly by 6 percentage points, other things equal; people who never got married are more likely by 7 percentage points. The effect of having children is positive but barely significant. The higher the experience, the higher the probability of being (or having been) underground, while the squared term has a negative sign; also having changed many jobs is positively related with the irregularity status. Employment in the informal sector is more likely among those who have had at least one experience as independent worker and among those employed in the tertiary and building sectors.

So far we have ignored the fact that pensions contributions are paid only by those working, and that participation in the labour market is usually not random. It might be the case that, especially in an advanced economy, very low skilled people are excluded from the labour market, even from the irregular segment. Simple probit estimates are thus likely to suffer from the sample selection bias

---

[8] The variable education is coded using the number of years given in parenthesis; for the highest class a value of 20 years is chosen. Our breakdown is more fine-grained with respect to that in Attanasio, Guiso and Jappelli (2002), including vocational high school and short college degree.

[9] Here and in the rest of the paper we will define as compulsory school the path including primary and junior high school. It is however important to note that since 1999, as anticipated in Section 3, the compulsory requirements have been extended further, up to the 15th birthday.

induced by the choice (or chance) to enter the labour market, either with a regular or an irregular status.

More formally, we model the worker's decision as follows, slightly modifying the model in Ahn and de la Rica (1997)[10].

Let $W_f$ be the wage[11] in the formal sector and $W_u$ that in the underground sector. Moreover, let $W_s$ be the value of being unemployed (or of search) and $W_r$ be the reservation wage, i.e. the value of leisure. In a first stage the individual decides to participate in the labour market if

$W_r < \max (W_f, W_u, W_s)$.

If wages offered in both the regular and the irregular sectors are too low, he will keep on searching. More formally, he stays unemployed as far as

$W_s > \max (W_f, W_u) \mid W_r < \max (W_f, W_u, W_s)$.

In the second stage he will choose the formal sector and not the irregular sector if the former provides the higher wage, i.e.

$W_f > W_u \mid W_s < \max (W_f, W_u) \ \& \ W_r < \max (W_f, W_u, W_s)$.

As we are analyzing two discrete decisions (working versus not working and being underground versus being regular) we can model them using a probit model with sample selection[12] (van de Ven and van Praag, 1981), an extension of the original Heckman selection model when the response variable is binary. This method allows using information from non-working individuals to improve the estimates of the parameters in the outcome equation. Our relationship of interest is

$$[3] \qquad y_i^* = x_i^{'}\beta + u_i$$

where y* is a latent (unobservable) variable. Equation *[3]* is the outcome equation, modelling the probability of working underground. We do not observe y*, but y, defined as

$$y_i = \begin{cases} 1 & y_i^* > 0 \\ 0 & y_i^* \leq 0 \end{cases}$$

---

[10] In that study there is a second stage in which the individual works in the formal sector if Wf > Ws | Wr < max (Wf, Wu, Ws) and a third stage in which if the worker does not find a job in the formal sector he turns to the underground sector provided that Wu > Ws | Wf < Ws & Wr < max (Wf, Wu, Ws).

[11] The term wage must be referred to a wider package of job characteristics including not only the current monetary compensation but also insurance and pension coverage, work environment, working time, location, etc..

[12] The estimation was performed using the Stata 'Heckprob' procedure.

In turn, y is observed if and only if another latent variable exceeds a certain threshold:

$$[4] \qquad s_i^* = z_i^{'}\gamma + e_i$$

where again $s_i = \begin{cases} 1 & s_i^* > 0 \\ 0 & s_i^* \leq 0 \end{cases}$. Equation *[4]* is the selection or participation equation.

In our case it models the probability of having ever worked (or the "propensity to work"), and can be estimated on the entire sample. The probability of working underground in equation *[1]*, instead, can be observed only for the sub-sample of individuals who have ever worked. The error terms *u* and *e* are bivariate normally distributed with correlation ρ. If this correlation is found to be different from zero standard probit techniques that ignore selection would yield biased estimates.

Simultaneous estimation of the participation equation and of the outcome equation is performed through a maximum likelihood approach. In order to strengthen identification we introduce some exclusion restrictions, that is variables featuring in the selection equation but not in the outcome equation. We assume that the probability of working also depends on job search[13] and on whether the individual receives either pensions or transfers (table 4, column D).

The correlation between the error terms[14] of the two equations is statistically different from zero, supporting our concern for a selection bias; despite this, results are fairly consistent with those from the previous specification. Also in this case, schooling is estimated to have a significant and negative impact on the probability of working underground, only marginally less pronounced than in the baseline case: continuing up to college degree instead of dropping out soon after the compulsory requirements reduces this probability by 10 percentage points for men and by 11 for women (picture 1). The benefit is, as one would expect, slightly higher when one completes the compulsory school track with respect to those with no formal education (respectively 13 and 14 percentage points). As soon as we control for selection the difference by gender in the impact of education found with the previous model vanishes; also the marginal effect of being male on the probability of working off the books more than halves, though being still highly significant.

---

[13] The job search variable is equal to one if the individual has ever put some effort into searching for a job, either in the past (if he or she is currently working) or now (if he or she is currently searching).

[14] According to the likelihood ratio test, the correlation ρ between the error terms is strongly significantly different from zero (see, in table 4, bottom part of column D); hence we strongly fail to accept the null of independent equations.

*4.3    Sensitivity analysis*

In this section we check the robustness of our results along different dimensions. In particular, we include information available at macro level in order to control for local labour demand conditions; we control for the potential endogeneity of the level of education; we introduce an alternative measure of education; we consider different ways of identifying underground workers. As a whole we derive consistent results.

a. <u>Influence of local demand conditions</u>.

As the SHIW supplies the residence of the respondent, we can merge information available at macro level with microeconomic data in order to control for local demand conditions. Specifically, we assign to each individual the unemployment rate and the irregularity rate in his/her region of residence. The latter in particular, apart from representing the extent of employment opportunities in the underground sector, is meant to account for the strength of the social norms, which reasonably depends on the fraction of the economy already situated in the informal sector. The larger this fraction, the lower is the stigma associated with that behaviour, the weaker the social norm. It is also likely that a larger fraction of workers in the irregular segment represents an incentive to go idle as it makes more difficult for the relevant authorities to detect these irregularities. The inclusion of this macro factors leaves our results unaltered, as coefficients are not statistically significant (table 4, column B).

b. <u>Endogeneity (if any) of education choices</u>.

As already said, one might wonder if students drop out of school as they are attracted by employment opportunities in the underground sector. If this is the case, the direction of causality would be the opposite, flowing from the underground status to the education attainment. As a first check we test whether high school dropouts are significantly associated with the rate of irregular employment in the region, which is meant to proxy for the demand of labour from the underground sector. We do this by restricting to the sub-sample of boys and girls aged between 14 and 19 years, and by defining our outcome variable – probability of dropping out - as equal to 1 if he/she does not hold a diploma and is not enrolled in education. On the right-hand side, apart from the rate of hidden labour in the region, we consider: gender; parents' education and parents' working status; province of residence and town size in terms of inhabitants; regional rate of unemployment. Town size should account for supply conditions, whereas availability of schools is arguably lower in smaller towns (e.g. rural or remote municipalities) but also in bigger towns (due to congestion). As table 5 shows, the size of the local irregular sector and the conditions of the local labour market, as

well as the town size, do not influence the choice of continuing secondary education. Boys are more likely to abandon. As supported by other studies (see, for Italy, Cingano and Cipollone, 2007), family background matters in shaping schooling choices of children: the probability of dropping out of secondary school is decreasing in both parents' educational achievement, with coefficients that are almost the same. While our evidence on parents' education is crystal clear, that on parents' working status is more blurred, though pointing to an increase of dropout probability when parents are out of the labour force. Unfortunately we miss some important variables that are likely to induce students to drop out, such as grade repetition or other indicators of poor school performance.

Furthermore, to credibly identify a causal impact of schooling on the probability of working underground on the whole sample we instrument each individual's number of years spent at school with his/her parents' education. We postulate, in order for our instrument to be valid, that parents' education does not affect the working status of children if not through children's education. We find that both father's and mother's education serve as relevant instruments, showing largely significant coefficients in the first-stage regression (table 4, column C)[15]. The coefficient for fathers is now twice that for mothers, suggesting – puzzlingly, though - that while the children's choice of remaining at school after the compulsory requirements is affected in equal parts by fathers and mothers, this is not also the case for the whole length of the educational track. Since our model is over-identified (two instruments and one potentially endogenous variable) we can adopt a two stage least squares (2SLS) approach and formally test for the validity of the instruments chosen through a test for over-identification. The Sargan test of overidentifying restrictions fails to reject the null hypothesis that our instruments are valid. Finally, the Wald test rejects the null hypothesis of exogeneity only at the 5 per cent but not at the 1 per cent confidence level.

Against this background, our results are broadly confirmed. There are no remarkable changes with respect to the previous specifications apart from the fact that moving to an instrumental variable estimate leads to a smaller effect (in absolute value) of schooling on our outcome. The effort exerted into studying for a further ten years after the compulsory requirements is still worthwhile, as the probability of working underground will shrink by 8 percentage points for men and by 11 for women. Achieving the compulsory schooling makes job irregularity less likely with respect to those with no formal education respectively by 11 and 13 percentage points (picture 1).

---

[15] Sample size is slightly lower with respect to the baseline estimation due to missing information regarding parents' education.

c. <u>An alternative measure of education</u>.

Instead of using education measured in years, we have checked our results against an alternative indicator of schooling, given by three dummies for the highest qualification attained: compulsory school or less, high school diploma, college degree and beyond. The close relationship between school attainment and working status is strongly confirmed. Our most conservative estimate, obtained with the Heckman selection model, suggests that the probability of working underground for high school or college graduates is lower by 8 percentage points with respect to those who at most achieved the compulsory education (table 6). Again we find an advantage for those with a secondary education compared with those with at most compulsory education, but in contrast with specifications featuring the number of years of schooling here we do not find an additional benefit from going beyond the high school.

d. <u>Different criteria for identifying underground workers from the SHIW</u>.

The adoption of the criteria described in Section 4.1 to identify the underground workers is clearly judgemental. Choosing a higher (lower) threshold in *[2]* produces the effect of increasing (reducing) the number of those considered as underground. We have first restricted ourselves to the narrow measure of underground, i.e. to those who have been engaged in the shadow sector throughout their whole career. We still find a sizeable effect of education on the probability of being underground (table 7). Men who achieve compulsory education face a probability of being stuck in an informality trap which is less than half that for those with no education at all (11 versus 5 per cent); this probability shrinks further for those taking a college degree (2 per cent). For women the dynamics is broadly the same, though the probability levels are remarkably higher: from 17 to 7 per cent by completing compulsory schooling with respect to those with no qualification, decreasing further to 3 per cent for college graduates. Finally, results have been checked against different values of the thresholds both in the relative and in the absolute criteria adopted to retrieve the broad measure. We find very mild changes in the extent of estimated coefficients and of their standard errors.[16]

## 5. Conclusions

The paper has tested the relationship between irregular labour and schooling achievement for the Italian economy. In particular we have used microeconomic data on households from the Bank of Italy's Survey of Households Income and

---

[16] Estimates are not reported for the sake of brevity but are available upon request.

Wealth, which allows identifying irregular workers by relying on individuals' self reported social security positions.

We show that having low education levels sizeably and significantly increases the probability of working permanently underground. In this regard, our contribution corroborates, on an empirical standpoint, the evidence found for other countries. It also adds to the existing literature referred to Italy, albeit very scarce, by explicitly modelling this relationship. Apart from controlling for a wide range of socio-demographic and economic variables, we have addressed potential endogeneity and selection issues and checked the consistency of our results along different dimensions.

The data lead to an estimated reduction by at least 10 and 11 percentage points of the probability of working off the books when respectively boys and girls reach a college degree instead of dropping out of school at the compulsory age. This compares to an average observed probability of 16 and 19 per cent respectively. The gain is slightly higher when one completes the compulsory school track with respect to those with no formal education at all (respectively 13 and 14 percentage points). Remarkable differences by gender have emerged instead in absence of a proper treatment of the selection bias induced by the non random participation into the labour market. We have not been able instead to support that job opportunities in the underground labour market induce students to drop out of school.

# References

Ahn, N. and S. de La Rica (1997) "The underground economy in Spain: an alternative to unemployment?", *Applied economics*, **29**, pp. 733-743.

Attanasio, O.P., Guiso, L. and T. Jappelli (2002) "The demand for money, financial innovation, and the welfare cost of inflation: an analysis with household data", *Journal of Political Economy*, **110**, pp. 317-351.

Banca d'Italia (2006) "I bilanci delle famiglie italiane nell'anno 2004", *Supplementi al Bollettino Statistico - Indagini campionarie*, 16, 7.

Barth, E. and T. Ognedal (2005) "Unreported labour", *IZA Discussion Paper*, No. 1893.

Boeri, T. and P. Garibaldi (2002) "Shadow activity and unemployment in a depressed labour market", *CEPR Discussion Paper*, No. 3433.

Boeri, T. and P. Garibaldi (2005) "Shadow sorting", paper presented at the NBER Macroeconomic Conference, Budapest, June.

Brandolini, A. and G. D'Alessio (2002) "Evidenze sul lavoro sommerso nell'Indagine della banca d'Italia sui bilanci delle famiglie italiane", *Bank of Italy*, manuscript.

Carillo, M.R. and E. Papagni (2002) "Economia sommersa e crescita economica in presenza di complementarieta' tra innovazione e capitale umano", *Rivista di Politica Economica*, **92**, pp. 249-280.

Carillo, M.R. and M. Pugno (2002) "The underground economy and the underdevelopment trap", *University of Trento, Department of Economics*, Discussion Paper No. 1.

Cingano, F. and P. Cipollone (2007) "University drop-out: the case of Italy", *Bank of Italy, Temi di discussione* No. 626.

European Commission (2004) "Undeclared work in an enlarged union", DG for Employment and Social Affairs, Brussels.

Fields, G. S. (2005) "A guide to multisector labour market models", *Social Protection Discussion Paper Series,* The World Bank, No. 505.

Gobbi, G. and R. Zizza (2007) "Does the underground economy hold back financial deepening? Evidence from the Italian credit market", *London School of Economics, Centre for Economic Performance Discussion Paper*, No. 789, April.

Gong, X. and A. van Soest (2001) "Wage differentials and mobility in the urban labour market: a panel data analysis for Mexico", *IZA Discussion Paper*, No. 329.

Isae (2002) "Quarterly report", Rome, April (in Italian).

Istat (2005) "Education of the population in 2001" (in Italian).

Istat (2006) "The measurement of the underground economy in official statistics. Years 2000-2004" (in Italian).

Kolm, A. and B. Larsen (2003) "Does tax evasion affect unemployment and educational choice?", *Copenhagen Business School, Department of Economics Working Paper*, No. 12.

Maloney, W. F. (2004) "Informality revisited", *World Development*, Elsevier, vol. 32(7), pp. 1159-1178.

OECD (2002) "*Measuring the non-observed economy. A handbook*", Paris.

OECD (2006) "*Education at a glance*", Paris.

Peracchi, F., Perotti, V., Scarpetta, S., Carletto, G., Pagés, C. and K. Scott (2007) "*Informality and social protection: preliminary results from pilot surveys in Bulgaria and Colombia*", paper presented at the IZA-World Bank Workshop "*The Informal Economy and Informal Labor Markets in Developing, Transition and Emerging Economies*", Bertinoro, 26-27 January 2007.

Rockwool Foundation (2005) "*News from the Rockwool Foundation Research Unit*", June.

Scarlato, M. (2001) "Settore sommerso ed equilibri multipli: uno schema analitico per il Mezzogiorno", *Rivista di Politica Economica*, **91**, pp. 37-59.

Schneider F. and D. H. Enste (2000) "Shadow economies: size, causes and consequences", *Journal of Economic Literature,* vol. XXXVIII, pp. 77-114.

van de Ven, W.P.M.M. and B.M.S. van Praag (1981) "The demand for deductibles in private health insurance. A probit model with sample selection", *Journal of Econometrics*, **17**, pp. 229-252.

**Appendix. Comparison between SHIW and Social Security records.**

In this paper we use data on pension contributions resulting from the SHIW; as for any other information gathered through surveys they can suffer from a number of sampling and non sampling errors. The availability of the same type of information from administrative sources such as the Social Security records collected by Inps (Italian National Social Security Institute) allows us to assess how reliable the data we use are. To ensure comparability here we have excluded from SHIW those working in the agricultural and in the public sectors, and have restricted to individuals working in 2004 and born after 1956, as the Inps archive starts recording only from 1970. The picture compares the median number of years in which pension contributions have been paid by age, as emerging from the SHIW and the Inps archives. "Declared" and "actual" pension contributions are very similar at most ages, supporting the use of survey data to perform our analysis.[17]

Chart A1
Median number of years when pension contributions have been paid by age. Men and women.



---
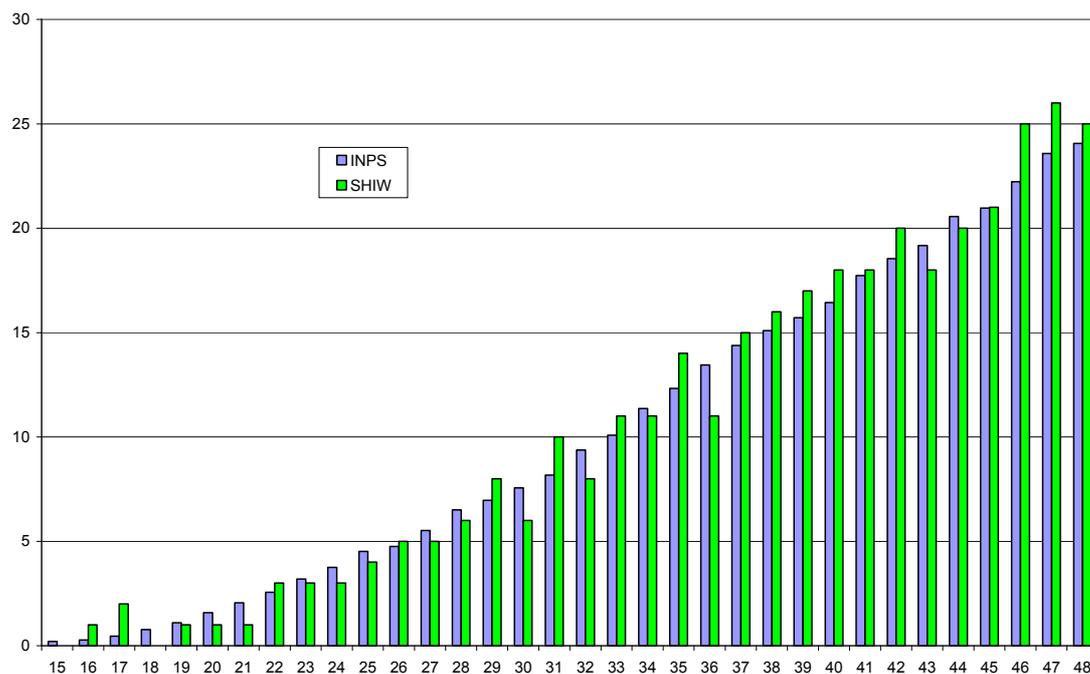
[17] We repeated the same exercise considering men and women separately and still obtained fairly similar results between SHIW and Inps years of contributions.

**Table 1. Descriptive statistics for regional data.**

| Variable | Mean | Median | Standard deviation | First quartile | Third quartile | Minimum | Maximum |
|---|---|---|---|---|---|---|---|
| IRR | 17.0 | 14.9 | 6.0 | 12.1 | 21.1 | 7.5 | 35.6 |
| DROPOUT | 4.5 | 4.6 | 2.1 | 3.0 | 5.7 | 0.0 | 11.2 |
| ENF | 7.8 | 7.8 | 0.2 | 7.6 | 8.0 | 7.3 | 8.2 |
| UNEM | 10.4 | 9.9 | 5.5 | 5.4 | 14.4 | 2.5 | 24.5 |
| GDPPC | 10.8 | 10.8 | 0.1 | 10.7 | 10.8 | 10.4 | 11.0 |
| SHAG | 3.5 | 3.3 | 1.6 | 2.1 | 4.4 | 1.2 | 7.6 |
| SHCO | 5.6 | 5.3 | 1.3 | 4.7 | 6.2 | 3.5 | 11.2 |
| SHSE | 70.9 | 70.3 | 6.0 | 66.1 | 76.0 | 59.4 | 81.9 |

*Notes*: our calculations on Istat data. Statistics calculated on the average over the sample period. IRR is the rate of irregularity in the private sector, calculated from that published by Istat excluding from the denominator the labour force employed in the Public Administration. ENF is computed as the average number of days in logs that it takes a bankruptcy procedure to be worked out in courts. GDPPC is the ratio between GDP and the population aged 15 years and above; in logs. The share of value added of the manufacturing sector is the omitted category.

**Table 2. Determinants of the underground employment**

| Explanatory variables | Two way fixed effect estimate | |
|---|---|---|
| DROPOUT | 0.26*** | (0.08) |
| ENF | 2.65*** | (0.93) |
| UNEM | 0.02 | (0.02) |
| GDPPC | 4.89 | (8.09) |
| SHAG | -0.60** | (0.28) |
| SHCO | -0.55*** | (0.16) |
| SHSE | -0.12 | (0.14) |
| Constant | -45.77 | (90.79) |
| *Number of observations* | 180 | |
| *R-square* | 0.41 | |
| *F-test* | 6.82 (0.00) | |

*Notes*: data for 20 Italian regions in the time span 1995-2004. Robust standard errors in parentheses. The symbols ***, **, * indicate a significance level of 1 per cent, 5 per cent and 10 per cent respectively.

**Table 3a. Percentage incidence of irregularity in selected groups. Narrow and broad measures.**

| Groups | Narrow | Broad |
|---|---|---|
| Men | 7.1 | 16.1 |
| Women | 9.3 | 19.2 |
| 14-30 years old | 18.4 | 19.5 |
| 31-40 years old | 8.2 | 17.8 |
| 41-50 years old | 4.8 | 18.4 |
| 51-65 years old | 4.4 | 15.3 |
| North | 4.7 | 12.0 |
| Centre | 6.3 | 15.5 |
| South and Isles | 14.7 | 27.7 |
| Compulsory school or below | 9.0 | 22.4 |
| High school | 6.7 | 12.5 |
| College degree or beyond | 8.0 | 11.9 |
| Always dependent worker | 7.4 | 16.1 |
| Always independent worker | 15.4 | 24.4 |
| Either independent or dependent | 4.5 | 18.5 |
| Only one job in working history | 10.5 | 18.2 |
| More than one job in working history | 5.0 | 16.5 |
| Total | 7.9 | 17.3 |

*Notes*: our estimations on 1998-2004 SHIW.

**Table 3b. Characteristics of the sample by occupational status.**

| Variable | Regular | Irregular | Not in professional status | All |
|---|---|---|---|---|
| Percentage of males | 57.7 | 52.3 | 30.8 | 49.3 |
| Percentage of singles | 25.0 | 29.0 | 60.1 | 35.5 |
| Percentage having children | 72.7 | 74.5 | 87.3 | 77.1 |
| Percentage achieved compulsory school or below (omitted category) | 48.2 | 65.8 | 64.5 | 55.0 |
| Percentage achieved high school degree | 41.3 | 26.8 | 31.5 | 36.7 |
| Percentage achieved college degree or above | 10.5 | 7.4 | 4.0 | 8.3 |
| Average years of (potential) experience | 22.9 | 23.1 | -- | 22.9 |
| Average number of jobs held | 1.96 | 2.05 | -- | 1.97 |
| Percentage always dependent workers (omitted category) | 76.0 | 69.6 | -- | 74.9 |
| Percentage always independent workers | 10.8 | 15.8 | -- | 11.7 |
| Percentage either dependent or independent workers | 13.2 | 14.6 | -- | 13.4 |
| Percentage in agriculture | 6.0 | 9.3 | -- | 6.5 |
| Percentage in manufacturing | 29.5 | 22.0 | -- | 28.5 |
| Percentage in building sector | 6.2 | 15.4 | -- | 7.4 |
| Percentage in tertiary sector (omitted category) | 58.3 | 53.3 | -- | 57.6 |

*Notes*: our calculations on 1998-2004 SHIW. Averages are computed using sample weights and using the entire sample of individuals between 14 and 65 years old.

**Table 4. Probability of working underground (broad measure). Measure of schooling: years.**

| Explanatory variables | Baseline probit model (A) | | | Baseline probit model with regional variables (B) | | | Probit model with endogenous regressor (C) | | | Probit model with sample selection (D) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Outcome equation | | | Selection eq. |
| | Coefficients | | Marg. Eff. | Coefficients | | Marg. Eff. | Coefficients | | Marg. Eff. | Coefficients | | Marg. Eff. | Coefficients |
| Male * | -0.265*** | (0.020) | -0.062 | -0.265*** | (0.039) | -0.062 | -0.289*** | (0.022) | -0.065 | -0.112*** | (0.021) | -0.027 | 0.90*** (0.02) |
| Never married * | 0.297*** | (0.027) | 0.073 | 0.297*** | (0.048) | 0.073 | 0.324*** | (0.029) | 0.077 | 0.236*** | (0.027) | 0.058 | -0.55*** (0.03) |
| Had children * | 0.044** | (0.023) | 0.010 | 0.043 | (0.033) | 0.010 | 0.093** | (0.025) | 0.020 | 0.017 | (0.022) | 0.004 | -0.13*** (0.03) |
| **Years at school** | **-0.065*** | **(0.003)** | **-0.015** | **-0.065*** | **(0.007)** | **-0.015** | **-0.050*** | **(0.007)** | **-0.011** | **-0.053*** | **(0.003)** | **-0.013** | **0.03*** (0.00)** |
| Experience | 0.009*** | (0.003) | 0.002 | 0.009*** | (0.004) | 0.002 | 0.009*** | (0.003) | 0.002 | 0.016*** | (0.003) | 0.004 | |
| $(Experience/10)^2$ | -0.023*** | (0.005) | -0.005 | -0.023*** | (0.007) | -0.005 | -0.017*** | (0.006) | -0.004 | -0.036*** | (0.005) | -0.008 | |
| Number of jobs | 0.031*** | (0.004) | 0.007 | 0.031*** | (0.007) | 0.007 | 0.036*** | (0.005) | 0.008 | 0.029*** | (0.004) | 0.007 | |
| Always independent * | 0.316*** | (0.028) | 0.081 | 0.315*** | (0.057) | 0.081 | 0.337*** | (0.030) | 0.085 | 0.331*** | (0.027) | 0.079 | |
| Either dep. or indep. * | 0.204*** | (0.028) | 0.050 | 0.202*** | (0.034) | 0.050 | 0.208*** | (0.030) | 0.049 | 0.226*** | (0.028) | 0.053 | |
| Agriculture * | -0.262*** | (0.042) | -0.053 | -0.260*** | (0.063) | -0.052 | -0.250*** | (0.050) | -0.048 | -0.220*** | (0.041) | -0.047 | |
| Manufacturing * | -0.442*** | (0.026) | -0.090 | -0.441*** | (0.053) | -0.090 | -0.432*** | (0.030) | -0.084 | -0.395*** | (0.025) | -0.082 | |
| Building sector * | 0.206*** | (0.036) | 0.051 | 0.208*** | (0.065) | 0.052 | 0.220*** | (0.042) | 0.053 | 0.248*** | (0.035) | 0.066 | |
| Rate of underg. labour | | | | -0.014 | (0.018) | -0.003 | | | | | | | |
| Rate of unemployment | | | | 0.016 | (0.013) | 0.004 | | | | | | | |
| Constant | -0.129 | (0.216) | | -0.378* | (0.208) | | -0.784*** | (0.292) | | -0.416** | (0.214) | | -1.20*** (0.41) |
| Father's education | | | | | | | § 0.259*** | (0.007) | | | | | |
| Mother's education | | | | | | | § 0.131*** | (0.008) | | | | | |
| Age | | | | | | | | | | | | | 0.08*** (0.01) |
| $(Age/10)^2$ | | | | | | | | | | | | | -0.03*** (0.01) |
| Pension or transfers * | | | | | | | | | | | | | -1.27*** (0.03) |
| Ever searched for job * | | | | | | | | | | | | | 1.65*** (0.03) |
| Province of residence * | yes | | | yes | | | yes | | | | yes | | |
| | | | | | | | | | | | | | |
| *Wald test of exogeneity* | | | | | | | 4.12 (0.04)^ | | | | | | |
| *Overidentif. restr. test* | | | | | | | 0.006 (0.94) ^ | | | | | | |
| *LR test indep. equations* | | | | | | | | | | 609.10 (0.00)^ | | | |
| | | | | | | | | | | | | | |
| *Number of observations* | 28,459 | | | 28,459 | | | 24,960 | | | 40,725 (uncensored: 28,459) | | | |
| *Observed probability* | 0.173 | | | 0.173 | | | 0.165 | | | 0.173 | | | |
| *Predicted probability* | 0.145 | | | 0.145 | | | 0.138 | | | 0.154 | | | |

*Notes*: an asterisk in the first column indicates that the variable is a dummy. See table 3 for omitted categories. Marginal effects evaluated at the sample mean. Standard errors are reported in parentheses; in column (B) errors are clustered by region. The symbols ***, **, * indicate a significance level of 1 per cent, 5 per cent and 10 per cent respectively. (§) Coefficient of the instruments in the first-stage regression. (^) P-value of the corresponding test statistics in parenthesis.

## Table 5. Probability of dropping out of school
*(Individuals between 14 and 19 years old)*

| Explanatory variables | Probit estimates | | |
|---|---|---|---|
| | Coefficients | | Marginal effects |
| Male * | 0.255*** | (0.057) | 0.039 |
| Father's education (in years) | -0.090*** | (0.010) | -0.014 |
| Mother's education (in years) | -0.095*** | (0.012) | -0.015 |
| Father: dependent worker * | -0.172*** | (0.063) | -0.027 |
| Father: independent worker * | -0.200** | (0.089) | -0.028 |
| Father: not working (omitted category) | | | |
| Mother: dependent worker * | -0.044 | (0.070) | -0.007 |
| Mother: independent worker * | -0.465*** | (0.169) | -0.054 |
| Mother: not working (omitted category) | | | |
| Town size: less than 20,000 * | -0.098 | (0.098) | -0.015 |
| Town size: 20,000-40,000 (omitted category) | | | |
| Town size: 40,000-500,000 * | 0.074 | (0.158) | 0.012 |
| Town size: more than 500,000 * | -0.091 | (0.127) | -0.013 |
| Rate of underground labour | -0.013 | (0.030) | -0.002 |
| Rate of unemployment | 0.010 | (0.029) | 0.002 |
| Province of residence * | | yes | |
| *Number of observations* | | 3,602 | |
| *Observed probability* | | 0.152 | |
| *Predicted probability* | | 0.084 | |

*Notes*: an asterisk in the first column indicates that the variable is a dummy. Marginal effects evaluated at the sample mean. Standard errors, reported in parentheses, are clustered by region. The symbols ***, **, * indicate a significance level of 1 per cent, 5 per cent and 10 per cent respectively.

**Table 6. Probability of working underground (broad measure). Measure of schooling: achieved qualification dummies.**

| Explanatory variables | Baseline probit model | | | Probit model with sample selection | | |
|---|---|---|---|---|---|---|
| | | | | Outcome equation | | Selection eq. |
| | Coefficients | | Marg. Eff. | Coefficients | | Marg. Eff. | Coefficients |

| Explanatory variables | Coefficients | | Marg. Eff. | Coefficients | | Marg. Eff. | Coefficients | |
|---|---|---|---|---|---|---|---|---|
| Male * | -0.280*** | (0.020) | -0.066 | -0.124*** | (0.021) | -0.029 | 0.91*** | (0.02) |
| Never married * | 0.299*** | (0.027) | 0.074 | 0.237*** | (0.027) | 0.058 | -0.55*** | (0.03) |
| Had children * | 0.040* | (0.022) | 0.009 | 0.012 | (0.022) | 0.003 | -0.12*** | (0.03) |
| **High school degree *** | **-0.433*** | (0.022)** | **-0.094** | **-0.358*** | (0.022)** | **-0.081** | 0.21*** | (0.02) |
| **College degree or above *** | **-0.551*** | (0.036)** | **-0.099** | **-0.439*** | (0.035)** | **-0.085** | 0.13*** | (0.04) |
| Experience | 0.008*** | (0.003) | 0.002 | 0.015*** | (0.003) | 0.004 | | |
| (Experience/10)² | -0.014*** | (0.005) | -0.003 | -0.029*** | (0.005) | -0.007 | | |
| Number of jobs | 0.032*** | (0.004) | 0.007 | 0.030*** | (0.004) | 0.007 | | |
| Always independent * | 0.305*** | (0.028) | 0.078 | 0.322*** | (0.027) | 0.076 | | |
| Either dep. or indep. * | 0.189*** | (0.028) | 0.047 | 0.213*** | (0.028) | 0.050 | | |
| Agriculture * | -0.209*** | (0.042) | -0.043 | -0.171*** | (0.041) | -0.037 | | |
| Manufacturing * | -0.430*** | (0.026) | -0.088 | -0.383*** | (0.025) | -0.080 | | |
| Building sector * | 0.232*** | (0.036) | 0.059 | 0.271*** | (0.035) | 0.072 | | |
| Constant | -0.832*** | (0.249) | | -1.097*** | (0.246) | | -1.06*** | (0.41) |
| Age | | | | | | | 0.08*** | (0.01) |
| (Age/10)² | | | | | | | -0.03*** | (0.01) |
| Pension or transfers * | | | | | | | -1.27*** | (0.03) |
| Ever searched for job * | | | | | | | 1.67*** | (0.03) |
| Province of residence * | yes | | | yes | | | | |
| | | | | | | | | |
| *LR test of indep. equations* | | | | 626.54 (0.00)^ | | | | |
| | | | | | | | | |
| *Number of observations* | 28,459 | | | 40,725 (uncensored: 28,459) | | | | |
| *Observed probability* | 0.173 | | | 0.173 | | | | |
| *Predicted probability* | 0.146 | | | 0.154 | | | | |

*Notes*: an asterisk in the first column indicates that the variable is a dummy. See table 3 for omitted categories. Marginal effects evaluated at the sample mean. Standard errors are reported in parentheses. The symbols ***, **, * indicate a significance level of 1 per cent, 5 per cent and 10 per cent respectively. (^) P-value of the corresponding test statistics in parenthesis.

**Table 7. Probability of working underground (narrow measure). Measure of schooling: years.**

| Explanatory variables | Baseline probit model | | Probit model with sample selection | | |
|---|---|---|---|---|---|
| | | | Outcome equation | | Selection eq. |
| | Coefficients | Marg. Eff. | Coefficients | Marg. Eff. | Coefficients |
| Male * | -0.286*** (0.027) | -0.029 | -0.138*** (0.027) | -0.014 | 0.90*** (0.02) |
| Never married * | 0.319*** (0.033) | 0.035 | 0.256*** (0.032) | 0.028 | -0.58*** (0.03) |
| Had children * | -0.060** (0.030) | -0.006 | -0.093*** (0.029) | -0.010 | -0.13*** (0.03) |
| **Years at school** | **-0.056*** (0.004)** | **-0.005** | **-0.043*** (0.004)** | **-0.004** | 0.03*** (0.00) |
| Experience | -0.057*** (0.003) | -0.005 | -0.049*** (0.003) | -0.005 | |
| (Experience/10)$^2$ | 0.073*** (0.007) | 0.007 | 0.061*** (0.007) | 0.006 | |
| Number of jobs | 0.009 (0.006) | 0.001 | 0.010* (0.006) | 0.001 | |
| Always independent * | 0.476*** (0.033) | 0.062 | 0.491*** (0.032) | 0.050 | |
| Either dep. or indep. * | 0.027 (0.043) | 0.003 | 0.050 (0.043) | 0.005 | |
| Agriculture * | -0.264*** (0.056) | -0.021 | -0.215*** (0.055) | -0.019 | |
| Manufacturing * | -0.416*** (0.035) | -0.034 | -0.363*** (0.034) | -0.031 | |
| Building sector * | 0.110** (0.046) | 0.011 | 0.155*** (0.046) | 0.018 | |
| Constant | -0.620 (0.458) | | -0.659** (0.330) | | -1.39*** (0.41) |
| Age | | | | | 0.08*** (0.01) |
| (Age/10)$^2$ | | | | | -0.04*** (0.01) |
| Pension or transfers * | | | | | -1.27*** (0.03) |
| Ever searched for job * | | | | | 1.65*** (0.03) |
| Province of residence * | yes | | yes | | |
| | | | | | |
| *LR test of indep. equations* | | | 332.58 (0.00)^ | | |
| | | | | | |
| *Number of observations* | 28,459 | | 40,725 (uncensored: 28,459) | | |
| *Observed probability* | 0.079 | | 0.079 | | |
| *Predicted probability* | 0.046 | | 0.050 | | |

*Notes*: an asterisk in the first column indicates that the variable is a dummy. See table 3 for omitted categories. Marginal effects evaluated at the sample mean. Standard errors are reported in parentheses. The symbols ***, **, * indicate a significance level of 1 per cent, 5 per cent and 10 per cent respectively. (^) P-value of the corresponding test statistics in parenthesis.

**Picture 1. Predicted probabilities of working underground by years of schooling according to different estimated models.**
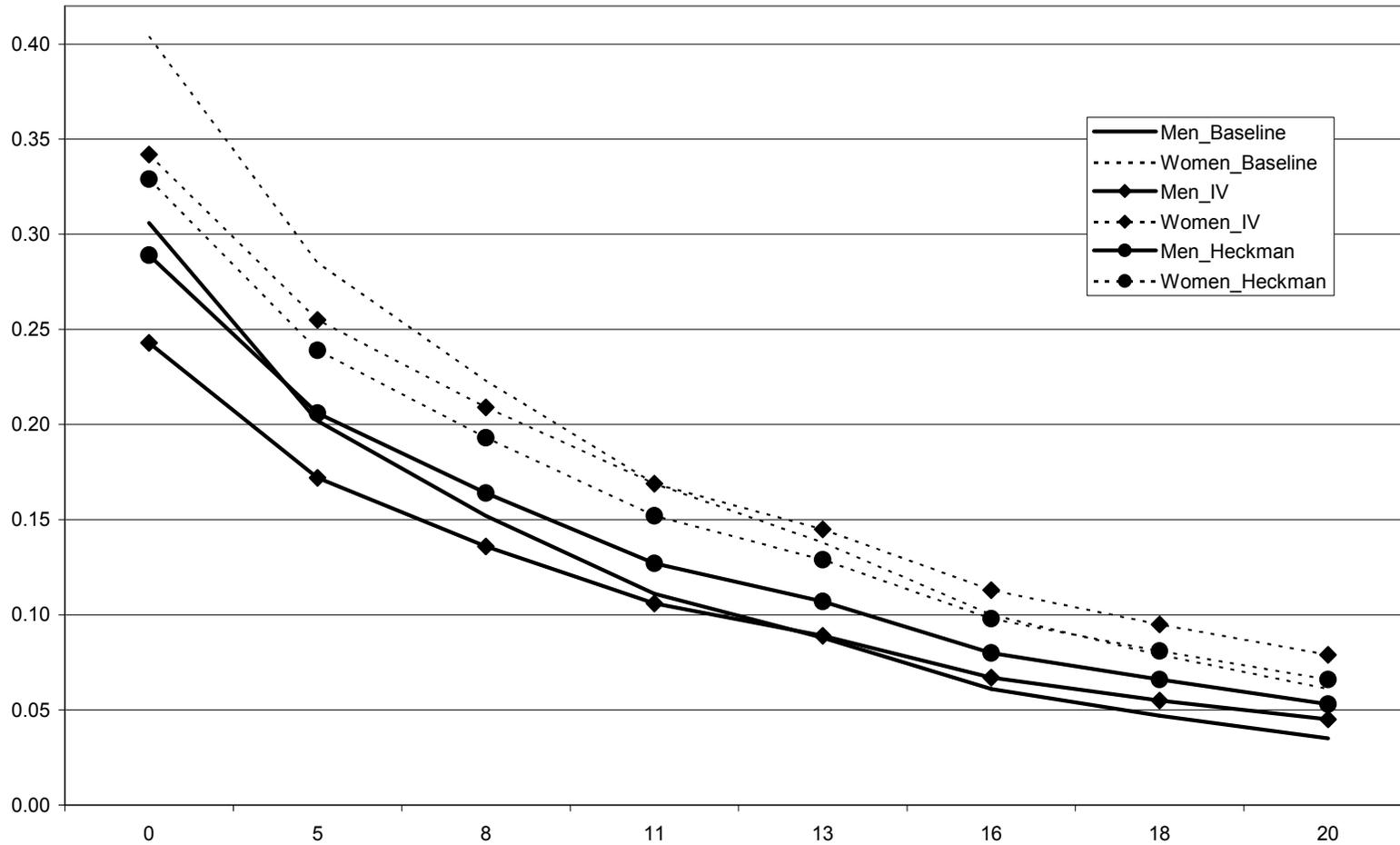


*Notes*: predicted probabilities are evaluated for men and women at the corresponding number of years of schooling and at the sample mean for the other independent variables. The broad measure of being underground is used (see Section 4.1 for a definition).