

DROPPING THE BOOKS AND WORKING OFF THE BOOKS

Rita Cappariello* and Roberta Zizza*

Abstract

The paper empirically tests the relationship between underground labour and schooling achievement for Italy, a country ranking badly in both respects when compared to other high-income economies, with a marked duality between North and South. In order to identify underground workers, we exploit in particular the information on individuals' social security positions available from the Bank of Italy's Survey on Household Income and Wealth. After controlling for a wide range of socio-demographic and economic variables and addressing potential endogeneity and selection issues, we show that a low level of education sizeably and significantly increases the probability of working underground. Switching from completing compulsory school to graduating at college more than halves this probability for both men and women. The gain is slightly higher for individuals completing the compulsory track with respect to those having no formal education at all. The different probabilities found for self-employed and dependent workers support the view of a dual informal sector, in which necessity and desirability coexist.

JEL classification: I21, J24, O17, R23.

Keywords: irregular employment, underground economy, dual informal sector, occupational choice, education, school drop-out, North and South divide.

* Bank of Italy, Economics, Research and International Relations. Via Nazionale, 91. 00184 Rome, Italy. We wish to thank Luca Arciero, Luigi Cannari, Maria Concetta Chiuri, Alessandra De Michele, Stefano Federico, Marco Magnani, Sauro Mocetti, Paolo Sestito, Luigi Federico Signorini, Roberto Tedeschi and seminar participants at XIX SIEP Conference, at VI Brucchi Luchino Labour Economics Workshop and at Bank of Italy for comments and suggestions and Ivan Faiella and Alfonso Rosolia for helping us with the data. The paper was in part written when Roberta was visiting the Centre for Economic Performance at LSE and benefited from the interaction with researchers in the Labour Market and Education programs. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Italy. Email: rita.cappariello@bancaditalia.it; roberta.zizza@bancaditalia.it.

*At 14 she left school on sick leave, never to go back...[].
"I didn't have any proper education or qualifications,
so I had to be a writer. What else would I have done?"*

From an interview to **Doris Lessing**,
Review Saturday Guardian, 20 January 2007

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 striking duality between the Northern and the Southern parts of the country: in 2004 irregular labour ranged from 8 per cent of total employment in Lombardy (North of Italy) to 33 per cent in Calabria (South of Italy). This geographical divide stands at odds with a substantial homogeneity, at least *de jure*, in terms of tax and audit/punishment policies, and in terms of burden of regulation and bureaucracy, which are 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).

Italy ranks badly also in terms of education if compared 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 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 as for education achievement of the labour force: the Southern regions are lagging behind according to many schooling indicators (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 regarded as more appropriate than firm level data for measuring the extent of informal labour. In fact,

a surplus of labour input derived from household surveys over that from enterprise sources is considered as 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, 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 with the actual provision of unreported work, show a negative effect on both demand and supply of irregular labour of different human capital variables for the Norwegian economy. A study for Germany by 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, or hold low-skill jobs 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 skill distribution, as the proportion of workers with a primary or lower level of education is much larger in the shadow sector than in the regular one. They argue, from these results as well as from those emerging from a survey conducted by 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).

We first attempt to evaluate the relationship between engagement in shadow labour and schooling attainment for Italy by using the official measures of hidden labour available at the regional level in the time span ranging from 1995 to 2004. We try to shed light on the role played by education in shaping the heterogeneity across regions in the diffusion of irregular labour by regressing a measure of shadow employment (the irregularity rate) on an indicator of education (the dropout rate at secondary school¹) and on some control variables (see table 1 for a description of variables used and for regression results). Dropping out of secondary school turns out to have a positive and significant impact on the extent of irregular employment. Among the control variables, only the efficiency of the judiciary system and the sectoral composition of local economy display coefficients which are statistically

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.

different from zero at standard confidence level and whose sign conforms to theory: in particular, the bigger the size of underground labour is, the less efficient the judiciary system. Results in terms of both sign and magnitude of the coefficient for dropout are robust to the exclusion of control variables and of regional fixed effects. This exercise confirms the hypothesis that underground economy and education are interconnected phenomena, but does not provide evidence on the direction of causality between them since it fails to address the issue of endogeneity (if any) of education choices. An approach based on individual data seems more suitable to our scopes, and is then attempted by using microeconomic data on households from the Bank of Italy's Survey on Household Income and Wealth. To our knowledge, only two papers provided descriptive evidence on our topic of interest 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 criteria introduced to identify underground workers in the survey. In Section 4 we model the probability of working underground. Section 5 focuses on selected groups within our sample. In Section 6 we perform some sensitivity analysis. Section 7 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 supply of underground labour is 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 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. On the other hand, if 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.

A way of combining these two polar views is to regard the underground sector as having its own internal duality; in this two-tier view, some informal jobs are preferred to formal jobs and some are not (Fields, 2005).

The framework is further complicated by the fact that the direction of causality may be the opposite one. It might be the opportunity to work in the irregular sector 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. The preferences of heterogeneous workers between formal or informal jobs are also modelled by Galiani and Weinschebaum (2007). In equilibrium they find that human capital is the factor that determines whether workers are employed formally or informally; in particular, low human capital workers operate in the informal sector while high human capital workers operate formally.

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

3. 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.

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*. Individuals 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). In principle, by comparing this information with that on the number of years worked (i.e. with experience; EXPER) it would be possible to retrieve a precise indication of social contribution evasion, that occurs when $YCONTR < EXPER$.² Unfortunately the SHIW does not ask 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. Since we are interested to permanent effects of the low level of education we are willing to select

² 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.

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 criterion³:

$$[1] \quad YCONTR / POTEXPER < p25$$

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 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.⁴ 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 fully comparable with that provided by Istat and used in Section 1, 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. Still, our measure broken down by region is quite strongly correlated (by more than two thirds) with that released by Istat as well as with the measure of tax evasion by Pisani and Polito (2006) and with the opinions on tax evasion surveyed by the 2004 SHIW (Fiorio and Zanardi, 2006; Cannari and D’Alessio, 2007). These checks corroborate the use of our measure 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.⁵ 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 2. The percentage of underground workers differs considerably according to geographical location and personal characteristics. As expected, it is remarkably higher in the South and in the Isles, irrespective of the measure considered. The proportion of people in the informal sector is higher among women, young and self-

³ 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).

⁴ 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.

⁵ 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.

employed individuals. The incidence of the underground sector is higher among individuals with a compulsory or lower level of education. Descriptive statistics reveal that the share of workers in the informal sector does not always decrease linearly with education: the reduction is relevant only between compulsory and secondary school but not for achievements beyond the secondary school.

4. 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 3. 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 or *diploma* (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 by considering the minimum number of years that it takes to achieve a certain qualification.⁶

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).⁷ Most of these 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

⁶ 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.

⁷ Results are not affected if we introduce other “non permanent” conditions, such as being the head of the household or the existence of at least another income recipient in the household.

the compulsory schooling level⁸ 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. The gain from the compulsory level to the diploma is more pronounced than that from the diploma to the college degree. 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 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).⁹

Let W_f be the wage¹⁰ 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

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

⁸ 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 precise that only since 1963 the junior high school is mandatory and since 1999 the compulsory requirements have been extended further, up to the 15th birthday.

⁹ In that study worker's decision is modelled using three stages. In particular, in the second stage the individual works in the formal sector if $W_f > W_s \mid W_r < \max (W_f, W_u, W_s)$; in the third stage if the worker does not find a job in the formal sector he turns to the underground sector provided that $W_u > W_s \mid W_f < W_s \ \& \ W_r < \max (W_f, W_u, W_s)$.

¹⁰ 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..

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 selection¹¹ (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

$$[2] \quad y_i^* = x_i' \beta + u_i$$

where y^* is a latent (unobservable) variable. Equation [2] is the outcome equation, modelling the probability of working underground. We do not observe y^* , but y , defined as

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

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

$$[3] \quad s_i^* = z_i' \gamma + e_i$$

where again $s_i = \begin{cases} 1 & s_i^* > 0 \\ 0 & s_i^* \leq 0 \end{cases}$. Equation [3] 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 [2], 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. Both job

¹¹ The estimation was performed using the Stata ‘Heckprob’ procedure.

search¹² and whether the individual receives either pensions or transfers seem appropriate variables to capture the individual's decision in participating in the labour market. We consider the effort put into searching for a job as a signal of labour supply. With respect to the second instrument, we expect that people who receive pensions or transfers have less incentives in participating in the labour market. Coefficients for both variables are statistically different from zero and signs are as expected (table 4, column D).

The correlation between the error terms¹³ 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). Thus, once we control for selection, the difference by gender in the impact of education found with the baseline model vanishes. The gain from achieving a college degree instead of dropping out after a diploma is now more balanced with that from getting a diploma instead of achieving only the compulsory level. 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). Even though the marginal effect of being male is now more than halved, the probability of working off the books is still constantly higher for women, irrespective of the level of education.

5. A look at selected subgroups

Descriptive statistics referred to in Section 3 showed how heterogeneous the diffusion of underground labour is, depending on individual characteristics. Thus now we deepen our analysis by focusing on some specific subgroups within our sample. Groups have been built considering one by one breakdowns by gender, age, geographic area and kind of occupation. Corresponding predicted probabilities are illustrated in picture 2.

The comparison by gender (north-west chart) confirms our previous findings, with a probability for women which is always above that for men; the divide lessens with the level of education. Differences by age groups are less marked, with a lower

¹² 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).

¹³ 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.

likelihood of being underground for the oldest people (i.e. for those, presumably, with longer experience in the labour market; north-east chart); the pattern followed by the youngsters (between 14 and 34 years old) is slightly different, showing a relatively low level of probability of being underground for low-medium levels of education. This is consistent with the fact that in most recent years young workers have been involved in atypical contracts, that allow for less expensive social security treatment. The probability of being underground in the South is at least twice that in the North for levels of schooling below the secondary school, and close to 50 per cent for those with no formal education; the divide is strongly reduced for college graduates and basically nil for subsequent levels of schooling (south-west chart).

The breakdown by kind of occupation provides very interesting insights in order to assess the validity of the theoretical framework we referred to in Section 2. The event of being underground is only barely influenced by the level of education for the self-employed, whose likelihood is always higher than that for dependent workers beyond the compulsory level of schooling (south-east chart). This is consistent with the hypothesis that being informal is a desirable option for the self-employed, while is worse-off for dependent workers, especially for those who achieved higher education. As a whole we find hence support to the view of an informal sector with its own internal duality.

6. 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 and the kind of occupation; we consider different ways of identifying underground workers. As a whole we derive fairly consistent results.¹⁴

a. Influence of local demand conditions.

As the SHIW supplies the residence of the respondent, we can merge microeconomic data with information available at macro level in order to control for local demand conditions. Specifically, we assign to each individual the unemployment rate and the irregularity rate in his/her province of residence.¹⁵ 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

¹⁴ Thereafter, all estimates which are not fully reported for the sake of brevity are available upon request.

¹⁵ Unemployment rates are published by Istat. Irregularity rates at provincial level are available in Censis (2004).

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 fairly unaltered; the event of being underground is positively affected only by the extent of unemployment at the local level, while the irregularity rate emerges as not significant (table 4, column B).

b. Endogeneity issues.

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 province, 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. Restricting the sample by age is intended to better control for the influence of *current* conditions of the local labour market (either formal or informal) on the choice of pursuing further education. The extension to the whole sample would require data on irregularity and unemployment by province in periods far in the past to proxy labour market conditions of the place of education (or the place of residence when in education); unluckily, both these pieces of information are not available.

On the right-hand side, apart from the rate of hidden labour in the province, we consider: gender; age and its square; parents' education and parents' working status; province of residence and town size in terms of inhabitants; provincial 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 remarkably similar. 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 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).¹⁶ 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).

In order for our instrument to be valid we have postulated that parents' education does not affect the working status of children if not through children's education. Arguably, parents' education could affect children's working status also through other channels, in particular the extent of social contacts, notably an important determinant of job search outcomes (see, for example, Cingano and Rosolia, 2006). As a robustness check we control for this network effect thanks to a specific question on job search channels included in the 2004 survey only, asking whether he or she had asked relatives, friends or acquaintances to find a job; our results are left unaffected by introducing this further control.

Apart from education choices, the kind of occupation is likely to be endogenous as well. Individuals might prefer an independent position if they expect this makes

¹⁶ Sample size is slightly lower with respect to the baseline estimation due to missing information regarding parents' education.

it easier to avoid taxation; education choices could also reflect these expectations. In order to control for this issue we instrumented the kind of occupation with parents' kind of occupation, presumably a good instrument as, at least in Italy, occupation persistence across generations is very high (Mocetti, 2007). Corresponding estimates still do not change substantially.

c. Different criteria for identifying underground workers from the SHIW.

The adoption of the criteria described in Section 3 to identify the underground workers is clearly judgemental. Choosing a higher (lower) threshold in [1] 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 6). 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.

7. 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 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. By explicitly modelling this relationship it also adds to the existing, albeit very scarce, literature referred to Italy. 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 for those completing 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.

By focusing on selected subgroups we have detected marked differences in the probability of working off the books by gender, geographic area and kind of occupation. The event of being underground is strongly influenced by the level of education attained in all cases but for the self-employed. The latter result is consistent with the view of an informal sector with its own internal duality. Working in the underground sector is worse-off for dependent workers, especially for those with higher education. For the self-employed this probability is only barely influenced by the extent of education, and beyond the compulsory schooling level this probability is consistently higher than for dependent workers, suggesting a desirability of the informality option.

References

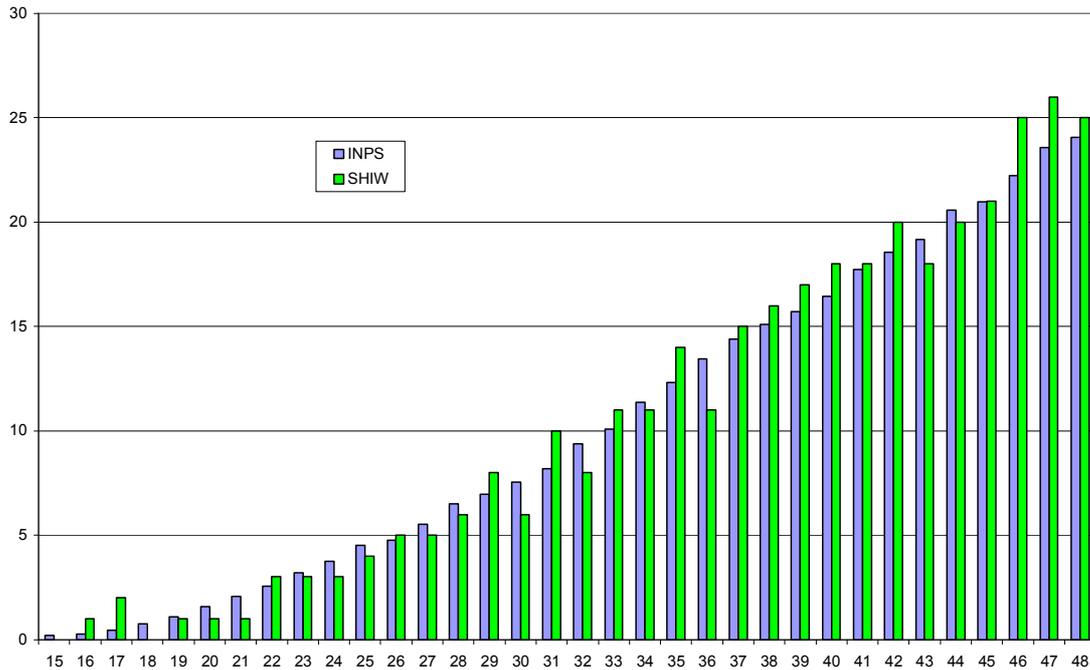
- Ahn, N. and S. de La Rica (1997) "The underground economy in Spain: an alternative to unemployment?", *Applied economics*, **29**, pp. 733-743.
- Attanasio, O.P., Guiso, L. and T. Jappelli (2002) "The demand for money, financial innovation, and the welfare cost of inflation: an analysis with household 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", in Pissarides, C. and Frenkel, J. (eds.), *NBER Macroeconomics Annual*, MIT Press, 2005.
- 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.
- Cannari, L. and G. D'Alessio (2007) "Le opinioni degli italiani sull'evasione fiscale," *Bank of Italy, Temi di discussione* No. 618.
- Censis (2004) "Understanding the underground economy", Rome (in Italian).
- Cingano, F. and P. Cipollone (2007) "University drop-out: the case of Italy", *Bank of Italy, Temi di discussione* No. 626.
- Cingano, F. and A. Rosolia (2006) "People I know: workplace network and job search outcomes", *Bank of Italy, Temi di discussione* No. 600.
- 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.
- C. Fiorio and A. Zanardi (2006) "It's a lot, but let it stay", paper presented at the XVIII Riunione Scientifica Siep, Pavia, 14-15 September 2006.
- Galiani, S. and F. Weinschelbaum (2007) "Modeling Informality Formally: Households and Firms", available at SSRN: <http://ssrn.com/abstract=957329>.

- Gobbi, G. and R. Zizza (2007) "Does the underground economy hold back financial deepening? Evidence from the Italian credit market", *Bank of Italy, Temi di discussione* No. 646.
- 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.
- Mocetti, S. (2007) "Intergenerational Earnings Mobility in Italy", *The B.E. Journal of Economic Analysis & Policy*, vol. 7: Iss. 2 (Contributions), Article 5.
- 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.
- Pisani S. and C. Polito (2006) "Analisi dell'evasione fondata sui dati IRAP. Anni 1998-2002", Documenti di lavoro dell'Ufficio Studi dell'Agenzia delle Entrate, Ministero dell'Economia e delle Finanze, Roma.
- Rockwool Foundation (2005) "News from the Rockwool Foundation Research Unit", June.
- 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 remarkably 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. Determinants of the underground employment

Explanatory variables	Dependent variable: IRR	
	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)
<i>Number of observations</i>	180	
<i>R-square</i>	0.41	
<i>F-test</i>	6.82 (0.00)	

Notes: our elaborations on Istat 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. 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. Hidden labour regional estimates 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). DROPOUT is the rate of dropout at the second year of the secondary school. ENF is computed as the average number of days in logs that it takes a bankruptcy procedure to be worked out in courts. UNEM is the rate of unemployment. GDPPC is the ratio between GDP and the population aged 15 years and above; in logs. SHAG, SHCO and SHSE are the shares of value added for the agriculture, the construction and the services sectors respectively, where the manufacturing sector is the omitted category.

Table 2. 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. See Section 3 for definitions of narrow and broad measures.

Table 3. 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. Broad measure of irregularity (see Section 3 for a definition).

Table 4. Probability of working underground (broad measure).

Explanatory variables	Baseline probit model (A)		Baseline probit model with provincial 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	-0.258*** (0.028)	-0.061	-0.289*** (0.022)	-0.065	-0.112*** (0.021)	-0.027	0.90*** (0.02)		
Never married *	0.297*** (0.027)	0.073	0.298*** (0.035)	0.074	0.324*** (0.029)	0.077	0.236*** (0.027)	0.058	-0.55*** (0.03)		
Had children *	0.044** (0.023)	0.010	0.048* (0.027)	0.011	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	-0.064*** (0.005)	-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	0.010*** (0.003)	0.002	0.009*** (0.003)	0.002	0.016*** (0.003)	0.004			
(Experience/10) ²	-0.023*** (0.005)	-0.005	-0.024*** (0.006)	-0.006	-0.017*** (0.006)	-0.004	-0.036*** (0.005)	-0.008			
Number of jobs	0.031*** (0.004)	0.007	0.031*** (0.008)	0.007	0.036*** (0.005)	0.008	0.029*** (0.004)	0.007			
Always independent *	0.316*** (0.028)	0.081	0.312*** (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	0.189*** (0.037)	0.047	0.208*** (0.030)	0.049	0.226*** (0.028)	0.053			
Agriculture *	-0.262*** (0.042)	-0.053	-0.278*** (0.055)	-0.056	-0.250*** (0.050)	-0.048	-0.220*** (0.041)	-0.047			
Manufacturing *	-0.442*** (0.026)	-0.090	-0.432*** (0.038)	-0.089	-0.432*** (0.030)	-0.084	-0.395*** (0.025)	-0.082			
Building sector *	0.206*** (0.036)	0.051	0.212*** (0.044)	0.054	0.220*** (0.042)	0.053	0.248*** (0.035)	0.066			
Rate of underg. labour			0.002 (0.007)	0.000							
Rate of unemployment			0.025*** (0.009)	0.006							
Constant	-0.129 (0.216)		-0.361*** (0.170)			-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)
Place of residence *	Province		Region		Province		Province		Province		
<i>Wald test of exogeneity</i>						4.12 (0.04) [^]					
<i>Overidentif. restr. test</i>						0.006 (0.94) [^]					
<i>LR test indep. equations</i>								609.10 (0.00) [^]			
<i>Number of observations</i>	28,459		28,459		24,960		40,725 (uncensored: 28,459)				
<i>Observed probability</i>	0.173		0.173		0.165		0.173				
<i>Predicted probability</i>	0.145		0.149		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 province. 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 *	0.238*** (0.073)	0.033
Age	1.973*** (0.362)	0.276
(Age/10) ²	-5.359*** (1.085)	-0.748
Father's education (in years)	-0.095*** (0.012)	-0.013
Mother's education (in years)	-0.088*** (0.011)	-0.012
Father: dependent worker *	-0.094 (0.063)	-0.013
Father: independent worker *	-0.148 (0.099)	-0.019
Father: not working* (omitted category)		
Mother: dependent worker *	-0.018 (0.082)	-0.003
Mother: independent worker *	-0.409*** (0.156)	-0.044
Mother: not working* (omitted category)		
Town size: less than 20,000 *	-0.055 (0.119)	-0.008
Town size: 20,000-40,000 *	-0.015 (0.119)	-0.002
Town size: 40,000-500,000 *	0.088 (0.121)	0.012
Town size: more than 500,000 * (omitted category)		
Rate of underground labour	0.026 (0.019)	0.004
Rate of unemployment	0.007 (0.019)	0.001
Region of residence *		yes
Constant	-18.71*** (3.097)	
<i>Number of observations</i>		3,802
<i>Observed probability</i>		0.144
<i>Predicted probability</i>		0.074

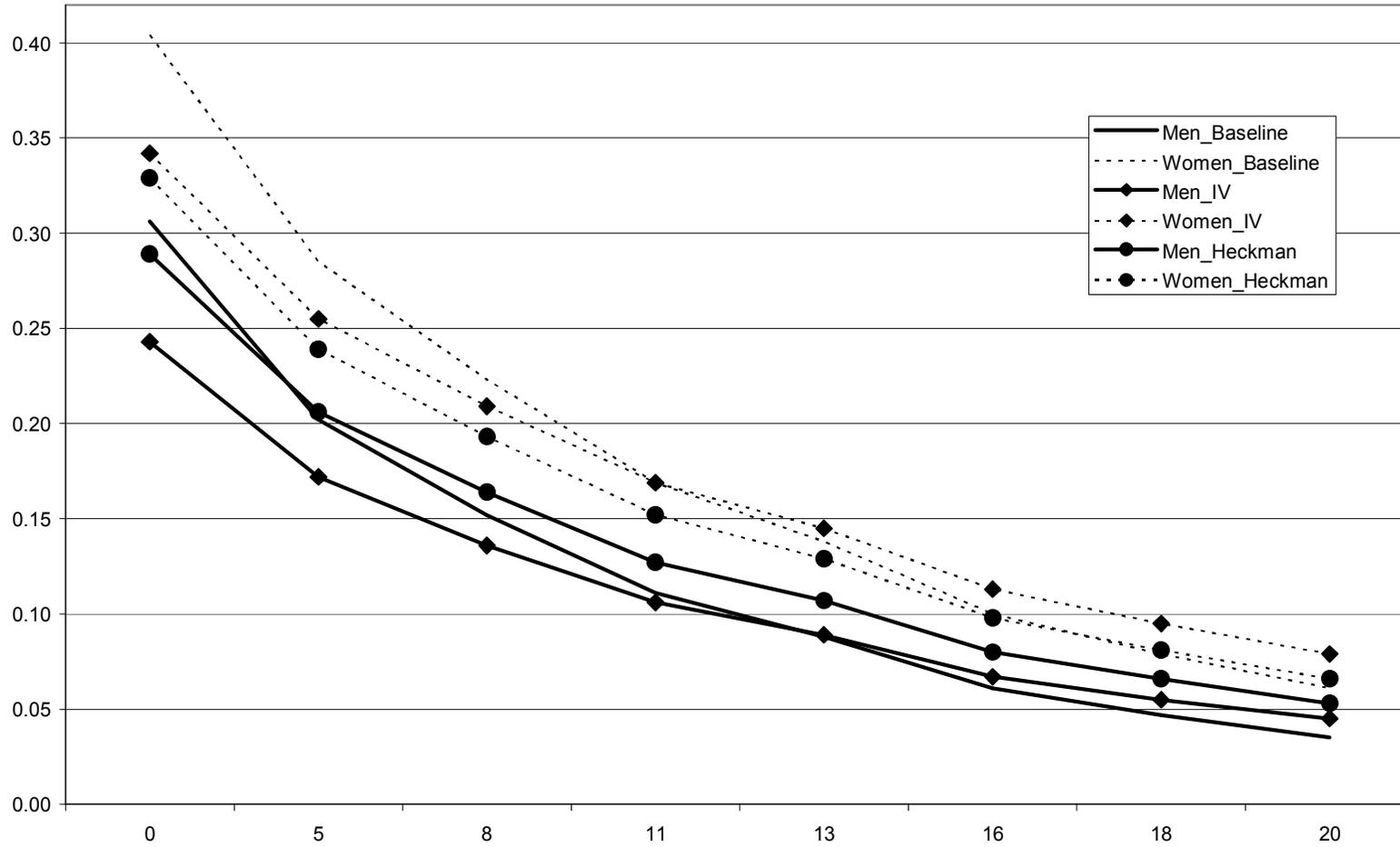
Notes: an asterisk in the first column indicates that the variable is a dummy. Marginal effects evaluated at the sample mean. Standard errors, reported in parentheses, are clustered by province. The symbols ***, **, * indicate a significance level of 1 per cent, 5 per cent and 10 per cent respectively.

Table 6. Probability of working underground (narrow measure).

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		
<i>LR test of indep. equations</i>					332.58 (0.00) [^]
<i>Number of observations</i>	28,459		40,725 (uncensored: 28,459)		
<i>Observed probability</i>	0.079		0.079		
<i>Predicted probability</i>	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 3 for a definition).

Picture 2. Predicted probabilities of working underground by years of schooling in selected subgroups.

