INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2007-06

Education in France during World War II and Subsequent Mortality

L. LEQUIEN [1]

[1] INSEE – CREST.
Telephone : +33 1 41 17 77 33,Courriel : Laurent.lequien@insee.fr
The author would like to thank participants of the CREST-LMI international seminar, March 2007, for their helpful comments. All remaining errors are mine.

Education in France during World War II and Subsequent Mortality

Laurent Lequien*

5th June 2007

Preliminary work, please do not quote

*INSEE - CREST
Email: laurent.lequien@insee.fr Telephone: + 33 1 41 17 77 33
The author would like to thank participants at CREST-LMI internal seminar, March 2007, and JMA, May 2007, for their helpful comments. All remaining errors are mine.

Abstract

We investigate the link between education and health. A strong correlation between those two components of human capital is now well established, and this result holds with various health measures. This study focuses on a potential causality running from education to mortality. Using a French longitudinal dataset, we compute a health measure at the same given age for everyone, namely being still alive at 70. We follow a two-stage estimation strategy, our instrumental variable exploits the fact that many French students were forced to leave school earlier than they wished during World War II while France was under foreign occupation. Moreover, this exogenous shock on education choices did not happen at the same time countrywide, and we use this temporal dimension as another source of identification.

The results suggest a positive and significant impact of education on survival at 70. While the magnitude of the effect is in line with the literature, returns to education are estimated more precisely than in previous studies.

Keywords: Return to education, health, mortality, instrumental variable

JEL Classification: I12, I20

Résumé

Nous étudions la nature du lien entre le niveau d'éducation d'un individu et son état de santé. Une forte corrélation positive a été constatée entre ces deux composantes du capital humain, et ce résultat reste valable avec différentes mesures de la santé. La présente étude tente de déterminer si cette corrélation empirique peut, au moins en partie, être expliquée par un mécanisme causal allant du niveau d'éducation vers la santé. En s'appuyant sur un panel provenant de source administrative française, l'Echantillon Démographique Permanent, nous abordons la santé sous l'angle de la mortalité, et construisons un index de survie à 70 ans. Une stratégie d'estimation en deux étapes est mise en oeuvre. Notre variable instrumentale exploite le fait que beaucoup d'étudiants français ont été obligés de quitter l'école plus tôt qu'ils ne l'auraient souhaité à cause de l'occupation durant la seconde guerre mondiale. De plus, ce choc exogène n'a pas eu lieu au même moment dans la France entière, et nous utilisons cette dimension temporelle comme autre source d'identification.

Les résultats suggèrent que le niveau d'éducation a un impact causal significatif et positif sur la survie à 70 ans. Nous estimons qu'une année d'études supplémentaire augmente de 2.3 points de pourcentage la probabilité d'être encore en vie à 70 ans. Tandis que l'ampleur de ce rendement de l'éducation est en accord avec la littérature existante, il est estimé avec une plus grande précision que dans des études similaires.

Mots clé : Rendement de l'éducation, santé, survie, variable instrumentale

Classification JEL : I12, I20

1 Introduction

Education and health are considered two of the main determinants of human capital, and for this reason they were used as explanatory variables in wage equations. Noticing a strong correlation between those two variables, Grossman (1972) was the first to build a theoretical framework modeling their interdependence. He postulated the existence of a health production function, education was one of its inputs and hence could affect one's health. As such, Grossman was describing a causal link running from education to health. But causality is not the only possible explanation to the empirical correlation between education and health. There could be a third factor influencing both education and health, and thus indirectly creating a link between them. For instance, a plausible candidate would be one's familial background: parents who stimulate their children to achieve successful studies may at the same time make them aware of the importance of behaving healthily. Time preferences may also be an unobserved characteristic simultaneously governing education and health choices, since investing in education and in health improves one's well-being mostly in the long term.

A growing literature focuses on determining whether any causal mechanism could explain - at least partly - the correlation between education and health. The most common strategy to deal with potential endogeneity of education is to use an instrumental variable. The idea behind IV is to isolate a specific event which induced a variation in educational level. As long as this event is not correlated with variables influencing both education and health, its sole possible way to affect health status is to trigger an indirect mechanism passing through education. Therefore this method allows to reveal causal paths running from education to health, when such paths do exist. Several types of instruments have been used in such a case. Lleras-Muney (2005), Glied and Lleras-Muney (2003), Spajosevic (2003), Oreopoulos (2003), and Arendt (2005) exploit changes in compulsory education laws as an exogenous source of variation in education. Adams (2002) uses quarter of birth, which is correlated to the number of years spent in school, and should not have any direct impact on health. Arkes (2003), Auld and Sidhu (2005) use local unemployment rate as an instrument. Eventually, a majority of studies show the existence of a causal impact of education on health, but no consensus has emerged yet[1].

We add to the existing literature on this topic by working on a French longitudinal dataset, which has not been done yet as far as we know. Moreover we propose a two-stage estimation using an innovative instrumental variable : France was invaded by the German army during World War II, and this led - among others consequences - to a decrease in school leaving age for those who left school during occupation. This exogenous drop in education allows us to identify a possible causality from education to health. The next paragraph describes the identification strategy and discuss the validity of our instrument, while section 3 presents the dataset as well as basic statistics. Section 4 shows our results, and the last section concludes.

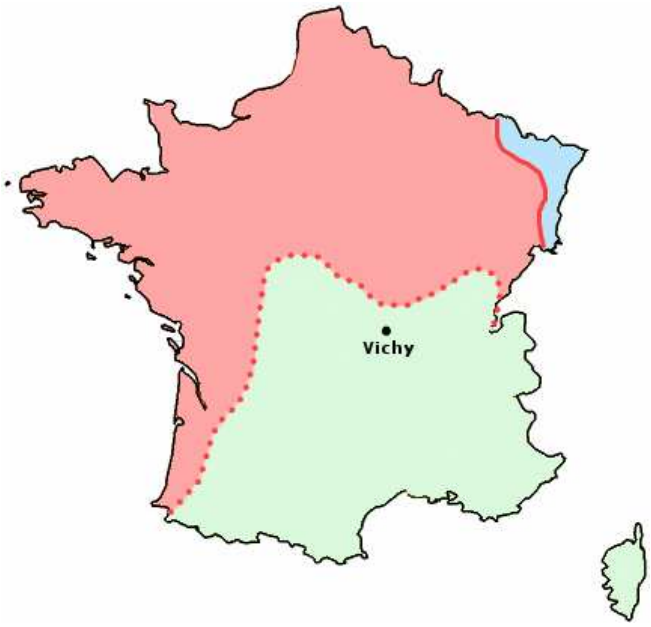
[1] See Grossman (2004) or Albouy and Lequien (2007) for a more comprehensive survey, as well as a discussion on the various measures of health used in the existing literature.

2 Historical background and identification strategy

France was invaded by the German army during World War II. This happened in two successive phases. From June 1940 to November 1942 roughly half of France was under the control of the French Government (the so-called free zone), whereas the north half was under the joint domination of the French government and the German army (the occupied zone). Then from November 1942 until the Liberation in 1944, both zones were under German control. So the occupied zone was run by Germans from 1940 to 1944, while the free zone was occupied only between 1943 and 1944. Figure 1 represents a map of France during World War II, with the demarcation line separating occupied zone from free zone.

Although daily life was tough for everyone during the war, living conditions were harder in places where the German army was in charge. One had to go through hardships like lack of food, fear of being arrested, and more generally all kinds of worries brought by a foreign military invasion. German occupation was likely to affect every aspect of French people lives. For example, finding food became a significant issue for many people who previously had enough money not to worry about food. There are several reasons to believe that occupation lowered education by disorganizing school environment: school may close, teacher's attendance may decline,

Figure 1: France during World War II



Note: The dotted red line is the demarcation line. It separated the occupied zone in north of France (in red) from the free zone (in green). The north-east part, in blue, was the territory annexed by Germany. Source: wikipedia.org, under the terms of the *GNU Free Documentation License*.

parents may flee the conflict and not necessarily send their children to school in their new town. Besides, German occupation certainly had an impact on time preferences as well, because immediate matters (such as managing to stay alive until the end of the day) took precedence more strongly than before over long term projects. Since relative importance of short-term and long-term investments was altered in decisions making processes, students in those days were likely to leave school earlier than what they would have chosen if there had been no German occupation.

We use occupation as an exogenous source of variation in school leaving age. Individuals affected by this instrument are those who gave up school during German occupation. It concerns either people living in occupied zone who left school between 1940 and 1944, or those living in free zone who dropped out between 1943 and 1944. As in any other study, an instrumental variable must satisfy two criteria to be valid. The first one is to be correlated with the endogenous variable, which here is the school leaving age. This is the case, as shown in the next section. The second condition is usually trickier, because its validity can only be theoretically assessed. It stipulates that the instrument must not be directly linked to the variable of interest, the health. In other words, the simple fact of leaving school under German occupation must have no other impact on subsequent health that the indirect one due to a lower education.

Let's focus on one example, in order to see what this exclusion condition implies. Consider two men born in occupied zone. The first one left school before the war in 1938, the second one stopped studying in 1941 and was potentially forced to drop out sooner than he wished because of the occupation. Unless one of them moved abroad or in free zone[2], they both lived in occupied zone during the war, and they both suffered from the same events which may have had some influence on their subsequent health (war, lack of food, stress, etc.). They differ in only two dimensions: one had already finished his schooling when the war broke out and the other not, one was on average older than the other during the war. Our instrument will be valid only if all possible events altering health - except occupation through its impact on education - have a constant effect over those two dimensions.

Few mechanisms could have affected the health of those two persons in a different way. The first to cross one's mind is the compulsory labor service (Service du Travail Obligatoire, STO). In 1942 Germany lacked labor supply and asked the French government to send 250 000 men to Germany. It was first set up on a voluntary basis, but failed to recruit the required number. Therefore it became mandatory in 1943, the enrollment mainly concerning men born between 1920 and 1922. As we ran our estimations on people born between 1923 and 1934, our sample does not contain cohorts which were the most affected by STO. However, it is still possible that some individuals among the oldest in our sample were indeed forced to go to Germany. These men most likely left school before the war (more than 80% of men

[2]This could be a real issue, since many people moved from occupied to free zone right after German settlement in occupied zone. Unfortunately, we can't take that into account, because we know where people were born, but not the actual place where they lived during the war. This exodus may bias downwards our estimation: those movers found milder living conditions in free zone, which might counterbalance some of the negative health effect due to an early drop out of school. Therefore it will strengthen our conclusions if we find a positive impact in spite of that bias.

born in 1923 had left school before 1939). Their health surely did not improve during their stay abroad, compared to those who stayed in France. It should then reduce survival differences -if they ever existed - otherwise attributed to occupation : those whose survival may lower because occupation forced them to drop out of school were too young to be sent to Germany, whereas some people among those who left school before the occupation era enrolled in STO and so have lower survival rates as well. Beside this direct effect of STO on health for those who worked in Germany, STO may also have indirectly lowered education during occupation, if fourteen-year-olds dropped out of school to replace in plants older workers sent to Germany. As such, STO could be described as one specific channel through which occupation led to lower school leaving age, and this channel would satisfy the exclusion condition (no other impact on health that via education for those who left school during the occupation). Since force workers sent to Germany were men in their vast majority, it is likely that they were replaced by other men at work; therefore STO would also account for a larger drop in education for men than for women during occupation. In fine any potential causal effect of education on survival might be biased downwards by STO, but STO will not invalidate the causal interpretation we suggest, should we find a positive and significant impact.

Another mechanism which could have altered health differently for our two representative persons is food rationing. Facing a severe lack of food supply during occupation, a system of ration cards was established in September 1940 in the occupied zone. Individual tickets entitled their owner to several kinds of food (meat, milk, bread, sugar, etc.). Quantities depended on age (child under 3, child between 3 and 12, adult, etc.), and type of work for adults (force worker or not, farmer, retiree). Age discontinuities created by this classification may have had an impact on health, because individuals in our sample did not have the same age during occupation and hence were not given the same amount of food. Nevertheless several reasons lead us to believe that this possible impact was all but minor. First of all, differences in amount of food between categories were not that important, at least they were negligible compared to the acute starvation the whole population was enduring. Moreover, family members certainly put together their individual rations to prepare domestic meals, which behavior smoothed differences between tickets. Lastly, food shortage became such a major issue from 1943 on that many shops just could not deliver enough food to meet ration cards demand. Black market became in those days a common way to find food, and that also contributed to the dilution of any potential age threshold effect. Finally, it seems that all individuals in our sample faced roughly the same conditions concerning food rationing during occupation. As long as the effect of starvation on health is homogeneous in the population[3], we feel confident that rationing had the same impact on health for all individuals in our sample.

Eventually we find no convincing explanation undermining the validity of our instrumental variable.

[3]One point is still pending though : it may be the case that rationing has more or less serious consequences on health depending on the age at which the individual suffered from it. We don't take into account this hypothesis in our model, because our sample contains individuals who were all teenagers during the war and thus were about the same age.

3 Data

We use information from the Echantillon Démographique Permanent (EDP thereafter), a longitudinal dataset containing administrative information on all French women and men born on one of the first four days of October. It compiles 1968, 1975, 1982, 1990 and 1999 census data with information from register of births, marriages and deaths until 2004. This database was created from the 1968 national census, which means that individuals born on one of those four days before 1968 had to be still alive in 1968 to appear in the EDP[4]. In particular, those who died during the war are not present in our sample. It is therefore possible that our sample is not a perfectly faithful picture of the French population living during World War II. The EDP contains approximately 5 000 persons per year of birth. For each one of them, censuses provide us with their school leaving age *sla*. We add it to the year of birth *year* to compute the year during which they ended their schooling $yearsch$. Place of birth tells us whether one lived in occupied ($zone = 1$) or free zone ($zone = 0$)[5].

The instrumental variable *occup* is a dummy equal to 1 if one left school during occupation. Hence it is set to 1 for those who lived in occupied zone if they left school between 1940 and 1944, to 0 otherwise. Similarly for inhabitants of free zone, *occup* equals 1 if they left school in 1943 or 1944.

The date of death is used to build a dummy indicating whether one is still alive at 70. According to existing literature, differences in mortality due to socio-economic status appear when people are in their sixties. Therefore 70 is a relevant age to show causal effects from education to health. Moreover, as an objective measure of one's health status, it does not suffer from bias inherent to all self-declared measures. Eventually the longitudinal structure of the EDP allows us to have a health measure at the same given age from everyone, and we can thus avoid the usual pitfall consisting in modeling the age dependence of health.

Our estimation sample is an extract from the EDP, from which we selected all individuals born between 1923 and 1934 (see descriptive statistics in Table 1). This sample contains 63 770 persons. Although populations of occupied and free zones are similar in many dimensions, one point worth noting is that inhabitants of occupied zone are twice as numerous as inhabitants of free zone.

Figure 2 pictures the mean school leaving age by generation. Separate curves are plotted for inhabitants of free and occupied zone. We notice the well-known positive trend over time, stating the rise in level of education among the population through generations[6]. There is a drop in education in occupied zone for the 1928 cohort,

[4] Everyone in our sample was born after 1923. In order to keep the selection bias constant throughout our sample, we excluded those who died before turning 45. Therefore, our health measure, being alive or not at 70, must be understood as a survival index *given that one is still alive at 45*. For the sake of simplicity, we will omit this precision in the remaining of the paper.

[5] A demarcation line separated occupied zone and free zone. Continental France counted 90 departments, and this line went through 13 of these departments. We excluded from our sample individuals born in one of these 13 departments, to avoid border effects, and also dropped those born overseas. So every individual in our sample was born in a department fully included either in occupied or free zone.

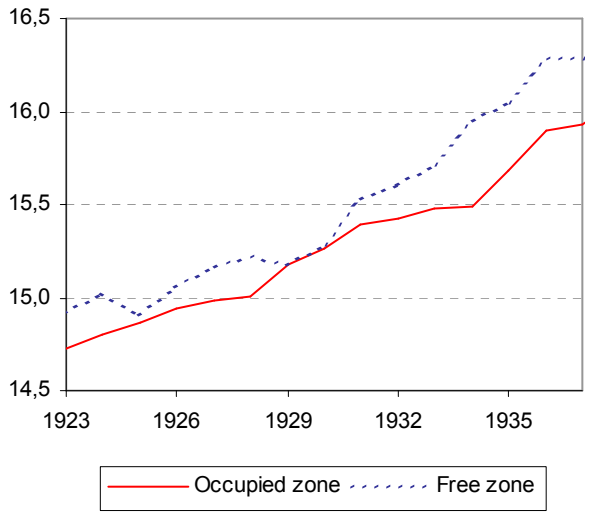
[6] Compulsory education laws were modified twice in France during the twentieth century. Educa-

Table 1: Descriptive statistics

	Total sample	Occupied zone	Free zone
Number of observations	63 770	45 562	18 208
Proportion of women	0.509	0.510	0.506
	(0.499)	(0.499)	(0.499)
Year of birth	1928.55	1928.54	1928.56
	(3.42)	(3.42)	(3.44)
School leaving age	15.18	15.13	15.30
	(2.74)	(2.71)	(2.82)
Survival at 70	0.79	0.77	0.82
	(0.41)	(0.42)	(0.38)

Notes: Standard errors in parenthesis.
Total sample consists of individuals who were born between 1923 and 1934.

Figure 2: School leaving age by birth cohort



Note: School leaving age as a function of the year one was born. Based on the 78 414 individuals who were born between 1923 and 1937, depending on whether they were born in occupied or free zone.

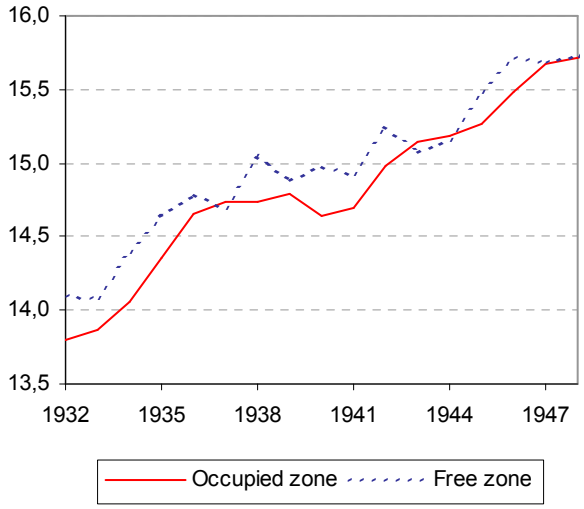
followed by a similar decrease in free zone one year later. Yet one must have in mind that the magnitude of these drops does not represent the actual effect of occupation on education, since individuals leaving school during occupation could belong to any generation from 1923 to 1934. Occupation had indeed an impact spread over several cohorts. These cohorts were unevenly affected, the intensity of the impact being related to the proportion of each generation still studying during occupation. Hence decreases in Figure 2 can only give a hint of the differential impact between cohorts. It is likely that the magnitude of the temporal trend pictured in Figure 2 was also affected by this composition effect.

An alternative way to present the effect of occupation on education is to show variations of school leaving age as a function of the year during which one ended its schooling. Such curves are drawn in Figure 3 for men and Figure 4 for women. Unlike year of birth, the year during one left school is the result of an individual choice. Therefore two "cohorts" in Figures 3 and 4 may not be comparable, as it was the case in Figure 2. The first reason is that they include individuals who faced different constraints regarding education choices in their last year of schooling. Moreover the number and level of schooling of persons likely to drop out one given year depends on the number and level of schooling of persons who left school the years before. Despite these possible biases, Figures 3 and 4 are still informative. Structural differences in education (prior to the war) exist between north and south of France: people attend school on average 2 months more in free zone than in occupied zone. Women's school leaving age rises at a steady pace, and evolutions are quite similar in occupied and free zone. There is nevertheless a slight decrease in education for inhabitant of occupied zone, corresponding to the years of occupation. The picture is more strongly marked for men. There is a sudden drop in men's education in 1940 and 1941 for those living in occupied zone, corresponding to the beginning of the occupation in that zone. At the same time, no noticeable change happens in free zone, which is not occupied yet. We find a inverted pattern in 1943 : education falls in free zone (though to a lesser extent) when it is invaded by the German army, while school leaving age resumes its rise in occupied zone. This last fact is hard to interpret : education should not have increased yet because occupied zone was still under German control in 1943. It certainly stems from the composition effect described above. The drop in the first year of occupation represents indeed the decrease in education due to occupation. But subsequent years mix occupation with consequences of the skimming that occured since the beginning of occupation. Another explanation is that perhaps daily pressure decreased there because the German army had to disperse its troops to cover a territory twice as large as before. Or it might have taken two years to the population to adapt its behavior to occupation constraints regarding education[7].

tion was mandatory until one turned 13, for children born before 1922. It was first extended in 1933 to 14, this reform being effective for children born after January 1^{st}, 1923. Then for those born from 1953 on, school has been mandatory until 16. So everyone in our sample grew up under the same legislative regime regarding education laws.

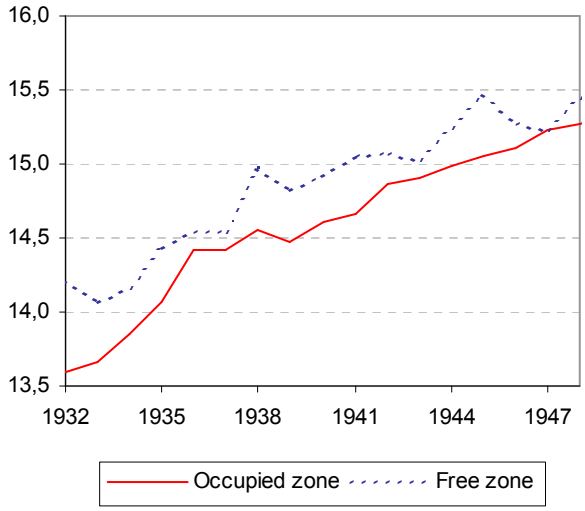
[7]See Sirinelli et al. (1995) for a detailed description of living conditions in France during the war. It seems that the population spirits were beginning to raise in 1942 in occupied zone. To illustrate this idea, the authors explain that the so-called *baby-boom* (high increase of the number of births in

Figure 3: Men school leaving age



Note: School leaving age as a function of the year one left school. Based on the 31 008 men who left school between 1932 and 1948, depending on whether they were born in occupied or free zone.

Figure 4: Women school leaving age

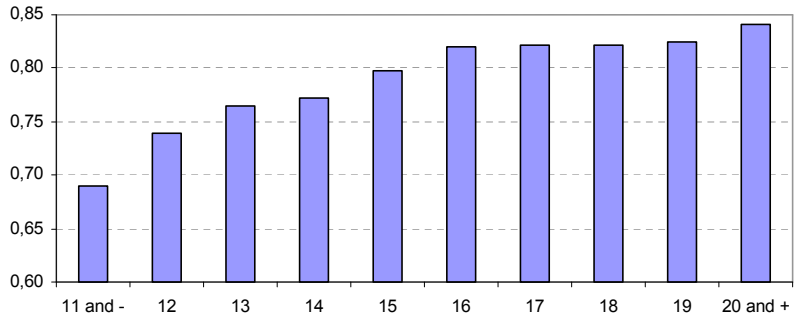


Note: School leaving age as a function of the year one left school. Based on the 32 306 women who left school between 1932 and 1948, depending on whether they were born in occupied or free zone.

From 1945 on, school leaving age rises in both zones, as it used to be the case before the war.

Empirical correlation between education and health is shown in Figure 5. A positive link is clearly visible : the longer one stays at school, the likelier one is to be still alive at 70. The next section tries to disentangle a possible causality from this correlation.

Figure 5: Link between school leaving age and probability of survival at 70



Note: Survival at 70 as a function of the school leaving age. Based on the 63 770 individuals born between 1923 and 1934.

4 Results

We aim at estimating the impact of education on health. If there were no unobserved heterogeneity altering both education choices and health, estimation of probit model (1) would give us an answer:

$$surv_i^* = a_1 + b_1\, sex_i + c_1\, sla_i + d_1\, zone_i + e_1\, zone_i \times sex_i + f_1\, year_i + u_{1i} \qquad (1)$$

Health is measured by $surv_i$, a dummy variable equal to 1 if individual i is still alive at 70. Its corresponding latent variable is $surv_i^*$. First explanatory variables are gender ($sex_i = 1$ if i is a man, 0 otherwise) and school leaving age sla_i. Zone of birth $zone_i$ equals 1 (resp. 0) if i was born in occupied (resp. free) zone. It should capture structural differences in health existing before the war between occupied and free zones, as well as effects of occupation on health common to inhabitants of a given zone. We allow the impact of zone on health to be different for men and women through $zone_i \times sex_i$. We use the year of birth $year_i$ to take into account gradual

the aftermath of the war) began as early as 1942 and not after the end of the war.

improvements in health over time (due to medicine progress, better living conditions, etc.). u_{0i} sums up all other (un)observed variables.

If education choices were not endogenous, this error term would not be correlated with explanatory variables, and a probit estimation would be convergent. Results of such an estimation are presented in column 1 of Table 2. As expected, there is a huge difference in survival between men and women: proportion of men still alive at 70 is 0.16 points lower than for women. Survival is also higher in free zone, either for men and for women. Many reasons can account for this. Among obvious explanations like way of life, economic or weather conditions, lies the fact that inhabitants of free zone suffered from occupation for only two years, whereas it lasted four years in occupied zone. As a result, it is possible that occupation affected people's health with a different intensity in free and occupied zones[8]. One extra year of education is associated with an increase by 1.2 percentage point of the probability of being still alive at 70.

The first stage of our instrumental variable strategy consists in introducing an exogenous source of variation among variables explaining education:

$$\begin{aligned} sla_i = a_2 + b_2\, sex_i + c_2\, zone_i + d_2\, occup_i + e_2\, year_i \\ + f_2\, occup_i \times sex_i + g_2\, zone_i \times sex_i + u_{2i} \end{aligned} \tag{2}$$

Level of education sla_i depends on gender, the zone one was born, and whether one dropped out of school during occupation. Interactions between those last two variables and gender are allowed. A temporal trend is also present. As can be seen in column 2 of Table 2, men stay longer than women at school, both in free zone (0.24 year) and in occupied zone (0.55 year). The effect of zone on education depends on gender: women drop out sooner in occupied zone than in free zone (-0.15 year), but it goes the other way for men (0.16 year)[9]. It appears that our instrument (leaving school under occupation) has a large and significant impact on education, its amplitude is twice more important for men than for women (-1.11 year vs. -0.54 year). It is worth noticing that the instrument has a huge impact on education, compared to other instruments in existing literature. One possible explanation is that occupation affected education choices of every individual studying during the

[8] Keep in mind that everyone in our sample suffered from the war, even those who left school before the war began. It is an important difference between education and health : occupation altered education of a small part of our sample (people who dropped out during occupation), while it had an impact on everyone's health. So health differences due different lengths of occupation should appear even for those who left school before or after the war.

[9] This is not what one would have expected, since Figure 3 seems to indicate that men have a higher level of education in free zone than in occupied zone. We ran the first stage estimation without the instrumental variables *occup* and $occup \times sex$, and it appeared that both men and women studied longer in free zone. Though significant, the difference in education for men between occupied zone and free zone is small (-0.06 year, versus -0.26 year for women). Part of this zone effect accounts for occupation and the fact that occupation lasted longer (and thus affected more people) in occupied zone than in free zone, since *occup* and $occup \times sex$ are not present in the regression. Adding *occup* and $occup \times sex$ as explanatory variables should logically make those coefficients increase. This is what happened, to the point that the zone effect became positive for men.

Table 2: Probit and IV Probit estimations

	Dependant variable				
	Survival at 70	School leaving age	Survival at 70	School leaving age	Survival at 70
	Total sample	Total sample	Total sample	Men only	Men only
	(1)	(2)	(3)	(4)	(5)
Intercept		14.16***		14.55***	
		(0.066)		(0.098)	
sla	0.012***				
	(0.00062)				
$\widehat{sla}$			0.023***		0.029***
			(0.0039)		(0.0051)
occup		-0.54***		-1.13***	
		(0.033)		(0.038)	
$occup \times sex$		-0.57***			
		(0.048)			
sex	-0.16***	0.24***	-0.16***		
	(0.0063)	(0.041)	(0.0063)		
zone	-0.032***	-0.15***	-0.030***	0.16***	-0.069***
	(0.0054)	(0.034)	(0.0055)	(0.038)	(0.0056)
$zone \times sex$	-0.025***	0.31***	-0.026***		
	(0.0075)	(0.048)	(0.0076)		
year	0.0014***	0.062***	-0.00066	0.054***	-0.0011
	(0.00047)	(0.0032)	(0.00056)	(0.0050)	(0.00086)
N	63 770	63 770	63 770	31 309	31 309

Notes: Estimations are run on individuals who were born between 1923 and 1934. We subtracted 1900 from *year* in order to have a better estimation of the intercept in the first stage. Column (1) is a probit regression, columns (2) and (3) are the two-stage estimation on the whole sample, columns (4) and (5) are the two-stage estimation run on the subsample composed of men only. Coefficients reported in columns (1), (3) and (5) are mean marginal effects. Standard errors in parenthesis. *, **, and *** mean that coefficients are significant at respectively 10%, 5% and 1% confidence levels.

occupation. Therefore the whole distribution of school leaving ages was potentially affected by occupation, whereas other instruments used in similar studies have an impact on a very limited part of the population (e.g. those who wish to leave school at an age between old and new mandatory schooling limits in case of a change in compulsory education laws).

We use the predicted school leaving age $\widehat{sla_i}$ from (2) into the second step of our estimation:

$$\begin{cases} surv_i^* = a_3 + b_3\, sex_i + c_3\, \widehat{sla_i} + d_3\, zone_i \\ \qquad\qquad + e_3\, zone_i \times sex_i + f_3\, year_i + u_{3i} \\ surv_i = \mathbb{1}(surv_i^* \geq 0) \end{cases} \tag{3}$$

Results are shown in column 3 of Table 2. Education has a positive and significant causal impact on survival: one additional year of education increases the probability of being alive at 70 by 2.3 percentage points. This coefficient is highly significant.

As impact of occupation on education choices is mostly focused on men, it seems legitimate to run the previous estimation on the subsample containing only men. First stage equation (2) is hence transformed into:

$$sla_i = a_4 + b_4\, zone_i + c_4\, occup_i + d_4\, year_i + u_{4i} \tag{4}$$

while equation (3) becomes:

$$\begin{cases} surv_i^* = a_5 + b_5\, \widehat{sla_i} + c_5\, zone_i + d_5\, year_i + u_{5i} \\ surv_i = \mathbb{1}(surv_i^* \geq 0) \end{cases} \tag{5}$$

Results are similar to those obtained with the whole sample (columns 4 and 5 of Table 2). Leaving school during occupation leads to a drop by 1.13 year. The causal effect of one year of education on survival at 70 is an increase by 2.9 percentage points. This is a sizeable impact: as men's survival is around 70%, it represents a 10% decrease in the probability of dying before one turns 70.

We tested the sensitivity of our results regarding the sample selection process (keeping individuals born between 1923 and 1934), and ran the two-stage estimation with various ranges of years of birth. Neither the magnitude nor the precision of the causal impact seem to depend on the widening of this interval.

One can note that the IV coefficient is larger than the one obtained by OLS, and that those two estimates are statistically different (a Wald test of exogeneity gives $\chi^2(1) = 6.77$, $pvalue < 0.01$). This confirms that education is indeed an endogenous variable in the health equation. On the other hand, one would have expected the IV coefficient to be smaller than the OLS one. A possible explanation is that education is measured with error, which biases downwards the OLS estimate. This bias disappears in IV estimations if the instrument is independant of the measurement error[10].

[10] See Grossman (2004) or Card (2001). See also Belzil and Hansen (2005) who provide an explanation using a dynamic framework.

5 Conclusion

We find a positive and significant impact of education on health. For those affected by our instrument, the probability of being alive at 70 increases by 2.3 percentage points with each extra year of schooling. The magnitude of the impact is in line with what is found in other studies relying on the same measure of health (Lleras-Muney (2005) and Cipollone et al. (2006)). Moreover, the precision of our estimations is excellent, even when using an instrumental variable. This allows us to reject the null hypothesis of exogeneity of education in the survival equation.

Although we do have a significant causal effect, we do not know precisely how a higher level of education leads to a better health. However, our identification strategy does allow us to provide a piece of answer to a possible heterogeneity of returns to education among individuals: are returns to education the same if one is forced to stay at school (e.g. via a raising of school leaving age) and if one willingly goes on studying? In other words, should we expect to find the same impact with our instrument that in studies where children have to attend school even if they don't want to? It may well be the case that (both monetary and non monetary) returns to education depend on students' motivation in classrooms, and that those interested in studying benefit more from an extra year of education. If heterogeneity among students' motivation indeed matters, we should find a larger causal impact than in other studies because our instrument prevented motivated students from going to school whereas compulsory education laws have an impact on children who don't want to study any more. Angrist and Imbens (1994) point out the importance of knowing what part of the population is affected by the instrument, since the causal effect (which they call *Local Average Treatment Effect*) is estimated solely on that part of the population. Comparing Lleras-Muney's (2005) results with ours suggests that differences in motivation may not be a source of heterogeneity in return to education on health.

Moreover, it would be interesting to know whether the magnitude of return to education depends on school leaving age. This assumption of constant return is almost always assumed in the literature. Unlike compulsory education laws targeted towards students of a specific age, our instrument affected the whole distribution of school leaving ages and is as such comparable to instruments like local unemployment rate (Arkes (2003), Auld and Sidhu (2005)) or quarter of birth (Adams (2002)). Hence a possible avenue for further research could be to use the widespread impact of our instrument to test the validity of this hypothesis.

References

Adams, S. J. (2002), 'Educational attainment and health: evidence from a sample of older adults', *Education Economics* **10**(1), 97–109.

Albouy, V. and Lequien, L. (2007), 'Les rendements non monétaires de l'éducation: le cas de la santé', *INSEE-D3E Working Paper* **G2007/02**.

Angrist, J. D. and Imbens, G. W. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–475.

Arendt, J. N. (2005), 'Education effects on health: A panel data analysis using school reform for identification', *Economics of Education Review* **24**(2), 149–160.

Arkes, J. (2003), 'Does schooling improve health?', *RAND Working Paper* .

Auld, M. C. and Sidhu, N. (2005), 'Schooling, cognitive ability, and health', *Health Economics* **14**, 1019–1034.

Belzil, C. and Hansen, J. (2005), 'A structural analysis of the correlated random coefficient wage regression model with an application to the ols/iv puzzle', *IZA Working Paper* .

Card, D. (2001), 'Estimating the return to schooling: progress on some persistent econometric problems', *Econometrica* **69**(5), 1127–1160.

Cipollone, P., Radicchia, D. and Rosolia, A. (2006), 'The effect of education on youth mortality', *Banca d'Italia mimeo* .

Glied, S. and Lleras-Muney, A. (2003), 'Health inequality, education and medical innovation', *NBER Working Paper* **9738**.

Grossman, M. (1972), 'On the concept of health capital and the demand for health', *The Journal of Political Economy* **80**(2), 223–255.

Grossman, M. (2004), 'The demand for health, 30 years later: a very personal retrospective and prospective reflection', *Journal of Health Economics* **23**, 629–636.

Lleras-Muney, A. (2005), 'The relationship between education and adult mortality in the united states', *Review of Economic Studies* **72**, 189–221.

Oreopoulos, P. (2003), 'Do dropouts drop out too soon? international evidence from changes in school-leaving laws', *NBER Working Paper* **10155**.

Sirinelli, J. F., Vandenbussche, R. and Vavasseur-Desperriers, J. (1995), 'La france de 1914 à nos jours', *Presses Universitaires de France* .

Spajosevic, J. (2003), 'Effect of education on adult health in sweden: results from a natural experiment', *Ph.D Dissertation, City University of New York Graduate Center* .